



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ & ΣΤΑΤΙΣΤΙΚΗΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

«ΜΟΝΤΕΛΑ ΑΚΟΛΟΥΘΙΑΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ
ΑΝΑΛΥΣΗΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ»

Αλέξανδρος Λ. Στεφανίδης

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
Σεπτέμβριος 2015



UNIVERSITY OF PIRAEUS
SCHOOL OF FINANCE & STATISTICS
DEPARTMENT OF STATISTICS AND INSURANCE
SCIENCE

**POSTGRADUATE PROGRAM
IN APPLIED STATISTICS**

**«SEQUENTIAL ANALYSIS MODELS
AND APPLICATIONS»**

Alexandros L. Stefanidis

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
in Master of Science in Applied Statistics

Piraeus
September 2015

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

Τα μέλη της Επιτροπής ήταν:

- Επίκουρος Καθηγητής Μ. Μπούτσικας (Επιβλέπων)
- Αναπληρωτής Καθηγητής Δ. Αντζουλάκος
- Καθηγητής Γ. Ηλιόπουλος

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

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον Επίκουρο Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς κ.Μ.Μπούτσικα για την επίβλεψη, τις οδηγίες και την καταλυτική συμβολή του στην συγγραφή της διπλωματικής εργασίας.

Σεπτέμβριος 2015

Στεφανίδης Αλέξανδρος

Περίληψη

Σκοπός της παρούσας εργασίας είναι η επισκόπηση των βασικών μαθηματικών μεθόδων της ακολουθιακής στατιστικής ανάλυσης. Ειδικότερα, παρουσιάζεται η μετρο-θεωρητική προσέγγιση στην απόδειξη των δύο εξισώσεων του Wald και της θεμελιώδους εξισώσεως, με χρήση της θεωρίας των martingales και των χρόνων στάσης. Μελετώνται οι ακριβείς και ασυμπτωτικές ιδιότητες των τεχνικών κατασκευής διαστήματος εμπιστοσύνης για το μέσο της κανονικής κατανομής, με σταθερό εκ των προτέρων εύρος, καθώς και η επέκταση της μεθοδολογίας στις p -διάστατες περιοχές εμπιστοσύνης. Δίνεται μία πλήρης περιγραφή του ακολουθιακού ελέγχου λόγου πιθανοφανειών και η προσέγγιση του Wald για την εύρεση των συναρτήσεων λειτουργικών χαρακτηριστικών και μέσου αριθμού δείγματος. Τέλος, εφαρμόζεται η ακολουθιακή μεθοδολογία με σκοπό τον αναλυτικό και γεωμετρικό ορισμό του στατιστικού αλγορίθμου σωρευτικών αθροισμάτων και την προσέγγιση στη συνάρτηση μέσου μήκους ροής. Για την εξαγωγή αριθμητικών αποτελεσμάτων χρησιμοποιήθηκε το στατιστικό πακέτο R.

Λέξεις-κλειδιά: Martingales, χρόνοι στάσης, ακολουθιακή διαδικασία, SPRT, CUSUM, μέσο μήκος ροής

Abstract

The purpose of this thesis is to review the basic methods of sequential analysis. In particular, we present the measure-theoretic approach for proving Wald's equations and the fundamental equation, via the martingale theory and the theory of stopping times. We study the exact and asymptotic properties of the techniques for constructing fixed width confidence intervals for the mean of the normal distribution and their extension to p -dimensional confidence regions. A complete description of the sequential probability ratio test is given along with Wald's approximation for estimating the operating characteristic function and the average sample number. Finally, applications of the sequential methodology are presented in order to define the analytical and geometric aspect of the cumulative sum statistical algorithm and the approximation of the average run length. The software package R was used for getting all the numerical results.

Keywords: Martingales, stopping times, sequential procedure, SPRT, CUSUM, average run length

Περιεχόμενα

Περιεχόμενα	xiii
Κατάλογος Πινάκων	xvi
Κατάλογος Σχημάτων	xviii
Κατάλογος Συντομογραφιών	xx
Εισαγωγή	1
1 Martingales και Χρόνοι Στάσης	3
1.1 Βασικές έννοιες	3
1.2 Χρόνοι στάσης	5
1.3 Το θεώρημα επιλεκτικής στάσης	7
1.4 Εξισώσεις του Wald	8
1.5 Η Θεμελιώδης εξίσωση	10
1.6 Θεωρία βέλτιστης στάσης	12
2 Ακολουθιακή Εκτιμητική κατά Διάστημα	15
2.1 Εύρεση βέλτιστου μεγέθους δείγματος	15
2.2 Μέθοδος του Stein	17
2.3 Κατανομή του μεγέθους του δείγματος	19
2.4 Πλήρως ακολουθιακή μέθοδος	20
2.5 Επέκταση στις p -διαστάσεις	21
2.5.1 Μέθοδος του Healy	23
2.5.2 Μέθοδος του Srivastava	23
2.6 Σύγκριση μεθόδων μέσω προσομοίωσης	24
3 Ακολουθιακός Έλεγχος Λόγου Πιθανοφαινειών	29
3.1 Εισαγωγή	29
3.2 Πεπερασμένος Τερματισμός	30
3.3 Υπολογισμός φραγμάτων	32

3.4	Βέλτιστη ιδιότητα	34
3.5	Συναρτήσεις OC και ASN	35
3.6	Σύγκριση μεθόδων μέσω προσομοίωσης	37
3.6.1	Σύγκριση με τον έλεγχο των Neyman-Pearson	37
3.6.2	Περίπτωση ανεξάρτητων μεταβλητών	39
3.6.3	Περίπτωση συσχετισμένων μεταβλητών	40
3.6.4	Σύγκριση των συναρτήσεων ASN και OC	42
4	Αλγόριθμος Σωρευτικών Αθροισμάτων	45
4.1	Γενικά περί στατιστικών αλγορίθμων ανίχνευσης	45
4.2	Αλγόριθμος CUSUM	47
4.2.1	Διασθητική ερμηνεία	47
4.2.2	Ο αλγόριθμος CUSUM ως επαναλαμβανόμενος έλεγχος SPRT	51
4.2.3	Μέθοδος V-mask	53
4.2.4	H-K μέθοδος	55
4.3	Συνάρτηση ARL	56
4.3.1	Προσεγγίσεις των Wald και Siegmund	58
4.3.2	Σύγκριση συναρτήσεων ARL	60
4.3.3	Υπολογισμός Παραμέτρων	63
4.4	Βέλτιστη ιδιότητα του CUSUM	64
	Σύνοψη	67
	Παράρτημα	68
	(I) Απόκλιση Kullback-Leibler	68
	(II) Θεωρία πινάκων	69
	(III) Κώδικας R για σχήματα	70
	(IV) Κώδικας R για προγράμματα	73
	Βιβλιογραφία	79

Κατάλογος Πινάκων

2.6.1	Ακριβείς και ασυμπτωτικές εκτιμήσεις των μεθόδων ($p=1$)	26
2.6.2	Ακριβείς και ασυμπτωτικές εκτιμήσεις των μεθόδων ($p=2$)	26
2.6.3	Ιδιότητες των μεθόδων εκτίμησης μέσου κατά διάστημα	27
3.1.1	SPRT στο n -στο βήμα	30
3.6.1	Σύγκριση στατιστικών ελέγχων Neyman-Pearson και SPRT	38
3.6.2	Σύγκριση τιμών OC και ASN ανάμεσα στην προσέγγιση του Wald και σε Monte Carlo προσομοίωση	43
4.3.1	Σύγκριση τιμών συναρτήσεων ARL μεταξύ των στατιστικών αλγορίθμων CUSUM και Shewhart	63

Κατάλογος Σχημάτων

1.5.1	Γραφική παράσταση της ροπογεννήτριας όταν η τ.μ. έχει αρνητική και θετική αντίστοιχα μέση τιμή	12
2.3.1	Ιστόγραμμα για το χρόνο στάσης T	20
3.3.1	Προσομοίωση υπό την \mathcal{H}_0 για $\theta_0 = 1$, $\theta_1 = 1.4$, $\sigma = 1$, $\alpha = 0.05$ και $\beta = 1$	34
3.6.1	Ιστογράμματα της τ.μ. T υπό την \mathcal{H}_0 και \mathcal{H}_1 αντίστοιχα	39
4.1	Μετατόπιση του μέσου κατανομής	46
4.2.1	Συμπεριφορά των σημείων (n, S_n) κατά την μετατόπιση μέσου	48
4.2.2	Συμπεριφορά των σημείων (n, S_n) κατά την μετατόπιση μέσου με χρήση του λόγου πιθανοφαιγιών	49
4.2.3	Συμπεριφορά των σημείων (n, G_n) κατά την μετατόπιση του μέσου.	50
4.2.4	Συμπεριφορά των σημείων (n, S_n) κατά την μετατόπιση του μέσου, με χρήση του SPRT	52
4.2.5	Περιγραφή του CUSUM ως V-mask	53
4.2.6	Έλεγχος του Armitage	54
4.2.7	H-K μέθοδος CUSUM για την ανίχνευσης μετατόπισης της μέσης τιμής όταν $\kappa = 0.5$ και $h = 5$	56
4.3.1	Συνάρτηση ARL για τις διάφορες τιμές του θ με προσομοίωση (x), με την προσέγγιση του Siegmund (διακεκομμένη γραμμή) και με την προσέγγιση του Wald	61
4.3.2	Συναρτήσεις ARL για θ_0 και θ_1 , σύμφωνα με την προσέγγιση του Siegmund, ως προς τις διάφορες τιμές του h	64

Κατάλογος Συντομογραφιών

τ.μ.	τυχαία μεταβλητή
i.i.d.	ανεξάρτητες και ισόνομες τυχαίες μεταβλητές
σ.π.π.	συνάρτηση πυκνότητας πιθανότητας
σ.κ.	συνάρτηση κατανομής
χ.σ.	χρόνος στάσης
δ.ε	διάστημα εμπιστοσύνης
σ.β.	σχεδόν βέβαια
σ.α.	στοχαστική ανέλιξη
θ.β.σ	Θεωρία Βέλτιστης Στάσης
K-L	Kullback-Leibler
AR(1)	first order autoregressive
OC	Operating Characteristic
ASN	Average Sample Number
SPRT	Sequential Probability Ratio Test
CUSUM	Cumulative Sum
ADD	Average Detection Delay
SADD	Supremum Average Detection Delay
ESADD	Essential Supremum Average Detection Delay
PFA	Probability of False Alarm

Εισαγωγή

Ακολουθιακή ανάλυση ονομάζεται ο κλάδος της στατιστικής που ασχολείται με προβλήματα όπου το μέγεθος του δείγματος δεν είναι σταθερό αλλά μεταβάλλεται σε κάθε πείραμα. Κατά τον Ghosh (1991), οι πρώτες ιδέες αυτού του κλάδου βρίσκονται στα μνημειώδη έργα των Huyghens, Bernoulli και De Moivre που αφορούσαν το πρόβλημα χρεωκοπίας του παίκτη (*gambler's ruin*). Οι αρχικές εφαρμογές των ακολουθιακών μεθόδων έγιναν στον έλεγχο ποιότητας και συγκεκριμένα στα διπλά δειγματοληπτικά σχέδια (Dodge & Romig, 1929) και στα διαγράμματα ελέγχου του Shewhart (1931). Αυτές ήταν οι τρεις σημαντικότερες περιοχές της στατιστικής στις οποίες το μέγεθος του δείγματος δεν ήταν σταθερό αλλά ήταν ουσιαστικά μία τυχαία μεταβλητή. Σκοπός της ακολουθιακής ανάλυσης είναι η εξαγωγή στατιστικών συμπερασμάτων με χρήση όσο το δυνατόν μικρότερου δείγματος, λαμβάνοντας υπόψιν τις ιδιότητες της τ.μ. που εκφράζει το μέγεθος του δείγματος.

Ο Wald (1947) ήταν ο πρώτος που θεμελίωσε έναν ακολουθιακό στατιστικό έλεγχο απλών υποθέσεων, ανάλογο με τον ομοιόμορφα ισχυρότατο έλεγχο των Neyman-Pearson (1933). Ο Wald όρισε την αποδοχή της \mathcal{H}_0 όταν ο λόγος των πιθανοφανειών γίνει μικρότερος από μία σταθερά έστω A , την απόρριψη της \mathcal{H}_0 όταν ο λόγος γίνει μεγαλύτερος από μία σταθερά B και την συνέχιση της δειγματοληψίας σε κάθε άλλη περίπτωση. Η προσπάθεια σύγκρισης αυτού του ελέγχου με τον κλασσικό έλεγχο των Neyman-Pearson οδήγησε τους Wald και Wolfowitz (1948) στη δημιουργία της θεωρίας βέλτιστης στάσης (θ.β.σ. - *optimal stopping theory*) και την μελέτη της μορφής των τυχαίων μεταβλητών που βελτιστοποιούν ορισμένες συνθήκες στα πλαίσια των ακολουθιακών ελέγχων. Η γενίκευση της θ.β.σ. βρίσκει σημαντικές εφαρμογές στη θεωρία παιγνίων (Ferguson, 2007), στα χρηματοοικονομικά μαθηματικά (Shiryaev, 2013) και στη θεωρία ταχύτατης ανίχνευσης αλλαγής κατανομής (*quickest change detection*) (Veeravalli, 2012).

Προβλήματα με μεταβαλλόμενο δείγμα παρουσιάζονται και στον κλάδο της εκτιμητικής θεωρίας (Mukhopadhyay, 2009), όπως η κατασκευή διαστήματος εμπιστοσύνης με σταθερό *εκ των προτέρων* εύρος και η εύρεση εκτιμητή του μέσου με *εκ των προτέρων* γνωστό φράγμα του ρίσκου αλλά και στον κλάδο της στοχαστικής προσέγγισης (Ghosh, 1997). Επίσης σημαντικές είναι οι συνεισφορές του Wald στη θεωρία των στοχαστικών ανεπίξεων, με την δημιουργία εξισώσεων που αφορούν σύνθετα αθροίσματα τυχαίων μεταβλητών. Θα δούμε πως η χρήση αυτών εξισώσεων χρησιμοποιείται στην εκτίμηση του μέσου μήκους ροής του

σωρευτικού αλγόριθμου ανίχνευσης μετατόπισης του μέσου (Page, 1954). Τέλος έχουν μελετηθεί εφαρμογές της ακολουθιακής ανάλυσης στις κλινικές δοκιμές, στην επεξεργασία σήματος και σε προβλήματα multiarmed bandit (Lai, 2001).

Martingales και Χρόνοι Στάσης

1.1 Βασικές έννοιες

Θα δανειστούμε κάποια βασικά αποτελέσματα από την θεωρία στοχαστικών ανελίξεων και την θεωρία των martingales, για την απόδειξη των εξισώσεων του Wald. Αυτό το κεφάλαιο αποτελεί το πιθανοθεωρητικό κομμάτι της ακολουθιακής στατιστικής και εμπεριέχει τα ουσιώδη εργαλεία για την εκτίμηση του χρόνου μίας διαδικασίας, όπως θα δούμε στα επόμενα κεφάλαια.

Έστω $(\Omega, \mathcal{F}, \mathbb{P})$ ο χώρος πιθανότητας, η τυχαία μεταβλητή (τ.μ.) $X: \Omega \rightarrow \mathbb{R}$ και $(\mathbb{R}, \mathcal{B}, \mu_X)$ ο επαγόμενος χώρος, όπου \mathcal{B} είναι η Borel σ -άλγεβρα του \mathbb{R} .

Ορισμός 1.1.1. Έστω \mathcal{C} μία κλάση υποσυνόλων του Ω . Η σ -άλγεβρα που παράγεται από την \mathcal{C} είναι η μικρότερη σ -άλγεβρα που περιέχει όλα τα στοιχεία της \mathcal{C} . Συμβολίζουμε με $\sigma(\mathcal{C})$.

