Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο   Π Ε Ι Ρ Α Ι Ω Σ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΜΗ–ΠΑΡΑΜΕΤΡΙΚΑ ΔΙΑΓΡΑΜΜΑΤΑ ΣΤΟ ΣΤΑΤΙΣΤΙΚΟ ΕΛΕΓΧΟ ΔΙΕΡΓΑΣΙΑΣ

Παναγιώτης Β. Ματζιώρος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική.

Πειραιάς,
Οκτώβριος 2008

Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο   Π Ε Ι Ρ Α Ι Ω Σ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΜΗ–ΠΑΡΑΜΕΤΡΙΚΑ ΔΙΑΓΡΑΜΜΑΤΑ ΣΤΟ ΣΤΑΤΙΣΤΙΚΟ ΕΛΕΓΧΟ ΔΙΕΡΓΑΣΙΑΣ

Παναγιώτης Β. Ματζιώρος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική.

Πειραιάς,
Οκτώβριος 2008

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ'άριθμ. ..... συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της τριμελούς επιτροπής ήταν :

Μάρκος Κούτρας, Καθηγητής, Επιβλέπων

Δημήτριος Αντζουλάκος, Αναπληρωτής Καθηγητής

Δημήτριος Στέγγος, Επίκουρος Καθηγητής

Η έγκριση της Διπλωματικής Εργασίας από το τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

U N I V E R S I T Y O F P I R A E U S

DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

NONPARAMETRIC CONTROL CHARTS IN STATISTICAL PROCESS CONTROL

By

Panayiotis V. Matzioros

MSc Dissertation
submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfillment of the requirements for the degree of Master of Science in Applied Statistics.

Piraeus, Greece
Oktober 2008

Στους γονείς μου,

Μυρσίνη και Βασίλειο

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον Επιβλέποντα Καθηγητή Κύριο Μάρκο Κούτρα, για την πολύτιμη βοήθεια και τη συμβουλευτική κατεύθυνση που μου παρείχε. Θα ήταν πολύ σημαντική παράλειψη να μην αναφέρω το ότι., με βοήθησε σημαντικά σε ολόκληρη την πορεία της διπλωματικής μου εργασίας και φυσικά στην τελική και επιτυχή περάτωσή της. Χωρίς τη δική του μεγάλη συμβολή, η συγκεκριμένη διπλωματική εργασία ίσως δε θα είχε τη σημερινή της μορφή.

Περίληψη

Σήμερα τα μη–παραμετρικά, ή αλλιώς «ελεύθερα κατανομής», διαγράμματα ελέγχου, παίζουν πολύ σημαντικό ρόλο στο Στατιστικό Έλεγχο Διεργασιών, τόσο ως προς το στατιστικό και πιθανοθεωρητικό μέρος τους, όσο και στο γραφικό και διαγραμματικό κομμάτι τους. Αυτό συμβαίνει, επειδή σε αρκετές περιπτώσεις δεν είναι γνωστό με μεγάλη ακρίβεια το είδος και η μορφή της (συνεχούς) κατανομής που ακολουθούν τα δεδομένα που συλλέγονται, ή μπορεί να είναι αρκετά δύσκολη η διαδικασία της εκτίμησης ή της προσέγγισης της κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται.

Όπως είναι φυσικό, τα μη–παραμετρικά δ.ε. παρουσιάζουν ορισμένα σημαντικά πλεονεκτήματα, αλλά και (λιγότερα) μειονεκτήματα έναντι των γνωστών κλασικών παραμετρικών δ.ε.. Έτσι λοιπόν, θα περίμενε κάποιος ότι, τα συγκεκριμένα δ.ε. θα κατείχαν πολύ σημαντική και περίοπτη θέση στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας. Παρ’όλα αυτά, σύμφωνα με τις πρόσφατες αναφορές στη διεθνή βιβλιογραφία σχετικά με τη σημερινή χρήση και εφαρμογή των μη–παραμετρικών δ.ε., προκύπτει το συμπέρασμα ότι τα συγκεκριμένα δ.ε. δε φαίνεται να χρησιμοποιούνται αρκετά συχνά στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας. Αντιθέτως, φαίνεται εκ πρώτης όψεως να κυριαρχούν τα κλασικά παραμετρικά δ.ε. τύπου Shewhart, καθώς και τα διαγράμματα *CUSUM* και *EWMA*. Έστω όμως και υπό αυτές τις συνθήκες, τα μη–παραμετρικά δ.ε., όχι μόνο δε χάνουν σε καμία περίπτωση τη θεωρητική και πρακτική τους αξία, αλλά συνεχίζουν να αποτελούν ένα αρκετά ενδιαφέρον, χρήσιμο, πρακτικό και ισχυρό γνωστικό υποπεδίο του Στατιστικού Ελέγχου Διεργασιών και του Στατιστικού Ελέγχου Ποιότητας και της Επιστήμης της Στατιστικής γενικότερα.

Ο κύριος σκοπός αυτής της διπλωματικής εργασίας είναι, η παρουσίαση των σημαντικότερων μη–παραμετρικών δ.ε. τύπου Shewhart και των μη–παραμετρικών δ.ε. τύπου *CUSUM*, *EWMA* και κανόνων ροών που έχουν προταθεί στην πρόσφατη διεθνή βιβλιογραφία. Επίσης, θα αναφερθούν οι βασικότερες στατιστικές ιδιότητες, καθώς και τα διάφορα πλεονεκτήματα και μειονεκτήματα που παρουσιάζουν τα συγκεκριμένα μη–παραμετρικά δ.ε..

Abstract

Today the nonparametric or, in other words, distribution–free control charts play a very important role in Statistical Process Control, either by their statistical and probabilistic developments or by their graphical and diagrammatical contribution. This happens mostly because, in a lot of cases, the type and the shape of the (continuous) distribution which follow the collected data, is not known precisely, or because it may be too difficult to estimate and approximate with high accuracy the exact type and form of the (continuous) distribution which describes the monitored observations.

As expected, the nonparametric control charts own some quite essential advantages, but also (fewer) disadvantages, as compared to the classical and more popular parametric control charts. Thus, one would expect that these control charts would occupy a much prominent and outstanding position in modern applications and implementations of Statistical Process Control. However, according to the current surveys about the use of nonparametric control charts, these charts are not broadly used in the modern implementations of Statistical Process Control. On the contrary, one may argue with high confidence that the Shewhart, *CUSUM* and *EWMA* parametric control charts are predominating over them. But even if the previous assertion is really true, the nonparametric control charts not only retain their theoretical and practical significance and value, but also consist a very interesting, useful, practical and powerful scientific subfield of Statistical Process Control and of Statistical Quality Control and Statistics, on the whole.

The main goal of this MSc Dissertation is the presentation of the most important nonparametric Shewhart, *CUSUM*, *EWMA* and runs–type control charts which have been proposed in the recent international bibliography. In addition, the basic statistical properties, the advantages and the disadvantages characterizing these nonparametric control charts will be outlined.

Περιεχόμενα

Κατάλογος Συντομογραφιών

α.ο.ε	άνω όριο ελέγχου ή άνω όρια ελέγχου
α.σ.κ..	αθροιστική συνάρτηση κατανομής ή αθροιστικές συναρτήσεις κατανομής
δ.ε.	διάγραμμα ελέγχου ή διαγράμματα ελέγχου
Θ.Ο.Π.	Θεώρημα Ολικής Πιθανότητας
Θ.Π.	Θεωρία Πιθανοτήτων
κ.ο.ε.	κάτω όριο ελέγχου ή κάτω όρια ελέγχου
ο.ε.	όριο ελέγχου ή όρια ελέγχου
σ.π.	συνάρτηση πιθανότητας ή συναρτήσεις πιθανότητας
σ.π.π.	συνάρτηση πυκνότητας πιθανότητας ή συναρτήσεις πυκνότητας πιθανότητας
τ.δ.	τυχαίο δείγμα ή τυχαία δείγματα
τ.μ.	τυχαία μεταβλητή ή τυχαίες μεταβλητές
ARL	μέσο μήκος ροής
ARL_{in}	εντός ελέγχου μέσο μήκος ροής
ARL_{out}	εκτός ελέγχου μέσο μήκος ροής
CL	κεντρική γραμμή στο διάγραμμα ελέγχου
FAR	πιθανότητα «λανθασμένου συναγερμού» στο διάγραμμα ελέγχου
LCL	κάτω όριο ελέγχου στο διάγραμμα ελέγχου
UCL	άνω όριο ελέγχου στο διάγραμμα ελέγχου

ΚΕΦΑΛΑΙΟ 1

Μη–παραμετρικά διαγράμματα ελέγχου

1.1 Εισαγωγή

Στα πλαίσια του Στατιστικού Ελέγχου Διεργασίας και γενικότερα του Στατιστικού Ελέγχου Ποιότητας, η υπόθεση της κανονικής κατανομής για τις (συνεχείς) παρατηρήσεις στα δείγματα που συλλέγονται, μέχρι πριν από κάποια έτη, έπαιζε πολύ σημαντικό και κυρίαρχο ρόλο. Δυστυχώς όμως, στην περίπτωση που παραβιάζεται η πολύ σημαντική αυτή υπόθεση της κανονικότητας για τα δεδομένα, τα στατιστικά αποτελέσματα και συμπεράσματα που προκύπτουν στο Στατιστικό Έλεγχο Διεργασίας δεν παρουσιάζουν μεγάλη ισχύ και αξιοπιστία. Δηλαδή, αν τα δεδομένα που συλλέγονται δεν ακολουθούν την κανονική κατανομή και αν κάποιος εφαρμόσει όλες εκείνες τις στατιστικές μεθόδους, οι οποίες στηρίζονται στην υπόθεση της κανονικότητας, είναι πολύ πιθανό ότι θα καταλήξει σε λανθασμένα στατιστικά συμπεράσματα (βλέπε Αντζουλάκος (2006)).

Για το σκοπό αυτό, τις τρεις τελευταίες δεκαετίες, στον τομέα του Στατιστικού Ελέγχου Διεργασίας, δημιουργήθηκαν και αναπτύχθηκαν ειδικές στατιστικές μέθοδοι και τεχνικές, οι οποίες δεν επηρεάζονται καθόλου από την υπόθεση της κανονικότητας των δεδομένων. Οι συγκεκριμένες μέθοδοι, δεν απαιτούν τα δεδομένα που συλλέγονται να ακολουθούν την κανονική κατανομή, αφού χρησιμοποιούν γνωστά αποτελέσματα από τον κλάδο της Απαραμετρικής Στατιστικής.

Η στατιστική δομή των μη–παραμετρικών δ.ε. έχει ως αποτέλεσμα αυτά, να παρουσιάζουν αφ'ενός μεν σημαντικά πλεονεκτήματα, αλλά ταυτόχρονα και ορισμένα μειονεκτήματα. Τα τελευταία χρόνια, πρέπει να σημειωθεί ότι, στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας, παρουσιάζεται ιδιαίτερα αυξημένη ανάγκη για τη χρησιμοποίηση μη–παραμετρικών (ελεύθερων κατανομής) δ.ε.. Αυτό συμβαίνει επειδή, αρκετά συχνά στην πράξη, η φύση και η στατιστική δομή των δεδομένων δεν ικανοποιεί τη βασική υπόθεση της κανονικότητας.

Πρέπει να σημειωθεί εδώ ότι, το γνωστικό αντικείμενο των μη–παραμετρικών δ.ε. εξακολουθεί ακόμη και σήμερα να είναι μη κορεσμένο, δηλαδή παρατηρείται ότι υπάρχουν ακόμη πολλά κίνητρα για εκτεταμένη και αναλυτική επιστημονική έρευνα πάνω στα μη–παραμετρικά δ.ε.. Σε κάθε περίπτωση πάντως, θεωρείται ότι τα συγκεκριμένα δ.ε. κατέχουν πολύ σημαντική θέση στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας και της Στατιστικής γενικότερα, αφού παρουσιάζουν εξαιρετικό επιστημονικό ενδιαφέρον, τόσο ως προς το πιθανοθεωρητικό μέρος τους, όσο και ως προς το στατιστικό και πρακτικό κομμάτι τους.

1.2 Ο ορισμός, οι βασικές ιδιότητες και τα πλεονεκτήματα ενός μη–παραμετρικού διαγράμματος ελέγχου

Στο Στατιστικό Έλεγχο Διεργασιών (βλέπε Chakraborti, Bakir and Van Der Laan (2001)), με τον όρο «μη–παραμετρικό δ.ε.», εννοούμε ένα δ.ε. στο οποίο η κατανομή του μήκους ροής, το εντός–ελέγχου μέσο μήκος ροής ARL_{in}, καθώς και η πιθανότητα «λανθασμένου συναγερμού» *FAR* είναι ίδιες και σταθερές για όλες τις συνεχείς κατανομές που είναι πιθανό να ακολουθούν τα δεδομένα. Επομένως, τα συγκεκριμένα στατιστικά χαρακτηριστικά του διαγράμματος δεν εξαρτώνται από το είδος της (συνεχούς) κατανομής την οποία ακολουθούν τα δεδομένα που συλλέγονται. Ίσως, ο όρος «ελεύθερο κατανομής» να ταίριαζε καλύτερα σε σύγκριση με τον όρο «μη–παραμετρικό» για τα συγκεκριμένα δ.ε.. Με άλλα λόγια, στα μη–παραμετρικά δ.ε., δεν απαιτείται κάποια συγκεκριμένη κατανομή για τις παρατηρήσεις. Το αποτέλεσμα αυτό, αποτελεί γενικά, ένα πολύ σημαντικό πλεονέκτημα για τα μη–παραμετρικά δ.ε..

Επίσης (βλέπε Chakraborti, Bakir and Van Der Laan (2001)), ένα ακόμη πλεονέκτημα των συγκεκριμένων διαγραμμάτων είναι ότι, η φύση και η στατιστική δομή τους τα καθιστά εξαιρετικά ανθεκτικά στις έκτροπες παρατηρήσεις (outliers) που είναι πιθανό να υπάρχουν. Αυτό συμβαίνει, κυρίως επειδή οι απαραμετρικές τεχνικές που χρησιμοποιούνται στα συγκεκριμένα δ.ε., συντελούν ώστε αυτά να μην επηρεάζονται σχεδόν καθόλου από τυχόν έκτροπες και ακραίες τιμές, χαρακτηριστικό που δεν παρουσιάζουν τα κλασικά παραμετρικά δ.ε. τύπου Shewhart και *CUSUM*. Ακόμη, στα μη–παραμετρικά δ.ε., για τον υπολογισμό των δύο ο.ε., δε χρησιμοποιούνται εκτιμήσεις των άγνωστων στατιστικών παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις, αλλά τα δύο ο.ε. συνήθως (στα μη–παραμετρικά δ.ε. τύπου Shewhart) υπολογίζονται χρησιμοποιώντας τις τιμές διατεταγμένων παρατηρήσεων, σε αντίθεση με τα κλασικά παραμετρικά δ.ε. τύπου Shewhart, στα οποία απαιτείται η εκτίμηση των άγνωστων παραμέτρων μ και σ της κανονικής κατανομής που υποτίθεται ότι ακολουθούν τα δεδομένα.

Όμως, στις απαραμετρικές μεθόδους που χρησιμοποιούνται, περιλαμβάνονται αρκετές πολύπλοκες μαθηματικές σχέσεις, οι οποίες μπορούν να επιλυθούν μόνο με τη βοήθεια μαθηματικών προγραμμάτων στον υπολογιστή ή με μεθόδους Αριθμητικής Ανάλυσης ή Προσομοίωσης και όχι με γνωστές και συνηθισμένες τεχνικές. Ακόμη, στα περισσότερα μη–παραμετρικά δ.ε., απαιτείται η χρήση ειδικών πινάκων που περιλαμβάνουν εξειδικευμένα μαθηματικά και στατιστικά δεδομένα, τα οποία δυστυχώς δεν είναι γνωστά σε όλους όσους ασχολούνται εντατικά με το αντικείμενο του Στατιστικού Ελέγχου Διεργασίας. Εκτός από τα παραπάνω, τα μη–παραμετρικά δ.ε., κατά τη χρήση και την πρακτική εφαρμογή τους, απαιτούν αρκετά καλές γνώσεις Απαραμετρικής Στατιστικής, τις οποίες συχνά δε διαθέτουν όσοι πιθανόν ασχοληθούν με τα συγκεκριμένα διαγράμματα, αφού το γνωστικό αντικείμενο που κυριαρχούσε και κυριαρχεί μέχρι σήμερα είναι η Κλασική Παραμετρική Στατιστική και όχι η Μη–Παραμετρική Στατιστική.

Στα πλαίσια της παρούσας διπλωματικής εργασίας, θα επικεντρωθούμε στα μονομεταβλητά (univariate) μη–παραμετρικά δ.ε. και δε θα ασχοληθούμε αρκετά με τα λεγόμενα πολυμεταβλητά (multivariate) μη–παραμετρικά δ.ε.. Για τα πολυμεταβλητά μη–παραμετρικά δ.ε., θα αναφέρουμε μόνο (βλέπε Qiu and Hawkins (2003)) ότι χρησιμοποιούνται για την παρακολούθηση μιας διεργασίας η οποία περιλαμβάνει περισσότερα του ενός μετρήσιμα χαρακτηριστικά ή ένα διάνυσμα συγκεκριμένου πλήθους μετρήσιμων χαρακτηριστικών, δεν υποθέτουν συγκεκριμένη πολυδιάστατη κατανομή για τα δεδομένα και χρησιμοποιούν ειδικές τεχνικές του κλάδου της Απαραμετρικής Στατιστικής, αλλά τις εφαρμόζουν σε πολυδιάστατα δεδομένα. Ο κύριος σκοπός των πολυμεταβλητών δ.ε. είναι, να ελέγξουν έγκαιρα, αν κάποια στατιστική παράμετρος (δηλαδή κάποιο διάνυσμα παραμέτρων) της πολυδιάστατης κατανομής των δεδομένων παρέμεινε η ίδια ή μεταβλήθηκε. Δηλαδή, τα πολυμεταβλητά μη–παραμετρικά δ.ε., είναι το αντίστοιχο των μονομεταβλητών μη–παραμετρικών δ.ε., για την περίπτωση κατά την οποία οι παρατηρήσεις που συλλέγονται είναι πολυδιάστατες.

Σε αυτή τη διπλωματική εργασία, θα παρουσιάσουμε τα σημαντικότερα μη– παραμετρικά δ.ε. που έχουν προταθεί και αναλυθεί στη διεθνή βιβλιογραφία τις τελευταίες τρεις δεκαετίες. Συγκεκριμένα, θα παρουσιάσουμε μη–παραμετρικά δ.ε. τύπου Shewhart, ***CUSUM***, *EWMA* και μη–παραμετρικά δ.ε. που βασίζονται σε κανόνες ροών. Τα περισσότερα από αυτά τα δ.ε. θα παρουσιαστούν αναλυτικά, μαζί με όλες τις γραφικές, στατιστικές και πιθανοθεωρητικές ιδιότητές τους.

1.3 Μειονεκτήματα των μη–παραμετρικών διαγραμμάτων ελέγχου

Το βασικότερο μειονέκτημα που παρουσιάζουν τα μη–παραμετρικά δ.ε. είναι (βλέπε Chakraborti, Bakir and Van Der Laan (2001)) η πολυπλοκότητα και η δυσκολία που υπάρχει στην επίλυση ορισμένων μαθηματικών τύπων και ολοκληρωτικών εκφράσεων που σχετίζονται με αυτά, αποτέλεσμα το οποίο καθιστά μάλλον αναγκαία τη χρήση ειδικών υπολογιστικών προγραμμάτων ή προσομοίωσης. Εκτός από αυτό, αρκετά συχνά, στις πρακτικές εφαρμογές αυτών των διαγραμμάτων, απαιτείται η χρήση ειδικών στατιστικών πινάκων, οι οποίοι περιέχουν αναλυτικά κατανομές πιθανότητας συγκεκριμένων διακριτών τ.μ.. Για αυτούς τους ειδικούς πίνακες, δεν υπάρχει πάντοτε διαθέσιμη βιβλιογραφία για να μπορέσει κανείς να τη χρησιμοποιήσει, προκειμένου να υλοποιήσει πρακτικά τα συγκεκριμένα διαγράμματα. Παράλληλα (βλέπε Chakraborti, Bakir and Van Der Laan (2001)), φαίνεται ότι δεν υπάρχουν πολλά άτομα που να έχουν ευρεία και «βαθιά» γνώση των μη–παραμετρικών δ.ε. και γενικότερα του κλάδου της Απαραμετρικής Στατιστικής, αφού η κλασική Παραμετρική Στατιστική έχει διδαχθεί πολύ πιο αναλυτικά και διεξοδικά σε σύγκριση με την Απαραμετρική Στατιστική. Έτσι, προκύπτει το συμπέρασμα ότι ορισμένοι επιστήμονες σήμερα, μάλλον δε θα έχουν αφομοιώσει ικανοποιητικά όλες εκείνες τις στατιστικές και πιθανοθεωρητικές έννοιες που απαιτούνται για την χρήση και την πρακτική εφαρμογή των μη–παραμετρικών δ.ε.. Ένα ακόμη μειονέκτημα για τα συγκεκριμένα διαγράμματα, είναι ότι δε μπορεί να υπολογιστεί εύκολα και άμεσα το εκτός–ελέγχου μέσο μήκος ροής ARL_{out}, αφού αυτή η ποσότητα δεν είναι ίδια και σταθερή για όλες τις συνεχείς κατανομές, αλλά διαφέρει από κατανομή σε κατανομή και κατά συνέπεια αν η διεργασία είναι εκτός στατιστικού ελέγχου και δεν είναι γνωστή η νέα συνεχής κατανομή που ακολουθούν τα δεδομένα, στο μη–παραμετρικό δ.ε. δε μπορεί να υπολογιστεί ή τουλάχιστον, να εκτιμηθεί ικανοποιητικά, το εκτός–ελέγχου μέσο μήκος ροής $\boldsymbol{ARL_{out}}$. Ένα τελευταίο μειονέκτημα, είναι ότι συνήθως απαιτούν τη συλλογή αρκετών δειγμάτων ή τη συλλογή ενός αρχικού δείγματος μεγάλου μεγέθους και αυτό μπορεί να προκαλέσει σημαντική δαπάνη χρόνου και κυρίως, μεγάλο οικονομικό κόστος.

ΚΕΦΑΛΑΙΟ 2

Τα είδη των μη–παραμετρικών δ.ε.

2.1 Δίπλευρα μη–παραμετρικά δ.ε. τύπου Shewhart που χρησιμοποιούν για ο.ε. διατεταγμένες παρατηρήσεις

Το πρώτο είδος μη–παραμετρικών δ.ε., προτάθηκε σχετικά πρόσφατα από τους Chakraborti, Van Der Laan and Van De Wiel (2000) και χρησιμοποιεί ένα προκαταρκτικό τ.δ. «αναφοράς» (reference sample) μεγέθους m, έστω X_1, X_2, ..., X_m από τη διεργασία, θεωρώντας ότι τη στιγμή που αυτό συλλέγεται η διεργασία είναι εντός στατιστικού ελέγχου. Αυτό το προκαταρκτικό τ.δ διατάσσεται κατά αύξουσα σειρά (ascending order) και ύστερα επιλέγονται δύο παρατηρήσεις, η α–διατεταγμένη παρατήρηση $X_{(\alpha;m)}$ και η b–διατεταγμένη παρατήρηση $X_{(b;m)}$, για να χρησιμοποιηθούν ως κ.ο.ε. LCL και α.ο.ε. UCL αντίστοιχα στο συγκεκριμένο μη–παραμετρικό δ.ε.. Στη συνέχεια, συλλέγονται και άλλα τ.δ. $Y_1, Y_2, ..., Y_n$ από τη διεργασία, μεγέθους n παρατηρήσεων το καθένα (δηλαδή τα νέα, ανεξάρτητα τ.δ. που συλλέγονται είναι ισοπληθή μεταξύ τους), υπολογίζεται το διατεταγμένο τ.δ. για κάθε ένα από αυτά και μετά απ'αυτό, η j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ συγκρίνεται με τα δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$. Αν η j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ βρίσκεται μέσα στα δύο ο.ε. $X_{(\alpha;m)}$ και $X_{(b;m)}$, τότε η διεργασία θεωρείται ότι έχει παραμείνει εντός στατιστικού ελέγχου. Στην αντίθετη περίπτωση, δηλαδή αν για κάποιο διατεταγμένο τ.δ. η j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ βρεθεί εκτός των δύο ο.ε. $X_{(\alpha;m)}$ και $X_{(b;m)}$, η διεργασία θεωρείται ότι βρίσκεται εκτός στατιστικού ελέγχου και θα πρέπει να αναζητηθούν ειδικές αιτίες μεταβλητότητας (βλέπε Αντζουλάκος (2006)).

2.1.1 Η ακριβής εντός–ελέγχου κατανομή της στατιστικής συνάρτησης W_j

Έστω W_j η διακριτή, θετική τ.μ. που δηλώνει τον αριθμό των παρατηρήσεων στο αρχικό τ.δ $X_1, X_2, \ldots, X_m$, οι οποίες είναι μικρότερες ή το πολύ ίσες από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ του νέου τ.δ. μεγέθους n. Η σ.π της τ.μ. W_j παίζει πολύ σημαντικό ρόλο στον υπολογισμό της πιθανότητας λανθασμένου συναγερμού *FAR* στο συγκεκριμένο μη–παραμετρικό δ.ε..

Η σ.π. της τ.μ. W_j μπορεί να υπολογιστεί με ακρίβεια, είτε με χρήση μαθηματικών μεθόδων, είτε με χρήση προσομοίωσης είτε με μεθόδους της Συνδυαστικής. Χρησιμοποιώντας τεχνικές της Συνδυαστικής, θα δείξουμε στη συνέχεια ότι η σ.π. της τ.μ. W_j, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, θα δίνεται από τον παρακάτω τύπο:

$$P_C(W_j = w) = \frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}},$$

για $w = 0, 1, \ldots, m$.

Έστω ότι η διεργασία βρίσκεται εντός στατιστικού ελέγχου και ας συμβολίσουμε με $Z_1, Z_2, Z_3, \ldots, Z_{n+m-1}, Z_{n+m}$ το συνολικό τ.δ. που προκύπτει από τις m παρατηρήσεις $X_1, X_2, \ldots, X_m$ του προκαταρκτικού τ.δ. και από τις n παρατηρήσεις $Y_1, Y_2, \ldots, Y_n$ του νέου τ.δ.. Η ζητούμενη πιθανότητα $P_C(W_j=w)$ μπορεί να υπολογιστεί, σύμφωνα με τον κλασικό ορισμό της πιθανότητας (βλέπε Κούτρας (2002) (α)), ως το πηλίκο του συνολικού πλήθους των «ευνοϊκών» ενδεχομένων προς το συνολικό πλήθος όλων των ενδεχομένων του δειγματικού χώρου. Τα «ευνοϊκά» ενδεχόμενα είναι εκείνα τα ενδεχόμενα για τα οποία στο τ.δ. $Z_1, Z_2, \ldots, Z_{n+m-1}, Z_{n+m}$ υπάρχουν συνολικά ακριβώς w παρατηρήσεις X_i μικρότερες από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ και συνολικά ακριβώς $m–w$ παρατηρήσεις X_i μεγαλύτερες από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$, έτσι ώστε ο συνολικός αριθμός των παρατηρήσεων X_i οι οποίες προηγούνται της j–διατεταγμένης παρατήρησης $Y_{(j;n)}$ να είναι ακριβώς ίσος με w.

Στο ολικό διατεταγμένο τ.δ. που αποτελείται από όλες τις παρατηρήσεις X_i και Y_i, το συνολικό πλήθος των «ευνοϊκών» ενδεχομένων για την πιθανότητα $P_C(W_j=w)$ μπορεί να υπολογιστεί (βλέπε Κούτρας (2001)) σύμφωνα με το εξής σκεπτικό: αν διαταχθούν κατά αύξουσα σειρά όλες οι $n+m$ συνολικά το πλήθος παρατηρήσεις, για να πραγματοποιηθεί το ενδεχόμενο $\{W_j=w\}$ θα πρέπει, πριν από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$, να υπάρχουν διατεταγμένες ακριβώς w παρατηρήσεις X_i και j–1 παρατηρήσεις Y_i. Δηλαδή, πριν από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$, θα πρέπει να υπάρχουν συνολικά ακριβώς $w+(j-1)$ «κενές θέσεις» τις οποίες θα καταλαμβάνουν συνολικά ακριβώς w παρατηρήσεις X_i. Άρα, οι w παρατηρήσεις X_i, θα μπορούν να «τοποθετηθούν» στις $w+(j-1)$ «κενές θέσεις» με $\binom{w+(j-1)}{w}$ συνολικά πιθανούς τρόπους. Δηλαδή, για κάθε μία «τοποθέτηση» των w παρατηρήσεων X_i στις $w+(j-1)$ συνολικά «κενές θέσεις», απομένει ένας μόνο τρόπος «τοποθέτησης» των υπόλοιπων j–1 παρατηρήσεων Y_i στις υπόλοιπες j–1 «κενές θέσεις». Επομένως, σύμφωνα με την Πολλαπλασιαστική Αρχή της Συνδυαστικής (βλέπε Κούτρας (2001)), προκύπτει το συμπέρασμα ότι οι w παρατηρήσεις X_i και οι j–1 παρατηρήσεις Y_i μπορούν να διαταχθούν πριν από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$, ώστε να μπορεί να πραγματοποιηθεί το ενδεχόμενο $\{W_j=w\}$ με συνολικά $\binom{w+(j-1)}{w}=\binom{w+j-1}{w}$ πιθανούς τρόπους.

Επίσης, υπάρχουν άλλες $(n+m)-(j+w)$ «κενές θέσεις» για τις υπόλοιπες $(n-j)$ παρατηρήσεις Y_i και $(m-w)$ παρατηρήσεις X_i στο συνολικό διατεταγμένο τ.δ.. Η «τοποθέτηση» των υπόλοιπων $(n-j)$ παρατηρήσεων Y_i και $(m-w)$ παρατηρήσεων X_i σε αυτές τις $(n+m)-(j+w)$ συνολικά «κενές θέσεις» που απέμειναν, μπορεί να γίνει ακριβώς με το ίδιο σκεπτικό (βλέπε Κούτρας (2001)) όπως προηγουμένως. Συγκεκριμένα, για να πραγματοποιηθεί το ενδεχόμενο $\{W_j=w\}$, θα πρέπει μετά τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ να υπάρχουν συνολικά ακριβώς $m-w$ διατεταγμένες παρατηρήσεις X_i και $n-j$ διατεταγμένες παρατηρήσεις Y_i. Δηλαδή, μετά τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$, θα πρέπει να υπάρχουν ακριβώς $(n+m)-(j+w)$ «κενές θέσεις» τις οποίες μπορούν να καταλάβουν οι $m-w$ παρατηρήσεις X_i. Αυτό, μπορεί να γίνει με $\binom{(n+m)-(j+w)}{m-w}$ συνολικά διαφορετικούς τρόπους και επειδή για κάθε μία «τοποθέτηση» των $m-w$ παρατηρήσεων X_i στις $(n+m)-(j+w)$ συνολικά «κενές θέσεις» που απομένουν, υπάρχει μόνο ένας τρόπος «τοποθέτησης» των υπόλοιπων $n-j$ παρατηρήσεων Y_i, σύμφωνα με την Πολλαπλασιαστική Αρχή της Συνδυαστικής (βλέπε Κούτρας (2001)), προκύπτει ότι οι $m-w$ παρατηρήσεις X_i και οι n–j παρατηρήσεις Y_i μπορούν να διαταχθούν μετά από τη j–διατεταγμένη παρατήρηση $Y_{(j;n)}$, ώστε να πραγματοποιηθεί το ενδεχόμενο $\{W_j=w\}$, με $\binom{(n+m)-(j+w)}{m-w}=\binom{n+m-j-w}{m-w}$ συνολικά διαφορετικούς πιθανούς τρόπους. Εφαρμόζοντας πάλι την Πολλαπλασιαστική Αρχή της Συνδυαστικής, συμπεραίνουμε ότι, αν η διεργασία είναι εντός στατιστικού ελέγχου, το ενδεχόμενο $\{W_j=w\}$ μπορεί να πραγματοποιηθεί με $\binom{w+(j-1)}{w}\binom{(n+m)-(j+w)}{m-w}=\binom{w+j-1}{w}\binom{n+m-j-w}{m-w}$ συνολικά διαφορετικούς (πιθανούς) τρόπους.

Επίσης, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η «τοποθέτηση» των m παρατηρήσεων X_i στο νέο τ.δ. που αποτελείται από τις $n+m$ παρατηρήσεις X_i και Y_i, μπορεί να γίνει με $\binom{n+m}{m}$ συνολικά διαφορετικούς (πιθανούς) τρόπους. Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, όλα τα «ευνοϊκά» ενδεχόμενα για το (σύνθετο) ενδεχόμενο $\{W_j=w\}$, είναι μεταξύ τους ισοπίθανα, αφού έχουμε ένα τυχαίο (συνολικό) δείγμα από όλες τις παρατηρήσεις. Άρα, εδώ μπορούμε να χρησιμοποιήσουμε τον κλασικό ορισμό της πιθανότητας ενός ενδεχομένου.

Επομένως, σύμφωνα με τον κλασικό ορισμό της πιθανότητας ενός ενδεχομένου (βλέπε Κούτρας (2002) (α)), η πιθανότητα του ενδεχομένου $\{W_j=w\}$, όταν η διεργασία είναι εντός στατιστικού ελέγχου, θα είναι ίση (βλέπε Κούτρας (2002) (α)) με

$$P_C(W_j = w) = \frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}},$$

$w = 0, 1, 2, \ldots, m$.

Στο σημείο αυτό, πρέπει να σημειωθεί ότι, η σ.π. της διακριτής τ.μ. W_j εξαρτάται μόνο από τις παραμέτρους n, m και j και όχι από το είδος και τη μορφή της (συνεχούς) κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται. Με την προϋπόθεση ότι οι παρατηρήσεις που συλλέγονται είναι συνεχείς και μεταξύ τους ανεξάρτητες και ισόνομες, προκύπτει ότι η σ.π. της διακριτής τ.μ. W_j δε θα εξαρτάται από το είδος και τη μορφή της (συνεχούς) κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται. Ως εκ τούτου, η πιθανότητα λανθασμένου συναγερμού *FAR* και το εντός–ελέγχου μέσο μήκος ροής ARL_{in} επίσης δε θα εξαρτώνται από την κατανομή που ακολουθούν τα δεδομένα, κάτι που τελικά καθιστά το συγκεκριμένο δ.ε. μη–παραμετρικό, ή αλλιώς, ελεύθερο κατανομής.

Η σ.π. της διακριτής τ.μ. W_j, ο ακριβής τύπος της οποίας αποδείχτηκε προηγουμένως, παίζει βασικό ρόλο στο συγκεκριμένο μη–παραμετρικό δ.ε., αφού από αυτήν προκύπτει (με έμμεσο τρόπο) η πιθανότητα λανθασμένου συναγερμού *FAR*. Συγκεκριμένα, αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η πιθανότητα η j–διατεταγμένη παρατήρηση $Y_{(j;n)}$ να βρεθεί μέσα στα δύο ο.ε. $LCL= X_{(a;m)}$ και $UCL= X_{(b;m)}$ θα είναι ίση με:

$$p_C = P_C(X_{(a;m)} \le Y_{(j;n)} \le X_{(b;m)}) = P_C(a \le W_j \le b-1) = \sum_{w=a}^{b-1} P_C(W_j = w) =$$

$$= \sum_{w=a}^{b-1} \frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}}.$$

Εδώ, χρησιμοποιήθηκε το ότι η κατανομή των παρατηρήσεων X_i και Y_i είναι συνεχής, επομένως θα ισχύει

$$P(Y_{(j;n)} = X_{(b;m)}) = 0$$

ή ισοδύναμα,

$$P_C(X_{(a;m)} \le Y_{(j;n)} \le X_{(b;m)}) = P_C(X_{(a;m)} \le Y_{(j;n)} < X_{(b;m)}) = P_C(a \le W_j \le b-1).$$

Επομένως, η πιθανότητα λανθασμένου συναγερμού *FAR* θα ισούται με:

$$FAR = P_C(Y_{(j;n)} \notin [X_{(a;m)}, X_{(b;m)}]) = 1 - p_C = 1 - \sum_{w=a}^{b-1} P_C(W_j = w) =$$

$$= 1 - \sum_{w=a}^{b-1} \frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}}.$$

Οι παράμετροι a και b προσδιορίζονται με τέτοιον τρόπο (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)), ώστε η πιθανότητα λανθασμένου συναγερμού *FAR* να είναι αρκετά μικρή ή το πολύ ίση με μία προκαθορισμένη (μικρή) τιμή p_0, δηλαδή να ισχύει
$FAR \le p_0$, ή ισοδύναμα

$$1-\sum_{w=a}^{b-1}\frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}}\leq p_0$$

ή αλλιώς

$$\sum_{w=a}^{b-1}\frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}}\geq 1-p_0.$$

Σε αυτό το σημείο, πρέπει να σημειωθεί ότι, οι δύο άγνωστες παράμετροι ελέγχου *α* και ***b***, δεν προσδιορίζονται μονοσήμαντα από την παραπάνω σχέση, αφού είναι πολύ πιθανό να υπάρχουν περισσότερα από ένα ζεύγη παραμέτρων (*α*,***b***), τα οποία θα ικανοποιούν την παραπάνω σχέση. Οι δύο άγνωστες παράμετροι *α* και *b*, προσδιορίζονται ως εκείνες οι θετικές ακέραιες τιμές, οι οποίες κάνουν το παραπάνω άθροισμα που γράφτηκε να είναι στη μικρότερη απόσταση από την πιθανότητα $1-p_0$ (δηλαδή να πλησιάζει όσο το δυνατόν περισσότερο την τιμή της πιθανότητας $1-p_0$).

Όταν προσδιοριστούν οι άγνωστες παράμετροι *α* και ***b***, τότε αυτόματα προσδιορίζονται και τα δύο ο.ε. *LCL* και *UCL* για αυτό το δ.ε., αφού τα δύο ο.ε. είναι, όπως αναφέρθηκε, η *α*–διατεταγμένη παρατήρηση $X_{(\alpha;m)}$ και η *b*–διατεταγμένη παρατήρηση $X_{(b;m)}$ του τ.δ. αναφοράς.

Στο συγκεκριμένο μη–παραμετρικό δ.ε., οι δύο παράμετροι *α* και ***b*** υπολογίζονται σύμφωνα με την παρακάτω διαδικασία (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)): αρχικά, η πιθανότητα «λανθασμένου συναγερμού» ***FAR*** τίθεται ίση με μία προκαθορισμένη και επιθυμητή αρκετά μικρή τιμή (για παράδειγμα ίση με 0.01 ή 0.05) και στη συνέχεια επιλύεται ως προς τις άγνωστες παραμέτρους α και ***b*** η εξίσωση

$$\sum_{w=a}^{b-1}\frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}}=1-FAR.$$

Προφανώς, η προηγούμενη εξίσωση είναι αρκετά δύσκολο να επιλυθεί αναλυτικά με τις γνωστές και συνηθισμένες μαθηματικές μεθόδους, μπορεί ωστόσο να επιλυθεί με μεθόδους

Αριθμητικής Ανάλυσης ή με χρήση ενός ειδικού μαθηματικού προγράμματος στον υπολογιστή.

Μετά τον προσδιορισμό των δύο αγνώστων παραμέτρων α και b, τα δύο ο.ε. LCL και UCL υπολογίζονται αρκετά εύκολα. Το κ.ο.ε. LCL είναι η α–διατεταγμένη παρατήρηση $X_{(\alpha;m)}$ και το α.ο.ε. UCL είναι η b–διατεταγμένη παρατήρηση $X_{(b;m)}$, που προκύπτουν από την συλλογή του τ.δ. αναφοράς $X_1, X_2,..., X_m$.

Σε αυτό το σημείο, πρέπει να σημειωθεί ότι, τα δύο ο.ε. LCL και UCL, ή ισοδύναμα οι δύο παράμετροι α και b, θα μπορούσαν θεωρητικά να υπολογιστούν και με βάση το εντός–ελέγχου μέσο μήκος ροής ARL_{in}, θέτοντας αρχικά το ARL_{in} ίσο με κάποια συγκεκριμένη αρκετά μεγάλη τιμή. Όμως, σε αυτό το μη–παραμετρικό δ.ε. που χρησιμοποιεί διατεταγμένα τ.δ., το ARL_{in} είναι αρκετά δύσκολο να υπολογιστεί, αφού η κατανομή του συνολικού αριθμού των σημείων που παρουσιάζονται στο δ.ε. μέχρι να προκύψει για πρώτη φορά ένα σημείο εκτός των δύο ο.ε. δε θα είναι η γεωμετρική κατανομή, όπως συμβαίνει στα κλασικά παραμετρικά δ.ε. τύπου Shewhart, επειδή τα σημεία που απεικονίζονται στο δ.ε. δεν είναι ανεξάρτητα μεταξύ τους. Επίσης, για το συγκεκριμένο μη–παραμετρικό δ.ε., δεν υπάρχει γενικά κάποιος «κλειστός» μαθηματικός τύπος ο οποίος να υπολογίζει ή να εκτιμάει το εντός–ελέγχου μέσο μήκος ροής ARL_{in}. Η δυσκολία λοιπόν του ακριβούς υπολογισμού ή έστω μιας ικανοποιητικής προσέγγισης του ARL_{in}, αποτελεί ένα σημαντικό μειονέκτημα για το συγκεκριμένο μη–παραμετρικό δ.ε..

2.1.2 Η δειγματική διάμεσος ως σημείο αναφοράς και κριτήριο

Η πιο συνηθισμένη περίπτωση για το δ.ε. που παρουσιάστηκε προηγουμένως, είναι εκείνη κατά την οποία (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) από τα καινούργια διατεταγμένα τ.δ. μεγέθους n, λαμβάνονται οι δειγματικές διάμεσοι $Y_{(j;n)}$ και συγκρίνονται με τα δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$. Έχει αποδειχθεί, θεωρητικά και πρακτικά, ότι (βλέπε Janacek and Meikle (1997)) η δειγματική διάμεσος ενός τ.δ., ως μέτρο θέσης της κατανομής που ακολουθούν οι παρατηρήσεις, παρουσιάζει πολύ ικανοποιητικές και χρήσιμες ιδιότητες. Για λόγους απλούστευσης, έστω ότι το μέγεθος n των νέων τ.δ. είναι περιττός αριθμός, δηλαδή $n = 2s+1$, όπου s είναι κάποιος θετικός ακέραιος αριθμός (αν το μέγεθος των νέων τ.δ. είναι άρτιος αριθμός $n = 2s$, σε αυτήν την περίπτωση η δειγματική διάμεσος θα είναι το πηλίκο των διατεταγμένων παρατηρήσεων $\frac{Y_s + Y_{s+1}}{2}$, κάτι που επιφέρει κάποια πολυπλοκότητα ή δυσκολία στους υπολογισμούς που απαιτούνται). Τότε, η δειγματική διάμεσος θα είναι η (s+1)–διατεταγμένη παρατήρηση $Y_{(s+1;n)}$, έτσι ώστε ακριβώς s παρατηρήσεις να είναι μικρότερες από την δειγματική διάμεσο και ακριβώς s παρατηρήσεις να είναι μεγαλύτερες από την δειγματική διάμεσο. Στην περίπτωση αυτή, θα ισχύει $j = s+1$ και η σ.π. της τ.μ. W_j, για $j = s+1$ και $n = 2s+1$, θα πάρει την ακόλουθη μορφή:

$$P_C(W_{s+1} = w) = \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}},$$

για $w = 0, 1, \ldots, m$.

