ΚΕΦΑΛΑΙΟ 5

ΜΗ ΠΑΡΑΜΕΤΡΙΚΑ ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΜΕ ΧΡΗΣΗ ΡΟΩΝ ΚΑΙ ΒΑΘΜΟΛΟΓΙΚΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΤΥΠΟΥ *WILCOXON*

5.1. Εισαγωγή

Έστω ότι $X_1, X_2, ..., X_m$ είναι ένα τυχαίο δείγμα μεγέθους m που προέρχεται από μια συνεχή κατανομή με αθροιστική συνάρτηση $F_X(x)$ και $Y_1, Y_2, ..., Y_n$ ένα δείγμα μεγέθους n από μια άλλη κατανομή $F_Y(x)$. Ένα πρόβλημα που παρουσιάζει ιδιαίτερο ενδιαφέρον είναι ο έλεγχος της μηδενικής υπόθεσης $H_0 : F_X(x) = F_Y(x)$, δηλαδή η εξέταση για το αν οι δύο διεργασίες έχουν την ίδια κατανομή. Στα πλαίσια της Θεωρίας Αξιοπιστίας και του Ελέγχου Ποιότητας, ο έλεγχος της παραπάνω υπόθεσης πραγματοποιείται συχνά έναντι της εναλλακτικής $H_1 : \overline{F}_X(x) >_{st} \overline{F}_Y(x)$, δηλαδή έναντι της υπόθεσης ότι η X είναι στοχαστικά μεγαλύτερη από την Y (ή της μονόπλευρης εναλλακτικής με αντίστροφη φορά $H_2 : \overline{F}_X(x) <_{st} \overline{F}_Y(x)$). Πρακτικά, η συγκεκριμένη εναλλακτική υπόθεση δηλώνει ότι η $X-$διεργασία παράγει πιο αξιόπιστα προϊόντα (δηλαδή που έχουν μεγαλύτερο χρόνο ζωής) σε σχέση με την $Y-$διεργασία. Για έναν τέτοιο έλεγχο μπορεί να χρησιμοποιηθεί η αθροιστική βαθμολογική συνάρτηση τύπου *Wilcoxon* (*Wilcoxon's rank–sum statistic*), η οποία είχε αρχικά προταθεί στην εργασία του Wilcoxon (1945). Για την κατασκευή της συνάρτησης αυτής, ενώνουμε τα δύο δείγματα και στη συνέχεια βαθμολογούμε τις παρατηρήσεις στο από κοινού δείγμα από 1 μέχρι $(m+n)$. Τότε, η αθροιστική βαθμολογική στατιστική συνάρτηση τύπου *Wilcoxon* δίνεται από τον τύπο

$$W_R = \sum_{i=1}^{m} Rank(X_i),$$

και εκφράζει το άθροισμα των βαθμών των $X-$παρατηρήσεων στο ενιαίο δείγμα. Δεδομένου ότι υπό την εναλλακτική υπόθεση πως η X είναι στοχαστικά μεγαλύτερη από την Y, αναμένουμε ότι το $X-$δείγμα θα λάβει κυρίως τους μεγάλους βαθμούς, έπεται ότι θα έχουμε ένδειξη για απόρριψη της μηδενικής υπόθεσης H_0 έναντι της εναλλακτικής H_1, αν $W_R \geq w$, όπου w είναι μια κατάλληλα επιλεγμένη κρίσιμη τιμή. Προφανώς, αν η εναλλακτική υπόθεση ήταν αμφίπλευρη ($H_1 : F_X(x) \neq F_Y(x)$), θα είχαμε ένδειξη απόρριψης της H_0 αν ίσχυε $W_R \geq w_2$ ή $W_R \leq w_1$, όπου $w_1 < w_2$ δύο κατάλληλα επιλεγμένες σταθερές. Οι Ng & Balakrishnan (2004) πρότειναν παραλλαγές της παραπάνω βαθμολογικής συνάρτησης για τον έλεγχο ισονομίας δύο δειγμάτων, κατασκευάζοντας κατάλληλους στατιστικούς ελέγχους προτεραιότητας, ενώ οι Balakrishnan *et al.* (2008b) τη χρησιμοποίησαν για το σχεδιασμό μη παραμετρικών διαγραμμάτων ελέγχου. Υπό τη μηδενική υπόθεση H_0, έχουμε ότι

$$P(Rank(X_i) = i_1, \ldots, Rank(X_m) = i_m \mid H_0 : F_X = F_Y) = \frac{n!}{(m+n)!}$$

για κάθε υποσύνολο $\{i_1, \ldots, i_m\}$ του συνόλου $\{1, 2, \ldots, m+n\}$. Συνεπώς, προκύπτει ότι

$$P(W_R = w \mid H_0 : F_X = F_Y) = \frac{N_w}{\binom{m+n}{m}},$$

όπου N_w είναι το πλήθος των υποσυνόλων $\{i_1, \ldots, i_m\}$ με $i_1 + \ldots + i_m = w$. Για τον υπολογισμό της κατανομής της βαθμολογικής στατιστικής συνάρτησης τύπου *Wilcoxon* μέσω αναδρομικών σχέσεων, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Brus (1989) και Chang (1992). Αξίζει να σημειωθεί ότι η βαθμολογική συνάρτηση W_R μπορεί να υπολογισθεί με τη βοήθεια των ποσοτήτων M_i, που ορίσθηκαν στο Κεφάλαιο 3.

Στο Κεφάλαιο 5, προτείνουμε τρία νέα ελεύθερα κατανομής διαγράμματα ελέγχου τύπου *Shewhart*, τα οποία χρησιμοποιούν συναρτήσεις ροών ή βαθμολογικές συναρτήσεις για να οδηγηθούμε σε απόφαση κατά το πόσο μια διεργασία παραμένει εντός ελέγχου ή έχει μετατοπιστεί σε εκτός ελέγχου κατάσταση. Η κατανομή των στατιστικών συναρτήσεων που απεικονίζονται στα νέα μη παραμετρικά διαγράμματα ελέγχου προσδιορίζεται με τη βοήθεια των ποσοτήτων M_i. Το πρώτο διάγραμμα ροών χρησιμοποιεί το μέγιστο μήκος ροής των παρατηρήσεων του εξεταζόμενου δείγματος στο από κοινού δείγμα, ενώ το δεύτερο προτεινόμενο διάγραμμα λαμβάνει υπόψιν το πλήθος των ροών των παρατηρήσεων του εξεταζόμενου δείγματος, των οποίων το μήκος υπερβαίνει μια προκαθορισμένη τιμή. Τέλος, το τρίτο διάγραμμα στηρίζεται στο άθροισμα των βαθμών των παρατηρήσεων του εξεταζόμενου δείγματος, οι οποίες βρίσκονται ανάμεσα στα όρια ελέγχου.

Στην Παράγραφο 5.3, εξάγουμε ακριβείς τύπους για την εντός ελέγχου κατανομή των προαναφερθεισών στατιστικών συναρτήσεων και υπολογίζουμε τις πιθανότητες λανθασμένου συναγερμού των τριών μη παραμετρικών διαγραμμάτων ελέγχου. Στην Παράγραφο 5.4, εξετάζουμε εκτός ελέγχου μετατοπίσεις της διεργασίας σε εναλλακτικές τύπου *Lehmann*, υπολογίζοντας την πιθανότητα ανίχνευσης της παραπάνω μετατόπισης στα προτεινόμενα διαγράμματα. Η ακριβής κατανομή του μήκους ροής, η οποία προκύπτει με την εφαρμογή της τεχνικής δέσμευσης που αρχικά είχε προταθεί στην εργασία του Chakraborti (2000) και χρησιμοποιήθηκε και στο προηγούμενο Κεφάλαιο, δίνεται στην Παράγραφο 5.5. Τέλος, στην Παράγραφο 5.6 παρουσιάζονται αριθμητικοί υπολογισμοί, οι οποίοι επαληθεύουν την αποτελεσματικότητα των νέων διαγραμμάτων, τόσο σε εντός ελέγχου καταστάσεις, όσο και σε περιπτώσεις εκτός ελέγχου μετατοπίσεων της διεργασίας.

5.2. Τρία νέα μη παραμετρικά διαγράμματα ελέγχου

Όπως έχει ήδη αναφερθεί, τα όρια ελέγχου των νέων μη παραμετρικών διαγραμμάτων που παρουσιάζονται στο Κεφάλαιο 5 του παρόντος κειμένου, προσδιορίζονται με τη βοήθεια ενός δείγματος αναφοράς που συλλέγεται από τη διεργασία, όταν βρίσκεται σε εντός ελέγχου κατάσταση. Έστω ότι $X_1, X_2, ..., X_m$ είναι ένα τυχαίο δείγμα μεγέθους m που προέρχεται από την εντός ελέγχου (αθροιστική) κατανομή $F_X(x) = F(x)$ και ας υποθέσουμε ότι χρησιμοποιούνται ως όρια ελέγχου του διαγράμματος δύο συγκεκριμένες διατεταγμένες παρατηρήσεις $X_{a:m}, X_{b:m}$, δηλαδή

$$LCL = X_{a:m} \ , \quad UCL = X_{b:m}$$

($1 \le a < b \le m$). Οι παράμετροι a, b είναι παράμετροι σχεδιασμού των προτεινόμενων διαγραμμάτων και ο προσδιορισμός τους επιτυγχάνεται μέσω δύο διαδικασιών. Η πρώτη διαδικασία απαιτεί την επίτευξη μιας συγκεκριμένης πιθανότητας λανθασμένου συναγερμού, ενώ η δεύτερη στηρίζεται σε μια προκαθορισμένη τιμή του εντός ελέγχου μέσου μήκους ροής (ARL_{in}), συνήθως ίση με 370 ή 500. Αξίζει να σημειωθεί ότι το ARL_{in} των τριών διαγραμμάτων ελέγχου παραμένει ίδιο για όλες τις (εντός ελέγχου) συνεχείς κατανομές.

Η κατασκευή των νέων διαγραμμάτων ελέγχου βασίζεται στην έννοια της ροής (*run*) ίδιων συμβόλων (ή αποτελεσμάτων). Πιο συγκεκριμένα, ως ροή αποτελεσμάτων ίδιου τύπου ορίζουμε την εμφάνιση διαδοχικών (χωρίς διακοπή) όμοιων αποτελεσμάτων τα οποία ακολουθούνται και έπονται από διαφορετικά αποτελέσματα. Το πλήθος k των όμοιων στοιχείων μιας ροής ονομάζεται μήκος ροής (k θετικός ακέραιος). Για παράδειγμα, στην ακολουθία δύο (διαφορετικών) συμβόλων X, Y, που φαίνεται παρακάτω

$$X\ X\ Y\ Y\ Y\ X\ Y\ Y\ X\ X\ X\ X,$$

έχουμε διαδοχικά μία ροή των X μήκους 2, ακολουθούμενη από μία ροή των Y μήκους 3, μία ροή των X μήκους 1, μία ροή των Y μήκους 2 και τέλος μία ροή των X μήκους 4. Για περισσότερες λεπτομέρειες σχετικά με τη Θεωρία Ροών και στατιστικών συναρτήσεων ροών (*run statistics*), ο ενδιαφερόμενος αναγνώστης παραπέμπεται στη μονογραφία των Balakrishnan & Koutras (2001) και τις εργασίες των Antzoulakos *et al*. (2008) και Μπερσίμη (2005).

Ας υποθέσουμε ότι, μετά τον καθορισμό των ορίων ελέγχου LCL, UCL, συλλέγονται ανεξάρτητα τυχαία δείγματα από τη διεργασία και θέλουμε να εξετάσουμε αν παραμένει εντός ελέγχου ή αν έχει μετατοπιστεί σε εκτός ελέγχου κατάσταση. Αν $Y_1, Y_2, ..., Y_n$ είναι ένα εξεταζόμενο δείγμα μεγέθους n και $F_Y(x) = G(x)$ η αντίστοιχη αθροιστική συνάρτηση κατανομής του, ο στόχος μας είναι να ανιχνεύσουμε μια πιθανή μετατόπιση της εντός ελέγχου κατανομής από $F(x)$ σε $G(x)$, δηλαδή να ελέγξουμε τη μηδενική υπόθεση $H_0 : F(x) = G(x)$ έναντι της δίπλευρης εναλλακτικής $H_1 : F(x) \neq G(x)$.

