

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΠΟΛΥΔΙΑΣΤΑΤΑ ΔΙΑΓΡΑΜΜΑΤΑ
ΕΛΕΓΧΟΥ ΔΙΕΡΓΑΣΙΩΝ
ΚΑΙ ΕΦΑΡΜΟΓΕΣ**

Ελισάβετ Μ. Σοφικίτου

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβλήθηκε στο
Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης
του Πανεπιστημίου Πειραιώς

Πειραιάς
Δεκέμβριος, 2017

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΠΟΛΥΔΙΑΣΤΑΤΑ ΔΙΑΓΡΑΜΜΑΤΑ
ΕΛΕΓΧΟΥ ΔΙΕΡΓΑΣΙΩΝ
ΚΑΙ ΕΦΑΡΜΟΓΕΣ**

Ελισάβετ Μ. Σοφικίτου

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβλήθηκε στο
Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης
του Πανεπιστημίου Πειραιώς

Πειραιάς
Δεκέμβριος, 2017

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**MULTIVARIATE PROCESS
CONTROL CHARTS
AND APPLICATIONS**

Elisavet M. Sofikitou

PhD Thesis

Submitted to

Department of Statistics and Insurance Science of
the University of Piraeus

Piraeus, Greece

December, 2017

*Στους γονείς μου
Μιχάλη, Κατερίνα
και στην αδελφή μου
Μαρία*

Ευχαριστίες

Θερμές και ειλικρινείς ευχαριστίες θα ήθελα να εκφράσω στον επιβλέποντα καθηγητή μου, κο Μάρκο Β. Κούτρα, καθώς η συμβολή του κατά την εκπόνηση της παρούσας διατριβής υπήρξε πολύτιμη, σε όλα τα στάδια: επιλογή θέματος, συλλογή και επεξεργασία στοιχείων - δεδομένων, μελέτη του αντικειμένου, εμβάθυνση στο γνωστικό πεδίο της ερευνητικής περιοχής, εξαγωγή και παρουσίαση συμπερασμάτων, συγγραφή και υποστήριξη της διατριβής.

Σε επιστημονικό επίπεδο ο καθηγητής, κος Κούτρας, με τις πολύπλευρες γνώσεις του και την επάρκειά του σε όλα τα επιστημονικά πεδία που άπτονται του θέματος της διατριβής μου, μου μετέδωσε τη γνώση, αλλά και την αγάπη για έρευνα και μελέτη. Μου δίδαξε καινούργιους επιστημονικούς δρόμους και άνοιξε τους ορίζοντές μου, εμφυσώντας τη μαθηματική και ερευνητική του σκέψη. Χάρη στην καθοριστική του βοήθεια προσδιορίστηκαν με σαφήνεια όχι μόνο οι ερευνητικοί στόχοι της παρούσας διατριβής, αλλά και η μεθοδολογία στα επιμέρους στάδια εκπόνησής της.

Επιπλέον, ο κος Κούτρας με στήριξε και σε ανθρώπινο επίπεδο, σε όλη την πορεία της διατριβής, κατευνάζοντας αρκετές φορές το άγχος και τις ανησυχίες μου, δίνοντας απαντήσεις σε όλα τα ερευνητικά ερωτήματα που με ταλάνισαν, επιδεικνύοντας, πάντα, αστείρευτη υπομονή κατά τη διάρκεια της προσπάθειάς μου. Χωρίς την εμπιστοσύνη και την αμέριστη συμπαράστασή του δε θα είχε καταστεί δυνατή η ολοκλήρωση της παρούσας διδακτορικής διατριβής. Για όλους τους παραπάνω λόγους ευχαριστώ πρωτίστως και ιδιαιτέρως τον καθηγητή μου.

Ευχαριστώ, επίσης, τα δύο άλλα μέλη της τριμελούς επιτροπής μου, τον Αναπληρωτή Καθηγητή κο Δ. Αντζουλάκο και τον Επίκουρο Καθηγητή κο Π. Μαραβελάκη, για την ενθάρρυνσή τους καθ' όλη τη διάρκεια της εκπόνησης της διατριβής, αλλά και για τον πολύτιμο χρόνο που αφιέρωσαν για τη μελέτη αυτής, καθώς και τα υπόλοιπα μέλη της επταμελούς επιτροπής για τις πολύτιμες παρατηρήσεις τους. Θα ήταν παράλειψη να μην ευχαριστήσω τους εξαιρετους καθηγητές που είχα στη διάρκεια των προπτυχιακών και μεταπτυχιακών μου σπουδών τόσο στο Πανεπιστήμιο Πατρών, όσο και στο Πανεπιστήμιο Πειραιώς, καθώς η παρούσα εργασία αποτέλεσε το επιστέγασμα όλων των γνώσεων που μου προσέφεραν.

Οφείλω, ακόμη, να ευχαριστήσω το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

του Πανεπιστημίου Πειραιά, το οποίο χρηματοδότησε τη συμμετοχή μου σε συνέδρια και ενίσχυσε οικονομικά τη δημοσίευση σε διεθνή περιοδικά σχετιζόμενων με τη διατριβή εργασιών στα πλαίσια του προγράμματος «Δράσεις Έρευνας και Προβολής του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης», καθώς και το διοικητικό προσωπικό, η βοήθεια του οποίου ήταν πολύτιμη στην επίλυση πολλών διαδικαστικών θεμάτων. Ολοκληρώνοντας, αισθάνομαι την ανάγκη να ευχαριστώ θερμά τους γονείς μου και την αδελφή μου, η συμπαράσταση και ενθάρρυνση των οποίων υπήρξαν πολύτιμες για μένα σε όλη τη διάρκεια των σπουδών μου (προπτυχιακών και μεταπτυχιακών).

Περίληψη

Η παρούσα διατριβή πραγματεύεται θέματα που σχετίζονται με τη Θεωρία των Διατεταγμένων Παρατηρήσεων και του Στατιστικού Ελέγχου Ποιότητας. Αρχικά, γίνεται αναφορά σε βασικές έννοιες που συνδέονται με τη Θεωρία των Διατεταγμένων Παρατηρήσεων και των Συμμεταβλητών τους (*Theory of Order Statistics and Concomitants*), καθώς και με τη Θεωρία των Συνδέσμων (*Theory of Copulas*). Στη συνέχεια, παρουσιάζονται νέες κατανομές δισδιάστατων διατεταγμένων παρατηρήσεων και απαριθμητριών αυτών, οι οποίες βρίσκουν πρακτική εφαρμογή στον έλεγχο ποιότητας.

Για την ακρίβεια, τα αποτελέσματα που εξάγονται αξιοποιούνται στην κατασκευή τεσσάρων νέων δισδιάστατων ημιπαραμετρικών διαγραμμάτων ελέγχου και εφαρμόζονται με σκοπό τον ακριβή προσδιορισμό εκφράσεων για τα χαρακτηριστικά αυτών, όπως για παράδειγμα η χαρακτηριστική συνάρτηση λειτουργίας και το ποσοστό (λανθασμένου) συναγερού. Αξίζει να σημειωθεί ότι δυο από τα προτεινόμενα διαγράμματα είναι αποτελεσματικά στην ανίχνευση μετατοπίσεων της μέσης τιμής της υποκείμενης διεργασίας, ενώ τα άλλα δύο είναι ευαίσθητα στον εντοπισμό πιθανών μετατοπίσεων τόσο της μέσης τιμής, όσο και της μεταβλητότητας των υπό μελέτη χαρακτηριστικών.

Ένα από τα διαγράμματα που προτείνονται για την παρακολούθηση της μέσης τιμής επηρεάζεται από τη δομή εξάρτησης των ποιοτικών χαρακτηριστικών, η οποία ουσιαστικά περιγράφεται από το σύνδεσμο που αντιστοιχεί στην από κοινού δισδιάστατη κατανομή της διεργασίας, αλλά όχι από τις περιθώριες (μονοδιάστατες) κατανομές. Αντίθετα, στα υπόλοιπα διαγράμματα, οι τιμές του ποσοστού λανθασμένου συναγερού και του εντός ελέγχου μέσου μήκους ροής παραμένουν σχεδόν αμετάβλητες όταν γίνεται χρήση διαφορετικών συνδέσμων. Συνεπώς, το βασικό τους πλεονέκτημα έγκειται στο ότι μπορούν να αξιοποιηθούν ως πλήρως μη παραμετρικά διαγράμματα ελέγχου.

Τέλος, ιδιαίτερη έμφαση θα πρέπει να δοθεί στο γεγονός ότι η λογική κατασκευής όλων των προαναφερθέντων δισδιάστατων διαγραμμάτων ελέγχου είναι αρκετά απλή. Ως αποτέλεσμα, τα νέα διαγράμματα ελέγχου μπορούν εύκολα να γενικευθούν σε μεγαλύτερες διαστάσεις, ώστε να είναι εφικτή η παρακολούθηση περισσότερων των δύο ποιοτικών χαρακτηριστικών, χρησιμοποιώντας αναπόφευκτα πιο πολύπλοκες εκφράσεις για τον προσδιορισμό των χαρακτηριστικών τους.

Abstract

The present PhD dissertation deals with topics related to the Theory of Order Statistics and Statistical Quality Control. Some important notions on basic concepts concerning the Theory of Order Statistics and their Concomitants, as well as the Theory of Copulas are first presented. Subsequently, new distributions for bivariate order statistics and enumerating variables for them, which have a practical application in quality control, are introduced.

The new results are used to construct four new bivariate semi-parametric control charts and are applied to establish expressions for their characteristics, such as the operating characteristic function and the (false) alarm rate. It is worth noting that two of the proposed schemes are capable of detecting mean shifts of the underlying process, while the rest are effective in tracking possible deviations from the null values of both the mean and the variability of the features under study.

One of the two charts proposed for the monitoring of process mean, is affected by the dependence structure of the quality characteristics, as reflected on the copula associated with the bivariate distribution that describes the monitored characteristics; it is not influenced by the (univariate) marginals. On the contrary, for the rest three charts the values of the false alarm rate and the in-control average run length do not change crucially when different copulas are used. Consequently, their key advantage is the fact that they can be exploited as fully non parametric schemes.

Finally, it should be stressed that the construction of the above bivariate control charts is quite simple. As a result, the new schemes can be easily extended to higher dimensions, so that the monitoring of more than two quality characteristics is maintained, at the cost of more complicated expressions for the determination of their characteristics.

Περιεχόμενα

| | |
|---|-------------|
| Κατάλογος Σχημάτων | xi |
| Κατάλογος Πινάκων | xiii |
| Κατάλογος Συντομογραφιών | xvii |
| 1 Εισαγωγή στο Στατιστικό Έλεγχο Ποιότητας | 1 |
| 1.1 Ιστορική Αναδρομή | 1 |
| 1.2 Περιγραφή Τυπικού Διαγράμματος Ελέγχου | 4 |
| 1.3 Εισαγωγικές Έννοιες Ελέγχου Ποιότητας | 6 |
| 1.4 Είδη Διαγραμμάτων Ελέγχου | 11 |
| 1.5 Ανασκόπηση Βιβλιογραφίας | 13 |
| 1.5.1 Διαγράμματα ελέγχου για την παρακολούθηση μέσης τιμής ή διασποράς | 15 |
| 1.5.2 Διαγράμματα ελέγχου για την ταυτόχρονη παρακολούθηση μέσης τιμής και διασποράς | 16 |
| 1.6 Σκοπός και Δομή Διδακτορικής Διατριβής | 19 |
| 2 Θεωρία Διατεταγμένων Παρατηρήσεων, Συμμεταβλητών και Συνδέσμων | 23 |
| 2.1 Εισαγωγή | 23 |
| 2.2 Διατεταγμένα δείγματα και εφαρμογές | 23 |
| 2.3 Μονοδιάστατες κατανομές διατεταγμένων παρατηρήσεων | 30 |
| 2.4 Συμμεταβλητές Διατεταγμένων Παρατηρήσεων | 45 |
| 2.5 Σύνδεσμοι | 49 |
| 3 Νέες Δισδιάστατες Κατανομές Διατεταγμένων Παρατηρήσεων | 57 |
| 3.1 Εισαγωγή | 57 |
| 3.2 Δισδιάστατες από κοινού κατανομές διατεταγμένων παρατηρήσεων | 58 |
| 3.3 Δισδιάστατες από κοινού κατανομές πλήθους διατεταγμένων παρατηρήσεων | 72 |

| | |
|---|------------|
| 4 Δισδιάστατο Διάγραμμα Ελέγχου με Χρήση Διατεταγμένων Παρατηρήσεων | 77 |
| 4.1 Εισαγωγή | 77 |
| 4.2 Περιγραφή ενός νέου διαγράμματος ελέγχου | 81 |
| 4.3 Χαρακτηριστικά διαγράμματος ελέγχου | 85 |
| 4.3.1 Δίπλευρα διαγράμματα ελέγχου | 85 |
| 4.3.2 Μονόπλευρα διαγράμματα ελέγχου | 90 |
| 4.4 Απόδοση διαγράμματος για συγκεκριμένες οικογένειες συνδέσμων | 91 |
| 4.5 Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου | 95 |
| 4.6 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα | 101 |
| 4.7 Εφαρμογή σε πραγματικά δεδομένα | 107 |
| 5 Δισδιάστατο Διάγραμμα Ελέγχου με Χρήση Διατεταγμένων Παρατηρήσεων και Συμμεταβλητών | 111 |
| 5.1 Εισαγωγή | 111 |
| 5.2 Περιγραφή ενός νέου διαγράμματος ελέγχου | 112 |
| 5.3 Χαρακτηριστικά διαγράμματος ελέγχου | 115 |
| 5.3.1 Δίπλευρα διαγράμματα ελέγχου | 117 |
| 5.3.2 Μονόπλευρα διαγράμματα ελέγχου | 121 |
| 5.4 Απόδοση διαγράμματος για συγκεκριμένους συνδέσμους | 122 |
| 5.5 Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου | 125 |
| 5.6 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα | 128 |
| 5.7 Εφαρμογή σε πραγματικά δεδομένα | 133 |
| 6 Δισδιάστατο Διάγραμμα Ελέγχου για τη Θέση και τη Διασπορά με Χρήση Διατεταγμένων Παρατηρήσεων | 139 |
| 6.1 Εισαγωγή | 139 |
| 6.2 Περιγραφή ενός νέου διαγράμματος ελέγχου | 144 |
| 6.3 Χαρακτηριστικά διαγράμματος ελέγχου | 146 |
| 6.4 Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου | 150 |
| 6.5 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα | 158 |
| 6.6 Εφαρμογή σε πραγματικά δεδομένα | 164 |
| 7 Δισδιάστατο Διάγραμμα Ελέγχου για τη Θέση και τη Διασπορά με Χρήση Απαριθμητριών Στατιστικών Συναρτήσεων | 167 |
| 7.1 Εισαγωγή | 167 |

| | |
|---|------------|
| 7.2 Περιγραφή ενός νέου διαγράμματος ελέγχου | 168 |
| 7.3 Χαρακτηριστικά διαγράμματος ελέγχου | 170 |
| 7.4 Μελέτη Απόδοσης και σχεδιασμός διαγράμματος ελέγχου | 173 |
| 7.5 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα | 182 |
| 7.6 Εφαρμογή σε πραγματικά δεδομένα | 187 |
| 8 Σύνοψη | 189 |
| 8.1 Συμπεράσματα | 189 |
| 8.2 Προτάσεις για Μελλοντική Έρευνα | 192 |
| Ελληνική Βιβλιογραφία | 195 |
| Ξένη Βιβλιογραφία | 217 |

Κατάλογος Σχημάτων

| | | |
|-----|---|-----|
| 2.1 | Γραφική απεικόνιση του ενδεχομένου $[X_{r:n} \leq x]$ | 31 |
| 2.2 | Γραφική απεικόνιση του ενδεχομένου $[X_{r_1:n} \leq x_1, X_{r_2:n} \leq x_2]$ | 34 |
| 3.1 | Γραφική απεικόνιση του ενδεχομένου $[U_{r:n} \leq u, V_{s:n} \leq v]$ | 59 |
| 3.2 | Γραφική απεικόνιση του ενδεχομένου $[U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2]$ | 64 |
| 3.3 | Γραφική απεικόνιση του ενδεχομένου $[U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1]$ | 67 |
| 3.4 | Γραφική απεικόνιση του ενδεχομένου $[U_{r_1:n} \leq u_1; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2]$ | 70 |
| 3.5 | Γραφική απεικόνιση πιθανότητας $Q_{O2N2}^{(D)} = Q_{O2N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y)$ | 75 |
| 4.1 | Εφαρμογή διαγράμματος $O2$ στα δεδομένα λευκού οίνου της ποικιλίας <i>Vinho Verde</i> με μεταβλητές τα Θεϊκά Άλατα και την Αλκοόλη (Γράφημα 3D) | 84 |
| 4.2 | Πιθανότητες Συναγερμού $AR_{O2}^{(2)}$ για τους συνδέσμους FGM & AMH και δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) | 96 |
| 4.3 | ARL_{out} των διαγραμμάτων $SN^2, SR^2, \chi^2, O2$ στην περίπτωση ίσων μετατοπίσεων της μέσης τιμής | 103 |
| 4.4 | ARL_{out} των διαγραμμάτων $SN^2, SR^2, \chi^2, O2$ στην περίπτωση άνισων μετατοπίσεων της μέσης τιμής | 104 |
| 4.5 | Εφαρμογή του διαγράμματος $O2$ στο σύνολο δεδομένων λευκού οίνου της ποικιλίας <i>Vinho Verde</i> με μεταβλητές τα Θεϊκά Άλατα και την Αλκοόλη (Γράφημα 2D με το πρώτο δείγμα ελέγχου ποιότητας $LV6$) | 110 |
| 4.6 | Εφαρμογή του διαγράμματος $O2$ στο σύνολο δεδομένων λευκού οίνου της ποικιλίας <i>Vinho Verde</i> με μεταβλητές τα Θεϊκά Άλατα και την Αλκοόλη (Γράφημα 2D με το δεύτερο δείγμα ελέγχου ποιότητας $LV5$) | 110 |
| 5.1 | Εφαρμογή διαγράμματος OC στα δεδομένα λευκού οίνου της ποικιλίας <i>Vinho Verde</i> με μεταβλητές το Ελεύθερο και Ολικό Διοξείδιο του Θείου (Γράφημα 3D) | 114 |
| 5.2 | Συντελεστές Συσχέτισης τ του <i>Kendall</i> και ρ_S του <i>Spearman</i> για τρεις συνδέσμους (GUH, CLP, FRA) | 127 |

| | | |
|-----|---|-----|
| 5.3 | Κοινό γράφημα σ.π.π. της $\Gamma(\alpha_1, \beta_1)$ και ιστογράμματος σχετικών συχνοτήτων για το Ελεύθερο Διοξείδιο του Θείου | 134 |
| 5.4 | Κοινό γράφημα σ.π.π. της $\Gamma(\alpha_2, \beta_2)$ και ιστογράμματος σχετικών συχνοτήτων για το Ολικό Διοξείδιο του Θείου | 134 |
| 5.5 | Διαγράμματα διασποράς των αρχικών και των μετασχηματισμένων τιμών των χαρακτηριστικών του Ελεύθερου και Ολικού Διοξειδίου του Θείου | 135 |
| 5.6 | Εφαρμογή του διαγράμματος <i>OC</i> στο σύνολο δεδομένων λευκού οίνου της ποικιλίας <i>Vinho Verde</i> με μεταβλητές το Ελεύθερο και το Ολικό Διοξείδιο του Θείου (Γράφημα 2D με το πρώτο δείγμα ελέγχου ποιότητας <i>LV6</i>) | 137 |
| 5.7 | Εφαρμογή του διαγράμματος <i>OC</i> στο σύνολο δεδομένων λευκού οίνου της ποικιλίας <i>Vinho Verde</i> με μεταβλητές το Ελεύθερο και το Ολικό Διοξείδιο του Θείου (Γράφημα 2D με το δεύτερο δείγμα ελέγχου ποιότητας <i>LV5</i>) | 137 |
| 6.1 | Εφαρμογή του διαγράμματος <i>O4</i> στο σύνολο δεδομένων WDBC | 165 |
| 7.1 | Εφαρμογή του διαγράμματος <i>O2N2</i> στο σύνολο δεδομένων WDBC | 188 |

Κατάλογος Πινάκων

| | | |
|-----|---|-----|
| 1.1 | Σφάλμα Τύπου I και Σφάλμα Τύπου II | 8 |
| 2.1 | Σχήμα δειγματοληψίας μιας διατεταγμένης παρατήρησης | 31 |
| 2.2 | Σχήμα δειγματοληψίας δύο διατεταγμένων παρατηρήσεων του ίδιου δείγματος | 34 |
| 2.3 | Περιθώριες και από κοινού αθροιστικές συναρτήσεις κατανομής διατεταγμένων παρατηρήσεων | 41 |
| 2.4 | Περιθώριες και από κοινού συναρτήσεις πυκνοτήτων διατεταγμένων παρατηρήσεων | 42 |
| 2.5 | Δισδιάστατοι Σύνδεσμοι | 54 |
| 3.1 | Σχήμα Δειγματοληψίας – (2×2) : Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i, j \in \{1, 2\}$ | 59 |
| 3.2 | Σχήμα Δειγματοληψίας – (3×3) : Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i, j \in \{1, 2, 3\}$ | 64 |
| 3.3 | Σχήμα Δειγματοληψίας – (3×2) : Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i \in \{1, 2, 3\}$ και $j \in \{1, 2\}$ | 67 |
| 3.4 | Σχήμα Δειγματοληψίας – (2×3) : Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i \in \{1, 2\}$ και $j \in \{1, 2, 3\}$ | 69 |
| 4.1 | Πιθανότητες Συναγερμού $AR_{O2}^{(2)}$ του δίπλευρου διαγράμματος $O2$ λόγω μετατοπίσεων μέσης τιμής για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) | 97 |
| 4.2 | Πιθανότητες Λανθασμένου Συναγερμού $FAR_{O2}^{(2)}$ του δίπλευρου διαγράμματος $O2$ για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) | 99 |
| 4.3 | Πιθανότητες Λανθασμένου Συναγερμού $FAR_{O2}^{(2)}$ του μονόπλευρου διαγράμματος $O2$ για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) | 100 |
| 4.4 | Αλγόριθμος υπολογισμού ARL_{in} και ARL_{out} του διαγράμματος $O2$ | 102 |

- 4.5 Απόδοση του δίπλευρου διαγράμματος $O2$ βάσει του ARL_{out} σε σχέση με 3 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων μέσης τιμής, χρησιμοποιώντας δεδομένα από κανονική και $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) 105
- 4.6 Απόδοση του δίπλευρου διαγράμματος $O2$ βάσει του ARL_{out} σε σχέση με 3 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση άνισων μετατοπίσεων μέσης τιμής, χρησιμοποιώντας δεδομένα από κανονική και $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) 106
- 5.1 Πιθανότητες Συναγερμού $AR_{OC}^{(2)}$ του δίπλευρου διαγράμματος OC για δεδομένο σχεδιασμό ($b = m - a + 1$ & $r = s = (n + 1)/2$), χρησιμοποιώντας το σύνδεσμο του *Gauss* με διαφορετικές περιθώριες κατανομές και $\rho_{in}^{(R)} = 0.9$ 126
- 5.2 Απόδοση του δίπλευρου διαγράμματος OC βάσει του AR σε σχέση με το διάγραμμα $O2$ με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων μέσης τιμής, χρησιμοποιώντας δεδομένα από $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) 131
- 5.3 Απόδοση του δίπλευρου διαγράμματος OC βάσει του ARL_{out} σε σχέση με 4 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων μέσης τιμής, χρησιμοποιώντας δεδομένα από $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) 132
- 5.4 Έλεγχοι καλής προσαρμογής για την κατανομή του Ελεύθερου Διοξειδίου του Θείου 134
- 5.5 Έλεγχοι καλής προσαρμογής για την κατανομή του Ολικού Διοξειδίου του Θείου 134
- 5.6 Έλεγχοι καλής προσαρμογής του συνδέσμου των GUH με παράμετρο $\alpha = 2.5$ στα ζεύγη των παρατηρήσεων Ελεύθερου και Ολικού Διοξειδίου του Θείου . . . 135
- 6.1 Πιθανότητες Συναγερμού AR_{O4} λόγω μετατοπίσεων μέσης τιμής για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$) 151
- 6.2 Πιθανότητες Συναγερμού AR_{O4} λόγω μετατοπίσεων τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$) 152

| | | |
|-----|---|-----|
| 6.3 | Πιθανότητες Συναγερμού AR_{O4} λόγω ταυτόχρονων μετατοπίσεων μέσης τιμής και τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$) | 153 |
| 6.4 | Πιθανότητες Λανθασμένου Συναγερμού FAR_{O4} για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$) | 155 |
| 6.5 | Πιθανότητες Λανθασμένου Συναγερμού FAR_{O4} για δεδομένο σχεδιασμό ($a > c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1, r_2 = s_2$) | 156 |
| 6.6 | Πιθανότητες Λανθασμένου Συναγερμού FAR_{O4} για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 < s_1, r_2 > s_2$) | 157 |
| 6.7 | Απόδοση του διαγράμματος $O4$ βάσει του ARL_{out} σε σχέση με 6 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων τυπικής απόκλισης χρησιμοποιώντας δεδομένα από κανονική κατανομή για δεδομένο σχεδιασμό | 160 |
| 6.8 | Απόδοση του διαγράμματος $O4$ βάσει του ARL_{out} σε σχέση με 6 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων τυπικής απόκλισης χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό | 161 |
| 6.9 | Απόδοση του διαγράμματος $O4$ βάσει του ARL_{out} σε σχέση με 2 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση ταυτόχρονων μετατοπίσεων μέσης τιμής και μεταβλητότητας χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό | 163 |
| 7.1 | Διάσπαση Ποσοστών Συναγερμού του διαγράμματος $O2N2$ για δεδομένο σχεδιασμό ($m = 500, n = 25, a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$) | 175 |
| 7.2 | Πιθανότητες Συναγερμού AR_{O2N2} λόγω μετατοπίσεων μέσης τιμής για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$) | 177 |
| 7.3 | Πιθανότητες Συναγερμού AR_{O2N2} λόγω μετατοπίσεων τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$) | 178 |
| 7.4 | Πιθανότητες Συναγερμού AR_{O2N2} λόγω ταυτόχρονων μετατοπίσεων μέσης τιμής και τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$) | 179 |
| 7.5 | Πιθανότητες Λανθασμένου Συναγερμού FAR_{O2N2} για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$) | 180 |

- 7.6 Πιθανότητες Λανθασμένου Συναγερμού FAR_{O2N2} για δεδομένο σχεδιασμό ($a > c$,
 $b = m - a + 1$, $d = m - c + 1$ & $r = s = (n + 1)/2$, $\ell = \ell_X = \ell_Y$) 181
- 7.7 Πιθανότητες Λανθασμένου Συναγερμού FAR_{O2N2} για δεδομένο σχεδιασμό ($a = c$,
 $b = m - a + 1$, $d = m - c + 1$ & $r = s = (n + 1)/2$, $\ell_X > \ell_Y$) 182
- 7.8 Απόδοση του διαγράμματος $O2N2$ βάσει του ARL_{out} σε σχέση με 7 ανταγωνιστικά
διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων τυπικής
απόκλισης χρησιμοποιώντας δεδομένα από κανονική κατανομή για δεδομένο
σχεδιασμό 184
- 7.9 Απόδοση του διαγράμματος $O2N2$ βάσει του ARL_{out} σε σχέση με 7 ανταγωνι-
στικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων τυ-
πικής απόκλισης χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο
σχεδιασμό 185
- 7.10 Απόδοση διαγραμμάτων $O4$ και $O2N2$ βάσει του ARL_{out} με κοινό $ARL_{in} \approx 200$
για την ανίχνευση ταυτόχρονων μετατοπίσεων μέσης τιμής και μεταβλητότητας
χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό 186

Κατάλογος Συντομογραφιών

| | |
|--------------------------|--|
| α.σ.κ. | αθροιστική συνάρτηση κατανομής |
| β.ε. | βαθμοί ελευθερίας |
| ΘΟΠ | Θεώρημα Ολικής Πιθανότητας |
| σ.π.π. | συνάρτηση πυκνότητας πιθανότητας |
| σ.σ. | στατιστική συνάρτηση |
| τ.μ. | τυχαία μεταβλητή |
| AMH | Ali-Mikhail-Haq Copula / Σύνδεσμος <i>AMH</i> |
| AR | Alarm Rate / Πιθανότητα Συναγερμού |
| ARL | Average Run Length / Μέσο Μήκος Ροής |
| ARL_{in} | In-Control Average Run Length / Εντός Ελέγχου Μέσο Μήκος Ροής |
| ARL_{out} | Out-of-Control Average Run Length / Εκτός Ελέγχου Μέσο Μήκος Ροής |
| ATS | Average Time to Signal / Μέσος Χρόνος Σήματος |
| CL | Center Line / Κεντρική Γραμμή |
| CLP | Clayton-Pareto Copula / Σύνδεσμος <i>CLP</i> |
| CUSUM | Cumulative Sum / Συσσωρευμένο Άθροισμα |
| DDMA | Data Depth Moving Average / Κινούμενος Μέσος που βασίζεται στο Βάθος Δεδομένων |
| EWMA | Exponentially Weighted Moving Average / Εκθετικά Σταθμισμένος Κινούμενος Μέσος |
| FAR | False Alarm Rate / Ποσοστό Λανθασμένου Συναγερμού |
| FGM | Farlie-Gumbel-Morgenstern Copula / Σύνδεσμος <i>FGM</i> |

| | |
|--|--|
| <i>FRA</i> | Frank's Copula / Σύνδεσμος <i>FRA</i> |
| <i>I</i> | Individual units sampled / Αναμενόμενος αριθμός μεμονωμένων μονάδων δειγματοληψίας |
| <i>IC</i> | In-Control / Εντός Ελέγχου |
| <i>ISO</i> | International Standards Organization / Διεθνής Οργανισμός Τυποποίησης |
| <i>IWL</i> | Inner Warning Limits / Εσωτερικά Προειδοποιητικά Όρια |
| <i>JUSE</i> | Japanese Union for Scientists and Engineers / Ιαπωνική Ένωση Επιστημόνων και Μηχανικών |
| <i>LCL</i> | Lower Control Limit / Κάτω Όριο Ελέγχου |
| <i>MA</i> | Moving Average / Κινούμενος Μέσος |
| <i>MD_F</i> | Mahalanobis Depth / Βάθος δεδομένων που βασίζεται στην απόσταση του Mahalanobis |
| <i>MD_{F_m}</i> | Sample Mahalanobis Depth / Εμπειρικό βάθος δεδομένων που βασίζεται στην απόσταση του Mahalanobis |
| <i>OCF</i> | Operating Characteristic Function / Χαρακτηριστική Συνάρτηση Λειτουργίας |
| <i>OWL</i> | Outer Warning Limits / Εξωτερικά Προειδοποιητικά Όρια |
| <i>RL</i> | Run Length / Μήκος Ροής |
| <i>SN test</i> | Sign Test / Προσημικός Έλεγχος |
| <i>SR test</i> | Signed-Rank Test / Βαθμολογικός Προσημικός Έλεγχος |
| <i>UCL</i> | Upper Control Limit / Άνω Όριο Ελέγχου |

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή στο Στατιστικό Έλεγχο Ποιότητας

1.1 Ιστορική Αναδρομή

Μπορεί να μην είναι ευρέως γνωστό, αλλά τα πρώτα πρώιμα σημάδια ύπαρξης βασικών μεθόδων ποιοτικού ελέγχου σημειώνονται στη μακρινή αρχαιότητα. Κατά τον Howell (1952), το πιο πρώιμο παράδειγμα Ποιοτικού Ελέγχου χρονολογείται εκατομμύρια χρόνια πριν (συγκεκριμένα την εποχή του λίθου), όταν ο προϊστορικός άνθρωπος κατασκεύασε τον πρώτο πέλεκυ χρησιμοποιώντας ως πρώτη ύλη το ξύλο και την πέτρα. Οι ιστορικοί υποστηρίζουν ότι η Μεγάλη Πυραμίδα της Γκίζας κατασκευάστηκε το 2584 π.Χ. σύμφωνα με ένα εγγυημένο σύστημα ποιότητας. Επίσης, κατά τους Gustafsson *et al.* (1999), ο Κώδικας του Χαμουραμί¹, καθώς και το Σύνταγμα του Τσου² μπορούν να χαρακτηρισθούν ως τα πρώτα συστήματα ποιότητας με οργανωτικές δομές.

Επιπλέον, στην Αρχαία Ελλάδα χρησιμοποιήθηκε κάποιο είδος συστήματος για τη διασφάλιση και τη διαχείριση της ποιότητας. Ειδικότερα, διάφορες επιγραφές που βρέθηκαν έφεραν στο φως ότι οι αρχαίοι Έλληνες χρησιμοποιούσαν αυστηρές προδιαγραφές στις παραγγελίες τους για χάλκινα, ασημένια ή χρυσά αντικείμενα, ακόμη και νομίσματα. Για παράδειγμα, στην επιγραφή της Ελευσίνας (4ος αι. π.Χ.) περιγράφεται ένας εμπειρικός τρόπος ελέγχου των τεχνικών και χημικών προδιαγραφών που απαιτούνταν για την κατασκευή συνδέσμων πόλων και εμπολίων, τα οποία θα έμπαιναν ανάμεσα στους σπονδύλους

¹ Ο κώδικας του Χαμουραμί αποτελεί μια από τις αρχαιότερες σειρές νόμων της αρχαίας Μεσοποταμίας. Συνολικά υπήρχαν 282 νόμοι, οι οποίοι έγιναν ιδιαίτερα γνωστοί για την ακρίβεια, αλλά και τη σκληρότητά τους, διότι περιγράφουν τιμωρίες που καταλήγουν συνήθως σε τραυματισμό ή ακόμη και στη θανατική ποινή. Βασική αρχή του ποινικού δικαίου στον Κώδικα του Χαμουραμί ήταν η αρχή της «ανταπόδοσης των ίσων».

² Το Σύνταγμα του Τσου περιέχει πληροφορίες και νόμους σχετικά με τη δημόσια διοίκηση και τον τρόπο με τον οποίο ο αρχηγός του κράτους ελέγχει την κυβέρνηση. Η αρχική οργανωτική δομή του συγκεκριμένου συστήματος χρησιμοποιείται ακόμα και σήμερα στην Κίνα.

κίωνων της Φιλώνειας Στοάς (Βαρουφάκης (2005), Παπαργύρης και Παπαργύρης (2010)).

Παράλληλα, υπήρχε ολόκληρος μηχανισμός ελέγχου της ποιότητας των προϊόντων (τροφίμων και των ποτών), ο οποίος διεξαγόταν μέσω δημόσιων λειτουργιών που ορίζονταν με κλήρο. Όπως αναφέρει ο Αριστοτέλης στο έργο του *Αθηναίων Πολιτεία*, τα προς πώληση αγαθά έπρεπε να είναι «καθαρά» και «ακίβδηλα», καθώς έπρεπε να είναι ανόθευτα και να πληρούν συγκεκριμένους όρους υγιεινής. Για το σκοπό αυτό, η πολιτεία όριζε 10 Αγορανόμους και 10 Μετρονόμους (5 στην Αθήνα και 5 στον Πειραιά) που ήταν υπεύθυνοι για τον έλεγχο σωστής χρήσης των μέτρων και των σταθμών, τα οποία έπρεπε να ανταποκρίνονται σε ορισμένες προδιαγραφές. Επιπλέον, ορίζονταν 35 Σιτοφύλακες (20 στην Αθήνα και 15 στον Πειραιά), οι οποίοι ήταν υπεύθυνοι για τον έλεγχο της ποιότητας του σιταριού και του κριθαριού (Champ (2004), Βαρουφάκης (2005)). Αξίζει να σημειωθεί ότι κάτι ανάλογο παρατηρείται και στη Μεσαιωνική Ευρώπη, καθώς τεχνίτες οργανώνονταν σε ειδικά συνδικάτα, που ονομάζονταν συντεχνίες, προκειμένου να επιθεωρήσουν την ποιότητα των παραγόμενων προϊόντων (Hellman and Liu (2013)).

Παρόλο που ο Έλεγχος Ποιότητας συνεχίζει να εξελίσσεται στο πέρασμα των αιώνων, πριν τη Βιομηχανική Επανάσταση δε δόθηκε έμφαση στις στατιστικές μεθόδους για τον ποιοτικό έλεγχο και τη βελτίωση ενός προϊόντος, καθώς και μιας ολόκληρης παραγωγικής διεργασίας. Μόνο τότε, η έννοια του Στατιστικού Ελέγχου Ποιότητας (*Statistical Quality Control*) και οι τεχνικές του απέκτησαν ευρεία αποδοχή. Η μεγαλύτερη ίσως καινοτομία που έγινε στη βιομηχανία χρονολογείται λίγο πριν το 1800, όταν ο E. Whitney δημιούργησε ανταλλάξιμα τμήματα όπλων, τα οποία έπρεπε να είναι πανομοιότυπα για να έχουν σωστή εφαρμογή (Howell (1952)). Το 1975, εισήχθησαν από τον F. W. Taylor, αρχές Επιστημονικής Διοίκησης (*Scientific Management*) με σκοπό τον καταμερισμό της εργασίας σε επιμέρους καθήκοντα, ώστε -μέσω εφαρμογής τυποποιημένων μεθόδων- να απλοποιηθεί η κατασκευή και η συναρμολόγηση ενός προϊόντος (Montgomery (2009)).

Κατά τη χρονική περίοδο 1900-1930, δίνεται περισσότερη έμφαση στις εργασιακές μεθόδους με αποτέλεσμα να παρατηρείται βελτίωση της παραγωγικότητας. Συγκεκριμένα, ο H. Ford εισάγει τις έννοιες του αυτοελέγχου και της επιθεώρησης κατά τη διάρκεια μιας παραγωγικής διαδικασίας. Το 1907 η εταιρεία *AT & T (American Telephone & Telegraph)* ξεκινά τη συστηματική επιθεώρηση στον έλεγχο προϊόντων και υλικών, ενώ 20 περίπου χρόνια αργότερα δημιουργεί τμήμα ποιοτικού ελέγχου που εστιάζει στην ποιότητα, επιθεώρηση και αξιοπιστία των παραγόμενων προϊόντων. Το 1908, ο W. S. Gosset εισάγει την κατανομή *t (student)*, τα αποτελέσματα της οποίας αξιοποιούνται από τη ζυθοποιία *Gui-*

ness. Μεταξύ 1915 και 1919, ιδρύεται στην Αγγλία ο Σύνδεσμος Τεχνικής Επιθεώρησης (*Technical Inspection Association*), ο οποίος μετονομάστηκε αργότερα σε Ινστιτούτο Διασφάλισης Ποιότητας (*Institute of Quality Assurance*). Παράλληλα, στην Αγγλία, η εταιρεία *General Electric* εφάρμοσε στατιστικές μεθόδους για τον έλεγχο της ποιότητας ηλεκτρικών λαμπτήρων. Αξίζει να σημειωθεί ότι τη διετία 1922-1923 δημοσιεύεται για πρώτη φορά μία σειρά θεμελιωδών μελετών στους Πειραματικούς Σχεδιασμούς από τον R. A. Fisher.

Το ορόσημο, όμως, στην ιστορία της ποιότητας εντοπίζεται το 1926, όταν ο W. Shewhart εισήγαγε την έννοια του διαγράμματος ελέγχου σε ένα υπόμνημα που κυκλοφόρησε από την εταιρεία στην οποία εργαζόταν, την *Bell Telephone Laboratories*. Το 1928, η Δειγματοληψία Αποδοχής (*Acceptance Sampling*) συστηματοποιείται από τους Dodge and Romig (1929). Το 1931, ο W. Shewhart εξέδωσε το κλασικό του βιβλίο “*Economic Control of Quality of Manufactured Product*”, στο οποίο περιγράφει μεθόδους που μπορούν να εφαρμοσθούν σε διαγράμματα ελέγχου και σε παραγωγικές διαδικασίες. Στη δεκαετία του 1930 και 1940, οι W. Shewhart και W. E. Deming έδιναν σεμινάρια όπου παρουσιάζονταν θέματα σχετικά με διαγράμματα ελέγχου και του Στατιστικού Ελέγχου Ποιότητας, αντίστοιχα. Σύμφωνα με τον Montgomery (2009), η έλευση του Δευτέρου Παγκοσμίου Πολέμου συντέλεσε στο να επεκταθεί η χρήση των μεθόδων Στατιστικού Ελέγχου Ποιότητας, οι οποίες μέχρι τότε εφαρμόζονταν μόνο από την εταιρεία *Western Electric* και κρίθηκαν απαραίτητες κατά τη διάρκεια του πολέμου λόγω έλλειψης εξειδικευμένου προσωπικού στον τομέα παραγωγής. Αξίζει να τονισθεί ότι το έτος 1946 ήταν πολύ σημαντική χρονιά στην ιστορία του ποιοτικού ελέγχου, διότι ιδρύθηκαν οι εξής οργανισμοί: η Αμερικανική Εταιρεία Ελέγχου Ποιότητας (*American Society for Quality Control*), η οποία το 1997 μετονομάστηκε σε Αμερικανική Εταιρεία Ποιότητας (*American Society for Quality*), η Ιαπωνική Ένωση Επιστημόνων και Μηχανικών (*Japanese Union for Scientists and Engineers, JUSE*) και ο Διεθνής Οργανισμός Τυποποίησης (*International Organization for Standardization, ISO*).

Τη δεκαετία του 1950, ο A. V. Feigenbaum (1951) εξέδωσε το βιβλίο “*Total Quality Control*” και οι Juran and Gryna (1957) το βιβλίο “*Quality Control Handbook*”, ενώ ο K. Ishikawa εισήγαγε το διάγραμμα αιτίας-αποτελέσματος (*cause and effect diagram*). Ταυτόχρονα, η ένωση *JUSE*, καθιέρωσε το βραβείο Deming με σκοπό την επιβράβευση σημαντικών επιτευγμάτων στον Έλεγχο Ποιότητας. Την επόμενη δεκαετία, ο K. Ishikawa εισήγαγε την έννοια του κύκλου ποιότητας (*Quality Control Circle*) και οι Box and Hunter (1961) μελέτησαν τους κλασματικούς παραγοντικούς σχεδιασμούς 2^{k-p} . Επιπρόσθετα, στη Μεγάλη Βρετανία σχηματίζεται το Εθνικό Συμβούλιο Ποιότητας Παραγωγικότητας

(*National Council of Quality and Productivity*) ως μέρος του Βρετανικού Συμβουλίου Παραγωγικότητας (*British Productivity Council*). Είναι σημαντικό να αναφερθεί ότι στο τέλος της δεκαετίας του 1960, διακόπηκε η έκδοση του περιοδικού *Industrial Quality Control* και στη θέση του άρχισαν να εκδίδονται τα περιοδικά *Quality Process* και *Journal of Quality Technology*.

Τη δεκαετία του 1970, ιδρύθηκε ο Βρετανικός Σύνδεσμος Ποιότητας (*British Quality Association*) και η εμφάνιση βιβλίων σχετικών με τους Πειραματικούς Σχεδιασμούς οδήγησε στην ανάπτυξη της Ολικής Ποιότητας (*Total Quality Management*). Επίσης, τη δεκαετία του 1980, οι Πειραματικοί Σχεδιασμοί αξιοποιούνται στις βιομηχανίες και για πρώτη φορά εμφανίζονται εργασίες του G. Taguchi. Στη διάρκεια της ίδιας δεκαετίας πρωτοεκδίδονται τα περιοδικά *Quality & Reliability Engineering International* και *Quality Engineering*. Είναι, επίσης, σημαντικό να υπογραμμισθεί ότι η Αμερικανική Στατιστική Ένωση (*American Statistical Society*) καθιέρωσε μία Ad Hoc επιτροπή για την Ποιότητα και την Παραγωγικότητα, η οποία αργότερα εξελίχθηκε σε ολόκληρο τομέα. Ταυτόχρονα, το Κογκρέσο των Η.Π.Α. καθιέρωσε το Βραβείο Ποιότητας *Malcom Baldrige*.

Ολοκληρώνοντας την ιστορική αναδρομή αναφέρεται ότι, τη δεκαετία του 1990, η εταιρεία *Motorola* εφάρμοσε το πρόγραμμα ποιότητας 6σ (*Six Sigma*), το οποίο έκτοτε άρχισε να διαδίδεται και σε άλλες εταιρείες. Το σημαντικότερο επίτευγμα θεωρείται η καθιέρωση διάφορων προτύπων ποιότητας, όπως λόγω χάρη το πρότυπο ISO 9000. Σύμφωνα με τα παραπάνω, γίνεται σαφές ότι η ανάπτυξη του Στατιστικού Ελέγχου Ποιότητας ήταν ραγδαία τις τελευταίες δεκαετίες, με αποτέλεσμα οι τεχνικές του να εφαρμόζονται σε ολόκληρο το φάσμα της βιομηχανίας και σχεδόν σε κάθε στάδιο της παραγωγικής διεργασίας. Αναμφισβήτητα, σε αυτό συνέβαλε και η ανάπτυξη της τεχνολογίας, η οποία συνετέλεσε, επίσης, στη βελτίωση των στατιστικών μεθόδων του παρελθόντος και ώθησε στη μελέτη και κατασκευή μοντέρνων-σύγχρονων στατιστικών εργαλείων με κυριότερο όλων το διάγραμμα ελέγχου.

1.2 Περιγραφή Τυπικού Διαγράμματος Ελέγχου

Στόχος κάθε παραγωγικής διαδικασίας είναι η έγκαιρη ανίχνευση πιθανών αλλαγών που προκύπτουν στη διάρκεια του χρόνου. Αυτό επιτυγχάνεται με την παρακολούθηση της συμπεριφοράς μιας κρίσιμης ποσότητας, η οποία αναφέρεται κατά κύριο λόγο σε ένα (μετρήσιμο) χαρακτηριστικό, έστω X , των παραγόμενων προϊόντων και ουσιαστικά χαρακτηρίζει την ποιότητά τους. Πιο συγκεκριμένα, λαμβάνονται τυχαία από τη διεργασία δείγματα της

μορφής X_1, X_2, \dots, X_n που προέρχονται από το υπό μελέτη χαρακτηριστικό X με σκοπό τον υπολογισμό της τιμής μιας στατιστικής συνάρτησης (σ.σ.), συμβ. $W_i = g(X_i)$, $i = 1, 2, \dots, n$. Συνήθως, αυτή η σ.σ. αποτελεί αμερόληπτη εκτιμήτρια μιας κρίσιμης ποσότητας (παραμέτρου) που καθορίζει την ποιότητα του παραγόμενου προϊόντος, η διαχρονική παρακολούθηση της οποίας γίνεται βάσει των παρατηρούμενων τιμών μέσω των δειγμάτων.

Το βασικό εργαλείο που χρησιμοποιείται σε αυτές τις περιπτώσεις είναι το διάγραμμα ελέγχου, το οποίο δεν είναι παρά μια διδιάστατη γραφική απεικόνιση της εξέλιξης της παραγωγικής διεργασίας στην πάροδο του χρόνου. Ο οριζόντιος άξονας αντιπροσωπεύει το χρόνο ή τα δείγματα που συλλέγονται με χρονολογική σειρά, ενώ στον κάθετο άξονα καταγράφονται οι τιμές της ποσότητας W_i , $i = 1, 2, \dots, n$, οι οποίες απεικονίζονται με σημεία και ενώνονται με μία τεθλασμένη γραμμή. Στο διάγραμμα υπάρχουν, επίσης, τρεις γραμμές: η **κεντρική γραμμή** (*Center Line, CL*) που αντιπροσωπεύει το επίπεδο της παραγωγικής διεργασίας υπό σταθερές συνθήκες και δύο όρια, το **κάτω όριο ελέγχου** (*Lower Control Limit, LCL*) και το **άνω όριο ελέγχου** (*Upper Control Limit, UCL*). Τα δύο αυτά όρια τοποθετούνται εκατέρωθεν της κεντρικής γραμμής και δηλώνουν το φυσικό εύρος της διακύμανσης της ποσότητας W .

Η κατασκευή ενός τυπικού διαγράμματος βασίζεται σε δύο κυρίως πρότυπα, το **μοντέλο ορίων $L\sigma$** (*sigma limits model*) και το **μοντέλο πιθανότητας α** (*probability limits model*) [βλ. Αντζουλάκος (2010)]. Σύμφωνα με το πρώτο πρότυπο, τα όρια τοποθετούνται σε απόσταση L μονάδων τυπικής απόκλισης από την κεντρική γραμμή, η οποία τίθεται ίση με το μέσο της διεργασίας, όπως φαίνεται παρακάτω

$$\text{Μοντέλο Ορίων } L\sigma : \begin{cases} LCL = \mu_W - L \cdot \sigma_W \\ CL = \mu_W \\ UCL = \mu_W + L \cdot \sigma_W. \end{cases}$$

Στο σημείο αυτό, θα πρέπει να σημειώσουμε ότι τα όρια που προκύπτουν για $L = 1$ αναφέρονται συνήθως ως **εσωτερικά προειδοποιητικά όρια** (*Inner Warning Limits, IWL*), ενώ τα όρια που βρίσκονται σε απόσταση $L = 2$ ονομάζονται **εξωτερικά προειδοποιητικά όρια** (*Outer Warning Limits, OWL*). Ωστόσο, η συνήθης επιλογή είναι αυτή των **φυσικών ορίων ανοχής** (*Natural Tolerance Limits*), η οποία προκύπτει για $L = 3$.

Εναλλακτικά, στην περίπτωση που η σ.σ. W ακολουθεί την Κανονική κατανομή, προ-

κύπτει το δεύτερο πρότυπο που περιγράφεται ως ακολούθως

$$\text{Μοντέλο Πιθανότητας } \alpha : \begin{cases} LCL = w_{\alpha/2} = \mu_W - z_{\alpha/2} \cdot \sigma_W \\ CL = w_{0.5} = \mu_W \\ UCL = w_{1-\alpha/2} = \mu_W + z_{\alpha/2} \cdot \sigma_W, \end{cases}$$

όπου w_α είναι το α άνω-ποσοστιαίο σημείο της κατανομής της σ.σ. W .

1.3 Εισαγωγικές Έννοιες Ελέγχου Ποιότητας

Όπως αναφέραμε και στην Ενότητα 1.2, ένα διάγραμμα ελέγχου χρησιμοποιείται γενικά για την έγκαιρη ανίχνευση της «αφύσικης» συμπεριφοράς μιας διεργασίας. Ανεξάρτητα, όμως, από το πόσο καλά έχει οργανωθεί ή συντηρηθεί μία παραγωγική διεργασία, πάντα υπάρχει η **φυσική μεταβλητότητα** (*natural variability*), η οποία είναι το σωρευτικό αποτέλεσμα πολλών επιμέρους, μικρότερων αιτιών. Οι τελευταίες επιδράσεις αναφέρονται στη βιβλιογραφία με τον όρο **κοινές ή τυχαίες αιτίες μεταβλητότητας** (*common or chance causes of variation*) και θεωρούνται αναπόσπαστο κομμάτι κάθε διεργασίας. Όταν είναι παρούσες μόνο τέτοιου είδους αιτίες, τότε η διεργασία κρίνεται **εντός στατιστικού ελέγχου** (*in-statistical control*) ή θεωρείται ότι λειτουργεί σε **ευσταθή κατάσταση** (*stable stage*).

Σε μια παραγωγική διαδικασία μπορεί, επίσης, να υπάρχει και η **ειδική μεταβλητότητα** (*special variability*), η οποία είναι μεγαλύτερη σε σχέση με τη φυσική μεταβλητότητα και προκαλείται από **ειδικές ή συστηματικές αιτίες μεταβλητότητας** (*special or assignable causes of variation*). Οι αιτίες αυτές μπορεί να είναι, λόγου χάρη, ελαττωματικές πρώτες ύλες, ακατάλληλα προσαρμοσμένες μηχανές, σφάλματα των χειριστών των μηχανημάτων, κ.ά., οι οποίες επηρεάζουν την ποιότητα του χαρακτηριστικού που μελετάται και οδηγούν σε ένα μη αποδεκτό επίπεδο απόδοσης της διεργασίας. Όταν μία διεργασία λειτουργεί με την παρουσία αυτών των πρόσθετων αιτιών μεταβλητότητας θεωρείται ότι είναι **εκτός στατιστικού ελέγχου** (*out-of-control*) ή ότι λειτουργεί σε **ασταθή κατάσταση** (*unstable stage*) και κατά συνέπεια θα πρέπει να ενεργοποιηθεί **συναγερμός** (*alarm*) στο διάγραμμα.

Συνεπώς, χρησιμοποιώντας ένα τυπικό διάγραμμα ελέγχου, εύκολα μπορούμε να αποφανθούμε σχετικά με το εάν η διεργασία βρίσκεται εντός ή εκτός ελέγχου. Ειδικότερα, λαμβάνοντας υπ' όψιν μας τα μοντέλα ορίων $L\sigma$ και πιθανότητας a που παρουσιάστηκαν στην Ενότητα 1.2, θεωρούμε ότι η διεργασία λειτουργεί σε ευσταθή κατάσταση όταν οι τιμές της ποσότητας W βρίσκονται μεταξύ των αντίστοιχων ορίων ελέγχου. Ωστόσο, σε αρκετές περιπτώσεις, τα σημεία ενός διαγράμματος μπορεί να εμφανίζουν **ειδικές ακο-**

λουθίες (*special patterns*) χωρίς απαραίτητα να βρίσκονται εκτός της περιοχής ελέγχου που καθορίζουν τα όρια. Αυτές οι ακολουθίες δημιουργούν συγκεκριμένα μοτίβα με αποτέλεσμα τα σημεία του διαγράμματος να συμπεριφέρονται με συστηματικό, μη τυχαίο τρόπο και σε αυτή την περίπτωση υπάρχει ένδειξη ότι η διεργασία πιθανώς βρίσκεται σε εκτός ελέγχου κατάσταση.

Τα βασικά πρότυπα που εμφανίζονται συχνότερα σε μία παραγωγική διαδικασία είναι το κυκλικό ή περιοδικό πρότυπο (*cycles*), το μικτό πρότυπο (*mixture pattern*), το πρότυπο αλμάτων (*shifts in process level*), το πρότυπο τάσης (*trend in process level*) και το πρότυπο έλλειψης μεταβλητότητας (*stratification pattern*). Σε αυτές τις περιπτώσεις, οι ειδικές ακολουθίες μπορούν να εντοπισθούν χάρη σε κανόνες που έχουν δημιουργηθεί με βάση τα προειδοποιητικά όρια *IWL* και *OWL* του μοντέλου ορίων 3σ . Οι τέσσερις πιο γνωστοί κανόνες αναφέρονται στη βιβλιογραφία ως *Western Electric Rules* (Western Electric (1956)), ενώ έχουν προταθεί και πρόσθετοι κανόνες ροών (βλ. για παράδειγμα Nelson (1984, 1985), Champ and Woodall (1987)). Αναλυτικότερη παρουσίαση των προαναφερθέντων προτύπων και κανόνων γίνεται στους Montgomery (2009) και Αντζουλάκος (2010) [βλ. επίσης Μπερσίμης (2005), Ρακιτζής (2008)].

Στη βιβλιογραφία, η μελέτη σχεδιασμού ενός διαγράμματος ελέγχου, καθώς και η εφαρμογή του περιλαμβάνει δύο ξεχωριστές φάσεις: τη Φάση I και τη Φάση II (Woodall (2000), Montgomery (2009), Αντζουλάκος (2010)). Στη **Φάση I**, η οποία συχνά καλείται και **Αναδρομική Φάση** (*Phase I or Off-Line Control Phase or Initial Study Phase or Retrospective Analysis*), τα διαγράμματα ελέγχου χρησιμοποιούνται για να ελέγξουν αναδρομικά εάν η διεργασία ήταν εντός ή εκτός ελέγχου, όταν συλλέχθηκαν τα ιστορικά δείγματα. Εφόσον η διεργασία κρίνεται εντός ελέγχου -σε διαφορετική περίπτωση λαμβάνονται διορθωτικές ενέργειες, ώστε η διεργασία να τεθεί σε εντός ελέγχου κατάσταση- καθορίζονται αξιόπιστα όρια ελέγχου, τα οποία χρησιμοποιούνται για τη μελλοντική παρακολούθηση της διεργασίας. Στη **Φάση II** (*Phase II or On-Line Control Phase or Control to Standard Phase or Prospective Analysis or Process Monitoring*), το ενδιαφέρον επικεντρώνεται στον έλεγχο των μελλοντικών δεδομένων, προκειμένου να ληφθεί απόφαση σχετικά με το εάν η διεργασία εξακολουθεί να παραμένει εντός ελέγχου ή έχει μετατοπισθεί σε ένα εκτός ελέγχου στάδιο.

Η ανεπαρκής μελέτη και ο μη προσεκτικός σχεδιασμός της Φάσης I μπορεί να επιφέρει σημαντικές επιπτώσεις στη Φάση II. Ως εκ τούτου συχνά υποστηρίζεται ότι η επιτυχής ανάλυση της Φάσης II επηρεάζεται σε σημαντικό βαθμό από τον επιτυχή σχεδιασμό της Φάσης

I. Σύμφωνα με αυτή την άποψη, τόσο οι ερευνητές, όσο και οι χρήστες των διαγραμμάτων θα πρέπει να κατανοήσουν πλήρως τη φύση των ειδικών αιτιών μεταβλητότητας πριν προχωρήσουν στον προσδιορισμό των ορίων ελέγχου. Η σημασία της ανάλυσης της Φάσης I τονίζεται στις εργασίες των Chakraborti *et al.* (2008), Jones-Farmer *et al.* (2014), οι οποίοι έκαναν και μία ανασκόπηση της βιβλιογραφίας που συνδέεται με διαγράμματα ελέγχου Φάσης I. Επιπλέον, οι Jensen *et al.* (2006) υπογράμμισαν ότι η ποιότητα και το μέγεθος του ιστορικού δείγματος είναι άκρως συνδεδεμένα με την αποδοτικότητα της ανάλυσης που γίνεται κατά τη Φάση II.

Για την επιλογή του κατάλληλου μεγέθους δείγματος και τον ακριβή προσδιορισμό των ορίων ελέγχου θα πρέπει, επίσης, να δίνεται ιδιαίτερη προσοχή στις πιθανότητες του Σφάλματος Τύπου I και Τύπου II. Οι πιθανότητες αυτές ορίζονται σε αναλογία με την έννοια των σφαλμάτων σε έναν στατιστικό έλεγχο υπόθεσης και συνοψίζονται στον Πίνακα 1.1. Πιο συγκεκριμένα, το **Σφάλμα Τύπου I** (*Type I Error* or *Producer's Risk* or *Error α*) είναι η πιθανότητα ένα σημείο να βρεθεί εκτός της περιοχής ελέγχου, δεδομένου ότι η διεργασία θεωρείται εντός ελέγχου. Αξίζει να σημειωθεί πως στον έλεγχο ποιότητας το Σφάλμα Τύπου I καλείται συνήθως και **ποσοστό λανθασμένου συναγερμού** (*False Alarm Rate, FAR*). Με τον όρο **Σφάλμα Τύπου II** (*Type II Error* or *Consumer's Risk* or *Error β*) αναφερόμαστε στην πιθανότητα ένα σημείο να βρεθεί μεταξύ των ορίων, ενώ στην πραγματικότητα η διεργασία είναι εκτός ελέγχου, δηλαδή ενώ η κατανομή του υπό μελέτη χαρακτηριστικού έχει αλλάξει.

| Κατάσταση Διεργασίας | Βάσει του Διαγράμματος Ελέγχου | |
|----------------------|---|---------------------------------------|
| | Εντός Ελέγχου (η παραγωγή συνεχίζεται) | Εκτός Ελέγχου (η παραγωγή σταματά) |
| Πραγματική | | |
| Εντός Ελέγχου | Δεν υπάρχει σφάλμα ($1 - \alpha$) | Σφάλμα Τύπου I (α) |
| Εκτός Ελέγχου | Σφάλμα Τύπου II (β) | Δεν υπάρχει σφάλμα ($1 - \beta$) |

Πίνακας 1.1: Σφάλμα Τύπου I και Σφάλμα Τύπου II

Σύμφωνα με τους παραπάνω ορισμούς, το μέγεθος δείγματος n μπορεί να προσδιορισθεί ελέγχοντας τις τιμές και των δύο σφαλμάτων ταυτόχρονα. Μία αύξηση του n συντελεί στη μείωση των πιθανοτήτων α και β . Λαμβάνοντας υπ' όψιν μας το γενικό μοντέλο ορίων $L\sigma$ της Ενότητας 1.2, για δεδομένο n , η χρήση πλατύτερων ορίων (η οποία επιτυγχάνεται

μεγαλώνοντας την απόσταση L των ορίων ελέγχου από την κεντρική γραμμή), συνεπάγεται τη μείωση της πιθανότητας του Σφάλματος Τύπου I και παράλληλα την αύξηση της πιθανότητας του Σφάλματος Τύπου II, ενώ το αντίθετο συμβαίνει όταν το L λαμβάνει μικρές τιμές.

Ας υποθέσουμε ότι έχουμε κατασκευάσει ένα διάγραμμα ελέγχου με βάση το μοντέλο ορίων $L\sigma$, ότι η διεργασία είναι εντός ελέγχου και ότι η κατανομή της σ.σ. W είναι κανονική, συμβ. $H_0 : W \sim N(\mu_w, \sigma_w^2)$. Επιπλέον, ας υποθέσουμε ότι η διεργασία τίθεται εκτός ελέγχου όταν ο μέσος μετατοπίζεται από τη θέση μ_w στη θέση μ^* , συμβ. $H_1 : W \sim N(\mu^*, \sigma_w^2)$. Σε αυτή την περίπτωση, τα Σφάλματα Τύπου I και II υπολογίζονται αντίστοιχα ως εξής

$$\alpha = P(W \notin [LCL, UCL] | H_0) = 1 - P(LCL \leq W \leq UCL | H_0) = 2\Phi(-L),$$

$$\beta = P(W \in [LCL, UCL] | H_1) = \Phi\left(\frac{\mu_w - \mu^*}{\sigma_w} + L\right) - \Phi\left(\frac{\mu_w - \mu^*}{\sigma_w} - L\right).$$

Προφανώς, στην ειδική περίπτωση που έχουμε θεωρήσει όρια ελέγχου 3σ ($L = 3$), το Σφάλμα Τύπου I ισούται με $\alpha = 2\Phi(-3) = 0.0027$.

Μία ακόμη πολύ χρήσιμη έννοια που σχετίζεται με τα διαγράμματα ελέγχου και συνδέει το σχεδιασμό τους με τις πιθανότητες α και β είναι το **μέσο μήκος ροής** (*Average Run Length, ARL*). Εάν q είναι η πιθανότητα ένα σημείο να βρεθεί εκτός των ορίων ελέγχου, τότε το μέσο μήκος ροής δηλώνει τον αναμενόμενο αριθμό σημείων (δειγμάτων) που τοποθετούνται σε ένα διάγραμμα μέχρις ότου βρεθεί το πρώτο σημείο που ενεργοποιεί συναγερμό. Η πιθανότητα ένα σημείο να βρεθεί εκτός των ορίων στο πρώτο δείγμα είναι απλά q , η πιθανότητα να βρεθεί εκτός των ορίων ελέγχου στο δεύτερο διαδοχικό δείγμα είναι $q(1 - q)$, στο τρίτο διαδοχικό δείγμα είναι $q(1 - q)^2$ κ.ο.κ. Συνεπώς, η πιθανότητα ένα σημείο να βρεθεί εκτός των ορίων στο r -οστό δείγμα είναι q φορές η πιθανότητα τα $r - 1$ προγενέστερα δείγματα να μην ενεργοποιούν συναγερμό, δηλαδή $q(1 - q)^{r-1}$. Προφανώς, εφόσον οι παρατηρήσεις που απεικονίζονται στο διάγραμμα είναι μεταξύ τους ανεξάρτητες, το **μήκος ροής** (*Run Length, RL*) ακολουθεί τη Γεωμετρική Κατανομή με πιθανότητα επιτυχίας q , συμβ. $RL \sim G(q)$, και έχει μέση τιμή ίση με $1/q$. Άρα, το μέσο μήκος ροής υπολογίζεται από τον τύπο

$$ARL = \sum_{r=1}^{\infty} r q (1 - q)^{r-1} = \frac{1}{q}.$$

Θεωρώντας το μοντέλο ορίων $L\sigma$, για μία διεργασία που βρίσκεται εντός στατιστι-

κού ελέγχου, το **εντός ελέγχου μέσο μήκος ροής** (*in-control Average Run Length, ARL_{in}*) υπολογίζεται ως εξής

$$ARL_{in} = ARL_0 = \frac{1}{\alpha},$$

όπου α είναι το Σφάλμα Τύπου I. Είναι εμφανές ότι στην περίπτωση ορίων ελέγχου 3σ ($L = 3$) και εφόσον ταυτόχρονα η κατανομή της σ.σ. W είναι κανονική, το εντός ελέγχου μέσο μήκος ροής είναι $ARL_{in} = 1/0.0027 \approx 370$.

Υπό την προϋπόθεση ότι η διεργασία τίθεται εκτός ελέγχου λόγω μετατόπισης του μέσου από τη θέση μ_w στη θέση μ^* και γίνεται χρήση του μοντέλου ορίων $L\sigma$, το **εκτός ελέγχου μέσο μήκος ροής** (*out-of-control Average Run Length, ARL_{out}*) δίνεται από τον τύπο

$$ARL_{out} = ARL_1 = \frac{1}{1 - \beta} = \frac{1}{power},$$

όπου β είναι το Σφάλμα Τύπου II, δηλαδή η πιθανότητα το διάγραμμα να μην είναι σε θέση να ανιχνεύσει τη μετατόπιση. Η ποσότητα $1 - \beta$ υποδηλώνει την πιθανότητα ένα σημείο να βρίσκεται εκτός των ορίων ελέγχου δεδομένου ότι η διεργασία έχει μετατοπισθεί, η οποία είναι ουσιαστικά η **ισχύς** (*power*) του διαγράμματος ελέγχου.

Τέλος, θα πρέπει να σημειωθεί ότι η απόδοση ενός διαγράμματος ελέγχου μπορεί να βασισθεί και σε άλλα μέτρα απόδοσης. Για παράδειγμα, πολλές φορές παρουσιάζει ενδιαφέρον η μελέτη του **μέσου χρόνου σήματος** (*Average Time to Signal, ATS*), δηλαδή του μέσου (πραγματικού) χρόνου που απαιτείται για να ενεργοποιηθεί συναγερμός στο διάγραμμα ελέγχου. Η ποσότητα *ATS* υπολογίζεται από τον τύπο

$$ATS = h \times ARL,$$

όπου h είναι ο χρόνος που μεσολαβεί για τη λήψη διαδοχικών δειγμάτων, ο οποίος θεωρείται σταθερός. Εναλλακτικά, μπορεί να είναι χρήσιμο το *ARL* να εκφρασθεί ως προς τον **αναμενόμενο αριθμό μεμονωμένων μονάδων δειγματοληψίας** (*Individual units sampled*), έστω I , αντί του αριθμού των δειγμάτων που ελήφθησαν για την ανίχνευση μιας μετατόπισης. Εάν το μέγεθος δείγματος είναι n , η σχέση που συνδέει το *ARL* με το I είναι η ακόλουθη

$$I = n \times ARL.$$

1.4 Είδη Διαγραμμάτων Ελέγχου

Η διάκριση των διαγραμμάτων ελέγχου γίνεται λαμβάνοντας υπ' όψιν ορισμένα χαρακτηριστικά, όπως για παράδειγμα το πλήθος, το είδος και την κατανομή των υπό μελέτη μεταβλητών, το μέγεθος δείγματος, καθώς και την πληροφορία, η οποία σχετίζεται με την κρίσιμη ποσότητα που απεικονίζεται στο διάγραμμα. Επιπλέον, σημαντικό ρόλο παίζει η πιθανή εξάρτηση των μετρήσεων που λαμβάνονται μεταξύ διαδοχικών χρονικών στιγμών, η (απόλυτη) διαφορά των οποίων δεν είναι απαραίτητα σταθερή.

Συγκεκριμένα, εάν οι μετρήσεις που λαμβάνονται αφορούν ένα μόνο ποιοτικό χαρακτηριστικό ($p = 1$), τότε τα διαγράμματα ελέγχου χαρακτηρίζονται ως **μονοδιάστατα** ή **μονομεταβλητά** (*univariate control charts*). Αντίθετα, όταν οι μετρήσεις αφορούν περισσότερα ποιοτικά χαρακτηριστικά ($p > 1$), τότε αναφερόμαστε σε **πολυδιάστατα** ή **πολυμεταβλητά** (*multivariate control charts*) διαγράμματα.

Μία ακόμη διάκριση μπορεί να γίνει ανάλογα με το είδος του χαρακτηριστικού, το οποίο μπορεί να είναι είτε συνεχές είτε διακριτό. Με βάση τη στατιστική ορολογία, εάν το υπό μελέτη χαρακτηριστικό, περιγράφεται από μία συνεχή τυχαία μεταβλητή (τ.μ.), τότε αναφερόμαστε σε **διάγραμμα ελέγχου για μεταβλητές** (*control charts for variables*), ενώ εάν το χαρακτηριστικό περιγράφεται από μία διακριτή τ.μ. αναφερόμαστε σε **διάγραμμα ελέγχου για ιδιότητες** (*control charts for attributes*).

Στην περίπτωση που οι μετρήσεις του χαρακτηριστικού προέρχονται από κάποια γνωστή κατανομή (συνήθως την Κανονική), γίνεται συνήθως εφαρμογή **παραμετρικών** διαγραμμάτων ελέγχου (*parametric control charts*). Στην περίπτωση, όμως, που υπάρχει ελλιπής πληροφορία σχετικά με την κατανομή του χαρακτηριστικού ή η κατανομή του δεν είναι εξ' ολοκλήρου γνωστή, πρέπει να εφαρμόζονται **μη παραμετρικά** ή **ελεύθερα κατανομής** διαγράμματα ελέγχου (*nonparametric or distribution-free control charts*).

Κατηγοριοποιήσεις μπορούν, επίσης, να προκύψουν βάσει του μεγέθους δείγματος του υπό μελέτη χαρακτηριστικού. Πιο συγκεκριμένα, εάν σε κάθε χρονική στιγμή λαμβάνεται μία μέτρηση/παρατήρηση ($n = 1$) από την παραγωγική διαδικασία, τότε χρησιμοποιούνται **διαγράμματα ελέγχου για μεμονωμένες παρατηρήσεις** (*control charts for individual observations*), ενώ εάν το μέγεθος των δειγμάτων των μετρήσεων είναι μεγαλύτερο της μονάδας ($n > 1$), τότε εφαρμόζονται **διαγράμματα ελέγχου για δείγματα ή ομάδες** (*control charts for rational subgroups*). Πολλές φορές, με σκοπό την ταχύτερη ανίχνευση πιθανών μετατοπίσεων, το μέγεθος δείγματος επιλέγεται να μην είναι σταθερό σε κάθε χρονική

στιγμή, αλλά μεταβλητό. Αυτό έχει ως αποτέλεσμα τη διάκριση σε **διαγράμματα ελέγχου σταθερού ή μεταβλητού μεγέθους δείγματος** (*control charts for fixed or variable sample size*), αντίστοιχα.

Επιπρόσθετα, η πληροφορία που σχετίζεται με το χρόνο που λαμβάνονται οι μετρήσεις μπορεί να καθορίσει νέες κατηγορίες διαγραμμάτων. Όταν τα δείγματα συλλέγονται σε σταθερούς χρόνους, τότε αναφερόμαστε σε **διαγράμματα ελέγχου σταθερού χρόνου** (*fixed sampling interval control charts*), διαφορετικά αναφερόμαστε σε **διαγράμματα μεταβλητού χρόνου** (*variable sampling interval control charts*). Επιπρόσθετα, ενδέχεται οι μετρήσεις που λαμβάνονται σε κάθε χρονική στιγμή t , να εξαρτώνται από τις μετρήσεις που ελήφθησαν κατά την προηγούμενη χρονική στιγμή $t - 1$. Ανάλογα, λοιπόν, με το εάν οι μετρήσεις αυτές είναι εξαρτημένες ή όχι, χρησιμοποιούνται αντίστοιχα **διαγράμματα ελέγχου για αυτοσυσχετισμένες ή ασυσχετιστες διεργασίες** (*control charts for autocorrelated or uncorrelated processes*). Στο σημείο αυτό θα πρέπει να επισημανθεί ότι τα διαγράμματα μεταβλητού ή σταθερού χρόνου και μεγέθους δείγματος ανήκουν σε μία ευρύτερη οικογένεια σχημάτων, γνωστών ως **προσαρμοστικά διαγράμματα ελέγχου** (*adaptive control charts*).

Συμπληρωματικά, αναφέρουμε ότι οι προαναφερθείσες κατηγορίες μπορούν να διαχωριστούν σε επιμέρους οικογένειες διαγραμμάτων. Για παράδειγμα, ανάλογα με το εάν η κρίσιμη ποσότητα (το σημείο) που απεικονίζεται σε ένα διάγραμμα κάθε χρονική στιγμή χρησιμοποιεί πληροφορία που συνδέεται με προγενέστερα δείγματα ή βασίζεται στο πιο πρόσφατο δείγμα (αγνοώντας την πληροφορία που περιέχεται σε προηγούμενα δείγματα), γίνεται διάκριση σε **διαγράμματα ελέγχου με μνήμη ή χωρίς μνήμη** (*control charts with or without memory*).

Στην τελευταία κατηγορία ανήκουν τα διαγράμματα ελέγχου τύπου *Shewhart* (*Shewhart type control charts*), τα οποία προτάθηκαν από τον Shewhart (1926, 1931) και θεωρούνται πιο αποτελεσματικά στην ανίχνευση σχετικά μεγάλων μετατοπίσεων (της τάξης μεταξύ 1.5σ και 2σ ή περισσότερο). Αντίθετα, στην κατηγορία των διαγραμμάτων με μνήμη ταξινομούνται τα **διαγράμματα ελέγχου συσσωρευμένων αθροισμάτων** (*Cumulative Sum, CUSUM type control charts*), τα **διαγράμματα ελέγχου για τον εκθετικά σταθμισμένο μέσο** (*Exponentially Weighted Moving Average, EWMA type control charts*) και τα **διαγράμματα ελέγχου κινούμενου μέσου** (*Moving Average, MA type control charts*), η μελέτη των οποίων αναπτύχθηκε πριν από 60 περίπου χρόνια. Συνήθως, τα διαγράμματα *CUSUM* εφαρμόζονται όταν η κατανομή των αρχικών παρατηρήσεων είναι γνωστή, ενώ τα διαγράμματα *EWMA* χρησιμοποιούνται όταν η κατανομή αυτή δεν είναι γνωστή. Το

βασικό πλεονέκτημα και των δύο σχημάτων είναι η ευαισθησία τους σε μικρές μετατοπίσεις μιας παραγωγικής διεργασίας, η οποία συντελεί στην ταχεία ανίχνευσή τους. Αξίζει να υπογραμμισθεί ότι η ευαισθησία των διαγραμμάτων ελέγχου τύπου *Shewhart* βελτιώνεται με τη χρήση κανόνων ροών (βλ. για παράδειγμα Champ and Woodall (1987), Koutras *et al.* (2006, 2007), Antzoulakos and Rakitzis (2008a,b), Rakitzis and Antzoulakos (2011)).

Τέλος, τα διαγράμματα ελέγχου για ιδιότητες περιλαμβάνουν τα **διαγράμματα ελέγχου για το ποσοστό ελαττωματικών προϊόντων** (*control charts for the proportion/fraction of nonconforming units, p-chart*), τα **διαγράμματα ελέγχου για τον αριθμό ελαττωματικών προϊόντων** (*control charts for the number of nonconforming units, np-chart*), τα **διαγράμματα ελέγχου για τον αριθμό ελαττωμάτων** (*control charts for the total number of nonconformities per unit/control charts for nonconformities, c-chart*) και τα **διαγράμματα ελέγχου για το μέσο αριθμό ελαττωμάτων** (*control charts for the average number of nonconformities per unit/control charts for nonconformities per unit, u-chart*). Επιπροσθέτως, τα διαγράμματα ελέγχου για μεταβλητές διαχωρίζονται σε **διαγράμματα ελέγχου για τη μέση τιμή** (*control charts for process mean*), τα **διαγράμματα ελέγχου για τη διασπορά ή διακύμανση/τυπική απόκλιση/μεταβλητότητα** (*control charts for process variance/standard deviation/variability*) ή/και συνδυασμό αυτών, δηλαδή **διαγράμματα ελέγχου για την ταυτόχρονη παρακολούθηση μέσης τιμής και διασποράς** μιας διεργασίας (*control charts for simultaneous monitoring of process mean and variance*).

Είναι περιττό να αναφερθεί ότι ένα διάγραμμα μπορεί να χαρακτηριστεί συνδυάζοντας κάποιες από τις παραπάνω κατηγορίες, ανάλογα με την περίπτωση στην οποία χρησιμοποιείται. Για παράδειγμα, στην ενότητα που ακολουθεί γίνεται ανασκόπηση των πολυδιάστατων, μη παραμετρικών διαγραμμάτων ελέγχου που έχουν προταθεί στη διεθνή βιβλιογραφία, εστιάζοντας κυρίως σε διαγράμματα ελέγχου τύπου *Shewhart* για τη μέση τιμή, καθώς και σε αντίστοιχα διαγράμματα για τη μέση τιμή και τη διασπορά.

1.5 Ανασκόπηση Βιβλιογραφίας

Τις τελευταίες δεκαετίες παρατηρείται μια διαρκώς αυξανόμενη ανάπτυξη ανταγωνιστικών διαγραμμάτων ελέγχου, τα οποία μπορούν να εφαρμοσθούν υπό διαφορετικές συνθήκες. Όπως για παράδειγμα αναφέραμε στην Ενότητα 1.4, όταν τίθεται υπό παρακολούθηση ένα μόνο ποιοτικό χαρακτηριστικό ($p = 1$), χρησιμοποιούνται μονοδιάστατα διαγράμματα ελέγχου, ενώ σε διαφορετική περίπτωση ($p > 1$), θα πρέπει να εφαρμοσθούν πολυδιάστατα διαγράμματα. Αξίζει να υπογραμμισθεί ότι στην τελευταία περίπτωση, η

χρήση p ξεχωριστών μονοδιάστατων διαγραμμμάτων ελέγχου μπορεί να είναι παραπλανητική, καθώς τα p το πλήθος, υπό μελέτη χαρακτηριστικά, είναι πιθανό να συνδέονται μεταξύ τους. Συνεπώς, η πιθανότητα α του Σφάλματος Τύπου I είναι στην πραγματικότητα πολύ μεγαλύτερη από την αντίστοιχη, από κοινού πιθανότητα του p -διάστατου διαγράμματος, η οποία είναι τουλάχιστον $\alpha' = 1 - (1 - \alpha)^p$.

Στο σημείο αυτό, θα πρέπει να τονισθεί ότι η χρήση p μονοδιάστατων διαγραμμμάτων ελέγχου δεν ενδείκνυται ούτε στην περίπτωση που καλούμαστε να ανιχνεύσουμε τη μεταβλητή ή τις μεταβλητές που είναι υπαίτιες για την ενεργοποίηση συναγερού σε ένα πολυδιάστατο διάγραμμα ελέγχου, δεδομένου ότι η συσχέτιση μεταξύ των μεταβλητών πρέπει να λαμβάνεται υπ' όψιν. Τελευταία, αρκετοί ερευνητές έχουν μελετήσει διαφορετικές μεθόδους ικανές για την ορθή ανίχνευση ενός ή περισσότερων χαρακτηριστικών που έθεσαν τη διεργασία εκτός ελέγχου, όπως για παράδειγμα, οι Murphy (1987), Adams (1994), Fuchs and Benjamini (1994), Mason *et al.* (1995), Maravelakis *et al.* (2002) και Maravelakis and Bersimis (2009) μεταξύ άλλων.

Γενικότερα, ερευνητικά άρθρα σχετικά με πολυδιάστατες μεθόδους που εφαρμόζονται στον έλεγχο ποιότητας έχουν δημοσιευθεί από τους Lowry and Montgomery (1995), MacGregor and Kourti (1995), Yeh *et al.* (2006) και Bersimis *et al.* (2007). Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην εργασία των Topalidou and Psarakis (2009) για ανασκόπηση πολυδιάστατων πολυωνυμικών διαγραμμμάτων ελέγχου (*multinomial control charts*) και πολυδιάστατων διαγραμμμάτων για ιδιότητες (*multiattribute control charts*), καθώς και στην εργασία των Saghir and Lin (2015) για επισκόπηση διαγραμμμάτων ελέγχου διασκορπισμένων δεδομένων απαρίθμησης (*control charts for dispersed count data*). Επιπλέον, πρόσφατα θέματα και ιδέες για μελλοντική έρευνα που συνδέονται με το στατιστικό έλεγχο διεργασιών μπορούν να βρεθούν στα άρθρα των Woodall and Montgomery (1999, 2014).

Υπογραμμίζεται ότι, όταν υπάρχει πλήρης (αντίστοιχα ελλιπής) γνώση της κατανομής της που ακολουθούν τα χαρακτηριστικά που μελετώνται, πρέπει να γίνει χρήση παραμετρικών (αντίστοιχα μη παραμετρικών) διαγραμμμάτων. Χωρίς αμφιβολία, η βιβλιογραφία σχετικά με μονοδιάστατα ή πολυδιάστατα παραμετρικά διαγράμματα ελέγχου είναι εκτενής, η βιβλιογραφία που συνδέεται με τον μη παραμετρικό έλεγχο ποιότητας είναι αρκετά περιορισμένη και ακόμη πιο περιορισμένη θεωρείται η βιβλιογραφία για πολυδιάστατα μη παραμετρικά διαγράμματα ελέγχου. Θα πρέπει να σημειωθεί ότι στον παραμετρικό έλεγχο ποιότητας, γίνεται συνήθως η υπόθεση ότι η υποκείμενη διεργασία προέρχεται από κανονικό πληθυσμό. Ως αποτέλεσμα, τόσο η σ.σ. όσο και τα όρια ελέγχου εξαρτώνται από

τις ιδιότητες της συγκεκριμένης (εξ ολοκλήρου) γνωστής κατανομής για την οποία έχουν σχεδιασθεί. Με αυτή την έννοια, τα διαγράμματα αυτά αναμένεται να είναι πιο αποτελεσματικά από τα αντίστοιχα μη παραμετρικά. Ωστόσο, η πραγματική κατανομή μιας παραγωγικής διεργασίας δεν είναι πάντα γνωστή ή δεν υπάρχει αρκετή πληροφορία σχετικά με το είδος και τη μορφή της κατανομής αυτής. Υπό αυτές τις συνθήκες, τα μη παραμετρικά διαγράμματα ελέγχου μπορούν να προσφέρουν χρήσιμες εναλλακτικές λύσεις. Στις εργασίες των Chakraborti *et al.* (2001), Chakraborti and Graham (2008), Chakraborti (2011) και Chakraborti *et al.* (2011) συνοψίζονται τα πλεονεκτήματα εφαρμογής μη παραμετρικών διαγραμμάτων ελέγχου και παράλληλα παρέχεται μία ολοκληρωμένη ανασκόπηση των μονοδιάστατων μη παραμετρικών διαγραμμάτων που έχουν προταθεί στη βιβλιογραφία.

Είναι γεγονός ότι η πλειοψηφία των διαγραμμάτων ελέγχου (μονοδιάστατων ή πολυδιάστατων) είναι σχεδιασμένη για την παρακολούθηση της μέσης τιμής ή της διασποράς (εναλλακτικά του πίνακα διακυμάνσεων-συνδιακυμάνσεων). Κατά συνέπεια, τα διαγράμματα αυτά είναι αποτελεσματικά στην ανίχνευση μετατοπίσεων είτε της θέσης (*location*) είτε της κλίμακας (*scale*) της εντός ελέγχου κατανομής. Αναμφισβήτητα, μέχρι σήμερα έχει μελετηθεί ένας μικρός μόνο αριθμός πολυδιάστατων μη παραμετρικών διαγραμμάτων, με μνήμη ή χωρίς, για τη μέση τιμή ή τη διασπορά, ενώ ακόμη λιγότερα σχήματα έχουν προταθεί με σκοπό την ανίχνευση ταυτόχρονων μετατοπίσεων θέσης και κλίμακας. Στα πλαίσια της παρούσας διατριβής, εστιάζουμε το ενδιαφέρον μας σε πολυδιάστατα, μη παραμετρικά διαγράμματα ελέγχου.

1.5.1 Διαγράμματα ελέγχου για την παρακολούθηση μέσης τιμής ή διασποράς

Στη διεθνή βιβλιογραφία, έχουν μελετηθεί κάποια **πολυδιάστατα, μη παραμετρικά σχήματα**, κατάλληλα για την ανίχνευση πιθανών μετατοπίσεων της μέσης τιμής της υποκείμενης διεργασίας. Κατ' αρχάς, οι Hayter and Tsui (1994) εισήγαγαν μία μη παραμετρική διαδικασία για την παρακολούθηση της θέσης μιας διεργασίας, η οποία βασίζεται στην κατασκευή ταυτόχρονων διαστημάτων εμπιστοσύνης για κάθε μέσο των υπό μελέτη μεταβλητών, αποτυγχάνει, όμως, να παρακολουθεί τις συσχετίσεις που αναπτύσσονται μεταξύ των χαρακτηριστικών. Παράλληλα προτάθηκαν αντίστοιχα **διαγράμματα ελέγχου, τύπου Shewhart**. Συγκεκριμένα, οι Abu-Shawiesh and Abdullah (2001) παρουσίασαν ένα εύρωστο διάγραμμα, το οποίο κάνει χρήση των εκτιμητριών των Hodges-Lehmann (1963) και Shamos-Bickel-Lehmann (1976). Πρόσφατα, οι Boone and Chakraborti (2012) πρότειναν δύο

διαγράμματα ελέγχου, τα οποία βασίζονται σε πολυδιάστατες γενικεύσεις του προσημικού ελέγχου (*sign test*) και του βαθμολογικού προσημικού ελέγχου του *Wilcoxon* (*Wilcoxon's signed-rank test*). Το πρώτο διάγραμμα αποτελεί γενίκευση του δισδιάστατου διαγράμματος που πρότεινε ο Das (2009). Ακόμη, οι Ghute and Shirke (2012) έδωσαν μία δισδιάστατη εκδοχή ενός μη παραμετρικού διαγράμματος που βασίζεται στον βαθμολογικό προσημικό έλεγχο.

Θα πρέπει να αναφερθεί ότι έχουν μελετηθεί αντίστοιχα **διαγράμματα με μνήμη**. Για παράδειγμα, οι Karatou and Reynolds (1994, 1998) [βλ. επίσης Karatou (1996)] πρότειναν την εφαρμογή *EWMA* στην κλασσική προσημική σ.σ. και στη βαθμολογική προσημική σ.σ., με σκοπό τη δημιουργία δύο πολυδιάστατων, διαγραμμάτων για μικρά δείγματα. Τα εν λόγω διαγράμματα έχουν προταθεί για την παρακολούθηση του διανύσματος των μέσων, ωστόσο το βασικό τους μειονέκτημα είναι ότι δεν παραμένουν αναλλοίωτα σε γραμμικούς μετασχηματισμούς (*affine invariance*). Συνεπώς, μέθοδοι όπως αυτή των Κυρίων Συνιστωσών δεν μπορούν να εφαρμοσθούν, αφού γραμμικοί μετασχηματισμοί μπορεί να επηρεάσουν το τελικό αποτέλεσμα. Επιπλέον, οι Qiu and Hawkins (2001, 2003) πρότειναν ένα πολυδιάστατο μη παραμετρικό διάγραμμα *CUSUM*, το οποίο βασίζεται στη διάταξη των μετρήσεων και στη διάταξη της πληροφορίας μεταξύ των συνιστωσών των μετρήσεων και των εντός ελέγχου μέσων τιμών, καθώς και ένα αντίστοιχο διάγραμμα που βασίζεται στο βαθμό των μετρήσεων.

Πρόσφατα, μελετήθηκαν και **πολυδιάστατα, μη παραμετρικά διαγράμματα ελέγχου**, τα οποία είναι **ευαίσθητα στην ανίχνευση μετατοπίσεων της διασποράς** μιας παραγωγικής διαδικασίας. Για παράδειγμα, οι Osei-Aning *et al.* (2017) πρότειναν τέσσερα (μόνο-πλευρα) δισδιάστατα διαγράμματα ελέγχου, τα οποία χρησιμοποιούν εκτιμήσεις απλών σ.σ. που εμπεριέχουν μέτρα διασποράς. Οι Riaz and Does (2008) πρότειναν το διάγραμμα $|G|$, το οποίο βασίζεται στις γενικευμένες μέσες διαφορές *Gini* $|G|$. Επιπλέον, ο Sagir (2015) μελέτησε τις επιδόσεις του διαγράμματος $|G|$ και του κλασσικού διαγράμματος $|S|$ στην περίπτωση δισδιάστατων μη κανονικών κατανομών, ενώ οι Sagir *et al.* (2017) προχώρησαν σε μία ανάλογη μελέτη για την απόδοση των ίδιων διαγραμμάτων στη Φάση I.

1.5.2 Διαγράμματα ελέγχου για την ταυτόχρονη παρακολούθηση μέσης τιμής και διασποράς

Είναι αξιοσημείωτο ότι, κατά την περίοδο των τελευταίων μόνο ετών, οι ερευνητές έδειξαν ενδιαφέρον για την κατασκευή κανόνων κατάλληλων για ταυτόχρονη παρακολούθηση

τόσο της μέσης τιμής όσο και της διακύμανσης της υποκείμενης διεργασίας. Ανασκόπηση των συγκεκριμένων διαγραμμάτων γίνεται στα άρθρα των Cheng and Thaga (2006) και McCracken and Chakraborti (2013). Παράλληλα, οι Reynolds and Stoumbos (2006) και Ou *et al.* (2012) σύγκριναν μερικά από αυτά τα διαγράμματα μεταξύ τους και μελέτησαν την απόδοσή τους. Τα περισσότερα σχήματα που περιλαμβάνονται στα προαναφερθέντα άρθρα είναι μονοδιάστατα και βασίζονται κατά ένα μεγάλο ποσοστό στην υπόθεση της κανονικότητας.

Αρκετές μονοδιάστατες διαδικασίες έχουν κατάλληλα επεκταθεί για την παρακολούθηση πολλαπλών ποιοτικών χαρακτηριστικών, τα οποία προέρχονται από πολυδιάστατη κανονική κατανομή. Για παράδειγμα, οι Cheng and Mao (2008) πρότειναν το διάγραμμα ελέγχου *Semi-Circe (SC)*, το οποίο μπορεί να θεωρηθεί γενίκευση των μονοδιάστατων διαγραμμάτων ελέγχου *SC* των Chao and Cheng (1996, 2008). Ο Khoo (2004) και οι Thaga and Gabaitiri (2006) δημιούργησαν, αντίστοιχα, τη δισδιάστατη και πολυδιάστατη εκδοχή του (μονοδιάστατου) διαγράμματος *MAQ* των Chen and Cheng (1998). Ενώ, ο Xie (1991) και οι Cheng and Thaga (2005) παρουσίασαν ένα πολυδιάστατο διάγραμμα *EWMA* και ένα διάγραμμα *CUSUM*. Περισσότερα πολυδιάστατα διαγράμματα ελέγχου *EWMA*, έχουν προταθεί από τους Jeong and Cho (2012) και Chen *et al.* (2005).

Οι Zhao and Mei (2014) πρότειναν ένα δισδιάστατο διάγραμμα για ταυτόχρονη παρακολούθηση, το οποίο βασίζεται σε μία συνδυασμένη σ.σ. που φέρει πληροφορία και για τη μέση τιμή και τη διασπορά. Αντίθετα, οι Spiring and Cheng (1998) εισήγαγαν ένα ταυτόχρονο διάγραμμα, στο οποίο τόσο η μέση τιμή όσο και η τυπική απόκλιση της διεργασίας σχεδιάζονται στο ίδιο γράφημα, όπου, όμως γίνεται χρήση διαφορετικών στατιστικών συναρτήσεων. Σε ορισμένες περιπτώσεις, έχουν κατασκευαστεί σχήματα παρακολούθησης μέσω δύο διαγραμμάτων (*two-chart monitoring schemes*), τα οποία αποτελούνται από χωριστά διαγράμματα, ένα για τη μέση τιμή και ένα για τη διακύμανση με όρια ελέγχου που είναι κατάλληλα προσαρμοσμένα στο συνολικό ποσοστό λανθασμένου συναγερμού (*FAR*) της διεργασίας. Τέτοιου είδους διαγράμματα έχουν κατασκευασθεί από τους Levinson *et al.* (2002), Reynolds and Stoumbos (2008) και Maboudou-Tchao and Hawkins (2011).

Θα πρέπει να αναφέρουμε ότι αναπτύχθηκαν, επίσης, μερικές προσεγγίσεις με σκοπό την κατασκευή διαγραμμάτων που μπορούν να εφαρμοσθούν σε ποικίλες περιπτώσεις, όπως για παράδειγμα τα διαγράμματα ελέγχου λόγου πιθανοφανειών (*likelihood-ratio based control charts*) των Zhang *et al.* (2010), Wang *et al.* (2014) και τα διαγράμματα των Sullivan and Woodall (2000) και Zamba and Hawkins (2009) που είναι ικανά να ανιχνεύσουν το

σημείο (χρονική στιγμή) που παρατηρείται η μετατόπιση (*change point detection models*). Εργαζόμενοι σε ένα εντελώς διαφορετικό πλαίσιο, οι Reynolds and Cho (2006) ανέπτυξαν κάποια προσαρμοστικά διαγράμματα, στα οποία ενσωμάτωσαν μεταβλητά διαστήματα δειγματοληψίας (*variable sampling intervals*) και ακολουθιακή δειγματοληψία.

Η ταυτόχρονη παρακολούθηση μέσης τιμής και διασποράς μιας διεργασίας με τη χρήση μη παραμετρικών διαγραμμάτων ελέγχου φαίνεται να είναι μία σχετικά νέα ερευνητική περιοχή και, ως εκ τούτου, εξακολουθούν να υπάρχουν διάφορα προβλήματα που σχετίζονται με αυτή, τα οποία θα χρειαστεί να αντιμετωπιστούν στο μέλλον. Μέχρι σήμερα, έχουν εμφανιστεί στη βιβλιογραφία μερικά **μονοδιάστατα διαγράμματα ελέγχου τύπου Shewhart**. Ένα από τα πρώτα διαγράμματα που προτάθηκαν είναι αυτό των Balakrishnan *et al.* (2010), οι οποίοι βελτίωσαν το κλασικό διάγραμμα για τη διάμεσο των Janacek and Meikle (1997) και την εκτεταμένη εκδοχή του που μελετήθηκε από τους Chakraborti *et al.* (2004). Οι Balakrishnan *et al.* (2010) ενσωμάτωσαν στο τελευταίο διάγραμμα έναν πρόσθετο κανόνα, ώστε να λαμβάνεται υπ' όψιν όχι μόνο η θέση μιας διατεταγμένης παρατήρησης (του δείγματος ελέγχου), αλλά και ο αριθμός των παρατηρήσεων από το δείγμα ελέγχου που βρίσκονται μεταξύ των ορίων. Αξίζει να σημειωθεί ότι πρόσφατα ο Triantafyllou (2017) πρότεινε μία παραλλαγή του προαναφερθέντος διαγράμματος ελέγχου. Στο νέο σχήμα, ο κανόνας ελέγχου βασίζεται στη θέση δύο διατεταγμένων παρατηρήσεων και επιπλέον στο πλήθος των παρατηρήσεων που ανήκουν στην περιοχή ελέγχου. Οι Mukherjee and Chakraborti (2012) και Chowdhury *et al.* (2014) εκμεταλλεύτηκαν τη σ.σ. *Lepage* (1971), η οποία συνδυάζει τη βαθμολογική σ.σ. ελέγχου του *Wilcoxon* (*Wilcoxon rank-sum statistic*) για τη θέση με τη σ.σ. των *Ansari-Bradley* για την κλίμακα, καθώς και τη συνάρτηση του *Cucconi* (1968) για την κατασκευή των διαγραμμάτων *Shewhart-Lepage* και *Shewhart-Cucconi*, αντίστοιχα.

Επιπρόσθετα, στη διεθνή βιβλιογραφία είναι διαθέσιμα και μερικά **(μονοδιάστατα) μη παραμετρικά διαγράμματα ελέγχου με μνήμη**. Για παράδειγμα, οι Reynolds and Stoumbos (2010) θεώρησαν δύο συνδυασμούς *CUSUM*, οι οποίοι είναι αρκετά ανθεκτικοί σε εφαρμογές με μη κανονικά δεδομένα και είναι αποτελεσματικοί στην ανίχνευση μετατοπίσεων της θέσης, της κλίμακας και του σχήματος (*shape*), ενώ οι Zou and Tsung (2010) πρότειναν ένα *EWMA* διάγραμμα που βασίζεται σε έναν έλεγχο καλής προσαρμογής. Επιπλέον, οι Yang and Arnold (2010, 2015) εξέτασαν δύο απλές προσεγγίσεις συνδυάζοντας δύο γραφήματα ελέγχου *EWMA*, ένα για τη μέση τιμή και ένα για τη διακύμανση με σκοπό την παρακολούθηση τόσο της μέσης τιμής όσο και της μεταβλητότητας της υποκείμενης

διαδικασίας. Στην πρώτη προσέγγιση, χρησιμοποίησαν ασύμμετρα όρια ελέγχου για να βελτιώσουν την απόδοση του διαγράμματος διακύμανσης της δεύτερης προσέγγισης. Δύο πιο πρόσφατα παραδείγματα αποτελούν τα διαγράμματα *CUSUM-Lepage* και *CUSUM-Ciccconi* Φάσης II, που εισήχθησαν από τους Chowdhury *et al.* (2015) και Mukherjee and Marozzi (2017), αντίστοιχα. Βασιζόμενοι σε ένα διαφορετικό πλαίσιο, οι Ross *et al.* (2011) παρουσίασαν ένα διάγραμμα ελέγχου για ταυτόχρονη παρακολούθηση μέσης τιμής και μεταβλητότητας για δεδομένα ροής (*data streams*).

Από όσα γνωρίζουμε, θα μπορούσαμε να ισχυριστούμε ότι η βιβλιογραφία σχετικά με τα **πολυδιάστατα μη παραμετρικά διαγράμματα ελέγχου**, τα οποία στοχεύουν στην **ανίχνευση αλλαγών τόσο στη μέση τιμή όσο και στη διακύμανση**, είναι περιορισμένη και εστιάζεται κυρίως σε **μονόπλευρα διαγράμματα ελέγχου με μνήμη**. Συγκεκριμένα, η Liu (1995) πρότεινε τρία μονόπλευρα διαγράμματα που βασίζονται στην έννοια του βάθους δεδομένων (*data depth*): τα διαγράμματα r , Q και S , τα οποία μπορούν να θεωρηθούν πολυδιάστατες γενικεύσεις των (μονοδιάστατων) διαγραμμάτων X , \bar{X} και *CUSUM*, αντίστοιχα. Έχει, επίσης, ενδιαφέρον να σημειώσουμε ότι τα διαγράμματα r και Q , ανήκουν στις κατηγορίες των διαγραμμάτων ελέγχου τύπου *Shewhart* για μεμονωμένες και ομαδοποιημένες παρατηρήσεις, αντίστοιχα. Βασιζόμενοι στην προαναφερθείσα μεθοδολογία, οι Hamurkaroglu *et al.* (2004) πρότειναν αντίστοιχα διαγράμματα χρησιμοποιώντας την απόσταση του *Mahalanobis*. Οι Liu *et al.* (2004) πρότειναν ένα διάγραμμα ελέγχου κινούμενου μέσου για το βάθος των δεδομένων, που ονομάζεται διάγραμμα *DDMA* και έχει ένα (κάτω) όριο ελέγχου. Πρόσφατα, οι Li *et al.* (2013) εισήγαγαν δύο διαδικασίες *CUSUM*, οι οποίες αξιοποιούν το χωρικό πρόσημο (*spatial sign*) και το βάθος των δεδομένων, για να κρίνουν εάν η διεργασία είναι εντός ή εκτός ελέγχου ορίζοντας ένα μόνο (άνω) όριο ελέγχου. Τέλος, οι Zou and Tsung (2011) παρουσίασαν ένα πολυδιάστατο μη παραμετρικό διάγραμμα *EWMA* με κάτω όριο ελέγχου, προτείνοντας τη χρήση της πολυδιάστατης γραμμικής ισοσταθμισμένης διαμέσου (*Affine Equivariant Multivariate median*) των Hettmansperger and Randles (2002).

1.6 Σκοπός και Δομή Διδακτορικής Διατριβής

Όπως γίνεται φανερό και από την Ενότητα 1.5, ελάχιστες πολυδιάστατες, μη παραμετρικές διαδικασίες έχουν προταθεί στη διεθνή βιβλιογραφία για την παρακολούθηση της συμπεριφοράς μίας παραγωγικής διεργασίας. Τα περισσότερα σχήματα που έχουν μελετηθεί εστιάζουν στην ανίχνευση πιθανών μετατοπίσεων είτε της μέσης τιμής είτε της διασποράς

της υποκείμενης διεργασίας, ενώ μόνο ένα διάγραμμα τύπου *Shewhart* (το διάγραμμα *Q* της Liu (1995)) φαίνεται να έχει προταθεί μέχρι σήμερα για την ταυτόχρονη ανίχνευση μετατοπίσεων της μέσης τιμής και της μεταβλητότητας. Επιπλέον, στα διαγράμματα ελέγχου με μνήμη, η σ.σ. εξαρτάται από την πληροφορία που λαμβάνεται από προγενέστερα δείγματα και ως εκ τούτου η κατασκευή τους είναι πολύ πιο περίπλοκη συγκριτικά με αυτή ενός διαγράμματος ελέγχου τύπου *Shewhart*.

Δεδομένης της έλλειψης που παρατηρείται στη σχετική βιβλιογραφία και της πολυπλοκότητας των παραπάνω διαγραμμάτων, σκοπός της παρούσας διδακτορικής εργασίας είναι η παρουσίαση νέων σχημάτων με σχετικά πιο εύκολη κατασκευή. Στα πλαίσια της διατριβής αυτής προτείνονται τέσσερα νέα δισδιάστατα, ημιπαραμετρικά διαγράμματα ελέγχου (τύπου *Shewhart*), δύο για τη μέση τιμή και δύο για την παράλληλη παρακολούθηση της μέσης τιμής και της διασποράς της διεργασίας, τα οποία όχι μόνο είναι απλά στην υλοποίηση, αλλά μπορούν εύκολα να γενικευθούν και σε περισσότερες διαστάσεις. Η συμβολή της παρούσας διατριβής έγκειται στο ότι αξιοποιεί τη Θεωρία των Διατεταγμένων Παρατηρήσεων σε συνδυασμό με τη Θεωρία των Συμμεταβλητών και τη Θεωρία των Συνδέσμων, με σκοπό να προτείνει νέες από κοινού δισδιάστατες κατανομές, οι οποίες μπορούν να εφαρμοσθούν στο Στατιστικό Έλεγχο Διεργασιών.

Οι προαναφερθείσες κατανομές χρησιμοποιούνται για τον προσδιορισμό των χαρακτηριστικών (χαρακτηριστική συνάρτηση λειτουργίας, ποσοστό συναγερμού και ποσοστό λανθασμένου συναγερμού) των προτεινόμενων διαγραμμάτων. Θα πρέπει, επίσης, να υπογραμμισθεί ότι στα πλαίσια της παρούσας διατριβής έχει μελετηθεί η απόδοση των τεσσάρων νέων κλάσεων διαγραμμάτων που προτάθηκαν βάσει τόσο των εντός όσο και των εκτός ελέγχου τιμών του ποσοστού συναγερμού και του μέσου μήκους ροής. Πιο συγκεκριμένα, έγιναν αριθμητικές συγκρίσεις μεταξύ των προαναφερθέντων διαγραμμάτων και άλλων αντίστοιχων διαγραμμάτων ελέγχου που είναι διαθέσιμα στη βιβλιογραφία, τα αποτελέσματα των οποίων έδειξαν ότι τα νέα σχήματα συμπεριφέρονται συνήθως καλύτερα. Κατά συνέπεια, μπορούν να αποτελέσουν χρήσιμες εναλλακτικές λύσεις, καθώς δεν απαιτούν καμία υπόθεση σχετικά με την κατανομή της διεργασίας και έχουν την πρόσθετη ευελιξία να χρησιμοποιούνται σε περιπτώσεις μη συμμετρικών κατανομών με τον προσδιορισμό κατάλληλων ορίων ελέγχου. Ακόμη, αξίζει να σημειωθεί ότι η πρακτική εφαρμογή των διαγραμμάτων παρουσιάζεται μέσω παραδειγμάτων στα οποία γίνεται χρήση πραγματικών δεδομένων από το χώρο της οινοποιίας και της βιοστατιστικής.