Ορισμός 1.1.2. Ονομάζουμε ως παραγόμενη σ -άλγεβρα από την τ.μ. X το σύνολο $\{\sigma(X) := X^{-1}(B): B \in \mathbb{R}\}$. Όμοια επεκτείνουμε την έννοια της παραγόμενης σ -άλγεβρας για τις τ.μ. X_1, X_2, \dots, X_n σύμφωνα με την σχέση $\sigma(X_1, X_2, \dots, X_n) := \sigma(\cup_{i=1}^n \sigma(X_i))$.

Άμεση συνέπεια είναι ότι η X είναι \mathcal{F} -μετρήσιμη αν-ν $\sigma(X) \subset \mathcal{F}$. Υπενθυμίζουμε ότι γενικά $\mathbb{E}(X \cdot I_A) = \int_A X d\mathbb{P}$, $\forall A \in \mathcal{F}$, ενώ $\mathbb{E}(X \cdot I_{[X \in B]}) = \int_B x f_X(x) dx$ όταν η τ.μ. X είναι συνεχής και $\sum_{x \in B} x \mathbb{P}(X = x)$ όταν η X είναι διακριτή τ.μ. Θα εξετάσουμε τα θεωρήματα για συνεχείς τ.μ. Τα ίδια αποτελέσματα προκύπτουν και για την διακριτή περίπτωση και μπορούν να βρεθούν σε εγχειρίδια στοχαστικών ανελίξεων (π.χ. βλ. Walsh, 2014). Διατυπώνουμε την παρακάτω επέκταση του ορισμού της δεσμευμένης μέσης τιμής.

Ορισμός 1.1.3. Έστω \mathcal{G} σ -άλγεβρα τέτοια ώστε $\mathcal{G} \subset \mathcal{F}$. Εάν για την τ.μ. X ισχύει $\mathbb{E}|X| < \infty$ τότε ορίζουμε ως δεσμευμένη μέση τιμή της X δοθείσης της \mathcal{G} μια τ.μ. Z ώστε:

- (I1) η Z είναι \mathcal{G} -μετρήσιμη
- (I2) $\mathbb{E}(Z \cdot I_A) = \mathbb{E}(X \cdot I_A)$, $\forall A \in \mathcal{G}$

Θα συμβολίζουμε $Z = \mathbb{E}(X|\mathcal{G})$. Αποδεικνύεται με χρήση της έννοιας της Radon-Nikodym παραγώγου, ότι η Z ορίζεται όταν η X είναι τ.μ. που ικανοποιεί την σχέση $\mathbb{E}(|X|) < \infty$. Αποδεικνύεται ότι εάν Z' είναι μία άλλη τ.μ. που ικανοποιεί τις (I1) και (I2) τότε $Z' = Z$ σ.β.

Θα παραλείψουμε τον συμβολισμό σ, β για την ισότητα των δεσμευμένων μέσων τιμών. Αναφέρουμε κάποιες από τις ιδιότητες που θα χρησιμοποιηθούν στη συνέχεια σε αποδείξεις.

Ορισμός 1.1.4. Δύο σ -άλγεβρες \mathcal{G} και \mathcal{G}' θα ονομάζονται ανεξάρτητες εάν $\forall A \in \mathcal{G}$ και $\forall B \in \mathcal{G}'$: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Θα θεωρούμε ότι μία τ.μ. X , είναι ανεξάρτητη με την \mathcal{G}' εάν οι σ -άλγεβρες $\mathcal{G} = \sigma(X)$ και \mathcal{G}' είναι ανεξάρτητες.

Θεώρημα 1.1.1. Εάν οι X και Y είναι τ.μ. με $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$, τότε ισχύουν τα παρακάτω:

- (i) $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$
- (ii) Εάν η X είναι \mathcal{G} -μετρήσιμη, τότε $\mathbb{E}(X|\mathcal{G}) = X$
- (iii) $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$, $\forall a, b \in \mathbb{R}$
- (iv) Εάν η τ.μ. X είναι ανεξάρτητη με την \mathcal{G} τότε $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$
- (v) Εάν η X είναι \mathcal{G} -μετρήσιμη τότε $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$

Απόδειξη. (i) Προκύπτει από την ιδιότητα (I2) του ορισμού 1.1.3 θέτοντας $A = \Omega$.

(ii) Αφού η X είναι μετρήσιμη εφαρμόζουμε την ιδιότητα (I2) για $Z = X$.

(iii). Έστω $ab \neq 0$. Θα δείξουμε πρώτα ότι η τ.μ. $aX + bY$ είναι μετρήσιμη. Πράγματι, αφού η Borel σ -άλγεβρα παράγεται από τα διαστήματα της μορφής $(-\infty, x]$, όπου $x \in \mathbb{R}$, αρκεί να δειχθεί ότι $\{aX + bY \leq x\} \in \mathcal{G}$. Ισχύει

$$aX + bY < x \Rightarrow aX < x - bY$$

και λόγω της πυκνότητας των ρητών στο \mathbb{R} , $\exists q \in \mathbb{Q}$ τέτοιος ώστε

$$aX < q < x - bY \Rightarrow \{aX + bY < x\} = \bigcup_{q \in \mathbb{Q}} \left\{ X < \frac{q}{a} \right\} \cap \left\{ Y > \frac{x - q}{b} \right\} \in \mathcal{G}$$

αφού οι X και Y είναι \mathcal{G} -μετρήσιμες τ.μ. Χρησιμοποιώντας τώρα την προσθετική ιδιότητα των ολοκληρωμάτων προκύπτει το ζητούμενο. Για τις υπόλοιπες περιπτώσεις των a και b εργαζόμαστε κάνοντας χρήση ότι οι X και Y είναι \mathcal{G} -μετρήσιμες τ.μ.

(iv) Η τ.μ. $g: \Omega \rightarrow \mathbb{R}$ με $g(\omega) = \mathbb{E}(X)$ είναι προφανώς \mathcal{G} -μετρήσιμη. Έστω $A \in \mathcal{G}$. Λόγω της ανεξαρτησίας θα ισχύει

$$\mathbb{E}(X \cdot I_A) = \mathbb{E}(X)\mathbb{E}(I_A) = \mathbb{E}(\mathbb{E}(X) \cdot I_A)$$

σύμφωνα με την (iii).

(v) Η απόδειξη μπορεί να βρεθεί στον Walsh (2004).

Μπορούμε να δούμε την σ -άλγεβρα ως ένα σύνολο ενδεχομένων που περιέχει όλη την «πληροφορία» για την τ.μ. X . Έτσι, όταν σε ένα πείραμα παίρνουμε κατά βήματα το τυχαίο δείγμα X_1, X_2, \dots και η $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ εκφράζει την πληροφορία που παίρνουμε από τις X_1, X_2, \dots, X_n , η \mathcal{F}_n να συγκρατεί όλη την πληροφορία των προηγούμενων τ.μ. Δηλαδή $\mathcal{F}_{n-1} \subset \mathcal{F}_n$, $\forall n \in \mathbb{N}$. Πρακτικά, αυτό κυρίως εμφανίζεται σε στοχαστικά παιχνίδια που πραγματοποιούνται σε μία σειρά βημάτων. Η τ.μ. X_n δηλώνει την περιουσία ενός παίκτη στο n -στο βήμα. Τότε το ενδεχόμενο να κερδίσει ή να αποχωρήσει από το παιχνίδι ο παίκτης σε αυτό το βήμα, εξαρτάται αποκλειστικά από τις τιμές των X_1, X_2, \dots, X_n που θα είναι γνωστές μέχρι και εκείνη την στιγμή.

Ορισμός 1.1.5. Μία ακολουθία σ -αλγεβρών $(\mathcal{F}_n)_{n \geq 1}$ θα ονομάζεται διήθηση εάν $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$.

Ορισμός 1.1.6. Η στοχαστική ανέλιξη (σ.α.) $(X_n)_{n \geq 1}$ θα ονομάζεται προσαρμοσμένη ως προς την διήθηση $(\mathcal{F}_n)_{n \geq 1}$ εάν οι τ.μ. X_n είναι \mathcal{F}_n -μετρήσιμες $\forall n \in \mathbb{N}$.

Παρατηρούμε ότι κάθε σ.α. $(X_n)_{n \geq 1}$ είναι προσαρμοσμένη ως προς την διήθηση $(\mathcal{F}_n)_{n \geq 1}$, όπου $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Η διήθηση αυτή θα ονομάζεται συνήθης διήθηση της $(X_n)_{n \geq 1}$.

Ορισμός 1.1.7. Η σ.α. $(X_n)_{n \geq 1}$ όπου $\mathbb{E}|X_n| < \infty$ θα ονομάζεται *martingale* ως προς την διήθηση $(\mathcal{F}_n)_{n \geq 1}$ εάν είναι προσαρμοσμένη και $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$, $\forall n \in \mathbb{N}$. Θα την συμβολίζουμε με $(X_n, \mathcal{F}_n)_{n \geq 1}$.

Παράδειγμα 1.1.1. Εάν $(X_n, \mathcal{F}_n)_{n \geq 1}$ είναι *martingale*, τότε όλες οι τ.μ. X_n έχουν την ίδια αναμενόμενη τιμή αφού ισχύει

$$\mathbb{E}(X_{n+1}) = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \mathbb{E}(X_n)$$

και επομένως αναδρομικά,

$$\mathbb{E}(X_n) = \mathbb{E}(X_1), \forall n \in \mathbb{N} \quad \square$$

1.2 Χρόνοι στάσης

Τόσο στη στατιστική ανάλυση όσο και στη θεωρία των *martingales*, χρειαζόμαστε να προσδιορίσουμε τον τερματισμό της ακολουθιακής δειγματοληψίας. Αυτός θα ορίζεται εν μέρει από τον κανόνα απόφασης, αλλά και από την μεταβλητή που δηλώνει το βήμα στο οποίο θα σταματήσουμε.

Ορισμός 1.2.1. Έστω διήθηση $(\mathcal{F}_n)_{n \geq 1}$. Μία τ.μ. $T: \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ θα ονομάζεται *χρόνος στάσης* ή *διακοπής* (χ.σ.) ως προς την διήθηση $(\mathcal{F}_n)_{n \geq 1}$, εάν $\{T = n\} \in \mathcal{F}_n \forall n \in \mathbb{N}$.

Παράδειγμα 1.2.1. Η τ.μ.

$$T = \inf \{n \geq 1: S_n = a \text{ ή } S_n = b\}, \quad \inf \emptyset = \infty$$

όπου $S_n = \sum_{i=1}^n X_i$, είναι χρόνος στάσης (χ.σ.) ως προς την συνήθη διήθηση $(\mathcal{F}_n)_{n \geq 1}$ των τ.μ. X_1, X_2, \dots, X_n γιατί

$$\{T = n\} = \{S_1 \notin \{a, b\}\} \cap \dots \cap \{S_{n-1} \notin \{a, b\}\} \cap \{S_n \in \{a, b\}\} \in \mathcal{F}_n. \quad \square$$

Παράδειγμα 1.2.2. Συμβολίζουμε με $x \wedge y = \min\{x, y\}$. Εάν T_1 και T_2 είναι χρόνοι στάσης ως προς την διήθηση $(\mathcal{F}_n)_{n \geq 1}$, τότε η τ.μ. $T_1 \wedge T_2$ είναι χ.σ. Πράγματι

$$\{T_1 \wedge T_2 = n\} = (\{T_1 = n\} \cap \{T_2 \geq n\}) \cup (\{T_1 \geq n\} \cap \{T_2 = n\}) \in \mathcal{F}_n$$

αφού ισχύει

$$\{T_i \geq n\} = \Omega \setminus \left\{ \bigcup_{j=1}^{n-1} \{T_i = j\} \right\} \in \mathcal{F}_n$$

Θα δείξουμε τώρα ότι αν η ακολουθία $(X_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale ως προς την συνήθη διήθηση, τότε είναι και η $(X_{T \wedge n}, \mathcal{F}_{T \wedge n})_{n \geq 1}$ όπου $\mathbb{P}(T < \infty) = 1$ (η X_T πρέπει να είναι καλώς ορισμένη). Η τ.μ. $X_{T \wedge n}$ είναι \mathcal{F}_n -μετρήσιμη γιατί είναι $\mathcal{F}_{T \wedge n}$ -μετρήσιμη και $T \wedge n \leq n$. Μπορούμε να δούμε πρακτικά την $X_{T \wedge n}$ για $T = 4$ ως $X_1, X_2, X_3, X_4, X_4, X_4, \dots$. Άρα

$$|X_{T \wedge n}| \leq \max_{1 \leq i \leq n} \{|X_i|\} \leq \sum_{i=1}^n |X_i| \Rightarrow \mathbb{E}|X_{T \wedge n}| < \infty.$$

Παρατηρούμε ότι ισχύει

$$X_{T \wedge (n+1)} = X_{T \wedge n} + (X_{n+1} - X_n)I_{\{T > n\}}$$

και επομένως,

$$\begin{aligned} \mathbb{E}(X_{T \wedge (n+1)} | \mathcal{F}_n) &= X_{T \wedge n} + \mathbb{E}\left((X_{n+1} - X_n)I_{\{T > n\}} | \mathcal{F}_n\right) \\ &= X_{T \wedge n} + I_{\{T > n\}} \mathbb{E}\left((X_{n+1} - X_n) | \mathcal{F}_n\right) \\ &= X_{T \wedge n} + I_{\{T > n\}} (\mathbb{E}(X_{n+1} | \mathcal{F}_n) - X_n) \\ &= X_{T \wedge n} + I_{\{T > n\}} \cdot 0 \\ &= X_{T \wedge n} \end{aligned}$$

όπου στην πρώτη και δεύτερη εξίσωση χρησιμοποιήσαμε το γεγονός ότι η $I_{\{T > n\}}$ είναι \mathcal{F}_n -μετρήσιμη, τις ιδιότητες (ii) και (v) του Θεωρήματος 1.1.1 και ότι η $(X_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale. \square

Στη στατιστική υπάρχουν πολλά είδη χρόνων στάσης. Ενδιαφέρον παρουσιάζουν οι *χρόνοι απορρόφησης* (βλ. Παράδειγμα 1.2.1) και οι *χρόνοι εξόδου* (βλ. Παράγραφο 3.2). Εξαρχής θα θέλαμε η διαδικασία την οποία εξετάζουμε να σταματήσει σε πεπερασμένο πλήθος βημάτων με πιθ. 1. Δηλαδή θα θέλαμε χρόνους για τους οποίους ισχύει $\mathbb{P}(T < \infty) = 1$ (*ομαλοί χρόνοι*). Στο σημείο αυτό αξίζει να σημειωθεί ότι όταν $\mathbb{E}(T) < \infty$ τότε κατ'ανάγκη $\mathbb{P}(T < \infty) = 1$. Παρατηρούμε ότι για τα σύνολα $A_n = \{T \geq n\}$ ισχύει

$$A_{n+1} \subset A_n \text{ και } \{T = \infty\} = \bigcap_{n=1}^{\infty} A_n$$

και άρα από την ανισότητα του Markov

$$\mathbb{P}(A_n) \leq \frac{1}{n} \mathbb{E}(T) \Rightarrow \mathbb{P}(T = \infty) \leq \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0$$

Το αντίστροφο δεν ισχύει εν γένει. Για τον *χρόνο απορρόφησης* ενός συμμετρικού τυχαίου περιπάτου η αναμενόμενη τιμή δεν είναι πεπερασμένη (π.χ. βλ. Ghosh 1997, σελ. 27).

1.3 Το θεώρημα επιλεκτικής στάσης

Ο θεμελιώδης τρόπος με τον οποίο συνδέονται οι χρόνοι στάσης με την υπάρχουσα θεωρία των martingales δίνεται από το παρακάτω θεώρημα που ονομάζεται στη βιβλιογραφία ως *το θεώρημα της επιλεκτικής στάσης*.