Οι στατιστικές συναρτήσεις που απεικονίζονται στα προτεινόμενα διαγράμματα ελέγχου ορίζονται με τη βοήθεια

α. των ροών των Y–παρατηρήσεων που βρίσκονται ανάμεσα στα όρια ελέγχου LCL, UCL,

β. των βαθμών των Y–παρατηρήσεων που βρίσκονται ανάμεσα στα όρια ελέγχου LCL, UCL.

Η μεθοδολογία για τα τρία διαγράμματα ελέγχου μπορεί να συνοψιστεί ως εξής: υπό τη μηδενική υπόθεση $H_0 : F = G$ (δηλαδή αν, τόσο το δείγμα αναφοράς, όσο και τα εξεταζόμενα τυχαία δείγματα που συλλέγονται από τη διεργασία, προέρχονται από την ίδια κατανομή), το πλήθος των Y–παρατηρήσεων που βρίσκονται μεταξύ δύο διαδοχικών X–παρατηρήσεων δεν θα πρέπει να λαμβάνει ακραίες τιμές, όπου οι ακραίες τιμές ορίζονται βάσει της αναλογίας n/m. Συνεπώς, δύο πιθανές στατιστικές συναρτήσεις που μπορεί να χρησιμοποιηθούν για να αποφασίζουμε αν η διεργασία παραμένει εντός ελέγχου ή όχι, έχουν ως εξής:

- το μέγιστο μήκος ροής των Y–παρατηρήσεων που παρατηρούνται μεταξύ των ορίων ελέγχου,
- το πλήθος των ροών των Y–παρατηρήσεων (μεταξύ των ορίων ελέγχου), των οποίων το μήκος υπερβαίνει μια προκαθορισμένη τιμή k.

Μια τρίτη επιλογή που διαφέρει από τις παραπάνω, είναι μέσω

- του αθροίσματος των βαθμών (στο από κοινού δείγμα των X και Y παρατηρήσεων) των Y–παρατηρήσεων που βρίσκονται ανάμεσα στα όρια ελέγχου.

Αξίζει να σημειωθεί ότι η τελευταία επιλογή αντιστοιχεί στη γνωστή αθροιστική βαθμολογική συνάρτηση τύπου *Wilcoxon*. Όπως αναφέρθηκε και προηγουμένως, οι τρεις στατιστικές συναρτήσεις που αναφέρθηκαν προηγουμένως, είναι δυνατόν να εκφρασθούν με τη βοήθεια των στατιστικών συναρτήσεων προτεραιότητας. Πιο συγκεκριμένα, έστω ότι M_i, i=1,2,…,m εκφράζει το πλήθος των Y–παρατηρήσεων του εξεταζόμενου δείγματος που βρίσκονται μεταξύ της (i–1)–οστής και i–οστής διατεταγμένης παρατήρησης του X–δείγματος (σύμβαση: $X_{(0)} = -\infty$). Με άλλα λόγια, τα M_i είναι τα μήκη των ροών των Y–παρατηρήσεων μεταξύ διαδοχικών X– παρατηρήσεων. Οι τρεις προαναφερθείσες στατιστικές συναρτήσεις δίνονται από τις ακόλουθες σχέσεις

$$R = \max(M_{a+1}, M_{a+2}, ..., M_b), \quad N_k = \left|\{M_i : a+1 \le i \le b \text{ and } M_i \ge k\}\right|, \quad W = \sum_{i=a+1}^{b} W_i$$

(k είναι μια πρόσθετη παράμετρος σχεδιασμού που λαμβάνει ακέραιες τιμές), όπου το W_i εκφράζει το άθροισμα των βαθμών των Y– παρατηρήσεων που βρίσκονται μεταξύ των $X_{(i-1)}$ και $X_{(i)}$. Προκειμένου να εκφράσουμε την αθροιστική βαθμολογική συνάρτηση τύπου *Wilcoxon* με τη βοήθεια των M_i, αρκεί να παρατηρήσουμε ότι το πλήθος των Y–παρατηρήσεων που βρίσκονται πριν από την $X_{(i-1)}$ είναι

$(M_1 + M_2 + ... + M_{i-1}) = \sum_{r=1}^{i-1} M_r$, συνεπώς οι βαθμοί των Y–παρατηρήσεων που βρίσκονται μεταξύ των $X_{(i-1)}$ και $X_{(i)}$, είναι $\sum_{r=1}^{i-1} M_r + (i-1) + 1$, $\sum_{r=1}^{i-1} M_r + (i-1) + 2$, ..., $\sum_{r=1}^{i-1} M_r + (i-1) + M_i$ αντίστοιχα. Με βάση τα παραπάνω, έχουμε ότι

$$W_i = \sum_{j=1}^{M_i} ((i-1) + \sum_{r=1}^{i-1} M_r + j) = M_i((i-1) + \sum_{r=1}^{i-1} M_r) + \frac{M_i(M_i + 1)}{2}$$

και υλοποιώντας αλγεβρικές πράξεις, καταλήγουμε στην ακόλουθη έκφραση

$$W = \frac{1}{2}\left(\sum_{i=a+1}^{b} M_i \right)^2 + \sum_{i=a+1}^{b} iM_i + (M_0 + a - \frac{3}{2}) \sum_{i=a+1}^{b} M_i . \tag{5.1}$$

Όταν χρησιμοποιείται η στατιστική συνάρτηση R, η διεργασία κρίνεται ότι είναι εντός ελέγχου, αν επαληθεύονται οι δύο ακόλουθες συνθήκες

$$R \le r \qquad \text{και} \qquad M_0 \le r_0 ,$$

όπου $M_0 = \sum_{i=1}^{a} M_i$ εκφράζει το πλήθος των παρατηρήσεων του Y–δείγματος πριν το LCL και r, r_0 είναι παράμετροι σχεδιασμού του διαγράμματος.

Όταν εφαρμόζεται η στατιστική συνάρτηση N_k, η διεργασία κρίνεται ότι είναι εντός ελέγχου, αν επαληθεύονται οι δύο ακόλουθες συνθήκες

$$N_k \le r_1 \quad \text{και} \quad M_0 \le r_0 ,$$

όπου r_1, r_0 αποτελούν παραμέτρους σχεδιασμού.

Τέλος, αν στο διάγραμμα απεικονίζεται η στατιστική συνάρτηση W, τότε δεν θα παράγεται σήμα για εκτός ελέγχου μετατόπιση, εφόσον επαληθεύονται οι ακόλουθες δύο συνθήκες

$$W \le w \quad \text{και} \quad M_0 \le r_0 ,$$

όπου w, r_0 είναι παράμετροι σχεδιασμού του διαγράμματος. Σχετικά με τον τελευταίο κανόνα, αξίζει να σημειωθεί ότι η μέγιστη δυνατή τιμή της W προκύπτει όταν όλες οι Y–παρατηρήσεις βρίσκονται στο διάστημα $(X_{b-1:m}, X_{b:m})$. Στη συγκεκριμένη περίπτωση, οι βαθμοί τους είναι ίσοι με $(b-1) + j$, $j = 1,2,...,n$ και η αθροιστική βαθμολογική συνάρτηση θα πάρει την τιμή

$$n(b-1)+\sum_{j=1}^{n} j = \frac{n(n+2b-1)}{2}.$$

Αντίστοιχα, η μικρότερη δυνατή τιμή της W είναι ίση με μηδέν (η οποία επιτυγχάνεται αν καμία Y–παρατήρηση δεν βρίσκεται μεταξύ των LCL και UCL) και συνεπώς το στήριγμα της συνάρτησης W είναι

$$R_W = \left\{0,1,\ldots,\frac{n(n+2b-1)}{2}\right\}. \tag{5.2}$$

Θα παρουσιάσουμε στη συνέχεια ένα παράδειγμα για να γίνουν πιο κατανοητές οι προτεινόμενες διαδικασίες ελέγχου καθώς και η λογική που τις διέπει. Ας υποθέσουμε ότι θέλουμε να κατασκευάσουμε μη παραμετρικά διαγράμματα ελέγχου χρησιμοποιώντας ένα δείγμα αναφοράς μεγέθους $m=10$ και τυχαία εξεταζόμενα δείγματα μεγέθους $n=4$. Στον Πίνακα 5.1 παρουσιάζουμε μια επιλογή των παραμέτρων σχεδιασμού για κάθε ένα από τα τρία διαγράμματα που περιγράφηκαν νωρίτερα, η οποία εξασφαλίζει πιθανότητα λανθασμένου συναγερμού 10% $(FAR \le 0.10)$. Για τον προσδιορισμό των παραμέτρων σχεδιασμού, χρησιμοποιούμε σχέσεις που θα αποδειχθούν στη συνέχεια του Κεφαλαίου 5.

Πίνακας 5.1. Τρεις σχεδιασμοί με $FAR \le 0.10$

R-διάγραμμα					N- διάγραμμα						W- διάγραμμα				
a	b	r_0	r	FAR	a	b	r_0	r_1	k	FAR	a	b	r_0	w	FAR
1	4	1	2	0.0989	3	6	2	0	2	0.0979	1	4	4	10	0.0919

Προκειμένου να αξιολογήσουμε την απόδοση των τριών διαγραμμάτων ελέγχου, θεωρούμε την περίπτωση που οι εντός ελέγχου παρατηρήσεις προέρχονται από την ομοιόμορφη κατανομή στο (0,1) (δηλαδή $F(x)=x,\ 0<x<1$). Οι δύο πρώτες γραμμές του Πίνακα 5.2 παρουσιάζουν ένα διατεταγμένο δείγμα αναφοράς που

Πίνακας 5.2. Δείγμα αναφοράς και εξεταζόμενα τυχαία δείγματα

Δείγμα αναφοράς	0.0547494	0.0915627	0.1925020	0.3298570	0.5872420
	0.6464540	0.7250190	0.7319490	0.8796790	0.9683790
Εντός ελέγχου δείγμα	0.1492300	0.3494970	0.6038480	0.6787060	
Εκτός ελέγχου δείγμα	0.000000007	0.000127800	0.127365000	0.144250000	

παράγεται με προσομοίωση και το οποίο χρησιμοποιείται για τον προσδιορισμό των ακόλουθων ορίων ελέγχου

R– διάγραμμα: $LCL = X_{1:10}$, $UCL = X_{4:10}$,

N– διάγραμμα: $LCL = X_{3:10}$, $UCL = X_{6:10}$,

W– διάγραμμα: $LCL = X_{1:10}$, $UCL = X_{4:10}$.

Αν υποθέσουμε ότι, αφού συλλέγεται το δείγμα αναφοράς, η διεργασία παραμένει εντός ελέγχου, τα τυχαία δείγματα μεγέθους $n = 4$ που εξάγονται θα προέρχονται από την ομοιόμορφη κατανομή στο (0,1). Ένα τέτοιο δείγμα έχει παραχθεί με προσομοίωση και παρουσιάζεται στην τρίτη γραμμή του Πίνακα 5.2. Οι $m + n = 14$ παρατηρήσεις του από κοινού διατεταγμένου (X, Y) δείγματος σχηματίζουν την ακόλουθη διαδοχή

$$X\ X\ Y\ X\ X\ Y\ X\ Y\ X\ Y\ X\ X\ X\ X$$

και οι ποσότητες M_i παίρνουν τις τιμές

$$M_1 = 0, M_2 = 0, M_3 = 1, M_4 = 0, M_5 = 1, M_6 = 1, M_7 = 1, M_8 = 0, M_9 = 0, M_{10} = 0 .$$

Αν το R–διάγραμμα εφαρμόζεται, η στατιστική συνάρτηση R είναι ίση με

$$R = \max(M_2, M_3, M_4) = 1$$

και αφού ισχύει ότι

$$R \le 2 = r \quad \text{και} \quad M_0 = M_1 = 0 \le 1 = r_0 ,$$

δεν θα παραχθεί σήμα για εκτός ελέγχου μετατόπιση. Αντιστοίχως, αν χρησιμοποιηθεί το N–διάγραμμα, παίρνουμε ότι

$$N_2 = \left|\{M_i : 4 \le i \le 6 \text{ and } M_i \ge 2\}\right| = 0$$

και η διεργασία κρίνεται ότι είναι εντός ελέγχου αν

$$N_2 \le 0 = r_1 \quad \text{και} \quad M_0 = \sum_{i=1}^{3} M_i = 1 \le 2 = r_0 .$$

Τέλος, η βαθμολογική συνάρτηση τύπου *Wilcoxon* μπορεί να υπολογισθεί με τη βοήθεια της (5.1) (για $a = 1$, $b = 4$) ως εξής

$$W = \frac{1}{2}\left(\sum_{i=2}^{4} M_i\right)^2 + \sum_{i=2}^{4} iM_i + (M_0 + a - \frac{3}{2})\sum_{i=2}^{4} M_i = 3 ,$$

ή απευθείας: μεταξύ της $1^{ης}$ και $2^{ης}$ παρατήρησης του δείγματος αναφοράς υπάρχει μόνο ένα Y (η $3^{η}$ παρατήρηση στο διατεταγμένο δείγμα των 14 παρατηρήσεων) συνεπώς $W = 3$. Με βάση τα παραπάνω, συμπεραίνουμε ότι η διεργασία δεν έχει μετατοπιστεί σε εκτός ελέγχου κατάσταση, αφού

$$W \le 10 = w \quad \text{και} \quad M_0 = M_1 = 0 \le 4 = r_0 .$$

Στη συνέχεια υποθέτουμε ότι η διεργασία μετατοπίζεται στην κατανομή $G(x) = x^{1/4}, 0 < x < 1$. Χρησιμοποιώντας το προσομοιωμένο τυχαίο δείγμα που παρουσιάζεται στην τέταρτη γραμμή του Πίνακα 5.2, προκύπτει η ακόλουθη σειρά των $X-Y$ παρατηρήσεων

$$Y\,Y\,X\,X\,Y\,Y\,X\,X\,X\,X\,X\,X\,X\,X$$

ενώ οι ποσότητες M_i δίνονται ως ακολούθως

$M_1 = 2, M_2 = 0, M_3 = 2, M_4 = 0, M_5 = 0, M_6 = 0, M_7 = 0, M_8 = 0, M_9 = 0, M_{10} = 0$.