Επιλογικά, περιγράφεται η δομή της παρούσας διατριβής. Αρχικά, στο Κεφάλαιο 1,

παρουσιάστηκε μια σύντομη ιστορική αναδρομή στον Έλεγχο Ποιότητας και ορίστηκαν κάποιες βασικές του έννοιες. Παράλληλα, έγινε ανασκόπηση της σχετικής βιβλιογραφίας, όπου δόθηκε έμφαση στα πολυδιάστατα, μη παραμετρικά διαγράμματα ελέγχου. Ακολούθως, παρουσιάζεται το περιεχόμενο των επόμενων κεφαλαίων. Συγκεκριμένα, στο Κεφάλαιο 2, αναφέρονται τα βασικά σημεία της Θεωρίας των Διατεταγμένων Παρατηρήσεων, των Συμμεταβλητών και των Συνδέσμων. Στο Κεφάλαιο 3 παρουσιάζονται νέα αποτελέσματα που συνδέονται με τις από κοινού κατανομές δισδιάστατων διατεταγμένων παρατηρήσεων και απαριθμητριών συναρτήσεων αυτών. Στη συνέχεια, στα Κεφάλαια 4–5, προτείνονται δύο νέα μη παραμετρικά διαγράμματα ελέγχου για την ανίχνευση πιθανών μετατοπίσεων της μέσης τιμής της παραγωγικής διαδικασίας που μελετάται, ενώ στα Κεφάλαια 6–7, παρουσιάζονται δύο ακόμη μη παραμετρικά διαγράμματα ελέγχου για την ταυτόχρονη παρακολούθηση της μέσης τιμής και της μεταβλητότητας της εν λόγω διεργασίας. Τέλος, στο Κεφάλαιο 8, συνοψίζονται τα κυριότερα πλεονεκτήματα χρήσης και εφαρμογής των προτεινόμενων διαγραμμάτων ελέγχου έναντι άλλων αντίστοιχων ανταγωνιστικών σχημάτων που έχουν προταθεί στη διεθνή βιβλιογραφία, ενώ παράλληλα περιγράφονται θέματα για περαιτέρω μελέτη και μελλοντική έρευνα.

ΚΕΦΑΛΑΙΟ 2

Θεωρία Διατεταγμένων Παρατηρήσεων, Συμμεταβλητών και Συνδέσμων

2.1 Εισαγωγή

Σκοπός του παρόντος κεφαλαίου είναι να παρουσιαστούν νέα αποτελέσματα σχετικά με τις από κοινού κατανομές παρατηρήσεων που προέρχονται από δύο διατεταγμένα δείγματα, τα οποία είναι συσχετισμένα μεταξύ τους. Συγκεκριμένα, στην Ενότητα 2.2, ορίζεται αρχικά το διατεταγμένο δείγμα και έπειτα γίνεται αναφορά στις περιοχές όπου έχουν πρακτική εφαρμογή οι κατανομές και οι ιδιότητες των διατεταγμένων παρατηρήσεων. Στη συνέχεια, παρουσιάζονται οι ήδη γνωστές κατανομές παρατηρήσεων που ανήκουν στο ίδιο διατεταγμένο δείγμα (Ενότητα 2.3). Επιπλέον, στις Ενότητες 2.4 και 2.5 εισάγονται όλες οι βασικές έννοιες, συμβολισμοί και ιδιότητες της Θεωρίας των Συμμεταβλητών των Διατεταγμένων Παρατηρήσεων (*Theory of Concomitants of Order Statistics*) και της Θεωρίας των Συνδέσμων (*Theory of Copulas*). Ουσιαστικά, στο κεφάλαιο αυτό παρουσιάζουμε όλα τα απαραίτητα εργαλεία, τα οποία θα αξιοποιήσουμε στο επόμενο κεφάλαιο για να βρούμε νέες από κοινού δισδιάστατες κατανομές που σχετίζονται με δύο ή περισσότερες διατεταγμένες παρατηρήσεις, οι οποίες προέρχονται από συσχετισμένα δείγματα.

2.2 Διατεταγμένα δείγματα και εφαρμογές

Για να εισάγουμε την έννοια του διατεταγμένου δείγματος, θεωρούμε αρχικά ότι έχουμε στη διάθεσή μας n τ.μ., έστω X_1, X_2, \dots, X_n , τις οποίες στη συνέχεια διατάσσουμε σε αύξουσα σειρά. Το δείγμα που προκύπτει ονομάζεται **διατεταγμένο** (*ordered sample*) και έχει την ακόλουθη μορφή

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n},$$

όπου με $X_{1:n}$ και $X_{n:n}$ συμβολίζονται αντίστοιχα η μικρότερη και μεγαλύτερη από τις n παρατηρήσεις, ενώ γενικότερα ο συμβολισμός $X_{r:n}$ αντιστοιχεί στην r -οστή **διατεταγμένη παρατήρηση** ενός δείγματος μεγέθους n . Συνήθως, οι αρχικές παρατηρήσεις X_1, X_2, \dots, X_n θεωρούμε ότι προέρχονται από ένα τυχαίο δείγμα (δηλαδή είναι ανεξάρτητες και ισόνομες). Ωστόσο, αποτελέσματα για το διατεταγμένο δείγμα έχουν αναπτυχθεί και στην περίπτωση που έχουμε εξάρτηση ή οι παρατηρήσεις προέρχονται από διαφορετικές κατανομές.

Η Θεωρία Διατεταγμένων Παρατηρήσεων (*Theory of Order Statistics*) αποτελεί αναπόσπαστο κομμάτι της Στατιστικής Θεωρίας και Πρακτικής (*Statistical Theory and Practice*). Τις τελευταίες δεκαετίες παρατηρείται έντονη ερευνητική δραστηριότητα σχετικά με τη μελέτη των διατεταγμένων παρατηρήσεων, καθώς βρίσκουν εφαρμογή σε ποικίλα προβλήματα πρακτικού ενδιαφέροντος. Ενδεικτικά αναφέρουμε τα βιβλία των Arnold and Balakrishnan (1989), Reiss (1989), Balakrishnan and Cohen (1991), Arnold *et al.* (1992), David and Nagaraja (2003), Balakrishnan *et al.* (2006) και τα εγχειρίδια των Balakrishnan and Rao (1998a,b), τα οποία καλύπτουν επαρκώς τις βασικές έννοιες των διατεταγμένων παρατηρήσεων τόσο σε θεωρητικό όσο και σε πρακτικό επίπεδο.

Οι συναρτήσεις των διατεταγμένων παρατηρήσεων που μελετώνται κυρίως στη βιβλιογραφία είναι η διάμεσος (*median*), συμβ.

$$\delta = \begin{cases} X_{(n+1)/2:n}, & \text{εάν το μέγεθος του δείγματος } n \text{ είναι περιττός,} \\ (X_{n/2:n} + X_{n/2+1:n})/2, & \text{εάν το μέγεθος του δείγματος } n \text{ είναι άρτιος,} \end{cases}$$

οι ακραίες παρατηρήσεις (*extreme values*) $X_{1:n}$ και $X_{n:n}$, το εύρος (*range*) $W = X_{n:n} - X_{1:n}$ και η ακραία απόκλιση (*extreme deviate*) από το δειγματικό μέσο (*sample mean*)¹. Επιπλέον, για ένα δείγμα που προέρχεται από την Κανονική κατανομή $N(\mu, \sigma^2)$ χρησιμοποιείται συνήθως το τυποποιημένο εύρος (*studentized range*) W/S_n , όπου S_n είναι η ρίζα μίας αμερόληπτης εκτιμήτριας του σ^2 . Μερικές ενδεικτικές περιοχές, όπου η εφαρμογή της Θεωρίας των Διατεταγμένων Παρατηρήσεων και η χρήση των παραπάνω συναρτήσεων μπορούν να διαδραματίσουν σημαντικό ρόλο, είναι οι ακόλουθες:

- **Προσομοίωση (Simulation)**. Μία συνήθης μεθοδολογία που ακολουθείται στην προσομοίωση για την παραγωγή ψευδο-τυχαίων αριθμών από μία αθροιστική συνάρτηση κατανομής (α.σ.κ.) $F_X(x)$, χρησιμοποιώντας έναν τυχαίο αριθμό $U \sim \mathcal{U}(0, 1)$, είναι η μέθοδος του αντίστροφου μετασχηματισμού. Για την ακρίβεια, στην περίπτωση

¹ $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_{i:n} = \frac{1}{n} \sum_{i=1}^n X_i$

όπου η $F_X(x)$ είναι συνεχής, ο μετασχηματισμός $U = F_X(X)$ παράγει την Ομοιόμορφη κατανομή στο $(0, 1)$ και ως αποτέλεσμα προκύπτει η παρακάτω ισότητα κατανομών

$$F_X(X_{i:n}) \stackrel{d}{=} U_{i:n}, \quad i = 1, 2, \dots, n.$$

Επιπλέον, εάν για μία τ.μ. X με οποιαδήποτε α.σ.κ. $F_X(x)$ θεωρήσουμε την αντίστροφη συνάρτησή της $F_X^{-1}(\cdot)$, η οποία ορίζεται ως εξής

$$F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}, \quad 0 < u < 1,$$

οδηγούμαστε στις σχέσεις

$$F_X(F_X^{-1}(u)) \geq u \quad \text{και} \quad F_X^{-1}(F_X(x)) \leq x.$$

Προφανώς, ισχύει η ισοδυναμία (βλ. Serfling (1980))

$$u \leq F_X(x) \Leftrightarrow F_X^{-1}(u) \leq x.$$

Επομένως, για $0 < F_X(x) < 1$ έχουμε

$$F_X(x) = P(X \leq x) = P(F_X(X) \leq F_X(x)) = P(U \leq F_X(x)) = P(F_X^{-1}(U) \leq x).$$

Είναι πλέον ξεκάθαρο ότι ισχύουν τα ακόλουθα (βλ. Scheffé and Tukey (1945))

$$F_X^{-1}(U_i) \stackrel{d}{=} X_i \quad \text{και} \quad F_X^{-1}(U_{i:n}) \stackrel{d}{=} X_{i:n}, \quad i = 1, 2, \dots, n.$$

Μέσω της τελευταίας σχέσης και της α.σ.κ. της $U_{i:n}$ μπορούμε να οδηγηθούμε άμεσα στην α.σ.κ. της διατεταγμένης παρατήρησης $X_{i:n}$ ($1 \leq i \leq n$). Περισσότερες πληροφορίες σχετικά με τη δημιουργία αλγορίθμων για την παραγωγή ψευδο-τυχαίων αριθμών -είτε η $F_X(x)$ είναι διακριτή είτε συνεχής- παρέχονται στα κείμενα των Ross (2013) [βλ. επίσης Μπούτσικας (2004)].

- **Κατασκευή Διαστημάτων Εμπιστοσύνης.** Με βάση τη μεθοδολογία που περιγράφηκε προηγουμένως, μπορούν να κατασκευασθούν διαστήματα εμπιστοσύνης για το ποσοστημόριο $F_X^{-1}(p)$ ενός πληθυσμού, τα οποία έχουν ως όρια συγκεκριμένες διατεταγμένες παρατηρήσεις. Το βασικό χαρακτηριστικό των παραγόμενων διαστημάτων εμπιστοσύνης είναι ότι δεν εξαρτώνται από την $F_X(x)$ και μπορούν να προσδιορισθούν μέσω των πινάκων της διωνυμικής κατανομής. Η διαδικασία αυτή μπορεί να επεκταθεί για την κατασκευή διαστημάτων εμπιστοσύνης για ποσοστημόρια (βλ. Thompson (1936)) ή συγκεκριμένα για τη διάμεσο (βλ. MacKinnon (1964), Hettmansperger and Sheather (1986)) και για διαφορές ποσοστημορίων, όπως το ενδοτεταρτημοριακό εύρος (*interquartile/quasi range*) που εξετάζεται στην εργασία του Chu (1957). Επιπλέον, με τη βοήθεια των διατεταγμένων παρατηρήσεων

μπορούν να κατασκευασθούν μη παραμετρικά διαστήματα ανοχής (*tolerance intervals*) και πρόβλεψης (*prediction intervals*). Και οι τρεις αυτές κατηγορίες διαστημάτων αναλύονται από τους Hahn and Meeker (1991).

- **Έλεγχοι Καλής Προσαρμογής (*Goodness of Fit Tests*)**. Οι πιο συνηθισμένοι έλεγχοι καλής προσαρμογής κάνουν χρήση διαφορών μεταξύ διαδοχικών διατεταγμένων παρατηρήσεων ($X_{i:n} - X_{i-1:n}$) ή συχνά εστιάζουν σε αποκλίσεις μεταξύ εμπειρικών ποσοστημορίων (*empirical quantiles*). Έλεγχοι όπως αυτοί των Kolmogorov (1933)-Smirnov (1939) και Anderson-Darling (1952) κάνουν χρήση της εμπειρικής α.σ.κ., ενώ αντίστοιχοι έλεγχοι έχουν προταθεί για συγκεκριμένες κατανομές, όπως η Εκθετική (βλ. Epstein (1960a,b), Stephens (1986a)), η Κανονική (βλ. Shapiro and Wilk (1965), D'Agostino (1986b)) και η Ομοιόμορφη (βλ. Stephens (1986b)). Επιπρόσθετα, έχουν κατασκευασθεί γραφικοί έλεγχοι καλής προσαρμογής με χαρακτηριστικότερο παράδειγμα το *Q-Q Plot* των Wilk and Gnanadesikan (1968), ενώ παραλλαγές του για Κανονικές και Λογαριθμοκανονικές (*lognormal*) κατανομές έχουν συζητηθεί στο κείμενο του D'Agostino (1986a).
- **Ανίχνευση Έκτροπων Παρατηρήσεων (*Outlier Detection*)**. Το γεγονός ότι μία παρατήρηση μπορεί να λαμβάνει ασυνήθιστα μικρή ή μεγάλη τιμή σε σχέση με αυτές των υπόλοιπων παρατηρήσεων του δείγματος αποτελεί ένδειξη ότι πρόκειται για έκτροπη παρατήρηση. Αυτό εύκολα διαπιστώνεται εάν, έχοντας ένα συγκεκριμένο δείγμα διατεταγμένων παρατηρήσεων, υπολογισθεί η πιθανότητα η μεγαλύτερη διατεταγμένη παρατήρηση να λαμβάνει τόσο ακραίες τιμές όσο αυτές των «ύποπτων» παρατηρήσεων. Παραδείγματος χάριν, μεγάλες τιμές της στατιστικής συνάρτησης $T = (X_{n:n} - \bar{X})/S$, όπου $S^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n - 1)$ είναι η δειγματική διακύμανση, υποδεικνύουν ύπαρξη έκτροπων παρατηρήσεων. Στη βιβλιογραφία έχουν κατασκευαστεί έλεγχοι για τη σωστή ανίχνευση τέτοιου είδους παρατηρήσεων βάσει της συνάρτησης T (βλ. Pearson and Sekar (1936), Grubbs (1969), Barnett and Lewis (1994)), ενώ διαφορετικές προσεγγίσεις δίνονται από τους Murphy (1951) και Hawkins (1980).
- **Ανθεκτικές Εκτιμήτριες Θέσης-Κλίμακας (*Robust Estimators for Location-Scale*)**. Χωρίς αμφιβολία, εκτιμήσεις που βασίζονται στη διάμεσο (*median*) ή γραμμικές εκτιμήσεις της μέσης τιμής, όπως για παράδειγμα ο μέσος όρος των κεντρικών παρατηρήσεων (*average of central order statistics*), ο συμμετρικός περικομμένος μέσος

(*symmetric trimmed mean*)¹, ο μέσος του Winsor (*Winsorized mean*)², ο γραμμικά σταθμισμένος μέσος (*linearly weighted mean*)³ n ανεξάρτητων μετρήσεων, καθώς και οι εκτιμήτριες L & M (*L & M estimates*) συνιστούν ανθεκτικές εκτιμήτριες. Αυτό συμβαίνει, διότι δίνουν λιγότερο βάρος στις ακραίες παρατηρήσεις (*extreme values*) του δείγματος και ως αποτέλεσμα προτιμώνται έναντι της συνήθους επιλογής του απλού δειγματικού μέσου. Πιο λεπτομερής επισκόπηση σχετικά με την εύρεση ανθεκτικών εκτιμητριών κλίμακας ή/και θέσης (δεδομένης της ύπαρξης ακραίων τιμών στο δείγμα), ο ενδιαφερόμενος αναγνώστης μπορεί να συμβουλευτεί τα κείμενα των Lloyd (1952), Andrews *et al.* (1972), Hogg (1974), David and Shu (1978), Shu (1978), David (1979), Barnett and Lewis (1994) και Huber and Ronchetti (2009).

- **Θεωρία Ακραίων Τιμών (*Extreme Value Theory*)**. Η στατιστική μελέτη των ακραίων τιμών (βλ. Gumbel (2004)), συμβάλλει στη μοντελοποίηση ακραίων φυσικών φαινομένων, όπως για παράδειγμα ανεμοστρόβιλων, σεισμών, ξηρασιών, πλημμυρών, πυρκαϊών (βλ. Buishand (1989), Alvarado *et al.* (1998), Katz *et al.* (2002), Reiss and Thomas (2007), Katz (2010)), ενώ εξίσου σημαντική κρίνεται η συνεισφορά στα Χρηματοοικονομικά (βλ. Embrechts *et al.* (1997), Krishna (2010), Novak (2011)) και στην αντοχή υλικών (*strength of materials*) ή γενικότερα στη Μηχανική (βλ. Castillo (1988), Castillo *et al.* (2005)).
- **Διαδικασίες Κατάταξης και Επιλογής (*Ranking and Selecting Procedures*)**. Ακολουθώντας παρόμοια λογική με αυτή που περιγράφηκε στην προηγούμενη παράγραφο, με τη διαφορά ότι πλέον οι έκτροπες παρατηρήσεις εκλαμβάνονται ως ακρότατες τιμές (*record values*), οι οποίες δε θεωρούνται εσφαλμένες μετρήσεις και συνεπώς δεν αποκλείονται από το δείγμα, λύση μπορεί να δοθεί σε προβλήματα που συνδέονται με το χώρο της βιοστατιστικής (ως μέτρο αποτελεσματικότητας μίας θεραπείας) ή

$$^1 T_n(r) = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} X_{i:n}, \quad 0 \leq r \leq (n-1)/2$$

$$^2 W_n(r) = \begin{cases} \frac{1}{n} \left(\sum_{i=r+2}^{n-r-1} X_{i:n} + (r+1)(X_{r+1:n} + X_{n-r:n}) \right), & 0 \leq r < (n-1)/2, \\ X_{(n+1)/2:n}, & \text{εάν το } n \text{ είναι περιττός και } r = (n-1)/2. \end{cases}$$

$$^3 L_n(r) = \begin{cases} \sum_{i=1}^{n/2-r} \frac{(2i-1)(X_{r+i:n} + X_{n-r+1-i:n})}{2(n/2-r)^2}, & \text{εάν το } n \text{ είναι άρτιος,} \\ \sum_{i=1}^{((n-1)/2-r)/2} \frac{(2i-1)(X_{r+i:n} + X_{n-r+1-i:n}) + (n-2r)X_{(n+1)/2:n}}{((n-1)/2-r)^2 + ((n+1)/2-r)^2}, & \text{εάν το } n \text{ είναι περιττός.} \end{cases}$$

σε διαδικασίες κατάταξης και επιλογής (π.χ. ολυμπιακά ρεκόρ, επίδοση σε διαγωνισμούς ή εξετάσεις). Σε αυτές τις περιπτώσεις, από ένα σύνολο n παρατηρήσεων, οι παρατηρήσεις με τους k μεγαλύτερους βαθμούς συνδέονται με τα βέλτιστα δυνατά αποτελέσματα ή έστω επιλέγεται η μεγαλύτερη διατεταγμένη παρατήρηση, η οποία αντιστοιχεί στην καλύτερη επίδοση (*high score*). Η χρήση της Θεωρίας των Διατεταγμένων Παρατηρήσεων στην εφαρμογή των μεθόδων αυτών απασχόλησε πολλούς ερευνητές, μεταξύ άλλων και τους Gupta (1984), Nagaraja (1988). Συγγενής μεθοδολογία θεωρείται η δειγματοληψία διατεταγμένων δειγμάτων (*ranked-set sampling*), η οποία εισήχθη από τον McIntyre (1952) και εφαρμόζεται σε περιπτώσεις όπου είναι δύσκολο ή δεν είναι οικονομικά συμφέρον να ληφθούν μετρήσεις του υπό μελέτη χαρακτηριστικού. Όταν, μάλιστα, σε τέτοιου είδους διαδικασίες εμπλέκονται περισσότερες από μία συσχετισμένες μεταβλητές είναι αναγκαία η χρήση της Θεωρίας των Συμμεταβλητών (βλ. Yeo and David (1984), Nagaraja and David (1994), Ke Wang (2008), Ke Wang and Nagaraja (2010a,b)).

- **Λογοκριμένη Δειγματοληψία (*Censored Sampling*)**. Σε ένα πείραμα ελέγχου χρόνου ζωής (*life testing experiment*) τίθενται υπό έλεγχο n παρατηρήσεις και η μελέτη σταματά εάν ένα προκαθορισμένο πλήθος r μετρήσεων αποτύχει. Ως αποτέλεσμα, οι r πρώτες διατεταγμένες παρατηρήσεις ($X_{1:n} \leq \dots \leq X_{r:n}$) συνιστούν τους παρατηρηθέντες χρόνους ζωής, ενώ οι υπόλοιποι χρόνοι ζωής ($X_{r+1:n} \leq \dots \leq X_{n:n}$) δε μελετώνται καθόλου. Τέτοιου είδους μελέτες συμβάλλουν ουσιαστικά και στην Ανάλυση Επιβίωσης (*Survival Analysis*), όπου αντικείμενο μελέτης είναι ο χρόνος ζωής ενός ανθρώπου ή ο χρόνος από την έναρξη μιας θεραπείας (στην οποία υποβλήθηκε ο ασθενής) μέχρι τη στιγμή απόκρισης σε αυτή (βλ. Αντζουλάκος (2009)). Ανάλογα με την πληροφορία που υπάρχει για τους παρατηρούμενους χρόνους ζωής, χρησιμοποιούνται διαφορετικά είδη λογοκρισίας, λόγου χάρη Τύπου I (*Type I:Time*), Τύπου II (*Type II:Failure*), Προοδευτική (*Progressive*) και Τυχαία (*Random*). Εκτιμήσεις παραμέτρων θέσης και κλίμακας για λογοκριμένα δεδομένα υπολογίζονται στις εργασίες των Gupta (1952), Sarhan and Greenberg (1956, 1959), Cohen (1959, 1961, 1963) και το βιβλίο του Cohen (1991).
- **Θεωρία Αξιοπιστίας (*Reliability Theory*)**. Στη Θεωρία Αξιοπιστίας, ο χρόνος ζωής ενός συστήματος r -από-τα- $n : G$, το οποίο λειτουργεί εάν και μόνο εάν τουλάχιστον r από τις n μονάδες του λειτουργούν, είναι ουσιαστικά η διατεταγμένη παρατήρηση

$X_{n-r+1:n}$. Στην ειδική περίπτωση όπου $r = 1$ ($r = n$), το σύστημα ονομάζεται παράλληλο (σειριακό) και λειτουργεί όταν τουλάχιστον μία από τις μονάδες (όλες οι μονάδες) του λειτουργούν. Η μελέτη χρόνων ζωής τέτοιου είδους απλών συστημάτων ή συστημάτων με πιο πολύπλοκες δομές απαραίτητα εμπλέκουν κατανομές διατεταγμένων παρατηρήσεων. Περισσότερες πληροφορίες σε θέματα αξιοπιστίας και χρόνων ζωής δίνονται στα κλασσικά βιβλία των Barlow and Proschan (1975), Sinha (1986) [βλ. επίσης Κούτρας (2008), Μηλιένος (2009), Τριανταφύλλου (2009)].

- **Έλεγχος Ποιότητας (*Quality Control*)**. Η συνεισφορά της Θεωρίας των Διατεταγμένων Παρατηρήσεων είναι καθοριστική στην επίλυση προβλημάτων Ελέγχου Ποιότητας, σκοπός του οποίου -όπως ήδη έχουμε αναφέρει- είναι η έγκαιρη ανίχνευση πιθανών ειδικών αιτών μεταβλητότητας του/των υπό μελέτη χαρακτηριστικού/κών. Η ύπαρξη των εν λόγω αιτιών σε μία διεργασία, μπορεί ενδεχομένως να προκαλέσει την «αφύσικη» συμπεριφορά της. Βασικά εργαλεία για την επαναφορά της διεργασίας στην αρχική της κατάσταση αποτελούν τα διαγράμματα ελέγχου, η κατασκευή πολλών εκ των οποίων κάνει χρήση διατεταγμένων παρατηρήσεων και συναρτήσεων αυτών. Χαρακτηριστικό παράδειγμα αποτελεί το μονοδιάστατο παραμετρικό διάγραμμα για το εύρος (*R chart*) που προτάθηκε από τον Shewhart (1931). Στο σημείο αυτό αξίζει να σημειωθεί ότι πρόσφατα οι Karagöz and Koyuncu (2017) χρησιμοποίησαν διατεταγμένα σχήματα δειγματοληψίας (*ranked set sampling schemes*) σε δείγματα που προέρχονται από δισδιάστατες ασύμμετρες κατανομές με σκοπό την κατασκευή ορίων ελέγχου σε διαγράμματα για το εύρος.

Παράλληλα, μελετήθηκαν μη παραμετρικά διαγράμματα, τα οποία χρησιμοποιούν διατεταγμένες παρατηρήσεις για την κατασκευή των σ.σ. ελέγχου ή/και τον προσδιορισμό των ορίων ελέγχου. Συγκεκριμένα, οι Janacek and Meikle (1997) παρουσίασαν ένα διάγραμμα για τη διάμεσο (*Median chart*), ενώ πολλές παραλλαγές αυτού έχουν προταθεί όχι μόνο για την παρακολούθηση της διαμέσου, αλλά γενικότερα μιας διατεταγμένης παρατήρησης (βλ. Chakraborti *et al.* (2004), Balakrishnan *et al.* (2010), Triantafyllou (2017)). Τα προαναφερθέντα διαγράμματα, όπως και αυτά των Balakrishnan *et al.* (2009), Chakraborti *et al.* (2009), θεωρούνται διαγράμματα ελέγχου προτεραιότητας (*precedence control charts*), διότι η στατιστική τους συνάρτηση δεν είναι παρά μία συνάρτηση προτεραιότητας (*precedence statistics*). Θα πρέπει να αναφερθεί ότι έχουν, επίσης, αναπτυχθεί έλεγχοι προτεραιότητας (*precedence tests*), οι οποίοι αξιοποιούν τις ιδιότητες των διατεταγμένων παρατηρήσεων και έχουν πολλές

πρακτικές εφαρμογές (βλ. Balakrishnan and Ng (2006)).

Εκτενής ανασκόπηση της βιβλιογραφίας σχετικά με την παραπάνω οικογένεια/κλάση διαγραμμάτων γίνεται στο επόμενο κεφάλαιο, ενώ στα Κεφάλαια 3 έως 6 παρουσιάζονται νέα διαγράμματα ελέγχου που βασίζονται σε διατεταγμένες παρατηρήσεις ή/και τις συμμεταβλητές τους, τα οποία ουσιαστικά αποτελούν γενικεύσεις σε δύο διαστάσεις του κλασσικού μονοδιάστατου διαγράμματος ελέγχου για τη διάμεσο και τροποποιήσεις αυτού για την παρακολούθηση ζεύγους (δισδιάστατων) διατεταγμένων παρατηρήσεων ή/και πλήθους συγκεκριμένων παρατηρήσεων. Πρώτα, όμως, κρίνεται αναγκαία η παρουσίαση βασικών εννοιών και η απόδειξη νέων αποτελεσμάτων σχετικά με δισδιάστατες κατανομές διατεταγμένων παρατηρήσεων που χρειάζονται απαραίτητως για την ανάπτυξη και θεωρητική μελέτη των προτεινόμενων διαγραμμάτων ελέγχου.

2.3 Μονοδιάστατες κατανομές διατεταγμένων παρατηρήσεων

Στην ενότητα αυτή θα παρουσιάσουμε κάποια ήδη γνωστά αποτελέσματα σχετικά με κατανομές διατεταγμένων παρατηρήσεων που προέρχονται από το ίδιο δείγμα. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα βιβλία των Arnold *et al.* (1992), David and Nagaraja (2003) για μία εκτενή παρουσίαση των κατανομών αυτών (βλ. επίσης Κούτρας (2005, 2016), Αντζουλάκος και Κούτρας (2016)).

Στο σημείο θα πρέπει να αναφερθεί ότι παρόλο που υπάρχουν αποτελέσματα για διακριτές τ.μ. σε διατεταγμένες παρατηρήσεις, στα πλαίσια της παρούσας διατριβής παρουσιάζονται αποτελέσματα μόνο για συνεχείς τ.μ.

Αρχικά, θεωρούμε ότι το διατεταγμένο δείγμα $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ αντιστοιχεί σε ένα τυχαίο δείγμα n ανεξάρτητων και ισόνομων (*independent and identically distributed*, i.i.d.) παρατηρήσεων X_1, \dots, X_n με α.σ.κ. $F_{X_i}(x) = F_X(x)$ και συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) $f_{X_i}(x) = f_X(x) = F'_X(x)$, $1 \leq i \leq n$.

Ξεκινώντας από τις ακραίες παρατηρήσεις, η α.σ.κ. της μεγαλύτερης διατεταγμένης παρατήρησης $X_{n:n} = \max(X_1, X_2, \dots, X_n)$ υπολογίζεται ως εξής

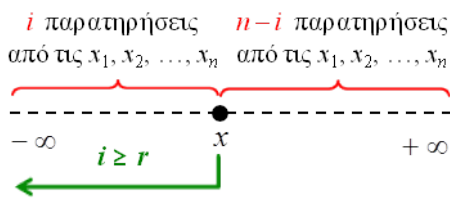
$$\begin{aligned} F_{n:n}(x) &= F_{X_{n:n}}(x) = P(X_{n:n} \leq x) \\ &= P(\text{όλες οι τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x) \\ &= P(X_1 \leq x) \dots P(X_n \leq x) \\ &= [F_X(x)]^n, \quad x \in \mathbb{R}, \end{aligned} \tag{2.1}$$

2.3. Μονοδιάστατες κατανομές διατεταγμένων παρατηρήσεων

ενώ με όμοιο τρόπο προκύπτει εύκολα και η α.σ.κ. της μικρότερης διατεταγμένης παρατήρησης $X_{1:n} = \min(X_1, X_2, \dots, X_n)$ όπως φαίνεται παρακάτω

$$\begin{aligned}
 F_{1:n}(x) &= F_{X_{1:n}}(x) = P(X_{1:n} \leq x) = 1 - P(X_{1:n} > x) \\
 &= 1 - P(\text{όλες οι τ.μ. } X_1, \dots, X_n \text{ είναι μεγαλύτερες από } x) \\
 &= 1 - P(X_1 > x) \dots P(X_n > x) \\
 &= 1 - [1 - P(X_1 \leq x)] \dots [1 - P(X_n \leq x)] \\
 &= 1 - [1 - F_X(x)]^n. \quad x \in \mathbb{R},
 \end{aligned} \tag{2.2}$$

Οι Σχέσεις (2.1) και (2.2) αποτελούν ειδικές περιπτώσεις ενός γενικότερου αποτελέσματος για την α.σ.κ. (συμβ. $F_{X_{r:n}}(x) = F_{r:n}(x)$) της r -οστής διατεταγμένης παρατήρησης $X_{r:n}$.



Σχήμα 2.1: Γραφική απεικόνιση του ενδεχομένου $[X_{r:n} \leq x]$

| | $i = 1$ | $i = 2$ |
|----------|--------------|-------------|
| A_i | $[X \leq x]$ | $[X > x]$ |
| $P(A_i)$ | p | $q = 1 - p$ |

Πίνακας 2.1: Σχήμα δειγματοληψίας μιας διατεταγμένης παρατήρησης

Το ενδεχόμενο που περιγράφεται στην έκφραση της $F_{X_{r:n}}(x)$ απεικονίζεται στο Σχήμα 2.1. Σύμφωνα με αυτό, για να πραγματοποιηθεί το ενδεχόμενο $[X_{r:n} \leq x]$ χρειάζεται τα ενδεχόμενα $A_1 : [X \leq x]$ και $A_2 : [X > x]$ να πραγματοποιηθούν i και $n - i$ φορές με πιθανότητες $p = P(X \leq x) = F_X(x)$ και $q = P(X > x) = 1 - F_X(x)$, αντίστοιχα (βλ. Πίνακα 2.1). Υπό την προϋπόθεση ότι το πείραμα επαναλαμβάνεται n φορές, για να συμβεί το ενδεχόμενο $[X_{r:n} \leq x]$ θα πρέπει να παρατηρηθούν

- i εμφανίσεις του ενδεχομένου A_1 (οι οποίες πραγματοποιούνται με $\binom{n}{i}$ τρόπους) και
- $n - i$ εμφανίσεις του ενδεχομένου A_2 (οι οποίες πραγματοποιούνται ακριβώς με $\binom{n-i}{n-i} = 1$ τρόπο), με το i να λαμβάνει τις τιμές $r, r + 1, \dots, n$.

Πιο συγκεκριμένα, έχουμε

$$\begin{aligned}
 F_{r:n}(x) &= P(X_{r:n} \leq x) \\
 &= P(\text{τουλάχιστον } r \text{ από τις τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x)
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=r}^n P(\text{ακριβώς } i \text{ από τις τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x) \\
 &= \sum_{i=r}^n P(\text{το ενδεχόμενο } [X \leq x] \text{ πραγματοποιείται ακριβώς } i \text{ φορές και} \\
 &\quad \text{το ενδεχόμενο } [X > x] \text{ πραγματοποιείται ακριβώς } n - i \text{ φορές}) \\
 &= \sum_{i=r}^n \binom{n}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i}, \quad x \in \mathbb{R}. \tag{2.3}
 \end{aligned}$$

Συνεπώς, είναι εμφανές ότι η α.σ.κ. της $X_{r:n}$ εκφράζεται μέσω της ουράς (*tail probability*) της διωνυμικής κατανομής $b(n, p)$ με πιθανότητα επιτυχίας p και πλήθος δοκιμών n . Στις ειδικές περιπτώσεις όπου $r = n$ και $r = 1$, εύκολα επαληθεύεται ότι η Σχέση (2.3) παίρνει τη μορφή των Σχέσεων (2.1) και (2.2), αντίστοιχα.

Εναλλακτικά, η α.σ.κ. της $X_{r:n}$ μπορεί να εκφραστεί με τη βοήθεια της κανονικοποιημένης μη-πλήρους συνάρτησης Βήτα (*regularized incomplete beta function*) $I_x(p, q)$, η οποία ορίζεται ως το πηλίκο της μη-πλήρους συνάρτησης Βήτα (*incomplete beta function*) του Pearson (1934) προς τη γνωστή συνάρτηση Βήτα (*beta function*) με αντίστοιχους τύπους

$$B_x(p, q) = \int_0^x t^{p-1}(1-t)^{q-1} dt \quad \text{και} \quad B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \frac{(p-1)!(q-1)!}{(p+q-1)!},$$

όπου $0 < x < 1$ και $p, q > 0$ (πραγματικοί). Για περαιτέρω μελέτη των προαναφερθεισών συναρτήσεων, καθώς και σχέσεων που συνδέουν άλλες ειδικές συναρτήσεις με ολοκληρώματα, αθροίσματα ή γινόμενα ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Gradshteyn and Ryzhik (2007). Αξιοποιώντας την παραπάνω έκφραση, μπορεί να διαπιστωθεί ότι η Σχέση (2.3) γράφεται ισοδύναμα ως εξής

$$F_{r:n}(x) = \int_0^{F_X(x)} \frac{1}{B(r, n-r+1)} t^{r-1}(1-t)^{n-r} dt = I_{F_X(x)}(r, n-r+1), \quad x \in \mathbb{R}. \tag{2.4}$$

Προφανώς, εκατοστιαία σημεία (*percentage points*) της $X_{r:n}$ μπορούν να υπολογισθούν απευθείας από τους αντίστοιχους πίνακες (εκατοστιαίων σημείων) της μη-πλήρους συνάρτησης Βήτα (βλ. Pearson and Hartley (1970)). Αξίζει να σημειωθεί ότι οι Epstein (1949), Gupta and Gupta (1984) και Rohatgi (1987) ήταν από τους πρώτους που μελέτησαν την κατανομή της $X_{r:n}$ στην ειδική περίπτωση που το μέγεθος δείγματος είναι τ.μ., έστω N , η οποία μπορεί να υπολογισθεί δεσμεύοντας ως προς το ενδεχόμενο $\{N = n\}$.

Θα ήταν παράλειψη να μην αναφερθεί ότι τύποι για την $F_{r:n}(x)$ προκύπτουν κάνοντας χρήση πιθανοτήτων που σχετίζονται με την αρνητική διωνυμική κατανομή. Αυτό μπορεί να διαπιστωθεί γράφοντας διαδοχικά το εξής

$$\begin{aligned}
 F_{r:n}(x) &= P(X_{r:n} \leq x) \\
 &= P(\text{επίτευξη } r \text{ επιτυχιών στο σύνολο το πολύ } n \text{ δοκιμών} \\
 &\quad \text{με πιθανότητα επιτυχίας } p = F_X(x)) \\
 &= \binom{r-1}{r-1} [F_X(x)]^r [1 - F_X(x)]^0 + \binom{r}{r-1} [F_X(x)]^r [1 - F_X(x)]^1 \\
 &\quad + \cdots + \binom{n-1}{r-1} [F_X(x)]^r [1 - F_X(x)]^{n-r} \\
 &= \sum_{i=0}^{n-r} \binom{n-i-1}{r-1} [F_X(x)]^r [1 - F_X(x)]^{n-r-i}, \quad x \in \mathbb{R} \tag{2.5\alpha}
 \end{aligned}$$

$$= [F_X(x)]^r \sum_{i=1}^{n-r+1} \binom{n-i}{r-1} [1 - F_X(x)]^{n-r-i+1}, \quad x \in \mathbb{R}, \tag{2.5\beta}$$

όπου οι Σχέσεις (2.5α) και (2.5β) αποδείχθηκαν από τους Pinsky *et al.* (1986) και οι David and Shu (1978), αντίστοιχα. Οι David and Nagaraja (2003) έδωσαν μία ακόμη εκδοχή της α.σ.κ. της $X_{r:n}$ μέσω της αρνητικής διωνυμικής μετρώντας όμως το πλήθος των αποτυχιών j ως ακολούθως

$$\begin{aligned}
 F_{r:n}(x) &= \sum_{j=0}^{n-r} P(\text{ακριβώς } r \text{ από τις τ.μ. } X_1, X_2, \dots, X_{r+j}, \text{ συμπεριλαμβανομένης της } X_{r+j}, \\
 &\quad \text{είναι μικρότερες ή ίσες από } x) \\
 &= [F_X(x)]^r \sum_{j=0}^{n-r} \binom{r+j-1}{r-1} [1 - F_X(x)]^j, \quad x \in \mathbb{R}. \tag{2.5\gamma}
 \end{aligned}$$

Επιπλέον, είναι γνωστό ότι η $F_{r:n}(x)$ ικανοποιεί τις παρακάτω αναδρομικές σχέσεις (βλ. David and Shu (1978))

$$F_{r:n}(x) = F_{r+1:n}(x) + \binom{n}{r} [F_X(x)]^r [1 - F_X(x)]^{n-r}, \quad r = 1, 2, \dots, n-1, \tag{2.6}$$

$$F_{r:n}(x) = F_{r:n-1}(x) + \binom{n-1}{r-1} [F_X(x)]^r [1 - F_X(x)]^{n-r}, \quad r = 1, 2, \dots, n-1, \tag{2.7}$$

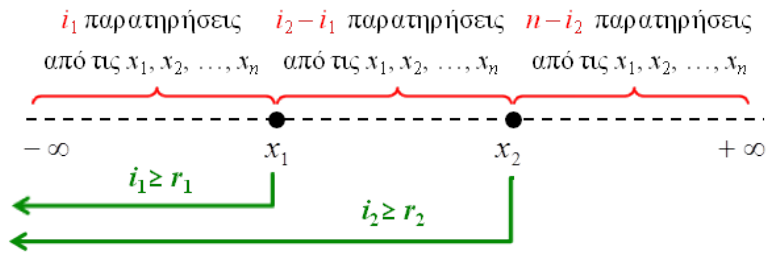
όπου η Σχέση (2.6) προκύπτει απευθείας από τη Σχέση (2.3), ενώ στη Σχέση (2.7) μπορούμε να οδηγηθούμε είτε μέσω της Σχέσης (2.3) είτε μέσω της Σχέσης (2.6) σε συνδυασμό με τη χρήση της παρακάτω βασικής αναδρομικής σχέσης

$$(n-r)F_{r:n}(x) + rF_{r+1:n}(x) = nF_{r:n-1}(x), \quad r = 1, 2, \dots, n-1.$$

Είναι περιττό να αναφερθεί ότι οι Εκφράσεις (2.5β) και (2.5γ) μπορούν να υπολογιστούν και ύστερα από επαναλαμβανόμενες εφαρμογές της Σχέσης (2.7).

Ενδιαφέρον παρουσιάζει η από κοινού α.σ.κ. κατανομή $F_{r_1, r_2; n}(x_1, x_2) = F_{X_{r_1; n}, X_{r_2; n}}(x_1, x_2)$ δύο διατεταγμένων παρατηρήσεων, έστω $X_{r_1; n}$ και $X_{r_2; n}$. Στην περίπτωση που ισχύει $x_1 \geq x_2$, από την ανισότητα $X_{r_2; n} \leq x_2$ συνεπάγεται ότι $X_{r_1; n} \leq x_1$ και επομένως

$$F_{r_1, r_2; n}(x_1, x_2) = F_{r_2; n}(x_2).$$



Σχήμα 2.2: Γραφική απεικόνιση του ενδεχομένου $[X_{r_1; n} \leq x_1, X_{r_2; n} \leq x_2]$

| | $i = 1$ | $i = 2$ | $i = 3$ |
|----------|----------------|----------------------|-------------|
| A_i | $[X \leq x_1]$ | $[x_1 < X \leq x_2]$ | $[X > x_2]$ |
| $P(A_i)$ | p_1 | p_2 | p_3 |

Πίνακας 2.2: Σχήμα δειγματοληψίας δύο διατεταγμένων παρατηρήσεων του ίδιου δείγματος

Ωστόσο, όταν $x_1 < x_2$, προκύπτει η διάταξη του Σχήματος 2.2. Ακολουθώντας παρόμοια λογική με αυτή που χρησιμοποιήθηκε στην απόδειξη της Σχέσης (2.3) για τον υπολογισμό της $F_{X_{r; n}}$, θεωρούμε αρχικά ότι τα ενδεχόμενα $A_1 : [X \leq x_1]$, $A_2 : [x_1 < X \leq x_2]$ και $A_3 : [X > x_2]$ εμφανίζονται i_1 , $i_2 - i_1$ και $n - i_2$ φορές με αντίστοιχες πιθανότητες $p_1 = P(X \leq x_1) = F_X(x_1)$, $p_2 = P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$, $p_3 = P(X > x_2) = 1 - F_X(x_2)$ (βλ. Πίνακα 2.2). Συνεπώς, σε n δοκιμές του πειράματος τύχης,

- i_1 αποτελέσματα του ενδεχομένου A_1 πραγματοποιούνται με $\binom{n}{i_1}$ τρόπους,
- $i_2 - i_1$ αποτελέσματα του ενδεχομένου A_2 παρατηρούνται με $\binom{n - i_1}{i_2 - i_1}$ τρόπους και
- $n - i_2$ αποτελέσματα του ενδεχομένου A_3 πραγματοποιούνται ακριβώς με $\binom{n - i_1 - (i_2 - i_1)}{n - i_2} = \binom{n - i_2}{n - i_2} = 1$ τρόπο.

Ως εκ τούτου, όλες οι πιθανές περιπτώσεις σε n δοκιμές απαριθμούνται ως ακολούθως

$$\binom{n}{i_1} \binom{n-i_1}{i_2-i_1} = \binom{n}{i_1, i_2-i_1, n-i_2} = \frac{n!}{i_1!(i_2-i_1)!(n-i_2)!} = \alpha_n(i_1, i_2).$$

Τελικά, η από κοινού α.σ.κ. των $X_{r_1:n}, X_{r_2:n}$ ($1 \leq r_1 \leq r_2 \leq n$) εκφράζεται μέσω της ουράς της τριωνυμικής κατανομής (*trinomial distribution*) με παραμέτρους n , p_1 και p_2 ως εξής

$$\begin{aligned} F_{r_1, r_2; n}(x_1, x_2) &= P(X_{r_1:n} \leq x_1, X_{r_2:n} \leq x_2) \\ &= P(\text{τουλάχιστον } r_1 \text{ από τις τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x_1, \\ &\quad \text{τουλάχιστον } r_2 \text{ από τις τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x_2) \end{aligned}$$

$$\begin{aligned} &\sum_{i_2=r_2}^n \sum_{i_1=r_1}^{i_2} P(\text{ακριβώς } i_1 \text{ από τις τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x_1, \\ &\quad \text{ακριβώς } i_2 \text{ από τις τ.μ. } X_1, \dots, X_n \text{ είναι μικρότερες ή ίσες από } x_2) \\ &= \sum_{i_2=r_2}^n \sum_{i_1=r_1}^{i_2} \alpha_n(i_1, i_2) [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1} [1 - F_X(x_2)]^{n-i_2}, \end{aligned}$$

όπου $-\infty < x_1 < x_2 < +\infty$. Αξίζει να σημειωθεί ότι όλοι οι παραπάνω τύποι, στους οποίους εμπλέκονται αθροίσματα, είναι έγκυροι και στη διακριτή περίπτωση.

Εναλλακτικά, η παραπάνω σχέση γράφεται μέσω της διμεταβλητής (*bivariate*) μη-πλήρους συνάρτησης Βήτα και της διμεταβλητής συνάρτησης Βήτα, συμβ.

$$B_{x_1, x_2}(p, q, r) = \int_0^1 \int_0^{x_2} t_1^{p-1} (t_2 - t_1)^{q-1} (1 - t_2)^{r-1} dt_1 dt_2, \quad 0 < x_1 \leq x_2 < 1,$$

και

$$B(p, q, r) = B(p, q) \cdot B(p + q, r) = \frac{\Gamma(p)\Gamma(q)\Gamma(r)}{\Gamma(p + q + r)}, \quad p, q, r > 0 \text{ (πραγματικοί)},$$

όπως φαίνεται παρακάτω

$$\begin{aligned} F_{r_1, r_2; n}(x_1, x_2) &= \int_0^1 \int_0^{F_X(x_2)} \frac{n!}{(r_1-1)!(r_2-r_1-1)!(n-r_2)!} t_1^{r_1-1} (t_2 - t_1)^{r_2-r_1-1} (1 - t_2)^{n-r_2} dt_1 dt_2 \\ &= B_{1, F_X(x_2)}(r_1, r_2 - r_1, n - r_2 + 1) / B(r_1, r_2 - r_1, n - r_2 + 1) \\ &= I_{1, F_X(x_2)}(r_1, r_2 - r_1, n - r_2 + 1), \quad -\infty < x_1 < x_2 < +\infty. \end{aligned}$$

Για τον υπολογισμό των συναρτήσεων πυκνότητας, αρκεί να παραγωγίσουμε κατάλληλα τις αντίστοιχες αθροιστικές κατανομές. Συνεπώς, οι σ.π.π. των $X_{n:n}$ και $X_{1:n}$ δίνονται από τους παρακάτω τύπους

$$f_{n:n}(x) = \frac{\partial}{\partial x} F_{n:n}(x) = n[F_X(x)]^{n-1} f_X(x), \quad x \in \mathbb{R},$$

$$f_{1:n}(x) = \frac{\partial}{\partial x} F_{1:n}(x) = n[1 - F_X(x)]^{n-1} f_X(x), \quad x \in \mathbb{R},$$

ενώ η σ.π.π. της $X_{r:n}$ υπολογίζεται ως εξής¹

$$\begin{aligned} f_{r:n}(x) &= \frac{\partial}{\partial x} F_{r:n}(x) = \frac{\partial}{\partial x} \sum_{i=r}^n \binom{n}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i} \\ &= \sum_{i=r}^n \binom{n}{i} i [F_X(x)]^{i-1} [1 - F_X(x)]^{n-i} f_X(x) - \sum_{i=r}^{n-1} \binom{n}{i} [F_X(x)]^i (n-i) [1 - F_X(x)]^{n-i-1} f_X(x) \\ &= n f_X(x) \left\{ \sum_{i=r}^n \binom{n-1}{i-1} [F_X(x)]^{i-1} [1 - F_X(x)]^{n-i} - \sum_{i=r}^{n-1} \binom{n-1}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i-1} \right\} \\ &= n f_X(x) \binom{n-1}{r-1} [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} \\ &= \frac{1}{B(r, n-r+1)} [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} f_X(x), \quad x \in \mathbb{R}. \end{aligned} \quad (2.8)$$

Με ανάλογο τρόπο υπολογίζεται ο τύπος της από κοινού σ.π.π. των διατεταγμένων παρατηρήσεων $X_{r_1:n}$ και $X_{r_2:n}$ ($1 \leq r_1 \leq r_2 \leq n$). Συγκεκριμένα, γράφοντας την $F_{r_1, r_2:n}(x_1, x_2)$ στην παρακάτω μορφή²

$$\begin{aligned} f_{r_1, r_2:n}(x_1, x_2) &= \frac{\partial^2}{\partial x_1 \partial x_2} F_{r_1, r_2:n}(x_1, x_2) \\ &= \frac{\partial^2}{\partial x_1 \partial x_2} \sum_{i_2=r_2}^n \sum_{i_1=r_1}^{i_2} \alpha_n(i_1, i_2) [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1} [1 - F_X(x_2)]^{n-i_2} \\ &= \frac{\partial}{\partial x_2} \sum_{i_2=r_2}^n \binom{n}{i_2} [1 - F_X(x_2)]^{n-i_2} \left\{ \frac{\partial}{\partial x_1} \sum_{i_1=r_1}^{i_2} \binom{i_2}{i_1} [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1} \right\}, \end{aligned}$$

¹ Στην απόδειξη απαιτείται η χρήση των γνωστών συνδυαστικών ταυτοτήτων $i \binom{n}{i} = n \binom{n-1}{i-1}$ και $(n-i) \binom{n}{i} = (n-1) \binom{n}{n-i} = n \binom{n-1}{n-1-i} = n \binom{n-1}{i}$ με $1 \leq i \leq n$ (βλ. Χαραλαμπίδης (2000) και Κούτρας (2001)), ενώ η απλοποίηση των αθροισμάτων προκύπτει άμεσα παρατηρώντας ότι $\sum_{i=r}^n \binom{n-1}{i-1} [F_X(x)]^{i-1} [1 - F_X(x)]^{n-i} = \sum_{i=r-1}^{n-1} \binom{n-1}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i-1} = \binom{n-1}{r-1} [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} + \sum_{i=r}^{n-1} \binom{n-1}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i-1}$.

² Αξιοποιώντας ταυτότητα της συνδυαστικής (βλ. Χαραλαμπίδης (2000), Κούτρας (2001)), είναι φανερό ότι για τον τριωνυμικό συντελεστή $a_n(i_1, i_2)$ ισχύει η ισότητα $\binom{n}{i_1} \binom{n-i_1}{i_2-i_1} = \binom{n}{i_2} \binom{i_2}{i_1}$.

η επόμενη ποσότητα μπορεί να υπολογισθεί ανεξάρτητα ως ακολούθως

$$\begin{aligned}
 \frac{\partial}{\partial x_1} \sum_{i_1=r_1}^{i_2} \binom{i_2}{i_1} [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1} &= \sum_{i_1=r_1}^{i_2} \binom{i_2}{i_1} i_1 [F_X(x_1)]^{i_1-1} f_X(x_1) [F_X(x_2) - F_X(x_1)]^{i_2-i_1} \\
 &\quad - \sum_{i_1=r_1}^{i_2-1} \binom{i_2}{i_1} [F_X(x_1)]^{i_1} (i_2 - i_1) [F_X(x_2) - F_X(x_1)]^{i_2-i_1-1} f_X(x_1) \\
 &= \sum_{i_1=r_1}^{i_2} \binom{i_2-1}{i_1-1} i_2 [F_X(x_1)]^{i_1-1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1} f_X(x_1) \\
 &\quad - \sum_{i_1=r_1}^{i_2-1} \binom{i_2-1}{i_1} i_2 [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1-1} f_X(x_1) \\
 &= i_2 f_X(x_1) \left\{ \sum_{i_1=r_1-1}^{i_2-1} \binom{i_2-1}{i_1} [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1-1} \right. \\
 &\quad \left. - \sum_{i_1=r_1}^{i_2-1} \binom{i_2-1}{i_1} [F_X(x_1)]^{i_1} [F_X(x_2) - F_X(x_1)]^{i_2-i_1-1} \right\} \\
 &= \binom{i_2-1}{r_1-1} i_2 [F_X(x_1)]^{r_1-1} [F_X(x_2) - F_X(x_1)]^{i_2-r_1} f_X(x_1),
 \end{aligned}$$

με αποτέλεσμα η από κοινού πυκνότητα να γράφεται ισοδύναμα στη μορφή¹

$$\begin{aligned}
 f_{r_1, r_2; n}(x_1, x_2) &= r_1 \binom{n}{r_1} [F_X(x_1)]^{r_1-1} f_X(x_1) \\
 &\quad \times \left\{ \frac{\partial}{\partial x_2} \sum_{i_2=r_2}^n \binom{n-r_1}{i_2-r_1} [1 - F_X(x_2)]^{n-i_2} [F_X(x_2) - F_X(x_1)]^{i_2-r_1} \right\}.
 \end{aligned}$$

Στη συνέχεια, εάν αντικαταστήσουμε την παράγωγο ως προς x_2 με την ποσότητα

$$\begin{aligned}
 \frac{\partial}{\partial x_2} \sum_{i_2=r_2}^n \binom{n-r_1}{i_2-r_1} [1 - F_X(x_2)]^{n-i_2} [F_X(x_2) - F_X(x_1)]^{i_2-r_1} \\
 &= \sum_{i_2=r_2}^n \binom{n-r_1}{i_2-r_1} [1 - F_X(x_2)]^{n-i_2} (i_2 - r_1) [F_X(x_2) - F_X(x_1)]^{i_2-r_1-1} f_X(x_2) \\
 &\quad - \sum_{i_2=r_2}^{n-1} \binom{n-r_1}{i_2-r_1} (n - i_2) [1 - F_X(x_2)]^{n-i_2-1} f_X(x_2) [F_X(x_2) - F_X(x_1)]^{i_2-r_1}
 \end{aligned}$$

¹ Λαμβάνεται υπ' όψιν ότι $\binom{n}{i_2} \binom{i_2-1}{r_1-1} i_2 = r_1 \binom{n}{r_1} \binom{n-r_1}{i_2-r_1}$.

$$\begin{aligned}
 &= \sum_{i_2=r_2}^n \binom{n-r_1-1}{i_2-r_1-1} (n-r_1) [1-F_X(x_2)]^{n-i_2} [F_X(x_2) - F_X(x_1)]^{i_2-r_1-1} f_X(x_2) \\
 &\quad - \sum_{i_2=r_2}^{n-1} \binom{n-r_1-1}{i_2-r_1} (n-r_1) [1-F_X(x_2)]^{n-i_2-1} [F_X(x_2) - F_X(x_1)]^{i_2-r_1} f_X(x_2) \\
 &= (n-r_1) f_X(x_2) \left\{ \sum_{i_2=r_2-1}^{n-1} \binom{n-r_1-1}{i_2-r_1} [1-F_X(x_2)]^{n-i_2-1} [F_X(x_2) - F_X(x_1)]^{i_2-r_1} \right. \\
 &\quad \left. - \sum_{i_2=r_2}^{n-1} \binom{n-r_1-1}{i_2-r_1} [1-F_X(x_2)]^{n-i_2-1} [F_X(x_2) - F_X(x_1)]^{i_2-r_1} \right\} \\
 &= \binom{n-r_1-1}{r_2-r_1-1} (n-r_1) [1-F_X(x_2)]^{n-r_2} [F_X(x_2) - F_X(x_1)]^{r_2-r_1-1} f_X(x_2)
 \end{aligned}$$

και παρατηρήσουμε ότι

$$\begin{aligned}
 r_1 (n-r_1) \binom{n}{r_1} \binom{n-r_1-1}{r_2-r_1-1} &= \frac{n!}{(n-r_1-1)!(r_1-1)!} \binom{n-r_1-1}{r_2-r_1-1} \\
 &= \binom{n}{r_1-1, 1, r_2-r_1-1, 1, n-r_2} = \frac{n!}{(r_1-1)!(r_2-r_1-1)!(n-r_2)!},
 \end{aligned}$$

η τελική έκφραση της από κοινού σ.π.π. δύο οποιωνδήποτε διατεταγμένων παρατηρήσεων $X_{r_1:n}, X_{r_2:n}$ ($1 \leq r_1 < r_2 \leq n$) παίρνει τη μορφή

$$\begin{aligned}
 f_{r_1, r_2:n}(x_1, x_2) &= \frac{n!}{(r_1-1)!(r_2-r_1-1)!(n-r_2)!} [F_X(x_1)]^{r_1-1} [F_X(x_2) - F_X(x_1)]^{r_2-r_1-1} \\
 &\quad \times [1-F_X(x_2)]^{n-r_2} f_X(x_1) f_X(x_2), \quad -\infty < x_1 < x_2 < +\infty.
 \end{aligned} \tag{2.9}$$

Ειδικότερα, θέτοντας $r_1 = r_i$ και $r_2 = r_i + 1$ στην παραπάνω σχέση, προκύπτει η από κοινού σ.π.π. δύο διαδοχικών διατεταγμένων παρατηρήσεων, η οποία λαμβάνει την πιο απλή έκφραση

$$f_{r_i, r_i+1:n}(x_i, x_{i+1}) = \frac{n!}{(r_i-1)!(n-r_i-1)!} [F_X(x_i)]^{r_i-1} [1-F_X(x_{i+1})]^{n-r_i-1} f_X(x_i) f_X(x_{i+1}),$$

όπου $-\infty < x_i < x_{i+1} < +\infty$. Επιπρόσθετα, στην ειδική περίπτωση όπου $r_1 = 1$ και $r_2 = n$, λαμβάνεται η από κοινού σ.π.π. των ακραίων διατεταγμένων παρατηρήσεων (μικρότερη και μεγαλύτερη διατεταγμένη παρατήρηση), η οποία δίνεται από τον ακόλουθο τύπο

$$f_{1,n:n}(x_1, x_n) = n(n-1) [F_X(x_n) - F_X(x_1)]^{n-2} f_X(x_1) f_X(x_n), \quad -\infty < x_1 < x_n < +\infty.$$

Γενικεύσεις της Σχέσης (2.9) για περισσότερες από δύο διατεταγμένες παρατηρήσεις είναι

πλέον εμφανείς. Για παράδειγμα, η από κοινού σ.π.π. των $X_{r_1:n}, X_{r_2:n}, \dots, X_{r_k:n}$ ($1 \leq r_1 < r_2 < \dots < r_k \leq n$, $1 \leq k \leq n$) δίνεται από τον τύπο

$$f_{r_1, r_2, \dots, r_k:n}(x_1, x_2, \dots, x_k) = \frac{n!}{(r_1 - 1)!(r_2 - r_1 - 1)! \dots (n - r_k)!} \\ \times [F_X(x_1)]^{r_1-1} [F_X(x_2) - F_X(x_1)]^{r_2-r_1-1} \dots [1 - F_X(x_k)]^{n-r_k} \\ \times f_X(x_1)f_X(x_2) \dots f_X(x_k), \quad -\infty < x_1 < x_2 < \dots < x_k < +\infty.$$

Ισοδύναμα, η παραπάνω σχέση γράφεται ως εξής

$$f_{r_1, r_2, \dots, r_k:n}(x_1, x_2, \dots, x_k) = n! \left[\prod_{i=0}^k \left\{ \frac{[F_X(x_{i+1}) - F_X(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1} - r_i - 1)!} \right\} \right] \times \left[\prod_{i=1}^k f_X(x_i) \right] \quad (2.10)$$

(σύμβαση: $x_0 = -\infty$, $x_{k+1} = +\infty$, $r_0 = 0$ και $r_{k+1} = n + 1$).

Τελικά, η από κοινού σ.π.π. όλων των n διατεταγμένων παρατηρήσεων έχει την απλή μορφή που ακολουθεί

$$f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f_X(x_i), \quad -\infty < x_1 < x_2 < \dots < x_n < +\infty. \quad (2.11)$$

Η παραπάνω έκφραση προκύπτει λαμβάνοντας υπ' όψιν τη διάταξη $x_{1:n} < x_{2:n} < \dots < x_{n:n}$ και την παραδοχή ότι οι αρχικές μεταβλητές X_i λαμβάνουν τις τιμές $x_{i:n}$ ($i = 1, 2, \dots, n$), το οποίο λόγω συμμετρίας έχει ως αποτέλεσμα να αποδίδονται ίσες πιθανότητες σε καθμία από τις $n!$ το πλήθος μεταθέσεις των αριθμών $1, 2, \dots, n$. Οι Hogg and Craig (1978) οδηγήθηκαν στη Σχέση (2.11) ακολουθώντας διαφορετική προσέγγιση χρησιμοποιώντας την Ιακωβιανή Μέθοδο (*Jacobian Method*) και έναν μετασχηματισμό $n!$ -προς-1 ($n!$ -to-1).

Αξίζει να σημειωθεί ότι ξεκινώντας από τη Σχέση (2.11) μπορεί κανείς να καταλήξει στις περιθώριες πυκνότητες που περιγράφονται από τους Τύπους (2.8) και (2.9) ύστερα από κατάλληλες πολλαπλές ολοκληρώσεις. Εναλλακτικά, μπορούμε να θεωρήσουμε ότι οι συναρτήσεις πυκνότητας των Σχέσεων (2.8) και (2.9) περιγράφονται από τα ενδεχόμενα $[x < X_{r:n} \leq x + \delta x]$ και $[x_1 < X_{r_1:n} \leq x_1 + \delta x_1, x_2 < X_{r_2:n} \leq x_2 + \delta x_2]$, τα οποία αντιστοιχούν σε συγκεκριμένες διατάξεις των αρχικών μεταβλητών του δείγματος. Στην περίπτωση αυτή, οι τύποι για τον υπολογισμό των προαναφερθεισών πυκνοτήτων μπορούν να δοθούν άμεσα δεδομένου ότι οι αντίστοιχες πιθανότητες να παρατηρηθούν τέτοιου είδους διατάξεις εκφράζονται μέσω της συνάρτησης πιθανότητας της πολυωνυμικής κατανομής. Για περισσότερες πληροφορίες, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα βιβλία των Balakrishnan and Cohen (1991), Arnold *et al.* (1992) και David and Nagaraja (2003).

Ωστόσο, αυτές οι προσεγγίσεις περιπλέκονται καθώς προχωράμε σε μεγαλύτερες διαστάσεις, όπου εμπλέκονται δύο ή περισσότερα συσχετισμένα χαρακτηριστικά. Για το λόγο αυτό είναι συνήθως προτιμότερο να ορίζονται αρχικά οι από κοινού α.σ.κ. των εμπλεκόμενων τ.μ. (βλ. για παράδειγμα το Κεφάλαιο 3) και στη συνέχεια να υπολογίζονται οι τύποι των αντίστοιχων πυκνοτήτων μέσω παραγωγίσης.

Αν έχουμε στη διάθεσή μας ένα τυχαίο δείγμα n ανεξάρτητων και ισόνομων παρατηρήσεων U_1, U_2, \dots, U_n που προέρχονται από την Ομοιόμορφη κατανομή στο $(0, 1)$, συμβ. $\mathcal{U}(0, 1)$, η σ.π.π. και η α.σ.κ. δίνονται αντίστοιχα από τους τύπους $f_U(u) = 1$ και $F_U(u) = u$ ($0 \leq u \leq 1$). Οπότε, οι τύποι για τις κατανομές που συνδέονται με το διατεταγμένο δείγμα $U_{1:n} \leq U_{2:n} \leq \dots \leq U_{n:n}$ προκύπτουν άμεσα από τους γενικούς τύπους που παρουσιάστηκαν προηγουμένως.

Για παράδειγμα, η α.σ.κ. της $U_{r:n}$, κάνοντας χρήση της Σχέσης (2.3), δίνεται από τον ακόλουθο τύπο

$$F_{r:n}(u) = \sum_{i=r}^n \binom{n}{i} u^i (1-u)^{n-i}, \quad 0 \leq u \leq 1$$

ή εναλλακτικά προκύπτει με τη βοήθεια της Σχέσης (2.4), όπως φαίνεται στην παρακάτω έκφραση

$$F_{r:n}(u) = \int_0^u \frac{1}{B(r, n-r+1)} t^{r-1} (1-t)^{n-r} dt = I_u(r, n-r+1), \quad 0 \leq u \leq 1.$$

Όμοια, βασιζόμενοι στη Σχέση (2.8), η σ.π.π. της $U_{r:n}$ υπολογίζεται ως εξής

$$f_{r:n}(u) = \frac{1}{B(r, n-r+1)} u^{r-1} (1-u)^{n-r}, \quad 0 \leq u \leq 1.$$

Προδήλως, η κατανομή της $U_{r:n}$ είναι μία κατανομή Βήτα με παραμέτρους r και $n-r+1$. Εκτενής μελέτη όχι μόνο της κατανομής Βήτα, αλλά και άλλων μονοδιάστατων (συνεχών ή διακριτών) κατανομών γίνεται στα βιβλία των Johnson *et al.* (1994, 1995), Johnson *et al.* (2005), Κούτρας (2004, 2016).

Με ανάλογο τρόπο προκύπτουν και οι υπόλοιπες αντιστοιχίες μεταξύ των τύπων για τον υπολογισμό των περιθώριων ή από κοινού κατανομών διατεταγμένων παρατηρήσεων που προέρχονται από το ίδιο δείγμα, όταν η κατανομή των αρχικών παρατηρήσεων είναι μία οποιαδήποτε συνεχής κατανομή (Περίπτωση Α) και όταν είναι η Ομοιόμορφη κατανομή στο $(0,1)$ (Περίπτωση Β). Τα αποτελέσματα παρουσιάζονται συγκεντρωτικά στους Πίνακες 2.3 και 2.4.

| Περίπτωση Α | Περίπτωση Β |
|---|--|
| $F_{n;n}(x) = [F_X(x)]^n, \quad x \in \mathbb{R}$ | $F_{n;n}(u) = u^n, \quad u \in [0, 1]$ |
| $F_{1;n}(x) = 1 - [1 - F_X(x)]^n, \quad x \in \mathbb{R}$ | $F_{1;n}(u) = 1 - (1 - u)^n, \quad u \in [0, 1]$ |
| $F_{r;n}(x) = \int_0^{F_X(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt$ $= \sum_{i=r}^n \binom{n}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i}$ $= [F_X(x)]^r \sum_{i=1}^{n-r+1} \binom{n-i}{r-1} [1 - F_X(x)]^{n-r-i+1}$ $= [F_X(x)]^r \sum_{j=0}^{n-r} \binom{r+j-1}{r-1} [1 - F_X(x)]^j, \quad x \in \mathbb{R}$ | $F_{r;n}(u) = \int_0^u \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt$ $= \sum_{i=r}^n \binom{n}{i} u^i (1-u)^{n-i}$ $= u^r \sum_{i=1}^{n-r+1} \binom{n-i}{r-1} (1-u)^{n-r-i+1}$ $= u^r \sum_{j=0}^{n-r} \binom{r+j-1}{r-1} (1-u)^j, \quad u \in [0, 1]$ |
| $F_{r_1, r_2; n}(x_1, x_2) = \sum_{i_2=i_1}^n \sum_{i_1=i_2}^n \frac{n!}{i_1!(i_2-i_1)!(n-i_2)!} [F_X(x_1)]^{i_1}$ $\times [F_X(x_2) - F_X(x_1)]^{i_2-i_1} [1 - F_X(x_2)]^{n-i_2}, \quad -\infty < x_1 < x_2 < +\infty$ | $F_{r_1, r_2; n}(u_1, u_2) = \sum_{i_2=i_1}^n \sum_{i_1=i_2}^n \frac{n!}{i_1!(i_2-i_1)!(n-i_2)!}$ $\times u_1^{i_1} (u_2 - u_1)^{i_2-i_1} (1 - u_2)^{n-i_2}, \quad 0 < u_1 < u_2 < 1$ |

Πίνακας 2.3: Περιθώριες και από κοινού αθροιστικές συναρτήσεις κατανομής διατεταγμένων παρατηρήσεων

| Περίπτωση Α | Περίπτωση Β |
|--|---|
| $f_{nn}(x) = n [F_X(x)]^{n-1} f_X(x), \quad x \in \mathbb{R}$ | $f_{nn}(u) = nu^{n-1} \quad u \in [0, 1]$ |
| $f_{1n}(x) = n [1 - F_X(x)]^{n-1} f_X(x), \quad x \in \mathbb{R}$ | $f_{1n}(u) = n(1-u)^{n-1} \quad u \in [0, 1]$ |
| $f_{rn}(x) = \frac{n!}{(r-1)!(n-r)!} [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} f_X(x), \quad x \in \mathbb{R}$ | $f_{rn}(u) = \frac{n!}{(r-1)!(n-r)!} u^{r-1} (1-u)^{n-r}, \quad u \in [0, 1]$ |
| $f_{r_1, r_2, n}(x_1, x_2) = \frac{n!}{(r_1-1)!(r_2-r_1-1)!(n-r_2)!} [F_X(x_1)]^{r_1-1} \times [F_X(x_2) - F_X(x_1)]^{r_2-r_1-1} [1 - F_X(x_2)]^{n-r_2} f_X(x_1) f_X(x_2),$ $-\infty < x_1 < x_2 < +\infty$ | $f_{r_1, r_2, n}(u_1, u_2) = \frac{n!}{(r_1-1)!(r_2-r_1-1)!(n-r_2)!} \times u_1^{r_1-1} (u_2 - u_1)^{r_2-r_1-1} (1-u_2)^{n-r_2},$ $0 \leq u_1 < u_2 \leq 1$ |
| $f_{r_i, r_{i+1}, n}(x_i, x_{i+1}) = \frac{n!}{(r_i-1)!(n-r_i-1)!} [F_X(x_i)]^{r_i-1} [1 - F_X(x_{i+1})]^{n-r_i-1} \times f_X(x_i) f_X(x_{i+1}),$ $-\infty < x_i < x_{i+1} < +\infty$ | $f_{r_i, r_{i+1}, n}(u_i, u_{i+1}) = \frac{n!}{(r_i-1)!(n-r_i-1)!} \times u_i^{r_i-1} (1-u_{i+1})^{n-r_i-1},$ $0 \leq u_i < u_{i+1} \leq 1$ |
| $f_{1, n:n}(x_1, x_n) = n(n-1) [F_X(x_n) - F_X(x_1)]^{n-2} f_X(x_1) f_X(x_n),$ $-\infty < x_1 < x_n < +\infty$ | $f_{1, n:n}(u_1, u_n) = n(n-1) (u_n - u_1)^{n-2}, \quad 1 \leq u_1 < u_n \leq 1$ |
| $f_{r_1, r_2, \dots, r_k, n}(x_1, x_2, \dots, x_k) = n! \left[\prod_{i=1}^k f_X(x_i) \right] \times \left[\prod_{i=0}^k \left\{ \frac{[F_X(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!} \right\} \right],$ $(1 \leq k \leq n, 1 \leq r_1 < \dots < r_k \leq n), -\infty < x_1 < \dots < x_k < +\infty$ <p style="text-align: center;">Σύμβαση: $x_0 = -\infty, x_{k+1} = +\infty, r_0 = 0, r_{k+1} = n+1$</p> | $f_{r_1, r_2, \dots, r_k, n}(u_1, u_2, \dots, u_k) = n! \prod_{i=0}^k \left\{ \frac{(u_{i+1} - u_i)^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!} \right\},$ $(1 \leq k \leq n, 1 \leq r_1 < \dots < r_k \leq n), 0 \leq u_1 < \dots < u_k \leq 1$ <p style="text-align: center;">Σύμβαση: $u_0 = 0, u_{k+1} = 1, r_0 = 0, r_{k+1} = n+1$</p> |
| $f_{r_1, r_2, \dots, r_n, n}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f_X(x_i),$ $-\infty < x_1 < \dots < x_n < +\infty$ | $f_{r_1, r_2, \dots, r_n, n}(u_1, u_2, \dots, u_n) = n!,$ $0 \leq u_1 < \dots < u_n \leq 1$ |

Πίνακας 2.4: Περιθώριες και από κοινού συναρτήσεις πυκνοτήτων διατεταγμένων παρατηρήσεων

Στο σημείο αυτό, θα πρέπει να αναφερθεί ότι από τον ορισμό του διατεταγμένου δείγματος γίνεται φανερό ότι οι διατεταγμένες παρατηρήσεις $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ είναι εξαρτημένες ακόμη και στην περίπτωση που οι αρχικές παρατηρήσεις X_1, X_2, \dots, X_n είναι ανεξάρτητες και ισόνομες. Ο συντελεστής γραμμικής συσχέτισης του *Pearson* δύο διατεταγμένων παρατηρήσεων $X_{r_1:n}$ και $X_{r_2:n}$ ($1 \leq r_1 < r_2 \leq n$) υπολογίζεται από τον τύπο

$$\rho(X_{r_1:n}, X_{r_2:n}) = \frac{\text{Cov}(X_{r_1:n}, X_{r_2:n})}{\sqrt{V(X_{r_1:n})} \sqrt{V(X_{r_2:n})}} = \frac{\sigma_{r_1, r_2:n}}{\sigma_{r_1:n} \sigma_{r_2:n}}.$$

Η ροπή k -τάξης (k -th raw moment) της διατεταγμένης παρατήρησης $X_{r_1:n}$ δίνεται από τον τύπο

$$\mu_{r_1:n}^{(k)} = E(X_{r_1:n}^k) = \int_{-\infty}^{\infty} x^k f_{r_1:n}(x) dx, \quad r_1 = 1, 2, \dots, n, \quad k \geq 1,$$

από τον οποίο (θέτοντας $k = 1$) οδηγούμαστε στην έκφραση για τη μέση τιμή της $X_{r_1:n}$, συμβ.

$$\mu_{r_1:n} = E(X_{r_1:n}) = \int_{-\infty}^{\infty} x f_{r_1:n}(x) dx, \quad r_1 = 1, 2, \dots, n.$$

Με τον ίδιο τρόπο μπορεί να ορισθεί η ροπή του γινομένου (*product moment*) των $X_{r_1:n}, X_{r_2:n}$ μέσω της παρακάτω έκφρασης

$$\mu_{r_1, r_2:n} = E(X_{r_1:n} X_{r_2:n}) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} x_1 x_2 f_{r_1, r_2:n}(x_1, x_2) dx_1 dx_2, \quad 1 \leq r_1 < r_2 \leq n,$$

ενώ η συνδιακύμανση (*covariance*) των $X_{r_1:n}, X_{r_2:n}$ δίνεται από τον τύπο που ακολουθεί

$$\begin{aligned} \sigma_{r_1, r_2:n} &= \text{Cov}(X_{r_1:n}, X_{r_2:n}) = E[(X_{r_1:n} - \mu_{r_1:n})(X_{r_2:n} - \mu_{r_2:n})] = \mu_{r_1, r_2:n} - \mu_{r_1:n} \mu_{r_2:n} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} (x_1 - \mu_{r_1:n})(x_2 - \mu_{r_2:n}) f_{r_1, r_2:n}(x_1, x_2) dx_1 dx_2, \quad 1 \leq r_1 < r_2 \leq n. \end{aligned}$$

Προφανώς, η διακύμανση (*variance*) της $X_{r_1:n}$ λαμβάνει την εξής μορφή

$$\sigma_{r_1:n}^{(2)} = \sigma_{r_1, r_1:n} = V(X_{r_1:n}) = \mu_{r_1:n}^{(2)} - \mu_{r_1:n}^2 = \int_{-\infty}^{\infty} (x - \mu_{r_1:n})^2 f_{r_1:n}(x) dx.$$

Με τη βοήθεια των παραπάνω τύπων, μπορεί να υπολογισθεί ο συντελεστής συσχέτισης για οποιαδήποτε συνεχή κατανομή. Σε περιπτώσεις στις οποίες η αναλυτική ολοκλήρωση δεν είναι εφικτή, μπορεί να εφαρμοστεί αριθμητική ολοκλήρωση. Γενικότερα, στη διεθνή βιβλιογραφία έχουν μελετηθεί αναδρομικές σχέσεις μεταξύ των ροπών διατεταγμένων παρατηρήσεων, ενώ παράλληλα έχουν δοθεί ακριβείς τύποι υπολογισμού των ροπών και άλλων αναμενόμενων τιμών διατεταγμένων παρατηρήσεων για συγκεκριμένες συνεχείς

κατανομές, όπως Ομοιόμορφη, Εκθετική, Λογιστική, Γάμμα, Κανονική, Weibull και άλλες. Ανασκόπηση των περιπτώσεων/μεθόδων αυτών γίνεται στις μονογραφίες των Balakrishnan and Cohen (1991), Arnold *et al.* (1992), David and Nagaraja (2003).

Αξίζει να σημειωθεί, ότι στην περίπτωση της Ομοιόμορφης κατανομής $\mathcal{U}(0, 1)$, η ροπή k -τάξης, η μέση τιμή και η διακύμανση της διατεταγμένης παρατήρησης $U_{r_1:n}$ δίνονται από τις εκφράσεις που ακολουθούν

$$E(U_{r_1:n}^k) = \frac{n!}{(r_1 - 1)!(n - r_1)!} \int_0^1 u^{k+r_1-1} (1 - u)^{n-r_1} du = \frac{B(r_1 + k, n - r_1 + 1)}{B(r_1, n - r_1 + 1)} = \frac{n! (k + r_1 - 1)!}{(n + k)! (r_1 - 1)!},$$

$$E(U_{r_1:n}) = \frac{n!}{(r_1 - 1)!(n - r_1)!} \int_0^1 u^{r_1} (1 - u)^{n-r_1} du = \frac{r_1}{n + 1},$$

$$V(U_{r_1:n}) = \int_0^1 (u - \mu_{r_1:n})^2 f_{r_1:n}(u) du = \frac{r_1(r_1 + 1)}{(n + 1)(n + 2)} - \frac{r_1^2}{(n + 1)^2},$$

ενώ η ροπή του γινομένου και η συνδιακύμανση των $U_{r_1:n}, U_{r_2:n}$ έχουν τη μορφή

$$E(U_{r_1:n} U_{r_2:n}) = \int_0^1 \int_0^{u_2} u_1 u_2 f_{r_1, r_2:n}(u_1, u_2) du_1 du_2 = \frac{r_1(r_2 + 1)}{(n + 1)(n + 2)},$$

$$\begin{aligned} Cov(U_{r_1:n}, U_{r_2:n}) &= \int_0^1 \int_0^{u_2} (u_1 - \mu_{r_1:n})(u_2 - \mu_{r_2:n}) f_{r_1, r_2:n}(u_1, u_2) du_1 du_2 \\ &= \frac{r_1(r_2 + 1)}{(n + 1)(n + 2)} - \frac{r_1 r_2}{(n + 1)^2}. \end{aligned}$$

Αξιοποιώντας τις προηγούμενες σχέσεις, εύκολα υπολογίζεται ο συντελεστής συσχέτισης δύο ομοιόμορφων διατεταγμένων παρατηρήσεων $\rho(U_{r_1:n}, U_{r_2:n})$. Χρησιμοποιώντας πολυώνυμα του Jacobi (*Jacobi polynomials*), οι Székely and Móri (1985) απέδειξαν το εξής

$$\rho(X_{r_1:n}, X_{r_2:n}) \leq \rho(U_{r_1:n}, U_{r_2:n}) = \sqrt{\frac{r_1(n + 1 - r_2)}{r_2(n + 1 - r_1)}}, \quad 1 \leq r_1 < r_2 \leq n,$$

ενώ μία εναλλακτική απόδειξη προτάθηκε από τους Rohatgi and Székely (1992). Επιπλέον, ο Terrell (1983) [βλ. επίσης Abdelhamid (1985), Papathanasiou (1990)] μέσω ορθογωνίων πολυωνύμων του Legendre (*Legendre orthogonal polynomials*) απέδειξε ότι

$$\rho(X_{1:2}, X_{2:2}) \leq \rho(U_{1:2}, U_{2:2}) = 1/2,$$

έκφραση, η οποία είναι ειδική περίπτωση της προηγούμενης.

2.4 Συμμεταβλητές Διατεταγμένων Παρατηρήσεων

Όπως αναφέραμε και στην Ενότητα 2.2, η πιο σημαντική εφαρμογή της Θεωρίας των Συμμεταβλητών ανακύπτει σε διαδικασίες επιλογής, όπου επιλέγονται k ζεύγη με βάση τις τιμές που λαμβάνει η μεταβλητή X , δεδομένου ότι οι τιμές του συσχετισμένου χαρακτηριστικού Y είναι δύσκολο να μετρηθούν ή παρατηρούνται σε μεταγενέστερο χρόνο. Για παράδειγμα, οι Yeο and David (1984) μελέτησαν το πρόβλημα επιλογής των καλύτερων k υποψηφίων (από ένα σύνολο n υποψηφίων) βάσει της βοηθητικής μεταβλητής X , όταν δεν είναι διαθέσιμες οι μετρήσεις του χαρακτηριστικού πρωταρχικού ενδιαφέροντος Y . Σύμφωνα με τον Ke Wang (2008), η μεταβλητή X μπορεί, επίσης, να αντιπροσωπεύει την επίδοση ενός υποψηφίου σε ένα διαγωνισμό και η μεταβλητή Y να μετρά την τελική επίδοση του υποψηφίου ή εναλλακτικά το χαρακτηριστικό X να συμβολίζει τα αποτελέσματα αναζήτησης που προκύπτουν μέσω μίας συγκεκριμένης μηχανής αναζήτησης (π.χ. Google, Yahoo) και το χαρακτηριστικό Y να είναι τα αποτελέσματα που βασίζονται σε μία εκτενέστερη αναζήτηση στο διαδίκτυο.

Σε προβλήματα παρόμοια με αυτά που προαναφέρθηκαν, λύση μπορεί να δοθεί μέσα από τη μελέτη των κατανομών που εμπλέκουν διατεταγμένες παρατηρήσεις και τις συμμεταβλητές τους, καθώς και των ιδιοτήτων αυτών. Εκτενής ανασκόπηση της Θεωρίας Συμμεταβλητών γίνεται στα κείμενα των Bhattacharya (1984), David (1993), David and Nagaraja (2003, 1998). Στη συνέχεια, παρουσιάζονται κάποια αποτελέσματα σχετικά με τις από κοινού και περιθώριες κατανομές των συμμεταβλητών.

Κατ' αρχάς, ας θεωρήσουμε ότι έχουμε στη διάθεσή μας ένα τυχαίο δείγμα αποτελούμενο από ζεύγη παρατηρήσεων της μορφής $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, τα οποία προέρχονται από μία δισδιάστατη κατανομή. Εάν το δείγμα διαταχθεί κατά αύξουσα σειρά μόνο ως προς τη μεταβλητή X , δηλαδή

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n},$$

τότε οι μεταβλητές Y που συνδέονται με τις παραπάνω διατεταγμένες παρατηρήσεις θα συμβολίζονται με

$$Y_{[1:n]}, Y_{[2:n]}, \dots, Y_{[n:n]},$$

και ονομάζονται **συμμεταβλητές** των διατεταγμένων παρατηρήσεων $X_{1:n}, X_{2:n}, \dots, X_{n:n}$, αντίστοιχα. Αξίζει να σημειωθεί ότι στην αγγλική ορολογία έχουν προταθεί δύο εκδοχές σχετικά με την ερμηνεία των συμμεταβλητών. Συγκεκριμένα, ο David (1973) χρησιμοποίησε τον όρο *concomitants* (ο οποίος φαίνεται να έχει επικρατήσει στη διεθνή βιβλιογραφία), ενώ

από τον Bhattacharya (1974) προτάθηκε ο εναλλακτικός όρος *induced order statistics*.

Υπό την προϋπόθεση ότι το δείγμα αποτελείται από n ανεξάρτητα και ισόνομα τυχαία ζεύγη παρατηρήσεων, έστω (X_i, Y_i) , $i = 1, 2, \dots, n$, τα οποία ακολουθούν μία συνεχή α.σ.κ. $F_{X,Y}(x, y) = F(x, y)$ και σ.π.π. $f_{X,Y}(x, y) = f(x, y)$ με αντίστοιχες περιθώριες $F_X(x)$, $F_Y(y)$ και $f_X(x)$, $f_Y(y)$, συνεπάγεται ότι

$$f_{Y_{[r:n]}|X_{r:n}}(y | X_{r:n} = x) = f_{Y|X}(y | x) = \frac{f(x, y)}{f_X(x)}.$$

Επομένως, η από κοινού σ.π.π. του ζεύγους $(X_{r:n}, Y_{[r:n]})$ θα δίνεται από τον τύπο

$$f_{X_{r:n}, Y_{[r:n]}}(x, y) = f_{Y|X}(y | x) f_{r:n}(x),$$

όπου η $f_{r:n}(x)$ είναι η πυκνότητα της r -οστής διατεταγμένης παρατήρησης. Χρησιμοποιώντας στη συνέχεια τον τύπο της $f_{r:n}(x)$, ο οποίος προσδιορίζεται στη Σχέση (2.8), μπορούμε απευθείας να συμπεράνουμε το εξής (βλ. David (1973), Yang (1977))

$$f_{X_{r:n}, Y_{[r:n]}}(x, y) = \frac{n!}{(r-1)!(n-r)!} [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} f(x, y), \quad x, y \in \mathbb{R}. \quad (2.12)$$

Σύμφωνα με τον Yang (1977), στην περίπτωση της από κοινού πυκνότητας των $(X_{r:n}, Y_{[s:n]})$ με $s < r$ ισχύει

$$f_{X_{r:n}, Y_{[s:n]}}(x, y) = \int_{-\infty}^x f_{Y|X}(y | t) f_{s,r:n}(t, x) dt, \quad x, y \in \mathbb{R}, \quad (2.13)$$

όπου $f_{s,r:n}(t, x)$ είναι η πυκνότητα που περιγράφεται στη Σχέση (2.9).

Οι Abo-Eleneen and Nagaraja (2002) έδειξαν ότι η Σχέση (2.12) μπορεί άμεσα να γενικευθεί για μία συλλογή από ζεύγη $(X_{r_1:n}, Y_{[r_1:n]})$, $(X_{r_2:n}, Y_{[r_2:n]})$, \dots , $(X_{r_k:n}, Y_{[r_k:n]})$ ως εξής

$$\begin{aligned} & f_{X_{r_1:n}, X_{r_2:n}, \dots, X_{r_k:n}; Y_{[r_1:n]}, Y_{[r_2:n]}, \dots, Y_{[r_k:n]}}(x_1, x_2, \dots, x_k; y_1, y_2, \dots, y_k) = \\ & = \frac{n!}{(r_1-1)!(n-r_k)!} [F_X(x_1)]^{r_1-1} [1 - F_X(x_k)]^{n-r_k} \\ & \times \prod_{i=2}^k \frac{[F_X(x_i) - F_X(x_{i-1})]^{r_i-r_{i-1}-1}}{(r_i - r_{i-1} - 1)!} \prod_{i=1}^k f(x_i, y_i), \end{aligned} \quad (2.14)$$

όπου $1 \leq r_1 < r_2 < \dots < r_k \leq n$ και $x_i, y_i \in \mathbb{R}$.

Δεδομένου ότι η σ.π.π. της συμμεταβλητής $Y_{[r:n]}$ αποτελεί περιθώρια συνάρτηση πυκνό-

τητας της $f_{X_{r:n}, Y_{[s:n]}}(x, y)$ μπορεί να εκφρασθεί ως εξής

$$f_{[r:n]}(y) = f_{Y_{[r:n]}}(y) = \int_{-\infty}^{\infty} f_{Y|X}(y | x) f_{r:n}(x) dx, \quad (2.15)$$

ενώ γενικότερα για την από κοινού σ.π.π. των συμμεταβλητών $Y_{[r_1:n]}, Y_{[r_2:n]}, \dots, Y_{[r_k:n]}$ ισχύει

$$\begin{aligned} & f_{Y_{[r_1:n]}, Y_{[r_2:n]}, \dots, Y_{[r_k:n]}}(y_1, y_2, \dots, y_k) = \\ & = \int_{-\infty}^{\infty} \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_2} \prod_{i=1}^k f_{Y|X}(y_i | x_i) f_{r_1, r_2, \dots, r_k:n}(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k. \end{aligned} \quad (2.16)$$

με $1 \leq r_1 < r_2 < \dots < r_k \leq n$ (βλ. Yang (1977)). Η τελευταία σχέση βασίζεται ουσιαστικά στο γεγονός ότι τα $Y_{[r_1:n]}, Y_{[r_2:n]}, \dots, Y_{[r_k:n]}$ είναι υπό συνθήκη ανεξάρτητα δοθέντος ότι (*conditionally independent given*) $X_{r_1:n} = x_1, X_{r_2:n} = x_2, \dots, X_{r_k:n} = x_k$, συμβ.

$$f_{Y_{[r_1:n]}, Y_{[r_2:n]}, \dots, Y_{[r_k:n]} | X_{r_1:n}=x_1, X_{r_2:n}=x_2, \dots, X_{r_k:n}=x_k}(y_1, y_2, \dots, y_k) = \prod_{i=1}^k f_{Y|X}(y_i | x_i),$$

ένα αποτέλεσμα που απέδειξε ο Bhattacharya (1974). Είναι περιττό να αναφερθεί ότι η Σχέση (2.4) αποτελεί γενίκευση της Σχέσης (2.15) για $1 \leq r_1 < r_2 < \dots < r_k \leq n$.

Από τις Σχέσεις (2.13) και (2.4) συνεπάγονται τα ακόλουθα (Yang (1977))

$$E(Y_{[r:n]}) = E(m(X_{r:n})), \quad (2.17\alpha)$$

$$V(Y_{[r:n]}) = V(m(X_{r:n})) + E(\sigma^2(X_{r:n})), \quad (2.17\beta)$$

$$\text{Cov}(X_{r:n}, Y_{[s:n]}) = \text{Cov}(X_{r:n}, m(X_{s:n})), \quad (2.17\gamma)$$

$$\text{Cov}(Y_{[r:n]}, Y_{[s:n]}) = \text{Cov}(m(X_{r:n}), m(X_{s:n})), \quad r \neq s. \quad (2.17\delta)$$

όπου $m(x) = E(Y | X = x)$ και $\sigma^2(x) = V(Y | X = x)$ [βλ. Bhattacharya (1974)].

Οι David *et al.* (1977) [βλ. επίσης Yang (1977)] μελέτησαν το βαθμό (*rank*) της συμμεταβλητής $Y_{[r:n]}$, συμβ.

$$R_{r,n} = R_{Y_{[r:n]}} = \sum_{i=1}^n I(Y_i - Y_{[r:n]}), \quad r = 1, 2, \dots, n,$$

όπου $I(\cdot)$ είναι μία δείκτρια συνάρτηση της μορφής

$$I(u) = \begin{cases} 1, & \text{για } u \geq 0, \\ 0, & \text{διαφορετικά.} \end{cases}$$

Συγκεκριμένα, πρότειναν την επόμενη έκφραση για τη συνάρτηση πυκνότητας του $R_{r,n}$

$$P(R_{r,n} = s) = n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=\max(0, r+s-n-1)}^{\min(r-1, s-1)} \frac{(n-1)!}{k!(r-1-k)!(s-1-k)!(n-r-s+1+k)!} \\ \times \theta_1^k \theta_2^{r-1-k} \theta_3^{s-1-k} \theta_4^{n-r-s+1+k} f(x, y) dx dy$$

και τον ακόλουθο τύπο υπολογισμού της αναμενόμενης τιμής του $R_{r,n}$

$$E(R_{r,n}) = 1 + n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta_1 f_{Y|X}(y | x) dy f_{r-1;n-1}(x) dx + n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta_3 f_{Y|X}(y | x) dy f_{r;n-1}(x) dx,$$

όπου $\theta_i = \theta_i(x, y)$, $i = 1, 2, 3, 4$ και

$$\theta_1(x, y) = P(X \leq x, Y \leq y), \quad \theta_2(x, y) = P(X \leq x, Y > y),$$

$$\theta_3(x, y) = P(X > x, Y \leq y), \quad \theta_4(x, y) = P(X > x, Y > y).$$

Θα πρέπει να τονισθεί ότι η βασική υπόθεση που γίνεται σε όλα τα παραπάνω αποτελέσματα είναι ότι το δείγμα αποτελείται από n ανεξάρτητα και ισόνομα ζεύγη παρατηρήσεων (X_i, Y_i) , $i = 1, 2, \dots, n$. Ωστόσο, ο Eryilmaz (2005) εξέτασε την περίπτωση όπου τα ζεύγη (X_i, Y_i) είναι ανεξάρτητα, αλλά όχι απαραίτητα ισόνομα. Η γενίκευση της Θεωρίας των Συμμεταβλητών σε περισσότερες διαστάσεις είναι, επίσης, ένα ζήτημα που κέντρισε το ενδιαφέρον αρκετών ερευνητών, όπως για παράδειγμα των Song *et al.* (1992), Song and Deddens (1993), Balakrishnan (1993), ενώ οι ασυμπτωτικές κατανομές και οι κατανομές των ακραίων τιμών των συμμεταβλητών έχουν μελετηθεί από τους David and Galambos (1974), David (1994).

Αξίζει να σημειωθεί ότι εξίσου σημαντικά αποτελέσματα σχετικά με τη Θεωρία των Συμμεταβλητών προκύπτουν όταν υποθέτουμε ότι τα X_i και Y_i ($i = 1, 2, \dots, n$) έχουν μέσες τιμές μ_X, μ_Y , διακυμάνσεις σ_X^2, σ_Y^2 και συνδέονται μέσω του ακόλουθου γραμμικού μοντέλου παλινδρόμησης ($|\rho| < 1$)

$$Y_i = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X_i - \mu_X) + \epsilon_i,$$

όπου τα X_i , ϵ_i είναι ανεξάρτητα. Από το παραπάνω μοντέλο συνεπάγεται ότι $E(\epsilon_i) = 0$, $V(\epsilon_i) = \sigma_Y^2(1 - \rho^2)$, $\rho = Corr(X, Y)$. Εάν επιπλέον τα X_i , ϵ_i έχουν κανονική κατανομή, τότε το ζευγάρι (X, Y) ακολουθεί τη διδιάστατη κανονική κατανομή με παραμέτρους

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \quad \text{και} \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

Διατάσσοντας ως προς τα X_i έχουμε

$$Y_{[r:n]} = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X_{r:n} - \mu_X) + \epsilon_{[r]}, \quad r = 1, 2, \dots, n,$$

όπου $\epsilon_{[r]}$ υποδηλώνει το συγκεκριμένο ϵ_i που σχετίζεται με την τ.μ. $X_{r:n}$. Επομένως, θέτοντας

$$\alpha_{r:n} = E\left(\frac{X_{r:n} - \mu_X}{\sigma_X}\right) \quad \text{και} \quad \beta_{r,s:n} = \text{Cov}\left(\frac{X_{r:n} - \mu_X}{\sigma_X}, \frac{X_{s:n} - \mu_X}{\sigma_X}\right),$$

προκύπτουν τα εξής

$$E(Y_{[r:n]}) = \mu_Y + \rho \sigma_Y \alpha_{r:n}, \quad (2.18\alpha)$$

$$V(Y_{[r:n]}) = \sigma_Y^2 (\rho^2 \beta_{r,r:n} + 1 - \rho^2), \quad (2.18\beta)$$

$$\text{Cov}(X_{r:n}, Y_{[s:n]}) = \rho \sigma_X \sigma_Y \beta_{r,s:n}, \quad (2.18\gamma)$$

$$\text{Cov}(Y_{[r:n]}, Y_{[s:n]}) = \rho^2 \sigma_Y^2 \beta_{r,s:n}, \quad r \neq s. \quad (2.18\delta)$$

για $r, s = 1, 2, \dots, n$. Προφανώς, οι Σχέσεις (2.17α-δ) είναι γενικεύσεις των αντίστοιχων Σχέσεων (2.18α-δ).

2.5 Σύνδεσμοι

Τα τελευταία χρόνια παρατηρείται έντονη ερευνητική δραστηριότητα σχετικά με τη μελέτη της Θεωρίας των Συνδέσμων, εκτενής μελέτη της οποίας γίνεται στα βιβλία των Joe (1997, Κεφ.5), Nelsen (2006), Balakrishnan and Lai (2009, Κεφ.1). Οι σύνδεσμοι ή συζεύξεις (*copulas*) βρίσκουν πρακτική εφαρμογή κυρίως στα Χρηματοοικονομικά (Cherubini *et al.* (2004), Cherubini *et al.* (2012)), αλλά και σε άλλες περιοχές όπως η υδρολογία (Favre *et al.* (2004), Salvadori and De Michele (2007)), η αξιοπιστία (Eryilmaz (2011, 2014), Achcar and Moala (2015), Zhang and Wilson (2017)) και ο έλεγχος ποιότητας (Fatahi *et al.* (2012), Dokouhaki and Noorossana (2013), Verdier (2013), Long and Emura (2014), Busababodhin and Amphanthong (2016), Emura *et al.* (2016), Koutras and Sofikitou (2017a,b,c)).

Παράλληλα, ιδιαίτερα σημαντική κρίνεται η συνεισφορά των συνδέσμων στην κατασκευή νέων πολυδιάστατων κατανομών με συγκεκριμένη δομή εξάρτησης δοθεισών των περιθώριων κατανομών. Ουσιαστικά, οι σύνδεσμοι αποτελούν συναρτήσεις, οι οποίες «συνδέουν» (“*couple*”) τις πολυδιάστατες κατανομές με τις μονοδιάστατες περιθώριες κατανομές τους και χρησιμοποιούνται για να περιγράψουν τη δομή εξάρτησης των εμπλεκόμενων μεταβλητών. Ισοδύναμα, μπορούμε να ισχυριστούμε ότι οι σύνδεσμοι είναι πολυδιάστατες κατανομές, των οποίων οι περιθώριες είναι Ομοιόμορφες στο διάστημα

(0, 1). Η αυστηρή διατύπωση δίνεται στο παρακάτω θεώρημα.

Θεώρημα 2.1 (Θεώρημα του Sklar, 1959). Έστω H η από κοινού α.σ.κ. δύο τ.μ. X και Y με περιθώριες $F_X(x)$, $F_Y(y)$, αντίστοιχα. Τότε, υπάρχει σύνδεσμος C τέτοιος, ώστε για όλα τα $x, y \in [-\infty, \infty]$, να ισχύει το εξής

$$H(x, y) = C(F_X(x), F_Y(y)). \quad (2.19)$$

Εάν οι $F_X(x)$ και $F_Y(y)$ είναι συνεχείς, τότε ο σύνδεσμος είναι μοναδικός. Σε διαφορετική περίπτωση, ο C είναι μοναδικά ορισμένος στο $\text{Range}(F_X) \times \text{Range}(F_Y)$. Αντίστροφα, εάν C είναι ένας σύνδεσμος και οι $F_X(x)$, $F_Y(y)$ είναι α.σ.κ., τότε η συνάρτηση H είναι μία από κοινού α.σ.κ. με περιθώριες $F_X(x)$, $F_Y(y)$.

Από τα παραπάνω γίνεται εμφανές ότι εάν οι περιθώριες $F_X(x)$, $F_Y(y)$ είναι Ομοιόμορφες, τότε η Σχέση (2.19) παίρνει τη μορφή $H(x, y) = C(x, y)$. Αυτό πρακτικά σημαίνει ότι αναπαριστά μία δισδιάστατη κατανομή της οποίας οι περιθώριες έχουν μετασχηματιστεί κατάλληλα, ώστε να είναι Ομοιόμορφες στο διάστημα (0, 1). Με άλλα λόγια, ένας δισδιάστατος σύνδεσμος είναι απλά η ομοιόμορφη αναπαράσταση της υπό μελέτη δισδιάστατης κατανομής. Επιπλέον, εάν συμβολίσουμε με $h(x, y)$, $f_X(x)$, $f_Y(y)$ και c τις συναρτήσεις πυκνότητας που αντιστοιχούν στις $H(x, y)$, $F_X(x)$, $F_Y(y)$ και C , από τη Σχέση (2.19) συνεπάγεται ότι

$$h(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y). \quad (2.20)$$

Επιπλέον, μπορεί κανείς εύκολα να κατασκευάσει ένα σύνδεσμο επιβίωσης (*survival copula*, *complementary copula*) $\hat{C}(u, v)$ εάν αντικαταστήσει τα C , u , v με \hat{C} , $1 - u$, $1 - v$, αντίστοιχα (Drouet-Mari and Kotz (2001), Nelsen (2006)). Σε αυτή την περίπτωση, ισχύουν τα εξής

$$\bar{H}(x, y) = \hat{C}(\bar{F}_X(x), \bar{F}_Y(y))$$

και

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

Αξίζει να τονιστεί ότι ο σύνδεσμος \hat{C} δεν αποτελεί συνάρτηση επιβίωσης. Αν συμβολίσουμε με το σύνδεσμο \bar{C} την από κοινού συνάρτηση επιβίωσης δύο ομοιόμορφων μεταβλητών, των οποίων η από κοινού κατανομή είναι ο σύνδεσμος C , τότε ο σύνδεσμος \bar{C} υπολογίζεται

από τη σχέση

$$\bar{C}(u, v) = 1 - u - v + C(u, v) = \hat{C}(1 - u, 1 - v).$$

Όλοι οι σύνδεσμοι πληρούν κάποιες βασικές ιδιότητες. Οι πιο σημαντικές από αυτές δίνονται παρακάτω:

- Για κάθε $u, v \in (0, 1)$ ισχύουν τα εξής

$$C(u, 0) = 0 = C(0, v) \quad \text{και} \quad C(u, 1) = u, \quad C(1, v) = v. \quad (2.21)$$

- Για κάθε $u_1, u_2, v_1, v_2 \in [0, 1] \times [0, 1]$ με $u_1 \leq u_2$ και $v_1 \leq v_2$ ισχύει

$$V_C([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0,$$

όπου η συνάρτηση V_C καλείται C -όγκος (C -volume) του ορθογωνίου $[u_1, u_2] \times [v_1, v_2]$ και είναι 2-αύξουσα (2 -increasing), καθώς δεν παίρνει αρνητικές τιμές. Αξίζει να σημειωθεί ότι ο σύνδεσμος C εκφράζει ένα μέτρο πιθανότητας στον χώρο $[0, 1] \times [0, 1]$, διότι $V_C([0, u] \times [0, v]) = C(u, v)$.

- Ο σύνδεσμος είναι ομοιόμορφα συνεχής συνάρτηση και ως αποτέλεσμα ικανοποιεί την πιο ισχυρή συνθήκη του Lipschitz, δηλαδή για $0 \leq u_1, u_2, v_1, v_2 \leq 1$ ισχύει (βλ. Balakrishnan and Lai (1983))

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

- Για κάθε σύνδεσμο C και για κάθε $u, v \in [0, 1] \times [0, 1]$ ισχύει

$$C^-(u, v) \leq C(u, v) \leq C^+(u, v),$$

όπου $C^+(u, v) = \min(u, v)$ και $C^-(u, v) = \max(u + v - 1, 0)$ είναι αντίστοιχα τα λεγόμενα άνω και κάτω Fréchet-Hoeffding όρια.

- Για κάθε $u, v \in [0, 1]$, οι μερικές παράγωγοι $\partial C(u, v)/\partial u$ και $\partial C(u, v)/\partial v$ υπάρχουν για σχεδόν όλα τα u, v και ικανοποιούν τη συνθήκη

$$0 \leq \frac{\partial C(u, v)}{\partial u}, \frac{\partial C(u, v)}{\partial v} \leq 1.$$

- Το $C^0(u, v) = uv$ είναι ο **σύνδεσμος της ανεξαρτησίας** (*independence copula*), καθώς συνδέεται με το ζευγάρι (U, V) δύο ανεξάρτητων παρατηρήσεων.
- Ένας κυρτός (*convex*) συνδυασμός δύο συνδέσμων C_1 και C_2 είναι επίσης σύνδεσμος. Για παράδειγμα,

$$C_a(u, v) = a C^+(u, v) + (1 - a) C^-(u, v), \quad 0 \leq a \leq 1,$$

$$C_{a,\beta}(u, v) = a C^+(u, v) + \beta C^-(u, v) + (1 - a - \beta) C_0(u, v).$$

Η τελευταία σχέση είναι γνωστή ως σύνδεσμος του Fréchet (Fréchet (1958)).