Θεώρημα 1.3.1. *Εστω $(X_n, \mathcal{F}_n)_{n \geq 1}$ martingale και T ένας χ.σ.. Εάν ισχύουν οι συνθήκες:*

- (i) $\mathbb{P}(T < \infty) = 1$
- (ii) $\mathbb{E}|X_T| < \infty$
- (iii) $\lim_{n \rightarrow \infty} \mathbb{E}(X_n I_{\{T > n\}}) = 0$

τότε

$$\mathbb{E}(X_T) = \mathbb{E}(X_1).$$

Μια πιο εύχρηστη μορφή του θεωρήματος επιλεκτικής στάσης δίνεται από το επόμενο θεώρημα.

Θεώρημα 1.3.2. *Εστω martingale $(X_n, \mathcal{F}_n)_{n \geq 1}$ και T ένας χ.σ.. Τότε θα ισχύει $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ εάν μία από τις παρακάτω συνθήκες είναι αληθείς:*

- (i) $\exists c \in \mathbb{N}: T \leq c$ σ.β.
- (ii) $\mathbb{E}(T) < \infty$ και $\exists c \in \mathbb{R}$ ώστε $\mathbb{E}(|X_{n+1} - X_n| | \mathcal{F}_n) \leq c, \forall n < T$.

Η απόδειξη βασίζεται στην χρήση του martingale $(X_{T \wedge n}, \mathcal{F}_{T \wedge n})_{n \geq 1}$ του Παραδείγματος 1.2.2. Γράφοντας την $X_{T \wedge n}$ ως άθροισμα $X_{T \wedge n} = X_1 + \sum_{n=1}^{T \wedge n} (X_{n+1} - X_n)$ και κάνοντας χρήση των θεωρημάτων της κυριαρχημένης και της μονότονης σύγκλισης (π.χ. βλ. Χελιώτης, 2014, σελ. 26), αποδεικνύεται ότι

$$\mathbb{E}(X_1) = \lim_{n \rightarrow \infty} \mathbb{E}(X_{T \wedge n}) = \mathbb{E}(X_T)$$

Θα δούμε αναλυτικά μία παρόμοια απόδειξη για την δεύτερη εξίσωση του Wald.

1.4 Εξισώσεις του Wald

Στη συνέχεια θα κάνουμε ευρεία χρήση σύνθετων αθροισμάτων τ.μ., στα οποία το πλήθος των προσθετέων είναι ένας χρόνος στάσης. Ένα τέτοιο παράδειγμα είναι η τ.μ. που δηλώνει το κέρδος ενός παίκτη όταν αποχωρήσει από ένα παιχνίδι, στο οποίο παίζει μέχρι να νικήσει ή μέχρι να χάσει όλο το ποσό με το οποίο ξεκίνησε (*πρόβλημα χρεωκοπίας του παίκτη*). Από τα βασικότερα αποτελέσματα είναι η θεμελιώδης εξίσωση της ακολουθιακής ανάλυσης και οι δύο εξισώσεις του Wald.

Θεώρημα 1.4.1. *Εστω i.i.d. τ.μ. $(X_n)_{n \geq 1}$ με $\mathbb{E}|X_i| < \infty$, μέση τιμή μ και διασπορά σ . Εστω ακόμα ότι ο T είναι χ.σ. τέτοιος ώστε $\mathbb{E}(T) < \infty$. Τότε για την σύνθετη τ.μ. $S_T = \sum_{i=1}^T X_i$ θα ισχύουν οι σχέσεις:*

$$(i) \mathbb{E}(S_T) = \mu \cdot \mathbb{E}(T)$$

$$(ii) \text{var}(S_T - T\mu) = \sigma^2 \mathbb{E}(T), \text{ εάν επιπρόσθετα } \text{var}(X_n) < \infty.$$

Απόδειξη. (i) Θέτουμε $Y_n = S_n - n\mu$ και $(\mathcal{F}_n)_{n \geq 1}$ ως είναι η συνήθης διήθηση των τ.μ. X_1, X_2, \dots, X_n . Θα δειχθεί ότι η ακολουθία $(Y_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale. Ισχύει

$$Y_{n+1} = S_{n+1} - (n+1)\mu = S_n + X_{n+1} - (n+1)\mu$$

και επομένως

$$\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = \mathbb{E}(S_n | \mathcal{F}_n) + \mathbb{E}(X_{n+1} | \mathcal{F}_n) - (n+1)\mathbb{E}(\mu \cdot I_\Omega | \mathcal{F}_n) \quad (1.1)$$

Παρατηρούμε ότι: η S_n είναι \mathcal{F}_n -μετρήσιμη και από την ιδιότητα (ii) του Θεωρήματος 1.1.1 παίρνουμε ότι

$$\mathbb{E}(S_n | \mathcal{F}_n) = S_n \quad (1.2)$$

Η X_{n+1} είναι ανεξάρτητη από τις τ.μ. X_1, X_2, \dots, X_n επομένως ανεξάρτητη του \mathcal{F}_n , άρα από την (iv) του Θεωρήματος 1.1.1 .

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \mathbb{E}(X_{n+1}) = \mu \quad (1.3)$$

Τέλος από την ιδιότητα (iii) του Θεωρήματος 1.1.1

$$\mathbb{E}(\mu \cdot I_\Omega | \mathcal{F}_n) = \mu \cdot \mathbb{E}(I_\Omega | \mathcal{F}_n) = \mu \quad (1.4)$$

Από τις (1.1)-(1.4) καταλήγουμε ότι $\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = Y_n$, δηλαδή η ακολουθία $(Y_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale. Ακόμα ισχύει ότι:

$$\mathbb{E}(|Y_{n+1} - Y_n| | \mathcal{F}_n) = \mathbb{E}(|X_{n+1} - \mu| | \mathcal{F}_n) \leq 2\mathbb{E}|X_1| < \infty$$

Άρα εφαρμόζοντας το Θεώρημα 1.2.2:

$$\mathbb{E}(Y_T) = \mathbb{E}(Y_1) = \mathbb{E}(S_1 - \mu) = 0$$

από όπου προκύπτει τελικά ότι

$$\mathbb{E}(S_T) = \mu \cdot \mathbb{E}(T).$$

(ii) Χωρίς βλάβη της γενικότητας υποθέτουμε ότι $\mathbb{E}(X_i) = 0$. Θέτουμε $Z_n = S_n^2 - n\sigma^2$ και $(\mathcal{F}_n)_{n \geq 1}$ όπως το (i). Θα δείξουμε πρώτα ότι η $(Z_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale. Οι τ.μ. Z_n είναι \mathcal{F}_n -μετρήσιμες, $\mathbb{E}|Z_n| < \infty$ και επιπλέον ισχύει:

$$\begin{aligned} Z_{n+1} &= S_{n+1}^2 - (n+1)\sigma^2 = (S_n + X_{n+1})^2 - (n+1)\sigma^2 \\ &= S_n^2 + X_{n+1}^2 + 2X_{n+1}S_n - (n+1)\sigma^2 \end{aligned}$$

Παρατηρούμε ότι $\mathbb{E}(S_n^2 | \mathcal{F}_n) = S_n^2$ γιατί η S_n^2 είναι \mathcal{F}_n -μετρήσιμη και λόγω ανεξαρτησίας

$$\mathbb{E}(X_{n+1}^2 | \mathcal{F}_n) = \mathbb{E}(X_{n+1}^2) = \sigma^2.$$

Επίσης,

$$\mathbb{E}(X_{n+1}S_n | \mathcal{F}_n) = \mathbb{E}(X_{n+1} | \mathcal{F}_n)S_n = \mathbb{E}(X_{n+1})S_n = 0$$

Δηλαδή $\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = S_n^2 + \sigma^2 - (n+1)\sigma^2 = Z_n$. Η $(Z_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale και άρα και η $(Z_{T \wedge n}, \mathcal{F}_{T \wedge n})_{n \geq 1}$ θα είναι martingale σύμφωνα με το Παράδειγμα 1.2.2. Από το Παράδειγμα 1.1.1

$$\mathbb{E}(Z_{T \wedge n}) = \mathbb{E}(Z_1) = 0$$

και άρα

$$\mathbb{E}(S_{T \wedge n}^2) = \sigma^2 \mathbb{E}(T \wedge n) \quad (1.5)$$

Για την τ.μ. $T \wedge n$ ισχύει ότι είναι αύξουσα ακολουθία τ.μ. και $T \wedge n \xrightarrow{n \rightarrow \infty} T$. Εφαρμόζοντας το θεώρημα μονότονης σύγκλισης παίρνουμε ότι

$$\lim_{n \rightarrow \infty} \mathbb{E}(T \wedge n) = \mathbb{E}(T)$$

Για να περάσουμε το όριο καθώς το $n \rightarrow \infty$ στο αριστερό μέρος της ισότητας (1.5) θα χρησιμοποιήσουμε το θεώρημα κυριαρχημένης σύγκλισης.

$$\begin{aligned}
S_{T \wedge n}^2 &\leq S_T^2 = \left(\sum_{i=1}^T X_i \right)^2 = \sum_{i=1}^T X_i^2 + 2 \sum_{j<i}^T X_i X_j \\
&= \sum_{i=1}^{\infty} X_i^2 I_{\{T \geq i\}} + 2 \sum_{j<i}^{\infty} X_i X_j I_{\{T \geq i\}} I_{\{T \geq j\}}
\end{aligned} \tag{1.6}$$

Παρατηρούμε ότι η τ.μ. $I_{\{T \geq i\}}$ όπως και η $X_j I_{\{T \geq i\}} I_{\{T \geq j\}}$ είναι \mathcal{F}_{i-1} -μετρήσιμες. Λόγω ανεξαρτησίας της τ.μ. X_i με την σ -άλγεβρα \mathcal{F}_{i-1} προκύπτει ότι

$$\mathbb{E}(S_T^2) = \sigma^2 \sum_{i=1}^{\infty} \mathbb{P}(T \geq i) = \sigma^2 \mathbb{E}(T) < \infty \tag{1.7}$$

Άρα από τις (1.6) και (1.7) και το θεώρημα κυριαρχημένης σύγκλισης προκύπτει το ζητούμενο. \square

1.5 Η Θεμελιώδης εξίσωση

Το επόμενο θεώρημα συνδέει τον χρόνο στάσης T , με το σύνθετο άθροισμα S_T και την ροπογεννήτρια των τ.μ. X_i και αναφέρεται στην βιβλιογραφία ως *θεμελιώδης εξίσωση της ακολουθιακής ανάλυσης* ή και ως *θεμελιώδης ταυτότητα του Wald*.

Θεώρημα 1.5.1. Έστω *i.i.d* τ.μ. $(X_n)_{n \geq 1}$ με ροπογεννήτρια $M(t) = \mathbb{E}(e^{tX})$. Εάν ικανοποιούνται οι προϋποθέσεις:

- (i) $\exists t_0 \neq 0$ τέτοιο ώστε $1 \leq M(t_0) < \infty$
- (ii) $\exists c \in \mathbb{R}_+$ και χ.σ. T ώστε $\mathbb{E}(T) < \infty$ και $|S_n| \leq c, \forall n < T$

τότε

$$\mathbb{E}(e^{t_0 S_T} M(t_0)^{-T}) = 1.$$

Απόδειξη. Θέτουμε $Y_n = \frac{e^{t_0 S_n}}{M(t_0)^n}$ και $(\mathcal{F}_n)_{n \geq 1}$ ως είναι η συνήθης διήθηση. Θα δειχθεί ότι η ακολουθία $(Y_n, \mathcal{F}_n)_{n \geq 1}$ είναι martingale. Πράγματι

$$\begin{aligned}
\mathbb{E}(Y_{n+1} | \mathcal{F}_n) &= \mathbb{E} \left(\frac{e^{t_0 S_{n+1}}}{M(t_0)^{n+1}} | \mathcal{F}_n \right) = \mathbb{E} \left(\frac{e^{t_0 S_n}}{M(t_0)^n} \cdot \frac{e^{t_0 X_{n+1}}}{M(t_0)} | \mathcal{F}_n \right) \\
&= Y_n \cdot \mathbb{E} \left(\frac{e^{t_0 X_{n+1}}}{M(t_0)} | \mathcal{F}_n \right) = Y_n M(t_0)^{-1} \mathbb{E}(e^{t_0 X_{n+1}}) = Y_n
\end{aligned}$$

κάνοντας χρήση των ιδιοτήτων (ii) και (v) του Θεωρήματος 1.1.1. Ισχύει

$$\mathbb{E}(Y_{n+1} - Y_n | \mathcal{F}_n) = Y_n \cdot \mathbb{E} \left(\left| \frac{e^{t_0 X_{n+1}}}{M(t_0)} - 1 \right| \right) \leq \frac{Y_n}{M(t_0)} \mathbb{E}(e^{t_0 X_{n+1}} + M(t_0)) = 2Y_n$$

Από την υπόθεση (ii) $\forall n < T$: $Y_n = \frac{e^{t_0 S_n}}{M(t_0)^n} \leq \frac{e^{t_0 |S_n|}}{M(t_0)^n} \leq e^{t_0 c}$. Επομένως εφαρμόζοντας το Θεώρημα 1.3.2 παίρνουμε ότι $\mathbb{E}(Y_T) = \mathbb{E}(Y_1) \Rightarrow \mathbb{E}(e^{t_0 S_T} M(t_0)^{-T}) = 1$. \square

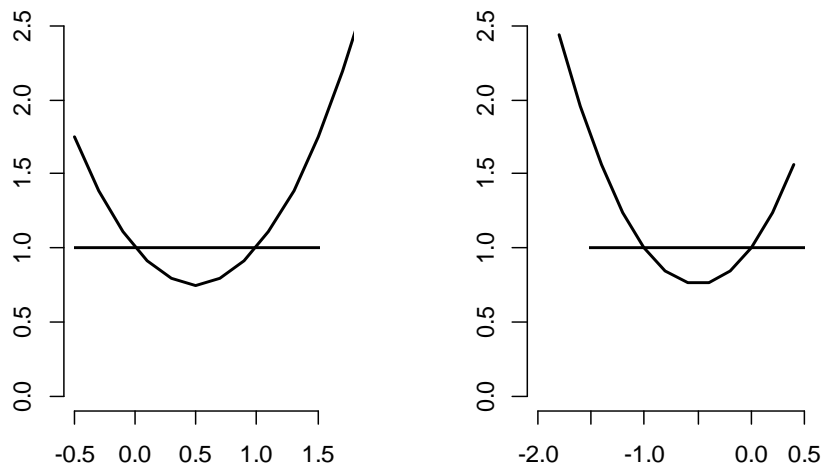
Για να απλουστεύσουμε την χρήση του παραπάνω θεωρήματος αρκεί να βρούμε κατάλληλες συνθήκες για την ροπογεννήτρια $M(t)$ έτσι να υπάρχει ένα ακριβώς $t_0 \neq 0$ με $M(t_0) = 1$. Παραγωγίζοντας δύο φορές την ροπογεννήτρια ως προς t , παίρνουμε

$$M''(t) = \mathbb{E}(X^2 e^{tX}) \quad (1.8)$$

Αν $\mathbb{P}(X > 0) > 0$, τότε κατ'ανάγκη $\exists \varepsilon > 0$ και $\delta \in (0,1)$ τέτοια ώστε $\mathbb{P}(X > \varepsilon) = \delta$. Άρα από την (1.8) $M''(t) > 0$, δηλαδή η καμπύλη γ της ροπογεννήτριας είναι κυρτή και άρα η $y = 1$ μπορεί να την τέμνει το πολύ σε δύο σημεία. Θα θέλαμε η καμπύλη γ να μην εφάπτεται της $y = 1$ στο σημείο με τετμημένη $t = 0$. Άρα ζητάμε $M'(0) = \mu \neq 0$. Διακρίνουμε δύο περιπτώσεις ανάλογα με το πρόσημο της μέσης τιμής της μ . Εάν $\mu < 0$ ($\mu > 0$) ζητάμε $\lim_{t \rightarrow +\infty} M(t) = +\infty$ ($\lim_{t \rightarrow -\infty} M(t) = +\infty$) ώστε η $y = 1$ να τέμνει την γ σε ένα τουλάχιστον σημείο διάφορο του $t = 0$. Για να πετύχουμε αυτό το όριο θέλουμε η ροπογεννήτρια να φράσσεται από μία συνάρτηση του t που απειρίζεται καθώς το $t \rightarrow +\infty$. Αν $\mathbb{P}(X > 0) > 0$ τότε

$$\forall t > 0: M(t) = \mathbb{E}(e^{tX}) \geq e^{t\varepsilon} \delta$$

και αφήνοντας το t να τείνει στο άπειρο παίρνουμε το ζητούμενο όριο για την γ . Με όμοιο τρόπο αποδεικνύεται και το αντίστοιχο όριο όταν $\mu > 0$. Καταλήξαμε στο παρακάτω αποτέλεσμα.