2.1.3 Η διαδικασία υπολογισμού των ο.ε. για τη συμμετρική περίπτωση

Είναι σχετικά εύκολο να διαπιστωθεί ότι, όταν γίνεται χρήση της δειγματικής διαμέσου, η κατανομή της διακριτής τ.μ. $W_j = W_{s+1}$ είναι συμμετρική ως προς m, δηλαδή θα ισχύει η ισότητα

$$P_C(W_{s+1} = w) = P_C(W_{s+1} = m-w),\ w = 0, 1, \ldots, m.$$

Η συμμετρικότητα αυτή, υποδεικνύει (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) «χωρίς βλάβη της γενικότητας» (δηλαδή χωρίς να παραβιάζεται κάποια συνθήκη ή κάποιος περιορισμός), η παράμετρος b να οριστεί ίση με $b = m-\alpha+1$ (αν ισχύει $b = m-\alpha+1$, τότε θα ισχύει η σχέση

$P(Y_{(j;n)} > X_{(b;m)} = UCL \mid$ η διεργασία εντός ελέγχου$)$ = $P(Y_{(j;n)} < X_{(a;m)} = LCL \mid$ η διεργασία εντός ελέγχου$)$

και σε αυτήν την περίπτωση, η άγνωστη παράμετρος α προσδιορίζεται πιο εύκολα και με λιγότερες αριθμητικές πράξεις) . Αν p_0 είναι μία προκαθορισμένη πιθανότητα με τιμή αρκετά μικρή, τότε οι παράμετροι α και b θα πρέπει να ικανοποιούν τη σχέση

$$P_C(X_{(\alpha;m)} \le Y_{(j;n)} \le X_{(b;m)}) \ge 1-p_0,$$

δηλαδή την εξής σχέση:

$$\sum_{w=a}^{b-1} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \geq 1-p_0.$$

Όμως, ισχύει $b = m–\alpha+1$, οπότε η προηγούμενη σχέση παίρνει την εξής μορφή:

$$\sum_{w=a}^{m-a} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \geq 1-p_0.$$

Για γνωστά m, s και p_0 η προηγούμενη σχέση, η οποία αποτελείται από ένα άθροισμα συνδυαστικών όρων, μπορεί να επιλυθεί (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) ως προς την άγνωστη παράμετρο α με τη χρήση ενός μαθηματικού υπολογιστικού προγράμματος και τελικά να καθοριστούν τα δύο ο.ε. του συγκεκριμένου μη–παραμετρικού δ.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(m-\alpha+1;m)}$. Αν ληφθεί υπ’όψιν και η προηγούμενη σχέση $P_C(W_{s+1} = w) = P_C(W_{s+1} = m–w)$, τότε συνεπάγεται ότι ισχύει

$$\sum_{w=0}^{a-1} P_C(W_{s+1}=w) = \sum_{w=0}^{a-1} P_C(W_{s+1}=m-w),$$

ή ισοδύναμα,

$$\sum_{w=0}^{a-1} P_C(W_{s+1}=w) = \sum_{w=m-a+1}^{m} P_C(W_{s+1}=w),$$

δηλαδή συνεπάγεται ότι

$$P_C(0 \leq W_{s+1} \leq \alpha–1) = P_C(m–\alpha+1 \leq W_{s+1} \leq m).$$

Όμως, ισχύει και η σχέση

$$P_C(0 \leq W_{s+1} \leq \alpha–1) + P_C(\alpha \leq W_{s+1} \leq m–\alpha) + P_C(m–\alpha+1 \leq W_{s+1} \leq m) = 1,$$

οπότε με συνδυασμό αυτών των δύο τελευταίων σχέσεων προκύπτει ότι

$$P_C(0 \leq W_{s+1} \leq \alpha–1) = \frac{1-P_C(a \leq W_{s+1} \leq m-a)}{2} \qquad (1)$$

Δεδομένου όμως ότι ισχύει

$$\sum_{w=a}^{m-a} P_C\,(W_{s+1} = w) \geq 1 - p_0 \;\Leftrightarrow P_C(\alpha \leq W_{s+1} \leq m\text{–}\alpha) \geq 1 - p_0,$$

με τη βοήθεια της (1), καταλήγουμε τελικά στην ανισότητα

$$P_C(0 \leq W_{s+1} \leq \alpha\text{–}1) \leq \frac{p_0}{2},$$

δηλαδή ισχύει επίσης και η παρακάτω σχέση:

$$\sum_{w=0}^{a-1} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \leq \frac{p_0}{2}.$$

Ανακαλώντας ότι η πιθανότητα p_0 είναι (όπως ορίστηκε), η μέγιστη επιθυμητή πιθανότητα λανθασμένου συναγερμού, δηλαδή ότι $p_0 = maxFAR$, η προηγούμενη ανισότητα μπορεί να γραφεί στην εξής μορφή:

$$\sum_{w=0}^{a-1} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \leq \frac{\max FAR}{2}.$$

Έτσι, για δοθέντα m, s και FAR, η αμέσως προηγούμενη σχέση, η οποία αποτελείται από ένα άθροισμα συνδυαστικών όρων, μπορεί να επιλυθεί αριθμητικά ως προς την άγνωστη παράμετρο α και τελικά να καθοριστούν τα δύο ο.ε. του συγκεκριμένου μη–παραμετρικού δ.ε., $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)} = X_{(m-\alpha+1;m)}$.

2.1.4 Ένα αριθμητικό παράδειγμα

Για να εφαρμοστούν και να κατανοηθούν οι έννοιες που αναλύθηκαν παραπάνω, δίνεται εδώ (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) ένα σχετικά απλουστευμένο αριθμητικό παράδειγμα. Έστω ότι $m = 50$, $s = 2$, $n = 5$ και $j = 3$, δηλαδή το μέγεθος του τ.δ. αναφοράς είναι ίσο με 50 και το κοινό μέγεθος των νέων τ.δ. που συλλέγονται, τα οποία είναι ανεξάρτητα μεταξύ τους και ανεξάρτητα επίσης από το τ.δ. αναφοράς, είναι ίσο με 5. Προφανώς, η δειγματική διάμεσος των νέων τ.δ., θα είναι η τρίτη διατεταγμένη παρατήρηση $Y_{(3;5)}$.

Έστω επίσης ότι η μέγιστη επιθυμητή πιθανότητα λανθασμένου συναγερμού *maxFAR* για το συγκεκριμένο δίπλευρο μη–παραμετρικό δ.ε. έχει καθοριστεί να είναι το πολύ 1%. Με άλλα λόγια, έχει καθοριστεί ότι, η μέγιστη πιθανότητα η δειγματική διάμεσος $Y_{(3;5)}$ να πάρει μια τιμή έξω από τα δύο ο.ε. $LCL = X_{(\alpha;50)}$ και $UCL = X_{(50-\alpha+1;50)}$, ενώ η διεργασία παραμένει εντός στατιστικού ελέγχου, είναι ίση με 0.01. Τότε, σύμφωνα με την παραπάνω ανάλυση, θα πρέπει να λυθεί αριθμητικά ως προς α η σχέση

$$\sum_{w=0}^{\alpha-1} \frac{\binom{2+w}{w}\binom{52-w}{50-w}}{\binom{55}{50}} \leq 0.005,$$

από την οποία προκύπτει ότι η παράμετρος α θα είναι ίση με $\alpha = 3$, επομένως θα ισχύει $50 - \alpha + 1 = 48$. Άρα, τα ζητούμενα δύο ο.ε. θα είναι ίσα με $LCL = X_{(3;50)}$, δηλαδή η τρίτη διατεταγμένη παρατήρηση από το τ.δ. αναφοράς και $UCL = X_{(48;50)}$, δηλαδή η τεσσαρακοστή όγδοη διατεταγμένη παρατήρηση από το τ.δ. αναφοράς. Με τις τιμές αυτές, θα συγκρίνονται οι δειγματικές διάμεσοι $Y_{(3;5)}$ των νέων τ.δ. που συλλέγονται, τα οποία θα είναι μεγέθους n = 5 παρατηρήσεων το καθένα. Σε αυτήν την περίπτωση, λόγω της συμμετρικότητας της σ.π. της τ.μ. W_3, η ακριβής πιθανότητα λανθασμένου συναγερμού *FAR* θα είναι ίση με:

$$FAR = P(Y_{(3;5)} \notin [X_{(3;50)}, X_{(48;50)}]) = P_C(Y_{(3;5)} < X_{(3;50)}) + P_C(Y_{(3;5)} > X_{(48;50)}) =$$

$$= P_C(0 \leq W_3 \leq 2) + P_C(48 \leq W_3 \leq 50) = P_C(0 \leq W_3 \leq 2) + P_C(0 \leq W_3 \leq 2) =$$

$$2\,P_C(0 \leq W_3 \leq 2) = 2\sum_{w=0}^{2} P(W_3 = w) = 2\sum_{w=0}^{2} \frac{\binom{2+w}{w}\binom{52-w}{50-w}}{\binom{55}{50}} = 2(0.0036) = 0.0072 = 0.72\ \%.$$

Η ακριβής πιθανότητα λανθασμένου συναγερμού *FAR*, προφανώς, θα είναι μικρότερη από τη μέγιστη πιθανότητα λανθασμένου συναγερμού *maxFAR*, που είχε καθοριστεί αρχικά στο επίπεδο 0.01 = 1 %.

2.2 Μονόπλευρα μη–παραμετρικά δ.ε. τύπου Shewhart

Για να διαπιστωθεί αν μία διεργασία βρίσκεται εντός ή εκτός στατιστικού ελέγχου, χωρίς να υποτεθεί κάποια συγκεκριμένη (συνεχής) κατανομή για τα δεδομένα και χρησιμοποιώντας μόνο τα διατεταγμένα και ανεξάρτητα μεταξύ τους τ.δ. που συλλέγονται, μπορούν να χρησιμοποιηθούν επίσης κατάλληλα μονόπλευρα μη–παραμετρικά δ.ε., αντί των δίπλευρων διαγραμμάτων που παρουσιάστηκαν στην Παράγραφο 2.1.. Σε αυτά τα δ.ε., υπάρχει μόνο ένα (και όχι δύο) ο.ε., δηλαδή είτε το κ.ο.ε. LCL, είτε το α.ο.ε. UCL, τα οποία είναι πάλι κάποιες διατεταγμένες παρατηρήσεις, που υπολογίζονται με βάση το διατεταγμένο τ.δ. αναφοράς μεγέθους m. Η δειγματική διάμεσος $Y_{(j;n)}$, που υπολογίζεται από κάθε νέο τ.δ. μεγέθους n (δηλαδή τα ανεξάρτητα τ.δ. που συλλέγονται είναι ισοπληθή μεταξύ τους), συγκρίνεται (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) με το ο.ε. και αν είναι μεγαλύτερη από το α.ο.ε. $UCL = X_{(b;m)}$, στην περίπτωση που υπάρχει μόνο το α.ο.ε. UCL, ή μικρότερη από το κ.ο.ε. $LCL = X_{(\alpha;m)}$, στην περίπτωση που υπάρχει μόνο το κ.ο.ε. LCL, η διεργασία θεωρείται ότι βρίσκεται εκτός στατιστικού ελέγχου και θα πρέπει να αναζητηθούν ειδικές αιτίες μεταβλητότητας (βλέπε Αντζουλάκος (2006)). Σε αυτήν την περίπτωση, παρουσιάζει ιδιαίτερο ενδιαφέρον η κατεύθυνση και το πρόσημο της μεταβολής, δηλαδή αν η διεργασία μετατοπίζεται σε υψηλότερες τιμές ή σε χαμηλότερες τιμές σε σύγκριση με τις εντός–ελέγχου τιμές των αρχικών παρατηρήσεων. Συνήθως, για λόγους πρακτικής ευκολίας και για να είναι πιο απλός ο υπολογισμός της δειγματικής διαμέσου $Y_{(j;n)}$, το μέγεθος n των νέων ανεξάρτητων τ.δ. που συλλέγονται καθορίζεται να είναι περιττός αριθμός, δηλαδή να ισχύει $n = 2s+1$, οπότε η δειγματική διάμεσος $Y_{(j;n)}$ θα είναι η $(s+1)$–διατεταγμένη παρατήρηση $Y_{(s+1;n)}$.

Για την πιθανότητα λανθασμένου συναγερμού FAR_1 σε αυτό το μονόπλευρο δ.ε. Shewhart, αποδεικνύεται ότι ισχύει η σχέση

$$FAR_1 = \frac{FAR_2}{2},$$

όπου FAR_2 είναι η πιθανότητα λανθασμένου συναγερμού για το αντίστοιχο δίπλευρο δ.ε. Shewhart. Η σχέση αυτή μπορεί να αποδειχθεί σχετικά εύκολα, αν ληφθεί υπ' όψιν ότι στο δίπλευρο διάγραμμα ισχύει

$$P_C(Y_{(s+1;n)} < X_{(\alpha;m)}) = P_C(Y_{(s+1;n)} > X_{(m-\alpha+1;m)}),$$

λόγω της συμμετρικότητας που παρουσιάζει η σ.π. της διακριτής τ.μ. W_{s+1}. Επομένως, θα ισχύει επίσης

$$P_C(Y_{(s+1;n)} < X_{(\alpha;m)}) = P_C(Y_{(s+1;n)} > X_{(m-\alpha+1;m)}) = \frac{P_C(Y_{(s+1;n)} < X_{(\alpha;m)}) + P_C(Y_{(s+1;n)} > X_{(m-\alpha+1;m)})}{2}.$$

Όμως, για το μονόπλευρο δ.ε. Shewhart, έχουμε

$$P_C(Y_{(s+1;n)} < X_{(\alpha;m)}) = P_C(Y_{(s+1;n)} > X_{(m-\alpha+1;m)}) = FAR_1,$$

ενώ για το αντίστοιχο δίπλευρο διάγραμμα, ισχύει

$$P_C(Y_{(s+1;n)} < X_{(\alpha;m)}) + P_C(Y_{(s+1;n)} > X_{(m-\alpha+1;m)}) = FAR_2.$$

Από τις τελευταίες σχέσεις, προκύπτει άμεσα ότι οι πιθανότητες λανθασμένου συναγερμού FAR_1 και FAR_2 για το μονόπλευρο και το δίπλευρο αντίστοιχα μη– παραμετρικό δ.ε. τύπου Shewhart, θα συνδέονται μέσω της σχέσης

$$FAR_1 = \frac{FAR_2}{2}.$$

2.2.1 Η διαδικασία υπολογισμού των ο.ε. για τη συμμετρική περίπτωση

Από τη σχέση $FAR_2 = 2\,FAR_1$ προκύπτει ότι, αν στο μονόπλευρο μη–παραμετρικό δ.ε. Shewhart η πιθανότητα λανθασμένου συναγερμού είναι αρχικά γνωστή, προκαθορισμένη και ίση με μία επιθυμητή αρκετά μικρή τιμή, η αντίστοιχη πιθανότητα λανθασμένου συναγερμού στο δίπλευρο διάγραμμα Shewhart θα είναι ακριβώς η διπλάσια. Στο μονόπλευρο δ.ε., η διαδικασία υπολογισμού της παραμέτρου α (η παράμετρος b για το α.ο.ε. υπολογίζεται άμεσα στη συμμετρική περίπτωση, όπου ισχύει $b = m-\alpha+1$) είναι ακριβώς η ίδια με την αντίστοιχη διαδικασία για το δίπλευρο δ.ε.. Η μόνη διαφορά, εντοπίζεται στις δύο πιθανότητες λανθασμένου συναγερμού *FAR*. Αν είναι γνωστή η *FAR* για το μονόπλευρο δ.ε., τότε αυτή η πιθανότητα διπλασιάζεται (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)), ώστε να υπολογιστεί η αντίστοιχη *FAR* για το δίπλευρο δ.ε.. Γνωρίζοντας λοιπόν ότι, στο δίπλευρο διάγραμμα, η παράμετρος *α* υπολογίζεται από τη λύση της ανίσωσης

$$\sum_{w=0}^{a-1} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \leq \frac{FAR_2}{2}$$

ή ισοδύναμα, από τη λύση της ανίσωσης

$$\sum_{w=m-a+1}^{m} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \leq \frac{FAR_2}{2},$$

συμπεραίνουμε ότι, αν στο μονόπλευρο δ.ε. Shewhart είναι γνωστή η πιθανότητα λανθασμένου συναγερμού FAR_1, για να προσδιοριστεί η παράμετρος *α*, αρκεί να λύσουμε ως προς *α* την ανίσωση

$$\sum_{w=0}^{a-1} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \leq FAR_1,$$

ή εναλλακτικά ως προς *α* την ανίσωση

$$\sum_{w=m-a+1}^{m} \frac{\binom{s+w}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \leq FAR_1.$$

Αν στο δ.ε. υπάρχει μόνο το α.ο.ε. *UCL*, αυτό θα ισούται με την (m–α+1)–διατεταγμένη παρατήρηση $X_{(m-\alpha+1;m)}$, ενώ αν στο δ.ε. υπάρχει μόνο το κ.ο.ε. *LCL*, αυτό θα ισούται με την *α*–διατεταγμένη παρατήρηση $X_{(\alpha;m)}$.

Προφανώς, οι παραπάνω ανισώσεις δε μπορούν να επιλυθούν αναλυτικά με τις γνωστές μαθηματικές μεθόδους, μπορούν όμως να επιλυθούν με χρήση ενός μαθηματικού προγράμματος στον υπολογιστή.

Ας σημειωθεί στο σημείο αυτό ότι, οι δύο προηγούμενες ανισώσεις, ουσιαστικά είναι μόνο μία, αφού λόγω της συμμετρικότητας που παρουσιάζει η σ.π. της διακριτής τ.μ. W_{s+1}, θα ισχύει

$$P_C(W_{s+1} = w) = P_C(W_{s+1} = m - w), \text{ για } w = 0, 1, 2, \ldots, m,$$

οπότε αθροίζοντας για όλα τα w, προκύπτει ότι

$$\sum_{w=0}^{a-1} P_C(W_{s+1} = w) = \sum_{w=0}^{a-1} P_C(W_{s+1} = m - w) \Leftrightarrow \sum_{w=0}^{a-1} P_C(W_{s+1} = w) = \sum_{w=m-a+1}^{m} P_C(W_{s+1} = w)$$

και έτσι οι δύο ανισώσεις γίνονται μία.

2.2.2 Ένα αριθμητικό παράδειγμα

Για να εφαρμοστούν και να κατανοηθούν όλες οι έννοιες που αναλύθηκαν, δίνεται (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) ένα σχετικά απλουστευμένο αριθμητικό παράδειγμα, με συγκεκριμένες τιμές για όλες τις παραμέτρους. Έστω ότι ισχύει $m = 75$, $s = 7$, $n = 15$ και $j = 8$, δηλαδή το μέγεθος του αρχικού τ.δ. είναι ίσο με 75 και το μέγεθος των νέων τ.δ. που συλλέγονται, τα οποία είναι ανεξάρτητα μεταξύ τους και ανεξάρτητα επίσης από το αρχικό τ.δ. είναι ίσο με 15, επομένως η δειγματική διάμεσος των νέων τ.δ. θα είναι η όγδοη διατεταγμένη παρατήρηση $Y_{(8;15)}$. Ακόμη, το μονόπλευρο μη–παραμετρικό δ.ε. Shewhart, περιέχει μόνο το α.ο.ε. *UCL*, δηλαδή μας ενδιαφέρει μόνο η μετατόπιση της διεργασίας προς υψηλότερες τιμές των παρατηρήσεων από τις αρχικές εντός–ελέγχου τιμές τους. Έστω επίσης ότι, η μέγιστη επιθυμητή πιθανότητα λανθασμένου συναγερμού *maxFAR* για το συγκεκριμένο μονόπλευρο δ.ε., έχει καθοριστεί να είναι το πολύ 0.0027 = 0.27%. Με άλλα λόγια, έχει καθοριστεί ότι όσο η διεργασία παραμένει εντός ελέγχου, η μέγιστη πιθανότητα η δειγματική διάμεσος $Y_{(8;15)}$ να πάρει μία τιμή μεγαλύτερη από το α.ο.ε. $UCL = X_{(75-\alpha+1;75)}$ θα είναι ίση με 0.0027. Τότε, σύμφωνα με όλη την παραπάνω ανάλυση, μετά από την επίλυση κατάλληλης ανίσωσης ως προς α, βρέθηκε ότι η παράμετρος α θα πρέπει να είναι ίση με $\alpha = 12$, επομένως θα ισχύει $75-\alpha+1 = 64$. Άρα, το ζητούμενο α.ο.ε. θα είναι ίσο με $UCL = X_{(64;75)}$, δηλαδή θα είναι η εξηκοστή τέταρτη διατεταγμένη παρατήρηση από το αρχικό διατεταγμένο τ.δ. και με αυτήν την τιμή θα συγκρίνονται οι δειγματικές διάμεσοι $Y_{(8;15)}$ των νέων τ.δ. που συλλέγονται, τα οποία θα είναι μεγέθους n = 15 παρατηρήσεων το καθένα. Σε αυτήν την περίπτωση, η ακριβής πιθανότητα λανθασμένου συναγερμού FAR_1 σε αυτό το μονόπλευρο μη–παραμετρικό δ.ε., θα είναι ίση με:

$$FAR_1 = P_C(Y_{(8;15)} > UCL) = P_C(Y_{(8;15)} > X_{(64;75)}) = P_C(64 \le W_8 \le 75) = \sum_{w=64}^{75} P_C(W_8 = w) =$$

$$= \sum_{w=64}^{75} \frac{\binom{7+w}{w}\binom{82-w}{75-w}}{\binom{90}{75}} = 0.002517 = 0.2517\ \%.$$

Η ακριβής πιθανότητα λανθασμένου συναγερμού FAR_1, προφανώς, θα είναι μικρότερη από τη μέγιστη επιθυμητή πιθανότητα λανθασμένου συναγερμού $maxFAR$ που είχε προκαθοριστεί αρχικά, η οποία ήταν ίση με $maxFAR$ = 0.0027 = 0.27%. Σε αυτό το παράδειγμα, αν το μονόπλευρο δ.ε. Shewhart περιέχει μόνο το κ.ο.ε. LCL, δηλαδή αν εξετάζεται και παρακολουθείται η μετατόπιση της διεργασίας μόνο προς χαμηλότερες τιμές από τις αρχικές εντός–ελέγχου τιμές των παρατηρήσεων, τότε το κ.ο.ε. LCL θα ήταν η α–διατεταγμένη παρατήρηση $X_{(\alpha;75)}$ από το τ.δ. αναφοράς και με αυτήν την παρατήρηση θα συγκρίνονταν οι δειγματικές διάμεσοι $Y_{(8;15)}$ των νέων τ.δ. που θα συλλέγονταν, τα οποία θα ήταν μεγέθους n = 15 παρατηρήσεων το καθένα. Αν η μέγιστη επιθυμητή πιθανότητα λανθασμένου συναγερμού $maxFAR$ για το συγκεκριμένο μονόπλευρο δ.ε. Shewhart είχε καθοριστεί να είναι το πολύ 0.0027 = 0.27%, μετά από την επίλυση κατάλληλης ανίσωσης ως προς α, θα προέκυπτε η εκτίμηση α = 12. Επομένως, το κ.ο.ε. LCL θα ήταν η δωδέκατη διατεταγμένη παρατήρηση $X_{(12;75)}$ από το αρχικό τ.δ. και με αυτήν την παρατήρηση θα συγκρίνονταν οι δειγματικές διάμεσοι $Y_{(8;15)}$ των νέων τ.δ. που θα συλλέγονταν. Σε αυτήν την περίπτωση, η ακριβής πιθανότητα λανθασμένου συναγερμού FAR_1 θα ήταν ίση με:

$$FAR_1 = P_C(Y_{(8;15)} < LCL) = P_C(Y_{(8;15)} < X_{(12;7)}) = P_C(0 \le W_8 \le 11) = \sum_{w=0}^{11} P_C(W_8 = w) =$$

$$= \sum_{w=0}^{11} \frac{\binom{7+w}{w}\binom{82-w}{75-w}}{\binom{90}{75}} = 0.002517 = 0.2517\ \%,$$

η οποία είναι ακριβώς η ίδια με την αμέσως προηγούμενη πιθανότητα λανθασμένου συναγερμού FAR_1 για το κ.ο.ε. *LCL*. Αυτό είναι μάλλον αναμενόμενο, λόγω της συμμετρικότητας (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) που παρουσιάζει η σ.π. της διακριτής τ.μ. W_8. Και σε αυτήν την περίπτωση, η ακριβής πιθανότητα λανθασμένου συναγερμού FAR_1 θα είναι μικρότερη από τη μέγιστη επιθυμητή πιθανότητα λανθασμένου συναγερμού ***maxFAR*** που είχε αρχικά καθοριστεί, η οποία ήταν ίση με ***maxFAR*** = 0.0027 = 0.27%.

ΚΕΦΑΛΑΙΟ 3

Η κατανομή του μήκους ροής στο δίπλευρο μη–παραμετρικό δ.ε. τύπου Shewhart

3.1 Το πρόβλημα της μη ανεξαρτησίας των στατιστικών συναρτήσεων ελέγχου

Όπως αναφέρθηκε σε προηγούμενη ενότητα, στο δίπλευρο μη–παραμετρικό δ.ε. τύπου Shewhart που εξετάσαμε, η πιθανότητα λανθασμένου συναγερμού FAR_2 θα είναι ίση με

$FAR_2 = P(Y_{(j;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]) = 1 - P(Y_{(j;n)} \in [X_{(\alpha;m)}, X_{(b;m)}]) =$

$= 1 - P(X_{(\alpha;m)} \leq Y_{(j;n)} \leq X_{(b;m)})$,

όπου $Y_{(j;n)}$ είναι η j–διατεταγμένη παρατήρηση από κάποιο τελικό τ.δ. που συλλέγεται. Όμως, σε αυτό το δ.ε. Shewhart, παρουσιάζεται το εξής φαινόμενο: παρόλο που τα νέα τ.δ. που συλλέγονται είναι ανεξάρτητα μεταξύ τους, ωστόσο τα ενδεχόμενα $\{Y_{(k;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]\}$ και $\{Y_{(l;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]\}$ δεν είναι στατιστικά ανεξάρτητα μεταξύ τους (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)). Ακριβώς το ίδιο ισχύει και για τα ενδεχόμενα $\{Y_{(k,n)} \in [X_{(\alpha;m)}, X_{(b;m)}]\}$ και $\{Y_{(l,n)} \in [X_{(\alpha;m)}, X_{(b;m)}]\}$, όπου $Y_{(k;n)}$ και $Y_{(l;n)}$ είναι παρατηρήσεις από δύο διαφορετικά, ανεξάρτητα μεταξύ τους τ.δ.. Η μη ανεξαρτησία των παραπάνω ενδεχομένων που γράφτηκαν, οφείλεται στο ότι τα ο.ε. στο συγκεκριμένο μη–παραμετρικό δ.ε. Shewhart δεν είναι σταθερές και γνωστές ποσότητες, αλλά είναι τυχαίες μεταβλητές (είναι οι διατεταγμένες τυχαίες μεταβλητές $X_{(a;m)}$ και $X_{(b;m)}$ από το τ.δ. αναφοράς).

Επομένως, στην περίπτωση αυτή, προκύπτει το συμπέρασμα ότι η διακριτή κατανομή του συνολικού αριθμού των σημείων που απεικονίζονται στο διάγραμμα, έως ότου εμφανιστεί για πρώτη φορά κάποιο σημείο που θα βρίσκεται εκτός των δύο ο.ε. δε θα είναι η γεωμετρική κατανομή (όπως στα κλασικά παραμετρικά δ.ε. τύπου Shewhart), αφού η γεωμετρική κατανομή απαιτεί την υπόθεση της πλήρους ανεξαρτησίας μεταξύ όλων των ενδεχομένων του δειγματικού χώρου (βλέπε Κούτρας (2002) (α)), υπόθεση η οποία δεν ικανοποιείται για τα σημεία που απεικονίζονται στο συγκεκριμένο δίπλευρο διάγραμμα. Λόγω του παραπάνω προβλήματος, κρίνεται απαραίτητη η χρήση κάποιας εναλλακτικής ειδικής τεχνικής δέσμευσης για την εύρεση της κατανομής που μας ενδιαφέρει.

3.2 Η ακριβής σ.π. του μήκους ροής

Έστω N η διακριτή, θετική, ακέραιη τ.μ., η οποία δηλώνει το συνολικό αριθμό των σημείων που απεικονίζονται στο δίπλευρο μη–παραμετρικό δ.ε. τύπου Shewhart έως ότου εμφανιστεί για πρώτη φορά ένα σημείο το οποίο θα βρίσκεται εκτός των δύο ο.ε. LCL και UCL, ενώ στην πραγματικότητα η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Τότε, όπως αναφέρθηκε και παραπάνω, η κατανομή της διακριτής τ.μ. N δε θα είναι η γεωμετρική, αφού τα σημεία που απεικονίζονται στο δίπλευρο δ.ε. δεν είναι ανεξάρτητα μεταξύ τους (ο λόγος για αυτό αναφέρθηκε προηγουμένως).

Όμως, για δοθέντα ο.ε. LCL και UCL, δηλαδή αν είναι αρχικά γνωστά τα δύο ο.ε. LCL και UCL, η δεσμευμένη κατανομή της τ.μ. N θα είναι η γεωμετρική κατανομή και στην περίπτωση αυτή όλα τα σημεία που απεικονίζονται στο διάγραμμα θα είναι τελείως ανεξάρτητα μεταξύ τους.

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου και συμβολίσουμε με p τη δεσμευμένη πιθανότητα ένα σημείο που απεικονίζεται στο διάγραμμα να βρίσκεται μέσα στα δύο ο.ε. LCL και UCL, δοθέντων όμως αυτών των δύο ορίων (δηλαδή γνωρίζοντας αρχικά τις τιμές που θα πάρουν οι δύο ποσότητες LCL και UCL), τότε (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) θα έχουμε:

$$p = P(X_{(\alpha;m)} \leq Y_{(j;n)} \leq X_{(b;m)} \mid X_{(\alpha;m)} = x, X_{(b;m)} = y) =$$

$$= P(F(X_{(\alpha;m)}) \leq F(Y_{(j;n)}) \leq F(X_{(b;m)}) \mid F(X_{(\alpha;m)}) = F(x), F(X_{(b;m)}) = F(y))$$

$$= P(U_{(\alpha;m)} \leq U_{(j;n)} \leq U_{(b;m)} \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = P(s \leq U_{(j;n)} \leq t \mid U_{(\alpha;m)} = s, U_{(b;m)} = t),$$

όπου F είναι η εντός–ελέγχου α.σ.κ. της τ.μ. X,

$$s = F(x) = P(X \leq x), t = F(y) = P(X \leq y) \text{ και } 0 < s < t < 1.$$

Το παραπάνω αποτέλεσμα προκύπτει, αν λάβουμε υπ'όψιν (βλέπε Στέγγος (2006)) ότι η (συνεχής) τ.μ. $U = F(X)$ ακολουθεί την ομοιόμορφη κατανομή στο διάστημα (0,1) και επίσης ότι η α.σ.κ. F είναι μία συνάρτηση γνησίως αύξουσα, οπότε «διατηρεί» την ανισότητα $\leq$.

Όμως, η τ.μ. $U_{(j;n)} = F(Y_{(j;n)})$, προέρχεται από το νέο τ.δ. μεγέθους n, το οποίο είναι ανεξάρτητο από το τ.δ. αναφοράς. Επομένως η (συνεχής) τ.μ. $U_{(j;n)} = F(Y_{(j;n)})$ θα είναι ανεξάρτητη από τις (συνεχείς) τ.μ. $U_{(\alpha;m)} = F(X_{(\alpha;m)})$ και $U_{(b;m)} = F(X_{(b;m)})$, άρα η δέσμευση στην προηγούμενη πιθανότητα p παύει να υπάρχει (βλέπε Κούτρας (2002) (β)), με αποτέλεσμα να μπορούμε να γράψουμε

$p = P(X_{(\alpha;m)} \leq Y_{(j;n} \leq X_{(b;m)} \mid X_{(\alpha;m)} = x, X_{(b;m)} = y) =$

$$= P(s \leq U_{(j;n)} \leq t \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = P(s \leq U_{(j;n)} \leq t) = \int_s^t f_{U_{(j;n)}}(w)dw,$$

όπου $0 < s < t < 1$. Όμως, η τ.μ. $U_{(j;n)}$ ακολουθεί (βλέπε Στέγγος (2006)) την κατανομή Βήτα με παραμέτρους j και $n\text{–}j+1$. Άρα, για την σ.π.π. $f_{U_{(j;n)}}$ της τ.μ. $U_{(j;n)}$, θα ισχύει (βλέπε Κούτρας (2002) (α))

$$f_{U_{(j;n)}}(w) = \frac{1}{B(j,n-j+1)} w^{j-1}(1-w)^{(n-j+1)-1} = \frac{n!}{(j-1)!\,(n-j)!}\int_s^t w^{j-1}(1-w)^{n-j}\,dw\ ,$$

$w \in (0,1)$ και τελικά η πιθανότητα $p(s,t)$ θα πάρει τη μορφή

$$p(s,t) = \int_s^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j}\,dw,\ 0 < s < t < 1.$$

Ο υπολογισμός της σ.π. της διακριτής τ.μ. N, μπορεί να γίνει (βλέπε Κούτρας (2002) (β)) με τη χρήση του Θ.Ο.Π., δεσμεύοντας (συνεχώς) ως προς τις τ.μ. $U_{(\alpha;m)}$ και $U_{(b;m)}$, οι οποίες ακολουθούν την κατανομή Βήτα και παίρνουν τιμές στο διάστημα (0,1). Οι τ.μ. $U_{(\alpha;m)}$ και $U_{(b;m)}$, ουσιαστικά αντιπροσωπεύουν τα δύο ο.ε. του διαγράμματος $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$, αφού ισχύει $U_{(\alpha;m)} = F(X_{(\alpha;m)})$ και $U_{(b;m)} = F(X_{(b;m)})$, όπου $F(\ )$ είναι η (συνεχής) α.σ.κ. που ακολουθούν οι παρατηρήσεις, όσο η διεργασία βρίσκεται εντός στατιστικού ελέγχου.

Επίσης, όταν τα δύο ο.ε. LCL και UCL είναι γνωστές και σταθερές ποσότητες και όχι τ.μ., όλα τα σημεία που απεικονίζονται στο δίπλευρο δ.ε. θα είναι ανεξάρτητα μεταξύ τους και επομένως, σε αυτήν την περίπτωση (βλέπε Αντζουλάκος (2006)), η δεσμευμένη κατανομή της τ.μ. N, δοθέντων των τιμών των δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$, όταν η διεργασία είναι εντός στατιστικού ελέγχου, θα είναι η γεωμετρική κατανομή με παράμετρο (πιθανότητα «επιτυχίας») ίση με την ποσότητα $p(s,t)$ που αναφέρθηκε προηγουμένως.

Τελικά, όταν η διεργασία είναι εντός στατιστικού ελέγχου, για τη σ.π. της τ.μ. N, σύμφωνα με το Θ.Ο.Π., θα ισχύει:

$$P_C[N=k] = \int_{-\infty}^{\infty}\int_{-\infty}^{y} P(N=k \mid X_{(a;m)}=x, X_{(b;m)}=y)\, f_{X_{(a;m)},X_{(b;m)}}(x,y)\,dx\,dy =$$

$$= \int_{0}^{1}\int_{s}^{1} P(N=k \mid U_{(a;m)}=s, U_{(b;m)}=t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dt\,ds,$$

$k = 1, 2, \ldots$.

Η δεσμευμένη κατανομή της τ.μ. N, δοθέντων των τιμών των τ.μ. $X_{(\alpha;m)}$ και $X_{(b;m)}$, όταν η διεργασία είναι εντός στατιστικού ελέγχου, θα είναι (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) η γεωμετρική κατανομή με παράμετρο (πιθανότητα «επιτυχίας») ίση με $1-p(s,t)$, δηλαδή θα ισχύει

$$N \mid X_{(\alpha;m)} = x, X_{(b;m)} = y \sim Geometric(1-p(s,t))$$

και επομένως, θα μπορούμε να γράψουμε

$$P_C(N=k \mid X_{(\alpha;m)}=x, X_{(b;m)}=y) = P_C(N=k \mid U_{(\alpha;m)}=s, U_{(b;m)}=t) =$$

$$= [p(s,t)]^{k-1}[1-p(s,t)] = [p(s,t)]^{k-1} - [p(s,t)]^{k},$$

$k = 1, 2, \ldots$, όπου

$$p(s,t) = \frac{n!}{(j-1)!\,(n-j)!}\int_{s}^{t} w^{j-1}(1-w)^{n-j}\,dw\ , \ 0 < s < t < 1$$

και $f_{U_{(a;m)},U_{(b;m)}}(s,t)$ είναι η από κοινού σ.π.π. των τ.μ. $U_{(\alpha;m)}$ και $U_{(b;m)}$, η οποία θα δίνεται (βλέπε Κούτρας (2002) (β)) από τον τύπο

$$f_{U_{(a;m)},U_{(b;m)}}(s,t) =$$

$$\frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!} f_U(s)\, f_U(t)[F_U(s)]^{a-1}[F_U(t)-F_U(s)]^{b-a-1}[1-F_U(t)]^{m-b}$$

$$= \frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\ , \ 0 < s < t < 1,$$

όπου $U = F(X)$ είναι μία συνεχής τ.μ. (βλέπε Στέγγος (2006)) που ακολουθεί την ομοιόμορφη κατανομή στο διάστημα (0,1).

Συνδυάζοντας όλες τις παραπάνω σχέσεις, συνεπάγεται ότι η σ.π. της τ.μ. N, όταν η διεργασία είναι εντός στατιστικού ελέγχου, θα είναι ίση με:

$$P_C[N=k] =$$

$$\frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!}\int_{0}^{1}\int_{s}^{1}\left([p(s,t)]^{k-1} - [p(s,t)]^{k}\right)s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dt\,ds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 [p(s,t)]^{k-1} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

$$- \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 [p(s,t)]^{k} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds \ ,$$

$k = 1, 2, \ldots$ και αντικαθιστώντας στην τελευταία σχέση την έκφραση

$$p(s,t) = \frac{n!}{(j-1)!(n-j)!}\int_s^t w^{j-1}(1-w)^{n-j}\,dw \ ,$$

προκύπτει ότι:

$P_C[N = k] =$

$$\int_0^1\int_s^1 [\int_s^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j}\,dw]^{k-1} \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

$-$

$$\int_0^1\int_s^1 [\int_s^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j}\,dw]^{k} \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds \ .$$

Επίσης, αποδεικνύεται (βλέπε Κούτρας (2001)) ότι ισχύει και η γενική σχέση

$$\int_s^t w^{j-1}(1-w)^{n-j}\,dw = \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{i+j}(t^{i+j}-s^{i+j}) = \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{i+j}t^{i+j} - \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{i+j}s^{i+j}$$

και έτσι, προκύπτει ότι η ακριβής σ.π. του μήκους ροής στο δίπλευρο δ.ε. Shewhart, θα είναι η εξής:

$P_C[N = k] =$

$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 ([p(s,t)]^{k-1} - [p(s,t)]^{k}) s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds \ ,$$

όπου $p(s,t) =$

$$= \frac{n!}{(j-1)!(n-j)!}\int_s^t w^{j-1}(1-w)^{n-j}\,dw = \frac{n!}{(j-1)!(n-j)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}(t^{j+i}-s^{j+i}) \ ,$$

$0 < s < t < 1$ και $k = 1, 2, \ldots$.

Από το παραπάνω αποτέλεσμα, είναι μάλλον φανερό ότι, ο ακριβής υπολογισμός της σ.π. $P_C[N=k]$ της τ.μ. N είναι πρακτικά αρκετά δύσκολος, επειδή το ορισμένο διπλό ολοκλήρωμα

$$\int_0^1\int_s^1 [\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}(t^{j+i}-s^{j+i})]^k s^{a-1}(t-s)^{b-a-1}dtds$$

που εμφανίστηκε παραπάνω, είναι αδύνατο να υπολογιστεί επ’ακριβώς μέσω ενός μαθηματικού τύπου σε «κλειστή» μορφή. Το γεγονός αυτό, ίσως να αποτελεί ένα πρόβλημα στο δίπλευρο μη–παραμετρικό δ.ε. τύπου Shewhart.

Σημειώνεται εδώ (βλέπε Κούτρας (2002) (α)) ότι η ποσότητα $p(s,t)$ είναι το ολοκλήρωμα στο διάστημα $[s,t] \subseteq (0,1)$ της σ.π.π. μιας τ.μ. η οποία ακολουθεί την κατανομή Βήτα με παραμέτρους j και $n-j+1$. Επομένως το $p(s,t)$ εκφράζει την πιθανότητα $P(s \leq Y \leq t)$, όπου Y είναι μία συνεχής τ.μ. η οποία ακολουθεί την κατανομή Βήτα με παραμέτρους j και $n-j+1$ και προφανώς, θα ισχύει $0 < p(s,t) < 1$.

3.3 Ο υπολογισμός της πιθανότητας λανθασμένου συναγερμού

Όσον αφορά την πιθανότητα λανθασμένου συναγερμού FAR για το συγκεκριμένο δίπλευρο δ.ε. τύπου Shewhart, αυτή θα είναι ίση με:

$FAR = P(Y_{(j;n)} \notin [LCL, UCL]$|η διεργασία εντός ελέγχου) =

$= P(Y_{(j;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]$ | η διεργασία εντός ελέγχου)

$= 1 - P(Y_{(j;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]$ | η διεργασία εντός ελέγχου)

$= 1 - P(X_{(\alpha;m)} \leq Y_{(j;n)} \leq X_{(b;m)}$ | η διεργασία εντός ελέγχου).

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η πιθανότητα $P(X_{(\alpha;m)} \leq Y_{(j;n)} \leq X_{(b;m)})$ μπορεί να υπολογιστεί (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) εφαρμόζοντας το Θ.Ο.Π., δεσμεύοντας συνολικά ως προς όλες τις πιθανές (συνεχείς) τιμές που μπορούν να πάρουν τα δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$. Θα είναι λοιπόν:

$P(X_{(\alpha;m)} \leq Y_{(j;n)} \leq X_{(b;m)}) =$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{y} P(X_{(a;m)} \leq Y_{(j;n)} \leq X_{(b;m)} \mid X_{(a;m)} = x, X_{(b;m)} = y) f_{X_{(a;m)},X_{(b;m)}}(x,y)dxdy$$

$$= \int_0^1\int_s^1 P(U_{(a;m)} \leq U_{(j;n)} \leq U_{(b;m)} \mid U_{(a;m)} = s, U_{(b;m)} = t) f_{U_{(a;m)},U_{(b;m)}}(s,t)dtds$$

$$= \int_0^1 \int_s^1 P(U_{(a;m)} \le U_{(j;n)} \le U_{(b;m)} \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds$$

$$= \int_0^1 \int_s^1 P(s \le U_{(j;n)} \le t \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds$$

$$= \int_0^1 \int_s^1 P(s \le U_{(j;n)} \le t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds$$

$$= \int_0^1 \int_s^1 [\int_s^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j}\,dw]\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds$$

$$= \int_0^1 \int_s^1 [p(s,t)]\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 [p(s,t)] s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

Αναπτύσσοντας σε δυνάμεις (βλέπε Κούτρας (2001)) το διωνυμικό ανάπτυγμα των όρων $(t-s)^{b-a-1}$ και $(1-t)^{m-b}$ και ολοκληρώνοντας στη συνέχεια ως προς t και ως προς s όλες τις συναρτήσεις που εμφανίζονται, κατόπιν πράξεων, προκύπτει ότι

$P(X_{(\alpha;m)} \le Y_{(j;n)} \le X_{(b;m)} \mid$ η διεργασία είναι εντός ελέγχου$) =$

$$\frac{n!}{(j-1)!}\frac{m!}{(a-1)!}\sum_{i=0}^{n-j}\sum_{k=0}^{b-a-1}\sum_{l=0}^{m-b}\frac{(-1)^{i+b+l-a-k-1}}{i!(n-j-i)!k!(b-a-k-1)!l!(m-b-l)!(b-k-1)(i+l+j+b)(i-k+j+b-1)}$$

Επομένως, η πιθανότητα λανθασμένου συναγερμού FAR στο δίπλευρο δ.ε. Shewhart, θα είναι ίση με:

$FAR = 1 - P(X_{(\alpha;m)} \le Y_{(j;n)} \le X_{(b;m)} \mid$ η διεργασία είναι εντός ελέγχου$)) =$

$$1 - \frac{n!}{(j-1)!}\frac{m!}{(a-1)!}\sum_{i=0}^{n-j}\sum_{k=0}^{b-a-1}\sum_{l=0}^{m-b}\frac{(-1)^{i+b+l-a-k-1}}{i!(n-j-i)!k!(b-a-k-1)!l!(m-b-l)!(b-k-1)(i+l+j+b)(i-k+j+b-1)}.$$

Από τις παραπάνω σχέσεις, είναι μάλλον φανερό ότι η πιθανότητα *FAR* μπορεί να υπολογιστεί άμεσα, για δοθείσες τιμές των παραμέτρων *m*, *n*, *α*, *b* και *j*, μετά όμως από αρκετές αριθμητικές πράξεις.