Υπάρχουν, ακόμη, περιπτώσεις στις οποίες η κατασκευή των συνδέσμων βασίζεται στην επόμενη συνάρτηση

$$\varphi(H(x, y)) = \varphi(F_X(x)) + \varphi(F_Y(y))$$

ή ισοδύναμα

$$\varphi(C(u, v)) = \varphi(u) + \varphi(v).$$

Για την ακρίβεια, μπορεί να κατασκευασθεί μία ολόκληρη οικογένεια συνδέσμων, γνωστοί ως **Αρχιμήδειοι σύνδεσμοι** (*Archimedean Copulas*), ορίζοντας κατάλληλα μία αντιστρέψιμη συνάρτηση φ και ορίζοντας το σύνδεσμο μέσω της ισότητας

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)).$$

Οι **Ελλειπτικοί σύνδεσμοι** (*Elliptical Copulas*), οι οποίοι συνδέονται με τις ελλειπτικές κατανομές (όπως η Κανονική Κατανομή ή η Κατανομή *t-Student*), αποτελούν άλλη μια γνωστή κατηγορία. Η χαρακτηριστική ιδιότητα της οικογένειας αυτής είναι η ακτινική συμμετρία. Ειδικότερα, ένας σύνδεσμος θεωρείται ακτινικά συμμετρικός, εάν για κάθε $(u, v) \in [0, 1]^2$ ικανοποιείται η σχέση

$$C(u, v) = \hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

Οι Drouet-Mari and Kotz (2001), βασιζόμενοι στη μέθοδο Rüschendorf, πρότειναν την κατασκευή πολυωνυμικών συνδέσμων. Πιο συγκεκριμένα, **ένας πολυωνυμικός σύνδεσμος m -τάξης** (*polynomial copula of power m*) περιγράφεται από τους παρακάτω τύπους

$$c(u, v) = 1 + \sum_{k \geq 1, q \geq 1} \theta_{kq} \left(u^k - \frac{1}{k+1} \right) \left(v^q - \frac{1}{q+1} \right),$$

$$C(u, v) = uv \left[1 + \sum_{k \geq 1, q \geq 1} \frac{\theta_{kq}}{(k+1)(q+1)} (u^k - 1)(v^q - 1) \right],$$

όπου

$$0 \leq \min \left(\sum_{k \geq 1, q \geq 1} \theta_{kq} \frac{q}{(k+1)(q+1)}, \sum_{k \geq 1, q \geq 1} \theta_{kq} \frac{k}{(q+1)(k+1)} \right) \leq 1.$$

Στην πραγματικότητα ένας σύνδεσμος μπορεί να γραφεί προσεγγιστικά στη μορφή ενός πολυωνυμικού συνδέσμου με τη βοήθεια του αναπτύγματος Taylor όπως φαίνεται παρακάτω

$$C_\theta(u, v) \approx C_{\theta_0}(u, v) + \sum_{i=1}^n \frac{C_{\theta_0}^{(i)}(u, v)(\theta - \theta_0)^i}{i!},$$

όπου

$$C_{\theta_0}^{(i)}(u, v) = \frac{\partial^i}{\partial \theta_0^i} C_{\theta_0}(u, v),$$

εάν η επιλογή του θ_0 είναι τέτοια, ώστε να αντιστοιχεί στην ανεξαρτησία (π.χ. $C_{\theta_0}(u, v) = uv$) και οι διαδοχικές παράγωγοι του C_θ ως προς θ είναι δυνάμεις στο uv .

Τέλος, έχουν μελετηθεί και σύνδεσμοι που σχετίζονται με διατεταγμένες παρατηρήσεις ή ακραίες τιμές (Nelsen (2006)). Ειδικότερα, έστω X_1, X_2, \dots, X_n ένα δείγμα ανεξάρτητων και ισόνομων παρατηρήσεων, τότε ο σύνδεσμος των ακραίων διατεταγμένων παρατηρήσεων $X_{1:n} = \min(X_1, X_2, \dots, X_n)$ και $X_{n:n} = \max(X_1, X_2, \dots, X_n)$ υπολογίζεται ως ακολούθως

$$C_{X_{1:n}, X_{n:n}}(u_1, u_2) = u_2 - [\max\{(1 - u_1)^{1/n} + u_2^{1/n} - 1, 0\}]^n.$$

Στην περίπτωση που έχουμε στη διάθεσή μας ένα δισδιάστατο δείγμα ανεξάρτητων και ισόνομων παρατηρήσεων της μορφής (X_i, Y_i) , $i = 1, 2, \dots, n$, τότε ο σύνδεσμος των ακραίων παρατηρήσεων $X_{n:n} = \max(X_1, X_2, \dots, X_n)$ και $Y_{n:n} = \max(Y_1, Y_2, \dots, Y_n)$ δίνεται από τον τύπο

$$C_{X_{n:n}, Y_{n:n}}(u, v) = C^n(u^{1/n}, v^{1/n}), \quad 0 \leq u, v \leq 1.$$

Ένας σύνδεσμος C_* θεωρείται σύνδεσμος ακραίας τιμής (*extreme value copula*) εάν υπάρχει σύνδεσμος τέτοιος, ώστε να ισχύει το εξής

$$C_*(u, v) = \lim_{n \rightarrow \infty} C^n(u^{1/n}, v^{1/n}), \quad 0 \leq u, v \leq 1.$$

Σε αυτό το σημείο έχει ενδιαφέρον να αναφερθεί ότι ο σύνδεσμος των τ.μ. $U_{r:n} = F_X(X_{r:n})$ και $V_{s:n} = F_Y(Y_{s:n})$ ισούται με το σύνδεσμο των τ.μ. $X_{r:n}$ και $Y_{s:n}$, δηλαδή $C_{X_{r:n}, Y_{s:n}}(u, v) = C_{U_{r:n}, V_{s:n}}(u, v)$. Περισσότερες πληροφορίες σχετικά με το αποτέλεσμα αυτό δίνονται στις εργασίες των Anjous *et al.* (2005) και Anjos *et al.* (2005).

Είναι γεγονός ότι στη βιβλιογραφία έχουν προταθεί πολλοί σύνδεσμοι. Στα πλαίσια, όμως, αυτής της εργασίας θα παρουσιαστούν μόνο οι σύνδεσμοι εκείνοι που θα χρησιμοποιηθούν στα επόμενα κεφάλαια, οι οποίοι δίνονται συγκεντρωτικά στον Πίνακα 2.5.

| Όνομα Συνδέσμου | $C(u, v)$ & $c(u, v)$ | τ & ρ_S | λ_U & λ_L | Γεννήτορας |
|---|--|---|---|--|
| Ali-Mikhail-Haq (Αρχιμήδειος Σύνδεσμος) | $C_{\theta}^{AMH}(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$ $c_{\theta}^{AMH}(u, v) = \frac{1 + \alpha(uv + u + v - 2) + \alpha^2(1-u)(1-v)}{[1 - \alpha(1-u)(1-v)]^3}$ $\theta \in [-1, 1]$ | $\tau = \frac{3\alpha - 2}{3\alpha} - \frac{2(1-\alpha)^2}{3\alpha^2} \ln(1-\alpha)$ $\rho_S = \frac{\alpha^2}{12(1+\alpha)} \operatorname{dli}\ln(1-\alpha) - \frac{24(1-\alpha)}{\alpha^2} \ln(1-\alpha) - \frac{3(12+\alpha)}{\alpha}$ όπου $\operatorname{dli}\ln(1-\alpha) = \int_0^{\alpha} x^{-1} \ln(1-x) dx$ | $\lambda_U = \lambda_L = 0$ | $\varphi(t) = \ln\left(\frac{1-\alpha(1-t)}{t}\right)$ $\varphi^{-1}(t) = \frac{1-\alpha}{e^t - \alpha}$ |
| Clayton-Pareto (Αρχιμήδειος Σύνδεσμος) | $C_{\theta}^{CLP}(u, v) = (u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta}$ $c_{\theta}^{CLP}(u, v) = \theta^{-1}(\theta + 1)(\alpha v)^{-1-1/\theta}(u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta-2}$ $\theta > 0$ | $\tau = \frac{1}{2\theta + 1}$ ρ_S : πολυπλοκή έκφραση | $\lambda_U = 0$ $\lambda_L = 2 - \theta$ | $\varphi(t) = t^{-1/\theta}$ $\varphi^{-1}(t) = (t + 1)^{-\theta}$ |
| Frank (Αρχιμήδειος Σύνδεσμος) | $C_{\alpha}^{FR\lambda}(u, v) = \hat{C}_{\alpha}(u, v) = \log_{\alpha} \left[1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1} \right]$ $c_{\alpha}^{FR\lambda}(u, v) = \frac{(\alpha - 1) \log_{\alpha} \alpha^{uv}}{[\alpha - 1 + (\alpha^u - 1)(\alpha^v - 1)]^2}$ $\alpha \in (0, 1) \cup (1, \infty)$ | $\tau = 1 + \frac{4[D_1(\alpha^*) - 1]}{\alpha^*}$ $\rho_S = 1 + \frac{12[D_2(\alpha^*) - D_1(\alpha^*)]}{\alpha^*}$ όπου $\alpha^* = -\log_{\alpha} \alpha$ & $D_k(\beta) = \frac{k}{\beta^k} \int_0^{\beta} \frac{t^k}{e^t - 1} dt$ | $\lambda_U = \lambda_L = 0$ | $\varphi(t) = \log\left(\frac{1-\alpha^t}{1-\alpha}\right)$ $\varphi^{-1}(t) = \log_{\alpha}(1 + \alpha^{1-t} - \alpha^{-t})$ |
| Gaussian (Ελλειπτικός Σύνδεσμος) | $C_{\rho}^{GAU}(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v))$ $= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right\} ds dt$ $c_{\rho}^{GAU}(u, v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2}{2(1-\rho^2)} [(\Phi^{-1}(u))^2 + (\Phi^{-1}(v))^2] + \frac{\rho}{1-\rho^2} \Phi^{-1}(u)\Phi^{-1}(v)\right\}$ $\rho \in (0, 1)$ | $\tau = \frac{2}{\pi} \sin^{-1}(\rho) = \frac{2}{\pi} \arcsin(\rho)$ $\rho_S = \frac{6}{\pi} \sin^{-1}\left(\frac{\rho}{2}\right) = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right)$ | $\lambda_U = \lambda_L = 0$ | |
| Gumbel-Barnett (Αρχιμήδειος Σύνδεσμος) | $C_{\theta}^{GUB}(u, v) = uv e^{-\theta \ln uv}$ $c_{\theta}^{GUB}(u, v) = (-\theta + (1 - \theta \ln u)(1 - \theta \ln v)) e^{-\theta \ln uv}$ $\theta \in [0, 1]$ | $\tau = e^{2/\theta} \operatorname{Ei}(-2/\theta)$ $\rho_S = -3 - [12e^{1/\theta} \operatorname{Ei}(-4/\theta)]/\theta$ όπου $\operatorname{Ei}(x) = \int_{-\infty}^x (e^t/t) dx$ | $\lambda_U = \lambda_L = 0$ | $\varphi(t) = \ln(1 - \theta \ln t)$ $\varphi^{-1}(t) = e^{(1-e^t)/\theta}$ |
| Gumbel-Hougaard (Αρχιμήδειος Σύνδεσμος & Σύνδεσμος Ακράτων Τιμών) | $C_{\alpha}^{GUH}(u, v) = \exp(-[-(\ln u)^{\alpha} + (-\ln v)^{\alpha}]^{1/\alpha})$ $c_{\alpha}^{GUH}(u, v)$: πολυπλοκή έκφραση $\alpha \geq 1$ | $\tau = \frac{\alpha - 1}{\alpha}$ ρ_S : δεν υπάρχει κλειστή μορφή | $\lambda_U = 2 - 2^{1/\alpha}$ $\lambda_L = 0$ | $\varphi(t) = (-\ln t)^{\alpha}$ $\varphi^{-1}(t) = \exp(-t^{1/\alpha})$ |
| Farlie-Gumbel-Morgenstern (Πολυωνυμικός Σύνδεσμος Τάξης $m = 4$) | $C_{\theta}^{FGM}(u, v) = \hat{C}_{\theta}(u, v) = uv[1 + \theta(1-u)(1-v)]$ $c_{\theta}^{FGM}(u, v) = 1 + \theta(1-2u)(1-2v)$ $\theta \in [-1, 1]$ | $\tau = 2\theta/9$ $\rho_S = \theta/3$ | $\lambda_U = \lambda_L = 0$ | |

Πίνακας 2.5: Διοδίστατοι Σύνδεσμοι

Πληροφορία σχετικά με το βαθμό εξάρτησης μεταξύ μεταβλητών με συνεχή α.σ.κ., μπορεί να δοθεί μέσω των συντελεστών συσχέτισης τάξεων (*rank correlation coefficients*). Σε αυτή την κατηγορία ανήκουν οι συντελεστές ρ_S του *Spearman* και το τ του *Kendall*, οι οποίοι βασίζονται στις συμφωνίες-ασυμφωνίες των τ.μ. και δεν εξαρτώνται από τις περιθώριες κατανομές (Nelsen (2006), Fredricks and Nelsen (2007)).

Πιο συγκεκριμένα, ας θεωρήσουμε ότι $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ είναι ανεξάρτητα ζεύγη παρατηρήσεων που το καθένα έχει από κοινού α.σ.κ. H και περιθώριες $F_X(x), F_Y(y)$. Τότε, οι παραπάνω συντελεστές εκφράζονται μέσω των πιθανοτήτων που ακολουθούν

$$\begin{aligned}\tau &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ \rho_S &= 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0])\end{aligned}$$

ή με τη βοήθεια του αντίστοιχου συνδέσμου C ως εξής

$$\begin{aligned}\tau &= 4E(C(u, v)) - 1 = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) dudv - 1 \\ \rho_S &= 12E(UV) - 3 = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3 = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv.^1\end{aligned}$$

όπου $-1 \leq \tau, \rho_S \leq 1$ και $-1 \leq 3\tau - 2\rho_S \leq 1$. Στην περίπτωση που το C ανήκει στην οικογένεια των Αρχιμήδειων Συνδέσμων, ο τύπος για το συντελεστή συσχέτισης τ του *Kendall* υπολογίζεται ως εξής

$$\tau = 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt + 1.$$

Τα παραπάνω μέτρα συσχέτισης έχουν σχεδιαστεί για να εξετάζουν κατά πόσο μεγάλες (αντίστοιχα μικρές) τιμές της μίας τ.μ. εμφανίζονται σε συνδυασμό με μεγάλες (αντίστοιχα μικρές) τιμές της άλλης με μεγάλη πιθανότητα. Ωστόσο, πολλές φορές ενδιαφερόμαστε για την εξάρτηση των μεταβλητών στο άνω-δεξί και κάτω-αριστερό τεταρτημόριο του διαστήματος $[0, 1] \times [0, 1]$, δηλαδή για την εξάρτηση στα άκρα-στις ουρές.

Ο δείκτης εξάρτησης της άνω ουράς (*upper tail dependence*) λ_U (αντίστοιχα κάτω ουράς (*lower tail dependence*) λ_L) είναι το όριο -εφόσον υπάρχει- της δεσμευμένης πιθανότητας ότι το Y είναι μεγαλύτερο ή ίσο (αντίστοιχα μικρότερο ή ίσο) του $100t$ -οστού ποσοστημορίου της F_Y , δεδομένου ότι το X είναι μεγαλύτερο ή ίσο (αντίστοιχα μικρότερο ή ίσο) του $100t$ -οστού ποσοστημορίου της F_X , καθώς το t τείνει στο 1 (αντίστοιχα τείνει στο 0), δηλαδή

² Γράφοντας το ρ του *Spearman* στη μορφή $\rho = (E(UV) - 1/4)/(1/12)$, εύκολα μπορεί να παρατηρήσει κανείς ότι $\rho_S(X, Y) = \rho(U, V)$, όπου ρ είναι ο συντελεστής συσχέτισης του *Pearson*.

$$\begin{aligned}
 \lambda_U &= \lim_{t \rightarrow 1^-} P[Y > F_Y^{-1}(t) \mid X > F_X^{-1}(t)] = \lim_{t \rightarrow 1^-} P[F_Y(Y) > t \mid F_X(X) > t] \\
 &= \lim_{t \rightarrow 1^-} \frac{P[F_Y(Y) > t, F_X(X) > t]}{P[F_X(X) > t]} = \lim_{t \rightarrow 1^-} \frac{\bar{C}(t, t)}{1 - t}, \\
 \lambda_L &= \lim_{t \rightarrow 0^+} P[Y \leq F_Y^{-1}(t) \mid X \leq F_X^{-1}(t)] = \lim_{t \rightarrow 0^+} P[F_Y(Y) \leq t \mid F_X(X) \leq t] \\
 &= \lim_{t \rightarrow 0^+} \frac{P[F_Y(Y) \leq t, F_X(X) \leq t]}{P[F_X(X) \leq t]} = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}.
 \end{aligned}$$

Οι παραπάνω δείκτες στην περίπτωση Αρχιμήδειου Συνδέσμου παίρνουν τη μορφή

$$\begin{aligned}
 \lambda_U &= 2- \lim_{t \rightarrow 1^-} \frac{1 - \varphi^{-1}(2\varphi(t))}{1 - t} = 2- \lim_{x \rightarrow 0^+} \frac{1 - \varphi^{-1}(2x)}{1 - \varphi^{-1}(x)}, \\
 \lambda_L &= \lim_{t \rightarrow 0^+} \frac{1 - \varphi^{-1}(2\varphi(t))}{t} = \lim_{x \rightarrow \infty} \frac{\varphi^{-1}(2x)}{\varphi^{-1}(x)}.
 \end{aligned}$$

ΚΕΦΑΛΑΙΟ 3

Νέες Δισδιάστατες Κατανομές Διατεταγμένων Παρατηρήσεων

3.1 Εισαγωγή

Στην ενότητα αυτή, παρουσιάζονται νέα αποτελέσματα σχετικά με τις από κοινού α.σ.κ. δύο ή περισσότερων διατεταγμένων παρατηρήσεων που προέρχονται από δύο συσχετισμένα δείγματα. Οι κατανομές αυτές παρουσιάζουν ιδιαίτερο ενδιαφέρον από τη σκοπιά της Θεωρίας Πολυδιάστατων Κατανομών, καθώς μπορούν να μελετηθούν ανεξάρτητα ή να χρησιμοποιηθούν, μεταξύ άλλων, σε πολυδιάστατα διαγράμματα ελέγχου, σύνθετα συστήματα αξιοπιστίας, κ.α. Παρόμοιες κατανομές, οι οποίες βασίζονται σε τετραπλά διωνυμικά μοντέλα, έχουν μελετηθεί πρόσφατα από τους Huang *et al.* (2013), Bayramoglu-Bairamov (2013), Bayramoglu-Bairamov and Kemalbay (2013), Bayramoglu and Bayramoglu-Bairamov (2014), Eryilmaz (2010, 2017), Kemalbay and Bayramoglu-Bairamov (2015).

Αρχικά, αξίζει να σημειωθεί ότι η α.σ.κ. του ζεύγους των (δισδιάστατων) διατεταγμένων παρατηρήσεων $(X_{r:n}, Y_{s:n})$ μπορεί εύκολα να κατασκευασθεί μέσω της δισδιάστατης διωνυμικής κατανομής, η οποία παρουσιάστηκε για πρώτη φορά από τους Aitken and Gonin (1936) σε συνδυασμό με ένα τετραπλό σχήμα δειγματοληψίας (βλ. επίσης Johnson *et al.* (1997)). Συγκεκριμένα, ο τύπος της προαναφερθείσας από κοινού κατανομής λαμβάνει τη μορφή που ακολουθεί

$$F_{r,s:n}(x, y) = P(X_{r:n} \leq x, Y_{s:n} \leq y) \\ = \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!} \theta_1^k \theta_2^{i-k} \theta_3^{j-k} \theta_4^{n-i-j+k}, \quad (3.1)$$

όπου τα $\theta_i = \theta_i(x, y)$, $i = 1, 2, 3, 4$ δίνονται από τους τύπους

$$\begin{aligned}\theta_1(x, y) &= P(X \leq x, Y \leq y), & \theta_2(x, y) &= P(X \leq x, Y > y), \\ \theta_3(x, y) &= P(X > x, Y \leq y), & \theta_4(x, y) &= P(X > x, Y > y),\end{aligned}$$

(David and Nagaraja (2003), Huang *et al.* (2013)).

Θα πρέπει να αναφέρουμε ότι οι Bofinger and Bofinger (1965) και Bofinger (1970) ασχολήθηκαν με τον υπολογισμό του συντελεστή συσχέτισης $\rho(X_{n:n}, Y_{n:n})$ στην περίπτωση της κανονικής ή άλλων κατανομών, αντίστοιχα. Παράλληλα, ο πίνακας διακυμάνσεων-συνδυακυμάνσεων του ζεύγους των διαμέσων εκτιμήθηκε από τον Maritz (1991) και στη συνέχεια επεκτάθηκε στην εκτίμηση του $Cov(X_{r:n}, Y_{s:n})$ από τον Hutson (2000). Επιπλέον, οι ασυμπτωτικές κατανομές διατεταγμένων παρατηρήσεων δειγμάτων που προέρχονται από πολυδιάστατες κατανομές μελετήθηκαν από τον Galambos (1975), ενώ γενικότερα συζήτηση σχετικά με τη διάταξη πολυδιάστατων δεδομένων μπορεί να βρεθεί στην εργασία του Barnett (1976).

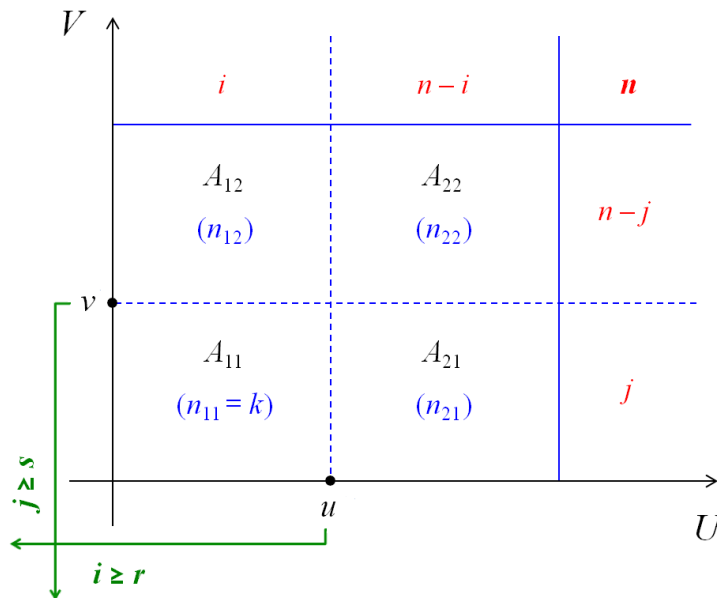
3.2 Δισδιάστατες από κοινού κατανομές διατεταγμένων παρατηρήσεων

Στη συνέχεια, ακολουθώντας παρόμοια λογική με αυτή που εφαρμόζεται για τον υπολογισμό της $F_{r,s:n}(x, y)$, παρουσιάζονται νέες α.σ.κ., οι οποίες εμπλέκουν συνδυασμούς με μία ή περισσότερες διατεταγμένες παρατηρήσεις ανά χαρακτηριστικό. Η βασική υπόθεση που γίνεται είναι ότι έχουμε στη διάθεσή μας ένα τυχαίο δείγμα που αποτελείται από n ανεξάρτητα ζεύγη παρατηρήσεων $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, τα οποία ακολουθούν μία (δισδιάστατη) συνεχή κατανομή με α.σ.κ. $F(x, y)$ και περιθώριες $F_X(x), F_Y(y)$. Προφανώς, οι τ.μ. $U_i = F_X(X_i)$ και $V_j = F_Y(Y_j)$, $i, j = 1, 2, \dots, n$, θα ακολουθούν την Ομοιόμορφη κατανομή στο $(0,1)$. Ως αποτέλεσμα, οι αντίστοιχες διατεταγμένες παρατηρήσεις $U_{i:n} = F_X(X_{i:n})$ και $V_{j:n} = F_Y(Y_{j:n})$ θα ακολουθούν την κατανομή Βήτα, συμβ. $U_{i:n} \sim B(i, n - i + 1)$ και $V_{j:n} \sim B(j, n - j + 1)$.

Προς διευκόλυνση της παρουσίασης των καινούργιων αποτελεσμάτων, θεωρούμε τους Πίνακες 3.1–3.4 σε συνδυασμό με τα αντίστοιχα Σχήματα 3.1–3.4, όπου για κάθε ένα από τα υπό μελέτη χαρακτηριστικά, έστω U και V , ένα ή δύο σημεία αποκοπής (*cut points*) χρησιμοποιούνται με σκοπό τον προσδιορισμό των πιθανοτήτων των συνδυασμένων περιοχών που προκύπτουν ανά περίπτωση. Κάθε πίνακας αντιπροσωπεύει ένα σχήμα δειγματοληψίας τύπου $i \times j$, όπου $i = 2, 3$ και $j = 2, 3$ είναι ο αριθμός των περιοχών που αντιστοιχούν στα χαρακτηριστικά U και V . Οι ποσότητες p_{ij} που εμφανίζονται στους πίνακες, υποδηλώνουν τις πιθανότητες των από κοινού ενδεχομένων A_{ij} . Για παράδειγμα,

εάν λάβουμε υπ' όψιν μας τον Πίνακα 3.1 έχουμε $A_{12} = [U \leq u, V > v]$ και $p_{12} = P(A_{12}) = P(U \leq u, V > v)$. Είναι περιττό να αναφερθεί ότι σε κάθε περίπτωση επαληθεύεται η συνθήκη $\sum_{i,j} p_{ij} = 1$.

Ο τελικός στόχος είναι να εκφραστούν διάφορες πιθανότητες που συνδέονται με τα διαγράμματα ελέγχου, τα οποία παρουσιάζονται αναλυτικά στα επόμενα κεφάλαια, μέσω της πολυωνυμικής κατανομής αξιοποιώντας τους προαναφερθέντες πίνακες και τις αντίστοιχες σχηματικές απεικονίσεις. Επομένως, εάν θεωρήσουμε ότι εκτελούνται n ανεξάρτητα πειράματα, τα αποτελέσματα των οποίων ταξινομούνται σύμφωνα με τα συστήματα δειγματοληψίας που ορίστηκαν, ο συμβολισμός n_{ij} θα χρησιμοποιείται στο εξής για να δηλώσει το πλήθος εμφάνισης των ενδεχομένων A_{ij} μεταξύ των n δοκιμών.



Σχήμα 3.1: Γραφική απεικόνιση του ενδεχομένου $[U_{r:n} \leq u, V_{s:n} \leq v]$

| Ενδεχόμενα | $A_1 : [U \leq u]$ | $A_2 : [U > u]$ | Περιθώριες Πιθανότητες |
|------------------------|-----------------------|-----------------------|------------------------|
| $B_2 : [V > v]$ | $p_{12} = P(A_1 B_2)$ | $p_{22} = P(A_2 B_2)$ | $p_{\bullet 2}$ |
| $B_1 : [V \leq v]$ | $p_{11} = P(A_1 B_1)$ | $p_{21} = P(A_2 B_1)$ | $p_{\bullet 1}$ |
| Περιθώριες Πιθανότητες | $p_{1\bullet}$ | $p_{2\bullet}$ | 1 |

Πίνακας 3.1: Σχήμα Δειγματοληψίας – (2×2) :

Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i, j \in \{1, 2\}$

Λήμμα 3.1. Έστω (X_i, Y_i) , $i = 1, 2, \dots, n$, n ανεξάρτητες τ.μ. που προέρχονται από μία διδιάστατη συνεχή κατανομή με α.σ.κ. $F(x, y)$, περιθώριες $F_X(x)$, $F_Y(y)$ και αντίστοιχο σύνδεσμο $D(u, v)$. Τότε η από κοινού α.σ.κ. και η συνάρτηση επιβίωσης των τ.μ.

$$U_{r:n} = F_X(X_{r:n}) \quad \text{και} \quad V_{s:n} = F_Y(Y_{s:n})$$

εκφράζονται αντίστοιχα ως εξής

$$\begin{aligned} F_{r,s:n}^{(D)}(u, v) &= P(U_{r:n} \leq u, V_{s:n} \leq v) \\ &= \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} a_n(i, j, k) \times q_n^{(D)}(u, v; i, j, k), \end{aligned} \quad (3.2)$$

$$\begin{aligned} \bar{F}_{r,s:n}^{(D)}(u, v) &= P(U_{r:n} > u, V_{s:n} > v) \\ &= \sum_{i=n-r+1}^n \sum_{j=n-s+1}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} a_n(i, j, k) \times w_n^{(D)}(u, v; i, j, k), \end{aligned} \quad (3.3)$$

όπου

$$a_n(i, j, k) = \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!}$$

και

$$\begin{aligned} q_n^{(D)}(i, j, k; u, v) &= [D(u, v)]^k [u - D(u, v)]^{i-k} [v - D(u, v)]^{j-k} [\bar{D}(u, v)]^{n-i-j+k}, \\ w_n^{(D)}(i, j, k; u, v) &= [\bar{D}(u, v)]^k [v - D(u, v)]^{i-k} [u - D(u, v)]^{j-k} [D(u, v)]^{n-i-j+k}, \end{aligned}$$

για $1 \leq r, s \leq n$ και $\bar{D}(u, v) = 1 - u - v + D(u, v)$.

Απόδειξη. Τα αποτελέσματα του συγκεκριμένου λήμματος μπορούν να αποδειχθούν με δύο διαφορετικούς τρόπους.

(Ai) Σύμφωνα με την πρώτη διαδικασία, η από κοινού α.σ.κ., έστω $F_{r,s:n}^{(D)}(u, v)$, μπορεί να παραχθεί άμεσα θεωρώντας το τετραπλό σχήμα δειγματοληψίας του Πίνακα 3.1 ως εξής

$$\begin{aligned} F_{r,s:n}^{(D)}(u, v) &= F_{U_{r:n}, V_{s:n}}^{(D)}(u, v) \\ &= P(U_{r:n} \leq u, V_{s:n} \leq v) \\ &= P(\text{τουλάχιστον } r \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μικρότερες ή ίσες από } u, \\ &\quad \text{τουλάχιστον } s \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μικρότερες ή ίσες από } v) \\ &= \sum_{i=r}^n \sum_{j=s}^n P(\text{το ενδεχόμενο } [U \leq u] \text{ συμβαίνει ακριβώς } i \text{ φορές και} \\ &\quad \text{το ενδεχόμενο } [V \leq v] \text{ συμβαίνει ακριβώς } j \text{ φορές}) \end{aligned}$$

$$= \sum_{i=r}^n \sum_{j=s}^n P \left(\begin{array}{l} \text{το ενδεχόμενο } A_{11}A_{12} \text{ συμβαίνει ακριβώς } i \text{ φορές και} \\ \text{το ενδεχόμενο } A_{11}A_{21} \text{ συμβαίνει ακριβώς } j \text{ φορές.} \end{array} \right)$$

Έχοντας ως δεδομένο ότι τα από κοινού ενδεχόμενα $A_{11} : [U \leq u, V \leq v]$, $A_{12} : [U \leq u, V > v]$, $A_{21} : [U > u, V \leq v]$, $A_{22} : [U > u, V > v]$ έχουν πραγματοποιηθεί n_{11} , n_{12} , n_{21} και n_{22} φορές, αντίστοιχα (βλ. Σχήμα 3.1), δημιουργείται ένα σύστημα με 4 εξισώσεις και 4 αγνώστους της μορφής

$$\begin{cases} n_{11} + n_{21} = j \\ n_{12} + n_{22} = n - j \\ n_{12} + n_{11} = i \\ n_{22} + n_{21} = n - i. \end{cases}$$

Σημειώνεται ότι, αφού το άθροισμα των αριστερών μελών των δύο πρώτων εξισώσεων είναι ίσο με το αντίστοιχο άθροισμα των δύο τελευταίων, η μία από τις εξισώσεις μπορεί να απαλειφθεί, οπότε μένουμε με 3 εξισώσεις και 4 αγνώστους. Επομένως, η λύση του παραπάνω συστήματος εξισώσεων μπορεί να βρεθεί, εάν θέσουμε $n_{11} = k$ και λύσουμε ως προς τους υπόλοιπους αγνώστους, οι οποίοι θα δίνονται από τις σχέσεις $n_{12} = i - k$, $n_{21} = j - k$ και $n_{22} = n - i - j + k$. Οι τέσσερις πιθανότητες που φαίνονται στον Πίνακα 3.1 μπορούν να περιγραφούν με χρήση του συνδέσμου D ως εξής

$$\begin{cases} p_{11} = P(A_{11}) = P(U \leq u, V \leq v) = D(u, v) \\ p_{12} = P(A_{12}) = P(U \leq u) - P(U \leq u, V \leq v) = D(u, 1) - D(u, v) = u - D(u, v) \\ p_{21} = P(A_{21}) = P(V \leq v) - P(U \leq u, V \leq v) = D(1, v) - D(u, v) = v - D(u, v) \\ p_{22} = P(A_{22}) = 1 - P(U \leq u) - P(V \leq v) + P(U \leq u, V \leq v) = 1 - u - v + D(u, v) = \bar{D}(u, v). \end{cases}$$

Συνεπώς, υπό την προϋπόθεση ότι το πείραμα επαναλαμβάνεται n φορές,

- οι k εμφανίσεις του ενδεχομένου A_{11} μπορούν να πραγματοποιηθούν με $\binom{n}{k}$ τρόπους,
- οι $i - k$ εμφανίσεις του ενδεχομένου A_{12} μπορούν να πραγματοποιηθούν με $\binom{n-k}{i-k}$ τρόπους,
- οι $j - k$ εμφανίσεις του ενδεχομένου A_{21} παρατηρούνται με $\binom{n-k-(i-k)}{n-j} = \binom{n-i}{n-j}$ τρόπους και τελικά

- οι $n-i-j+k$ εμφανίσεις του ενδεχομένου A_{22} παρατηρούνται ακριβώς με $\binom{n-i-(j-k)}{n-i-j+k} = \binom{n-i-j+k}{n-i-j+k} = 1$ τρόπο.

Ως αποτέλεσμα, όλες οι δυνατές περιπτώσεις σε n ανεξάρτητες δοκιμές απαριθμούνται όπως φαίνεται παρακάτω

$$\begin{aligned} a_n(i, j, k) &= \binom{n}{k} \binom{n-k}{i-k} \binom{n-i}{n-j} = \binom{n}{k, i-k, j-k, n-i-j+k} \\ &= \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!} = \frac{n!}{\prod_{\tau,\rho=1}^2 n_{\tau\rho}!} \end{aligned}$$

Συνδυάζοντας όλα τα παραπάνω, μπορούμε να συμπεράνουμε ότι η από κοινού α.σ.κ. $F_{r,s;n}^{(D)}(u, v)$ δίνεται από τη σχέση

$$\begin{aligned} F_{r,s;n}^{(D)}(u, v) &= \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} P(\text{τα από κοινού ενδεχόμενα } A_{11}, A_{12}, A_{21}, A_{22} \\ &\quad \text{πραγματοποιούνται ακριβώς } k, i-k, j-k \\ &\quad \text{και } n-i-j+k \text{ φορές, αντίστοιχα}) \\ &= \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} a_n(i, j, k) \times q_n^{(D)}(i, j, k; u, v), \end{aligned}$$

όπου $q_n^{(D)}(i, j, k; u, v) = \prod_{i,j=1}^2 p_{ij}^{n_{ij}} = [D(u, v)]^k [u - D(u, v)]^{i-k} [v - D(u, v)]^{j-k} [\bar{D}(u, v)]^{n-i-j+k}$ και $1 \leq r, s \leq n$.

(Aii) Εναλλακτικά, η από κοινού α.σ.κ. $F_{r,s;n}^{(D)}(u, v)$ μπορεί να παραχθεί έμμεσα με τη βοήθεια της Σχέσης (3.1), η οποία μπορεί να γραφεί στην ακόλουθη μορφή

$$\begin{aligned} F_{r,s;n}(x, y) &= P(X_{r:n} \leq x, Y_{s:n} \leq y) \\ &= \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!} \\ &\quad \times [F(x, y)]^k [F_X(x) - F(x, y)]^{i-k} [F_Y(y) - F(x, y)]^{j-k} [1 - F_X(x) - F_Y(y) + F(x, y)]^{n-i-j+k}. \end{aligned}$$

Στη συνέχεια, παρατηρώντας ότι η από κοινού α.σ.κ. των τ.μ. $U_{r:n}$ και $V_{s:n}$ μπορεί να εκφραστεί ισοδύναμα ως εξής

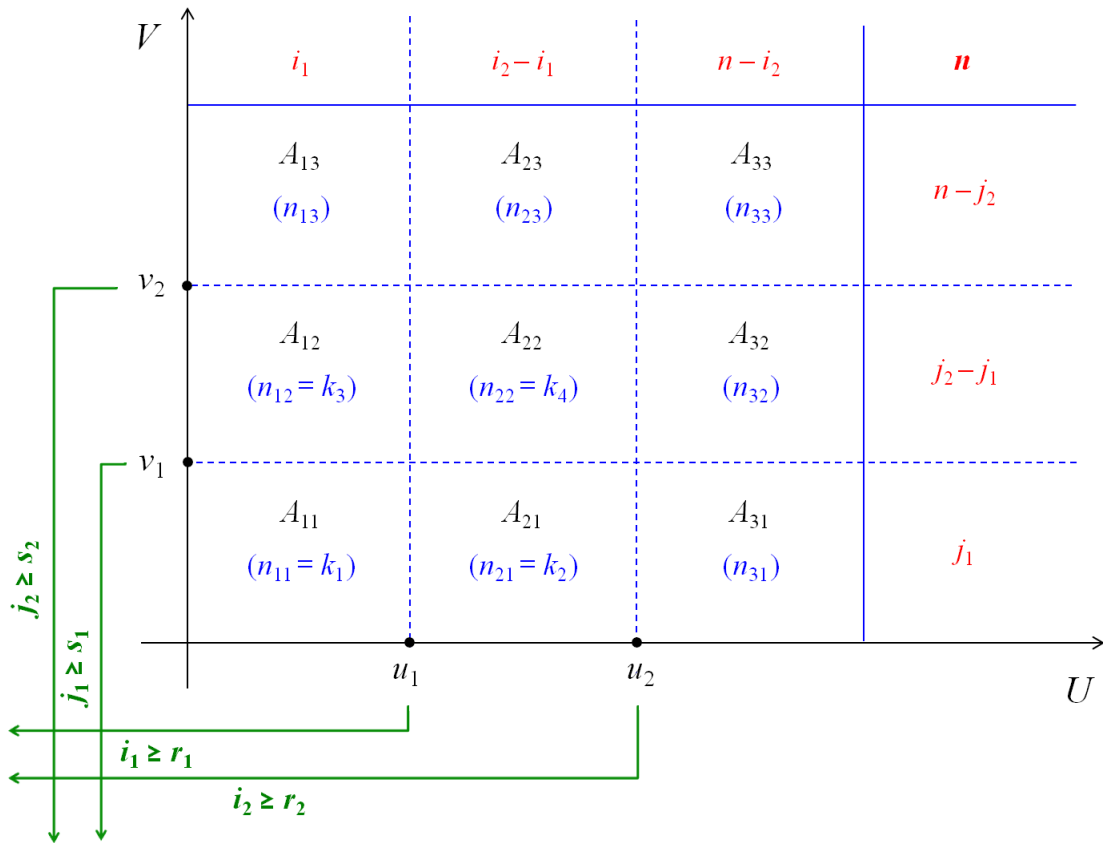
$$\begin{aligned}
 F_{r,s;n}^{(D)}(u, v) &= P(U_{r:n} \leq u, V_{s:n} \leq v) \\
 &= P(F_X(X_{r:n}) \leq u, F_Y(Y_{s:n}) \leq v) \\
 &= P(X_{r:n} \leq F_X^{-1}(u), Y_{s:n} \leq F_Y^{-1}(v)),
 \end{aligned}$$

εύκολα καταλήγουμε στην τελική έκφραση της Σχέσης (3.2), λαμβάνοντας υπ' όψιν ότι $F_X(F_X^{-1}(u)) = u$, $F_Y(F_Y^{-1}(v)) = v$ και $F(F_X^{-1}(u), F_Y^{-1}(v)) = D(F_X(F_X^{-1}(u)), F_Y(F_Y^{-1}(v))) = D(u, v)$.

(B) Όμοια, η συνάρτηση επιβίωσης των τ.μ. $U_{r:n}$ και $V_{s:n}$ ($1 \leq r, s \leq n$) μπορεί να εκφραστεί θεωρώντας το ίδιο σχήμα δειγματοληψίας, αλλά με διαφορετικές πιθανότητες επιτυχίας. Σε αυτή την περίπτωση, η συνάρτηση $\bar{F}_{r,s;n}(u, v)$ λαμβάνει τη μορφή

$$\begin{aligned}
 \bar{F}_{r,s;n}^{(D)}(u, v) &= \bar{F}_{U_{r:n}, V_{s:n}}^{(D)}(u, v) = P(U_{r:n} \geq u, V_{s:n} \geq v) \\
 &= P(\text{τουλάχιστον } n - r + 1 \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μεγαλύτερες ή ίσες από } u, \\
 &\quad \text{τουλάχιστον } n - s + 1 \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μεγαλύτερες ή ίσες από } v) \\
 &= \sum_{i=n-r+1}^n \sum_{j=n-s+1}^n P(\text{ακριβώς } i \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μεγαλύτερες ή ίσες από } u, \\
 &\quad \text{ακριβώς } j \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μεγαλύτερες ή ίσες από } v) \\
 &= \sum_{i=n-r+1}^n \sum_{j=n-s+1}^n P(\text{το ενδεχόμενο } [U \geq u] \text{ συμβαίνει ακριβώς } i \text{ φορές και} \\
 &\quad \text{το ενδεχόμενο } [V \geq v] \text{ συμβαίνει ακριβώς } j \text{ φορές}). \\
 &= \sum_{i=n-r+1}^n \sum_{j=n-s+1}^n P(\text{το ενδεχόμενο } A_{21}A_{22} \text{ συμβαίνει ακριβώς } i \text{ φορές και} \\
 &\quad \text{το ενδεχόμενο } A_{12}A_{22} \text{ συμβαίνει ακριβώς } j \text{ φορές}) \\
 &= \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} P(\text{τα από κοινού ενδεχόμενα } A_{11}, A_{12}, A_{21}, A_{22} \\
 &\quad \text{πραγματοποιούνται ακριβώς } n - i - j + k, j - k, \\
 &\quad i - k \text{ και } k \text{ φορές, αντίστοιχα}) \\
 &= \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} a_n(i, j, k) \times w_n^{(D)}(i, j, k; u, v). \quad \square
 \end{aligned}$$

Στο λήμμα που ακολουθεί αποδεικνύεται έκφραση για την από κοινού α.σ.κ. των διατεταγμένων παρατηρήσεων $U_{r_1:n}, U_{r_2:n}$ και $V_{s_1:n}, V_{s_2:n}$. Για το σκοπό αυτό, στο Σχήμα 3.2 που ακολουθεί γίνεται απεικόνιση του ενδεχομένου $[U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2]$.



Σχήμα 3.2: Γραφική απεικόνιση του ενδεχομένου

$$[U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2]$$

| Ενδεχόμενα | $A_1 : [U \leq u_1]$ | $A_2 : [u_1 < U \leq u_2]$ | $A_3 : [U > u_2]$ | Περιθώριες Πιθανότητες |
|----------------------------|-----------------------|----------------------------|-----------------------|---------------------------|
| $B_3 : [V > v_2]$ | $p_{13} = P(A_1 B_3)$ | $p_{23} = P(A_2 B_3)$ | $p_{33} = P(A_3 B_3)$ | $p_{\bullet 3}$ |
| $B_2 : [v_1 < V \leq v_2]$ | $p_{12} = P(A_1 B_2)$ | $p_{22} = P(A_2 B_2)$ | $p_{32} = P(A_3 B_2)$ | $p_{\bullet 2}$ |
| $B_1 : [V \leq v_1]$ | $p_{11} = P(A_1 B_1)$ | $p_{21} = P(A_2 B_1)$ | $p_{31} = P(A_3 B_1)$ | $p_{\bullet 1}$ |
| Περιθώριες Πιθανότητες | $p_{1\bullet}$ | $p_{2\bullet}$ | $p_{3\bullet}$ | 1 |

Πίνακας 3.2: Σχήμα Δειγματοληψίας – (3 × 3):

Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i, j \in \{1, 2, 3\}$

Λήμμα 3.2. Έστω (X_i, Y_i) , $i = 1, 2, \dots, n$, n ανεξάρτητες τ.μ. που προέρχονται από μία δισδιάστατη συνεχή κατανομή με α.σ.κ. $F(x, y)$, περιθώριες $F_X(x)$, $F_Y(y)$ και αντίστοιχο σύνδεσμο $D(u, v)$. Τότε η από κοινού α.σ.κ. των τ.μ.

$$U_{r_1:n} = F_X(X_{r_1:n}), \quad U_{r_2:n} = F_X(X_{r_2:n}) \quad \text{και} \quad V_{s_1:n} = F_Y(Y_{s_1:n}), \quad V_{s_2:n} = F_Y(Y_{s_2:n})$$

εκφράζεται ως εξής

$$\begin{aligned}
 F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2) &= P(U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2) \\
 &= \sum_{\substack{i_1=r_1 \\ i_1+i_2 \leq n}}^n \sum_{\substack{i_2=r_2 \\ j_1+j_2 \leq n}}^n \sum_{j_1=s_1}^n \sum_{j_2=s_2}^n \sum_{k_1, k_2, k_3, k_4} \frac{n!}{\prod_{i,j=1}^3 n_{ij}!} \prod_{i,j=1}^3 p_{ij}^{n_{ij}}, \quad (3.4)
 \end{aligned}$$

όπου

$$p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1}),$$

για $1 \leq r_1 \leq r_2 \leq n$, $1 \leq s_1 \leq s_2 \leq n$ και $i, j \in \{1, 2, 3\}$ (σύμβαση: $u_0 = v_0 = 0$ και $u_3 = v_3 = 1$).

Απόδειξη. Ο υπολογισμός της από κοινού α.σ.κ. των τεσσάρων διατεταγμένων παρατηρήσεων $U_{r_1:n} = F_X(X_{r_1:n})$, $U_{r_2:n} = F_X(X_{r_2:n})$, $V_{s_1:n} = F_Y(Y_{s_1:n})$, $V_{s_2:n} = F_Y(Y_{s_2:n})$ βασίζεται ουσιαστικά στην ίδια τεχνική με αυτή που ακολουθήθηκε στο Λήμμα 3.1. Λαμβάνοντας υπ' όψιν τον Πίνακα 3.2, παρατηρούμε ότι

$$\begin{aligned}
 F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2) &= F_{U_{r_1:n}, U_{r_2:n}; V_{s_1:n}, V_{s_2:n}}^{(D)}(u_1, u_2; v_1, v_2) \\
 &= P(U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2) \\
 &= P(\text{τουλάχιστον } r_1 \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μικρότερες ή ίσες από } u_1, \\
 &\quad \text{τουλάχιστον } r_2 \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μικρότερες ή ίσες από } u_2, \\
 &\quad \text{τουλάχιστον } s_1 \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μικρότερες ή ίσες από } v_1, \\
 &\quad \text{τουλάχιστον } s_2 \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μικρότερες ή ίσες από } v_2) \\
 &= \sum_{\substack{i_1=r_1 \\ i_1+i_2 \leq n}}^n \sum_{\substack{i_2=r_2 \\ j_1+j_2 \leq n}}^n \sum_{j_1=s_1}^n \sum_{j_2=s_2}^n P(\text{τα ενδεχόμενα } [U \leq u_1], [U \leq u_2], [V \leq v_1], [V \leq v_2] \\
 &\quad \text{πραγματοποιούνται ακριβώς } i_1, i_2, j_1, j_2 \text{ φορές, αντίστοιχα}) \\
 &= \sum_{\substack{i_1=r_1 \\ i_1+i_2 \leq n}}^n \sum_{\substack{i_2=r_2 \\ j_1+j_2 \leq n}}^n \sum_{j_1=s_1}^n \sum_{j_2=s_2}^n \sum_{k_1, k_2, k_3, k_4} P(\text{τα από κοινού ενδεχόμενα } A_{ij} \text{ πραγματοποιούνται} \\
 &\quad \text{ακριβώς } n_{ij} \text{ φορές, αντίστοιχα, για } i, j \in \{1, 2, 3\}).
 \end{aligned}$$

Έχοντας συμβολίσει με n_{ij} τον αριθμό των πραγματοποιήσεων των ενδεχομένων A_{ij} (βλ. Σχήμα 3.2), προκύπτει το ακόλουθο σύστημα 6 εξισώσεων και 9 αγνώστων

$$\left\{ \begin{array}{l} n_{11} + n_{21} + n_{31} = j_1 \\ n_{12} + n_{22} + n_{32} = j_2 - j_1 \\ n_{13} + n_{23} + n_{33} = n - j_2 \\ n_{11} + n_{12} + n_{13} = i_1 \\ n_{21} + n_{22} + n_{23} = i_2 - i_1 \\ n_{31} + n_{32} + n_{33} = n - i_2. \end{array} \right.$$

Και εδώ μπορούμε να παραλείψουμε μία από τις παραπάνω εξισώσεις, οπότε θα έχουμε 5 εξισώσεις με 9 αγνώστους. Για να οδηγηθούμε στη λύση του συστήματος, αρκεί να ορίσουμε/σταθεροποιήσουμε 4 από τα n_{ij} και να λύσουμε ως προς τα υπόλοιπα. Θεωρώντας ότι $n_{11} = k_1, n_{21} = k_2, n_{12} = k_3$ και $n_{22} = k_4$, εύκολα μπορεί να διαπιστώσει κανείς ότι

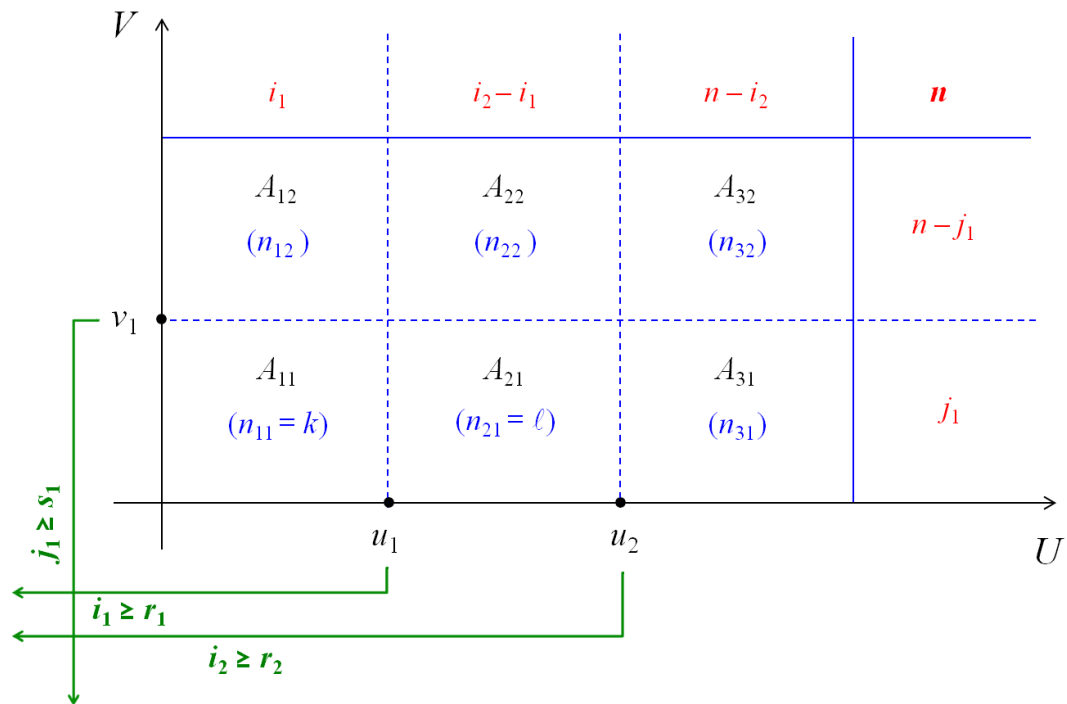
$$\left\{ \begin{array}{l} n_{31} = j_1 - k_1 - k_2 \\ n_{32} = j_2 - j_1 - k_3 - k_4 \\ n_{13} = i_1 - k_1 - k_3 \\ n_{23} = i_2 - i_1 - k_2 - k_4 \\ n_{33} = n - i_2 - j_2 + k_1 + k_2 + k_3 + k_4. \end{array} \right.$$

Επομένως, η από κοινού α.σ.κ. $F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2)$ λαμβάνει την παρακάτω μορφή με τη βοήθεια του πολυωνυμικού μοντέλου

$$\begin{aligned} F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2) &= P(U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2) \\ &= \sum_{\substack{i_1=r_1 \\ i_1+i_2 \leq n}}^n \sum_{\substack{i_2=r_2 \\ j_1+j_2 \leq n}}^n \sum_{j_1=s_1}^n \sum_{j_2=s_2}^n \sum_{k_1, k_2, k_3, k_4} \frac{n!}{\prod_{i,j=1}^3 n_{ij}!} \prod_{i,j=1}^3 p_{ij}^{n_{ij}}, \end{aligned}$$

όπου $p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1})$ για $i, j \in \{1, 2, 3\}$ υπό τη σύμβαση ότι $u_0 = v_0 = 0$ και $u_3 = v_3 = 1$. □

Στο επόμενο λήμμα προσδιορίζεται η από κοινού α.σ.κ. των διατεταγμένων παρατηρήσεων $U_{r_1:n}, U_{r_2:n}$ και $V_{s_1:n}$, η οποία εκφράζεται με τη βοήθεια του ενδεχομένου $[U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1]$ που απεικονίζεται στο Σχήμα 3.3.



Σχήμα 3.3: Γραφική απεικόνιση του ενδεχομένου

$$[U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1]$$

| Ενδεχόμενα | $A_1 : [U \leq u_1]$ | $A_2 : [u_1 < U \leq u_2]$ | $A_3 : [U > u_2]$ | Περιθώριες Πιθανότητες |
|---------------------------|-----------------------|----------------------------|-----------------------|---------------------------|
| $B_2 : [V > v_1]$ | $p_{12} = P(A_1 B_2)$ | $p_{22} = P(A_2 B_2)$ | $p_{32} = P(A_3 B_2)$ | $p_{\bullet 2}$ |
| $B_1 : [V \leq v_1]$ | $p_{11} = P(A_1 B_1)$ | $p_{21} = P(A_2 B_1)$ | $p_{31} = P(A_3 B_1)$ | $p_{\bullet 1}$ |
| Περιθώριες Πιθανότητες | $p_{1\bullet}$ | $p_{2\bullet}$ | $p_{3\bullet}$ | 1 |

Πίνακας 3.3: Σχήμα Δειγματοληψίας – (3 × 2):

Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i \in \{1, 2, 3\}$ και $j \in \{1, 2\}$

Λήμμα 3.3. Έστω (X_i, Y_i) , $i = 1, 2, \dots, n$, n ανεξάρτητες τ.μ. που προέρχονται από μία δισδιάστατη συνεχή κατανομή με α.σ.κ. $F(x, y)$, περιθώριες $F_X(x)$, $F_Y(y)$ και αντίστοιχο σύνδεσμο $D(u, v)$. Τότε η από κοινού α.σ.κ. των τ.μ.

$$U_{r_1:n} = F_X(X_{r_1:n}), \quad U_{r_2:n} = F_X(X_{r_2:n}) \quad \text{και} \quad V_{s_1:n} = F_Y(Y_{s_1:n})$$

εκφράζεται ως εξής

$$\begin{aligned}
 F_{r_1, r_2; s_1; n}^{(D)}(u_1, u_2; v_1) &= P(U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1) \\
 &= \sum_{\substack{i_1=r_1 \\ i_2 \geq i_1}}^n \sum_{i_2=r_2}^n \sum_{j_1=s_1}^n \sum_{k, \ell} \frac{n!}{\prod_{i=1}^3 \prod_{j=1}^2 n_{ij}!} \prod_{i=1}^3 \prod_{j=1}^2 p_{ij}^{n_{ij}}, \quad (3.5)
 \end{aligned}$$

όπου

$$p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1}),$$

για $1 \leq r_1 \leq r_2 \leq n$, $1 \leq s_1 \leq n$ και $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$ (σύμβαση : $u_0 = v_0 = 0$ και $u_3 = v_2 = 1$).

Απόδειξη. Η από κοινού α.σ.κ. των τ.μ. $U_{r_1:n} = F_X(X_{r_1:n})$, $U_{r_2:n} = F_X(X_{r_2:n})$, $V_{s_1:n} = F_Y(Y_{s_1:n})$ είναι στην πραγματικότητα μία περιθώρια συνάρτηση της α.σ.κ. που περιγράφηκε στο Λήμμα 3.2 και ως αποτέλεσμα μπορεί να προκύψει από την έκφραση

$$F_{r_1, r_2; s_1; n}^{(D)}(u_1, u_2; v_1) = \lim_{v_2 \rightarrow 1} F_{r_1, r_2; s_1; s_1; n}^{(D)}(u_1, u_2; v_1, v_2).$$

Παρόλα αυτά, εύκολα μπορεί κανείς να καταλήξει σε μία πιο ελκυστική έκφραση αξιοποιώντας το σχήμα δειγματοληψίας του Πίνακα 3.3 σε συνδυασμό με το πολυωνυμικό θεώρημα. Πιο συγκεκριμένα, η $F_{r_1, r_2; s_1; n}^{(D)}(u_1, u_2; v_1)$ παίρνει τη μορφή

$$\begin{aligned}
 F_{r_1, r_2; s_1; n}^{(D)}(u_1, u_2; v_1) &= F_{U_{r_1:n}, U_{r_2:n}; V_{s_1:n}}^{(D)}(u_1, u_2; v_1) \\
 &= P(U_{r_1:n} \leq u_1, U_{r_2:n} \leq u_2; V_{s_1:n} \leq v_1) \\
 &= P(\text{τουλάχιστον } r_1 \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μικρότερες ή ίσες από } u_1, \\
 &\quad \text{τουλάχιστον } r_2 \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μικρότερες ή ίσες από } u_2, \\
 &\quad \text{τουλάχιστον } s_1 \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μικρότερες ή ίσες από } v_1) \\
 &= \sum_{\substack{i_1=r_1 \\ i_2 \geq i_1}}^n \sum_{i_2=r_2}^n \sum_{j_1=s_1}^n \sum_{k, \ell} P(\text{τα ενδεχόμενα } [U \leq u_1], [U \leq u_2], [V \leq v_1] \\
 &\quad \text{πραγματοποιούνται ακριβώς } i_1, i_2, j_1 \text{ φορές, αντίστοιχα}) \\
 &= \sum_{\substack{i_1=r_1 \\ i_2 \geq i_1}}^n \sum_{i_2=r_2}^n \sum_{j_1=s_1}^n \sum_{k, \ell} P(\text{τα από κοινού ενδεχόμενα } A_{ij} \text{ πραγματοποιούνται} \\
 &\quad \text{ακριβώς } n_{ij} \text{ φορές, αντίστοιχα, για } i \in \{1, 2, 3\}, j \in \{1, 2\}) \\
 &= \sum_{\substack{i_1=r_1 \\ i_2 \geq i_1}}^n \sum_{i_2=r_2}^n \sum_{j_1=s_1}^n \sum_{k, \ell} \frac{n!}{\prod_{i=1}^3 \prod_{j=1}^2 n_{ij}!} \prod_{i=1}^3 \prod_{j=1}^2 p_{ij}^{n_{ij}},
 \end{aligned}$$

όπου $p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1})$ για $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$ υπό τη σύμβαση ότι $u_0 = v_0 = 0$ και $u_3 = v_2 = 1$. Σε αυτή την περίπτωση (βλ. Σχήμα 3.3), δημιουργείται το ακόλουθο σύστημα με 5 εξισώσεις και 6 αγνώστους

$$\begin{cases} n_{11} + n_{21} + n_{31} = j_1 \\ n_{12} + n_{22} + n_{32} = n - j_1 \\ n_{11} + n_{12} = i_1 \\ n_{21} + n_{22} = i_2 - i_1 \\ n_{31} + n_{32} = n - i_2, \end{cases}$$

από το οποίο μπορούμε να απαλείψουμε μία εξίσωση και θα έχουμε πλέον 4 εξισώσεις με 6 αγνώστους. Οι ακέραιοι n_{ij} που εμφανίζονται στη Σχέση (3.5) δίνονται από την επόμενη έκφραση, η οποία προκύπτει θέτοντας $n_{11} = k$, $n_{21} = \ell$ και λύνοντας το σύστημα για τα υπόλοιπα n_{ij}

$$\begin{cases} n_{12} = i_1 - k \\ n_{22} = i_2 - i_1 - \ell \\ n_{31} = j_1 - k - \ell \\ n_{32} = n - i_2 - j_1 + k + \ell. \end{cases}$$

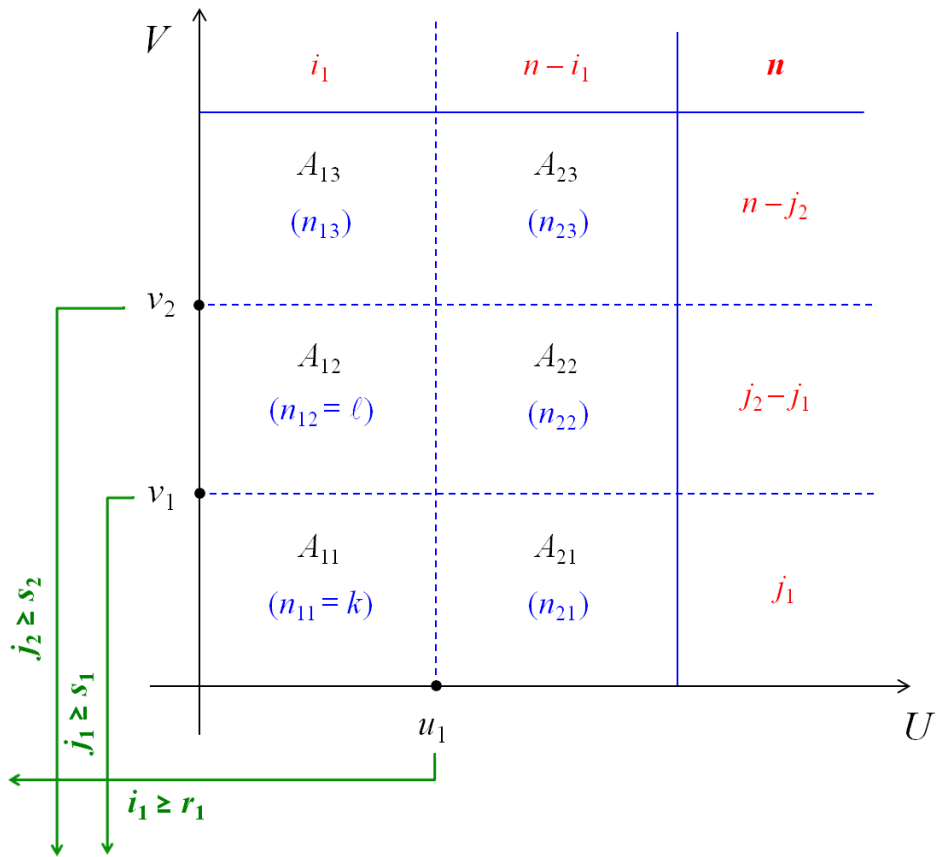
□

Στο Λήμμα 3.4 δίνεται έκφραση για την από κοινού α.σ.κ. των διατεταγμένων παρατηρήσεων $U_{r_1:n}$ και $V_{s_1:n}, V_{s_2:n}$, η οποία μπορεί να προσδιορισθεί μέσω των πιθανοτήτων των συνδυασμένων περιοχών που προκύπτουν εάν θεωρήσουμε ένα σημείο αποκοπής για το χαρακτηριστικό U και δύο για το χαρακτηριστικό V . Οι περιοχές που δημιουργούνται απεικονίζονται στο Σχήμα 3.4.

| Ενδεχόμενα | $A_1 : [U \leq u_1]$ | $A_2 : [U > u_1]$ | Περιθώριες Πιθανότητες |
|----------------------------|-----------------------|-----------------------|------------------------|
| $B_3 : [V > v_2]$ | $p_{13} = P(A_1 B_3)$ | $p_{23} = P(A_2 B_3)$ | $p_{\bullet 3}$ |
| $B_2 : [v_1 < V \leq v_2]$ | $p_{12} = P(A_1 B_2)$ | $p_{22} = P(A_2 B_2)$ | $p_{\bullet 2}$ |
| $B_1 : [V \leq v_1]$ | $p_{11} = P(A_1 B_1)$ | $p_{21} = P(A_2 B_1)$ | $p_{\bullet 1}$ |
| Περιθώριες Πιθανότητες | $p_{1\bullet}$ | $p_{2\bullet}$ | 1 |

Πίνακας 3.4: Σχήμα Δειγματοληψίας – (2×3) :

Τα ενδεχόμενα A_{ij} και οι αντίστοιχες πιθανότητες p_{ij} για $i \in \{1, 2\}$ και $j \in \{1, 2, 3\}$



Σχήμα 3.4: Γραφική απεικόνιση του ενδεχομένου

$$[U_{r_1:n} \leq u_1; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2]$$

Λήμμα 3.4. Έστω (X_i, Y_i) , $i = 1, 2, \dots, n$, n ανεξάρτητες τ.μ. που προέρχονται από μία δισδιάστατη συνεχή κατανομή με α.σ.κ. $F(x, y)$, περιθώριες $F_X(x)$, $F_Y(y)$ και αντίστοιχο σύνδεσμο $D(u, v)$. Τότε η από κοινού α.σ.κ. των τ.μ.

$$U_{r_1:n} = F_X(X_{r_1:n}) \quad \text{και} \quad V_{s_1:n} = F_Y(Y_{s_1:n}), \quad V_{s_2:n} = F_Y(Y_{s_2:n})$$

εκφράζεται ως εξής

$$\begin{aligned} F_{r_1; s_1, s_2; n}^{(D)}(u_1; v_1, v_2) &= P(U_{r_1:n} \leq u_1; V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2) \\ &= \sum_{i_1=r_1}^n \sum_{\substack{j_1=s_1 \\ j_2 \geq j_1}}^n \sum_{j_2=s_2}^n \sum_{k, \ell} \frac{n!}{\prod_{i=1}^2 \prod_{j=1}^3 n_{ij}!} \prod_{i=1}^2 \prod_{j=1}^3 p_{ij}^{n_{ij}}, \end{aligned} \quad (3.6)$$

όπου

$$p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1}),$$

για $1 \leq r_1 \leq n$, $1 \leq s_1 \leq s_2 \leq n$ και $i \in \{1, 2\}$, $j \in \{1, 2, 3\}$ (σύμβαση : $u_0 = v_0 = 0$ και $u_2 = v_3 = 1$),

Απόδειξη. Εργαζόμενοι με λογική παρόμοια αυτής που ακολουθήθηκε προηγουμένως, μπορούμε να χρησιμοποιήσουμε τον Πίνακα 3.4, ώστε να εκφράσουμε την από κοινού α.σ.κ. των τ.μ. $U_{r_1:n} = F_X(X_{r_1:n})$, $V_{s_1:n} = F_Y(Y_{s_1:n})$, $V_{s_2:n} = F_Y(Y_{s_2:n})$ με τη μορφή πολυωνυμικών πιθανοτήτων ως εξής

$$\begin{aligned}
 F_{r_1:s_1,s_2:n}^{(D)}(u_1; v_1, v_2) &= F_{U_{r_1:n}; V_{s_1:n}, V_{s_2:n}}^{(D)}(u_1; v_1, v_2) \\
 &= P(U_{r_1:n} \leq u_1, V_{s_1:n} \leq v_1, V_{s_2:n} \leq v_2) \\
 &= P(\text{τουλάχιστον } r_1 \text{ από τις τ.μ. } U_1, \dots, U_n \text{ είναι μικρότερες ή ίσες από } u_2, \\
 &\quad \text{τουλάχιστον } s_1 \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μικρότερες ή ίσες από } v_1, \\
 &\quad \text{τουλάχιστον } s_2 \text{ από τις τ.μ. } V_1, \dots, V_n \text{ είναι μικρότερες ή ίσες από } v_2) \\
 &= \sum_{i_1=r_1}^n \sum_{\substack{j_1=s_1 \\ j_2 \geq j_1}}^n \sum_{j_2=s_2}^n P(\text{τα ενδεχόμενα } [U \leq u_1], [V \leq v_1], [V \leq v_2] \\
 &\quad \text{πραγματοποιούνται ακριβώς } i_1, j_1, j_2 \text{ φορές, αντίστοιχα}) \\
 &= \sum_{i_1=r_1}^n \sum_{\substack{j_1=s_1 \\ j_2 \geq j_1}}^n \sum_{j_2=s_2}^n \sum_{k,\ell} P(\text{τα από κοινού ενδεχόμενα } A_{ij} \text{ πραγματοποιούνται} \\
 &\quad \text{ακριβώς } n_{ij} \text{ φορές, αντίστοιχα, για } i \in \{1, 2\}, j \in \{1, 2, 3\}) \\
 &= \sum_{i_1=r_1}^n \sum_{\substack{j_1=s_1 \\ j_2 \geq j_1}}^n \sum_{j_2=s_2}^n \sum_{k,\ell} \frac{n!}{\prod_{i=1}^2 \prod_{j=1}^3 n_{ij}!} \prod_{i=1}^2 \prod_{j=1}^3 p_{ij}^{n_{ij}},
 \end{aligned}$$

όπου $p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1})$ για $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$ με τη σύμβαση ότι $u_0 = v_0 = 0$ και $u_2 = v_3 = 1$. Σε αυτή την περίπτωση (βλ. Σχήμα 3.4), προκύπτει το εξής σύστημα 5 εξισώσεων και 6 αγνώστων

$$\left\{ \begin{array}{l}
 n_{11} + n_{21} = j_1 \\
 n_{12} + n_{22} = j_2 - j_1 \\
 n_{13} + n_{23} = n - j_2 \\
 n_{11} + n_{12} + n_{13} = i_1 \\
 n_{21} + n_{22} + n_{23} = n - i_1.
 \end{array} \right.$$

Προφανώς, παραλείποντας μία από τις παραπάνω προκύπτουν 4 εξισώσεις και 6 άγνωστοι,

οπότε θέτοντας $n_{11} = k$ και $n_{12} = \ell$ έχουμε

$$\begin{cases} n_{21} = j_1 - k \\ n_{22} = j_2 - j_1 - \ell \\ n_{31} = i_1 - k - \ell \\ n_{32} = n - i_1 - j_2 + k + \ell. \end{cases} \quad \square$$

3.3 Δισδιάστατες από κοινού κατανομές πλήθους διατεταγμένων παρατηρήσεων

Στη συνέχεια, παρουσιάζονται δύο ακόμη από κοινού δισδιάστατες κατανομές για το πλήθος διατεταγμένων ομοιόμορφων παρατηρήσεων στο διάστημα $(0,1)$. Συγκεκριμένα, οι καινούργιες κατανομές διαφέρουν από αυτές που προτάθηκαν στην Ενότητα 3.2, διότι τώρα το ενδιαφέρον μας δεν εστιάζεται μόνο στη θέση που έχει το ζευγάρι $(U_{r:n}, V_{s:n})$, αλλά και το πλήθος παρατηρήσεων (που συνδέεται με κάθε ένα από τα υπό μελέτη χαρακτηριστικά), το οποίο ανήκει σε μία ορθογώνια περιοχή. Το πλήθος αυτό μπορεί να ορίζεται από τον αριθμό παρατηρήσεων που ανήκουν είτε σε συγκεκριμένο διάστημα διατεταγμένων παρατηρήσεων είτε σε όλο το εύρος του αρχικού δείγματος. Τα αποτελέσματα παρουσιάζονται στα επόμενα λήμματα.

Στο λήμμα που ακολουθεί δίνεται έκφραση για τον υπολογισμό της πιθανότητας συγκεκριμένο πλήθος ομοιόμορφων, διαδοχικών, διατεταγμένων παρατηρήσεων να ανήκει μεταξύ ορίων που ορίζουν μία ορθογώνια περιοχή.

Λήμμα 3.5. Έστω U_1, U_2, \dots, U_n και V_1, V_2, \dots, V_n τυχαία δείγματα εξαρτημένων ζευγαριών που προέρχονται από την Ομοιόμορφη κατανομή στο διάστημα $(0, 1)$ και ας συμβολίσουμε με $U_{r_1:n}, U_{r_2:n}$ και $V_{s_1:n}, V_{s_2:n}$ τις διατεταγμένες παρατηρήσεις που συνδέονται αντίστοιχα με τα δείγματα αυτά, όπου $1 \leq r_1 \leq r_2 \leq n$ και $1 \leq s_1 \leq s_2 \leq n$. Τότε, η πιθανότητα

$$Q_{O4}^{(D)} = Q_{O4}^{(D)}(u_1, u_2; v_1, v_2) = P [u_1 \leq U_{r_1:n} \leq U_{r_2:n} \leq u_2, v_1 \leq V_{s_1:n} \leq V_{s_2:n} \leq v_2]$$

εκφράζεται ως εξής

$$Q_{O4}^{(D)} = F_{r_2;s_2:n}^{(D)}(u_2; v_2) - F_{r_1;r_2;s_2:n}^{(D)}(u_1, u_2; v_2) - F_{r_2;s_1;s_2:n}^{(D)}(u_2; v_1, v_2) + F_{r_1;r_2;s_1;s_2:n}^{(D)}(u_1, u_2; v_1, v_2). \quad (3.7)$$

Απόδειξη. Αρχικά, γράφουμε τη ζητούμενη ποσότητα στην ισοδύναμη μορφή που φαίνεται παρακάτω

$$\begin{aligned} Q_{04}^{(D)} &= Q_{04}^{(D)}(u_1, u_2; v_1, v_2) = P [u_1 \leq U_{r_1:n} \leq U_{r_2:n} \leq u_2, v_1 \leq V_{s_1:n} \leq V_{s_2:n} \leq v_2] \\ &= P [u_1 \leq U_{r_1:n} \text{ και } U_{r_2:n} \leq u_2 \text{ και } v_1 \leq V_{s_1:n} \text{ και } V_{s_2:n} \leq v_2]. \end{aligned}$$

Στη συνέχεια, εισάγουμε τα ενδεχόμενα

$$A_1 : [U_{r_1:n} \leq u_1], \quad B_1 : [V_{s_1:n} \leq v_1],$$

$$A_2 : [U_{r_2:n} \leq u_2], \quad B_2 : [V_{s_2:n} \leq v_2],$$

ώστε να μπορέσουμε να εκφράσουμε την ποσότητα $Q_{04}^{(D)}$ ως εξής

$$\begin{aligned} P(A_1' A_2 B_1' B_2) &= P((A_1 \cup B_1)' A_2 B_2) \\ &= P(A_2 B_2) - P((A_1 \cup B_1) A_2 B_2) \\ &= P(A_2 B_2) - P(A_1 A_2 B_2) - P(A_2 B_1 B_2) + P(A_1 A_2 B_1 B_2). \end{aligned}$$

Προφανώς, προκύπτει άμεσα το επόμενο αποτέλεσμα

$$Q_{04}^{(D)} = F_{r_2; s_2; n}^{(D)}(u_2; v_2) - F_{r_1, r_2; s_2; n}^{(D)}(u_1, u_2; v_2) - F_{r_2; s_1, s_2; n}^{(D)}(u_2; v_1, v_2) + F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2),$$

όπου $F_{r_2; s_2; n}^{(D)}(u_2; v_2)$ είναι η α.σ.κ. που περιγράφεται μέσω της Σχέσης (3.2) για $r = r_2$ και $s = s_2$, ενώ $F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2)$ είναι η α.σ.κ. που περιγράφεται στη Σχέση (3.4), $F_{r_1, r_2; s_2; n}^{(D)}(u_1, u_2; v_2)$ είναι η α.σ.κ. της Σχέσης (3.5) εάν το s_2 αντικατασταθεί με s_1 και $F_{r_2; s_1, s_2; n}^{(D)}(u_2; v_1, v_2)$ είναι η α.σ.κ. της Σχέσης (3.6) εάν το r_2 αντικατασταθεί με r_1 . \square

Τέλος, στο παρακάτω λήμμα δίνεται έκφραση για την από κοινού α.σ.κ. της θέσης του ζεύγους $(U_{r:n}, V_{s:n})$ και του πλήθους ομοιόμορφων διατεταγμένων παρατηρήσεων από όλο το εύρος του δείγματος, για κάθε ένα από τα υπό μελέτη χαρακτηριστικά.

Λήμμα 3.6. Έστω $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, n ανεξάρτητες τ.μ. που προέρχονται από μία δισδιάστατη συνεχή κατανομή με α.σ.κ. $F(x, y)$, περιθώριες $F_X(x)$, $F_Y(y)$ και αντίστοιχο σύνδεσμο $D(u, v)$. Ας συμβολίσουμε με $U_{r:n}$ και $V_{s:n}$ την r -οστή και s -οστή διατεταγμένη παρατήρηση των διατεταγμένων δειγμάτων $U_{1:n} \leq U_{2:n} \leq \dots \leq U_{n:n}$ και $V_{1:n} \leq V_{2:n} \leq \dots \leq V_{n:n}$, όπου οι $U_i = F_X(X_i)$ και $V_j = F_Y(Y_j)$, για $i, j = 1, 2, \dots, n$, ακολουθούν την Ομοιόμορφη

κατανομή στο διάστημα $(0, 1)$. Επιπλέον, ορίζουμε την ακόλουθη απαριθμητρία συνάρτηση

$$L(z_1, z_2, \dots, z_n; w_1, w_2) = |\{i \in \{1, 2, \dots, n\} : w_1 \leq z_i \leq w_2\}|$$

η οποία για δεδομένα $z_1, z_2, \dots, z_n \in \mathbb{R}$ και $w_1, w_2 \in \mathbb{R}$, επιστρέφει τον αριθμό των z_i που βρίσκονται μεταξύ των τιμών w_1 και w_2 . Στην περίπτωση αυτή, η πιθανότητα

$$\begin{aligned} Q_{O2N2}^{(D)} &= Q_{O2N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y) \\ &= P(u_1 \leq U_{r:n} \leq u_2, v_1 \leq V_{s:n} \leq v_2 \quad \text{και} \\ &\quad L(U_1, U_2, \dots, U_n; u_1, u_2) \geq \ell_X, L(V_1, V_2, \dots, V_n; v_1, v_2) \geq \ell_Y), \end{aligned}$$

μπορεί να εκφραστεί με τη βοήθεια των πολυωνυμικών πιθανοτήτων ως εξής

$$Q_{O2N2}^{(D)} = Q_{O2N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y) = \sum_{i_1, i_2} \sum_{j_1, j_2} \sum_{k_1, k_2, k_3, k_4} \frac{n!}{\prod_{i,j=1}^3 n_{ij}!} \prod_{i,j=1}^3 p_{ij}^{n_{ij}}, \quad (3.8)$$

όπου $1 \leq u_1 < u_2 \leq n$, $1 \leq v_1 < v_2 \leq n$, $1 \leq r, s \leq n$ και

$$p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1})$$

για $i, j \in \{1, 2, 3\}$ (σύμβαση : $u_0 = v_0 = 0$ και $u_3 = v_3 = 1$).

Απόδειξη. Η ποσότητα $Q_{O2N2}^{(D)}$ μπορεί να περιγραφεί με τη βοήθεια του Σχήματος 3.5 ως ακολούθως

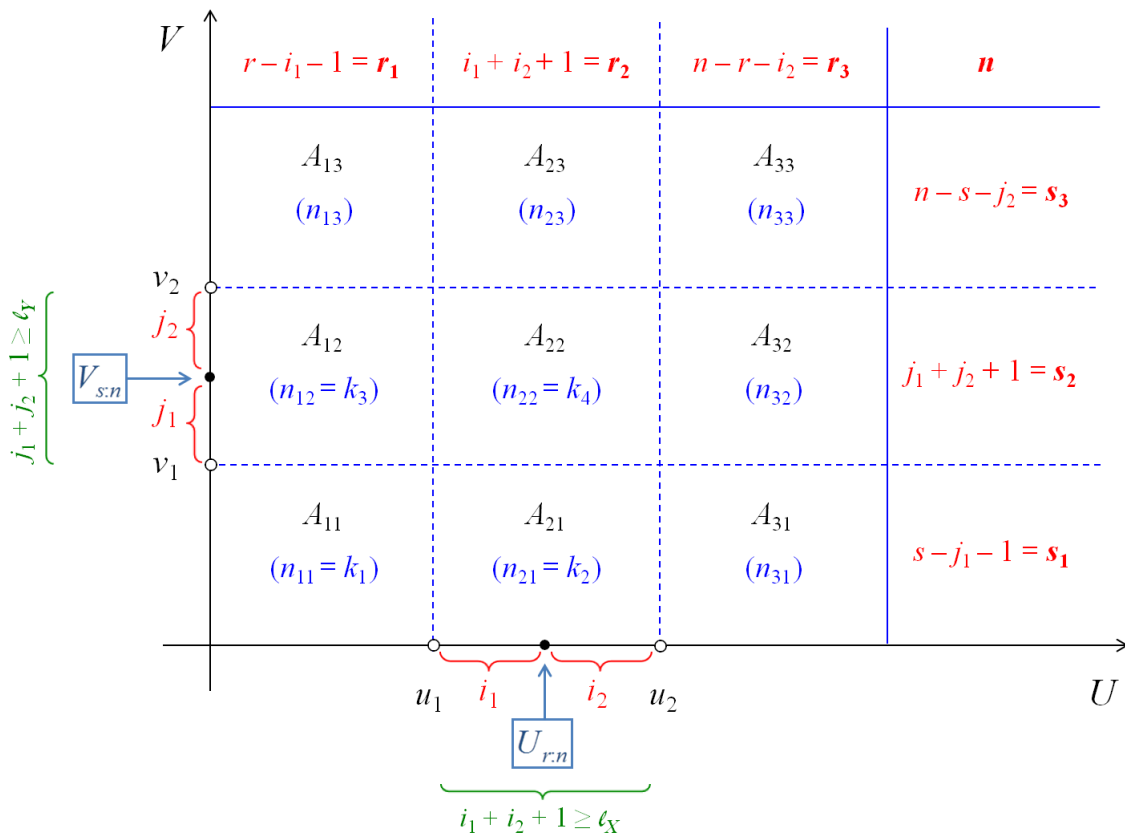
$$\begin{aligned} Q_{O2N2}^{(D)} &= Q_{O2N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y) \\ &= P(u_1 \leq U_{r:n} \leq u_2, v_1 \leq V_{s:n} \leq v_2 \quad \text{και} \\ &\quad L(U_1, U_2, \dots, U_n; u_1, u_2) \geq \ell_X, L(V_1, V_2, \dots, V_n; v_1, v_2) \geq \ell_Y) \\ &= \sum_{\substack{\ell_X - 1 \leq i_1 + i_2 \leq n - 1 \\ 0 \leq i_1 \leq r - 1, 0 \leq i_2 \leq n - r}} \sum_{\substack{\ell_Y - 1 \leq j_1 + j_2 \leq n - 1 \\ 0 \leq j_1 \leq s - 1, 0 \leq j_2 \leq n - s}} P(\text{πραγματοποίησης της απεικόνισης του} \\ &\quad \text{Σχήματος 3.5 για σταθερές τιμές των } i_1, i_2, j_1, j_2) \end{aligned}$$

$$= \sum_{i_1, i_2} \sum_{j_1, j_2} P(n_{ij} \text{ από τα ζεύγη } (U, V) \text{ ανήκουν στην περιοχή } A_{ij} \text{ για } i, j \in \{1, 2, 3\})$$

και τα n_{ij} ικανοποιούν τις συνθήκες $\sum_{j=1}^3 n_{ij} = r_i, \sum_{i=1}^3 n_{ij} = s_j,$

όπου $s_1 = s - j_1 - 1, s_2 = j_1 + j_2 + 1, s_3 = n - s - j_2$

$r_1 = r - i_1 - 1, r_2 = i_1 + i_2 + 1, r_3 = n - r - i_2).$



Σχήμα 3.5: Γραφική απεικόνιση πιθανότητας $Q_{02N2}^{(D)} = Q_{02N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y)$

Τελικά, η ποσότητα $Q_{02N2}^{(D)}$ μπορεί να υπολογισθεί ακολουθώντας το ίδιο σκεπτικό με αυτό που εφαρμόστηκε για τον υπολογισμό των α.σ.κ. που παρουσιάστηκαν στην Ενότητα 3.2.

Επομένως,

$$Q_{02N2}^{(D)} = Q_{02N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y) = \sum_{i_1, i_2} \sum_{j_1, j_2} \sum_{k_1, k_2, k_3, k_4} \frac{n!}{\prod_{i,j=1}^3 n_{ij}!} \prod_{i,j=1}^3 p_{ij}^{n_{ij}},$$

όπου $p_{ij} = P(A_{ij}) = D(u_i, v_j) - D(u_{i-1}, v_j) - D(u_i, v_{j-1}) + D(u_{i-1}, v_{j-1})$ για $i, j \in \{1, 2, 3\}$ με τη σύμβαση ότι $u_0 = v_0 = 0$ και $u_3 = v_3 = 1$). Στη συνέχεια, λαμβάνοντας υπ' όψιν μας την απεικόνιση του Σχήματος 3.5, παράγεται το επόμενο σύστημα με 6 εξισώσεις και 9 αγνώστους

$$\begin{cases} n_{11} + n_{21} + n_{31} = s - j_1 - 1 \\ n_{12} + n_{22} + n_{32} = j_1 + j_2 + 1 \\ n_{13} + n_{23} + n_{33} = n - s - j_2 \\ n_{11} + n_{12} + n_{13} = r - i_1 - 1 \\ n_{21} + n_{22} + n_{23} = i_1 + i_2 + 1 \\ n_{31} + n_{32} + n_{33} = n - r - i_2. \end{cases}$$

Είναι περιττό να αναφερθεί ότι μία από τις παραπάνω εξισώσεις μπορεί να παραλειφθεί με αποτέλεσμα να οδηγούμαστε σε ένα σύστημα 5 εξισώσεων με 9 αγνώστους. Θέτοντας $n_{11} = k_1$, $n_{21} = k_2$, $n_{12} = k_3$, $n_{22} = k_4$ και αντικαθιστώντας τις ποσότητες αυτές στο τελικό σύστημα, ο αριθμός των εξισώσεων γίνεται ίσος με τον αριθμό των αγνώστων και μπορούμε πλέον εύκολα να υπολογίσουμε τις τιμές των υπόλοιπων n_{ij} ως ακολούθως

$$\begin{cases} n_{31} = s - j_1 - k_1 - k_2 - 1 \\ n_{32} = j_1 + j_2 - k_3 - k_4 + 1 \\ n_{13} = r - i_1 - k_1 - k_3 - 1 \\ n_{23} = i_1 + i_2 - k_2 - k_4 + 1 \\ n_{33} = n - r - s - i_2 - j_2 + k_1 + k_2 + k_3 + k_4. \end{cases} \quad \square$$

ΚΕΦΑΛΑΙΟ 4

Δισδιάστατο Διάγραμμα Ελέγχου με Χρήση Διατεταγμένων Παρατηρήσεων

4.1 Εισαγωγή

Η βασική υπόθεση που γίνεται σε ένα πολυδιάστατο (p -διάστατο) παραμετρικό διάγραμμα ελέγχου είναι ότι τα υπό μελέτη χαρακτηριστικά προέρχεται από κάποια συγκεκριμένη κατανομή (συνήθως πολυδιάστατη κανονική) και ως αποτέλεσμα τόσο η επιλογή της σ.σ. όσο και ο υπολογισμός των ορίων ελέγχου εξαρτώνται από τη μορφή και τις ιδιότητες της κατανομής αυτής. Το διάγραμμα χ^2 θεωρείται το πλέον κλασσικό πολυδιάστατο παραμετρικό διάγραμμα ελέγχου Φάσης II, το οποίο χρησιμοποιείται για την παρακολούθηση της μέσης τιμής μιας διεργασίας μέσω m δειγμάτων μεγέθους n το καθένα. Η μορφή της σ.σ. που χρησιμοποιείται σε αυτό δίνεται από τις εκφράσεις

$$\chi_k^2 = n(\bar{\mathbf{X}}_k - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}}_k - \boldsymbol{\mu}), \quad k = 1, 2, \dots, m,$$

για την περίπτωση μεμονωμένων παρατηρήσεων ($n = 1$) και υποομάδων ($n > 1$). Δεδομένου ότι οι παραπάνω σ.σ. αποτελούν τετραγωνικές μορφές, οι οποίες ακολουθούν την κατανομή χ^2 με p βαθμούς ελευθερίας (β.ε.), τα όρια ελέγχου του (δίπλευρου) διαγράμματος χ^2 είναι

$$LCL = 0 \quad \text{και} \quad UCL = \chi_{p;\alpha}^2.$$

Στην πράξη, όμως, η γνώση που υπάρχει για την κατανομή των χαρακτηριστικών που μελετώνται μπορεί να είναι ελλιπής ή ενδέχεται η πραγματική κατανομή τους να είναι εντελώς άγνωστη. Η κατασκευή μη παραμετρικών διαγραμμάτων αποτελεί χρήσιμη εναλλακτική λύση σε αυτές τις περιπτώσεις. Παρ' όλα αυτά, είναι λίγα σε αριθμό τα

πολυδιάστατα απαραμετρικά διαγράμματα ελέγχου που έχουν προταθεί στη διεθνή βιβλιογραφία. Πρόσφατα, οι Boone and Chakraborti (2012) πρότειναν τα διαγράμματα SN^2 και SR^2 , τα οποία βασίζονται σε πολυδιάστατες μορφές του προσημικού ελέγχου και του βαθμολογικού προσημικού ελέγχου του *Wilcoxon* που δόθηκαν από τον Hettmansperger (2006) [βλ. επίσης Puri and Sen (1976)]. Πιο συγκεκριμένα, για κάθε χαρακτηριστικό υπολογίζεται η μονοδιάστατη προσημική σ.σ.

$$S_j = \sum_{i=1}^n \text{sgn}(X_{ji} - \theta_{j0}), \quad j = 1, 2, \dots, p,$$

όπου X_{ji} είναι η i παρατήρηση του j χαρακτηριστικού, n ο αριθμός των παρατηρήσεων, θ_{j0} η καθορισμένη εντός ελέγχου τιμή της διαμέσου και

$$\text{sgn}(X_{ji} - \theta_{j0}) = \begin{cases} +1, & \text{για } X_{ji} - \theta_{j0} > 0, \\ -1, & \text{για } X_{ji} - \theta_{j0} < 0. \end{cases}$$

Συνεπώς, $S_j = \#(X_{ji} > \theta_{j0}) - \#(X_{ji} < \theta_{j0})$ για $j = 1, 2, \dots, p$ και $i = 1, 2, \dots, n$. Εάν \mathbf{S} είναι το $(p \times 1)$ διάνυσμα των S_j , τότε η ποσότητα $n^{-1/2}\mathbf{S}$ ακολουθεί ασυμπτωτικά την πολυδιάστατη κανονική κατανομή με μέση τιμή το διάνυσμα $\mathbf{0}$ και πίνακα διακυμάνσεων-συνδιακυμάνσεων $n^{-1}\mathbf{V}$. Η εκτίμηση $\hat{\mathbf{V}} = ((\hat{v}_{jl}))$, $j, l = 1, 2, \dots, p$ του πίνακα \mathbf{V} προκύπτει μέσω των τύπων

$$\hat{v}_{jj} = n \quad \text{και} \quad \hat{v}_{jl} = \sum_{i=1}^n \text{sgn}(X_{ji} - \theta_{j0}) \cdot \text{sgn}(X_{li} - \theta_{l0}).$$

Τελικά, η σ.σ. του διαγράμματος SN^2 είναι

$$SN^2 = \mathbf{S}'\hat{\mathbf{V}}^{-1}\mathbf{S}.$$

Με ανάλογο τρόπο, για κάθε χαρακτηριστικό, υπολογίζεται η μονοδιάστατη βαθμολογική προσημική σ.σ. του *Wilcoxon*

$$W_j = \sum_{i=1}^n R(|X_{ji} - \theta_{j0}|) \cdot \text{sgn}(X_{ji} - \theta_{j0}), \quad j = 1, 2, \dots, p,$$

όπου $R(|X_{ji} - \theta_{j0}|)$ είναι ο βαθμός του $|X_{ji} - \theta_{j0}|$ μεταξύ των $|X_{j1} - \theta_{j0}|, \dots, |X_{jn} - \theta_{j0}|$. Εάν \mathbf{W} είναι το $(p \times 1)$ διάνυσμα των W_j , τότε η ασυμπτωτική κατανομή της ποσότητας $n^{-3/2}\mathbf{W}$ είναι πολυδιάστατη κανονική κατανομή με μέση τιμή το διάνυσμα $\mathbf{0}$ και πίνακα διακυμάνσεων-συνδιακυμάνσεων $n^{-3}\mathbf{L}$. Η εκτίμηση $\hat{\mathbf{L}} = ((\hat{l}_{jl}))$, $j, l = 1, 2, \dots, p$ του πίνακα

\mathbf{L} προκύπτει μέσω των τύπων

$$\hat{l}_{jj} = \frac{n(n+1)(2n+1)}{6}, \quad j = 1, 2, \dots, p,$$

και

$$\hat{l}_{jl} = \sum_{i=1}^n R(|X_{ji} - \theta_{j0}|) \cdot R(|X_{li} - \theta_{l0}|) \cdot \text{sgn}(X_{ji} - \theta_{j0}) \cdot \text{sgn}(X_{li} - \theta_{l0}), \quad j = 1, 2, \dots, p.$$

Σε αυτή την περίπτωση, η σ.σ. για το k δείγμα του διαγράμματος SR^2 είναι η εξής

$$SR^2 = \mathbf{W}'\hat{\mathbf{L}}^{-1}\mathbf{W}.$$

Οι παραπάνω σ.σ. είναι τετραγωνικές μορφές, οι οποίες ακολουθούν ασυμπτωτικά την κατανομή χ^2 . Συνεπώς, τα όρια ελέγχου θα υπολογίζονται ως εξής

$$LCL = 0 \quad \text{και} \quad UCL = \chi_{p;a}^2.$$

Δεδομένης της έλλειψης εκτενούς βιβλιογραφίας που συνδέεται με πολυδιάστατα μη παραμετρικά διαγράμματα ελέγχου, στα πλαίσια της παρούσας διατριβής εξετάστηκε η δυνατότητα γενίκευσης, σε μεγαλύτερες διαστάσεις, κάποιων μονοδιάστατων μη παραμετρικών σχημάτων, εκτενής ανασκόπηση των οποίων μπορεί να βρεθεί στα κείμενα των Chakraborti *et al.* (2001), Chakraborti and Graham (2008) και Chakraborti (2011).

Ένα από τα διαγράμματα που μελετήθηκαν στις προαναφερθείσες εργασίες είναι το διάγραμμα ελέγχου για τη διάμεσο που πρότειναν οι Janacek and Meikle (1997). Ουσιαστικά, πρόκειται για ένα δίπλευρο διάγραμμα ελέγχου, τύπου *Shewhart*, τα όρια του οποίου είναι συγκεκριμένες διατεταγμένες παρατηρήσεις που επιλέγονται από ένα εντός ελέγχου δείγμα αναφοράς (*reference sample*), έστω $X_1^{(R)}, X_2^{(R)}, \dots, X_m^{(R)} \sim F_X(x)$, ενώ η σ.σ. που απεικονίζεται είναι η διάμεσος δ του δείγματος ελέγχου (*test sample*), έστω $X_1^{(T)}, X_2^{(T)}, \dots, X_n^{(T)} \sim G_X(x)$.

Για την ακρίβεια, επιλέγονται διαδοχικά δείγματα ελέγχου, που είναι ανεξάρτητα μεταξύ τους και από το δείγμα αναφοράς, με τη βοήθεια των οποίων εξετάζεται κατά πόσο οι διάμεσοι εμπίπτουν στην περιοχή ελέγχου που έχει καθορισθεί από τα όρια. Σκοπός είναι η έγκαιρη ανίχνευση πιθανών μετατοπίσεων της θέσης της κατανομής $F_X(x)$. Βάσει της στατιστικής ορολογίας, γίνεται έλεγχος της μηδενικής υπόθεσης $H_0 : F_X(x) = G_X(x)$ έναντι της δίπλευρης εναλλακτικής $H_1 : F_X(x) \neq G_X(x)$.

Το βασικό πλεονέκτημα του διαγράμματος της διαμέσου είναι ότι δεν επηρεάζεται από πιθανά σφάλματα μετρήσεων και μπορεί να χρησιμοποιηθεί σε περιπτώσεις που η

μέτρηση του χαρακτηριστικού περιλαμβάνει υποκειμενική αξιολόγηση ή διατάξιμα δεδομένα. Λαμβάνοντας υπ' όψιν ότι η εντός ελέγχου κατανομή είναι άγνωστη, είναι λογικό να χρησιμοποιούνται διατεταγμένες παρατηρήσεις ως όρια ελέγχου. Επιπλέον, λόγω της χρήσης της διαμέσου, τα όρια που επιλέγονται είναι συμμετρικά μεταξύ τους, πιο συγκεκριμένα

$$LCL = X_{a:m}^{(R)} \quad \text{και} \quad UCL = X_{m-a+1:m}^{(R)}$$

Η παράμετρος σχεδιασμού $1 \leq a \leq m$ προσδιορίζεται, ώστε το ποσοστό λανθασμένου συναγερού FAR (δηλαδή η πιθανότητα το διάγραμμα να παράγει σήμα για εκτός ελέγχου μετατόπιση μιας διεργασίας που στην πραγματικότητα είναι εντός ελέγχου), να μην υπερβαίνει ένα προκαθορισμένο επίπεδο f , δηλαδή

$$FAR = 1 - p_0 = 1 - P(LCL \leq \delta \leq UCL | H_0) \leq f$$

όπου

$$p_0 = P(X_{a:m}^{(R)} \leq X_{median:n}^{(T)} \leq X_{m-a+1:m}^{(R)} | H_0) = 1 - 2 \sum_{i=[n/2]+1}^n \frac{\binom{a+i-1}{i} \binom{m+n-a-i}{n-i}}{\binom{m+n}{n}}$$

Οι Chakraborti *et al.* (2000, 2004) επέκτειναν την παραπάνω ιδέα, προτείνοντας μία κλάση δίπλευρων και μονόπλευρων μη παραμετρικών διαγραμμάτων ελέγχου, τύπου *Shewhart*, για την παρακολούθηση ποσοστημορίων. Στα διαγράμματα αυτά επιλέγονται κατάλληλα δύο διατεταγμένες παρατηρήσεις $X_{a:m}^{(R)}$ και $X_{b:m}^{(R)}$ ($1 \leq a < b \leq m$) από το δείγμα αναφοράς, οι οποίες χρησιμοποιούνται ως όρια ελέγχου. Η διεργασία θεωρείται ότι βρίσκεται εκτός ελέγχου, εάν η j -οστή διατεταγμένη παρατήρηση του δείγματος ελέγχου, έστω $X_{j:n}^{(T)}$, δεν τοποθετείται μεταξύ των ορίων ελέγχου. Σε αυτή την περίπτωση, απαιτείται διερεύνηση για την ύπαρξη πιθανών ειδικών αιτιών που προκάλεσαν τη μετατόπιση, και εφόσον υπάρχουν, λαμβάνονται άμεσα διορθωτικές ενέργειες ώστε η διεργασία να επανέλθει στην εντός ελέγχου κατάσταση.

Σύμφωνα με τους συγγραφείς, η πιθανότητα το διάγραμμα να μην παράγει σήμα για μετατοπίσεις της θέσης της κατανομής $F_X(x)$ μπορεί να προσδιοριστεί μέσω της πυκνότητας της σ.σ. προτεραιότητας (*precedence statistic*) W_j , η οποία μετρά τον αριθμό των παρατηρήσεων του δείγματος αναφοράς που προηγούνται της διατεταγμένης παρατήρησης $X_{j:n}^{(T)}$ του δείγματος ελέγχου. Αυτό συμβαίνει, διότι το ενδεχόμενο η τ.μ. $X_{j:n}^{(T)}$ να βρίσκεται μεταξύ

των διατεταγμένων παρατηρήσεων $X_{a:m}^{(R)}$ και $X_{b:m}^{(R)}$ ισοδυναμεί με το ενδεχόμενο η σ.σ. W_j να ανήκει στο διάστημα $[a, b - 1]$. Συνεπώς, ισχύει το εξής

$$p_0 = P(X_{a:m}^{(R)} \leq X_{j:n}^{(T)} \leq X_{b:m}^{(R)} | H_0) = P_{IC}(a \leq W_{j:n} \leq b - 1),$$

όπου ο συμβολισμός IC υποδηλώνει την εντός ελέγχου κατάσταση (*In-Control state*) της διεργασίας.

Πράγματι, εάν η $F_X(x)$ είναι συνεχής, η ανίσωση $X_{j:n}^{(T)} \leq X_{b:m}^{(R)}$ είναι αληθής εάν και μόνο εάν το πολύ $b - 1$ παρατηρήσεις του δείγματος αναφοράς προηγούνται της παρατήρησης $X_{j:n}^{(T)}$, ενώ με την ίδια λογική μπορεί κανείς να διαπιστώσει ότι η ανίσωση $X_{a:m}^{(R)} \leq X_{j:n}^{(T)}$ ισχύει εάν και μόνο εάν τουλάχιστον a παρατηρήσεις του δείγματος αναφοράς προηγούνται της $X_{j:n}^{(T)}$. Αξίζει να σημειωθεί ότι ο προσδιορισμός των παραμέτρων σχεδιασμού a, b γίνεται με τη βοήθεια της ακόλουθης συνθήκης

$$\sum_{w=a}^{b-1} P_{IC}(W_j = w) = \sum_{w=a}^{b-1} \frac{\binom{j+w-1}{w} \binom{m+n-j-w}{m-w}}{\binom{m+n}{n}} \geq 1 - f.$$

Θα πρέπει να τονιστεί ότι η λογική κατασκευής των παραπάνω σχημάτων είναι σχετικά απλή και μπορεί εύκολα να γενικευθεί σε μεγαλύτερες διαστάσεις. Στην επόμενη ενότητα περιγράφεται αναλυτικά η διαδικασία κατασκευής δισδιάστατων διαγραμμάτων ελέγχου για την παρακολούθηση της μέσης τιμής μιας διεργασίας. Τα νέα, δίπλευρα και μονόπλευρα, διαγράμματα που προτείνονται (Koutras and Sofikitou (2015, 2017a)) αποτελούν άμεσες γενικεύσεις των αντίστοιχων διαγραμμάτων που εισήγαγαν οι Chakraborti *et al.* (2000, 2004).

4.2 Περιγραφή ενός νέου διαγράμματος ελέγχου

Σε αυτή την ενότητα περιγράφουμε τη διαδικασία για την κατασκευή ενός δισδιάστατου διαγράμματος ελέγχου, η οποία δε διαφέρει με τη λογική που περιγράφηκε προηγουμένως για την κατασκευή ενός μονοδιάστατου μη παραμετρικού διαγράμματος. Αρχικά, υποθέτουμε ότι ένα ικανοποιητικά επαρκές δείγμα αναφοράς $(X_1^{(R)}, Y_1^{(R)}), (X_2^{(R)}, Y_2^{(R)}), \dots, (X_m^{(R)}, Y_m^{(R)})$ μεγέθους m συλλέγεται από μία άγνωστη (τουλάχιστον εν μέρει), εντός ελέγχου, δισδιάστατη κατανομή με συνεχή από κοινού α.σ.κ. $F_{X,Y}^{(R)}(x, y) = F(x, y)$. Αξίζει να σημειωθεί ότι ένα προκαταρκτικό δείγμα μεγέθους m περίπου 1000 θεωρείται επαρκές. Η

επίδραση του μεγέθους του δείγματος αναφοράς στην απόδοση ενός διαγράμματος ελέγχου έχει μελετηθεί από τους Chakraborti *et al.* (2000, 2004), Quesenberry (1993) στη μονοδιάστατη περίπτωση και από τους Rigdon (1995), Stoumbos *et al.* (2001), μεταξύ άλλων, στην πολυδιάστατη περίπτωση.

Τα δεδομένα που συλλέγονται από το δείγμα αναφοράς θα χρησιμοποιηθούν για τη δημιουργία του κανόνα απόφασης του νέου διαγράμματος ελέγχου. Πιο συγκεκριμένα, προκειμένου να αποφασίσουμε εάν η διαδικασία βρίσκεται εντός ελέγχου, συλλέγονται διαδοχικά δείγματα ελέγχου από την υποκείμενη διαδικασία, τα οποία είναι ανεξάρτητα μεταξύ τους και επιπλέον από το δείγμα αναφοράς. Χρησιμοποιώντας στατιστικούς συμβολισμούς, με $(X_1^{(T)}, Y_1^{(T)})$, $(X_2^{(T)}, Y_2^{(T)})$, ..., $(X_n^{(T)}, Y_n^{(T)})$ υποδηλώνεται το δείγμα ελέγχου (μεγέθους n), ενώ η από κοινού α.σ.κ. συμβολίζεται με $F_{X,Y}^{(T)}(x, y) = G(x, y)$. Σκοπός μας είναι να διαπιστώσουμε εάν η διεργασία έχει μετατοπισθεί από την εντός ελέγχου κατανομή $F(x, y)$ σε μία εκτός ελέγχου κατανομή $G(x, y)$. Με άλλα λόγια, για να το θέσουμε πιο επίσημα, ενδιαφερόμαστε να ελέγξουμε τη μηδενική υπόθεση $H_0 : F(x, y) = G(x, y)$ έναντι της δίπλευρης εναλλακτικής $H_1 : F(x, y) \neq G(x, y)$.