Σχήμα 1.5.1 Γραφική παράσταση της ροπογεννήτριας όταν η τ.μ. έχει αρνητική και θετική αντίστοιχα μέση τιμή

Πρόταση 1.5.1. Εάν για μία τ.μ. X ισχύει ότι (i) $\mathbb{P}(X > 0) > 0$, (ii) $\mathbb{P}(X < 0) > 0$, (iii) $\mathbb{E}(X) \neq 0$ και (iv) $M(t) < \infty$, τότε $\exists! t_0 \in \mathbb{R} \setminus \{0\} : M(t_0) = 1$.

Η πρόταση αυτή σε συνδυασμό με την θεμελιώδη εξίσωση της ακολουθιακής ανάλυσης είναι γενικά χρήσιμη για τον υπολογισμό πιθανοτήτων της μορφής $\mathbb{P}(S_T > k)$ όπου ο T είναι χρόνος απορρόφησης και k σταθερά. Ακόμα εφαρμογές υπάρχουν και στην προσέγγιση των καμπύλων OC και ASN, στην προσέγγιση της συνάρτησης ARL του αλγορίθμου CUSUM και σε προβλήματα που εμφανίζονται στην ανανεωτική θεωρία (π.χ. βλ. Ghosh 1997, σελ. 50-55).

1.6 Θεωρία βέλτιστης στάσης

Είδαμε μέχρι τώρα πως συνδέεται η αναμενόμενη τιμή ενός χρόνου στάσης με την αντίστοιχη ενός σύνθετου αθροίσματος τ.μ. Σε πολλά προβλήματα στατιστικής δίνεται ο χ.σ. και ερευνώνται οι ιδιότητες του με χρήση των εξισώσεων του Wald. Ωστόσο, μπορεί να τεθεί και το αντίστροφο πρόβλημα: να παρουσιαστούν κάποιες ιδιότητες-συνθήκες που αφορούν τον χ.σ. και να ζητηθεί να βρεθεί η μορφή του.

Έστω η ακολουθία ανεξάρτητων τ.μ. $(X_n)_{n \geq 1}$, $(\mathcal{F}_n)_{n \geq 1}$ ας είναι η συνήθης διήθηση και T ένας χ.σ. ως προς τη διήθηση αυτή τέτοιος ώστε

$$\mathbb{P}(1 \leq T < \tau) = 1 \tag{1.9}$$

όπου $\tau \in [0 + \infty)$. Συμβολίζουμε με Δ_τ το σύνολο όλων των χ.σ. ως προς τη συνήθη διήθηση που ικανοποιούν την (1.9). Εάν $(Y_n)_{n \geq 1}$ είναι μία στοχαστική ανέλιξη στο χώρο $(\Omega, \mathcal{F}, \mathbb{P})$ τότε το πρόβλημα της θ.β.σ. είναι να υπολογιστεί η ποσότητα

$$V = \sup_{T \in \Delta_\tau} \mathbb{E}(Y_T) \quad (1.10)$$

και να βρεθεί ο τύπος του χ.σ. T που την μεγιστοποιεί. Προφανώς θα μπορούσαμε να διατυπώσουμε το ίδιο πρόβλημα και για το κάτω πέρας στην σχέση (1.10). Ο Shiryaev (2013) αναφέρει την μεθοδολογία εύρεσης λύσης της σχέσης (1.10) με χρήση της θεωρίας martingales. Θα δούμε στο Κεφάλαιο 4 κάποια προβλήματα της θ.β.σ. που αφορούν στατιστικούς αλγορίθμους.

ΚΕΦΑΛΑΙΟ 2

Ακολουθιακή Εκτιμητική κατά Διάστημα

2.1 Εύρεση βέλτιστου μεγέθους δείγματος

Σε αυτήν την παράγραφο θα αποδείξουμε ότι το πρόβλημα εύρεσης $100(1-a)\%$ διαστήματος εμπιστοσύνης, για την μέση τιμή κανονικής κατανομής, που να έχει εύρος ίσο με ένα δοσμένο εκ των προτέρων εύρος, δεν επιλύεται με κάποια μεθοδολογία σταθερού δείγματος.

Συμβολίζουμε με $\mathbf{X} = (X_1, \dots, X_n)$ ένα τυχαίο δείγμα μεγέθους n , με $\mathbf{x} = (x_1, \dots, x_n)$ μία πραγματοποίηση του και με $\delta(\mathbf{X})$ έναν εκτιμητή της μέσης τιμής μ της κατανομής του δείγματος. Τότε η $0-1$ συνάρτηση απώλειας (βλ. Mukhopadhyay, 2009, σελ. 16) για το συγκεκριμένο πρόβλημα θα γράφεται

$$W(\delta(\mathbf{X}), \mu) = \begin{cases} 0, & |\delta(\mathbf{X}) - \mu| \leq d \\ 1, & |\delta(\mathbf{X}) - \mu| > d \end{cases} \quad (2.1)$$

όπου $d > 0$ είναι το δοσμένο εύρος του δ.ε. Θα δείξουμε ότι

$$\exists \delta(\mathbf{X}): \mathbb{P}(|\delta(\mathbf{X}) - \mu| \leq d) \geq 1 - a \quad (2.2)$$

Το επόμενο θεώρημα μπορεί να βρεθεί στον Lehman (1951).

Θεώρημα 2.1.1 *Έστω οι i.i.d. τ.μ.*

$$X_1, \dots, X_n \sim \sigma^{-1} f\left(\frac{x - \theta}{\sigma}\right),$$

όπου $\theta \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$, είναι άγνωστες παράμετροι. Έστω ότι η συνάρτηση απώλειας δίνεται από την σχέση $W(\delta(\mathbf{X}), \theta) = H(|\delta(\mathbf{X}) - \theta|)$, όπου η $H(|y|)$ είναι αύξουσα συνάρτηση του $|y|$. Εάν $M = \sup\{H(|y|): y \in \mathbb{R}\}$, τότε $\forall L < M$ δεν υπάρχει εκτιμητής $\delta(\mathbf{X})$ τέτοιος ώστε

$$\sup\{\mathbb{E}[W(\delta(\mathbf{X}), \theta)]: (\theta, \sigma) \in (\mathbb{R}, \mathbb{R}_+)\} \leq L.$$

Παρατηρούμε ότι για την συνάρτηση απώλειας (2.1) ισχύει ότι

$$\mathbb{E}[W(\delta(\mathbf{X}), \theta)] = \mathbb{P}(|\delta(\mathbf{X}) - \mu| > d)$$

Για $M = 1, L = a$, από το Θεώρημα 2.1.1 παίρνουμε ότι

$$\exists \delta(\mathbf{X}): \mathbb{P}(|\delta(\mathbf{X}) - \mu| > d) \leq a$$

και καταλήγουμε στην (2.2).

Έχοντας, με βάση το παραπάνω αποτέλεσμα, αποδεχθεί ότι το μέγεθος n του δείγματος θα πρέπει να είναι μεταβαλλόμενο, θα θέλαμε να προσδιορίσουμε εκείνο το n και το διάστημα $I_n = [\bar{X}_n - d, \bar{X}_n + d]$ για τα οποία ισχύει

$$\mathbb{P}(\mu \in I_n) \geq 1 - a \quad (2.3)$$

Θα εξετάσουμε την περίπτωση που έχουμε δείγμα από την κανονική κατανομή. Η γενική θεωρία για άλλες κατανομές μπορεί να βρεθεί στον Ghosh (1997).

Έστω $X_1, \dots, X_n \sim N(\mu, \sigma^2), 0 < \sigma < \infty$. Υποθέτουμε αρχικά ότι η διασπορά είναι γνωστή. Με πράξεις παίρνουμε ότι

$$\mathbb{P}(\mu \in I_n) = \mathbb{P}(|\bar{X}_n - \mu| \leq d) = 2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right) - 1 \quad (2.4)$$

Για λύσουμε την ανίσωση (2.3) με χρήση της (2.4), αρκεί να εκφράσουμε το a με τη βοήθεια της συνάρτησης Φ . Δηλαδή αρκεί να βρεθεί x τέτοιο ώστε

$$1 - a = 2\Phi(x) - 1$$

και επομένως,

$$x = \Phi^{-1}\left(1 - \frac{a}{2}\right) = z_{a/2} \quad (2.5)$$

άρα από τις (2.3), (2.4) και (2.5):

$$2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right) - 1 \geq 2\Phi(z_{a/2}) - 1$$

από όπου προκύπτει ότι

$$n \geq \frac{\sigma^2}{d^2} z_{a/2}^2 = k \quad (2.6)$$

Σύμφωνα με όλα τα παραπάνω καταλήξαμε στην σχέση

$$\mathbb{P}(\mu \in I_n) \geq 1 - a \Leftrightarrow n \geq k \quad (2.7)$$

Επειδή για τον υπολογισμό του k η διασπορά σ^2 ήταν γνωστή, το k θα είναι και το βέλτιστο μέγεθος του δείγματος για την κατασκευή του διαστήματος εμπιστοσύνης.

2.2 Μέθοδος του Stein

Το k στη σχέση (2.7) μπορεί πρακτικά να εκτιμηθεί κάνοντας χρήση της δειγματικής διασποράς s^2 όταν η διασπορά σ^2 είναι άγνωστη. Το $a/2$ άνω ποσοστιαίο σημείο της κανονικής κατανομής θα αντικατασταθεί με το αντίστοιχο $t_{n_0-1; a/2}^2 \equiv t_{n_0-1}^2$ της κατανομής Student, με βαθμούς ελευθερίας μειωμένους κατά 1, και θα πάρουμε έναν εκτιμητή \hat{k} του k . Εκτιμώντας την διασπορά με χρήση ενός δείγματος μεγέθους n_0 , μπορεί αυτό το n_0 να ικανοποιεί την σχέση (2.7). Εάν δεν την ικανοποιεί, τότε επιλέγουμε $\hat{k} - n_0$ επιπλέον παρατηρήσεις και το τελικό δ.ε. θα είναι $[\bar{X}_{\hat{k}} - d, \bar{X}_{\hat{k}} + d]$. Ο Stein (1949), πρότεινε τον εξής εκτιμητή του απαιτούμενου μεγέθους δείγματος :

$$T = \max \left\{ n_0, \left\lfloor \frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \right\rfloor + 1 \right\} \quad (2.8)$$

όπου με $\lfloor x \rfloor$ συμβολίζουμε τον μεγαλύτερο ακέραιο που είναι μικρότερος ή ίσος του x . Παρατηρούμε ότι η τ.μ. T εξαρτάται από την διασπορά των μεταβλητών. Άρα αν $(\mathcal{F}_n)_{n \geq 1}$ είναι η συνήθης διήθηση, τότε η τ.μ. T είναι χ.σ. Προφανώς $\mathbb{P}(T < \infty)$.

Για να αποδείξουμε ότι η μέθοδος του Stein (1949) επιλύει το πρόβλημα που αναφέρθηκε στην Παράγραφο (2.1), θα χρειαστούμε το παρακάτω γνωστό αποτέλεσμα.

Θεώρημα 2.2.1. Έστω X_1, \dots, X_n τυχαίο δείγμα από μία κανονική κατανομή. Οι τ.μ. $\bar{X}_n = \sum_{i=1}^n X_i$ και $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ είναι στοχαστικά ανεξάρτητες.

Προκύπτει ακόμα ότι και οι τ.μ. \bar{X}_n και $I(T = n)$ είναι επίσης ανεξάρτητες $\forall n \geq n_0$ (βλ. Mukhopadhyay, 2009, σελ. 101). Άρα:

$$\begin{aligned} \mathbb{P}(\mu \in I_T) &= \sum_{n=n_0}^{\infty} \mathbb{P}(\{|\bar{X}_n - \mu| \leq d\} \cap \{T = n\}) \\ &= \sum_{n=n_0}^{\infty} \mathbb{P}(|\bar{X}_n - \mu| \leq d) \mathbb{P}(T = n) \\ &= \mathbb{E} \left[2\Phi \left(\frac{\sqrt{T}d}{\sigma} \right) - 1 \right] \end{aligned} \quad (2.9)$$

Σημειώνουμε ότι όταν $d \rightarrow 0$, δηλαδή όταν το εύρος του δ.ε. που ζητάμε μειώνεται, θα χρειαζόμαστε περισσότερες παρατηρήσεις για την κατασκευή του άρα θα αναφερόμαστε σε ασυμπτωτική ιδιότητα. Στο επόμενο θεώρημα δίνουμε τις ιδιότητες της μεθόδου, υπό την έννοια των Chow και Robbins (1965).

Θεώρημα 2.2.2. Για τον χρόνο στάσης T και για κάθε μ και σ^2 , ισχύουν τα επόμενα:

- (i) $\lim_{d \rightarrow 0} \mathbb{E} \left(\frac{T}{k} \right) > 1$ (ασυμπτωτική επάρκεια)
- (ii) $\mathbb{P}(\mu \in I_T) \geq 1 - a$ (ακριβής συνέπεια)
- (iii) $\lim_{d \rightarrow 0} \mathbb{P}(\mu \in I_T) = 1 - a$ (ασυμπτωτική συνέπεια)

Απόδειξη: (i). Από τον ορισμό (2.8) της τ.μ. T ισχύει

$$\frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \leq T \leq n_0 + \frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \quad (2.10)$$

και παίρνοντας την αναμενόμενη τιμή κατά μέρη προκύπτει ότι

$$\frac{t_{n_0-1}^2 \sigma^2}{d^2} \leq \mathbb{E}(T) \leq n_0 + \frac{t_{n_0-1}^2 \sigma^2}{d^2}$$

Διαιρώντας με k την παραπάνω σχέση, όπου το k ορίστηκε στην (2.6), και αφήνοντας το d να τείνει στο μηδέν

$$\lim_{d \rightarrow 0} \mathbb{E} \left(\frac{T}{k} \right) = \frac{t_{n_0-1}^2}{z_{a/2}^2} \quad (2.11)$$

Με χρήση της επέκτασης των Cornish-Fisher (Johnson, 1970) για τα ποσοστιαία σημεία, μπορούμε να εκφράσουμε το άνω $a/2$ ποσοστιαίο σημείο της κατανομής του Student συναρτήσει του αντίστοιχου της τυπικής κανονικής κατανομής ως εξής:

$$t_{n_0-1} = z_{a/2} + \frac{z_{a/2}(z_{a/2}^2 + 1)}{4(n_0 - 1)} + O(n_0^{-2})$$

Άρα $\frac{t_{n_0-1}}{z_{a/2}} > 1$ και με χρήση της (2.11) αποδεικνύεται η (i).