Ο τελευταίος τύπος που γράφτηκε, είναι ταυτόσημος με τον τύπο στη σελίδα 11, που δίνει την πιθανότητα λανθασμένου συναγερμού *FAR* για το δίπλευρο μη–παραμετρικό δ.ε. Shewhart, μέσω της κατανομής της διακριτής τ.μ. W_j. Δηλαδή, οι δύο τύποι που δίνουν την *FAR*, θα ταυτίζονται μεταξύ τους και θα ισχύουν για κάθε πιθανή τιμή των θετικών ακέραιων παραμέτρων *n, m, j, a* και *b,* με $1 \le a < b \le m$ και $1 \le j \le n$.

3.4 Ο υπολογισμός του εντός–ελέγχου μέσου μήκους ροής ARL_{in}

Το εντός–ελέγχου μέσο μήκος ροής ARL_{in} για το δίπλευρο μη–παραμετρικό δ.ε. Shewhart, θα είναι (βλέπε Αντζουλάκος (2006)) η αναμενόμενη τιμή $E(N)$ της διακριτής θετικής τ.μ. *N*, η οποία δηλώνει τον συνολικό αριθμό σημείων που απεικονίζονται έως ότου εμφανιστεί για πρώτη φορά ένα σημείο το οποίο θα βρίσκεται εκτός των δύο ο.ε. *LCL* και *UCL*, ενώ η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Η διακριτή τ.μ. *N* παίρνει τις τιμές 1, 2, ... και από γνωστό θεώρημα της Θ.Π., (βλέπε Κούτρας (2002) (α)) προκύπτει ότι η αναμενόμενη τιμή $E(N)$ θα είναι ίση με το άπειρο άθροισμα

$$E(N) = \sum_{k=0}^{\infty} P(N > k).$$

Η πιθανότητα $P(N>k)$, σύμφωνα με το Θ.Ο.Π., δεσμεύοντας (βλέπε Κούτρας (2002) (β)) ως προς όλες τις (συνεχείς) τιμές που μπορούν να πάρουν τα δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$, θα είναι ίση με:

$$P(N > k) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} P(N > k \mid X_{(a;m)} = x, X_{(b;m)} = y)\, f_{X_{(a;m)},X_{(b;m)}}(x,y)dxdy =$$

$$= \int_{0}^{1}\int_{s}^{1} P(N > k \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)dtds$$

$k = 0, 1, \ldots$ (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)).

Όμως, η τ.μ. $N \mid U_{(\alpha;m)} = s, U_{(b;m)} = t$, ακολουθεί τη γεωμετρική κατανομή με παράμετρο $1 - p(s,t)$, όπου

$$p(s,t) = P(U_{(\alpha;m)} \le U_{(j;n)} \le U_{(b;m)} \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) =$$

$$= \frac{n!}{(j-1)!(n-j)!}\left[\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}t^{j+i} - \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}s^{j+i}\right],\ 0 < s < t < 1$$

και επειδή ισχύει (βλέπε Κούτρας (2002) (α))

$P(N > k \mid U_{(a;m)} = s, U_{(b;m)} = t) = [p(s,t)]^k$,

προκύπτει ότι:

$$P(N > k) = \int_0^1 \int_s^1 P(N > k \mid U_{(a;m)} = s, U_{(b;m)} = t) f_{U_{(a;m)},U_{(b;m)}}(s,t) dtds =$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 p(s,t)^k s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds ,$$

$k = 0, 1, \ldots$. Σύμφωνα με όλα τα παραπάνω, για το ARL_{in} θα ισχύει

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N > k) =$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \sum_{k=0}^{\infty} \int_0^1 \int_s^1 [p(s,t)]^k s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \sum_{k=0}^{\infty} [p(s,t)]^k s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \{\sum_{k=0}^{\infty} [p(s,t)]^k\} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds .$$

Επειδή η ποσότητα $p(s,t)$ είναι μία πιθανότητα και παίρνει τιμές στο διάστημα (0,1), η σειρά $\sum_{k=0}^{\infty} [p(s,t)]^k$ συγκλίνει και συνεπώς (βλέπε Thomas and Finney (1990)) θα ισχύει ο μαθηματικός τύπος

$$\sum_{k=0}^{\infty} [p(s,t)]^k = \frac{1}{1 - p(s,t)}, \ p(s,t) \in (0,1),$$

οπότε, αντικαθιστώντας την τελευταία σχέση στο παραπάνω διπλό ολοκλήρωμα, προκύπτει ότι

$$ARL_{in} = E(N) =$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \{\sum_{k=0}^{\infty} [p(s,t)]^k\} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \frac{1}{1 - p(s,t)} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds$$

$$= \frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!}\int_0^1\int_s^1 \frac{s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}}{1-\dfrac{n!}{(j-1)!}\displaystyle\sum_{i=0}^{n-j}\frac{(-1)^i\,(t^{j+i}-s^{j+i})}{i!\,(n-j-i)!}}dtds\ .$$

Το παραπάνω ορισμένο διπλό ολοκλήρωμα, δηλαδή το εντός–ελέγχου μέσο μήκος ροής ARL_{in}, λόγω της ιδιαίτερης μαθηματικής του πολυπλοκότητας, δεν είναι δυνατόν να απλοποιηθεί περισσότερο με τις γνωστές μαθηματικές μεθόδους και τεχνικές, μπορεί όμως να υπολογιστεί με χρήση ενός μαθηματικού προγράμματος σε υπολογιστή.

Ο υπολογισμός του ARL_{in} θα μπορούσε να γίνει και με έναν δεύτερο τρόπο, χρησιμοποιώντας γνωστό θεώρημα από τη Θ.Π. για δεσμευμένες μέσες τιμές. Συγκεκριμένα, ισχύει (βλέπε Κούτρας (2002) (β)) η σχέση

$ARL_{in} = E(N) = E(E(N \mid X_{(\alpha;m)}, X_{(b;m)})) = E(E(N \mid U_{(\alpha;m)}, U_{(b;m)}))$,

όπου $X_{(\alpha;m)}$ και $X_{(b;m)}$ είναι το κάτω και το άνω ο.ε., $U_{(\alpha;m)} = F(X_{(\alpha;m)})$, $U_{(b;m)} = F(X_{(b;m)})$ και F είναι η εντός–ελέγχου α.σ.κ. των τιμών των αρχικών παρατηρήσεων X_i, που συλλέχθηκαν ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου. Επίσης, όπως αναφέρθηκε προηγουμένως, η δεσμευμένη κατανομή $N \mid X_{(\alpha;m)} = x,\ X_{(b;m)} = y$, δηλαδή η δεσμευμένη κατανομή $N \mid U_{(\alpha;m)} = s,\ U_{(b;m)} = t$, θα είναι η γεωμετρική κατανομή $Geometric(1–p(s,t))$, όπου $p(s,t)$ είναι η πιθανότητα

$p(s,t) = 1 - FAR = P(U_{(\alpha;m)} \le U_{(j;n)} \le U_{(b;m)} \mid U_{(\alpha;m)} = s,\ U_{(b;m)} = t)$.

Επομένως, αφού η τ.μ. $N \mid U_{(\alpha;m)} = s,\ U_{(b;m)} = t$ ακολουθεί τη γεωμετρική κατάνομή με παράμετρο (πιθανότητα «επιτυχίας») ίση με $1–p(s,t)$, η δεσμευμένη μέση τιμή $E(N \mid U_{(\alpha;m)} = s,\ U_{(b;m)} = t)$ θα είναι ίση με τη μέση τιμή αυτής της κατανομής, δηλαδή (βλέπε Κούτρας (2002) (α)) θα ισχύει

$$E(N \mid U_{(\alpha;m)} = s,\ U_{(b;m)} = t) = \frac{1}{1-p(s,t)}\ ,\ p(s,t) \in (0,1)\,.$$

Άρα, η αναμενόμενη τιμή $E(N)$, σύμφωνα με όλα τα παραπάνω, θα είναι ίση με:

$ARL_{in} = E(N) = E(E(N \mid U_{(\alpha;m)}, U_{(b;m)})) =$

$$= \int_0^1\int_s^1 E(N \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)dtds = \int_0^1\int_s^1 \frac{1}{1-p(s,t)}\, f_{U_{(a;m)},U_{(b;m)}}(s,t)dtds$$

$$= \frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!}\int_0^1\int_s^1 \frac{1}{1-p(s,t)}\, s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\, dtds$$

$$= \frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!}\int_0^1\int_s^1 \frac{s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}}{1-\dfrac{n!}{(j-1)!\,(n-j)!}\displaystyle\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i\,(t^{i+j}-s^{i+j})}{i+j}}\,dtds$$

και έτσι, καταλήγουμε ξανά στον τύπο που βρήκαμε προηγουμένως.

3.5 Η σ.π. του μήκους ροής όταν η διεργασία είναι εκτός στατιστικού ελέγχου

Αν κάποια χρονική στιγμή η (συνεχής) κατανομή που ακολουθούν οι παρατηρήσεις που συλλέγονται αλλάξει και μετατοπιστεί είτε προς μία άλλη κατανομή είτε προς την ίδια αρχική κατανομή αλλά με διαφορετικές παραμέτρους, ή ισοδύναμα αν μετατοπιστεί η αρχική α.σ.κ. F των τιμών των παρατηρήσεων προς μία άλλη α.σ.κ. κατανομής G, τότε η διεργασία βρίσκεται εκτός στατιστικού ελέγχου και θα πρέπει (βλέπε Αντζουλάκος (2006)) να βρεθούν οι ειδικές αιτίες μεταβλητότητας οι οποίες προκάλεσαν αυτήν την αλλαγή. Σε αυτήν την περίπτωση φυσικά, η κατανομή του εκτός–ελέγχου μήκους ροής δε θα είναι ίδια με την κατανομή του εντός–ελέγχου μήκους ροής. Όταν η νέα, εκτός στατιστικού ελέγχου, κατανομή που ακολουθούν τα δεδομένα δεν είναι γνωστή ή έστω προσδιορίσιμη μέσω ενός «κλειστού» μαθηματικού τύπου, δεν είναι δυνατόν να υπολογιστεί η σ.π. του εκτός–ελέγχου μήκους ροής, ούτε μπορεί να υπολογιστεί με ακρίβεια το εκτός–ελέγχου μέσο μήκος ροής ARL_{out}, ούτε όμως και η πιθανότητα σφάλματος (λανθασμένης ένδειξης) τύπου 2 στο δίπλευρο μη–παραμετρικό δ.ε. Shewhart. Το γεγονός αυτό, αποτελεί γενικά, ένα σημαντικό μειονέκτημα των μη– παραμετρικών δ.ε. σε σύγκριση με τα κλασικά παραμετρικά δ.ε. τύπου Shewhart, αφού στα τελευταία (βλέπε Αντζουλάκος (2006)), η κατανομή που ακολουθούν οι παρατηρήσεις είναι η κανονική και κατά συνέπεια η ακριβής σ.π. του εκτός–ελέγχου μήκους ροής, το ARL_{out} και η πιθανότητα λανθασμένης ένδειξης στο δ.ε. μπορούν να υπολογιστούν, ή έστω να προσεγγιστούν ικανοποιητικά.

Έστω N_0 η διακριτή, θετική τ.μ. η οποία δηλώνει το συνολικό αριθμό σημείων που απεικονίζονται στο δίπλευρο μη–παραμετρικό δ.ε., έως ότου εμφανιστεί για πρώτη φορά ένα σημείο το οποίο θα βρίσκεται εκτός των δύο ο.ε. LCL και UCL, όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου. Τότε, η κατανομή της τ.μ. N_0 δε θα είναι η γεωμετρική, αφού (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) όλα τα σημεία που απεικονίζονται στο δίπλευρο δ.ε. δε θα είναι ανεξάρτητα μεταξύ τους. Όμως, για δοθέντα ο.ε. *LCL* και *UCL*, η δεσμευμένη κατανομή της τ.μ. N_0 θα είναι η γεωμετρική κατανομή και στην περίπτωση αυτή όλα τα σημεία που απεικονίζονται στο διάγραμμα θα είναι στατιστικά ανεξάρτητα μεταξύ τους. Πρέπει εδώ να σημειωθεί και να τονιστεί ότι, τα δύο ο.ε. *LCL* και *UCL*, δεν είναι κάποιες σταθερές ποσότητες, γνωστές εκ των προτέρων, αλλά είναι τ.μ. και το γεγονός αυτό κάνει τα σημεία που απεικονίζονται στο διάγραμμα να συσχετίζονται.

Αν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου και συμβολίσουμε με G τη νέα εκτός–ελέγχου α.σ.κ. των παρατηρήσεων και με F την αρχική εντός–ελέγχου α.σ.κ. των παρατηρήσεων, τότε η δεσμευμένη πιθανότητα p_0 για ένα σημείο που απεικονίζεται στο διάγραμμα να βρίσκεται μέσα στα δύο ο.ε. *LCL* και *UCL*, δοθέντων των τιμών των δύο αυτών ορίων, θα δίνεται από τον τύπο

$$
\begin{aligned}
p_0 &= P(X_{(\alpha;m)} \le Y_{(j;n)} \le X_{(b;m)} \mid X_{(\alpha;m)} = x, X_{(b;m)} = y) = \\
&= P(F(X_{(\alpha;m)}) \le F(Y_{(j;n)}) \le F(X_{(b;m)}) \mid F(X_{(\alpha;m)}) = F(x), F(X_{(b;m)}) = F(y)) = \\
&= P(U_{(\alpha;m)} \le F(Y_{(j;n)}) \le U_{(b;m)} \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = \\
&= P(F^{-1}(U_{(\alpha;m)}) \le Y_{(j;n)} \le F^{-1}(U_{(b;m)}) \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = \\
&= P(G(F^{-1}(U_{(\alpha;m)})) \le G(Y_{(j;n)}) \le G(F^{-1}(U_{(b;m)})) \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = \\
&= P(G(F^{-1}(U_{(\alpha;m)})) \le U_{(j;n)} \le G(F^{-1}(U_{(b;m)})) \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = \\
&= P(G(F^{-1}(s)) \le U_{(j;n)} \le G(F^{-1}(t)) \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = \\
&= P(G(F^{-1}(s)) \le U_{(j;n)} \le G(F^{-1}(t)) = p_0(s,t),
\end{aligned}
$$

όπου $s = F(x) = P(X \le x)$, $t = F(y) = P(X \le y)$, $0 < s < t < 1$ και $U = G(X)$ είναι μία (συνεχής) τ.μ. που ακολουθεί (βλέπε Στέγγος (2006)) την ομοιόμορφη κατανομή στο διάστημα (0,1). Επίσης, η α.σ.κ. F είναι γνησίως αύξουσα συνάρτηση, επομένως η ανισότητα $\le$ διατηρείται στην F. Όμως, η τ.μ. $U_{(j;n)} = G(Y_{(j;n)})$ προέρχεται από το τελικό τ.δ. μεγέθους n, το οποίο είναι ανεξάρτητο από το αρχικό τ.δ. μεγέθους m. Άρα, η τ.μ. $U_{(j;n)} = G(Y_{(j;n)})$, θα είναι ανεξάρτητη από τις τ.μ. $U_{(\alpha;m)} = F(X_{(\alpha,m)})$ και $U_{(b;m)} = F(X_{(b;m)})$, οπότε στην προηγούμενη πιθανότητα p_0 η δέσμευση παύει να υπάρχει, με αποτέλεσμα να μπορούμε να γράψουμε (βλέπε Κούτρας (2002) (β))

$$
\begin{aligned}
p_0 &= p_0(s,t) = P(G(F^{-1}(s)) \le U_{(j;n)} \le G(F^{-1}(t)) \mid U_{(\alpha;m)} = s, U_{(b;m)} = t, \text{ διεργασία εκτός ελέγχου}) \\
&= P(G(F^{-1}(s)) \le U_{(j;n)} \le G(F^{-1}(t))) = \int_{G(F^{-1}(s))}^{G(F^{-1}(t))} f_{U_{(j;n)}}(w)\,dw, \quad 0 < s < t < 1.
\end{aligned}
$$

Όπως αναφέρθηκε σε προηγούμενη ενότητα, η τ.μ. $U_{(j;n)}$ ακολουθεί (βλέπε Στέγγος (2006)) την κατανομή Βήτα με παραμέτρους j και $n-j+1$. Άρα, για την σ.π.π. $f_{U_{(j;n)}}$ της τ.μ. $U_{(j;n)}$ θα ισχύει (βλέπε Κούτρας (2002) (α))

$$f_{U_{(j;n)}}(w) = \frac{1}{B(j,n-j+1)} w^{j-1}(1-w)^{n-j} = \frac{n!}{(j-1)!\,(n-j)!} w^{j-1}(1-w)^{n-j}, \; w \in (0,1).$$

Λαμβάνοντας υπ'όψιν ότι (βλέπε Κούτρας (2001)), όπως αναφέρθηκε σε προηγούμενη ενότητα, ισχύει ο γενικός τύπος

$$\int_s^t w^{j-1}(1-w)^{n-j}dw = \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}(t^{j+i}-s^{j+i}) = \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}t^{j+i} - \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}s^{j+i},$$

προκύπτει ότι η πιθανότητα $p_0 = p_0(s,t)$ θα ισούται με:

$$p_0 = p_0(s,t) = \int_{G(F^{-1}(s))}^{G(F^{-1}(t))} f_{U(j,n)}(w)dw = \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}\{[G(F^{-1}(t))]^{j+i} - [G(F^{-1}(s))]^{j+i}\},$$

$0 < s < t < 1$.

Υπενθυμίζουμε ότι, η από κοινού σ.π.π. $f_{U_{(a;m)},U_{(b;m)}}$ των τ.μ. $U_{(a;m)}$ και $U_{(b;m)}$, όπως αναφέρθηκε σε προηγούμενη ενότητα, θα είναι ίση με:

$$f_{U_{(a;m)},U_{(b;m)}}(s,t) = \frac{m!}{(a-1)!\,(b-a-1)!\,(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}, \; 0 < s < t < 1.$$

Προφανώς, η πιθανότητα $p_0 = p_0(s,t)$, θα είναι η δεσμευμένη πιθανότητα σφάλματος τύπου 2 στο δίπλευρο μη–παραμετρικό δ.ε., δοθέντων των τιμών των δύο ο.ε. $LCL = X_{(a;m)}$ και $UCL = X_{(b;m)}$. Δηλαδή, η ποσότητα $p_0 = p_0(s,t)$, θα δηλώνει την πιθανότητα το δ.ε., λανθασμένα, να μη δώσει ένδειξη ότι η διεργασία είναι εκτός στατιστικού ελέγχου, όταν πράγματι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, δοθέντων των τιμών των δύο ο.ε. $LCL = X_{(a;m)}$ και $UCL = X_{(b;m)}$. Σε αυτό το σημείο, είναι μάλλον φανερό ότι η δεσμευμένη πιθανότητα $p_0 = p_0(s,t)$ είναι αδύνατο να υπολογιστεί ή έστω να προσεγγιστεί, αφού στην περίπτωση του συγκεκριμένου μη–παραμετρικού δ.ε., η αρχική α.σ.κ. F και η νέα α.σ.κ. G είναι άγνωστες, δηλαδή ο ακριβής συναρτησιακός τους τύπος δεν είναι γνωστός, ή έστω προσεγγίσιμος.

Ο υπολογισμός της σ.π. της τ.μ. N_0, δηλαδή ο υπολογισμός της πιθανότητας $P[N_0 = k]$, μπορεί να γίνει με τη χρήση του Θ.Ο.Π., δεσμεύοντας ως προς τις τ.μ. $U_{(\alpha;m)}$ και $U_{(b;m)}$, οι οποίες ακολουθούν την κατανομή Βήτα (βλέπε Στέγγος (2006)) και παίρνουν τιμές στο διάστημα (0,1). Οι τ.μ. $U_{(\alpha;m)}$ και $U_{(b;m)}$, ουσιαστικά αντιπροσωπεύουν τα δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$, αφού ισχύει $U_{(\alpha;m)} = F(X_{(\alpha;m)})$ και $U_{(b;m)} = F(X_{(b;m)})$. Επίσης, όταν τα δύο ο.ε. LCL και UCL είναι γνωστές και σταθερές ποσότητες και όχι τυχαίες μεταβλητές, όλα τα σημεία που απεικονίζονται στο δ.ε. θα είναι ανεξάρτητα μεταξύ τους και σε αυτήν την περίπτωση (βλέπε Αντζουλάκος (2006)) η δεσμευμένη κατανομή της τ.μ. N_0, δοθέντων των τιμών των δύο ο.ε. $LCL = X_{(\alpha;m)}$ και $UCL = X_{(b;m)}$, όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, θα είναι η γεωμετρική κατανομή με παράμετρο $1-p_0(s,t)$, όπου $p_0(s,t)$ είναι η πιθανότητα που ορίστηκε και υπολογίστηκε προηγουμένως. Σύμφωνα με το Θ.Ο.Π., η σ.π. της τ.μ. N_0 θα γράφεται (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) στην εξής μορφή:

$$P[N_0 = k] = \int_{-\infty}^{\infty}\int_{-\infty}^{y} P(N_0 = k \mid X_{(a;m)} = x, X_{(b;m)} = y)\, f_{X_{(a;m)},X_{(b;m)}}(x,y)\,dxdy \;=$$

$$= \int_0^1\int_s^1 P(N_0 = k \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds,$$

$k = 1, 2, \ldots$.

Όμως, όπως ήδη αναφέρθηκε, η δεσμευμένη κατανομή της τ.μ. N_0, δοθέντων των τιμών των τ.μ. $X_{(\alpha;m)}$ και $X_{(b;m)}$, όταν η διεργασία είναι εκτός στατιστικού ελέγχου, θα είναι η γεωμετρική κατανομή με παράμετρο ίση με $1-p_0(s,t)$, επομένως θα ισχύει (βλέπε Κούτρας (2002) (α))

$$P[N_0 = k \mid X_{(\alpha;m)} = x, X_{(b;m)} = y] = P[N_0 = k \mid U_{(\alpha;m)} = s, U_{(b;m)} = t] =$$

$$= [p_0(s,t)]^{k-1}[1-p_0(s,t)] = [p_0(s,t)]^{k-1} - [p_0(s,t)]^k$$

$k = 1, 2, \ldots$.

Άρα, η σ.π. της τ.μ N_0, θα γράφεται τελικά στην εξής μορφή:

$$P[N_0 = k] = \int_0^1\int_s^1 P(N_0 = k \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds \;=$$

$$= \int_0^1\int_s^1 ([p_0(s,t)]^{k-1} - [p_0(s,t)]^k)\, f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds$$

$$= \int_0^1\int_s^1 [p_0(s,t)]^{k-1} f_{U_{(a;m)},U_{(b;m)}}(s,t)dtds \; - \; \int_0^1\int_s^1 [p_0(s,t)]^{k} f_{U_{(a;m)},U_{(b;m)}}(s,t)dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 [p_0(s,t)]^{k-1} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

$$- \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 [p_0(s,t)]^{k} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 [\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}((G(F^{-1}(t)))^{j+i}-(G(F^{-1}(s)))^{j+i})]^{k-1} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

$$- \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 [\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{j+i}((G(F^{-1}(t)))^{j+i}-(G(F^{-1}(s)))^{j+i})]^{k} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds$$

Τα δύο παραπάνω ορισμένα διπλά ολοκληρώματα που προέκυψαν δε μπορούν να υπολογιστούν, αφού λόγω των άγνωστων κατανομών που ακολουθούν οι παρατηρήσεις, ο ακριβής συναρτησιακός τύπος των α.σ.κ. F και G δεν είναι γνωστός, ούτε και προσεγγίσιμος. Επομένως, προκύπτει το συμπέρασμα ότι, στο δίπλευρο μη–παραμετρικό δ.ε. τύπου Shewhart, η ακριβής σ.π. του μήκους ροής δε μπορεί να υπολογιστεί και είναι άγνωστη (απροσδιόριστη).

3.6 Ο υπολογισμός του εκτός–ελέγχου μέσου μήκους ροής

Όπως αναφέρθηκε σε προηγούμενη ενότητα, το εκτός–ελέγχου μέσο μήκος ροής ARL_{out}, δηλαδή η αναμενόμενη τιμή της διακριτής τ.μ. N_0, όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, μπορεί να γραφεί στην εξής μορφή:

$$ARL_{out} = E(N_0) = \sum_{k=0}^{\infty} P(N_0 > k).$$

Ο υπολογισμός της πιθανότητας $P(N_0 > k)$ γίνεται ακριβώς με την ίδια διαδικασία, όπως στην περίπτωση κατά την οποία η διεργασία ήταν εντός στατιστικού ελέγχου. Συγκεκριμένα, σύμφωνα με το Θ.Ο.Π., θα ισχύει (βλέπε Chakraborti, Bakir and Van Der Laan (2001))

$$P(N_0 > k) = \int_0^1\int_s^1 P(N_0 > k \mid U_{(a;m)} = s, U_{(b;m)} = t)\, f_{U_{(a;m)},U_{((b;m)}}(s,t)dtds,$$

$k = 0, 1, \ldots$ και όπως αναφέρθηκε προηγουμένως, η διακριτή τ.μ.

$$N_0 \mid U_{(\alpha;m)} = s,\ U_{(b;m)} = t \ \sim\ Geometric(1 - p_0(s,t)).$$

Τότε (βλέπε Κούτρας (2002) (α)), θα ισχύει

$$P(N > k \mid U_{(a,m)} = s, U_{(b,m)} = t) = [p_0(s,t)]^k,$$

οπότε

$$P(N_0 > k) = \int_0^1 \int_s^1 [p_0(s,t)]^k f_{U_{(a;m)},U_{(b;m)}}(s,t)\,dtds =$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 p_0(s,t) s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b}\, dtds$$

$k = 0, 1, \ldots$, όπου $p_0(s,t)$ συμβολίζει την ποσότητα

$$p_0(s,t) = \frac{n!}{(j-1)!(n-j)!} \int_{G(F^{-1}(s))}^{G(F^{-1}(t))} w^{j-1}(1-w)^{n-j}\, dw =$$

$$= \frac{n!}{(j-1)!(n-j)!} \sum_{i=0}^{n-j} \binom{n-j}{i} \frac{(-1)^i}{i+j} [(G(F^{-1}(t)))^{j+i} - (G(F^{-1}(s)))^{j+i}] .$$

Επομένως, προκύπτει (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) για το ARL_{out} ότι

$$ARL_{out} = E(N_0) = \sum_{k=0}^{\infty} P(N_0 > k) =$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \sum_{k=0}^{\infty} \int_0^1 \int_s^1 p_0(s,t) s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b}\, ds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \sum_{k=0}^{\infty} p_0(s,t) s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b}\, dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \{\sum_{k=0}^{\infty} p_0(s,t)\} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b}\, dtds$$

και αντικαθιστώντας στην τελευταία έκφραση τη σχέση (βλέπε Thomas and Finney (1990))

$$\sum_{k=0}^{\infty} [p_0(s,t)]^k = \frac{1}{1 - p_0(s,t)},\ p_0(s,t) \in (0,1),$$

προκύπτει ότι

$ARL_{out} = E(N_0) =$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 \frac{1}{1-p_0(s,t)} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dtds =$$

$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!}\int_0^1\int_s^1 \frac{s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}}{1-\frac{n!}{(j-1)!(n-j)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{i+j}([G(F^{-1}(t))]^{j+i}-[G(F^{-1}(s))]^{j+i})}dtds$$

.

Το παραπάνω ορισμένο διπλό ολοκλήρωμα δεν είναι δυνατόν να υπολογιστεί ή έστω να προσεγγιστεί ικανοποιητικά, εκτός αν είναι γνωστός ο ακριβής συναρτησιακός τύπος της αρχικής α.σ.κ. F και της νέας α.σ.κ. G. Έτσι, είναι προφανές ότι το ακριβές εκτός–ελέγχου μέσο μήκος ροής ARL_{out} δε μπορεί να υπολογιστεί, ούτε και να προσεγγιστεί ικανοποιητικά, αφού στην περίπτωση του μη–παραμετρικού δ.ε. δεν είναι γνωστή η συναρτησιακή (μαθηματική) μορφή των α.σ.κ. F και G. Αυτό το αποτέλεσμα, ίσως να αποτελεί ένα σημαντικό μειονέκτημα για το συγκεκριμένο δίπλευρο μη–παραμετρικό δ.ε..

Ο υπολογισμός του εκτός–ελέγχου μέσου μήκους ροής ARL_{out}, θα μπορούσε να γίνει και με έναν δεύτερο τρόπο, χρησιμοποιώντας γνωστό θεώρημα από την Θ.Π. για δεσμευμένες μέσες τιμές. Συγκεκριμένα, (βλέπε Κούτρας (2002) (β)) θα ισχύει

$ARL_{out} = E(N_0) = E(E(N_0 \mid X_{(\alpha;m)}, X_{(b;m)})) = E(E(N_0 \mid U_{(\alpha;m)}, U_{(b;m)}))$,

όπου $X_{(\alpha;m)}$, $X_{(b;m)}$ είναι το κ.ο.ε. και το α.ο.ε. αντίστοιχα, $U_{(\alpha;m)} = F(X_{(\alpha;m)})$, $U_{(b;m)} = F(X_{(b;m)})$ και F είναι η εντός–ελέγχου α.σ.κ. των τιμών των αρχικών παρατηρήσεων X_i, που συλλέχθηκαν ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου. Επίσης, όπως ήδη αναφέρθηκε, η δεσμευμένη κατανομή $N_0 \mid X_{(\alpha;m)} = x$, $X_{(b;m)} = y$ ή ισοδύναμα, η δεσμευμένη κατανομή $N_0 \mid U_{(\alpha;m)} = s$, $U_{(b;m)} = t$, θα είναι η γεωμετρική κατανομή με παράμετρο 1 – $p_0(s,t)$. Επομένως, αφού η τ.μ. $N_0 \mid U_{(\alpha;m)} = s$, $U_{(b;m)} = t$ ακολουθεί τη γεωμετρική κατανομή με παράμετρο 1– $p_0(s,t)$, η δεσμευμένη μέση τιμή $E(N_0 \mid U_{(\alpha;m)} = s, U_{(b;m)} = t)$, θα είναι ίση με τη μέση τιμή αυτής της γεωμετρικής κατανομής, δηλαδή θα ισχύει (βλέπε Κούτρας (2002) (α))

$$E(N_0 \mid U_{(\alpha;m)} = s, U_{(b;m)} = t) = \frac{1}{1-p_0(s,t)}.$$

Επομένως, η αναμενόμενη τιμή $E(N_0)$, σύμφωνα με όλα τα παραπάνω, θα είναι ίση με:

$ARL_{out} = E(N_0) = E(E(N_0 \mid U_{(a;m)}, U_{(b;m)})) =$

$$= \int_0^1 \int_s^1 E(N_0 \mid U_{(a;m)} = s, U_{(b;m)} = t) f_{U_{(a;m)},U_{(b;m)}}(s,t) dtds = \int_0^1 \int_s^1 \frac{1}{1 - p_0(s,t)} f_{U_{(a;m)},U_{(b;m)}}(s,t) dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \frac{1}{1 - p_0(s,t)} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} dtds$$

$$= \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \int_0^1 \int_s^1 \frac{s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}}{1 - \frac{n!}{(j-1)!(n-j)!} \sum_{i=0}^{n-j} \binom{n-j}{i} \frac{(-1)^i}{i+j} ([G(F^{-1}(t))]^{j+i} - [G(F^{-1}(s))]^{j+i})} dtds$$

τύπος ο οποίος, όπως ήταν αναμενόμενο, συμπίπτει με αυτόν που βρέθηκε προηγουμένως.

ΚΕΦΑΛΑΙΟ 4

Η κατανομή του μήκους ροής στο μονόπλευρο μη–παραμετρικό δ.ε. τύπου Shewhart

4.1 Ο υπολογισμός της πιθανότητας λανθασμένου συναγερμού

Όπως στο δίπλευρο μη–παραμετρικό δ.ε. Shewhart, έτσι και στο αντίστοιχο μονόπλευρο, παρατηρείται ακριβώς το ίδιο φαινόμενο στα σημεία που απεικονίζονται. Δηλαδή, (βλέπε Chakraborti, Bakir and Van Der Laan (2001)) αυτά δεν είναι ανεξάρτητα μεταξύ τους, αλλά συσχετίζονται (πιθανόν και σε σημαντικό βαθμό). Με άλλα λόγια, το ενδεχόμενο $\{Y_{(k;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]\}$ δε θα είναι στατιστικά ανεξάρτητο από το ενδεχόμενο $\{Y_{(l;n)} \notin [X_{(\alpha;m)}, X_{(b;m)}]\}$ και το ενδεχόμενο $\{Y_{(k,n)} \in [X_{(\alpha;m)}, X_{(b;m)}]\}$ δε θα είναι στατιστικά ανεξάρτητο από το ενδεχόμενο $\{Y_{(l,n)} \in [X_{(\alpha;m)}, X_{(b;m)}]\}$, όπου $Y_{(k;n)}$ και $Y_{(l;n)}$ είναι παρατηρήσεις που προκύπτουν από δύο διαφορετικά και ανεξάρτητα μεταξύ τους (νέα) τ.δ.. Αυτό συμβαίνει, κυρίως λόγω του ότι, το όριο ελέγχου, δεν είναι κάποια γνωστή και σταθερή ποσότητα, αλλά είναι τυχαία μεταβλητή.

Επομένως και πάλι, προκύπτει το συμπέρασμα ότι, η διακριτή κατανομή του συνολικού αριθμού των σημείων που απεικονίζονται στο μονόπλευρο δ.ε., έως ότου για πρώτη φορά εμφανιστεί κάποιο σημείο που θα βρίσκεται εκτός των δύο ο.ε., ασφαλώς δε θα είναι η γεωμετρική κατανομή (όπως στα κλασικά παραμετρικά δ.ε. τύπου Shewhart), αφού η γεωμετρική κατανομή απαιτεί την υπόθεση της πλήρους ανεξαρτησίας μεταξύ όλων των ενδεχομένων του δειγματικού χώρου (βλέπε Κούτρας (2002) (α)), κάτι που δεν ισχύει στο συγκεκριμένο μονόπλευρο διάγραμμα.

Στην περίπτωση κατά την οποία στο διάγραμμα Shewhart υπάρχει μόνο ένα ο.ε., η διαδικασία υπολογισμού της πιθανότητας λανθασμένου συναγερμού FAR θα είναι ίδια με αυτή για το αντίστοιχο δίπλευρο δ.ε.. Συγκεκριμένα, αν στο διάγραμμα υπάρχει μόνο το α.ο.ε. $UCL = X_{(b;m)}$, η πιθανότητα λανθασμένου συναγερμού FAR θα ισούται με

$$FAR = P(Y_{(j;n)} > X_{(b;m)} \mid \text{η διεργασία εντός–ελέγχου})$$

και σύμφωνα με το Θ.Ο.Π., δεσμεύοντας ως προς όλες τις (συνεχείς) τιμές για την τ.μ. $X_{(b;m)}$, η συγκεκριμένη πιθανότητα θα γράφεται (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) στην εξής μορφή:

$FAR = P(Y_{(j;n)} > X_{(b;m)} \,|$ η διεργασία εντός ελέγχου$) = P(F(Y_{(j;n)}) > F(X_{(b;m)})) =$

$$= P(U_{(j;n)} > U_{(b;m)}) = \int_0^1 P(U_{(j;n)} > U_{(b;m)} \,|\, U_{(b;m)} = s) f_{U_{(b;m)}}(s)ds$$

$$= \int_0^1 P(U_{(j;n)} > s \,|\, U_{(b;m)} = s) f_{U_{(b;m)}}(s)ds = \int_0^1 P(U_{(j;n)} > s) f_{U_{(b;m)}}(s)ds$$

και λαμβάνοντας υπ'όψιν εδώ ότι (βλέπε Στέγγος (2006)) η τ.μ. $U_{(j;n)}$ ακολουθεί την κατανομή Βήτα με παραμέτρους j και $n\text{–}j+1$ και ότι η τ.μ. $U_{(b;m)}$ ακολουθεί την κατανομή Βήτα με παραμέτρους b και $m\text{–}b+1$, μπορούμε να γράψουμε (βλέπε Κούτρας (2002) (α))

$FAR = P(Y_{(j;n)} > X_{(b;m)} \,|$ η διεργασία εντός ελέγχου$) =$

$$= \int_0^1 \Big[\int_s^1 \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j} dw\Big] \frac{m!}{(b-1)!(m-b)!} s^{a-1}(1-s)^{m-b} ds$$

$$= \frac{n!}{(j-1)!(n-j)!} \frac{m!}{(b-1)!(m-b)!} \int_0^1 \int_s^1 w^{j-1}(1-w)^{n-j} s^{b-1}(1-s)^{m-b} dwds \ .$$

Αποδεικνύεται, κατόπιν πράξεων, ότι η παραπάνω σχέση γράφεται ως

$$\frac{n!}{(j-1)!(n-j)!} \frac{m!}{(b-1)!(m-b)!} \int_0^1 \int_s^1 w^{j-1}(1-w)^{n-j} s^{b-1}(1-s)^{m-b} dwds =$$

$$= 1 - \frac{n!}{(j-1)!(n-j)!} \frac{m!}{(b-1)!} \sum_{i=0}^{n-j} \binom{n-j}{i} \frac{(-1)^i}{i+j} \frac{(i+j+b-1)!}{(i+j+m)!} ,$$

οπότε η πιθανότητα λανθασμένου συναγερμού FAR για το μονόπλευρο δ.ε. Shewhart, στην περίπτωση που υπάρχει το α.ο.ε. $UCL = X_{(b;m)}$, θα πάρει τελικά τη μορφή

$FAR = P(Y_{(j;n)} > X_{(b,m)} \,|$ η διεργασία εντός ελέγχου$) =$

$$= 1 - \frac{n!}{(j-1)!(n-j)!} \frac{m!}{(b-1)!} \sum_{i=0}^{n-j} \binom{n-j}{i} \frac{(-1)^i}{i+j} \frac{(i+j+b-1)!}{(i+j+m)!} .$$

Όμως, όπως διαπιστώθηκε σε προηγούμενη ενότητα, η πιθανότητα λανθασμένου συναγερμού *FAR* για το μονόπλευρο δ.ε. Shewhart, στην περίπτωση που υπάρχει το α.ο.ε. $UCL = X_{(b;m)}$, μπορεί να υπολογιστεί, μέσω της κατανομής της διακριτής τ.μ. W_j και από τον τύπο

$FAR = P(Y_{(j;n)} > X_{(b,m)} \mid$ η διεργασία εντός ελέγχου$)$ =

$$= \sum_{w=b}^{m} \frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}}.$$

Οι δύο τελευταίοι τύπο που γράφτηκαν, αναφέρονται στην ίδια ακριβώς ποσότητα (στην πιθανότητα λανθασμένου συναγερμού *FAR*) και ισχύουν για κάθε πιθανή τιμή των θετικών και ακέραιων παραμέτρων *n, m, j* και *b*, με $1 \le b \le m$ και $1 \le j \le n$. Άρα, συμπεραίνουμε εδώ ότι, οι δύο μαθηματικοί τύποι που δίνουν την πιθανότητα *FAR*, θα ταυτίζονται μεταξύ τους.

Στην περίπτωση που στο διάγραμμα υπάρχει μόνο το κ.ο.ε. $LCL = X_{(\alpha;m)}$, η πιθανότητα λανθασμένου συναγερμού *FAR* θα είναι ίση με

$$FAR = P(Y_{(j;n)} < X_{(\alpha;m)} \mid \text{η διεργασία εντός ελέγχου})$$

και σύμφωνα με το Θ.Ο.Π., δεσμεύοντας ως προς όλες τις (συνεχείς) τιμές για την τ.μ. $X_{(\alpha;m)}$, η *FAR* θα είναι (βλέπε Chakraborti, Van Der Laan and Van De Wiel (2004)) ίση με:

$FAR = P(Y_{(j;n)} < X_{(\alpha;m)} \mid$ η διεργασία εντός ελέγχου$) = P(F(Y_{(j;n)}) < F(X_{(\alpha;m)}))$ =

$$= P(U_{(j;n)} < U_{(\alpha;m)}) = \int_0^1 P(U_{(j;n)} < U_{(a;m)} \mid U_{(a;m)} = t) f_{U_{(a;m)}}(t)dt$$

$$= \int_0^1 P(U_{(j;n)} < t \mid U_{(a;m)} = t) f_{U_{(a;m)}}(t)dt = \int_0^1 P(U_{(j;n)} < t) f_{U_{(a;m)}}(t)dt$$

και λαμβάνοντας υπ' όψιν εδώ ότι η τ.μ. $U_{(j;n)}$ ακολουθεί την κατανομή Βήτα με παραμέτρους *j* και $n-j+1$ και ότι η τ.μ. $U_{(\alpha;m)}$ ακολουθεί την κατανομή Βήτα με παραμέτρους *α* και $m-\alpha+1$, η πιθανότητα *FAR* θα ισούται με

$FAR = P(Y_{(j;n)} < X_{(\alpha;m)} \mid$ η διεργασία εντός ελέγχου$)$ =

$$= \int_0^1 [\int_0^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j} dw] \frac{m!}{(a-1)!(m-a)!} t^{a-1}(1-t)^{m-a} dt =$$

$$= \frac{n!}{(j-1)!(n-j)!}\frac{m!}{(a-1)!(m-a)!}\int_0^1\int_0^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j} t^{a-1}(1-t)^{m-a}\, dwdt \ .$$

Αποδεικνύεται, κατόπιν πράξεων, ότι η παραπάνω σχέση γράφεται ως

$$\frac{n!}{(j-1)!(n-j)!}\frac{m!}{(a-1)!(m-a)!}\int_0^1\int_0^t \frac{n!}{(j-1)!(n-j)!} w^{j-1}(1-w)^{n-j} t^{a-1}(1-t)^{m-a}\, dwdt =$$

$$= \frac{n!}{(j-1)!(n-j)!}\frac{m!}{(a-1)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{i+j}\frac{(i+j+a-1)!}{(i+j+m)!},$$

οπότε η πιθανότητα λανθασμένου συναγερμού *FAR* για το μονόπλευρο δ.ε. τύπου Shewhart, στην περίπτωση που υπάρχει το κ.ο.ε. $LCL = X_{(\alpha;m)}$, θα είναι τελικά ίση με:

$FAR = P(Y_{(j;n)} < X_{(\alpha;m)} \mid$ η διεργασία είναι εντός ελέγχου$) =$

$$= \frac{n!}{(j-1)!(n-j)!}\frac{m!}{(a-1)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i}{i+j}\frac{(i+j+a-1)!}{(i+j+m)!}.$$

Είναι μάλλον προφανές ότι, επειδή η παραπάνω ποσότητα περιέχει αρκετούς όρους με παραγοντικά και ένα διπλό άθροισμα, αυτή (δηλαδή η *FAR*) θα είναι αρκετά δύσκολο και επίπονο να υπολογιστεί μέσω των παραπάνω αριθμητικών πράξεων. Μπορεί όμως να υπολογιστεί πιο απλά και με λιγότερο κόπο, με χρήση ενός μαθηματικού προγράμματος σε υπολογιστή.