Χάρη στην ιδιότητα της ευρωστίας (*robustness property*), η χρήση του δειγματικού διάνυσματος των διαμέσων δ είναι συνήθως προτιμότερη από το δειγματικό διάνυσμα των μέσων μ σε ένα μη παραμετρικό πολυδιάστατο περιβάλλον. Για το λόγο αυτό, στον κανόνα απόφασης, θα χρησιμοποιήσουμε το τυχαίο διάνυσμα δ και θα το συγκρίνουμε με κατάλληλα επιλεγμένα όρια ελέγχου από το δείγμα αναφοράς, ώστε να αποφασίσουμε εάν η διεργασία παραμένει εντός ελέγχου ή έχει μετατοπισθεί σε μία εκτός ελέγχου κατανομή καθώς συλλέγονται μελλοντικά δείγματα. Τα όρια ελέγχου, έστω **LCL** και **UCL**, επιλέγονται συνήθως έτσι, ώστε το *FAR* να μην υπερβαίνει ένα προκαθορισμένο επίπεδο f , δηλαδή

$$P(\mathbf{LCL} \leq \delta \leq \mathbf{UCL} \mid H_0) \geq 1 - f$$

Στη δισδιάστατη περίπτωση, ο κανόνας απόφασης που προτείνουμε βασίζεται σε τέσσερις συγκεκριμένες διατεταγμένες παρατηρήσεις, οι οποίες θα χρησιμοποιηθούν ως όρια ελέγχου του προτεινόμενου (δίπλευρου) διαγράμματος, πιο συγκεκριμένα

$$LCL_X = X_{a:m}^{(R)}, \quad UCL_X = X_{b:m}^{(R)} \quad \text{και} \quad LCL_Y = Y_{c:m}^{(R)}, \quad UCL_Y = Y_{d:m}^{(R)},$$

με $1 \leq a < b \leq m$ και $1 \leq c < d \leq m$. Ειδικότερα, συμβολίζοντας με $X_{r:n}^{(T)}$ και $Y_{s:n}^{(T)}$ την r -οστή και s -οστή διατεταγμένη παρατήρηση του δείγματος ελέγχου, η υποκείμενη διεργασία θα

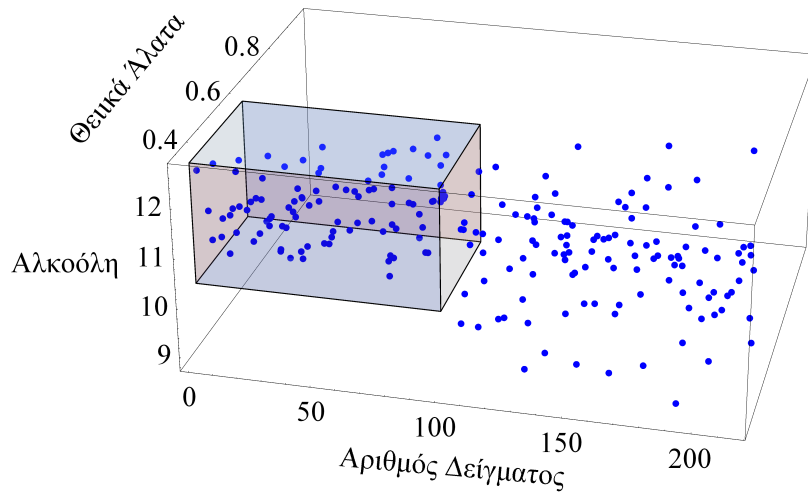
χαρακτηρίζεται εντός ελέγχου εάν επαληθεύονται οι επόμενες δύο συνθήκες

$$X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \quad \text{και} \quad Y_{c:m}^{(R)} \leq Y_{s:n}^{(T)} \leq X_{d:m}^{(R)}. \quad (4.1)$$

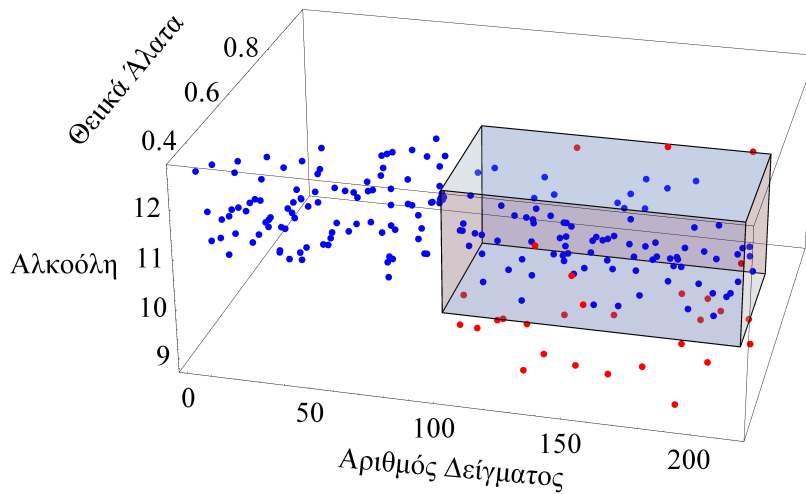
Επειδή γίνεται χρήση δύο διατεταγμένων παρατηρήσεων του δείγματος ελέγχου (2 *Order statistics*), το νέο διάγραμμα θα ονομάζεται στο εξής διάγραμμα *O2*.

Η συνήθης επιλογή των παραμέτρων r, s είναι $[(n + 1)/2]$. Σε αυτή την περίπτωση, οι σ.σ. $X_{r:n}^{(T)}, Y_{s:n}^{(T)}$ δεν είναι παρά οι διάμεσοι των χαρακτηριστικών X και Y , αντίστοιχα. Ωστόσο, η επιλογή των τιμών των r, s μπορεί να επιλεγθεί λαμβάνοντας υπ' όψιν την (α)συμμετρία της κατανομής ή/και το είδος της μετατόπισης που επιθυμούμε να ανιχνεύσουμε. Οι παράμετροι m, n, r, s, a, b, c και d ονομάζονται παράμετροι σχεδιασμού (*design parameters*) και επιλέγονται κατάλληλα, ώστε το *FAR* να λαμβάνει μια προκαθορισμένη επιθυμητή τιμή. Στην Ενότητα 4.5, περιγράφεται λεπτομερώς η λογική προσδιορισμού των παραμέτρων αυτών.

Στο Σχήμα 4.1 απεικονίζεται το τρισδιάστατο (3D) γράφημα του προτεινόμενου δι-διάστατου (δίπλευρου) διαγράμματος για ένα συγκεκριμένο σύνολο δεδομένων, το οποίο θα αναλυθεί αργότερα στην Ενότητα 4.7. Ο οριζόντιος άξονας χρησιμοποιείται για να περιγράψει τη χρονολογική σειρά κάθε δείγματος ή υποομάδας, ενώ στους άλλους δύο άξονες καταχωρούνται τα ζεύγη των τιμών των υπό μελέτη χαρακτηριστικών. Η περιοχή ελέγχου του νέου διαγράμματος αντιπροσωπεύεται από το σκιασμένο ορθογώνιο παραλληλεπίπεδο, τα όρια του οποίου έχουν καθοριστεί στη Φάση I με τη βοήθεια ενός ιστορικού συνόλου δεδομένων (δείγμα αναφοράς). Δεδομένου ότι δεν παρουσιάσθηκαν ειδικές αιτίες μεταβλητότητας, τα όρια αυτά χρησιμοποιούνται για τη μελλοντική παρακολούθηση της διεργασίας. Στη Φάση II, το δείγμα ελέγχου αξιολογείται για να ελεγχθεί εάν επαληθεύεται ο Κανόνας 4.1. Κάθε σημείο που βρίσκεται εκτός της σκιασμένης περιοχής υποδηλώνει την ύπαρξη ισχυρών ενδείξεων ότι η διεργασία είναι εκτός ελέγχου και, επομένως, σημειώνεται με κόκκινο χρώμα. Στο Σχήμα 4.1 χρησιμοποιείται ένα σύνολο πραγματικών δεδομένων από μία διεργασία παραγωγής κρασιού για λόγους απεικόνισης. Περισσότερες λεπτομέρειες σχετικά με την επιλογή των παραμέτρων σχεδιασμού, καθώς και του συγκεκριμένου συνόλου δεδομένων παρέχονται στην Ενότητα 4.7.



(α') Φάση I: Περιοχή ελέγχου από το δείγμα αναφοράς



(β') Φάση II: Απεικονιζόμενα σημεία του δείγματος ελέγχου στην περιοχή ελέγχου που προσδιορίστηκε στη Φάση I

Σχήμα 4.1: Εφαρμογή διαγράμματος O_2 στα δεδομένα λευκού οίνου της ποικιλίας *Vinho Verde* με μεταβλητές τα Θειικά Άλατα και την Αλκοόλη (Γράφημα 3D)

4.3 Χαρακτηριστικά διαγράμματος ελέγχου

Στην παρούσα ενότητα, αποδεικνύουμε μερικές εκφράσεις για τον προσδιορισμό των χαρακτηριστικών του διαγράμματος ελέγχου. Για το σκοπό αυτό, χρειάζεται να καταφύγουμε στη Θεωρία των Δισδιάστατων Συνδέσμων. Ας υποθέσουμε ότι με $C(u, v)$, $D(u, v)$ δηλώνουμε τους δισδιάστατους συνδέσμους που σχετίζονται με τις από κοινού α.σ.κ. $F(x, y)$, $G(x, y)$, ενώ με $F_X(x)$, $F_Y(y)$ και $G_X(x)$, $G_Y(y)$ συμβολίζουμε τις περιθώριες αθροιστικές συναρτήσεις. Σύμφωνα με το Θεώρημα του Sklar (1959), οι από κοινού συναρτήσεις $F(x, y)$, $G(x, y)$ μπορούν να εκφραστούν ως εξής

$$F(x, y) = C(F_X(x), F_Y(y)) \quad \text{και} \quad G(x, y) = D(G_X(x), G_Y(y)).$$

Από αυτό το σημείο, θα θεωρούμε ότι τόσο το δείγμα αναφοράς όσο και το δείγμα ελέγχου προέρχονται από συνεχείς δισδιάστατες κατανομές.

4.3.1 Δίπλευρα διαγράμματα ελέγχου

Σε αυτή την περίπτωση, η πιθανότητα το διάγραμμα $O2$ να μην παράγει σήμα για εκτός ελέγχου μετατοπίσεις ορίζεται ως εξής

$$p_{O2}^{(2)} = P(X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \quad \text{και} \quad Y_{c:m}^{(R)} \leq Y_{s:n}^{(T)} \leq Y_{d:m}^{(R)}). \quad (4.2)$$

Παρατηρούμε ότι η πιθανότητα $OCF_{O2}^{(2)} = p_{O2}^{(2)}$ δηλώνει τη χαρακτηριστική συνάρτηση λειτουργίας (*Operating Characteristic Function, OCF*) του δίπλευρου διαγράμματος $O2$, ενώ η ποσότητα $AR_{O2}^{(2)} = 1 - p_{O2}^{(2)}$ δηλώνει την πιθανότητα συναγεμού (*Alarm Rate, AR*), δηλαδή την πιθανότητα να ενεργοποιηθεί συναγεμός στο διάγραμμα. Στη συνέχεια αποδεικνύουμε μία έκφραση για την πιθανότητα $p_{O2}^{(2)}$, όπως αυτή ορίζεται στη Σχέση (4.2).

Πρόταση 4.1. Η χαρακτηριστική συνάρτηση λειτουργίας του δίπλευρου διαγράμματος $O2$ υπολογίζεται από την έκφραση

$$p_{O2}^{(2)} = p_{O2}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b, c, d; r, s) \\ = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u)), G_Y(F_Y^{-1}(v))) \times \frac{\partial^2}{\partial u \partial v} H^{(C)}(m; a, b, c, d; u, v) \, dudv, \quad (4.3)$$

όπου

$$H^{(C)}(m; a, b, c, d; u, v) = F_{b,d;m}^{(C)}(u, v) + F_{a,c;m}^{(C)}(u, v) - F_{a,d;m}^{(C)}(u, v) - F_{b,c;m}^{(C)}(u, v)$$

και $F_{r,s;n}^{(D)}(\cdot, \cdot)$ είναι η από κοινού α.σ.κ. του Λήμματος 3.1, ενώ οι ποσότητες $F_{b,d;m}^{(C)}(u, v)$,

$F_{a,c;m}^{(C)}(u, v)$, $F_{a,d;m}^{(C)}(u, v)$, $F_{b,c;m}^{(C)}(u, v)$ υπολογίζονται με τη βοήθεια της Σχέσης (3.2) εάν το $D(u, v)$ αντικατασταθεί με $C(u, v)$.

Απόδειξη. Η πιθανότητα που εμφανίζεται στο δεξί μέλος της Σχέσης (4.2) γράφεται στην ισοδύναμη μορφή

$$P_{O_2}^{(2)} = P(F_X(X_{a;m}^{(R)}) \leq F_X(X_{r;n}^{(T)}) \leq F_X(X_{b;m}^{(R)}) \text{ και } F_Y(Y_{c;m}^{(R)}) \leq F_Y(Y_{s;n}^{(T)}) \leq F_Y(Y_{d;m}^{(R)}))$$

και εφαρμόζοντας το Θεώρημα Ολικής Πιθανότητας (ΘΟΠ) για συνεχείς τ.μ., πιο συγκεκριμένα τις

$$U_{a;m}^{(R)} = F_X(X_{a;m}^{(R)}), \quad U_{b;m}^{(R)} = F_X(X_{b;m}^{(R)}) \quad \text{και} \quad V_{c;m}^{(R)} = F_Y(Y_{c;m}^{(R)}), \quad V_{d;m}^{(R)} = F_Y(Y_{d;m}^{(R)}),$$

προκύπτει η επόμενη έκφραση

$$P_{O_2}^{(2)} = \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} P[u_1 \leq F_X(X_{r;n}^{(T)}) \leq u_2 \text{ και } v_1 \leq F_Y(Y_{s;n}^{(T)}) \leq v_2] \\ \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2,$$

όπου $f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. που δίνεται στον Τύπο (3.4) εάν τα r_1, r_2, s_1, s_2, n και C χρησιμοποιηθούν αντί των ποσοτήτων a, b, c, d, m και D , αντίστοιχα.

Στη συνέχεια, ας γράψουμε την πιθανότητα που εμφανίζεται στο τετραπλό ολοκλήρωμα στην ακόλουθη μορφή

$$P[F_X^{-1}(u_1) \leq X_{r;n}^{(T)} \leq F_X^{-1}(u_2), F_Y^{-1}(v_1) \leq Y_{s;n}^{(T)} \leq F_Y^{-1}(v_2)] \\ = P[G_X(F_X^{-1}(u_1)) \leq U_{r;n}^{(T)} \leq G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1)) \leq V_{s;n}^{(T)} \leq G_Y(F_Y^{-1}(v_2))] \\ = + F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) + F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \\ - F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) - F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1)))$$

(η τελευταία έκφραση κάνει χρήση της από κοινού α.σ.κ. που εμφανίζεται στη Σχέση (3.2)). Αντικαθιστώντας το παραπάνω αποτέλεσμα στο τετραπλό ολοκλήρωμα και αλλάζοντας τη σειρά ολοκλήρωσης, προκύπτουν τα επόμενα τέσσερα ολοκληρώματα

$$I_1 = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \times f_{b,d;m}^{(C)}(u_2, v_2) du_2 dv_2, \\ I_2 = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \times f_{a,c;m}^{(C)}(u_1, v_1) du_1 dv_1,$$

$$I_3 = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \times f_{a,d;m}^{(C)}(u_1, v_2) du_1 dv_2,$$

$$I_4 = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \times f_{b,c;m}^{(C)}(u_2, v_1) du_2 dv_1,$$

όπου $f_{b,d;m}^{(C)}(u, v)$, $f_{a,c;m}^{(C)}(u, v)$, $f_{a,d;m}^{(C)}(u, v)$, $f_{b,c;m}^{(C)}(u, v)$ είναι οι από κοινού πυκνότητες της α.σ.κ. που περιγράφεται στο Λήμμα 3.1. Λαμβάνοντας υπ' όψιν όλα τα παραπάνω, η πιθανότητα $p_{O2}^{(2)}$ περιορίζεται στην έκφραση

$$p_{O2}^{(2)} = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u)), G_Y(F_Y^{-1}(v)))$$

$$\times [f_{b,d;m}^{(C)}(u, v) + f_{a,c;m}^{(C)}(u, v) - f_{a,d;m}^{(C)}(u, v) - f_{b,c;m}^{(C)}(u, v)] dudv$$

$$= \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u)), G_Y(F_Y^{-1}(v))) \times \frac{\partial^2}{\partial u \partial v} H^{(C)}(m; a, b, c, d; u, v) dudv$$

όπου

$$H^{(C)}(m; a, b, c, d; u, v) = F_{b,d;m}^{(C)}(u, v) + F_{a,c;m}^{(C)}(u, v) - F_{a,d;m}^{(C)}(u, v) - F_{b,c;m}^{(C)}(u, v).$$

Αξίζει να σημειωθεί ότι, ύστερα από απλοποιήσεις και με τη βοήθεια του Λήμματος 3.1, η ποσότητα $H^{(C)}(m; a, b, c, d; u, v)$ μπορεί ισοδύναμα να υπολογισθεί ως εξής

$$H^{(C)}(m; a, b, c, d; u, v) = \sum_{i_1=a}^{b-1} \sum_{j_1=c}^{d-1} \sum_{k_1=\max(0, i_1+j_1-m)}^{\min(i_1, j_1)} a_m(i_1, j_1, k_1) \times q_m^{(C)}(i_1, j_1, k_1; u, v), \quad (4.4)$$

όπου

$$a_m(i_1, j_1, k_1) = \frac{m!}{k_1!(i_1 - k_1)!(j_1 - k_1)!(m - i_1 - j_1 + k_1)!}$$

και

$$q_m^{(C)}(i_1, j_1, k_1; u, v) = [C(u, v)]^{k_1} [u - C(u, v)]^{i_1 - k_1} [v - C(u, v)]^{j_1 - k_1} [\bar{C}(u, v)]^{m - i_1 - j_1 + k_1} \quad (4.5)$$

με $\bar{C}(u, v) = 1 - u - v + C(u, v)$. Στο σημείο αυτό, ολοκληρώνεται η απόδειξη της Πρότασης. \square

Εάν θέσουμε $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, προκύπτει ένας ακριβής τύπος για τον υπολογισμό της πιθανότητας συναγεμού του δίπλευρου διαγράμματος $O2$ που δίνεται από τη Σχέση (4.3). Στην περίπτωση που επιπλέον ισχύει $C(u, v) = D(u, v)$, προκύπτει ο τύπος για τον υπολογισμό της πιθανότητας λανθασμένου συναγεμού του νέου διαγράμματος. Τα αποτελέσματα παρουσιάζονται στα ακόλουθα πορίσματα.

Πόρισμα 4.1. Εάν $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, η πιθανότητα συναγερού του διαγράμματος O_2 , το οποίο ορίζεται από τον Κανόνα 4.1, δεν εξαρτάται από τις περιθώριες κατανομές και υπολογίζεται μέσω της έκφρασης

$$\begin{aligned} AR_{O_2}^{(2)} &= AR_{O_2}^{(C,D)}(m, n; a, b, c, d; r, s) = 1 - p_{O_2}^{(F_X, F_X; F_Y, F_Y; C, D)}(m, n; a, b, c, d; r, s) \\ &= 1 - \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(u, v) \times \frac{\partial^2}{\partial u \partial v} H^{(C)}(m; a, b, c, d; u, v) dudv, \end{aligned} \quad (4.6\alpha)$$

ή ισοδύναμα από τη σχέση

$$\begin{aligned} AR_{O_2}^{(2)} &= 1 + \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} a_n(i, j, k) \\ &\quad \times \sum_{i_1=a}^{b-1} \sum_{j_1=c}^{d-1} \sum_{k_1=\max(0, i_1+j_1-m)}^{\min(i_1, j_1)} a_m(i_1, j_1, k_1) \times Q^{(C,D)}(i, j, k; i_1, j_1, k_1), \end{aligned} \quad (4.6\beta)$$

όπου

$$Q^{(C,D)}(i, j, k; i_1, j_1, k_1) = \int_0^1 \int_0^1 \frac{\partial}{\partial u} q_n^{(D)}(i, j, k; u, v) \times \frac{\partial}{\partial v} q_m^{(C)}(i_1, j_1, k_1; u, v) dudv. \quad (4.7)$$

Απόδειξη. Επειδή $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, συνεπάγεται ότι $G_X(F_X^{-1}(u)) = u$ και $G_Y(F_Y^{-1}(v)) = v$. Σε αυτή την περίπτωση η Σχέση (4.3) παίρνει τη μορφή

$$AR_{O_2}^{(2)} = 1 - \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(u, v) \times \frac{\partial^2}{\partial u \partial v} H^{(C)}(m; a, b, c, d; u, v) dudv,$$

και με παραγοντική ολοκλήρωση στο εσωτερικό ολοκλήρωμα έχουμε

$$\begin{aligned} AR_{O_2}^{(2)} &= 1 - \int_0^1 \left(F_{r,s;n}^{(D)}(u, v) \times \frac{\partial^2}{\partial u \partial v} H^{(C)}(m; a, b, c, d; u, v) \right) \Big|_{u=0}^1 dv \\ &\quad + \int_0^1 \int_0^1 \frac{\partial}{\partial u} F_{r,s;n}^{(D)}(u, v) \times \frac{\partial}{\partial v} H^{(C)}(m; a, b, c, d; u, v) dudv. \end{aligned} \quad (4.8)$$

Στη συνέχεια, παρατηρούμε ότι

$$\begin{aligned} H^{(C)}(m; a, b, c, d; u, v) &= F_{b,d;m}^{(C)}(u, v) + F_{a,c;m}^{(C)}(u, v) - F_{a,d;m}^{(C)}(u, v) - F_{b,c;m}^{(C)}(u, v) \\ &= C_{U_{b,m}, V_{d,m}}(I_u(b, m - b + 1), I_v(d, m - d + 1)) \\ &\quad + C_{U_{a,m}, V_{c,m}}(I_u(a, m - a + 1), I_v(c, m - c + 1)) \\ &\quad - C_{U_{a,m}, V_{d,m}}(I_u(a, m - a + 1), I_v(d, m - d + 1)) \\ &\quad - C_{U_{b,m}, V_{c,m}}(I_u(b, m - b + 1), I_v(c, m - c + 1)), \end{aligned}$$

όπου

$$I_x(a, b) = \frac{B_x(a, b)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt$$

δηλώνει την κανονικοποιημένη μη-πλήρη συνάρτηση Βήτα. Λαμβάνοντας υπ' όψιν τις ιδιότητες των συνδέσμων $C(u, 1) = u, C(1, v) = v, C(u, 0) = C(0, v) = 0$ (βλ. Σχέσεις (2.21)), καταλήγουμε στο εξής

$$H^{(C)}(m; a, b, c, d; 1, v) = H^{(C)}(m; a, b, c, d; 0, v) = 0.$$

Συνεπώς, το απλό ολοκλήρωμα που εμφανίζεται στη Σχέση (4.8) μηδενίζεται και τελικά η έκφραση του $AR_{O2}^{(2)}$ παίρνει τη μορφή

$$AR_{O2}^{(2)} = 1 + \int_0^1 \int_0^1 \frac{\partial}{\partial u} F_{r,s;n}^{(D)}(u, v) \times \frac{\partial}{\partial v} H^{(C)}(m; a, b, c, d; u, v) dudv.$$

Τελικά, με βάση τα παραπάνω αποτελέσματα, αξιοποιώντας ταυτόχρονα τη Σχέση (4.4) και το Λήμμα 3.1, οδηγούμαστε στη Σχέση (4.6β). \square

Πόρισμα 4.2. Στην ειδική περίπτωση που ισχύει $F_X(x) = G_X(x), F_Y(y) = G_Y(y)$ και συγχρόνως $C(u, v) = D(u, v)$, προκύπτει η πιθανότητα λανθασμένου συναγερομού του δίπλευρου διαγράμματος $O2$, η οποία ορίζεται από την έκφραση

$$\begin{aligned} FAR_{O2}^{(2)} &= FAR_{O2}^{(C)}(m, n; a, b, c, d; r, s) \\ &= 1 - P(X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)}, Y_{c:m}^{(R)} \leq Y_{s:n}^{(T)} \leq Y_{d:m}^{(R)} | H_0) \\ &= AR_{O2}^{(C,C)}(m, n; a, b, c, d; r, s), \end{aligned}$$

όπου

$$AR_{O2}^{(C,C)}(m, n; a, b, c, d; r, s) = 1 + \int_0^1 \int_0^1 \frac{\partial}{\partial u} F_{r,s;n}^{(C)}(u, v) \times \frac{\partial}{\partial v} H^{(C)}(m; a, b, c, d; u, v) dudv$$

ή ισοδύναμα

$$\begin{aligned} AR_{O2}^{(C,C)}(m, n; a, b, c, d; r, s) &= 1 + \sum_{i=r}^n \sum_{j=s}^n \sum_{k=\max(0, i+j-n)}^{\min(i, j)} a_n(i, j, k) \\ &\quad \times \sum_{i_1=a}^{b-1} \sum_{j_1=c}^{d-1} \sum_{k=\max(0, i_1+j_1-m)}^{\min(i_1, j_1)} a_m(i_1, j_1, k) \times Q^{(C,C)}(i, j, k; i_1, j_1, k_1). \end{aligned}$$

Απόδειξη. Τα αποτελέσματα προκύπτουν άμεσα εάν θέσουμε $D = C$ στις Σχέσεις (4.6α) και (4.6β). \square

Θα πρέπει να τονίσουμε το γεγονός ότι το $FAR_{O2}^{(2)}$ δεν επηρεάζεται από τις περιθώριες κατανομές των δειγμάτων αναφοράς και ελέγχου. Τυπικά, όμως, επηρεάζεται από τη δομή εξάρτησης των υπό μελέτη χαρακτηριστικών, όπως ορίζεται από το σύνδεσμο $C(u, v)$. Για το λόγο αυτό, τα διαγράμματα που μελετάμε εδώ χαρακτηρίζονται ως ημιπαραμετρικά. Ωστόσο, οι τιμές του ποσοστού λανθασμένου συναγερού δεν αλλάζει δραματικά για διαφορετικές επιλογές συνδέσμων, όπως θα διαπιστώσουμε από τα αριθμητικά αποτελέσματα της Ενότητας 4.5. Επομένως, μπορούν να χρησιμοποιηθούν σχεδόν ως μη παραμετρικά.

4.3.2 Μονόπλευρα διαγράμματα ελέγχου

Στο Στατιστικό Έλεγχο Ποιότητας, υπάρχουν πολλές εφαρμογές στις οποίες δεν έχει νόημα να χρησιμοποιηθούν δύο όρια ελέγχου (άνω και κάτω) για να διαπιστώσουμε ότι μία διεργασία είναι εκτός ελέγχου. Αντίθετα, μας ενδιαφέρει να ελέγξουμε μόνο εάν η διεργασία μετατοπίστηκε σε μία εκτός ελέγχου κατάσταση και επομένως λαμβάνουμε υπ' όψιν μας μόνο τις τιμές της σ.σ. ελέγχου, οι οποίες είναι μεγαλύτερες/μικρότερες από μία τιμή κατώφλι (*threshold value*). Σε τέτοιου είδους περιπτώσεις, τα μονόπλευρα διαγράμματα είναι πιο επιθυμητά, καθώς μπορούν να προσδιοριστούν με εφαρμόσιμα όρια μόνο στη μία πλευρά.

Ας υποθέσουμε ότι για το δισδιάστατο σχήμα που προτείνουμε, επιθυμούμε να ελέγξουμε εάν οι μεταβλητές X, Y έχουν μετατοπισθεί δεξιά. Σε αυτή την περίπτωση, υποθέτοντας βέβαια ότι οι μεταβλητές είναι θετικά συσχετισμένες, είναι λογικό να θεωρήσουμε ως εκτός ελέγχου ένδειξη την εμφάνιση μεγάλων τιμών σε ένα τουλάχιστον χαρακτηριστικό. Συνεπώς, η πιθανότητα το μονόπλευρο διάγραμμα $O2$ να μην παράγει σήμα για εκτός ελέγχου μετατοπίσεις δίνεται από την έκφραση

$$p_{O2}^{(1)} = p_{b,d}^{(1)} = P(X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \text{ και } Y_{s:n}^{(T)} \leq Y_{d:m}^{(R)}).$$

Χρησιμοποιώντας όμοια επιχειρήματα με αυτά που εφαρμόστηκαν στο δίπλευρο διάγραμμα ελέγχου μπορούμε, χωρίς ιδιαίτερη δυσκολία, να διαπιστώσουμε ότι η χαρακτηριστική συνάρτηση λειτουργίας, το ποσοστό συναγερού, καθώς και το ποσοστό λανθασμένου συναγερού του μονόπλευρου διαγράμματος ελέγχου (με άνω όρια), δίδονται αντίστοιχα από τις ακόλουθες εκφράσεις

$$\begin{aligned} p_{b,d}^{(1)} &= P_{O2}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; b, d; r, s) \\ &= \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u)), G_Y(F_Y^{-1}(v))) \times \frac{\partial^2}{\partial u \partial v} F_{b,d;m}^{(C)}(u, v) \, dudv \end{aligned}$$

και

$$AR_{O2}^{(1)} = AR_{b,d}^{(1)} = AR_{O2}^{(C,D)}(m, n; b, d; r, s) = \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(u, v) \times \frac{\partial^2}{\partial u \partial v} F_{b,d;m}^{(C)}(u, v) dudv,$$

$$FAR_{O2}^{(1)} = FAR_{b,d}^{(1)} = AR_{O2}^{(C,C)}(m, n; b, d; r, s) = \int_0^1 \int_0^1 F_{r,s;n}^{(C)}(u, v) \times \frac{\partial^2}{\partial u \partial v} F_{b,d;m}^{(C)}(u, v) dudv.$$

Είναι προφανές ότι η χαρακτηριστική συνάρτηση λειτουργίας $p_{O2}^{(2)}$ του δίπλευρου διαγράμματος ελέγχου μπορεί να εκφραστεί μέσω των μονόπλευρων πιθανοτήτων $p_{i,j}^{(1)}$ ως εξής

$$p_{O2}^{(2)} = p_{b,d}^{(1)} + p_{a,c}^{(1)} - p_{a,d}^{(1)} - p_{b,c}^{(1)}$$

ενώ όμοια αποτελέσματα μπορούν να προκύψουν αντίστοιχα για τα $AR_{O2}^{(1)}$ και $FAR_{O2}^{(1)}$.

4.4 Απόδοση διαγράμματος για συγκεκριμένες οικογένειες συνδέσμων

Σε αυτή την ενότητα, θα παρουσιάσουμε πώς μπορούν οι τύποι που αναπτύχθηκαν νωρίτερα να εφαρμοσθούν σε συγκεκριμένες οικογένειες συνδέσμων. Για το σκοπό αυτό χρησιμοποιούμε τους συνδέσμους *FGM* και *AMH* που αντιστοιχούν στις ομώνυμες διδιάστατες οικογένειες κατανομών των Farlie, Gumbel, Morgenstern (Farlie (1960), Gumbel (1958), Morgenstern (1956)) και Ali-Mikhail-Haq (1978) με τύπους

$$C_{\theta}^{FGM}(u, v) = uv [1 + \theta(1 - u)(1 - v)], \quad -1 \leq \theta \leq 1,$$

$$C_{\theta}^{AMH}(u, v) = uv/[1 - \theta(1 - u)(1 - v)], \quad -1 \leq \theta \leq 1.$$

Έχει ενδιαφέρον να σημειώσουμε ότι ο σύνδεσμος *AMH* μπορεί να εκφρασθεί ως γεωμετρική σειρά όπως φαίνεται παρακάτω

$$C_{\theta}^{AMH}(u, v) = uv \left[1 + \sum_{k=1}^{\infty} (\theta(1 - u)(1 - v))^k \right] = C_{\theta}^{FGM}(u, v) + \sum_{k=2}^{\infty} uv(1 - u)^k(1 - v)^k \theta^k, \quad |\theta| \leq 1,$$

το οποίο σημαίνει ότι ο σύνδεσμος *FGM* μπορεί να θεωρηθεί ως προσέγγιση πρώτης τάξης του συνδέσμου *AMH*.

Ας προχωρήσουμε στην αξιολόγηση των χαρακτηριστικών του δίπλευρου διαγράμματος ελέγχου *O2*. Για λόγους τυπογραφικής ευκολίας, εξετάζουμε την περίπτωση όπου τόσο τα δείγματα αναφοράς όσο και τα δείγματα ελέγχου έχουν τις ίδιες περιθώριες κατανομές και τους συνδέσμους που αντιστοιχούν σε αυτές, οι οποίες προέρχονται είτε από την οικογένεια *FGM* είτε από την οικογένεια *AMH* με διαφορετικές παραμέτρους, έστω θ_1, θ_2 .

Ο υπολογισμός του ποσοστού συναγερμού μπορεί να πραγματοποιηθεί με τη βοήθεια

του Τύπου (4.6β), ο οποίος στην πραγματικότητα απαιτεί τον προσδιορισμό της ποσότητας $Q^{(C,D)}(i, j, k; i_1, j_1, k_1)$ που δίνεται στη Σχέση (4.7). Ας ξεκινήσουμε με την οικογένεια FGM . Σε αυτή την περίπτωση, υποθέτουμε ότι

$$C(u, v) = C_{\theta_1}^{FGM}(u, v) \quad \text{και} \quad D(u, v) = C_{\theta_2}^{FGM}(u, v).$$

Αντικαθιστώντας τον τύπο του $C(u, v)$ στη Σχέση (4.5), έχουμε

$$q_m^{(C)}(i_1, j_1, k_1; u, v) = u^{i_1} v^{j_1} (1-u)^{m-i_1} (1-v)^{m-j_1} (1+\theta_1(1-u)(1-v))^{k_1} \\ \times (1-v\theta_1(1-u))^{i_1-k_1} (1-u\theta_1(1-v))^{j_1-k_1} (1+uv\theta_1)^{m-i_1-j_1+k_1}$$

και αναπτύσσοντας τους τέσσερις τελευταίους όρους μέσω του διωνυμικού αναπτύγματος, συμπεραίνουμε ότι

$$q_m^{(C)}(i_1, j_1, k_1; u, v) = \sum_{r_1=0}^{k_1} \sum_{r_2=0}^{i_1-k_1} \sum_{r_3=0}^{j_1-k_1} \sum_{r_4=0}^{m-i_1-j_1+k_1} \binom{k_1}{r_1} \binom{i_1-k_1}{r_2} \binom{j_1-k_1}{r_3} \binom{m-i_1-j_1+k_1}{r_4} \\ \times (-1)^{r_2+r_3} \theta_1^{r_1+r_2+r_3+r_4} u^{i_1+r_3+r_4} v^{j_1+r_2+r_4} (1-u)^{m-i_1+r_1+r_2} (1-v)^{m-j_1+r_1+r_3}.$$

Με την ίδια λογική, μπορούμε να γράψουμε το εξής

$$q_n^{(D)}(i, j, k; u, v) = \sum_{s_1=0}^k \sum_{s_2=0}^{i-k} \sum_{s_3=0}^{j-k} \sum_{s_4=0}^{n-i-j+k} \binom{k}{s_1} \binom{i-k}{s_2} \binom{j-k}{s_3} \binom{n-i-j+k}{s_4} \\ \times (-1)^{s_2+s_3} \theta_2^{s_1+s_2+s_3+s_4} u^{i+s_3+s_4} v^{j+s_2+s_4} (1-u)^{n-i+s_1+s_2} (1-v)^{n-j+s_1+s_3}$$

και με αντικατάσταση των προηγούμενων εκφράσεων στη Σχέση (4.7), καταλήγουμε στο ότι η ποσότητα $Q^{(C,D)}(i, j, k; i_1, j_1, k_1)$ γράφεται ως πολλαπλό άθροισμα ολοκληρωμάτων (τα οποία είναι πολλαπλασιασμένα με κατάλληλους διωνυμικούς συντελεστές) της μορφής

$$I = (-1)^{r_2+r_3+s_2+s_3} \theta_1^{r_1+r_2+r_3+r_4} \theta_2^{s_1+s_2+s_3+s_4} \int_0^1 \int_0^1 \left[\frac{\partial}{\partial u} \left(u^{i+s_3+s_4} v^{j+s_2+s_4} (1-u)^{n-i+s_1+s_2} (1-v)^{n-j+s_1+s_3} \right) \right. \\ \left. \times \frac{\partial}{\partial v} \left(u^{i+r_3+r_4} v^{j_1+r_2+r_4} (1-u)^{m-i_1+r_1+r_2} (1-v)^{m-j_1+r_1+r_3} \right) \right] dudv.$$

Η τελευταία ποσότητα μπορεί εύκολα να εκφρασθεί μέσω της συνάρτησης Βήτα ως ακολούθως

$$I = (-1)^{r_2+r_3+s_2+s_3} \theta_1^{r_1+r_2+r_3+r_4} \theta_2^{s_1+s_2+s_3+s_4} (I_1 - I_2 - I_3 + I_4),$$

όπου

$$\begin{aligned}
 I_1 &= (j_1 + r_2 + r_4)B(j_1 + j + r_2 + r_4 + s_2 + s_4, m + n - j_1 - j + r_1 + r_3 + s_1 + s_3 + 1) \\
 &\quad \times (i + s_3 + s_4)B(i_1 + i + r_3 + r_4 + s_3 + s_4, m + n - i_1 - i + r_1 + r_2 + s_1 + s_2 + 1), \\
 I_2 &= (m - j_1 + r_1 + r_3)B(j_1 + j + r_2 + r_4 + s_2 + s_4 + 1, m + n - j_1 - j + r_1 + r_3 + s_1 + s_3) \\
 &\quad \times (i + s_3 + s_4)B(i_1 + i + r_3 + r_4 + s_3 + s_4, m + n - i_1 - i + r_1 + r_2 + s_1 + s_2 + 1), \\
 I_3 &= (j_1 + r_2 + r_4)B(j_1 + j + r_2 + r_4 + s_2 + s_4, m + n - j_1 - j + r_1 + r_3 + s_1 + s_3 + 1) \\
 &\quad \times (n - i + s_1 + s_2)B(i_1 + i + r_3 + r_4 + s_3 + s_4 + 1, m + n - i_1 - i + r_1 + r_2 + s_1 + s_2), \\
 I_4 &= (m - j_1 + r_1 + r_3)B(j_1 + j + r_2 + r_4 + s_2 + s_4 + 1, m + n - j_1 - j + r_1 + r_3 + s_1 + s_3) \\
 &\quad \times (n - i + s_1 + s_2)B(i_1 + i + r_3 + r_4 + s_3 + s_4 + 1, m + n - i_1 - i + r_1 + r_2 + s_1 + s_2).
 \end{aligned}$$

Μία εναλλακτική έκφραση για την ποσότητα I δίνεται και μέσω της συνάρτησης Γάμμα

$$\begin{aligned}
 I &= (-1)^{r_2+r_3+s_2+s_3} \theta_1^{r_1+r_2+r_3+r_4} \theta_2^{s_1+s_2+s_3+s_4} \\
 &\quad \times [(j_1 + r_2 + r_4)(n - j + s_1 + s_3) - (m - j_1 + r_1 + r_3)(j + s_2 + s_4)] \Gamma(j_1 + j + r_2 + r_4 + s_2 + s_4) \\
 &\quad \times [(i + s_3 + s_4)(m - i_1 + r_1 + r_2) - (n - i + s_1 + s_2)(i_1 + r_3 + r_4)] \Gamma(i_1 + i + r_3 + r_4 + s_3 + s_4) \\
 &\quad \times \frac{\Gamma(m + n - j_1 - j + r_1 + r_3 + s_1 + s_3) \Gamma(m + n - i_1 - i + r_1 + r_2 + s_1 + s_2)}{[\Gamma(m + n + r_1 + r_2 + r_3 + r_4 + s_1 + s_2 + s_3 + s_4 + 1)]^2}.
 \end{aligned}$$

Είναι περιττό να ειπωθεί πως όταν $\theta_1 = \theta_2 = \theta$, οι παραπάνω υπολογισμοί αναφέρονται στην περίπτωση στην οποία τόσο το δείγμα αναφοράς όσο και το δείγμα ελέγχου προέρχονται από την ίδια διδιάστατη κατανομή. Αυτό πρακτικά σημαίνει ότι καλύπτουν την ίδια εντός ελέγχου κατάσταση.

Ας προχωρήσουμε με την περίπτωση του συνδέσμου AMH , όπου θα υποθέσουμε ότι

$$C(u, v) = C_{\theta_1}^{AMH}(u, v) \quad \text{και} \quad D(u, v) = C_{\theta_2}^{AMH}(u, v).$$

και, ως αποτέλεσμα, η ποσότητα $q_m^{(C)}(i_1, j_1, k_1; u, v)$ θα λάβει τη μορφή

$$\begin{aligned}
 q_m^{(C)}(i_1, j_1, k_1; u, v) &= u^{i_1} v^{j_1} (1 - u)^{m-i_1} (1 - v)^{m-j_1} (1 - \theta_1(1 - u))^{i_1-k_1} (1 - \theta_1(1 - v))^{j_1-k_1} \\
 &\quad \times (1 - \theta_1(1 - u - v))^{m-i_1-j_1+k_1} (1 - \theta_1(1 - u)(1 - v))^{-m}
 \end{aligned}$$

ή ισοδύναμα (χρησιμοποιώντας το διωνυμικό ανάπτυγμα)

$$\begin{aligned}
 q_m^{(C)}(i_1, j_1, k_1; u, v) &= \sum_{r_1=0}^{i_1-k_1} \sum_{r_2=0}^{j_1-k_1} \sum_{r_3=0}^{m-i_1-j_1+k_1} \sum_{r_4=0}^{r_3} \sum_{r_5=0}^{\infty} \binom{i_1-k_1}{r_1} \binom{j_1-k_1}{r_2} \binom{m-i_1-j_1+k_1}{r_3} \binom{r_3}{r_4} \binom{r_3}{r_5} \binom{m+r_5-1}{r_5} \\
 &\quad \times (-1)^{r_1+r_2+r_3+r_4} \theta_1^{r_1+r_2+r_3+r_5} u^{i_1} v^{j_1+r_4} (1 - u)^{m-i_1+r_1+r_3-r_4+r_5} (1 - v)^{m-j_1+r_2+r_5}.
 \end{aligned}$$

Παρόμοια αποτελέσματα ισχύουν και για την ποσότητα $q_n^{(C)}(i, j, k; u, v)$, η οποία δίνεται από την έκφραση

$$q_n^{(D)}(i, j, k; u, v) = \sum_{s_1=0}^{i-k} \sum_{s_2=0}^{j-k} \sum_{s_3=0}^{n-i-j+k} \sum_{s_4=0}^{s_3} \sum_{s_5=0}^{\infty} \binom{i-k}{s_1} \binom{j-k}{s_2} \binom{n-i-j+k}{s_3} \binom{s_3}{s_4} \binom{n+s_5-1}{s_5} \\ \times (-1)^{s_1+s_2+s_3+s_4} \theta_2^{s_1+s_2+s_3+s_5} u^i v^{j+s_4} (1-u)^{n-i+s_1+s_3-s_4+s_5} (1-v)^{m-j+s_2+s_5}.$$

Τελικά, η ποσότητα $Q^{(C,D)}(i, j, k; i_1, j_1, k_1)$ μπορεί να γραφεί ως πολλαπλό άθροισμα από ολοκληρώματα της μορφής

$$I = (-1)^{r_1+r_2+r_3+r_4+s_1+s_2+s_3+s_4} \theta_1^{r_1+r_2+r_3+r_5} \theta_2^{s_1+s_2+s_3+s_5} \\ \times \int_0^1 \int_0^1 \left[\frac{\partial}{\partial u} \left(u^i v^{j+s_4} (1-u)^{n-i+s_1+s_3-s_4+s_5} (1-v)^{m-j+s_2+s_5} \right) \right. \\ \left. \times \frac{\partial}{\partial v} \left(u^{i_1} v^{j_1+r_4} (1-u)^{m-i_1+r_1+r_3-r_4+r_5} (1-v)^{m-j_1+r_2+r_5} \right) \right] dudv,$$

ενώ καθαρά αλγεβρικοί χειρισμοί οδηγούν στην επόμενη έκφραση

$$I = (-1)^{r_1+r_2+r_3+r_4+s_1+s_2+s_3+s_4} \theta_1^{r_1+r_2+r_3+r_5} \theta_2^{s_1+s_2+s_3+s_5} (I_1 - I_2 - I_3 + I_4),$$

όπου

$$I_1 = i(j_1 + r_4) \mathbf{B}(j_1 + j + r_4 + s_4, m + n - j_1 - j + r_2 + r_5 + s_2 + s_5 + 1)$$

$$\times \mathbf{B}(i_1 + i, m + n - i_1 - i + r_1 + r_3 - r_4 + r_5 + s_1 + s_3 - s_4 + s_5 + 1),$$

$$I_2 = i(m - j_1 + r_2 + r_5) \mathbf{B}(j_1 + j + r_4 + s_4 + 1, m + n - j_1 - j + r_2 + r_5 + s_2 + s_5)$$

$$\times \mathbf{B}(i_1 + i, m + n - i_1 - i + r_1 + r_3 - r_4 + r_5 + s_1 + s_3 - s_4 + s_5 + 1),$$

$$I_3 = (j_1 + r_4) \mathbf{B}(j_1 + j + r_4 + s_4, m + n - j_1 - j + r_2 + r_5 + s_2 + s_5 + 1)$$

$$\times (n - i + s_1 + s_3 - s_4 + s_5) \mathbf{B}(i_1 + i + 1, m + n - i_1 - i + r_1 + r_3 - r_4 + r_5 + s_1 + s_3 - s_4 + s_5),$$

$$I_4 = (m - j_1 + r_2 + r_5) \mathbf{B}(j_1 + j + r_4 + s_4 + 1, m + n - j_1 - j + r_2 + r_5 + s_2 + s_5)$$

$$\times (n - i + s_1 + s_3 - s_4 + s_5) \mathbf{B}(i_1 + i + 1, m + n - i_1 - i + r_1 + r_3 - r_4 + r_5 + s_1 + s_3 - s_4 + s_5).$$

Επιπλέον, εύκολα μπορεί να παραχθεί μία εναλλακτική έκφραση για την ποσότητα I μέσω της συνάρτησης Γάμμα

$$I = (-1)^{r_1+r_2+r_3+r_4+s_1+s_2+s_3+s_4} \theta_1^{r_1+r_2+r_3+r_5} \theta_2^{s_1+s_2+s_3+s_5} \\ \times [i(m - i_1 + r_3 - r_4 + r_5) - i_1(n - i + s_1 + s_3 - s_4 + s_5)] \\ \times [(j_1 + r_4)(n - j + s_2 + s_5) - (j + s_4)(m - j_1 + r_2 + r_5)] \\ \times \frac{\Gamma(j_1 + j + r_4 + s_4) \Gamma(m + n - j_1 - j + r_2 + r_5 + s_2 + s_5) \Gamma(i_1 + i)}{\Gamma(m + n + r_2 + r_4 + r_5 + s_2 + s_4 + s_5 + 1)} \\ \times \frac{\Gamma(m + n - i_1 - i + r_1 + r_3 - r_4 + r_5 + s_1 + s_3 - s_4 + s_5)}{\Gamma(m + n + r_1 + r_3 - r_4 + r_5 + s_1 + s_3 - s_4 + s_5 + 1)}.$$

4.5 Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

Σε αυτή την ενότητα, παρουσιάζονται πίνακες για ορισμένες τυπικές τιμές του εντός ελέγχου μέσου μήκους ροής (ARL_{in}) και του ποσοστού λανθασμένου συναγερωμού (FAR) για τη διευκόλυνση του σχεδιασμού του νέου διαγράμματος. Αν και τα ARL_{in} , FAR μπορούν να προσδιορισθούν αριθμητικά για οποιαδήποτε επιλογή των παραμέτρων σχεδιασμού, τα r και s καθορίστηκαν και τα δύο ίσα με $(n+1)/2$ και ως εκ τούτου χρησιμοποιήθηκαν οι διάμεσοι του δείγματος ελέγχου, ώστε να μειωθεί το πλήθος των παραμέτρων. Για τον ίδιο λόγο, οι παράμετροι σχεδιασμού a , b , c και d επιλέχθηκαν συμμετρικά ($b = m - a + 1$, $d = m - c + 1$) και έτσι τα όρια ελέγχου που προκύπτουν κάνουν χρήση συμμετρικών διατεταγμένων παρατηρήσεων, δηλαδή

$$LCL_X = X_{a:m}^{(R)}, \quad UCL_X = X_{b:m}^{(R)} \quad \text{και} \quad LCL_Y = Y_{c:m}^{(R)}, \quad UCL_Y = Y_{d:m}^{(R)}.$$

Υπάρχουν δύο τυπικές προσεγγίσεις που παραδοσιακά χρησιμοποιούνται για τον προσδιορισμό των παραμέτρων σχεδιασμού. Σύμφωνα με την πρώτη, πρέπει να προσδιοριστεί το FAR (συνήθως 1, 5 ή 10%), ενώ η δεύτερη προσέγγιση απαιτεί τη χρήση μιας συγκεκριμένης τιμής του ARL_{in} (όπως 200, 370 ή 500). Υπενθυμίζουμε ότι, στην περίπτωση των μη παραμετρικών διαγραμμάτων, η κατανομή του μήκους ροής συνήθως δεν είναι πάντα γεωμετρική και ως εκ τούτου το ARL_{in} δεν είναι ίσο με το αντίστροφο της πιθανότητας συναγερωμού (Chakraborti *et al.* (2004)).

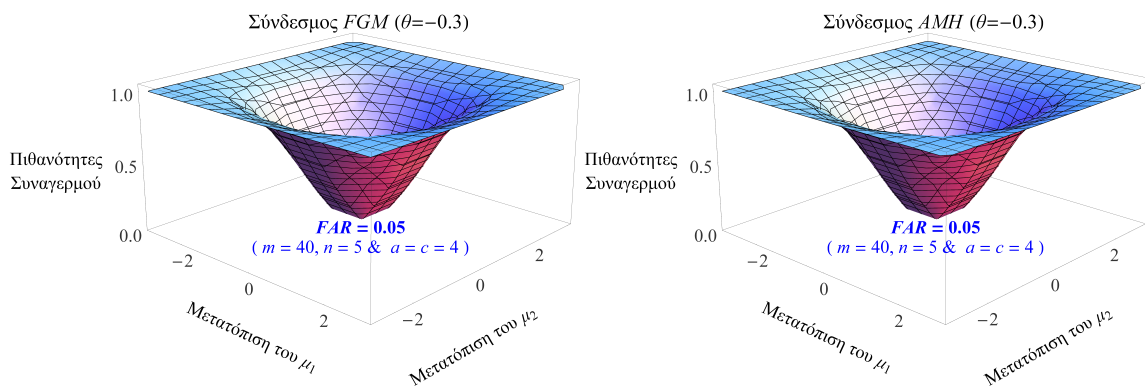
Το βασικό χαρακτηριστικό όλων των απαραμετρικών διαγραμμάτων ελέγχου είναι ότι τόσο οι τιμές του FAR όσο και οι τιμές του ARL_{in} είναι οι ίδιες για όλες τις συνεχείς, εντός ελέγχου, κατανομές. Όπως έχει ήδη αναφερθεί, η απόδοση του προτεινόμενου διαγράμματος είναι η ίδια για όλες τις συνεχείς περιθώριες κατανομές. Ωστόσο, επηρεάζεται θεωρητικά από τη δομή εξάρτησης των χαρακτηριστικών ενδιαφέροντος. Πρακτικά, όμως, η διαφορά μεταξύ των τιμών FAR και ARL_{in} , είναι αμελητέα, όταν χρησιμοποιούνται διαφορετικές κατανομές ή διαφορετικοί σύνδεσμοι. Το γεγονός αυτό καθιστά το διάγραμμα ελέγχου σχεδόν αμιγώς μη παραμετρικό.

Για τους αριθμητικούς μας υπολογισμούς, χρησιμοποιήσαμε τους συνδέσμους FGM και AMH , οι οποίοι έχουν ήδη εισαχθεί στην Ενότητα 4.4, καθώς και τους συνδέσμους των *Clayton-Pareto (CLP)*, *Gumbel-Hougaard (GUH)* και *Frank (FRA)*. Οι αθροιστικές συναρτήσεις όλων των προαναφερθέντων συνδέσμων, καθώς και άλλες πληροφορίες που σχετίζονται με αυτούς, δίνονται συγκεντρωτικά στον Πίνακα 2.5. Για κάθε σύνδεσμο, επιλέχθηκε μια συγκεκριμένη τιμή για την παράμετρο που συνδέεται με αυτόν, τόσο για

την εντός ελέγχου όσο και για την εκτός ελέγχου περίπτωση, ενώ έχουμε υποθέσει ότι οι περιθώριες κατανομές είναι κανονικές με μέση τιμή 0 και τυπική απόκλιση 1.

Το μέγεθος του δείγματος αναφοράς που επιλέχθηκε είναι ίσο με $m = 40$, ενώ εξετάστηκαν τρεις διαφορετικές επιλογές για το μέγεθος του δείγματος ελέγχου. Έχοντας καθορίσει όλες τις παραμέτρους που αναφέρονται παραπάνω, υπολογίσαμε τις κατάλληλες διατεταγμένες παρατηρήσεις (δηλαδή τις παραμέτρους $a = c$ και $b = d = m - a + 1$) με τις οποίες επιτυγχάνεται το επιθυμητό FAR (εφαρμόζοντας τρία διαφορετικά επίπεδα που κυμαίνονταν από 5 έως 10%. Το επόμενο βήμα ήταν να αξιολογηθεί το AR , καθώς η μέση τιμή των περιθωρίων κατανομών μετατοπιζόταν από $\mu_X^{(R)} = \mu_Y^{(R)} = 0$ σε μία εκτός ελέγχου τιμή $\mu_X^{(T)} = \mu_Y^{(T)} = \mu_{out}$. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 4.1, όπου οι τιμές με έντονους χαρακτήρες υποδηλώνουν τα FAR .

Στο Σχήμα 4.2, οι τιμές του AR απεικονίζονται σε τρισδιάστατα γραφήματα, όταν εμφανίζονται ίσες και άνισες μετατοπίσεις στη μέση τιμή των υπό μελέτη χαρακτηριστικών για ορισμένες σταθερές τιμές των υπόλοιπων παραμέτρων (δηλαδή για $m = 40, n = 5, a = c = 4$ και $b = d = 37$). Στα πλαίσια της παρούσας διατριβής παρουσιάζονται μόνο τα γραφήματα για τους συνδέσμους FGM και AMH , δεδομένου ότι στις υπόλοιπες περιπτώσεις παράγονται παρόμοια αποτελέσματα.



Σχήμα 4.2: Πιθανότητες Συναγερμού $AR_{O2}^{(2)}$ για τους συνδέσμους FGM & AMH και δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$)

Είναι προφανές, τόσο από τον Πίνακα 4.1 όσο και από το Σχήμα 4.2, ότι το (δίπλευρο) διάγραμμα $O2$ είναι ικανό να ανιχνεύσει μεταβολές της διεργασίας γρήγορα. Επιπλέον, όπως σαφέστατα παρουσιάζεται στον Πίνακα 4.1, το FAR παραμένει σχεδόν αμετάβλητο ανεξάρτητα από την επιλογή του συνδέσμου, έχοντας φυσικά διατηρήσει σταθερές τις τιμές των παραμέτρων m, n, r, s, a, b, c και d . Συνεπώς, παρόλο που η επίδοση του διαγράμματος

4.5. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | Μεταποπίσεις στη Μέση Τιμή : $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ | | | | | | | | |
|-------------------------------|-----|--------------------|--|------|------|------|-------------|------|------|------|------|
| Παράμετροι Σχεδιασμού | | | -2 | -1.5 | -1 | 0.5 | 0 | 0.5 | 1 | 1.5 | 2 |
| m | n | $(a, b) \& (c, d)$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | |
| 40 | 5 | (4, 37) | 0.98 | 0.85 | 0.49 | 0.15 | 0.05 | 0.16 | 0.49 | 0.85 | 0.98 |
| | | (5, 36) | 0.99 | 0.92 | 0.60 | 0.23 | 0.08 | 0.23 | 0.61 | 0.92 | 0.99 |
| | | (6, 35) | 1 | 0.95 | 0.71 | 0.30 | 0.13 | 0.30 | 0.70 | 0.95 | 1 |
| | 11 | (7, 34) | 1 | 0.99 | 0.78 | 0.26 | 0.05 | 0.25 | 0.79 | 0.99 | 1 |
| | | (8, 33) | 1 | 1 | 0.85 | 0.36 | 0.09 | 0.35 | 0.86 | 1 | 1 |
| | | (9, 32) | 1 | 1 | 0.91 | 0.44 | 0.13 | 0.44 | 0.91 | 1 | 1 |
| | 25 | (9, 32) | 1 | 1 | 0.94 | 0.34 | 0.03 | 0.35 | 0.94 | 1 | 1 |
| | | (10, 31) | 1 | 1 | 0.97 | 0.47 | 0.06 | 0.46 | 0.97 | 1 | 1 |
| | | (11, 30) | 1 | 1 | 0.98 | 0.58 | 0.10 | 0.58 | 0.99 | 1 | 1 |
| m | n | $(a, b) \& (c, d)$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | |
| 40 | 5 | (4, 37) | 0.97 | 0.81 | 0.46 | 0.14 | 0.05 | 0.15 | 0.47 | 0.81 | 0.97 |
| | | (5, 36) | 0.99 | 0.88 | 0.56 | 0.21 | 0.08 | 0.22 | 0.57 | 0.88 | 0.98 |
| | | (6, 35) | 0.99 | 0.93 | 0.66 | 0.28 | 0.12 | 0.28 | 0.66 | 0.92 | 0.99 |
| | 11 | (7, 34) | 1 | 0.98 | 0.74 | 0.24 | 0.05 | 0.25 | 0.75 | 0.98 | 1 |
| | | (8, 33) | 1 | 0.99 | 0.81 | 0.32 | 0.08 | 0.33 | 0.82 | 0.99 | 1 |
| | | (9, 32) | 1 | 0.99 | 0.88 | 0.41 | 0.13 | 0.42 | 0.87 | 0.99 | 1 |
| | 25 | (9, 32) | 1 | 1 | 0.90 | 0.32 | 0.03 | 0.33 | 0.91 | 1 | 1 |
| | | (10, 31) | 1 | 1 | 0.90 | 0.43 | 0.06 | 0.44 | 0.95 | 1 | 1 |
| | | (11, 30) | 1 | 1 | 0.97 | 0.53 | 0.10 | 0.54 | 0.98 | 1 | 1 |
| m | n | $(a, b) \& (c, d)$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | |
| 40 | 5 | (4, 37) | 0.98 | 0.85 | 0.50 | 0.16 | 0.05 | 0.16 | 0.50 | 0.85 | 0.98 |
| | | (5, 36) | 1 | 0.91 | 0.62 | 0.23 | 0.08 | 0.23 | 0.60 | 0.91 | 0.99 |
| | | (6, 35) | 1 | 0.95 | 0.70 | 0.31 | 0.12 | 0.30 | 0.71 | 0.95 | 1 |
| | 11 | (7, 34) | 1 | 0.99 | 0.79 | 0.26 | 0.05 | 0.26 | 0.79 | 1 | 1 |
| | | (8, 33) | 1 | 1 | 0.86 | 0.35 | 0.08 | 0.35 | 0.86 | 1 | 1 |
| | | (9, 32) | 1 | 1 | 0.91 | 0.44 | 0.13 | 0.45 | 0.91 | 1 | 1 |
| | 25 | (9, 32) | 1 | 1 | 0.93 | 0.35 | 0.04 | 0.35 | 0.94 | 1 | 1 |
| | | (10, 31) | 1 | 1 | 0.97 | 0.46 | 0.06 | 0.46 | 0.97 | 1 | 1 |
| | | (11, 30) | 1 | 1 | 0.99 | 0.54 | 0.10 | 0.58 | 0.99 | 1 | 1 |
| m | n | $(a, b) \& (c, d)$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 2$) | | | | | | | | |
| 40 | 5 | (4, 37) | 0.99 | 0.86 | 0.50 | 0.16 | 0.05 | 0.15 | 0.50 | 0.86 | 0.99 |
| | | (5, 36) | 0.99 | 0.92 | 0.61 | 0.30 | 0.08 | 0.23 | 0.62 | 0.92 | 0.99 |
| | | (6, 35) | 1 | 0.95 | 0.70 | 0.31 | 0.12 | 0.30 | 0.70 | 0.95 | 1 |
| | 11 | (7, 34) | 1 | 0.99 | 0.79 | 0.26 | 0.05 | 0.26 | 0.79 | 0.99 | 1 |
| | | (8, 33) | 1 | 1 | 0.86 | 0.35 | 0.08 | 0.35 | 0.86 | 1 | 1 |
| | | (9, 32) | 1 | 1 | 0.91 | 0.45 | 0.13 | 0.46 | 0.91 | 1 | 1 |
| | 25 | (9, 32) | 1 | 1 | 0.94 | 0.35 | 0.04 | 0.36 | 0.94 | 1 | 1 |
| | | (10, 31) | 1 | 1 | 0.97 | 0.46 | 0.06 | 0.47 | 0.97 | 1 | 1 |
| | | (11, 30) | 1 | 1 | 0.98 | 0.57 | 0.10 | 0.58 | 0.99 | 1 | 1 |
| m | n | $(a, b) \& (c, d)$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 2$) | | | | | | | | |
| 40 | 5 | (4, 37) | 0.95 | 0.76 | 0.43 | 0.14 | 0.05 | 0.13 | 0.42 | 0.76 | 0.95 |
| | | (5, 36) | 0.97 | 0.84 | 0.52 | 0.20 | 0.08 | 0.19 | 0.51 | 0.83 | 0.97 |
| | | (6, 35) | 0.98 | 0.87 | 0.60 | 0.26 | 0.12 | 0.25 | 0.59 | 0.88 | 0.99 |
| | 11 | (7, 34) | 1 | 0.96 | 0.70 | 0.23 | 0.05 | 0.21 | 0.68 | 0.95 | 1 |
| | | (8, 33) | 1 | 0.98 | 0.77 | 0.30 | 0.08 | 0.29 | 0.77 | 0.98 | 1 |
| | | (9, 32) | 1 | 0.98 | 0.83 | 0.38 | 0.11 | 0.37 | 0.83 | 0.99 | 1 |
| | 25 | (9, 32) | 1 | 1 | 0.88 | 0.31 | 0.03 | 0.29 | 0.86 | 1 | 1 |
| | | (10, 31) | 1 | 1 | 0.92 | 0.40 | 0.05 | 0.39 | 0.92 | 1 | 1 |
| | | (11, 30) | 1 | 1 | 0.96 | 0.49 | 0.09 | 0.50 | 0.95 | 1 | 1 |

Πίνακας 4.1: Πιθανότητες Συναγερμού $AR_{02}^{(2)}$ του δίπλευρου διαγράμματος $O2$ λόγω μεταποπίσεων μέσης τιμής για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$)

επιηρεάζεται τυπικά από τη δομή εξάρτησης των υποκείμενων χαρακτηριστικών (υπό την έννοια ότι οι τύποι του Πορίσματος 4.2 εξαρτώνται από το σύνδεσμο $C(u, v)$), η τιμή του FAR δε διαφοροποιείται σημαντικά, όταν χρησιμοποιούνται διαφορετικοί σύνδεσμοι. Το ίδιο ισχύει και για τις τιμές του ARL_{in} , όπως θα δούμε στη συνέχεια.

Προκειμένου να διευκολυνθεί η διαδικασία δημιουργίας ενός κατάλληλου δισδιάστατου διαγράμματος ελέγχου, όπως αυτό περιγράφεται στο παρόν κεφάλαιο, δημιουργήσαμε τους Πίνακες 4.2 και 4.3, όπου παρουσιάζονται οι τιμές του FAR των αντιστοίχων προτεινόμενων δίπλευρων και μονόπλευρων διαγραμμάτων για διάφορες επιλογές των παραμέτρων σχεδιασμού m, n, r, s, a, b, c και d . Οι υπολογισμοί έγιναν με τη χρήση των διαμέσων και στα δύο χαρακτηριστικά έτσι, ώστε να μειωθεί ο αριθμός των επιλογών για τις τιμές των παραμέτρων. Σε αυτή την περίπτωση, είναι εύλογο να χρησιμοποιηθούν συμμετρικά όρια ελέγχου, επομένως $b = m - a + 1$ και $d = m - c + 1$.

Παρά το γεγονός ότι οι τιμές του FAR υπολογίστηκαν υποθέτοντας ότι η υποκείμενη δομή εξάρτησης περιγράφεται από το σύνδεσμο FGM , όπου $C_\theta(u, v) = uv[1 + \theta(1 - u)(1 - v)]$ με $\theta = 1$, μπορεί να χρησιμοποιηθεί οποιαδήποτε δομή εξάρτησης επειδή, όπως ήδη αναφέρθηκε προηγουμένως, το FAR παραμένει πρακτικά ανεπηρέαστο από την επιλογή του συνδέσμου. Αν θέλουμε να ανιχνεύσουμε ίσες μέσες μετατοπίσεις και στα δύο χαρακτηριστικά, το a πρέπει να είναι ίσο με c ($a = c$), διαφορετικά τα a και c θα πρέπει να διαφέρουν ($a \neq c$).

Δεδομένου ότι το προτεινόμενο διάγραμμα δεν απαιτεί καμία υπόθεση σχετικά με το είδος της κατανομής, μπορεί να χρησιμοποιηθεί είτε για συμμετρικές είτε για ασύμμετρες κατανομές. Στην τελευταία περίπτωση, προτείνουμε οι παράμετροι σχεδιασμού να μην επιλέγονται συμμετρικά από τα άκρα του διαστήματος του δείγματος αναφοράς. Για παράδειγμα, η επιλογή των ορίων θα μπορούσε να γίνει έτσι, ώστε να ισχύει μία αναλογία μεταξύ των διαστημάτων που ορίζονται από τα r, s προς τα αντίστοιχα διαστήματα που καθορίζουν οι διατεταγμένες παρατηρήσεις του δείγματος αναφοράς, δηλαδή να ισχύουν τα εξής

$$\frac{r}{a} = \frac{n + 1}{m + a - b + 1} \quad \text{και} \quad \frac{s}{c} = \frac{n + 1}{m + c - d + 1}.$$

Βάσει των παραπάνω σχέσεων, στην ειδική περίπτωση που $r = s = (n + 1)/2$ και το n είναι περιττός αριθμός, επαληθεύεται ότι $b = m - a + 1$ και $d = m - c + 1$. Σημειώνεται ότι η διαδικασία προσδιορισμού μη συμμετρικών ορίων ελέγχου που περιγράφηκε παραπάνω μπορεί να εφαρμοσθεί έχοντας προκαθορίσει μία τιμή είτε για το FAR είτε για το ARL_{in} .

4.5. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| m | n | $a = c$ | FAR | $a > c$ | FAR | $a < c$ | FAR | | | |
|------|-----|---------|--------|---------|-------|---------|--------|-----|--------|--------|
| 50 | 5 | 5 | 0.0466 | 6 | 3 | 0.0446 | 2 | 7 | 0.0517 | |
| | | 6 | 0.0740 | 7 | 4 | 0.0659 | 3 | 8 | 0.0799 | |
| | | 7 | 0.1010 | 8 | 5 | 0.0975 | 3 | 9 | 0.1049 | |
| | 11 | 9 | 0.0458 | 11 | 5 | 0.0580 | 7 | 10 | 0.0490 | |
| | | 10 | 0.0748 | 12 | 7 | 0.0867 | 8 | 11 | 0.0679 | |
| | | 11 | 0.1043 | 12 | 8 | 0.0941 | 9 | 12 | 0.1004 | |
| | 25 | 12 | 0.0404 | 14 | 10 | 0.0507 | 8 | 14 | 0.0524 | |
| | | 13 | 0.0565 | 15 | 11 | 0.0819 | 10 | 15 | 0.0799 | |
| | | 14 | 0.1002 | 15 | 12 | 0.0977 | 12 | 15 | 0.0987 | |
| | 100 | 5 | 11 | 0.0511 | 12 | 10 | 0.0516 | 8 | 13 | 0.0537 |
| | | | 13 | 0.0741 | 14 | 11 | 0.0733 | 11 | 14 | 0.0727 |
| | | | 14 | 0.0957 | 16 | 12 | 0.1020 | 13 | 16 | 0.1041 |
| | | 11 | 19 | 0.0479 | 22 | 11 | 0.0512 | 14 | 22 | 0.0537 |
| | | | 21 | 0.0723 | 24 | 12 | 0.0702 | 15 | 24 | 0.0777 |
| 23 | | | 0.1071 | 26 | 13 | 0.1018 | 16 | 26 | 0.1081 | |
| 25 | | 27 | 0.0481 | 29 | 12 | 0.0447 | 14 | 29 | 0.0459 | |
| | | 29 | 0.0868 | 31 | 13 | 0.0718 | 15 | 31 | 0.0708 | |
| | | 30 | 0.1081 | 33 | 15 | 0.1078 | 16 | 33 | 0.1021 | |
| | | | | | | | | | | |
| 200 | 5 | 23 | 0.0513 | 25 | 18 | 0.0480 | 13 | 27 | 0.0503 | |
| | | 26 | 0.0749 | 30 | 22 | 0.0772 | 18 | 33 | 0.0782 | |
| | | 29 | 0.0951 | 32 | 27 | 0.1033 | 20 | 37 | 0.1098 | |
| | 11 | 39 | 0.0504 | 43 | 31 | 0.0490 | 34 | 44 | 0.0545 | |
| | | 43 | 0.0766 | 46 | 34 | 0.0654 | 38 | 46 | 0.0702 | |
| | | 47 | 0.1070 | 51 | 35 | 0.1020 | 42 | 50 | 0.1035 | |
| | 25 | 55 | 0.0462 | 61 | 36 | 0.0483 | 47 | 59 | 0.0458 | |
| | | 58 | 0.0702 | 65 | 40 | 0.0787 | 51 | 63 | 0.0781 | |
| | | 61 | 0.1009 | 67 | 41 | 0.1028 | 55 | 65 | 0.1048 | |
| | | | | | | | | | | |
| 500 | 5 | 56 | 0.0480 | 60 | 50 | 0.0470 | 45 | 65 | 0.0529 | |
| | | 65 | 0.0714 | 70 | 63 | 0.0708 | 52 | 74 | 0.0727 | |
| | | 74 | 0.1028 | 76 | 69 | 0.0986 | 60 | 80 | 0.1006 | |
| | 11 | 100 | 0.0503 | 110 | 88 | 0.0468 | 90 | 112 | 0.0521 | |
| | | 109 | 0.0730 | 120 | 94 | 0.0770 | 100 | 116 | 0.0720 | |
| | | 118 | 0.1040 | 125 | 100 | 0.0972 | 110 | 120 | 0.0981 | |
| | 25 | 142 | 0.0486 | 147 | 135 | 0.0476 | 130 | 150 | 0.0485 | |
| | | 149 | 0.0731 | 153 | 149 | 0.0793 | 136 | 156 | 0.0686 | |
| | | 156 | 0.1017 | 157 | 152 | 0.0959 | 143 | 163 | 0.0982 | |
| | | | | | | | | | | |
| 1000 | 5 | 116 | 0.0502 | 127 | 100 | 0.0498 | 100 | 130 | 0.0531 | |
| | | 133 | 0.0793 | 136 | 126 | 0.0720 | 116 | 142 | 0.0708 | |
| | | 150 | 0.1025 | 160 | 133 | 0.1024 | 132 | 160 | 0.0970 | |
| | 11 | 202 | 0.0508 | 208 | 195 | 0.0484 | 180 | 220 | 0.0520 | |
| | | 221 | 0.0777 | 232 | 197 | 0.0721 | 200 | 230 | 0.0716 | |
| | | 240 | 0.1074 | 260 | 200 | 0.1033 | 224 | 245 | 0.0979 | |
| | 25 | 286 | 0.0503 | 312 | 210 | 0.0554 | 262 | 300 | 0.0480 | |
| | | 301 | 0.0766 | 316 | 270 | 0.0709 | 286 | 310 | 0.0713 | |
| | | 316 | 0.1080 | 320 | 303 | 0.0988 | 290 | 330 | 0.1027 | |
| | | | | | | | | | | |

Πίνακας 4.2: Πιθανότητες Λανθασμένου Συναγερμού $FAR_{O2}^{(2)}$ του δίπλευρου διαγράμματος $O2$ για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$)

| m | n | $b = d$ | FAR | $b > d$ | FAR | $b < d$ | FAR | | |
|------|-----|---------|--------|---------|-------|---------|-------|-----|--------|
| 50 | 5 | 44 | 0.0505 | 47 | 43 | 0.0459 | 42 | 48 | 0.0520 |
| | | 43 | 0.0735 | 46 | 41 | 0.0733 | 41 | 47 | 0.0705 |
| | | 42 | 0.0952 | 45 | 40 | 0.0972 | 40 | 44 | 0.1063 |
| | 11 | 40 | 0.0511 | 46 | 38 | 0.0536 | 38 | 50 | 0.0504 |
| | | 39 | 0.0740 | 45 | 37 | 0.0728 | 37 | 39 | 0.0719 |
| | | 38 | 0.0968 | 44 | 36 | 0.1007 | 36 | 47 | 0.0959 |
| | 25 | 37 | 0.0477 | 48 | 35 | 0.0546 | 35 | 43 | 0.0580 |
| | | 36 | 0.0711 | 46 | 34 | 0.0745 | 34 | 42 | 0.0758 |
| | | 35 | 0.1053 | 44 | 33 | 0.1050 | 33 | 41 | 0.1059 |
| 100 | 5 | 86 | 0.0498 | 89 | 85 | 0.0520 | 83 | 94 | 0.0488 |
| | | 84 | 0.0763 | 87 | 83 | 0.0717 | 80 | 90 | 0.0784 |
| | | 82 | 0.1005 | 85 | 80 | 0.0982 | 77 | 89 | 0.1058 |
| | 11 | 78 | 0.0501 | 85 | 70 | 0.0507 | 76 | 82 | 0.0507 |
| | | 76 | 0.0777 | 83 | 73 | 0.0722 | 74 | 81 | 0.0709 |
| | | 74 | 0.1064 | 77 | 72 | 0.0998 | 72 | 80 | 0.0951 |
| | 25 | 71 | 0.0498 | 75 | 70 | 0.0474 | 69 | 87 | 0.0457 |
| | | 70 | 0.0681 | 75 | 67 | 0.0750 | 66 | 85 | 0.0726 |
| | | 68 | 0.0979 | 75 | 65 | 0.1009 | 64 | 83 | 0.1043 |
| 200 | 5 | 172 | 0.0520 | 174 | 170 | 0.0511 | 165 | 180 | 0.0511 |
| | | 167 | 0.0741 | 170 | 165 | 0.0708 | 161 | 175 | 0.0725 |
| | | 162 | 0.1034 | 166 | 160 | 0.1036 | 157 | 170 | 0.1017 |
| | 11 | 155 | 0.0502 | 160 | 150 | 0.0534 | 150 | 171 | 0.0468 |
| | | 151 | 0.0719 | 158 | 148 | 0.0657 | 145 | 168 | 0.0692 |
| | | 147 | 0.1033 | 152 | 142 | 0.1064 | 140 | 166 | 0.0955 |
| | 25 | 140 | 0.0501 | 160 | 135 | 0.0492 | 136 | 150 | 0.0481 |
| | | 138 | 0.0673 | 155 | 130 | 0.0774 | 132 | 147 | 0.0692 |
| | | 134 | 0.0994 | 152 | 128 | 0.0972 | 128 | 145 | 0.1034 |
| 500 | 5 | 427 | 0.0506 | 437 | 420 | 0.0484 | 410 | 445 | 0.0530 |
| | | 416 | 0.0737 | 428 | 409 | 0.0729 | 399 | 435 | 0.0795 |
| | | 405 | 0.0974 | 419 | 398 | 0.0980 | 388 | 425 | 0.1036 |
| | 11 | 383 | 0.0519 | 450 | 365 | 0.0511 | 373 | 409 | 0.0490 |
| | | 374 | 0.0674 | 425 | 357 | 0.0715 | 360 | 400 | 0.0729 |
| | | 365 | 0.1015 | 400 | 348 | 0.0984 | 345 | 396 | 0.1013 |
| | 25 | 344 | 0.0520 | 370 | 335 | 0.0498 | 335 | 360 | 0.0506 |
| | | 337 | 0.0772 | 360 | 327 | 0.0687 | 328 | 355 | 0.0705 |
| | | 330 | 0.1031 | 350 | 320 | 0.0981 | 318 | 350 | 0.1069 |
| 1000 | 5 | 853 | 0.0498 | 865 | 845 | 0.0497 | 840 | 865 | 0.0504 |
| | | 830 | 0.0735 | 842 | 817 | 0.0750 | 825 | 855 | 0.0706 |
| | | 807 | 0.1047 | 820 | 790 | 0.1041 | 800 | 815 | 0.1017 |
| | 11 | 771 | 0.0478 | 775 | 755 | 0.0506 | 750 | 800 | 0.0474 |
| | | 748 | 0.0713 | 765 | 735 | 0.0685 | 730 | 770 | 0.0682 |
| | | 725 | 0.1066 | 760 | 705 | 0.1034 | 710 | 740 | 0.1032 |
| | 25 | 693 | 0.0447 | 700 | 680 | 0.0495 | 670 | 720 | 0.0504 |
| | | 674 | 0.0684 | 680 | 665 | 0.0721 | 660 | 700 | 0.0713 |
| | | 655 | 0.0969 | 660 | 650 | 0.1061 | 640 | 680 | 0.1043 |

Πίνακας 4.3: Πιθανότητες Λανθασμένου Συναγερμού $FAR_{O2}^{(2)}$ του μονόπλευρου διαγράμματος $O2$ για δεδομένο σχεδιασμό ($b = m - a + 1$, $d = m - c + 1$ & $r = s = (n + 1)/2$)

Είναι αυτονόητο ότι η επιλογή των παραμέτρων για την επίτευξη ενός προκαθορισμένου επιθυμητού FAR δεν είναι μοναδική. Έτσι, εκτός από τις επιλογές που παρουσιάζονται στους Πίνακες 4.2 και 4.3 μπορεί κανείς να πάρει πρόσθετες τιμές διατηρώντας το ίδιο (περίπου) FAR . Ως εκ τούτου, το τμήμα ελέγχου ποιότητας μιας βιομηχανίας, αν επιθυμεί να εκμεταλλευτεί το νέο σχήμα, έχει την ευελιξία να δημιουργήσει έναν αποδεκτό συνδυασμό όλων των παραμέτρων σύμφωνα με τις τρέχουσες ερευνητικές του ανάγκες. Συνολικά, οι προαναφερθέντες πίνακες μπορούν να αποτελέσουν οδηγούς για τη διαμόρφωση του νέου δισδιάστατου διαγράμματος λαμβάνοντας υπ' όψιν τα δίπλευρα και μονόπλευρα όρια ελέγχου στην περίπτωση συμμετρικών ή μη συμμετρικών κατανομών.

4.6 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

Προκειμένου να συγκριθεί η απόδοση του προτεινόμενου (δίπλευρου) διαγράμματος με αντίστοιχα παραμετρικά και μη παραμετρικά διαγράμματα ελέγχου, πραγματοποιήθηκε μια μελέτη προσομοίωσης χρησιμοποιώντας τη δεύτερη μέθοδο που περιγράφηκε στην αρχή της Ενότητας 4.5, δηλαδή την προσέγγιση βάσει του ARL . Τα βήματα της αλγοριθμικής διαδικασίας σύμφωνα με την οποία υπολογίζονται το εντός και το εκτός ελέγχου μέσο μήκος ροής παρουσιάζονται στον Πίνακα 4.4.

Για το σκοπό αυτό, προσομοιώσαμε δεδομένα από μια δισδιάστατη κανονική κατανομή και μία κατανομή $t(5)$, δηλαδή της κατανομής t (*student*) 5 β.ε. με τις εξής παραμέτρους

$$\mu_{in} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma_{in} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \text{και} \quad \Sigma_{t(5)} = \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.6 \end{pmatrix} = \frac{3}{5} \Sigma_{in}.$$

Η επιλογή των παραπάνω παραμέτρων βασίστηκε στη δημοσίευση των Boone and Chakraborti (2012), οι οποίοι παρουσίασαν μια σχετική μελέτη προσομοίωσης για δισδιάστατα μη παραμετρικά διαγράμματα ελέγχου. Αξίζει να σημειωθεί ότι η κατανομή t είναι μια από τις πιο δημοφιλείς πολυδιάστατες μη κανονικές κατανομές που χρησιμοποιούνται σε μελέτες προσομοίωσης. Θα πρέπει, επίσης, να τονίσουμε ότι ο πίνακας $\Sigma_{t(5)}$ επιλέχθηκε έτσι, ώστε να επιτευχθεί ο ίδιος πίνακας διακυμάνσεων-συνδιακυμάνσεων και για τις δύο κατανομές, εξασφαλίζοντας μια δίκαιη σύγκριση μεταξύ των δύο εναλλακτικών. Στη μελέτη μας χρησιμοποιήσαμε ένα μέγεθος δείγματος αναφοράς $m = 1000$ για να αναπαραχθούν τα αποτελέσματα των Boone and Chakraborti (2012). Επιπλέον, επιλέξαμε $a = c$, όταν συμβαίνουν ίσες μετατοπίσεις και $a \neq c$, στην περίπτωση άνισων μετατοπίσεων.

- Βήμα 0.** Θέτουμε $ARL = 0$, $SUM = 0$ και δηλώνουμε τις αρχικές τιμές των παραμέτρων $m, n, a, b, c, d, r, s, N_0$ (π.χ. $N_0 = 10000$).
- Βήμα 1.** Θέτουμε $RL = 0$, $alarm = False$ και παράγουμε m ζεύγη παρατηρήσεων του δείγματος αναφοράς από την εντός ελέγχου κατανομή F , συμβ. $(X_i^{(R)}, Y_i^{(R)}) \sim F(x, y)$, $i = 1, 2, \dots, m$.
- Βήμα 2.** Δημιουργούμε τα διατεταγμένα ζεύγη $(X_{i:m}^{(R)}, Y_{i:m}^{(R)})$, $i = 1, 2, \dots, m$, και υπολογίζουμε τα όρια ελέγχου LCL_X , UCL_X , LCL_Y , UCL_Y επιλέγοντας κατάλληλα τις τιμές των παραμέτρων a, b, c, d , ώστε να επιτευχθεί η προκαθορισμένη τιμή του ARL_{in} .
- Βήμα 3.** Εάν $alarm = True$, τότε παράγουμε n ζεύγη παρατηρήσεων του δείγματος ελέγχου από την κατανομή G , συμβ. $(X_i^{(T)}, Y_i^{(T)}) \sim G(x, y)$, $i = 1, 2, \dots, n$.
- Βήμα 4.** Υπολογίζουμε τις διατεταγμένες παρατηρήσεις $X_{r:n}^{(T)}$, $Y_{s:n}^{(T)}$ του δείγματος ελέγχου και θέτουμε $RL = RL + 1$.
- Βήμα 5.** Διαπιστώνουμε εάν παραβιάζεται ή όχι ο κανόνας ελέγχου της Σχέσης (4.2). Εάν ο κανόνας παραβιάζεται, τότε θέτουμε $alarm = True$ και σταματάμε. Εάν ο κανόνας δεν παραβιάζεται, επιστρέφουμε στο Βήμα 3.
- Βήμα 6.** Υπολογίζουμε το μήκος ροής θέτοντας $SUM = SUM + RL$.
- Βήμα 7.** Επαναλαμβάνουμε N_0 φορές τα Βήματα 1 έως 6.
- Βήμα 8.** Τυπώνουμε την εκτίμηση του μέσου μήκους ροής $ARL = SUM/N_0$.
 Προφανώς, εάν $F(x, y) = G(x, y)$, τότε $ARL \equiv ARL_{in}$.
 Διαφορετικά, εάν $F(x, y) \neq G(x, y)$, τότε $ARL \equiv ARL_{out}$.

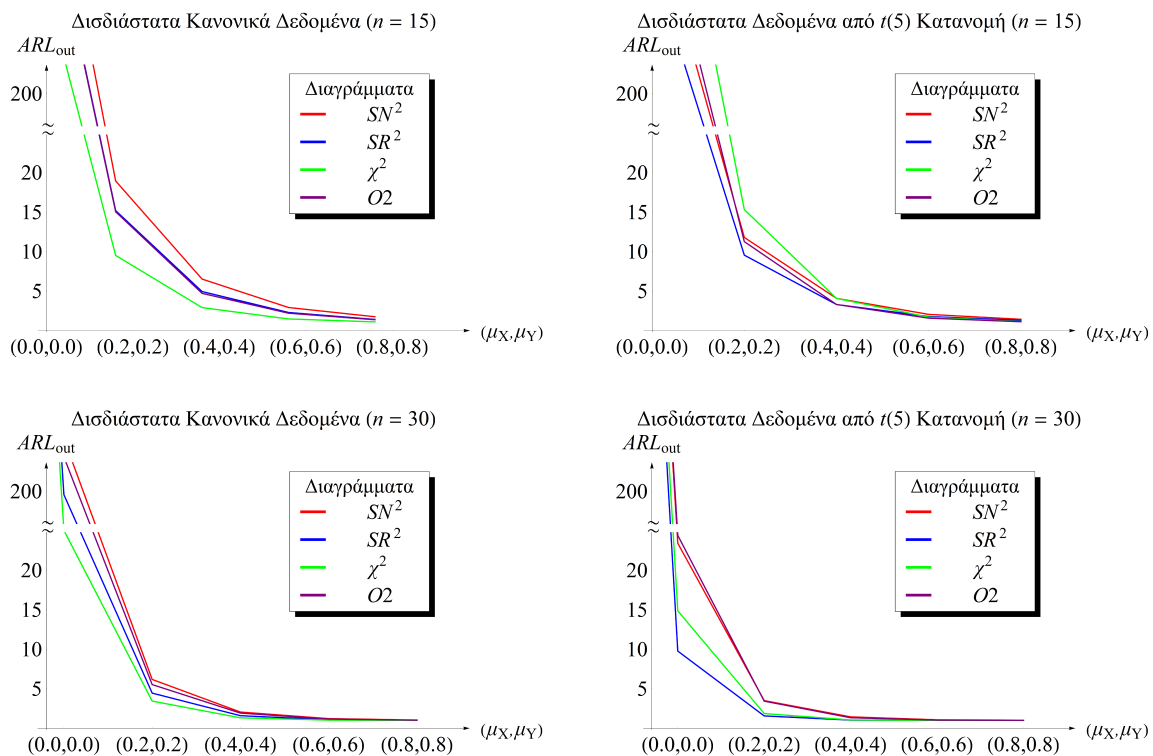
Πίνακας 4.4: Αλγόριθμος υπολογισμού ARL_{in} και ARL_{out} του διαγράμματος O2

Σχετικά με την επιλογή των παραμέτρων a και c , πρέπει να ληφθεί υπ' όψιν η μετατόπιση του χαρακτηριστικού που επιθυμεί κανείς να ανιχνεύσει. Για να το θέσουμε διαφορετικά, εάν κάποιος ενδιαφέρεται να ανιχνεύσει μια πιθανή μετατόπιση μόνο στη μεταβλητή X (ή πολύ μεγαλύτερη μετατόπιση στη μεταβλητή X σε σχέση με τη μετατόπιση της μεταβλητής Y), τότε θα ήταν προτιμότερο να επιλέξει $a > c$, ώστε με αποτέλεσμα τα όρια ελέγχου του συγκεκριμένου χαρακτηριστικού να είναι στενότερα. Στην πραγματικότητα, όπως θα διαπιστώσουμε και στην Ενότητα 4.6, το νέο σχήμα φαίνεται να λειτουργεί ελαφρώς καλύτερα όταν $\mu_1 > \mu_2$ και $a > c$ σε σχέση με ανταγωνιστικά διαγράμματα ελέγχου.

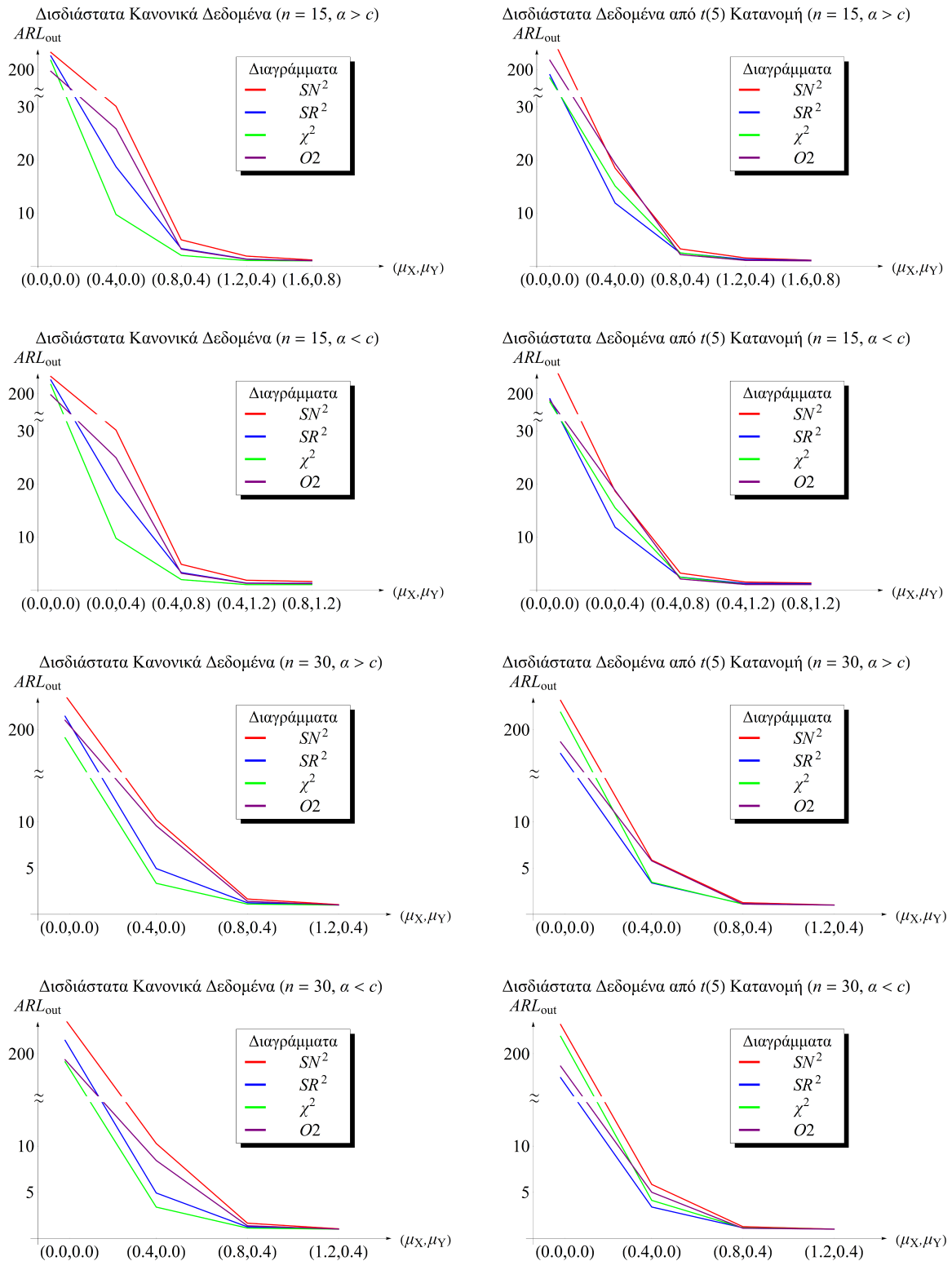
Στα Σχήματα 4.3–4.4 και στους Πίνακες 4.5–4.6, το νέο διάγραμμα συγκρίνεται με τα διαγράμματα SN^2 και SR^2 των Boone and Chakraborti (2012), καθώς και με το παραδοσιακό

4.6. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

χ^2 όσον αφορά τις εκτός ελέγχου τιμές του ARL . Θα πρέπει να αναφέρουμε ότι τα όρια ελέγχου για τα διαγράμματα SN^2 , SR^2 , χ^2 που χρησιμοποιήθηκαν δεν ήταν τα ασυμπτωτικά όρια $\chi^2_{p,a}$, αλλά αυτά που εκμεταλλεύτηκαν οι Boone and Chakraborti (2012) για να επιτύχουν το επιθυμητό ARL_{in} . Για να εκτιμηθούν οι εντός και εκτός ελέγχου τιμές του ARL για το προτεινόμενο διάγραμμα, χρησιμοποιήθηκαν οι διάμεσοι και στα δύο δείγματα, ενώ ταυτόχρονα έγιναν οι ίδιες υποθέσεις, όπως και προηγουμένως, σχετικά με τις παραμέτρους σχεδιασμού. Οι αριθμητικές τιμές του ARL παρουσιάζονται στον Πίνακα 4.3 και στον Πίνακα 4.4, ενώ τα αποτελέσματα απεικονίζονται στο Σχήμα 4.3 και στο Σχήμα 4.4, ώστε να έχουμε μια καλύτερη γραφική εικόνα. Πρέπει να αναφερθεί ότι οι έντονες τιμές, που εμφανίζονται στους Πίνακες 4.3 και 4.4, υποδεικνύουν τη χαμηλότερη τιμή του ARL και ως εκ τούτου το διάγραμμα με τις καλύτερες επιδόσεις.



Σχήμα 4.3: ARL_{out} των διαγραμμάτων SN^2 , SR^2 , χ^2 , $O2$ στην περίπτωση ίσων μετατοπίσεων της μέσης τιμής



Σχήμα 4.4: ARL_{out} των διαγραμμάτων $SN^2, SR^2, \chi^2, O2$ στην περίπτωση άνισων μετατοπίσεων της μέσης τιμής

4.6. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

| n | Μεταποτίσεις Μέσης Τιμής | Διοδίαστατα Δεδομένα από Κανονική Κατανομή | | | Διοδίαστατα Δεδομένα από $t(5)$ Κατανομή | | |
|-----------|---|--|------------------|--------------------|--|------------------|--------------------|
| | | Διάγραμμα SN^2 | Διάγραμμα SR^2 | Διάγραμμα χ^2 | Διάγραμμα SN^2 | Διάγραμμα SR^2 | Διάγραμμα χ^2 |
| 15 | $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ | $UCL = 8.69$ | $UCL = 8.78$ | $UCL = 10.60$ | $UCL = 8.70$ | $UCL = 8.73$ | $UCL = 12.22$ |
| | (0,0,0,0) | 203.27 | 202.60 | 201.76 | 207.02 | 199.07 | 198.44 |
| | (0,2,0,2) | 75.35 | 66.42 | 51.40 | 53.46 | 46.73 | 74.60 |
| | (0,4,0,4) | 18.94 | 15.19 | 9.52 | 11.79 | 9.55 | 15.32 |
| | (0,6,0,6) | 6.52 | 4.94 | 2.91 | 4.08 | 3.30 | 4.07 |
| | (0,8,0,8) | 2.92 | 2.29 | 1.47 | 2.05 | 1.78 | 1.70 |
| (1,0,1,0) | 1.75 | 1.41 | 1.10 | 1.43 | 1.29 | 1.15 | |
| 30 | $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ | $UCL = 9.80$ | $UCL = 9.72$ | $UCL = 10.60$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ |
| | (0,0,0,0) | 203.74 | 201.46 | 199.13 | 203.18 | 197.42 | 201.90 |
| | (0,2,0,2) | 36.66 | 29.57 | 25.05 | 23.42 | 19.71 | 32.56 |
| | (0,4,0,4) | 6.18 | 4.45 | 3.44 | 3.50 | 3.07 | 4.11 |
| | (0,6,0,6) | 2.05 | 1.59 | 1.32 | 1.44 | 1.34 | 1.41 |
| | (0,8,0,8) | 1.22 | 1.08 | 1.03 | 1.07 | 1.04 | 1.04 |
| (1,0,1,0) | 1.04 | 1.01 | 1.00 | 1.01 | 1.00 | 1.00 | |

Πίνακας 4.5: Απόδοση του δίπλευρου διαγράμματος $O2$ βάσει του ARL_{out} σε σχέση με 3 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μεταποτίσεων ίσων μέσης τιμής, χρησιμοποιώντας δεδομένα από κανονική και $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$)

| n | Μεταποπίσεις | Δισδιάστατα Δεδομένα από Κανονική Κατανομή | | | | Δισδιάστατα Δεδομένα από $t(5)$ Κατανομή | | | | | | | | | | | | | | | |
|------------|-----------------------------|--|--------------|--------------------|---------------|--|--------------|--------------------|---------------|--------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|
| | | Διάγραμμα SN^2 | | Διάγραμμα χ^2 | | Διάγραμμα SN^2 | | Διάγραμμα χ^2 | | | | | | | | | | | | | |
| | | $UCL = 8.69$ | $UCL = 8.68$ | $UCL = 10.60$ | $UCL = 10.60$ | $UCL = 8.70$ | $UCL = 8.73$ | $UCL = 12.22$ | $UCL = 12.22$ | | | | | | | | | | | | |
| 15 | Μέσης Τιμής | | | | | | | | | | | | | | | | | | | | |
| | $\mu_X^{(T)} > \mu_Y^{(T)}$ | 203.27 | 202.60 | 201.76 | 199.71 | 207.02 | 199.07 | 198.44 | 201.79 | 18.47 | 11.89 | 15.10 | 19.33 | 3.25 | 2.47 | 2.50 | 1.54 | 1.27 | 1.09 | 1.08 | |
| | (0.0, 0.0) | 30.10 | 18.71 | 9.72 | 25.88 | 18.65 | 11.88 | 15.52 | 18.80 | 3.22 | 2.47 | 2.43 | 2.12 | 1.54 | 1.28 | 1.10 | 1.24 | 1.09 | 1.07 | 1.09 | |
| | (0.4, 0.4) | 4.96 | 3.31 | 2.02 | 3.19 | 3.22 | 2.47 | 2.43 | 2.12 | 1.54 | 1.28 | 1.10 | 1.08 | 1.54 | 1.28 | 1.10 | 1.24 | 1.09 | 1.07 | 1.09 | |
| | (0.8, 0.4) | 1.89 | 1.33 | 1.06 | 1.30 | 1.37 | 1.24 | 1.09 | 1.07 | 1.13 | 1.07 | 1.00 | 1.00 | 1.13 | 1.07 | 1.00 | 1.07 | 1.00 | 1.00 | 1.00 | |
| | (1.2, 0.4) | 1.17 | 1.04 | 1.00 | 1.02 | 1.37 | 1.24 | 1.09 | 1.07 | 1.13 | 1.07 | 1.00 | 1.00 | 1.13 | 1.07 | 1.00 | 1.07 | 1.00 | 1.00 | 1.00 | |
| | (1.6, 0.8) | | | | | | | | | | | | | | | | | | | | |
| | $\mu_X^{(T)} < \mu_Y^{(T)}$ | $UCL = 8.69$ | $UCL = 8.78$ | $UCL = 10.60$ | $UCL = 10.60$ | $UCL = 8.70$ | $UCL = 8.73$ | $UCL = 12.22$ | $UCL = 12.22$ | $UCL = 8.70$ | $UCL = 8.73$ | $UCL = 12.22$ | $UCL = 12.22$ | $UCL = 8.70$ | $UCL = 8.73$ | $UCL = 12.22$ | $UCL = 12.22$ | $UCL = 8.70$ | $UCL = 8.73$ | $UCL = 12.22$ | $UCL = 12.22$ |
| | (0.0, 0.0) | 203.27 | 202.60 | 201.76 | 197.80 | 207.02 | 199.07 | 198.44 | 198.78 | 18.47 | 11.89 | 15.10 | 19.33 | 3.25 | 2.47 | 2.50 | 1.54 | 1.27 | 1.09 | 1.08 | 1.09 |
| | (0.4, 0.4) | 30.16 | 18.80 | 9.78 | 24.95 | 18.65 | 11.88 | 15.52 | 18.80 | 3.22 | 2.47 | 2.43 | 2.12 | 1.54 | 1.28 | 1.10 | 1.24 | 1.09 | 1.07 | 1.09 | |
| (0.4, 0.8) | 4.88 | 3.33 | 2.00 | 3.19 | 3.22 | 2.47 | 2.43 | 2.12 | 1.54 | 1.28 | 1.10 | 1.08 | 1.54 | 1.28 | 1.10 | 1.24 | 1.09 | 1.07 | 1.09 | | |
| (0.4, 1.2) | 1.88 | 1.34 | 1.06 | 1.31 | 1.37 | 1.24 | 1.09 | 1.07 | 1.13 | 1.07 | 1.00 | 1.00 | 1.13 | 1.07 | 1.00 | 1.07 | 1.00 | 1.00 | 1.00 | | |
| (0.8, 1.2) | 1.64 | 1.31 | 1.06 | 1.24 | 1.37 | 1.24 | 1.09 | 1.07 | 1.13 | 1.07 | 1.00 | 1.00 | 1.13 | 1.07 | 1.00 | 1.07 | 1.00 | 1.00 | 1.00 | | |
| 30 | $\mu_X^{(T)} > \mu_Y^{(T)}$ | $UCL = 9.80$ | $UCL = 9.72$ | $UCL = 10.60$ | $UCL = 10.60$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 11.49$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 11.49$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 11.49$ | |
| | (0.0, 0.0) | 203.74 | 201.46 | 199.13 | 201.02 | 203.18 | 197.42 | 201.90 | 198.68 | 10.25 | 3.40 | 3.48 | 5.79 | 1.66 | 1.14 | 1.11 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | (0.4, 0.0) | 10.25 | 4.96 | 3.36 | 9.59 | 5.88 | 3.40 | 3.48 | 5.79 | 1.66 | 1.14 | 1.11 | 1.11 | 1.66 | 1.14 | 1.11 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | (0.8, 0.4) | 1.66 | 1.25 | 1.11 | 1.41 | 1.26 | 1.14 | 1.11 | 1.11 | 1.66 | 1.14 | 1.11 | 1.11 | 1.66 | 1.14 | 1.11 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | (1.2, 0.4) | 1.05 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.05 | 1.00 | 1.00 | 1.00 | 1.05 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | $\mu_X^{(T)} < \mu_Y^{(T)}$ | $UCL = 9.80$ | $UCL = 9.72$ | $UCL = 10.60$ | $UCL = 10.60$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 11.49$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 11.49$ | $UCL = 9.76$ | $UCL = 9.64$ | $UCL = 11.49$ | $UCL = 11.49$ | |
| | (0.0, 0.0) | 203.74 | 201.46 | 199.13 | 199.36 | 203.18 | 197.42 | 201.90 | 198.70 | 10.28 | 3.40 | 4.12 | 5.00 | 1.65 | 1.15 | 1.10 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | (0.4, 0.4) | 10.28 | 4.92 | 3.39 | 8.44 | 5.85 | 3.40 | 4.12 | 5.00 | 1.65 | 1.15 | 1.10 | 1.10 | 1.65 | 1.15 | 1.10 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | (0.4, 0.8) | 1.65 | 1.25 | 1.12 | 1.37 | 1.27 | 1.15 | 1.10 | 1.10 | 1.65 | 1.15 | 1.10 | 1.10 | 1.65 | 1.15 | 1.10 | 1.01 | 1.00 | 1.00 | 1.00 | |
| | (0.4, 1.2) | 1.04 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.04 | 1.00 | 1.00 | 1.00 | 1.04 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 | |

Πίνακας 4.6: Απόδοση του διπλευρου διαγράμματος $O2$ βάσει του ARL_{out} σε σχέση με 3 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση άνωτων μεταποπίσεων μέσης τιμής· χρησιμοποιώντας δεδομένα από κανονική και $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$)

Η μελέτη προσομοίωσης που πραγματοποιήθηκε αποκαλύπτει ότι η απόδοση των συγκρίσιμων διαγραμμάτων είναι αρκετά παρόμοια με το διάγραμμα χ^2 που είναι το καλύτερο, όπως αναμενόταν, όταν τα δεδομένα που χρησιμοποιούνται είναι κανονικά. Για τα μη κανονικά δεδομένα, όπου η χρήση ενός μη παραμετρικού διαγράμματος είναι απαραίτητη, το διάγραμμα SR^2 ανιχνεύει μετατοπίσεις πιο γρήγορα από το διάγραμμα SN^2 , όπως θα περίμενε κανείς, υπενθυμίζοντας ότι η βαθμολογική συνάρτηση του *Wilcoxon* είναι ισχυρότερη από αυτή της προσημικής συνάρτησης σε περιπτώσεις συμμετρικών κατανομών και κατανομών με λεπτές ουρές. Συγκρίνοντας το SR^2 με το νέο διάγραμμα, είναι προφανές ότι η απόδοση δεν είναι ομοίμορφα καλύτερη για κανένα από αυτά, καθώς το SR^2 είναι πιο ευαίσθητο στην ανίχνευση μικρών μετατοπίσεων και το νέο διάγραμμα αποδίδει καλύτερα σε μεγάλες μετατοπίσεις και άνισες μετατοπίσεις της μέσης τιμής.

4.7 Εφαρμογή σε πραγματικά δεδομένα

Σε αυτή την ενότητα, το νέο διάγραμμα εφαρμόζεται σε πραγματικά δεδομένα χρησιμοποιώντας ένα δημοφιλές σύνολο δεδομένων που σχετίζεται με το κρασί *Vinho Verde*, ένα μοναδικό πορτογαλικό προϊόν. Το σύνολο δεδομένων αποτελείται από δύο διαφορετικές ποικιλίες οίνου (κόκκινου και λευκού). Με βάση τις φυσικοχημικές ιδιότητες, οι Cortez *et al.* (2009) ήταν οι πρώτοι που μελέτησαν τα προαναφερθέντα δεδομένα εκμεταλλευόμενοι τις τεχνικές ομαδοποιήσεις δεδομένων για τη μοντελοποίηση και πρόβλεψη της προτίμησης οίνου. Το ίδιο σύνολο δεδομένων χρησιμοποιήθηκε στο πλαίσιο του πολυμεταβλητού ελέγχου διεργασιών όχι μόνο για την απεικόνιση ενός συστήματος ελέγχου πολλαπλών μεταβλητών, αλλά και για την επίδειξη της αποτελεσματικότητας της συγκεκριμένης προτεινόμενης μεθόδου (βλ. για παράδειγμα Zou *et al.* (2011), Zou and Tsung (2012), Cheng and Shiau (2015), Ou *et al.* (2015), Zhou *et al.* (2015)).