(ii). Σύμφωνα με την ανισότητα (2.10) και την σχέση (2.9) παίρνουμε

$$\begin{aligned} \mathbb{P}(\mu \in I_T) &\geq \mathbb{E} \left[2\Phi \left(\frac{t_{n_0-1} S_{n_0}}{\sigma} \right) - 1 \right] \\ &= \mathbb{E} \left[\mathbb{P}(|Z| \leq t_{n_0-1} S_{n_0} \sigma^{-1} | S_{n_0}) \right], \quad Z \sim N(0,1) \\ &= \mathbb{E} \left[\mathbb{E}(I(|Z| \leq t_{n_0-1} S_{n_0} \sigma^{-1}) | S_{n_0}) \right], \\ &= \mathbb{E} [I(|Z| \leq t_{n_0-1} S_{n_0} \sigma^{-1})] = \mathbb{P}(|Z| \leq t_{n_0-1} S_{n_0} \sigma^{-1}) \\ &= \mathbb{P}(Y \leq t_{n_0-1}), \quad Y \sim t(n_0 - 1) \end{aligned}$$

$$= 1 - a \quad (2.12)$$

(iii). Ισχύει ότι $\sqrt{T}d \xrightarrow{d \rightarrow 0} t_{n_0-1} S_{n_0}$, $\sqrt{T}d \leq 2t_{n_0-1} S_{n_0-1} = W$ για d αρκούντως μικρό, σύμφωνα με την ανισότητα (2.10), και $\mathbb{E}[|W|] < \infty$. Άρα από το θεώρημα κυριαρχημένης σύγκλισης προκύπτει ότι

$$\mathbb{E} \left(2\Phi \left(\frac{\sqrt{T}d}{\sigma} \right) - 1 \right) \xrightarrow{d \rightarrow 0} \mathbb{E} \left(2\Phi \left(\frac{t_{n_0-1} S_{n_0}}{\sigma} \right) - 1 \right) = 1 - a.$$

όπου η τελευταία ισότητα προέκυψε από την (2.12). \square

2.3 Κατανομή του μεγέθους του δείγματος

Για την κατανομή του μεγέθους του δείγματος μπορούμε να πάρουμε αναλυτικά αποτελέσματα. Ειδικότερα, υπολογίζουμε τις πιθανότητες

$$\begin{aligned} \mathbb{P}(T = n_0) &= \mathbb{P} \left(\left\lfloor \frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \right\rfloor + 1 \leq n_0 \right) = \mathbb{P} \left(0 \leq \frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \leq n_0 \right) \\ &= \mathbb{P} \left(0 < Y \leq \frac{n_0(n_0 - 1)d^2}{\sigma^2 t_{n_0-1}^2} \right) \end{aligned}$$

όπου $Y \sim \chi_{n_0-1}^2$. Όμοια $\forall m \in \mathbb{N}$:

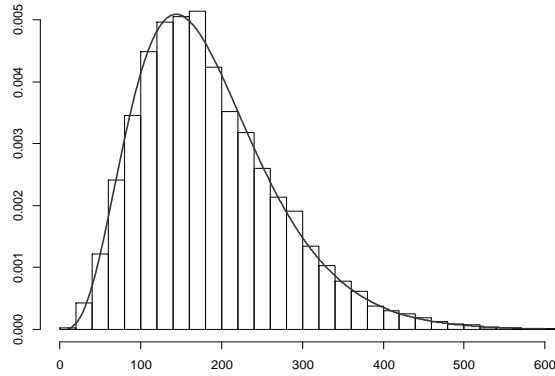
$$\begin{aligned} \mathbb{P}(T = n_0 + m) &= \mathbb{P} \left(\left\lfloor \frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \right\rfloor = n_0 + m - 1 \right) \\ &= \mathbb{P} \left(n_0 + m - 1 < \frac{t_{n_0-1}^2 S_{n_0}^2}{d^2} \leq n_0 + m \right) \\ &= \mathbb{P} \left(\frac{(n_0 + m - 1)(n_0 - 1)d^2}{\sigma^2 t_{n_0-1}^2} < Y \leq \frac{(n_0 + m)(n_0 - 1)d^2}{\sigma^2 t_{n_0-1}^2} \right), \end{aligned}$$

όπου $Y \sim \chi_{n_0-1}^2$. Εάν θέσουμε

$$g_{n_0}(y) = \left\{ 2^{(n_0-1)/2} \Gamma \left(\frac{n_0-1}{2} \right) \right\}^{-1} \exp(-y/2) y^{(n_0-3)/2}, \quad y > 0$$

και $c_m = \frac{(n_0+m)(n_0-1)d^2}{\sigma^2 t_{n_0-1}^2}$, $k = 0, 1, 2, \dots$, τότε οι παραπάνω πιθανότητες γράφονται

$$\mathbb{P}(T = n_0) = \int_0^{c_0} g_{n_0}(y) dy, \quad \text{και} \quad \mathbb{P}(T = n_0 + m) = \int_{c_{m-1}}^{c_m} g_{n_0}(y) dy.$$



Σχήμα 2.3.1. Ιστόγραμμα για τον χρόνο στάσης T

Παρατηρούμε ότι οι θεωρητικές τιμές συμπίπτουν με τις αντίστοιχες προσομοιωμένες τιμές της τ.μ. T . Για τον κώδικα που αφορά την κατασκευή του παραπάνω ιστογράμματος (Σχ. 2.3.1) σχετικών συχνοτήτων βλ. Παράρτημα III. Συγκεκριμένα εφαρμόστηκαν τα εξής: Δημιουργούμε έναν βρόχο $n = 10^4$ επαναλήψεων όπου σε κάθε επανάληψη παράγουμε $n_0 = 10$ τυχαίους αριθμούς από την $N(1, 3^2)$, υπολογίζουμε την τ.μ. T όταν $\alpha = 0.05$ και την αποθηκεύουμε σε έναν πίνακα T_s . Για τον υπολογισμό των πιθανοτήτων $\mathbb{P}(T = j)$ διακρίνουμε δύο περιπτώσεις: Εάν $j = n_0$ και εάν $j = n_0 + k$, όπου το k παίρνει τιμές από 1 ως 590. Κάνοντας χρήση των παραπάνω τύπων, αποθηκεύουμε τις τιμές των πιθανοτήτων σε έναν πίνακα $Prob$ και παίρνουμε τις θεωρητικές τιμές.

2.4 Πλήρως ακολουθιακή μέθοδος

Ο Anscombe (1952) είχε ασχοληθεί πρώτος με την κατασκευή δ.ε. με γνωστό εκ των προτέρων εύρος, εφαρμόζοντας τεχνικές που έχουν πλήρη ακολουθιακό χαρακτήρα. Ο χ.σ. που θα εξετάζεται τώρα είναι ο

$$T = \min \left\{ n \geq n_0 : n \geq \frac{z_{\alpha/2}^2 S_n^2}{a^2} \right\} \quad (2.13)$$

όπου $n_0 \geq 2$ είναι το αρχικό μέγεθος του δείγματος που ξεκινάει η διαδικασία. Το τελικό δ.ε. θα είναι το $I = [\bar{X}_T - d, \bar{X}_T + d]$. Για την ακολουθιακή μέθοδο ισχύουν τα παρακάτω αποτελέσματα που μπορούν να βρεθούν στους Woodroffe (1977) και Chow και Robbins (1965).

Θεώρημα 2.4.1 Για τον χρόνο στάσης T ισχύουν τα επόμενα:

(i) $\mathbb{E}(T) - k = g(1) + o(1)$, όταν $n_0 \geq 4$

(ii) $\mathbb{P}(\mu \in I_T) = 1 - \alpha + \frac{z_{\alpha/2}^2}{2k} (2g(1) - 1 - z_{\alpha/2}^2) f(z_{\alpha/2}^2; 1) + o(C^{-1})$

$$(iii) \lim_{d \rightarrow 0} \mathbb{E} \left(\frac{T}{k} \right) = 1$$

$$(iv) \lim_{d \rightarrow 0} \mathbb{P}(\mu \in I_T) = 1 - a$$

όπου $g(x) = \frac{1}{2} - \frac{1}{x} - \frac{1}{x} \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}(\max\{0, \chi_{nx}^2 - 2nx\})$, χ_{nx}^2 είναι μία τ.μ. που ακολουθεί χ^2 κατανομή με nx βαθμούς ελευθερίας και f είναι η σ.π. της χ_1^2 κατανομής.

Επειδή $g(1) < 0$ με χρήση του (i) και (ii) προκύπτει ότι $\mathbb{E}(T) < k$ και $\mathbb{P}(\mu \in I_T) < 1 - a$. Άρα η ακολουθιακή μέθοδος δεν ικανοποιεί την ιδιότητα της ακριβούς συνέπειας. Μπορούμε με χρήση της εξίσωσης του Wald να βρούμε αναλυτικά, ένα μεγαλύτερο φράγμα για τον χρόνο T . Από τον ορισμό:

$$T - 1 < n_0 - 1 + \frac{z_{a/2}^2 S_{T-1}^2}{d^2}$$

και επομένως,

$$(T - n_0)(T - 2) < \frac{z_{a/2}^2}{d^2} \sum_{i=1}^{T-1} (X_i - \bar{X}_{T-1})^2 < \frac{z_{a/2}^2}{d^2} \sum_{i=1}^T (X_i - \mu)^2 = \frac{z_{a/2}^2}{d^2} \sum_{i=1}^T Y_i^2.$$

Άρα,

$$T^2 - (n_0 + 2)T < (T - n_0)(T - 2) < \frac{z_{a/2}^2}{d^2} \sum_{i=1}^T Y_i^2$$

και

$$\mathbb{E}(T^2) - (n_0 + 2)\mathbb{E}(T) < \frac{z_{a/2}^2}{d^2} \sigma^2 \mathbb{E}(T) = k\mathbb{E}(T).$$

Όμως επειδή $0 < \sigma^2 = \mathbb{E}(T^2) - (\mathbb{E}(T))^2$, καταλήγουμε στην σχέση

$$\mathbb{E}(T) < k + n_0 + 2, \quad k = \frac{z_{a/2}^2}{d^2} \sigma^2.$$

2.5 Επέκταση στις p -διαστάσεις

Θα επεκτείνουμε τις μεθόδους εύρεσης δ.ε. με γνωστό εκ των προτέρων μήκος, όταν έχουμε τυχαία διανύσματα.

Έστω $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \sim N_p(\boldsymbol{\mu}, \Sigma)$, όπου $\boldsymbol{\mu}$ είναι άγνωστο διάνυσμα και Σ είναι ο συμμετρικός και θετικά ορισμένος πίνακας διασπορών-συνδιασπορών. Συμβολίζουμε επίσης με

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, S_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)^t,$$

με λ_i τις ιδιοτιμές του Σ και με $\lambda_{(p)}$ την μεγαλύτερη ιδιοτιμή του. Θα χρειαστούμε την επόμενη πρόταση για την απόδειξη βασικών αποτελεσμάτων.

Πρόταση 2.5.1. *Ισχύουν τα ακόλουθα (π.χ. βλ. Johnson 2007, σελ. 78, 163):*

(i) $\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^t \Sigma \mathbf{x}}{\mathbf{x}^t \mathbf{x}} = \lambda_{(p)}$

(ii) ο Σ είναι διαγωνιοποιήσιμος

(iii) $n(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \Sigma^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \sim \chi_p^2$

Θέλουμε να βρούμε το κατάλληλο μέγεθος του δείγματος ώστε η σφαιρική περιοχή με διάμετρο d που θα κατασκευάσουμε να έχει συντελεστή εμπιστοσύνης τουλάχιστον $1 - \alpha$, $\alpha \in (0,1)$. Έστω

$$\Pi_n = \{\mathbf{x} \in \mathbb{R}^p : (\bar{\mathbf{X}}_n - \mathbf{x})^t (\bar{\mathbf{X}}_n - \mathbf{x}) \leq d^2\}$$

Από την (ii): $\exists L, D \in \mathcal{M}_{p \times p}(\mathbb{R})$, όπου ο L είναι ορθογώνιος και $D = \text{diag}(\lambda_1, \dots, \lambda_p)$ τέτοιοι ώστε $\Sigma = L^t D L$. Θέτοντας $\mathbf{Y} = L(\bar{\mathbf{X}}_n - \boldsymbol{\mu})$ θα ισχύουν τα παρακάτω:

$$(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \Sigma^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) = \mathbf{Y}^t D^{-1} \mathbf{Y} \leq \frac{\mathbf{Y}^t \mathbf{Y}}{\lambda_{(p)}} = \frac{(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t (\bar{\mathbf{X}}_n - \boldsymbol{\mu})}{\lambda_{(p)}}$$

από όπου προκύπτει ότι

$$\begin{aligned} \mathbb{P}(\boldsymbol{\mu} \in \Pi_n) &\geq \mathbb{P}(\lambda_{(p)} (\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \Sigma^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \leq d^2) \\ &= \mathbb{P}\left(n(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \Sigma^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \leq \frac{nd^2}{\lambda_{(p)}}\right) = F\left(\frac{nd^2}{\lambda_{(p)}}\right) \end{aligned}$$

όπου F είναι η σ.κ. της χ_p^2 . Έστω a το άνω α -ποσοστιαίο σημείο της κατανομής χ_p^2 . Δηλαδή $F(a) = 1 - \alpha$. Καταλήξαμε ότι

$$\mathbb{P}(\boldsymbol{\mu} \in \Pi_n) \geq 1 - \alpha \Leftrightarrow n \geq \frac{a\lambda_{(p)}}{d^2} = C$$

Για την κατασκευή σφαιρικής περιοχής με συντελεστή εμπιστοσύνης $1 - \alpha$, θα πρέπει το μέγεθος του δείγματος να είναι τουλάχιστον $n = \left\lceil \frac{a\lambda_{(p)}}{d^2} \right\rceil + 1$.

2.5.1 Μέθοδος του Healy

Ο Healy (1956) πρότεινε μία ανάλογη μέθοδο με αυτήν του Stein (1949). Παίρνουμε αρχικό δείγμα $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$, όπου $m \geq p + 1$. Υπολογίζουμε τα $\bar{\mathbf{X}}_m$ και S_m καθώς και την μέγιστη ιδιοτιμή $\lambda_{(p)m}$ του πίνακα S_m . Ορίζουμε τον χρόνο στάσης

$$T = \max \left\{ m, \left\lceil \frac{u \lambda_{(p)}}{d^2} \right\rceil + 1 \right\}$$

όπου $u = T_\alpha^2 = \frac{p(m-1)}{m-p} F_{p, m-p, \alpha}$ είναι το άνω α -ποσοστιαίο σημείο της κατανομής $T_{p, m-p}^2$ του Hotelling και $F_{p, m-p, \alpha}$ το άνω α -ποσοστιαίο σημείο της κατανομής $F_{p, m-p}$. Ουσιαστικά εάν $T > m$ τότε συλλέγουμε $T - m$ επιπλέον τυχαία διανύσματα και βρίσκουμε την τελική $100(1 - \alpha)\%$ περιοχή εμπιστοσύνης

$$\Pi_T = \{ \mathbf{x} \in \mathbb{R}^p : (\bar{\mathbf{X}}_T - \mathbf{x})^t (\bar{\mathbf{X}}_T - \mathbf{x}) \leq d^2 \}.$$

Θεώρημα 2.5.1. Για κάθε $\boldsymbol{\mu}, \Sigma, p, m, d$ και α ισχύει ότι

- (i) $T(\bar{\mathbf{X}}_T - \boldsymbol{\mu})^t S_m^{-1}(\bar{\mathbf{X}}_T - \boldsymbol{\mu}) \sim T^2(p, m - p)$
- (ii) $\mathbb{P}(\boldsymbol{\mu} \in \Pi_T) \geq 1 - \alpha$ (ακριβής συνέπεια)

Απόδειξη. (i). Η απόδειξη βασίζεται στην ανεξαρτησία της τ.μ. $\bar{\mathbf{X}}_n$ με την τ.μ. $I(T = n)$ (π.χ. βλ. Mukhopadhyay, 2009).