Οι τιμές των τριών στατιστικών συναρτήσεων είναι τώρα ίσες με

$$R = 2\,,\ N_2 = 1\,,\ W = 11$$

και μπορεί εύκολα να ελεγχθεί ότι οι συνθήκες για εντός ελέγχου διεργασία δεν ικανοποιούνται (έχουμε $M_0 = 2 > 1 = r_0$ για το *R*–διάγραμμα, $N_2 = 1 > 0 = r_1$ για το *N*–διάγραμμα και $W = 11 > 10 = w$ για το *W*–διάγραμμα). Συνεπώς, όλα τα προτεινόμενα μη παραμετρικά διαγράμματα ελέγχου, στη συγκεκριμένη περίπτωση, πέτυχαν να ανιχνεύσουν τη μετατόπιση.

5.3. Μελέτη της πιθανότητας λανθασμένου συναγερμού των νέων διαγραμμάτων ελέγχου

Στη συγκεκριμένη Παράγραφο θα υπολογίσουμε την πιθανότητα λανθασμένου συναγερμού των τριών νέων διαγραμμάτων ελέγχου, δηλαδή την πιθανότητα να παραχθεί σήμα για εκτός ελέγχου μετατόπιση της διεργασίας, ενώ στην πραγματικότητα η διεργασία παραμένει εντός ελέγχου. Αρχικά, θα μελετήσουμε την εντός ελέγχου κατανομή των στατιστικών συναρτήσεων που απεικονίζονται στα τρία διαγράμματα ελέγχου που εισήχθησαν στην Παράγραφο 5.2. Έστω ότι $X_1, X_2, ..., X_m$ είναι ένα δείγμα αναφοράς μεγέθους m που έχει συλλεχθεί από την εντός ελέγχου κατανομή $F_X(x) = F(x)$ και $Y_1, Y_2, ..., Y_n$ ένα τυχαίο δείγμα που είναι διαθέσιμο από την ίδια κατανομή, δηλαδή $F_X(x) = F_Y(x)$ για κάθε x.

Οι ακριβείς κατανομές των στατιστικών συναρτήσεων R και N_k μπορούν να μελετηθούν με τη βοήθεια της από κοινού κατανομής των $(M_{a+1}, M_{a+2}, ..., M_b)$, ενώ η συνάρτηση W μπορεί να αναλυθεί με χρήση της από κοινού κατανομής των $(M_0, M_{a+1}, M_{a+2}, ..., M_b)$. Οι Balakrishnan & Frattina (2000) και οι Balakrishnan & Ng (2001), πρότειναν, για την εφαρμογή ενός ελέγχου προτεραιότητας στη σύγκριση δύο κατανομών F_X, F_Y, το μέγιστο στατιστικό έλεγχο προτεραιότητας (*maximal precedence test*). Η στατιστική συνάρτηση M που χρησιμοποιείται στον παραπάνω έλεγχο, ορίζεται ως το μέγιστο πλήθος των Y–παρατηρήσεων που συμβαίνουν πριν την πρώτη, μεταξύ πρώτης και δεύτερης,…, μεταξύ της (r-1)–οστής και r–οστής διατεταγμένης παρατήρησης του X–δείγματος. Συνεπώς, ισχύει ότι $M = \max(M_1, M_2, ..., M_r)$ και η κατανομή της, υπό τη μηδενική υπόθεση $H_0 : F_X = F_Y$, μπορεί να δοθεί με τη βοήθεια της από κοινού κατανομής των $(M_1, M_2, ..., M_r)$, η οποία περιγράφεται στο ακόλουθο Λήμμα (για την απόδειξη του Λήμματος 5.1, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στη μονογραφία των Balakrishnan & Ng (2006)).

Λήμμα 5.1. *Η από κοινού συνάρτηση πιθανότητας των* $(M_1, M_2, ..., M_r)$, $r \geq 1$ *υπό τη μηδενική υπόθεση* $H_0 : F_X = F_Y$ *δίνεται στην ακόλουθη σχέση*

$$P(M_1 = m_1, M_2 = m_2, ..., M_r = m_r | H_0) = \frac{\binom{m + n - \sum_{i=1}^{r} m_i - r}{m - r}}{\binom{m+n}{n}}, \qquad \sum_{i=1}^{r} m_i \leq n .$$

Στη συνέχεια, θα αποδείξουμε ακριβείς εκφράσεις για τις από κοινού κατανομές των $(M_0, M_{a+1}, ..., M_b)$ και $(M_{a+1}, ..., M_b)$, οι οποίες όπως αναφέρθηκε προηγουμένως, μπορούν να χρησιμοποιηθούν για την εύρεση των εντός ελέγχου κατανομών των στατιστικών συναρτήσεων R, N_k και W.

Πρόταση 5.1. *Έστω ότι* $p_{a,b}(m_0, m_{a+1}, m_{a+2}, ..., m_b)$ *εκφράζει την ποσότητα*

$$p_{a,b}(m_0, m_{a+1}, m_{a+2}, ..., m_b) = \frac{\binom{m_0 + a - 1}{a - 1}\binom{m + n - m_0 - \sum_{i=a+1}^{b} m_i - b}{m - r}}{\binom{m+n}{n}}$$

για κάθε υποσύνολο μη αρνητικών ακεραίων αριθμών $m_0, m_{a+1}, ..., m_b$ *που ικανοποιούν τη συνθήκη* $m_0 + \sum_{i=a+1}^{b} m_i \leq n$ *(διαφορετικά,* $p_{a,b}(m_0, m_{a+1}, m_{a+2}, ..., m_b) = 0$*). Τότε,*

α. Η από κοινού συνάρτηση πιθανότητας των $(M_0, M_{a+1}, ..., M_b)$ *υπό την υπόθεση ότι η διεργασία βρίσκεται εντός ελέγχου είναι ίση με*

$$P(M_0 = m_0, M_{a+1} = m_{a+1}, ..., M_b = m_b) = p_{a,b}(m_0, m_{a+1}, m_{a+2}, ..., m_b)$$

β. Η από κοινού συνάρτηση πιθανότητας των $(M_{a+1},...,M_b)$ υπό την υπόθεση ότι η διεργασία βρίσκεται εντός ελέγχου δίνεται στην ακόλουθη σχέση

$$P(M_{a+1} = m_{a+1}, ..., M_b = m_b) = \sum_{m_0=0}^{n} p_{a,b}(m_0, m_{a+1}, m_{a+2}, ..., m_b).$$

Απόδειξη. Αν η διεργασία είναι εντός ελέγχου, τότε τα $X-$ και $Y-$δείγματα προέρχονται από την ίδια κατανομή δηλαδή $F_X = F_Y$. Λαμβάνοντας υπόψιν ότι $M_0 = \sum_{i=1}^{a} M_i$, η πιθανότητα

$$p = P(M_0 = m_0, M_{a+1} = m_{a+1}, ..., M_b = m_b) = P(\sum_{i=1}^{a} M_i = m_0, M_{a+1} = m_{a+1}, ..., M_b = m_b)$$

μπορεί να υπολογισθεί με τη βοήθεια του Λήμματος 5.1 για $r = b$, αθροίζοντας για όλες τις τιμές $m_1, m_2, ..., m_a$ που ικανοποιούν τη συνθήκη $\sum_{i=1}^{a} m_i = m_0$, δηλαδή

$$p = \binom{m+n}{n}^{-1} \sum_{\substack{m_1, m_2, ..., m_a \\ \sum_{i=1}^{a} m_i = m_0}} \binom{m+n-\sum_{i=1}^{b} m_i - b}{m-b}.$$

Στη συνέχεια, γράφοντας την πιθανότητα p στη μορφή

$$p = \binom{m+n}{n}^{-1} \sum_{\substack{m_1, m_2, ..., m_a \\ \sum_{i=1}^{a} m_i = m_0}} \binom{m+n-m_0-\sum_{i=a+1}^{b} m_i - b}{m-b}$$

είναι σαφές ότι

$$p = c \cdot \binom{m+n}{n}^{-1} \binom{m+n-m_0-\sum_{i=a+1}^{b} m_i - b}{m-b}$$

όπου c εκφράζει το πλήθος των μη αρνητικών ακεραίων λύσεων της γραμμικής εξίσωσης $\sum_{i=1}^{a} m_i = m_0$. Η απόδειξη ολοκληρώνεται, χρησιμοποιώντας τη σχέση (βλ. Charalambides (2002), σελ. 69)

$$c = \binom{m_0+a-1}{a-1}.$$

β. Το ζητούμενο αποτέλεσμα προκύπτει άμεσα υπολογίζοντας την περιθώρια κατανομή των $(M_{a+1},...,M_b)$ από τη συνάρτηση πιθανότητας που δίνεται στο μέρος (α). □

Με τη βοήθεια της Πρότασης 5.1, μπορούμε να εκφράσουμε την αθροιστική συνάρτηση κατανομής της στατιστικής συνάρτησης $R = \max(M_{a+1}, M_{a+2},...,M_b)$ ως εξής

$$P(R \le r) = P(M_{a+1} \le r, M_{a+2} \le r,..., M_b \le r) = \sum_{m_0=0}^{n} \sum_{m_{a+1},...,m_b} p_{a,b}(m_0, m_{a+1},...,m_b)$$

όπου το εσωτερικό άθροισμα εκτείνεται σε όλες τις τιμές των m_i, $i = a+1,...,b$ που ικανοποιούν τις συνθήκες

$$0 \le m_i \le r,\ i = a+1,..,b \quad \text{και} \quad \sum_{i=a+1}^{b} m_i \le n - m_0 .$$

Μια αντίστοιχη έκφραση μπορεί να δοθεί για την $P(N_k = r)$ (το εσωτερικό άθροισμα θα περιλαμβάνει εκείνους τους συνδυασμούς των $(m_{a+1},...,m_b)$ που έχουν ακριβώς *r* συντεταγμένες μεγαλύτερες ή ίσες από την τιμή *k*, ενώ ταυτόχρονα πρέπει να ικανοποιείται η συνθήκη $\sum_{i=a+1}^{b} m_i \le n - m_0$).