Ο τελευταίος τύπος που γράφτηκε, είναι ταυτόσημος με τον τύπο

$$FAR = \sum_{w=0}^{a-1}\frac{\binom{j+w-1}{w}\binom{n+m-j-w}{m-w}}{\binom{n+m}{m}},$$

που δίνει την πιθανότητα λανθασμένου συναγερμού *FAR* για το μονόπλευρο μη–παραμετρικό δ.ε. Shewhart, μέσω της κατανομής της διακριτής τ.μ. W_j. Δηλαδή, οι δύο τύποι που δίνουν την *FAR*, θα ταυτίζονται μεταξύ τους και θα ισχύουν για κάθε πιθανή τιμή των θετικών ακέραιων παραμέτρων n, m, j και a, με $1 \le a \le m$ και $1 \le j \le n$.

4.2 Η ακριβής σ.π. του εντός–ελέγχου μήκους ροής

Όπως στο δίπλευρο μη–παραμετρικό δ.ε., έτσι και στο μονόπλευρο, ο υπολογισμός της σ.π. του μήκους ροής, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, θα γίνει ακριβώς με τον ίδιο τρόπο. Υπολογίζεται αρχικά η δεσμευμένη σ.π. του μήκους ροής, για δεδομένη τιμή του ο.ε.. Όπως έχει ήδη αναφερθεί, στην περίπτωση που είναι γνωστή και σταθερή η τιμή του ο.ε., όλα τα σημεία που απεικονίζονται στο δ.ε. θα είναι ανεξάρτητα μεταξύ τους και επομένως, η κατανομή που θα ακολουθεί το μήκος ροής στο διάγραμμα (βλέπε Αντζουλάκος (2006)) θα είναι η γεωμετρική. Στη συνέχεια, για να υπολογιστεί η σ.π. του μήκους ροής, γίνεται χρήση του Θ.Ο.Π., δεσμεύοντας ως προς όλες τις (συνεχείς) τιμές για το ο.ε., δηλαδή ισοδύναμα δεσμεύοντας τη σ.π. μιας γεωμετρικής κατανομής ως προς όλες τις πιθανές τιμές για το ο.ε..

Έστω λοιπόν ότι στο διάγραμμα υπάρχει μόνο το α.ο.ε. $UCL = X_{(b;m)}$. Τότε, σύμφωνα με τα παραπάνω, αν N είναι η διακριτή τ.μ. η οποία δηλώνει το εντός–ελέγχου μήκος ροής, για την εντός–ελέγχου σ.π. της τ.μ. N, σύμφωνα με το Θ.Ο.Π., θα ισχύει η εξής σχέση:

$$P_C[N = k] = \int_{-\infty}^{\infty} P_C(N = k \mid X_{(b;m)} = x)\, f_{X_{(b;m)}}(x)dx \;=\; \int_0^1 P_C(N = k \mid U_{(b;m)} = s)\, f_{U_{(b;m)}}(s)ds,$$

$k = 1, 2, \ldots$ και όπως προηγουμένως αναφέρθηκε, η διακριτή τ.μ. $N \mid X_{(b;m)} = x$, ή ισοδύναμα η τ.μ. $N \mid U_{(b;m)} = s$, θα ακολουθεί τη γεωμετρική κατανομή με παράμετρο (ή πιθανότητα «επιτυχίας») $p(s)$, όπου $p(s)$ είναι η δεσμευμένη πιθανότητα

$$p(s) = P(Y_{(j,n)} > X_{(b;m)} \mid X_{(b;m)} = x, \text{ η διεργασία εντός ελέγχου}) =$$

$$= P(U_{(j;n)} > U_{(b;m)} \mid U_{(b;m)} = s) = P(U_{(j;n)} > s \mid U_{(b;m)} = s) = P(U_{(j;n)} > s),$$

λόγω της ανεξαρτησίας. Δηλαδή, θα ισχύει

$$p(s) = P(U_{(j;n)} > s) = \int_s^1 f_{U(j;n)}(w)dw = \int_s^1 \frac{n!}{(j-1)!\,(n-j)!} w^{j-1}(1-w)^{n-j}\,dw =$$

$$= \frac{n!}{(j-1)!\,(n-j)!} \int_s^1 w^{j-1}(1-w)^{n-j}\,dw$$

και επειδή ισχύει η γενική σχέση (βλέπε Κούτρας (2001))

$$\int_s^1 w^{j-1}(1-w)^{n-j}dw = \sum_{i=0}^{n-j} \binom{n-j}{i} \frac{(-1)^i}{i+j}(1-s^{i+j}),$$

προκύπτει ότι

$p(s) = P(Y_{(j;n)} \geq X_{(b;m)} \,|\, X_{(b;m)} = x$, η διεργασία εντός ελέγχου) =

$$= P(U_{(j;n)} \geq U_{(b;m)} \,|\, U_{(b;m)} = s) = \frac{n!}{(j-1)!(n-j)!}\int_s^1 w^{j-1}(1-w)^{n-j}\,dw =$$

$$= \frac{n!}{(j-1)!(n-j)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i(1-s^{i+j})}{(i+j)}$$

$$= \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i(1-s^{i+j})}{i!(n-j-i)!(i+j)}$$

και επομένως (βλέπε Κούτρας (2002) (α))

$$P_C(N = k \,|\, U_{(b;m)} = s) = [1-p(s)]^{k-1}[p(s)] = [q(s)]^{k-1}[1-q(s)] = [q(s)]^{k-1} - [q(s)]^k,$$

$k = 1, 2, \ldots$, όπου

$$q(s) = 1 - p(s) = 1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i(1-s^{i+j})}{i!(n-j-i)!(i+j)}.$$

Λαμβάνοντας υπ'όψιν όλα τα παραπάνω, τελικά προκύπτει ότι η σ.π. της τ.μ. N θα είναι ίση με:

$$P_C[N = k] = \frac{m!}{(b-1)!(m-b)!}\int_0^1 P_C(N = k \,|\, U_{(b;m)} = s)s^{b-1}(1-s)^{m-b}\,ds =$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 ([q(s)]^{k-1} - [q(s)]^k)s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 [q(s)]^{k-1}s^{b-1}(1-s)^{m-b}\,ds$$

$$- \frac{m!}{(b-1)!(m-b)!}\int_0^1 [q(s)]^{k}s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 [1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i(1-s^{i+j})}{i!(n-j-i)!(i+j)}]^{k-1}s^{b-1}(1-s)^{m-b}\,ds$$

$$- \frac{m!}{(b-1)!(m-b)!}\int_0^1 [1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i(1-s^{i+j})}{i!(n-j-i)!(i+j)}]^{k}s^{b-1}(1-s)^{m-b}\,ds,$$

$k = 1, 2, \ldots$ και σε αυτό το σημείο είναι μάλλον φανερό ότι, τα δύο παραπάνω ορισμένα ολοκληρώματα που εμφανίστηκαν σε σύνθετη μορφή (δηλαδή η σ.π. $P_C[N = k]$), λόγω της ιδιαίτερης μαθηματικής τους πολυπλοκότητας, δεν είναι δυνατόν να υπολογιστούν με τις γνωστές μαθηματικές μεθόδους ολοκλήρωσης, αλλά υπολογίζονται αναλυτικά και με ακρίβεια μόνο με χρήση ενός μαθηματικού προγράμματος σε υπολογιστή.

Παρόμοια διαδικασία ακολουθείται στην περίπτωση που στο δ.ε. υπάρχει μόνο το κ.ο.ε. $LCL = X_{(\alpha;m)}$. Συγκεκριμένα, σε αυτήν την περίπτωση, σύμφωνα με το Θ.Ο.Π., η σ.π. της τ.μ. N θα είναι

$$P_C[N=k] = \int_{-\infty}^{\infty} P_C(N=k \mid X_{(a;m)}=y) f_{X_{(a;m)}}(y)dy = \int_0^1 P_C(N=k \mid U_{(a;m)}=t) f_{U_{(a;m)}}(t)dt =$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 P_C(N=k \mid U_{(a;m)}=t)t^{a-1}(1-t)^{m-a}dt \ ,$$

$k = 1, 2, \ldots$. Επίσης, η τ.μ. $N \mid X_{(\alpha;m)} = x$, ή ισοδύναμα η τ.μ. $N \mid U_{(\alpha;m)} = t$, θα ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p_2(t)$, όπου $p_2(t)$ είναι η δεσμευμένη πιθανότητα

$p_2(t) = P(Y_{(j;n)} \leq X_{(\alpha;m)} \mid X_{(\alpha;m)} = x$, η διεργασία εντός ελέγχου) =

$$= P(U_{(j;n)} \leq U_{(\alpha;m)} \mid U_{(\alpha,m)} = t) = P(U_{(j,n)} \leq t \mid U_{(b;m)} = t) = P(U_{(j;n)} \leq t),$$

λόγω της ανεξαρτησίας. Δηλαδή,

$$p_2(t) = P(U_{(j;n)} \leq t) = \int_o^t f_{U_{(j;n)}}(w)dw = \frac{n!}{(j-1)!(n-j)!}\int_0^t w^{j-1}(1-w)^{n-j}dw$$

και αν λάβουμε υπ'όψιν ότι ισχύει η γενική σχέση

$$\int_0^t w^{j-1}(1-w)^{n-j}dw = \sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i t^{i+j}}{i+j}$$

προκύπτει ότι

$$p_2(t) = P(U_{(j;n)} \leq t) = \frac{n!}{(j-1)!(n-j)!}\int_0^t w^{j-1}(1-w)^{n-j}dw =$$

$$= \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i t^{i+j}}{i!(n-j-i)!(i+j)}$$

και επομένως, θα ισχύει

$$P_C(N=k \mid U_{(\alpha;m)}=t) = [1-p_2(t)]^{k-1}[p_2(t)] = [q_2(t)]^{k-1}[1-q_2(t)] = [q_2(t)]^{k-1}-[q_2(t)]^k,$$

όπου

$$q_2(t) = 1-p_2(t) = 1-\frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i t^{i+j}}{i!(n-j-i)!(i+j)}.$$

Λαμβάνοντας υπ'όψιν όλα τα παραπάνω, καταλήγουμε τελικά στο ότι η σ.π. $P_C[N = k]$ της τ.μ. N, θα είναι ίση με:

$$P_C[N = k] = \frac{m!}{(a-1)!(m-a)!}\int_0^1 P_C(N = k \mid U_{(a;m)} = t)t^{a-1}(1-t)^{m-a}\,dt =$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 ([q_2(t)]^{k-1} - [q_2(t)]^{k})t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 [q_2(t)]^{k-1}t^{a-1}(1-t)^{m-a}\,dt$$

$$- \frac{m!}{(a-1)!(m-a)!}\int_0^1 [q_2(t)]^{k}t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 [1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i t^{i+j}}{i!(n-j-i)!(i+j)}]^{k-1} t^{a-1}(1-t)^{m-a}\,dt$$

$$- \frac{m!}{(a-1)!(m-a)!}\int_0^1 [1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i t^{i+j}}{i!(n-j-i)!(i+j)}]^{k} t^{a-1}(1-t)^{m-a}\,dt ,$$

$k = 1, 2, \ldots$ και σε αυτό το σημείο προκύπτουν τα ίδια ακριβώς αποτελέσματα και συμπεράσματα, όπως στην περίπτωση που στο μονόπλευρο διάγραμμα υπήρχε μόνο το α.ο.ε. *UCL*.

4.3 Ο υπολογισμός του εντός–ελέγχου μέσου μήκους ροής

Η διαδικασία υπολογισμού του εντός–ελέγχου μέσου μήκους ροής ARL_{in} για το μονόπλευρο μη–παραμετρικό δ.ε. Shewhart, είναι ακριβώς η ίδια με αυτήν για το αντίστοιχο δίπλευρο διάγραμμα. Συγκεκριμένα, το εντός–ελέγχου μέσο μήκος ροής ARL_{in}, θα είναι (βλέπε Αντζουλάκος (2006)) η μέση τιμή της διακριτής τ.μ. N, η οποία παριστάνει το μήκος ροής στο δ.ε. όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Επειδή η διακριτή τ.μ. N παίρνει τις τιμές 1, 2, ... , σύμφωνα με γνωστό θεώρημα της Θ.Π., η μέση τιμή της θα ισούται (βλέπε Κούτρας (2002) (α)) με το άπειρο άθροισμα

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N > k)\,.$$

Αν στο διάγραμμα υπάρχει μόνο το α.ο.ε. $UCL = X_{(b;m)}$, εφαρμόζοντας το Θ.Ο.Π. για την πιθανότητα $P(N > k)$, προκύπτει ότι

$$P(N > k) = \int_{-\infty}^{\infty} P(N > k \mid X_{(b;m)} = x)\, f_{X_{(b;m)}}(x)dx = \int_{0}^{1} P(N > k \mid U_{(b;m)} = s)\, f_{U_{(b;m)}}(s)ds$$

και επειδή η τ.μ. $N \mid U_{(b;m)} = s$ ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p(s)$, όπου $p(s)$ είναι η πιθανότητα που ορίστηκε και υπολογίστηκε προηγουμένως, θα ισχύει (βλέπε Κούτρας (2002) (α)) για τη δεσμευμένη πιθανότητα $P(N > k \mid U_{(b;m)} = s)$ το εξής:

$$P(N > k \mid U_{(b;m)} = s) = [1-p(s)]^k.$$

Επομένως, σύμφωνα με το Θ.Ο.Π. παίρνουμε ότι

$P(N > k) =$

$$= \int_{0}^{1} P(N > k \mid U_{(b;m)} = s)\, f_{U_{(b;m)}}(s)ds = \frac{m!}{(b-1)!(m-b)!}\int_{0}^{1}[1-p(s)]^k s^{b-1}(1-s)^{m-b}ds\,,$$

$k = 0, 1, \ldots$, οπότε, λαμβάνοντας υπ'όψιν όλα τα παραπάνω, καταλήγουμε στο ότι το εντός–ελέγχου μέσο μήκος ροής ARL_{in} θα είναι ίσο με:

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N > k) = \sum_{k=0}^{\infty} \frac{m!}{(b-1)!(m-b)!}\int_{0}^{1}[1-p(s)]^k s^{b-1}(1-s)^{m-b}ds =$$

$$= \frac{m!}{(b-1)!(m-b)!}\sum_{k=0}^{\infty}\int_{0}^{1}[1-p(s)]^k s^{b-1}(1-s)^{m-b}ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_{0}^{1}\sum_{k=0}^{\infty}[1-p(s)]^k s^{b-1}(1-s)^{m-b}ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_{0}^{1}\{\sum_{k=0}^{\infty}[1-p(s)]^k\} s^{b-1}(1-s)^{m-b}ds\,.$$

Επειδή η ποσότητα $p(s)$ είναι μία πιθανότητα και παίρνει τιμές στο διάστημα (0,1), η σειρά $\sum_{k=0}^{\infty}[1-p(s)]^k$ συγκλίνει και ισχύει (βλέπε Thomas and Finney (1990))

$$\sum_{k=0}^{\infty}[1-p(s)]^k = \frac{1}{1-[1-p(s)]} = \frac{1}{p(s)}\,,\ p(s) \in (0,1)\,,$$

οπότε, αντικαθιστώντας την τελευταία σχέση στο παραπάνω ορισμένο ολοκλήρωμα και λαμβάνοντας υπ'όψιν ότι ισχύει

$$p(s) = \frac{n!}{(j-1)!(n-j)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i(1-s^{i+j})}{i+j} = \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i(1-s^{i+j})}{i!(n-j-i)!(i+j)},$$

προκύπτει ότι:

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N > k) = \frac{m!}{(b-1)!(m-b)!}\int_0^1 \{\sum_{k=0}^{\infty}[1-p(s)]^k\} s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 \frac{1}{p(s)} s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{(j-1)!}{n!}\frac{m!}{(b-1)!(m-b)!}\int_0^1 \frac{s^{b-1}(1-s)^{m-b}}{\sum_{i=0}^{n-j}\frac{(-1)^i(1-s^{i+j})}{i!(n-j-i)!(i+j)}}ds\,.$$

Είναι μάλλον φανερό ότι η τελευταία έκφραση δεν επιδέχεται περισσότερη απλοποίηση, επομένως το εντός–ελέγχου μέσο μήκος ροής ARL_{in} δεν είναι δυνατόν να υπολογιστεί αναλυτικά με τις γνωστές μαθηματικές μεθόδους ολοκλήρωσης, αλλά υπολογίζεται μόνο με χρήση ενός ειδικού μαθηματικού προγράμματος σε υπολογιστή.

Ανάλογη διαδικασία ακολουθείται και για τον υπολογισμό του εντός–ελέγχου μέσου μήκους ροής ARL_{in}, στην περίπτωση που στο διάγραμμα υπάρχει μόνο το κ.ο.ε. $LCL = X_{(\alpha;m)}$. Συγκεκριμένα, το ARL_{in}, σύμφωνα με θεώρημα της Θ.Π., θα είναι ίσο με

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N > k)$$

και εφαρμόζοντας το Θ.Ο.Π. για την πιθανότητα $P(N > k)$ προκύπτει ότι

$$P(N > k) = \int_{-\infty}^{\infty} P(N > k \mid X_{(\alpha;m)} = x) f_{X_{(\alpha;m)}}(x)dx = \int_0^1 P(N > k \mid U_{(\alpha;m)} = t) f_{U_{(\alpha;m)}}(t)dt\,,$$

$k = 0, 1, \ldots$. Επειδή η τ.μ. $N \mid U_{(\alpha;m)} = t$ ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p_2(t)$, όπου $p_2(t)$ είναι η πιθανότητα που ορίστηκε και υπολογίστηκε σε προηγούμενη ενότητα, θα ισχύει

$$P(N > k \mid U_{(\alpha;m)} = t) = [1 - p_2(t)]^k.$$

Εφαρμόζοντας το Θ.Ο.Π. για την πιθανότητα $P(N > k)$, προκύπτει ότι

$$P(N > k) = \int_0^1 P(N > k \mid U_{(\alpha;m)} = t) f_{U_{(\alpha;m)}}(t)dt =$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 [1-p_2(t)]^k t^{a-1}(1-t)^{m-a}\,dt\ ,$$

k = 0, 1, … και επομένως, το ARL_{in} θα ισούται με:

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N>k) = \sum_{k=0}^{\infty}\frac{m!}{(a-1)!(m-a)!}\int_0^1 [1-p_2(t)]^k t^{a-1}(1-t)^{m-a}\,dt =$$

$$= \frac{m!}{(a-1)!(m-a)!}\sum_{k=0}^{\infty}\int_0^1 [1-p_2(t)]^k t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 \sum_{k=0}^{\infty}[1-p_2(t)]^k t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 \{\sum_{k=0}^{\infty}[1-p_2(t)]^k\} t^{a-1}(1-t)^{m-a}\,dt\ .$$

Επειδή η ποσότητα $p_2(t)$ είναι μία πιθανότητα και παίρνει τιμές στο διάστημα (0,1), η σειρά $\sum_{k=0}^{\infty}[1-p_2(t)]^k$ συγκλίνει και άρα θα ισχύει (βλέπε Thomas and Finney (1990)) ο μαθηματικός τύπος

$$\sum_{k=0}^{\infty}[1-p_2(t)]^k = \frac{1}{1-[1-p_2(t)]} = \frac{1}{p_2(t)},\ p_2(t)\in(0,1).$$

Αντικαθιστώντας την τελευταία έκφραση στο παραπάνω ολοκλήρωμα και λαμβάνοντας επίσης υπ’όψιν ότι ισχύει η σχέση

$$p_2(t) = \frac{n!}{(j-1)!(n-j)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i t^{i+j}}{(i+j)} = \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i t^{i+j}}{i!(n-j-i)!(i+j)}\ ,$$

προκύπτει ότι το ARL_{in} θα ισούται με

$$ARL_{in} = E(N) = \sum_{k=0}^{\infty} P(N>k) = \frac{m!}{(a-1)!(m-a)!}\int_0^1 \{\sum_{k=0}^{\infty}[1-p_2(t)]^k\} t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 \frac{1}{p_2(t)} t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{(j-1)!}{n!}\frac{m!}{(a-1)!(m-a)!}\int_0^1 \frac{t^{a-1}(1-t)^{m-a}}{\sum_{i=0}^{n-j}\frac{(-1)^i t^{i+j}}{i!(n-j-i)!(i+j)}}dt\ .$$

Για το παραπάνω ολοκλήρωμα (δηλαδή για το εντός–ελέγχου μέσο μήκος ροής ARL_{in}), θα ισχύουν ακριβώς τα ίδια αποτελέσματα και συμπεράσματα, όπως στο αντίστοιχο ολοκλήρωμα που εμφανίστηκε σε σύνθετη μορφή στην περίπτωση που στο μονόπλευρο δ.ε. υπήρχε μόνο το α.ο.ε. UCL.

4.4 Η σ.π. του μήκους ροής όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου

Στην περίπτωση που η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, η διαδικασία υπολογισμού της σ.π. του μήκους ροής για το μονόπλευρο δ.ε. είναι ίδια με αυτή για το αντίστοιχο δίπλευρο διάγραμμα που αναφέρθηκε σε προηγούμενη ενότητα. Συγκεκριμένα, αν στο διάγραμμα υπάρχει μόνο το α.ο.ε. $UCL = X_{(b;m)}$, F είναι η α.σ.κ. των τιμών των αρχικών παρατηρήσεων που συλλέχθηκαν ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου, G είναι η νέα α.σ.κ. των τιμών των παρατηρήσεων που συλλέχθηκαν όταν η διεργασία βρέθηκε εκτός στατιστικού ελέγχου (λόγω μετατόπισης της αρχικής κατανομής των τιμών των παρατηρήσεων) και N_2 είναι η διακριτή τ.μ. που δηλώνει το εκτός–ελέγχου μήκος ροής στο δ.ε., η διακριτή τ.μ. $N_2 \mid X_{(b;m)} = x$ θα ακολουθεί τη γεωμετρική κατανομή με παράμετρο (πιθανότητα «επιτυχίας»)

$$
\begin{aligned}
p_2 = {} & P(Y_{(j;n)} > X_{(b;m)} \mid X_{(b;m)} = x, \text{η διεργασία εκτός ελέγχου}) = \\
& = P(F(Y_{(j;n)}) > F(X_{(b;m)}) \mid F(X_{(b;m)}) = F(x), \text{η διεργασία εκτός ελέγχου}) \\
& = P(F(Y_{(j;n)}) > U_{(b;m)} \mid U_{(b;m)} = s, \text{η διεργασία εκτός ελέγχου}) \\
& = P(F^{-1}(F(Y_{(j;n)})) > F^{-1}(U_{(b;m)}) \mid U_{(b;m)} = s, \text{η διεργασία εκτός ελέγχου}) \\
& = P(Y_{(j;n)} > F^{-1}(U_{(b;m)}) \mid U_{(b;m)} = s, \text{η διεργασία εκτός ελέγχου}) \\
& = P(G(Y_{(j;n)}) > G(F^{-1}(U_{(b;m)})) \mid U_{(b;m)} = s) = P(U_{(j;n)} > G(F^{-1}(U_{(b;m)})) \mid U_{(b;m)} = s) \\
& = P(U_{(j;n)} > G(F^{-1}(s)) \mid U_{(b;m)} = s) = P(U_{(j;n)} > G(F^{-1}(s))) = p_2(s),
\end{aligned}
$$

λόγω της ανεξαρτησίας και επειδή η α.σ.κ. F είναι γνησίως αύξουσα συνάρτηση και διατηρεί την ανισότητα $>$. Η τ.μ. $U_{(j;n)}$ θα ακολουθεί (βλέπε Στέγγος (2006)) την κατανομή Βήτα με παραμέτρους j και $n–j+1$, επομένως η πιθανότητα $p_2(s)$ θα είναι ίση με

$$p_2(s) = P(U_{(j;n)} > G(F^{-1}(s))) =$$

$$= \int_{G(F^{-1}(s))}^{1} f_{U_{(j;n)}}(w)\,dw \;= \frac{n!}{(j-1)!(n-j)!}\int_{G(F^{-1}(s))}^{1} w^{j-1}(1-w)^{n-j}\,dw$$

(βλέπε Κούτρας (2002) (α)).

Αποδεικνύεται (βλέπε Κούτρας (2001)) ότι η πιθανότητα $p_2(s)$ μπορεί να γραφεί στη μορφή

$$p_2(s) = \frac{n!}{(j-1)!(n-j)!}\sum_{i=0}^{n-j}\binom{n-j}{i}\frac{(-1)^i[1-(G(F^{-1}(s)))^{i+j}]}{(i+j)} =$$

$$= \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i[1-(G(F^{-1}(s)))^{i+j}]}{i!(n-j-i)!(i+j)}$$

και επομένως, σύμφωνα με όλα τα παραπάνω, θα ισχύει

$$P(N_2=k \mid X_{(b;m)}=x) = P(N_2=k \mid U_{(b;m)}=s) = [1-p_2(s)]^{k-1}[p_2(s)] = [q_2(s)]^{k-1}[1-q_2(s)]$$

$$= [q_2(s)]^{k-1}-[q_2(s)]^{k},$$

όπου

$$q_2(s) = 1-p_2(s) = 1-\frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i[1-(G(F^{-1}(s)))^{i+j}]}{i!(n-j-i)!(i+j)}.$$

Άρα, εφαρμόζοντας το Θ.Ο.Π., η σ.π. $P(N_2=k)$ της τ.μ. N_2 θα ισούται με:

$$P(N_2=k) = \int_{-\infty}^{\infty} P(N_2=k \mid X_{(b;m)}=x) f_{X_{(b;m)}}(x)dx = \int_0^1 P(N_2=k \mid U_{(b;m)}=s) f_{U_{(b;m)}}(s)ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 \{[q_2(s)]^{k-1}-[q_2(s)]^{k}\} s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 [q_2(s)]^{k-1} s^{b-1}(1-s)^{m-b}\,ds$$

$$- \frac{m!}{(b-1)!(m-b)!}\int_0^1 [q_2(s)]^{k} s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 \Big[1-\frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i\;[1-(G(F^{-1}(s)))^{i+j}]}{i!(n-j-i)!(i+j)}\Big]^{k-1} s^{b-1}(1-s)^{m-b}\,ds$$

$$- \frac{m!}{(b-1)!(m-b)!}\int_0^1 \Big[1-\frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i\;[1-(G(F^{-1}(s)))^{i+j}]}{i!(n-j-i)!(i+j)}\Big]^{k} s^{b-1}(1-s)^{m-b}\,ds\,,$$

$k = 1, 2, \ldots$ και εδώ, καταλήγουμε στο ότι, τα δύο παραπάνω ορισμένα διπλά ολοκληρώματα που εμφανίστηκαν δεν είναι δυνατόν να υπολογιστούν, αφού ο ακριβής συναρτησιακός τύπος των α.σ.κ. F και G δεν είναι γνωστός ή έστω προσεγγίσιμος. Επομένως, στο συγκεκριμένο μονόπλευρο δ.ε., προκύπτει το συμπέρασμα ότι η ακριβής σ.π. του συνολικού μήκους ροής δε μπορεί να υπολογιστεί και είναι απροσδιόριστη.

Ανάλογη διαδικασία ακολουθείται και για τον υπολογισμό της σ.π. του εκτός–ελέγχου μήκους ροής στην περίπτωση που στο διάγραμμα υπάρχει μόνο το κ.ο.ε. $LCL = X_{(\alpha;m)}$. Συγκεκριμένα, αν F είναι η α.σ.κ. των τιμών των αρχικών παρατηρήσεων που συλλέχθηκαν ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου, G είναι η νέα α.σ.κ. των τιμών των παρατηρήσεων που συλλέχθηκαν όταν η διεργασία βρέθηκε εκτός στατιστικού ελέγχου (λόγω μετατόπισης της αρχικής κατανομής των παρατηρήσεων) και N_3 είναι η διακριτή τ.μ. η οποία δηλώνει το εκτός–ελέγχου μήκος ροής στο δ.ε., η τ.μ. $N_3 \mid X_{(\alpha;m)} = x$ θα ακολουθεί τη γεωμετρική κατανομή με παράμετρο

$$\begin{aligned}
p_3 &= P(Y_{(j;n)} < X_{(\alpha;m)} \mid X_{(\alpha;m)} = x, \text{ η διεργασία εκτός ελέγχου}) = \\
&= P(F(Y_{(j;n)}) < F(X_{(\alpha;m)}) \mid F(X_{(\alpha;m)}) = F(x), \text{ η διεργασία εκτός ελέγχου}) \\
&= P(F(Y_{(j;n)}) < U_{(\alpha;m)} \mid U_{(\alpha;m)} = t, \text{ η διεργασία εκτός ελέγχου}) \\
&= P(F^{-1}(F(Y_{(j;n)})) < F^{-1}(U_{(\alpha;m)}) \mid U_{(\alpha;m)} = t, \text{ η διεργασία εκτός ελέγχου}) \\
&= P(Y_{(j;n)} < F^{-1}(U_{(\alpha;m)}) \mid U_{(\alpha;m)} = t, \text{ η διεργασία εκτός ελέγχου}) \\
&= P(G(Y_{(j;n)}) < G(F^{-1}(U_{(\alpha;m)})) \mid U_{(\alpha;m)} = t) = P(U_{(j;n)} < G(F^{-1}(U_{(\alpha;m)})) \mid U_{(\alpha;m)} = t) \\
&= P(U_{(j;n)} < G(F^{-1}(t)) \mid U_{(\alpha;m)} = t) = P(U_{(j;n)} < G(F^{-1}(t))) = p_3(t),
\end{aligned}$$

λόγω της ανεξαρτησίας και λόγω του ότι η α.σ.κ. F είναι γνησίως αύξουσα συνάρτηση και διατηρεί την ανισότητα $<$. Άρα, η πιθανότητα $p_3 = p_3(t)$, θα είναι ίση με

$$\begin{aligned}
p_3 = p_3(t) = P(U_{(j;n)} < G(F^{-1}(t))) &= \int_0^{G(F^{-1}(t))} f_{U_{(j;n)}}(w)dw \\
&= \frac{n!}{(j-1)!(n-j)!} \int_0^{G(F^{-1}(t))} w^{j-1}(1-w)^{n-j}\, dw,
\end{aligned}$$

επειδή η τ.μ. $U_{(j;n)}$ ακολουθεί την κατανομή Βήτα με παραμέτρους j και $n\text{–}j+1$.

Αποδεικνύεται (βλέπε Κούτρας (2001)) ότι η πιθανότητα $p_3(t)$ μπορεί να γραφτεί στην εξής μορφή:

$$p_3(t) = \frac{n!}{(j-1)!(n-j)!} \sum_{i=0}^{n-j} \binom{n-j}{i} \frac{(-1)^i [G(F^{-1}(t))]^{i+j}}{(i+j)} =$$

$$= \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i[G(F^{-1}(t))]^{i+j}]}{i!(n-j-i)!(i+j)}$$

και επομένως θα ισχύει, σύμφωνα με τα παραπάνω,

$$P(N_3 = k \mid X_{(\alpha;m)} = x) = P(N_3 = k \mid U_{(\alpha;m)} = t) = [1 - p_3(t)]^{k-1}[p_3(t)] = [q_3(t)]^{k-1}[1 - q_3(t)] =$$

$$= [q_3(t)]^{k-1} - [q_3(t)]^{k},$$

όπου

$$q_3(t) = 1 - p_3(t) = 1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i[G(F^{-1}(t))]^{i+j}]}{i!(n-j-i)!(i+j)}.$$

Άρα, εφαρμόζοντας το Θ.Ο.Π., η σ.π. $P(N_3 = k)$ της τ.μ N_3, θα ισούται με:

$$P(N_3 = k) = \int_{-\infty}^{\infty} P(N_3 = k \mid X_{(a;m)} = x) f_{X_{(a;m)}}(x)dx = \int_0^1 P(N_3 = k \mid U_{(a;m)} = t) f_{U_{(a;m)}}(t)dt =$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 \{[q_3(t)]^{k-1} - [q_3(t)]^{k}\} t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 [q_3(t)]^{k-1} t^{a-1}(1-t)^{m-a}\,dt$$

$$- \frac{m!}{(a-1)!(m-a)!}\int_0^1 [q_3(t)]^{k} t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1 [1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i\ [G(F^{-1}(t))]^{i+j}}{i!(n-j-i)!(i+j)}]^{k-1} t^{a-1}(1-t)^{m-a}\,dt$$

$$- \frac{m!}{(a-1)!(m-a)!}\int_0^1 [1 - \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i\ [G(F^{-1}(t))]^{i+j}}{i!(n-j-i)!(i+j)}]^{k} t^{a-1}(1-t)^{m-a}\,dt$$

και εδώ, ισχύουν ακριβώς τα ίδια αποτελέσματα και συμπεράσματα (ότι δηλαδή η ακριβής σ.π. του μήκους ροής δε μπορεί να υπολογιστεί και είναι απροσδιόριστη), όπως στο μονόπλευρο δ.ε. τύπου Shewhart στο οποίο υπήρχε μόνο το α.ο.ε. *UCL*.

4.5 Ο υπολογισμός του εκτός–ελέγχου μέσου μήκους ροής

Η διαδικασία υπολογισμού του εκτός–ελέγχου μέσου μήκους ροής ARL_{out} στο μονόπλευρο μη–παραμετρικό δ.ε. είναι ίδια με αυτή για το αντίστοιχο δίπλευρο διάγραμμα. Συγκεκριμένα, το ARL_{out} θα είναι η μέση τιμή της διακριτής τ.μ. N_1, η οποία δηλώνει το μήκος ροής στο δ.ε. όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου. Επειδή η τ.μ. N_1 παίρνει τις τιμές 1, 2, … , η μέση τιμή της $E(N_1)$ θα ισούται, σύμφωνα με θεώρημα της Θ.Π., με το άπειρο άθροισμα (βλέπε Κούτρας (2002) (α))

$$E(N_1) = \sum_{k=0}^{\infty} P(N_1 > k).$$

Αν στο διάγραμμα υπάρχει μόνο το α.ο.ε. $UCL = X_{(b;m)}$, εφαρμόζοντας το Θ.Ο.Π. για την πιθανότητα $P(N_1 > k)$ θα ισχύει

$$P(N_1 > k) = \int_{-\infty}^{\infty} P(N_1 > k \mid X_{(b;m)} = x) f_{X_{(b;m)}}(x)dx = \int_0^1 P(N_1 > k \mid U_{(b;m)} = s) f_{U_{(b;m)}}(s)ds,$$

$k = 0, 1, \ldots$.

Η τ.μ. $N_1 \mid U_{(b;m)} = s$ θα ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p_2(s)$, όπου $p_2(s)$ είναι η πιθανότητα που ορίστηκε και υπολογίστηκε αναλυτικά σε προηγούμενη ενότητα μέσω της σχέσης

$$p_2(s) = \frac{n!}{(j-1)!} \sum_{i=0}^{n-j} \frac{(-1)^i [1-(G(F^{-1}(t)))^{i+j}]}{i!(n-j-i)!(i+j)}.$$

Επομένως, (βλέπε Κούτρας (2002) (α)) θα ισχύει

$$P(N_1 > k \mid U_{(b;m)} = s) = [1 - p_2(s)]^k.$$

Οπότε, η πιθανότητα $P(N_1 > k)$, θα ισούται σύμφωνα με τα παραπάνω, με

$$P(N_1 > k) = \int_0^1 P(N_1 > k \mid U_{(b;m)} = s) f_{U_{(b;m)}}(s)ds =$$

$$= \frac{m!}{(b-1)!(m-b)!} \int_0^1 [1 - p_2(s)]^k s^{b-1} (1-s)^{m-b} ds,$$

$k = 0, 1, \ldots$ και επομένως, σύμφωνα με τα παραπάνω, προκύπτει για το ARL_{out} ότι

$$ARL_{out} = E(N_1) = \sum_{k=0}^{\infty} P(N_1 > k) = \sum_{k=0}^{\infty} \frac{m!}{(b-1)!(m-b)!} \int_0^1 [1 - p_2(s)]^k s^{b-1} (1-s)^{m-b} ds$$

$$= \frac{m!}{(b-1)!(m-b)!} \sum_{k=0}^{\infty} \int_0^1 [1 - p_2(s)]^k s^{b-1} (1-s)^{m-b} ds$$

$$= \frac{m!}{(b-1)!(m-b)!} \int_0^1 \sum_{k=0}^{\infty} [1 - p_2(s)]^k s^{b-1} (1-s)^{m-b} ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1 \{\sum_{k=0}^{\infty}[1-p_2(s)]^k\} s^{b-1}(1-s)^{m-b}\,ds\ .$$

Επίσης, επειδή $0 < p_2(s) < 1$, θα ισχύει (βλέπε Thomas and Finney (1990)) ο μαθηματικός τύπος

$$\sum_{k=0}^{\infty}[1-p_2(s)]^k \ = \ \frac{1}{1-[1-p_2(s)]} \ = \frac{1}{p_2(s)}\ ,\ p_2(s)\in(0,1)\,.$$

Αντικαθιστώντας τη συγκεκριμένη έκφραση στο παραπάνω ολοκλήρωμα και λαμβάνοντας υπ'όψιν ότι ισχύει η σχέση

$$p_2(s) \ = \ \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i[1-(G(F^{-1}(t)))^{i+j}]}{i!(n-j-i)!(i+j)},$$

παίρνουμε ότι

$$ARL_{out} = E(N_1) \ = \ \sum_{k=0}^{\infty}P(N_1>k) \ = \ \frac{m!}{(b-1)!(m-b)!}\int_0^1\{\sum_{k=0}^{\infty}[1-p_2(s)]^k\}s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{m!}{(b-1)!(m-b)!}\int_0^1\frac{1}{p_2(s)}s^{b-1}(1-s)^{m-b}\,ds$$

$$= \frac{(j-1)!}{n!}\frac{m!}{(b-1)!(m-b)!}\int_0^1\frac{s^{b-1}(1-s)^{m-b}}{\sum_{i=0}^{n-j}\frac{(-1)^i[1-(G(F^{-1}(s)))^{i+j}]}{i!(n-j-i)!(i+j)}}ds$$

και εδώ, είναι μάλλον εύκολο να διαπιστωθεί ότι, το παραπάνω ορισμένο ολοκλήρωμα, δηλαδή το εκτός–ελέγχου μέσο μήκος ροής ARL_{out}, δε μπορεί να υπολογιστεί, αφού δεν είναι γνωστός ο ακριβής συναρτησιακός τύπος των α.σ.κ. F και G, οι οποίες εμφανίζονται στο παραπάνω ορισμένο ολοκλήρωμα. Άρα, στο μονόπλευρο δ.ε., το εκτός–ελέγχου μέσο μήκος ροής ARL_{out} είναι άγνωστο (απροσδιόριστο).

Στην περίπτωση που στο δ.ε. υπάρχει μόνο το κ.ο.ε. $LCL = X_{(\alpha;m)}$, η διαδικασία υπολογισμού του εκτός–ελέγχου μέσου μήκους ροής είναι ακριβώς η ίδια με αυτή για το α.ο.ε.. Συγκεκριμένα, το ARL_{out} θα είναι (βλέπε Αντζουλάκος (2006)) η μέση τιμή της διακριτής τ.μ. N_2, η οποία δηλώνει το μήκος ροής στο δ.ε. όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου. Επειδή η τ.μ. N_2 παίρνει τις τιμές 1, 2, ... , η μέση τιμή της $E(N_2)$ θα ισούται, σύμφωνα με θεώρημα της Θ.Π., με

$$ARL_{out} = E(N_2) = \sum_{k=0}^{\infty} P(N_2 > k).$$

Η πιθανότητα $P(N_2 > k)$, σύμφωνα με το Θ.Ο.Π., θα είναι ίση με:

$$P(N_2 > k) = \int_{-\infty}^{\infty} P(N_2 > k \mid X_{(a;m)} = x) f_{X_{(a;m)}}(x)dx = \int_{0}^{1} P(N_2 > k \mid U_{(a;m)} = t) f_{U_{(a;m)}}(t)dt,$$

$k = 0, 1, \ldots$.

Η τ.μ. $N_2 \mid U_{(\alpha;m)} = t$ θα ακολουθεί τη γεωμετρική κατανομή με παράμετρο ίση με $p_3(t)$, όπου $p_3(t)$ είναι η πιθανότητα που ορίστηκε και υπολογίστηκε αναλυτικά σε προηγούμενη ενότητα, μέσω της σχέσης

$$p_3(t) = \frac{n!}{(j-1)!} \sum_{i=0}^{n-j} \frac{(-1)^i [G(F^{-1}(t))]^{i+j}}{i!(n-j-i)!(i+j)}.$$

Επομένως, θα ισχύει

$$P(N_2 > k \mid U_{(a;m)} = t) = [1 - p_3(t)]^k.$$

Οπότε, η πιθανότητα $P(N_2 > k)$, θα ισούται σύμφωνα με τα παραπάνω, με

$$P(N_2 > k) = \int_{0}^{1} P(N_2 > k \mid U_{(a;m)} = t) f_{U_{(a;m)}}(t)dt =$$

$$= \frac{m!}{(a-1)!(m-a)!} \int_{0}^{1} [1 - p_3(t)]^k t^{a-1} (1-t)^{m-a} dt,$$

$k = 0, 1, \ldots$ και επομένως, για το ARL_{out} σύμφωνα με τα παραπάνω, προκύπτει ότι

$$ARL_{out} = E(N_2) = \sum_{k=0}^{\infty} P(N_2 > k) = \sum_{k=0}^{\infty} \frac{m!}{(a-1)!(m-a)!} \int_{0}^{1} [1 - p_3(t)]^k t^{a-1} (1-t)^{m-a} dt$$

$$= \frac{m!}{(a-1)!(m-a)!} \sum_{k=0}^{\infty} \int_{0}^{1} [1 - p_3(t)]^k t^{a-1} (1-t)^{m-a} dt$$

$$= \frac{m!}{(a-1)!(m-a)!} \int_{0}^{1} \sum_{k=0}^{\infty} [1 - p_3(t)]^k t^{a-1} (1-t)^{m-a} dt$$

$$= \frac{m!}{(a-1)!(m-a)!} \int_{0}^{1} \{\sum_{k=0}^{\infty} [1 - p_3(t)]^k\} t^{a-1} (1-t)^{m-a} dt.$$

Χρησιμοποιώντας το μαθηματικό τύπο

$$\sum_{k=0}^{\infty}[1-p_3(t)]^k = \frac{1}{1-[1-p_3(t)]} = \frac{1}{p_3(t)}, \; 0 < p_3(t) < 1,$$

αντικαθιστώντας την έκφραση αυτή στο παραπάνω ολοκλήρωμα και λαμβάνοντας υπ'όψιν ότι ισχύει η σχέση

$$p_3(t) = \frac{n!}{(j-1)!}\sum_{i=0}^{n-j}\frac{(-1)^i[G(F^{-1}(t))]^{i+j}}{i!(n-j-i)!(i+j)},$$

προκύπτει ότι:

$$ARL_{out} = E(N_2) = \sum_{k=0}^{\infty}P(N_2>k) = \frac{m!}{(a-1)!(m-a)!}\int_0^1\{\sum_{k=0}^{\infty}[1-p_3(t)]^k\}t^{a-1}(1-t)^{m-a}\,dt =$$

$$= \frac{m!}{(a-1)!(m-a)!}\int_0^1\frac{1}{p_3(t)}t^{a-1}(1-t)^{m-a}\,dt$$

$$= \frac{m!}{(a-1)!(m-a)!}\frac{(j-1)!}{n!}\int_0^1\frac{t^{a-1}(1-t)^{m-a}}{\sum_{i=0}^{n-j}\frac{(-1)^i[G(F^{-1}(t))]^{i+j}}{i!(n-j-i)!(i+j)}}dt.$$

Για το παραπάνω ορισμένο ολοκλήρωμα που εμφανίστηκε, ισχύουν ακριβώς τα ίδια αποτελέσματα και συμπεράσματα, όπως για το αντίστοιχο ολοκλήρωμα που εμφανίστηκε στην περίπτωση που στο μονόπλευρο δ.ε. τύπου Shewhart υπήρχε μόνο το α.ο.ε. *UCL* (ότι δηλαδή αυτό το ορισμένο ολοκλήρωμα είναι απροσδιόριστο).