(ii) Ισχύει ότι

$$\begin{aligned} \mathbb{P}((\bar{\mathbf{X}}_T - \boldsymbol{\mu})^t (\bar{\mathbf{X}}_T - \boldsymbol{\mu}) \leq d^2) &\geq \mathbb{P} \left((\bar{\mathbf{X}}_T - \boldsymbol{\mu})^t S_m^{-1} (\bar{\mathbf{X}}_T - \boldsymbol{\mu}) \leq \frac{d^2}{\lambda_{(p)m}} \right) \\ &= \mathbb{P} \left(T(\bar{\mathbf{X}}_T - \boldsymbol{\mu})^t S_m^{-1} (\bar{\mathbf{X}}_T - \boldsymbol{\mu}) \leq \frac{T d^2}{\lambda_{(p)m}} \right) \\ &\geq \mathbb{P}(T(\bar{\mathbf{X}}_T - \boldsymbol{\mu})^t S_m^{-1} (\bar{\mathbf{X}}_T - \boldsymbol{\mu}) \leq u) = 1 - \alpha \end{aligned}$$

όπου η πρώτη ανισότητα προκύπτει όπως ακριβώς και για τον πίνακα Σ ενώ η δεύτερη ανισότητα ισχύει επειδή $\frac{T d^2}{\lambda_{(p)m}} \geq u$ σύμφωνα με τον ορισμό της τ.μ. T και λόγω του (i). \square

Αποδείχθηκε λοιπόν, ότι η μέθοδος του Healy (1956) έχει την ιδιότητα της ακριβούς συνέπειας.

2.5.2 Μέθοδος του Srivastava

Έστω τώρα ότι λαμβάνουμε ένα αρχικό τυχαίο δείγμα $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$ μεγέθους $m \geq p + 1$. Ο Srivastava (1967) όρισε τον χρόνο στάσης

$$T = \min \left\{ n \geq m : n \geq \left(\frac{a}{d^2} \right) \lambda_{(p)n} \right\} \quad (2.14)$$

όπου a είναι το άνω α -ποσοστιαίο σημείο της χ_p^2 κατανομής. Κάθε φορά παίρνουμε ένα τυχαίο διάνυσμα, υπολογίζουμε τον πίνακα S_n καθώς και την μέγιστη ιδιοτιμή του, και εξετάζουμε εάν ισχύει η συνθήκη $n \geq \left(\frac{a}{d^2} \right) \lambda_{(p)n}$. Η τελική περιοχή εμπιστοσύνης θα είναι η Π_T .

Θεώρημα 2.5.2. Για τον χρόνο στάσης (2.14) και για κάθε $\mu, \Sigma, p, m, \alpha$ ισχύει ότι:

- (i) $\mathbb{P}(T < \infty) = 1$
- (ii) $\lim_{d \rightarrow 0} \mathbb{E}(T/C) = 1$ (ασυμπτωτική επάρκεια)
- (iii) $\lim_{d \rightarrow 0} \mathbb{P}(\mu \in \Pi_n) = 1 - \alpha$ (ασυμπτωτική συνέπεια)

Η μέθοδος του Srivastava (1967) κληρονομεί τις ασυμπτωτικές ιδιότητες πρώτης τάξης αλλά δεν έχει την ακριβή συνέπεια της μεθόδου του Healy. Η απόδειξη της θεωρήματος βασίζεται σε παρόμοια μεθοδολογία που αναπτύχθηκε στη μονοδιάστατη περίπτωση (βλ. Mukhopadhyay, 2009, σελ. 288-289).

2.6 Σύγκριση μεθόδων μέσω προσομοίωσης

Για να επαληθεύσουμε τις ασυμπτωτικές και τις ακριβείς ιδιότητες της ακολουθιακής μεθόδου και της μεθόδου του Stein καθώς και των μεθόδων κατασκευής σφαιρικών περιοχών, θα πραγματοποιήσουμε Monte Carlo προσομοιώσεις για την εκτίμηση του μέσου αριθμού δείγματος για $\alpha = 0.05$.

Για τη μέθοδο του Stein (Παράρτημα IV, Πρόγραμμα 1) δημιουργούμε ένα βρόχο $n = 10^3$ επαναλήψεων όπου κάθε φορά παράγουμε $n_0 = 10$ τυχαίους αριθμούς από την κατανομή $N(1, 3^2)$, υπολογίζουμε την τιμή της τ.μ. T και την αποθηκεύουμε σε έναν πίνακα T_s . Εάν ισχύει $T \leq n_0$, το τελικό δ.ε. θα είναι $I = [\bar{X}_T - d, \bar{X}_T + d]$. Εάν όμως $T > n_0$ παράγουμε επιπλέον $T - n_0$ τυχαίους αριθμούς από την $N(1, 3^2)$ και υπολογίζουμε το \bar{X}_T . Έχοντας τώρα το τελικό δείγμα και το δ.ε. I , ελέγχουμε εάν $\mu \in I$. Εάν ναι, αυξάνουμε την τιμή ενός μετρητή s κατά ένα ώστε να τον χρησιμοποιήσουμε στον υπολογισμό της πιθανότητας $p = \mathbb{P}(\mu \in I)$. Τα παραπάνω επαναλαμβάνονται για έξι τιμές της ακτίνας d . Ο πίνακας mat.Stein θα περιέχει την τιμή του βέλτιστου δείγματος k , την μέση τιμή \bar{T} , την τυπική απόκλιση $SD(T)$, το πηλίκο \bar{T}/k και την εκτίμηση της πιθανότητας p για τις διάφορες τιμές του d .

Για την πλήρως ακολουθιακή μέθοδο (βλ. Παράρτημα IV, Πρόγραμμα 2) χρησιμοποιούμε ένα βρόχο $n = 10^3$ επαναλήψεων για να πάρουμε τιμές της τ.μ. T . Κάθε φορά παράγουμε

$n_0 = 10$ τυχαίους αριθμούς από την $N(1, 3^2)$ και υπολογίζουμε την ποσότητα $k_0 = \frac{z_{\alpha/2}^2 s_{n_0}^2}{d^2}$. Για να βρούμε το ελάχιστο i που ικανοποιεί την σχέση $i \geq k_0$ (βλ. τον ορισμό της τ.μ. T) χρησιμοποιούμε την εντολή επανάληψης `while`, παράγοντας κάθε φορά έναν τυχαίο αριθμό από την $N(1, 3^2)$ και υπολογίζοντας το καινούργιο k_0 , μέχρις ότου $i \geq k_0$. Έχοντας λοιπόν εκτιμήσει τον χρόνο $T = i$, εξετάζουμε εάν $\mu \in I$ όπως και στην περίπτωση της μεθόδου του Stein. Τα παραπάνω επαναλαμβάνονται για 6 τιμές της ακτίνας d . Ο τελικός πίνακας `mat.Seq` θα περιέχει αντίστοιχες ποσότητες με τον πίνακα `mat.Stein`.

Όσον αφορά την μέθοδο του Healy (βλ. Παράρτημα IV, Πρόγραμμα 3) εφαρμόζουμε τα εξής: Υπολογίζουμε αρχικά την μέγιστη ιδιοτιμή του πίνακα $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$ καθώς και το βέλτιστο μέγεθος δείγματος για δοσμένη τιμή του d . Κατασκευάζουμε έναν βρόχο $n = 10^3$ επαναλήψεων, όπου σε κάθε επανάληψη παράγουμε αρχικά $m = 10$ τυχαίους αριθμούς από την κατανομή $N(\mu, \Sigma)$, όπου $\mu = [1 \ 2]$, και υπολογίζουμε την μέγιστη ιδιοτιμή του δειγματικού πίνακα διασπορών S και μετά την τ.μ. T σύμφωνα με τον ορισμό. Εάν $T \leq m$ η περιοχή εμπιστοσύνης θα είναι $\Pi = \{\mathbf{x} \in \mathbb{R}^p: (\bar{\mathbf{X}}_T - \mathbf{x})^t (\bar{\mathbf{X}}_T - \mathbf{x}) \leq d^2\}$. Εάν όμως $T > m$, παράγουμε επιπλέον $T - m$ τυχαία διανύσματα, υπολογίζουμε το καινούργιο $\bar{\mathbf{X}}_T$ και την περιοχή Π . Εξετάζουμε εάν $\mu \in \Pi$ αυξάνοντας την τιμή ενός μετρητή `sum` ώστε να εκτιμήσουμε την πιθανότητα $\mathbb{P}(\mu \in \Pi)$. Όλα τα παραπάνω υλοποιούνται για 6 τιμές της ακτίνας d . Ο πίνακας `mat.Healy` θα έχει την ίδια μορφή με την μονοδιάστατη περίπτωση.

Τέλος για τη μέθοδο του Srivastava (βλ. Παράρτημα IV, Πρόγραμμα 4) υπολογίζουμε αρχικά τη μέγιστη ιδιοτιμή του πίνακα Σ και την τιμή του βέλτιστου δείγματος C . Σε έναν βρόχο $n = 10^3$ επαναλήψεων πραγματοποιούμε τα ακόλουθα: Παράγουμε $m = 10$ τυχαία διανύσματα από την $N(\mu, \Sigma)$ και υπολογίζουμε την μέγιστη ιδιοτιμή του δειγματικού πίνακα διασπορών S . Στον βρόχο επανάληψης `while` με συνθήκη $N < \frac{\alpha}{d^2} \lambda_n$ παράγουμε κάθε φορά τυχαίο διάνυσμα από την $N(\mu, \Sigma)$ και υπολογίζουμε την μέγιστη ιδιοτιμή του καινούργιου δειγματικού πίνακα S . Στο τέλος του βρόχου θα έχουμε πάρει μία εκτίμηση του χρόνου T . Όμοια με την προηγούμενη μέθοδο χρησιμοποιούμε έναν μετρητή `sum` για τον υπολογισμό της πιθανότητας $\mathbb{P}(\mu \in \Pi)$ και αποθηκεύουμε τα αποτελέσματα στον πίνακα `mat.Sriva` για 6 τιμές του d . Έτσι, καταλήγουμε στους επόμενους δύο πίνακες.

ΠΙΝΑΚΑΣ 2.6.1
Ακριβείς και ασυμπτωτικές εκτιμήσεις των μεθόδων ($p=1$)

d	k	Μέθοδος του Stein				Ακολουθιακή Μέθοδος			
		\bar{T}	$SD(T)$	\bar{T}/k	$\Pr(\mu \in I_T)$	\bar{T}	$SD(T)$	\bar{T}/k	$\Pr(\mu \in I_T)$
1	34,6	46,8	21,6	1,354	0,950	32,9	9,6	0,952	0,940
0,75	61,5	81,5	37,2	1,326	0,952	60,2	12,1	0,979	0,948
0,5	138,3	183,5	82,7	1,327	0,948	136,8	17,5	0,989	0,941
0,25	553,2	757,4	348,3	1,369	0,952	551,3	33,8	0,997	0,940
0,15	1537,6	2034,6	971,3	1,324	0,951	1538,1	56,7	1,001	0,951
0,05	13829,3	18445,1	8906,1	1,334	0,944	13830,0	163,4	1,000	0,959

ΠΙΝΑΚΑΣ 2.6.2
Ακριβείς και ασυμπτωτικές εκτιμήσεις των μεθόδων ($p=2$)

d	C	Μέθοδος του Healy				Μέθοδος του Srivastava			
		\bar{T}	$SD(T)$	\bar{T}/k	$\Pr(\mu \in I_T)$	\bar{T}	$SD(T)$	\bar{T}/k	$\Pr(\mu \in I_T)$
1	13,2	24,9	10,64	1,880	0,994	14,0	3,74	1,062	0,985
0,75	23,5	43,1	18,27	1,835	0,995	23,1	6,76	0,982	0,975
0,5	52,9	94,6	41,62	1,788	0,989	52,0	10,92	0,982	0,971
0,25	221,6	390,9	165,75	1,847	0,993	211,1	21,08	0,998	0,977
0,15	587,7	1060,1	460,51	1,804	0,987	588,2	34,17	1,001	0,985
0,05	5289,5	9590,7	4049,95	1,813	0,990	5288,9	101,79	1,000	0,980

Η ακολουθιακή μέθοδος δείχνει ότι η διασπορά της τ.μ. T είναι σαφώς μικρότερη (όπως δείχνουν οι τυπικές αποκλίσεις στην τέταρτη και όγδοη στήλη του πίνακα), δίνοντας έτσι πιο αξιόπιστη εκτίμηση του μεγέθους k . Σύμφωνα με τον πίνακα, επιβεβαιώνονται όλες οι ασυμπτωτικές ιδιότητες των δύο μεθόδων. Για την διαδικασία του Stein βλέπουμε ότι η εκτίμηση του αναμενόμενου αριθμού του δείγματος είναι πάντοτε μεγαλύτερη του k . Επίσης, για όλες τις τιμές του d η πιθανότητα $p = \mathbb{P}(\mu \in I_T)$ είναι περίπου 0.95. Για την ακολουθιακή μέθοδο ο αναμενόμενος αριθμός του δείγματος είναι μικρότερος του k για μεγάλες τιμές του d ενώ ασυμπτωτικά $\bar{T} \approx k$. Η πιθανότητα p ολοένα και προσεγγίζει την τιμή 0.95 (θεώρημα 2.4.1 (ii)-(iv)), ξεκινώντας από τιμές μικρότερες από αυτήν.

Γενικά, συμπεραίνουμε ότι η ακολουθιακή μέθοδος απαιτεί λιγότερες παρατηρήσεις, δίνει πιο αξιόπιστη εκτίμηση του βέλτιστου μεγέθους k , αλλά δεν έχει την ιδιότητα της ακριβούς συνέπειας. Αντίθετα, η μέθοδος του Stein χρειάζεται περισσότερες παρατηρήσεις για την εκτίμηση του k , αλλά επιλύει το θεμελιώδες πρόβλημα εύρεσης δ.ε. με σταθερό εύρος, για οποιαδήποτε τιμή του εύρους.

Όσον αφορά τις σφαιρικές περιοχές, παρατηρούμε παρόμοια συμπεράσματα με τη μονοδιάστατη περίπτωση. Η μέθοδος του Srivastava ωστόσο, δεν ικανοποιεί επακριβώς την

ιδιότητα της ασυμπτωτικής συνέπειας, για τις δοσμένες ακτίνες των περιοχών. Επίσης, αξίζει να σημειωθεί η μέθοδος του Stein έχει την ιδιότητα της ασυμπτωτικής συνέπειας σε αντίθεση με τη μέθοδο του Healy.

Διάφορες παραλλαγές των μονοδιάστατων μεθόδων έχουν δημιουργηθεί με σκοπό να βελτιωθεί η ταχύτητα τερματισμού της ακολουθιακής διαδικασίας αλλά και άλλες ασυμπτωτικές ιδιότητες. Ενδεικτικά έχουν κατασκευαστεί η επιταχυνόμενη μέθοδος καθώς και η μέθοδος των τριών βημάτων οι οποίες βασίζονται σε μία αρχική εκτίμηση ενός κλάσματος ρk , όπου $\rho \in (0,1)$. Ωστε όταν το ρ θα είναι κοντά στο μηδέν οι μέθοδοι θα συμπεριφέρονται όπως και η μέθοδος του Stein, ενώ για τιμές κοντά στο ένα, όπως η πλήρως ακολουθιακή μέθοδος.

Αποδεικνύεται (βλ. Mukhopadhyay, 2009, σελ.130-131) ότι η επιταχυνόμενη μέθοδος είναι ταχύτερη από την ακολουθιακή μέθοδο, κληρονομεί τις ασυμπτωτικές ιδιότητες, αλλά δεν καταφέρνει να ικανοποιήσει την ιδιότητα της ακριβούς συνέπειας. Παρόμοια αποτελέσματα ισχύουν και για την μέθοδο των τριών βημάτων. Στον επόμενο πίνακα δίνουμε τις ιδιότητες των τεσσάρων μεθόδων, όπου η ασυμπτωτική επάρκεια δεύτερης τάξεως επιτυγχάνεται όταν

$$\lim_{d \rightarrow 0} \mathbb{E}(T - k) < \infty.$$

ΠΙΝΑΚΑΣ 2.6.3
Ιδιότητες των μεθόδων εκτίμησης μέσου κατά διάστημα

Μέθοδος	Ιδιότητα			
	Ακριβής		Ασυμπτωτική	
	Συνέπεια	Συνέπεια	Επάρκεια πρώτης τάξης	Επάρκεια δεύτερης τάξης
<i>Stein</i>	✓	✓	✗	✗
<i>Ακολουθιακή</i>	✗	✓	✓	✓
<i>Επιταχυνόμενη</i>	✗	✓	✓	✓
<i>Τριών βημάτων</i>	✗	✓	✓	✓
<i>Healy</i>	✓	✗	✗	✗
<i>Srivastava</i>	✗	✓	✓	✓

Σημειώνουμε ότι οι τεχνικές που παρουσιάστηκαν σε αυτήν την παράγραφο, μπορούν να επεκταθούν και για την εύρεση εκτιμητή του μέσου μίας κανονικής κατανομής, που να έχει φραγμένο ρίσκο για εκ των προτέρων δοσμένο άνω φράγμα. Αρκεί να χρησιμοποιηθεί μία ικανή και αναγκαία συνθήκη της μορφής (2.1.7).