Και τα δύο σύνολα δεδομένων για τις κόκκινες και τις λευκές ποικιλίες κρασιού είναι διαθέσιμα για ερευνητική χρήση στο UC Irvine Machine Learning Repository και υπάρχουν σε ηλεκτρονική μορφή στη διεύθυνση <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>.

Τα στοιχεία συγκεντρώθηκαν από τον Μάιο του 2004 έως το Φεβρουάριο του 2007 και καταγράφηκαν με τη βοήθεια ηλεκτρονικού συστήματος για την αυτόματη διαχείριση των δοκιμών της αμπελοοινικής διαδικασίας λαμβάνοντας υπ' όψιν όχι μόνο τα αιτήματα του παραγωγού, αλλά και την εργαστηριακή και αισθητηριακή ανάλυση (*sensory analysis*). Τα δεδομένα συνίστανται από τις εξής 11 συνεχείς μεταβλητές εισροών: σταθερή οξύτητα

(fixed acidity), πτητική οξύτητα (volatile acidity), κιτρικό οξύ (citric acid), υπολειμματική ζάχαρη (residual sugar), χλωρίδια (chlorides), ελεύθερο διοξείδιο του θείου (free sulfur dioxide), ολικό διοξείδιο του θείου (total sulfur dioxide), πυκνότητα (density), pH, θειικά άλατα (sulphates), αλκοόλη (alcohol), οι οποίες βασίζονται σε φυσικοχημικές δοκιμές (physico-chemical tests). Επιπλέον, υπάρχει μία κατηγορική μεταβλητή, η οποία υποδεικνύει την ποιότητα του οίνου. Η τελευταία μεταβλητή υπολογίστηκε ως ο μέσος όρος τουλάχιστον τριών αισθητηριακών αξιολογήσεων (sensory evaluations), που πραγματοποιήθηκαν από εμπειρογνώμονες, χρησιμοποιώντας μια κλίμακα από 0 (πολύ κακή) έως 10 (εξαιρετική). Λεπτομερής μελέτη και περιγραφή σχετικά με το σύνολο δεδομένων μπορεί να βρεθεί στην εργασία των Cortez *et al.* (2009) και στις αναφορές της.

Για την απεικόνιση του διαγράμματός μας, χρησιμοποιήσαμε τα δεδομένα λευκού οίνου που αποτελούνται από 4898 παρατηρήσεις. Δεδομένου ότι τα ακατέργαστα δεδομένα συλλέγονται με χρονολογική σειρά, θεωρήσαμε τις πρώτες 3788 παρατηρήσεις ως το ιστορικό σύνολο δεδομένων βάσει του οποίου προσδιορίστηκαν τα όρια ελέγχου. Στη συνέχεια, ελήφθησαν υπ' όψιν μόνο οι 500 πρώτες εντός ελέγχου παρατηρήσεις, στις οποίες το επίπεδο ποιότητας είναι τουλάχιστον 7, επειδή σύμφωνα με τους Cortez *et al.* (2009) το τυπικό αποδεκτό επίπεδο ποιότητας οίνου ήταν αυτό με την ένδειξη 7. Για τους σκοπούς της ανάλυσης Φάσης I, οι παρατηρήσεις του ιστορικού συνόλου δεδομένων ομαδοποιήθηκαν σε υποομάδες μεγέθους $n = 5$. Έτσι, τα 100 δείγματα αποτελούν το δείγμα αναφοράς μας. Στο σημείο αυτό θα πρέπει να αναφερθεί ότι τα όρια χρειάστηκε να επαναπροδιορισθούν μέχρις ότου τα δείγματα Φάσης I να μην παράγουν συναγερμό. Ακολούθως, οι υπόλοιπες 1110 παρατηρήσεις χρησιμοποιήθηκαν στην ανάλυση της Φάσης II. Πιο συγκεκριμένα, για να προχωρήσουμε στην ανάλυση της Φάσης II, επιλέξαμε για δύο διαφορετικά δείγματα δοκιμής ένα επίπεδο ποιότητας ίσο με 6 (LV6) και 5 (LV5). Οι 580 και 310 παρατηρήσεις που προέκυψαν, ομαδοποιήθηκαν επίσης σε υποομάδες μεγέθους $n = 5$.

Μια προκαταρκτική ανάλυση των διαθέσιμων δεδομένων, με εφαρμογή τυποποιημένων στατιστικών τεχνικών, μπορεί εύκολα να αποκαλύψει ότι οι περιθώριες κατανομές των χαρακτηριστικών που παρακολουθούνται απέχουν πολύ από το κανονικό μοντέλο. Επομένως, απαιτείται η χρήση μη παραμετρικών διαγραμμάτων ελέγχου.

Σύμφωνα με τους Cortez *et al.* (2009) που εφάρμοσαν τις τεχνικές εξόρυξης δεδομένων (data mining techniques) βασισμένες σε μηχανές φορέων υποστήριξης (support vector machines), τα θειικά άλατα, η αλκοόλη και το υπολειμματικό σάκχαρο είναι οι μεταβλητές μείζονος σημασίας. Ας υποθέσουμε ότι θέλουμε να δημιουργήσουμε ένα διάγραμμα ελέγχου

4.7. Εφαρμογή σε πραγματικά δεδομένα

με $FAR_{02}^{(2)} = 5\%$ για να παρακολουθήσουμε τα θειικά άλατα και την αλκοόλη με δεδομένα που προέρχονται από την ποιότητα του οίνου 6 (LV6). Χρησιμοποιώντας τα αποτελέσματα που παρουσιάστηκαν στις προηγούμενες ενότητες, μπορούμε να επαληθεύσουμε ότι είναι ένα σύνολο παραμέτρων που διατηρούν το επιθυμητό FAR είναι το εξής

$$a = c = 56 \quad \text{και} \quad b = d = 500 - 56 + 1 = 445.$$

Επομένως, έχοντας ένα δείγμα αναφοράς μεγέθους 500 ($m = 100 \times 5$), τα όρια ελέγχου των δύο χαρακτηριστικών καθορίζονται από την 56η και την 445η παρατήρηση. Ως αποτέλεσμα, το διάγραμμα ελέγχου θα παράγει εκτός ελέγχου σήμα εάν παραβιάζεται τουλάχιστον μία από τις επόμενες δύο συνθήκες

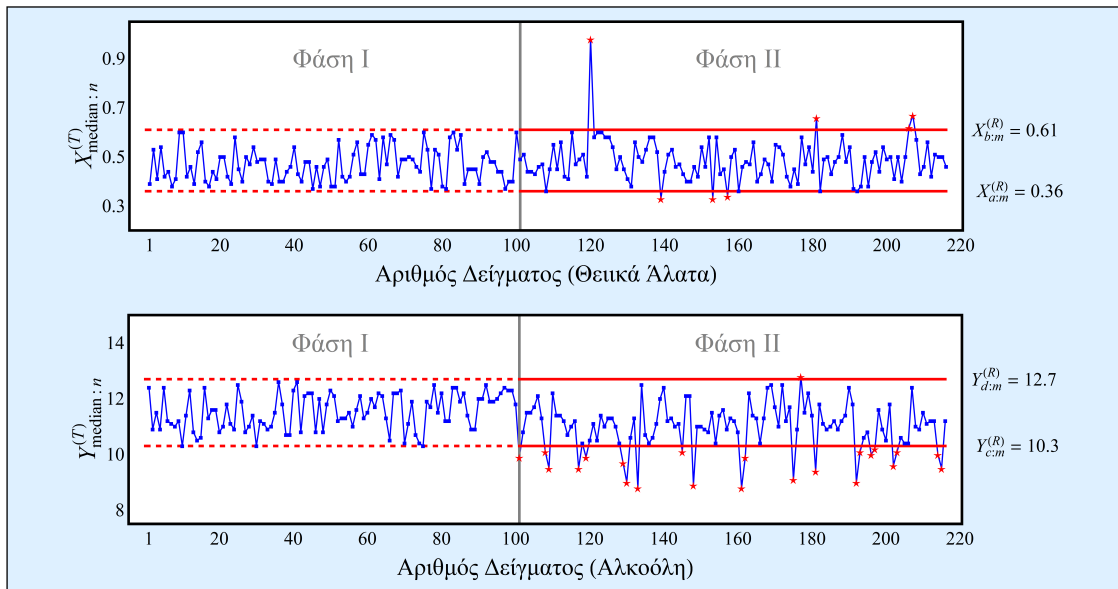
$$X_{56:500}^{(R)} \leq X_{3:5}^{(T)} \leq X_{445:500}^{(R)} \quad \text{και} \quad Y_{56:500}^{(R)} \leq Y_{3:5}^{(T)} \leq Y_{445:500}^{(R)}.$$

Στο Σχήμα 4.1, απεικονίζεται το προτεινόμενο διάγραμμα για τα δεδομένα LV6 με τη βοήθεια ενός τρισδιάστατου γραφήματος. Τα σημεία που δεν εμπίπτουν στα όρια ελέγχου επισημαίνονται με κόκκινο χρώμα. Ένας εναλλακτικός τρόπος απεικόνισης του γραφήματος μας είναι η εξέταση των δισδιάστατων προβολών (2D) που απεικονίζονται στο Σχήμα 4.5, το οποίο προσδιορίζει με σαφήνεια τα σημεία που τίθενται εκτός ελέγχου. Το Σχήμα 4.6 απεικονίζει το αντίστοιχο ημιπαραμετρικό διάγραμμα για το δείγμα με ποιότητα LV5.

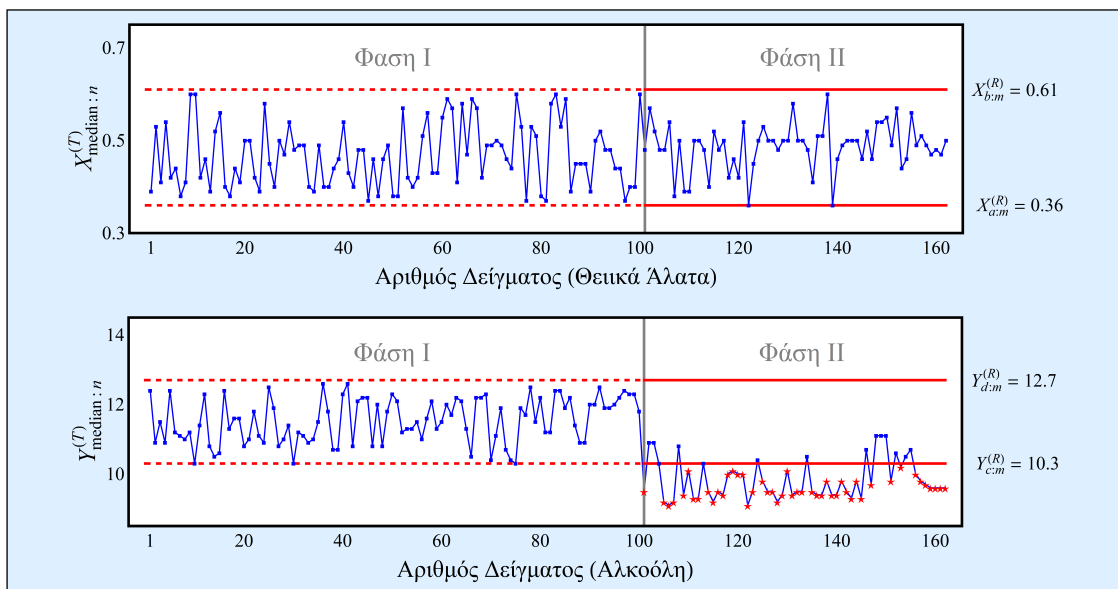
Λαμβάνοντας υπ' όψιν το πρώτο σενάριο, είναι προφανές ότι τόσο τα θειικά άλατα όσο και η αλκοόλη είναι υπαίτια για την εκτός ελέγχου κατάσταση της διεργασίας. Συγκεκριμένα, το διάγραμμα δημιουργεί σήματα συναγερμού στο 120ο, 139ο, 153ο, 157ο, 206ο, 207ο δείγμα εξαιτίας της πρώτης μεταβλητής, στο 101ο, 108ο, 109ο, 117ο, 119ο, 129ο, 130ο, 133ο, 145ο, 148ο, 161ο, 162ο, 175ο, 177ο, 192ο, 193ο, 196ο, 197ο, 202ο, 203ο, 214ο και στο 215ο δείγμα εξαιτίας της δεύτερης μεταβλητής, ενώ στο 181ο δείγμα ενεργοποιείται συναγερμός λόγω και των δύο μεταβλητών. Εάν τα θειικά άλατα όσο και η αλκοόλη δίνουν σήμα για εκτός ελέγχου μετατόπιση στο ίδιο δείγμα, το διάγραμμά μας είναι ικανό να αναγνωρίζει μετατοπίσεις ξεχωριστά για κάθε ένα από τα χαρακτηριστικά αυτά αναφορικά με την ποιότητα των δειγμάτων οίνου.

Εφαρμόζοντας το δεύτερο σενάριο, παρατηρούμε ότι τα διαγράμματα σηματοδοτούνται σχεδόν σε κάθε δείγμα, όπως φαίνεται και από το Σχήμα 4.6. Παρότι τα δείγματα Φάσης II είναι λιγότερα από αυτά του προηγούμενου σεναρίου, δεν επηρεάζεται η ευαισθησία του κανόνα. Αυτό θα μπορούσε να αποδοθεί στο ότι όσο χαμηλότερο είναι το επίπεδο ποιότητας των δειγμάτων ελέγχου, τόσο πιο γρήγορα το γράφημα ανιχνεύει τη μετατόπιση. Το γεγονός ότι μόνο μία μεταβλητή υπερβαίνει τα όρια ελέγχου μπορεί να ερμηνευθεί ως δείκτης της τάσης της συγκεκριμένης μεταβλητής να οδηγεί σε υψηλότερη ποιότητα οίνου.

Η πειραματική μας μελέτη αποκάλυψε ότι, κατά την εξέταση εναλλακτικών συνδυασμών μεταξύ των μεταβλητών που είναι διαθέσιμες στο αρχικό σύνολο δεδομένων, τα αποτελέσματα ήταν παρόμοια. Θα πρέπει να αναφερθεί ότι όταν επιλέχθηκαν λιγότερο σημαντικές μεταβλητές (σύμφωνα με το μοντέλο των Cortez *et al.* (2009), το σήμα για εκτός ελέγχου κατάσταση ενεργοποιήθηκε πολύ αργότερα, όπως αναμενόταν.



Σχήμα 4.5: Εφαρμογή του διαγράμματος $O2$ στο σύνολο δεδομένων λευκού οίνου της ποικιλίας *Vinho Verde* με μεταβλητές τα Θευικά Άλατα και την Αλκοόλη (Γράφημα 2D με το πρώτο δείγμα ελέγχου ποιότητας *LV6*)



Σχήμα 4.6: Εφαρμογή του διαγράμματος $O2$ στο σύνολο δεδομένων λευκού οίνου της ποικιλίας *Vinho Verde* με μεταβλητές τα Θευικά Άλατα και την Αλκοόλη (Γράφημα 2D με το δεύτερο δείγμα ελέγχου ποιότητας *LV5*)

ΚΕΦΑΛΑΙΟ 5

Δισδιάστατο Διάγραμμα Ελέγχου με Χρήση Διατεταγμένων Παρατηρήσεων και Συμμεταβλητών

5.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο προτάθηκε ένα ημιπαραμετρικό διάγραμμα ελέγχου το $O2$, στο οποίο η παρακολούθηση της διαδικασίας βασίζεται στη δισδιάστατη σ.σ. ελέγχου $(X_{r:n}^{(T)}, Y_{s:n}^{(T)})$. Τα όρια ελέγχου του διαγράμματος αυτού αποτελούν τέσσερις κατάλληλα επιλεγμένες διατεταγμένες παρατηρήσεις του ιστορικού δείγματος και η απόφαση σχετικά με το εάν η διεργασία βρίσκεται εντός ή εκτός ελέγχου λαμβάνεται ανάλογα με το εάν τα συγκεκριμένα στατιστικά στοιχεία του δείγματος ελέγχου βρίσκονται μεταξύ των ορίων. Το κύριο πλεονέκτημα του διαγράμματος $O2$ είναι ότι οι τιμές του ποσοστού λανθασμένου συναγερμού, καθώς και οι τιμές του εντός ελέγχου μέσου μήκους ροής δεν επηρεάζονται από τις περιθώριες κατανομές και παραμένουν σχεδόν αμετάβλητες για διαφορετικούς συνδέσμους. Συνεπώς, το διάγραμμα $O2$ διατηρεί την ελκυστική ιδιότητα όλων των μη παραμετρικών διαγραμμάτων ελέγχου της μη εξάρτησης από την κατανομή της διεργασίας και έτσι μπορεί να χρησιμοποιηθεί εξ ολοκλήρου ως απαραμετρικό διάγραμμα.

Στο παρόν κεφάλαιο εισάγουμε μία νέα κλάση δισδιάστατων ημιπαραμετρικών διαγραμμάτων ελέγχου, τύπου *Shewhart*, η οποία βασίζεται στην ίδια σ.σ. με αυτή του διαγράμματος $O2$, αλλά η διαδικασία επιλογής των ορίων ελέγχου γίνεται έτσι ώστε να διατηρούνται τα ζεύγη παρατηρήσεων (X, Y) του ιστορικού δείγματος. Συγκεκριμένα, τα νέα όρια ελέγχου αποτελούν δύο ζεύγη που μπορούν να επιλεγούν ελέγχοντας/επιθεωρώντας τις διατεταγμένες παρατηρήσεις της μεταβλητής X και τις αντίστοιχες τιμές της μεταβλητής

Y (συμμεταβλητές). Είναι προφανές ότι ο κανόνας απόφασης θα εμπλέκει διατεταγμένες παρατηρήσεις και συμμεταβλητές (*Order statistics & Concomitants*). Για αυτόν ακριβώς το λόγο, η προτεινόμενη κλάση διαγραμμάτων (Koutras and Sofikitou (2017b)) θα συμβολίζεται στο εξής με τα αρχικά OC .

Βασικό πλεονέκτημα του νέου διαγράμματος ελέγχου είναι ότι για την κατασκευή του δεν απαιτείται καμία υπόθεση για το είδος ή τη μορφή της δισδιάστατης κατανομής που ακολουθούν τα προς μελέτη χαρακτηριστικά. Επιπλέον, οι τιμές των FAR και ARL_{in} δεν επηρεάζονται από την επιλογή των περιθωρίων κατανομών. Εν τούτοις, επηρεάζονται από τη δομή εξάρτησης των δισδιάστατων παρατηρήσεων (όπως φυσικά ανακλάται στο σύνδεσμο). Εν όψει αυτής της χαρακτηριστικής ιδιότητας, τα προτεινόμενα μονόπλευρα και δίπλευρα διαγράμματα χαρακτηρίζονται ως ημιπαραμετρικά. Πρέπει, επίσης, να σημειωθεί ότι η απόδοση του νέου (δίπλευρου) διαγράμματος/σχήματος είναι αρκετά ικανοποιητική για μέτριες έως ισχυρές συσχετίσεις μεταξύ των ποιοτικών χαρακτηριστικών και η απόδοσή του βελτιώνεται σημαντικά όταν τίθενται σε ισχύ μη συμμετρικά όρια ελέγχου, η εφαρμογή των οποίων στο αντίστοιχο (δίπλευρο) διάγραμμα $O2$ δεν είναι λογική.

5.2 Περιγραφή ενός νέου διαγράμματος ελέγχου

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, παραδοσιακά, η κατασκευή ενός μη παραμετρικού διαγράμματος ελέγχου απαιτεί δύο ανεξάρτητα δείγματα, το δείγμα αναφοράς και το δείγμα ελέγχου. Αρχικά, ως υποθέσουμε ότι $(X_1^{(R)}, Y_1^{(R)})$, $(X_2^{(R)}, Y_2^{(R)})$, \dots , $(X_m^{(R)}, Y_m^{(R)})$ είναι το δείγμα αναφοράς που περιέχει ζεύγη που συλλέγονται από μια εντός ελέγχου συνεχή κατανομή, της οποίας η από κοινού α.σ.κ. συμβολίζεται με $F_{X,Y}^{(R)} = F(x, y)$ και οι αντίστοιχες περιθώριες συναρτήσεις κατανομής δηλώνονται με $F_X(x)$ και $F_Y(y)$. Εάν οι παρατηρήσεις της μεταβλητής X διαταχθούν σε αύξουσα σειρά ως εξής

$$X_{1:m}^{(R)} \leq X_{2:m}^{(R)} \leq \dots \leq X_{m:m}^{(R)},$$

τότε οι μεταβλητές Y που συνδέονται με τις παραπάνω διατεταγμένες παρατηρήσεις $(X_{1:m}^{(R)}, X_{2:m}^{(R)}, \dots, X_{m:m}^{(R)})$ συμβολίζονται με

$$Y_{[1:m]}^{(R)}, Y_{[2:m]}^{(R)}, \dots, Y_{[m:m]}^{(R)}$$

και ονομάζονται συμμεταβλητές των

$$X_{1:m}^{(R)}, X_{2:m}^{(R)}, \dots, X_{m:m}^{(R)},$$

αντίστοιχα. Οι βασικές έννοιες και ιδιότητες των Συμμεταβλητών παρουσιάσθηκαν ανα-

λυτικά στην Ενότητα 2.4.

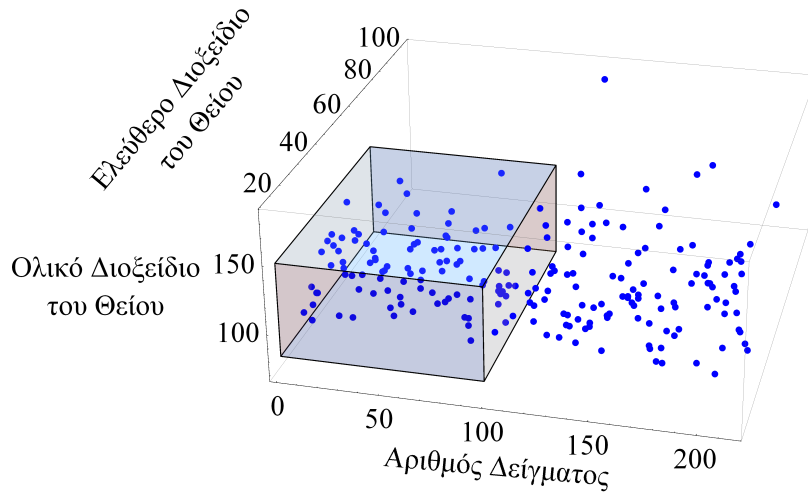
Στη συνέχεια, ας υποθέσουμε ότι αντλούνται ανεξάρτητα διαδοχικά δείγματα ελέγχου από την υποκείμενη διεργασία, τα οποία είναι επίσης ανεξάρτητα από το δείγμα αναφοράς, προκειμένου να αποφασιστεί εάν η αρχική κατανομή της διεργασίας έχει μετατοπιστεί σε μία εκτός ελέγχου κατανομή $F_{X,Y}^{(T)} = G(x, y)$ με περιθώριες κατανομές $G_X(x)$ και $G_Y(y)$. Για να το επιτύχουμε αυτό, επιλέγουμε δύο συγκεκριμένα ζεύγη από το δείγμα αναφοράς, για παράδειγμα $(X_{a:m}^{(R)}, Y_{[a:m]}^{(R)})$ και $(X_{b:m}^{(R)}, Y_{[b:m]}^{(R)})$ με $1 \leq a < b \leq m$, τα οποία χρησιμοποιούμε για την εκτίμηση των ορίων ελέγχου του προτεινόμενου διαγράμματος.

Μετά τη συλλογή των δειγμάτων ελέγχου, οι παρατηρούμενες τιμές της δισδιάστατης σ.σ. $(X_{r:n}^{(T)}, Y_{s:n}^{(T)})$, όπου με $X_{r:n}^{(T)}$ και $Y_{s:n}^{(T)}$ δηλώνουμε την r -οστή και s -οστή διατεταγμένη παρατήρηση των αντίστοιχων μεταβλητών X και Y (του δείγματος ελέγχου), συγκρίνονται με τα παραπάνω όρια ελέγχου. Συγκεκριμένα, η υποκείμενη διαδικασία δηλώνεται ότι είναι εντός ελέγχου, εάν επαληθεύεται ο ακόλουθος κανόνας

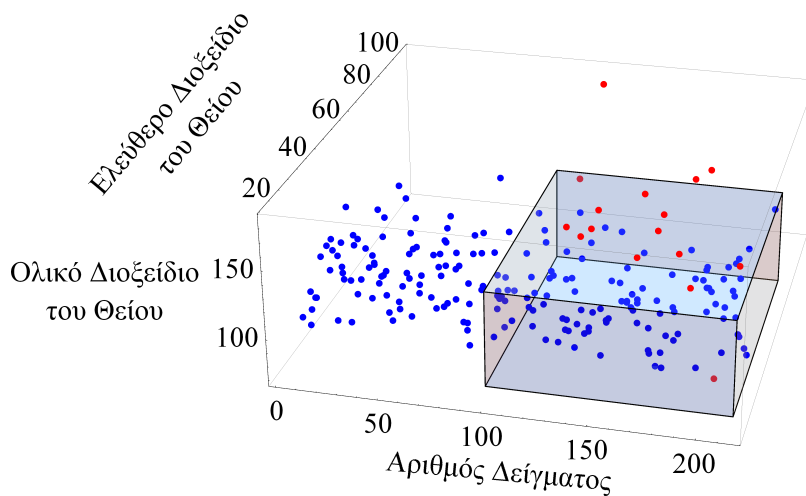
$$X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \quad \text{και} \quad \min\{Y_{[a:m]}^{(R)}, Y_{[b:m]}^{(R)}\} \leq Y_{s:n}^{(T)} \leq \max\{Y_{[a:m]}^{(R)}, Y_{[b:m]}^{(R)}\} \quad (5.1)$$

Ο προσδιορισμός του χαρακτηριστικού ή των χαρακτηριστικών που ενεργοποίησαν το εκτός ελέγχου σήμα μπορεί εύκολα να γίνει παρατηρώντας απλά ποια από τις παραπάνω συνθήκες παραβιάζεται.

Στο Σχήμα 5.1 παρουσιάζεται η τρισδιάστατη μορφή του νέου διαγράμματος εφαρμόζοντας τον κανόνα ελέγχου στο πραγματικό σύνολο δεδομένων λευκού οίνου της ποικιλίας *Vinho Verde*, το οποίο περιγράφηκε αναλυτικά στην Ενότητα 4.7 του προηγούμενου κεφαλαίου. Περισσότερες πληροφορίες σχετικά με την επιλογή των κατάλληλων μεταβλητών και το σχεδιασμό του διαγράμματος *OC* δίνονται στην Ενότητα 5.7. Στο τρισδιάστατο (3D) γράφημα, ο οριζόντιος άξονας χρησιμοποιείται για να περιγράψει τη χρονολογική σειρά των δειγμάτων, ενώ κάθε ένας από τους δύο άλλους άξονες καταγράφει τις τιμές των υπό μελέτη μεταβλητών. Με τη βοήθεια των δύο αυτών αξόνων, τοποθετούνται στο χώρο τα σημεία που αντιστοιχούν στα ζευγάρια της μορφής $(X_{r:n}^{(T)}, Y_{s:n}^{(T)})$. Προφανώς, η σκιασμένη περιοχή υποδηλώνει την περιοχή ελέγχου, η οποία καθορίζεται από τα όρια ελέγχου που προσδιορίστηκαν στη Φάση I και εμπλέκουν δύο ζευγάρια διατεταγμένων παρατηρήσεων με τις συμμεταβλητές τους, δηλαδή $(X_{a:m}^{(R)}, Y_{[a:m]}^{(R)})$ και $(X_{b:m}^{(R)}, Y_{[b:m]}^{(R)})$. Στη Φάση II, λαμβάνεται ένα δείγμα ελέγχου με σκοπό να διαπιστωθεί κατά πόσο επαληθεύεται ή όχι ο Κανόνας 5.1. Το σημείο για το οποίο παραβιάζεται έστω μία από τις συνθήκες του Κανόνα 5.1, σημειώνεται με κόκκινο χρώμα, καθώς δεν ανήκει στην εντός ελέγχου περιοχή.



(α') Φάση I: Περιοχή ελέγχου από το δείγμα αναφοράς



(β') Φάση II: Απεικονιζόμενα σημεία του δείγματος ελέγχου
στην περιοχή ελέγχου που προσδιορίστηκε στη Φάση I

Σχήμα 5.1: Εφαρμογή διαγράμματος *OC* στα δεδομένα λευκού οίνου της ποικιλίας *Vinho Verde* με μεταβλητές το Ελεύθερο και Ολικό Διοξείδιο του Θείου (Γράφημα 3D)

5.3 Χαρακτηριστικά διαγράμματος ελέγχου

Στην παρούσα ενότητα αποδεικνύουμε τύπους για τον ακριβή προσδιορισμό των χαρακτηριστικών του διαγράμματος ελέγχου. Υπενθυμίζουμε ότι, εάν υποθέσουμε πως $C(u, v)$ και $D(u, v)$ είναι οι δισδιάστατοι σύνδεσμοι που αντιστοιχούν στις κατανομές $F(x, y)$ και $G(x, y)$, αντίστοιχα, τότε οι από κοινού συναρτήσεις κατανομής $F(x, y), G(x, y)$ μπορούν να εκφραστούν ως εξής

$$F(x, y) = C(F_X(x), F_Y(y)) \quad \text{και} \quad G(x, y) = D(G_X(x), G_Y(y)).$$

Προφανώς, η από κοινού πυκνότητα του ζεύγους (X, Y) υπολογίζεται από την έκφραση

$$f(x, y) = \frac{\partial^2 C(F_X(x), F_Y(y))}{\partial x \partial y} = c(F_X(x), F_Y(y)) f_X(x) f_Y(y),$$

όπου

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad (5.2)$$

δηλώνει την πυκνότητα του συνδέσμου $C(u, v)$. Στο εξής θα θεωρούμε ότι οι κατανομές των δειγμάτων αναφοράς και ελέγχου είναι συνεχείς.

Για να μελετήσουμε την απόδοση του προτεινόμενου σχήματος, χρειάζεται να καταφύγουμε σε έννοιες που συνδέονται όχι μόνο με τη Θεωρία των Συνδέσμων, αλλά και με τη Θεωρία των Συμμεταβλητών (βλ. Ενότητα 2.4). Τα παρακάτω λήμματα θα μας φανούν χρήσιμα στη συνέχεια.

Λήμμα 5.7. *Ας υποθέσουμε ότι (X_i, Y_i) , $i = 1, 2, \dots, m$ είναι m ανεξάρτητες τ.μ. που έχουν από κοινού α.σ.κ. $F(x, y)$ με περιθώριες $F_X(x), F_Y(y)$ και αντίστοιχο σύνδεσμο $C(x, y)$. Τότε, η από κοινού σ.π.π. των τ.μ. $U_{r:m} = F_X(X_{r:m})$ και $V_{[r:m]} = F_Y(Y_{[r:m]})$ εκφράζεται από τη σχέση*

$$f_{U_{r:m}, V_{[r:m]}}^{(C)}(u, v) = \frac{1}{B(r, m - r + 1)} u^{r-1} (1 - u)^{m-r} c(u, v)$$

για $u, v \in [0, 1]$ και $1 \leq r \leq m$.

Απόδειξη. Ανακαλώντας, πρώτα, τη Σχέση (2.12) και αντικαθιστώντας το n με m έχουμε

$$f_{X_{r:m}, Y_{[r:m]}}(x, y) = \frac{1}{B(r, m - r + 1)} F_X(x)^{r-1} (1 - F_X(x))^{m-r} f_{X,Y}(x, y),$$

όπου $x, y \in (-\infty, \infty)$. Έπειτα, παρατηρώντας ότι οι τ.μ. $U_{r:m} = F_X(X_{r:m})$ και $V_{[r:m]} = F_Y(Y_{[r:m]})$ προέρχονται από ομοιόμορφες κατανομές στο διάστημα $[0, 1]$, μπορούμε άμεσα να ισχυρι-

στούμε ότι η από κοινού πυκνότητά τους υπολογίζεται μέσω του τύπου

$$f_{U_{r:m}, V_{[r:m]}}^{(C)}(u, v) = \frac{n!}{(r-1)!(n-r)!} u^{r-1} (1-u)^{m-r} c(u, v) = \frac{1}{B(r, m-r+1)} u^{r-1} (1-u)^{m-r} c(u, v).$$

Το παραπάνω αποτέλεσμα προκύπτει άμεσα επειδή $F_U(u) = u$ ($0 \leq u \leq 1$) και $f_{U,V}(u, v) = c(u, v)$, όπου το $c(u, v)$ ορίζεται στη Σχέση (5.2). \square

Λήμμα 5.8. Έστω (X_i, Y_i) , $i = 1, 2, \dots, m$, m ανεξάρτητες τ.μ. με από κοινού α.σ.κ. $F(x, y)$, περιθώριες $F_X(x)$, $F_Y(y)$ και αντίστοιχο σύνδεσμο $C(x, y)$. Τότε, η από κοινού σ.π.π. των $(U_{a:m}, V_{[a:m]})$, $(U_{b:m}, V_{[b:m]})$ δίνεται από την έκφραση

$$\frac{f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) = \frac{m!}{(a-1)!(m-b)!(b-a-1)!} u_1^{a-1} (1-u_2)^{m-b} (u_2-u_1)^{b-a-1} c(u_1, v_1) c(u_2, v_2)$$

για $0 \leq u_1 \leq u_2 \leq 1$, $0 \leq v_1, v_2 \leq 1$ και $1 \leq a \leq b \leq m$.

Απόδειξη. Σύμφωνα με τους Abo-Eleneen and Nagaraja (2002), η από κοινού σ.π.π. της συλλογής $(X_{r_1:m}, Y_{[r_1:m]})$, $(X_{r_2:m}, Y_{[r_2:m]})$, \dots , $(X_{r_k:m}, Y_{[r_k:m]})$ δίνεται από τη Σχέση (2.4) χρησιμοποιώντας το m στη θέση του n , δηλαδή από τον παρακάτω τύπο

$$\begin{aligned} f_{X_{r_1:m}, X_{r_2:m}, \dots, X_{r_k:m}; Y_{[r_1:m]}, Y_{[r_2:m]}, \dots, Y_{[r_k:m]}}(x_1, x_2, \dots, x_k; y_1, y_2, \dots, y_k) &= \\ &= \frac{m!}{(r_1-1)!(m-r_k)!} [F_X(x_1)]^{r_1-1} [1-F_X(x_k)]^{m-r_k} \\ &\times \prod_{i=2}^k \frac{[F_X(x_i) - F_X(x_{i-1})]^{r_i-r_{i-1}-1}}{(r_i-r_{i-1}-1)!} \prod_{i=1}^k f(x_i, y_i), \end{aligned}$$

όπου $1 \leq r_1 < r_2 < \dots < r_k \leq m$ και $x_i, y_i \in \mathbb{R}$. Προφανώς, για $0 \leq u_1 \leq u_2 \leq 1$, $0 \leq v_1, v_2 \leq 1$ και $1 \leq a < b \leq m$, η από κοινού πυκνότητα των ζευγαριών $(U_{a:m}, V_{[a:m]})$, $(U_{b:m}, V_{[b:m]})$, όπου $U, V \sim U(0, 1)$, είναι

$$f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(D)} = \frac{m!}{(a-1)!(m-b)!(b-a-1)!} u_1^{a-1} (1-u_2)^{m-b} (u_2-u_1)^{b-a-1} c(u_1, v_1) c(u_2, v_2),$$

με το $c(u, v)$ να ορίζεται στη Σχέση (5.2). Η απόδειξη ολοκληρώνεται λαμβάνοντας υπ' όψιν ότι $F_U(u) = u$ ($0 \leq u \leq 1$) και $f_{U,V}(u, v) = c(u, v)$. \square

5.3.1 Δίπλευρα διαγράμματα ελέγχου

Η πιθανότητα το διάγραμμα OC να μην ενεργοποιεί συναγερμό για εκτός ελέγχου μετατοπίσεις ορίζεται ως ακολούθως

$$p_{OC}^{(2)} = P\left(X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \text{ και } \min\{Y_{[a:m]}^{(R)}, Y_{[b:m]}^{(R)}\} \leq Y_{s:n}^{(T)} \leq \max\{Y_{[a:m]}^{(R)}, Y_{[b:m]}^{(R)}\}\right). \quad (5.3)$$

Προφανώς, η πιθανότητα $OCF_{OC}^{(2)} = p_{OC}^{(2)}$ δηλώνει τη χαρακτηριστική συνάρτηση λειτουργίας του δίπλευρου διαγράμματος OC και η ποσότητα $AR_{OC}^{(2)} = 1 - p_{OC}^{(2)}$ την πιθανότητα να ενεργοποιηθεί συναγερμός. Παρακάτω, αποδεικνύουμε μία έκφραση για την πιθανότητα $p_{OC}^{(2)}$, όπως αυτή ορίζεται στη Σχέση (5.3).

Πρόταση 5.2. Η χαρακτηριστική συνάρτηση λειτουργίας του δίπλευρου διαγράμματος OC προσδιορίζεται από την έκφραση

$$\begin{aligned} p_{OC}^{(2)} &= p_{OC}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b; r, s) \\ &= + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \times f_{U_{a:m}; V_{[a:m]}}^{(C)}(u_1, v_1) du_1 dv_1 \\ &\quad - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \times \left(\int_0^{v_1} f_{U_{a:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1; v_1, v_2) dv_2 \right) du_1 dv_1 \\ &\quad + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \times f_{U_{b:m}; V_{[b:m]}}^{(C)}(u_2, v_2) du_2 dv_2 \\ &\quad - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \times \left(\int_{v_2}^1 f_{U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_2; v_1, v_2) dv_1 \right) du_2 dv_2 \\ &\quad + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \times f_{U_{a:m}; V_{[b:m]}}^{(C)}(u_1, v_2) du_1 dv_2 \\ &\quad - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \times \left(\int_0^{v_2} f_{U_{a:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1; v_1, v_2) dv_1 \right) du_1 dv_2 \\ &\quad + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \times f_{U_{b:m}; V_{[a:m]}}^{(C)}(u_2, v_1) du_2 dv_1 \\ &\quad - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \times \left(\int_{v_1}^0 f_{U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_2; v_1, v_2) dv_2 \right) du_2 dv_1, \end{aligned}$$

όπου η $F_{r,s;n}^{(D)}(u, v)$ δηλώνει την α.σ.κ. του Λήμματος 3.1, $f_{U_{k:m}; V_{[k:m]}}^{(C)}(u, v)$, είναι η σ.π.π. που παρουσιάστηκε στο Λήμμα 5.7 για $k \in \{a, b\}$ και $0 \leq u, v \leq 1$, ενώ οι ποσότητες $f_{U_{b:m}; V_{[a:m]}}^{(C)}(u_2; v_1)$, $f_{U_{a:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1; v_1, v_2)$, $f_{U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_2; v_1, v_2)$ είναι οι περιθώριες συναρτήσεις πυκνότητας της σ.π.π. που προσδιορίστηκε στο Λήμμα 5.8.

Απόδειξη. Γράφουμε πρώτα την Έκφραση 5.3 στην ισοδύναμη μορφή

$$P \left[F_X(X_{a:m}^{(R)}) \leq F_X(X_{r:n}^{(T)}) \leq F_X(X_{b:m}^{(R)}) \text{ και} \right. \\ \left. \min \{F_Y(Y_{[a:m]}^{(R)}), F_Y(Y_{[b:m]}^{(R)})\} \leq F_Y(Y_{s:n}^{(T)}) \leq \max \{F_Y(Y_{[a:m]}^{(R)}), F_Y(Y_{[b:m]}^{(R)})\} \right]$$

και έπειτα εφαρμόζουμε την αρχή δέσμευσης-αποδέσμευσης (δηλαδή τη συνεχή έκδοση του ΘΟΠ (Κούτρας ((2004, 2016)) στις τ.μ.

$$U_{a:m} = F_X(X_{a:m}^{(R)}), \quad U_{b:m} = F_X(X_{b:m}^{(R)}) \quad \text{και} \quad V_{[a:m]} = F_Y(Y_{[a:m]}^{(R)}), \quad V_{[b:m]} = F_Y(Y_{[b:m]}^{(R)}),$$

ώστε να λάβουμε την ακόλουθη έκφραση

$$p_{OC}^{(2)} = \int_0^1 \int_0^{u_2} \int_0^1 \int_0^1 P[u_1 \leq F_X(X_{r:n}^{(T)}) \leq u_2 \text{ και } \min\{v_1, v_2\} \leq F_Y(Y_{s:n}^{(T)}) \leq \max\{v_1, v_2\}] \\ \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}; V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2.$$

Δεδομένου ότι η διάταξη των v_1, v_2 δεν είναι γνωστή, θα πρέπει να διακρίνουμε τις περιπτώσεις $v_1 \leq v_2$ και $v_2 \leq v_1$. Με τον τρόπο αυτό, η πιθανότητα $p_{OC}^{(2)}$ μπορεί εναλλακτικά να γραφεί στη μορφή

$$p_{OC}^{(2)} = \int_0^1 \int_0^{u_2} \int_0^1 \int_0^{v_2} P[u_1 \leq F_X(X_{r:n}^{(T)}) \leq u_2 \text{ και } v_1 \leq F_Y(Y_{s:n}^{(T)}) \leq v_2] \\ \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}; V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2 \\ + \int_0^1 \int_0^{u_2} \int_0^1 \int_{v_2}^1 P[u_1 \leq F_X(X_{r:n}^{(T)}) \leq u_2 \text{ και } v_2 \leq F_Y(Y_{s:n}^{(T)}) \leq v_1] \\ \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}; V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2,$$

όπου $f_{U_{a:m}, U_{b:m}; V_{[a:m]}; V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η από κοινού σ.π.π. των $U_{a:m}, U_{b:m}, V_{[a:m]}, V_{[b:m]}$ όπως περιγράφεται στο Λήμμα 5.8

Στη συνέχεια, λαμβάνοντας υπ' όψιν τη σχέση

$$P[F_X(X_{r:n}^{(T)}) \leq u_i, F_Y(Y_{s:n}^{(T)}) \leq v_j] = P[X_{r:n}^{(T)} \leq F_X^{-1}(u_i), Y_{s:n}^{(T)} \leq F_Y^{-1}(v_j)] \\ = P[G_X(X_{r:n}^{(T)}) \leq G_X(F_X^{-1}(u_i)), G_Y(Y_{s:n}^{(T)}) \leq G_Y(F_Y^{-1}(v_j))] \\ = P[U_{r:n}^{(T)} \leq G_X(F_X^{-1}(u_i)), V_{s:n}^{(T)} \leq G_Y(F_Y^{-1}(v_j))] \\ = F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_i)), G_Y(F_Y^{-1}(v_j))), \quad \text{για } i, j = 1, 2,$$

παρατηρούμε ότι οι πιθανότητες που εμφανίζονται μέσα στο τετραπλό ολοκλήρωμα μπο-

ρούν να εκφραστούν ως αλγεβρικό άθροισμα των παραπάνω ποσοτήτων. Συγκεκριμένα, ισχύουν τα εξής

$$\begin{aligned} & P\left[u_1 \leq F_X(X_{r:n}^{(T)}) \leq u_2 \text{ και } v_1 \leq F_Y(Y_{s:n}^{(T)}) \leq v_2\right] = \\ & = + F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) + F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \\ & \quad - F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) - F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \end{aligned}$$

και

$$\begin{aligned} & P\left[u_1 \leq F_X(X_{r:n}^{(T)}) \leq u_2 \text{ και } v_2 \leq F_Y(Y_{s:n}^{(T)}) \leq v_1\right] = \\ & = + F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) + F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \\ & \quad - F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) - F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))), \end{aligned}$$

όπου $F_{r,s:n}^{(D)}(u, v)$ είναι η από κοινού α.σ.κ που δίνεται στο Λήμμα 3.1. Αξιοποιώντας τα προηγούμενα αποτελέσματα, προκύπτουν τα ολοκληρώματα που ακολουθούν

$$\begin{aligned} I_1 & = + \int_0^1 \int_0^{u_2} \int_0^1 \int_0^{v_2} F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2 \\ I_2 & = + \int_0^1 \int_0^{u_2} \int_0^1 \int_0^{v_2} F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2 \\ I_3 & = - \int_0^1 \int_0^{u_2} \int_0^1 \int_0^{v_2} F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2 \\ I_4 & = - \int_0^1 \int_0^{u_2} \int_0^1 \int_0^{v_2} F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2 \\ I_5 & = + \int_0^1 \int_0^{u_2} \int_0^1 \int_{v_2}^1 F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2, \\ I_6 & = + \int_0^1 \int_0^{u_2} \int_0^1 \int_{v_2}^1 F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2, \\ I_7 & = - \int_0^1 \int_0^{u_2} \int_0^1 \int_{v_2}^1 F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \\ & \quad \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2, \end{aligned}$$

$$I_8 = - \int_0^1 \int_0^{u_2} \int_0^1 \int_{v_2}^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \\ \times f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) dv_1 dv_2 du_1 du_2.$$

Τέλος, αλλάζοντας τη σειρά ολοκλήρωσης προκύπτει η ζητούμενη έκφραση

$$p_{OC}^{(2)} = p_{OC}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b; r, s) = (I_1 + I_7) + (I_2 + I_8) + (I_3 + I_5) + (I_4 + I_6),$$

όπου

$$I_1 + I_7 = + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \times f_{U_{a:m}, V_{[a:m]}}^{(C)}(u_1, v_1) du_1 dv_1 \\ - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_1))) \times \left(\int_0^{v_1} f_{U_{a:m}, V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1; v_1, v_2) dv_2 \right) du_1 dv_1, \\ I_2 + I_8 = + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \times f_{U_{b:m}, V_{[b:m]}}^{(C)}(u_2, v_2) du_2 dv_2 \\ - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_2))) \times \left(\int_{v_2}^1 f_{U_{b:m}, V_{[a:m]}, V_{[b:m]}}^{(C)}(u_2; v_1, v_2) dv_1 \right) du_2 dv_2, \\ I_3 + I_5 = + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \times f_{U_{a:m}, V_{[b:m]}}^{(C)}(u_1, v_2) du_1 dv_2 \\ - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_1)), G_Y(F_Y^{-1}(v_2))) \times \left(\int_0^{v_2} f_{U_{a:m}, V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1; v_1, v_2) dv_1 \right) du_1 dv_2, \\ I_4 + I_6 = + \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \times f_{U_{b:m}, V_{[a:m]}}^{(C)}(u_2, v_1) du_2 dv_1 \\ - 2 \int_0^1 \int_0^1 F_{r,s;n}^{(D)}(G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1))) \times \left(\int_{v_1}^0 f_{U_{b:m}, V_{[a:m]}, V_{[b:m]}}^{(C)}(u_2; v_1, v_2) dv_2 \right) du_2 dv_1$$

και η απόδειξη της πρότασης ολοκληρώνεται. \square

Εργαζόμενοι με την ίδια λογική που ακολουθήσαμε στο προηγούμενο κεφάλαιο, θέτοντας $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$ στην έκφραση της Πρότασης 5.2, προκύπτει ένας τύπος για τον ακριβή υπολογισμό της πιθανότητας συναγεμού του δίπλευρου διαγράμματος OC . Όταν ισχύει ταυτόχρονα $C(u, v) = D(u, v)$, προκύπτει ο τύπος της πιθανότητας λανθασμένου συναγεμού. Στα πορίσματα που ακολουθούν παρουσιάζονται τα προαναφερθέντα αποτελέσματα.

Πόρισμα 5.3. Εάν $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, η πιθανότητα συναγεμού του διαγράμματος OC , που ορίζεται στον Κανόνα 5.1, δεν εξαρτάται από τις περιθώριες κατανομές και δίνεται από την έκφραση

$$AR_{OC}^{(2)} = AR_{OC}^{(C, D)}(m, n; a, b; r, s) = 1 - p_{OC}^{(F_X, F_X; F_Y, F_Y; C, D)}(m, n; a, b; r, s). \quad (5.4)$$

Απόδειξη. Στην περίπτωση που έχουμε $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, προφανώς ισχύει $G_X(F_X^{-1}(u)) = u$ και $G_Y(F_Y^{-1}(v)) = v$. Συνεπώς, οι ποσότητες $F_{r,s:n}^{(D)}(G_X(F_X^{-1}(u_i)), G_Y(F_Y^{-1}(v_j)))$ που εμφανίζονται στην πιθανότητα $p_{OC}^{(2)}$ οδηγούν στις αθροιστικές συναρτήσεις $F_{r,s:n}^{(D)}(u_i, v_j)$ για $i, j = 1, 2$. \square

Πόρισμα 5.4. Στην ειδική περίπτωση που έχουμε $F_X(x) = G_X(x)$, $F_Y(y) = G_Y(y)$ και συγχρόνως $C(u, v) = D(u, v)$, προκύπτει η πιθανότητα λανθασμένου συναγερομού του δίπλευρου διαγράμματος OC, η οποία ορίζεται από την έκφραση

$$\begin{aligned} FAR_{OC}^{(2)} &= FAR_{OC}^{(C)}(m, n; a, b; r, s) \\ &= 1 - P\left(X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \text{ και } \min\{Y_{[a:m]}^{(R)}, Y_{[b:m]}^{(R)}\} \leq Y_{s:n}^{(T)} \leq \max\{Y_{[a:m]}^{(R)}, Y_{[b:m]}^{(R)}\} \mid H_0\right) \\ &= AR_{OC}^{(C,C)}(m, n; a, b; r, s). \end{aligned}$$

Απόδειξη. Το αποτέλεσμα προκύπτει απευθείας από τη Σχέση (5.4) αντικαθιστώντας το σύνδεσμο $D(u, v)$ με το σύνδεσμο $C(u, v)$. \square

5.3.2 Μονόπλευρα διαγράμματα ελέγχου

Θα πρέπει να αναφερθεί ότι υπάρχουν περιπτώσεις στις οποίες το τμήμα ποιοτικού ελέγχου μιας βιομηχανίας μπορεί να μην ενδιαφέρεται να ανιχνεύσει τις μεταβολές μιας διεργασίας και προς τις δύο κατευθύνσεις. Για παράδειγμα, η μετατόπιση προς μια κατεύθυνση θα μπορούσε να αποτελέσει σοβαρότερη αιτία ανησυχίας σε σχέση με μια μετατόπιση προς την άλλη κατεύθυνση, καθώς μπορεί να οδηγήσει σε υποβάθμιση της ποιότητας της υποκείμενης διεργασίας. Όταν το ενδιαφέρον μας επικεντρώνεται στην παρακολούθηση πιθανών μετατοπίσεων προς τα πάνω ή προς τα κάτω, πρέπει να εφαρμοστούν μονόπλευρα διαγράμματα ελέγχου, όπου μόνο ένα άνω ή ένα κάτω όριο είναι αντίστοιχα επαρκές. Πρόσφατα, οι Eryilmaz and Bairamov (2003) μελέτησαν τις περιθώριες και τις από κοινού κατανομές του βαθμού μιας διατεταγμένης παρατήρησης και της συμμεταβλητής της, έστω $X_{r:n}$ και $Y_{[r:n]}$, μεταξύ ενός νέου δείγματος $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \dots, (X_{n+m}, Y_{n+m})$.

Το προτεινόμενο σχήμα μπορεί εύκολα να τροποποιηθεί με σκοπό να δημιουργηθούν μονόπλευρα διαγράμματα ελέγχου. Ας υποθέσουμε ότι οι χαμηλότερες τιμές σε ένα τουλάχιστον χαρακτηριστικό παρέχουν ένδειξη μετατόπισης της διεργασίας σε μία εκτός ελέγχου κατάσταση (στην οποία περίπτωση η παράμετρος ενδιαφέροντος μετατοπίστηκε προς τα αριστερά). Συγκεκριμένα, θεωρούμε ότι το ζευγάρι $(X_{a:m}^{(R)}, Y_{[a:m]}^{(R)})$ συνιστά ένα κατώφλι, όπου

η μεταβλητή X αναπαριστά την τελική απόδοση ενός υποψηφίου/ατόμου και η μεταβλητή Y υποδηλώνει το βαθμό του σε μια δοκιμασία εξέτασης. Σε αυτή την περίπτωση, ενδιαφερόμαστε για τον υπολογισμό της πιθανότητας του ενδεχομένου ότι τουλάχιστον $n - r + 1$ από τις παρατηρήσεις X του δείγματος ελέγχου (του δεύτερου δείγματος) βρίσκονται μεταξύ των a άνω καλύτερων τελικών επιδόσεων του δείγματος αναφοράς (του πρώτου δείγματος) και τουλάχιστον $n - s + 1$ των παρατηρήσεων Y του δείγματος ελέγχου έχουν βαθμολογία (στη δοκιμασία εξέτασης) μεγαλύτερη ή ίση από την τιμή $Y_{[a:m]}^{(R)}$. Τότε, η χαρακτηριστική συνάρτηση λειτουργίας του μονόπλευρου διαγράμματος με κάτω όρια (θεωρώντας θετική εξάρτηση μεταξύ των δύο χαρακτηριστικών) δίνεται από τη σχέση

$$\begin{aligned} p_{OC}^{(1)} &= p_{OC}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a; r, s) \\ &= P[X_{r:n}^{(T)} \geq X_{a:m}^{(R)} \text{ και } Y_{s:n}^{(T)} \geq Y_{[a:m]}^{(R)}] \\ &= \int_0^1 \int_0^1 P[X_{r:n}^{(T)} \geq F_X^{-1}(u) \text{ και } Y_{s:n}^{(T)} \geq F_Y^{-1}(v)] \times f_{U_{a:m}, V_{[a:m]}}^{(C)}(u, v) \, dudv \\ &= \int_0^1 \int_0^1 P[U_{r:n}^{(T)} \geq G_X(F_X^{-1}(u)) \text{ και } V_{s:n}^{(T)} \geq G_Y(F_Y^{-1}(v))] \times f_{U_{a:m}, V_{[a:m]}}^{(C)}(u, v) \, dudv \\ &= \int_0^1 \int_0^1 \bar{F}_{r,s;n}^{(D)}(G_X(F_X^{-1}(u)), G_Y(F_Y^{-1}(v))) \times f_{U_{a:m}, V_{[a:m]}}^{(C)}(u, v) \, dudv, \end{aligned}$$

όπου $\bar{F}_{r,s;n}^{(D)}(u, v)$ είναι η συνάρτηση επιβίωσης του Λήμματος 3.1, ενώ $f_{U_{a:m}, V_{[a:m]}}^{(C)}(u, v)$ είναι η πυκνότητα που περιγράφηκε στο Λήμμα 5.7.

Δοθείσης της παραπάνω έκφρασης για την πιθανότητα $p_{OC}^{(1)}$, μπορούν εύκολα να προσδιορισθούν το ποσοστό συναγερμού $AR_{OC}^{(1)}$ και το ποσοστό λανθασμένου συναγερμού $FAR_{OC}^{(1)}$ από τις αντίστοιχες εκφράσεις που ακολουθούν

$$\begin{aligned} AR_{OC}^{(1)} &= AR_{OC}^{(C,D)}(m, n; a; r, s) \\ &= 1 - \int_0^1 \int_0^1 \bar{F}_{r,s;n}^{(D)}(u, v) \times f_{U_{a:m}, V_{[a:m]}}^{(C)}(u, v) \, dudv, \\ FAR_{OC}^{(1)} &= FAR_{OC}^{(C)}(m, n; a; r, s) = AR_{OC}^{(C,C)}(m, n; a; r, s) \\ &= 1 - \int_0^1 \int_0^1 \bar{F}_{r,s;n}^{(C)}(u, v) \times f_{U_{a:m}, V_{[a:m]}}^{(C)}(u, v) \, dudv. \end{aligned}$$

5.4 Απόδοση διαγράμματος για συγκεκριμένους συνδέσμους

Στην τρέχουσα ενότητα, θα δώσουμε μία ακριβή έκφραση για το ποσοστό συναγερμού στην περίπτωση που τόσο το δείγμα αναφοράς όσο και το δείγμα ελέγχου προέρχονται από την

ίδια δισδιάστατη κατανομή και έχουν τις ίδιες περιθώριες (μονοδιάστατες) κατανομές. Ειδικότερα, υποθέτουμε ότι και τα δύο δείγματα ακολουθούν την κατανομή των *FGM* (Morgenstern (1956), Gumbel (1958), Farlie (1960)) με από κοινού α.σ.κ.

$$F_{\theta}(x, y) = F_X(x)F_Y(y)[1 + \theta(1 - F_X(x))(1 - F_Y(y))], \quad x, y \geq 0$$

και διαφορετικές παραμέτρους συσχέτισης. Τότε, οι σύνδεσμοι του δείγματος αναφοράς και του δείγματος ελέγχου, οι οποίοι σχετίζονται με την παραπάνω κατανομή, δίνονται από τις αντίστοιχες εκφράσεις που ακολουθούν

$$C(u, v) = C_{\theta_1}^{FGM}(u, v) = uv[1 + \theta_1(1 - u)(1 - v)], \quad -1 \leq \theta_1 \leq 1,$$

$$D(u, v) = D_{\theta_2}^{FGM}(u, v) = uv[1 + \theta_2(1 - u)(1 - v)], \quad -1 \leq \theta_2 \leq 1.$$

Στο σημείο αυτό, θα πρέπει να αναφερθεί ότι ο σύνδεσμος *FGM* είναι ένας πολυωνυμικός σύνδεσμος τέταρτης ($m = 4$) τάξης (βλ. Ενότητα 2.5).

Αντικαθιστώντας την παραπάνω έκφραση για το $D(u, v)$ στην α.σ.κ. (Σχέση (3.2)) του Λήμματος 3.1, η ποσότητα $q_n^{(D)}(i, j, k; u, v)$ μπορεί να γραφεί με τη βοήθεια διωνυμικών αθροισμάτων ως εξής

$$q_n^{(D)}(i, j, k; u, v) = \sum_{r_1=0}^k \sum_{r_2=0}^{i-k} \sum_{r_3=0}^{j-k} \sum_{r_4=0}^{n-i-j+k} \binom{k}{r_1} \binom{i-k}{r_2} \binom{j-k}{r_3} \binom{n-i-j+k}{r_4} \\ \times (-1)^{r_2+r_3} \theta^{r_1+r_2+r_3+r_4} u^{j+r_3+r_4} v^{j+r_2+r_4} (1-u)^{m-i-r_1+r_2} (1-v)^{m-j+r_1+r_3}.$$

Έπειτα, παρατηρούμε ότι η πυκνότητα του συνδέσμου $C(u, v)$ γράφεται στη μορφή

$$c(u, v) = c_{\theta_1}(u, v) = 1 + \theta_1(1 - 2u)(1 - 2v) \\ = 1 + \theta_1 \sum_{x=0}^1 \sum_{y=0}^1 (-1)^{x+y} u^x (1-u)^{1-x} v^y (1-v)^{1-y} \\ = 1 + \theta_1 C_1(u, v).$$

Σε αυτή την περίπτωση, έχουμε

$$c(u_1, v_1) c(u_2, v_2) = (1 + \theta_1 C_1(u_1, v_1))(1 + \theta_1 C_1(u_2, v_2)) \\ = 1 + \theta_1 C_1(u_1, v_1) + \theta_1 C_1(u_2, v_2) + \theta_1^2 C_2(u_1, u_2; v_1, v_2),$$

όπου

$$C_2(u_1, u_2; v_1, v_2) = \sum_{x_1=0}^1 \sum_{y_1=0}^1 \sum_{x_2=0}^1 \sum_{y_2=0}^1 (-1)^{x_1+y_1+x_2+y_2} \\ \times u_1^{x_1} (1-u_1)^{1-x_1} v_1^{y_1} (1-v_1)^{1-y_1} u_2^{x_2} (1-u_2)^{1-x_2} v_2^{y_2} (1-v_2)^{1-y_2}.$$

Λαμβάνοντας υπ' όψιν τα παραπάνω αποτελέσματα, η έκφραση του Λήμματος 5.8 λαμβάνει τη μορφή

$$f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) = \frac{m!}{(a-1)!(m-b)!(b-a-1)!} \\ \times [S_1(u_1, u_2) + \theta_1 S_2(u_1, u_2; v_1) + \theta_1 S_3(u_1, u_2; v_2) + \theta_1^2 S_4(u_1, u_2; v_1, v_2)],$$

όπου

$$S_1(u_1, u_2) = u_1^{a-1} (1-u_2)^{m-b} (u_2-u_1)^{b-a-1}$$

και οι ποσότητες $S_2(u_1, u_2; v_1)$, $S_3(u_1, u_2; v_2)$ γράφονται ως αθροίσματα τεσσάρων προσθετέων της μορφής

$$s_2(u_1, u_2; v_1) = u_1^{a-1+x_1} (1-u_1)^{1-x_1} (1-u_2)^{m-b} (u_2-u_1)^{b-a-1} v_1^{y_1} (1-v_1)^{1-y_1}, \\ s_3(u_1, u_2; v_2) = u_1^{a-1} u_2^{x_2} (1-u_2)^{m-b+1-x_2} (u_2-u_1)^{b-a-1} v_2^{y_2} (1-v_2)^{1-y_2}.$$

Όμοια, η ποσότητα $S_4(u_1, u_2; v_1, v_2)$ μπορεί να γραφεί ως άθροισμα οκτώ προσθετέων της μορφής

$$s_4(u_1, u_2; v_1, v_2) = u_1^{a-1+x_1} (1-u_1)^{1-x_1} v_1^{y_1} (1-v_1)^{1-y_1} u_2^{x_2} (1-u_2)^{m-b+1-x_2} \\ \times (u_2-u_1)^{b-a-1} v_2^{y_2} (1-v_2)^{1-y_2}.$$

Στη συνέχεια, προκειμένου να υπολογίσουμε τις περιθώριες κατανομές της πυκνότητας του Λήμματος 5.8, θα πρέπει να ολοκληρωθούν κατάλληλα οι επιμέρους προσθετέοι. Αυτό μπορεί εύκολα να πραγματοποιηθεί παρατηρώντας το εξής

$$f_{U_{a:m}; V_{[a:m]}}^{(C)}(u_1, v_1) = \frac{m!}{(a-1)!(m-a)!} u_1^{a-1} (1-u_1)^{m-a} (1 + \theta_1 C_1(u_1, v_1)), \\ f_{U_{b:m}; V_{[b:m]}}^{(C)}(u_2, v_2) = \frac{m!}{(b-1)!(m-b)!} u_2^{b-1} (1-u_2)^{m-b} (1 + \theta_1 C_1(u_2, v_2)),$$

$$f_{U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1; v_1, v_2) = \int_{u_1}^1 f_{U_{a:m}, U_{b:m}; V_{[a:m]}, V_{[b:m]}}^{(C)}(u_1, u_2; v_1, v_2) du_2,$$

όπου η $f_{U_{k:m}; V_{[k:m]}}^{(C)}(u, v)$, για $k \in \{a, b\}$ και $0 \leq u, v \leq 1$, είναι η σ.π.π. που περιγράφεται στο Λήμμα 5.7 και

$$\int_{u_1}^1 S_1(u_1, u_2) du_2 = u_1^{a-1} (1-u_1)^{m-a} \sum_{k=0}^{b-a-1} \binom{b-a-1}{k} \frac{(-1)^{b-a-1-k}}{m-a-k},$$

$$\int_{u_1}^1 s_2(u_1, u_2; v_1) du_2 = u_1^{a-1+x_1} (1-u_1)^{m-a+1-x_1} v_1^{y_1} (1-v_1)^{1-y_1} \sum_{k=0}^{b-a-1} \binom{b-a-1}{k} \frac{(-1)^{b-a-1-k}}{m-a-k},$$

$$\int_{u_1}^1 s_3(u_1, u_2; v_2) du_2 = u_1^{a-1} (1-u_1)^k v_2^{y_2} (1-v_2)^{1-y_2} \sum_{k=0}^{b-a-1} \binom{b-a-1}{k} (-1)^{b-a-1-k}$$

$$\times [\mathbf{B}(x_2+1, m-x_2-a-k+1) - \mathbf{B}_{u_1}(x_2+1, m-x_2-a-k+1)],$$

$$\int_{u_1}^1 s_4(u_1, u_2; v_1, v_2) du_2 = u_1^{a-1+x_1} (1-u_1)^{1-x_1+k} v_1^{y_1} (1-v_1)^{1-y_1}$$

$$\times v_2^{y_2} (1-v_2)^{1-y_2} \sum_{k=0}^{b-a-1} \binom{b-a-1}{k} (-1)^{b-a-1-k}$$

$$\times [\mathbf{B}(x_2+1, m-x_2-a-k+1) - \mathbf{B}_{u_1}(x_2+1, m-x_2-a-k+1)].$$

Με ανάλογο τρόπο μπορούν να υπολογισθούν και οι υπόλοιπες περιθώριες κατανομές που εμφανίζονται στον τύπο του AR_{OC} . Στην πραγματικότητα, η παραπάνω διαδικασία μπορεί να επεκταθεί για τον υπολογισμό του ποσοστού λανθασμένου συναγερμού γενικότερα για άλλες οικογένειες πολυωνυμικών συνδέσμων ή ειδικότερα για πολυωνυμικούς συνδέσμους μεγαλύτερης τάξης ($m > 4$). Προφανώς, εάν οι παράμετροι των συνδέσμων (που αντιστοιχούν στο δείγμα αναφοράς και ελέγχου) είναι κοινές, η προαναφερθείσα μεθοδολογία υπολογίζει το ποσοστό λανθασμένου συναγερμού της υποκείμενης διεργασίας.

5.5 Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

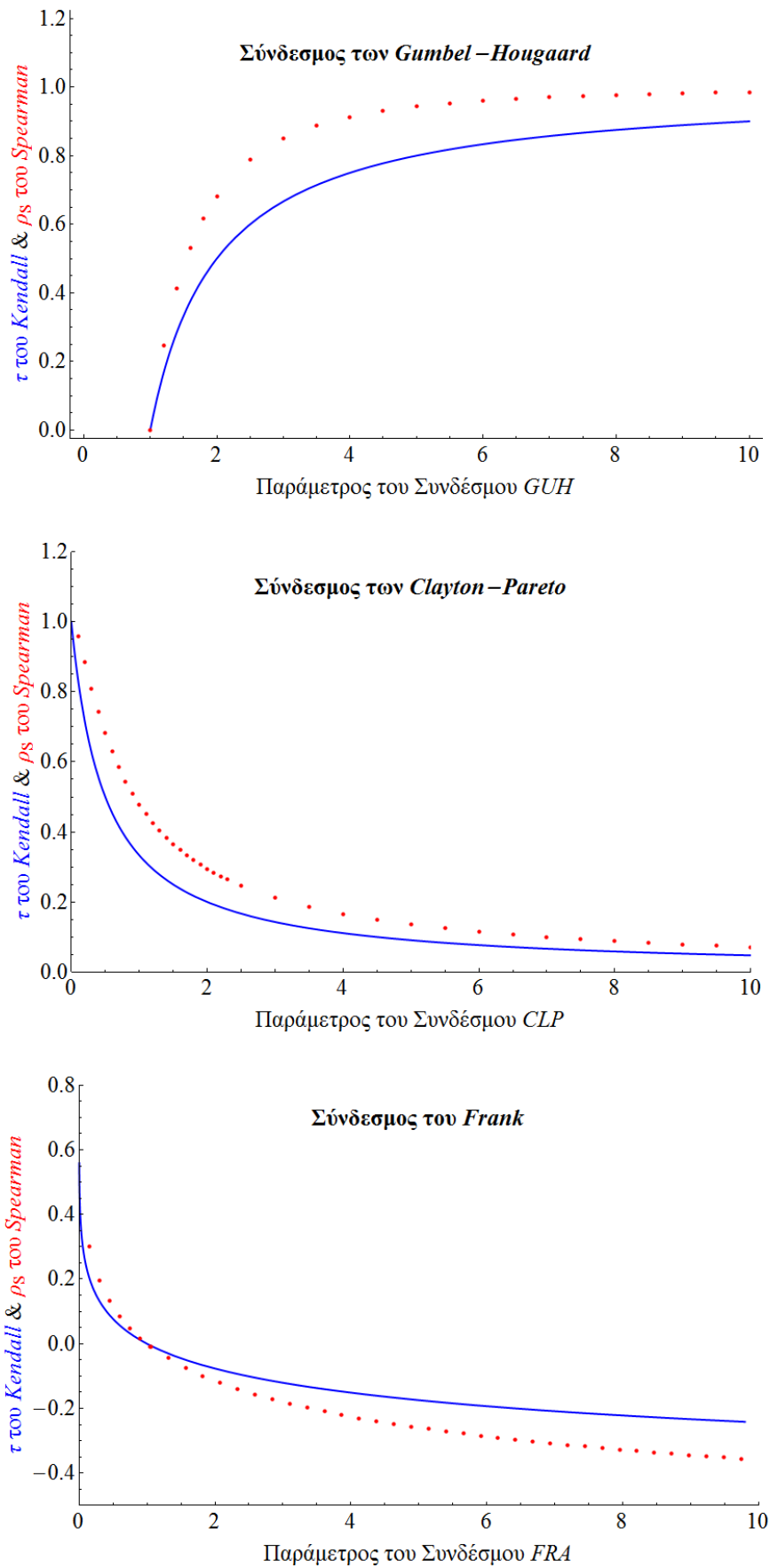
Στην τρέχουσα ενότητα, γίνεται μελέτη του σχεδιασμού του διαγράμματος OC . Για το σκοπό αυτό, αρχικά παρουσιάζεται ο Πίνακας 5.1, όπου δίνονται οι πιθανότητες συναγερμού του προτεινόμενου δίπλευρου διαγράμματος ελέγχου ($AR_{OC}^{(2)}$) ξεκινώντας από δεδομένο ποσοστό λανθασμένου συναγερμού $FAR_{OC}^{(2)}$, το οποίο κυμαίνεται από 5% έως 10% και σημειώνεται με έντονη γραφή. Για τον υπολογισμό των τιμών του $AR_{OC}^{(2)}$ χρησιμοποιήθηκε ο σύνδεσμος του *Gauss* (βλ. Πίνακα 2.5) με διαφορετικές περιθώριες κατανομές (τυπική Κανονική, Ομοιόμορφη στο (0,1), Εκθετική με παράμετρο 1) και με διάφορες δομές συσχέτισης (που περιγράφονται από την παράμετρο ρ του συγκεκριμένου συνδέσμου). Σημειώνεται ότι η εντός ελέγχου τιμή της παραμέτρου συσχέτισης ορίσθηκε ίση με $\rho_{in}^{(R)} = 0.9$. Επιπλέον, επιλέχθηκαν τρία μεγέθη για το δείγμα ελέγχου ($n = 5, 11, 25$), ενώ έγινε χρήση ενός δείγματος αναφοράς μεγέθους $m = 500$. Με σκοπό να μειωθεί το πλήθος των παραμέτρων

σχεδιασμού, δεδομένου ότι ο προσδιορισμός των τιμών τους δε γίνεται με μοναδικό τρόπο, χρησιμοποιήθηκαν οι διάμεσοι του δείγματος ελέγχου ($r = s = (n + 1)/2$). Αυτό οδήγησε στην επιλογή συμμετρικών ορίων ελέγχου από το δείγμα αναφοράς, σε αντιστοιχία πάντα με τη μεταβλητή X , οπότε ορίστηκε $b = m - a + 1$.

| Παράμετροι | | Τιμές συντελεστή συσχέτισης ρ | | | | | | | | | |
|------------|-----------|--|--------|--------|--------|--------|--------|--------|--------|-------------------|---------------|
| Σχεδιασμού | | $\rho_{out}^{(T)}$ | | | | | | | | $\rho_{in}^{(R)}$ | |
| $m = 500$ | | -0.9 | -0.8 | -0.7 | -0.6 | -0.5 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| n | (a, b) | Τυπικές Κανονικές Περιθώριες Κατανομές | | | | | | | | | |
| 5 | (30, 471) | 0.0462 | 0.1263 | 0.2217 | 0.3106 | 0.3935 | 0.3921 | 0.3108 | 0.2204 | 0.1230 | 0.0437 |
| | (40, 461) | 0.0718 | 0.1667 | 0.2642 | 0.3520 | 0.4234 | 0.4259 | 0.3502 | 0.2558 | 0.1665 | 0.0705 |
| | (50, 451) | 0.1024 | 0.2133 | 0.3007 | 0.3856 | 0.4490 | 0.4532 | 0.3900 | 0.3095 | 0.2104 | 0.1067 |
| 11 | (55, 446) | 0.0557 | 0.1627 | 0.2666 | 0.3553 | 0.4144 | 0.4072 | 0.3550 | 0.2584 | 0.1635 | 0.0538 |
| | (65, 436) | 0.0768 | 0.1959 | 0.2959 | 0.3803 | 0.4363 | 0.4339 | 0.3834 | 0.2901 | 0.1952 | 0.0748 |
| | (75, 426) | 0.1062 | 0.2260 | 0.3266 | 0.3940 | 0.4485 | 0.4557 | 0.3894 | 0.3251 | 0.2321 | 0.1041 |
| 25 | (70, 431) | 0.0532 | 0.1689 | 0.2783 | 0.3548 | 0.4123 | 0.4158 | 0.3553 | 0.2824 | 0.1715 | 0.0515 |
| | (80, 421) | 0.0762 | 0.2056 | 0.3184 | 0.3801 | 0.4244 | 0.4333 | 0.3767 | 0.3071 | 0.2036 | 0.0787 |
| | (90, 411) | 0.1051 | 0.2304 | 0.3344 | 0.4044 | 0.4434 | 0.4449 | 0.4002 | 0.3321 | 0.2431 | 0.1060 |
| n | (a, b) | Ομοιόμορφες Περιθώριες Κατανομές στο (0, 1) | | | | | | | | | |
| 5 | (30, 471) | 0.0478 | 0.1234 | 0.2177 | 0.3188 | 0.4035 | 0.3901 | 0.3144 | 0.2143 | 0.1232 | 0.0453 |
| | (40, 461) | 0.0721 | 0.1678 | 0.2655 | 0.3481 | 0.4252 | 0.4230 | 0.3477 | 0.2688 | 0.1698 | 0.0698 |
| | (50, 451) | 0.1030 | 0.2103 | 0.3077 | 0.3938 | 0.4510 | 0.4418 | 0.3934 | 0.3104 | 0.2127 | 0.1035 |
| 11 | (55, 446) | 0.0552 | 0.1579 | 0.2589 | 0.3468 | 0.4129 | 0.4155 | 0.3532 | 0.2673 | 0.1677 | 0.0538 |
| | (65, 436) | 0.0833 | 0.2005 | 0.3010 | 0.3728 | 0.4399 | 0.4285 | 0.3746 | 0.2906 | 0.2010 | 0.0788 |
| | (75, 426) | 0.1004 | 0.2353 | 0.3307 | 0.3995 | 0.4604 | 0.4420 | 0.3992 | 0.3292 | 0.2281 | 0.1061 |
| 25 | (70, 431) | 0.0547 | 0.1668 | 0.2722 | 0.3553 | 0.4112 | 0.4157 | 0.3573 | 0.2814 | 0.1723 | 0.0516 |
| | (80, 421) | 0.0795 | 0.2049 | 0.3054 | 0.3840 | 0.4195 | 0.4319 | 0.3857 | 0.3095 | 0.2072 | 0.0738 |
| | (90, 411) | 0.1009 | 0.2428 | 0.3398 | 0.3930 | 0.4392 | 0.4461 | 0.3985 | 0.3244 | 0.2371 | 0.0999 |
| n | (a, b) | Τυπικές Εκθετικές Περιθώριες Κατανομές | | | | | | | | | |
| 5 | (30, 471) | 0.0439 | 0.1207 | 0.2252 | 0.3247 | 0.3836 | 0.3938 | 0.3101 | 0.2186 | 0.1202 | 0.0491 |
| | (40, 461) | 0.0741 | 0.1679 | 0.2641 | 0.3510 | 0.4304 | 0.4291 | 0.3463 | 0.2580 | 0.1623 | 0.0740 |
| | (50, 451) | 0.1087 | 0.2103 | 0.3032 | 0.3778 | 0.4467 | 0.4518 | 0.3923 | 0.2986 | 0.2134 | 0.1071 |
| 11 | (55, 446) | 0.0555 | 0.1598 | 0.2674 | 0.3505 | 0.4095 | 0.4124 | 0.3448 | 0.2681 | 0.1605 | 0.0544 |
| | (65, 436) | 0.0814 | 0.1995 | 0.3001 | 0.3751 | 0.4209 | 0.4304 | 0.3797 | 0.2983 | 0.1962 | 0.0769 |
| | (75, 426) | 0.1074 | 0.2246 | 0.3304 | 0.3937 | 0.4439 | 0.4499 | 0.3969 | 0.3219 | 0.2312 | 0.1020 |
| 25 | (70, 431) | 0.0527 | 0.1784 | 0.2765 | 0.3573 | 0.4124 | 0.4209 | 0.3607 | 0.2766 | 0.1734 | 0.0531 |
| | (80, 421) | 0.0744 | 0.2020 | 0.3114 | 0.3773 | 0.4291 | 0.4284 | 0.3787 | 0.3139 | 0.2027 | 0.0772 |
| | (90, 411) | 0.1039 | 0.2331 | 0.3311 | 0.3988 | 0.4445 | 0.4461 | 0.3970 | 0.3370 | 0.2379 | 0.0980 |

Πίνακας 5.1: Πιθανότητες Συναγερμού $AR_{OC}^{(2)}$ του δίπλευρου διαγράμματος OC για δεδομένο σχεδιασμό ($b = m - a + 1$ & $r = s = (n + 1)/2$), χρησιμοποιώντας το σύνδεσμο του Gauss με διαφορετικές περιθώριες κατανομές και $\rho_{in}^{(R)} = 0.9$

5.5. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου



Σχήμα 5.2: Συντελεστές Συσχέτισης τ του Kendall και ρ_S του Spearman για τρεις συνδέσμους (GUH,CLP,FRA)

Από τον Πίνακα 5.1 είναι εμφανές ότι το ποσοστό λανθασμένου συναγεμού του διαγράμματος *OC* δεν εξαρτάται από τις περιθώριες, το οποίο φυσικά έχει ήδη αποδειχθεί στο Πόρισμα 5.4. Θα πρέπει να αναφερθεί ότι παρόλο που το *FAR* δεν αλλάζει για απόλυτες τιμές του συντελεστή συσχέτισης ρ (η παρατηρούμενη συμμετρία των τιμών του οφείλεται στις συναρτήσεις *min* και *max*), οι τιμές του *FAR* διαφέρουν ανάλογα με την επιλογή του επιπέδου συσχέτισης. Κατά συνέπεια, η απόδοση του διαγράμματος επηρεάζεται από τη δομή εξάρτησης των υπό μελέτη χαρακτηριστικών που παρακολουθούνται. Είναι προφανές ότι όσο πιο ισχυρή είναι η συσχέτιση (θετική ή αρνητική), τόσο πιο ισχυρή γίνεται η απόδοση του προτεινόμενου σχήματος.

Η τελευταία παρατήρηση υποδεικνύει ότι το νέο διάγραμμα δεν αναμένεται να είναι αποτελεσματικό στην περίπτωση συνδέσμων, όπως ο σύνδεσμος *FGM*, των *AMH* ή των *Gumbel-Barnett (GUB)*, καθώς το εύρος των συσχετίσεών τους είναι αρκετά περιορισμένο. Για παράδειγμα, είναι ευρέως γνωστό ότι ο σύνδεσμος *FGM* μπορεί να μοντελοποιήσει σχετικά αδύναμες εξαρτήσεις, ενώ αντίθετα, σύνδεσμοι όπως αυτοί των *CLP*, *GUH* και *FRA* μπορούν να μοντελοποιήσουν υψηλές συσχετίσεις.

Στο Σχήμα 5.2, απεικονίζονται τα γραφήματα των συντελεστών συσχέτισης τ του *Kendall* και ρ_S του *Spearman* έναντι των παραμέτρων που συνδέονται με τους τρεις τελευταίους συνδέσμους. Δεδομένου ότι εμπλέκονται διατεταγμένες παρατηρήσεις στον κανόνα απόφασης του προτεινόμενου σχήματος, είναι λογικό να εστιάσουμε το ενδιαφέρον μας στο συντελεστή συσχέτισης του *Spearman*. Συνεπώς, για να επιτευχθεί μία σχετικά υψηλή τιμή του προαναφερθέντος συντελεστή, για παράδειγμα $\rho_S \geq 0.8$, μπορεί κανείς να επιλέξει μια τιμή μεγαλύτερη από 2.5 και μικρότερη από 1/3 για την παράμετρο του συνδέσμου που σχετίζεται με την οικογένεια *GUH* και *CLP*, αντίστοιχα, ενώ μια παράμετρος που λαμβάνει τιμή μεγαλύτερη από 10^4 (αντίστοιχα μικρότερη από 10^{-4}), θεωρώντας ισχυρή αρνητική (αντίστοιχα θετική) συσχέτιση, θα μπορούσε να εφαρμοστεί στην περίπτωση του συνδέσμου του *FRA*.

5.6 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

Η επίδοση ενός διαγράμματος ελέγχου αξιολογείται λαμβάνοντας συνήθως υπ' όψιν ορισμένα χαρακτηριστικά που συνδέονται με την κατανομή του μήκους ροής ή/και της δυναμικής του να ενεργοποιεί συναγεμό όταν η διεργασία μετατοπίζεται σε μία εκτός ελέγχου κατάσταση, καθώς και την ικανότητά του να μην παράγει σήματα λανθασμένου συναγεμού, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Στον έλεγχο ποιό-

5.6. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

τητας, τόσο η εντός ελέγχου όσο και η εκτός ελέγχου απόδοση του διαγράμματος έχουν μεγάλη σημασία. Σε αυτή την ενότητα, παρουσιάζουμε δύο διαφορετικές προσεγγίσεις για τη σύγκριση διαφορετικών διαγραμμάτων. Σύμφωνα με την πρώτη, μια προκαθορισμένη τιμή για το FAR επιτυγχάνεται για κάθε γράφημα (π.χ. 5% ή 10%) και στη συνέχεια συγκρίνονται οι πιθανότητες συναγεμού. Με βάση τη δεύτερη προσέγγιση, μπορεί κάποιος εναλλακτικά να χρησιμοποιήσει μία τυπική τιμή για το ARL_{in} (π.χ. 200, 370 ή 500) και να αξιολογήσει τις αντίστοιχες τιμές για το ARL_{out} , δεδομένου ότι συγκεκριμένες μετατοπίσεις λαμβάνουν χώρα στην υποκείμενη διεργασία. Ως αποτέλεσμα, αυτό που έχει ενδιαφέρον είναι να εστιάσουμε στη μέγιστη δυνατή απομάκρυνση των πιθανοτήτων συναγεμού (AR) και των τιμών του ARL_{out} από τις εντός ελέγχου τιμές τους (FAR και ARL_{in} , αντίστοιχα) και ταυτόχρονα στην ταχύτητα ανίχνευσης ενός εκτός ελέγχου σήματος.

Ακολουθώντας αρχικά την πρώτη προσέγγιση, μελετάται η επίδοση του δίπλευρου διαγράμματος ελέγχου OC , καθώς συγκρίνεται με το δίπλευρο διάγραμμα $O2$ που προτάθηκε στο προηγούμενο κεφάλαιο. Τα αποτελέσματα της μεταξύ τους σύγκρισης παρουσιάζονται στον Πίνακα 5.2. Πιο συγκεκριμένα, θεωρώντας τα ίδια μεγέθη για τα δείγματα αναφοράς και ελέγχου που χρησιμοποιήθηκαν στην Ενότητα 4.6 και διατηρώντας σταθερή την τιμή των υπόλοιπων παραμέτρων (σχεδιασμού), υπολογίσθηκε το FAR (τιμές με έντονη γραφή) για το σύνδεσμο του FRA με $a \geq 10^4$ και το σύνδεσμο των GUH με $a \geq 4$. Στη συνέχεια, προσδιορίσθηκε το ποσοστό συναγεμού για διαφορετικές μετατοπίσεις της μέσης τιμής κανονικών περιθώριων κατανομών της μορφής $N(\mu, 1)$. Το νέο διάγραμμα, τα όρια του οποίου εμπλέκουν διατεταγμένες παρατηρήσεις και συμμεταβλητές, ανιχνεύει τις μετατοπίσεις αρκετά γρήγορα. Επιπλέον, η απόδοσή του βελτιώνεται καθώς η συσχέτιση γίνεται πιο ισχυρή μέχρι να φτάσει την απόδοση του διαγράμματος $O2$ (η οποία υποδεικνύεται με σκίαση) και από εκεί και έπειτα το νέο διάγραμμα OC ξεπερνά την επίδοση του διαγράμματος $O2$. Αυτό μπορεί να αποδοθεί στο γεγονός ότι οι συμμεταβλητές φαίνεται να συμπεριφέρονται σχεδόν όπως οι διατεταγμένες παρατηρήσεις που εμφανίζονται στον κανόνα απόφασης του διαγράμματος $O2$ και ως αποτέλεσμα, τα όρια ελέγχου που σχετίζονται με τη μεταβλητή Y να γίνονται πλατύτερα. Κατά συνέπεια, η δεύτερη συνθήκη στον κανόνα απόφασης του διαγράμματος OC δεν ενεργοποιεί πολλούς συναγεμούς και το προτεινόμενο σχήμα γίνεται πιο ανθεκτικό.

Στη συνέχεια, ακολουθώντας τη δεύτερη προσέγγιση¹, πρώτα καθορίστηκε η τιμή του

¹ Τα βήματα της αλγοριθμικής διαδικασίας σύμφωνα με την οποία υπολογίζονται το εντός και το εκτός ελέγχου μέσο μήκος ροής είναι παρόμοια με αυτά που παρουσιάζονται στον Πίνακα 4.4, με τη διαφορά ότι στο Βήμα 5 διαπιστώνεται εάν παραβιάζεται ή όχι ο κανόνας ελέγχου της Σχέσης (5.3).

ARL_{in} ώστε να είναι περίπου ίση με 200 και ύστερα η απόδοση του δίπλευρου διαγράμματος OC συγκρίθηκε όχι μόνο με το ημιπαραμετρικό διάγραμμα $O2$, αλλά και με το κλασικό παραμετρικό διάγραμμα χ^2 , καθώς και με τα μη παραμετρικά διαγράμματα SN^2 και SR^2 των Boone and Chakraborti (2012). Τα αποτελέσματα παρουσιάζονται στον Πίνακα 5.3. Για το σκοπό αυτό, δημιουργήθηκαν δισδιάστατα δεδομένα από την κατανομή t με 5 β.ε. και παραμέτρους μ , Σ , $\Sigma_{(5)}$. Η αρχική επιλογή των παραμέτρων βασίστηκε στις εργασίες των Boone and Chakraborti (2012) και Koutras and Sofikitou (2017a), όπου $\rho = 0.50$. Στη συνέχεια, ο πίνακας $\Sigma_{(5)}$ επιλέχθηκε έτσι, ώστε να επιτευχθούν τρία ακόμη επίπεδα συσχέτισης, δηλαδή $\rho = 0.70, 0.90, 0.95$ και τα όρια ελέγχου προσδιορίστηκαν μέσω προσομοίωσης για την επίτευξη του επιθυμητού ARL_{in} . Στη μελέτη προσομοίωσης που διεξήχθη, χρησιμοποιήθηκε ένα δείγμα αναφοράς μεγέθους $m = 1000$ και για τα δύο διαγράμματα $O2, OC$. Η καλύτερη εκτός ελέγχου τιμή ARL (ARL_{out}) είναι υπογραμμισμένη (υποδεικνύμενη με έντονους χαρακτήρες), όταν για το διάγραμμα OC θεωρούνται συμμετρικά (μη συμμετρικά) όρια ελέγχου, τα οποία συμβολίζονται με a και b (a^* και b^*).

Η μελέτη προσομοίωσης που πραγματοποιήθηκε αποκαλύπτει ότι το διάγραμμα χ^2 δε φαίνεται να έχει καλή απόδοση σε σύγκριση με τα άλλα διαγράμματα, το οποίο είναι λογικό, δεδομένου ότι τα δεδομένα δεν προέρχονται από κανονική κατανομή. Όταν συμβαίνουν ίσες μετατοπίσεις ($a = c$), το διάγραμμα ελέγχου SR^2 είναι ανώτερο στην ανίχνευση μικρών μετατοπίσεων, ενώ το διάγραμμα $O2$ είναι πιο ευαίσθητο στην ανίχνευση μεγάλων μετατοπίσεων. Παρόλα αυτά, το διάγραμμα ελέγχου SR^2 μπορεί να εφαρμοστεί μόνο για διαγώνια συμμετρικές κατανομές (*diagonally symmetric distributions*) [βλ. Serfling (2006), Jurečková *et al.* (2012)] ή όταν η εκτίμηση του πίνακα της υποκείμενης διεργασίας είναι αναστρέψιμη. Η τελευταία παρατήρηση είναι επίσης απαραίτητη για την εφαρμογή του διαγράμματος SN^2 . Η περίπτωση των άνισων μέσων μετατοπίσεων ($a \neq b$) έχει μελετηθεί λεπτομερώς στην Ενότητα 4.6.

Αξίζει να τονισθεί ότι κάνοντας χρήση συμμετρικών ορίων ελέγχου, το διάγραμμα OC έχει πάντα καλύτερη απόδοση από το διάγραμμα χ^2 , όπως αναμενόταν, αλλά ελαφρώς χειρότερη σε σύγκριση με τα υπόλοιπα διαγράμματα. Ωστόσο, καθώς η συσχέτιση γίνεται ισχυρότερη, η αποτελεσματικότητα του διαγράμματος OC φαίνεται να μοιάζει με εκείνη του διαγράμματος $O2$. Εφόσον η κατανομή μιας διατεταγμένης παρατήρησης δεν είναι συμμετρική, ακόμη και αν η κατανομή του αρχικού χαρακτηριστικού είναι συμμετρική, φαίνεται λογικό να χρησιμοποιηθούν ασύμμετρα όρια ελέγχου. Το βασικό χαρακτηριστικό του νέου σχήματος είναι ότι η απόδοσή του είναι ομοιόμορφα καλύτερη στην περίπτωση

5.6. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

μη συμμετρικών ορίων. Παρόμοια αποτελέσματα παρατηρούνται για μεγαλύτερα μεγέθη δείγματος.

| Παράμετροι | | | | Διαγράμματα O2 & OC | | | | | | | | |
|-------------------|----------|---------------------|-----------------------|--|------|------|------|-------------|------|------|------|------|
| Σχεδιασμού | | Όρια | Συνδέσμου <i>a</i> | Μετατοπίσεις Μέσης Τιμής | | | | | | | | |
| <i>m</i> | <i>n</i> | | | -2 | -1.5 | -1 | -0.5 | 0 | 0.5 | 1 | 1.5 | 2 |
| Σύνδεσμος GUH : | | | | $C_a^{GUH}(u, v) = \exp(-[(-\log u)^a + (-\log v)^a]^{1/a}), a \geq 1$ | | | | | | | | |
| 40 | 5 | (2,39) | 4 | 0.81 | 0.55 | 0.29 | 0.11 | 0.04 | 0.06 | 0.18 | 0.46 | 0.75 |
| | | (3,38) | 6 | 0.87 | 0.63 | 0.33 | 0.11 | 0.04 | 0.08 | 0.26 | 0.57 | 0.84 |
| | | (4,37) | 8 | 0.90 | 0.76 | 0.43 | 0.14 | 0.05 | 0.11 | 0.34 | 0.66 | 0.89 |
| | | Διάγραμμα O2 | (3,37) | 0.95 | 0.76 | 0.43 | 0.14 | 0.05 | 0.13 | 0.42 | 0.76 | 0.95 |
| | | (5,36) | 10 | 0.94 | 0.75 | 0.44 | 0.17 | 0.06 | 0.15 | 0.42 | 0.74 | 0.93 |
| 40 | 5 | (3, 38) | 4 | 0.88 | 0.67 | 0.40 | 0.16 | 0.07 | 0.11 | 0.30 | 0.60 | 0.87 |
| | | (4, 37) | 6 | 0.91 | 0.72 | 0.42 | 0.16 | 0.06 | 0.12 | 0.36 | 0.68 | 0.90 |
| | | (5, 36) | 8 | 0.94 | 0.76 | 0.46 | 0.18 | 0.07 | 0.16 | 0.43 | 0.75 | 0.94 |
| | | Διάγραμμα O2 | (5,36) | 0.97 | 0.84 | 0.52 | 0.20 | 0.08 | 0.19 | 0.51 | 0.83 | 0.97 |
| | | (6,35) | 10 | 0.96 | 0.82 | 0.52 | 0.22 | 0.09 | 0.20 | 0.50 | 0.80 | 0.96 |
| 40 | 25 | (7,34) | 5 | 1 | 0.97 | 0.70 | 0.22 | 0.05 | 0.18 | 0.66 | 0.97 | 1 |
| | | (9,32) | 10 | 1 | 0.99 | 0.81 | 0.28 | 0.04 | 0.25 | 0.79 | 0.99 | 1 |
| | | (10,31) | 15 | 1 | 1 | 0.85 | 0.33 | 0.05 | 0.31 | 0.84 | 0.99 | 1 |
| | | (11,30) | 20 | 1 | 1 | 0.89 | 0.40 | 0.07 | 0.39 | 0.89 | 1 | 1 |
| | | Διάγραμμα O2 | (10,31) | 1 | 1 | 0.92 | 0.40 | 0.05 | 0.39 | 0.92 | 1 | 1 |
| 40 | 25 | (9, 32) | 5 | 1 | 0.99 | 0.85 | 0.37 | 0.10 | 0.33 | 0.83 | 0.99 | 1 |
| | | (11, 30) | 10 | 1 | 1 | 0.91 | 0.45 | 0.10 | 0.43 | 0.90 | 1 | 1 |
| | | (11, 30) | 15 | 1 | 1 | 0.09 | 0.44 | 0.08 | 0.44 | 0.96 | 1 | 1 |
| | | (10, 31) | 20 | 1 | 1 | 0.98 | 0.54 | 0.09 | 0.54 | 0.99 | 1 | 1 |
| | | Διάγραμμα O2 | (9,32) | 1 | 1 | 0.98 | 0.57 | 0.10 | 0.58 | 0.99 | 1 | 1 |
| Σύνδεσμος FRA : | | | | $C_a^{FRA}(u, v) = \log_a[1 + (a^u - 1)(a^v - 1)/(a - 1)], 0 < a \neq 1$ | | | | | | | | |
| 40 | 25 | (2,39) | 10^4 | 0.97 | 0.75 | 0.37 | 0.11 | 0.04 | 0.12 | 0.37 | 0.77 | 0.97 |
| | | (6,35) | 10^6 | 1 | 0.99 | 0.74 | 0.20 | 0.04 | 0.21 | 0.74 | 0.99 | 1 |
| | | (7,34) | 10^8 | 1 | 1 | 0.84 | 0.24 | 0.04 | 0.23 | 0.84 | 1 | 1 |
| | | (8,33) | 10^{10} | 1 | 1 | 0.92 | 0.31 | 0.04 | 0.31 | 0.92 | 1 | 1 |
| | | Διάγραμμα O2 | (9,32) | 1 | 1 | 0.94 | 0.35 | 0.04 | 0.36 | 0.94 | 1 | 1 |
| | | (9,32) | 10^{12} | 1 | 1 | 0.97 | 0.41 | 0.05 | 0.41 | 0.97 | 1 | 1 |
| 40 | 25 | (5,36) | 10^4 | 1 | 0.97 | 0.66 | 0.22 | 0.09 | 0.21 | 0.66 | 0.96 | 1 |
| | | (8,33) | 10^6 | 1 | 1 | 0.91 | 0.36 | 0.09 | 0.37 | 0.90 | 1 | 1 |
| | | (9,32) | 10^8 | 1 | 1 | 0.96 | 0.44 | 0.08 | 0.44 | 0.96 | 1 | 1 |
| | | (10,31) | 10^{10} | 1 | 1 | 0.98 | 0.54 | 0.09 | 0.54 | 0.99 | 1 | 1 |
| | | Διάγραμμα O2 | (9,32) | 1 | 1 | 0.98 | 0.57 | 0.10 | 0.58 | 0.99 | 1 | 1 |
| | | (11,30) | 10^{12} | 1 | 1 | 1 | 0.66 | 0.10 | 0.67 | 1 | 1 | 1 |

Πίνακας 5.2: Απόδοση του δίπλευρου διαγράμματος OC βάσει του AR σε σχέση με το διάγραμμα O2 με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων μέσης τιμής, χρησιμοποιώντας δεδομένα από $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2$)

Δισδιάστατο Διάγραμμα Ελέγχου με Χρήση
Διατεταγμένων Παρατηρήσεων και Συμμεταβλητών

| | | | | | | | |
|-----------------------------|---|------------------------------|--------------------------------|--------------------------------|------------------------|---------------------------|----------------------------|
| $n = 15$ | Δισδιάστατα Δεδομένα από κατανομή $t(5)$ με $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & 0.50 \\ 0.50 & 1 \end{pmatrix}$ & $\Sigma_{t(5)} = \begin{pmatrix} 0.60 & 0.30 \\ 0.30 & 0.60 \end{pmatrix}$ | | | | | | |
| (μ_x, μ_y) | χ^2 | SN^2 | SR^2 | $O2$ | OC | | |
| Μετατοπίσεις Μέσης Τιμής | $LCL = 0.00$ $UCL = 12.22$ | $LCL = 0.00$ $UCL = 8.70$ | $LCL = 0.00$ $UCL = 8.73$ | $a = c = 166$ $b = d = 835$ | $a = 131$ $b = 870$ | $a^* = 48$ $b^* = 845$ | $a^* = 108$ $b^* = 857$ |
| (0.0, 0.0) | 198.44 | 207.02 | 199.07 | 201.14 | 201.01 | 199.50 | 198.86 |
| (0.2, 0.2) | 74.60 | 53.46 | <u>46.73</u> | 57.85 | 67.69 | 34.25 | 45.94 |
| (0.4, 0.4) | 15.32 | 11.79 | 9.55 | 11.26 | 15.15 | 7.53 | 10.64 |
| (0.6, 0.6) | 4.07 | 4.08 | <u>3.30</u> | <u>3.30</u> | 4.19 | 2.76 | 3.30 |
| (0.8, 0.8) | 1.70 | 2.05 | 1.78 | <u>1.57</u> | 1.84 | 1.49 | 1.63 |
| (1.0, 1.0) | 1.15 | 1.43 | 1.29 | 1.12 | 1.26 | 1.12 | 1.17 |
| $n = 15$ | Δισδιάστατα Δεδομένα από κατανομή $t(5)$ με $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & 0.70 \\ 0.70 & 1 \end{pmatrix}$ & $\Sigma_{t(5)} = \begin{pmatrix} 0.60 & 0.42 \\ 0.42 & 0.60 \end{pmatrix}$ | | | | | | |
| (μ_x, μ_y) | χ^2 | SN^2 | SR^2 | $O2$ | OC | | |
| Μετατοπίσεις Μέσης Τιμής | $LCL = 0.00$ $UCL = 12.33$ | $LCL = 0.00$ $UCL = 8.60$ | $LCL = 0.000$ $UCL = 8.743$ | $a = c = 166$ $b = d = 835$ | $a = 139$ $b = 862$ | $a^* = 54$ $b^* = 835$ | $a^* = 125$ $b^* = 851$ |
| (0.0, 0.0) | 208.03 | 218.57 | 206.42 | 206.53 | 203.17 | 201.68 | 204.05 |
| (0.2, 0.2) | 87.67 | <u>51.55</u> | 51.61 | 60.65 | 70.77 | 35.10 | 50.35 |
| (0.4, 0.4) | 19.06 | 10.95 | 11.26 | 11.78 | 15.54 | 7.85 | 11.03 |
| (0.6, 0.6) | 5.00 | <u>3.72</u> | 3.83 | <u>3.49</u> | 4.31 | 2.69 | 3.45 |
| (0.8, 0.8) | 2.01 | 1.99 | 2.03 | <u>1.64</u> | 1.89 | 1.46 | 1.64 |
| (1.0, 1.0) | 1.26 | 1.42 | 1.41 | 1.16 | 1.25 | 1.12 | 1.18 |
| $n = 15$ | Δισδιάστατα Δεδομένα από κατανομή $t(5)$ με $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & 0.90 \\ 0.90 & 1 \end{pmatrix}$ & $\Sigma_{t(5)} = \begin{pmatrix} 0.60 & 0.54 \\ 0.54 & 0.60 \end{pmatrix}$ | | | | | | |
| (μ_x, μ_y) | χ^2 | SN^2 | SR^2 | $O2$ | OC | | |
| Μετατοπίσεις Μέσης Τιμής | $LCL = 0.00$ $UCL = 12.22$ | $LCL = 0.00$ $UCL = 8.37$ | $LCL = 0.000$ $UCL = 8.635$ | $a = c = 169$ $b = d = 832$ | $a = 152$ $b = 849$ | $a^* = 50$ $b^* = 818$ | $a^* = 130$ $b^* = 835$ |
| (0.0, 0.0) | 197.31 | 213.53 | 193.66 | 205.19 | 202.45 | 201.32 | 199.87 |
| (0.2, 0.2) | 90.31 | 58.56 | <u>53.37</u> | 61.70 | 70.47 | 31.96 | 48.07 |
| (0.4, 0.4) | 21.53 | 13.61 | <u>11.84</u> | 12.22 | 15.48 | 7.13 | 10.39 |
| (0.6, 0.6) | 5.86 | 4.89 | 4.18 | <u>3.70</u> | 4.28 | 2.54 | 3.22 |
| (0.8, 0.8) | 2.29 | 2.47 | 2.18 | <u>1.75</u> | 1.86 | 1.43 | 1.61 |
| (1.0, 1.0) | 1.36 | 1.68 | 1.50 | <u>1.20</u> | 1.24 | 1.24 | 1.18 |
| $n = 15$ | Δισδιάστατα Δεδομένα από κατανομή $t(5)$ με $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}$ & $\Sigma_{t(5)} = \begin{pmatrix} 0.60 & 0.57 \\ 0.57 & 0.60 \end{pmatrix}$ | | | | | | |
| (μ_x, μ_y) | χ^2 | SN^2 | SR^2 | $O2$ | OC | | |
| Μετατοπίσεις Μέσης Τιμής | $LCL = 0.00$ $UCL = 12.18$ | $LCL = 0.00$ $UCL = 8.33$ | $LCL = 0.000$ $UCL = 8.609$ | $a = c = 171$ $b = d = 830$ | $a = 158$ $b = 843$ | $a^* = 60$ $b^* = 815$ | $a^* = 121$ $b^* = 822$ |
| (0.0, 0.0) | 197.91 | 194.18 | 196.72 | 200.16 | 202.84 | 206.18 | 197.17 |
| (0.2, 0.2) | 91.94 | 58.01 | <u>55.43</u> | 61.30 | 72.92 | 32.70 | 38.12 |
| (0.4, 0.4) | 22.09 | 14.09 | <u>12.22</u> | 12.53 | 15.37 | 7.27 | 8.78 |
| (0.6, 0.6) | 6.10 | 5.22 | 4.33 | <u>3.79</u> | 4.22 | 2.64 | 2.91 |
| (0.8, 0.8) | 2.41 | 2.71 | 2.23 | <u>1.80</u> | 1.89 | 1.44 | 1.53 |
| (1.0, 1.0) | 1.40 | 1.77 | 1.52 | <u>1.21</u> | 1.25 | 1.12 | 1.14 |

Πίνακας 5.3: Απόδοση του δίπλευρου διαγράμματος OC βάσει του ARL_{out} σε σχέση με 4 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων μέσης τιμής, χρησιμοποιώντας δεδομένα από $t(5)$ κατανομή για δεδομένο σχεδιασμό ($b = m - a + 1$, $d = m - c + 1$ & $r = s = (n + 1)/2$)

Συνοψίζοντας, θα πρέπει να σημειωθεί ότι οι τιμές των μη συμμετρικών ορίων a^* και b^* μπορούν εύκολα να προσδιορισθούν εκτελώντας επαναλαμβανόμενες δοκιμές και επιλέγοντας τις πλέον κατάλληλες παραμέτρους σχεδιασμού, ώστε να επιτευχθεί το επιθυμητό ARL_{in} . Για να διευκολυνθεί η κατασκευή του διαγράμματος $O2$, θα πρέπει να δημιουργηθούν πίνακες για συγκεκριμένες ομάδες, ούτως ώστε ο χρήστης του διαγράμματος ελέγχου να είναι σε θέση να επιλέξει τις παραμέτρους σύμφωνα με το επίπεδο του FAR ή ARL_{in} που επιδιώκει και τις τρέχουσες ερευνητικές του ανάγκες (π.χ. μεγέθη δείγματος, δομή εξάρτησης, τύπος μετατοπίσεων που επιθυμεί να ανιχνεύσει).