4.6 Άλλα μη–παραμετρικά δ.ε. τύπου Shewhart

Στην παράγραφο αυτή, θα παρουσιάσουμε με συντομία, ορισμένα άλλα μη–παραμετρικά δ.ε. τύπου Shewhart, τα οποία έχουν εντοπισθεί σχετικά πρόσφατα στη διεθνή βιβλιογραφία.

4.6.1 Μη–παραμετρικά δ.ε. τύπου Shewhart που χρησιμοποιούν απλές (μη διατεταγμένες) παρατηρήσεις της υπό έλεγχο διεργασίας

Εκτός από τα μη–παραμετρικά δ.ε. που πρότειναν (όπως αναφέρθηκε) οι Chakraborti, Van Der Laan and Van De Wiel (2004), τα οποία χρησιμοποιούν το διατεταγμένο τ.δ. των παρατηρήσεων, έχουν προταθεί στην πρόσφατη βιβλιογραφία από τους Willemain and Runger (1996) και κάποια άλλα μη–παραμετρικά δ.ε. που βασίζονται στο διατεταγμένο τ.δ. των παρατηρήσεων. Αυτά, χρησιμοποιούν για ο.ε. δύο διατεταγμένες παρατηρήσεις $X_{(k)}$ και $X_{(b+k)}$, οι οποίες προκύπτουν από ένα αρχικό διατεταγμένο τ.δ. που συλλέγεται σε χρονική στιγμή κατά την οποία η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Όμως, η βασική διαφορά σε σύγκριση με τα προηγούμενα μη–παραμετρικά δ.ε. είναι ότι, δε συγκρίνεται με τα ο.ε. κάποια διατεταγμένη παρατήρηση από ένα νέο τ.δ. (όπως συνέβαινε προηγουμένως), αλλά μία απλή παρατήρηση Y_i, που προκύπτει από τη συλλογή ενός τελικού τ.δ.. Όπως στα δ.ε. των Chakraborti, Van Der Laan and Van De Wiel (2004), έτσι και σε αυτά, η διεργασία θεωρείται ότι βρίσκεται εκτός στατιστικού ελέγχου αν μία απλή παρατήρηση βρεθεί έξω από τα δύο ή το ένα ο.ε. $LCL = X_{(k)}$ και $UCL = X_{(b+k)}$, ενώ η διεργασία θεωρείται ότι βρίσκεται εντός στατιστικού ελέγχου αν μία απλή παρατήρηση βρεθεί εντός των δύο ο.ε. $LCL= X_{(k)}$ και $UCL= X_{(b+k)}$ ή του ενός ο.ε. Έχει αποδειχθεί ότι, σε αυτά τα δ.ε., η ακριβής κατανομή του εντός–ελέγχου μήκους ροής είναι μία υπεργεωμετρική κατανομή με σχετικά «βαριά» δεξιά ουρά. Αποδεικνύεται επίσης ότι, σε αυτό το μη–παραμετρικό διάγραμμα, το εντός–ελέγχου μέσο μήκος ροής ARL_{in} θα είναι ίσο με

$$ARL_{in} = \frac{m}{m-b},$$

όπου m είναι το μέγεθος του αρχικού τ.δ.. Στην πράξη, συνήθως, οι δύο παράμετροι k και b υπολογίζονται σύμφωνα με την παρακάτω διαδικασία: αρχικά, προσδιορίζεται μια επιθυμητή «αρκετά μεγάλη» τιμή για το εντός–ελέγχου μέσο μήκος ροής ARL_{in} και για δοθέν μέγεθος m του αρχικού τ.δ., η παράμετρος b προσδιορίζεται αυτόματα από την επίλυση ως προς b της σχέσης $ARL_{in} = \frac{m}{m-b}$. Στη συνέχεια, η άλλη παράμετρος k, προσδιορίζεται υποθέτοντας ότι αυτό το δίπλευρο μη–παραμετρικό δ.ε. είναι συμμετρικό, δηλαδή ότι τα δύο ο.ε. είναι μεταξύ τους συμμετρικά ως προς το αρχικό τ.δ. Αυτή η συμμετρικότητα των δύο ορίων οδηγεί τελικά στη σχέση

$$k = m-(b+k)+1,$$

από την οποία προκύπτει ότι, η παράμετρος k θα είναι ίση με

$$k = \frac{m-b+1}{2}.$$

Για παράδειγμα, έστω ότι το συνολικό μέγεθος του προκαταρκτικού τ.δ. είναι m = 300 παρατηρήσεις και ότι είναι επιθυμητή, αρχικά, μία τιμή για το εντός–ελέγχου μέσο μήκος ροής ARL_{in} ίση με ARL_{in} = 100. Τότε, από την σχέση $ARL_{in} = \frac{m}{m-b}$, προκύπτει ότι η παράμετρος b θα είναι ίση με b = 297 και για να είναι μεταξύ τους συμμετρικά τα δύο ο.ε. στο διάγραμμα. θα πρέπει να ισχύει η σχέση $k = \frac{m-b+1}{2}$, οπότε η παράμετρος k θα είναι ίση με k = 2. Επομένως, στο συγκεκριμένο μη–παραμετρικό δ.ε., το κ.ο.ε. θα είναι ίσο με $LCL = X_{(k)} = X_{(2)}$ και το α.ο.ε. με $UCL = X_{(b+k)} = X_{(299)}$. Με αυτά τα δύο ο.ε. λοιπόν, θα συγκρίνεται μία οποιαδήποτε απλή παρατήρηση Y από ένα τελικό τ.δ., ανεξάρτητο από το αρχικό τ.δ..

Τα συγκεκριμένα μη–παραμετρικά δ.ε. εμφανίζουν πλεονεκτήματα ανάλογα με εκείνα των μη–παραμετρικών δ.ε. που προτάθηκαν από τους Chakraborti, Van Der Laan and Van De Wiel (2004). Το κυριότερο μειονέκτημά τους είναι ότι, για να επιτευχθεί ένα αρκετά μεγάλο εντός–ελέγχου μέσο μήκος ροής ARL_{in} ή μία αρκετά μικρή πιθανότητα λανθασμένου συναγερμού *FAR*, απαιτείται συνήθως πολύ μεγάλο μέγεθος δείγματος m (συνήθως πάνω από 300 παρατηρήσεις στο αρχικό τ.δ.), κάτι που προφανώς είναι οικονομικά ασύμφορο, δαπανηρό και σε μερικές περιπτώσεις εξαιρετικά δύσκολο ή και ανέφικτο. Ακόμα όμως και αν είναι πρακτικά εφικτό να συγκεντρωθεί μεγάλο δείγμα, συνήθως οι δύο διατεταγμένες παρατηρήσεις $LCL = X_{(k)}$ και $UCL = X_{(b+k)}$ θα απέχουν πολύ μεταξύ τους (στο παράδειγμα που αναφέρθηκε πριν, για m = 300, τα ο.ε. ήταν η 2η και η 299η κατά σειρά παρατήρηση) και επομένως η πιθανότητα σφάλματος τύπου 2 (δηλαδή η πιθανότητα η διεργασία να βρεθεί εκτός στατιστικού ελέγχου, αλλά το δ.ε. λανθασμένα να μην δώσει αυτήν την ένδειξη) είναι πολύ πιθανό να πάρει μεγάλη τιμή, κάτι που μάλλον δεν είναι επιθυμητό, στα πλαίσια του Στατιστικού Ελέγχου Διεργασίας.

4.6.2 Μη–παραμετρικά δ.ε. με χρήση της εκτιμήτριας Hodges–Lehmann

Τα συγκεκριμένα μη–παραμετρικά δ.ε. προτάθηκαν από τους Alloway and Raghavachari (1991) και παρακολουθούν τη διάμεσο μιας συνεχούς συμμετρικής κατανομής, χρησιμοποιώντας μη–παραμετρικά διαστήματα εμπιστοσύνης για αυτήν, τα οποία δημιουργούνται μέσω της εκτιμήτριας Hodges–Lehmann. Η τελευταία, υπολογίζεται σύμφωνα με την παρακάτω διαδικασία: έστω ότι από τη διεργασία λαμβάνονται m συνολικά ανεξάρτητα μεταξύ τους τ.δ. και το καθένα από αυτά είναι μεγέθους n. Για το i τ.δ. $X_{i1}, X_{i2}, ..., X_{in}$, υπολογίζονται οι μέσοι όροι των n παρατηρήσεων όταν λαμβάνονται ανά δύο και συγκεκριμένα οι ποσότητες

$$W_{ir} = \frac{X_{ij} + X_{ih}}{2}, \text{ για } r = 1, 2, ..., M,\ h = 1, 2, \ldots, n \text{ και } j = 1, 2, \ldots, h,$$

οι οποίες ονομάζονται “Walsh averages“. Εδώ, το σύμβολο M, εκφράζει το πλήθος των επαναληπτικών συνδυασμών των n παρατηρήσεων του δείγματος, όταν αυτές επιλέγονται ανά δύο, με όλους τους δυνατούς τρόπους, δηλαδή θα ισχύει

$$M = \frac{n(n+1)}{2}.$$

Τότε, η εκτιμήτρια Hodges–Lehmann για το i τ.δ., θα είναι η δειγματική διάμεσος των M συνολικά τιμών “Walsh averages“. Αν $W_{i(1)}, W_{i(2)}, W_{i(3)}, ..., W_{i(M)}$ είναι οι M διατεταγμένες ποσότητες “Walsh averages“ για το i τ.δ., η εκτιμήτρια Hodges–Lehmann $\tilde{q}_i$ για το i τ.δ., θα δίνεται από τον τύπο $\tilde{q}_i = W_{i(\frac{M+1}{2})}$, αν ο αριθμός M είναι περιττός ή $\tilde{q}_i = \frac{1}{2}[W_{i(\frac{M}{2})} + W_{i(\frac{M}{2}+1)}]$, αν ο αριθμός M είναι άρτιος. Ένα μη–παραμετρικό 100(1–α)% διάστημα εμπιστοσύνης για τη διάμεσο θ του πληθυσμού που βασίζεται στο i δείγμα, θα είναι ένα διάστημα της μορφής $[W_{i(a_i)}, W_{i,(M-a_i+1)}]$, για το οποίο θα ισχύει η εξής σχέση:

$$P(W_{i(a_i)} \le q \le W_{i(M-a_i+1)}) = 1 – \alpha,$$

όπου α_i είναι μία θετική παράμετρος που παίρνει τις τιμές 1, 2,…, M–1, M και α είναι μία θετική ποσότητα με αρκετά μικρή τιμή (που εκφράζει την πιθανότητα του στατιστικού σφάλματος για το διάστημα εμπιστοσύνης).

Τα δύο ο.ε. για την παρακολούθηση της διαμέσου θ του πληθυσμού, προσδιορίζονται ως εξής: το κ.ο.ε. θα είναι η δειγματική διάμεσος από τα m συνολικά κάτω όρια

$W_{1(a_1)}, W_{2(a_2)}, ..., W_{m(a_m)}$ των διαστημάτων εμπιστοσύνης που υπολογίστηκαν από τα m συνολικά τ.δ., το α.ο.ε. θα είναι η δειγματική διάμεσος από τα m συνολικά άνω όρια $W_{1(M-a_1+1)}$, $W_{2(M-a_2+1)}$, ..., $W_{m(M-a_m+1)}$ των διαστημάτων εμπιστοσύνης που υπολογίστηκαν από τα m συνολικά τ.δ. και η κεντρική γραμμή στο δ.ε. θα είναι ο δειγματικός μέσος όρος από τις m εκτιμήτριες Hodges–Lehmann που υπολογίστηκαν από όλα τα νέα τ.δ..

Στη συνέχεια, για κάθε ένα από τα m συνολικά ανεξάρτητα τ.δ., σχεδιάζεται το δ.ε. για την παρακολούθηση της διαμέσου θ του πληθυσμού και οι δειγματικές εκτιμήτριες Hodges–Lehmann $\tilde{q}_i$ συγκρίνονται με τα δύο ο.ε. που αναφέρθηκαν παραπάνω. Αν για κάποιο i τ.δ. διαπιστωθεί ότι η δειγματική εκτιμήτρια Hodges–Lehmann $\tilde{q}_i$ βρίσκεται έξω από τα δύο ο.ε., τότε προκύπτει το συμπέρασμα ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, επειδή το δ.ε. έδωσε γραφική ένδειξη ότι η διάμεσος θ του πληθυσμού μετατοπίστηκε (είτε αυξήθηκε, είτε μειώθηκε). Εάν όμως και για τα m συνολικά τ.δ., οι δειγματικές εκτιμήτριες Hodges–Lehmann $\tilde{q}_i$ βρίσκονται ανάμεσα στα δύο ο.ε., τότε προκύπτει το συμπέρασμα ότι η διεργασία βρίσκεται εντός στατιστικού ελέγχου και ότι η διάμεσος δε μεταβλήθηκε.

Για αρκετά μεγάλα μεγέθη δειγμάτων, αποδεικνύεται ότι αυτό το μη–παραμετρικό δ.ε. που παρακολουθεί τη διάμεσο θ του πληθυσμού, είναι (στατιστικά) ισοδύναμο με ένα κλασικό παραμετρικό δ.ε. τύπου Shewhart με 3σ ο.ε..

Τα κυριότερα πλεονεκτήματα για το συγκεκριμένο μη–παραμετρικό δ.ε. είναι ότι, δεν απαιτεί οι παρατηρήσεις που συλλέγονται να ακολουθούν την κανονική κατανομή και είναι ανθεκτικό απέναντι σε έκτροπες παρατηρήσεις. Από την άλλη πλευρά, υπάρχει το μειονέκτημα ότι απαιτεί η (συνεχής) κατανομή των παρατηρήσεων να είναι συμμετρική γύρω από τη διάμεσο ή τη μέση τιμή της κατανομής, ενώ αυτό είναι πιθανό να μην ισχύει. Επίσης, ένα άλλο σημαντικό μειονέκτημα είναι ότι, η ακριβής πιθανότητα λανθασμένου συναγερμού ***FAR***, καθώς και το εντός–ελέγχου μέσο μήκος ροής $\boldsymbol{ARL_{in}}$, δε μπορούν να υπολογιστούν. Αυτό, οφείλεται κυρίως στο γεγονός ότι, τα ο.ε. στο διάγραμμα δε βασίζονται στην εντός–ελέγχου κατανομή που ακολουθεί η εκτιμήτρια Hodges–Lehmann $\tilde{q}_i$. Ακόμη, δεν είναι αρκετά σαφές αν τα ο.ε. αυτού του δ.ε. χρησιμοποιούνται προοπτικά ή αναδρομικά, δηλαδή δεν είναι ξεκάθαρο αν τα ο.ε. υπολογίζονται και σχεδιάζονται για την παρακολούθηση του αρχικού δείγματος ή των τελικών δειγμάτων ή όλων συνολικά των δειγμάτων. Σε κάθε περίπτωση πάντως, για τα μη–παραμετρικά δ.ε. που προτάθηκαν από τους Alloway and Raghavachari, είναι μάλλον απαραίτητο να διεξαχθεί περισσότερη επιστημονική έρευνα και θεωρητική ανάλυση, ώστε να αναπτυχθεί περισσότερο η πρακτική εφαρμογή τους και η χρησιμότητά τους.

4.6.3 Μη–παραμετρικά δ.ε. με χρήση προσημικών στατιστικών συναρτήσεων

Τα συγκεκριμένα μη–παραμετρικά δ.ε. προτάθηκαν από τους Amin, Reynolds and Bakir (1995) και παρακολουθούν τη διάμεσο ή τη μέση τιμή μιας συνεχούς κατανομής. Χρησιμοποιούν τη γνωστή από την Απαραμετρική Στατιστική (βλέπε Στέγγος (2006)) προσημική στατιστική συνάρτηση (sign test statistic) για κάθε ένα από *m* ανεξάρτητα μεταξύ τους τ.δ. που συλλέγονται, τα οποία έχουν μέγεθος *n* παρατηρήσεις (το κάθε ένα). Η στατιστική συνάρτηση (τ.μ.) που χρησιμοποιείται είναι η εξής:

$$SN_i = \sum_{j=1}^{n} sign(X_{ij} - q_0)$$

για το *i* τ.δ., όπου θ_0 είναι η εκτίμηση για την αρχική, εντός–ελέγχου τιμή της παραμέτρου του πληθυσμού και sign( ) είναι μία διακριτή τ.μ., η οποία παίρνει την τιμή –1, αν η παρατήρηση X_{ij} είναι μικρότερη από την εκτιμημένη παράμετρο θ_0, την τιμή +1, αν η παρατήρηση X_{ij} είναι μεγαλύτερη από το θ_0 και την τιμή 0, αν η παρατήρηση X_{ij} είναι ακριβώς ίση με το θ_0.

Προφανώς, αφού η κατανομή που ακολουθούν τα δεδομένα είναι συνεχής, η πιθανότητα μία παρατήρηση X_{ij} να είναι ακριβώς ίση με την εντός–ελέγχου τιμή θ_0 θα είναι ίση με μηδέν, δηλαδή η πιθανότητα η διακριτή τ.μ. sign( ) να πάρει την τιμή 0 θα είναι ίση με μηδέν (βλέπε Κούτρας (2002) (α)). Επομένως, στο συνολικό τ.δ. των παρατηρήσεων, πιθανοθεωρητικά, δεν αναμένεται καμία παρατήρηση X_{ij} να είναι ακριβώς ίση με την τιμή θ_0.

Πρέπει εδώ να σημειωθεί ότι, τα συγκεκριμένα μη–παραμετρικά δ.ε., δεν απαιτούν η κατανομή των παρατηρήσεων να είναι συμμετρική γύρω από τη μέση τιμή ή τη διάμεσο της κατανομής. Η στατιστική συνάρτηση (τ.μ.) SN_i που αναφέρθηκε παραπάνω, συνδέεται άμεσα με τη διακριτή τ.μ. K_i, η οποία δηλώνει το συνολικό αριθμό των παρατηρήσεων X_{ij}

στο *i* τ.δ., οι οποίες είναι μεγαλύτερες από την τιμή θ_0. Συγκεκριμένα, οι δύο (διακριτές) τ.μ. SN_i και K_i συνδέονται μέσω της σχέσης

$$2K_i = SN_i + n\,.$$

Έτσι, η εντός–ελέγχου σ.π. της τ.μ. SN_i, μπορεί να υπολογιστεί σχετικά εύκολα, αν βεβαίως είναι γνωστή η εντός–ελέγχου σ.π. της τ.μ. K_i.

Όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου και η παράμετρος του πληθυσμού που παρακολουθείται είναι η διάμεσος, αποδεικνύεται (βλέπε Amin, Reynolds and Bakir (1995)) ότι η τ.μ. K_i θα ακολουθεί τη διωνυμική κατανομή με παραμέτρους *n* και $p = \dfrac{1}{2}$. Επομένως, αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η σ.π. της διακριτής τ.μ. SN_i θα είναι γνωστή και θα προκύπτει έμμεσα από την διωνυμική κατανομή την οποία ακολουθεί η τ.μ. K_i.

Τα συγκεκριμένα μη–παραμετρικά δ.ε. χρησιμοποιούν ο.ε. τύπου Shewhart, μαζί με προειδοποιητικά ο.ε. (warning limits) και κανόνες ροών (runs rules). Αν a_1, a_2 είναι τα δύο ο.ε., w_1, w_2 τα δύο προειδοποιητικά ο.ε. (εάν υπάρχουν), r είναι ο αριθμός των συνεχόμενων σημείων (ροών) που είναι πιθανό να βρεθούν μεταξύ ενός προειδοποιητικού ο.ε. και ενός ο.ε. (και συνεπώς να δώσουν γραφική ένδειξη για εκτός ελέγχου διεργασία) και *n* είναι το συνολικό μέγεθος του κάθε τ.δ., για δοθείσες τιμές των παραμέτρων a_1, a_2, w_1, w_2, r και *n*, έχουν κατασκευαστεί ειδικοί στατιστικοί πίνακες (βλέπε Amin, Reynolds and Bakir (1995)), από τους οποίους υπολογίζονται, σχετικά εύκολα, η ακριβής πιθανότητα λανθασμένου συναγερμού *FAR*, καθώς και το εντός–ελέγχου μέσο μήκος ροής ARL_{in} για αυτό το δ.ε..

Στην πράξη, συνήθως, τα ο.ε. ή και τα προειδοποιητικά ο.ε., υπολογίζονται σύμφωνα με την ακόλουθη διαδικασία: αρχικά, προσδιορίζεται μέσω των ειδικών πινάκων που αναφέρθηκαν παραπάνω, μία αρκετά μικρή και επιθυμητή τιμή για την πιθανότητα λανθασμένου συναγερμού ή μια αρκετά μεγάλη και επιθυμητή τιμή για το εντός–ελέγχου μέσο μήκος ροής και στη συνέχεια, σύμφωνα με αυτές τις προκαθορισμένες τιμές, προσδιορίζονται οι υπόλοιπες παράμετροι a_1, a_2, w_1, w_2 και *r*. Τελικά, μετά τον υπολογισμό και των τελευταίων παραμέτρων, κατασκευάζεται το συγκεκριμένο μη–παραμετρικό δ.ε. για την παρακολούθηση της παραμέτρου θ (δηλαδή της διαμέσου ή της μέσης τιμής) της κατανομής των δεδομένων.

Οι Amin, Reynolds and Bakir (1995) παρουσιάζουν επίσης, ένα μη–παραμετρικό δ.ε. για την παρακολούθηση της διασποράς (δηλαδή της μεταβλητότητας) της διεργασίας, χρησιμοποιώντας έναν ειδικό έλεγχο για το ενδοτεταρτημοριακό εύρος (interquartile range) του συνολικού τ.δ. που συλλέγεται. Μέχρι σήμερα πάντως, δεν έχει αναπτυχθεί αρκετά μεγάλη έρευνα, ούτε θεωρητική ούτε και πρακτική, πάνω στα μη–παραμετρικά δ.ε. που παρακολουθούν τη διασπορά μιας διεργασίας.

Οι ίδιοι συγγραφείς, έχουν επίσης προτείνει, να πραγματοποιείται περικομμένη δειγματοληψία για αυτό το προσημικό δ.ε., στην περίπτωση που η συνολική δειγματοληψία όλων των μονάδων κάθε τ.δ. παρουσιάζει μεγάλο κόστος και είναι οικονομικά ασύμφορη. Η περικομμένη δειγματοληψία που έχουν προτείνει, σημαίνει ουσιαστικά τη διακοπή της δειγματοληψίας πριν αυτή ολοκληρωθεί, για όλες τις μονάδες καθενός τ.δ. που συλλέγεται, δίνοντας έτσι τη δυνατότητα στους ερευνητές, αμέσως μετά από αυτήν τη διακοπή, να διαπιστώσουν αν η διεργασία συνεχίζει να βρίσκεται εντός στατιστικού ελέγχου ή αν βρέθηκε εκτός στατιστικού ελέγχου λόγω μετατόπισης της μέσης τιμής, της διαμέσου ή της διασποράς της (συνεχούς) κατανομής που ακολουθούν οι παρατηρήσεις που συλλέχθηκαν.

Το συγκεκριμένο δ.ε. παρουσιάζει αρκετά σημαντικά πλεονεκτήματα. Τα κυριότερα από αυτά είναι ότι, δεν προϋποθέτει οι παρατηρήσεις που συλλέγονται να ακολουθούν την κανονική κατανομή, δεν προϋποθέτει η κατανομή των παρατηρήσεων να είναι συμμετρική, παρουσιάζει ίδια και σταθερή πιθανότητα λανθασμένου συναγερμού *FAR*, καθώς επίσης ίδιο και σταθερό εντός–ελέγχου μέσο μήκος ροής ARL_{in} για όλες τις (συνεχείς) κατανομές που είναι πιθανό να ακολουθούν τα δεδομένα και ακόμη είναι εξαιρετικά ανθεκτικό απέναντι σε τυχόν έκτροπες παρατηρήσεις.

Εκτός από τα παραπάνω, έχει αποδειχθεί (θεωρητικά και πρακτικά) ότι, σε περίπτωση που τα δεδομένα ακολουθούν κάποια κατανομή με έντονη ασυμμετρία και «βαριά» δεξιά ουρά (όπως είναι για παράδειγμα οι κατανομές Γάμμα και Weibull), αυτό το δ.ε. είναι περισσότερο αποτελεσματικό και αποδοτικό σε σύγκριση με το πιο γνωστό παραμετρικό $\overline{X}$ τύπου Shewhart. Έτσι, σε περιπτώσεις κατανομών που εμφανίζουν τα στατιστικά χαρακτηριστικά που αναφέρθηκαν παραπάνω, το προσημικό μη–παραμετρικό δ.ε. παρουσιάζει μικρότερη πιθανότητα λανθασμένου συναγερμού, μεγαλύτερο εντός–ελέγχου μέσο μήκος ροής και μικρότερο εκτός–ελέγχου μέσο μήκος ροής συγκριτικά με το (πιο γνωστό) παραμετρικό δ.ε.

τύπου Shewhart. Επίσης, πλεονεκτεί στο γεγονός ότι, η στατιστική συνάρτηση (διακριτή τ.μ.) SN_i που χρησιμοποιείται είναι αρκετά απλή και εύκολη στον υπολογισμό της, αφού περιλαμβάνει μόνο τις αποκλίσεις $X_{ij} - q_0$ και προσθαφαιρέσεις μέσω των προσήμων +1, –1 και όχι πολύπλοκες μαθηματικές συναρτήσεις και δύσκολους υπολογισμούς.

Τέλος, σε αυτό το σημείο, πρέπει να σημειωθεί ότι, τα ο.ε. σε αυτό το δ.ε., υπολογίζονται σχετικά εύκολα, με χρήση όμως ειδικών πινάκων (βλέπε Amin, Reynolds and Bakir (1995)), στην περίπτωση φυσικά που είναι γνωστή (αρχικά προσδιορισμένη) η πιθανότητα λανθασμένου συναγερμού ή το εντός–ελέγχου μέσο μήκος ροής.

Από την άλλη πλευρά, το βασικότερο μειονέκτημα για το συγκεκριμένο δ.ε. εντοπίζεται στο ότι δεν παρουσιάζει αρκετά μεγάλη στατιστική ισχύ και αξιοπιστία στην περίπτωση που αρκετές παρατηρήσεις X_{ij} είναι ακριβώς ίσες με την εντός–ελέγχου τιμή της διαμέσου θ_0 (δηλαδή στην περίπτωση που υπάρχουν αρκετοί δεσμοί). Επίσης, ένα άλλο μειονέκτημα είναι ότι, απαιτεί να είναι αρχικά γνωστή η εντός–ελέγχου παράμετρος θ_0, δηλαδή απαιτεί την αρχική εκτίμηση του θ_0, κάτι το οποίο, ανάλογα με τη στατιστική μορφή των δεδομένων σε ορισμένες περιπτώσεις δεν είναι εφικτό, ή είναι αρκετά δύσκολο να γίνει (για παράδειγμα, στην περίπτωση που τα δεδομένα είναι λογοκριμένα, περικομμένα ή μετατοπισμένα).

4.7 Γενικές παρατηρήσεις και σχόλια στα μη–παραμετρικά δ.ε. Shewhart

Δε θα ήταν λανθασμένος ο ισχυρισμός ότι, τα μη–παραμετρικά δ.ε. τύπου Shewhart α-ποτελούν ένα σχετικά πρόσφατο αποτέλεσμα εκτεταμένων και αναλυτικών επιστημονικών ερευνών πάνω στο Στατιστικό Έλεγχο Διεργασίας. Παρ’ολ’αυτά, γενικά, θεωρείται (βλέπε Chakraborti, Bakir and Van Der Laan (2001)) ότι υπάρχουν ακόμη πολλές προοπτικές μελλοντικής αναλυτικής έρευνας για αυτά τα διαγράμματα. Σε γενικές γραμμές, τα συγκεκριμένα δ.ε., παρουσιάζουν όλα τα πλεονεκτήματα και όλα τα μειονεκτήματα που αναφέρθηκαν στις προηγούμενες ενότητες. Οι πιο σημαντικές και χρήσιμες στατιστικές τους ιδιότητες είναι ότι, δεν υποθέτουν συγκεκριμένη κατανομή για τις παρατηρήσεις που συλλέγονται, τα στατιστικά χαρακτηριστικά τους δεν εξαρτώνται και δεν επηρεάζονται από την κατανομή που ακολουθούν τα δεδομένα και επίσης δεν επηρεάζονται σχεδόν καθόλου από τυχόν έκτροπες παρατηρήσεις. Όμως, τα συγκεκριμένα διαγράμματα, δε μπορούν να μειώσουν σημαντικά την πιθανότητα σφάλματος τύπου 2 (δηλαδή την πιθανότητα το δ.ε.

εσφαλμένα να μη δώσει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, όταν πράγματι βρίσκεται εκτός ελέγχου) στο διάγραμμα. Παρ'όλο που μπορούν να περιορίσουν την πιθανότητα λανθασμένου συναγερμού *FAR* σε ένα αρκετά χαμηλό επίπεδο, ωστόσο δε μπορούν να κάνουν το ίδιο για την αντίστοιχη πιθανότητα σφάλματος τύπου 2, αφού αν η μία από τις δύο πιθανότητες αυξηθεί, η άλλη θα μειωθεί και το αντίστροφο. Δηλαδή, σε αυτά τα διαγράμματα, η πιθανότητα σφάλματος τύπου 2 δε θα είναι μικρή. Επίσης, αυτά τα δ.ε., εμφανίζουν ένα ακόμη (στατιστικό) μειονέκτημα: αυτό, είναι ότι δε μπορούν να εντοπίσουν αρκετά εύκολα και έγκαιρα μικρές ή πολύ μικρές μεταβολές της κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται, στην περίπτωση που η κατανομή των δεδομένων είναι άγνωστη (βλέπε Chakraborti, Bakir and Van Der Laan (2001)). Συγκεκριμένα, τα διαγράμματα αυτά, δε δίνουν εύκολα γραφική ένδειξη ότι η διεργασία είναι εκτός στατιστικού ελέγχου, στην περίπτωση που υπάρχει σχετικά μικρή μετατόπιση της μέσης τιμής ή της διαμέσου ή κάποιου άλλου στατιστικού χαρακτηριστικού της (πληθυσμιακής) κατανομής των δεδομένων. Για να εξαλειφθεί το παραπάνω μειονέκτημα, έχουν προταθεί στην πρόσφατη βιβλιογραφία ορισμένα άλλα μη–παραμετρικά δ.ε., τα οποία είναι τα μη–παραμετρικά δ.ε. τύπου *CUSUM* και χρησιμοποιούνται για να ανιχνεύσουν έγκαιρα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών (πληθυσμιακών) παραμέτρων της κατανομής των δεδομένων (βλέπε Chakraborti, Bakir and Van Der Laan (2001)). Τα συγκεκριμένα δ.ε., θα παρουσιαστούν και θα αναλυθούν στο αμέσως επόμενο κεφάλαιο.

ΚΕΦΑΛΑΙΟ 5

Μη–παραμετρικά δ.ε. τύπου *CUSUM*

5.1 Δ.ε. που χρησιμοποιούν προσημικές–βαθμολογικές στατιστικές συναρτήσεις

Όπως αναφέρθηκε προηγουμένως, εκτός από τα μη–παραμετρικά δ.ε. τύπου Shewhart, στην πρόσφατη βιβλιογραφία έχουν προταθεί επίσης και τα μη–παραμετρικά δ.ε. τύπου *CUSUM*, τα οποία χρησιμοποιούνται (βλέπε Chakraborti, Bakir and Van Der Laan (2001)) κυρίως για την ανίχνευση μικρών ή πολύ μικρών μετατοπίσεων της κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται. Τα μη–παραμετρικά δ.ε. τύπου *CUSUM* που χρησιμοποιούν προσημικές–βαθμολογικές στατιστικές συναρτήσεις (signed–rank statistics) προτάθηκαν από τους Bakir and Reynolds (1979). Σε αυτά τα διαγράμματα, αρχικά συλλέγεται ένα τ.δ. αναφοράς, σε χρονική στιγμή κατά την οποία η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Στη συνέχεια, συλλέγονται και άλλα, ανεξάρτητα μεταξύ τους, τ.δ. και ανεξάρτητα επίσης από το αρχικό τ.δ. και για κάθε ένα από αυτά υπολογίζεται η στατιστική συνάρτηση ελέγχου (τ.μ.)

$$SR_i = \sum_{j=1}^{g} sign(X_{ij} - q_0) R_{ij} \ ,$$

όπου g είναι το μέγεθος των νέων τ.δ., θ_0 είναι η εκτίμηση για την εντός–ελέγχου τιμή κάποιας στατιστικής (πληθυσμιακής) παραμέτρου της κατανομής των αρχικών δεδομένων, που συλλέχθηκαν ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου, R_{ij} είναι ο βαθμός (rank) που αντιστοιχεί στην παρατήρηση $| X_{ij} - q_0 |$ για το i τ.δ. (δηλαδή είναι ο βαθμός που αντιστοιχεί για το i τ.δ. στην απόλυτη τιμή της απόκλισης της παρατήρησης X_{ij} από την εντός–ελέγχου εκτιμημένη παράμετρο θ_0) και sign( ) είναι μία στατιστική συνάρτηση (τ.μ.) που παίρνει την τιμή +1 αν η ποσότητα $X_{ij} - q_0$ είναι μεγαλύτερη από το μηδέν, την τιμή –1, αν η ποσότητα $X_{ij} - q_0$ είναι μικρότερη από το μηδέν και την τιμή 0, αν η ποσότητα $X_{ij} - q_0$ είναι ακριβώς ίση με το μηδέν.

Η στατιστική και πιθανοθεωρητική ανάλυση σε αυτό το μη–παραμετρικό δ.ε., βασίζεται στην προσημική– βαθμολογική στατιστική συνάρτηση του Wilcoxon (βλέπε Στέγγος (2006)). Η στατιστική συνάρτηση (τ.μ.) SR_i είναι άμεσα συνδεδεμένη με τη στατιστική συνάρτηση (τ.μ.) του Wilcoxon V_g, που δηλώνει το συνολικό άθροισμα των βαθμών μόνο για τις θετικές αποκλίσεις $X_{ij} - q_0$, μέσω της σχέσης

$$SR_i = 2V_g - \frac{g(g+1)}{2}.$$

Επομένως, η σ.π. της διακριτής τ.μ. SR_i, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, μπορεί να υπολογιστεί σχετικά εύκολα, γνωρίζοντας την εντός–ελέγχου σ.π. της διακριτής τ.μ. V_g. Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η σ.π. της τ.μ. V_g θα είναι γνωστή και θα μπορεί να βρεθεί από ειδικούς στατιστικούς πίνακες (βλέπε Bakir and Reynolds (1979)).

Το δ.ε. που περιγράφηκε παραπάνω είναι μη–παραμετρικό, αφού έχει βρεθεί (βλέπε Bakir and Reynolds (1979)) ότι, όταν η διεργασία είναι εντός στατιστικού ελέγχου, η σ.π. της τ.μ. SR_i δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις και αυτή είναι η ίδια, για οποιαδήποτε κατανομή είναι πιθανό να ακολουθούν αυτές.

Το συγκεκριμένο μη–παραμετρικό δ.ε., ως *CUSUM* διάγραμμα, απεικονίζει τα συσσωρευμένα αθροίσματα των τ.μ. SR_i από κάθε ένα από τα τ.δ. που συλλέγονται και δίνει ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου στο *n*–οστό τ.δ. στο οποίο ισχύει για πρώτη φορά η σχέση

$$\sum_{i=1}^{n}(SR_i - k) - \min_{0 \le m \le n} \sum_{i=1}^{m}(SR_i - k) \ge h$$

ή στο *n*–οστό τ.δ. στο οποίο ισχύει για πρώτη φορά η σχέση

$$\max_{0 \le m \le n} \sum_{i=1}^{m}(SR_i + k) - \sum_{i=1}^{n}(SR_i + k) \ge h,$$

όπου k και h είναι δύο θετικές παράμετροι. Οι δύο αυτές παράμετροι k και h, οι οποίες διαμορφώνουν την κατασκευή και την εφαρμογή αυτού του διαγράμματος, προσδιορίζονται ως εκείνες οι τιμές, οι οποίες για γνωστό μέγεθος g των (νέων) τ.δ., αντιστοιχούν σε ένα αρχικά προσδιορισμένο και αρκετά μεγάλο εντός–ελέγχου μέσο μήκος ροής ARL_{in}.

Για τον προσδιορισμό των παραμέτρων k και h, υπάρχουν ειδικοί αναλυτικοί πίνακες (βλέπε Bakir and Reynolds (1979)) οι οποίοι, για συγκεκριμένες τιμές των k και h και για δεδομένη τιμή της παραμέτρου g, παρέχουν την τιμή του εντός–ελέγχου μέσου μήκους ροής ARL_{in} που αντιστοιχεί σε αυτές. Για παράδειγμα (βλέπε Bakir and Reynolds (1979)), αν το μέγεθος των δειγμάτων είναι g = 6, τότε η επιλογή k =11 και h =10, οδηγεί σε ένα εντός–ελέγχου μέσο μήκος ροής ARL_{in} ίσο με $ARL_{in} \cong 301$, τιμή η οποία θεωρείται μάλλον αρκετά ικανοποιητική για το ARL_{in}.

Έχει διαπιστωθεί (βλέπε Bakir and Reynolds (1979)) με χρήση προσομοίωσης ότι, αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, το συγκεκριμένο δ.ε. παρουσιάζει μικρότερη πιθανότητα λανθασμένου συναγερμού *FAR* και μεγαλύτερο εντός–ελέγχου μέσο μήκος ροής ARL_{in}, σε σύγκριση με το αντίστοιχο παραμετρικό δ.ε. τύπου *CUSUM*, όπως και με το κλασικό παραμετρικό δ.ε. τύπου Shewhart. Επίσης, με την παραπάνω μέθοδο έχει βρεθεί ότι, όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, το συγκεκριμένο δ.ε. παρουσιάζει μικρότερη πιθανότητα σφάλματος τύπου 2 και μικρότερο εκτός ελέγχου μέσο μήκος ροής ARL_{out} σε σύγκριση με το αντίστοιχο παραμετρικό δ.ε. τύπου *CUSUM* και με το (πιο γνωστό) παραμετρικό δ.ε. τύπου Shewhart, δηλαδή ανιχνεύει πιο γρήγορα πιθανές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των παρατηρήσεων, αν αυτές οι αλλαγές συμβούν. Με λίγα λόγια, αυτό το μη–παραμετρικό δ.ε. είναι πιο αποτελεσματικό, αποδοτικό και ισχυρό από το αντίστοιχο παραμετρικό δ.ε. *CUSUM* και από το παραμετρικό δ.ε. Shewhart. Αυτά, είναι και τα σημαντικότερα στατιστικά πλεονεκτήματα που έχει το συγκεκριμένο δ.ε. τύπου *CUSUM*.

Όμως, ένα σημαντικό μειονέκτημα για το συγκεκριμένο μη–παραμετρικό δ.ε. είναι ότι, ο τύπος της στατιστικής συνάρτησης που χρησιμοποιείται, δε λαμβάνει υπ'όψιν την ακριβή αριθμητική τιμή των παρατηρήσεων (όπως συμβαίνει στα μη–παραμετρικά Shewhart δ.ε.), αλλά μόνο τους βαθμούς (ranks) των παρατηρήσεων και τα πρόσημα (signs) των αποκλίσεων των παρατηρήσεων από την εκτιμημένη παράμετρο θ_0. Αυτό έχει ως συνέπεια, να χάνεται στατιστική πληροφορία από τα τ.δ. που συλλέγονται. Ακόμη ένα μειονέκτημα, είναι ότι απαιτεί την αρχική εκτίμηση της εντός–ελέγχου παραμέτρου θ_0, κάτι που ίσως να μην είναι απλό ή εφικτό, ανάλογα με τη στατιστική μορφή του δείγματος που συλλέγεται (για παράδειγμα, στις περιπτώσεις των λογοκριμένων, μετατοπισμένων και περικομμένων δεδομένων). Ακόμα όμως και στην περίπτωση που αυτό μπορεί να γίνει, είναι πιθανό η εκτίμηση που προκύπτει για την παράμετρο θ_0 να μην είναι αρκετά αξιόπιστη και ακριβής, ιδιαίτερα όταν το μέγεθος του αρχικού τ.δ. δεν είναι αρκετά μεγάλο, ώστε να ισχύουν τα γνωστά ασυμπτωτικά στατιστικά αποτελέσματα.

5.2 Δ.ε. που χρησιμοποιούν διαδοχικούς βαθμούς

Εκτός από τα δ.ε. τύπου *CUSUM* που χρησιμοποιούν την προσημική–βαθμολογική στατιστική συνάρτηση του Wilcoxon, έχουν προταθεί στην πρόσφατη βιβλιογραφία και κάποια άλλα δ.ε. *CUSUM*, τα οποία χρησιμοποιούν διαδοχικούς βαθμούς (sequential ranks). Τα διαγράμματα αυτά, έχουν προταθεί από τον McDonald (1990) και έχουν το χαρακτηριστικό ότι δε χρησιμοποιούν κάποιο αρχικό δείγμα, αλλά μόνο ένα τελικό. Σε αυτό το (τελικό) τ.δ., για κάθε παρατήρηση X_i που συλλέγεται, αντιστοιχεί ένας διαδοχικός βαθμός R_i. Ο διαδοχικός βαθμός R_i, που είναι η στατιστική συνάρτηση ελέγχου σε αυτό το διάγραμμα, για μία παρατήρηση X_i, ορίζεται ως εξής:

$$R_i = 1 + \sum_{j=1}^{i-1} I(X_j < X_i),$$

όπου $I(\ )$ είναι η γνωστή δείκτρια συνάρτηση που παίρνει τις τιμές 0 και 1. Επίσης, για την παρατήρηση X_i, υπολογίζεται και η ποσότητα $U_i = \frac{R_i}{i+1}$, όπου R_i είναι ο διαδοχικός βαθμός που ορίστηκε παραπάνω. Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, έχει βρεθεί (βλέπε McDonald (1990)) ότι οι ποσότητες U_i είναι μεταξύ τους ανεξάρτητες τ.μ. και ακολουθούν τη διακριτή ομοιόμορφη κατανομή στο σύνολο $\{\frac{1}{i+1}, \frac{2}{i+1}, \ldots, \frac{i}{i+1}\}$. Στη συνέχεια, απεικονίζονται στο διάγραμμα οι ποσότητες

$$T_i = \max\{T_{i-1} + U_i - k, 0\}, \text{ για } i = 1, 2, \ldots,$$

όπου $T_0 = 0$ και k είναι μία θετική σταθερά.