Τέλος, η κατανομή της βαθμολογικής συνάρτησης τύπου *Wilcoxon*, όπως αυτή ορίζεται στην (5.1), μπορεί να εκφρασθεί ως

$$P(W = w) = \sum_{m_0, m_{a+1}, m_b} p_{a,b}(m_0, m_{a+1},...,m_b),$$

όπου $w \in R_W$ και το άθροισμα υπολογίζεται για όλες τις μη αρνητικές τιμές $m_0, m_{a+1},...,m_b$ που ικανοποιούν τις συνθήκες

$$m_0 + \sum_{i=a+1}^{b} m_i \le n \quad \text{και} \quad \frac{1}{2}\left(\sum_{i=a+1}^{b} m_i\right)^2 + \sum_{i=a+1}^{b} i m_i + (m_0 + a - \frac{3}{2}) \sum_{i=a+1}^{b} m_i = w. \qquad (5.3)$$

Κάνοντας χρήση της από κοινού κατανομής των $(M_0, M_{a+1},...,M_b)$ (υπό τη μηδενική υπόθεση $H_0 : F_X = F_Y$), μπορούμε άμεσα να καταλήξουμε σε εκφράσεις για την πιθανότητα *FAR* των διαγραμμάτων ελέγχου *R*, *N* και *W* (στη συνέχεια θα χρησιμοποιούμε την ορολογία *R*–διάγραμμα, *N*– διάγραμμα και *W*– διάγραμμα αντίστοιχα). Για παράδειγμα, στο *R*–διάγραμμα, η πιθανότητα να μην παραχθεί σήμα ότι η διεργασία έχει μετατοπιστεί σε εκτός ελέγχου κατάσταση δίνεται στην ακόλουθη σχέση

$$P(R \le r \text{ and } M_0 \le r_0) = P(M_0 \le r_0, M_{a+1} \le r, M_{a+2} \le r, ..., M_b \le r)$$
$$= \sum_{m_0, m_{a+1}, ..., m_b} p_{a,b}(m_0, m_{a+1}, m_{a+2}, ..., m_b)$$

όπου το άθροισμα υπολογίζεται για όλες τις μη αρνητικές τιμές $m_0, m_{a+1}, m_{a+2}, ..., m_b$ που ικανοποιούν τις συνθήκες

$$m_0 \le r_0,\ m_{a+1} \le r,\ m_{a+2} \le r, ..., m_b \le r\ ,\ \sum_{i=a+1}^{b} m_i \le n - m_0\ .$$

Συνεπώς, η πιθανότητα λανθασμένου συναγερμού θα βρίσκεται άμεσα από την έκφραση

$$FAR = 1 - P(R \le r \text{ και } M_0 \le r_0)\,.$$

Αντίστοιχες εκφράσεις είναι δυνατόν να δοθούν και για τα άλλα δύο μη παραμετρικά διαγράμματα ελέγχου (N–διάγραμμα και W–διάγραμμα).

5.4. Μελέτη των νέων διαγραμμάτων ελέγχου για εναλλακτικές υποθέσεις τύπου *Lehmann*

Στη συγκεκριμένη Παράγραφο θα μελετήσουμε την εκτός ελέγχου κατανομή των τριών στατιστικών συναρτήσεων που ορίσθηκαν προηγουμένως. Για το λόγο αυτό, θεωρούμε ότι οι παρατηρήσεις $Y_1, Y_2, ..., Y_n$ ενός τυχαίου δείγματος προέρχονται από μια συνεχή κατανομή $F_Y(x) = G(x)$ και εκφράζουμε την πιθανότητα ότι ο κανόνας δεν παράγει σήμα, στην ακόλουθη μορφή

$$p(m, n, a, b; \mathbf{c}; F, G) = \sum_{(m_0, m_{a+1}, ..., m_b) \in A} P(M_0 = m_0 \text{ και } M_j = m_j \text{ για } a+1 \le j \le b) \qquad (5.4)$$

όπου $\mathbf{c}$ είναι ένα διάνυσμα σταθερών που πρέπει να προσδιορισθούν ώστε ο κανόνας να ικανοποιεί ορισμένες απαιτήσεις (π.χ. μια προκαθορισμένη πιθανότητα λανθασμένου συναγερμού ή ένα δεδομένο εντός ελέγχου μέσο μήκος ροής) και A είναι το σύνολο τιμών για το διάνυσμα $(M_0, M_{a+1}, M_{a+2}, ..., M_b)$ ώστε το διάγραμμα ελέγχου να μην παράγει σήμα (προφανώς το A εξαρτάται από τις σταθερές $\mathbf{c}$). Είναι σαφές ότι η πιθανότητα λανθασμένου συναγερμού είναι η συμπληρωματική πιθανότητα του $p(m, n, a, b; \mathbf{c}; F, G)$ για την ειδική περίπτωση $G = F$, δηλαδή

$$FAR = 1 - p(m,n,a,b;\mathbf{c};F,F).$$

Η τελευταία ποσότητα μπορεί να υπολογισθεί με τη βοήθεια των σχέσεων που δόθηκαν νωρίτερα.

Ας θεωρήσουμε τη γενική περίπτωση που οι X– και Y–παρατηρήσεις δεν ακολουθούν την ίδια κατανομή, δηλαδή $G(x) \neq F(x)$. Ας συμβολίσουμε με $\mathbf{Y}$ το τυχαίο διάνυσμα $(Y_1, Y_2, ..., Y_n)$ και με $\mathbf{X}_{a:b}$ το διατεταγμένο X–δείγμα που περιλαμβάνει τις X–παρατηρήσεις από την a–οστή έως την b–οστή διατεταγμένη παρατήρηση, δηλαδή $\mathbf{X}_{a:b} = (X_{a:m}, X_{a+1:m}, ..., X_{b:m})$. Τότε, οι ποσότητες $M_0, M_{a+1}, ..., M_b$ είναι συναρτήσεις των $\mathbf{Y}$ και $\mathbf{X}_{a:b}$ και επομένως οι προσθετέοι του δεξιού μέλους στην (5.4) παίρνουν τη μορφή

$$p_{\mathbf{m}} = P(M_0(\mathbf{Y};\mathbf{X}_{a:b}) = m_0 \text{ και } M_j(\mathbf{Y};\mathbf{X}_{a:b}) = m_j \text{ για } a+1 \le j \le b), \qquad (5.5)$$

όπου $\mathbf{m} = (m_0, m_{a+1}, ..., m_b)$ το διάνυσμα των δεικτών που αθροίζονται.

Χρησιμοποιώντας το γεγονός ότι

$$F(\mathbf{X}_{a:b}) = (F(X_{a:m}), F(X_{a+1:m}), ..., F(X_{b:m})) = (U_{a:m}, U_{a+1:m}, ..., U_{b:m}) = \mathbf{U}_{a:b},$$

όπου $\mathbf{U}_{a:b}$ περιλαμβάνει τις ενδιάμεσες διατεταγμένες παρατηρήσεις (από την $a-$οστή έως την $b-$οστή) ενός τυχαίου δείγματος μεγέθους m από την ομοιόμορφη κατανομή στο (0,1), φτάνουμε στην επόμενη ισοδύναμη έκφραση

$$p_{\mathbf{m}} = P(M_0(F(\mathbf{Y});\mathbf{U}_{a:b}) = m_0 \text{ και } M_j(F(\mathbf{Y});\mathbf{U}_{a:b}) = m_j \text{ για } a+1 \le j \le b). \qquad (5.6)$$

Δεδομένου ότι η από κοινού συνάρτηση πυκνότητας του διανύσματος $\mathbf{U}_{a:b}$ είναι (βλ. επίσης Κεφ. 1)

$$f_{a:b}(u_a, u_{a+1}, ..., u_b) = f_{a:b}(\mathbf{u}_{a:b}) = \frac{m!}{(a-1)!(m-b)!}(u_a)^{a-1}(1-u_b)^{m-b}, \; 0 \le u_a \le u_{a+1} \le ... \le u_b \le 1 \qquad (5.7)$$

μπορούμε να εκφράσουμε την πιθανότητα $p_{\mathbf{m}}$ μέσω του ακόλουθου ολοκληρώματος

$$p_{\mathbf{m}} = \iint\limits_{0 \le u_a \le u_{a+1} \le ... \le u_b \le 1} \!\!\!\cdots\!\int P(M_0(F(\mathbf{Y});\mathbf{u}_{a:b}) = m_0 \text{ και } M_j(F(\mathbf{Y});\mathbf{u}_{a:b}) = m_j \text{ για } a+1 \le j \le b)$$

$$\times f_{\mathbf{U}_{a:b}}(\mathbf{u}_{a:b}) du_a du_{a+1} \cdots du_b \qquad (5.8)$$

όπου $\mathbf{u}_{a:b} = (u_a, u_{a+1}, ..., u_b)$. Η πιθανότητα που βρίσκεται μέσα στο ολοκλήρωμα γράφεται στη μορφή

$$P(M_0(\mathbf{Y};F^{-1}(\mathbf{u}_{a:b})) = m_0 \text{ και } M_j(\mathbf{Y};F^{-1}(\mathbf{u}_{a:b})) = m_j \text{ για } a+1 \le j \le b) =$$

$$P(M_0(G(\mathbf{Y});GF^{-1}(\mathbf{u}_{a:b}))=m_0 \text{ και } M_j(G(\mathbf{Y});GF^{-1}(\mathbf{u}_{a:b}))=m_j \text{ για } a+1\le j\le b)$$

και δεδομένου ότι η τυχαία μεταβλητή $G(\mathbf{Y})=(G(Y_1),G(Y_2),...,G(Y_n))=$ $=(U_1,U_2,...,U_n)=\mathbf{U}$ αποτελεί ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή στο διάστημα (0,1), προκύπτει ότι

$$P(M_0(\mathbf{U};GF^{-1}(\mathbf{u}_{a:b}))=m_0 \text{ και } M_j(\mathbf{U};GF^{-1}(\mathbf{u}_{a:b}))=m_j \text{ για } a+1\le j\le b).$$

Συνεπώς, απομένει να δοθεί η απάντηση στο ακόλουθο ερώτημα: αν λάβουμε ένα τυχαίο δείγμα $U_1,U_2,...,U_n$ από την ομοιόμορφη κατανομή στο (0,1), ποια είναι η πιθανότητα ότι m_0 $U-$παρατηρήσεις θα βρεθούν πριν την $GF^{-1}(u_a)$, m_j $U-$παρατηρήσεις θα βρεθούν μεταξύ των $GF^{-1}(u_{j-1})$ και $GF^{-1}(u_j)$ για $j=a+1,a+2,...,b$ και οι υπόλοιπες $n-m_0-\sum_{j=a+1}^{b}m_j$ μετά την $GF^{-1}(u_b)$. Είναι σαφές ότι η πιθανότητα να παρατηρηθεί μια τέτοια διάταξη είναι ίση με $q(GF^{-1}(u_a),GF^{-1}(u_{a+1}),...,GF^{-1}(u_b))$, όπου

$$q(v_a,v_{a+1},...,v_b)=\frac{n!}{m_0!(\prod_{j=a+1}^{b}m_j!)(n-m_0-\sum_{j=a+1}^{b}m_j)!}v_a^{m_0}\prod_{j=a+1}^{b}(v_j-v_{j-1})^{m_j}(1-v_b)^{n-m_0-\sum_{j=a+1}^{b}m_j}, \tag{5.9}$$

είναι η πολυωνυμική πιθανότητα με $0\le v_a\le v_{a+1}\le ...\le v_b\le 1$. Συνδυάζοντας τις σχέσεις (5.6), (5.8) και (5.9) μπορούμε να διατυπώσουμε την επόμενη Πρόταση.

Πρόταση 5.2. *Η από κοινού συνάρτηση πιθανότητας των* $(M_0,M_{a+1},...,M_b)$ *υπό την υπόθεση ότι οι X–παρατηρήσεις ακολουθούν μια κατανομή με αθροιστική συνάρτηση* $F(x)$ *και οι Y–παρατηρήσεις προέρχονται από την κατανομή* $G(x)$, *δίνεται στην ακόλουθη σχέση*

$$P(M_0=m_0 \text{ και } M_j=m_j \text{ για } a+1\le j\le b)=$$

$$=\iint\limits_{0\le u_a\le u_{a+1}\le...\le u_b\le 1}\!\!...\!\int q(GF^{-1}(u_a),GF^{-1}(u_{a+1}),...,GF^{-1}(u_b))\, f_{a:b}(\mathbf{u}_{a:b})du_a du_{a+1}\cdots du_b \tag{5.10}$$

όπου $f_{a:b}$ *και q είναι οι συναρτήσεις που έχουν ορισθεί στις* (5.7) *και* (5.9) *αντίστοιχα.*

Αξίζει να σημειωθεί ότι αν θέσουμε $F=G$ η παραπάνω έκφραση παίρνει τη μορφή

$$P(M_0=m_0 \text{ και } M_j=m_j \text{ για } a+1\le j\le b)=$$

$$= \int\int\dots\int_{0\le u_a \le u_{a+1}\le \dots \le u_b \le 1} q(u_a, u_{a+1},\dots,u_b)\ f_{a:b}(u_a,u_{a+1},\dots,u_b)du_a du_{a+1}\cdots du_b$$

και ύστερα από αλγεβρικές πράξεις μπορούμε να αποδείξουμε με διαφορετικό τρόπο το αποτέλεσμα της Πρότασης 5.1. Συνεπώς, η εντός ελέγχου κατανομή των $(M_0, M_{a+1},\dots,M_b)$ μπορεί να θεωρηθεί ως μια ειδική περίπτωση του γενικότερου αποτελέσματος που παρουσιάζεται στην Πρόταση 5.2.