Ακολουθιακός Έλεγχος Λόγου Πιθανοφαινειών

3.1 Εισαγωγή

Έστω i.i.d. τ.μ. $(X_n)_{n \geq 1}$ όπου $X_i \sim f(x|\theta)$ με μέση τιμή θ και διασπορά $0 < \sigma < \infty$. Θα ασχοληθούμε αρχικά με τον έλεγχο της υπόθεσης

$$\mathcal{H}_0: \theta = \theta_0 \text{ έναντι της } \mathcal{H}_1: \theta = \theta_1 \text{ (} \theta_1 > \theta_0 \text{)} \quad (3.1)$$

όπου $\theta_0, \theta_1 \in \Theta \subset \mathbb{R}$. Το μέγεθος του δείγματος σύμφωνα με την κλασική θεωρία των Neyman-Pearson (1933) είναι σταθερό και γνωστό εκ των προτέρων. Όμως, τώρα θα παίρνουμε μία-μία τις παρατηρήσεις και κάθε φορά θα ελέγχουμε για απόρριψη ή αποδοχή της \mathcal{H}_0 . Ορίζουμε το λόγο των πιθανοφαινειών

$$R_n := \frac{L(\theta_1|X_1, \dots, X_n)}{L(\theta_0|X_1, \dots, X_n)} = \prod_{i=1}^n \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}.$$

Μεγάλες τιμές του R_n θα δίνουν ένδειξη για να απορρίψουμε την \mathcal{H}_0 , ενώ μικρές τιμές αντίστοιχα για να αποδεχθούμε την \mathcal{H}_0 . Αρκεί επομένως να θεωρήσουμε κάποια ιδανικά φράγματα τα οποία θα προσδιορίζουν τον κανόνα του τερματισμού. Έστω $A, B \in \mathbb{R}_+$ με $A < B$. Ο ακολουθιακός έλεγχος του λόγου πιθανοφαινειών (*sequential probability ratio test-SPRT*) βασίζεται στην εξής αναδρομική διαδικασία: εάν στο n -στο βήμα ισχύει

$$A < R_n < B \quad (3.2)$$

τότε παίρνουμε ακόμα μία παρατήρηση X_{n+1} , υπολογίζουμε το R_{n+1} και ελέγχουμε εκ νέου την συνθήκη (3.2). Θα σταματήσουμε την δειγματοληψία όταν $R_n \leq A$, οπότε και θα αποδεχθούμε την \mathcal{H}_0 , ή όταν $R_n \geq B$ απορρίπτοντας την \mathcal{H}_0 . Μπορούμε να πάρουμε μία πιο βολική μορφή της (3.2) που θα μας διευκολύνει παρακάτω γιατί θα κάνουμε χρήση των αποτελεσμάτων από το δεύτερο κεφάλαιο. Λογαριθμίζοντας κατά μέρη την (3.2) θα πάρουμε την ισοδύναμη ανισότητα

$$a < S_n < b \quad (3.3)$$

όπου $a = \log A$, $b = \log B$, $S_n = \sum_{i=1}^n Z_i$ και $Z_i = \log \left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \right)$.

ΠΙΝΑΚΑΣ 3.1.1
SPRT στο n -στο βήμα

Συνθήκη	Απόφαση
$S_n < a$	αποδοχή της μηδενικής υπόθεσης
$S_n > b$	απόρριψη της μηδενικής υπόθεσης
$a < S_n < b$	εκ νέου δειγματοληψία

Σύμφωνα με τα παραπάνω δημιουργούνται τρία βασικά ερωτήματα: Πως υπολογίζονται τα φράγματα A και B ; Στον έλεγχο SPRT λαμβάνουμε απόφαση σε πεπερασμένου πλήθους βήματα; Τι πλεονεκτήματα ή μειονεκτήματα έχει ο SPRT σε σχέση με τον ισχυρότατο έλεγχο των Neyman-Pearson; Πως μπορεί να επεκταθεί η παρούσα θεωρία σε ελέγχους σύνθετων υποθέσεων;

3.2 Πεπερασμένος Τερματισμός

Ορίζουμε την τ.μ.

$$T := \inf \{n \geq 1: S_n \notin (a, b)\}, \quad \inf \emptyset = \infty \quad (3.4)$$

Η T είναι χ.σ. Πράγματι, αν θεωρήσουμε την συνήθη διήθηση των τ.μ. X_1, X_2, \dots παίρνουμε ότι:

$$\{T = n\} = \{S_1 \in (a, b)\} \cap \dots \cap \{S_{n-1} \in (a, b)\} \cap \{S_n \notin (a, b)\} \in \mathcal{F}_n$$

γιατί οι τ.μ. S_i είναι \mathcal{F}_i -μετρήσιμες. Μπορούμε να δούμε αρχικά ότι η τ.μ. T είναι πεπερασμένη σ.β., δηλαδή $\mathbb{P}(T < \infty) = 1$, αν κάνουμε χρήση του κεντρικού οριακού θεωρήματος:

$$\mathbb{P}(a < S_n < b) \approx \Phi\left(\frac{b - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a - n\mu}{\sigma\sqrt{n}}\right)$$

Αφήνοντας το n να τείνει στο ∞ , η τελευταία ποσότητα τείνει στο μηδέν. Ωστόσο, θα μπορούσαμε να πάρουμε ισχυρότερα αποτελέσματα για την τ.μ. T , βασιζόμενοι στο επόμενο θεώρημα (βλ. Govindarajulu, 1974, σελ 23-25).

Θεώρημα 3.2.1. *Εάν οι i.i.d. τ.μ. $(Z_n)_{n \geq 1}$ ικανοποιούν την συνθήκη $\mathbb{P}(Z_i = 0) < 1$, τότε υπάρχουν σταθερές $q \in (0, 1)$ και $c \in \mathbb{R}_+$ τέτοιες ώστε $\mathbb{P}(T > n) \leq cq^n, \forall n \geq 1$.*

Απόδειξη. Επειδή ισχύει $\mathbb{P}(Z_i = 0) < 1$ υποθέτουμε, χωρίς βλάβη της γενικότητας, ότι $\mathbb{P}(Z_i > 0) > 0$ (ανάλογα αποτελέσματα παίρνουμε και στην περίπτωση $\mathbb{P}(Z_i > 0) < 0$).

$$\exists \varepsilon, \delta > 0 \text{ τέτοια ώστε } \mathbb{P}(Z > \varepsilon) \geq \delta \quad (3.5)$$

Έστω $r \in \mathbb{N}$ ώστε

$$r\varepsilon > \log(B/A) \quad (3.6)$$

και ας είναι $n = kr$, όπου $k \in \mathbb{N}$ σταθερά. Ορίζουμε τα αθροίσματα

$$S_1 = \sum_{i=1}^r Z_i, \quad S_2 = \sum_{i=r}^{2r} Z_i, \dots, \quad S_k = \sum_{i=(k-1)r}^{kr} Z_i.$$

Εφ' όσον οι Z_i είναι i.i.d. τ.μ. τότε και οι S_i θα είναι i.i.d. και

$$\begin{aligned} \mathbb{P}\left(|S_1| > \log\left(\frac{B}{A}\right)\right) &= \mathbb{P}\left(\left|\sum_{i=1}^r Z_i\right| > \log\left(\frac{B}{A}\right)\right) \geq \mathbb{P}\left(\sum_{i=1}^r Z_i > \log\left(\frac{B}{A}\right)\right) \\ &\geq \mathbb{P}\left(\sum_{i=1}^r Z_i > r\varepsilon\right) \geq \prod_{i=1}^r \mathbb{P}(Z_i > \varepsilon) \geq \delta^r \end{aligned}$$

όπου για την δεύτερη ανισότητα κάναμε χρήση της (3.6). Για να τερματίσει ο SPRT μετά από kr βήματα θα πρέπει κατ' ανάγκη όλες οι μεταβλητές $|S_i|$ να είναι μικρότερες του $\log(B/A)$. Άρα

$$\mathbb{P}(T > kr) \leq \prod_{j=1}^k \mathbb{P}\left(|S_j| < \log\left(\frac{B}{A}\right)\right) \leq (1 - \delta^r)^k.$$

Θέτοντας $c = 1/(1 - \delta^r)$ και $q = (1 - \delta^r)^{1/r}$ προκύπτει μετά από πράξεις η ζητούμενη ανισότητα. \square

Η παραπάνω ιδιότητα δείχνει ότι η τ.μ. T έχει θετικά ασύμμετρη κατανομή.

Πόρισμα 3.2.1. Για το χρόνο στάσης T ισχύει ότι $\mathbb{E}(T) < \infty$.

Απόδειξη. Στην περίπτωση του SPRT, το σύνολο $\{f(X_i|\theta_1) = f(X_i|\theta_0)\}$ έχει πιθανότητα μηδέν αφού $\theta_1 \neq \theta_0$. Άρα $\mathbb{P}(Z_i = 0) < 1$ και από το Θεώρημα 3.2.1:

$$\mathbb{E}(T) = \sum_{n=1}^{\infty} \mathbb{P}(T \geq n) = \mathbb{P}(T = 1) + \sum_{n=1}^{\infty} \mathbb{P}(T > n) \leq c_0 + c \sum_{n=1}^{\infty} q^n = \frac{c'}{(1-q)} < \infty. \quad \square$$

Αφού λοιπόν δείξαμε ότι $\mathbb{E}(T) < \infty$ άρα $\mathbb{P}(T < \infty) = 1$. Σημειώνουμε ότι με την παραπάνω ανισότητα αποδεικνύεται ότι όλες οι ροπές της τ.μ. T είναι πεπερασμένες.

3.3 Υπολογισμός φραγμάτων

Θεωρούμε τώρα τον κανόνα απόφασης $d: \mathcal{X} \rightarrow \{0,1\}$, που ορίζεται από την σχέση

$$d = \begin{cases} 0, & S_T \leq a \\ 1, & S_T \geq b \end{cases} \quad (3.7)$$

Τότε ο SPRT θα καθορίζεται πλήρως από το ζεύγος (T, d) . Προφανώς, όταν ο έλεγχος αφορά σταθερό δείγμα μεγέθους n τότε $T \equiv n$. Συμβολίζουμε με $\mathbb{P}_\theta(\cdot)$ την πιθανότητα ενός ενδεχομένου, όταν η παράμετρος της σ.π.π. είναι θ . Για τα σφάλματα τύπου I και II έχουμε:

$$\alpha = \mathbb{P}(I) = \mathbb{P}_{\theta_0}(d = 1)$$

$$\beta = \mathbb{P}(II) = \mathbb{P}_{\theta_1}(d = 0)$$

Έχουμε την δυνατότητα να βρούμε συνθήκες για τις τιμές των A και B , εργαζόμενοι κατά τρόπο ανάλογο της περιπτώσεως σταθερού δείγματος.

Θεώρημα 3.3.1. *Για τα φράγματα A και B του SPRT ισχύουν οι σχέσεις*

$$A \geq \frac{\beta}{1-\alpha} \text{ και } B \leq \frac{1-\beta}{\alpha}$$

Οι αριθμοί $\tilde{A} = \frac{\beta}{1-\alpha}$ και $\tilde{B} = \frac{1-\beta}{\alpha}$ θα ονομάζονται προσεγγίσεις του Wald.

Απόδειξη. Έστω $\mathcal{A}_n^i \subset \mathbb{R}^n$ το σύνολο των σημείων για τα οποία απορρίπτουμε την υπόθεση \mathcal{H}_i όπου $i = 0,1$ στο βήμα n . Παρατηρούμε ότι $\forall n \neq m : \mathcal{A}_n^i \cap \mathcal{A}_m^i = \emptyset$. Τότε

$$\begin{aligned} \mathbb{P}_{\theta_0}(d = 1) &= \sum_{1 \leq n < \infty} \mathbb{P}_{\theta_0}(d = 1 | T = n) \\ &= \sum_{1 \leq n < \infty} \int_{\mathcal{A}_n^0} L(\theta_0 | \mathbf{x}) d\mathbf{x} \leq B^{-1} \sum_{1 \leq n < \infty} \int_{\mathcal{A}_n^0} L(\theta_1 | \mathbf{x}) d\mathbf{x} = B^{-1}(1 - \beta) \end{aligned}$$

άρα $B \leq \frac{1-\beta}{\alpha}$.

Όμοια

$$\begin{aligned} \mathbb{P}_{\theta_1}(d = 0) &= \sum_{1 \leq n < \infty} \mathbb{P}_{\theta_1}(d = 0 | T = n) \\ &= \sum_{1 \leq n < \infty} \int_{\mathcal{A}_n^1} L(\theta_1 | \mathbf{x}) d\mathbf{x} \leq A \sum_{1 \leq n < \infty} \int_{\mathcal{A}_n^1} L(\theta_0 | \mathbf{x}) d\mathbf{x} = A(1 - \alpha) \end{aligned}$$

άρα $A \geq \frac{\beta}{1-\alpha}$. \square

Επισημαίνουμε πως οι προσεγγίσεις του Wald δεν εγγυώνται ότι οι πιθανότητες σφαλμάτων I και II θα είναι ακριβώς ίσες με τα δοσμένα α και β . Εάν α' και β' είναι οι πιθανότητες των σφαλμάτων που παίρνουμε τότε από το Θεώρημα 3.3.1 ότι

$$\frac{\beta'}{(1-\alpha')} \leq \frac{\beta}{(1-\alpha)} \Rightarrow (1-\alpha)\beta' \leq \beta(1-\alpha')$$

$$\frac{1-\beta'}{\alpha'} \geq \frac{1-\beta}{\alpha} \Rightarrow (1-\beta)\alpha' \leq (1-\beta')\alpha \Rightarrow \alpha' + \beta' \leq \alpha + \beta$$

Αυτό πρακτικά σημαίνει ότι τουλάχιστον μία από τις ανισότητες $\alpha' \leq \alpha$ ή $\beta' \leq \beta$ θα πρέπει να ισχύει, δηλαδή από την εφαρμογή του SPRT το πολύ μία πιθανότητα σφάλματος θα γίνει μεγαλύτερη από αυτήν που ζητήσαμε. Επίσης, συχνά επιλέγουμε τις τιμές των α και β αρκετά μικρές. Χωρίς βλάβη της γενικότητας θα θεωρούμε ότι $\alpha + \beta < 1$. Τότε προκύπτει ότι $B > 1 > A$ και επαληθεύουμε ότι όσο πιο κοντά είναι το θ_0 στο θ_1 , άρα και ο λόγος των πιθανοφανειών κοντά στο 1, τόσο περισσότερες παρατηρήσεις θα χρειαζόμαστε για να πάρουμε απόφαση. Θα δούμε ένα παρόμοιο αποτέλεσμα με χρήση της απόκλισης Kullback-Leibler (βλ. Nowak, 2010, σελ 1-3).