5.7 Εφαρμογή σε πραγματικά δεδομένα

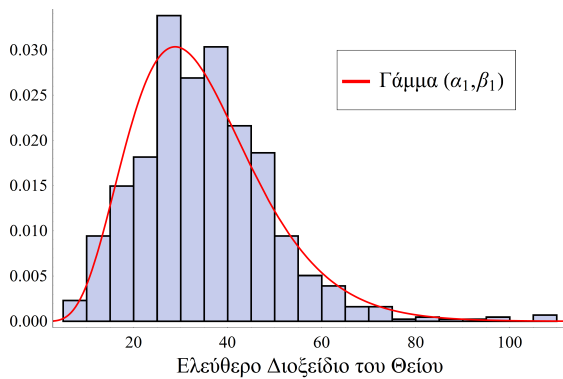
Στην παρούσα ενότητα, το διάγραμμα OC εφαρμόζεται στα πραγματικά δεδομένα λευκού οίνου της ποικιλίας *Vinho Verde* που χρησιμοποιήθηκαν και παρουσιάστηκαν στην Ενότητα 4.7. Ακολουθώντας, αρχικά, παρόμοια λογική με αυτή του προηγούμενου κεφαλαίου και θεωρώντας ότι τα δεδομένα συλλέχθηκαν με χρονολογική σειρά, οι πρώτες 3788 παρατηρήσεις (από τις 4898 παρατηρήσεις του συνόλου δεδομένων) αξιοποιήθηκαν ως ιστορικά δεδομένα, συνιστώντας ουσιαστικά το δείγμα αναφοράς μας. Στη συνέχεια, ελήφθησαν υπ' όψιν μόνο οι 500 πρώτες εντός ελέγχου παρατηρήσεις, στις οποίες το επίπεδο ποιότητας είναι τουλάχιστον 7. Υπενθυμίζεται ότι η ένδειξη 7 αντιστοιχεί στο τυπικό αποδεκτό επίπεδο ποιότητας οίνου (βλ. Cortez *et al.* (2009)). Για τους σκοπούς της γραφικής απεικόνισης διαγράμματος OC , το σύνολο δεδομένων ομαδοποιήθηκε σε υποομάδες μεγέθους $n = 5$. Συνεπώς, τα 100 δείγματα αποτελούν το δείγμα αναφοράς Φάσης I, βάσει του οποίου καθορίστηκαν τα όρια ελέγχου (τα οποία επαναπροσδιορίζονταν μέχρις ότου και τα 100 σημεία της Φάσης I να είναι εντός ελέγχου), ενώ οι υπόλοιπες 1110 παρατηρήσεις χρησιμοποιήθηκαν στην ανάλυση της Φάσης II. Ειδικότερα, στην ανάλυση της Φάσης II, επιλέχθηκαν δύο διαφορετικά δείγματα δοκιμής: ένα με επίπεδο ποιότητας 6 ($LV6$) και ένα με επίπεδο ποιότητας 5 ($LV5$). Οι 580 και 310 παρατηρήσεις που προέκυψαν, ομαδοποιήθηκαν, επίσης, σε υποομάδες μεγέθους $n = 5$.

Στο σημείο αυτό αυτό, θα πρέπει να τονισθεί ότι η επιλογή των προς μελέτη μεταβλητών (χαρακτηριστικών) διαφέρει από αυτή του διαγράμματος $O2$, καθώς το ποσοστό λανθασμένου συναγεμού του διαγράμματος OC επηρεάζεται από την επιλογή του συνδέσμου. Αυτό έχει ως αποτέλεσμα, ο προσδιορισμός των ορίων ελέγχου να μην μπορεί να γίνει χωρίς να διερευνηθεί πρώτα εάν τα ζεύγη των παρατηρήσεων της Φάσης I μοντελοποιούνται μέσω ενός δισδιάστατου συνδέσμου από κάποια γνωστή οικογένεια, κατά προτίμηση με

σχετικά υψηλό εύρος συσχέτισης (όπως για παράδειγμα οι σύνδεσμοι των *CLP, GUH, FRA*). Ως εκ τούτου, από τις 11 συνεχείς μεταβλητές του αρχικού συνόλου δεμένων, επιλέχθηκαν το ελεύθερο και ολικό διοξείδιο του θείου, στα οποία -όπως θα δούμε παρακάτω- ο σύνδεσμος των *GUH* έχει καλή προσαρμογή.

Η διαδικασία για την εύρεση του συνδέσμου που προσαρμόζεται επαρκώς στα χαρακτηριστικά μας έχει ως εξής. Πρώτα, διερευνάται εάν οι περιθώριες (μονοδιάστατες) κατανομές ταιριάζουν με κάποιες από τις ήδη γνωστές συνεχείς κατανομές που είναι διαθέσιμες στη βιβλιογραφία. Η προκαταρκτική ανάλυση αποκάλυψε ότι οι κατανομές του ελεύθερου και του ολικού διοξειδίου του θείου, ταιριάζουν με την κατανομή Γάμμα, συμβ. $\Gamma(\alpha, \beta)$, με σ.π.π.

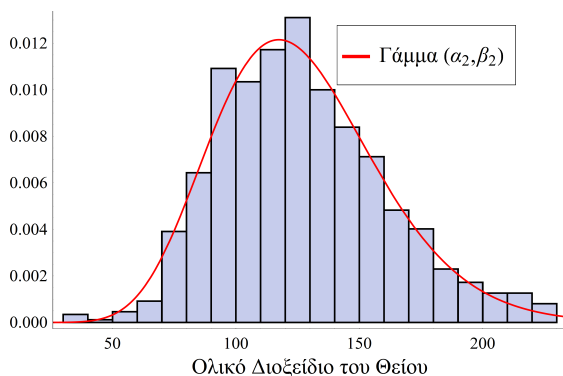
$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$



Σχήμα 5.3: Κοινό γράφημα σ.π.π. της $\Gamma(\alpha_1, \beta_1)$ και ιστογράμματος σχετικών συχνοτήτων για το Ελεύθερο Διοξείδιο του Θείου

| Έλεγχος | Τιμή σ.σ. | <i>p</i> - value |
|--------------------|-----------|------------------|
| Anderson-Darling | 0.5073 | 0.7394 |
| Cramér-von Mises | 0.0728 | 0.7345 |
| Kolmogorov-Smirnov | 0.0230 | 0.7356 |

Πίνακας 5.4: Έλεγχοι καλής προσαρμογής για την κατανομή του Ελεύθερου Διοξειδίου του Θείου



Σχήμα 5.4: Κοινό γράφημα σ.π.π. της $\Gamma(\alpha_2, \beta_2)$ και ιστογράμματος σχετικών συχνοτήτων για το Ολικό Διοξείδιο του Θείου

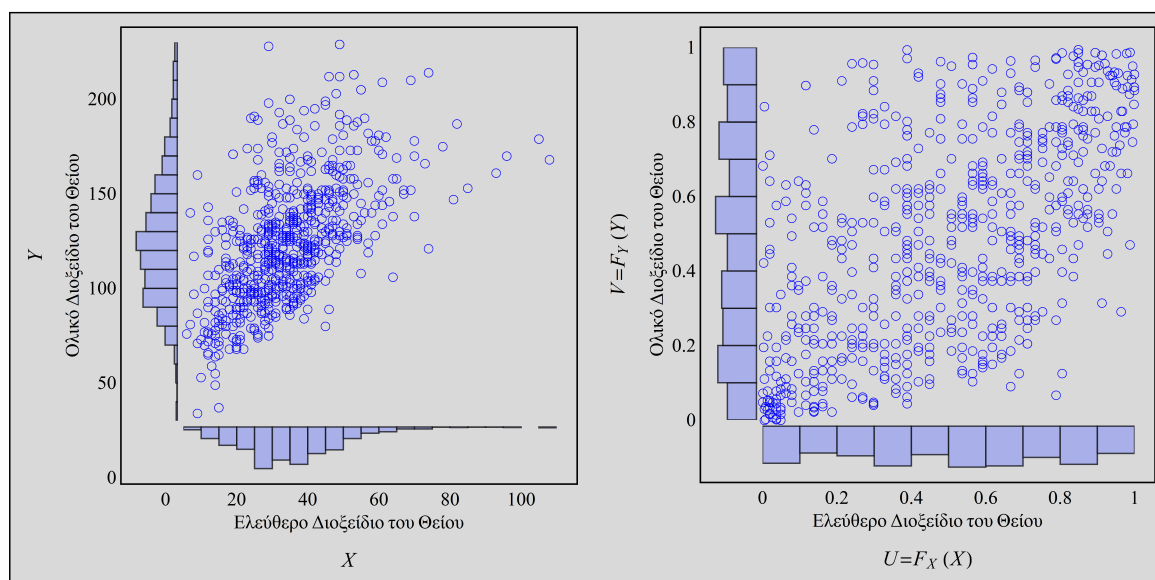
| Έλεγχος | Τιμή σ.σ. | <i>p</i> - value |
|--------------------|-----------|------------------|
| Anderson-Darling | 0.3315 | 0.9127 |
| Cramér-von Mises | 0.0573 | 0.8313 |
| Kolmogorov-Smirnov | 0.0213 | 0.8116 |

Πίνακας 5.5: Έλεγχοι καλής προσαρμογής για την κατανομή του Ολικού Διοξειδίου του Θείου

5.7. Εφαρμογή σε πραγματικά δεδομένα

Συγκεκριμένα, η συνάρτηση πυκνότητας του Ελεύθερου και Ολικού Διοξειδίου του Θείου (έστω χαρακτηριστικά X και Y , αντίστοιχα) είναι Γάμμα με παραμέτρους $\alpha_1 = 5.974$, $\beta_1 = 5.796$ και $\alpha_2 = 13.953$, $\beta_2 = 9.063$, συμβ. $X \sim \Gamma(\alpha_1, \beta_1)$ και $Y \sim \Gamma(\alpha_2, \beta_2)$. Μία πρώτη ένδειξη για τα προαναφερθέντα αποτελέσματα δίνεται μέσω των γραφημάτων που απεικονίζονται στα Σχήματα 5.3 και 5.4, ενώ οι γραφικές ενδείξεις επιβεβαιώνονται και μέσω των (στατιστικών) ελέγχων καλής προσαρμογής που παρουσιάζονται στους Πίνακες 5.4 και 5.5.

Στη συνέχεια, οι τιμές των δύο χαρακτηριστικών μετασχηματίζονται στην κλίμακα $(0, 1)$ των συνδέσμων, μέσω των (μονοδιάστατων) περιθώριων κατανομών Γάμμα κάνοντας χρήση των αντίστοιχων εκτιμήσεων που προέκυψαν από την παραπάνω ανάλυση για τις παραμέτρους α_1 , β_1 και α_2 , β_2 . Στο Σχήμα 5.5, απεικονίζονται τα διαγράμματα διασποράς των αρχικών μεταβλητών X, Y και των μετασχηματισμένων μεταβλητών $U, V \sim (0, 1)$.



Σχήμα 5.5: Διαγράμματα διασποράς των αρχικών και των μετασχηματισμένων τιμών των χαρακτηριστικών του Ελεύθερου και Ολικού Διοξειδίου του Θείου

| Έλεγχος | Τιμή σ.σ. | $p - value$ |
|------------------|-----------|-------------|
| Anderson-Darling | 1.4472 | 0.1760 |
| Cramér-von Mises | 0.2365 | 0.1920 |

Πίνακας 5.6: Έλεγχοι καλής προσαρμογής του συνδέσμου των GUH με παράμετρο $\alpha = 2.5$ στα ζεύγη των παρατηρήσεων Ελεύθερου και Ολικού Διοξειδίου του Θείου

Τέλος, επιλέγεται ο σύνδεσμος που έχει την καλύτερη προσαρμογή στις μετασχηματισμένες πλέον μεταβλητές U, V . Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην εργασία των Genest *et al.* (2009) για μια ανασκόπηση ελέγχων καλής προσαρμογής για συνδέσμους. Από τον Πίνακα 5.6, γίνεται φανερό ότι ο σύνδεσμος των GUH με παράμετρο $\alpha = 2.5$ προσαρμόζεται αρκετά ικανοποιητικά στα δεδομένα μας ($p - value > 10\%$).

Συνεπώς, εάν υποθέσουμε ότι θέλουμε να δημιουργήσουμε ένα διάγραμμα ελέγχου με σκοπό την παρακολούθηση του Ελεύθερου και του Ολικού Διοξειδίου του Θείου, προκαθορίζουμε αρχικά το επίπεδο του ποσοστού λανθασμένου συναγερού, έστω ίσο με $FAR_{OC}^{(2)} = 5\%$. Το επιθυμητό FAR επιτυγχάνεται, μεταξύ άλλων, από τις παραμέτρους

$$a = 9 \quad \text{και} \quad b = 500 - 9 + 1 = 492.$$

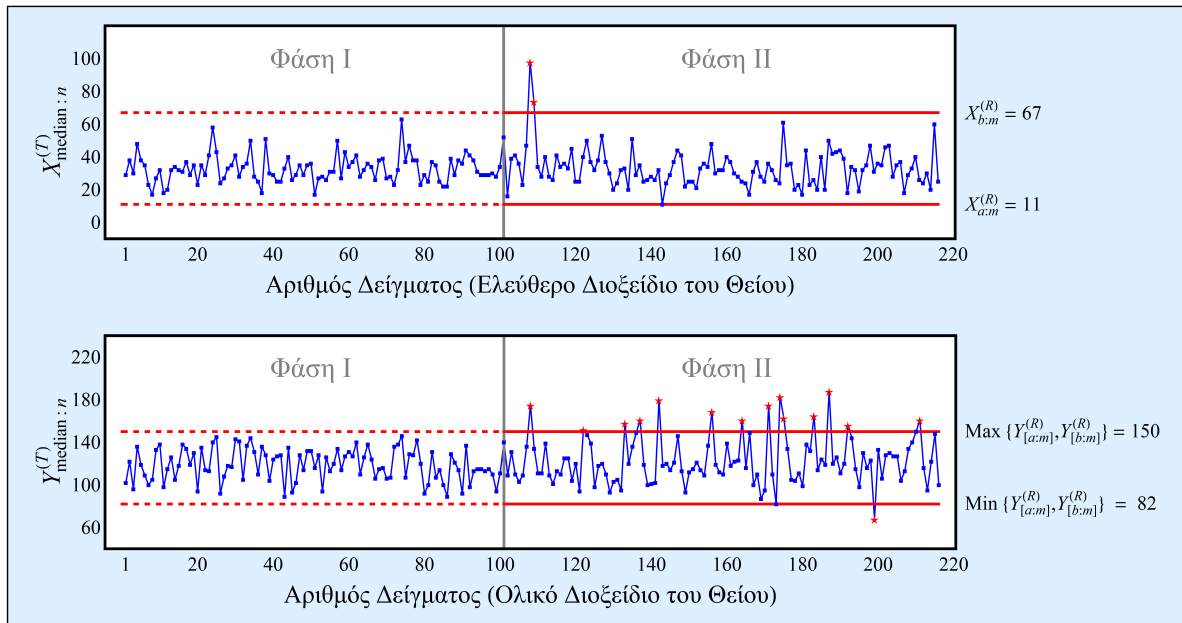
Επομένως, έχοντας ένα δείγμα αναφοράς μεγέθους 500 ($m = 100 \times 5$), τα όρια ελέγχου των δύο χαρακτηριστικών καθορίζονται από την 9η και την 492η παρατήρηση. Κατά συνέπεια, το διάγραμμα ελέγχου θα παράγει εκτός ελέγχου σήμα όταν παραβιάζεται τουλάχιστον μία από τις επόμενες δύο συνθήκες

$$X_{9:500}^{(R)} \leq X_{3:5}^{(T)} \leq X_{492:500}^{(R)} \quad \text{και} \quad \min \{ Y_{[9:500]}^{(R)}, Y_{[492:500]}^{(R)} \} \leq Y_{3:5}^{(T)} \leq \max \{ Y_{[9:500]}^{(R)}, Y_{[492:500]}^{(R)} \}.$$

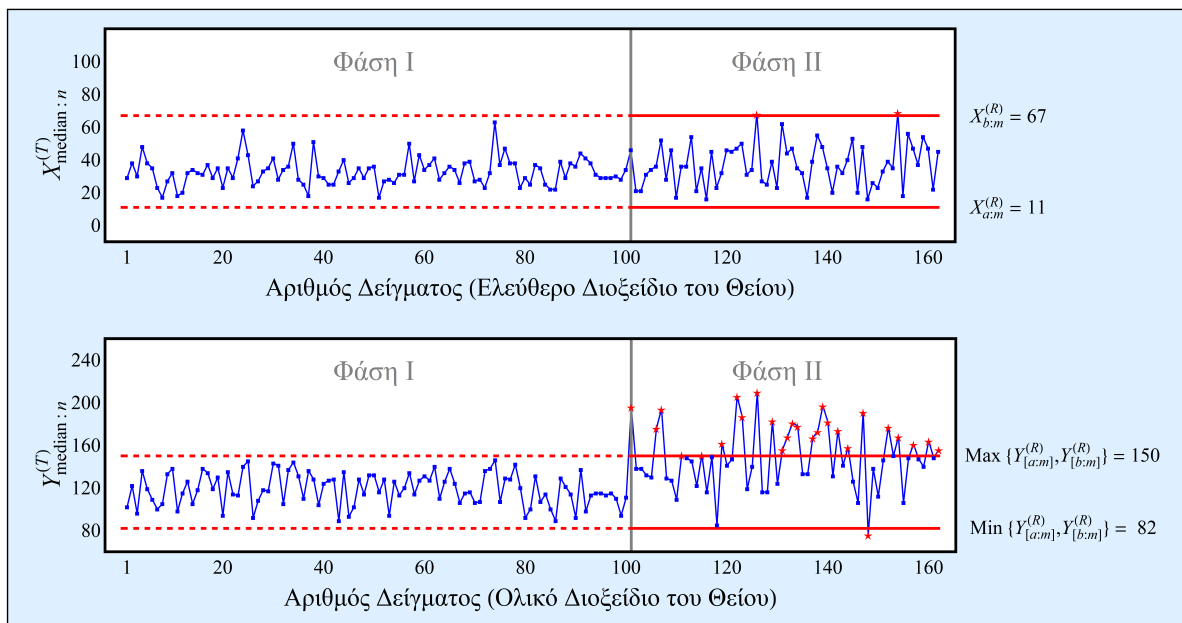
Στο Σχήμα 5.1, που παρουσιάστηκε στην Ενότητα 5.2, απεικονίζεται το διάγραμμα OC για τα δεδομένα $LV6$ με τη βοήθεια ενός τρισδιάστατου γραφήματος. Τα σημεία που δεν εμπίπτουν στην περιοχή ελέγχου επισημαίνονται με κόκκινο χρώμα. Εναλλακτικά, το γράφημα μπορεί να απεικονισθεί εξετάζοντας τις δισδιάστατες προβολές (2D), όπως φαίνεται στο Σχήμα 5.6. Με τον τρόπο αυτό, προσδιορίζονται με σαφήνεια τα σημεία που τίθενται εκτός ελέγχου. Παρόμοια, το Σχήμα 5.7 απεικονίζει το αντίστοιχο ημιπαραμετρικό διάγραμμα για το δείγμα με ποιότητα $LV5$.

Σύμφωνα με το πρώτο σενάριο, είναι προφανές ότι το Ολικό Διοξείδιο του Θείου είναι -κατά κύριο λόγο- το υπαίτιο χαρακτηριστικό για την εκτός ελέγχου κατάσταση της διεργασίας, ενώ μόνο ένα σημείο τίθεται εκτός των ορίων λόγω του Ελεύθερου Διοξειδίου του Θείου. Συγκεκριμένα, το διάγραμμα δημιουργεί σήματα συναγερού στο 109ο δείγμα εξαιτίας μόνο της πρώτης μεταβλητής, στο 122ο, 133ο, 137ο, 142ο, 156ο, 164ο, 171ο, 174ο, 175ο, 183ο, 187ο, 192ο, 199ο και στο 211ο δείγμα εξαιτίας μόνο της δεύτερης μεταβλητής, ενώ στο 108ο δείγμα ενεργοποιείται συναγερός λόγω και των δύο μεταβλητών. Παρά το γεγονός ότι στην τελευταία περίπτωση το σήμα οφείλεται τόσο στο Ελεύθερο όσο και στο Ολικό Διοξείδιο του Θείου, το διάγραμμά μας είναι ικανό να αναγνωρίζει μετατοπίσεις ξεχωριστά για κάθε ένα από τα χαρακτηριστικά αυτά αναφορικά με την ποιότητα των δειγμάτων οίνου.

5.7. Εφαρμογή σε πραγματικά δεδομένα



Σχήμα 5.6: Εφαρμογή του διαγράμματος *OC* στο σύνολο δεδομένων λευκού οίνου της ποικιλίας *Vinho Verde* με μεταβλητές το Ελεύθερο και το Ολικό Διοξείδιο του Θείου (Γράφημα 2D με το πρώτο δείγμα ελέγχου ποιότητας *LV6*)



Σχήμα 5.7: Εφαρμογή του διαγράμματος *OC* στο σύνολο δεδομένων λευκού οίνου της ποικιλίας *Vinho Verde* με μεταβλητές το Ελεύθερο και το Ολικό Διοξείδιο του Θείου (Γράφημα 2D με το δεύτερο δείγμα ελέγχου ποιότητας *LV5*)

Με την εφαρμογή του δεύτερου σεναρίου, παρατηρούμε με τη βοήθεια του Σχήματος 5.7 ότι το διάγραμμα, που αντιστοιχεί στο Ολικό Διοξείδιο του Θείου, σηματοδοτείται ταχύτερα και σε περισσότερα δείγματα (συγκεκριμένα ενεργοποιείται συναγερός στο 101ο, 106ο, 107ο, 111ο, 115ο, 119ο, 122ο, 123ο, 129ο, 131ο, 132ο, 133ο, 134ο, 137ο, 138ο, 139ο, 140ο, 142ο, 144ο, 147ο, 148ο, 152ο, 157ο, 160ο και στο 162ο λόγω του δεύτερο χαρακτηριστικού, ενώ στο 126ο και στο 154ο λόγω μετατοπίσεων και των δύο χαρακτηριστικών). Παρότι τα δείγματα Φάσης II είναι λιγότερα από αυτά του προηγούμενου σεναρίου, δεν επηρεάζεται η ευαισθησία του κανόνα. Αυτό θα μπορούσε να αποδοθεί στο ότι όσο χαμηλότερο είναι το επίπεδο ποιότητας των δειγμάτων ελέγχου, τόσο πιο γρήγορα το γράφημα ανιχνεύει τη μετατόπιση.

Θα πρέπει να επισημανθεί ότι σύμφωνα με το μοντέλο εξόρυξης δεδομένων που προτάθηκε από τους Cortez *et al.* (2009), οι μεταβλητές του Ελεύθερου και Ολικού Διοξειδίου του Θείου έχουν μικρότερη προβλεπτική ικανότητα (σε σχέση με άλλες μεταβλητές που είναι διαθέσιμες στο αρχικό σύνολο δεδομένων) όσον αφορά στην ανίχνευση της ποιότητας του οίνου. Ωστόσο, η πειραματική μελέτη που προηγήθηκε, αποκάλυψε ότι οι συγκεκριμένες μεταβλητές ενεργοποίησαν έγκαιρα σήμα για εκτός ελέγχου κατάσταση της διεργασίας. Τέλος, το γεγονός ότι σε κάποιες περιπτώσεις ενδέχεται μόνο μία μεταβλητή να υπερβαίνει τα όρια ελέγχου μπορεί να ερμηνευθεί ως δείκτης της τάσης της συγκεκριμένης μεταβλητής να οδηγεί σε υψηλότερη ποιότητα οίνου.

ΚΕΦΑΛΑΙΟ 6

Δισδιάστατο Διάγραμμα Ελέγχου για τη Θέση και τη Διασπορά με Χρήση Διατεταγμένων Παρατηρήσεων

6.1 Εισαγωγή

Όπως αναφέρθηκε και στο Κεφάλαιο 1, τα περισσότερα διαγράμματα που έχουν προταθεί στη βιβλιογραφία είναι κατάλληλα για την παρακολούθηση μόνο της μέσης ή μόνο της μεταβλητότητας μιας διεργασίας με την πλειοψηφία αυτών να εστιάζεται σε παραμετρικά σχήματα. Για παράδειγμα, το διάγραμμα $|S|$ Φάσης II που πρότεινε ο Alt (1985), για την παρακολούθηση δυο ποιοτικών χαρακτηριστικών, θεωρείται ανάλογο του κλασσικού (παραμετρικού) διαγράμματος χ^2 , με τη διαφορά ότι είναι ευαίσθητο στην ανίχνευση πιθανών μετατοπίσεων της διασποράς της υποκείμενης διεργασίας και όχι της μέσης τιμής. Η σ.σ. που απεικονίζεται στο εν λόγω διάγραμμα είναι η γενικευμένη δειγματική διακύμανση $|S|$, όπου S είναι ο δειγματικός πίνακας διακύμανσης-συνδιακύμανσης, και τα όρια πιθανότητας του διαγράμματος υπολογίζονται ως εξής

$$LCL = \frac{|\Sigma|}{4(n-1)^2} (\chi_{2n-4; 1-\alpha/2}^2)^2 \quad \text{και} \quad UCL = \frac{|\Sigma|}{4(n-1)^2} (\chi_{2n-4; \alpha/2}^2)^2.$$

Αξίζει να σημειωθεί ότι στη μονόπλευρη εκδοχή του, υπάρχει μόνο άνω όριο, το οποίο λαμβάνει τη μορφή

$$UCL = \frac{|\Sigma|}{4(n-1)^2} (\chi_{2n-4; \alpha}^2)^2.$$

Στη βιβλιογραφία έχουν, επίσης, προταθεί αρκετές παραλλαγές του προαναφερθέντος διαγράμματος για δύο ή περισσότερα ποιοτικά χαρακτηριστικά. Ενδεικτικά αναφέρονται οι εργασίες των Alt and Smith (1998), Aparisi *et al.* (1999, 2001), Khoo and Quah (2004),

Riaz and Does (2008) και Costa and Neto (2017).

Επιπλέον, έχουν μελετηθεί και πολυδιάστατα παραμετρικά σχήματα, η κατασκευή των οποίων βασίζεται σε μία σχετικά απλή λογική. Πιο συγκεκριμένα, οι Costa and Machado (2009) πρότειναν το διάγραμμα $Rmax$, το οποίο κάνει χρήση του μέγιστου δειγματικού εύρους που υπολογίζεται για κάθε μία από τις προς μελέτη μεταβλητές. Στην περίπτωση δύο ποιοτικών χαρακτηριστικών η σ.σ. γράφεται ως εξής

$$RMAX = \max\{R_X, R_Y\},$$

όπου $R_X = \max\{X_1, \dots, X_n\} - \min\{X_1, \dots, X_n\}$ και $R_Y = \max\{Y_1, \dots, Y_n\} - \min\{Y_1, \dots, Y_n\}$. Η διεργασία τίθεται εκτός ελέγχου εάν $RMAX > UCL$. Το όριο ελέγχου εκτιμάται, ώστε να επιτευχθεί το επιθυμητό ARL_{in} .

Πρόσφατα, οι Osei-Aning *et al.* (2017), παρακινούμενοι από την παραπάνω ιδέα, προτείνουν τέσσερα νέα δισδιάστατα μη παραμετρικά διαγράμματα ελέγχου που βασίζονται στις δειγματικές εκτιμήσεις της τυπικής απόκλισης, του ενδοτεταρτημοριακού εύρους, της απόλυτης διαφοράς από τη δειγματική διάμεσο και της απόλυτης απόκλισης από τη διάμεσο (*median absolute deviation*). Οι εκτιμήσεις αυτές ορίζονται αντίστοιχα από τις ακόλουθες σχέσεις

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

$$Q = (Q_3 - Q_1)/1.34898,$$

$$MD = \frac{1}{n} \sum_{i=1}^n |X_i - \delta|,$$

$$MAD = 1.4826 \text{ med } |X_i - \delta|,$$

όπου δ είναι η δειγματική διάμεσος και η ποσότητα Q έχει αναπροσαρμοσθεί με βάση την εκτίμηση που χρησιμοποιείται συνήθως από στατιστικά πακέτα, όπως το *SPSS* και το *Minitab* (βλ. Abbasi *et al.* (2015)). Λαμβάνοντας υπ' όψιν τα παραπάνω αποτελέσματα, οι σ.σ. των μη παραμετρικών διαγραμμάτων υπολογίζονται αντίστοιχα από τους τύπους

$$S MAX = \max\{S_X, S_Y\},$$

$$Q MAX = \max\{Q_X, Q_Y\},$$

$$M D MAX = \max\{M D_X, M D_Y\},$$

$$M A D MAX = \max\{M A D_X, M A D_Y\}.$$

Σε κάθε μία από αυτές τις περιπτώσεις παράγεται εκτός ελέγχου σήμα, όταν η εκάστοτε σ.σ. υπερβεί το όριο ελέγχου, το οποίο επιλέγεται κατάλληλα με σκοπό την επίτευξη του

επιθυμητού ARL_{in} .

Αξίζει να τονιστεί ότι τα τελευταία χρόνια παρατηρείται ερευνητική δραστηριότητα στην κατασκευή κανόνων για ταυτόχρονη παρακολούθηση της μέσης τιμής και της διασποράς (ή εναλλακτικά του πίνακα διακυμάνσεων-συνδιακυμάνσεων) μιας διεργασίας, η οποία όμως είναι σχετικά περιορισμένη όσον αφορά τα παραμετρικά σχήματα και σχεδόν ανύπαρκτη για μη παραμετρικά σχήματα.

Για παράδειγμα, ο Khoo (2004) πρότεινε ένα παραμετρικό δισδιάστατο (μονόπλευρο) διάγραμμα ελέγχου για ταυτόχρονες μετατοπίσεις, το οποίο στο εξής θα ονομάζουμε διάγραμμα $BV MAX$. Ειδικότερα, υποθέτοντας ότι X_{ij} , $i = 1, 2, \dots$ και $j = 1, 2, \dots, n$ είναι πολυδιάστατες παρατηρήσεις που προέρχονται από δισδιάστατη ($p = 2$) κανονική κατανομή, συμβ. $X_{ij} \sim N_p(\mu_0, \Sigma_0)$, και λαμβάνοντας υπ' όψιν ότι

$$T_i^2 \sim \chi_2^2 \quad \text{και} \quad 2(n-1) |\mathbf{S}_i|^{1/2} / |\Sigma_0|^{1/2} \sim \chi_{2n-4}^2,$$

όπου

$$T_i^2 = n(\bar{X}_i - \mu_0)' |\Sigma_0|^{-1} (\bar{X}_i - \mu_0) \quad \text{και} \quad \mathbf{S}_i = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)', \quad i = 1, 2, \dots,$$

προκύπτουν αντίστοιχα τα εξής

$$\begin{aligned} M_i &= \Phi^{-1} [H_2(T_i^2)] \sim N(0, 1), & i = 1, 2, \dots \\ V_i &= \Phi^{-1} [H_{2n-4} \{2(n-1) |\mathbf{S}_i|^{1/2} / |\Sigma_0|^{1/2}\}] \sim N(0, 1), & i = 1, 2, \dots \end{aligned}$$

με τα $\Phi(\cdot)$, $H_2(\cdot)$, $H_{2n-4}(\cdot)$ να χρησιμοποιούνται για να δηλώσουν τις αθροιστικές κατανομές της τυπικής κανονικής και της κατανομής χ^2 με 2 και $2n - 4$ β.ε., αντίστοιχα.

Αξιοποιώντας τα παραπάνω, η σ.σ. του διαγράμματος $BV MAX$ ορίζεται ως εξής

$$C_i = \max\{|M_i|, |V_i|\},$$

και λαμβάνει μεγάλες τιμές όταν παρατηρούνται μεγάλες μετατοπίσεις της μέσης τιμής ή/και του πίνακα διακυμάνσεων-συνδιακυμάνσεων της διεργασίας. Το όριο ελέγχου του διαγράμματος καθορίζεται έτσι, ώστε να επιτευχθεί ένα συγκεκριμένο προκαθορισμένο επίπεδο για το ποσοστό λανθασμένου συναγεμμού, δηλαδή

$$\int_{UCL}^{\infty} 4\phi(c)[2\Phi(c) - 1]dc = FAR, \quad c \geq 0,$$

(όπου τα $\Phi(\cdot)$, $\phi(\cdot)$ δηλώνουν αντίστοιχα την αθροιστική κατανομή και την πυκνότητα της τυπικής κανονικής κατανομής) και υπολογίζεται άμεσα από τον τύπο που ακολουθεί

$$UCL = \int_{UCL}^{\infty} \sqrt{\frac{8}{\pi}} \times (e^{-c^2/2}) \times \left(\sqrt{\frac{2}{\pi}} \int_{-\infty}^c e^{-t^2/2} dt - 1 \right) dc, \quad c \geq 0.$$

Εάν υποθέσουμε ότι η διεργασία είναι εκτός ελέγχου ($C_i > UCL$), εξετάζουμε τις τιμές των $|M_i|, |V_i|$ σε σχέση με το όριο ελέγχου, ώστε να αναγνωρίσουμε το είδος της μετατόπισης. Στην περίπτωση που ισχύει $|M_i| > UCL$, η διεργασία έχει τεθεί εκτός ελέγχου λόγω μετατόπισης της μέσης τιμής, στην περίπτωση που έχουμε $|V_i| > UCL$ είναι ένδειξη μετατόπισης του πίνακα διακυμάνσεων-συνδιακυμάνσεων, ενώ όταν και η σ.σ. $|M_i|$ και η σ.σ. $|V_i|$, υπερβαίνουν το όριο ελέγχου, τότε ενεργοποιείται συναγερμός εξαιτίας ταυτόχρονων μετατοπίσεων.

Ιδιαίτερη έμφαση πρέπει να δοθεί στο γεγονός ότι έχουν μελετηθεί μερικά μη παραμετρικά διαγράμματα για την ανίχνευση ταυτόχρονων μετατοπίσεων, τα οποία είναι κυρίως μονόπλευρα διαγράμματα ελέγχου με μνήμη. Ένα από τα τρία σχήματα που πρότεινε στην εργασία της η Liu (1995) είναι ένα διάγραμμα τύπου *Shewhart* για ομαδοποιημένες παρατηρήσεις. Για την ακρίβεια, πρόκειται για ένα πολυμεταβλητό μη παραμετρικό (μονόπλευρο) διάγραμμα, το οποίο βασίζεται στο βάθος δεδομένων και ονομάζεται διάγραμμα *Q*.

Ας υποθέσουμε, αρχικά, ότι έχουμε στη διάθεση μας ένα δείγμα αναφοράς, έστω X_1, X_2, \dots, X_m , το οποίο προέρχεται από μία m -διάστατη κατανομή F και η εμπειρική του κατανομή συμβολίζεται με $\hat{F}_m(\cdot)$. Στη συνέχεια, συλλέγουμε ένα δείγμα ελέγχου $Y_1, Y_2, \dots, Y_n \sim G$, με το οποίο συγκρίνουμε το δείγμα αναφοράς, ώστε να αποφανθούμε για το εάν η διεργασία βρίσκεται εντός ή εκτός ελέγχου. Χωρίς βλάβη της γενικότητας, το βάθος των δεδομένων μπορεί να βασιστεί στην απόσταση του *Mahalanobis*, η οποία μετρά το βάθος του σημείου x σε σχέση με το πόσο μικρή είναι η τετραγωνική απόστασή του από τον μέσο, δεδομένης της κατανομής F . Η απόσταση αυτή συμβολίζεται με $\mathbf{MD}_F(x)$ και υπολογίζεται μέσω του τύπου που ακολουθεί

$$\mathbf{MD}_F(x) = \frac{1}{1 + (x - \mu_F)' \Sigma_F^{-1} (x - \mu_F)},$$

όπου μ_F και Σ_F είναι η μέση τιμή και ο πίνακας διακυμάνσεων-συνδιακυμάνσεων της κατανομής F , αντίστοιχα. Η εμπειρική μορφή της ποσότητας $\mathbf{MD}_F(x)$ δίνεται από τον τύπο

$$\mathbf{MD}_{F_m}(x) = \frac{1}{1 + (x - \bar{X})' \mathbf{S}^{-1} (x - \bar{X})},$$

όπου \bar{X} είναι ο δειγματικός μέσος των X_1, X_2, \dots, X_m και \mathbf{S}_F είναι ο δειγματικός πίνακας διακυμάνσεων-συνδιακυμάνσεων.

Διατάσσουμε, αρχικά, τα βάρη των X_1, X_2, \dots, X_m σε αύξουσα σειρά και με $X_{[1]}, X_{[2]}, \dots, X_{[m]}$ συμβολίζουμε τη διάταξη αυτή. Το $X_{[m]}$ είναι το πιο κεντρικό σημείο. Όσο μικρότερος ο βαθμός του σημείου, τόσο πιο απομακρυσμένο είναι το σημείο αυτό ως προς την κατανομή F . Στη συνέχεια, παρατίθενται οι σ.σ.

$$r_F(x) = P(D_F(X) \leq D_F(x) | X \sim F),$$

$$r_{\hat{F}_m}(x) = \frac{\#\{X_j | D_{\hat{F}_m}(X_j) \leq D_{\hat{F}_m}(x), j = 1, 2, \dots, m\}}{m},$$

με τις οποίες υπολογίζουμε αντίστοιχα τα $\{r_F(X_1), r_F(X_2), \dots\}$ ή τα $\{r_{\hat{F}_m}(X_1), r_{\hat{F}_m}(X_2), \dots\}$ εάν είναι διαθέσιμα μόνο τα X_1, X_2, \dots, X_m και όχι η κατανομή F .

Συμβολίζοντας με $\hat{G}_n(\cdot)$ την εμπειρική κατανομή του δείγματος $Y_1, Y_2, \dots, Y_n \sim G$, ορίζουμε τις ποσότητες

$$Q(F, G) = P\{D_F(X) \leq D_F(Y) | X \sim F, Y \sim G\},$$

$$Q(F, \hat{G}_n) = \frac{1}{n} \sum_{i=1}^n r_F(Y_i) \quad \text{και} \quad Q(\hat{F}_m, \hat{G}_n) = \frac{1}{n} \sum_{i=1}^n r_{\hat{F}_m}(Y_i).$$

Για την κατασκευή του διαγράμματος Q , η Liu (1995) προτείνει την απεικόνιση των μέσων των υποσυνόλων των $\{r_F(Y_1), r_F(Y_2), \dots\}$ ή των $\{r_{\hat{F}_m}(Y_1), r_{\hat{F}_m}(Y_2), \dots\}$. Εάν θεωρήσουμε ότι κάθε υποσύνολο έχει μέγεθος n , οι μέσοι των $r_F(Y_i)$ ή των $r_{\hat{F}_m}(Y_i)$ υπολογίζονται από τους τύπους $Q(F, \hat{G}_n^j)$ και $Q(\hat{F}_m, \hat{G}_n^j)$, όπου \hat{G}_n^j είναι η εμπειρική κατανομή των Y_i του j υποσυνόλου για $j = 1, 2, \dots$. Εάν είναι διαθέσιμα μόνο τα X_1, X_2, \dots, X_m , στο διάγραμμα Q μπορούν να απεικονιστούν $\{r_F(X_1), r_F(X_2), \dots\}$ και τα $\{r_{\hat{F}_m}(X_1), r_{\hat{F}_m}(X_2), \dots\}$. Τα όρια ελέγχου του διαγράμματος Q που αντιστοιχούν στις σ.σ. $Q(F, \hat{G}_n^j)$ και $Q(\hat{F}_m, \hat{G}_n^j)$ υπολογίζονται ως εξής

$$LCL = \frac{1}{2} - z_a \sqrt{\frac{1}{12n}} \quad \text{και} \quad LCL = \frac{1}{2} - z_a \sqrt{\frac{1}{12} \left(\frac{1}{m} + \frac{1}{n} \right)},$$

και προφανώς εξαρτώνται από την επιλογή των n ή/και m . Σε κάθε περίπτωση, υπάρχει επίσης μία κεντρική γραμμή $CL = 1/2$ και τα όρια είναι εφαρμόσιμα μόνο όταν το n λαμβάνει μεγάλες τιμές (π.χ. $n \geq 5$). Στην περίπτωση που το n παίρνει μικρές τιμές και ταυτόχρονα ικανοποιείται η σχέση $a \leq 1/n!$, τα όρια ελέγχου που προκύπτουν είναι

$$CL = 1/2 \quad \text{και} \quad LCL = (n!a)^{1/n}/n.$$

Το a που εμφανίζεται στους παραπάνω τύπους είναι το ποσοστό λανθασμένου συναγερμού. Επειδή το διάγραμμα Q βασίζεται στο βάθος δεδομένων, απαιτεί μεγάλη υπολογιστική ισχύ για να εκτιμήσει το ARL , ειδικά για μεγάλα μεγέθη δείγματος. Οι Stoumbos *et al.* (2001)

μελέτησαν τις απαιτήσεις του δείγματος αναφοράς για τη δημιουργία διαγραμμάτων που βασίζονται στο βάθος των δεδομένων με σκοπό να επιτευχθεί μια λογική τιμή ARL_{in} . Όπως αναφέρεται στην εργασία τους, θα πρέπει να είναι κανείς πολύ προσεκτικός σχετικά με τη σωστή επιλογή των ορίων ελέγχου σε συνδυασμό πάντα με το μέγεθος δείγματος.

Στα δύο προηγούμενα κεφάλαια, παρουσιάστηκαν δύο νέα ημιπαραμετρικά διαγράμματα ελέγχου ($O2, OC$) για την ανίχνευση πιθανών μετατοπίσεων της μέσης τιμής μιας διεργασίας. Τα σχήματα αυτά είχαν κοινή σ.σ., το ζευγάρι των διατεταγμένων παρατηρήσεων $(X_{r:n}^{(T)}, Y_{s:n}^{(T)})$ του δείγματος ελέγχου. Ωστόσο, υπήρχε διαφοροποίηση στη διαδικασία επιλογής των ορίων ελέγχου, όπου γινόταν χρήση διατεταγμένων παρατηρήσεων ή/και συμμεταβλητών του δείγματος αναφοράς. Η διαφοροποίηση των ορίων, είχε αντίκτυπο στην ευαισθησία των διαγραμμάτων ως προς τη δομή εξάρτησης των υπό μελέτη χαρακτηριστικών.

Σκοπός αυτού του κεφαλαίου είναι η δημιουργία ενός σχήματος αποτελεσματικού στην ανίχνευση πιθανών μετατοπίσεων, όχι μόνο στη μέση τιμή της διεργασίας, αλλά και στη διακύμανση. Αξιοποιώντας τη βασική ιδέα του πλήθους παρατηρήσεων που παρουσιάζεται στο (μονοδιάστατο) μη παραμετρικό διάγραμμα των Balakrishnan *et al.* (2010), τροποποιούμε τον κανόνα απόφασης του διαγράμματος $O2$ (βλ. Κεφάλαιο 4), ώστε να ενσωματώσουμε ένα μέτρο διασποράς λαμβάνοντας υπ' όψιν την πληροφορία που σχετίζεται με το πλήθος διαδοχικών διατεταγμένων παρατηρήσεων.

6.2 Περιγραφή ενός νέου διαγράμματος ελέγχου

Χρησιμοποιώντας στατιστική ορολογία, το νέο διάγραμμα (Koutras and Sofikitou (2017c)) θα ελέγχει εάν ισχύει ή όχι η παρακάτω μηδενική υπόθεση

$$H_0 : \mu = \mu_0 \quad \text{και} \quad \Sigma = \Sigma_0, \quad (6.1)$$

όπου μ_0 και Σ_0 είναι οι εντός ελέγχου τιμές της μέσης τιμής και του πίνακα διακυμάνσεων-συνδιακυμάνσεων της υποκείμενης διεργασίας, αντίστοιχα. Η λήψη απόφασης σχετικά με το κατά πόσο η διεργασία βρίσκεται εντός ή εκτός ελέγχου βασίζεται σε μία συνδυασμένη σ.σ. από την οποία λαμβάνουμε πληροφορία για το εάν η μέση τιμή και η διακύμανση ανήκουν ταυτόχρονα στην κρίσιμη περιοχή ελέγχου.

Η γενική ιδέα για την κατασκευή του κανόνα απόφασης είναι αρκετά απλή. Συγκεκριμένα, η διεργασία θεωρείται ότι παραμένει εντός ελέγχου όσο ένα πλήθος διατεταγμένων παρατηρήσεων (ξεχωριστά για κάθε ένα από τα υπό μελέτη χαρακτηριστικά) βρίσκεται

μεταξύ των ορίων ελέγχου.

Η κατασκευή του νέου διαγράμματος απαιτεί την εφαρμογή της ίδιας διαδικασίας με αυτής που περιγράφηκε στο Κεφάλαιο 4. Αρχικά, με $(X_1^{(R)}, Y_1^{(R)}), (X_2^{(R)}, Y_2^{(R)}), \dots, (X_m^{(R)}, Y_m^{(R)})$ συμβολίζουμε το δείγμα αναφοράς, το οποίο προέρχεται από μία εντός ελέγχου κατανομή $F_{X,Y}^{(R)}(x, y) = F(x, y)$ με περιθώριες $F_X(x)$ και $F_Y(y)$. Η περιοχή ελέγχου ορίζεται βάσει του δείγματος αυτού, χρησιμοποιώντας ως όρια κατάλληλα επιλεγμένες διατεταγμένες παρατηρήσεις της μορφής

$$LCL_X = X_{a:m}^{(R)}, \quad UCL_X = X_{b:m}^{(R)} \quad \text{και} \quad LCL_Y = Y_{c:m}^{(R)}, \quad UCL_Y = Y_{d:m}^{(R)},$$

όπου $1 \leq a < b \leq m$ και $1 \leq c < d \leq m$.

Στη συνέχεια, θεωρώντας ότι έχουμε στη διάθεσή μας ένα δείγμα ελέγχου $(X_1^{(T)}, Y_1^{(T)}), (X_2^{(T)}, Y_2^{(T)}), \dots, (X_n^{(T)}, Y_n^{(T)})$ με από κοινού α.σ.κ. $F_{X,Y}^{(T)}(x, y) = G(x, y)$ και περιθώριες $G_X(x)$, $G_Y(y)$, εξετάζουμε εάν ικανοποιούνται οι παρακάτω συνθήκες

$$LCL_X \leq X_{r_1:n}^{(T)} \leq X_{r_2:n}^{(T)} \leq UCL_X \quad \text{και} \quad LCL_Y \leq Y_{s_1:n}^{(T)} \leq Y_{s_2:n}^{(T)} \leq UCL_Y \quad (6.2)$$

ή ισοδύναμα

$$\left\{ \begin{array}{l} X_{a:m}^{(R)} \leq X_{r_1:n}^{(T)} \leq X_{r_2:n}^{(T)} \leq X_{b:m}^{(R)} \\ Y_{c:m}^{(R)} \leq Y_{s_1:n}^{(T)} \leq Y_{s_2:n}^{(T)} \leq Y_{d:m}^{(R)} \end{array} \right\} \quad (6.2\alpha)$$

$$\left\{ \begin{array}{l} X_{a:m}^{(R)} \leq X_{r_1:n}^{(T)} \leq X_{r_2:n}^{(T)} \leq X_{b:m}^{(R)} \\ Y_{c:m}^{(R)} \leq Y_{s_1:n}^{(T)} \leq Y_{s_2:n}^{(T)} \leq Y_{d:m}^{(R)} \end{array} \right\} \quad (6.2\beta)$$

όπου $1 \leq r_1 \leq r_2 \leq n$ και $1 \leq s_1 \leq s_2 \leq n$. Για να το θέσουμε διαφορετικά, η διαδικασία κρίνεται εντός ελέγχου εάν διαδοχικές διατεταγμένες παρατηρήσεις του δείγματος ελέγχου (δηλαδή εκείνες που ανήκουν στα διαστήματα $I_X = [X_{r_1:n}^{(T)}, X_{r_2:n}^{(T)}]$ και $I_Y = [Y_{s_1:n}^{(T)}, Y_{s_2:n}^{(T)}]$) βρίσκονται μεταξύ των ορίων ελέγχου του δείγματος αναφοράς. Προφανώς, η επιλογή των παραμέτρων (σχεδιασμού) $m, n, a, b, c, d, r_1, r_2, s_1, s_2$ γίνεται σύμφωνα με τις ανάγκες ή απαιτήσεις του χρήστη. Προς διευκόλυνση του σχεδιασμού, στην Ενότητα 6.4 αναλύεται η μεθοδολογία σχετικά με τον προσδιορισμό των παραμέτρων αυτών.

Δεδομένου ότι ο κανόνας χρησιμοποιεί τέσσερις διατεταγμένες παρατηρήσεις από το δείγμα ελέγχου, θα ονομάζεται στο εξής διάγραμμα *O4*. Αξίζει να αναφερθεί ότι στην περίπτωση που $r_1 = r_2$ και $s_1 = s_2$, ο Κανόνας 6.2 συμπίπτει με τον Κανόνα 4.1, δηλαδή το διάγραμμα *O4* ταυτίζεται με το διάγραμμα *O2*.

Είναι σημαντικό να τονισθεί ότι το νέο διάγραμμα ελέγχου έχει τη δυνατότητα να προσδιορίσει ποιο από τα χαρακτηριστικά είναι υπαίτιο για την ενεργοποίηση συναγερμού. Ειδικότερα, εάν παραβιάζεται μόνο η Συνθήκη 6.2α (6.2β), τότε συνεπάγεται ότι έχει μετατοπισθεί η μέση τιμή και η διασπορά του πρώτου (δεύτερου) χαρακτηριστικού, ενώ

εάν παραβιάζονται και οι δύο συνθήκες ταυτόχρονα, μπορούμε να συμπεράνουμε ότι η μετατόπιση έχει προκληθεί εξαιτίας και των δύο χαρακτηριστικών.

Στο σημείο αυτό, θα πρέπει να σημειωθεί ότι ο Κανόνας 6.2 δεν μπορεί (από την κατασκευή του) να προσδιορίσει εάν η μετατόπιση της διεργασίας προήλθε λόγω της μέσης τιμής ή της διασποράς. Ωστόσο, είναι ξεκάθαρο ότι στην περίπτωση που τα διαστήματα I_X και I_Y είναι στενά (*narrow*), ο κανόνας θα είναι πιο ευαίσθητος σε μετατοπίσεις της μέσης τιμής, ενώ στην αντίθετη περίπτωση το διάγραμμα ελέγχου θα είναι πιο ευαίσθητο στην ανίχνευση μετατοπίσεων της διασποράς.

6.3 Χαρακτηριστικά διαγράμματος ελέγχου

Στην ενότητα αυτή, θα αποδείξουμε τύπους για τα βασικά χαρακτηριστικά του νέου διαγράμματος ελέγχου. Για το σκοπό αυτό, χρειάζεται να καταφύγουμε στη Θεωρία των Συνδέσμων και να αξιοποιήσουμε αποτελέσματα από το Κεφάλαιο 3. Αρχικά, θεωρούμε ότι $C(u, v)$ και $D(u, v)$ είναι οι σύνδεσμοι που συνδέονται με τις εντός και εκτός ελέγχου από κοινού α.σ.κ. $F(x, y)$ και $G(x, y)$, αντίστοιχα. Οι συναρτήσεις αυτές μπορούν να εκφραστούν μέσω των αντίστοιχων συνδέσμων με τη βοήθεια των περιθώριων συναρτήσεων $G_X(x)$, $G_Y(y)$ και $F_X(x)$, $F_Y(y)$ ως εξής

$$F(x, y) = C(F_X(x), F_Y(y)) \quad \text{και} \quad G(x, y) = D(G_X(x), G_Y(y)).$$

Εάν θεωρήσουμε ότι οι αρχικές παρατηρήσεις προέρχονται από συνεχείς κατανομές, τότε η πιθανότητα το διάγραμμα $O4$ να μην ενεργοποιεί συναγερμό ορίζεται ως εξής

$$p_{O4} = P(X_{a:m}^{(R)} \leq X_{r_1:n}^{(T)} \leq X_{r_2:n}^{(T)} \leq X_{b:m}^{(R)} \quad \text{και} \quad Y_{c:m}^{(R)} \leq Y_{s_1:n}^{(T)} \leq Y_{s_2:n}^{(T)} \leq Y_{d:m}^{(R)}), \quad (6.3)$$

η οποία δηλώνει, στην πραγματικότητα, τη χαρακτηριστική συνάρτηση λειτουργίας, συμβ. OCF_{O4} , ενώ η ποσότητα $AR_{O4} = 1 - p_{O4}$ δηλώνει την πιθανότητα να παραχθεί σήμα για εκτός ελέγχου μετατοπίσεις. Στη συνέχεια αποδεικνύουμε μία έκφραση για την πιθανότητα p_{O4} , όπως αυτή ορίζεται στη Σχέση (6.3).

Πρόταση 6.3. Η χαρακτηριστική συνάρτηση λειτουργίας του διαγράμματος $O4$ δίνεται από την έκφραση

$$\begin{aligned} p_{O4} &= p_{O4}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b, c, d; r_1, r_2, s_1, s_2) \\ &= \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} Q_{O4}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))) \\ &\quad \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2, \end{aligned}$$

6.3. Χαρακτηριστικά διαγράμματος ελέγχου

όπου $Q_{04}^{(D)}(\cdot)$ είναι η Πιθανότητα (3.7) του Λήμματος 3.5 και με $f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ συμβολίζεται η από κοινού σ.π.π. των $U_{a:m}^{(R)}, U_{b:m}^{(R)}, V_{c:m}^{(R)}, V_{d:m}^{(R)}$, η οποία είναι απλά η συνάρτηση πυκνότητας που αντιστοιχεί στην α.σ.κ. $F_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ της Σχέσης (3.4) και προκύπτει αντικαθιστώντας τις ποσότητες n, r_1, r_2, s_1, s_2, D με m, a, b, c, d, C .

Απόδειξη. Αρχικά, γράφουμε τη Σχέση (6.3) εναλλακτικά ως εξής

$$p_{04} = P(U_{a:m}^{(R)} \leq F_X(X_{r_1:n}^{(T)}) \leq F_X(X_{r_2:n}^{(T)}) \leq U_{b:m}^{(R)} \text{ και } V_{c:m}^{(R)} \leq F_Y(Y_{s_1:n}^{(T)}) \leq F_Y(Y_{s_2:n}^{(T)}) \leq V_{d:m}^{(R)})$$

και εφαρμόζουμε την αρχή δέσμευσης-αποδέσμευσης στις τ.μ.

$$U_{a:m}^{(R)} = F_X(X_{a:m}^{(R)}), \quad U_{b:m}^{(R)} = F_X(X_{b:m}^{(R)}) \quad \text{και} \quad V_{c:m}^{(R)} = F_Y(Y_{c:m}^{(R)}), \quad V_{d:m}^{(R)} = F_Y(Y_{d:m}^{(R)})$$

με σκοπό να λάβουμε την επόμενη ολοκληρωτική μορφή

$$p_{04} = \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} P[u_1 \leq F_X(X_{r_1:n}^{(T)}) \leq F_X(X_{r_2:n}^{(T)}) \leq u_2 \text{ και } v_1 \leq F_Y(Y_{s_1:n}^{(T)}) \leq F_Y(Y_{s_2:n}^{(T)}) \leq v_2] \\ \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2,$$

όπου $f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. που δίνεται στον Τύπο (3.4) εάν τα r_1, r_2, s_1, s_2, n και C χρησιμοποιηθούν αντί των ποσοτήτων a, b, c, d, m και D , αντίστοιχα.

Στη συνέχεια, παρατηρώντας ότι η πιθανότητα που εμφανίζεται στο τετραπλό ολοκλήρωμα μπορεί να γραφεί στην εξής ισοδύναμη μορφή

$$P[F_X^{-1}(u_1) \leq X_{r_1:n}^{(T)} \leq X_{r_2:n}^{(T)} \leq F_X^{-1}(u_2), F_Y^{-1}(v_1) \leq Y_{s_1:n}^{(T)} \leq Y_{s_2:n}^{(T)} \leq F_Y^{-1}(v_2)] \\ = P[G_X(F_X^{-1}(u_1)) \leq U_{r_1:n}^{(T)} \leq U_{r_2:n}^{(T)} \leq G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1)) \leq V_{s_1:n}^{(T)} \leq V_{s_2:n}^{(T)} \leq G_Y(F_Y^{-1}(v_2))] \\ = Q_{04}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2)))$$

και με απευθείας εφαρμογή του Λήμματος 3.5, προκύπτει η τελική έκφραση που ακολουθεί

$$p_{04} = p_{04}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b, c, d; r_1, r_2; s_1, s_2) \\ = \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} Q_{04}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))) \\ \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2.$$

Στο σημείο αυτό, η απόδειξη της Πρότασης ολοκληρώνεται. □

Στη συνέχεια, θέτοντας $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$ στην Έκφραση (6.4) της Πρότασης 6.3, προκύπτει ένας ακριβής τύπος για τον υπολογισμό της πιθανότητας συναγερωμού του διαγράμματος $O4$, ενώ εάν επιπρόσθετα το δείγμα αναφοράς και το δείγμα ελέγχου έχουν τον ίδιο σύνδεσμο ($C(u, v) = D(u, v)$) προκύπτει ο τύπος για τον υπολογισμό της πιθανότητας λανθασμένου συναγερωμού του νέου διαγράμματος. Τα αποτελέσματα παρουσιάζονται στα πορίσματα που ακολουθούν.

Πόρισμα 6.5. *Εάν $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, η πιθανότητα συναγερωμού του διαγράμματος $O4$, το οποίο ορίζεται από τον Κανόνα 6.2 δεν εξαρτάται από τις περιθώριες κατανομές και υπολογίζεται ως εξής*

$$\begin{aligned} AR_{O4}^{(C,D)} &= 1 - p_{O4}^{(F_X, F_X; F_Y, F_Y; C, D)}(m, n; a, b, c, d; r_1, r_2, s_1, s_2) \\ &= 1 - \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} Q_{O4}^{(D)}(u_1, u_2; v_1, v_2) \\ &\quad \times f_{a,b,c,d,m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2, \end{aligned} \quad (6.4)$$

όπου η ποσότητα $Q_{O4}^{(D)}(\cdot)$ δίνεται στη Σχέση (3.7) και η $f_{a,b,c,d,m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. της Σχέσης (3.4) με τις ποσότητες m, a, b, c, d, C να έχουν χρησιμοποιηθεί στη θέση των n, r_1, r_2, s_1, s_2, D , αντίστοιχα.

Απόδειξη. Σε αυτή την περίπτωση, δεδομένου ότι $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, θα ισχύει $G_X(F_X^{-1}(u_i)) = u_i$ και $G_Y(F_Y^{-1}(v_j)) = v_j$, για $i, j = 1, 2$. Παρόμοια, οι ποσότητες

$$\begin{aligned} &F_{r_1, r_2; s_2; n}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_2))), \\ &F_{r_2; s_1, s_2; n}^{(D)}(G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))), \end{aligned}$$

$$F_{r_1, r_2; s_1, s_2; n}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))),$$

οι οποίες εμφανίζονται στην έκφραση για την $Q_{O4}^{(D)}$ (βλ. Σχέση (3.7)), οδηγούν στις αντίστοιχες αθροιστικές συναρτήσεις $F_{r_1, r_2; s_2; n}^{(D)}(u_1, u_2; v_2)$, $F_{r_2; s_1, s_2; n}^{(D)}(u_2; v_1, v_2)$, $F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2)$. \square

Αξίζει να αναφερθεί ότι μία ισοδύναμη έκφραση για την πιθανότητα συναγερωμού AR_{O4} μπορεί να παραχθεί εάν εκμεταλλευτούμε τη Σχέση (3.7). Τότε, για την ποσότητα $p_{O4}^{(F_X, F_X; F_Y, F_Y; C, D)}$ θα έχουμε

$$p_{O4}^{(F_X, F_X; F_Y, F_Y; C, D)}(m, n; a, b, c, d; r_1, r_2, s_1, s_2) = I_1 - I_2 - I_3 + I_4,$$

όπου

$$\begin{aligned}
 I_1 &= \int_0^1 \int_0^1 F_{r_2; s_2; n}^{(D)}(u_2; v_2) \left(\int_0^{v_2} \int_0^{u_2} f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 dv_1 \right) du_2 dv_2 \\
 &= \int_0^1 \int_0^1 F_{r_2; s_2; n}^{(D)}(u_2; v_2) \times f_{b;d;m}^{(C)}(u_2; v_2) du_2 dv_2, \\
 I_2 &= \int_0^1 \int_0^1 \int_0^{u_2} F_{r_1, r_2; s_2; n}^{(D)}(u_1, u_2; v_2) \left(\int_0^{v_2} f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) dv_1 \right) du_1 du_2 dv_2 \\
 &= \int_0^1 \int_0^1 \int_0^{u_2} F_{r_1, r_2; s_2; n}^{(D)}(u_1, u_2; v_2) \times f_{a,b;d;m}^{(C)}(u_1, u_2; v_2) du_1 du_2 dv_2, \\
 I_3 &= \int_0^1 \int_0^{v_2} \int_0^1 F_{r_2; s_1, s_2; n}^{(D)}(u_2; v_1, v_2) \left(\int_0^{u_2} f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 \right) du_1 dv_1 dv_2 \\
 &= \int_0^1 \int_0^{v_2} \int_0^1 F_{r_2; s_1, s_2; n}^{(D)}(u_2; v_1, v_2) \times f_{b;c,d;m}^{(C)}(u_2; v_1, v_2) du_2 dv_1 dv_2, \\
 I_4 &= \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} F_{r_1, r_2; s_1, s_2; n}^{(D)}(u_1, u_2; v_1, v_2) \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2.
 \end{aligned}$$

Είναι προφανές ότι ο υπολογισμός κάθε ολοκληρώματος μπορεί να πραγματοποιηθεί ανακαλώντας τους τύπους που αναπτύχθηκαν στα Λήμματα 3.1, 3.2, 3.3 και 3.4.

Πόρισμα 6.6. Στην ειδική περίπτωση που ισχύει $F_X(x) = G_X(x)$, $F_Y(y) = G_Y(y)$ και $C(u, v) = D(u, v)$, προκύπτει η πιθανότητα λανθασμένου συναγερομού του διαγράμματος $O4$, η οποία δίνεται από την έκφραση

$$\begin{aligned}
 FAR_{O4}^{(C)} &= 1 - P(X_{a:m}^{(R)} \leq X_{r_1:n}^{(T)} \leq X_{r_2:n}^{(T)} \leq X_{b:m}^{(R)}, Y_{c:m}^{(R)} \leq Y_{s_1:n}^{(T)} \leq Y_{s_2:n}^{(T)} \leq Y_{d:m}^{(R)} | H_0) \\
 &= AR_{O4}^{(C,C)}(m, n; a, b, c, d; r_1, r_2; s_1, s_2),
 \end{aligned}$$

όπου η ποσότητα $Q_{O4}^{(C)}(\cdot)$ δίνεται στη Σχέση (3.7) για $D = C$ και η $f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. $F_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ της Σχέσης (3.4) που υπολογίζεται αντικαθιστώντας τις ποσότητες n, r_1, r_2, s_1, s_2, D με m, a, b, c, d, C , αντίστοιχα.

Απόδειξη. Το αποτέλεσμα είναι άμεσο βάσει της Έκφρασης (6.4). □

Από το παραπάνω πόρισμα μπορούμε να συμπεράνουμε ότι το νέο σχήμα αποτελεί ένα ημιπαραμετρικό διάγραμμα ελέγχου, το οποίο είναι μη παραμετρικό ως προς τις (μονοδιάστατες) περιθώριες κατανομές, αλλά τυπικά παραμετρικό όσον αφορά στη δομή εξάρτησης του ζεύγους (X, Y) . Παρόλα αυτά, στην επόμενη ενότητα θα διαπιστώσουμε ότι οι αλλαγές στις τιμές του FAR_{O4} είναι αμελητέες για διαφορετικούς συνδέσμους, γεγονός που μας επιτρέπει να χρησιμοποιούμε το διάγραμμα $O4$ εξ ολοκλήρου ως μη παραμετρικό.

6.4 Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

Στην παρούσα ενότητα, το νέο διάγραμμα εξετάζεται ως προς την εντός και εκτός ελέγχου απόδοσή του, όταν χρησιμοποιούνται διαφορετικά είδη συνδέσμων. Επιπλέον, μελετάται το πρόβλημα του σχεδιασμού του προτεινόμενου διαγράμματος. Για το σκοπό αυτό, παρουσιάζονται πίνακες βασισμένοι σε μερικές τυπικές τιμές του FAR_{O4} , με τη βοήθεια των οποίων μπορεί να δοθεί καθοδήγηση για τη σωστή επιλογή των παραμέτρων σχεδιασμού.

Στην πράξη, ο προσδιορισμός του FAR_{O4} μπορεί να γίνει για διαφορετικές τιμές των παραμέτρων σχεδιασμού ή για οποιοδήποτε αποδεκτό συνδυασμό ορισμένων παραμέτρων δεδομένου ότι οι υπόλοιπες έχουν λάβει μία προκαθορισμένη τιμή. Προκειμένου να μειωθεί το πλήθος των παραμέτρων, χρησιμοποιούνται τα κεντρικά ζευγάρια, δηλαδή τα ζευγάρια γύρω από το ζευγάρι των διαμέσων του δείγματος ελέγχου. Συνεπώς, στην περίπτωση που το n είναι περιττό, τα r_1, s_1 και r_2, s_2 επιλέγονται έτσι, ώστε να υπάρχει ίσος αριθμός παρατηρήσεων, έστω t , πριν και μετά τις διαμέσους, οπότε $r_1 = s_1 = (n + 1)/2 - t$ και $r_2 = s_2 = (n + 1)/2 + t$. Αντίθετα, όταν το n είναι άρτιο, μία λογική επιλογή είναι να θέσουμε $r_1 = s_1 = n/2 - t$ και $r_2 = s_2 = (n/2 + 1) + t$.

Λόγω της χρήσης των διαμέσων, οι παράμετροι σχεδιασμού a, b, c, d επιλέγονται συμμετρικά σε σχέση με τα άκρα τους, οπότε $b = m - a + 1$, $d = m - c + 1$. Όταν επιθυμούμε να ανιχνεύσουμε ίσες μετατοπίσεις και στα δύο χαρακτηριστικά, είναι εύλογο να επιλέξουμε $a = c$ και $b = d$, διαφορετικά τα a, c και b, d μπορεί να μην είναι ίσα μεταξύ τους, ανάλογα με το είδος της μετατόπισης που θέλουμε να εντοπίσουμε. Φυσικά, ο καθορισμός των παραμέτρων σχεδιασμού απαιτεί τον προκαθορισμό της ολικής πιθανότητας λανθασμένου συναγερμού (συνήθως 5 % ή 10 %) ή της τιμής του ARL_{in} (συνήθως 200, 370 ή 500).

Όπως έχει ήδη αναφερθεί στην προηγούμενη ενότητα, το προτεινόμενο διάγραμμα είναι ημιπαραμετρικό, ωστόσο το FAR (και το ARL_{in} , όπως θα δούμε αργότερα) παραμένει σχεδόν το ίδιο ανεξάρτητα από την επιλογή του συνδέσμου. Υπενθυμίζουμε, επίσης, ότι οι περιθώριες κατανομές δεν επηρεάζουν καθόλου τις εντός ελέγχου κατανομές. Κατά συνέπεια, το διάγραμμα $O4$ συμπεριφέρεται ως (σχεδόν) καθαρά μη παραμετρικό και αυτό φαίνεται να αποτελεί ένα από τα σημαντικότερα πλεονεκτήματά του. Αυτή η ελκυστική ιδιότητα διασαφηνίζεται καλύτερα στους Πίνακες 6.1, 6.2 και 6.3, όπου υπολογίζονται οι πιθανότητες συναγερμού AR_{O4} για πέντε διαφορετικούς συνδέσμους. Συγκεκριμένα, χρησιμοποιούμε τους συνδέσμους των AMH, CLP, FGM, GUH και FRA με τις αντίστοιχες α.σ.κ. που δίνονται στον Πίνακα 2.5.

6.4. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | | Μετατοπίσεις στη Μέση Τιμή : $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ | | | | | | | | |
|-------------------------------|-----|-----|---------|--|------|------|-------------|-------------|------|------|------|------|
| Παράμετροι Σχεδιασμού | | | | -2 | -1.5 | -1 | -0.5 | 0 | 0.5 | 1 | 1.5 | 2 |
| m | n | t | $a = c$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.97 | 0.79 | 0.41 | 0.12 | 0.04 | 0.13 | 0.41 | 0.79 | 0.97 |
| | | | 4 | 0.99 | 0.86 | 0.52 | 0.18 | 0.07 | 0.18 | 0.51 | 0.87 | 0.99 |
| | | | 5 | 1 | 0.91 | 0.61 | 0.23 | 0.10 | 0.24 | 0.61 | 0.92 | 0.99 |
| | 11 | 1 | 13 | 1 | 1 | 0.83 | 0.28 | 0.05 | 0.28 | 0.83 | 1 | 1 |
| | | | 14 | 1 | 1 | 0.87 | 0.33 | 0.06 | 0.33 | 0.86 | 1 | 1 |
| | | | 16 | 1 | 1 | 0.92 | 0.42 | 0.10 | 0.43 | 0.92 | 1 | 1 |
| | 25 | 1 | 24 | 1 | 1 | 0.99 | 0.56 | 0.05 | 0.57 | 0.99 | 1 | 1 |
| 25 | | | 1 | 1 | 1 | 0.62 | 0.07 | 0.62 | 1 | 1 | 1 | |
| 26 | | | 1 | 1 | 1 | 0.67 | 0.10 | 0.67 | 1 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.95 | 0.73 | 0.37 | 0.11 | 0.04 | 0.13 | 0.40 | 0.75 | 0.95 |
| | | | 4 | 0.97 | 0.82 | 0.48 | 0.17 | 0.07 | 0.18 | 0.50 | 0.83 | 0.98 |
| | | | 5 | 0.99 | 0.88 | 0.56 | 0.22 | 0.10 | 0.24 | 0.58 | 0.88 | 0.99 |
| | 11 | 1 | 13 | 1 | 0.99 | 0.78 | 0.26 | 0.05 | 0.27 | 0.80 | 0.99 | 1 |
| | | | 14 | 1 | 0.99 | 0.83 | 0.30 | 0.06 | 0.31 | 0.83 | 0.99 | 1 |
| | | | 16 | 1 | 1 | 0.88 | 0.39 | 0.10 | 0.40 | 0.89 | 1 | 1 |
| | 25 | 1 | 24 | 1 | 1 | 0.98 | 0.53 | 0.05 | 0.54 | 0.98 | 1 | 1 |
| 25 | | | 1 | 1 | 0.99 | 0.58 | 0.07 | 0.58 | 0.99 | 1 | 1 | |
| 26 | | | 1 | 1 | 0.99 | 0.62 | 0.10 | 0.64 | 0.99 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = 0.5$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.96 | 0.77 | 0.41 | 0.13 | 0.04 | 0.13 | 0.41 | 0.76 | 0.95 |
| | | | 4 | 0.98 | 0.84 | 0.50 | 0.18 | 0.07 | 0.17 | 0.50 | 0.84 | 0.98 |
| | | | 5 | 0.99 | 0.89 | 0.60 | 0.23 | 0.10 | 0.23 | 0.60 | 0.89 | 0.99 |
| | 11 | 1 | 13 | 1 | 0.99 | 0.80 | 0.27 | 0.05 | 0.28 | 0.8 | 0.99 | 1 |
| | | | 14 | 1 | 0.99 | 0.84 | 0.32 | 0.06 | 0.32 | 0.84 | 0.99 | 1 |
| | | | 16 | 1 | 1 | 0.90 | 0.41 | 0.10 | 0.40 | 0.90 | 1 | 1 |
| | 25 | 1 | 24 | 1 | 1 | 0.99 | 0.55 | 0.05 | 0.53 | 0.99 | 1 | 1 |
| 25 | | | 1 | 1 | 0.99 | 0.59 | 0.07 | 0.59 | 0.99 | 1 | 1 | |
| 26 | | | 1 | 1 | 0.99 | 0.64 | 0.10 | 0.64 | 0.99 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 3$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.97 | 0.80 | 0.42 | 0.13 | 0.05 | 0.12 | 0.42 | 0.79 | 0.97 |
| | | | 4 | 0.99 | 0.87 | 0.52 | 0.18 | 0.07 | 0.18 | 0.52 | 0.87 | 0.99 |
| | | | 5 | 1 | 0.92 | 0.62 | 0.24 | 0.10 | 0.24 | 0.62 | 0.92 | 1 |
| | 11 | 1 | 13 | 1 | 1 | 0.84 | 0.29 | 0.05 | 0.28 | 0.83 | 1 | 1 |
| | | | 14 | 1 | 1 | 0.88 | 0.33 | 0.06 | 0.33 | 0.87 | 1 | 1 |
| | | | 16 | 1 | 1 | 0.92 | 0.43 | 0.10 | 0.44 | 0.92 | 1 | 1 |
| | 25 | 1 | 24 | 1 | 1 | 0.99 | 0.57 | 0.05 | 0.57 | 0.99 | 1 | 1 |
| 25 | | | 1 | 1 | 1 | 0.62 | 0.08 | 0.63 | 1 | 1 | 1 | |
| 26 | | | 1 | 1 | 1 | 0.68 | 0.10 | 0.68 | 1 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 1$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.97 | 0.78 | 0.41 | 0.13 | 0.05 | 0.12 | 0.41 | 0.77 | 0.96 |
| | | | 4 | 0.98 | 0.86 | 0.51 | 0.18 | 0.07 | 0.18 | 0.51 | 0.86 | 0.98 |
| | | | 5 | 0.99 | 0.90 | 0.60 | 0.23 | 0.10 | 0.24 | 0.61 | 0.91 | 0.99 |
| | 11 | 1 | 13 | 1 | 0.99 | 0.82 | 0.28 | 0.05 | 0.28 | 0.82 | 0.99 | 1 |
| | | | 14 | 1 | 1 | 0.85 | 0.33 | 0.07 | 0.32 | 0.86 | 1 | 1 |
| | | | 16 | 1 | 1 | 0.91 | 0.42 | 0.10 | 0.42 | 0.91 | 1 | 1 |
| | 25 | 1 | 24 | 1 | 1 | 0.99 | 0.56 | 0.05 | 0.56 | 0.99 | 1 | 1 |
| 25 | | | 1 | 1 | 0.99 | 0.61 | 0.07 | 0.61 | 0.99 | 1 | 1 | |
| 26 | | | 1 | 1 | 1 | 0.66 | 0.10 | 0.66 | 1 | 1 | 1 | |

Πίνακας 6.1: Πιθανότητες Συναγερμού AR_{04} λόγω μετατοπίσεων μέσης τιμής για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$)

Δισδιάστατο Διάγραμμα Ελέγχου
για τη Θέση και τη Διασπορά με Χρήση Διατεταγμένων Παρατηρήσεων

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | | Μετατοπίσεις στην Τυπική Απόκλιση : $\sigma_{out} = \sigma_X^{(T)} = \sigma_Y^{(T)}$ | | | | | | | | |
|-------------------------------|-----|-----|---------|--|------|------|------|------|------|------|------|------|
| Παράμετροι Σχεδιασμού | | | | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 |
| m | n | t | $a = c$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.04 | 0.31 | 0.62 | 0.80 | 0.90 | 0.95 | 0.97 | 0.98 | 0.99 |
| | | | 4 | 0.07 | 0.38 | 0.68 | 0.84 | 0.93 | 0.96 | 0.98 | 0.99 | 0.99 |
| | | | 5 | 0.10 | 0.47 | 0.74 | 0.88 | 0.95 | 0.97 | 0.99 | 0.99 | 1 |
| | 11 | | 13 | 0.05 | 0.39 | 0.73 | 0.90 | 0.96 | 0.99 | 0.99 | 1 | 1 |
| | | | 14 | 0.07 | 0.45 | 0.78 | 0.92 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 16 | 0.10 | 0.51 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 |
| | 25 | | 24 | 0.05 | 0.36 | 0.72 | 0.90 | 0.96 | 0.99 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.47 | 0.80 | 0.93 | 0.98 | 0.99 | 1 | 1 | 1 |
| | | | 26 | 0.11 | 0.53 | 0.84 | 0.95 | 0.98 | 1 | 1 | 1 | 1 |
| m | n | t | $a = c$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.04 | 0.30 | 0.61 | 0.80 | 0.90 | 0.95 | 0.97 | 0.98 | 0.99 |
| | | | 4 | 0.07 | 0.40 | 0.68 | 0.85 | 0.92 | 0.96 | 0.98 | 0.99 | 0.99 |
| | | | 5 | 0.10 | 0.46 | 0.74 | 0.88 | 0.94 | 0.97 | 0.99 | 0.99 | 1 |
| | 11 | | 13 | 0.05 | 0.39 | 0.72 | 0.89 | 0.96 | 0.98 | 0.99 | 1 | 1 |
| | | | 14 | 0.07 | 0.44 | 0.77 | 0.92 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 16 | 0.10 | 0.50 | 0.81 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 |
| | 25 | | 24 | 0.05 | 0.36 | 0.71 | 0.90 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.48 | 0.80 | 0.93 | 0.98 | 0.99 | 1 | 1 | 1 |
| | | | 26 | 0.11 | 0.52 | 0.83 | 0.95 | 0.99 | 1 | 1 | 1 | 1 |
| m | n | t | $a = c$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = 0.5$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.04 | 0.30 | 0.62 | 0.80 | 0.89 | 0.95 | 0.97 | 0.98 | 0.99 |
| | | | 4 | 0.07 | 0.38 | 0.69 | 0.84 | 0.93 | 0.96 | 0.98 | 0.99 | 0.99 |
| | | | 5 | 0.10 | 0.45 | 0.75 | 0.88 | 0.94 | 0.97 | 0.99 | 0.99 | 1 |
| | 11 | | 13 | 0.05 | 0.39 | 0.74 | 0.90 | 0.96 | 0.99 | 0.99 | 1 | 1 |
| | | | 14 | 0.07 | 0.45 | 0.78 | 0.92 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 16 | 0.10 | 0.51 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 |
| | 25 | | 24 | 0.05 | 0.36 | 0.72 | 0.90 | 0.96 | 0.99 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.47 | 0.80 | 0.93 | 0.98 | 0.99 | 1 | 1 | 1 |
| | | | 26 | 0.11 | 0.53 | 0.83 | 0.95 | 0.99 | 1 | 1 | 1 | 1 |
| m | n | t | $a = c$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 3$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.04 | 0.31 | 0.61 | 0.80 | 0.90 | 0.94 | 0.97 | 0.98 | 0.99 |
| | | | 4 | 0.07 | 0.38 | 0.68 | 0.85 | 0.93 | 0.96 | 0.98 | 0.99 | 0.99 |
| | | | 5 | 0.10 | 0.45 | 0.74 | 0.88 | 0.94 | 0.97 | 0.99 | 0.99 | 0.99 |
| | 11 | | 13 | 0.05 | 0.39 | 0.73 | 0.90 | 0.96 | 0.99 | 0.99 | 1 | 1 |
| | | | 14 | 0.07 | 0.45 | 0.77 | 0.92 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 16 | 0.10 | 0.50 | 0.81 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 |
| | 25 | | 24 | 0.05 | 0.36 | 0.72 | 0.90 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.47 | 0.80 | 0.93 | 0.98 | 0.99 | 1 | 1 | 1 |
| | | | 26 | 0.11 | 0.53 | 0.84 | 0.95 | 0.99 | 1 | 1 | 1 | 1 |
| m | n | t | $a = c$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 1$) | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.04 | 0.31 | 0.62 | 0.81 | 0.90 | 0.95 | 0.97 | 0.98 | 0.99 |
| | | | 4 | 0.07 | 0.39 | 0.69 | 0.85 | 0.93 | 0.96 | 0.98 | 0.99 | 0.99 |
| | | | 5 | 0.10 | 0.46 | 0.75 | 0.89 | 0.94 | 0.97 | 0.99 | 0.99 | 0.99 |
| | 11 | | 13 | 0.05 | 0.40 | 0.74 | 0.90 | 0.96 | 0.98 | 0.99 | 1 | 1 |
| | | | 14 | 0.07 | 0.46 | 0.78 | 0.92 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 16 | 0.10 | 0.50 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 |
| | 25 | | 24 | 0.05 | 0.37 | 0.72 | 0.90 | 0.97 | 0.99 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.49 | 0.79 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 |
| | | | 26 | 0.11 | 0.53 | 0.83 | 0.95 | 0.99 | 1 | 1 | 1 | 1 |

152 **Πίνακας 6.2:** Πιθανότητες Συναγεμμού AR_{04} λόγω μετατοπίσεων τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$)

6.4. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | | Ταυτόχρονες Μετατοπίσεις στη Μέση Τιμή και την Τυπική Απόκλιση | | | | | | | | | |
|-------------------------------|-----|-----|---------|--|------|------|------|-------------|------|------|------|------|---|
| Παράμετροι Σχεδιασμού | | | | Μέση Τιμή: | -2 | -1.5 | -1 | -0.5 | 0 | 0.5 | 1 | 1.5 | 2 |
| | | | | Τυπ. Αποκλ.: | 3 | 2.5 | 2 | 1.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| m | n | t | $a = c$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.98 | 0.94 | 0.80 | 0.41 | 0.04 | 0.41 | 0.80 | 0.94 | 0.98 | |
| | | | 4 | 0.99 | 0.96 | 0.85 | 0.50 | 0.07 | 0.50 | 0.85 | 0.96 | 0.99 | |
| | | | 5 | 0.99 | 0.97 | 0.88 | 0.56 | 0.10 | 0.57 | 0.88 | 0.97 | 0.99 | |
| | 11 | 3 | 4 | 1 | 1 | 0.97 | 0.64 | 0.05 | 0.64 | 0.96 | 1 | 1 | |
| | | | 5 | 1 | 1 | 0.98 | 0.73 | 0.08 | 0.73 | 0.98 | 1 | 1 | |
| | | | 6 | 1 | 1 | 0.99 | 0.79 | 0.11 | 0.79 | 0.99 | 1 | 1 | |
| | 25 | 5 | 12 | 1 | 1 | 1 | 0.83 | 0.05 | 0.84 | 1 | 1 | 1 | |
| | | | 13 | 1 | 1 | 1 | 0.87 | 0.08 | 0.87 | 1 | 1 | 1 | |
| | | | 13 | 1 | 1 | 1 | 0.90 | 0.11 | 0.90 | 1 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.97 | 0.92 | 0.75 | 0.39 | 0.04 | 0.41 | 0.78 | 0.92 | 0.97 | |
| | | | 4 | 0.98 | 0.94 | 0.81 | 0.46 | 0.07 | 0.48 | 0.83 | 0.94 | 0.98 | |
| | | | 5 | 0.98 | 0.95 | 0.84 | 0.53 | 0.10 | 0.56 | 0.87 | 0.96 | 0.98 | |
| | 11 | 3 | 4 | 1 | 1 | 0.94 | 0.60 | 0.05 | 0.63 | 0.95 | 1 | 1 | |
| | | | 6 | 1 | 1 | 0.97 | 0.68 | 0.08 | 0.71 | 0.97 | 1 | 1 | |
| | | | 12 | 1 | 1 | 0.98 | 0.75 | 0.11 | 0.78 | 0.98 | 1 | 1 | |
| | 25 | 5 | 13 | 1 | 1 | 1 | 0.80 | 0.05 | 0.81 | 1 | 1 | 1 | |
| | | | 14 | 1 | 1 | 1 | 0.84 | 0.08 | 0.86 | 1 | 1 | 1 | |
| | | | 24 | 1 | 1 | 1 | 0.87 | 0.11 | 0.88 | 1 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = 0.5$) | | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.97 | 0.92 | 0.78 | 0.40 | 0.04 | 0.41 | 0.78 | 0.93 | 0.97 | |
| | | | 4 | 0.98 | 0.95 | 0.83 | 0.49 | 0.07 | 0.50 | 0.83 | 0.94 | 0.98 | |
| | | | 5 | 0.99 | 0.96 | 0.86 | 0.56 | 0.10 | 0.57 | 0.86 | 0.96 | 0.98 | |
| | 11 | 3 | 4 | 1 | 1 | 0.96 | 0.62 | 0.05 | 0.63 | 0.95 | 1 | 1 | |
| | | | 5 | 1 | 1 | 0.97 | 0.71 | 0.08 | 0.71 | 0.97 | 1 | 1 | |
| | | | 12 | 1 | 1 | 0.98 | 0.77 | 0.11 | 0.77 | 0.98 | 1 | 1 | |
| | 25 | 5 | 13 | 1 | 1 | 1 | 0.82 | 0.05 | 0.82 | 1 | 1 | 1 | |
| | | | 14 | 1 | 1 | 1 | 0.86 | 0.08 | 0.85 | 1 | 1 | 1 | |
| | | | 24 | 1 | 1 | 1 | 0.89 | 0.11 | 0.89 | 1 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 3$) | | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.98 | 0.94 | 0.81 | 0.42 | 0.04 | 0.42 | 0.80 | 0.95 | 0.98 | |
| | | | 4 | 0.99 | 0.96 | 0.85 | 0.50 | 0.07 | 0.49 | 0.85 | 0.96 | 0.99 | |
| | | | 5 | 0.99 | 0.97 | 0.89 | 0.58 | 0.10 | 0.58 | 0.89 | 0.97 | 0.99 | |
| | 11 | 3 | 4 | 1 | 1 | 0.97 | 0.65 | 0.05 | 0.63 | 0.97 | 1 | 1 | |
| | | | 5 | 1 | 1 | 0.98 | 0.73 | 0.08 | 0.73 | 0.98 | 1 | 1 | |
| | | | 12 | 1 | 1 | 0.99 | 0.79 | 0.11 | 0.79 | 0.99 | 1 | 1 | |
| | 25 | 5 | 13 | 1 | 1 | 1 | 0.82 | 0.05 | 0.81 | 1 | 1 | 1 | |
| | | | 14 | 1 | 1 | 1 | 0.85 | 0.08 | 0.86 | 1 | 1 | 1 | |
| | | | 24 | 1 | 1 | 1 | 0.89 | 0.11 | 0.89 | 1 | 1 | 1 | |
| m | n | t | $a = c$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 1$) | | | | | | | | | |
| 100 | 5 | 1 | 3 | 0.98 | 0.94 | 0.79 | 0.41 | 0.04 | 0.41 | 0.79 | 0.94 | 0.97 | |
| | | | 4 | 0.98 | 0.96 | 0.84 | 0.50 | 0.07 | 0.50 | 0.84 | 0.95 | 0.98 | |
| | | | 5 | 0.99 | 0.97 | 0.88 | 0.57 | 0.10 | 0.58 | 0.87 | 0.96 | 0.99 | |
| | 11 | 3 | 4 | 1 | 1 | 0.96 | 0.63 | 0.05 | 0.64 | 0.96 | 1 | 1 | |
| | | | 5 | 1 | 1 | 0.98 | 0.71 | 0.08 | 0.72 | 0.98 | 1 | 1 | |
| | | | 6 | 1 | 1 | 0.98 | 0.79 | 0.11 | 0.80 | 0.98 | 1 | 1 | |
| | 25 | 5 | 12 | 1 | 1 | 1 | 0.83 | 0.05 | 0.82 | 1 | 1 | 1 | |
| | | | 13 | 1 | 1 | 1 | 0.87 | 0.08 | 0.87 | 1 | 1 | 1 | |
| | | | 14 | 1 | 1 | 1 | 0.90 | 0.11 | 0.90 | 1 | 1 | 1 | |

Πίνακας 6.3: Πιθανότητες Συναγερμού AR_{04} λόγω ταυτόχρονων μετατοπίσεων μέσης τιμής και τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$)

Η επιλογή των παραμέτρων των συνδέσμων είναι κοινή τόσο για την εντός ελέγχου όσο και για την εκτός ελέγχου περίπτωση. Για τον υπολογισμό του FAR_{O4} , έχουν χρησιμοποιηθεί τυπικές κανονικές περιθώριες κατανομές, ενώ οι πιθανότητες συναγερμού AR_{O4} , έχουν υπολογισθεί καθώς η μέση τιμή και η τυπική απόκλιση της διεργασίας μετατοπίζονται από τις εντός ελέγχου τιμές $\mu_{in} = \mu_X^{(R)} = \mu_Y^{(R)} = 0$ και $\sigma_{in} = \sigma_X^{(R)} = \sigma_Y^{(R)} = 1$ στις εκτός ελέγχου τιμές $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ και $\sigma_{out} = \sigma_X^{(T)} = \sigma_Y^{(T)}$ (που εμφανίζονται στο άνω τμήμα των πινάκων). Ειδικότερα, πρέπει να αναφερθεί ότι στον Πίνακα 6.1 μετατοπίζεται μόνο η μέση τιμή, ενώ η τυπική απόκλιση παραμένει σταθερή, στον Πίνακα 6.2 μετατοπίζεται η τυπική απόκλιση διατηρώντας σταθερή τη μέση τιμή και στον Πίνακα 6.3 βλέπουμε πως συμπεριφέρεται το διάγραμμα $O4$ όταν προκαλούνται ταυτόχρονες μετατοπίσεις.

Επιπλέον, σε κάθε περίπτωση, το μέγεθος του δείγματος αναφοράς καθορίζεται ίσο με $m = 100$, ενώ εξετάζονται τρία διαφορετικά μεγέθη του δείγματος ελέγχου ($n = 5, 11, 25$). Δεδομένου ότι χρησιμοποιούνται οι δειγματικές διάμεσοι και ταυτόχρονα πραγματοποιούνται ίσες μετατοπίσεις στα δύο χαρακτηριστικά, χρησιμοποιούνται συμμετρικά όρια ελέγχου με τα a, b να τίθενται ίσα με τα c, d , αντίστοιχα. Τέλος, γίνονται διαφορετικές επιλογές της παραμέτρου σχεδιασμού t έτσι, ώστε να επιτευχθεί το επιθυμητό FAR_{O4} (τιμές με έντονους χαρακτήρες), το οποίο κυμαίνεται από 5% έως 10%.

Είναι σαφές από τους Πίνακες 6.1, 6.2 και 6.3 ότι, για σταθερές τιμές των παραμέτρων σχεδιασμού, οι τιμές των FAR_{O4} είναι ακριβώς οι ίδιες ανεξάρτητα από την επιλογή του συνδέσμου. Επομένως, το προτεινόμενο διάγραμμα μπορεί να χρησιμοποιηθεί σε ένα καθαρά μη παραμετρικό πλαίσιο. Επιπρόσθετα, σε όλα τα είδη μετατοπίσεων, το προτεινόμενο διάγραμμα είναι ικανό να παράγει εκτός ελέγχου σήμα πολύ γρήγορα.

Για να διευκολύνουμε τους χρήστες στην προσαρμογή του διαγράμματος ανάλογα με τις ανάγκες τους, απεικονίζουμε το FAR_{O4} στον Πίνακα 6.4, παρέχοντας έτσι περισσότερες επιλογές των παραμέτρων σχεδιασμού. Οι υπολογισμοί έγιναν με τη χρήση των διαμέσων του δείγματος ελέγχου και συμμετρικά τοποθετημένων ορίων ελέγχου. Παρόλα αυτά, μπορούν εύκολα να προκύψουν ανάλογα αποτελέσματα για μη συμμετρικά όρια ελέγχου με στόχο τη διευκόλυνση του σχεδιασμού του διαγράμματος $O4$ σε περιπτώσεις κατανομών με ασύμμετρες ή βαριές ουρές.

Ένα ακόμη σημείο στο οποίο αξίζει να δοθεί έμφαση σχετικά με την επιλογή των παραμέτρων σχεδιασμού είναι όταν το ενδιαφέρον επικεντρώνεται στην ανίχνευση ίσων μετατοπίσεων της μέσης τιμής και της διακύμανσης των δύο χαρακτηριστικών. Σε αυτή την περίπτωση, μπορούμε να ορίσουμε $a = c$ και $r_1 = s_1, r_2 = s_2$, διαφορετικά εάν ο κάποιος

6.4. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

ενδιαφέρεται να εντοπίσει μεγαλύτερες μετατοπίσεις στη μέση τιμή της μεταβλητής X , θα ήταν πιο ενδεδειγμένο να χρησιμοποιήσει $a > c$ και $r_1 = s_1, r_2 = s_2$. Παρόμοια, αν ενδιαφέρεται να ανιχνεύσει μεγαλύτερες μεταβολές της διακύμανσης του πρώτο χαρακτηριστικού, θα ήταν προτιμότερο να επιλέξει $a = c$ και $r_1 < s_1, r_2 > s_2$.

| Παράμετροι Σχεδιασμού | | | Μέγεθος Δείγματος Αναφοράς m | | | | | | | |
|-----------------------|-----|--|--------------------------------|--------|---------|--------|---------|--------|---------|--------|
| n | t | # Παρ. \in Όρια (ακριβώς $2t + 1$) | 100 | | 200 | | 500 | | 1000 | |
| | | | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR |
| 5 | 1 | 60% | 3 | 0.0408 | 7 | 0.0508 | 18 | 0.0507 | 37 | 0.0508 |
| | | | 4 | 0.0709 | 8 | 0.0620 | 22 | 0.0714 | 45 | 0.0755 |
| | | | 5 | 0.0988 | 10 | 0.0959 | 26 | 0.1014 | 53 | 0.0978 |
| 11 | 1 | 27.27% | 13 | 0.0479 | 27 | 0.0483 | 70 | 0.0523 | 142 | 0.0526 |
| | | | 14 | 0.0627 | 30 | 0.0741 | 77 | 0.0723 | 156 | 0.0771 |
| | | | 16 | 0.1017 | 33 | 0.1027 | 84 | 0.1008 | 170 | 0.1041 |
| | 2 | 45.45% | 8 | 0.0477 | 17 | 0.0506 | 43 | 0.0489 | 90 | 0.0536 |
| | | | 9 | 0.0714 | 19 | 0.0710 | 49 | 0.0723 | 100 | 0.0756 |
| | | | 10 | 0.0952 | 21 | 0.0984 | 55 | 0.1060 | 110 | 0.0984 |
| | 3 | 63.64% | 4 | 0.0466 | 9 | 0.0525 | 22 | 0.0501 | 46 | 0.0531 |
| | | | 5 | 0.0844 | 11 | 0.0879 | 26 | 0.0694 | 53 | 0.0695 |
| | | | 6 | 0.1156 | 12 | 0.1052 | 30 | 0.1041 | 60 | 0.0977 |
| 25 | 1 | 12% | 24 | 0.0544 | 50 | 0.0538 | 126 | 0.0458 | 255 | 0.0515 |
| | | | 25 | 0.0708 | 52 | 0.0794 | 132 | 0.0689 | 267 | 0.0720 |
| | | | 26 | 0.0982 | 54 | 0.1023 | 138 | 0.1015 | 279 | 0.1011 |
| | 2 | 20% | 20 | 0.0477 | 43 | 0.0524 | 109 | 0.0505 | 223 | 0.0531 |
| | | | 22 | 0.0788 | 45 | 0.0720 | 115 | 0.0729 | 234 | 0.0727 |
| | | | 23 | 0.1037 | 47 | 0.0983 | 121 | 0.1017 | 245 | 0.1043 |
| | 3 | 28% | 17 | 0.0472 | 36 | 0.0505 | 95 | 0.0516 | 190 | 0.0505 |
| | | | 19 | 0.0829 | 39 | 0.0789 | 100 | 0.0758 | 200 | 0.0693 |
| | | | 20 | 0.1078 | 41 | 0.1014 | 105 | 0.1021 | 210 | 0.1005 |
| | 4 | 36% | 14 | 0.0457 | 31 | 0.0539 | 80 | 0.0518 | 160 | 0.0515 |
| | | | 16 | 0.0812 | 33 | 0.0759 | 85 | 0.0796 | 170 | 0.0751 |
| | | | 17 | 0.1105 | 35 | 0.1048 | 90 | 0.1027 | 180 | 0.1035 |
| | 5 | 44% | 12 | 0.0538 | 25 | 0.0527 | 65 | 0.0512 | 130 | 0.0523 |
| | | | 13 | 0.0812 | 27 | 0.0721 | 70 | 0.0749 | 140 | 0.0704 |
| | | | 14 | 0.1093 | 29 | 0.1035 | 74 | 0.1001 | 150 | 0.1026 |
| | 6 | 52% | 9 | 0.0479 | 20 | 0.0537 | 51 | 0.0479 | 105 | 0.0498 |
| | | | 10 | 0.0721 | 22 | 0.0829 | 55 | 0.0659 | 113 | 0.0723 |
| | | | 11 | 0.0999 | 23 | 0.0989 | 59 | 0.0965 | 121 | 0.0971 |
| | 7 | 60% | 7 | 0.0553 | 15 | 0.0539 | 40 | 0.0530 | 80 | 0.0519 |
| | | | 8 | 0.0874 | 17 | 0.0841 | 43 | 0.0752 | 87 | 0.0755 |
| | | | 9 | 0.1292 | 18 | 0.1055 | 46 | 0.0969 | 94 | 0.1005 |

Πίνακας 6.4: Πιθανότητες Λανθασμένου Συναγερμού FAR_{04} για δεδομένο σχεδιασμό
($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1 = (n + 1)/2 - t, r_2 = s_2 = (n + 1)/2 + t$)

Για την καθοδήγηση του σχεδιασμού του διαγράμματος σε τέτοιου είδους καταστάσεις, οι οποίες χρήζουν αντιμετώπισης σε πραγματικές εφαρμογές της καθημερινής ζωής, παρέχονται οι Πίνακες 6.5 και 6.6.

| Παράμετροι Σχεδιασμού | | | Μέγεθος Δείγματος Αναφοράς m | | | | | | | |
|-----------------------|-----|--|--------------------------------|--------|---------|--------|-----------|--------|-----------|--------|
| n | t | # Παρ. \in Όρια (ακριβώς $2t + 1$) | 100 | | 200 | | 500 | | 1000 | |
| | | | $a > c$ | FAR | $a > c$ | FAR | $a > c$ | FAR | $a > c$ | FAR |
| 5 | 1 | 60% | 4 > 2 | 0.0472 | 9 > 5 | 0.0515 | 20 > 16 | 0.0511 | 41 > 33 | 0.0507 |
| | | | 5 > 3 | 0.0684 | 11 > 6 | 0.0722 | 25 > 19 | 0.0715 | 51 > 39 | 0.0796 |
| | | | 6 > 4 | 0.1000 | 13 > 7 | 0.1010 | 30 > 22 | 0.1001 | 61 > 45 | 0.0970 |
| 11 | 1 | 27.27 % | 16 > 7 | 0.0513 | 28 > 25 | 0.0492 | 74 > 65 | 0.0509 | 146 > 136 | 0.0504 |
| | | | 17 > 10 | 0.0761 | 32 > 27 | 0.0724 | 81 > 72 | 0.0726 | 159 > 150 | 0.0760 |
| | | | 18 > 13 | 0.1028 | 36 > 29 | 0.1002 | 88 > 79 | 0.0987 | 172 > 162 | 0.1008 |
| | 2 | 45.45 % | 10 > 4 | 0.0511 | 20 > 13 | 0.0496 | 49 > 40 | 0.0509 | 93 > 83 | 0.0510 |
| | | | 11 > 6 | 0.0735 | 22 > 15 | 0.0727 | 54 > 44 | 0.0742 | 104 > 93 | 0.0708 |
| | | | 12 > 8 | 0.1041 | 24 > 17 | 0.1026 | 59 > 48 | 0.1003 | 115 > 103 | 0.1010 |
| | 3 | 63.63 % | 5 > 3 | 0.0537 | 10 > 6 | 0.0474 | 24 > 21 | 0.0496 | 50 > 41 | 0.0510 |
| | | | 5 > 4 | 0.0648 | 12 > 7 | 0.0724 | 29 > 23 | 0.0718 | 57 > 49 | 0.0724 |
| | | | 6 > 5 | 0.0966 | 14 > 8 | 0.0987 | 34 > 25 | 0.1026 | 64 > 57 | 0.0986 |
| 25 | 1 | 12% | 26 > 18 | 0.0516 | 51 > 45 | 0.0493 | 130 > 121 | 0.0503 | 257 > 250 | 0.0495 |
| | | | 27 > 21 | 0.0723 | 54 > 48 | 0.0730 | 137 > 126 | 0.0720 | 269 > 262 | 0.0697 |
| | | | 28 > 24 | 0.1030 | 57 > 51 | 0.1034 | 144 > 131 | 0.1010 | 281 > 274 | 0.1010 |
| | 2 | 20% | 23 > 13 | 0.0527 | 46 > 38 | 0.0515 | 113 > 104 | 0.0509 | 228 > 210 | 0.0498 |
| | | | 24 > 16 | 0.0721 | 48 > 40 | 0.0740 | 118 > 110 | 0.0725 | 239 > 225 | 0.0731 |
| | | | 25 > 19 | 0.1030 | 50 > 42 | 0.0980 | 123 > 116 | 0.0976 | 250 > 240 | 0.1028 |
| | 3 | 28% | 19 > 15 | 0.0495 | 39 > 33 | 0.0497 | 97 > 90 | 0.0513 | 196 > 180 | 0.0514 |
| | | | 20 > 16 | 0.0706 | 41 > 36 | 0.0705 | 102 > 94 | 0.0703 | 207 > 192 | 0.0702 |
| | | | 21 > 17 | 0.0956 | 43 > 39 | 0.1034 | 109 > 98 | 0.1001 | 218 > 204 | 0.1028 |
| | 4 | 36% | 16 > 11 | 0.0505 | 31 > 29 | 0.0502 | 82 > 77 | 0.0510 | 166 > 154 | 0.0510 |
| | | | 17 > 13 | 0.0725 | 33 > 31 | 0.0659 | 87 > 82 | 0.0775 | 176 > 163 | 0.0750 |
| | | | 18 > 15 | 0.1051 | 35 > 33 | 0.0955 | 92 > 87 | 0.1025 | 186 > 172 | 0.1024 |
| | 5 | 44% | 13 > 9 | 0.0483 | 26 > 21 | 0.0469 | 66 > 62 | 0.0497 | 140 > 124 | 0.0525 |
| | | | 14 > 10 | 0.0709 | 29 > 22 | 0.0693 | 72 > 65 | 0.0730 | 145 > 136 | 0.0701 |
| | | | 15 > 11 | 0.0968 | 32 > 23 | 0.1018 | 78 > 68 | 0.1010 | 150 > 148 | 0.1016 |
| | 6 | 52% | 10 > 8 | 0.0502 | 21 > 18 | 0.0508 | 55 > 48 | 0.0512 | 111 > 93 | 0.0502 |
| | | | 11 > 9 | 0.0718 | 23 > 19 | 0.0738 | 59 > 50 | 0.0684 | 117 > 105 | 0.0695 |
| | | | 12 > 9 | 0.1007 | 25 > 20 | 0.1015 | 63 > 52 | 0.0958 | 123 > 117 | 0.1002 |
| | 7 | 60% | 8 > 3 | 0.0458 | 16 > 12 | 0.0488 | 41 > 35 | 0.0507 | 84 > 72 | 0.0486 |
| | | | 9 > 4 | 0.0735 | 18 > 13 | 0.0676 | 46 > 37 | 0.0679 | 90 > 82 | 0.0737 |
| | | | 10 > 5 | 0.1032 | 20 > 14 | 0.1014 | 51 > 39 | 0.1030 | 96 > 92 | 0.1002 |

Πίνακας 6.5: Πιθανότητες Λανθασμένου Συναγερμού FAR_{04} για δεδομένο σχεδιασμό
($a > c, b = m - a + 1, d = m - c + 1$ & $r_1 = s_1, r_2 = s_2$)

6.4. Μελέτη απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| Παράμετροι Σχεδιασμού | | | | Μέγεθος Δείγματος Αναφοράς m | | | | | | | |
|-----------------------|-------------|--------------------------------------|--------------------------|--------------------------------|--------|---------|--------|---------|--------|---------|--------|
| n | $t_X > t_Y$ | # Παρ. ∈ Όρια (ακριβώς) (ακριβώς) | | 100 | | 200 | | 500 | | 1000 | |
| | | $\binom{2t_X+1}{2t_X+1}$ | $\binom{2t_Y+1}{2t_Y+1}$ | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR |
| 5 | 2 > 1 | 100% | 60% | 1 | 0.0946 | 1 | 0.0489 | 3 | 0.0595 | 5 | 0.0490 |
| | | | | 2 | 0.1878 | 2 | 0.0949 | 4 | 0.0767 | 7 | 0.0662 |
| | | | | 3 | 0.2761 | 3 | 0.1372 | 5 | 0.0982 | 10 | 0.0978 |
| 11 | 3 > 1 | 63.63% | 27.27% | 5 | 0.0438 | 11 | 0.0482 | 29 | 0.0514 | 59 | 0.0498 |
| | | | | 6 | 0.0627 | 13 | 0.0662 | 33 | 0.0672 | 67 | 0.0668 |
| | | | | 7 | 0.0846 | 15 | 0.0999 | 37 | 0.0884 | 75 | 0.0879 |
| | 3 > 2 | 63.63% | 45.45% | 5 | 0.0476 | 11 | 0.0513 | 29 | 0.0499 | 59 | 0.0540 |
| | | | | 6 | 0.0720 | 13 | 0.0777 | 33 | 0.0720 | 67 | 0.0769 |
| | | | | 7 | 0.1056 | 15 | 0.1056 | 37 | 0.1003 | 75 | 0.1018 |
| | 4 > 2 | 81.81% | 45.45% | 2 | 0.0510 | 4 | 0.0470 | 12 | 0.0543 | 23 | 0.0519 |
| | | | | 3 | 0.0989 | 5 | 0.0661 | 14 | 0.0743 | 29 | 0.0776 |
| | | | | 4 | 0.1456 | 6 | 0.0930 | 16 | 0.0947 | 35 | 0.1099 |
| | 4 > 1 | 81.81% | 27.27% | 1 | 0.0195 | 4 | 0.0491 | 12 | 0.0547 | 23 | 0.0507 |
| | | | | 2 | 0.0534 | 5 | 0.0658 | 14 | 0.0771 | 29 | 0.0758 |
| | | | | 3 | 0.0970 | 6 | 0.0911 | 16 | 0.0984 | 35 | 0.1047 |
| 25 | 2 > 1 | 20% | 12% | 22 | 0.0523 | 46 | 0.0544 | 116 | 0.0528 | 234 | 0.0534 |
| | | | | 23 | 0.0772 | 48 | 0.0736 | 122 | 0.0701 | 246 | 0.0761 |
| | | | | 24 | 0.0983 | 50 | 0.1009 | 128 | 0.1036 | 258 | 0.1029 |
| | 3 > 2 | 28% | 20% | 19 | 0.0542 | 39 | 0.0510 | 100 | 0.0507 | 202 | 0.0504 |
| | | | | 20 | 0.0755 | 41 | 0.0706 | 105 | 0.0741 | 213 | 0.0708 |
| | | | | 21 | 0.1046 | 43 | 0.1016 | 110 | 0.1003 | 224 | 0.0990 |
| | 5 > 1 | 44% | 12% | 14 | 0.0526 | 29 | 0.0506 | 74 | 0.0531 | 150 | 0.0519 |
| | | | | 15 | 0.0837 | 31 | 0.0795 | 79 | 0.0761 | 160 | 0.0777 |
| | | | | 16 | 0.1030 | 33 | 0.0997 | 84 | 0.1025 | 170 | 0.1019 |
| | 5 > 3 | 44% | 28% | 13 | 0.0459 | 28 | 0.0500 | 72 | 0.0503 | 146 | 0.0517 |
| | | | | 14 | 0.0655 | 30 | 0.0736 | 77 | 0.0717 | 156 | 0.0714 |
| | | | | 15 | 0.0919 | 32 | 0.0980 | 82 | 0.0987 | 166 | 0.1027 |
| | 6 > 2 | 44% | 20% | 11 | 0.0531 | 23 | 0.0536 | 60 | 0.0530 | 118 | 0.0516 |
| | | | | 12 | 0.0794 | 25 | 0.0773 | 64 | 0.0700 | 128 | 0.0725 |
| | | | | 13 | 0.1049 | 27 | 0.1061 | 68 | 0.0970 | 138 | 0.1002 |
| | 6 > 4 | 44% | 36% | 11 | 0.0560 | 22 | 0.0490 | 58 | 0.0497 | 118 | 0.0503 |
| | | | | 12 | 0.0815 | 24 | 0.0661 | 63 | 0.0707 | 127 | 0.0726 |
| | | | | 13 | 0.1124 | 26 | 0.0993 | 68 | 0.1042 | 136 | 0.1038 |
| | 7 > 3 | 60% | 28% | 8 | 0.0459 | 17 | 0.0477 | 46 | 0.0498 | 93 | 0.0538 |
| | | | | 9 | 0.0682 | 19 | 0.0646 | 50 | 0.0752 | 102 | 0.0746 |
| | | | | 10 | 0.0999 | 21 | 0.0964 | 54 | 0.0992 | 111 | 0.1031 |
| | 7 > 5 | 60% | 44% | 8 | 0.0508 | 17 | 0.0499 | 45 | 0.0500 | 92 | 0.0528 |
| | | | | 9 | 0.0807 | 19 | 0.0694 | 49 | 0.0719 | 100 | 0.0747 |
| | | | | 10 | 0.1023 | 21 | 0.1039 | 53 | 0.1011 | 108 | 0.1005 |

Πίνακας 6.6: Πιθανότητες Λανθασμένου Συναγερμού FAR_{04} για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r_1 < s_1, r_2 > s_2$)

6.5 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

Στην ενότητα αυτή, εξετάζεται η απόδοση του διαγράμματος $O4$, όταν αυτό συγκρίνεται με άλλα ανταγωνιστικά σχήματα που έχουν προταθεί στη βιβλιογραφία. Αρχικά, η σύγκριση γίνεται με δύο παραμετρικά σχήματα, τη μονόπλευρη εκδοχή του κλασσικού δισδιάστατου διαγράμματος $|S|$ Φάσης II του Alt (1985), καθώς και με το πολυδιάστατο διάγραμμα $Rmax$ που αναπτύχθηκε από τους Costa and Machado (2009). Επιπλέον, η απόδοση του νέου διαγράμματος εξετάζεται και σε σχέση με τέσσερα δισδιάστατα μη παραμετρικά διαγράμματα ελέγχου, τα οποία μελετήθηκαν πρόσφατα από τους Osei-Aning *et al.* (2017). Θα πρέπει να αναφερθεί ότι όλα τα προαναφερθέντα ανταγωνιστικά σχήματα είναι ευαίσθητα στην ανίχνευση πιθανών μετατοπίσεων της μεταβλητότητας μόνο.