Επιπρόσθετα, αν η διεργασία μετατοπιστεί σε εκτός ελέγχου κατάσταση, η πιθανότητα το διάγραμμα να μην παράγει σήμα για εκτός ελέγχου μετατόπιση, εξαρτάται τόσο από την εντός ελέγχου, όσο και από την εκτός ελέγχου κατανομή ($F(x)$ και $G(x)$ αντίστοιχα). Δεδομένου όμως ότι το τελικό αποτέλεσμα εξαρτάται από τις F και G, μόνο μέσω της συνάρτησης $G\circ F^{-1}$, θα μπορούσαμε εύκολα να λάβουμε ακριβείς εκφράσεις για την εκτός ελέγχου πιθανότητα συναγερμού στην περίπτωση που έχουμε εναλλακτικές τύπου *Lehmann*. Η ακόλουθη Πρόταση παρουσιάζει την από κοινού κατανομή των $(M_0, M_{a+1},\dots,M_b)$ στην περίπτωση των εναλλακτικών τύπου *Lehmann* και μπορεί να χρησιμοποιηθεί για τη μελέτη της χαρακτηριστικής καμπύλης όλων των κανόνων που παρουσιάζονται στο Κεφάλαιο 5 του παρόντος κειμένου.

Πρόταση 5.3. *Η από κοινού συνάρτηση πιθανότητας των* $(M_0, M_{a+1},\dots,M_b)$ *υπό την εναλλακτική υπόθεση τύπου Lehmann* $G(x)=(F(x))^{\gamma}$ *δίνεται στην ακόλουθη σχέση*

$$P(M_0 = m_0 \text{ και } M_j = m_j \text{ για } a+1\le j\le b) =$$

$$= c_1 c_2 \gamma^{-(b-a+1)} \prod_{j=a}^{b-1} B\Big(\frac{j}{\gamma} + m_0 + \sum_{i=a+1}^{j} m_i, m_{j+1}+1\Big)\times$$

$$\times \sum_{\ell=0}^{m-b}(-1)^{\ell}\binom{m-b}{j} B\Big(\frac{b+\ell}{\gamma} + m_0 + \sum_{i=a+1}^{b} m_i, n-m_0-\sum_{i=a+1}^{b} m_i + 1\Big)$$

όπου

$$c_1 = \frac{m!}{(a-1)!(m-b)!}, \qquad c_2 = \frac{n!}{m_0!(\prod_{j=a+1}^{b} m_j!)(n-m_0-\sum_{i=a+1}^{b} m_j)!}.$$

Απόδειξη. Υπό την εναλλακτική τύπου *Lehmann* $G(x)=(F(x))^{\gamma}$, είναι σαφές ότι $GF^{-1}(x) = G(F^{-1}(x)) = x^{\gamma}$ και με άμεση αντικατάσταση στην (5.10) προκύπτει ότι

$$P(M_0 = m_0 \text{ και } M_j = m_j \text{ για } a+1\le j\le b) =$$

$$= \int\int\dots\int_{0\le u_a \le u_{a+1}\le \dots \le u_b \le 1} q(u_a^{\gamma}, u_{a+1}^{\gamma},\dots,u_b^{\gamma})\ f_{a:b}(u_a,u_{a+1},\dots,u_b)du_a du_{a+1}\cdots du_b .$$

Με τη βοήθεια των (5.7) και (5.9) παίρνουμε την ακόλουθη σχέση

$$P(M_0 = m_0 \text{ και} M_j = m_j \text{ για } a+1 \le j \le b) = c_1 c_2 \sum_{\ell=0}^{m-b} \binom{m-b}{\ell} (-1)^{\ell} \times$$

$$\times \iint\limits_{0 \le u_a \le u_{a+1} \le \ldots \le u_b \le 1} \!\!\ldots\!\int u_a^{a+m_0\gamma-1} \prod_{j=a+1}^{b} (u_j^{\gamma} - u_{j-1}^{\gamma})^{m_j} u_b^{\ell} (1-u_b^{\gamma})^{n-m_0-\sum_{i=a+1}^{b} m_i} du_a du_{a+1} \cdots du_b .$$

Εφαρμόζοντας το μετασχηματισμό $t = (u_a / u_{a+1})^{\gamma}$ στο εσωτερικό ολοκλήρωμα προκύπτει ότι

$$P(M_0 = m_0 \text{ και} M_j = m_j \text{ για } a+1 \le j \le b) = c_1 c_2 \sum_{\ell=0}^{m-b} \binom{m-b}{\ell} (-1)^{\ell} \times$$

$$\times \iint\limits_{0 \le u_a \le u_{a+1} \le \ldots \le u_b \le 1} \!\!\ldots\!\int \gamma^{-1} B(\frac{a}{\gamma} + m_0, m_{a+1}+1) \prod_{j=a+2}^{b} (u_j^{\gamma} - u_{j-1}^{\gamma})^{m_j} u_b^{\ell} (1-u_b^{\gamma})^{n-m_0-\sum_{i=a+1}^{b} m_i} du_{a+1} \cdots du_b .$$

Στη συνέχεια, χρησιμοποιώντας διαδοχικά τους μετασχηματισμούς $t = (u_j / u_{j+1})^{\gamma}$, $j = a+1, a+2, \ldots, b-1$ καταλήγουμε στην ακόλουθη έκφραση

$$P(M_0 = m_0 \text{ και} M_j = m_j \text{ για } a+1 \le j \le b) =$$

$$= \frac{m!n!}{(a-1)!(m-b)!m_0!(\prod_{j=a+1}^{b} m_j!)(n-m_0-\sum_{j=a+1}^{b} m_j)!} \gamma^{-(b-a+1)} \prod_{j=a}^{b-1} B(\frac{j}{\gamma} + m_0 + \sum_{i=a+1}^{j} m_i, m_{j+1}+1) \times$$

$$\times \sum_{\ell=0}^{m-b} (-1)^{\ell} \binom{m-b}{j} B(\frac{b+\ell}{\gamma} + m_0 + \sum_{i=a+1}^{b} m_i, n - m_0 - \sum_{i=a+1}^{b} m_i + 1)$$

και η απόδειξη ολοκληρώθηκε. □

5.5. Μελέτη της κατανομής του μήκους ροής των νέων διαγραμμάτων ελέγχου

Στα νέα μη παραμετρικά διαγράμματα ελέγχου τύπου *Shewhart* που μελετώνται στο Κεφάλαιο 5, η κατανομή του μήκους ροής τους δεν είναι γεωμετρική, διότι τα συμβάντα παραγωγής σήματος για εκτός ελέγχου μετατόπιση της διεργασίας, δεν είναι ανεξάρτητα (βλ. επίσης Κεφ.4). Ωστόσο, κάνοντας χρήση της τεχνικής δέσμευσης του Chakraborti (2000), μπορούμε να καταλήξουμε σε μια ακριβή μορφή της κατανομής του μήκους ροής.

Με τη βοήθεια των (5.4) και (5.5) είναι σαφές ότι, υπό τη συνθήκη $\mathbf{X}_{a:b} = (x_a, x_{a+1}, \ldots, x_b) = \mathbf{x}_{a:b} \in \Re^{b-a+1}$, η τυχαία μεταβλητή N, η οποία περιγράφει το πλήθος των Y–δειγμάτων μέχρι να παραχθεί σήμα συναγερμού για εκτός ελέγχου μετατόπιση, ακολουθεί γεωμετρική κατανομή με πιθανότητα επιτυχίας

$$p(\mathbf{x}_{a:b}) = \sum_{(m_0, m_{a+1} \ldots, m_b) \in A} P(M_0(\mathbf{Y}; \mathbf{x}_{a:b}) = m_0 \text{ και } M_j(\mathbf{Y}; \mathbf{x}_{a:b}) = m_j \text{ για } a+1 \le j \le b)$$

και συνάρτηση πιθανότητας

$$(1 - p(\mathbf{x}_{a:b}))(p(\mathbf{x}_{a:b}))^{k-1} = (p(\mathbf{x}_{a:b}))^{k-1} - (p(\mathbf{x}_{a:b}))^k, \quad k = 1,2,\ldots.$$

Συνεπώς, η μη δεσμευμένη κατανομή της N μπορεί να εκφρασθεί ως εξής

$$P(N = k) = D(k-1) - D(k), \tag{5.11}$$

όπου $D(0) = 1$ και

$$D(k) = E_{\mathbf{x}_{a:b}}\{(p(\mathbf{X}_{a:b}))^k\}, \quad k = 1,2,\ldots.$$

Μετασχηματίζοντας τις τυχαίες μεταβλητές $\mathbf{X}_{a:b}$ με τη βοήθεια της (5.6), παίρνουμε ότι

$$D(k) = E_{\mathbf{U}_{a:b}}\{(p(\mathbf{U}_{a:b}))^k\} = \iint\limits_{0 \le u_a \le u_{a+1} \le \ldots \le u_b \le 1} \cdots \int (p(\mathbf{u}_{a:b}))^k f_{a:b}(u_{a:b}) du_a du_{a+1} \cdots du_b,$$

όπου $f_{a:b}$ είναι η από κοινού συνάρτηση πιθανότητας, όπως αυτή ορίζεται στην (5.7), των διατεταγμένων παρατηρήσεων $(U_a, U_{a+1}, \ldots, U_b)$ και

$$p(\mathbf{u}_{a:b}) = \sum_{(m_0, m_{a+1} \ldots, m_b) \in A} P(M_0(\mathbf{Y}; \mathbf{u}_{a:b}) = m_0 \text{ και } M_j(\mathbf{Y}; \mathbf{u}_{a:b}) = m_j \text{ για } a+1 \le j \le b).$$

Ακολουθώντας τα βήματα της απόδειξης της Πρότασης 5.2, προκύπτει ότι

$$p(\mathbf{u}_{a:b}) = \sum_{(m_0, m_{a+1} \ldots, m_b) \in A} q(GF^{-1}(u_a), GF^{-1}(u_{a+1}), \ldots, GF^{-1}(u_b)),$$

όπου q είναι η συνάρτηση που ορίζεται στην (5.9).

Από την ανάλυση που προηγήθηκε, καταλήγουμε στο ακόλουθο γενικό αποτέλεσμα που ισχύει για την κατανομή του μήκους ροής και τη μέση τιμή ενός διαγράμματος ελέγχου.