Το συγκεκριμένο δ.ε., δίνει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, στην πρώτη παρατήρηση X_i για την οποία ισχύει $T_i \geq h$, όπου h είναι μία θετική σταθερά που δηλώνει το ο.ε. σε αυτό το διάγραμμα. Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, έχει διαπιστωθεί (βλέπε McDonald (1990)) ότι η πιθανότητα λανθασμένου συναγερμού *FAR* και το εντός–ελέγχου μέσο μήκος ροής ARL_{in}, δεν εξαρτώνται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από τις θετικές παραμέτρους k και h, επομένως αυτό το δ.ε. είναι μη–παραμετρικό.

Για την κατασκευή αυτού του δ.ε. (ή ισοδύναμα, για τον προσδιορισμό της θετικής παραμέτρου h), συνήθως προκαθορίζεται μία επιθυμητή και αρκετά χαμηλή τιμή για την πιθανότητα λανθασμένου συναγερμού ή μία επιθυμητή και αρκετά υψηλή τιμή για το εντός–ελέγχου μέσο μήκος ροής και με βάση αυτή, προσδιορίζονται οι παράμετροι k και h που απαιτούνται για την κατασκευή και εφαρμογή αυτού του δ.ε. *CUSUM*. Συγκεκριμένα, υπάρχουν αναλυτικοί στατιστικοί πίνακες (βλέπε McDonald (1990)) οι οποίοι, για δεδομένες τιμές των θετικών παραμέτρων k και h, παρέχουν την τιμή του εντός–ελέγχου μέσου μήκους ροής ARL_{in} που αντιστοιχεί σε αυτές. Σημειώνεται εδώ ότι, αυτό το δ.ε., ουσιαστικά απεικονίζει τους αθροιστικούς διαδοχικούς βαθμούς (cumulative sequential ranks) όλων των παρατηρήσεων που συλλέγονται, δηλαδή απεικονίζει τα μερικά αθροίσματα (partial sums) των διαδοχικών βαθμών που προκύπτουν από κάθε μία παρατήρηση που συλλέγεται και για αυτόν το λόγο (βλέπε Αντζουλάκος (2006)), το συγκεκριμένο διάγραμμα είναι ένα δ.ε. τύπου *CUSUM*.

Ένα βασικό πλεονέκτημα για το συγκεκριμένο δ.ε. είναι ότι, δε χρησιμοποιεί κάποιο αρχικό τ.δ. όπως τα προηγούμενα δ.ε., αλλά μόνο ένα τελικό τ.δ. για την κατασκευή του δ.ε.. Αυτό έχει ως συνέπεια μεγάλη εξοικονόμηση χρόνου και χρήματος και σημαντική μείωση του συνολικού κόστους. Ακόμη, αυτό το δ.ε. *CUSUM*, μπορεί να ανιχνεύσει αρκετά εύκολα και έγκαιρα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των δεδομένων, κάτι που είναι μάλλον επιθυμητό για όσους ασχολούνται συστηματικά με το Στατιστικό Έλεγχο Διεργασίας. Επίσης, αυτό το διάγραμμα είναι απλό και εύκολο στην κατασκευή του, αφού δεν περιλαμβάνει πολύπλοκους μαθηματικούς τύπους και δύσκολες αριθμητικές πράξεις. Εκτός από τα παραπάνω, οι δύο παράμετροι ελέγχου k και h, υπολογίζονται χωρίς ιδιαίτερη δυσκολία από ειδικούς στατιστικούς πίνακες, αν φυσικά δοθεί μία αρχική τιμή για την πιθανότητα λανθασμένου συναγερμού FAR ή για το εντός–ελέγχου μέσο μήκος ροής ARL_{in}.

Από την άλλη πλευρά, το βασικότερο μειονέκτημα για το συγκεκριμένο δ.ε. είναι ότι, ο τύπος της στατιστικής συνάρτησης που χρησιμοποιείται, δε λαμβάνει υπ’όψιν την ακριβή αριθμητική τιμή των παρατηρήσεων (όπως συμβαίνει στα μη–παραμετρικά δ.ε. τύπου Shewhart που αναλύθηκαν σε προηγούμενες ενότητες), αλλά χρησιμοποιεί μόνο τους διαδοχικούς βαθμούς των παρατηρήσεων. Αυτό, έχει ως συνέπεια στο συγκεκριμένο δ.ε., να χάνεται σημαντική στατιστική πληροφορία από το τελικό τ.δ. που συλλέγεται.

Ο Yashchin (1992) πρότεινε ένα μη–παραμετρικό δ.ε. τύπου *CUSUM* στην περίπτωση που η κατανομή που ακολουθούν τα δεδομένα είναι άγνωστη. Συγκεκριμένα, σε αυτό το δ.ε., ο Yashchin (βλέπε Yashchin (1992)) αναλύει το μήκος ροής και ορισμένα άλλα στατιστικά χαρακτηριστικά του διαγράμματος, αντικαθιστώντας την άγνωστη κατανομή που ακολουθούν οι παρατηρήσεις με την εμπειρική (δειγματική) συνάρτηση κατανομής (empirical distribution function) ενός αρχικού τ.δ.. Ο Yaschin, αναλύει διεξοδικά τα χαρακτηριστικά των στατιστικών συναρτήσεων ελέγχου που χρησιμοποιεί, εφαρμόζοντας κυρίως βασικές ιδιότητες ανελίξεων Markov και επίσης παρουσιάζει αρκετά χρήσιμα στατιστικά αποτελέσματα, που προκύπτουν με χρήση προσομοίωσης.

5.3 Δ.ε. που χρησιμοποιούν προσημικές στατιστικές συναρτήσεις

Οι Amin, Reynolds and Bakir (1995) πρότειναν ορισμένα άλλα μη–παραμετρικά *CUSUM* δ.ε. για την παρακολούθηση κάποιας στατιστικής παραμέτρου της κατανομής που ακολουθούν τα δεδομένα, τα οποία βασίζονται στη γνωστή από την Απαραμετρική Στατιστική προσημική στατιστική συνάρτηση (βλέπε Στέγγος (2006)). Σε αυτά τα δ.ε., χρησιμοποιείται η στατιστική συνάρτηση ελέγχου (τ.μ.)

$$SN_i = \sum_{j=1}^{n} sign(X_{ij} - q_0),$$

όπου n είναι το μέγεθος των τ.δ. που συλλέγονται, θ_0 είναι η εντός–ελέγχου εκτίμηση για τη στατιστική (πληθυσμιακή) παράμετρο της κατανομής των δεδομένων, η οποία εκτίμηση προέκυψε από ένα αρχικό τ.δ. που συλλέχθηκε ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου και sign( ) είναι η συνάρτηση (τ.μ.) προσήμου που ορίστηκε σε προηγούμενη ενότητα. Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η σ.π. της διακριτής τ.μ. SN_i μπορεί να υπολογιστεί σχετικά εύκολα με χρήση ειδικών στατιστικών πινάκων (βλέπε Amin, Reynolds and Bakir (1995)), οι οποίοι κατασκευάστηκαν με τη βοήθεια συγκεκριμένων αποτελεσμάτων της Απαραμετρικής Στατιστικής. Το συγκεκριμένο δ.ε., απεικονίζει τα συσσωρευμένα αθροίσματα των τ.μ. SN_i που προκύπτουν από κάθε ένα τ.δ. που συλλέγεται. Δίνει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, στο *r*–οστό τ.δ. για το οποίο ισχύει για πρώτη φορά η σχέση

$$\sum_{i=1}^{r}(SN_i - k) - \min_{0\le m\le r}\sum_{i=1}^{m}(SN_i - k) \ge h$$

ή στο *k*–οστό τ.δ. για το οποίο ισχύει για πρώτη φορά η σχέση

$$\max_{0\le m\le r}\sum_{i=1}^{m}(SN_i + k) - \sum_{i=1}^{r}(SN_i + k) \ge h\,,$$

όπου *k* και *h* είναι δύο θετικές παράμετροι. Οι δύο αυτές παράμετροι *k* και *h*, οι οποίες ρυθμίζουν την κατασκευή και την εφαρμογή αυτού του μη–παραμετρικού δ.ε., προσδιορίζονται ως εκείνες οι τιμές οι οποίες, για γνωστό μέγεθος *n* των τ.δ. που συλλέγονται, αντιστοιχούν σε ένα αρχικά προσδιορισμένο και αρκετά μεγάλο εντός–ελέγχου μέσο μήκος ροής ARL_{in}. Για τον προσδιορισμό των παραμέτρων *k* και *h*, υπάρχουν ειδικοί αναλυτικοί πίνακες (βλέπε Amin, Reynolds and Bakir (1995)) οι οποίοι, για συγκεκριμένες τιμές των παραμέτρων αυτών, παρέχουν την τιμή του ARL_{in} που αντιστοιχεί σε αυτές. Έχει αποδειχθεί, ότι (βλέπε Amin, Reynolds and Bakir (1995)) η σ.π. της τ.μ. SN_i, δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από το μέγεθος *n* των τ.δ.. Επομένως, το συγκεκριμένο δ.ε. *CUSUM* είναι μη–παραμετρικό.

Τα πλεονεκτήματα και τα μειονεκτήματα που εμφανίζει το συγκεκριμένο δ.ε. είναι, σε γενικές γραμμές, τα ίδια με αυτά που αναφέρθηκαν στην παράγραφο 5.1 για το δ.ε. *CUSUM* που χρησιμοποιούσε προσημικές–βαθμολογικές στατιστικές συναρτήσεις (βλέπε σελίδα 69).

5.4 Ένα άλλο μη–παραμετρικό δ.ε. τύπου *CUSUM*

Εκτός από τα προηγούμενα μη–παραμετρικά δ.ε. που αναφέρθηκαν, οι Bhattacharyya and Frierson (1981) πρότειναν ένα άλλο μη–παραμετρικό δ.ε. τύπου *CUSUM*, το οποίο χρησιμοποιείται στα πλαίσια του Στατιστικού Ελέγχου Διεργασίας για έναν συγκεκριμένο σκοπό: για να ανιχνεύσει εύκολα και αρκετά έγκαιρα μία αλλαγή (αν φυσικά αυτή συμβεί) στην κατανομή που ακολουθούν οι παρατηρήσεις. Συγκεκριμένα, στην πρόσφατη βιβλιογραφία, αναφέρεται ότι αυτό το μη–παραμετρικό δ.ε. εφαρμόζει τη στατιστική γνώση που υπάρχει σχετικά με το λεγόμενο πρόβλημα «αλλαγής σημείου» (change–point problem) (ή αλλιώς με το πρόβλημα «ανίχνευσης» (detection problem)) στο Στατιστικό Έλεγχο Διεργασίας (βλέπε Zacks (1991)).

Ειδικότερα, έστω $X_1, X_2, ..., X_N$ ένα τ.δ. από παρατηρήσεις που συλλέγονται και ότι η (συνεχής) κατανομή που αυτές ακολουθούν, μετατοπίζεται από την αρχική, εντός–ελέγχου α.σ.κ. *F* στην α.σ.κ. *G*, μετά από $[N\theta]$ το πλήθος παρατηρήσεις, όπου N είναι το συνολικό μέγεθος του τ.δ. που συλλέγεται και θ είναι μία άγνωστη θετική παράμετρος, η οποία παίρνει τιμές στο διάστημα (0,1). Ο βασικός στόχος αυτού του δ.ε. είναι, να ανιχνεύσει εύκολα και αρκετά έγκαιρα (με κάποια συγκεκριμένη στατιστική μέθοδο) την αλλαγή της κατανομής (από *F* σε *G*) των παρατηρήσεων, με σχετικά μικρό στατιστικό σφάλμα και λίγες λανθασμένες γραφικές ενδείξεις και χωρίς να υποτεθεί, αρχικά, κάποια συγκεκριμένη κατανομή για τα δεδομένα που συλλέγονται.

Αυτό το δ.ε. χρησιμοποιεί ως στατιστική συνάρτηση ελέγχου τους διαδοχικούς βαθμούς (sequential ranks)

$$R_i = 1 + \sum_{j=1}^{i-1} I(X_j < X_i)$$

των παρατηρήσεων $X_1, X_2, ...$ (η συγκεκριμένη στατιστική συνάρτηση αναφέρθηκε σε προηγούμενη ενότητα) και απεικονίζει γραφικά τα συσσωρευμένα αθροίσματα αυτών των διαδοχικών βαθμών, που προκύπτουν για κάθε μία μεμονωμένη παρατήρηση (individual observation) που συλλέγεται. Έχει βρεθεί (βλέπε Bhattacharyya and Frierson (1981)) ότι, αυτό το διάγραμμα, αποτελεί μία αρκετά χρήσιμη και ικανοποιητική εναλλακτική επιλογή για συγκεκριμένες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας σε περιπτώσεις που οι ερευνητές έχουν ως στόχο να ανιχνεύσουν μικρές ή πολύ μικρές μεταβολές των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι μεμονωμένες παρατηρήσεις.

Το σημαντικότερο πλεονέκτημα για το συγκεκριμένο μη–παραμετρικό δ.ε. τύπου *CUSUM* είναι ότι, σχεδόν πάντα, κατορθώνει να ανιχνεύσει έγκαιρα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι μεμονωμένες παρατηρήσεις. Επίσης, ένα άλλο πλεονέκτημα είναι ότι, δε χρησιμοποιεί κάποιο αρχικό τ.δ., παρά μόνο μεμονωμένες παρατηρήσεις και έτσι υπάρχει σημαντική εξοικονόμηση χρήματος και χρόνου.

Από την άλλη πλευρά, αυτό το δ.ε., συνήθως δεν καταφέρνει να ανιχνεύσει έγκαιρα και εύκολα μεγάλες μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις. Εκτός από αυτό, υποθέτει τη συλλογή μεμονωμένων παρατηρήσεων και όχι δειγμάτων και αυτό στην πράξη δεν είναι πάντοτε εφικτό ή εύκολο να πραγματοποιηθεί, για διάφορους πρακτικούς λόγους.

5.5 Γενικές παρατηρήσεις και σχόλια στα μη–παραμετρικά δ.ε. τύπου *CUSUM*

Τα μη–παραμετρικά δ.ε. τύπου *CUSUM*, κρίνονται αρκετά σημαντικά στο Στατιστικό Έλεγχο Διεργασίας, αφού μπορούν να ανιχνεύσουν αρκετά εύκολα και έγκαιρα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν τα δεδομένα. Αυτό, είναι ένα πολύ σημαντικό πλεονέκτημά τους, όσον αφορά τις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας. Αυτά τα διαγράμματα είναι μη–παραμετρικά, δηλαδή δεν προϋποθέτουν κάποια συγκεκριμένη κατανομή για τις παρατηρήσεις που συλλέγονται, παρουσιάζουν την ίδια πιθανότητα λανθασμένου συναγερμού ***FAR*** και το ίδιο εντός–ελέγχου μέσο μήκος ροής ARL_{in} για όλες τις πιθανές κατανομές που μπορεί να ακολουθούν τα δεδομένα, είναι αρκετά ανθεκτικά σε τυχόν έκτροπες παρατηρήσεις και ακόμη είναι σχετικά απλά και εύκολα στην κατασκευή τους, αφού δεν περιλαμβάνουν ιδιαίτερα δύσκολες και πολύπλοκες μαθηματικές πράξεις.

Από την άλλη πλευρά, τα συγκεκριμένα δ.ε. μειονεκτούν στο ότι, οι στατιστικές συναρτήσεις (τ.μ.) που χρησιμοποιούνται, δε λαμβάνουν υπ'όψιν τις ακριβείς αριθμητικές τιμές των παρατηρήσεων, αλλά μόνο τους βαθμούς και τα πρόσημα των παρατηρήσεων και κατά συνέπεια χάνεται σημαντική στατιστική πληροφορία από τα τ.δ. που συλλέγονται. Ακόμη, τα διαγράμματα αυτά, συνήθως δεν καταφέρνουν να ανιχνεύσουν εύκολα και έγκαιρα πιθανές μεγάλες μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των δεδομένων, όπως μπορούν να κάνουν τα μη–παραμετρικά δ.ε. Shewhart που παρουσιάστηκαν σε προηγούμενες ενότητες. Επίσης, τα συγκεκριμένα δ.ε., συνήθως απαιτούν την αρχική εκτίμηση των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις. Αυτό ίσως να μην είναι αρκετά εύκολο ή εφικτό, ανάλογα με τη στατιστική μορφή που έχουν τα τ.δ. που συλλέγονται (στην περίπτωση που έχουμε ελλιπή, λογοκριμένα, περικομμένα ή μετατοπισμένα δεδομένα). Εκτός από αυτό όμως, η αρχική εκτίμηση των πληθυσμιακών παραμέτρων της κατανομής των δεδομένων, ίσως να μην είναι πολύ ακριβής και αξιόπιστη, ιδιαίτερα στην περίπτωση που το μέγεθος του αρχικού τ.δ. δεν είναι αρκετά μεγάλο, ώστε να ισχύουν τα γνωστά ασυμπτωτικά στατιστικά αποτελέσματα.

ΚΕΦΑΛΑΙΟ 6

Μη–παραμετρικά δ.ε. τύπου *EWMA*

Εκτός από τα μη–παραμετρικά δ.ε. τύπου Shewhart και *CUSUM*, στην πρόσφατη βιβλιογραφία έχουν προταθεί, για συγκεκριμένους λόγους και ορισμένα άλλα μη–παραμετρικά δ.ε., τα οποία είναι τύπου *EWMA*. Τα διαγράμματα αυτά είναι τα αντίστοιχα των πιο γνωστών παραμετρικών δ.ε. τύπου *EWMA*, χρησιμοποιούνται όμως στην ειδική περίπτωση που δεν είναι γνωστή η κατανομή που ακολουθούν οι παρατηρήσεις που συλλέγονται. Όπως όλα τα δ.ε. στο Στατιστικό Έλεγχο Διεργασίας, έτσι και τα μη–παραμετρικά δ.ε. τύπου *EWMA*, παρουσιάζουν θεωρητικά και πρακτικά ορισμένα βασικά πλεονεκτήματα, αλλά και κάποια μειονεκτήματα.

6.1 Δ.ε. που χρησιμοποιούν τους τυποποιημένους βαθμούς μεμονωμένων παρατηρήσεων

Αυτά τα δ.ε., προτάθηκαν από τους Hackl and Ledolter (1991). Εφαρμόζονται σε μεμονωμένες παρατηρήσεις και χρησιμοποιούν ως στατιστική συνάρτηση ελέγχου τους τυποποιημένους βαθμούς (standardized ranks) των μεμονωμένων παρατηρήσεων. Για μία μεμονωμένη παρατήρηση X_i, ο τυποποιημένος βαθμός της R_i, ορίζεται ως εξής:

$$R_i = 2[F_0(X_i) - \frac{1}{2}],$$

όπου F_0 είναι η εντός–ελέγχου α.σ.κ. των τιμών των μεμονωμένων παρατηρήσεων που συλλέχθηκαν ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου. Αν η F_0 είναι γνωστή ή έστω προσεγγίσιμη, οι τυποποιημένοι βαθμοί R_i των μεμονωμένων παρατηρήσεων μπορούν να υπολογιστούν σχετικά εύκολα. Συνήθως όμως, η εντός–ελέγχου α.σ.κ. F_0 είναι άγνωστη ή δε μπορεί να προσεγγιστεί ικανοποιητικά και στην περίπτωση αυτή οι τυποποιημένοι βαθμοί

R_i δε μπορούν να υπολογιστούν σύμφωνα με τον παραπάνω τύπο. Τότε, ορίζονται με διαφορετικό τρόπο, ως εξής:

$$R_i^{\bullet} = \frac{2}{g}[R_i^{*} - \frac{g+1}{2}],$$

όπου g–1 είναι το μέγεθος ενός προκαταρκτικού τ.δ. $Y_1, Y_2, ..., Y_{g-1}$, το οποίο συλλέγεται ενώ η διεργασία βρίσκεται εντός στατιστικού ελέγχου και R_i^* είναι ο βαθμός της μεμονωμένης παρατήρησης X_i σε σύγκριση με το προκαταρκτικό τ.δ. $Y_1, Y_2, ..., Y_{g-1}$. Συγκεκριμένα, θα ισχύει

$$R_i^{*} = 1 + \sum_{j=1}^{g-1} I(X_i > Y_j),$$

όπου I( ) είναι η γνωστή δίτιμη συνάρτηση που παίρνει τις τιμές 0 και 1. Για ένα δεδομένο αρχικό τ.δ. $Y_1, Y_2, ..., Y_{g-1}$, αποδεικνύεται (βλέπε Hackl and Ledolter (1991)) ότι οι τυποποιημένοι βαθμοί $R_i^{\bullet}$ των μεμονωμένων παρατηρήσεων είναι ανεξάρτητες και ισόνομες μεταξύ τους τ.μ.. Επίσης, έχει βρεθεί (βλέπε Hackl and Ledolter (1991)) ότι οι τυποποιημένοι βαθμοί R_i ακολουθούν την ομοιόμορφη κατανομή στο διάστημα [–1,1] και οι τυποποιημένοι βαθμοί $R_i^{\bullet}$ των μεμονωμένων παρατηρήσεων ακολουθούν τη διακριτή ομοιόμορφη κατανομή στο σύνολο $\{ \frac{1}{g} - 1, \frac{3}{g} - 1, ..., 1 - \frac{3}{g}, 1 - \frac{1}{g} \}$.

Το συγκεκριμένο δ.ε. *EWMA*, απεικονίζει γραφικά, για το *i* τ.δ. που συλλέγεται, τη στατιστική συνάρτηση ελέγχου (διακριτή τ.μ.)

$$T_i = (1 - l)T_{i-1} + l R_i^{\bullet},$$

όπου $R_i^{\bullet}$ είναι οι τυποποιημένοι βαθμοί που αναφέρθηκαν παραπάνω, $T_0 = 0$ και λ είναι μία θετική παράμετρος που παίρνει τιμές στο διάστημα (0,1] και συνήθως προτείνεται να παίρνει τιμές μεταξύ των αριθμών 0.1 και 0.3. Από τον παραπάνω τύπο, είναι μάλλον φανερό ότι η στατιστική συνάρτηση ελέγχου για το *i* τ.δ., δε χρησιμοποιεί στατιστική πληροφορία μόνο από αυτό το δείγμα, αλλά χρησιμοποιεί πληροφορία και από το προηγούμενο *i*–1 τ.δ.. Επομένως, εδώ προκύπτει το συμπέρασμα ότι οι στατιστικές συναρτήσεις ελέγχου σε αυτό το διάγραμμα, δεν είναι ανεξάρτητες μεταξύ τους, αλλά συσχετίζονται.

Αυτό το δ.ε., δίνει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν για οποιαδήποτε μεμονωμένη παρατήρηση X_i ισχύει η σχέση $|T_i| > h$, όπου το h είναι μία θετική παράμετρος που δηλώνει το ο.ε. στο διάγραμμα *EWMA* και T_i είναι η στατιστική συνάρτηση ελέγχου που αναφέρθηκε παραπάνω. Εχει διαπιστωθεί (βλέπε Hackl and Ledolter (1991)) ότι η σ.π. της τ.μ. T_i, επομένως και η πιθανότητα λανθασμένου συναγερμού και το εντός–ελέγχου μέσο μήκος ροής, δεν εξαρτώνται από την κατανομή που ακολουθούν οι μεμονωμένες παρατηρήσεις X_i. Επομένως, αυτό το δ.ε. *EWMA* είναι μη–παραμετρικό, ή αλλιώς, ελεύθερο κατανομής. Ουσιαστικά, αυτό το δ.ε., απεικονίζει γραφικά τους τυποποιημένους βαθμούς $R_i^{\bullet}$ που προκύπτουν για κάθε μία παρατήρηση που συλλέγεται. Οι δύο θετικές παράμετροι *λ* και *h*, οι οποίες καθορίζουν την κατασκευή και την εφαρμογή αυτού του δ.ε., προσδιορίζονται ως εκείνες οι τιμές που για γνωστό μέγεθος g–1 του αρχικού τ.δ. $Y_1, Y_2, ..., Y_{g-1}$, αντιστοιχούν σε ένα αρχικά προσδιορισμένο και αρκετά μεγάλο εντός–ελέγχου μέσο μήκος ροής ARL_{in} ή σε μία αρχικά προσδιορισμένη και αρκετά μικρή πιθανότητα λανθασμένου συναγερμού *FAR*. Για τον προσδιορισμό των παραμέτρων *λ* και *h* υπάρχουν αναλυτικοί στατιστικοί πίνακες (βλέπε Hackl and Ledolter (1991)), οι οποίοι για δεδομένες τιμές των *λ* και *h*, παρέχουν την τιμή του εντός–ελέγχου μέσου μήκους ροής ARL_{in} ή την τιμή της πιθανότητας λανθασμένου συναγερμού *FAR* που αντιστοιχεί σε αυτές. Από αποτελέσματα προσομοίωσης, έχει διαπιστωθεί ότι, το συγκεκριμένο μη–παραμετρικό δ.ε. είναι αρκετά ανθεκτικό σε πιθανές έκτροπες παρατηρήσεις και επίσης είναι αρκετά αποδοτικό και αποτελεσματικό σε μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των δεδομένων στην περίπτωση που η διεργασία βρεθεί εκτός στατιστικού ελέγχου (δηλαδή αυτό το δ.ε. ανιχνεύει πολύ εύκολα και γρήγορα αυτές τις μικρές μετατοπίσεις).

Το κυριότερο πλεονέκτημα για το συγκεκριμένο δ.ε. είναι ότι, μπορεί να ανιχνεύσει αρκετά εύκολα, μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν τα δεδομένα. Επίσης, το διάγραμμα αυτό δεν υποθέτει συγκεκριμένη κατανομή για τις παρατηρήσεις, τα στατιστικά χαρακτηριστικά του δεν επηρεάζονται από την κατανομή που αυτές ακολουθούν και ακόμη είναι αρκετά ανθεκτικό σε πιθανές έκτροπες παρατηρήσεις.

Από την άλλη πλευρά, το βασικότερο μειονέκτημα για αυτό το δ.ε. τύπου *EWMA*, είναι ότι δεν καταφέρνει να εντοπίσει εύκολα και έγκαιρα μεγάλες μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν τα δεδομένα. Ακόμη, το συγκεκριμένο διάγραμμα απαιτεί τη συλλογή μεμονωμένων παρατηρήσεων X_i και όχι τη συλλογή δειγμάτων, κάτι που ίσως να μην είναι εύκολο ή εφικτό στην πράξη, ιδιαίτερα σε περιπτώσεις που παρουσιάζει μεγάλη πρακτική δυσκολία η συλλογή μεμονωμένων παρατηρήσεων. Επίσης, υποθέτει τη συλλογή ενός αρχικού τ.δ., κάτι που μπορεί να προκαλέσει δαπάνη χρόνου και κυρίως, σημαντικό οικονομικό κόστος, ιδιαίτερα αν το μέγεθος αυτού του δείγματος είναι μεγάλο.

6.2 Δ.ε. που χρησιμοποιούν τους διαδοχικούς βαθμούς μεμονωμένων παρατηρήσεων

Εκτός από τα προηγούμενα μη–παραμετρικά δ.ε. τύπου *EWMA*, έχουν προταθεί στην πρόσφατη βιβλιογραφία και ορισμένα άλλα δ.ε. αυτού του τύπου, τα οποία χρησιμοποιούν τους διαδοχικούς βαθμούς μεμονωμένων παρατηρήσεων. Τα διαγράμματα αυτά προτάθηκαν, όπως και τα αμέσως προηγούμενα που αναλύθηκαν, από τους Hackl and Ledolter (1992) και χρησιμοποιούν ένα αρχικό τ.δ., που συλλέγεται ενώ η διεργασία βρίσκεται εντός στατιστικού ελέγχου. Αυτό το δ.ε. *EWMA*, χρησιμοποιεί ως στατιστική συνάρτηση ελέγχου τους διαδοχικούς βαθμούς $R_i^{\bullet}$ μεμονωμένων παρατηρήσεων X_i, *i*= 1, 2, … και απεικονίζει γραφικά τη στατιστική συνάρτηση

$$T_i = (1-\lambda)T_{i-1} + \lambda R_i^{\bullet},$$

όπου $R_i^{\bullet}$ είναι ο διαδοχικός βαθμός της παρατήρησης που ορίστηκε σε προηγούμενη ενότητα, $T_0 = 0$ και λ είναι μία θετική παράμετρος που παίρνει τιμές στο διάστημα (0,1] και συνήθως προτείνεται να παίρνει τιμές μεταξύ των αριθμών 0.1 και 0.3. Η βασική διαφορά του συγκεκριμένου διαγράμματος από το αμέσως προηγούμενο διάγραμμα που παρουσιάστηκε, είναι ότι σε αυτό το δ.ε. χρησιμοποιούνται και απεικονίζονται γραφικά οι διαδοχικοί βαθμοί των μεμονωμένων παρατηρήσεων και όχι οι τυποποιημένοι βαθμοί τους, όπως συνέβαινε στο προηγούμενο διάγραμμα.

Το συγκεκριμένο δ.ε., δίνει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν για κάποια μεμονωμένη παρατήρηση X_i ισχύει η σχέση $|T_i| > h$, όπου h είναι μία θετική παράμετρος που δηλώνει το όριο ελέγχου σε αυτό το δ.ε. τύπου *EWMA* και T_i είναι η στατιστική συνάρτηση που αναφέρθηκε παραπάνω. Έχει αποδειχθεί (βλέπε Hackl and

Ledolter (1992)) ότι, η σ.π. της διακριτής τ.μ. T_i, άρα και η πιθανότητα λανθασμένου συναγερμού και το εντός–ελέγχου μέσο μήκος ροής, δεν εξαρτώνται από την κατανομή που ακολουθούν οι μεμονωμένες παρατηρήσεις X_i. Επομένως, αυτό το δ.ε. είναι μη–παραμετρικό. Ουσιαστικά, αυτό το δ.ε. *EWMA*, απεικονίζει γραφικά τους διαδοχικούς βαθμούς $R_i^{\bullet}$ που προκύπτουν για κάθε μία παρατήρηση που συλλέγεται, χρησιμοποιώντας όμως στατιστική πληροφορία όχι μόνο από κάθε δείγμα ξεχωριστά, αλλά και από το αμέσως προηγούμενο του καθενός δείγματος που συλλέγεται. Οι δύο θετικές παράμετροι λ και *h*, οι οποίες καθορίζουν την κατασκευή και την εφαρμογή αυτού του διαγράμματος, προσδιορίζονται ως εκείνες οι τιμές οι οποίες, για γνωστό μέγεθος *g*–1 του αρχικού τ.δ. $Y_1, Y_2, ..., Y_{g-1}$, αντιστοιχούν σε ένα αρχικά προσδιορισμένο και αρκετά μεγάλο εντός–ελέγχου μέσο μήκος ροής ARL_{in} ή σε μία αρχικά προσδιορισμένη και αρκετά μικρή πιθανότητα λανθασμένου συναγερμού *FAR*. Για τον προσδιορισμό των παραμέτρων λ και *h*, υπάρχουν ειδικοί αναλυτικοί πίνακες (βλέπε Hackl and Ledolter (1992)) οι οποίοι, για συγκεκριμένες τιμές των λ και *h*, παρέχουν την τιμή του εντός–ελέγχου μέσου μήκους ροής ή την τιμή της πιθανότητας λανθασμένου συναγερμού που αντιστοιχεί σε αυτές. Από αποτελέσματα προσομοίωσης, έχει βρεθεί ότι το συγκεκριμένο μη–παραμετρικό δ.ε. είναι αρκετά ανθεκτικό σε πιθανές έκτροπες παρατηρήσεις και αρκετά αποδοτικό και αποτελεσματικό σε μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των δεδομένων, σε περίπτωση που η διεργασία βρίσκεται εκτός στατιστικού ελέγχου (δηλαδή αυτό το δ.ε. ανιχνεύει αρκετά εύκολα και έγκαιρα αυτές τις μικρές μετατοπίσεις, αν φυσικά αυτές συμβούν). Παρ’ολ’αυτά, έχει διαπιστωθεί ότι, το συγκεκριμένο δ.ε., συνήθως δεν κατορθώνει να ανιχνεύσει εύκολα και έγκαιρα μεγάλες μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται.

Τα πλεονεκτήματα και τα μειονεκτήματα του συγκεκριμένου δ.ε., είναι σε γενικές γραμμές, παρόμοια με τα πλεονεκτήματα και τα μειονεκτήματα του διαγράμματος *EWMA* που χρησιμοποιεί τους τυποποιημένους βαθμούς των μεμονωμένων παρατηρήσεων και παρουσιάστηκε στην αμέσως προηγούμενη ενότητα (δηλαδή στην ενότητα 6.1).

6.3 Δ.ε. που χρησιμοποιούν προσημικές–βαθμολογικές στατιστικές συναρτήσεις

Εκτός από τα παραπάνω δ.ε. τύπου *EWMA*, έχουν προταθεί σχετικά πρόσφατα και ορισμένα άλλα δ.ε. τύπου *EWMA*, τα οποία χρησιμοποιούν τη γνωστή (βλέπε Στέγγος (2006)) από την Απαραμετρική Στατιστική προσημική–βαθμολογική στατιστική συνάρτηση. Τα διαγράμματα αυτά, προτάθηκαν από τους Amin and Searcy (1991) και χρησιμοποιούν για κάθε i τ.δ. που συλλέγεται τη στατιστική συνάρτηση ελέγχου (τ.μ.)

$$SR_i = \sum_{j=1}^{g} sign(X_{ij} - q_0) R_{ij},$$

όπου: g είναι το μέγεθος του καθενός τ.δ. που συλλέγεται, θ_0 είναι η εκτίμηση για την εντός–ελέγχου στατιστική παράμετρο της κατανομής των δεδομένων, η οποία εκτίμηση προέκυψε από τη συλλογή ενός αρχικού τ.δ. $Y_1, Y_2, \ldots\ldots\ldots, Y_{g-1}$, ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου, R_{ij} είναι ο βαθμός που αντιστοιχεί στην παρατήρηση $| X_{ij} - q_0 |$ για το i τ.δ και sign( ) είναι η συνάρτηση προσήμου με τιμές +1, –1 και 0. Σε αυτό το δ.ε. *EWMA*, απεικονίζεται γραφικά η στατιστική συνάρτηση ελέγχου (διακριτή τ.μ.)

$$Z_i = \lambda SR_i + (1 - \lambda) Z_{i-1},$$

όπου: SR_i είναι η στατιστική συνάρτηση που αναφέρθηκε παραπάνω, Z_0 είναι μία τιμή–στόχος για τη διεργασία (η τιμή–στόχος είναι συνήθως η εντός–ελέγχου μέση τιμή της κατανομής των παρατηρήσεων που προσπαθούμε, εκτιμώντας την, να την προσεγγίσουμε αρκετά ικανοποιητικά) και λ είναι μία θετική παράμετρος που παίρνει τιμές στο διάστημα (0,1] και συνήθως προτείνεται να παίρνει τιμές μεταξύ των αριθμών 0.1 και 0.3. Το συγκεκριμένο δ.ε έχει δύο ο.ε., το κ.ο.ε. $LCL = m_0 - L$ και το α.ο.ε. $UCL = m_0 + L$, όπου m_0 είναι μία άγνωστη παράμετρος που δηλώνει την κεντρική γραμμή σε αυτό το δ.ε. και L είναι μία άγνωστη θετική παράμετρος που δηλώνει την απόσταση των δύο ο.ε. από την κεντρική γραμμή του δ.ε. (ή αλλιώς το μισό του εύρους του δ.ε.). Οι δύο αυτές άγνωστες παράμετροι m_0 και *L*, εκτιμώνται από τα δεδομένα που συλλέγονται (βλέπε Amin and Searcy (1991)). Το συγκεκριμένο δ.ε., δίνει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν για κάποιο i τ.δ. η στατιστική συνάρτηση $Z_i = \lambda SR_i + (1-\lambda)Z_{i-1}$ πάρει τιμή έξω από τα δύο ο.ε., δηλαδή αν για κάποιο i τ.δ. ισχύει $Z_i > UCL = m_0 + L$ ή $Z_i < LCL = m_0 - L$. Η στατιστική ανάλυση σε αυτό το δ.ε., βασίζεται στη γνωστή από την Απαραμετρική Στατιστική (βλέπε Στέγγος (2006)) προσημική–βαθμολογική στατιστική συνάρτηση του Wilcoxon. Αυτό το δ.ε. *EWMA* είναι μη–παραμετρικό, αφού έχει βρεθεί (βλέπε Amin and Searcy (1991)) ότι, όταν η διεργασία είναι εντός στατιστικού ελέγχου, η σ.π. της διακριτής τ.μ. Z_i δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις και αυτή, είναι ίδια και σταθερή για όλες τις κατανομές που είναι πιθανό να ακολουθούν οι παρατηρήσεις.

Οι Amin and Searcy (1991) προτείνουν επίσης, μαζί με το συγκεκριμένο δ.ε., να χρησιμοποιείται και ένα άλλο μη–παραμετρικό δ.ε. για την παρακολούθηση της διασποράς (δηλαδή της μεταβλητότητας) της διεργασίας. Από συγκεκριμένες μελέτες, έχει αποδειχθεί ότι τα στατιστικά χαρακτηριστικά του συγκεκριμένου δ.ε. δεν επηρεάζονται από την επιλογή της θετικής παραμέτρου λ, ούτε και από την αυτοσυσχέτιση που εμφανίζεται μεταξύ των στατιστικών συναρτήσεων Z_i. Έχει διαπιστωθεί επίσης (βλέπε Amin and Searcy (1991)), ότι το συγκεκριμένο δ.ε. βελτιώνεται σημαντικά με την προσθήκη προειδοποιητικών ο.ε. (warning limits).

Το σημαντικότερο πλεονέκτημα για αυτό το διάγραμμα είναι ότι, καταφέρνει να ανιχνεύσει αρκετά εύκολα και έγκαιρα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των δεδομένων. Επίσης, είναι μη–παραμετρικό, δηλαδή δεν προϋποθέτει συγκεκριμένη κατανομή για τις παρατηρήσεις που συλλέγονται και παρουσιάζει την ίδια πιθανότητα λανθασμένου συναγερμού *FAR* και το ίδιο εντός–ελέγχου μέσο μήκος ροής ARL_{in} για όλες τις πιθανές κατανομές των παρατηρήσεων.

Από την άλλη πλευρά, το σημαντικότερο μειονέκτημα για αυτό το διάγραμμα είναι ότι, για να κατασκευαστεί, συλλέγονται αρκετά ή πολλά τ.δ. και αυτό προκαλεί σημαντική δαπάνη χρόνου και κυρίως αυξημένο οικονομικό κόστος, ιδιαίτερα εάν τα μεγέθη των τ.δ. είναι αρκετά μεγάλα. Εκτός από αυτό, ο τύπος της στατιστικής συνάρτησης που χρησιμοποιείται, δε λαμβάνει υπ'όψιν την ακριβή αριθμητική τιμή των παρατηρήσεων (όπως συμβαίνει στα μη–παραμετρικά δ.ε. τύπου Shewhart που παρουσιάστηκαν σε προηγούμενες ενότητες), αλλά χρησιμοποιεί μόνο τους βαθμούς των παρατηρήσεων και τα πρόσημα των αποκλίσεων των παρατηρήσεων από την εντός–ελέγχου εκτιμημένη παράμετρο θ_0. Αυτό, έχει ως συνέπεια σε αυτό το δ.ε., να χάνεται σημαντική στατιστική πληροφορία στα τ.δ. που συλλέγονται. Επίσης, συνήθως δεν καταφέρνει να ανιχνεύσει εύκολα και έγκαιρα μεγάλες μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις. Ακόμη ένα μειονέκτημα είναι ότι, απαιτεί την αρχική εκτίμηση της εντός–ελέγχου παραμέτρου θ_0 ή είναι πιθανό η εκτίμηση που προκύπτει για το θ_0 να μην είναι αρκετά αξιόπιστη και ακριβής, στην περίπτωση που το μέγεθος του αρχικού τ.δ. δεν είναι αρκετά μεγάλο. Τέλος, απαιτεί την αρχική εκτίμηση των ο.ε. $LCL = m_0 - L$ και $UCL = m_0 + L$, κάτι που δε συνέβαινε στα προηγούμενα μη–παραμετρικά δ.ε. που αναλύθηκαν.

6.4 Γενικές παρατηρήσεις και σχόλια στα μη–παραμετρικά δ.ε. *EWMA*

Όπως τα μη–παραμετρικά δ.ε. τύπου *CUSUM*, έτσι και τα αντίστοιχα διαγράμματα *EWMA* αποτελούν μία αρκετά καλή και χρήσιμη εναλλακτική επιλογή για όσους ασχολούνται συστηματικά με το Στατιστικό Έλεγχο Διεργασίας, αν φυσικά ικανοποιούνται όλες οι βασικές υποθέσεις για την εφαρμογή και τη χρήση τους. Το βασικότερο πλεονέκτημα που παρουσιάζουν είναι ότι, μπορούν να εντοπίσουν εύκολα και έγκαιρα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν τα δεδομένα. Επίσης, το ότι αυτά τα διαγράμματα είναι μη–παραμετρικά, τα καθιστά ανεξάρτητα από οποιαδήποτε κατανομή και αν ακολουθούν τα δεδομένα.

Από την άλλη πλευρά, το πιο σημαντικό μειονέκτημα για τα συγκεκριμένα διαγράμματα, είναι ότι συνήθως δεν καταφέρνουν να ανιχνεύσουν εύκολα μεγάλες μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των δεδομένων, όπως μπορούν να κάνουν τα μη–παραμετρικά δ.ε. Shewhart που παρουσιάστηκαν σε προηγούμενες ενότητες. Εκτός από αυτό, ορισμένα από αυτά τα δ.ε., απαιτούν τη συλλογή μεμονωμένων παρατηρήσεων και όχι δειγμάτων, κάτι που ίσως να είναι πρακτικά δύσκολο ή και ανέφικτο. Ακόμη, η συλλογή ενός αρχικού τ.δ., συνήθως προκαλεί αυξημένο οικονομικό κόστος, ιδιαίτερα στην περίπτωση που το μέγεθός του είναι σχετικά μεγάλο.

Σε κάθε περίπτωση πάντως, θεωρείται γενικά ότι, τα μη– παραμετρικά δ.ε. *EWMA* παίζουν σημαντικό ρόλο στο Στατιστικό Έλεγχο Διεργασίας και ότι (βλέπε Chakraborti, Bakir and Van Der Laan (2001)) υπάρχουν ακόμη πολλές δυνατότητες εκτεταμένης μελλοντικής επιστημονικής έρευνας πάνω στα συγκεκριμένα διαγράμματα. Η στατιστική και πρακτική σπουδαιότητα αυτών των διαγραμμάτων, εντοπίζεται στην έγκαιρη ανίχνευση μικρών ή πολύ μικρών μετατοπίσεων των στατιστικών (πληθυσμιακών) παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις που συλλέγονται.