Για τις αριθμητικές συγκρίσεις, δημιουργήθηκαν δισδιάστατα δεδομένα από την κανονική κατανομή και την κατανομή t με 5 β.ε. Οι εντός ελέγχου τιμές των παραμέτρων είναι $\rho_{in} = 0.5$ ($= \rho_{out}$), $\mu_X^{(R)} = \mu_Y^{(R)} = 0$ και $\sigma_X (= \sigma_X^{(R)}) = \sigma_Y (= \sigma_Y^{(R)}) = 1$. Θα πρέπει να αναφερθεί ότι ο πίνακας συνδιακυμάνσεων $\Sigma_{t(5)}$ αρχικά κανονικοποιήθηκε κατάλληλα έτσι, ώστε να προκύπτει ο ίδιος πίνακας διακυμάνσεων-συνδιακυμάνσεων και για τις δύο κατανομές. Με βάση ένα κοινό ARL_{in} ίσο με 200, το εκτός ελέγχου ARL (ARL_{out}) εκτιμήθηκε για όλα τα διαγράμματα¹, μετατοπίζοντας τον πίνακα διακυμάνσεων-συνδιακυμάνσεων από

$$\Sigma_{in} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \text{σε} \quad \Sigma_{out} = \begin{pmatrix} \delta_X^2 \sigma_X^2 & \delta_X \delta_Y \sigma_{XY} \\ \delta_X \delta_Y \sigma_{XY} & \delta_Y^2 \sigma_Y^2 \end{pmatrix},$$

όπου τα δ_X, δ_Y ελέγχουν τις μετατοπίσεις της μεταβλητότητας των χαρακτηριστικών X και Y , αντίστοιχα.

Η μελέτη προσομοίωσης που πραγματοποιήθηκε αποκαλύπτει ότι το διάγραμμα $|S|$ έχει καλύτερη απόδοση σε σχέση με τα υπόλοιπα διαγράμματα, όταν προκαλούνται ίσες μετατοπίσεις και στα δύο χαρακτηριστικά, δεδομένου ότι η υποκείμενη κατανομή είναι κανονική. Επιπλέον, το διάγραμμα $Smax$ ανιχνεύει τις μεταβολές πιο γρήγορα από τα υπόλοιπα διαγράμματα. Ωστόσο, όταν χρησιμοποιούνται μη κανονικά δεδομένα, το διάγραμμα $|S|$ είναι πιο ευαίσθητο μόνο σε μεγαλύτερες μετατοπίσεις, ενώ το προτεινόμενο διάγραμμα $O4$

¹ Τα βήματα της αλγοριθμικής διαδικασίας σύμφωνα με την οποία υπολογίζονται το εντός και το εκτός ελέγχου μέσο μήκος ροής είναι παρόμοια με αυτά που παρουσιάζονται στον Πίνακα 4.4, με τη διαφορά ότι στο Βήμα 5 διαπιστώνεται εάν παραβιάζεται ή όχι ο κανόνας ελέγχου της Σχέσης (6.3).

είναι πιο αποτελεσματικό έναντι των μη παραμετρικών ομολόγων του. Αξίζει να σημειωθεί ότι η απόδοση του προαναφερθέντος διαγράμματος μπορεί να βελτιωθεί με τη χρήση μεγαλύτερων μεγεθών του δείγματος αναφοράς, καθώς τα όρια ελέγχου που προκύπτουν είναι πιο ακριβή με αποτέλεσμα να κάνουν τον κανόνα ελέγχου πιο ανθεκτικό (*robust*). Επιπρόσθετα, ο κανόνας απόφασης που χρησιμοποιείται στο διάγραμμα *O4* μπορεί να θεωρηθεί αυστηρός, διότι ελέγχει μόνο εάν τα κεντρικά ζευγάρια βρίσκονται εντός των ορίων σε ένα δείγμα ελέγχου μεγέθους $n = 5$.

Πρέπει, επίσης, να υπογραμμιστεί ότι τα όρια ελέγχου του νέου διαγράμματος *O4* μπορούν να επιλεγούν ασύμμετρα σε σχέση με τα παρακολουθούμενα ποιοτικά χαρακτηριστικά, σε αντίθεση με τα υπόλοιπα διαγράμματα στα οποία δεν υπάρχει παρόμοια δυνατότητα. Αυτό παρέχει πρόσθετη ευελιξία για την αντιμετώπιση περιπτώσεων στις οποίες πρέπει να εντοπιστούν άνισες μετατοπίσεις. Για παράδειγμα, όταν η εκτός ελέγχου κατάσταση οφείλεται σε μεγαλύτερες μετατοπίσεις του πρώτου χαρακτηριστικού, είναι εύλογο να επιλέξουμε τις παραμέτρους έτσι, ώστε $r_1 < s_1$, $r_2 > s_2$. Σε αυτές τις περιπτώσεις, όταν τα δεδομένα προέρχονται από κανονική κατανομή, η απόδοση του προτεινόμενου διαγράμματος είναι στις περισσότερες περιπτώσεις καλύτερη από την απόδοση των υπόλοιπων διαγραμμάτων, με το *Smax* να έχει παρόμοια απόδοση μόνο για μεγάλες μετατοπίσεις. Αντίθετα, η απόδοση του διαγράμματος *O4* είναι ομοιόμορφα η καλύτερη για δεδομένα από την κατανομή $t_{(5)}$. Τα αποτελέσματα συνοψίζονται στους Πίνακες 6.7 και 6.8, όπου οι τιμές με έντονους χαρακτήρες δείχνουν τη μικρότερη τιμή του ARL_{out} , υποδεικνύοντας το διάγραμμα με την ταχύτερη ικανότητα ανίχνευσης μετατοπίσεων. Το μέγεθος δείγματος είναι κοινό σε όλα τα διαγράμματα ($n = 5$), οι παράμετροι σχεδιασμού για τον κανόνα ελέγχου *O4* είναι $b = m - a + 1$, $d = m - c + 1$, $r_1 = (n + 1)/2 - t_X$, $s_1 = (n + 1)/2 - t_Y$, $r_2 = (n + 1)/2 + t_X$, $s_2 = (n + 1)/2 + t_Y$, ενώ το μέγεθος του δείγματος αναφοράς έχει προσδιορισθεί, ώστε το ARL_{in} να είναι περίπου ίσο με 200.

Στο σημείο αυτό, θα πρέπει να σημειωθεί ότι δε γίνονται στην παρούσα ενότητα περαιτέρω συγκρίσεις για την ανίχνευση πιθανών μετατοπίσεων της μέσης τιμής. Εναλλακτικά, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην Ενότητα 4.6, όπου γίνονται συγκρίσεις του Κανόνα 4.1 (ο οποίος ταυτίζεται με τον κανόνα του διαγράμματος *O4* για $t = 1$) με υπάρχοντα παραμετρικά και μη παραμετρικά γραφήματα αποτελεσματικά μόνο στην ανίχνευση μετατοπίσεων στη θέση της εντός ελέγχου κατανομής. Παρόμοια αποτελέσματα μπορούν να αναπαραχθούν χρησιμοποιώντας το νέο διάγραμμα για μικρές τιμές της παραμέτρου t .

| Μεταπολίσεις Τυπ. Απόκλ. | Παραμετρικά Διαγράμματα | | Μη Παραμετρικά Διαγράμματα | | | Ημιπαραμετρικό Διάγραμμα | |
|-----------------------------|-------------------------|---------------|----------------------------|---------------|---------------|--------------------------|---|
| | S | RMAX | QMAX | S MAX | MDMAX | MADMAX | O4 |
| $\delta_X = \delta_Y$ | $UCL = 4.0314$ | $UCL = 5.145$ | $UCL = 2.238$ | $UCL = 2.024$ | $UCL = 1.476$ | $UCL = 2.567$ | $m = 245$ $a = c = 3$ $t = t_X = t_Y = 1$ |
| 1.00 | 200.92 | 199.78 | 200.40 | 201.84 | 200.72 | 201.71 | 200.28 |
| 1.25 | 15.45 | 17.93 | 28.30 | 16.16 | 17.88 | 30.13 | 24.18 |
| 1.50 | 4.53 | 5.13 | 9.94 | 4.65 | 5.32 | 10.15 | 7.66 |
| 1.75 | 2.37 | 2.64 | 4.96 | 2.44 | 2.65 | 5.15 | 3.88 |
| 2.00 | 1.67 | 1.80 | 3.21 | 1.69 | 1.80 | 3.39 | 2.52 |
| $\delta_X > \delta_Y$ | $UCL = 4.0314$ | $UCL = 5.145$ | $UCL = 2.238$ | $UCL = 2.024$ | $UCL = 1.476$ | $UCL = 2.567$ | $m = 535$ $a = c = 9$ $t_X = 1 > 0 = t_Y$ |
| 1.00 | 200.60 | 199.18 | 199.36 | 200.16 | 200.41 | 201.99 | 200.40 |
| 1.25 | 46.43 | 31.49 | 48.91 | 28.65 | 31.91 | 29.85 | 29.66 |
| 1.50 | 18.66 | 9.06 | 17.62 | 8.15 | 9.20 | 10.25 | 9.87 |
| 1.75 | 9.82 | 4.34 | 8.79 | 3.95 | 4.41 | 5.32 | 5.23 |
| 2.00 | 6.27 | 2.75 | 5.53 | 2.83 | 3.67 | 2.61 | 3.47 |
| 2.25 | 4.50 | 2.05 | 4.04 | 1.93 | 2.09 | 2.52 | 2.55 |
| 2.50 | 3.46 | 1.69 | 3.21 | 1.61 | 1.72 | 2.05 | 2.11 |

Πίνακας 6.7: Απόδοση του διαγράμματος O4 βάσει του ARL_{out} σε σχέση με 6 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μεταπολίσεων τυπικής απόκλισης χρησιμοποιώντας δεδομένα από κανονική κατανομή για δεδομένο σχεδιασμό

6.5. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

| Μεταπολίσεις Τυπ. Απόκλ. | Παραμετρικά Διαγράμματα | | Μη Παραμετρικά Διαγράμματα | | | | Ημιπαραμετρικό Διάγραμμα | |
|-----------------------------|-------------------------|---------------|----------------------------|---------------|---------------|---------------|--------------------------|---|
| | S | RMAX | QMAX | S MAX | MDMAX | MADMAX | O4 | |
| $\delta_X = \delta_Y$ | $UCL = 7.512$ | $UCL = 7.320$ | $UCL = 2.206$ | $UCL = 2.907$ | $UCL = 1.809$ | $UCL = 2.594$ | | $m = 245$ $a = c = 3$ $t = t_X = t_Y = 1$ |
| 1.00 | 200.71 | 199.76 | 200.84 | 201.21 | 200.36 | 200.49 | 201.86 | |
| 1.25 | 38.31 | 59.00 | 41.04 | 59.75 | 45.82 | 41.98 | 50.35 | |
| 1.50 | 12.23 | 21.93 | 14.85 | 23.68 | 15.78 | 15.67 | 17.83 | |
| 1.75 | 5.78 | 10.94 | 7.45 | 11.09 | 7.35 | 7.90 | 8.92 | |
| 2.00 | 3.42 | 6.25 | 4.55 | 6.30 | 4.25 | 4.93 | 5.27 | |
| $\delta_X > \delta_Y$ | $UCL = 7.512$ | $UCL = 7.320$ | $UCL = 2.206$ | $UCL = 2.907$ | $UCL = 1.809$ | $UCL = 2.594$ | | $m = 535$ $a = c = 9$ $t_X = 4 > 2 = t_Y$ |
| 1.00 | 200.62 | 200.67 | 200.42 | 199.41 | 201.32 | 199.06 | 201.41 | |
| 1.25 | 84.79 | 85.68 | 67.59 | 89.44 | 71.59 | 68.45 | 51.43 | |
| 1.50 | 43.80 | 37.39 | 26.38 | 38.73 | 26.60 | 27.64 | 19.70 | |
| 1.75 | 26.11 | 18.41 | 13.29 | 18.94 | 12.31 | 14.31 | 10.09 | |
| 2.00 | 17.14 | 10.44 | 8.08 | 10.45 | 6.91 | 8.84 | 6.23 | |
| 2.25 | 12.20 | 6.58 | 5.62 | 6.68 | 4.53 | 6.07 | 4.38 | |
| 2.50 | 9.11 | 4.30 | 4.30 | 4.60 | 3.32 | 4.55 | 3.32 | |

Πίνακας 6.8: Απόδοση του διαγράμματος O4 βάσει του ARL_{out} σε σχέση με 6 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μεταπολίσεων τυπικής απόκλισης δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό

Τέλος, το νέο διάγραμμα ελέγχου συγκρίνεται με δύο ακόμη διαγράμματα, τα οποία είναι ικανά να ανιχνεύουν ταυτόχρονες μετατοπίσεις της μέσης τιμής και της μεταβλητότητας της διεργασίας. Ειδικότερα, γίνεται χρήση του δισδιάστατου παραμετρικού διαγράμματος ελέγχου *BV MAX* του Khoo (2004) και το μη παραμετρικό διάγραμμα *Q* για ομαδοποιημένες παρατηρήσεις της Liu (1995), το οποίο κάνει χρήση του βάθους του *Mahalanobis*. Δεδομένου ότι το διάγραμμα *Q* βασίζεται στο βάθος δεδομένων, απαιτείται σχετικά μεγάλη υπολογιστική ισχύς για την εκτίμηση του *ARL*, κυρίως στην περίπτωση μεγάλων μεγεθών δείγματος. Οι Stoumbos *et al.* (2001) μελέτησαν τις απαιτήσεις του δείγματος αναφοράς ώστε να γίνεται σωστή επιλογή των ορίων ελέγχου.

Για να δημιουργήσουμε μία δίκαιη μελέτη σύγκρισης, εκμεταλλευτήκαμε ένα από τα σενάρια που εφαρμόστηκαν από τους Zamba and Hawkins (2009) για να ελέγξουν την αποτελεσματικότητα ενός σχήματος για την ανίχνευση του σημείου, όπου παρατηρείται η αλλαγή στη μέση τιμή ή/και στη διακύμανση. Σύμφωνα με αυτό το σενάριο, ο πίνακας διακυμάνσεων-συνδυακυμάνσεων της μορφής

$$\Sigma = \begin{pmatrix} 1 + \rho^2 & \rho^2 \\ \rho^2 & 1 + \rho^2 \end{pmatrix} = (1 + \rho^2) \begin{pmatrix} 1 & t^2 \\ t^2 & 1 \end{pmatrix},$$

όπου $t^2 = \rho^2 / (1 + \rho^2)$ και με ρ συμβολίζεται ο συντελεστής συσχέτισης μεταξύ των μεταβλητών.

Στη μελέτη μας, για την απόδοση ενός διαγράμματος με βάση *ARL*, υποθέτουμε ότι τα χαρακτηριστικά ποιότητας ήταν αρχικά ασυσχέτιστα, δηλαδή $\rho = 0$ και η μέση τιμή της διεργασίας ήταν $\mu_X = \mu_Y = 0$ (εντός ελέγχου κατάσταση). Στη συνέχεια εξετάστηκαν αρκετές μετατοπίσεις της μέσης τιμής σε συνδυασμό με διάφορες επιλογές του ρ που έχουν ως αποτέλεσμα αλλαγές σε όλες τις θέσεις του πίνακα Σ , δηλαδή αλλαγές στη διακύμανση και συσχέτιση των υπό μελέτη χαρακτηριστικών.

Για άλλη μια φορά, στο διάγραμμα *O4* χρησιμοποιούνται οι διάμεσοι, και έτσι επιλέγονται συμμετρικά όρια ελέγχου. Το μέγεθος του δείγματος ελέγχου είναι $n = 5$ σε όλες τις περιπτώσεις, ενώ χρησιμοποιείται ένα κατάλληλο δείγμα αναφοράς για κάθε γράφημα με σκοπό την επίτευξη του επιθυμητού ARL_{in} . Από τα αποτελέσματα που παρουσιάζονται στον Πίνακα 6.9 είναι σαφές ότι, όπως αναμενόταν, το παραμετρικό διάγραμμα έχει καλύτερη απόδοση, δεδομένου ότι τα δεδομένα που χρησιμοποιούνται σε αυτή την περίπτωση προέρχονται από κανονική κατανομή. Ωστόσο, συγκρίνοντας το νέο διάγραμμα *O4* με το διάγραμμα *Q*, εύκολα μπορεί να επιβεβαιωθεί ότι γίνονται αρκετά ανταγωνιστικά. Θα

6.5. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

πρέπει να σημειωθεί ότι με έντονη γραφή σημειώνεται η μικρότερη τιμή του ARL_{out} συγκρίνοντας και τα τρία σχήματα, ενώ η καλύτερη τιμή του ARL_{out} απλά υπογραμμίζεται, όταν συγκρίνονται μεταξύ τους μόνο τα δύο μη παραμετρικά διαγράμματα.

Λαμβάνοντας υπ' όψιν μας το γεγονός ότι η εφαρμογή των νέων διαγραμμάτων ελέγχου είναι πολύ εύκολη σε αντίθεση με αυτή του διαγράμματος Q που απαιτεί χρονοβόρους υπολογιστικούς χειρισμούς, μπορεί κανείς να δηλώσει, ότι από πρακτική άποψη το προτεινόμενο διάγραμμα αποτελεί χρήσιμο εργαλείο για την παρακολούθηση δισδιάστατων χαρακτηριστικών χωρίς να γίνεται καμία υπόθεση σχετικά με τις περιθώριες κατανομές.

| Μετατοπίσεις | | Διάγραμμα BV MAX | | | | |
|-----------------------|--------------------|------------------|--------------|--------------|--------------|--------------|
| Μέσης Τιμής | Σχεδιασμός | ρ | | | | |
| $\mu = \mu_X = \mu_Y$ | $ARL_{in} = 200.9$ | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 |
| 0.00 | $UCL = 3.023$ | 195.11 | 169.52 | 104.36 | 49.12 | 22.58 |
| 0.25 | $n = 5$ | 106.94 | 86.03 | 51.50 | 29.42 | 15.79 |
| 0.50 | | 26.20 | 22.38 | 15.86 | 11.27 | 7.90 |
| 0.75 | | 7.00 | 6.53 | 5.60 | 4.78 | 4.03 |
| 1.00 | | 2.71 | 2.66 | 2.59 | 2.51 | 2.35 |

| Μετατοπίσεις | | Διάγραμμα Q | | | | |
|-----------------------|--------------------|--------------|---------------|--------------|--------------|--------------|
| Μέσης Τιμής | Σχεδιασμός | ρ | | | | |
| $\mu = \mu_X = \mu_Y$ | $ARL_{in} = 198.9$ | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 |
| 0.00 | $UCL = 0.18$ | 190.84 | 154.94 | 83.30 | <u>38.32</u> | <u>18.08</u> |
| 0.25 | $n = 5$ | 126.35 | 110.39 | 65.08 | <u>32.03</u> | <u>16.25</u> |
| 0.50 | $m = 100$ | 50.95 | 47.54 | 33.18 | 19.91 | 11.73 |
| 0.75 | $a = 0.005$ | <u>16.83</u> | <u>16.68</u> | 14.29 | 10.59 | 7.70 |
| 1.00 | | <u>6.49</u> | <u>6.47</u> | 6.41 | 5.72 | 4.87 |

| Μετατοπίσεις | | Διάγραμμα O4 | | | | |
|-----------------------|--------------------|---------------|---------------|--------------|--------------|--------------|
| Μέσης τιμής | Σχεδιασμός | ρ | | | | |
| $\mu = \mu_X = \mu_Y$ | $ARL_{in} = 200.2$ | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 |
| 0.00 | $a = c = 3$ | 188.64 | 158.66 | 90.47 | 43.58 | 20.91 |
| 0.25 | $t = 1$ | <u>124.07</u> | <u>105.82</u> | <u>63.63</u> | <u>32.03</u> | 16.52 |
| 0.50 | $n = 5$ | <u>48.10</u> | <u>41.93</u> | <u>28.26</u> | <u>16.76</u> | <u>10.11</u> |
| 0.75 | $m = 255$ | 18.50 | 16.90 | <u>12.41</u> | <u>8.45</u> | <u>5.92</u> |
| 1.00 | | 7.62 | 7.24 | <u>5.95</u> | <u>4.52</u> | <u>3.66</u> |

Πίνακας 6.9: Απόδοση του διαγράμματος O4 βάσει του ARL_{out} σε σχέση με 2 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση ταυτόχρονων μετατοπίσεων μέσης τιμής και μεταβλητότητας χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό

6.6 Εφαρμογή σε πραγματικά δεδομένα

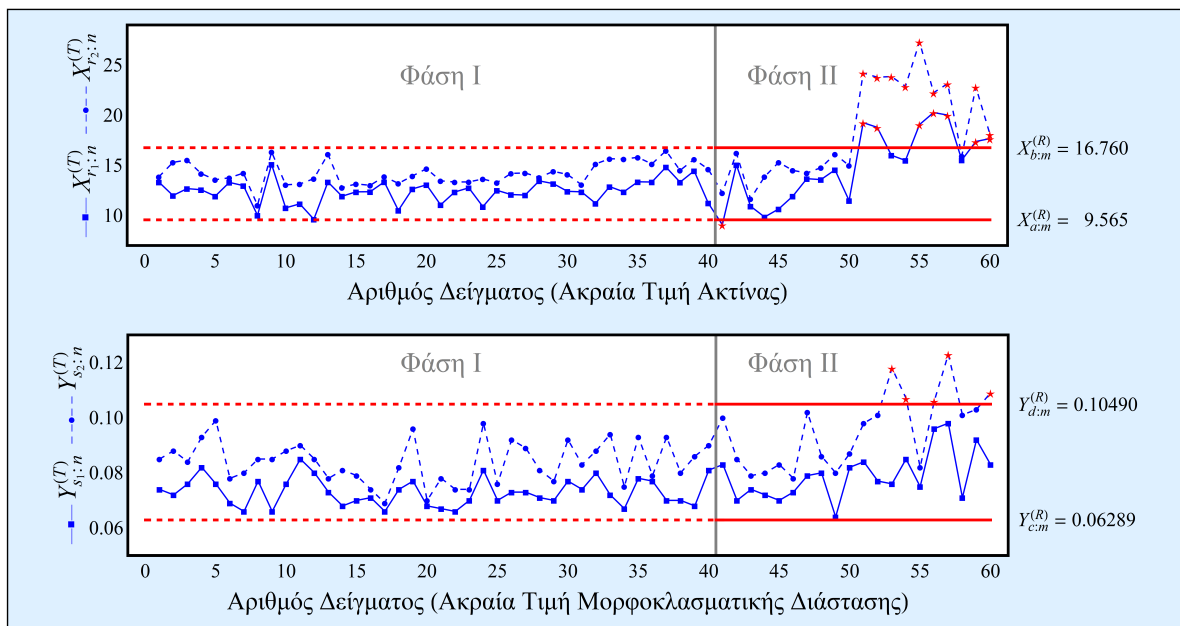
Κατά τη διάρκεια των τελευταίων ετών, εφαρμόστηκαν επιτυχώς σε μελέτες περιπτώσεων που σχετίζονταν με παρακολούθηση ασθενών στατιστικές τεχνικές, οι οποίες κάνουν χρήση διαγραμμάτων ελέγχου. Η χρήση, καθώς και η αναγκαιότητα εφαρμογής των μεθόδων αυτών για την παρακολούθηση και τη βελτίωση της ποιότητας σε ένα κλινικό περιβάλλον (*clinical setting*) και την υγειονομική περίθαλψη (*healthcare*) έχει ενδεικτικά υπογραμμιστεί από τους Carey (2003), Woodall (2006), Waterhouse *et al.* (2010), Woodall *et al.* (2012), Woodall and Montgomery (2014).

Προκειμένου να επεξηγηθεί η εφαρμογή του προτεινόμενου διαγράμματος ελέγχου, χρησιμοποιείται ένα πραγματικό σύνολο δεδομένων από μία διαγνωστική μελέτη καρκίνου του μαστού (Wisconsin Diagnostic Cancer Breast, WDBC), η οποία είναι διαθέσιμη για ερευνητική χρήση στο UC Irvine Machine Learning Repository και μπορεί να βρεθεί ηλεκτρονικά στην ιστοσελίδα [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)). Το σύνολο δεδομένων WDBC χρησιμοποιήθηκε για πρώτη φορά από τους Street *et al.* (1993) (βλ. επίσης την εργασία των Mangasarian *et al.* (1995)), οι οποίοι εφάρμοσαν μια διαδικασία ταξινόμησης για να διαχωρίσουν τις καλοήθειες από τις κακοήθειες παρατηρήσεις. Το ίδιο σύνολο δεδομένων αναλύθηκε αργότερα από τους Sukchotrat *et al.* (2008), οι οποίοι εφάρμοσαν μια διαδικασία ταξινόμησης που κάνει χρήση ενός διαγράμματος ελέγχου.

Τα δεδομένα χωρίζονται σε δύο κατηγορίες παρατηρήσεων ανάλογα με τη διάγνωση: 212 κακοήθειες (*Malignant*) και 357 καλοήθειες παρατηρήσεις (*Benign*). Επιπλέον, τα δεδομένα αποτελούνται από δέκα μεταβλητές: ακτίνα (*radius*), περίμετρος (*perimeter*), περιοχή (*area*), συμπάγεια/συμπαγότητα (*compactness*), ομαλότητα (*smoothness*), κοιλότητα (*concavity*), κοίλα σημεία (*concave points*), συμμετρία (*symmetry*), μορφοκλασματική διάσταση (*fractal dimension*) και υφή (*texture*), για τις οποίες υπολογίστηκαν ο μέσος όρος, το τυπικό σφάλμα και οι χειρότερες (ακραίες) τιμές για κάθε πυρήνα κυττάρων. Λεπτομερής περιγραφή των δεδομένων γίνεται στην εργασία των Street *et al.* (1993), σύμφωνα με τους οποίους το ζευγάρι μεταβλητών με το καλύτερο ποσοστό ορθής διάγνωσης (79.8%) είναι η ακραία τιμή ακτίνας (*worst radius*) και η ακραία τιμή μορφοκλασματικής διάστασης (*worst fractal dimension*). Οι μεταβλητές αυτές χρησιμοποιούνται ως χαρακτηριστικά παρακολούθησης του νέου διαγράμματος ελέγχου, προκειμένου να εξετάσουμε κατά πόσο διακρίνονται σωστά οι καλοήθειες όγκοι από τους κακοήθειες.

6.6. Εφαρμογή σε πραγματικά δεδομένα

Χωρίς βλάβη της γενικότητας, θεωρούμε ότι ένα μέρος από τις καλοήθειες παρατηρήσεις χρησιμοποιείται ως το δείγμα Φάσης I (δείγμα αναφοράς) βάσει του οποίου καθορίζονται τα όρια ελέγχου. Έπειτα, συλλέγονται τα εκτός ελέγχου δείγματα και τα όρια επαναξιολογούνται έως ότου τα υπόλοιπα δείγματα να μην ενεργοποιούν συναγερμό. Στην ανάλυσή μας, τα δείγματα ομαδοποιούνται σε υποομάδες μεγέθους $n = 5$. Επιπλέον, υποθέτουμε ότι θέλουμε να κατασκευάσουμε το διάγραμμα με $FAR_{O4} = 5\%$ έχοντας στη διάθεσή μας ένα δείγμα αναφοράς μεγέθους $m = 200 (= 40 \times 5)$. Αξιοποιώντας τον Πίνακα 6.4, μπορούμε να επιλέξουμε κατάλληλες παραμέτρους σχεδιασμού με τιμές $a = c = 7$, $b = d = 200 - 7 + 1 = 194$, $r_1 = s_1 = (n + 1)/2 - t = 2$, $r_2 = s_2 = (n + 1)/2 + t = 4$ για $t = 1$. Στη συνέχεια, συλλέγονται 10 επιπλέον καλοήθειες και 10 κακοήθειες παρατηρήσεις, οι οποίες συνιστούν το δείγμα Φάσης II (δείγμα ελέγχου). Τα αποτελέσματα της παραπάνω ανάλυσης παρουσιάζονται στο Σχήμα 6.1.



Σχήμα 6.1: Εφαρμογή του διαγράμματος $O4$ στο σύνολο δεδομένων WDBC

Από το Σχήμα 6.1 είναι εμφανές ότι μόνο μία από τις καλοήθειες παρατηρήσεις βρίσκεται εκτός ελέγχου και μία από τις κακοήθειες τοποθετείται εντός των ορίων. Πιο συγκεκριμένα, ενεργοποιείται συναγερμός στο 41ο δείγμα (που αντιστοιχεί σε καλοήθη όγκο) λόγω της τιμής που λαμβάνει η ακραία μορφοκλασματική διάσταση. Αντίθετα, το 58ο κακοήθες δείγμα θεωρείται ότι βρίσκεται εντός ελέγχου. Τα υπόλοιπα κακοήθη δείγματα είναι εκτός ελέγχου λόγω μετατόπισης που παρατηρείται είτε στο πρώτο είτε και στα δύο χαρακτηριστικά.

Η ανάλυσή μας αποκαλύπτει ότι παρόλο που υποθέτουμε ύπαρξη ίσων μετατοπίσεων και στα δύο χαρακτηριστικά, συναγεμρός φαίνεται να προκαλείται εξαιτίας του πρώτου χαρακτηριστικού και όχι λόγω του δεύτερου. Έτσι, θα μπορούσαμε εναλλακτικά να εξετάσουμε μία ενδεχομένως πιο λογική επιλογή των παραμέτρων σχεδιασμού με $r_1 \neq s_1$, $r_2 \neq s_2$ ή $a \neq c$. Πιο συγκεκριμένα, εάν είχαμε ενδείξεις για μετατοπίσεις της μέσης τιμής του πρώτου χαρακτηριστικού, θα ήταν πιο λογικό να επιλέξουμε $a > c$, $r_1 = s_1$, $r_2 = s_2$, ενώ εάν η εκτός ελέγχου κατάσταση οφειλόταν σε αποκλίσεις της μεταβλητότητας, θα ήταν πιο ενδεδειγμένη η χρήση $a = c$, $r_1 < s_1$, $r_2 > s_2$. Σε κάθε μία από αυτές τις περιπτώσεις, ανάλογα φυσικά με το είδος της μετατόπισης που θέλουμε να ανιχνεύσουμε, τα αποτελέσματα που προκύπτουν είναι ελαφρώς καλύτερα.

ΚΕΦΑΛΑΙΟ 7

Δισδιάστατο Διάγραμμα Ελέγχου για τη Θέση και τη Διασπορά με Χρήση Απαριθμητριών Στατιστικών Συναρτήσεων

7.1 Εισαγωγή

Ένα από τα πρώτα μονοδιάστατα μη παραμετρικά διαγράμματα ελέγχου για την ανίχνευση πιθανών μετατοπίσεων στη μέση τιμή ή/και στη διακύμανση μιας διεργασίας είναι αυτό που μελετήθηκε από τους Balakrishnan *et al.* (2010). Το διάγραμμα αυτό θεωρείται βελτίωση των διαγραμμάτων προτεραιότητας (βλ. Ενότητα 4.1), δηλαδή του κλασσικού διαγράμματος της διαμέσου των Janacek and Meikle (1997), καθώς και της κλάσης διαγραμμάτων για την παρακολούθηση ενός ποσοστημορίου που πρότειναν οι Chakraborti *et al.* (2000, 2004).

Τα όρια ελέγχου του διαγράμματος των Balakrishnan *et al.* (2010) είναι και πάλι δύο διατεταγμένες παρατηρήσεις του δείγματος αναφοράς, έστω

$$LCL = X_{a:m}^{(R)} \quad \text{και} \quad UCL = X_{b:m}^{(R)},$$

όπου $1 \leq a < b \leq m$. Ο κανόνας απόφασης διατυπώνεται κάνοντας χρήση της j -οστής διατεταγμένης παρατήρησης $X_{j:n}^{(T)}$ σε συνδυασμό με την ακόλουθη σ.σ.

$$R = R(X_1^{(T)}, X_2^{(T)}, \dots, X_n^{(T)}; X_{a:m}^{(R)}, X_{b:m}^{(R)}) = \left| \left\{ j \in \{1, 2, \dots, n\} : X_{a:m}^{(R)} \leq X_{j:n}^{(T)} \leq X_{b:m}^{(R)} \right\} \right|.$$

Αξίζει να σημειωθεί ότι η συνάρτηση R συμβολίζει απλά το πλήθος των παρατηρήσεων του δείγματος ελέγχου, οι οποίες βρίσκονται μεταξύ των ορίων ελέγχου που προσδιορίστηκαν από το δείγμα αναφοράς.

Ειδικότερα, η διεργασία κρίνεται εντός ελέγχου εάν επαληθεύονται οι επόμενες δύο συνθήκες

$$X_{a:m}^{(R)} \leq X_{j:n}^{(T)} \leq X_{b:m}^{(R)} \quad \text{και} \quad R \geq r.$$

Οι ποσότητες m, n, a, b, j, r είναι οι παράμετροι σχεδιασμού του διαγράμματος ελέγχου, οι τιμές των οποίων προσδιορίζονται βάσει του ποσοστού λανθασμένου συναγερμού

$$FAR = 1 - \sum_{r-1 \leq c+d \leq n-1} \frac{\binom{j-c+a-2}{a-1} \binom{m+n-b-d-j}{n-j-d} \binom{b+c+d-a}{c+d+1}}{\binom{m+n}{n}}$$

έτσι, ώστε να επιτευχθεί μία προκαθορισμένη τιμή του.

Το κύριο πλεονέκτημα του παραπάνω διαγράμματος είναι ότι ουσιαστικά λαμβάνεται υπ' όψιν όχι μόνο η θέση μιας συγκεκριμένης διατεταγμένης παρατήρησης του δείγματος ελέγχου, αλλά και το πλήθος των παρατηρήσεων που βρίσκονται μεταξύ των ορίων ελέγχου. Δεδομένου ότι η λογική κατασκευής του παραπάνω διαγράμματος ελέγχου είναι αρκετά απλή, μπορεί εύκολα να γενικευθεί για την παρακολούθηση περισσότερων ποιοτικών χαρακτηριστικών που είναι συσχετισμένα μεταξύ τους.

Στο τρέχον κεφάλαιο μελετάται ένα νέο ημιπαραμετρικό διάγραμμα ελέγχου για την ταυτόχρονη παρακολούθηση της μέσης τιμής και τη μεταβλητότητας μιας διεργασίας (Koutras and Sofikitou (2017c)), το οποίο αποτελεί άμεση γενίκευση του διαγράμματος των Balakrishnan *et al.* (2010) σε δύο διαστάσεις.

7.2 Περιγραφή ενός νέου διαγράμματος ελέγχου

Όπως έχει ήδη αναφερθεί στην εισαγωγή, στο παρόν κεφάλαιο γίνεται μελέτη ενός διαγράμματος ελέγχου για ανίχνευση πιθανών μετατοπίσεων στη μέση τιμή ή/και στη διασπορά μιας διεργασίας. Στην πραγματικότητα, ο έλεγχος υπόθεσης που γίνεται και σε αυτή την περίπτωση είναι ο ίδιος με αυτόν του προηγούμενου κεφαλαίου (βλ. Σχέση (6.1)). Ωστόσο, η σ.σ. ελέγχου που χρησιμοποιείται στο νέο προτεινόμενο διάγραμμα διαφοροποιείται.

Πιο συγκεκριμένα, στον Κανόνα 4.1 του διαγράμματος $O2$ προστίθεται μία ακόμη συνθήκη που εμπλέκει πλήθος παρατηρήσεων, οι οποίες προέρχονται από όλο το εύρος του αρχικού συνόλου παρατηρήσεων και όχι από μέρος αυτού. Ως αποτέλεσμα, η διεργασία θεωρείται ότι βρίσκεται εντός ελέγχου εάν μία διατεταγμένη παρατήρηση (για παράδειγμα ένα ποσοστημόριο ή συγκεκριμένα η διάμεσος) και ταυτόχρονα ένα προκαθορισμένο πλήθος παρατηρήσεων του δείγματος ελέγχου βρίσκεται μεταξύ των ορίων. Είναι περιττό να αναφερθεί ότι οι συνθήκες αυτές πρέπει να τηρούνται για κάθε ένα από τα προς μελέτη χαρακτηριστικά.

Κατ' αρχάς, ας υποθέσουμε ότι συλλέγεται ένα δείγμα αναφοράς $(X_1^{(R)}, Y_1^{(R)}), (X_2^{(R)}, Y_2^{(R)})$,

$\dots, (X_m^{(R)}, Y_m^{(R)})$ από μία εντός ελέγχου διεργασία με από κοινού α.σ.κ. $F_{X,Y}^{(R)}(x,y) = F(x,y)$ και περιθώριες $F_X(x), F_Y(y)$. Τα όρια ελέγχου, όπως και προηγουμένως, συνιστούν οι ακόλουθες διατεταγμένες παρατηρήσεις

$$LCL_X = X_{a:m}^{(R)}, \quad UCL_X = X_{b:m}^{(R)} \quad \text{και} \quad LCL_Y = Y_{c:m}^{(R)}, \quad UCL_Y = Y_{d:m}^{(R)},$$

όπου $1 \leq a < b \leq m$ και $1 \leq c < d \leq m$.

Έπειτα, θεωρούμε ότι έχουμε στη διάθεσή μας ένα δείγμα ελέγχου, έστω $(X_1^{(T)}, Y_1^{(T)})$, $(X_2^{(T)}, Y_2^{(T)})$, \dots , $(X_n^{(T)}, Y_n^{(T)})$ με από κοινού α.σ.κ. $F_{X,Y}^{(T)}(x,y) = G(x,y)$ και περιθώριες $G_X(x), G_Y(y)$, το οποίο χρησιμοποιούμε για να ελέγξουμε εάν η διεργασία παραμένει εντός ελέγχου. Πρώτα, όμως, ας ορίσουμε την απαριθμήτρια συνάρτηση

$$L(z_1, z_2, \dots, z_n; w_1, w_2) = \left| \{i \in \{1, 2, \dots, n\} : w_1 \leq z_i \leq w_2\} \right|,$$

η οποία για δεδομένα $z_1, z_2, \dots, z_n \in \mathbb{R}$ και $w_1, w_2 \in \mathbb{R}$, επιστρέφει τον αριθμό των z_i που βρίσκονται μεταξύ των τιμών w_1 και w_2 . Για την κατασκευή του καινούργιου σχήματος, θα αξιοποιήσουμε την r -οστή και s -οστή διατεταγμένη παρατήρηση του δείγματος ελέγχου (συμβ. $X_{r:n}^{(T)}$ και $Y_{s:n}^{(T)}$, αντίστοιχα, για $1 \leq r, s \leq n$) σε συνδυασμό με τις απαριθμήτριες συναρτήσεις που ακολουθούν

$$L_X = L(X_1^{(T)}, X_2^{(T)}, \dots, X_n^{(T)}; X_{a:m}^{(R)}, X_{b:m}^{(R)}) = \left| \{i \in \{1, 2, \dots, n\} : X_{a:m}^{(R)} \leq X_i^{(T)} \leq X_{b:m}^{(R)}\} \right|,$$

$$L_Y = L(Y_1^{(T)}, Y_2^{(T)}, \dots, Y_n^{(T)}; Y_{c:m}^{(R)}, Y_{d:m}^{(R)}) = \left| \{j \in \{1, 2, \dots, n\} : Y_{c:m}^{(R)} \leq Y_j^{(T)} \leq Y_{d:m}^{(R)}\} \right|.$$

Οι συναρτήσεις L_X και L_Y συμβολίζουν τον αριθμό των παρατηρήσεων X και Y , από όλο το εύρος του δείγματος ελέγχου, οι οποίες βρίσκονται μεταξύ των ορίων ελέγχου που προσδιορίστηκαν από το δείγμα αναφοράς.

Τελικά, το νέο διάγραμμα ελέγχου θα χαρακτηρίζει τη διεργασία εντός στατιστικού ελέγχου, εάν δεν παραβιάζεται καμία από τις παρακάτω συνθήκες

$$LCL_X \leq X_{r:n}^{(T)} \leq UCL_X, \quad LCL_Y \leq Y_{s:n}^{(T)} \leq UCL_Y \quad \text{και} \quad L_X \geq \ell_X, \quad L_Y \geq \ell_Y, \quad (7.1)$$

ή ισοδύναμα

$$\left\{ \begin{array}{l} X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \\ Y_{c:m}^{(R)} \leq Y_{r:n}^{(T)} \leq Y_{d:m}^{(R)} \end{array} \right\} \quad (7.1\alpha) \quad \text{και} \quad \left\{ \begin{array}{l} L_X \geq \ell_X \\ L_Y \geq \ell_Y \end{array} \right\} \quad (7.1\gamma)$$

$$\left\{ \begin{array}{l} X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)} \\ Y_{c:m}^{(R)} \leq Y_{r:n}^{(T)} \leq Y_{d:m}^{(R)} \end{array} \right\} \quad (7.1\beta) \quad \text{και} \quad \left\{ \begin{array}{l} L_X \geq \ell_X \\ L_Y \geq \ell_Y \end{array} \right\} \quad (7.1\delta)$$

Εναλλακτικά, μπορούμε να ισχυριστούμε ότι το διάγραμμα δεν παράγει συναγερομό όσο οι

διατεταγμένες παρατηρήσεις $X_{r:n}^{(T)}$ και $Y_{s:n}^{(T)}$ ανήκουν στην περιοχή ελέγχου και ταυτόχρονα τουλάχιστον ℓ_X και ℓ_Y από τις παρατηρήσεις X και Y του δείγματος ελέγχου τοποθετούνται στην ίδια περιοχή. Οι ποσότητες $m, n, a, b, c, d, r, s, \ell_X, \ell_Y$ είναι οι παράμετροι σχεδιασμού του διαγράμματος $O2N2$, η μεθοδολογία προσδιορισμού των οποίων αναλύεται στην Ενότητα 7.4.

Το νέο διάγραμμα θα καλείται στο εξής διάγραμμα $O2N2$ (*Order-Order & Number-Number*). Υπενθυμίζουμε ότι το προτεινόμενο διάγραμμα αποτελεί γενίκευση σε δύο διαστάσεις του μονοδιάστατου διαγράμματος που προτάθηκε από τους Balakrishnan *et al.* (2010), καθώς και ειδική περίπτωση του διαγράμματος $O2$, αφού οι κανόνες απόφασης (βλ. Κανόνες 4.1 και 7.1) ταυτίζονται επιλέγοντας $\ell_X = \ell_Y = 1$.

Ένα από τα βασικά πλεονεκτήματα του διαγράμματος $O2N2$ είναι η ικανότητά του να ανιχνεύει το είδος της μετατόπισης (θέσης ή κλίμακας/διασποράς) και ταυτόχρονα να προσδιορίζει ποια από τα χαρακτηριστικά μετατοπίστηκαν σε μία εκτός ελέγχου κατάσταση. Ειδικότερα, εάν η Συνθήκη 7.1α ή 7.1β (αντίστοιχα η Συνθήκη 7.1γ ή 7.1δ) παραβιάζεται, τότε ο συναγερμός ενεργοποιείται εξαιτίας μετατόπισης της μέσης τιμής (αντίστοιχα διασποράς). Αντίθετα, εάν δεν ικανοποιούνται μόνο οι Συνθήκες 7.1α ή/και 7.1γ (αντίστοιχα μόνο οι Συνθήκες 7.1β ή/και 7.1δ), αυτό πρακτικά σημαίνει ότι έχει προκληθεί μετατόπιση στη μέση τιμή ή/και τη διασπορά του πρώτου (αντίστοιχα δεύτερου) χαρακτηριστικού.

7.3 Χαρακτηριστικά διαγράμματος ελέγχου

Στη συνέχεια, αποδεικνύουμε τύπους υπολογισμού για τα χαρακτηριστικά του προτεινόμενου διαγράμματος ελέγχου, αξιοποιώντας τη Θεωρία των Συνδέσμων, καθώς και κάποια αποτελέσματα από αυτά που παρουσιάστηκαν στο Κεφάλαιο 3. Για το λόγο αυτό, θεωρούμε ότι $C(u, v)$ και $D(u, v)$ είναι οι σύνδεσμοι που αντιστοιχούν στις εντός και εκτός ελέγχου από κοινού α.σ.κ. $F(x, y)$ και $G(x, y)$. Επιπλέον, υποθέτουμε ότι οι αθροιστικές συναρτήσεις, οι οποίες εκφράζονται μέσω των περιθωρίων συναρτήσεων $F_X(x), F_Y(y)$ και $G_X(x), G_Y(y)$ από τους τύπους

$$F(x, y) = C(F_X(x), F_Y(y)) \quad \text{και} \quad G(x, y) = D(G_X(x), G_Y(y)),$$

είναι συνεχείς. Σε αυτή την περίπτωση, η πιθανότητα το διάγραμμα $O2N2$ να μη δίνει εκτός ελέγχου σήμα ορίζεται ως ακολούθως

$$p_{O2N2} = P(X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)}, Y_{c:m}^{(R)} \leq Y_{s:n}^{(T)} \leq Y_{d:m}^{(R)}, L_X \geq \ell_X, L_Y \geq \ell_Y), \quad (7.2)$$

Η παραπάνω πιθανότητα υποδηλώνει την OCF_{O2N2} και η ποσότητα $1 - p_{O2N2} = AR_{O2N2}$ είναι η πιθανότητα να ενεργοποιηθεί συναγερμός στη διεργασία. Στην επόμενη πρόταση, δίνεται μία έκφραση για την πιθανότητα p_{O2N2} , όπως αυτή ορίζεται στη Σχέση (7.2).

Πρόταση 7.4. Η χαρακτηριστική συνάρτηση λειτουργίας του διαγράμματος $O2N2$ δίνεται από την έκφραση

$$\begin{aligned} p_{O2N2} &= p_{O2N2}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b, c, d; r, s; \ell_X, \ell_Y) \\ &= \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} Q_{O2N2}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))); \ell_X, \ell_Y) \\ &\quad \times f_{a,b,c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2, \end{aligned} \quad (7.3)$$

όπου $Q_{O2N2}^{(D)}(\cdot)$ είναι η Πιθανότητα (3.8) του Λήμματος 3.6 και με $f_{a,b,c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ συμβολίζεται η από κοινού σ.π.π. των $U_{a:m}^{(R)}, U_{b:m}^{(R)}, V_{c:m}^{(R)}, V_{d:m}^{(R)}$ η οποία είναι η συνάρτηση πυκνότητας που αντιστοιχεί στην α.σ.κ. $F_{a,b,c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ της Σχέσης (3.4) και προκύπτει αντικαθιστώντας τις ποσότητες n, r_1, r_2, s_1, s_2, D με m, a, b, c, d, C .

Απόδειξη. Αρχικά, παρατηρούμε ότι η Πιθανότητα (7.2) γράφεται στην εναλλακτική μορφή που δίνεται παρακάτω

$$\begin{aligned} p_{O2N2} &= P [U_{a:m}^{(R)} \leq F_X(X_{r:n}^{(T)}) \leq U_{b:m}^{(R)}, V_{c:m}^{(R)} \leq F_Y(Y_{s:n}^{(T)}) \leq V_{d:m}^{(R)} \text{ και} \\ &\quad L(F_X(X_1^{(T)}), F_X(X_2^{(T)}), \dots, F_X(X_n^{(T)}); U_{a:m}^{(R)}, U_{b:m}^{(R)}) \geq \ell_X, \\ &\quad L(F_Y(Y_1^{(T)}), F_Y(Y_2^{(T)}), \dots, F_Y(Y_n^{(T)}); V_{c:m}^{(R)}, V_{d:m}^{(R)}) \geq \ell_Y]. \end{aligned}$$

Εφαρμόζοντας την αρχή δέσμευσης-αποδέσμευσης στις τ.μ.

$$U_{a:m}^{(R)} = F_X(X_{a:m}^{(R)}), \quad U_{b:m}^{(R)} = F_X(X_{b:m}^{(R)}) \quad \text{και} \quad V_{c:m}^{(R)} = F_Y(Y_{c:m}^{(R)}), \quad V_{d:m}^{(R)} = F_Y(Y_{d:m}^{(R)}),$$

η Σχέση (3.4) μπορεί να εκφραστεί μέσω του τετραπλού ολοκληρώματος που ακολουθεί

$$\begin{aligned} p_{O2N2} &= \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} P [u_1 \leq F_X(X_{r:n}^{(T)}) \leq u_2, v_1 \leq F_Y(Y_{s:n}^{(T)}) \leq v_2 \text{ και} \\ &\quad L(F_X(X_1^{(T)}), F_X(X_2^{(T)}), \dots, F_X(X_n^{(T)}); u_1, u_2) \geq \ell_X, \\ &\quad L(F_Y(Y_1^{(T)}), F_Y(Y_2^{(T)}), \dots, F_Y(Y_n^{(T)}); v_1, v_2) \geq \ell_Y] \\ &\quad \times f_{a,b,c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2, \end{aligned}$$

όπου $f_{a,b,c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. που δίνεται στον Τύπο (3.4) εάν τα r_1, r_2, s_1, s_2, n και C χρησιμοποιηθούν στη θέση των ποσοτήτων a, b, c, d, m και D , αντίστοιχα.

Η πιθανότητα που εμφανίζεται στο τετραπλό ολοκλήρωμα γράφεται ισοδύναμα ως εξής

$$\begin{aligned}
 & P [F_X^{-1}(u_1) \leq X_{r:n}^{(T)} \leq F_X^{-1}(u_2), F_Y^{-1}(v_1) \leq Y_{s:n}^{(T)} \leq F_Y^{-1}(v_2) \text{ και} \\
 & L(X_1^{(T)}, X_2^{(T)}, \dots, X_n^{(T)}; F_X^{-1}(u_1), F_X^{-1}(u_2)) \geq \ell_X, L(Y_1^{(T)}, Y_2^{(T)}, \dots, Y_n^{(T)}; F_Y^{-1}(v_1), F_Y^{-1}(v_2)) \geq \ell_Y] \\
 & = P [G_X(F_X^{-1}(u_1)) \leq G_X(X_{r:n}^{(T)}) \leq G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1)) \leq G_Y(Y_{s:n}^{(T)}) \leq G_Y(F_Y^{-1}(v_2)) \text{ και} \\
 & L(G_X(X_1^{(T)}), G_X(X_2^{(T)}), \dots, G_X(X_n^{(T)}); G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2))) \geq \ell_X, \\
 & L(G_Y(Y_1^{(T)}), G_Y(Y_2^{(T)}), \dots, G_Y(Y_n^{(T)}); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))) \geq \ell_Y].
 \end{aligned}$$

Με τη βοήθεια του Λήμματος 3.6, δεδομένου ότι οι ποσότητες $G_X(X_1^{(T)})$, $G_X(X_2^{(T)})$, \dots , $G_X(X_n^{(T)})$ και $G_Y(Y_1^{(T)})$, $G_Y(Y_2^{(T)})$, \dots , $G_Y(Y_n^{(T)})$ ακολουθούν την Ομοιόμορφη κατανομή στο διάστημα $(0, 1)$, έχουμε

$$\begin{aligned}
 & P [G_X(F_X^{-1}(u_1)) \leq U_{r:n}^{(T)} \leq G_X(F_X^{-1}(u_2)), G_Y(F_Y^{-1}(v_1)) \leq V_{s:n}^{(T)} \leq G_Y(F_Y^{-1}(v_2)) \text{ και} \\
 & L(U_1^{(T)}, U_2^{(T)}, \dots, U_n^{(T)}; G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2))) \geq \ell_X, \\
 & L(V_1^{(T)}, V_2^{(T)}, \dots, V_n^{(T)}; G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2))) \geq \ell_Y] \\
 & = Q_1^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2)); \ell_X, \ell_Y).
 \end{aligned}$$

Αντικαθιστώντας τα παραπάνω αποτελέσματα μέσα στο τετραπλό ολοκλήρωμα, προκύπτει η τελική έκφραση που ακολουθεί

$$\begin{aligned}
 p_{O2N2} & = p_{O2N2}^{(F_X, G_X; F_Y, G_Y; C, D)}(m, n; a, b, c, d; r, s; \ell_X, \ell_Y) \\
 & = \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} Q_{O2N2}^{(D)}(G_X(F_X^{-1}(u_1)), G_X(F_X^{-1}(u_2)); G_Y(F_Y^{-1}(v_1)), G_Y(F_Y^{-1}(v_2)); \ell_X, \ell_Y) \\
 & \quad \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2
 \end{aligned}$$

και η απόδειξη της Πρότασης ολοκληρώνεται. □

Εάν στην Έκφραση (7.3) της Πρότασης 7.4 θέσουμε $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$ παράγεται μία ακριβής έκφραση υπολογισμού της πιθανότητας συναγεμού του διαγράμματος $O2N2$, ενώ εάν επιπλέον ισχύει $C(u, v) = D(u, v)$ προκύπτει έκφραση για τον υπολογισμό της πιθανότητας λανθασμένου συναγεμού του προτεινόμενου διαγράμματος. Τα αποτελέσματα παρουσιάζονται στα δύο επόμενα πορίσματα.

Πόρισμα 7.7. *Εάν $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, η πιθανότητα συναγεμού του διαγράμματος $O2N2$, το οποίο ορίζεται στη Σχέση (7.1) δεν εξαρτάται από τις περιθώριες κατανομές και δίνεται από την έκφραση*

$$\begin{aligned}
 AR_{O2N2}^{(C,D)} &= 1 - P_{O2N2}^{(F_X, F_X; F_Y, F_Y; C, D)}(m, n; a, b, c, d; r, s; \ell_X, \ell_Y) \\
 &= 1 - \int_0^1 \int_0^{v_2} \int_0^1 \int_0^{u_2} Q_{O2N2}^{(D)}(u_1, u_2; v_1, v_2; \ell_X, \ell_Y) \\
 &\quad \times f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2) du_1 du_2 dv_1 dv_2, \quad (7.4)
 \end{aligned}$$

όπου $Q_{O2N2}^{(D)}(\cdot)$ είναι η πιθανότητα της Σχέσης (3.8) και $f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. της Σχέσης (3.4) με τις ποσότητες m, a, b, c, d, C να έχουν χρησιμοποιηθεί στη θέση των n, r_1, r_2, s_1, s_2, D , αντίστοιχα.

Απόδειξη. Αφού $F_X(x) = G_X(x)$ και $F_Y(y) = G_Y(y)$, είναι προφανές ότι $G_X(F_X^{-1}(u_i)) = u_i$ και $G_Y(F_Y^{-1}(v_j)) = v_j$, για $i, j = 1, 2$, το οποίο ολοκληρώνει την απόδειξη του πορίσματος. \square

Πόρισμα 7.8. Στην ειδική περίπτωση που ισχύει $F_X(x) = G_X(x)$, $F_Y(y) = G_Y(y)$ και $C(u, v) = D(u, v)$, προκύπτει η πιθανότητα λανθασμένου συναγερωμού του διαγράμματος $O2N2$, η οποία ορίζεται ως ακολούθως

$$\begin{aligned}
 FAR_{O2N2}^{(C)} &= 1 - P(X_{a:m}^{(R)} \leq X_{r:n}^{(T)} \leq X_{b:m}^{(R)}, Y_{c:m}^{(R)} \leq Y_{s:n}^{(T)} \leq Y_{d:m}^{(R)}, L_X \geq \ell_X, L_Y \geq \ell_Y | H_0) \\
 &= AR_{O2N2}^{(C,C)}(m, n; a, b, c, d; r, s; \ell_X, \ell_Y),
 \end{aligned}$$

όπου η ποσότητα $Q_{O4}^{(C)}(\cdot)$ δίνεται στη Σχέση (3.7) για $D = C$ και η $f_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ είναι η πυκνότητα της α.σ.κ. $F_{a,b;c,d;m}^{(C)}(u_1, u_2; v_1, v_2)$ της Σχέσης (3.4) που υπολογίζεται αντικαθιστώντας τις ποσότητες n, r_1, r_2, s_1, s_2, D με m, a, b, c, d, C , αντίστοιχα.

Απόδειξη. Το αποτέλεσμα προκύπτει άμεσα με τη βοήθεια της Σχέσης (7.4). \square

Από το τελευταίο πόρισμα είναι φανερό ότι το διάγραμμα $O2N2$ χαρακτηρίζεται ως ημιπαραμετρικό, καθώς η πιθανότητα λανθασμένου συναγερωμού δεν επηρεάζεται από τις (μονοδιάστατες) περιθώριες κατανομές, αλλά η επίδοσή του εξαρτάται τυπικά από τη δομή εξάρτησης των υπό μελέτη χαρακτηριστικών. Ωστόσο, οι τιμές του FAR_{O2N2} δε διαφοροποιούνται δραματικά ανεξάρτητα από την επιλογή του συνδέσμου, όπως θα δούμε στην επόμενη ενότητα.

7.4 Μελέτη Απόδοσης και σχεδιασμός διαγράμματος ελέγχου

Ο σκοπός της τρέχουσας ενότητας είναι τριμερής και συνοψίζεται στα εξής:

- (i) ικανότητα του διαγράμματος $O2N2$ να κάνει διάκριση μεταξύ μεταβολών μέσης τιμής ή/και τυπικής απόκλισης,

- (ii) μελέτη απόδοσης του διαγράμματος $O2N2$ για διαφορετικά είδη συνδέσμων και
- (iii) επιλογή παραμέτρων σχεδιασμού του διαγράμματος $O2N2$.

Αρχικά, παρουσιάζουμε τον Πίνακα 7.1, με τη βοήθεια του οποίου διασαφηνίζεται ότι το διάγραμμα $O2N2$ είναι ικανό να κάνει διάκριση μεταξύ μεταβολών της μέσης τιμής και της διασποράς. Συγκεκριμένα, ο Πίνακας 7.1 απεικονίζει τη διάσπαση του συνολικού ποσοστού συναγερμού σε επιμέρους ποσοστά συναγερμού που σχετίζονται με κάθε διαφορετική μετατόπιση (στη μέση τιμή/διακύμανση). Για να συγκρίνουμε τα στοιχεία του πίνακα, κατασκευάστηκαν δισδιάστατα δεδομένα από το σύνδεσμο του *Gauss* (βλ. Πίνακα 2.5) με $\rho = 0.5$ και κανονικές περιθώριες κατανομές. Ξεκινώντας με ποσοστό λανθασμένου συναγερμού ίσο με $FAR = 5.05\%$ ¹, όταν ως εντός ελέγχου περιθώρια κατανομή χρησιμοποιήθηκε η τυπική κανονική $N(0, 1)$, τα ποσοστά συναγερμού υπολογίσθηκαν καθώς η μέση τιμή και η τυπική απόκλιση μεταβάλλονταν από τις εντός ελέγχου τιμές τους $\mu_{in} = 0$ και $\sigma_{in} = 1$ (και για τα δύο χαρακτηριστικά).

Για κάθε επιλογή των μ και σ , παρέχονται τέσσερις αριθμοί: το συνολικό ποσοστό συναγερμού (%) και τρεις επιπλέον τιμές, οι οποίες υποδηλώνουν τα ποσοστά συναγερμού που προκαλούνται αποκλειστικά και μόνο λόγω της Συνθήκης 7.1α ή 7.1β (γεγονός το οποίο υποδεικνύει μετατόπιση της μέσης τιμής), της Συνθήκης 7.1γ ή 7.1δ (γεγονός το οποίο υποδεικνύει μετατόπιση της τυπικής απόκλισης) και τα ποσοστά συναγερμού που ενεργοποιούνται από παραβίαση των Συνθηκών 7.1α ή 7.1β και 7.1γ ή 7.1δ (υποδεικνύοντας ταυτόχρονες μετατοπίσεις στη μέση τιμή και την τυπική απόκλιση). Ειδικότερα, εάν A , B είναι τα ενδεχόμενα ότι έχει αντίστοιχα προκληθεί μία παραβίαση στη Συνθήκη 7.1α ή 7.1β, στη Συνθήκη 7.1γ ή 7.1δ, τότε και οι τέσσερις αριθμοί που απεικονίζονται στον Πίνακα 7.1 είναι οι $P(A \cup B)$ και $(P(AB')/P(A \cup B))$, $P(A'B)/P(A \cup B)$, $P(AB)/P(A \cup B)$.

Οι τιμές που σημειώνονται με έντονη γραφή υποδεικνύουν τη μέγιστη πιθανότητα μεταξύ των τριών διαφορετικών ποσοστών συναγερμού. Είναι προφανές ότι το προτεινόμενο διάγραμμα $O2N2$ λειτουργεί αρκετά ικανοποιητικά. Αφενός, όταν προκαλούνται ίσες μετατοπίσεις στη μέση τιμή (αντίστοιχα στην τυπική απόκλιση) σε αμφότερα τα χαρακτηριστικά, η πλειοψηφία των συναγερμών ενεργοποιείται επειδή παραβιάζεται τουλάχιστον ένας από τους Κανόνες 7.1α ή 7.1β (αντίστοιχα ένας από τους Κανόνες 7.1γ ή 7.1δ). Αφετέρου, όταν προκαλούνται μετατοπίσεις στη μέση τιμή και στην τυπική απόκλιση, τότε οι αντίστοιχες συνθήκες του Κανόνα 7.1 παραβιάζονται ταυτόχρονα.

¹ Το FAR είναι περίπου ίσο και όχι ακριβώς ίσο με 5% λόγω της διακριτής φύσης των τιμών των παραμέτρων σχεδιασμού.

| Μετατοπίσεις Μέσης Τιμής | | Μετατοπίσεις Τυπικής Απόκλισης: $\sigma = \sigma_x = \sigma_y = 1(0.25)3$ | | | | |
|-------------------------------------|-------|---|-------|-------------------------------|-------|-------------------------------|
| $\mu = \mu_x = \mu_y = 0(0.20)0.80$ | | 1.00 | 1.25 | 1.50 | | |
| 0.00 | 5.05 | (14.85, 67.92, 17.23) | 24.27 | (6.26, 76.89 , 16.85) | 53.00 | (3.19, 74.64 , 22.17) |
| 0.20 | 8.58 | (43.83 , 36.83, 19.35) | 28.65 | (15.08, 57.94 , 26.98) | 55.86 | (6.03, 64.45 , 29.52) |
| 0.40 | 25.61 | (65.40 , 8.20, 26.40) | 42.38 | (29.47, 28.13, 42.40) | 63.19 | (11.95, 41.75 46.30) |
| 0.60 | 56.13 | (69.04 , 1.53, 29.43) | 64.53 | (38.96, 8.97, 52.07) | 75.72 | (16.52, 19.76, 63.72) |
| 0.80 | 85.05 | (60.29 , 0.25, 39.46) | 84.19 | (37.17, 2.42, 60.41) | 87.40 | (18.40, 6.92, 74.68) |
| Μετατοπίσεις Μέσης Τιμής | | Μετατοπίσεις Τυπικής Απόκλισης: $\sigma = \sigma_x = \sigma_y = 1(0.25)3$ | | | | |
| $\mu = \mu_x = \mu_y = 0(0.20)0.80$ | | 1.75 | 2.00 | 2.25 | | |
| 0.00 | 74.66 | (1.68, 70.88 , 27.44) | 88.09 | (0.87, 65.06 , 34.07) | 94.64 | (0.34, 58.99 , 40.67) |
| 0.20 | 76.11 | (2.34, 65.01 , 32.65) | 88.52 | (1.03, 60.72 , 38.25) | 94.74 | (0.54, 55.49 , 43.97) |
| 0.40 | 79.93 | (4.29, 46.64, 49.07) | 90.26 | (1.60, 46.85, 51.55) | 95.59 | (0.73, 45.28, 53.99) |
| 0.60 | 85.77 | (6.48, 27.96, 65.56) | 91.86 | (2.62, 31.57, 65.81) | 96.08 | (1.00, 31.85, 67.15) |
| 0.80 | 91.77 | (8.38, 13.52, 78.10) | 95.08 | (3.01, 18.30, 78.69) | 97.42 | (1.17, 19.70, 79.13) |
| Μετατοπίσεις Μέσης Τιμής | | Μετατοπίσεις Τυπικής Απόκλισης: $\sigma = \sigma_x = \sigma_y = 1(0.25)3$ | | | | |
| $\mu = \mu_x = \mu_y = 0(0.20)0.80$ | | 2.50 | 2.75 | 3.00 | | |
| 0.00 | 97.53 | (0.20, 52.71 , 47.09) | 98.83 | (0.10, 46.14, 53.76) | 99.49 | (0.01, 41.83, 58.16) |
| 0.20 | 97.65 | (0.25, 49.53, 50.22) | 98.94 | (0.11, 44.26, 55.63) | 99.52 | (0.03, 39.92, 60.05) |
| 0.40 | 97.87 | (0.27, 42.30, 57.43) | 99.10 | (0.16, 38.68, 61.16) | 99.59 | (0.04, 35.05, 64.91) |
| 0.60 | 98.44 | (0.48, 32.05, 67.47) | 99.21 | (0.17, 30.08, 69.75) | 99.61 | (0.07, 28.50, 71.43) |
| 0.80 | 98.79 | (0.54, 21.79, 77.67) | 99.53 | (0.19, 21.52, 78.29) | 99.72 | (0.11, 21.19, 78.70) |

Πίνακας 7.1: Διάσπαση Ποσοστών Συνεργισμού του διαγράμματος $O2N2$ για δεδομένο σχεδιασμό

$$(m = 500, n = 25, a = c, b = m - a + 1, d = m - c + 1 \ \& \ r = s = (n + 1)/2)$$

Στη συνέχεια, μελετάται η ικανότητα του διαγράμματος $O2N2$ να ανιχνεύει μετατοπίσεις της μέσης τιμής ή/και της τυπικής απόκλισης της διεργασίας για διαφορετικά είδη συνδέσμων, ενώ παράλληλα δίνονται οδηγίες για τον κατάλληλο σχεδιασμό του προτεινόμενου διαγράμματος διακρίνοντας τρία πιθανά σενάρια για προκαθορισμένο FAR_{O2N2} .

Όπως έχουμε ήδη αναφέρει, οι επιλογές των παραμέτρων σχεδιασμού δεν είναι μοναδικές. Ωστόσο, για λόγους περιορισμού του αριθμού των παραμέτρων θα χρησιμοποιούμε το ζευγάρι των διαμέσων του δείγματος ελέγχου και στην περίπτωση που το μέγεθος δείγματος n είναι άρτιο θέτουμε $r = s = (n + 1)/2$, διαφορετικά $r = s = n/2 + 1$. Δεδομένου ότι η r -οστή και s -οστή παρατήρηση αντιστοιχούν στις διαμέσους του δείγματος ελέγχου για τις μεταβλητές X και Y , επιλέγονται συμμετρικά όρια (δηλαδή $b = m - a + 1$, $d = m - c + 1$) και σε συνδυασμό με το ότι ίσες μετατοπίσεις προκαλούνται στα υπό μελέτη χαρακτηριστικά λαμβάνονται οι ισότητες $a = c$ και $b = d$. Αξίζει να σημειωθεί ότι οι τιμές των υπόλοιπων παραμέτρων σχεδιασμού ℓ_X και ℓ_Y καθορίζονται προκειμένου να επιτευχθεί το επιθυμητό FAR_{O2N2} (τιμές με έντονους χαρακτήρες), το οποίο κυμαίνεται μεταξύ 5% και 10%.

Παρότι το διάγραμμα $O2N2$ χαρακτηρίζεται ως ημιπαραμετρικό, το ποσοστό λανθασμένου συναγερού παραμένει σχεδόν αμετάβλητο ανεξάρτητα από την επιλογή του συνδέσμου. Αυτό επιβεβαιώνεται και από τους Πίνακες 7.2, 7.3, 7.4, όπου ξεκινώντας με δεδομένο FAR_{O2N2} , υπολογίζονται οι πιθανότητες συναγερού AR_{O2N2} καθώς μεταβάλλεται αντίστοιχα η μέση τιμή, η τυπική απόκλιση ή ταυτόχρονα η μέση τιμή και η τυπική απόκλιση των περιθώριων κατανομών. Οι εντός ελέγχου περιθώριες κατανομές που χρησιμοποιούνται είναι τυπικές κανονικές, οπότε $\mu_{in} = \mu_X^{(R)} = \mu_Y^{(R)} = 0$ και $\sigma_{in} = \sigma_X^{(R)} = \sigma_Y^{(R)} = 1$, ενώ η μεταβολή των εκτός ελέγχου τιμών $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ και $\sigma_{out} = \sigma_X^{(T)} = \sigma_Y^{(T)}$ παρουσιάζεται στους πίνακες. Θα πρέπει να αναφερθεί ότι σε κάθε μία από τις παραπάνω περιπτώσεις γίνεται χρήση των συνδέσμων AMH , CLP , FGM , GUH και FRA (βλ. Πίνακα 2.5), όπου για τις παραμέτρους τους επιλέγεται κοινή εντός και εκτός ελέγχου τιμή. Η απόδοση του προτεινόμενου διαγράμματος είναι αρκετά ικανοποιητική ανεξάρτητα από το είδος της μετατόπισης που πραγματοποιείται.

Παράλληλα, στον Πίνακα 7.5 παρουσιάζονται οι δυνατές επιλογές των παραμέτρων σχεδιασμού που μπορεί κανείς να χρησιμοποιήσει για την πρακτική εφαρμογή του διαγράμματος $O2N2$ ανάλογα με τις ανάγκες του σχεδιασμού βάσει του FAR_{O2N2} . Και σε αυτή την περίπτωση, οι υπολογισμοί έγιναν χρησιμοποιώντας τις διαμέσους του δείγματος ελέγχου που είχε ως αποτέλεσμα την επιλογή συμμετρικών ορίων ελέγχου. Ανάλογες επιλογές μπορούν να προκύψουν σε περιπτώσεις μη συμμετρικών κατανομών.

7.4. Μελέτη Απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | | Μετατοπίσεις στη Μέση Τιμή : $\mu_{out} = \mu_X^{(T)} = \mu_Y^{(T)}$ | | | | | | | | |
|-------------------------------|-----|--------------------------|---------|--|------|------|------|-------------|------|------|------|------|
| Παράμετροι Σχεδιασμού | | | | -2 | -1.5 | -1 | -0.5 | 0 | 0.5 | 1 | 1.5 | 2 |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | |
| 100 | 5 | 1 | 11 | 0.99 | 0.90 | 0.56 | 0.17 | 0.05 | 0.18 | 0.55 | 0.90 | 0.99 |
| | | | 13 | 1 | 0.94 | 0.64 | 0.24 | 0.08 | 0.24 | 0.65 | 0.94 | 1 |
| | | | 14 | 1 | 0.95 | 0.68 | 0.27 | 0.10 | 0.27 | 0.68 | 0.95 | 1 |
| | 11 | 4 | 17 | 1 | 0.99 | 0.80 | 0.24 | 0.05 | 0.23 | 0.79 | 0.99 | 1 |
| | | | 19 | 1 | 1 | 0.85 | 0.31 | 0.08 | 0.32 | 0.86 | 1 | 1 |
| | | | 20 | 1 | 1 | 0.88 | 0.36 | 0.10 | 0.36 | 0.88 | 1 | 1 |
| | | | 25 | 1 | 1 | 0.98 | 0.43 | 0.05 | 0.43 | 0.97 | 1 | 1 |
| | 25 | 7 | 26 | 1 | 1 | 0.99 | 0.51 | 0.06 | 0.51 | 0.99 | 1 | 1 |
| | | | 27 | 1 | 1 | 0.99 | 0.57 | 0.10 | 0.57 | 0.99 | 1 | 1 |
| | | | | | | | | | | | | |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | |
| 100 | 5 | 1 | 11 | 0.99 | 0.87 | 0.51 | 0.16 | 0.05 | 0.17 | 0.52 | 0.86 | 0.98 |
| | | | 13 | 0.99 | 0.91 | 0.60 | 0.22 | 0.08 | 0.22 | 0.60 | 0.91 | 0.99 |
| | | | 14 | 0.99 | 0.92 | 0.64 | 0.26 | 0.10 | 0.26 | 0.66 | 0.92 | 0.99 |
| | 11 | 4 | 17 | 1 | 0.99 | 0.75 | 0.22 | 0.05 | 0.23 | 0.75 | 0.98 | 1 |
| | | | 19 | 1 | 0.99 | 0.81 | 0.30 | 0.08 | 0.30 | 0.81 | 0.99 | 1 |
| | | | 20 | 1 | 0.99 | 0.84 | 0.33 | 0.10 | 0.94 | 0.85 | 0.99 | 1 |
| | | | 25 | 1 | 1 | 0.97 | 0.42 | 0.05 | 0.43 | 0.97 | 1 | 1 |
| | 25 | 7 | 26 | 1 | 1 | 0.98 | 0.46 | 0.06 | 0.49 | 0.98 | 1 | 1 |
| | | | 27 | 1 | 1 | 0.98 | 0.53 | 0.10 | 0.84 | 0.99 | 1 | 1 |
| | | | | | | | | | | | | |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = 0.5$) | | | | | | | | |
| 100 | 5 | 1 | 11 | 0.99 | 0.88 | 0.53 | 0.18 | 0.05 | 0.17 | 0.53 | 0.88 | 0.99 |
| | | | 13 | 0.99 | 0.92 | 0.61 | 0.23 | 0.08 | 0.23 | 0.62 | 0.91 | 1 |
| | | | 14 | 1 | 0.93 | 0.65 | 0.26 | 0.10 | 0.26 | 0.65 | 0.93 | 1 |
| | 11 | 4 | 17 | 1 | 0.99 | 0.76 | 0.22 | 0.05 | 0.23 | 0.76 | 0.99 | 1 |
| | | | 19 | 1 | 0.99 | 0.82 | 0.31 | 0.08 | 0.31 | 0.83 | 0.99 | 1 |
| | | | 20 | 1 | 1 | 0.86 | 0.35 | 0.10 | 0.35 | 0.85 | 1 | 1 |
| | | | 25 | 1 | 1 | 0.98 | 0.43 | 0.05 | 0.43 | 0.98 | 1 | 1 |
| | 25 | 7 | 26 | 1 | 1 | 0.98 | 0.49 | 0.07 | 0.48 | 0.98 | 1 | 1 |
| | | | 27 | 1 | 1 | 1 | 0.54 | 0.10 | 0.54 | 1 | 1 | 1 |
| | | | | | | | | | | | | |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 3$) | | | | | | | | |
| 100 | 5 | 1 | 11 | 0.99 | 0.91 | 0.55 | 0.17 | 0.05 | 0.17 | 0.56 | 0.90 | 0.99 |
| | | | 13 | 1 | 0.94 | 0.66 | 0.24 | 0.08 | 0.24 | 0.67 | 0.94 | 1 |
| | | | 14 | 1 | 0.95 | 0.68 | 0.27 | 0.10 | 0.26 | 0.69 | 0.95 | 1 |
| | 11 | 4 | 17 | 1 | 0.99 | 0.80 | 0.24 | 0.05 | 0.23 | 0.79 | 0.99 | 1 |
| | | | 19 | 1 | 1 | 0.85 | 0.31 | 0.08 | 0.32 | 0.86 | 1 | 1 |
| | | | 20 | 1 | 1 | 0.88 | 0.35 | 0.10 | 0.36 | 0.89 | 1 | 1 |
| | | | 25 | 1 | 1 | 0.98 | 0.45 | 0.05 | 0.45 | 0.98 | 1 | 1 |
| | 25 | 7 | 26 | 1 | 1 | 0.99 | 0.51 | 0.06 | 0.51 | 0.99 | 1 | 1 |
| | | | 27 | 1 | 1 | 0.99 | 0.56 | 0.10 | 0.57 | 0.99 | 1 | 1 |
| | | | | | | | | | | | | |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 1$) | | | | | | | | |
| 100 | 5 | 1 | 11 | 0.99 | 0.88 | 0.53 | 0.17 | 0.05 | 0.17 | 0.54 | 0.88 | 0.99 |
| | | | 13 | 0.99 | 0.93 | 0.64 | 0.24 | 0.08 | 0.23 | 0.64 | 0.93 | 1 |
| | | | 14 | 1 | 0.94 | 0.67 | 0.27 | 0.10 | 0.27 | 0.68 | 0.95 | 1 |
| | 11 | 4 | 17 | 1 | 0.99 | 0.77 | 0.23 | 0.05 | 0.23 | 0.78 | 0.99 | 1 |
| | | | 19 | 1 | 1 | 0.85 | 0.31 | 0.08 | 0.31 | 0.85 | 1 | 1 |
| | | | 20 | 1 | 1 | 0.87 | 0.35 | 0.10 | 0.35 | 0.87 | 1 | 1 |
| | | | 25 | 1 | 1 | 0.98 | 0.44 | 0.05 | 0.45 | 0.98 | 1 | 1 |
| | 25 | 7 | 26 | 1 | 1 | 0.99 | 0.50 | 0.07 | 0.51 | 0.99 | 1 | 1 |
| | | | 27 | 1 | 1 | 0.99 | 0.56 | 0.10 | 0.55 | 0.99 | 1 | 1 |
| | | | | | | | | | | | | |

Πίνακας 7.2: Πιθανότητες Συναγερμού $AR_{0.2N_2}$ λόγω μετατοπίσεων μέσης τιμής για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$)

Δισδιάστατο Διάγραμμα Ελέγχου
για τη Θέση και τη Διασπορά με Χρήση Απαριθμητριών Στατιστικών Συναρτήσεων

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | | Μετατοπίσεις στην Τυπική Απόκλιση : $\sigma_{out} = \sigma_X^{(T)} = \sigma_Y^{(T)}$ | | | | | | | | | |
|-------------------------------|-----|--------------------------|---------|--|------|------|------|------|------|------|------|------|---|
| Παράμετροι Σχεδιασμού | | | | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | | |
| 100 | 5 | 2 | 10 | 0.05 | 0.25 | 0.51 | 0.67 | 0.79 | 0.86 | 0.91 | 0.93 | 0.95 | |
| | | | 12 | 0.07 | 0.33 | 0.57 | 0.73 | 0.84 | 0.89 | 0.92 | 0.95 | 0.96 | |
| | | | 14 | 0.10 | 0.40 | 0.64 | 0.79 | 0.86 | 0.91 | 0.94 | 0.96 | 0.97 | |
| | 11 | 5 | 15 | 0.05 | 0.45 | 0.78 | 0.93 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 16 | 0.08 | 0.49 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 17 | 0.10 | 0.56 | 0.85 | 0.99 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | 25 | 8 | 24 | 0.05 | 0.47 | 0.84 | 0.96 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.54 | 0.97 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | 26 | 0.10 | 0.61 | 0.91 | 0.98 | 1 | 1 | 1 | 1 | 1 | 1 |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | | |
| 100 | 5 | 2 | 10 | 0.05 | 0.25 | 0.49 | 0.67 | 0.78 | 0.85 | 0.90 | 0.93 | 0.95 | |
| | | | 12 | 0.07 | 0.32 | 0.56 | 0.72 | 0.83 | 0.89 | 0.92 | 0.94 | 0.96 | |
| | | | 14 | 0.10 | 0.40 | 0.63 | 0.77 | 0.86 | 0.90 | 0.94 | 0.96 | 0.97 | |
| | 11 | 5 | 15 | 0.05 | 0.44 | 0.78 | 0.93 | 0.97 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 16 | 0.08 | 0.50 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 17 | 0.10 | 0.55 | 0.85 | 0.95 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | 25 | 8 | 24 | 0.05 | 0.48 | 0.84 | 0.96 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.54 | 0.88 | 0.97 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | 26 | 0.10 | 0.61 | 0.90 | 0.98 | 1 | 1 | 1 | 1 | 1 | 1 |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = 0.5$) | | | | | | | | | |
| 100 | 5 | 2 | 10 | 0.05 | 0.26 | 0.49 | 0.67 | 0.79 | 0.86 | 0.90 | 0.93 | 0.95 | |
| | | | 12 | 0.08 | 0.33 | 0.57 | 0.74 | 0.84 | 0.89 | 0.93 | 0.95 | 0.96 | |
| | | | 14 | 0.10 | 0.40 | 0.63 | 0.78 | 0.87 | 0.92 | 0.94 | 0.96 | 0.97 | |
| | 11 | 5 | 15 | 0.05 | 0.44 | 0.78 | 0.93 | 0.97 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 16 | 0.08 | 0.50 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 17 | 0.10 | 0.55 | 0.85 | 0.95 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | 25 | 8 | 24 | 0.05 | 0.49 | 0.84 | 0.96 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.55 | 0.88 | 0.98 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 26 | 0.10 | 0.62 | 0.91 | 0.98 | 1 | 1 | 1 | 1 | 1 | 1 |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 3$) | | | | | | | | | |
| 100 | 5 | 2 | 10 | 0.05 | 0.29 | 0.53 | 0.69 | 0.82 | 0.88 | 0.91 | 0.94 | 0.96 | |
| | | | 12 | 0.07 | 0.32 | 0.56 | 0.73 | 0.83 | 0.88 | 0.93 | 0.95 | 0.96 | |
| | | | 14 | 0.11 | 0.40 | 0.63 | 0.78 | 0.86 | 0.91 | 0.93 | 0.96 | 0.97 | |
| | 11 | 5 | 15 | 0.05 | 0.44 | 0.78 | 0.93 | 0.97 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 16 | 0.08 | 0.50 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 17 | 0.10 | 0.55 | 0.84 | 0.95 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | 25 | 8 | 24 | 0.05 | 0.47 | 0.84 | 0.96 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.55 | 0.88 | 0.97 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 26 | 0.10 | 0.62 | 0.91 | 0.98 | 1 | 1 | 1 | 1 | 1 | 1 |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 1$) | | | | | | | | | |
| 100 | 5 | 2 | 10 | 0.05 | 0.25 | 0.51 | 0.68 | 0.79 | 0.85 | 0.90 | 0.93 | 0.95 | |
| | | | 12 | 0.07 | 0.32 | 0.56 | 0.73 | 0.83 | 0.89 | 0.92 | 0.94 | 0.96 | |
| | | | 14 | 0.10 | 0.40 | 0.64 | 0.78 | 0.87 | 0.91 | 0.94 | 0.96 | 0.97 | |
| | 11 | 5 | 15 | 0.05 | 0.45 | 0.79 | 0.92 | 0.97 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 16 | 0.08 | 0.50 | 0.82 | 0.94 | 0.98 | 0.99 | 1 | 1 | 1 | 1 |
| | | | 17 | 0.10 | 0.56 | 0.85 | 0.96 | 0.98 | 1 | 1 | 1 | 1 | 1 |
| | 25 | 8 | 24 | 0.05 | 0.48 | 0.84 | 0.96 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 25 | 0.08 | 0.54 | 0.88 | 0.97 | 0.99 | 1 | 1 | 1 | 1 | 1 |
| | | | 26 | 0.10 | 0.61 | 0.91 | 0.98 | 1 | 1 | 1 | 1 | 1 | 1 |

Πίνακας 7.3: Πιθανότητες Συναγερμού $AR_{0.2N_2}$ λόγω μετατοπίσεων τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$)

7.4. Μελέτη Απόδοσης και σχεδιασμός διαγράμματος ελέγχου

| $F_X(x), F_Y(y) \sim N(0, 1)$ | | | | Ταυτόχρονες Μετατοπίσεις στη Μέση Τιμή και την Τυπική Απόκλιση | | | | | | | | | |
|-------------------------------|-----|--------------------------|--------------------------|--|---|------|------|-------------|-------------|------|------|------|------|
| Παράμετροι Σχεδιασμού | | | | Μέση Τιμή: | -2 | -1.5 | -1 | -0.5 | 0 | 0.5 | 1 | 1.5 | 2 |
| | | | | Τυπ. Αποκλ.: | 3 | 2.5 | 2 | 1.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος AMH ($\theta_1 = \theta_2 = -0.3$) | | | | | | | | | |
| 100 | 5 | 3 | 7 | 0.97 | 0.91 | 0.76 | 0.40 | 0.05 | 0.42 | 0.77 | 0.92 | 0.97 | |
| | | | 9 | 0.98 | 0.95 | 0.83 | 0.53 | 0.09 | 0.53 | 0.83 | 0.94 | 0.98 | |
| | | | 10 | 0.99 | 0.95 | 0.85 | 0.57 | 0.12 | 0.57 | 0.86 | 0.94 | 0.98 | |
| | 11 | 6 | 11 | 1 | 0.99 | 0.92 | 0.55 | 0.05 | 0.55 | 0.92 | 0.99 | 1 | |
| | | | 12 | 1 | 0.99 | 0.93 | 0.61 | 0.07 | 0.62 | 0.94 | 0.99 | 1 | |
| | | | 13 | 1 | 0.99 | 0.95 | 0.67 | 0.10 | 0.66 | 0.95 | 0.99 | 1 | |
| | 25 | 9 | 22 | 1 | 1 | 0.97 | 0.66 | 0.05 | 0.67 | 0.97 | 1 | 1 | |
| | | | 23 | 1 | 1 | 0.98 | 0.73 | 0.07 | 0.72 | 0.98 | 1 | 1 | |
| | | | 24 | 1 | 1 | 0.99 | 0.77 | 0.10 | 0.78 | 0.99 | 1 | 1 | |
| | m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος CLP ($\theta_1 = \theta_2 = 2$) | | | | | | | | |
| | 100 | 5 | 3 | 7 | 0.96 | 0.89 | 0.74 | 0.39 | 0.05 | 0.39 | 0.75 | 0.91 | 0.96 |
| | | | | 9 | 0.97 | 0.94 | 0.81 | 0.50 | 0.09 | 0.51 | 0.82 | 0.93 | 0.98 |
| 10 | | | | 0.98 | 0.94 | 0.84 | 0.55 | 0.12 | 0.55 | 0.85 | 0.95 | 0.98 | |
| 11 | | 6 | 11 | 1 | 0.98 | 0.90 | 0.54 | 0.05 | 0.54 | 0.91 | 0.99 | 1 | |
| | | | 12 | 1 | 0.99 | 0.92 | 0.60 | 0.07 | 0.60 | 0.93 | 0.99 | 1 | |
| | | | 13 | 1 | 0.99 | 0.94 | 0.66 | 0.10 | 0.66 | 0.94 | 0.99 | 1 | |
| 25 | | 9 | 22 | 1 | 1 | 0.96 | 0.65 | 0.05 | 0.65 | 0.97 | 1 | 1 | |
| | | | 23 | 1 | 1 | 0.98 | 0.70 | 0.07 | 0.71 | 0.98 | 1 | 1 | |
| | | | 24 | 1 | 1 | 0.98 | 0.76 | 0.10 | 0.76 | 0.98 | 1 | 1 | |
| m | | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος FGM ($\theta_1 = \theta_2 = 0.5$) | | | | | | | | |
| 100 | | 5 | 3 | 7 | 0.96 | 0.91 | 0.75 | 0.40 | 0.05 | 0.41 | 0.75 | 0.91 | 0.96 |
| | | | | 9 | 0.89 | 0.94 | 0.83 | 0.52 | 0.09 | 0.51 | 0.83 | 0.94 | 0.97 |
| | 10 | | | 0.98 | 0.95 | 0.84 | 0.56 | 0.12 | 0.57 | 0.85 | 0.94 | 0.98 | |
| | 11 | 6 | 11 | 1 | 0.99 | 0.92 | 0.55 | 0.05 | 0.55 | 0.91 | 0.98 | 1 | |
| | | | 12 | 1 | 0.99 | 0.93 | 0.61 | 0.06 | 0.62 | 0.93 | 0.99 | 1 | |
| | | | 13 | 1 | 0.99 | 0.95 | 0.67 | 0.10 | 0.66 | 0.95 | 0.99 | 1 | |
| | 25 | 9 | 22 | 1 | 1 | 0.97 | 0.65 | 0.05 | 0.65 | 0.97 | 1 | 1 | |
| | | | 23 | 1 | 1 | 0.98 | 0.71 | 0.07 | 0.71 | 0.98 | 1 | 1 | |
| | | | 24 | 1 | 1 | 0.99 | 0.76 | 0.10 | 0.77 | 0.98 | 1 | 1 | |
| | m | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος FRA ($\alpha_1 = \alpha_2 = 3$) | | | | | | | | |
| | 100 | 5 | 3 | 7 | 0.97 | 0.92 | 0.77 | 0.41 | 0.05 | 0.41 | 0.76 | 0.92 | 0.97 |
| | | | | 9 | 0.98 | 0.95 | 0.83 | 0.52 | 0.09 | 0.52 | 0.84 | 0.94 | 0.98 |
| 10 | | | | 0.98 | 0.96 | 0.86 | 0.57 | 0.12 | 0.57 | 0.86 | 0.96 | 0.98 | |
| 11 | | 6 | 11 | 1 | 0.99 | 0.91 | 0.54 | 0.05 | 0.55 | 0.91 | 0.99 | 1 | |
| | | | 12 | 1 | 0.99 | 0.94 | 0.61 | 0.07 | 0.62 | 0.94 | 0.99 | 1 | |
| | | | 13 | 1 | 0.99 | 0.95 | 0.67 | 0.10 | 0.68 | 0.95 | 0.99 | 1 | |
| 25 | | 9 | 22 | 1 | 1 | 0.97 | 0.66 | 0.05 | 0.66 | 0.97 | 1 | 1 | |
| | | | 23 | 1 | 1 | 0.98 | 0.71 | 0.07 | 0.72 | 0.98 | 1 | 1 | |
| | | | 24 | 1 | 1 | 0.99 | 0.78 | 0.10 | 0.78 | 0.99 | 1 | 1 | |
| m | | n | $\ell = \ell_X = \ell_Y$ | $a = c$ | Σύνδεσμος GUH ($\alpha_1 = \alpha_2 = 1$) | | | | | | | | |
| 100 | | 5 | 3 | 7 | 0.97 | 0.92 | 0.76 | 0.41 | 0.05 | 0.41 | 0.76 | 0.91 | 0.97 |
| | | | | 9 | 0.98 | 0.94 | 0.83 | 0.52 | 0.09 | 0.52 | 0.83 | 0.94 | 0.98 |
| | 10 | | | 0.98 | 0.95 | 0.86 | 0.57 | 0.12 | 0.57 | 0.86 | 0.95 | 0.99 | |
| | 11 | 6 | 11 | 1 | 0.99 | 0.91 | 0.56 | 0.05 | 0.55 | 0.91 | 0.99 | 1 | |
| | | | 12 | 1 | 0.99 | 0.93 | 0.63 | 0.07 | 0.62 | 0.94 | 0.99 | 1 | |
| | | | 13 | 1 | 0.99 | 0.95 | 0.67 | 0.10 | 0.67 | 0.95 | 0.99 | 1 | |
| | 25 | 9 | 22 | 1 | 1 | 0.97 | 0.66 | 0.05 | 0.65 | 0.97 | 1 | 1 | |
| | | | 23 | 1 | 1 | 0.98 | 0.72 | 0.07 | 0.72 | 0.98 | 1 | 1 | |
| | | | 24 | 1 | 1 | 0.98 | 0.77 | 0.10 | 0.77 | 0.99 | 1 | 1 | |

Πίνακας 7.4: Πιθανότητες Συναγερμού $AR_{0.2N_2}$ λόγω ταυτόχρονων μετατοπίσεων μέσης τιμής και τυπικής απόκλισης για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$)

*Δισδιάστατο Διάγραμμα Ελέγχου
για τη Θέση και τη Διασπορά με Χρήση Απαριθμητριών Στατιστικών Συναρτήσεων*

| Παράμετροι Σχεδιασμού | | | Μέγεθος Δείγματος Αναφοράς m | | | | | | | |
|-----------------------|--------------------------|--|--------------------------------|--------|---------|--------|---------|--------|---------|--------|
| n | $\ell = \ell_X = \ell_Y$ | # Παρ. \in Όρια (τουλάχιστον ℓ) | 100 | | 200 | | 500 | | 1000 | |
| | | | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR |
| 5 | 1 | 20% | 11 | 0.0511 | 23 | 0.0513 | 57 | 0.0493 | 116 | 0.0502 |
| | | | 13 | 0.0741 | 26 | 0.0749 | 66 | 0.0780 | 133 | 0.0793 |
| | | | 14 | 0.0957 | 29 | 0.0951 | 74 | 0.1028 | 150 | 0.1025 |
| | 2 | 40% | 10 | 0.0494 | 22 | 0.0541 | 54 | 0.0508 | 110 | 0.0523 |
| | | | 12 | 0.0721 | 25 | 0.0754 | 62 | 0.0704 | 125 | 0.0742 |
| | | | 14 | 0.1045 | 28 | 0.1044 | 70 | 0.0991 | 140 | 0.1046 |
| | 3 | 60% | 7 | 0.0483 | 15 | 0.0537 | 37 | 0.0512 | 75 | 0.0524 |
| | | | 9 | 0.0839 | 17 | 0.0746 | 42 | 0.0735 | 85 | 0.0708 |
| | | | 10 | 0.1180 | 19 | 0.1038 | 48 | 0.1033 | 95 | 0.1002 |
| | 4 | 80% | 2 | 0.0373 | 5 | 0.0458 | 13 | 0.0480 | 27 | 0.0540 |
| | | | 3 | 0.0703 | 7 | 0.0863 | 16 | 0.0723 | 33 | 0.0728 |
| | | | 4 | 0.1059 | 8 | 0.1040 | 19 | 0.1029 | 39 | 0.1025 |
| 11 | 4 | 36.4% | 17 | 0.0488 | 35 | 0.0470 | 90 | 0.0487 | 180 | 0.0501 |
| | | | 19 | 0.0790 | 38 | 0.0699 | 96 | 0.0687 | 193 | 0.0689 |
| | | | 20 | 0.0998 | 41 | 0.1026 | 102 | 0.0981 | 206 | 0.1033 |
| | 5 | 45.5% | 15 | 0.0560 | 30 | 0.0522 | 76 | 0.0491 | 152 | 0.0487 |
| | | | 16 | 0.0771 | 32 | 0.0749 | 81 | 0.0675 | 162 | 0.0703 |
| | | | 17 | 0.1007 | 34 | 0.0998 | 86 | 0.0976 | 172 | 0.0976 |
| | 6 | 54.5% | 11 | 0.0465 | 23 | 0.0530 | 59 | 0.0502 | 118 | 0.0537 |
| | | | 12 | 0.0667 | 25 | 0.0738 | 64 | 0.0784 | 127 | 0.0718 |
| | | | 13 | 0.0972 | 27 | 0.1039 | 68 | 0.1016 | 136 | 0.0979 |
| | 7 | 63.6% | 8 | 0.0511 | 16 | 0.0469 | 42 | 0.0484 | 84 | 0.0542 |
| | | | 9 | 0.0764 | 18 | 0.0738 | 46 | 0.0721 | 95 | 0.0841 |
| | | | 1 | 0.1071 | 20 | 0.1047 | 50 | 0.1036 | 100 | 0.1004 |
| 25 | 7 | 28% | 25 | 0.0500 | 52 | 0.0524 | 131 | 0.0519 | 264 | 0.0465 |
| | | | 26 | 0.0629 | 54 | 0.0759 | 136 | 0.0746 | 274 | 0.0718 |
| | | | 27 | 0.0966 | 56 | 0.1030 | 141 | 0.1010 | 284 | 0.1043 |
| | 8 | 23% | 24 | 0.0529 | 49 | 0.0526 | 124 | 0.0500 | 249 | 0.0494 |
| | | | 25 | 0.0779 | 51 | 0.0777 | 128 | 0.0721 | 257 | 0.0734 |
| | | | 26 | 0.1081 | 52 | 0.0984 | 132 | 0.0998 | 265 | 0.1013 |
| | 9 | 36% | 22 | 0.0532 | 45 | 0.0496 | 115 | 0.0520 | 229 | 0.0476 |
| | | | 23 | 0.0728 | 47 | 0.0730 | 119 | 0.0709 | 233 | 0.0727 |
| | | | 24 | 0.0984 | 49 | 0.1040 | 123 | 0.1008 | 247 | 0.0997 |
| | 10 | 40% | 20 | 0.0473 | 41 | 0.0478 | 106 | 0.0543 | 210 | 0.0476 |
| | | | 21 | 0.0652 | 43 | 0.0730 | 110 | 0.0756 | 219 | 0.0681 |
| | | | 22 | 0.1028 | 45 | 0.1016 | 114 | 0.1021 | 228 | 0.1038 |

Πίνακας 7.5: Πιθανότητες Λανθασμένου Συναγερμού FAR_{02N2} για δεδομένο σχεδιασμό ($a = c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$)

Ιδιαίτερη προσοχή πρέπει να δίνεται στην επιλογή των παραμέτρων σχεδιασμού ανάλογα με το είδος της μετατόπισης που επιθυμούμε να ανιχνεύσουμε. Ειδικότερα, στην περίπτωση που θέλουμε να εντοπίσουμε μεγαλύτερες μετατοπίσεις στη μέση τιμή του πρώ-

7.4. Μελέτη Απόδοσης και σχεδιασμός διαγράμματος ελέγχου

του χαρακτηριστικού, είναι λογικό να οριστεί $a > c$ και $\ell_X = \ell_Y$, ενώ όταν αναμένονται μεγαλύτερες μετατοπίσεις στην τυπική απόκλιση του ίδιου χαρακτηριστικού είναι προτιμότερη η επιλογή των $a = c$ και $\ell_X > \ell_Y$. Τα αποτελέσματα αυτά παρουσιάζονται στους Πίνακες 7.6 και 7.7, αντίστοιχα.

| Παράμετροι Σχεδιασμού | | | Μέγεθος Δείγματος Αναφοράς m | | | | | | | |
|-----------------------|--------------------------|--|--------------------------------|--------|---------|--------|-----------|--------|-----------|--------|
| n | $\ell = \ell_X = \ell_Y$ | # Παρ. \in Όρια (τουλάχιστον ℓ) | 100 | | 200 | | 500 | | 1000 | |
| | | | $a > c$ | FAR | $a > c$ | FAR | $a > c$ | FAR | $a > c$ | FAR |
| 5 | 1 | 20% | 12 > 9 | 0.0501 | 25 > 20 | 0.0523 | 60 > 55 | 0.0498 | 120 > 110 | 0.0503 |
| | | | 14 > 11 | 0.0748 | 30 > 20 | 0.0723 | 70 > 63 | 0.0794 | 140 > 125 | 0.0729 |
| | | | 16 > 13 | 0.1013 | 35 > 20 | 0.1032 | 80 > 71 | 0.1005 | 160 > 140 | 0.1023 |
| | 2 | 40% | 12 > 8 | 0.0510 | 23 > 19 | 0.0500 | 58 > 50 | 0.0518 | 120 > 100 | 0.0501 |
| | | | 14 > 8 | 0.0710 | 27 > 21 | 0.0749 | 72 > 56 | 0.0851 | 137 > 110 | 0.0741 |
| | | | 16 > 8 | 0.0989 | 31 > 23 | 0.1006 | 76 > 62 | 0.1004 | 154 > 120 | 0.1022 |
| | 3 | 60% | 8 > 6 | 0.0502 | 18 > 9 | 0.0506 | 39 > 33 | 0.0501 | 80 > 70 | 0.0506 |
| | | | 9 > 7 | 0.0740 | 20 > 11 | 0.0736 | 43 > 37 | 0.0701 | 90 > 80 | 0.0716 |
| | | | 10 > 8 | 0.0968 | 22 > 13 | 0.0979 | 53 > 41 | 0.1036 | 100 > 90 | 0.1000 |
| | 4 | 80% | 3 > 2 | 0.0489 | 6 > 4 | 0.0503 | 15 > 12 | 0.0523 | 32 > 20 | 0.0508 |
| | | | 4 > 2 | 0.0703 | 7 > 5 | 0.0678 | 20 > 13 | 0.0802 | 40 > 25 | 0.0780 |
| | | | 5 > 2 | 0.0995 | 9 > 6 | 0.1039 | 25 > 14 | 0.1044 | 48 > 30 | 0.1018 |
| 11 | 4 | 36.36% | 19 > 14 | 0.0498 | 38 > 30 | 0.0503 | 96 > 83 | 0.0509 | 200 > 160 | 0.0509 |
| | | | 20 > 16 | 0.0693 | 42 > 32 | 0.0767 | 102 > 90 | 0.0753 | 205 > 180 | 0.0701 |
| | | | 21 > 18 | 0.0976 | 44 > 34 | 0.1017 | 108 > 97 | 0.0987 | 210 > 200 | 0.0973 |
| | 5 | 45.45% | 16 > 14 | 0.0556 | 33 > 21 | 0.0510 | 80 > 68 | 0.0505 | 154 > 150 | 0.0508 |
| | | | 17 > 15 | 0.0823 | 35 > 25 | 0.0656 | 87 > 72 | 0.0727 | 167 > 160 | 0.0801 |
| | | | 18 > 16 | 0.0996 | 37 > 29 | 0.1010 | 94 > 76 | 0.1034 | 180 > 170 | 0.0989 |
| | 6 | 54.54% | 12 > 10 | 0.0502 | 26 > 19 | 0.0519 | 60 > 55 | 0.0509 | 123 > 110 | 0.0511 |
| | | | 13 > 11 | 0.0689 | 28 > 21 | 0.0779 | 66 > 58 | 0.0732 | 133 > 119 | 0.0691 |
| | | | 14 > 12 | 0.1008 | 30 > 23 | 0.1030 | 72 > 61 | 0.0982 | 143 > 128 | 0.1026 |
| | 7 | 63.63% | 9 > 6 | 0.0511 | 19 > 12 | 0.0525 | 44 > 38 | 0.0494 | 81 > 85 | 0.0509 |
| | | | 10 > 7 | 0.0726 | 21 > 13 | 0.0731 | 50 > 43 | 0.0802 | 94 > 89 | 0.0718 |
| | | | 11 > 8 | 0.1011 | 23 > 14 | 0.1014 | 52 > 48 | 0.1014 | 107 > 93 | 0.1014 |
| 25 | 7 | 28% | 26 > 24 | 0.0501 | 53 > 51 | 0.0519 | 120 > 105 | 0.0487 | 275 > 250 | 0.0513 |
| | | | 27 > 25 | 0.0709 | 55 > 52 | 0.0701 | 124 > 110 | 0.0647 | 283 > 260 | 0.0719 |
| | | | 28 > 26 | 0.1005 | 57 > 53 | 0.0986 | 128 > 115 | 0.1013 | 291 > 270 | 0.1013 |
| | 8 | 32% | 28 > 24 | 0.0493 | 52 > 41 | 0.0523 | 125 > 123 | 0.0500 | 264 > 210 | 0.0504 |
| | | | 30 > 24 | 0.0739 | 54 > 43 | 0.0764 | 130 > 125 | 0.0700 | 273 > 215 | 0.0753 |
| | | | 32 > 24 | 0.1095 | 56 > 45 | 0.1001 | 135 > 127 | 0.1036 | 282 > 220 | 0.1028 |
| | 9 | 36% | 23 > 20 | 0.0482 | 48 > 38 | 0.0499 | 120 > 105 | 0.0509 | 242 > 210 | 0.0498 |
| | | | 25 > 20 | 0.0849 | 50 > 40 | 0.0702 | 124 > 110 | 0.0682 | 251 > 215 | 0.0754 |
| | | | 26 > 20 | 0.1062 | 52 > 42 | 0.1027 | 128 > 115 | 0.0992 | 260 > 220 | 0.1013 |
| | 10 | 40% | 22 > 16 | 0.0535 | 44 > 36 | 0.0510 | 112 > 88 | 0.0499 | 219 > 201 | 0.0514 |
| | | | 23 > 17 | 0.0781 | 46 > 38 | 0.0736 | 116 > 94 | 0.0734 | 226 > 211 | 0.0696 |
| | | | 24 > 18 | 0.1041 | 48 > 40 | 0.1031 | 120 > 100 | 0.0970 | 233 > 221 | 0.1010 |

Πίνακας 7.6: Πιθανότητες Λανθασμένου Συναγερμού FAR_{O2N2} για δεδομένο σχεδιασμό ($a > c, b = m - a + 1, d = m - c + 1$ & $r = s = (n + 1)/2, \ell = \ell_X = \ell_Y$)

Δισδιάστατο Διάγραμμα Ελέγχου
για τη Θέση και τη Διασπορά με Χρήση Απαριθμητριών Στατιστικών Συναρτήσεων

| Παράμετροι Σχεδιασμού | | | | Μέγεθος Δείγματος Αναφοράς m | | | | | | | |
|-----------------------|-------------------|--------------------------------|----------|--------------------------------|--------|---------|--------|---------|--------|---------|--------|
| n | $\ell_X > \ell_Y$ | # Παρ. ∈ Όρια (τουλάχιστον) | | 100 | | 200 | | 500 | | 1000 | |
| | | ℓ_X | ℓ_Y | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR | $a = c$ | FAR |
| 5 | 2 > 1 | 40% | 20% | 11 | 0.0521 | 22 | 0.0502 | 57 | 0.0514 | 114 | 0.0519 |
| | | | | 13 | 0.0866 | 25 | 0.0728 | 65 | 0.0825 | 129 | 0.0747 |
| | | | | 14 | 0.1000 | 28 | 0.0997 | 73 | 0.1027 | 144 | 0.0995 |
| | 3 > 1 | 60% | 20% | 9 | 0.0574 | 18 | 0.0581 | 44 | 0.0524 | 90 | 0.0518 |
| | | | | 10 | 0.0791 | 20 | 0.0768 | 50 | 0.0768 | 101 | 0.0818 |
| | | | | 11 | 0.1002 | 22 | 0.1018 | 56 | 0.1014 | 112 | 0.1028 |
| | 3 > 2 | 60% | 40% | 8 | 0.0477 | 16 | 0.0452 | 40 | 0.0499 | 86 | 0.0504 |
| | | | | 10 | 0.0782 | 19 | 0.0699 | 47 | 0.0650 | 99 | 0.0754 |
| | | | | 11 | 0.1056 | 22 | 0.0984 | 54 | 0.0966 | 112 | 0.0996 |
| | 4 > 2 | 80% | 40% | 4 | 0.0549 | 7 | 0.0509 | 18 | 0.0501 | 38 | 0.0500 |
| | | | | 5 | 0.0824 | 9 | 0.0692 | 23 | 0.0744 | 47 | 0.0701 |
| | | | | 6 | 0.1178 | 11 | 0.1010 | 28 | 0.1012 | 56 | 0.1000 |
| 11 | 4 > 2 | 36.36% | 18.18% | 18 | 0.0502 | 37 | 0.0498 | 94 | 0.0510 | 186 | 0.0490 |
| | | | | 19 | 0.0663 | 40 | 0.0710 | 101 | 0.0756 | 201 | 0.0715 |
| | | | | 21 | 0.1016 | 43 | 0.1014 | 108 | 0.1002 | 216 | 0.0979 |
| | 5 > 3 | 45.45% | 27.27% | 16 | 0.0508 | 32 | 0.0483 | 83 | 0.0497 | 166 | 0.0500 |
| | | | | 17 | 0.0748 | 35 | 0.0682 | 89 | 0.0707 | 177 | 0.0730 |
| | | | | 18 | 0.1006 | 38 | 0.1044 | 95 | 0.1017 | 188 | 0.0975 |
| | 6 > 2 | 54.54% | 18.18% | 13 | 0.0506 | 27 | 0.0543 | 68 | 0.0538 | 132 | 0.0516 |
| | | | | 14 | 0.0710 | 29 | 0.0783 | 73 | 0.0746 | 145 | 0.0724 |
| | | | | 15 | 0.0965 | 31 | 0.1025 | 78 | 0.1017 | 158 | 0.1042 |
| | 7 > 4 | 63.63% | 36.36% | 10 | 0.0572 | 20 | 0.0546 | 49 | 0.0501 | 100 | 0.0517 |
| | | | | 11 | 0.0779 | 22 | 0.0786 | 54 | 0.0754 | 110 | 0.0765 |
| | | | | 12 | 0.1095 | 24 | 0.1009 | 59 | 0.0957 | 120 | 0.1024 |
| 25 | 7 > 3 | 28% | 12% | 26 | 0.0524 | 53 | 0.0501 | 136 | 0.0501 | 274 | 0.0521 |
| | | | | 27 | 0.0734 | 55 | 0.0647 | 141 | 0.0726 | 284 | 0.0754 |
| | | | | 28 | 0.0955 | 57 | 0.0951 | 146 | 0.1017 | 294 | 0.1003 |
| | 8 > 4 | 32% | 16% | 25 | 0.0507 | 51 | 0.0513 | 129 | 0.0526 | 261 | 0.0507 |
| | | | | 26 | 0.0754 | 53 | 0.0727 | 134 | 0.0744 | 270 | 0.0757 |
| | | | | 27 | 0.1009 | 55 | 0.1019 | 139 | 0.1031 | 279 | 0.1012 |
| | 9 > 5 | 36% | 20% | 23 | 0.0478 | 48 | 0.0516 | 121 | 0.0501 | 243 | 0.0514 |
| | | | | 24 | 0.0656 | 50 | 0.0754 | 126 | 0.0758 | 252 | 0.0687 |
| | | | | 25 | 0.0941 | 52 | 0.1049 | 131 | 0.1004 | 261 | 0.1006 |
| | 10 > 6 | 40% | 24% | 22 | 0.0534 | 44 | 0.0505 | 112 | 0.0499 | 226 | 0.0503 |
| | | | | 23 | 0.0797 | 46 | 0.0708 | 117 | 0.0718 | 235 | 0.0706 |
| | | | | 24 | 0.1065 | 48 | 0.0956 | 122 | 0.1013 | 244 | 0.1003 |

Πίνακας 7.7: Πιθανότητες Λανθασμένου Συναγεرمού FAR_{O2N2} για δεδομένο σχεδιασμό ($a = c$, $b = m - a + 1$, $d = m - c + 1$ & $r = s = (n + 1)/2$, $\ell_X > \ell_Y$)

7.5 Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

Στην τρέχουσα ενότητα το διάγραμμα $O2N2$ συγκρίνεται με το διάγραμμα $O4$, ακολουθώντας την ίδια λογική με αυτή της Ενότητας 6.5. Αρχικά, γίνεται μελέτη της απόδοσης των δύο σχημάτων με σκοπό να μελετηθεί η ικανότητα ανίχνευσης μετατοπίσεων μόνο στην τυπική απόκλιση. Για το σκοπό αυτό, χρησιμοποιήθηκαν κανονικά δεδομένα και δεδομένα από την κατανομή $t(5)$ με εντός ελέγχου τιμές $\rho_{in} = 0.5$ ($= \rho_{out}$), $\mu_X^{(R)} = \mu_Y^{(R)} = 0$ και $\sigma_X (= \sigma_X^{(R)}) = \sigma_Y (= \sigma_Y^{(R)}) = 1$.