Πρόταση 5.4. *Έστω ότι N είναι το μήκος ροής για ένα διάγραμμα ελέγχου που παράγει σήμα για εκτός ελέγχου μετατόπιση, όταν το τυχαίο διάνυσμα $(M_0, M_{a+1}, \ldots, M_b)$ λαμβάνει τιμές έξω από το υποσύνολο $A \subseteq \{0,1,2,\ldots\}^{b-a}$. Αν*

$$Q(v_a, v_{a+1}, \ldots, v_b) = \sum_{(m_0, m_{a+1}, \ldots, m_b) \in A} q(v_a, v_{a+1}, \ldots, v_b), \quad 0 \le v_a \le v_{a+1} \le \ldots \le v_b \le 1$$

τότε

α. η συνάρτηση πιθανότητας της N δίνεται από τον τύπο $P(N = k) = D(k-1) - D(k)$, $k = 1,2,\ldots$, όπου $D(0) = 1$ και

$$D(k) = \iint\limits_{0 \le u_a \le u_{a+1} \le \ldots \le u_b \le 1} \cdots \int (Q^k(GF^{-1}(u_a), GF^{-1}(u_{a+1}), \ldots, GF^{-1}(u_b)) \times$$

$$\times f_{a:b}(u_a, u_{a+1}, ..., u_b) du_a du_{a+1} \cdots du_b.$$

β. το μέσο μήκος ροής $ARL = E(N)$ μπορεί να γραφεί στη μορφή

$$ARL = \sum_{k=0}^{\infty} D(k) = \iint\limits_{0 \le u_a \le u_{a+1} \le ... \le u_b \le 1} \!\!...\!\int \frac{1}{1 - Q(GF^{-1}(u_a), GF^{-1}(u_{a+1}), ..., GF^{-1}(u_b))} \times$$

$$\times f_{a:b}(u_a, u_{a+1}, ..., u_b) du_a du_{a+1} \cdots du_b.$$

Απόδειξη. Το μέρος (α) έχει ήδη αποδειχθεί με βάση την ανάλυση που αναπτύχθηκε πριν τη διατύπωση της Πρότασης 5.4. Το μέρος (β) προκύπτει άμεσα με τη βοήθεια της ακόλουθης σχέσης

$$E(N) = \sum_{k=0}^{\infty} P(N > k) = E_{U_{a:b}}(\sum_{k=0}^{\infty} p^k(\mathbf{U}_{a:b})) = E_{U_{a:b}}(\frac{1}{1 - p(\mathbf{U}_{a:b})}). \qquad \square$$

Η εντός ελέγχου κατανομή και το αντίστοιχο μέσο μήκος ροής (ARL_{in}) προκύπτουν θέτοντας $G = F$. Έτσι για τη δεύτερη ποσότητα μπορούμε να γράψουμε

$$ARL_{in} = \iint\limits_{0 \le u_a \le u_{a+1} \le ... \le u_b \le 1} \!\!...\!\int \frac{1}{1 - Q(u_a, u_{a+1}, ..., u_b)} f_{a:b}(u_a, u_{a+1}, ..., u_b) du_a du_{a+1} \cdots du_b. \quad (5.12)$$

Το πολλαπλό ολοκλήρωμα μπορεί να υπολογισθεί αριθμητικά για κάθε επιλογή των παραμέτρων σχεδιασμού (δηλαδή των *a*, *b* και των παραμέτρων που προσδιορίζουν το σύνολο *A*). Πρόσθετα, μπορούμε να κάνουμε το ίδιο σχόλιο για το εκτός ελέγχου μέσο μήκος ροής (ARL_{out}), υπό την προϋπόθεση ότι έχουν επιλεγεί συγκεκριμένες εντός και εκτός ελέγχου κατανομές F και G αντίστοιχα. Αξίζει να αναφερθεί ότι, όταν η εκτός ελέγχου κατανομή είναι τύπου *Lehmann*, δηλαδή είναι της μορφής $G(x) = (F(x))^{\gamma}$, δεν χρειαζόμαστε τη συναρτησιακή μορφή της $F(x)$, αφού $(GF^{-1}(x)) = x^{\gamma}$ και συνεπώς

$$ARL_{out} = \iint\limits_{0 \le u_a \le u_{a+1} \le ... \le u_b \le 1} \!\!...\!\int \frac{1}{1 - Q(u_a^{\gamma}, u_{a+1}^{\gamma}, ..., u_b^{\gamma})} f_{a:b}(u_a, u_{a+1}, ..., u_b) du_a du_{a+1} \cdots du_b.$$

5.6. Αριθμητικά αποτελέσματα για τα νέα διαγράμματα ελέγχου

Η απόδοση ενός διαγράμματος ελέγχου κρίνεται συνήθως με τη βοήθεια της κατανομής του μήκους ροής και την ικανότητα ανίχνευσης μιας πιθανής μετατόπισης της διεργασίας. Στη συγκεκριμένη Παράγραφο, υλοποιούμε αριθμητικούς υπολογισμούς για τα προτεινόμενα διαγράμματα που αποδεικνύουν την αποτελεσματικότητα και την αντοχή τους, τόσο σε εντός όσο και σε εκτός ελέγχου καταστάσεις. Στον Πίνακα 5.3, παρουσιάζουμε τις πιθανότητες συναγερμού (*alarm rates*, *AR*) των τριών διαγραμμάτων ελέγχου για διαφορετικές επιλογές των m, n. Οι παράμετροι σχεδιασμού a, b, r_0, r, k, w καθορίζονται κατάλληλα ώστε οι κανόνες να επιτυγχάνουν μια συγκεκριμένη πιθανότητα λανθασμένου συναγερμού (*false alarm rate*) $5‰$ $(FAR = 0.005)$. Στη συνέχεια, υπολογίζονται, με τη βοήθεια της Πρότασης 5.3, οι τιμές του *AR* για $\gamma = 1/2, 1/4, 1/8$.

Πίνακας 5.3. Σύγκριση των *AR* των τριών διαγραμμάτων ελέγχου με *FAR*=0.005 ($\gamma = 1/2, 1/4$ και $1/8$)

		R–διάγραμμα					*N_2– διάγραμμα*					*W- διάγραμμα*				
m	*n*	(*LCL*,*UCL*)	r_0	*r*	*Ακριβές FAR*	*AR*	(*LCL*,*UCL*)	r_0	r_1	*Ακριβές FAR*	*AR*	(*LCL*,*UCL*)	*w*	r_0	*Ακριβές FAR*	*AR*
100	5	(7, 10)	2	2	0.0043	**0.1201**	(7, 10)	2	1	0.0041	**0.1193**	(7, 10)	70	2	0.0041	**0.1193**
						0.5130					**0.5123**					**0.5123**
						0.8486					**0.8483**					**0.8483**
	11	(10, 13)	4	3	0.0049	**0.2461**	(10, 13)	4	1	0.0047	**0.2457**	(10, 13)	95	4	0.0048	**0.2475**
						0.8279					**0.8277**					**0.8288**
						0.9901					**0.9901**					**0.9902**
	25	(9, 12)	7	4	0.0045	**0.4668**	(9, 12)	7	1	0.0050	**0.4680**	(9, 12)	123	7	0.0049	**0.4753**
						0.9872					**0.9872**					**0.9879**
						0.9999					**0.9999**					**0.9999**
200	5	(16, 19)	2	2	0.0050	**0.1409**	(16, 19)	2	1	0.0050	**0.1408**	(15, 18)	65	2	0.0046	**0.1326**
						0.5526					**0.5526**					**0.5399**
						0.8689					**0.8689**					**0.8631**
	11	(21, 24)	4	3	0.0045	**0.2647**	(21, 24)	4	2	0.0044	**0.2646**	(21, 24)	135	4	0.0048	**0.2669**
						0.8488					**0.8488**					**0.8497**
						0.9922					**0.9922**					**0.9923**
	25	(16, 19)	6	5	0.0049	**0.5678**	(16, 19)	6	2	0.0049	**0.5677**	(16, 19)	190	6	0.0050	**0.5688**
						0.9952					**0.9952**					**0.9953**
						0.9999					**0.9999**					**0.9999**
500	5	(39, 42)	2	1	0.0047	**0.1374**	(39, 42)	2	2	0.0044	**0.1367**	(39, 42)	100	2	0.0049	**0.1381**
						0.5510					**0.5504**					**0.5515**
						0.8690					**0.8689**					**0.8692**
	11	(33, 36)	3	2	0.0048	**0.3042**	(33, 36)	3	2	0.0048	**0.3041**	(33, 36)	3	2	0.0049	**0.3045**
						0.8911					**0.8911**					**0.8912**
						0.9965					**0.9965**					**0.9965**
	25	(21, 24)	4	4	0.0045	**0.5907**	(21, 24)	4	2	0.0045	**0.5907**	(21, 24)	150	4	0.0046	**0.5915**
						0.9971					**0.9971**					**0.9971**
						0.9999					**0.9999**					**0.9999**

Από τον Πίνακα 5.3, είναι σαφές ότι τα τρία μη παραμετρικά διαγράμματα ελέγχου που μελετώνται στο παρόν Κεφάλαιο, παρουσιάζουν παρόμοια απόδοση, αφού, για την ίδια *FAR* (σχεδόν), πετυχαίνουν σχεδόν ίσες *AR* τιμές. Για το λόγο αυτό, οι υπολογισμοί που ακολουθούν, θα γίνουν μόνο σε ένα από αυτά (το $W-$διάγραμμα), μιας και έχουμε ισχυρές ενδείξεις ότι τα άλλα δύο διαγράμματα, με κατάλληλες επιλογές των παραμέτρων σχεδιασμού, θα δώσουν ισοδύναμα αποτελέσματα. Στον Πίνακα 5.4, παρουσιάζουμε τις πιθανότητες *FAR* του *W*–διαγράμματος για διάφορους σχεδιασμούς και επιλογές των παραμέτρων m, n, a, b, w, r_0. Οι υπολογισμοί πραγματοποιούνται με τη βοήθεια της Πρότασης 5.1.

Πίνακας 5.4. Πιθανότητες λανθασμένου συναγερμού για δεδομένο σχεδιασμό

		Δείγμα αναφοράς μεγέθους *m*											
n	r_0	40			60			100			200		
		(a,b)	w	FAR	(a,b)	w	FAR	(a,b)	w	FAR	(a,b)	w	FAR
5	2	(7,11)	200	0.0471	(10,13)	45	0.0466	(15,19)	150	0.0292	(20,24)	80	0.0140
			40	0.0597		30	0.0669		50	0.0464		42	0.0453
			33	0.0900		23	0.0747		33	0.0696		40	0.0890
	3	(13,17)	90	0.0499	(20,23)	150	0.0492	(15,19)	50	0.0201	(20,24)	43	0.0143
			63	0.0601		80	0.0761		33	0.0451		42	0.0369
			58	0.0923		44	0.0832		32	0.0857		27	0.0953
	4	(16,20)	100	0.0249	(25,28)	55	0.0439	(25,29)	60	0.0185	(30,34)	63	0.0185
			72	0.0430		54	0.0581		54	0.0323		62	0.0410
			65	0.0987		53	0.0943		53	0.0604		27	0.0948
11	4	(7,12)	170	0.0443	(11,15)	140	0.0469	(15,19)	78	0.0331	(20,24)	85	0.0164
			100	0.0506		90	0.0556		68	0.0499		44	0.0416
			66	0.0978		55	0.0987		53	0.0960		43	0.0693
	6	(10,15)	170	0.0152	(11,15)	100	0.0104	(18,22)	90	0.0143	(25,29)	55	0.0301
			1022	0.0471		58	0.0486		79	0.0427		54	0.0458
			79	0.0929		52	0.0913		42	0.0873		53	0.0770
	8	(15,20)	143	0.0472	(17,21)	130	0.0101	(21,25)	100	0.0130	(26,30)	70	0.0227
			115	0.0749		84	0.0477		91	0.0457		56	0.0471
			111	0.0961		78	0.0906		48	0.0916		55	0.0789
25	8	(5,8)	85	0.0496	(10,13)	250	0.0337	(15,18)	155	0.0198	(19,22)	120	0.0104
			70	0.0834		130	0.0481		105	0.0498		81	0.0464
			66	0.0966		85	0.0998		73	0.0906		43	0.0877
	14	(8,12)	150	0.0428	(15,18)	200	0.0115	(28,31)	240	0.0166	(30,33)	130	0.0330
			140	0.0567		146	0.0486		187	0.0471		127	0.0470
			120	0.0972		114	0.0963		130	0.0960		67	0.0821
	16	(12,16)	210	0.0444	(19,22)	210	0.0234	(32,35)	240	0.0155	(33,36)	160	0.0105
			190	0.0665		185	0.0469		215	0.0431		140	0.0468
			170	0.0998		143	0.0961		148	0.0946		73	0.0902

Χρησιμοποιώντας τον Πίνακα 5.4, μπορούμε να κατασκευάσουμε διαγράμματα ελέγχου ελεύθερα κατανομής που πετυχαίνουν μια προκαθορισμένη πιθανότητα λανθασμένου συναγερμού f. Η χρήση έξι παραμέτρων m, n, a, b, w, r_0 στο διάγραμμα προσφέρει την ευελιξία να σταθεροποιήσουμε ορισμένες από αυτές και στη συνέχεια να αναζητήσουμε τη βέλτιστη επιλογή για τις άλλες, ή εναλλακτικά για έναν αποδεκτό συνδυασμό τους που ικανοποιεί ορισμένες απαιτήσεις. Για παράδειγμα, αν έχουμε συλλέξει δείγμα αναφοράς μεγέθους $m = 200$ και επιθυμούμε να εξετάσουμε τυχαία δείγματα μεγέθους $n = 25$ που εξάγονται από τη διεργασία, τότε η πιθανότητα λανθασμένου συναγερμού ίση με $f = 0.05$ (σχεδόν) μπορεί να επιτευχθεί χρησιμοποιώντας

- την 19$^{\eta}$ και 22$^{\eta}$ μικρότερη παρατήρηση στο εξεταζόμενο δείγμα $(a = 19, b = 22)$, $w = 81$ και $r_0 = 8$ (με *FAR*=0.0464) ή

- την $30^{η}$ και $33^{η}$ μικρότερη παρατήρηση στο εξεταζόμενο δείγμα $(a = 30, b = 33)$, $w = 127$ και $r_0 = 14$ (με *FAR*=0.0470) ή
- την $33^{η}$ και $36^{η}$ μικρότερη παρατήρηση στο εξεταζόμενο δείγμα $(a = 33, b = 36)$, $w = 140$ και $r_0 = 16$ (με *FAR*=0.0468).