ΚΕΦΑΛΑΙΟ 7

Μη–παραμετρικά δ.ε. που βασίζονται σε ροές

Εκτός από τα μη–παραμετρικά δ.ε. τύπου Shewhart, *CUSUM* και *EWMA* που αναλύθηκαν σε προηγούμενες ενότητες, στην πρόσφατη βιβλιογραφία έχουν προταθεί και μελετηθεί ορισμένα άλλα μη–παραμετρικά δ.ε., τα οποία χρησιμοποιούν τις γνωστές από την Απαραμετρική Στατιστική ροές (runs). Ο κυριότερος λόγος για τον οποίο αναπτύχθηκαν τα συγκεκριμένα διαγράμματα, ήταν ότι τα δ.ε. που βασίζονται σε ροές, παρουσιάζουν (βλέπε Chakraborti and Eryilmaz (2007)) σημαντικά μικρότερη πιθανότητα λανθασμένου συναγερμού *FAR* και αρκετά μεγαλύτερο εντός–ελέγχου μέσο μήκος ροής ARL_{in}, σε σύγκριση με τα μη–παραμετρικά δ.ε. τύπου Shewhart, *CUSUM* και *EWMA*, καθώς και σε σύγκριση με τα κλασικά παραμετρικά δ.ε. τύπου Shewhart. Επίσης, έχει διαπιστωθεί με χρήση προσομοίωσης ότι, διατηρώντας σταθερό το εντός–ελέγχου μέσο μήκος ροής, τα συγκεκριμένα δ.ε. πλεονεκτούν όσον αφορά το εκτός ελέγχου μέσο μήκος ροής ARL_{out} (παρουσιάζουν δηλαδή μικρότερο εκτός–ελέγχου μέσο μήκος ροής) έναντι των μη–παραμετρικών δ.ε. τύπου Shewhart, *CUSUM* και *EWMA*, καθώς και σε σύγκριση με τα πιο γνωστά παραμετρικά δ.ε. τύπου Shewhart. Έχει διαπιστωθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι, τα συγκεκριμένα δ.ε., μπορούν να ανιχνεύσουν εύκολα και γρήγορα πιθανές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν τα δεδομένα και ιδιαίτερα μικρές ή πολύ μικρές μεταβολές των παραμέτρων αυτών. Επομένως, αυτά τα διαγράμματα, αποτελούν μία πολύ ικανοποιητική εναλλακτική επιλογή για τις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας, κυρίως επειδή σπάνια οδηγούν σε λανθασμένες γραφικές ενδείξεις, είτε όταν η διεργασία βρίσκεται εντός είτε εκτός στατιστικού ελέγχου.

Τα διαγράμματα αυτά, δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, στην περίπτωση που k συνεχόμενα σημεία (ροές) βρεθούν έξω από τα δύο ο.ε. ή έξω από το ο.ε., ανάλογα με την περίπτωση που στο δ.ε. υπάρχουν ένα ή δύο ο.ε.. Στην περίπτωση που μόνο ένα σημείο στο δ.ε. βρεθεί έξω από τα δύο ο.ε. ή έξω από το ο.ε., τα συγκεκριμένα δ.ε. δε δίνουν ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, όπως συμβαίνει στα μη–παραμετρικά δ.ε. τύπου Shewhart, *CUSUM* και *EWMA*, καθώς και στα παραμετρικά δ.ε. τύπου Shewhart. Έχει αποδειχθεί επίσης, με χρήση προσομοίωσης ότι, στην περίπτωση που η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, τα μη–παραμετρικά δ.ε. που βασίζονται σε ροές είναι αρκετά πιο αποδοτικά και αποτελεσματικά σε σύγκριση με τα υπόλοιπα μη–παραμετρικά δ.ε. που αναφέρθηκαν και σε σύγκριση με τα κλασικά παραμετρικά δ.ε. τύπου Shewhart.

Στη συνέχεια, θα παρουσιάσουμε μη–παραμετρικά δ.ε. που χρησιμοποιούν κανόνες ροών, σε συνδυασμό με τη γνωστή από την Απαραμετρική Στατιστική προσημική–βαθμολογική στατιστική συνάρτηση του Wilcoxon (βλέπε Στέγγος (2006)) που αναφέρθηκε σε προηγούμενη ενότητα.

Στα συγκεκριμένα διαγράμματα, η στατιστική μεθοδολογία που ακολουθείται, είναι η εξής: για το i τ.δ. που συλλέγεται, αρχικά υπολογίζεται η στατιστική συνάρτηση

$$y_i = \sum_{j=1}^{n} sign(X_{ij} - q_0) R_{ij} ,$$

όπου: n είναι το μέγεθος του καθενός τ.δ., θ_0 είναι η εκτίμηση για την εντός–ελέγχου τιμή της στατιστικής (πληθυσμιακής) παραμέτρου της κατανομής των δεδομένων, η οποία εκτίμηση προέκυψε από τη συλλογή ενός τ.δ. αναφοράς $Y_1, Y_2, ..., Y_{g-1}$ ενώ η διεργασία βρισκόταν εντός στατιστικού ελέγχου, R_{ij} είναι ο βαθμός που αντιστοιχεί στην παρατήρηση $| X_{ij} - q_0 |$ για το i τ.δ. και sign( ) είναι η στατιστική συνάρτηση προσήμου (διακριτή τ.μ.) που αναφέρθηκε σε προηγούμενη ενότητα. Αν W_i^+ είναι η διακριτή τ.μ. που δηλώνει το άθροισμα των βαθμών μόνο των θετικών παρατηρήσεων $X_{ij} - q_0$ (δηλαδή το άθροισμα των βαθμών μόνο των ποσοτήτων $| X_{ij} - q_0 |$ για τις οποίες οι παρατηρήσεις $X_{ij} - q_0$ είναι θετικές), τότε αποδεικνύεται (βλέπε Chakraborti and Eryilmaz (2007)) ότι οι διακριτές τ.μ. y_i και W_i^+ θα συνδέονται μεταξύ τους γραμμικά, μέσω της σχέσης

$$y_i = 2W_i^+ - \frac{n(n+1)}{2}.$$

Μία πολύ σημαντική και βασική προϋπόθεση για τα συγκεκριμένα διαγράμματα ροών, είναι ότι οι παρατηρήσεις ακολουθούν μία συνεχή κατανομή, η οποία είναι συμμετρική γύρω από τη διάμεσό της. Αλλιώς, τα αποτελέσματα που θα προκύψουν στη συνέχεια, δε θα ισχύουν.

Σε αυτά τα διαγράμματα, απεικονίζεται γραφικά, για το i τ.δ. που συλλέγεται, η τιμή της στατιστικής συνάρτησης y_i. Η τιμή της y_i προκύπτει έμμεσα από την τιμή της στατιστικής συνάρτησης W_i^+, μέσω της παραπάνω σχέσης. Για συγκεκριμένο μέγεθος n των τ.δ., η σ.π. της τ.μ. W_i^+ έχει πινακοποιηθεί αναλυτικά από διάφορους συγγραφείς (βλέπε Chakraborti and Eryilmaz (2007)).

Έχει διαπιστωθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η σ.π. της τ.μ. W_i^+ δεν εξαρτάται από τη κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από το μέγεθος n των τ.δ. που συλλέγονται. Επομένως, το συγκεκριμένο δ.ε. είναι μη–παραμετρικό (ή ελεύθερο κατανομής) και παρουσιάζει την ίδια πιθανότητα λανθασμένου συναγερμού και το ίδιο εντός–ελέγχου μέσο μήκος ροής για όλες τις κατανομές που είναι πιθανό να ακολουθούν τα δεδομένα.

Πρέπει να σημειωθεί εδώ, ότι τα ο.ε. (ή το ο.ε.) στα μη–παραμετρικά δ.ε. που χρησιμοποιούν κανόνες ροών, δεν είναι τυχαίες μεταβλητές (όπως είδαμε ότι συμβαίνει στα μη–παραμετρικά δ.ε. τύπου Shewhart που παρουσιάστηκαν σε προηγούμενες ενότητες), αλλά είναι κάποιες γνωστές και δεδομένες ποσότητες. Για το λόγο αυτό, τα σημεία (τιμές των στατιστικών συναρτήσεων ελέγχου) y_i είναι ανεξάρτητα μεταξύ τους και δεν συσχετίζονται (όπως συνέβαινε στα μη–παραμετρικά δ.ε. τύπου Shewhart).

Θα παρουσιάσουμε στη συνέχεια διάφορες περιπτώσεις, στις οποίες η στατιστική συνάρτηση y_i συνδυάζεται με κανόνες ροών (runs rules).

7.1 Μονόπλευρα δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*k*

Τα διαγράμματα αυτά, έχουν προταθεί και μελετηθεί αναλυτικά από τους Chakraborti and Eryilmaz (2007). Αυτά, έχουν ένα μόνο ο.ε., είτε το α.ο.ε. *UCL*, είτε το κ.ο.ε. *LCL*. Με αυτό, συγκρίνεται η τιμή της στατιστικής συνάρτησης y_i που αναφέρθηκε παραπάνω. Δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν k συνεχόμενα σημεία (ροές) στο διάγραμμα πάρουν όλα τιμές μεγαλύτερες από το α.ο.ε. *UCL*, στην περίπτωση υπάρχει μόνο το α.ο.ε. *UCL*, ή αν και μόνο αν k συνεχόμενα σημεία στο διάγραμμα πάρουν όλα τιμές μικρότερες από το κ.ο.ε. *LCL*, στην περίπτωση που υπάρχει μόνο το κ.ο.ε. *LCL*.

Θα ξεκινήσουμε, εξετάζοντας αρχικά, την περίπτωση που στο μονόπλευρο δ.ε. χρησιμοποιείται μόνο το α.ο.ε. *UCL*. Ας συμβολίσουμε με X_i τη δίτιμη τ.μ. που παίρνει την τιμή 1, αν κάποια τιμή (σημείο στο διάγραμμα) y_i είναι μεγαλύτερη από το α.ο.ε. *UCL* και την τιμή 0, στην αντίθετη περίπτωση. Τότε, η διακριτή τ.μ. T_k^+, η οποία δηλώνει το συνολικό αριθμό σημείων που απεικονίζονται στο μονόπλευρο δ.ε. έως ότου εμφανιστούν k συνεχόμενα σημεία που όλα θα πάρουν τιμές μεγαλύτερες του *UCL* (δηλαδή η τ.μ. T_k^+ θα δηλώνει το μήκος ροής σε αυτό το δ.ε.) μπορεί να γραφεί (βλέπε Chakraborti and Eryilmaz (2007)) στη μορφή

$$T_k^+ = \min\{t : \sum_{j=t-k+1}^{t} X_j = k\}.$$

Έχει διαπιστωθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η σ.π. της τ.μ. T_k^+ εξαρτάται μόνο από τον θετικό ακέραιο k και από την πιθανότητα

$$p^+ = P(y_i > UCL) = P(X_i = 1).$$

Αποδεικνύεται (βλέπε Balakrishnan and Koutras (2002)) ότι η σ.π. της T_k^+ θα δίνεται από τον παρακάτω τύπο:

$P(T_k^+ = x) = 0$, αν ισχύει $0 \le x < k$,

$P(T_k^+ = x) = (p^+)^k$, αν ισχύει $x = k$ και

$$P(T_k^+ = x) = \sum_{j=1}^{[\frac{x+1}{k+1}]} (-1)^{j-1} (p^+)^{jk} (1-p^+)^{j-1} \left[\binom{x-jk-1}{j-2} + (1-p^+)\binom{x-jk-1}{j-1}\right], \text{ αν ισχύει } x \ge k+1.$$

Η μέση τιμή της διακριτής τ.μ. T_k^+ (δηλαδή το μέσο μήκος ροής) και η διακύμανσή της θα δίνονται από τους εξής τύπους (βλέπε Balakrishnan and Koutras (2002)):

$$E(T_k^+) = ARL = \frac{1-(p^+)^k}{(1-p^+)(p^+)^k}, \quad Var(T_k^+) = \frac{1-(2k+1)(1-p^+)(p^+)^k - (p^+)^{2k+1}}{(1-p^+)^2 (p^+)^{2k}}.$$

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου (δηλαδή αν η κατανομή που ακολουθούν τα δεδομένα δεν έχει αλλάξει), έστω p_0^+ η πιθανότητα ένα οποιοδήποτε σημείο στο μονόπλευρο δ.ε. να πάρει τιμή μεγαλύτερη από το α.ο.ε., δηλαδή έστω ότι ισχύει το εξής:

$p_0^+ = P(y_i > UCL$| η διεργασία εντός ελέγχου) = $P(X_i = 1$ | η διεργασία εντός ελέγχου)

Τότε, έχει αποδειχθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι αυτή η πιθανότητα δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από τον θετικό ακέραιο *k*, από το μέγεθος των τ.δ. και από την τιμή του α.ο.ε. *UCL*.

Επίσης, ας συμβολίσουμε με $T_{k,0}^+$ τη διακριτή τ.μ. που δηλώνει το εντός–ελέγχου μήκος ροής σε αυτό το δ.ε.. Είναι μάλλον προφανές, λαμβάνοντας υπ'όψιν όλα τα προηγούμενα αποτελέσματα, ότι η σ.π. της τ.μ. $T_{k,0}^+$, θα δίνεται από τον παρακάτω τύπο:

$P(T_{k,0}^+ = x) = 0$, αν $0 \le x < k$,

$P(T_{k,0}^+ = x) = (p_0^+)^k$, αν $x = k$ και

$P(T_{k,0}^+ = x) = \sum_{j=1}^{[\frac{x+1}{k+1}]} (-1)^{j-1} (p_0^+)^{jk} (1-p_0^+)^{j-1} [\binom{x-jk-1}{j-2} + (1-p_0^+)\binom{x-jk-1}{j-1}]$, αν $x \ge k+1$.

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η εντός–ελέγχου μέση τιμή της τ.μ. $T_{k,0}^+$ (δηλαδή το εντός–ελέγχου μέσο μήκος ροής) και η εντός–ελέγχου διακύμανσή της θα δίνονται από τους παρακάτω τύπους:

$E(T_{k,0}^+) = ARL = \frac{1-(p_0^+)^k}{(1-p_0^+)(p_0^+)^k}$ και $Var(T_{k,0}^+) = \frac{1-(2k+1)(1-p_0^+)(p_0^+)^k - (p_0^+)^{2k+1}}{(1-p_0^+)^2 (p_0^+)^{2k}}$

και αυτοί οι τύποι, προκύπτουν προφανώς, από τους δύο προηγούμενους τύπους που γράφτηκαν, απλώς αντικαθιστώντας την πιθανότητα p^+ με την (εντός–ελέγχου) πιθανότητα p_0^+.

Επίσης, έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η πιθανότητα λανθασμένου συναγερμού *FAR* θα είναι ίση με

$$FAR = (p_0^+)^k.$$

Αποδεικνύεται (βλέπε Balakrishnan and Koutras (2002)) ότι η σ.π. της τ.μ. $T_{k,0}^{+}$ δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από τον θετικό ακέραιο k, από το μέγεθος των τ.δ. και από την τιμή του α.ο.ε. UCL. Άρα και η πιθανότητα λανθασμένου συναγερμού, καθώς και το εντός–ελέγχου μέσο μήκος ροής, δε θα εξαρτώνται από την κατανομή που ακολουθούν οι παρατηρήσεις, επομένως συνεπάγεται εδώ ότι το συγκεκριμένο μονόπλευρο δ.ε. είναι μη–παραμετρικό ή αλλιώς, ελεύθερο κατανομής.

Σε αυτό το σημείο, αξίζει να σημειωθεί ότι το ο.ε. σε αυτό το δ.ε. δε χρειάζεται να υπολογιστεί με κάποια μέθοδο, ούτε να εκτιμηθεί ή να προσεγγιστεί. Αυτό συμβαίνει, επειδή για συγκεκριμένη τιμή του α.ο.ε., καθώς και για δεδομένη τιμή του αριθμού k και για γνωστό μέγεθος των τ.δ. που συλλέγονται, υπάρχουν ειδικοί αναλυτικοί πίνακες (βλέπε Chakraborti and Eryilmaz (2007)) οι οποίοι, για αυτές τις δεδομένες τιμές, παρέχουν την τιμή της πιθανότητας λανθασμένου συναγερμού και του εντός–ελέγχου μέσου μήκους ροής που αντιστοιχούν σε αυτές. Έτσι λοιπόν, για την κατασκευή του μονόπλευρου δ.ε., το α.ο.ε. UCL μπορεί να βρεθεί, προσδιορίζοντας αρχικά μία επιθυμητή και αρκετά μικρή τιμή για την πιθανότητα λανθασμένου συναγερμού (συνήθως ίση με 5% ή 1%), ή ισοδύναμα, προσδιορίζοντας αρχικά μία επιθυμητή και αρκετά μεγάλη τιμή για το εντός– ελέγχου μέσο μήκος ροής (συνήθως ίσο με 300 ή 400) και στη συνέχεια, εντοπίζοντας από τους συγκεκριμένους πίνακες σε ποια τιμή του α.ο.ε. UCL αντιστοιχεί αυτή η τιμή.

Στην περίπτωση που στο διάγραμμα υπάρχει μόνο το κ.ο.ε. LCL, έχουμε την ίδια ακριβώς στατιστική ανάλυση και παρόμοια στατιστικά αποτελέσματα. Ειδικότερα, (βλέπε Chakraborti and Eryilmaz (2007)), αν X_i είναι μία δίτιμη τ.μ. που παίρνει την τιμή 1, αν κάποιο σημείο (τιμή της στατιστικής συνάρτησης ελέγχου) y_i είναι μικρότερο από το κ.ο.ε. LCL και την τιμή 0 στην αντίθετη περίπτωση και αν T_k^{-} είναι η διακριτή τ.μ. που δηλώνει το μήκος ροής σε αυτό το δ.ε., έχει βρεθεί ότι η τ.μ. T_k^{-} μπορεί να γραφεί στην εξής μορφή:

$$T_k^{-} = \min\{t : \sum_{j=t-k+1}^{t} X_j = k\}.$$

Επίσης, έστω p^{-} η πιθανότητα

$$p^{-} = P(y_i < LCL) = P(X_i = 1).$$

Έχει διαπιστωθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η σ.π. της τ.μ. T_k^- εξαρτάται μόνο από τον θετικό ακέραιο k και από την πιθανότητα $p^- = P(y_i < LCL) = P(X_i = 1)$. Ειδικότερα, (βλέπε Balakrishnan and Koutras (2002)) η σ.π. της τ.μ. T_k^- θα δίνεται από τον παρακάτω τύπο:

$P(T_k^- = x) = 0$, αν ισχύει $0 \le x < k$,

$P(T_k^- = x) = (p^-)^k$, αν ισχύει $x = k$ και

$P(T_k^- = x) = \sum_{j=1}^{[\frac{x+1}{k+1}]} (-1)^{j-1} (p^-)^{jk} (1-p^-)^{j-1} [\binom{x-jk-1}{j-2} + (1-p^-)\binom{x-jk-1}{j-1}]$, αν ισχύει $x \ge k+1$.

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, έχει βρεθεί (βλέπε Balakrishnan and Koutras (2002)) ότι η εντός–ελέγχου μέση τιμή της τ.μ. $T_{k,0}^-$ (δηλαδή το εντός–ελέγχου μέσο μήκος ροής) και η εντός–ελέγχου διακύμανσή της θα δίνονται από τους παρακάτω τύπους:

$$E(T_{k,0}^-) = ARL = \frac{1-(p_0^-)^k}{(1-p_0^-)(p_0^-)^k} \quad \text{και} \quad Var(T_{k,0}^-) = \frac{1-(2k+1)(1-p_0^-)(p_0^-)^k - (p_0^-)^{2k+1}}{(1-p_0^-)^2 (p_0^-)^{2k}}$$

και εδώ, είναι μάλλον φανερό ότι οι τελευταίοι τύποι, προκύπτουν από τους δύο τύπους που ήδη γράφτηκαν, απλώς αντικαθιστώντας την πιθανότητα p^- με την πιθανότητα p_0^-.

Επίσης, έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η πιθανότητα λανθασμένου συναγερμού *FAR* σε αυτό το μονόπλευρο δ.ε., θα είναι ίση με

$$FAR = (p_0^-)^k,$$

όπου p_0^- είναι η πιθανότητα

$p_0^- = P(y_i < LCL$ | η διεργασία εντός ελέγχου$) = P(X_i = 1$ | η διεργασία εντός ελέγχου$)$.

Έχει διαπιστωθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η σ.π. της τ.μ. $T_{k,0}^-$ δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις που συλλέγονται, αλλά μόνο από τον αριθμό *k*, από το μέγεθος των τ.δ. και από την τιμή του κ.ο.ε. *LCL*. Ακριβώς το ίδιο ισχύει και για την πιθανότητα $p_0^- = P(y_i < LCL$ | η διεργασία εντός ελέγχου$) = P(X_i = 1$ | η διεργασία εντός ελέγχου$)$. Άρα και η πιθανότητα λανθασμένου συναγερμού, καθώς και το εντός–ελέγχου μέσο μήκος ροής, δε θα εξαρτώνται από την κατανομή που ακολουθούν οι παρατηρήσεις. Επομένως, συνδυάζοντας όλα τα παραπάνω, καταλήγουμε τελικά στο συμπέρασμα ότι το συγκεκριμένο μονόπλευρο δ.ε. είναι μη–παραμετρικό.

Σε αυτό το σημείο, αξίζει να σημειωθεί ότι, το κ.ο.ε. σε αυτό το μονόπλευρο διάγραμμα δε χρειάζεται να υπολογιστεί με κάποια μέθοδο, ούτε να εκτιμηθεί ή να προσεγγιστεί. Το κ.ο.ε. *LCL*, προσδιορίζεται ακριβώς με την ίδια στατιστική μεθοδολογία, όπως προσδιορίστηκε το α.ο.ε. *UCL*, στην περίπτωση που στο μονόπλευρο διάγραμμα υπήρχε μόνο το α.ο.ε. (βλέπε σελίδα 90). Η σχετική άνεση λοιπόν που υπάρχει (δηλαδή η απουσία μαθηματικών πράξεων και πολύπλοκων υπολογισμών), όσον αφορά τον προσδιορισμό του ο.ε. (και συνεπώς την κατασκευή και την πρακτική εφαρμογή αυτού του διαγράμματος), αποτελεί ένα μάλλον σημαντικό στατιστικό πλεονέκτημα για αυτό το μονόπλευρο δ.ε..

7.2 Η ειδική περίπτωση *k*= 2 (μονόπλευρο δ.ε. που χρησιμοποιεί τον κανόνα ροών 2–από–2)

Στο μονόπλευρο μη–παραμετρικό δ.ε. που χρησιμοποιεί τον κανόνα ροών *k*–από–*k*, η ειδική περίπτωση $k = 2$ παρουσιάζει ιδιαίτερο πρακτικό ενδιαφέρον, αφού αυτή η περίπτωση είναι που χρησιμοποιείται και εφαρμόζεται συχνότερα στα δ.ε. που βασίζονται σε ροές. Συγκεκριμένα, όπως θα αναφερθεί σε επόμενη ενότητα, τα μονόπλευρα δ.ε. που βασίζονται σε ροές και χρησιμοποιούν τον κανόνα ροών 2–από–2, παρουσιάζουν αρκετά πλεονεκτήματα έναντι των υπόλοιπων μη–παραμετρικών δ.ε. που έχουν ήδη παρουσιαστεί και επίσης έναντι των πιο γνωστών παραμετρικών δ.ε. τύπου Shewhart.

Τα συγκεκριμένα δ.ε. (βλέπε Chakraborti and Eryilmaz (2007)) δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν δύο οποιαδήποτε συνεχόμενα σημεία στο δ.ε. πάρουν τιμές μεγαλύτερες από το α.ο.ε. *UCL*, στην περίπτωση που στο διάγραμμα υπάρχει μόνο το α.ο.ε., ή δύο οποιαδήποτε συνεχόμενα σημεία στο δ.ε. πάρουν τιμές μικρότερες από το κ.ο.ε. *LCL*, στην περίπτωση που στο διάγραμμα υπάρχει μόνο το κ.ο.ε.. Δηλαδή, αυτά τα δ.ε. δίνουν ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν ισχύει $y_{i-1} > UCL$ και $y_i > UCL$, στην περίπτωση που στο δ.ε. υπάρχει μόνο το α.ο.ε., ή $y_{i-1} < LCL$ και $y_i < LCL$, στην περίπτωση που στο δ.ε. υπάρχει μόνο το κ.ο.ε., όπου y_i είναι η προσημική–βαθμολογική στατιστική συνάρτηση του Wilcoxon που αναφέρθηκε σε προηγούμενη ενότητα.

Στην περίπτωση που στο δ.ε. υπάρχει μόνο το α.ο.ε. *UCL*, έστω T_2^+ η διακριτή τ.μ. που δηλώνει το μήκος ροής σε αυτό το δ.ε.. Αν X_i είναι μία δίτιμη τ.μ. η οποία παίρνει την τιμή 1, αν η παρατήρηση y_i είναι μεγαλύτερη από το α.ο.ε. *UCL* και την τιμή 0 στην αντίθετη περίπτωση, η τ.μ. T_2^+ μπορεί να γραφεί (βλέπε Chakraborti and Eryilmaz (2007)) στην εξής μορφή:

$$T_2^+ = \min\{t : X_{t-1} + X_t = 2\}.$$

Επίσης, αν p^+ είναι η πιθανότητα

$$p^+ = P(y_i > UCL) = P(X_i = 1),$$

η σ.π. της τ.μ. T_2^+ θα δίνεται από τον παρακάτω τύπο:

$P(T_2^+ = x) = 0$, αν $0 \le x < 2$,

$P(T_2^+ = x) = (p^+)^2$, αν $x = 2$ και

$$P(T_2^+ = x) = \sum_{j=1}^{[\frac{x+1}{3}]} (-1)^{j-1} (p^+)^{2j} (1-p^+)^{j-1} [\binom{x-2j-1}{j-2} + (1-p^+)\binom{x-2j-1}{j-1}], \text{ αν } x \ge 3.$$

Είναι προφανές εδώ, ότι ο παραπάνω τύπος που δίνει τη σ.π. της τ.μ. T_2^+, προκύπτει άμεσα από τον τύπο που δίνει τη σ.π. της τ.μ. T_k^+ και αναφέρθηκε σε προηγούμενη ενότητα, απλά θέτοντας $k = 2$.

Η μέση τιμή της τ.μ. T_2^+ (δηλαδή το μέσο μήκος ροής) και η διακύμανσή της, θα δίνονται από τους εξής τύπους:

$$E(T_2^+) = ARL = \frac{1+p^+}{(p^+)^2} \quad \text{και} \quad Var(T_2^+) = \frac{1-5(1-p^+)(p^+)^2-(p^+)^5}{(1-p^+)^2(p^+)^4},$$

οι οποίοι είναι μάλλον φανερό ότι προκύπτουν από τη μέση τιμή και τη διακύμανση της τ.μ. T_k^+ που αναφέρθηκε σε προηγούμενη ενότητα, απλά θέτοντας $k = 2$.

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η σ.π. της τ.μ. $T_{2,0}^+$ (που δηλώνει το εντός–ελέγχου μήκος ροής στο διάγραμμα), το εντός–ελέγχου μέσο μήκος ροής και η εντός–

ελέγχου διακύμανση της τ.μ. $T_{2,0}^{+}$ θα δίνονται από τους ίδιους ακριβώς τύπους που αναφέρθηκαν σε αυτήν την ενότητα, απλώς αντικαθιστώντας την πιθανότητα

$$p^{+} = P(y_i > UCL) = P(X_i = 1)$$

με την πιθανότητα

$$p_0^{+} = P(y_i > UCL \mid \text{η διεργασία εντός ελέγχου}) = P(X_i = 1 \mid \text{η διεργασία εντός ελέγχου}).$$

Επίσης, η πιθανότητα λανθασμένου συναγερμού FAR έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι είναι ίση με

$$FAR = (p_0^{+})^2.$$

Ακόμη, έχει διαπιστωθεί (βλέπε Balakrishnan and Koutras (2002)) ότι, αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου (δηλαδή αν δεν υπάρχει αλλαγή της κατανομής των δεδομένων), η σ.π. του εντός–ελέγχου μήκους ροής, το εντός–ελέγχου μέσο μήκος ροής και η εντός–ελέγχου διακύμανση του μήκους ροής δεν εξαρτώνται από την κατανομή που ακολουθούν τα δεδομένα. Άρα, προκύπτει ότι το συγκεκριμένο μονόπλευρο δ.ε., είναι μη–παραμετρικό.

Αν στο μονόπλευρο δ.ε. υπάρχει μόνο το κ.ο.ε. LCL, η στατιστική ανάλυση είναι ακριβώς η ίδια, όπως και τα στατιστικά αποτελέσματα. Ουσιαστικά, το μόνο που αλλάζει είναι ο συμβολισμός των διάφορων ποσοτήτων, αφού η στατιστική μορφή των τύπων είναι ακριβώς η ίδια. Συγκεκριμένα, η πιθανότητα p^{+} αλλάζει σε p^{-}, η τ.μ. T_2^{+} αλλάζει σε T_2^{-}, η πιθανότητα p_0^{+} σε p_0^{-} και η τ.μ. $T_{2,0}^{+}$ σε $T_{2,0}^{-}$. Οι τύποι που δίνουν τις διάφορες στατιστικές ποσότητες έχουν ακριβώς την ίδια μορφή. Επίσης, αποδεικνύεται με την ίδια στατιστική μεθοδολογία, ότι το μονόπλευρο δ.ε. που χρησιμοποιεί τον κανόνα ροών 2–από–2 είναι μη–παραμετρικό. Τέλος, η διαδικασία κατασκευής του μονόπλευρου διαγράμματος (δηλαδή η διαδικασία προσδιορισμού της τιμής του κ.ο.ε.) είναι ακριβώς η ίδια με το αντίστοιχο μονόπλευρο δ.ε. το οποίο περιέχει μόνο το α.ο.ε. UCL.

7.3 Δίπλευρα δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*k*

Αυτά τα δ.ε. είναι δίπλευρα, δηλαδή έχουν ένα α.ο.ε. *UCL* και ένα κ.ο.ε. *LCL*. Δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν *k* συνεχόμενα σημεία (παρατηρήσεις) y_i πάρουν τιμές έξω από τα δύο ο.ε.. Σε αυτά τα δ.ε., οι πιθανές μετατοπίσεις ορίζονται προς δύο κατευθύνσεις, είτε μετατόπιση της στατιστικής συνάρτησης y_i προς μεγάλες τιμές, δηλαδή τιμές μεγαλύτερες του α.ο.ε., είτε μετατόπιση της y_i προς μικρές τιμές, δηλαδή τιμές μικρότερες του κ.ο.ε.. Εδώ, η στατιστική ανάλυση και τα στατιστικά αποτελέσματα που προκύπτουν, είναι παρόμοια με εκείνα για τα μονόπλευρα δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*k*. Το μόνο που αλλάζει, είναι ο συμβολισμός των διάφορων στατιστικών ποσοτήτων. Συγκεκριμένα, η πιθανότητα ένα σημείο να προκύψει έξω από τα δύο ο.ε. συμβολίζεται με p, το μήκος ροής στο διάγραμμα συμβολίζεται με T_k, η προηγούμενη πιθανότητα, στην περίπτωση που η διεργασία είναι εντός ελέγχου με p_0 και το εντός–ελέγχου μήκος ροής με $T_{k,0}$. Οι τύποι που δίνουν τις πιθανότητες p και p_0 είναι ακριβώς οι ίδιοι, όπως στην περίπτωση του μονόπλευρου δ.ε. *k*–από–*k*. Το ίδιο ισχύει και για τους τύπους που δίνουν τη σ.π. των τ.μ. T_k και $T_{k,0}$, όπως και για αυτούς που δίνουν τη μέση τιμή και τη διακύμανση των τ.μ. T_k και $T_{k,0}$. Επίσης, το αποτέλεσμα ότι το δίπλευρο δ.ε. *k*–από–*k* είναι μη–παραμετρικό, προκύπτει ακριβώς με την ίδια στατιστική μεθοδολογία, όπως στο αντίστοιχο μονόπλευρο δ.ε.. Τέλος, τα δύο ο.ε. για αυτό το διάγραμμα υπολογίζονται ακριβώς με την ίδια διαδικασία, όπως υπολογίζεται το ο.ε. για το αντίστοιχο μονόπλευρο δ.ε. *k*–από–*k*. Πρέπει εδώ να σημειωθεί ότι, συνήθως στην πράξη, για το δίπλευρο διάγραμμα, το κ.ο.ε. *LCL* λαμβάνεται ως το αντίθετο του α.ο.ε. *UCL*, δηλαδή ισχύει η σχέση $LCL = -UCL$, λόγω της συμμετρικότητας που παρουσιάζει η σ.π. της τ.μ. y_i στην περίπτωση που η διεργασία βρίσκεται εντός στατιστικού ελέγχου (βλέπε Chakraborti and Eryilmaz (2007)).

7.4 Δίπλευρα δ.ε. τύπου *DR* που χρησιμοποιούν τον κανόνα ροών 2–από–2

Τα διαγράμματα αυτά, είναι ειδική περίπτωση των δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*k*, για $k = 2$. Αυτά, προτάθηκαν για πρώτη φορά από τους Derman and Ross

(1997), από τα αρχικά γράμματα των οποίων πήραν την συγκεκριμένη ονομασία. Τα διαγράμματα *DR* χρησιμοποιούν την προσημική–βαθμολογική στατιστική συνάρτηση y_i (που παρουσιάστηκε σε προηγούμενη ενότητα) για το i τ.δ. που συλλέγεται και δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, είτε αν δύο συνεχόμενα σημεία (παρατηρήσεις) y_i στο δ.ε. βρεθούν πάνω από το α.ο.ε. *UCL*, είτε αν αυτά βρεθούν κάτω από το κ.ο.ε. *LCL*, είτε αν το ένα βρεθεί πάνω από το α.ο.ε. *UCL* και το άλλο κάτω από το κ.ο.ε. *LCL*, με οποιαδήποτε σειρά μεταξύ τους.

Αν x_i είναι μία δίτιμη τ.μ. που παίρνει την τιμή 1, στην περίπτωση που το αντίστοιχο σημείο (τιμή της στατιστικής συνάρτησης ελέγχου) y_i από το i τ.δ. πάρει τιμή έξω από τα δύο ο.ε. *LCL* και *UCL* και την τιμή 0 στην αντίθετη περίπτωση και επίσης T_2 είναι η διακριτή τ.μ. που δηλώνει το μήκος ροής σε αυτό το δίπλευρο δ.ε., η τ.μ. T_2 μπορεί να γραφεί (βλέπε Derman and Ross (1997)) στη μορφή

$$T_2 = \min\{t : x_{t-1} + x_t = 2\}.$$

Ακόμη, αν p είναι η πιθανότητα $p = P(x_i = 1) = P(y_i < LCL$ ή $y_i > UCL)$, η σ.π. της διακριτής τ.μ. T_2 θα δίνεται από τον παρακάτω τύπο:

$P(T_2 = x) = 0$, αν ισχύει $0 \le x < 2$,

$P(T_2 = x) = p^2$, αν ισχύει $x = 2$ και

$$P(T_2 = x) = \sum_{j=1}^{[\frac{x+1}{3}]} (-1)^{j-1} p^{2j} (1-p)^{j-1} \left[\binom{x-2j-1}{j-2} + (1-p)\binom{x-2j-1}{j-1} \right], \text{ αν ισχύει } x \ge 3.$$

Είναι μάλλον προφανές ότι, ο τύπος που δίνει τη σ.π. της τ.μ. T_2, προκύπτει άμεσα από τον τύπο που δίνει την σ.π. της τ.μ. T_k στον κανόνα ροών *k*–από–*k*, απλά θέτοντας $k = 2$. Η μέση τιμή της T_2 (δηλαδή το μέσο μήκος ροής *ARL*) και η διακύμανσή της θα δίνονται (βλέπε Derman and Ross (1997)) από τους παρακάτω τύπους:

$$E(T_2) = ARL = \frac{1-p^2}{(1-p)p^2} = \frac{1+p}{p^2} = \frac{1}{p^2} + \frac{1}{p} \quad \text{και} \quad Var(T_2) = \frac{1-5(1-p)p^2 - p^5}{(1-p)^2 p^4}.$$

Επειδή έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η σ.π. της τ.μ. y_i είναι συμμετρική ως προς τα δύο ο.ε. *LCL* και *UCL*, το κ.ο.ε. *LCL* συνήθως ορίζεται ως το αντίθετο του α.ο.ε. *UCL*, δηλαδή τα δύο ο.ε. συνδέονται μέσω της σχέσης $LCL = -UCL$. Αν η διεργασία είναι εντός στατιστικού ελέγχου, θα ισχύει:

$$\begin{aligned} p_0 &= P(y_i < LCL \text{ ή } y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}) = \\ &= P(y_i < LCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}) \\ &\quad + P(y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου})) \\ &= P(y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}) \\ &\quad + P(y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}) \\ &= 2\,P(y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}) = 2\,p_0^+ . \end{aligned}$$

Άρα, θα ισχύει η σχέση

$$p_0 = 2\,p_0^+ ,$$

όπου p_0^+ είναι η πιθανότητα λανθασμένου συναγερμού στο αντίστοιχο μονόπλευρο δ.ε. με κανόνα ροών 2–από–2.

Αν η διεργασία είναι εντός στατιστικού ελέγχου, η σ.π. της τ.μ. $T_{2,0}$, η οποία δηλώνει το εντός–ελέγχου μήκος ροής σε αυτό το μονόπλευρο δ.ε. *DR*, θα δίνεται από τον παρακάτω τύπο (βλέπε Balakrishnan and Koutras (2002)):

$P(T_{2,0} = x) = 0$, αν $0 \le x < 2$,

$P(T_{2,0} = x) = p_0^2$, αν $x = 2$ και

$$P(T_{2,0} = x) = \sum_{j=1}^{[\frac{x+1}{3}]} (-1)^{j-1} p_0^{2j} (1-p_0)^{j-1} [\binom{x-2j-1}{j-2} + (1-p_0)\binom{x-2j-1}{j-1}], \text{ αν } x \ge 3,$$

όπου p_0 είναι η πιθανότητα

$$\begin{aligned} p_0 &= P(y_i < LCL \text{ ή } y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}) \\ &= 2\,P(y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}). \end{aligned}$$

Επίσης, σε αυτήν την περίπτωση, η εντός–ελέγχου μέση τιμή της τ.μ. $T_{2,0}$ (δηλαδή το εντός–ελέγχου μέσο μήκος ροής ARL_{in}) και η εντός–ελέγχου διακύμανσή της θα δίνονται από τους παρακάτω τύπους:

$$E(T_{2,0}) = ARL_{in} = \frac{1}{p_0^2} + \frac{1}{p_0} = \frac{1}{4(p_0^+)^2} + \frac{1}{2p_0^+} \quad \text{και} \quad Var(T_{2,0}) = \frac{1-5(1-p_0)p_0^2 - p_0^5}{(1-p_0)^2 p_0^4},$$

όπου p_0^+ είναι η αντίστοιχη πιθανότητα της πιθανότητας p_0, για το μονόπλευρο δ.ε. με κανόνα ροών 2–από–2.

Ακόμη, η πιθανότητα λανθασμένου συναγερμού *FAR* σε αυτό το δ.ε., θα είναι ίση με:

FAR = $p_{2,0}$ = $P(y_{i-1} < LCL$ | εντός ελέγχου) $P(y_i < LCL$ | εντός ελέγχου)

+ $P(y_{i-1} < LCL$ | εντός ελέγχου) $P(y_i > UCL$ | εντός ελέγχου)

+ $P(y_i > UCL$ | εντός ελέγχου) $P(y_{i-1} < LCL$ | εντός ελέγχου)

+ $P(y_i > UCL$ | εντός ελέγχου) $P(y_i > UCL$ | εντός ελέγχου)

= $(p_0^+)^2 + (p_0^+)^2 + (p_0^+)^2 + (p_0^+)^2 = 4(p_0^+)^2$,

οπότε, καταλήγουμε στη σχέση

$$FAR = p_{2,0} = 4(p_0^+)^2,$$

όπου $(p_0^+)^2$ είναι η πιθανότητα λανθασμένου συναγερμού για το αντίστοιχο μονόπλευρο δ.ε. με κανόνα ροών 2–από–2 και το κ.ο.ε. *LCL* είναι το αντίθετο του α.ο.ε. *UCL*, δηλαδή ισχύει *LCL* = –*UCL*. Άρα, η πιθανότητα λανθασμένου συναγερμού για το δίπλευρο δ.ε. *DR* με κανόνα ροών 2–από–2, είναι τετραπλάσια της πιθανότητας λανθασμένου συναγερμού για το μονόπλευρο διάγραμμα με κανόνα ροών 2–από–2.

Στην περίπτωση της εντός–ελέγχου διεργασίας, έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η σ.π. της τ.μ. $T_{2,0}$ (που δηλώνει το εντός–ελέγχου μήκος ροής σε αυτό το δίπλευρο διάγραμμα) δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από το μέγεθος των τ.δ. που συλλέγονται. Άρα και η πιθανότητα λανθασμένου συναγερμού *FAR* = $p_{2,0}$, καθώς και το εντός–ελέγχου μέσο μήκος ροής ARL_{in}, δε θα εξαρτώνται από την κατανομή που ακολουθούν οι παρατηρήσεις. Επομένως, σε αυτό το σημείο, μπορούμε να ισχυριστούμε ότι το δίπλευρο δ.ε. *DR* που χρησιμοποιεί τον κανόνα ροών 2–από–2 είναι μη–παραμετρικό.

Συνήθως στην πράξη, το δίπλευρο δ.ε. *DR* κατασκευάζεται (βλέπε Chakraborti and Eryilmaz (2007)) σύμφωνα με την ακόλουθη διαδικασία: αρχικά, προσδιορίζεται μία επιθυμητή και αρκετά μικρή τιμή για την πιθανότητα

$p_0 = P(y_i < LCL$ ή $y_i > UCL$ | η διεργασία εντός στατιστικού ελέγχου) ή μία επιθυμητή και αρκετά μεγάλη τιμή για το εντός–ελέγχου μέσο μήκος ροής ARL_{in}. Στην τελευταία περίπτωση, η πιθανότητα p_0, για δεδομένη τιμή του ARL_{in}, προσδιορίζεται μέσω της σχέσης

$$ARL_{in} = \frac{1+p_0}{p_0^2} = \frac{1}{p_0^2} + \frac{1}{p_0}.$$

Στη συνέχεια, υπολογίζεται η πιθανότητα p_0^+ για το αντίστοιχο μονόπλευρο δ.ε. *DR*, μέσω της σχέσης

$$p_0 = 2\, p_0^+ ,$$

δηλαδή υπολογίζεται η πιθανότητα

$$p_0^+ = \frac{p_0}{2},$$

όπου

$$p_0 = P(y_i < LCL \text{ ή } y_i > UCL \mid \text{η διεργασία εντός στατιστικού ελέγχου}).$$

Μετά από αυτόν τον υπολογισμό, εξετάζεται μέσω ειδικών στατιστικών πινάκων (βλέπε Chakraborti and Eryilmaz (2007)) σε ποια τιμή του *UCL* αντιστοιχεί αυτή η τιμή $p_0^+ = \frac{p_0}{2}$ (η οποία αναφέρεται στο αντίστοιχο μονόπλευρο δ.ε. *DR*). Τότε, για το δίπλευρο δ.ε., μετά από τον προσδιορισμό του α.ο.ε., το κ.ο.ε. *LCL* θα οριστεί ως το αντίθετο του α.ο.ε., μέσω της σχέσης *LCL* = –*UCL* (λόγω της συμμετρικότητας που παρουσιάζει η σ.π. της τ.μ. y_i). Έτσι, μετά τον προσδιορισμό των δύο ο.ε., η πιθανότητα $P(y_i < LCL$ ή $y_i > UCL \mid$ εντός ελέγχου) στο δίπλευρο δ.ε. *DR* 2–από–2 θα είναι ίση με:

$$\begin{aligned} P(y_i < LCL \text{ ή } y_i > UCL \mid \text{εντός ελέγχου}) &= \\ &= P(y_i < LCL \mid \text{εντός ελέγχου}) + P(y_i > UCL \mid \text{εντός ελέγχου}) \\ &= P(y_i < -UCL \mid \text{εντός ελέγχου}) + P(y_i > UCL \mid \text{εντός ελέγχου}) \\ &= 2\,P(y_i > UCL \mid \text{εντός ελέγχου}) = 2\,p_0^+ = 2\frac{p_0}{2} = p_0. \end{aligned}$$

Τα δίπλευρα δ.ε. *DR* που χρησιμοποιούν τον κανόνα ροών 2–από–2, παρουσιάζουν αρκετά ικανοποιητικές και χρήσιμες στατιστικές ιδιότητες. Συγκεκριμένα (βλέπε Chakraborti and Eryilmaz (2007)), εμφανίζουν αρκετά μικρή πιθανότητα λανθασμένου συναγερμού και αρκετά μεγάλο εντός–ελέγχου μέσο μήκος ροής, κάτι που είναι μάλλον επιθυμητό στο

Στατιστικό Έλεγχο Διεργασίας. Έχει αποδειχθεί (βλέπε Derman and Ross (1997)) ότι, τα δίπλευρα δ.ε. *DR*, παρουσιάζουν μικρότερη πιθανότητα λανθασμένου συναγερμού και μεγαλύτερο εντός–ελέγχου μέσο μήκος ροής σε σύγκριση με τα υπόλοιπα μη–παραμετρικά δ.ε., καθώς και σε σύγκριση με τα κλασικά παραμετρικά δ.ε. τύπου Shewhart. Επίσης, έχει διαπιστωθεί (βλέπε Derman and Ross (1997)) ότι, για μικρές ή αρκετά μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των παρατηρήσεων, τα συγκεκριμένα δ.ε. εμφανίζουν μικρότερο εκτός–ελέγχου μέσο μήκος ροής σε σύγκριση με τα υπόλοιπα μη–παραμετρικά δ.ε.. Δηλαδή, τα συγκεκριμένα δ.ε., εντοπίζουν πιο γρήγορα μικρές ή πολύ μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής που ακολουθούν οι παρατηρήσεις σε σύγκριση με τα υπόλοιπα μη–παραμετρικά δ.ε.. Οι στατιστικές ιδιότητες που αναφέρθηκαν, προσδίδουν στα συγκεκριμένα διαγράμματα σημαντική ισχύ και αποτελεσματικότητα και κατά συνέπεια αυτά, μπορούν να αποτελέσουν μία πολύ καλή και ισχυρή εναλλακτική επιλογή για όσους ασχολούνται συστηματικά με το Στατιστικό Έλεγχο Διεργασίας. Από την άλλη πλευρά, τα συγκεκριμένα δ.ε. μειονεκτούν στο ότι, συνήθως, για την πρακτική εφαρμογή τους, απαιτούν τη συλλογή αρκετών ή πολλών τ.δ., κάτι που μπορεί να προκαλέσει μεγάλη δαπάνη χρόνου και κυρίως σημαντικό οικονομικό κόστος, ιδιαίτερα όταν το μέγεθος του καθενός τ.δ. που συλλέγεται είναι μεγάλο. Στην τελευταία περίπτωση όμως, βελτιώνεται σημαντικά η στατιστική ισχύς και αξιοπιστία των αποτελεσμάτων που προκύπτουν.