7.5. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

Ο πίνακας συνδιακυμάνσεων $\Sigma_{t(5)}$ έχει ορισθεί κατάλληλα, ώστε να προκύπτει ο ίδιος πίνακας διακυμάνσεων-συνδιακυμάνσεων και για τις δύο κατανομές. Στη συνέχεια, ξεκινώντας με ένα κοινό ARL_{in} περίπου ίσο με 200, εκτιμήθηκε το εκτός ελέγχου ARL (ARL_{out})¹, μετατοπίζοντας τον πίνακα διακυμάνσεων-συνδιακυμάνσεων από

$$\Sigma_{in} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \text{σε} \quad \Sigma_{out} = \begin{pmatrix} \delta_X^2 \sigma_X^2 & \delta_X \delta_Y \sigma_{XY} \\ \delta_X \delta_Y \sigma_{XY} & \delta_Y^2 \sigma_Y^2 \end{pmatrix},$$

όπου τα δ_X, δ_Y χρησιμοποιούνται για τον έλεγχο/καθορισμό των μετατοπίσεων της μεταβλητότητας των χαρακτηριστικών X και Y , αντίστοιχα.

Από τα αποτελέσματα των αριθμητικών συγκρίσεων είναι εμφανές ότι το διάγραμμα $O2N2$ έχει μονίμως καλύτερη απόδοση από αυτή του διαγράμματος $O4$, ανεξάρτητα από την κατανομή των δεδομένων και το είδος των μετατοπίσεων (ίσων, άνισων). Στην περίπτωση που λάβουμε υπ' όψιν μας και τις τιμές του εκτός ελέγχου μέσου μήκους ροής που συνδέονται με τα διαγράμματα των Alt (1985), Costa and Machado (2009), Osei-Aning *et al.* (2017), ύστερα από σύγκριση που έγινε στην Ενότητα 6.5, παρατηρούμε ότι όταν τα δεδομένα είναι κανονικά το διάγραμμα $|S|$ είναι καλύτερο στην ανίχνευση ίσων μετατοπίσεων, ενώ το διάγραμμα $O2N2$ είναι πιο ευαίσθητο σε μικρότερες μετατοπίσεις έναντι του διαγράμματος S_{max} . Όταν τα δεδομένα προέρχονται από την κατανομή $t(5)$, το διάγραμμα $|S|$ είναι αποτελεσματικότερο στην ανίχνευση ίσων και μεγάλων μετατοπίσεων, ενώ το διάγραμμα $O2N2$ είναι πιο ευαίσθητο σε μικρότερες μετατοπίσεις. Όταν τα δεδομένα δεν είναι κανονικά, τα διαγράμματα $O2N2$ και $|S|$ συμπεριφέρονται σχεδόν το ίδιο.

Αξίζει να σημειωθεί ότι η απόδοση του προαναφερθέντος διαγράμματος $O4$ μπορεί να βελτιωθεί χρησιμοποιώντας μεγαλύτερα μεγέθη του δείγματος αναφοράς, καθώς τα όρια που προκύπτουν είναι πιο ακριβή και ως αποτέλεσμα ο κανόνας ελέγχου θεωρείται πιο ανθεκτικός. Προφανώς, τα όρια ελέγχου του νέου διαγράμματος $O2N2$ μπορούν να επιλεγούν ασύμμετρα σε σχέση με τα ποιοτικά χαρακτηριστικά με σκοπό την αντιμετώπιση περιπτώσεων στις οποίες εντοπίζονται άνισες μετατοπίσεις. Για παράδειγμα, όταν η εκτός ελέγχου κατάσταση οφείλεται σε μεγαλύτερες μετατοπίσεις του πρώτου χαρακτηριστικού, είναι εύλογο να επιλέξουμε τις παραμέτρους έτσι, ώστε $\ell_X < \ell_Y$. Τα αποτελέσματα συνοψίζονται στους Πίνακες 7.8 και 7.9, όπου οι τιμές με έντονους χαρακτήρες δείχνουν τη μικρότερη τιμή του ARL_{out} . Το μέγεθος δείγματος που χρησιμοποιήθηκε ήταν $n = 5$ και οι παράμετροι σχεδιασμού ήταν $b = m - a + 1, d = m - c + 1, r = s = (n + 1)/2$.

¹ Τα βήματα της αλγοριθμικής διαδικασίας σύμφωνα με την οποία υπολογίζονται το εντός και το εκτός ελέγχου μέσο μήκος ροής είναι παρόμοια με αυτά που παρουσιάζονται στον Πίνακα 4.4, με τη διαφορά ότι στο Βήμα 5 διαπιστώνεται εάν παραβιάζεται ή όχι ο κανόνας ελέγχου της Σχέσης (7.2).

| Μετατοπίσεις Τυπ. Απόκλ. | Παραμετρικά Διαγράμματα | | | Μη Παραμετρικά Διαγράμματα | | | Ημιπαραμετρικά Διαγράμματα | | |
|-----------------------------|-------------------------|---------------|---------------|----------------------------|---------------|---------------|---|---|--|
| | $ S $ | R_{max} | Q_{max} | S_{max} | MD_{max} | MAD_{max} | $O2N2$ | $O4$ | |
| $\delta_X > \delta_Y$ | $UCL = 4.0314$ | $UCL = 5.145$ | $UCL = 2.238$ | $UCL = 2.024$ | $UCL = 1.476$ | $UCL = 2.567$ | $m = 138$ $a = c = 3$ $\ell_X = \ell_Y = 4$ | $m = 245$ $a = c = 3$ $t = t_X = t_Y = 1$ | |
| 1.00 | 200.92 | 199.78 | 200.40 | 201.84 | 200.72 | 201.71 | 199.37 | 200.28 | |
| 1.25 | 15.45 | 17.93 | 28.30 | 16.16 | 17.88 | 30.13 | 22.17 | 24.18 | |
| 1.50 | 4.53 | 5.13 | 9.94 | 4.65 | 5.32 | 10.15 | 7.23 | 7.66 | |
| 1.75 | 2.37 | 2.64 | 4.96 | 2.44 | 2.65 | 5.15 | 3.37 | 3.88 | |
| 2.00 | 1.67 | 1.80 | 3.21 | 1.69 | 1.80 | 3.39 | 2.44 | 2.52 | |
| $\delta_X > \delta_Y$ | $UCL = 4.0314$ | $UCL = 5.145$ | $UCL = 2.238$ | $UCL = 2.024$ | $UCL = 1.476$ | $UCL = 2.567$ | $m = 375$ $a = c = 5$ $\ell_X = 4 > 2 = \ell_Y$ | $m = 535$ $a = c = 9$ $t_X = 1 > 0 = t_Y$ | |
| 1.00 | 200.60 | 199.18 | 199.36 | 200.16 | 200.41 | 201.99 | 201.50 | 200.40 | |
| 1.25 | 46.43 | 31.49 | 48.91 | 28.65 | 31.91 | 29.85 | 23.72 | 29.66 | |
| 1.50 | 18.66 | 9.06 | 17.62 | 8.15 | 9.20 | 10.25 | 7.62 | 9.87 | |
| 1.75 | 9.82 | 4.34 | 8.79 | 3.95 | 4.41 | 5.32 | 3.94 | 5.23 | |
| 2.00 | 6.27 | 2.75 | 5.53 | 2.55 | 2.83 | 3.67 | 2.61 | 3.47 | |
| 2.25 | 4.50 | 2.05 | 4.04 | 1.93 | 2.09 | 2.52 | 2.00 | 2.55 | |
| 2.50 | 3.46 | 1.69 | 3.21 | 1.61 | 1.72 | 2.05 | 1.66 | 2.11 | |

Πίνακας 7.8: Απόδοση του διαγράμματος $O2N2$ βάσει του ARL_{out} σε σχέση με 7 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων τυπικής απόκλισης χρησιμοποιώντας δεδομένα από κανονική κατανομή για δεδομένο σχεδιασμό

7.5. Σύγκριση με υπάρχοντα ανταγωνιστικά σχήματα

| Μετατοπίσεις | | Παραμετρικά Διαγράμματα | | | Μη Παραμετρικά Διαγράμματα | | | Ημιπαραμετρικά Διαγράμματα | | |
|-----------------------|---------------|-------------------------|---------------|---------------|----------------------------|---------------|---|---|--|--|
| Τυπ. Απόκλ. | $ S $ | $Rmax$ | $Qmax$ | $Smax$ | $MDmax$ | $MADmax$ | $O2N2$ | $O4$ | | |
| $\delta_X = \delta_Y$ | $UCL = 7.512$ | $UCL = 7.320$ | $UCL = 2.206$ | $UCL = 2.907$ | $UCL = 1.809$ | $UCL = 2.594$ | $m = 138$ $a = c = 3$ $\ell_X = \ell_Y = 4$ | $m = 245$ $a = c = 3$ $t = t_X = t_Y = 1$ | | |
| 1.00 | 200.71 | 199.76 | 200.84 | 201.21 | 200.36 | 200.49 | 201.64 | 201.86 | | |
| 1.25 | 38.31 | 59.00 | 41.04 | 59.75 | 45.82 | 41.98 | 37.58 | 50.35 | | |
| 1.50 | 12.23 | 21.93 | 14.85 | 23.68 | 15.78 | 15.67 | 12.61 | 17.83 | | |
| 1.75 | 5.78 | 10.94 | 7.45 | 11.09 | 7.35 | 7.90 | 6.25 | 8.92 | | |
| 2.00 | 3.42 | 6.25 | 4.55 | 6.30 | 4.25 | 4.93 | 3.83 | 5.27 | | |
| $\delta_X > \delta_Y$ | $UCL = 7.512$ | $UCL = 7.320$ | $UCL = 2.206$ | $UCL = 2.907$ | $UCL = 1.809$ | $UCL = 2.594$ | $m = 375$ $a = c = 5$ $\ell_X = 4 > 2 = \ell_Y$ | $m = 535$ $a = c = 9$ $t_X = 4 > 2 = t_Y$ | | |
| 1.00 | 200.62 | 200.67 | 200.42 | 199.41 | 201.32 | 199.06 | 199.37 | 201.41 | | |
| 1.25 | 84.79 | 85.68 | 67.59 | 89.44 | 71.59 | 68.45 | 46.46 | 51.43 | | |
| 1.50 | 43.80 | 37.39 | 26.38 | 38.73 | 26.60 | 27.64 | 17.54 | 19.70 | | |
| 1.75 | 26.11 | 18.41 | 13.29 | 18.94 | 12.31 | 14.31 | 8.44 | 10.09 | | |
| 2.00 | 17.14 | 10.44 | 8.08 | 10.45 | 6.91 | 8.84 | 5.12 | 6.23 | | |
| 2.25 | 12.20 | 6.58 | 5.62 | 6.68 | 4.53 | 6.07 | 3.49 | 4.38 | | |
| 2.50 | 9.11 | 4.30 | 4.30 | 4.60 | 3.32 | 4.55 | 2.71 | 3.32 | | |

Πίνακας 7.9: Απόδοση του διαγράμματος $O2N2$ βάσει του ARL_{out} σε σχέση με 7 ανταγωνιστικά διαγράμματα με κοινό $ARL_{in} \approx 200$ για την ανίχνευση μετατοπίσεων τυπικής απόκλισης χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό

| Μετατοπίσεις | | Διάγραμμα O4 | | | | |
|-----------------------|--------------------|---------------|---------------|--------------|--------------|--------------|
| Μέσης τιμής | Σχεδιασμός | ρ | | | | |
| $\mu = \mu_X = \mu_Y$ | $ARL_{in} = 200.2$ | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 |
| 0.00 | $a = c = 3$ | 188.64 | 158.66 | 90.47 | 43.58 | 20.91 |
| 0.25 | $t = 1$ | 124.07 | 105.82 | 63.63 | 32.03 | 16.52 |
| 0.50 | $n = 5$ | 48.10 | 41.93 | 28.26 | 16.76 | 10.11 |
| 0.75 | $m = 255$ | 18.50 | 16.90 | 12.41 | 8.45 | 5.92 |
| 1.00 | | 7.62 | 7.24 | 5.95 | 4.52 | 3.66 |

| Μετατοπίσεις | | Διάγραμμα O2N2 | | | | |
|-----------------------|-----------------------|----------------|---------------|--------------|-------------|-------|
| Μέσης τιμής | Σχεδιασμός | ρ | | | | |
| $\mu = \mu_X = \mu_Y$ | $ARL_{in} = 205.1$ | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 |
| 0.00 | $a = c = 5$ | 198.34 | 170.00 | 104.04 | 55.48 | 28.25 |
| 0.25 | $\ell_X = \ell_Y = 2$ | 116.34 | 103.23 | 67.91 | 38.53 | 21.77 |
| 0.50 | $n = 5$ | 39.71 | 36.50 | 27.16 | 18.46 | 12.26 |
| 0.75 | $m = 100$ | 13.76 | 12.71 | 10.79 | 8.69 | 6.62 |
| 1.00 | | 5.56 | 5.46 | 4.94 | 4.35 | 3.85 |

Πίνακας 7.10: Απόδοση διαγραμμάτων O4 και O2N2 βάσει του ARL_{out} με κοινό $ARL_{in} \approx 200$ για την ανίχνευση ταυτόχρονων μετατοπίσεων μέσης τιμής και μεταβλητότητας χρησιμοποιώντας δεδομένα από κατανομή $t(5)$ για δεδομένο σχεδιασμό

Στον Πίνακα 7.10, έγινε σύγκριση των δύο διαγραμμάτων ακολουθώντας το σενάριο των Zamba and Hawkins (2009) που περιγράφηκε στην Ενότητα 6.5. Στον πίνακα αυτό φαίνονται οι τιμές των ταυτόχρονων μετατοπίσεων της μέσης τιμής και του πίνακα διακυμάνσεων-συνδιακυμάνσεων

$$\Sigma = \begin{pmatrix} 1 + \rho^2 & \rho^2 \\ \rho^2 & 1 + \rho^2 \end{pmatrix},$$

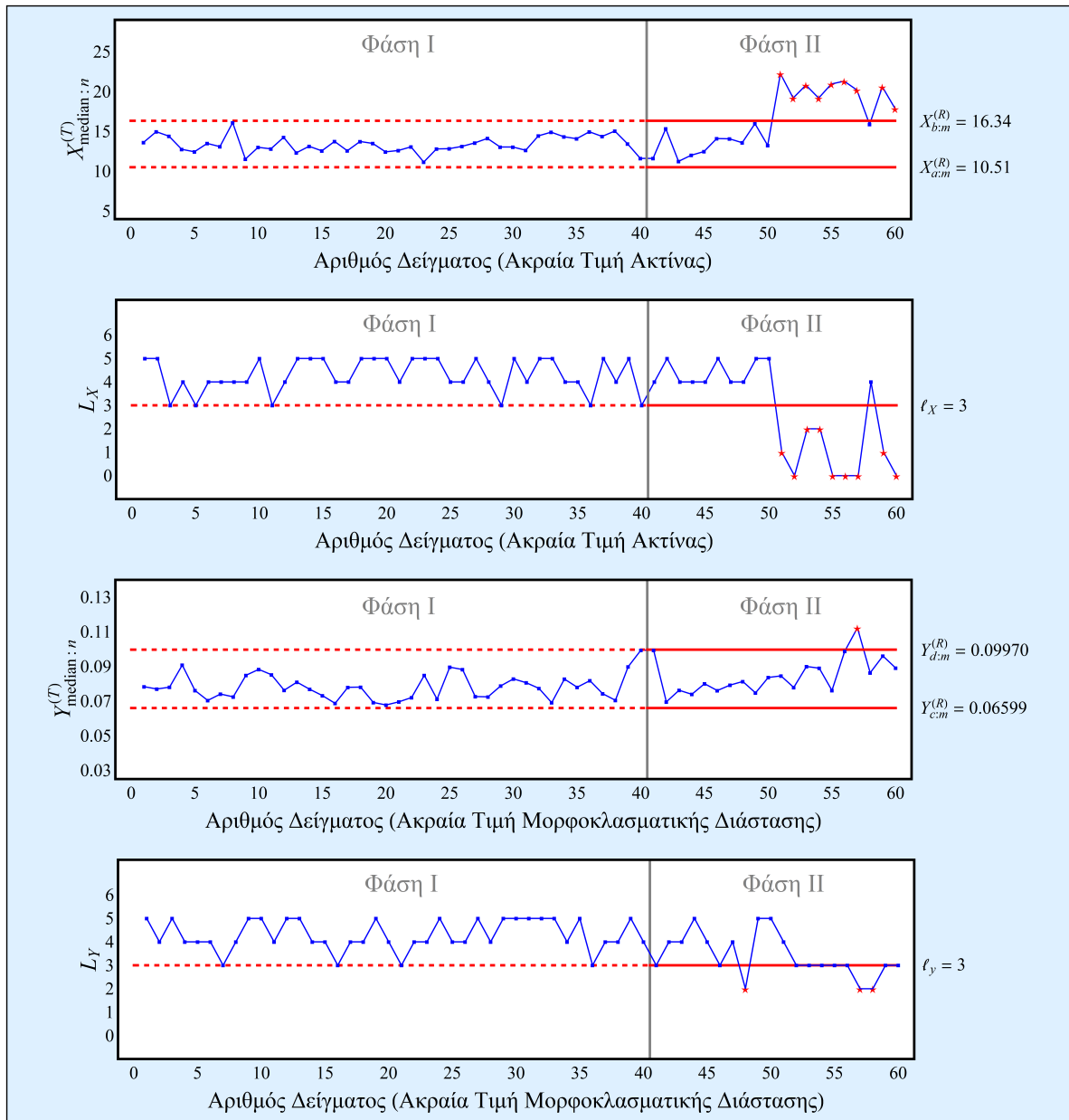
σε συνδυασμό με τις μετατοπίσεις της παραμέτρου συσχέτισης ρ . Παρατηρούμε ότι κανένα από τα διαγράμματα δεν έχει ομοιόμορφα καλύτερη συμπεριφορά. Συγκεκριμένα, το διάγραμμα O4 είναι πιο ευαίσθητο σε μεγάλες μετατοπίσεις του ρ (στοιχεία στοιχεία με έντονη γραφή πάνω από την κύρια διαγώνιο), ενώ ο διάγραμμα O2N2 φαίνεται καταλληλότερο στην ανίχνευση μετατοπίσεων της μέσης τιμής (μικρών ή μεγάλων), δεδομένου ότι το ρ λαμβάνει σχετικά μικρές τιμές (στοιχεία με έντονη γραφή κάτω από την κύρια διαγώνιο). Βέβαια, τα αποτελέσματα μπορεί να διαφοροποιούνται εάν για παράδειγμα επιλέξουμε άλλες παραμέτρους σχεδίασης με σκοπό να δώσουμε περισσότερο βάρος στον εντοπισμό μετατοπίσεων είτε της μέσης τιμής είτε της μεταβλητότητας. Συνολικά, θα μπορούσαμε να ισχυρισθούμε ότι οι διαφορές αυτές μεταξύ των δύο διαγραμμάτων είναι μικρές, γεγονός που υποδεικνύει ότι τα διαγράμματα O2N2 και O4 συμπεριφέρονται παρόμοια. Τέλος, σε

συνδυασμό με το ότι δεν απαιτούν κάποια υπόθεση σχετικά με την κατανομή της διεργασίας, μπορούν να χρησιμοποιηθούν άνισα όρια ελέγχου, κάτι το οποίο δεν είναι σύνηθες σε άλλα αντίστοιχα διαγράμματα ελέγχου. Ως αποτέλεσμα, τα προτεινόμενα διαγράμματα μπορούν να γίνουν ανταγωνιστικά και να προτιμηθούν έναντι άλλων ομολόγων τους.

7.6 Εφαρμογή σε πραγματικά δεδομένα

Στην ενότητα αυτή, εφαρμόζεται το νέο διάγραμμα κάνοντας χρήση του συνόλου πραγματικών δεδομένων WDBC της έρευνας για τη διάγνωση του καρκίνου του μαστού που χρησιμοποιήθηκε στο προηγούμενο κεφάλαιο (βλ. Ενότητα 6.6). Βασιζόμενοι στην ίδια λογική, 40 καλοήθεις όγκοι συνιστούν το δείγμα Φάσης I (δείγμα αναφοράς μεγέθους $m = 200 = 40 \times 5$), με τη βοήθεια του οποίου καθορίζονται κατάλληλα όρια ελέγχου έτσι, ώστε να μην παράγεται σήμα για εκτός ελέγχου μετατοπίσεις. Στη συνέχεια, επιλέγονται 10 επιπλέον καλοήθεις και 10 κακοήθεις ομάδες όγκων μεγέθους $n = 5$ με σκοπό τη δημιουργία του δείγματος Φάσης II (δείγμα ελέγχου). Δεδομένου ότι επιθυμούμε ένα συγκεκριμένο ποσοστό λανθασμένου συναγερμού, ίσο με $FAR_{02N2} = 5\%$, οι τιμές των παραμέτρων σχεδίασης που επιλέγονται βάσει του Πίνακα 7.5 είναι $a = c = 15$, $b = d = 200 - 15 + 1 = 186$, $\ell_X = \ell_Y = 3$ και $r = s = (n + 1)/2 = 3$. Το γράφημα που προκύπτει απεικονίζεται στο Σχήμα 7.1.

Από το Σχήμα 7.1 παρατηρούμε ότι μόνο ένα από τα 10 καλοήθη δείγματα δεν έχει αναγνωριστεί σωστά (βλ. παρατήρηση 48 στο τελευταίο γράφημα), ενώ όλα τα κακοήθη δείγματα έχουν σωστά τεθεί εκτός ελέγχου λόγω του ότι η μέση τιμή και η διασπορά του πρώτου χαρακτηριστικού έχει μετατοπισθεί. Προφανώς, το νέο διάγραμμα παράγει εκτός ελέγχου σήμα στο 48ο και 58ο δείγμα εξαιτίας μετατόπισης της διασποράς του δεύτερου χαρακτηριστικού, καθώς και στο 57ο δείγμα λόγω ταυτόχρονων μετατοπίσεων στη θέση και κλίμακα τόσο της ακραίας τιμής που λαμβάνει η ακτίνα όσο και η μορφοκλασματική διάσταση του κυτταρικού πυρήνα. Είναι περιττό να αναφερθεί ότι προκύπτουν ακόμη καλύτερα αποτελέσματα στην περίπτωση που επιλέξουμε $\ell_X \neq \ell_Y$ και $a \neq c$, διότι η παραπάνω ανάλυση αποκαλύπτει πως δεν προκαλούνται ίσες μετατοπίσεις στα δύο χαρακτηριστικά.



Σχήμα 7.1: Εφαρμογή του διαγράμματος $O2N2$ στο σύνολο δεδομένων WDBC

ΚΕΦΑΛΑΙΟ 8

Σύνοψη

8.1 Συμπεράσματα

Το αντικείμενο της διδακτορικής διατριβής είναι η μελέτη νέων κατανομών δισδιάστατων διατεταγμένων παρατηρήσεων και πλήθους αυτών, οι οποίες βρίσκουν πρακτική εφαρμογή στην κατασκευή τεσσάρων ημιπαραμετρικών διαγραμμάτων ελέγχου διεργασιών που προτείνουμε. Αρχικά, στο Κεφάλαιο 1, παρουσιάζονται κάποιες βασικές έννοιες σχετικά με τον ποιοτικό έλεγχο και το βασικό του εργαλείο, τα διαγράμματα ελέγχου. Επίσης, γίνεται ανασκόπηση της βιβλιογραφίας που συνδέεται με διαγράμματα ελέγχου, εστιάζοντας σε πολυδιάστατα μη παραμετρικά σχήματα τύπου *Shewhart*.

Στη συνέχεια, στο Κεφάλαιο 2, εισάγονται οι βασικές έννοιες που σχετίζονται με τη Θεωρία των Διατεταγμένων Παρατηρήσεων και των Συμμεταβλητών τους, καθώς επίσης και με τη Θεωρία των Συνδέσμων. Στο Κεφάλαιο 3, παρουσιάζονται νέες κατανομές δισδιάστατων διατεταγμένων παρατηρήσεων και απεριθμητριών αυτών, οι οποίες αξιολογούνται στα επόμενα κεφάλαια για τον ακριβή προσδιορισμό εκφράσεων των χαρακτηριστικών (π.χ. χαρακτηριστική συνάρτηση λειτουργίας, ποσοστό συναγερού και ποσοστό λανθασμένου συναγερού) τεσσάρων νέων ημιπαραμετρικών διαγραμμάτων ελέγχου.

Συγκεκριμένα, στο Κεφάλαιο 4, προτείνεται ένα νέο ημιπαραμετρικό δισδιάστατο διάγραμμα ελέγχου, συμβ. *O2*, για την παρακολούθηση της μέσης τιμής μιας διεργασίας. Το *O2* βασίζεται σε διατεταγμένες παρατηρήσεις και αποτελεί γενίκευση σε δύο διαστάσεις των κλασικών μονοδιάστατων διαγραμμάτων ελέγχου για τη διάμεσο (*Janacek and Meikle (1997)*) ή για οποιοδήποτε ποσοστημόριο της υποκείμενης διεργασίας (*Chakraborti et al. (2000, 2004)*). Το διάγραμμα ελέγχου *O2* προσφέρει ένα απλό εργαλείο, η απόδοση του οποίου είναι ισοδύναμη ή καλύτερη (σε περιπτώσεις ασύμμετρων κατανομών) από την απόδοση των ανταγωνιστικών μη παραμετρικών διαγραμμάτων ελέγχου (SR^2 και SN^2

των Boone and Chakraborti (2012)).

Παρά το γεγονός ότι το νέο διάγραμμα ελέγχου επηρεάζεται τυπικά από τη δομή εξάρτησης των υπό μελέτη χαρακτηριστικών, δε μεταβάλλονται σημαντικά οι τιμές των FAR και ARL_{in} για όλους τους συνδέσμους που χρησιμοποιούνται στην αριθμητική μας μελέτη. Αυτή είναι μια ελκυστική ιδιότητα που επιτρέπει τη χρήση του νέου διαγράμματος ελέγχου ως πλήρως μη παραμετρικό. Το κύριο πλεονέκτημα του νέου διαγράμματος σε σύγκριση με άλλα μη παραμετρικά διαγράμματα που είναι διαθέσιμα στη βιβλιογραφία είναι η απλότητα κατασκευής του και το γεγονός ότι δε γίνονται πρόσθετες υποθέσεις σχετικά με τις κατανομές των δισδιάστατων μεταβλητών. Για παράδειγμα, το διάγραμμα SR^2 μπορεί να χρησιμοποιηθεί μόνο υπό την προϋπόθεση ότι η κατανομή είναι διαγώνια συμμετρική, διαφορετικά η σ.σ. δεν μπορεί να υπολογισθεί, ενώ και στα δύο διαγράμματα SR^2 και SN^2 απαιτείται η αντιστροφή της εκτίμησης του πίνακα διακύμανσης-συνδιακύμανσης της διεργασίας.

Το Κεφάλαιο 5, πραγματεύεται την κατασκευή ενός δισδιάστατου ημιπαραμετρικού σχήματος, συμβ. OC , το οποίο χρησιμοποιεί την ίδια σ.σ. με αυτή του διαγράμματος $O2$, αλλά εφαρμόζει διαφορετική διαδικασία για την επιλογή των ορίων ελέγχου. Ειδικότερα, το διάγραμμα OC είναι κατάλληλο για την ανίχνευση πιθανών μετατοπίσεων της μέσης τιμής και ταυτόχρονα επηρεάζεται από τη δομή εξάρτησης των ποιοτικών χαρακτηριστικών (όπως αυτή περιγράφεται μέσω του συνδέσμου που αντιστοιχεί στη δισδιάστατη από κοινού α.σ.κ. της διεργασίας), αλλά όχι από τις μονοδιάστατες περιθώριες κατανομές. Αυτό οφείλεται στο ότι για τον καθορισμό/προσδιορισμό των ορίων ελέγχου γίνεται χρήση ζευγαριών, που ορίζουν συγκεκριμένες διατεταγμένες παρατηρήσεις με τις συμμεταβλητές τους. Συνεπώς, διατηρείται η δομή συσχέτισης των αρχικών ζευγαριών που συνιστούν τις δισδιάστατες παρατηρήσεις προς μελέτη. Λόγω αυτής της ιδιότητας το διάγραμμα είναι πιο αποτελεσματικό στην περίπτωση συνδέσμων με μέτριο έως υψηλό εύρος συσχέτισης.

Αξίζει να σημειωθεί ότι η απόδοση του διαγράμματος OC είναι καλύτερη από αυτή του κλασσικού παραμετρικού διαγράμματος χ^2 όταν γίνεται χρήση μη κανονικών δεδομένων, ενώ όταν παράγονται δεδομένα από την κατανομή t , η συμπεριφορά του είναι παρόμοια με αυτή των διαγραμμάτων SR^2 , SN^2 και $O2$. Ιδιαίτερη, όμως, έμφαση θα πρέπει να δοθεί στο γεγονός ότι η επίδοση του διαγράμματος OC βελτιώνεται θεαματικά όταν στον κανόνα απόφασης εφαρμόζονται μη συμμετρικά όρια ελέγχου. Τέτοια όρια μπορούν εύκολα να προσδιορισθούν, ώστε να επιτευχθούν οι επιθυμητές τιμές των χαρακτηριστικών FAR και ARL_{in} , εκτελώντας επαναλαμβανόμενες δοκιμές και επιλέγοντας τις πλέον κατάλληλες

τιμές για τις παραμέτρους σχεδίασης.

Στο Κεφάλαιο 6, τροποποιείται ο κανόνας του διαγράμματος $O2$ με σκοπό την κατασκευή ενός νέου δισδιάστατου ημιπαραμετρικού διαγράμματος ελέγχου για την ανίχνευση μετατοπίσεων τόσο στη μέση τιμή όσο και στη διασπορά μιας παραγωγικής διαδικασίας. Πιο συγκεκριμένα, το νέο διάγραμμα (συμβ. $O4$) βασίζεται σε τέσσερις διατεταγμένες παρατηρήσεις και το πλεονέκτημά του είναι ότι στον κανόνα απόφασης λαμβάνεται υπ' όψιν η θέση δύο ζευγαριών δισδιάστατων διατεταγμένων παρατηρήσεων που τοποθετούνται μεταξύ των ορίων ελέγχου.

Τέλος, στο Κεφάλαιο 7, παρουσιάζεται ένα ακόμη δισδιάστατο ημιπαραμετρικό διάγραμμα ελέγχου (συμβ. $O2N2$) για ταυτόχρονη παρακολούθηση της μέσης τιμής και της μεταβλητότητας της υποκείμενης διεργασίας, το οποίο είναι γενίκευση του μονοδιάστατου σχήματος που προτάθηκε από τους Balakrishnan *et al.* (2010). Η κατασκευή του διαγράμματος $O2N2$ είναι σχετικά απλή, καθώς στον κανόνα απόφασης του διαγράμματος $O2$ προστίθεται μία επιπλέον συνθήκη που ενσωματώνει ένα μέτρο διασποράς. Κατά συνέπεια, λαμβάνεται πληροφορία όχι μόνο για τη θέση ενός ζεύγους δισδιάστατων διατεταγμένων παρατηρήσεων, αλλά και για το πλήθος παρατηρήσεων που ανήκουν στην περιοχή ελέγχου.

Τα αποτελέσματα της αριθμητικής σύγκρισης μεταξύ των δύο τελευταίων διαγραμμάτων ($O4$, $O2N2$) και αντίστοιχων ανταγωνιστικών παραμετρικών διαγραμμάτων (Abdelhamid (1985), Khoo (2004), Costa and Machado (2009)) ή μη παραμετρικών διαγραμμάτων ελέγχου (Liu (1995), Osei-Aning *et al.* (2017)) αποκαλύπτει ότι τα προτεινόμενα διαγράμματα $O4$ και $O2N2$ έχουν καλύτερη απόδοση, κυρίως όταν παρατηρούνται άνισες μετατοπίσεις μεταξύ των χαρακτηριστικών. Είναι αξιοπρόσεκτο ότι τα παραμετρικά διαγράμματα που χρησιμοποιήθηκαν στις προαναφερθείσες συγκρίσεις, με δεδομένα που προέρχονταν από κανονική κατανομή, δεν ήταν ομοιόμορφα πιο αποτελεσματικά, όπως κανείς θα περίμενε λόγω της κανονικότητας. Στην περίπτωση που χρησιμοποιήθηκαν δεδομένα από την κατανομή t , τα διαγράμματα $O4$ και $O2N2$ είχαν τις περισσότερες φορές καλύτερη απόδοση σε σχέση με τα άλλα μη παραμετρικά που είναι διαθέσιμα στη διεθνή βιβλιογραφία και έχουν προταθεί για την παρακολούθηση της θέσης, της κλίμακας ή την ταυτόχρονη παρακολούθηση θέσης και κλίμακας της κατανομής μιας διεργασίας.

Ενδιαφέρον παρουσιάζει το γεγονός ότι, το διάγραμμα $O2$ είναι ουσιαστικά ειδική περίπτωση των διαγραμμάτων $O4$ και $O2N2$. Και τα τρία αυτά διαγράμματα διατηρούν τη χαρακτηριστική ιδιότητα όλων των μη παραμετρικών διαγραμμάτων και μπορούν να

αξιοποιηθούν ως πλήρως μη παραμετρικά διαγράμματα ελέγχου. Αυτό συμβαίνει διότι οι τιμές του FAR και του ARL_{in} δεν επηρεάζονται καθόλου από τις περιθώριες κατανομές και την ίδια στιγμή παραμένουν σχεδόν αμετάβλητες όταν χρησιμοποιούνται διαφορετικοί σύνδεσμοι. Αντίθετα, ο κανόνας του διαγράμματος $O2$ ενώ δεν επηρεάζεται από την επιλογή των περιθώριων κατανομών, επηρεάζεται από την επιλογή του συνδέσμου (δηλαδή από τη δομή εξάρτησης). Συνεπώς, όλα τα προτεινόμενα διαγράμματα, χαρακτηρίζονται ως ημιπαραμετρικά.

Επιπρόσθετα, ιδιαίτερα σημαντικό είναι το γεγονός ότι με τα νέα διαγράμματα που προτάθηκαν και μελετήθηκαν στην παρούσα διατριβή είναι εφικτή η ανίχνευση του χαρακτηριστικού, το οποίο πιθανώς να είναι υπαίτιο για την εκτός ελέγχου κατάσταση της διεργασίας. Το διάγραμμα $O2N2$ έχει την πρόσθετη ευελιξία να προσδιορίζει και το είδος της μετατόπισης, δηλαδή εντοπίζει αν υπάρχει μετατόπιση στη μέση τιμή, στη μεταβλητότητα ή και στα δύο. Θα πρέπει, επίσης, να σημειωθεί ότι σε όλα τα διαγράμματα ελέγχου μπορούν να εφαρμοσθούν ασύμμετρα όρια ελέγχου, το οποία μπορούν να προσδιορισθούν ανάλογα με το είδος της μετατόπισης που επιθυμούμε να ανιχνεύσουμε (π.χ. μετατόπιση στο πρώτο ή/και δεύτερο χαρακτηριστικό, μετατόπιση στη μέση τιμή ή/και στη διασπορά).

Λαμβάνοντας υπ' όψιν ότι τα διαγράμματα $O2$, OC , $O4$, $O2N2$ δεν απαιτούν καμία υπόθεση αναφορικά με την κατανομή των υπό μελέτη μεταβλητών ή τις ιδιότητες της κατανομής αυτής, φαίνεται να είναι χρήσιμα εργαλεία για τα τμήματα ποιοτικού ελέγχου που ενδιαφέρονται να τα αξιοποιήσουν, καθώς μπορούν να δώσουν εναλλακτική λύση σε περιπτώσεις στις οποίες ο τύπος των χαρακτηριστικών που παρακολουθούνται δεν επιτρέπει την εφαρμογή άλλων αντίστοιχων διαγραμμάτων. Τέλος, το βασικότερο πλεονέκτημα όλων των διαγραμμάτων ελέγχου που προτάθηκαν στα πλαίσια της παρούσας διδακτορικής εργασίας (πέρα από την υψηλή απόδοση συγκριτικά με άλλα διαγράμματα) είναι η ευκολία κατασκευής τους. Αυτό έχει ως αποτέλεσμα να μπορούν εύκολα να γενικευθούν, ώστε να είναι ικανά για την ταυτόχρονη παρακολούθηση περισσότερων των δύο ποιοτικών χαρακτηριστικών με το μόνο κόστος ότι προκύπτουν πιο περίπλοκες μαθηματικές εκφράσεις για τη μελέτη των χαρακτηριστικών τους.

8.2 Προτάσεις για Μελλοντική Έρευνα

Σε κάθε πειραματική διαδικασία το ενδιαφέρον εστιάζεται στην παρατήρηση, μελέτη και κατανόηση φαινομένων, τα οποία είναι συνήθως από τη φύση τους πολυμεταβλητά. Η διάδοση της χρήσης των ηλεκτρονικών υπολογιστών και η ραγδαία ανάπτυξη της υπο-

λογιστικής ισχύος, σε συνδυασμό με την αφθονία διαφορετικών στατιστικών πακέτων, διευκόλυνε τη συλλογή δεδομένων και οδήγησε στην ανάπτυξη στατιστικών τεχνικών για την ανάλυση των παρατηρήσεων, καθώς και την παρακολούθηση/επιτήρηση των διεργασιών που εμπλέκονται με τα φαινόμενα αυτά. Δεδομένου ότι στην πραγματικότητα τα υπό μελέτη χαρακτηριστικά είναι συσχετισμένα μεταξύ τους, η εφαρμογή πολλαπλών μονοδιάστατων τεχνικών μπορεί να είναι παραπλανητική, διότι οι σχέσεις που αναπτύσσονται μεταξύ των χαρακτηριστικών δε λαμβάνονται υπ' όψιν.

Στον ποιοτικό έλεγχο διεργασιών, όταν εμπλέκονται δύο ή περισσότερες μεταβλητές, τα συνήθη στατιστικά εργαλεία που χρησιμοποιούνται είναι τα πολυμεταβλητά διαγράμματα ελέγχου. Όπως έγινε σαφές και από το Κεφάλαιο 1 (Ενότητα 1.5), ο αριθμός των μη παραμετρικών σχημάτων που έχουν αναπτυχθεί στη διεθνή βιβλιογραφία είναι περιορισμένος. Κατά συνέπεια, η κατασκευή νέων πολυδιάστατων μη παραμετρικών διαγραμμάτων ελέγχου μπορεί να αποτελέσει ένα πεδίο που χρήζει περαιτέρω μελέτης για μελλοντική έρευνα.

Στο παραπάνω πλαίσιο, η παρούσα έρευνα θα μπορούσε να επεκταθεί στο σχεδιασμό-γενίκευση των προτεινόμενων σχημάτων σε μεγαλύτερες διαστάσεις. Για το σκοπό αυτό, απαραίτητη κρίνεται η μελέτη νέων κατανομών που σχετίζονται με πολυδιάστατες διατεταγμένες παρατηρήσεις ή/και απαριθμήτριες συναρτήσεις αυτών, οι οποίες θα αξιοποιηθούν στον προσδιορισμό μαθητικών εκφράσεων για τα χαρακτηριστικά των νέων διαγραμμάτων. Βέβαια, λόγω της πολυπλοκότητας των προαναφερθεισών κατανομών, έμφαση θα πρέπει να δοθεί στην απόδοση/εύρεση εκφράσεων με λιγότερο περίπλοκες δομές.

Επιπλέον, αξίζει να σημειωθεί ότι στα Κεφάλαια 4, 6 και 7 παρουσιάσθηκαν τα διαγράμματα $O2$, $O4$ και $O2N2$, αντίστοιχα, τα οποία -ενώ τυπικά είναι ημιπαραμετρικά- στην πραγματικότητα συμπεριφέρονται ως πλήρως μη παραμετρικά διαγράμματα ελέγχου. Αυτό συμβαίνει, διότι οι τιμές των FAR και ARL_{in} δεν αλλάζουν δραματικά όταν χρησιμοποιούνται διαφορετικοί σύνδεσμοι. Θα μπορούσε, λοιπόν, κανείς να ισχυρισθεί ότι τα χαρακτηριστικά των διαγραμμάτων αυτών φαίνεται να μην επηρεάζονται από την επιλογή του συνδέσμου. Συνεπώς, μία κατεύθυνση που θα είχε ενδιαφέρον να μελετηθεί προσεχώς, θα ήταν η προσπάθεια προσέγγισης των εκφράσεων που προσδιορίζουν τα χαρακτηριστικά τους (π.χ. χαρακτηριστική συνάρτηση λειτουργίας, ποσοστό συναγερμού, ποσοστό λανθασμένου συναγερμού) με νέους τύπους, οι οποίοι θα έχουν απαλλαχθεί από τους συνδέσμους. Πιο συγκεκριμένα, στις μαθηματικές εκφράσεις δεν θα εμφανίζονται οι ποσότητες $C(\cdot, \cdot)$ και $D(\cdot, \cdot)$, οι οποίες αντιστοιχούν στους συνδέσμους που σχετίζονται με τις

από κοινού α.σ.κ. των δειγμάτων αναφοράς και ελέγχου, αντίστοιχα (δηλ. των αντίστοιχων α.σ.κ. $F(x, y)$ και $G(x, y)$).

Παράλληλα, νέα ημιπαραμετρικά σχήματα θα μπορούσαν να προταθούν για το σχεδιασμό και κατασκευή νέων δισδιάστατων ή ακόμη και καλύτερα και πολυδιάστατων διαγραμμμάτων ελέγχου, στα όρια ελέγχου των οποίων θα εμπλέκονται συμμεταβλητές. Οι κανόνες ελέγχου θα μπορούσαν να περιλαμβάνουν και πλήθος συμμεταβλητών, ενδεχομένως όχι απαραίτητα σε συνδυασμό με το πλήθος των αντίστοιχων διατεταγμένων παρατηρήσεών τους, ανάλογα με το εάν θα χρησιμοποιούνται για την παρακολούθηση του μέσου μίας διεργασίας ή για την ταυτόχρονη επιτήρηση της μέσης τιμής και της διασποράς. Σε κάθε περίπτωση, οι κανόνες που θα προκύψουν, θα είναι ευαίσθητοι στην ανίχνευση πιθανής μεταβολής στη δομή εξάρτησης των υπό μελέτη μεταβλητών, διότι θα διατηρούν τη συσχέτιση των αρχικών παρατηρήσεων.

Τέλος, εργαζόμενοι σε ένα διαφορετικό πλαίσιο, η παρούσα διατριβή θα μπορούσε να αποτελέσει ερέθισμα για τον προσδιορισμό νέων πολυδιάστατων βαθμολογικών σ.σ. (*ranked-based statistics*) ή σ.σ. ροών (*runs statistics*) για τη δημιουργία κανόνων ελέγχου με σκοπό την κατασκευή νέων διαγραμμμάτων ελέγχου. Ο προσδιορισμός των ορίων ελέγχου σε συνδυασμό πάντα με τη μορφή της εκάστοτε ελεγχουσυνάρτησης, θα έχει ως αποτέλεσμα τη δημιουργία σχημάτων για την παρακολούθηση της θέσης ή/και της μεταβλητότητας μίας διεργασίας και κατ' επέκταση την κατασκευή νέων πολυδιάστατων είτε ημιπαραμετρικών είτε μη παραμετρικών διαγραμμμάτων ελέγχου.

Ολοκληρώνοντας την παρούσα διδακτορική εργασία, ιδιαίτερη έμφαση αξίζει να δοθεί στο γεγονός ότι με το πέρασμα του χρόνου έχει σημειωθεί μεγάλη πρόοδος στον κλάδο της Μη Παραμετρικής Στατιστικής. Οι απαραμετρικές μέθοδοι αποτελούν πλέον αναπόσπαστο εργαλείο των ερευνητών που ασχολούνται με πραγματικά δεδομένα, όπου η στατιστική συμπερασματολογία δεν μπορεί να εξαχθεί βάσει συγκεκριμένων παραμετρικών παραδοχών σχετικά με τον υπό μελέτη πληθυσμό, όπως για παράδειγμα την κανονικότητα. Παρόλα αυτά, ενώ η πρόοδος συνεχίστηκε ερευνητικά στο συγκεκριμένο πεδίο, οι μεθοδολογίες αυτές δε φαίνεται να έχουν εναρμονισθεί πλήρως με τον Έλεγχο Ποιότητας και κυρίως με τον Πολυμεταβλητό Έλεγχο Διεργασιών. Για το λόγο αυτό, η έρευνα που σχετίζεται με την τρέχουσα διατριβή προσανατολίσθηκε προς αυτή την κατεύθυνση. Θα ήταν παράλειψη να μην αναφερθεί ότι ταυτόχρονα και όλα τα θέματα που συζητήθηκαν παραπάνω στην ενότητα αυτή, μπορούν μελλοντικά να αποτελέσουν κατευθύνσεις και προτάσεις για περαιτέρω έρευνα.

Ελληνική Βιβλιογραφία

- Camp, J. M. (2004). *Η Αρχαία Αγορά της Αθήνας: Οι Ανασκαφές στην Καρδιά της Κλασικής Πόλης*, Εκδόσεις Milet, Μορφωτικό Ίδρυμα Εθνικής Τραπέζης, Αθήνα (Μετάφραση Κλεώπα, Μ.).
- Αντζουλάκος, Δ. Λ. και Κούτρας, Μ. Β. (2016). *Ασκήσεις Πιθανοτήτων: Μέρος II*, Εκδόσεις Σταμούλη, Αθήνα.
- Αντζουλάκος, Δ. Λ. (2009). *Ανάλυση Επιβίωσης, Σημειώσεις Παραδόσεων του μαθήματος της Ανάλυσης Επιβίωσης στα πλαίσια του ΠΜΣ στην «Εφαρμοσμένη Στατιστική»*, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Αντζουλάκος, Δ. Λ. (2010). *Στατιστικός Έλεγχος Ποιότητας, Σημειώσεις Παραδόσεων του μαθήματος του Στατιστικού Ελέγχου Ποιότητας στα πλαίσια του ΠΜΣ στην «Εφαρμοσμένη Στατιστική»*, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Βαρουφάκης, Γ. (2005). Ο Έλεγχος Ποιότητας των Προϊόντων και η Προστασία του Καταναλωτή στην Αρχαιότητα, *Αρχαιολογία και Τέχνες*, **95**, 81-87.
- Κούτρας, Μ. Β. (2001). *Εισαγωγή στη Συνδυαστική*, Εκδόσεις Σταμούλη, Αθήνα.
- Κούτρας, Μ. Β. (2004). *Εισαγωγή στις Πιθανότητες, Θεωρία και Εφαρμογές: Μέρος I*, Εκδόσεις Σταμούλη, Αθήνα.
- Κούτρας, Μ. Β. (2005). *Εισαγωγή στις Πιθανότητες, Θεωρία και Εφαρμογές: Μέρος II*, Εκδόσεις Σταμούλη, Αθήνα.
- Κούτρας, Μ. Β. (2008). *Στατιστική Θεωρία Αξιοπιστίας και Έλεγχοι Χρόνων Ζωής*, Σημειώσεις Παραδόσεων του μαθήματος της Στατιστικής Θεωρίας Αξιοπιστίας και Ελέγχων Χρόνων Ζωής στα πλαίσια του ΠΜΣ στην «Εφαρμοσμένη Στατιστική», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Κούτρας, Μ. Β. (2016). *Εισαγωγή στη Θεωρία Πιθανοτήτων και Εφαρμογές*, Εκδόσεις Σταμούλη, Αθήνα.

- Μηλιένος, Φ. Σ. (2009). *Ακριβείς και Προσεγγιστικές Μέθοδοι για τη Μελέτη Συστημάτων Αξιοπιστίας και Προβλημάτων Ελέγχου Ποιότητας*, Διδακτορική Διατριβή, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Μπερσίμης, Σ. (2005). *Θεωρία Ροών Επιτυχιών και Εφαρμογές*, Διδακτορική Διατριβή, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Μπούτσικας, Μ. Β. (2004). *Μέθοδοι Προσομοίωσης και Υπολογιστικές Στατιστικές Τεχνικές*, Σημειώσεις Παραδόσεων του μαθήματος των Μεθόδων Προσομοίωσης και Υπολογιστικών Στατιστικών Τεχνικών στα πλαίσια του ΠΜΣ στην «Εφαρμοσμένη Στατιστική», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Παπαργύρης, Α. και Παπαργύρης, Δ. (2010). *Ποιοτικός Έλεγχος Παραγωγής*, Εκδόσεις Ζήτη, Θεσσαλονίκη.
- Ρακιτζής, Α. Χ. (2008). *Στατιστικός Έλεγχος Ποιότητας και Θεωρία Ροών και Σχηματισμών*, Διδακτορική Διατριβή, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Τριανταφύλλου, Ι. Σ. (2009). *Συμβολή στη Μελέτη Ιδιοτήτων Γήρανσης Συστημάτων Αξιοπιστίας και Μη Παραμετρικών Διαγραμμμάτων Ελέγχου με Χρήση της Θεωρίας Διατεταγμένων Παρατηρήσεων*, Διδακτορική Διατριβή, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Πειραιάς.
- Χαραλαμπίδης, Χ. Α. (2000). *Συνδυαστική: Τεύχος 1*, Εκδόσεις Συμμετρία, Αθήνα.

Ξένη Βιβλιογραφία

- Abbasi, S., Riaz, M., Miller, A. and Ahmad, S. (2015). On the Performance of Phase I Dispersion Control Charts for Process Monitoring. *Quality and Reliability Engineering International*, **31**, 1705–1716.
- Abdelhamid, S. N. (1985). On a Characterization of Rectangular Distributions. *Statistics and Probability Letters*, **3**, 235–238.
- Abo-Eleneen, Z. A. and Nagaraja, H. N. (2002). Fisher Information in an Order Statistic and its Concomitant. *Annals of the Institute of Statistical Mathematics*, **54**, 667–680.
- Abu-Shawiesh, M. O. and Abdullah, M. B. (2001). A New Robust Bivariate Control Chart for Location. *Communications in Statistics - Simulation and Computation*, **30**, 513–529.
- Achcar, J. A. and Moala, F. A. (2015). Use of Copula Functions for the Reliability of Series Systems. *International Journal of Quality & Reliability Management*, **32**, 617–634.
- Adams, B. M. (1994). The Multivariate Control Web. *Quality Engineering*, **6**, 533–545.
- Aitken, A. C. and Gonin, H. T. (1936). XI.-On Fourfold Sampling With and Without Replacement. *Proceedings of the Royal Society of Edinburgh*, **55**, 114–125.
- Ali, M. M., Mikhail, N. N. and Haq, M. S. (1978). A Class of Bivariate Distributions Including the Bivariate Logistic. *Journal of Multivariate Analysis*, **8**, 405–412.
- Alt, F. B. (1985). Multivariate Quality Control, In *Encyclopedia of Statistical Sciences* (Eds., S. Kotz, N. L. Johnson and C. R. Read), pp. 110–122. New York: John Wiley & Sons.
- Alt, F. B. and Smith, N. D. (1998). Multivariate Process Control, In *Handbook of Statistics Vol. 7* (Eds., P. R. Krishnaiah and C. R. Rao), pp. 333–351, .
- Alvarado, E., Sandberg, D. V. and Pickford, S. G. (1998). Modeling Large Forest Fires as Extreme Events. *Northwest Science*, **72**, 66–75.
- Anderson, T. W. and Darling, D. A. (1952). Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, **23**, 193–212.

- Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W. and Tukey, J. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press.
- Anjos, U. U., Kolev, N. and Tanaka, N. I. (2005). Copula Associated to Order Statistics. *Brazilian Journal of Probability and Statistics*, **19**, 111–123.
- Antzoulakos, D. L. and Rakitzis, A. C. (2008a). The Modified r out of m Control Chart. *Communications in Statistics - Simulation and Computation*, **37**, 396–408.
- Antzoulakos, D. L. and Rakitzis, A. C. (2008b). The Revised m -out-of- k Runs Rule. *Quality Engineering*, **20**, 75–81.
- Aparisi, F., Jabaloyes, J. and Carrión, A. (1999). Statistical Properties of the $IS|$ Multivariate Control Chart. *Communications in Statistics - Theory and Methods*, **28**, 2671–2686.
- Aparisi, F., Jabaloyes, J. and Carrión, A. (2001). Generalized Variance Chart Design with Adaptive Sample Sizes. The Bivariate Case. *Communications in Statistics - Theory and Methods*, **30**, 931–948.
- Arnold, B. and Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics*. Lecture Notes in Statistics No 53. New York: Springer-Verlag.
- Arnold, B., Balakrishnan, N. and Nagaraja, H. (1992). *A First Course in Order Statistics*. New York: John Wiley & Sons.
- Avérous, J., Genest, C. and Kochar, S. C. (2005). On the Dependence Structure of Order Statistics. *Journal of Multivariate Analysis*, **94**, 159–171.
- Balakrishnan, N. (1993). Multivariate Normal Distribution and Multivariate Order Statistics Induced by Ordering Linear Combinations. *Statistics and Probability Letters*, **17**, 343–350.
- Balakrishnan, N., Castillo, E. and Sarabia, J. M. (2006). *Advances in Distribution Theory, Order Statistics and Inference*. Statistics for Industry and Technology. Boston: Birkhäuser.
- Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference. Estimation Methods. Statistical Modeling and Decision Science*. Academic Press.
- Balakrishnan, N. and Lai, C.-D. (1983). *Probabilistic Metric Spaces*. New York: North-Holland.
- Balakrishnan, N. and Lai, C.-D. (2009). *Continuous Bivariate Distributions*. (2nd ed.). London: Springer.
- Balakrishnan, N. and Ng, H. K. T. (2006). *Precedence-Type Tests and Applications*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons.

- Balakrishnan, N. and Rao, C. R. (Eds.) (1998a). *Order Statistics: Applications*. Handbook of Statistics Vol. 17. Amsterdam: Elsevier Science B. V.
- Balakrishnan, N. and Rao, C. R. (Eds.) (1998b). *Order Statistics: Theory & Methods*. Handbook of Statistics Vol. 16. Amsterdam: Elsevier Science B. V.
- Balakrishnan, N., Triantafyllou, I. S. and Koutras, M. V. (2009). Nonparametric Control Charts Based on Runs and Wilcoxon-Type Rank-Sum Statistics. *Journal of Statistical Planning and Inference*, **139**, 3177–3192.
- Balakrishnan, N., Triantafyllou, I. S. and Koutras, M. V. (2010). A Distribution-Free Control Chart Based on Order Statistics. *Communications in Statistics - Theory and Methods*, **39**, 3652–3677.
- Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. New York: Holt, Rinehart & Winston.
- Barnett, V. (1976). The Ordering of Multivariate Data. *Journal of the Royal Statistical Society: Series A (General)*, **139**, 318–355.
- Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. Wiley Series in Probability and Statistics (3rd ed.). New York: John Wiley & Sons.
- Bayramoglu, K. and Bayramoglu-Bairamov, I. (2014). Baker-Lin-Huang Type Bivariate Distributions Based on Order Statistics. *Communications in Statistics - Theory and Methods*, **43**, 1992–2006.
- Bayramoglu-Bairamov, I. (2013). Reliability and Mean Residual Life of Complex Systems with Two Dependent Components Per Element. *IEEE Transactions on Reliability*, **62**, 276–285.
- Bayramoglu-Bairamov, I. and Kemalbay, G. (2013). Some Novel Discrete Distributions under Fourfold Sampling Schemes and Conditional Bivariate Order Statistics. *Journal of Computational and Applied Mathematics*, **248**, 1–14.
- Bersimis, S., Psarakis, S. and Panaretos, J. (2007). Multivariate Statistical Process Control Charts: An Overview. *Quality and Reliability Engineering International*, **23**, 517–543.
- Bhattacharya, P. K. (1974). Convergence of Sample Paths of Normalized Sums of Induced Order Statistics. *The Annals of Statistics*, **2**, 1034–1039.
- Bhattacharya, P. K. (1984). Induced Order Statistics. Theory and Application, In *Nonparametric Methods*. Handbook of Statistics Vol. 4 (Eds., P. R. Krishnaiahgerger and P. K. Sen), pp. 383–403. Amsterdam: Elsevier Science B. V.

- Bickel, P. J. and Lehmann, E. L. (1976). Descriptive Statistics for Nonparametric Models. III. Dispersion. *The Annals of Statistics*, **4**, 1139–1158.
- Bofinger, E. and Bofinger, V. J. (1965). The Correlation of Maxima in Samples Drawn from a Bivariate Normal Distribution. *Australian Journal of Statistics*, **7**, 57–61.
- Bofinger, V. J. (1970). The Correlation of Maxima in Samples Bivariate Non-Normal Distributions. *Australian Journal of Statistics*, **12**, 1–7.
- Boone, J. M. and Chakraborti, S. (2012). Two Simple Shewhart-Type Multivariate Nonparametric Control Charts. *Applied Stochastic Models in Business and Industry*, **28**, 130–140.
- Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} Fractional Factorial Designs Part 1. *Technometrics*, **3**, 311–351.
- Buishand, T. A. (1989). Statistics of Extremes in Climatology. *Statistica Neerlandica*, **43**, 1–30.
- Busababodhin, P. and Amphanthong, P. (2016). Copula Modelling for Multivariate Statistical Process Control: A Review. *Communications for Statistical Applications and Methods*, **23**, 497–515.
- Carey, R. G. (2003). *Improving Healthcare with Control Charts: Basic and Advanced SPC Methods and Case Studies*. Milwaukee, Wisconsin: ASQ Quality Press.
- Castillo, E. (1988). *Extreme Value Theory in Engineering*. New York: Academic Press.
- Castillo, E., Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons.
- Chakraborti, S. (2011). Nonparametric (Distribution-Free) Quality Control Charts, In *Encyclopedia of Statistical Sciences* (Eds., S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic). New Jersey: John Wiley & Sons.
- Chakraborti, S., Eryilmaz, S. and Human, S. W. (2009). A Phase II Nonparametric Control Chart Based on Precedence Statistics with Runs-Type Signaling Rules. *Computational Statistics and Data Analysis*, **53**, 1054–1065.
- Chakraborti, S. and Graham, M. A. (2008). Control Charts, Nonparametric, In *Encyclopedia of Statistics in Quality and Reliability* (Eds., F. Ruggeri, R. S. Kenett and F. W. Faltin). New York: John Wiley & Sons.
- Chakraborti, S., Human, S. W. and Graham, M. A. (2008). Phase I Statistical Process Control Charts: An Overview and Some Results. *Quality Engineering*, **21**, 52–62.

- Chakraborti, S., Human, S. W. and Graham, M. A. (2011). Nonparametric (Distribution-Free) Quality Control Charts, In *Handbook of Methods and Applications of Statistics: Engineering, Quality Control and Physical Sciences* (Ed., N. Balakrishnan), pp. 298–329. New York: John Wiley & Sons.
- Chakraborti, S., van der Laan, P. and van de Wiel, M. A. (2000). *A Class of Distribution-Free Control Charts*. Technical Report SPOR-Report 2000-11 Technische Universiteit Eindhoven.
- Chakraborti, S., van der Laan, P. and van de Wiel, M. A. (2004). A Class of Distribution-Free Control Charts. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **53**, 443–462.
- Chakraborti, S., Van Der Laan, P. and Bakir, S. (2001). Nonparametric Control Charts: An Overview and Some Results. *Journal of Quality Technology*, **33**, 304–315.
- Champ, C. W. and Woodall, W. H. (1987). Exact Results for Shewhart Control Charts with Supplementary Runs Rules. *Technometrics*, **29**, 393–399.
- Chao, M. T. and Cheng, S. W. (1996). Semi-Circle Control Chart for Variables Data. *Quality Engineering*, **8**, 441–446.
- Chao, M. T. and Cheng, S. W. (2008). On 2-D Control Charts. *Quality Technology and Quantitative Management*, **5**, 243–262.
- Chen, G. and Cheng, S. W. (1998). Max Chart: Combining \bar{X} -bar Chart and S Chart. *Statistica Sinica*, **8**, 263–271.
- Chen, G., Cheng, S. W. and Xie, H. (2005). A New Multivariate Control Chart for Monitoring Both Location and Dispersion. *Communications in Statistics - Simulation and Computation*, **34**, 203–217.
- Cheng, C.-R. and Shiau, J.-J. (2015). A Distribution-Free Multivariate Control Chart for Phase I Applications. *Quality and Reliability Engineering International*, **31**, 97–111.
- Cheng, S. W. and Mao, H. (2008). A Multivariate Semi-Circle Control Chart for Variables Data. *Quality Technology and Quantitative Management*, **5**, 331–338.
- Cheng, S. W. and Thaga, K. (2005). Multivariate Max-CUSUM Chart. *Quality Technology and Quantitative Management*, **21**, 221–235.
- Cheng, S. W. and Thaga, K. (2006). Single Variables Control Charts: An Overview. *Quality and Reliability Engineering International*, **22**, 811–820.
- Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Chichester: John Wiley & Sons.

- Cherubini, U., Mulinacci, S., Gobbi, F. and Romagnoli, S. (2012). *Dynamic Copula Methods in Finance*. Chichester: John Wiley & Sons.
- Chowdhury, S., Mukherjee, A. and Chakraborti, S. (2014). A New Distribution-Free Control Chart for Joint Monitoring of Unknown Location and Scale Parameters of Continuous Distributions. *Quality and Reliability Engineering International*, **30**, 191–204.
- Chowdhury, S., Mukherjee, A. and Chakraborti, S. (2015). Distribution-Free Phase II CUSUM Control Chart for Joint Monitoring of Location and Scale. *Quality and Reliability Engineering International*, **31**, 135–151.
- Chu, J. T. (1957). Some Uses of Quasi-Ranges. *The Annals of Mathematical Statistics*, **28**, 173–180.
- Cohen, A. C. J. (1959). Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated. *Technometrics*, **1**, 217–237.
- Cohen, A. C. J. (1961). Tables for Maximum Likelihood Estimates: Slightly Truncated and Censored Samples. *Technometrics*, **3**, 535–541.
- Cohen, A. C. J. (1963). Progressively Censored Samples in Life Testing. *Technometrics*, **5**, 327–339.
- Cohen, A. C. J. (1991). *Truncated and Censored Tables: Theory and Applications*. Statistics: Textbooks and Monographs Vol. 119. New York: Marcel Dekker.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T. and Reis, J. (2009). Modeling Wine Preferences by Data Mining from Physicochemical Properties. *Decision Support Systems*, **47**, 547–553.
- Costa, A. F. B. and Machado, M. A. G. (2009). A New Chart Based on Sample Variances for Monitoring the Covariance Matrix of Multivariate Processes. *The International Journal of Advanced Manufacturing Technology*, **41**, 770–779.
- Costa, A. F. B. and Neto, A. F. (2017). The S Chart with Variable Charting Statistic to Control Bi and Trivariate Processes. *Computers & Industrial Engineering*, **113**, 27–34.
- Cucconi, O. (1968). Un Nuovo Test non Parametrico per il Confronto tra Due Gruppi Campionari. *Giornale degli Economisti*, **XXVII**, 225–248.
- D’Agostino, M. A. (1986a). Graphical Analysis, In *Goodness-of-Fit Techniques*. Statistics: Textbooks and Monographs Vol. 67 (Eds., R. B. D’Agostino and M. A. Stephens), pp. 7-62. New York: Marcel Dekker.
- D’Agostino, M. A. (1986b). Tests for the Normal Distribution, In *Goodness-of-Fit Techniques*. Statistics: Textbooks and Monographs Vol. 67 (Eds., R. B. D’Agostino and M. A. Stephens), pp. 367-420. New York: Marcel Dekker.

- Das, N. (2009). A New Multivariate Non-Parametric Control Chart Based on Sign Test. *Quality Technology and Quantitative Management*, **6**, 155–169.
- David, H. A. (1973). Concomitants of Order Statistics. *Bulletin of the International Statistical Institute*, **45**, 295–300.
- David, H. A. (1979). Robust Estimation in the Presence of Outliers, In *Robustness in Statistics* (Eds., R. L. Launer and G. N. Wilkinson), pp. 61-74. New York: Academic Press.
- David, H. A. (1993). Concomitants of Order Statistics: Review and Recent Developments, In *Multiple Comparisons, Selection and Application in Biometry*. Statistics: Textbooks and Monographs Vol. 134 (Ed., F. M. Hoppe), pp. 507–518. New York: Marcel Dekker.
- David, H. A. (1994). Concomitants of Extreme Order Statistics, In *Extreme Value Theory and Applications: Proceedings of the Conference on Extreme Value Theory and Applications Vol. 1* (Eds. Galambos, J., Lechner, J. and Simiu, E.), pp. 211–224.
- David, H. A. and Galambos, J. (1974). The Asymptotic Theory of Concomitants of Order Statistics. *Journal of Applied Probability*, **11**, 762–770.
- David, H. A. and Nagaraja, H. (2003). *Order Statistics*. (3rd ed.). Hoboken, New Jersey: John Wiley & Sons.
- David, H. A. and Nagaraja, H. N. (1998). Concomitants of Order Statistics, In *Order Statistics: Theory & Methods*. Handbook of Statistics Vol. 16 (Eds., N. Balakrishnan and C. R. Rao), pp. 487–513. Amsterdam: Elsevier Science B. V.
- David, H. A., O'Connell, M. J. and Yang, S. S. (1977). *The Annals of Statistics*, **7**, 216–223.
- David, H. A. and Shu, V. S. (1978). Robustness of Location Estimators in the Presence of an Outlier, In *Contributions to Survey Sampling and Applied Statistics: Papers Honor of H. O. Hartley* (Ed., H. A. David), pp. 235–250. New York: Academic Press.
- Dodge, H. F. and Romig, H. G. (1929). A method of Sampling Inspection. *Bell System Technical Journal*, **8**, 613–631.
- Dokouhaki, P. and Noorossana, R. (2013). A Copula Markov CUSUM Chart for Monitoring the Bivariate Autocorrelated Binary Observations. *Quality and Reliability Engineering International*, **29**, 911–919.
- Drouet-Mari, D. and Kotz, S. (2001). *Correlation and Dependence*. London: Imperial College Press.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Applications of Mathematics, Stochastic Modelling and Applied Probability No 33. New York: Springer-Verlag.

- Emura, T., Long, T.-H. and Sun, L.-H. (2016). *R* Routines for Performing Estimation and Statistical Process Control under Copula-Based Time Series Models. *Communications in Statistics - Simulation and Computation*, **46**, 3067–3087.
- Epstein, B. (1949). A Modified Extreme Value Problem. *The Annals of Mathematical Statistics*, **20**, 99–103.
- Epstein, B. (1960a). Estimation from Life Test Data. *Technometrics*, **2**, 447–454.
- Epstein, B. (1960b). Statistical Life Test Acceptance Procedures. *Technometrics*, **2**, 435–446.
- Eryilmaz, S. (2005). Concomitants in a Sequence of Independent Nonidentically Distributed Random Vectors. *Communications in Statistics - Theory and Methods*, **34**, 1925–1933.
- Eryilmaz, S. (2010). Conditional Lifetimes of Consecutive k -out-of- n Systems. *IEEE Transactions on Reliability*, **59**, 178–182.
- Eryilmaz, S. (2011). Estimation in Coherent Reliability Systems Through Copulas. *Reliability Engineering and System Safety*, **96**, 564–568.
- Eryilmaz, S. (2014). Multivariate Copula Based Dynamic Reliability Modeling with Application to Weighted- k -out-of- n Systems of Dependent Components. *Structural Safety*, **51**, 23–28.
- Eryilmaz, S. (2017). Reliability Analysis of Systems with Components Having Two Dependent Subcomponents. *Communications in Statistics - Simulation and Computation*, **46**, 8005–8013.
- Eryilmaz, S. and Bairamov, I. G. (2003). On a New Sample Rank of an Order Statistics and its Concomitant. *Statistics and Probability Letters*, **63**, 123–131.
- Farlie, D. J. G. (1960). The Performance of Some Correlation Coefficients for a General Bivariate Distribution. *Biometrika*, **47**, 307–323.
- Fatahi, A. A., Noorossana, R., Dokouhaki, P. and Moghaddam, B. F. (2012). Copula-Based Bivariate ZIP Control Chart for Monitoring Rare Events. *Communications in Statistics - Theory and Methods*, **41**, 2699–2716.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémondge, N. and Bobée, B. (2004). Multivariate Hydrological Frequency Analysis Using Copulas. *Water Resources Research*, **40**, 1–12.
- Feigenbaum, A. V. (1951). *Total Quality Control: Engineering and Management*. New York: McGraw-Hill.
- Fréchet, M. (1958). Remarque au Sujet de la Note Précédente. *Rendus de l'Académie des Sciences, Série I. Mathématique*, **246**, 2719–2720.

- Fredricks, G. A. and Nelsen, R. B. (2007). On the Relationship Between Spearman's Rho and Kendall's Tau for Pairs of Continuous Random Variables. *Journal of Statistical Planning and Inference*, **137**, 2143–2150.
- Fuchs, C. and Benjamini, Y. (1994). Multivariate Profile Charts for Statistical Process Control. *Technometrics*, **36**, 182–195.
- Galambos, J. (1975). Order Statistics of Samples from Multivariate Distributions. *Journal of the American Statistical Association*, **70**, 674–680.
- Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-Fit Tests for Copulas: A Review and a Power Study. *Insurance: Mathematics and Economics*, **44**, 199–213.
- Ghute, V. B. and Shirke, D. T. (2012). A Nonparametric Signed-Rank Control Chart for Bivariate Process Location. *Quality Technology and Quantitative Management*, **9**, 317–328.
- Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. (7th ed.). Amsterdam: Elsevier/Academic Press. (Translated from Russian, Translation edited by A. Jeffrey and D. Zwillinger).
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, **11**, 1–21.
- Gumbel, E. J. (1958). Distributions à Plusieurs Variables Dont les Marges Sont Données (with Remarks by M. Fréchet), In *Comptes Rendus de l'Académie des Science*, vol. 246. (pp. 2717–2720).
- Gumbel, E. J. (2004). *Statistics of Extremes*. Mineola, New York: Dover. (Reprint of the Columbia University Press, New York, 1958 edition).
- Gupta, A. K. (1952). Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample. *Biometrika*, **39**, 260–273.
- Gupta, D. and Gupta, R. C. (1984). On the Distribution of Order Statistics for a Random Sample Size. *Statistica Neerlandica*, **38**, 13–19.
- Gupta, R. C. (1984). Relationships Between Order Statistics and Record Values and Some Characterization results. *Journal of Applied Probability*, **21**, 425–430.
- Gustafsson, A., Edvardsson, B. and Gustafsson, A. (1999). *The Nordic School of Quality Management*. Studenlitteratur.
- Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics. New York: John Wiley & Sons.

- Hamurkaroglu, C., Mert, M. and Saykan, Y. (2004). Nonparametric Control Charts Based on Mahalanobis Depth. *Hacettepe Journal of Mathematics and Statistics*, **33**, 57–67.
- Hawkins, D. M. (1980). *Identification of Outliers*. Monographs on Applied Probability and Statistics. London: Chapman and Hall.
- Hayter, A. J. and Tsui, K.-L. (1994). Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Technology*, **26**, 197–208.
- Hellman, P. and Liu, Y. (2013). Development of Quality Management Systems: How Have Disruptive Technological Innovations in Quality Management Affected Organizations? *Quality Innovation Prosperity*, **17**, 104–119.
- Hettmansperger, T. P. (2006). Multivariate location tests, In *Encyclopedia of Statistical Sciences Vol. 8* (Eds., S. Kotz, C. B. Read, N. balakrishnan and B. vidakovic). New Jersey: John Wiley & Sons.
- Hettmansperger, T. P. and Randles, R. H. (2002). A Practical Affine Equivariant Multivariate Median. *Biometrika*, **89**, 851–860.
- Hettmansperger, T. P. and Sheather, S. J. (1986). Confidence Intervals Based on Interpolated Order Statistics. *Statistics and Probability Letters*, **4**, 75–79.
- Hodges, J. L. and Lehmann, E. L. (1963). Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics*, **34**, 598–611.
- Hogg, R. V. and Craig, A. T. (1978). *Introduction to Mathematical Statistics*. New York: MacMillan.
- Hogg, V., R. (1974). Adaptive Robust Procedures: A partial Review and Some Suggestions for Future Applications and Theory. *Journal of the American Statistical Association*, **69**, 909–923.
- Howell, J. M. (1952). Statistical Quality Control. *Mathematics Magazine*, **25**, 155–157.
- Huang, J. S., Dou, X., Kuriki, S. and Lin, G. D. (2013). Dependence Structure of Bivariate Order Statistics with Applications to Bayramoglu's Distributions. *Journal of Multivariate Analysis*, **114**, 201–208.
- Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*. Wiley Series in Probability and Statistics (2nd ed.). New York: John Wiley & Sons.
- Hutson, A. D. (2000). Estimating the Covariance of Bivariate Order Statistics with Applications. *Statistics and Probability Letters*, **48**, 195–203.

- Janacek, G. J. and Meikle, S. E. (1997). Control Charts Based on Medians. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **46**, 19–31.
- Jensen, W. A., Jones-Farmer, L. A., Champ, C. W. and Woodall, W. H. (2006). Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology*, **38**, 349–364.
- Jeong, J.-I. and Cho, G.-Y. (2012). Multivariate EWMA Control Charts for Monitoring the Variance-Covariance Matrix. *Journal of the Korean Data and Information Science Society*, **23**, 807–814.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability No 73. London: Chapman and Hall/CRC.
- Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics (3rd ed.). New York: John Wiley & Sons.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 1*. Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics (2nd ed.). New York: John Wiley & Sons.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2*. Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics (2nd ed.). New York: John Wiley & Sons.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics. New York: John Wiley & Sons.
- Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H. and Champ, C. W. (2014). Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology*, **46**, 265–280.
- Juran, J. M. and Gryna, F. M. (1957). *Quality Control Handbook*. New York: McGraw-Hill.
- Jurečková, J., Sen, P. K. and Picek, J. (2012). *Methodology in Robust and Nonparametric Statistics*. Boca Raton: CRC Press.
- Kapatou, A. (1996). *Multivariate Nonparametric Control Charts Using Small Samples*. Ph.D. thesis Virginia Polytechnical Institute and State University.
- Kapatou, A. and Reynolds, M. R. J. (1994). Multivariate Nonparametric Control Charts Using Small Samples, In *ASA Proceedings of the Section on Quality and Productivity*, Joint Statistical Meetings, American Statistical Association, Alexandria, Virginia, pp. 241–246.

- Kapatou, A. and Reynolds, M. R. J. (1998). Multivariate Nonparametric Control Charts For the Case of Unknown Sigma, In *ASA Proceedings of the Section on Quality and Productivity*, Joint Statistical Meetings, American Statistical Association, Alexandria, Virginia, pp. 77–82.
- Karagöz, D. and Koyuncu, N. (2017). New Ranked Set Sampling Schemes for Range Charts Limits Under Bivariate Skewed Distributions. *Soft Computing*, . doi:0.1007/s00500-017-2880-4.
- Katz, R. W. (2010). Statistics of Extremes in Climate Change. *Climatic Change*, **100**, 71–76.
- Katz, R. W., Parlange, M. B. and Naveau, P. (2002). Statistics of Extremes in Climate Hydrology. *Advances in Water Resources*, **25**, 8–12.
- Ke Wang, M. S. (2008). *On Concomitants of Order Statistics*. Ph.D. thesis Ohio State University.
- Ke Wang, M. S. and Nagaraja, H. N. (2010a). Distribution of Extremal Order Statistics from Large Subsets of Concomitants. *Statistics and Probability Letters*, **80**, 534–539.
- Ke Wang, M. S. and Nagaraja, H. N. (2010b). Distribution of Order Statistics from Selected Subsets of Concomitants. *Journal of Statistical Planning and Inference*, **140**, 3076–3087.
- Kemalbay, G. and Bayramoglu-Bairamov, I. (2015). Joint Distribution of New Sample Rank of Bivariate Order Statistics. *Journal of Applied Statistics*, **42**, 2280–2289.
- Khoo, M. B. C. (2004). A New Bivariate Control Chart to Monitor the Multivariate Process Mean and Variance Simultaneously. *Quality Engineering*, **17**, 109–118.
- Khoo, M. B. C. and Quah, S. H. (2004). Alternatives to the Multivariate Control Chart for Process Dispersion. *Quality Engineering*, **16**, 423–435.
- Kolmogorov, A. N. (1933). Sulla Determinazione Empirica di una Legge di Distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 83–91.
- Koutras, M. V., Bersimis, S. and Antzoulakos, D. L. (2006). Improving the Performance of the Chi-square Control Charts via Runs Rules. *Methodology and Computing in Applied Probability*, **8**, 409–426.
- Koutras, M. V., Bersimis, S. and Maravelakis, P. (2007). Statistical Process Control using Shewhart Control Charts with Supplementary Runs Rules. *Methodology and Computing in Applied Probability*, **9**, 207–224.
- Koutras, M. V. and Sofikitou, E. M. (2015). A Bivariate Semiparametric Control Chart Based on Order Statistics, In *Statistical Stochastic and Data Analysis Methods & Applications* (Eds., A. Karagrigoriou, T. Oliveira and C. Skiadas), pp. 543–553. ISAST: International Society for the Advancement of Science and Technology.

- Koutras, M. V. and Sofikitou, E. M. (2017a). A Bivariate Semiparametric Control Chart Based on Order Statistics. *Quality and Reliability Engineering International*, **33**, 183–202.
- Koutras, M. V. and Sofikitou, E. M. (2017b). A New Bivariate Semiparametric Control Chart Based on Order Statistics and Concomitants. *Statistics and Probability Letters*, **129**, 340–347.
- Koutras, M. V. and Sofikitou, E. M. (2017c). Bivariate Semiparametric Control Charts for Simultaneous Monitoring of Process Mean and Variance. *submitted for publication*.
- Krishna, V. (2010). *Auction Theory*. (2nd ed.). Amsterdam: Elsevier/Academic Press.
- Lepage, Y. (1971). A Combination of Wilcoxon's and Ansari-Bradley's Statistics. *Biometrika*, **58**, 213–7.
- Levinson, W. A., Holmes, D. S. and Mergen, A. E. (2002). Variation Charts for Multivariate Processes. *Quality Engineering*, **14**, 539–545.
- Li, J., Zhang, X. and Jeske, D. R. (2013). Nonparametric Multivariate CUSUM Control Charts for Location and Scale Changes. *Journal of Nonparametric Statistics*, **25**, 1–20.
- Liu, R. Y. (1995). Control Charts for Multivariate Processes. *Journal of the American Statistical Association*, **90**, 1380–1387.
- Liu, R. Y., Singh, K. and Teng, J. H. (2004). DDMA-Charts: Nonparametric Multivariate Moving Average Control Charts Based on Data Depth. *Allgemeines Statistisches ARCHIV*, **88**, 235–258.
- Lloyd, E. H. (1952). Least-Squares Estimation of Location and Scale Parameters Using Order Statistics. *Biometrika*, **39**, 88–95.
- Long, T.-H. and Emura, T. (2014). A Control Chart Using Copula-Based Markov Chain Models. *Journal of the Chinese Statistical Association*, **52**, 466–496.
- Lowry, C. A. and Montgomery, D. C. (1995). A Review of Multivariate Control Charts. *IIE Transactions*, **27**, 800–810.
- Maboudou-Tchao, E. and Hawkins, D. M. (2011). Self-Starting Multivariate Control Charts for Location and Scale. *Journal of Quality Technology*, **43**, 113–126.
- MacGregor, J. F. and Kourti, T. (1995). Statistical Process Control of Multivariate Processes. *Control Engineering Practice*, **3**, 403–414.
- MacKinnon, W. J. (1964). Table for Both the Sign Test and Distribution-Free Confidence Intervals of the Median for Sample Sizes to 1,000. *Journal of the American Statistical Association*, **59**, 935–956.

- Mangasarian, O. L., Street, W. N. and Wolberg, W. H. (1995). Breast Cancer Diagnosis and Prognosis via Linear Programming. *Operations Research*, **43**, 570–577.
- Maravelakis, P. E. and Bersimis, S. (2009). The Use of Andrews Curves for Detecting the Out-of-Control Variables When a Multivariate Control Chart Signals. *Statistical Papers*, **31**, 51–65.
- Maravelakis, P. E., Bersimis, S., Panaretos, J. and Psarakis, S. (2002). Identifying the Out-of-Control Variable in a Multivariate Control Chart. *Communications in Statistics-Theory and Methods*, **31**, 2391–2408.
- Maritz, J. S. (1991). Estimating the Covariance Matrix of Bivariate Medians. *Statistics and Probability Letters*, **12**, 305–309.
- Mason, R. L., Tracy, N. D. and Young, J. C. (1995). Decomposition of T^2 for Multivariate Control Chart Interpretation. *Journal of Quality Technology*, **27**, 99–108.
- McCracken, A. K. and Chakraborti, S. (2013). Control Charts for Joint Monitoring of Mean and Variance: An Overview. *Quality Technology and Quantitative Management*, **10**, 17–36.
- McIntyre, G. A. (1952). A Method for Unbiased Selective Sampling, Using Ranked Sets. *Australian Journal of Agricultural Research*, **3**, 385–390.
- Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. (6th ed.). New York: John Wiley & Sons.
- Morgenstern, D. (1956). Einfache Beispiele Zweidimensionaler Verteilungen, *Mitteilungsblatt für Mathematische Statistik*. *Mitteilungsblatt für Mathematische Statistik*, **8**, 234–235.
- Mukherjee, A. and Chakraborti, S. (2012). A Distribution-Free Control Chart for the Joint Monitoring of Location and Scale. *Quality and Reliability Engineering International*, **28**, 335–352.
- Mukherjee, A. and Marozzi, M. (2017). A Distribution-Free Phase-II Procedure for Monitoring Service Quality. *Total Quality Management and Business Excellence*, **28**, 1227–1263.
- Murphy, B. J. (1987). Selecting Out of Control Variables with the T^2 Multivariate Quality Control Procedure. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **36**, 571–581.
- Murphy, R. B. (1951). *On Tests for Outlying Observations*. Ph.D. thesis Princeton University.
- Nagaraja, H. N. (1988). Record Values and Related Statistics - A Review. *Communications in Statistics - Theory and Methods*, **17**, 2223–2238.

- Nagaraja, H. N. and David, H. A. (1994). Distribution of the Maximum of Concomitants of Selected Order Statistics. *The Annals of Statistics*, **22**, 478–494.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. (2nd ed.). New York: Springer.
- Nelson, L. S. (1984). The Shewhart Control Chart - Tests for Special Causes. *Journal of Quality Technology*, **16**, 337–339.
- Nelson, L. S. (1985). Interpreting Shewhart \bar{X} Control Charts. *Journal of Quality Technology*, **17**, 114–116.
- Novak, S. Y. (2011). *Extreme Value Methods with Applications to Finance*. Monographs on Applied Probability and Statistics No 122. London: Chapman and Hall.
- Osei-Aning, R., Abbasi, S. A. and Riaz, M. (2017). Bivariate Dispersion Control Charts for Monitoring Non-Normal Processes. *Quality and Reliability Engineering International*, **33**, 515–529.
- Ou, Y., Chen, N. and Khoo, M. B. C. (2015). An Efficient Multivariate Control Charting Mechanism Based on SPRT. *International Journal of Production Research*, **53**, 1937–1949.
- Ou, Y., Wen, D., Wu, Z. and Khoo, M. B. C. (2012). Comparison Study on Effectiveness and Robustness of Control Charts for Monitoring Process Mean and Variance. *Quality and Reliability Engineering International*, **28**, 2–17.
- Papathanasiou, V. (1990). Some Characterization of Distributions Based on Order Statistics. *Statistics and Probability Letters*, **9**, 145–147.
- Pearson, E. S. and Sekar, C. C. (1936). The Efficiency of Statistical Tools and A Criterion for the Rejection of Outlying Observations. *Biometrika*, **28**, 308–320.
- Pearson, K. (1934). *Tables of the Incomplete B-Function*. Cambridge University Press.
- Pearson, K. and Hartley, H. O. (1970). *Biometrika Tables for Statisticians*. Vol. 1. (3rd ed.). Cambridge University Press.
- Pinsker, I. S., Kipnis, V. and Grechanovsky, E. (1986). A Recursive Formula for the Probability of Occurrence of at Least m out of N Events. *The American Statistician*, **40**, 275–276.
- Puri, M. L. and Sen, P. K. (1976). *Nonparametric Methods in Multivariate Analysis*. New York: John Wiley & Sons.
- Qiu, P. and Hawkins, D. (2001). A Rank Based Multivariate CUSUM Procedure. *Technometrics*, **43**, 120–132.

- Qiu, P. and Hawkins, D. (2003). A Nonparametric Multivariate Cumulative Sum Procedure for Detecting Shifts in all Directions. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **52**, 151–164.
- Quesenberry, C. P. (1993). The Effect of Sample Size on Estimated Control Limits for \bar{X} and X Control Charts. *Journal of Quality Technology*, **25**, 237–247.
- Rakitzis, A. C. and Antzoulakos, D. L. (2011). Chi-square Control Charts with Runs Rules. *Methodology and Computing in Applied Probability*, **13**, 657–669.
- Reiss, R.-D. (1989). *Approximate Distributions Order Statistics: With Applications to Nonparametric Statistics*. New York: Springer-Verlag.
- Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Statistics for Industry and Technology (3rd ed.). Basel, Switzerland: Birkhäuser Verlag AG.
- Reynolds, M. R. J. and Cho, G.-Y. (2006). Multivariate Control Charts for Monitoring the Mean Vector and Covariance Matrix. *Journal of Quality Technology*, **38**, 230–253.
- Reynolds, M. R. J. and Stoumbos, Z. G. (2006). Comparisons of Some Exponentially Weighted Moving Average Control Charts for Monitoring the Process Mean and Variance. *Technometrics*, **48**, 550–567.
- Reynolds, M. R. J. and Stoumbos, Z. G. (2008). Combinations of Multivariate Shewhart and MEWMA Control Charts for Monitoring the Mean Vector and Covariance Matrix. *Journal of Quality Technology*, **40**, 381–393.
- Reynolds, M. R. J. and Stoumbos, Z. G. (2010). Robust CUSUM Charts for Monitoring the Process Mean and Variance. *Quality and Reliability Engineering International*, **26**, 453–473.
- Riaz, M. and Does, R. J. (2008). An Alternative to the Bivariate Control Chart for Process Dispersion. *Quality Engineering*, **21**, 63–71.
- Rigdon, S. E. (1995). The Effect of Sample Size on Estimated Control Limits for \bar{X} and x Control Charts, In *Proceedings of the 26th Annual Meeting of the Midwest Decision Sciences Institute*, pp. 240–243.
- Rohatgi, V. K. (1987). Distribution of Order Statistics with Random Sample Size. *Communications in Statistics - Theory and Methods*, **16**, 3739–3743.
- Rohatgi, V. K. and Székely, G. J. (1992). C385. On the Background of Some Correlation Inequalities. *Journal of Statistical Computation and Simulation*, **40**, 260–262.

- Ross, G. J., Tasoulis, D. K. and Adams, N. M. (2011). Nonparametric Monitoring of Data Streams for Changes in Location and Scale. *Technometrics*, **53**, 379–389.
- Ross, S. M. (2013). *Simulation*. (5th ed.). Waltham, MA: Academic Press.
- Saghir, A. and Lin, Z. (2015). Control Charts for Dispersed Count Data: An Overview. *Quality and Reliability Engineering International*, **31**, 725–739.
- Sagir, A. (2015). The Bivariate Dispersion Control Charts for Non-Normal Processes. *International Journal of Production Research*, **55**, 1964–1979.
- Sagir, A., Chakraborti, S. and Ahmad, I. (2017). On the Performance of Phase-I Bivariate Dispersion Charts to Non-Normality. *Quality and Reliability Engineering International*, **33**, 637–656.
- Salvadori, G. and De Michele, C. (2007). On the Use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrologic Engineering*, **12**, 369–380.
- Sarhan, A. E. and Greenberg, B. G. (1956). Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples. *The Annals of Mathematical Statistics*, **27**, 427–451.
- Sarhan, A. E. and Greenberg, B. G. (1959). Estimation of Location and Scale Parameters for the Rectangular Population from Censored Samples. *Journal of the Royal Statistical Society: Series B (Methodological)*, **21**, 356–363.
- Scheffé, H. and Tukey, J. W. (1945). Non-Parametric Estimation. I. Validation of Order Statistics. *The Annals of Mathematical Statistics*, **16**, 187–192.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons.
- Serfling, R. J. (2006). Multivariate Symmetry and Asymmetry, In *Encyclopedia of Statistical Sciences: Regressograms to St. Petersburg Paradox Vol. 8* (Eds., S. Kotz, C.B. Read, N. Balakrishnan and B. Vidakovic), pp. 5338–5345. New York: John Wiley & Sons.
- Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance Test for Normality. *Biometrika*, **52**, 591–611.
- Shewhart, W. A. (1926). Quality Control Charts. *Bell System Technical Journal*, **5**, 593–603.
- Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. New York: D. Van Nostrand Company.
- Shu, V. S. (1978). *Robust Estimation of a Location Parameter in the Presence of Outliers*. Ph.D. thesis Iowa State University.

- Sinha, S. K. (1986). *Reliability and Life Testing*. New York: John Wiley & Sons.
- Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.
- Smirnov, N. (1939). Sur les Écarts de la Courbe de Distribution Empirique. *Matematicheskii Sbornik*, **6**, 3–26.
- Song, R., Buchberger, S. G. and Deddens, J. A. (1992). Moments of Variables Summing to Normal Order Statistics. *Statistics and Probability Letters*, **15**, 203–208.
- Song, R. and Deddens, J. A. (1993). A Note on moments of Variables Summing to Normal Order Statistics. *Statistics and Probability Letters*, **17**, 337–341.
- Spiring, F. A. and Cheng, S. (1998). An Alternate Variables Control Chart: The Univariate and Multivariate Case. *Statistica Sinica*, **8**, 273–287.
- Stephens, M. A. (1986a). Tests for the Exponential Distribution, In *Goodness-of-Fit Techniques*. Statistics: Textbooks and Monographs Vol. 67 (Eds., R. B. D'Agostino and M. A. Stephens), pp. 421-460. New York: Marcel Dekker.
- Stephens, M. A. (1986b). Tests for the Uniform Distribution, In *Goodness-of-Fit Techniques*. Statistics: Textbooks and Monographs Vol. 67 (Eds., R. B. D'Agostino and M. A. Stephens), pp. 331-366. New York: Marcel Dekker.
- Stoumbos, Z. G., Jones, A., Woodall, W. H. and Reynolds, M. R. (2001). On nonparametric multivariate control charts based on data depth, In *Frontiers in Statistical Quality Control* Vol. 6 (Eds., H.-J. Lenz and P.-T. Wilrich), pp. 207–227. New York: Physica-Verlag (Springer-Verlag).
- Street, W. N., Wolberg, W. H. and Mangasarian, O. L. (1993). Nuclear Feature Extraction for Breast Tumor Diagnosis, In *Proceedings of the Biomedical Image Processing and Biomedical Visualization SPIE 1905* (Eds., R. S. Acharya and D. B. Goldgof), pp. 861–870. San Jose, CA: John Wiley & Sons.
- Sukchotrat, T., Kim, S. B., Tsui, K.-L. and Chen, V. C. P. (2008). *Classification-Based Control (C^2) Charts for Monitoring Multivariate Processes*. Technical Report No 08-01 The University of Texas at Arlington.
- Sullivan, J. H. and Woodall, W. H. (2000). Change-Point Detection of Mean Vector or Covariance Matrix Shifts Using Multivariate Individual Observations. *IIE Transactions*, **32**, 537–549.
- Székely, G. J. and Móri, T. F. (1985). An Extremal Property of Rectangular Distributions. *Statistics and Probability Letters*, **3**, 107–109.

- Terrell, G. R. (1983). A Characterization of Rectangular Distributions. *The Annals of Probability*, **11**, 823–826.
- Thaga, K. and Gabaitiri, L. (2006). Multivariate Max-Chart. *Economic Quality Control*, **21**, 113–125.
- Thompson, W. R. (1936). On Confidence Ranges for the Median and Other Expectation Distributions for Populations of Unknown Distribution Form. *The Annals of Mathematical Statistics*, **7**, 122–128.
- Topalidou, E. and Psarakis, S. (2009). Review of Multinomial and Multiattribute Quality Control Charts. *Quality and Reliability Engineering International*, **25**, 773–804.
- Triantafyllou, I. S. (2017). Nonparametric Control Charts Based on Order Statistics: Some Advances. *Communications in Statistics - Simulation and Computation*, . doi:10.1080/03610918.2017.1359283.
- Verdier, G. (2013). Application of Copulas to Multivariate Control Charts. *Journal of Statistical Planning and Inference*, **143**, 2151–2159.
- Wang, K., Yeh, A. B. and Li, B. (2014). Simultaneous Monitoring of Process Mean Vector and Covariance Matrix via Penalized Likelihood Estimation. *Computational Statistics and Data Analysis*, **78**, 206–217.
- Waterhouse, M., Smith, I., Assareh, H. and Mergersen, K. (2010). Implementation of Multivariate Control Charts in a Clinical Setting. *International Journal for Quality in Health Care*, **22**, 408–414.
- Western Electric (1956). *Statistical Quality Control Handbook*. Indianapolis, IN: Western Electric Corporation.
- Wilk, M. B. and Gnanadesikan, R. (1968). Probability Plotting Methods for the Analysis of Data. *Biometrika*, **55**, 1–17.
- Woodall, W. H. (2000). Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology*, **32**, 341–350.
- Woodall, W. H. (2006). Use of Control Charts in Health-Care and Public-Health Surveillance (with Discussion and Rejoinder). *Journal of Quality Technology*, **38**, 89–134.
- Woodall, W. H., Adams, B. M. and Benneyan, J. C. (2012). The Use of Control Charts in Healthcare, In *Statistical Methods in Healthcare* (Eds., F. W. Faltin, R. S. Kenett and F. Ruggeri), pp. 253–267. Chichester, UK: John Wiley & Sons.

- Woodall, W. H. and Montgomery, D. C. (1999). Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology*, **31**, 376–386.
- Woodall, W. H. and Montgomery, D. C. (2014). Some Current Directions in the Theory and Application of Statistical Process Monitoring. *Journal of Quality Technology*, **46**, 78–94.
- Xie, H. (1991). *Contributions to Qualimetry*. Ph.D. thesis University of Manitoba, Winnipeg, Canada.
- Yang, S. F. and Arnold, B. C. (2010). A Simple Approach for Monitoring Business Service Time Variation. *Scientific World Journal*, . doi:<http://dx.doi.org/10.1155/2014/238719>. Article ID 238719, 16 pages.
- Yang, S. F. and Arnold, B. C. (2015). A Simple Approach for Monitoring Process Mean and Variance Simultaneously, In *Frontiers in Statistical Quality Control Vol. 11* (Eds., S. Knoth and W. Schmid), pp. 135–149. Switzerland: Springer International Publishing.
- Yang, S. S. (1977). General Distribution Theory of the Concomitants of Order Statistics. *The Annals of Statistics*, **5**, 996–1002.
- Yeh, A. B., Lin, D. and McGrath, R. N. (2006). Multivariate Control Charts for Monitoring Covariance Matrix: A Review. *Quality Technology and Quantitative Management*, **3**, 415–416.
- Yeo, W. B. and David, H. A. (1984). Selection Through an Associated Characteristic, with Applications to the Random Effects Model. *Journal of the American Statistical Association*, **79**, 399–405.
- Zamba, K. D. and Hawkins, D. M. (2009). A Multivariate Change-Point Model for Change in Mean Vector and/or Covariance Structure. *Journal of Quality Technology*, **41**, 285–303.
- Zhang, J., Li, Z. and Wang, Z. (2010). A Multivariate Control Chart for Simultaneously Monitoring Process Mean and Variability. *Computational Statistics and Data Analysis*, **54**, 2244–2252.
- Zhang, X. and Wilson, A. (2017). System Reliability and Component Importance Under Dependence: A Copula Approach. *Technometrics*, **59**, 215–224.
- Zhao, Y. and Mei, W. (2014). A New Bivariate Control Chart for Monitoring the Mean Vector and Covariance Matrix Simultaneously. *Computer Modelling and New Technologies*, **18**, 301–307.
- Zhou, M., Zi, X., Geng, W. and Li, Z. (2015). A Distribution-Free Multivariate Change-Point Model for Statistical Process Control. *Communications in Statistics - Simulation and Computation*, **44**, 1975–1987.

Zou, C., Jiang, W. and Tsung, F. (2011). A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control. *Technometrics*, **53**, 297–309.

Zou, C. and Tsung, F. (2010). Likelihood Ratio-Based Distribution-Free EWMA Control Charts. *Journal of Quality Technology*, **42**, 174–196.

Zou, C. and Tsung, F. (2011). A Multivariate Sign EWMA Control Chart. *Technometrics*, **53**, 84–97.

Zou, C. and Tsung, F. (2012). A Spatial Rank-Based Multivariate EWMA Control Chart. *Naval Research Logistics*, **59**, 91–110.