Πίνακας 5.5. Πιθανότητες συναγερμού (*AR*) για δεδομένο σχεδιασμό. Κάθε κελί περιέχει τις τιμές *AR* για γ =1/3 (άνω είσοδος) και γ =1/5 (κάτω είσοδος)

		Δείγμα αναφοράς μεγέθους *m*											
		40			60			100			200		
n	r_0	(*a*,*b*)	*w*	*AR*	(*a*,*b*)	*w*	*AR*	(*a*,*b*)	*w*	*AR*	(*a*,*b*)	*w*	*AR*
5	2	(7,11)	200	0.5858 0.8251	(10,13)	45	0.5910 0.8296	(15,19)	150	0.5484 0.8067	(20,24)	80	0.4367 0.7333
			40	0.5967 0.8300		30	0.6011 0.8333		50	0.5661 0.8152		42	0.4907 0.7643
			33	0.6340 0.8491		23	0.6228 0.8452		33	0.6054 0.8354		40	0.5259 0.7804
	3	(13,17)	90	0.4886 0.7189	(20,23)	150	0.5037 0.7320	(15,19)	50	0.2474 0.5012	(20,24)	43	0.2071 0.4433
			63	0.4937 0.7217		80	0.5168 0.7386		33	0.3253 0.5641		42	0.2373 0.4658
			58	0.5211 0.7373		44	0.5352 0.7515		32	0.3660 0.5921		27	0.2752 0.4824
	4	(16,20)	100	0.2156 0.3912	(25,28)	55	0.2529 0.4295	(25,29)	60	0.1108 0.2537	(30,34)	63	0.1028 0.2079
			72	0.2319 0.4007		54	0.2837 0.4552		54	0.1643 0.3062		62	0.1262 0.2236
			65	0.2468 0.4067		53	0.3233 0.4844		53	0.1981 0.3323		27	0.1503 0.2331
11	4	(7,12)	170	0.8018 0.9690	(11,15)	140	0.8309 0.9764	(15,19)	78	0.7820 0.9670	(20,24)	85	0.6539 0.9341
			100	0.8088 0.9706		90	0.8351 0.9771		68	0.8059 0.9716		44	0.7107 0.9477
			66	0.8301 0.9745		55	0.8616 0.9818		53	0.8107 0.9721		43	0.7362 0.9532
	6	(10,15)	170	0.5838 0.8739	(11,15)	100	0.4374 0.7967	(18,22)	90	0.4394 0.8021	(25,29)	55	0.4097 0.7605
			102	0.6049 0.8803		58	0.4998 0.8260		79	0.4864 0.8198		54	0.4403 0.7698
			79	0.6477 0.8955		52	0.5514 0.8416		42	0.5717 0.8584		53	0.4589 0.7732
	8	(15,20)	143	0.3686 0.6687	(17,21)	130	0.2187 0.5307	(21,25)	100	0.1411 0.4167	(26,30)	70	0.0715 0.2504
			115	0.3959 0.6830		84	0.2712 0.5660		91	0.1762 0.4291		56	0.2419 0.3954
			111	0.4057 0.6855		78	0.3109 0.5789		48	0.3433 0.5689		55	0.2595 0.3985
25	8	(5,8)	85	0.9181 0.9978	(10,13)	250	0.9636 0.9995	(15,18)	155	0.9602 0.9995	(19,22)	120	0.8681 0.9972
			70	0.9336 0.9984		130	0.9672 0.9996		105	0.9649 0.9996		81	0.8868 0.9977
			66	0.9347 0.9984		85	0.9756 0.9997		73	0.9730 0.9997		43	0.9301 0.9987
	14	(8,12)	150	0.5543 0.9212	(15,18)	200	0.6696 0.9630	(28,31)	240	0.7627 0.9810	(30,33)	130	0.3777 0.8703
			140	0.5636 0.9219		146	0.6848 0.9650		187	0.7736 0.9820		127	0.3777 0.8703
			120	0.6135 0.9347		114	0.7200 0.9693		130	0.8066 0.9848		67	0.5636 0.9110
	16	(12,16)	210	0.3975 0.8187	(19,22)	210	0.5866 0.9283	(32,35)	240	0.6028 0.9385	(33,36)	160	0.1742 0.6775
			190	0.4130 0.8210		185	0.5890 0.9284		215	0.6090 0.9390		140	0.2026 0.6803
			170	0.4326 0.8311		143	0.6240 0.9348		148	0.6562 0.9466		73	0.4241 0.7692

Από τον Πίνακα 5.5 είναι δυνατόν να εξαχθούν ενδιαφέρουσες παρατηρήσεις σχετικά με την απόδοση του προτεινόμενου διαγράμματος για την περίπτωση των εναλλακτικών τύπου *Lehmann*. Στο συγκεκριμένο Πίνακα, υπολογίζονται οι πιθανότητες συναγερμού για διάφορες τιμές των παραμέτρων σχεδιασμού m, n, a, b, w, r_0 και της παραμέτρου μετατόπισης $\gamma > 0$. Στην πραγματικότητα, οι σχεδιασμοί που παρουσιάζονται στον Πίνακα 5.4, ελέγχονται για την ικανότητα τους να ανιχνεύουν μετατοπίσεις από την εντός ελέγχου κατανομή.

Σύμφωνα με τα αποτελέσματα του Πίνακα 5.5, για την περίπτωση $m = 200$, $n = 25$ που περιγράφηκε νωρίτερα, ο σχεδιασμός που αντιστοιχεί στις τιμές $a = 19, b = 22, w = 81, r_0 = 8$ (από τους τρεις εναλλακτικούς σχεδιασμούς που πετυχαίνουν πιθανότητα λανθασμένου συναγερμού ίση (σχεδόν) με 5%), παρουσιάζει την μεγαλύτερη πιθανότητα συναγερμού για ανίχνευση μιας μετατόπισης σε εναλλακτικές τύπου *Lehmann* με $\gamma = 1/3$ $(AR = 0.8868)$ και $\gamma = 1/5$ $(AR = 0.9977)$.

Στον Πίνακα 5.6, συγκρίνουμε την απόδοση του $W-$διαγράμματος και του νέου μη παραμετρικού διαγράμματος που εισήχθηκε στο Κεφάλαιο 4. Πιο συγκεκριμένα, χρησιμοποιώντας τρία διαφορετικά επίπεδα *FAR* και διάφορες τιμές για τις παράμετρους m, n, προσδιορίζουμε με τη βοήθεια του Πορίσματος 4.1 και της Πρότασης 5.1, τις παραμέτρους σχεδιασμού των διαγραμμάτων έτσι ώστε να επιτυγχάνεται πιθανότητα *FAR* ίση με την τιμή στόχο που έχουμε επιλέξει σε κάθε περίπτωση. Στη συνέχεια, υπολογίζουμε αριθμητικά τις *AR* τιμές (με τη βοήθεια της Πρότασης 5.3) για δύο συγκεκριμένες εναλλακτικές τύπου *Lehmann* $(\gamma = 0.4$ και $\gamma = 0.2)$. Τα αποτελέσματα αυτά συνοψίζονται στο τμήμα του Πίνακα 5.6 με επικεφαλίδα «$W-$*διάγραμμα*». Τα αντίστοιχα αποτελέσματα για το νέο μη παραμετρικό διάγραμμα που εισήχθηκε στο Κεφάλαιο 4 δίνονται στο τμήμα με επικεφαλίδα «*Διάγραμμα Κεφ. 4*».

Πίνακας 5.6. Σύγκριση των *AR* των διαγραμμάτων ελέγχου με την ίδια *FAR* ($\gamma = 0.4$ και 0.2)

FAR	*m*	*n*	*W-διάγραμμα*					*Διάγραμμα Κεφ.4* ($j=(n+1)/2$)			
			(*LCL,UCL*)	*w*	r_0	*Ακριβές FAR*	*AR*	(*LCL,UCL*)	*r*	*Ακριβές FAR*	*AR*
0.01	50	5	(4,7)	31	2	0.0082	**0.2538** **0.6660**	(3,48)	1	0.0072	**0.1888** **0.5956**
		11		41	4	0.0097	**0.3848** **0.8853**	(7,44)	1	0.0093	**0.3651** **0.8638**
		25		90	7	0.0096	**0.7018** **0.9974**	(6,45)	13	0.0097	**0.3886** **0.9388**
	100	5	(10,13)	65	2	0.0100	**0.3089** **0.9957**	(7,94)	1	0.0086	**0.2230** **0.6455**
		11		66	4	0.0095	**0.4620** **0.9249**	(6,95)	7	0.0091	**0.3179** **0.8539**
		25		115	7	0.0096	**0.8113** **0.9990**	(8,93)	16	0.0082	**0.5294** **0.9848**
	500	5	(50,53)	120	2	0.0094	**0.3139** **0.7327**	(38,43)	2	0.0091	**0.2472** **0.6760**
		11		104	4	0.0073	**0.4847** **0.9398**	(33,68)	7	0.0099	**0.3663** **0.8894**
		25		207	7	0.0091	**0.8401** **0.9995**	(120,81)	7	0.0095	**0.7421** **0.9965**
	1000	5	(100,103)	203	2	0.0091	**0.3164** **0.7354**	(79,92)	2	0.0099	**0.2568** **0.6856**
		11		203	4	0.0072	**0.4828** **0.9369**	(96,95)	6	0.0098	**0.3188** **0.8423**
		25		206	7	0.0086	**0.8542** **0.9996**	(243,78)	7	0.0099	**0.7545** **0.9969**
0.005	50	5	(3,6)	45	2	0.0036	**0.1885** **0.5956**	(2,49)	1	0.0030	**0.1273** **0.5053**
		11		39	4	0.0036	**0.2878** **0.8287**	(5,46)	2	0.0025	**0.2377** **0.7804**
		25		80	7	0.0049	**0.5868** **0.9896**	(7,44)	11	0.0048	**0.3547** **0.9410**
	100	5	(7,10)	40	2	0.0046	**0.2254** **0.6471**	(5,96)	1	0.0035	**0.1618** **0.5730**
		11		49	4	0.0048	**0.3348** **0.8729**	(5,96)	7	0.0046	**0.2538** **0.8141**
		25		83	7	0.0049	**0.6661** **0.9962**	(15,86)	11	0.0045	**0.4112** **0.9650**
	500	5	(40,43)	165	2	0.0048	**0.2569** **0.6861**	(20,481)	3	0.0048	**0.1197** **0.5581**
		11		84	4	0.0046	**0.3973** **0.9100**	(56,454)	5	0.0049	**0.3044** **0.8462**
		25		168	7	0.0048	**0.7460** **0.9988**	(90,411)	10	0.0050	**0.5391** **0.9865**
	1000	5	(82,85)	350	2	0.0050	**0.2627** **0.6922**	(62,99)	2	0.0049	**0.2043** **0.6347**
		11		167	4	0.0044	**0.4040** **0.9131**	(136,865)	4	0.0049	**0.3715** **0.8839**
		25		171	7	0.0044	**0.7665** **0.9990**	(129,872)	13	0.0050	**0.4760** **0.9766**
0.0027	50	5	(2,5)	19	2	0.0026	**0.1404** **0.5207**	(1,50)	1	0.0008	**0.0634** **0.3617**
		11	(3,6)	43	4	0.0027	**0.2788** **0.8264**	(5,46)	2	0.0025	**0.2377** **0.7804**
		25		100	7	0.0025	**0.5514** **0.9873**	(7,44)	10	0.0021	**0.3309** **0.9381**
	100	5	(5,8)	26	2	0.0026	**0.1730** **0.5854**	(4,97)	2	0.0021	**0.1304** **0.5262**
		11		39	3	0.0027	**0.2437** **0.8134**	(9,92)	5	0.0023	**0.2185** **0.7756**
		25		73	6	0.0026	**0.6461** **0.9960**	(12,89)	12	0.0019	**0.3028** **0.9364**
	500	5	(32,35)	134	2	0.0026	**0.2089** **0.6395**	(16,485)	3	0.0026	**0.1197** **0.5116**
		11	(38,41)	155	4	0.0025	**0.3534** **0.8952**	(64,437)	2	0.0026	**0.3386** **0.8693**
		25		162	7	0.0026	**0.7216** **0.9985**	(101,400)	8	0.0026	**0.6078** **0.9915**
	1000	5	(66,69)	330	2	0.0027	**0.2151** **0.6470**	(49,952)	2	0.0025	**0.1626** **0.5862**
		11	(64,67)	131	4	0.0027	**0.3186** **0.8778**	(49,952)	7	0.0027	**0.2536** **0.8355**
		25	(66,69)	268	7	0.0026	**0.6456** **0.9975**	(189,812)	9	0.0027	**0.5626** **0.9891**