7.5 Δίπλευρα δ.ε. τύπου *KL* που χρησιμοποιούν τον κανόνα ροών 2–από–2

Τα διαγράμματα αυτά, είναι ειδική περίπτωση των δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*k*, για $k = 2$. Προτάθηκαν για πρώτη φορά από τον Klein (2000), από τα αρχικά γράμματα του οποίου πήραν τη συγκεκριμένη ονομασία. Τα διαγράμματα *KL* χρησιμοποιούν ως στατιστική συνάρτηση ελέγχου την προσημική–βαθμολογική στατιστική συνάρτηση y_i (που παρουσιάστηκε σε προηγούμενη ενότητα) για κάθε τ.δ. που συλλέγεται και δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν δύο συνεχόμενα σημεία (παρατηρήσεις) y_i στο δ.ε. βρεθούν είτε πάνω από το α.ο.ε. *UCL*, είτε κάτω από το κ.ο.ε. *LCL*. Σε περίπτωση όμως που ένα σημείο βρεθεί πάνω από το α.ο.ε. *UCL* και το αμέσως επόμενο σημείο βρεθεί κάτω από το κ.ο.ε. *LCL* ή το αντίστροφο, τότε το συγκεκριμένο δ.ε. δε θα δώσει γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, όπως συνέβαινε στο δίπλευρο δ.ε. DR που αναφέρθηκε σε προηγούμενη ενότητα. Αυτή ακριβώς είναι και η βασική (γραφική) διαφορά μεταξύ του διαγράμματος DR και του διαγράμματος KL.

Στο σημείο αυτό, πρέπει να σημειωθεί ότι, οι μαθηματικοί τύποι και τα στατιστικά αποτελέσματα που αναφέρθηκαν στο προηγούμενο δίπλευρο δ.ε. DR σχετικά με την ακριβή σ.π. του μήκους ροής και τα στατιστικά μεγέθη που συνδέονται με αυτήν (δηλαδή τη μέση τιμή και τη διακύμανσή της), δεν ισχύουν σε καμία περίπτωση στο δίπλευρο δ.ε. KL με κανόνα ροών 2–από–2, αφού το διάγραμμα DR και το διάγραμμα KL δεν ταυτίζονται διαγραμματικά, δηλαδή δε δίνουν τις ίδιες γραφικές ενδείξεις. Άρα και οι ακριβείς τύποι που δίνουν το μέσο μήκος ροής ARL και το εντός–ελέγχου μέσο μήκος ροής ARL_{in} στα δύο συγκεκριμένα δίπλευρα δ.ε., προφανώς δε θα ταυτίζονται μεταξύ τους.

Αποδεικνύεται (βλέπε Klein (2000)) ότι, το μέσο μήκος ροής ARL_2^* στο δ.ε. KL 2–από–2 θα δίνεται από τον παρακάτω τύπο:

$$ARL_2^* = \frac{1}{p - \frac{p^+}{1+p^+} - \frac{p^-}{1+p^-}} ,$$

όπου p είναι η πιθανότητα

$$p = P(y_i < LCL \text{ ή } y_i > UCL) = P(y_i < LCL) + P(y_i > UCL),$$

p^+ είναι η πιθανότητα

$$p^+ = P(y_i > UCL)$$

και p^- είναι η πιθανότητα

$$p^- = P(y_i < LCL).$$

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, από τον παραπάνω τύπο που δίνει το μέσο μήκος ροής ARL_2^*, προκύπτει σχετικά εύκολα ότι το εντός–ελέγχου μέσο μήκος ροής $ARL_{2,0}^*$ θα δίνεται από τον παρακάτω τύπο:

$$ARL_{2,0}^* = \frac{1}{2}[\frac{1}{(p_0^+)^2} + \frac{1}{p_0^+}] ,$$

όπου p_0^+ είναι η πιθανότητα

$p_0^+ = P(y_i > UCL \mid \text{εντός ελέγχου}) = P(y_i < -UCL \mid \text{εντός ελέγχου}) = P(y_i < LCL \mid$ εντός ελέγχου), λόγω της συμμετρικότητας που εμφανίζει η κατανομή πιθανότητας της διακριτής τ.μ. y_i και για αυτόν το λόγο το κ.ο.ε. *LCL* τέθηκε ίσο με *LCL* = –*UCL*. Και σε αυτό το δίπλευρο δ.ε. θα ισχύει η σχέση

$$p_0 = 2\, p_0^{+},$$

όπου

$$p_0 = P(y_i < LCL \text{ ή } y_i > UCL \mid \text{η διεργασία εντός ελέγχου}).$$

Επιπλέον, η πιθανότητα λανθασμένου συναγερμού *FAR*, θα είναι ίση με:

$FAR = p_{2,0}^{*} = P(y_{i-1} < LCL \mid \text{εντός ελέγχου})\, P(y_i < LCL \mid \text{εντός ελέγχου})$

$+\ P(y_{i-1} > UCL \mid \text{εντός ελέγχου})\, P(y_i > UCL \mid \text{εντός ελέγχου})$

$= (p_0^{+})(p_0^{+}) + (p_0^{+})(p_0^{+}) = 2(p_0^{+})^2 .$

Έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι, όταν η διεργασία είναι εντός στατιστικού ελέγχου, η σ.π. της τ.μ. y_i δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις. Επομένως, μπορούμε εδώ να ισχυριστούμε ότι, το συγκεκριμένο δίπλευρο δ.ε. *KL* με κανόνα ροών 2–από–2 , είναι μη–παραμετρικό.

Συνήθως στην πράξη, το δίπλευρο δ.ε. *KL* κατασκευάζεται ακριβώς με την ίδια διαδικασία, όπως κατασκευάζεται το δίπλευρο δ.ε. *DR* που παρουσιάστηκε στην αμέσως προηγούμενη ενότητα (βλέπε σελίδες 98 και 99).

Τα δίπλευρα δ.ε. *KL* 2–από–2, εμφανίζουν γενικά, τα ίδια πλεονεκτήματα με τα αντίστοιχα δ.ε. *DR* 2–από–2. Δηλαδή, τα διαγράμματα αυτά, παρουσιάζουν μικρότερη πιθανότητα λανθασμένου συναγερμού και μεγαλύτερο εντός–ελέγχου μέσο μήκος ροής συγκριτικά με τα υπόλοιπα μη–παραμετρικά δ.ε.. Επίσης, αν είναι σταθερό το εντός–ελέγχου μέσο μήκος ροής και η πιθανότητα λανθασμένου συναγερμού, τότε για μικρές μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των παρατηρήσεων, αυτά τα διαγράμματα εμφανίζουν μικρότερο εκτός–ελέγχου μέσο μήκος ροής συγκριτικά με τα υπόλοιπα μη–παραμετρικά δ.ε., καθώς και με τα παραμετρικά δ.ε. τύπου Shewhart, δηλαδή ανιχνεύουν πιο εύκολα και έγκαιρα αυτές τις μετατοπίσεις από τα υπόλοιπα δ.ε., με σταθερές όμως τις παραπάνω ποσότητες. Τα συγκεκριμένα δ.ε., υπερτερούν σε όλους τους παραπάνω τομείς, απέναντι και στα δίπλευρα δ.ε. *DR* 2–από–2 που αναφέρθηκαν σε προηγούμενη ενότητα και αυτό, τους προσδίδει μεγάλη στατιστική σημαντικότητα, αποτελεσματικότητα και ισχύ.

Το μοναδικό μειονέκτημα που παρουσιάζουν, είναι ότι απαιτούν την συλλογή αρκετών ή πολλών δειγμάτων, κάτι που μπορεί να προκαλέσει μεγάλη δαπάνη χρόνου και κυρίως σημαντικό οικονομικό κόστος, ιδιαίτερα αν το μέγεθος του καθενός τ.δ. που συλλέγεται είναι μεγάλο.

7.6 Δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*w*

Και αυτά τα διαγράμματα προτάθηκαν και μελετήθηκαν από τους Chakraborti and Eryilmaz (2007). Δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν στα *w* τελευταία σημεία (τιμές της στατιστικής συνάρτησης ελέγχου) y_i που απεικονίζονται στο διάγραμμα υπάρχουν k σημεία y_i που παίρνουν τιμές έξω από τα δύο ο.ε. LCL και UCL, αν το δ.ε. είναι δίπλευρο, ή έξω από το ο.ε., αν το δ.ε. είναι μονόπλευρο. Αναφέρεται ξανά ότι η ποσότητα y_i παριστάνει την προσημική–βαθμολογική στατιστική συνάρτηση του Wilcoxon (βλέπε Στέγγος (2006)) που προκύπτει από το i τ.δ. που συλλέγεται.

Αν $T_k^{(w)}$ είναι η διακριτή τ.μ. που δηλώνει το μήκος ροής στο *k*–από–*w* δ.ε., έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η τ.μ. $T_k^{(w)}$ μπορεί να γραφεί στη μορφή

$$T_k^{(w)} = \min\{t : \sum_{j=\max(t-w+1,1)}^{t} X_j \geq k\},$$

όπου X_j είναι μία δίτιμη τ.μ. που παίρνει την τιμή 1 αν η παρατήρηση y_i από το i τ.δ. πάρει τιμή έξω από τα δύο ο.ε. LCL και UCL (ή έξω από το ο.ε. στο μονόπλευρο διάγραμμα) και την τιμή 0 στην αντίθετη περίπτωση. Η κατανομή πιθανότητας της διακριτής τ.μ. $T_k^{(w)}$ ονομάζεται «γεωμετρική κατανομή τάξης *k*/*m*» (βλέπε Balakrishnan and Koutras (2002)).

Για λόγους απλότητας στους συμβολισμούς (και στους αντίστοιχους τύπους), θα εξετάσουμε στη συνέχεια, την ειδική περίπτωση των δ.ε. που χρησιμοποιούν τον κανόνα ροών *k*–από–*w*, για $k = 2$. Αυτά, δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν στα *w* τελευταία σημεία (παρατηρήσεις) y_i που απεικονίζονται στο διάγραμμα υπάρχουν 2 σημεία y_i που παίρνουν τιμές έξω από τα δύο ο.ε. LCL και UCL, στην περίπτωση που το δ.ε. είναι δίπλευρο, ή έξω από το ο.ε., στην περίπτωση που το δ.ε. είναι μονόπλευρο, όπου y_i είναι η στατιστική συνάρτηση που αναφέρθηκε προηγουμένως.

Αν $T_2^{(w)}$ είναι η διακριτή τ.μ. που δηλώνει το μήκος ροής στο 2–από–*w* δ.ε., η τ.μ. $T_2^{(w)}$ μπορεί (βλέπε Chakraborti and Eryilmaz (2007)) να γραφτεί στη μορφή

$$T_2^{(w)} = \min\{t : \sum_{j=\max(t-w+1,1)}^{t} X_j \geq 2\},$$

όπου X_j είναι μία δίτιμη τ.μ. που παίρνει την τιμή 1, αν η παρατήρηση y_i από το *i* τ.δ. πάρει τιμή έξω από τα δύο ο.ε. *LCL* και *UCL* (ή έξω από το ο.ε. στην περίπτωση του μονόπλευρου διαγράμματος) και την τιμή 0 στην αντίθετη περίπτωση. Η κατανομή πιθανότητας της διακριτής τ.μ. $T_2^{(w)}$ ονομάζεται «γεωμετρική κατανομή τάξης 2/*m*» (βλέπε Balakrishnan and Koutras (2002)). Αν p_2 είναι η πιθανότητα ένα οποιοδήποτε σημείο που απεικονίζεται στο δ.ε. να πάρει τιμή έξω από τα δύο ο.ε. *LCL* και *UCL* (ή έξω από το ο.ε. στο μονόπλευρο διάγραμμα), έχει βρεθεί (βλέπε Balakrishnan and Koutras (2002)) ότι η μέση τιμή και η διακύμανση της τ.μ. $T_2^{(w)}$ θα δίνονται από τους παρακάτω τύπους:

$$E(T_2^{(w)}) = \frac{2-(1-p_2)^{w-1}}{p_2(1-(1-p_2)^{w-1})}$$

και

$$Var(T_2^{(w)}) = \frac{1-p_2}{p_2^2} + (2w-1)\frac{(1-p_2)^{w-1}}{p_2(1-(1-p_2)^{w-1})^2} + \frac{1-p_2}{p_2^2(1-(1-p_2)^{w-1})^2}.$$

Στην περίπτωση της εντός–ελέγχου διεργασίας, έχει διαπιστωθεί ότι η σ.π. της τ.μ. $T_2^{(w)}$ δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από τον θετικό ακέραιο *w*, από το μέγεθος των τ.δ. που συλλέγονται και από τις τιμές που επιλέγονται για τα όρια ελέγχου στο διάγραμμα. Άρα και η πιθανότητα λανθασμένου συναγερμού, καθώς και το εντός–ελέγχου μέσο μήκος ροής, δε θα εξαρτώνται από την κατανομή που ακολουθούν οι παρατηρήσεις. Επομένως, μπορούμε να ισχυριστούμε ότι το δ.ε. 2–από–*w* είναι μη–παραμετρικό.

Όπως θα αναφερθεί στη συνέχεια, τα συγκεκριμένα δ.ε., παρουσιάζουν αρκετά χρήσιμες και σημαντικές στατιστικές και πιθανοθεωρητικές ιδιότητες, οι οποίες τους προσδίδουν μεγάλη στατιστική ισχύ, αποδοτικότητα και αποτελεσματικότητα και τα καθιστούν ένα πολύ χρήσιμο στατιστικό εργαλείο για όσους ασχολούνται συστηματικά με το Στατιστικό Έλεγχο Διεργασίας.

7.7 Δ.ε. που χρησιμοποιούν τον κανόνα ροών 2–από–3

Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα δ.ε. που χρησιμοποιούν τον κανόνα ροών 2–από–3. Τα διαγράμματα αυτά, είναι προφανώς, ειδική περίπτωση των δ.ε. που χρησιμοποιούν τον κανόνα ροών 2–από–*w*, για *w* = 3. Δίνουν γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, αν και μόνο αν στα 3 τελευταία σημεία (παρατηρήσεις) y_i που απεικονίζονται στο διάγραμμα υπάρχουν 2 σημεία y_i που παίρνουν τιμές έξω από τα δύο ο.ε. LCL και UCL, αν το δ.ε. είναι δίπλευρο, ή έξω από το ο.ε., αν το δ.ε. είναι μονόπλευρο. Η ποσότητα (τ.μ.) y_i έχει οριστεί σε προηγούμενες ενότητες. Αν $T_2^{(3)}$ είναι η διακριτή τ.μ. που δηλώνει το μήκος ροής στο δ.ε. 2–από–3, έχει βρεθεί (βλέπε Chakraborti and Eryilmaz (2007)) ότι η τ.μ. $T_2^{(3)}$ μπορεί να γραφτεί στη μορφή

$$T_2^{(3)} = \min\{t : \sum_{j=\max(t-2,1)}^{t} X_j \geq 2\},$$

όπου X_j είναι μία δίτιμη τ.μ. που παίρνει την τιμή 1, αν η παρατήρηση y_i από το ***i*** τ.δ. πάρει τιμή έξω από τα δύο ο.ε. LCL και UCL (ή έξω από το ο.ε. στο μονόπλευρο διάγραμμα) και την τιμή 0 στην αντίθετη περίπτωση. Η κατανομή πιθανότητας της διακριτής τ.μ. $T_2^{(3)}$ ονομάζεται «γεωμετρική κατανομή τάξης 2/3» (βλέπε Balakrishnan and Koutras (2002)).

Αν στο μονόπλευρο δ.ε. υπάρχει μόνο το α.ο.ε. UCL και p_0^+ είναι η πιθανότητα

$$p_0^+ = P(y_i > UCL \mid \text{η διεργασία είναι εντός ελέγχου}),$$

στην περίπτωση της εντός–ελέγχου διεργασίας, η εντός–ελέγχου μέση τιμή και διακύμανση της τ.μ. $T_{2,0}^{(3)}$ (που δηλώνει το εντός–ελέγχου μήκος ροής στο συγκεκριμένο διάγραμμα) θα δίνονται (βλέπε Chakraborti and Eryilmaz (2007)) από τους παρακάτω τύπους:

$$ARL_{2,0}^{3,+} = E(T_{2,0}^{(3)}) = \frac{1+2p_0^+ - (p_0^+)^2}{(p_0^+)^2(2-p_0^+)}$$

και

$$Var(T_{2,0}^{(3)}) = \frac{(1-p_0^+)(1+p_0^+)[(p_0^+)^3 - 5(p_0^+)^2 + 4p_0^+ + 1]}{(p_0^+)^4(2-p_0^+)^2}.$$

Επίσης, η πιθανότητα λανθασμένου συναγερμού *FAR*, θα είναι ίση με:

$FAR = p_{2,0}^{3,+} = P(y_{i-2} > UCL,\ y_{i-1} > UCL$ | εντός ελέγχου)

$+\ P(y_{i-2} > UCL,\ y_{i-1} \le UCL,\ y_i > UCL$ | εντός ελέγχου)

$+\ P(y_{i-2} \le UCL,\ y_{i-1} > UCL,\ y_i > UCL$ | εντός ελέγχου)

$= (p_0^+)(p_0^+) + (p_0^+)(1-p_0^+)(p_0^+) + (1-p_0^+)(p_0^+)(p_0^+)$.

Άρα, θα ισχύει

$$FAR = (p_0^+)^2 + 2(p_0^+)^2(1-p_0^+) = (p_0^+)^2(3-2p_0^+),$$

λόγω της ανεξαρτησίας που υπάρχει μεταξύ των σημείων (παρατηρήσεων) y_i στο δ.ε.. Αναφέρουμε ξανά ότι p_0^+ είναι η πιθανότητα

$$p_0^+ = P(y_i > UCL \mid \text{η διεργασία είναι εντός ελέγχου}).$$

Αν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, έχει διαπιστωθεί ότι (βλέπε Chakraborti and Eryilmaz (2007)) η σ.π. της τ.μ. $T_{2,0}^{(3)}$ δεν εξαρτάται από την κατανομή που ακολουθούν οι παρατηρήσεις, αλλά μόνο από το μέγεθος των τ.δ. που συλλέγονται. Οπότε, η πιθανότητα λανθασμένου συναγερμού $p_{2,0}^{3,+}$ και το εντός–ελέγχου μέσο μήκος ροής $ARL_{2,0}^{3,+}$, επίσης δε θα εξαρτώνται από την κατανομή που ακολουθούν οι παρατηρήσεις. Επομένως, στο σημείο αυτό συμπεραίνουμε ότι, το μονόπλευρο δ.ε. που χρησιμοποιεί τον κανόνα ροών 2–από–3, θα είναι μη–παραμετρικό.

Στην περίπτωση που στο μονόπλευρο δ.ε. υπάρχει μόνο το κ.ο.ε. *LCL*, p_0^- είναι η πιθανότητα

$$p_0^- = P(y_i < LCL \mid \text{η διεργασία είναι εντός ελέγχου}),$$

και η διεργασία βρίσκεται εντός στατιστικού ελέγχου, η μέση τιμή και η διακύμανση της τ.μ. $T_{2,0}^{(3)}$ (που δηλώνει το εντός–ελέγχου μήκος ροής σε αυτό το μονόπλευρο δ.ε.) θα δίνονται από τους ίδιους ακριβώς τύπους, όπως στην περίπτωση που στο μονόπλευρο διάγραμμα υπάρχει μόνο το α.ο.ε. *UCL*. Το μόνο που αλλάζει είναι ότι, στη θέση της πιθανότητας p_0^+, εμφανίζεται η πιθανότητα p_0^- που ορίστηκε προηγουμένως.

Επίσης, στην περίπτωση αυτή, η πιθανότητα λανθασμένου συναγερμού FAR θα δίνεται από τον ίδιο ακριβώς τύπο, όπως στην περίπτωση που υπάρχει μόνο το α.ο.ε. UCL. Το μόνο που αλλάζει είναι ότι, στη θέση της πιθανότητας p_0^+, εμφανίζεται η πιθανότητα p_0^- που ορίστηκε προηγουμένως.

Στην περίπτωση που στο μονόπλευρο διάγραμμα υπάρχει μόνο το κ.ο.ε. LCL, το αποτέλεσμα ότι το συγκεκριμένο δ.ε. είναι μη–παραμετρικό, προκύπτει με την ίδια ακριβώς διαδικασία, όπως στην περίπτωση που στο διάγραμμα υπήρχε μόνο το α.ο.ε. UCL.

Συνήθως στην πράξη, στο μονόπλευρο δ.ε. που χρησιμοποιεί τον κανόνα ροών 2–από–3, το ο.ε. υπολογίζεται με παρόμοια διαδικασία, όπως στα υπόλοιπα μη–παραμετρικά δ.ε. που χρησιμοποιούν άλλους κανόνες ροών και παρουσιάστηκαν σε προηγούμενες ενότητες.

Σε γενικές γραμμές, τα μονόπλευρα δ.ε. που χρησιμοποιούν τον κανόνα ροών 2–από–3, αποτελούν μία αρκετά καλή εναλλακτική επιλογή και ένα πολύ χρήσιμο και αξιόπιστο στατιστικό εργαλείο στα πλαίσια του Στατιστικού Ελέγχου Διεργασίας και των πρακτικών εφαρμογών του. Ο λόγος είναι ότι, τα συγκεκριμένα δ.ε. παρουσιάζουν αρκετά μικρή πιθανότητα λανθασμένου συναγερμού $p_{2,0}^{3,+}$ και αρκετά μεγάλο–εντός ελέγχου μέσο μήκος ροής $ARL_{2,0}^{3,+}$, αποτελέσματα που είναι επιθυμητά (και ίσως απαραίτητα) στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας. Μάλιστα, έχει διαπιστωθεί με χρήση προσομοίωσης ότι, για δεδομένη τιμή του ο.ε., τα συγκεκριμένα δ.ε. εμφανίζουν μεγαλύτερο εντός–ελέγχου μέσο μήκος ροής, αλλά και μικρότερη πιθανότητα λανθασμένου συναγερμού σε σύγκριση με τα μονόπλευρα δ.ε. DR που χρησιμοποιούν τον κανόνα ροών 2–από–2 και σε σύγκριση με τα μονόπλευρα δ.ε. KL που χρησιμοποιούν τον ίδιο κανόνα ροών.

Ακόμη, για δεδομένη τιμή του ο.ε., τα μονόπλευρα δ.ε. 2–από–3 πλεονεκτούν στα συγκεκριμένα στατιστικά χαρακτηριστικά και έναντι των μονόπλευρων δ.ε. που χρησιμοποιούν τον κανόνα ροών 1–από–1. Από όλα τα παραπάνω, προκύπτει το συμπέρασμα ότι αυτά τα δ.ε. παρουσιάζουν μεγάλη στατιστική ισχύ και αποτελεσματικότητα και φαίνονται να είναι περισσότερο αποδοτικά σε σύγκριση με τα υπόλοιπα μονόπλευρα μη–παραμετρικά δ.ε. που παρουσιάστηκαν.

Από την άλλη πλευρά, τα βασικότερα μειονεκτήματα για τα συγκεκριμένα δ.ε. είναι ότι, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, απαιτούν την αρχική εκτίμηση κάποιας στατιστικής (πληθυσμιακής) παραμέτρου της κατανομής των παρατηρήσεων (η οποία συνήθως είναι η διάμεσος ή η μέση τιμή) και επίσης υποθέτουν τη συλλογή αρκετών ή πολλών τ.δ., κάτι που μπορεί να προκαλέσει δαπάνη χρόνου και κυρίως μεγάλο οικονομικό κόστος, ιδιαίτερα αν το μέγεθος του καθενός τ.δ. που συλλέγεται είναι μεγάλο. Στην τελευταία περίπτωση όμως, βελτιώνεται αρκετά η στατιστική ισχύς και αξιοπιστία των αποτελεσμάτων που προκύπτουν.

7.8 Γενικές παρατηρήσεις και σχόλια στα μη–παραμετρικά δ.ε. που βασίζονται σε κανόνες ροών

Σε γενικές γραμμές, όπως ήδη αναφέρθηκε στις προηγούμενες ενότητες, τα μη–παραμετρικά δ.ε. που βασίζονται σε κανόνες ροών, παρουσιάζουν πολύ ικανοποιητικές και χρήσιμες στατιστικές ιδιότητες (βλέπε Chakraborti and Eryilmaz (2007)). Εκτός του ότι δεν υποθέτουν καμία συγκεκριμένη κατανομή για τις παρατηρήσεις που συλλέγονται και ότι τα στατιστικά χαρακτηριστικά τους (πιθανότητα λανθασμένου συναγερμού και εντός–ελέγχου μέσο μήκος ροής) δεν εξαρτώνται από την κατανομή των παρατηρήσεων, έχει αποδειχθεί ότι τα συγκεκριμένα δ.ε. εμφανίζουν αρκετά μικρή πιθανότητα λανθασμένου συναγερμού και αρκετά μεγάλο εντός –ελέγχου μέσο μήκος ροής, ιδιότητες που είναι μάλλον χρήσιμες στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας. Δηλαδή, τα συγκεκριμένα δ.ε., σπανίως δίνουν λανθασμένη γραφική ένδειξη ότι η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, κάτι που προφανώς είναι επιθυμητό και ίσως απαραίτητο στο Στατιστικό Έλεγχο Διεργασίας. Έχει βρεθεί επίσης ότι, για σταθερή τιμή της πιθανότητας λανθασμένου συναγερμού ή του εντός–ελέγχου μέσου μήκους ροής, τα δ.ε. που βασίζονται σε κανόνες ροών, παρουσιάζουν μικρότερο εκτός–ελέγχου μέσο μήκος ροής σε σύγκριση με τα υπόλοιπα μη–παραμετρικά δ.ε. που έχουν παρουσιαστεί. Έτσι, όταν η διεργασία βρίσκεται εκτός στατιστικού ελέγχου, τα συγκεκριμένα δ.ε., για σταθερές τις τιμές των παραπάνω ποσοτήτων, ανιχνεύουν πιο εύκολα και πιο γρήγορα τις μετατοπίσεις των στατιστικών παραμέτρων της κατανομής των παρατηρήσεων σε σύγκριση με τα υπόλοιπα μη–παραμετρικά δ.ε.. Επομένως, τα δ.ε. που βασίζονται σε κανόνες ροών, πλεονεκτούν έναντι των υπόλοιπων μη–παραμετρικών διαγραμμάτων και σε αυτόν τον τομέα.

Όμως, για να εφαρμοστούν τα συγκεκριμένα δ.ε., θα πρέπει, όταν η διεργασία βρίσκεται εντός στατιστικού ελέγχου, να δοθεί μία αρχική εκτίμηση της διαμέσου ή της μέσης τιμής της κατανομής των παρατηρήσεων. Επίσης, τα συγκεκριμένα δ.ε., προϋποθέτουν τη συλλογή αρκετών ή πολλών τ.δ., κάτι που μπορεί να προκαλέσει δαπάνη χρόνου και σημαντικό οικονομικό κόστος, ιδιαίτερα αν το μέγεθος του καθενός τ.δ. είναι μεγάλο. Εκτός από αυτό, η κατανομή πιθανότητας του εντός–ελέγχου μήκους ροής, το εντός–ελέγχου μέσο μήκος ροής και η διακύμανση του μήκους ροής δεν υπολογίζονται αρκετά εύκολα στα συγκεκριμένα δ.ε., αφού οι αντίστοιχοι μαθηματικοί τους τύποι είναι σχετικά πολύπλοκοι και με αρκετές μαθηματικές πράξεις. Εξάλλου, τα μη–παραμετρικά δ.ε. που χρησιμοποιούν κανόνες ροών, απαιτούν η (συνεχής) κατανομή των δεδομένων να είναι συμμετρική γύρω από τη διάμεσό της και αυτό είναι αρκετά πιθανό να μην ισχύει σε ορισμένες περιπτώσεις.

Σε κάθε περίπτωση πάντως, μπορούμε να ισχυριστούμε (βλέπε Chakraborti and Eryilmaz (2007)) ότι, τα μη–παραμετρικά δ.ε. που βασίζονται σε κανόνες ροών παρουσιάζουν αρκετά πλεονεκτήματα που προσδίδουν σε αυτά μεγάλη στατιστική ισχύ, αποδοτικότητα και αποτελεσματικότητα. Έτσι, προκύπτει τελικά το συμπέρασμα ότι, τα συγκεκριμένα διαγράμματα, αποτελούν και θα συνεχίσουν μάλλον να αποτελούν μία ισχυρή εναλλακτική επιλογή και ένα χρήσιμο στατιστικό εργαλείο στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας και του Στατιστικού Ελέγχου Ποιότητας, στις περιπτώσεις που υπάρχει ισχυρή αμφιβολία για την κανονικότητα των δεδομένων που συλλέγονται ή στις περιπτώσεις που η κατανομή των δεδομένων είναι άγνωστη ή μη προσεγγίσιμη.

ΚΕΦΑΛΑΙΟ 8

Σύνοψη

Τα μη–παραμετρικά δ.ε., σε γενικές γραμμές, αποτελούν μία αρκετά καλή εναλλακτική επιλογή όσον αφορά το Στατιστικό Έλεγχο Διεργασίας και το Στατιστικό Έλεγχο Ποιότητας γενικότερα. Ο βασικός λόγος για αυτό είναι ότι, τα συγκεκριμένα διαγράμματα εμφανίζουν πολύ ικανοποιητικά και χρήσιμα στατιστικά χαρακτηριστικά, τα οποία τα καθιστούν αρκετά ευέλικτα σε ορισμένες ειδικές περιπτώσεις που παρουσιάζονται κατά τη συλλογή των παρατηρήσεων. Συγκεκριμένα (βλέπε Chakraborti, Bakir and Van Der Laan (2001)), τα μη–παραμετρικά δ.ε. δεν απαιτούν κάποια συγκεκριμένη (συνεχή) κατανομή για τα δεδομένα, σε αντίθεση με τα κλασικά παραμετρικά δ.ε., που υποθέτουν ότι οι παρατηρήσεις που συλλέγονται ακολουθούν όλες την κανονική κατανομή. Επίσης, αυτά τα δ.ε., εμφανίζουν την ίδια πιθανότητα λανθασμένου συναγερμού *FAR*, καθώς και το ίδιο εντός–ελέγχου μέσο μήκος ροής ARL_{in} για όλες τις κατανομές που είναι πιθανό να ακολουθούν οι παρατηρήσεις. Τέλος, έχει αποδειχθεί ότι τα συγκεκριμένα δ.ε. είναι εξαιρετικά ανθεκτικά (robust) απέναντι στην παρουσία τυχόν έκτροπων παρατηρήσεων.

Εκτός από αυτό, η κατασκευή τους είναι σχετικά απλή και εύκολη, αφού τα δύο ο.ε. ή το ο.ε. δε χρειάζεται να προσδιοριστούν ή να εκτιμηθούν μέσω ενός συγκεκριμένου μαθηματικού τύπου, αλλά προσδιορίζονται μέσω ειδικών στατιστικών πινάκων που έχουν κατασκευαστεί. Οι πίνακες αυτοί, για γνωστό μέγεθος των τ.δ. που συλλέγονται και για δεδομένη πιθανότητα λανθασμένου συναγερμού ή για δεδομένο εντός–ελέγχου μέσο μήκος ροής, παρέχουν άμεσα τις τιμές που αντιστοιχούν στα δύο ο.ε. ή στο ο.ε..

Παρ’ολ’αυτά, τα συγκεκριμένα δ.ε. εμφανίζουν και ορισμένα μειονεκτήματα. Κατ’αρχήν, απαιτούν συνήθως τη συλλογή πολλών τ.δ. ή ενός τ.δ. μεγάλου μεγέθους ή σε κάποιες περιπτώσεις την συλλογή δύο τ.δ. μεγάλου μεγέθους. Αυτό, μπορεί να προκαλέσει σπατάλη χρόνου και κυρίως σημαντικό οικονομικό κόστος, ιδιαίτερα σε περιπτώσεις όπου τα μεγέθη των τ.δ. είναι σχετικά μεγάλα. Επίσης, ορισμένα από τα μη– παραμετρικά δ.ε., απαιτούν την αρχική εκτίμηση μιας στατιστικής παραμέτρου της κατανομής των δεδομένων, όπως για παράδειγμα της διαμέσου ή της μέσης τιμής. Αν το μέγεθος του προκαταρκτικού τ.δ. δεν

είναι μεγάλο (δηλαδή είναι μικρότερο από 20 ή 30 παρατηρήσεις), οι στατιστικές εκτιμήσεις που προκύπτουν για τις άγνωστες παραμέτρους της διεργασίας μάλλον δε θα έχουν μεγάλη στατιστική ισχύ και αξιοπιστία. Ένα άλλο μειονέκτημα για τα συγκεκριμένα διαγράμματα είναι ότι, ο ακριβής ή ο προσεγγιστικός, έστω, υπολογισμός του εκτός–ελέγχου μέσου μήκους ροής ARL_{out} είναι πρακτικά αδύνατος, αφού στα μη–παραμετρικά δ.ε. δεν είναι γνωστή η κατανομή που ακολουθούν οι παρατηρήσεις, επομένως δε θα είναι γνωστή και η νέα κατανομή των παρατηρήσεων, αφού η διεργασία βρεθεί εκτός στατιστικού ελέγχου. Έτσι λοιπόν, το εκτός–ελέγχου μέσο μήκος ροής ARL_{out} δε μπορεί να υπολογιστεί και είναι άγνωστο (απροσδιόριστο). Επιπλέον, στα περισσότερα μη–παραμετρικά δ.ε., είναι απαραίτητη η χρήση ειδικών στατιστικών πινάκων, χωρίς τους οποίους δεν είναι δυνατόν ή είναι αρκετά επίπονο, να προσδιοριστούν τα ο.ε. ή το ο.ε. για την κατασκευή αυτών των διαγραμμάτων. Δηλαδή, η αναζήτηση, η εύρεση και η τελική επεξεργασία των συγκεκριμένων πινάκων, ίσως να είναι μία επίπονη και χρονοβόρα διαδικασία για όσους πιθανόν ασχοληθούν με τα μη–παραμετρικά δ.ε. στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας. Τέλος, ορισμένα από τα συγκεκριμένα διαγράμματα, υποθέτουν ότι η (συνεχής) κατανομή των δεδομένων είναι συμμετρική γύρω από τη διάμεσό της (αν φυσικά υπάρχει η διάμεσος της κατανομής), κάτι που σε αρκετές περιπτώσεις δε συμβαίνει.

Σε κάθε περίπτωση πάντως, τα μη–παραμετρικά δ.ε. θεωρούνται ένα πολύ χρήσιμο στατιστικό και γραφικό εργαλείο στις σύγχρονες εφαρμογές του Στατιστικού Ελέγχου Διεργασίας και του Στατιστικού Ελέγχου Ποιότητας γενικότερα. Ο λόγος είναι ότι, σε ορισμένες περιπτώσεις, κατά τη συλλογή των παρατηρήσεων, υπάρχει ισχυρή αμφιβολία για την κανονικότητά τους ή δεν είναι γνωστή η ακριβής κατανομή που αυτές ακολουθούν ή δεν είναι γνωστός ο ακριβής συναρτησιακός (μαθηματικός) τύπος της κατανομής που αυτές ακολουθούν. Σε αυτές ακριβώς τις περιπτώσεις, η εφαρμογή και η χρήση ενός μη–παραμετρικού δ.ε. κρίνεται αρκετά χρήσιμη, αν όχι απολύτως απαραίτητη για όσους ασχολούνται συστηματικά με το αντικείμενο του Στατιστικού Ελέγχου Διεργασίας. Η (κάθε μορφής) έρευνα όμως πάνω στα συγκεκριμένα δ.ε., δεν έχει ολοκληρωθεί ακόμη, αφού έχει διαπιστωθεί ότι ήδη υπάρχουν πολλά περιθώρια μελλοντικής επιστημονικής και ακαδημαϊκής έρευνας πάνω στα συγκεκριμένα δ.ε. (βλέπε Chakraborti, Bakir and Van Der Laan (2001)). Στα επόμενα έτη μάλιστα, αναμένονται (βλέπε Chakraborti, Bakir and Van Der Laan (2001)) αρκετά καινούργια θεωρητικά και πρακτικά αποτελέσματα, σε ό,τι αφορά ολόκληρη τη στατιστική θεωρία, αλλά και τη γραφική ανάλυση των μη–παραμετρικών δ.ε., που μάλλον για αρκετό χρονικό διάστημα ακόμα θα συνεχίσουν να αποτελούν ένα πολύ ενδιαφέρον γνωστικό αντικείμενο συστηματικής ενασχόλησης και επιστημονικής έρευνας για τους ερευνητές και τους επιστήμονες παγκοσμίως.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Αντζουλάκος, Δ. (2006). *Στατιστικός Έλεγχος Ποιότητας*, Πανεπιστημιακές σημειώσεις για το ΠΜΣ «Εφαρμοσμένη Στατιστική», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Κούτρας, Μ. (2001). *Εισαγωγή στην Συνδυαστική*, Εκδόσεις Αθ. Σταμούλης, Αθήνα.

Κούτρας, Μ. (2002) (α). *Εισαγωγή στις Πιθανότητες, Θεωρία και Εφαρμογές, μέρος Ι*, Εκδόσεις Αθ. Σταμούλης, Αθήνα.

Κούτρας, Μ. (2002) (β). *Εισαγωγή στις Πιθανότητες, Θεωρία και Εφαρμογές, μέρος ΙΙ*, Εκδόσεις Αθ. Σταμούλης, Αθήνα.

Στέγγος, Δ. (2006). *Απαραμετρική Στατιστική*, Πανεπιστημιακές σημειώσεις για το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Thomas, G. B. and Finney, R. L. (1990). *Απειροστικός Λογισμός, Τόμος Β', Συναρτήσεις πολλών μεταβλητών, Διανυσματικές συναρτήσεις και σειρές*, Πανεπιστημιακές Εκδόσεις Κρήτης, Απόδοση στα Ελληνικά–Επιστημονική Επιμέλεια: Κανάρης Τσίγκανος.

Ξένη

Alloway, J. A. and Raghavachari, M. (1991). Control Chart Based on the Hodges–Lehmann estimator, *Journal of Quality Technology*, **23**, 336–347.

Amin, R. and Searcy, A. J. (1991). A Nonparametric Exponentially Weighted Moving Average Control Scheme, *Communications in Statistics–Simulation and Computation*, **20**, 1049–1072.

Amin, R. W., Reynolds, M. R. J$_{R}$. and Bakir, S. T. (1995). Nonparametric Quality Control Charts Based on the Sign Statistic, *Communications in Statistics–Theory and Methods*, **24**, 1597– 1624.

Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*, John Wiley, New York.

Bakir, S. T. and Reynolds, M. R. J$_{R}$. (1979). A Nonparametric Procedure for Process Control Based on Within–Group Ranking, *Technometrics*, **21**, 175–183.

Bhattacharyya, P. K. and Frierson, D. (1981). A Nonparametric Control Chart for Detecting Small Disorders, *Annals of Statistics*, **9**, 544–554.

Chakraborti, S. and Eryilmaz, S. (2007). A nonparametric Shewhart–type signed–rank control chart based on runs, *Communications in Statistics–Simulation and Computation*, **36**, 335–356.

Chakraborti, S., Bakir, S. T. and Van Der Laan P. (2001). Nonparametric Control Charts: an overview and some results, *Journal of Quality Technology*, **33**, 304–315.

Chakraborti, S., Van Der Laan, P. and Van De Wiel, M. A. (2004). A class of distribution–free control charts, *Applied Statistics*, **53**, 443–462.

Derman, C. and Ross, S. M. (1997). *Statistical Aspects of Quality Control*, Academic Press, San Diego.

Hackl, P. and Ledolter, J. (1991). A Control Chart Based on Ranks, *Journal of Quality Technology*, **23**, 117–124.

Hackl, P. and Ledolter, J. (1992). A New Nonparametric Quality Control Technique, *Communications in Statistics–Simulation and Computation*, **21**, 423–443.

Janacek, G. J. and Meikle, S. E. (1997). Control Charts Based on Medians, *The Statistician*, **46**, 19–31.

Klein, M. (2000). Two alternatives to the Shewhart $\overline{X}$ chart, *Journal of Quality Technology*, **32**, 427–431.

McDonald, D. (1990). A *CUSUM* Procedure Based on Sequential Ranks, *Naval Research Logistics*, **37**, 627–646.

Qiu, P. and Hawkins, D. (2003). A Nonparametric Multivariate *CUSUM* Procedure for Detecting Shifts in all Directions, *The Statistician*, **52**, part 2, 151–164.

Willemain, T. R. and Runger, G. C. (1996). Designing Control Charts Using an Empirical Reference Distribution, *Journal of Quality Technology*, **28**, 31–38.

Yashchin, E. (1992). Analysis of *CUSUM* and other Markov–Type Control Schemes by Using Empirical Distributions, *Technometrics*, **34**, 54–63.

Zacks, S. (1991). Detection and change–point problems, in *Handbook of Sequential Analysis*, Edited by B. K. Ghosh and P. K. Sen, Marcel Dekker, New York.