Σύμφωνα με τον Πίνακα 5.6, για την ίδια τιμή FAR, το νέο διάγραμμα έχει καλύτερη απόδοση από το διάγραμμα που ορίζεται στην (4.3), αφού παρουσιάζει μεγαλύτερη πιθανότητα AR για όλες τις περιπτώσεις που εξετάσθηκαν. Για παράδειγμα, χρησιμοποιώντας δείγμα αναφοράς μεγέθους $m = 500$, εξεταζόμενα δείγματα μεγέθους $n = 5$ και προκαθορισμένο επίπεδο $FAR = 0.0027$, το διάγραμμα που ορίζεται στην (4.3) παρουσιάζει για $\gamma = 0.4$ ($\gamma = 0.2$), πιθανότητες συναγερμού ίσες με 0.1197 (0.5116), ενώ οι αντίστοιχες τιμές για το W–chart είναι 0.2089 (0.6395).

Πίνακας 5.7. Σύγκριση των AR των διαγραμμάτων ελέγχου με την ίδια τιμή ARL_{in}. Κάθε AR κελί περιέχει τις τιμές που επιτυγχάνονται για $\gamma = 0.5$ και 0.2 αντίστοιχα

ARL_{in}	m	n	*W-διάγραμμα*					*Διάγραμμα Κεφ.4* ($j=(n+1)/2$)			
			(LCL,UCL)	w	r_0	*Ακριβές* ARL_{in}	AR	(LCL,UCL)	r	*Ακριβές* ARL_{in}	AR
370	100	5	(8,11)	51	2	370.30	**0.1424** **0.6757**	(6,95)	1	359.64	**0.0990** **0.6121**
		7	(6,9)	44	2	379.04	**0.2325** **0.8578**	(9,92)	2	369.35	**0.1279** **0.7283**
		11	(11,14)	75	4	358.82	**0.2937** **0.9364**	(14,87)	3	360.04	**0.1952** **0.8770**
	200	5	(14,17)	60	2	367.31	**0.1240** **0.6575**	(11,190)	2	363.60	**0.0895** **0.6028**
		7	(9,12)	40	2	379.44	**0.1823** **0.8288**	(16,185)	3	372.86	**0.1105** **0.7103**
		11	(7,10)	50	2	376.58	**0.3359** **0.9673**	(22,179)	5	367.80	**0.1410** **0.8370**
500	100	5	(8,11)	88	2	499.46	**0.1404** **0.6744**	(4,97)	3	499.88	**0.0752** **0.5407**
		7	(6,9)	74	2	500.62	**0.2280** **0.8545**	(8,93)	3	490.92	**0.1116** **0.7019**
		11	(11,14)	80	4	494.65	**0.2883** **0.9360**	(13,88)	4	462.98	**0.1749** **0.8621**
	200	5	(14,17)	93	2	498.29	**0.1195** **0.6532**	(7,194)	3	500.20	**0.0621** **0.5229**
		7	(9,12)	41	2	505.36	**0.1772** **0.8266**	(16,185)	2	490.76	**0.1062** **0.7083**
		11	(18,21)	59	4	504.50	**0.2405** **0.9192**	(21,180)	5	496.80	**0.1274** **0.8255**

Συγκρίνουμε τέλος την απόδοση των δύο μη παραμετρικών διαγραμμάτων ελέγχου καθορίζοντας ένα κοινό εντός ελέγχου μέσο μήκος ροής ARL_{in} και υπολογίζοντας τις αντίστοιχες πιθανότητες συναγερμού (AR) για συγκεκριμένες εκτός ελέγχου (τύπου *Lehmann*) εναλλακτικές. Στον Πίνακα 5.7, παρουσιάζουμε τις τιμές AR των δύο διαγραμμάτων για $ARL_{in} = 370, 500$, $m = 100, 200$, $n = 5, 7, 11$ και

$\gamma = 0.2, 0.5$. Οι παράμετροι σχεδιασμού $LCL = X_{a:m}$, $UCL = X_{b:m}$, w, r_0 προσδιορίζονται με τη βοήθεια της (5.12), έτσι ώστε τα ARL_{in} να είναι όσο το δυνατόν πλησιέστερα στην τιμή στόχο που έχουμε επιλέξει σε κάθε περίπτωση. Τα αποτελέσματα του Πίνακα 5.7 αποκαλύπτουν την υπεροχή του W–διαγράμματος σύμφωνα και με αυτό τον τρόπο σύγκρισης.

Σύνοψη

Στα πλαίσια της παρούσας διδακτορικής διατριβής μελετώνται θέματα και εφαρμογές της Θεωρίας Διατεταγμένων Παρατηρήσεων στα πεδία της Θεωρίας Αξιοπιστίας και του Στατιστικού Ελέγχου Ποιότητας. Ειδικότερα, μελετήθηκε εκτενώς το πρόβλημα της διατήρησης των ιδιοτήτων γήρανσης κατά το σχηματισμό ενός μονότονου συστήματος αξιοπιστίας με τη χρήση του διανύσματος της υπογραφής του. Πιο συγκεκριμένα, στο Κεφάλαιο 2 παρουσιάζεται μια νέα μέθοδος εύρεσης της γεννήτριας των συντεταγμένων της υπογραφής ενός συστήματος με τη βοήθεια της συνάρτησης αξιοπιστίας του. Στη συνέχεια, διατυπώνονται συνθήκες, οι οποίες αν ισχύουν, εξασφαλίζουν τη διατήρηση του είδους γήρανσης των μονάδων και στο σύστημα που σχηματίζουν.

Επιπρόσθετα, εξετάζεται το σύστημα συνεχόμενο k–από–τα–$n:F$ (ευθύγραμμο και κυκλικό), για το οποίο υπολογίζονται αναδρομικές σχέσεις για τις συντεταγμένες της υπογραφής του. Στη συνέχεια, θεμελιώνεται μια σχέση ανάμεσα στις υπογραφές ευθύγραμμου και κυκλικού συστήματος, η οποία μας απαλλάσσει από την πολυπλοκότητα των υπολογισμών της μιας εκ των δύο περιπτώσεων. Με τη βοήθεια των προαναφερθεισών αναδρομικών σχέσεων, αποδεικνύονται αποτελέσματα σχετικά με τη διατήρηση της ιδιότητας *IFR* στα συστήματα συνεχόμενα k–από–τα–$n:F$. Τέλος, μελετάται αναλυτικά η οικογένεια των συστημάτων συνεχόμενα 2–από–τα–$n:F$, για τα οποία υπολογίζεται η υπογραφή τους και διατυπώνονται συνθήκες σχετικά με τη διατήρηση της ιδιότητας *IFR*, ξεχωριστά για άρτιο και περιττό πλήθος μονάδων n.

Επιπλέον, μελετώνται τα συστήματα r–μεταξύ–συνεχόμενων–k–από–τα–$n:F$ και m–συνεχόμενα–k–από–τα–$n:F$, για τα οποία υπολογίζονται οι συντεταγμένες των υπογραφών τους μέσω αναδρομικών σχέσεων. Με τη βοήθεια αυτών, αποδεικνύονται αποτελέσματα σχετικά με τη διατήρηση της ιδιότητας *IFR* στα παραπάνω συστήματα. Επιπρόσθετα, πραγματοποιούνται στοχαστικές συγκρίσεις για τους χρόνους ζωής διαφόρων συστημάτων αξιοπιστίας μέσω της υπογραφής τους. Αξίζει να σημειωθεί ότι η μελέτη και άλλων συστημάτων αξιοπιστίας με χρόνους ζωής εμφυτεύσιμους σε Μαρκοβιανή αλυσίδα (*Markov chain imbeddable structures*) μπορεί να υλοποιηθεί με την εφαρμογή της μεθοδολογίας που θεμελιώνεται στην παρούσα διδακτορική διατριβή.

Η Θεωρία των Διατεταγμένων Παρατηρήσεων βρίσκει εφαρμογές και στον Στατιστικό Ποιοτικό Έλεγχο και ειδικότερα στην κατασκευή μη παραμετρικών διαγραμμάτων ελέγχου. Στα πλαίσια της διδακτορικής διατριβής (βλ. Κεφάλαιο 4) προτείνονται δύο νέα ελεύθερα κατανομής διαγράμματα ελέγχου τύπου *Shewhart* με χρήση της διαμέσου του εξεταζόμενου δείγματος που εξάγεται από την υπό παρακολούθηση παραγωγική διεργασία. Για τα νέα αυτά διαγράμματα ελέγχου, υπολογίζεται η χαρακτηριστική καμπύλη και η πιθανότητα λανθασμένου συναγερμού τους, που επιτρέπουν τον κατάλληλο σχεδιασμό τους για την επίτευξη της επιθυμητής (σε κάθε περίπτωση) απόδοσης και λειτουργίας τους. Στη συνέχεια, μελετάται η κατανομή του μήκους ροής των νέων μη παραμετρικών διαγραμμάτων ελέγχου και προσδιορίζεται το μέσο μήκος ροής (ARL), τόσο σε εντός όσο και σε εκτός ελέγχου διεργασίες. Τέλος, υλοποιούνται αριθμητικές συγκρίσεις που επιβεβαιώνουν την αποτελεσματικότητα και αντοχή των δύο νέων διαγραμμάτων ελέγχου, όταν αυτά συγκρίνονται τόσο με γνωστά παραμετρικά διαγράμματα (όπως το $\overline{X}-$διάγραμμα ελέγχου), όσο και με το ελεύθερο κατανομής διάγραμμα των Chakraborti *et al.* (2004).

Στο Κεφάλαιο 5, προτείνουμε τρία νέα ελεύθερα κατανομής διαγράμματα ελέγχου τύπου *Shewhart*, τα οποία χρησιμοποιούν συναρτήσεις ροών ή βαθμολογικές συναρτήσεις για να οδηγηθούμε σε απόφαση κατά το πόσο μια διεργασία παραμένει εντός ελέγχου ή έχει μετατοπιστεί σε εκτός ελέγχου κατάσταση.

Στη συνέχεια, εξάγουμε ακριβείς τύπους για την εντός ελέγχου κατανομή των προαναφερθεισών στατιστικών συναρτήσεων και υπολογίζουμε τις πιθανότητες λανθασμένου συναγερμού των τριών μη παραμετρικών διαγραμμάτων ελέγχου, ενώ εξετάζουμε εκτός ελέγχου μετατοπίσεις της διεργασίας σε εναλλακτικές τύπου *Lehmann*, υπολογίζοντας την πιθανότητα ανίχνευσης της παραπάνω μετατόπισης στα προτεινόμενα διαγράμματα. Με τη βοήθεια της ακριβούς κατανομής του μήκους ροής των νέων διαγραμμάτων ελέγχου, η οποία προκύπτει με την εφαρμογή της τεχνικής δέσμευσης που αρχικά είχε προταθεί στην εργασία του Chakraborti (2000), υπολογίζουμε το μέσο μήκος ροής (ARL), τόσο σε εντός όσο και σε εκτός ελέγχου διεργασίες. Τέλος, παρουσιάζουμε αριθμητικούς υπολογισμούς, οι οποίοι επαληθεύουν την αποτελεσματικότητα των νέων διαγραμμάτων, τόσο σε εντός ελέγχου καταστάσεις, όσο και σε περιπτώσεις εκτός ελέγχου μετατοπίσεων της διεργασίας.