Παράδειγμα 3.3.1. Έστω i.i.d. τ.μ. $(X_n)_{n \geq 1}$ όπου $X_i \sim N(\theta, \sigma^2)$ και η μέση τιμή θ είναι άγνωστη ενώ η διασπορά σ^2 είναι γνωστή σταθερά. Η σ.π.π. είναι

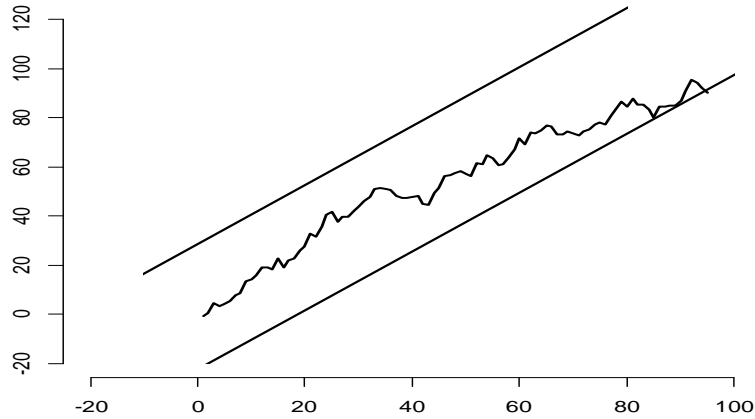
$$f(x|\theta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}, -\infty < x < +\infty$$

Υπολογίζουμε ότι $Z_i = \log \left[\frac{f(X|\theta_1, \sigma^2)}{f(X|\theta_0, \sigma^2)} \right] = \frac{\theta_1 - \theta_0}{\sigma^2} \left(X - \frac{\theta_1 + \theta_0}{2} \right)$.

Για τον έλεγχο της υπόθεσης (3.1), από την συνθήκη (3.3) παίρνουμε ότι

$$C_A + nD < \sum_{i=1}^n X_i < C_B + nD$$

όπου $C_A = (\theta_1 - \theta_0)^{-1} \sigma^2 \log(A)$, $C_B = (\theta_1 - \theta_0)^{-1} \sigma^2 \log(B)$ και $D = (\theta_0 + \theta_1)/2$



Σχήμα 3.3.1 Προσομοίωση υπό την \mathcal{H}_0 για $\theta_0 = 1$, $\theta_1 = 1.4$, $\sigma = 1$, $\alpha = 0.05$ και $\beta = 0.1$.

Σύμφωνα με το σχήμα θα πάρουμε απόφαση υπέρ ή κατά της \mathcal{H}_0 όταν ο τυχαίος περίπατος $Y_n = \sum_{i=1}^n X_i$ διασχίσει μία από τις δύο παράλληλες ευθείες $\varepsilon_A = C_A + nD$ ή $\varepsilon_B = C_B + nD$. Για την κατασκευή του Σχήματος 3.3.1 (βλ. Παράρτημα ΙΙΙ) θέτουμε ως $A = \frac{\beta}{1-\alpha}$ και $B = \frac{1-\beta}{\alpha}$. Υπολογίζουμε αρχικά τα C_A, C_B και D για τις δοσμένες τιμές των θ_1 και θ_0 . Όσο το άθροισμα Y_n ικανοποιεί την σχέση $C_A + nD < Y_n < C_B + nD$, παράγουμε τυχαίο αριθμό από την $N(\theta_0, \sigma)$, υπολογίζουμε το καινούργιο άθροισμα Y_{n+1} και εξετάζουμε εάν ικανοποιείται η παραπάνω ανισότητα για $n+1$. Έτσι απεικονίζουμε τις τιμές Y_n που βρίσκονται εντός των ευθειών μέχρι να συμβεί η πρώτη έξοδος ενός σημείου από τις παράλληλες ευθείες.

3.4 Βέλτιστη ιδιότητα

Το επόμενο θεώρημα οφείλεται στους Wald και Wolfowitz (1948) και χαρακτηρίζεται ως «βέλτιστη ιδιότητα του SPRT» γιατί υπερτερεί, ως προς το μέγεθος του δείγματος, έναντι του ελέγχου των Neyman-Pearson.

Θεώρημα 3.4.1. Έστω ότι θέλουμε να κάνουμε τον έλεγχο της υπόθεσης (3.1) και (T, d) ας είναι ο SPRT όπως ορίστηκε στην (3.3) με $\mathbb{P}_{\theta_0}(d = 1) = \alpha$, $\mathbb{P}_{\theta_1}(d = 0) = \beta$ και $\alpha + \beta < 1$. Εάν (T', d') είναι ένας ακολουθιακός έλεγχος που ικανοποιεί τις σχέσεις

- (i) $\mathbb{E}_{\theta_0}(T'), \mathbb{E}_{\theta_1}(T') < \infty$
- (ii) $\mathbb{P}_{\theta_0}(d' = 1) \leq \alpha$ και $\mathbb{P}_{\theta_1}(d' = 0) \leq \beta$

Τότε $\mathbb{E}_{\theta_0}(T') \geq \mathbb{E}_{\theta_0}(T)$ και $\mathbb{E}_{\theta_1}(T') \geq \mathbb{E}_{\theta_1}(T)$.

Η απόδειξη του θεωρήματος βασίζεται σε έννοιες της θεωρίας αποφάσεων καθώς και της θ.β.σ.

Έστω ότι το θ είναι τ.μ. που παίρνει δύο τιμές θ_0, θ_1 με εκ των προτέρων κατανομή $\pi = \mathbb{P}(\theta = \theta_0)$ και $\mathbb{P}(\theta = \theta_1) = 1 - \pi$. Θεωρούμε ότι κάθε επιπλέον παρατήρηση που συλλέγουμε έχει κόστος $c > 0$ και ότι η λανθασμένη απόρριψη της \mathcal{H}_0 έχει κόστος $c_0 > 0$ ενώ της \mathcal{H}_1 έχει κόστος $c_1 > 0$. Η συνάρτηση απώλειας (βλ. Shiryaev, 2007, σελ 165) είναι

$$L(\theta, d) = \begin{cases} c_0, & (\theta = \theta_0, d = 1) \\ c_1, & (\theta = \theta_1, d = 0) \\ 0, & (\theta = \theta_i, d = i, i = 0,1) \end{cases}$$

Η αναμενόμενη τιμή της L είναι:

$$\begin{aligned} \mathbb{E}(L(\theta, d)) &= c_0 \mathbb{P}(\{\theta = \theta_0\} \cap \{d = 1\}) + c_1 \mathbb{P}(\{\theta = \theta_1\} \cap \{d = 0\}) \\ &= c_0 \pi \mathbb{P}(d = 1 | \theta = \theta_0) + c_1 (1 - \pi) \mathbb{P}(d = 0 | \theta = \theta_1) \end{aligned}$$

σύμφωνα με τον τύπο του Bayes. Τότε το ρίσκο του κανόνα απόφασης δ είναι ίσο με

$$\rho_\pi(\delta) = \mathbb{E}_\pi(L(\theta, d)) + c \mathbb{E}_\pi(T)$$

Συμβολίζουμε με $\pi_n = \mathbb{P}(\theta = \theta_0 | X_1, \dots, X_n)$ και ως $\Delta := \{\delta = (T, d) : T \in \mathcal{T}_n \text{ και } d \in \mathcal{D}\}$, όπου \mathcal{T}_n είναι το σύνολο των ομαλών χρόνων στάσης ως προς τη διήθηση \mathcal{F}_n και \mathcal{D} το σύνολο των κανόνων απόφασης. Αποδεικνύεται (Shiryaev, 2007, σελ 165-170) ότι ο κανόνας δ που ελαχιστοποιεί το ρίσκο $\rho_{\pi_n}(\delta)$ αποτελείται από χρόνο στάσης της μορφής (3.7), δηλαδή πρόκειται για τον έλεγχο SPRT.

3.5 Συναρτήσεις OC και ASN

Κατά την κατασκευή ακολουθιακών ελέγχων έχουν δημιουργηθεί κάποια μέτρα που εκφράζουν την αποδοτικότητα τους και δρουν ως μέτρα σύγκρισης. Το πρώτο είναι μία συνάρτηση που δείχνει την αξιοπιστία του ελέγχου, όπως η συνάρτηση ισχύος στους ελέγχους σταθερού μεγέθους δείγματος ενώ το δεύτερο αφορά την ταχύτητα του ελέγχου, υπό την έννοια του αναμενόμενου αριθμού παρατηρήσεων που θα χρειαστούν. Δημιουργούμε λοιπόν δύο καμπύλες που θα συνοδεύουν κάθε έλεγχο (T, d) .

Ορισμός 3.5.1. Ονομάζουμε λειτουργική καμπύλη χαρακτηριστικών (*operating characteristic curve-OC*) τη συνάρτηση $Q(\theta) = \mathbb{P}_\theta(d = 0)$.

Προφανώς $Q(\theta_0) = 1 - \alpha$ και $Q(\theta_1) = \beta$ και $Q(\theta) = 1 - \pi(\theta)$, όπου $\pi(\theta)$ είναι η συνάρτηση ισχύος του ελέγχου. Θα ασχοληθούμε με την περίπτωση των απλών ελέγχων υποθέσεων. Για να βρούμε μία αναλυτική έκφραση της συνάρτησης Q θα κάνουμε χρήση του Θεωρήματος (1.5.1). Εάν η τ.μ. $Z_i = \log \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}$ ικανοποιεί τις τέσσερις συνθήκες της Πρότασης (1.5.1) τότε $\exists t_0 \neq 0$ τέτοιο ώστε $\mathbb{E}_\theta(e^{-t_0 S_T}) = 1$. Από αυτήν την σχέση και λόγω της μοναδικότητας συμπεραίνουμε ότι το t_0 είναι συνάρτηση του θ . Κάνοντας την παραδοχή ότι η τ.μ. S_T δεν υπερβαίνει κατά πολύ τα άκρα a και b δηλαδή $S_T \approx a$ ή $S_T \approx b$ τότε

$$\begin{aligned} 1 &= \mathbb{E}_\theta(e^{-t_0 S_T}) = \mathbb{E}_\theta(e^{-t_0 S_T} | S_T \leq a) + \mathbb{E}_\theta(e^{-t_0 S_T} | S_T \geq b) \\ &\approx e^{-t_0 a} \mathbb{P}_\theta(S_T \leq a) + e^{-t_0 b} \mathbb{P}_\theta(S_T \geq b) \\ &= e^{-t_0 a} Q(\theta) + e^{-t_0 b} (1 - Q(\theta)) \end{aligned}$$

άρα $Q(\theta) \approx \frac{e^{-t_0 b} - 1}{e^{-t_0 b} - e^{-t_0 a}}$ για $t_0 \neq 0$. Επεκτείνουμε τη συνάρτηση Q σε όλο το \mathbb{R} ώστε να είναι συνεχής:

$$Q(\theta) \approx \begin{cases} \frac{e^{-t_0(\theta)b} - 1}{e^{-t_0(\theta)b} - e^{-t_0(\theta)a}}, & t_0(\theta) \in \mathbb{R} \setminus \{0\} \\ \frac{b}{b-a}, & t_0(\theta) = 0 \end{cases} \quad (3.8)$$

Με όμοιο τρόπο επεκτείνουμε και την συνάρτηση $t_0 = t_0(\theta)$ (το t_0 μπορεί να επιλυθεί ως προς θ , π.χ. βλ. σχέση 3.12) στο \mathbb{R} , δίνοντας τιμή ίση με το όριο της όταν $t = 0$.

Ορισμός 3.5.2. Ονομάζουμε μέσο αριθμό δείγματος (*average sample number-ASN*) την αναμενόμενη τιμή $\mathbb{E}_\theta(T)$ ως συνάρτηση του θ .

Αφού ο T είναι χ.σ. με $\mathbb{E}(T) < \infty$ διακρίνουμε δύο περιπτώσεις:

(i) Εάν $\mathbb{E}_\theta(Z_1) \neq 0$, από την πρώτη εξίσωση του Wald ισχύει

$$\mathbb{E}_\theta(T) = \frac{\mathbb{E}_\theta(S_T)}{\mathbb{E}_\theta(Z_1)} \approx \frac{aQ(\theta) + b(1 - Q(\theta))}{\mathbb{E}_\theta(Z_1)} \quad (3.9)$$

(ii) Εάν $\mathbb{E}_\theta(Z_1) = 0$, εφαρμόζοντας την δεύτερη εξίσωση του Wald

$$\mathbb{E}_\theta(T) = \frac{\mathbb{E}_\theta(S_T^2)}{\mathbb{E}_\theta(Z_1^2)} \approx \frac{a^2 Q(\theta) + b^2 (1 - Q(\theta))}{\mathbb{E}_\theta(Z_1^2)} \quad (3.10)$$

Σημειώνουμε ότι όταν η συνάρτηση ASN είναι συνεχής στο διάστημα $[\theta_0, \theta_1]$, θα έχει μία τουλάχιστον ρίζα $\tilde{\theta} \in (\theta_0, \theta_1)$, σύμφωνα με τις ιδιότητες της απόκλισης K-L (Παράρτημα I).

Αν και για τους απλούς ελέγχους υποθέσεων χρησιμοποιούμε δύο μόνο σημεία του παραμετρικού χώρου Θ , θα κατασκευάσουμε τις συναρτήσεις OC και ASN για διάφορες

τιμές του παραμετρικού χώρου (βλ. Πίνακα 3.6.2), καθώς υπάρχουν ακολουθιακοί έλεγχοι σύνθετων υποθέσεων (βλ. Wald, 1947) που μπορούν να αναχθούν, με χρήση κατάλληλων μετασχηματισμών, σε ελέγχους της μορφής (3.1). Αρκετοί από αυτούς τους ελέγχους υπάγονται στη γενική θεωρία ακολουθιακών αναλλοίωτων ελέγχων (*sequential invariant tests*) (βλ. Baseville, 1993, σελ. 148). Ωστόσο δεν θα μελετηθούν στην παρούσα εργασία καθώς

- (i) οι σύνθετοι ακολουθιακοί έλεγχοι απαιτούν εκ των προτέρων γνώση των δεδομένων
- (ii) δεν έχουν κάποια βέλτιστη ιδιότητα που να τους καταστεί συγκρίσιμους με τους αντίστοιχους ελέγχους σταθερού δείγματος
- (iii) οι μετασχηματισμοί δεν διατηρούν την ανεξαρτησία των δεδομένων
- (iv) έχουν μεγάλη υπολογιστική πολυπλοκότητα.

3.6 Σύγκριση μεθόδων μέσω προσομοίωσης

Σε αυτήν την παράγραφο θα εξετάσουμε τον έλεγχο της υπόθεσης (3.1) με χρήση προσομοιωμένων δεδομένων όταν $\theta_0 = 1$, $\theta_1 = 1.4$, $\sigma = 2$ και $\alpha = 0.05$ (καθώς και $\beta = 0.1$ για τις Παραγράφους 3.6.2 και 3.6.3).

3.6.1 Σύγκριση με τον έλεγχο των Neyman-Pearson

Για να συγκρίνουμε τους δύο ελέγχους απλών υποθέσεων, θα χρησιμοποιήσουμε την βέλτιστη ιδιότητα του SPRT. Εάν για δοσμένο μέγεθος δείγματος n και σφάλμα τύπου I, βρούμε από τον έλεγχο των Neyman-Pearson το σφάλμα τύπου II, τότε χρησιμοποιώντας τον SPRT με αυτά τα δύο σφάλματα θα πρέπει τα αναμενόμενα μεγέθη των δειγμάτων που θα προκύψουν να είναι μικρότερα του n .

Ας υποθέσουμε ότι παίρνουμε δείγμα μεγέθους n από την κανονική κατανομή $N(\theta, \sigma^2)$. Τότε για τον έλεγχο της υπόθεσης (3.1), ο ομοιόμορφα ισχυρότατος έλεγχος σύμφωνα με τους Neyman-Pearson θα έχει απορριπτική περιοχή (π.χ. βλ. Hogg, 1970, σελ 243)

$$R = \{\mathbf{X} \in \mathcal{X} : T(\mathbf{X}) > z_\alpha\}$$

όπου $T(\mathbf{X}) = \frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma}$. Το σφάλμα τύπου II θα δίνεται από τον τύπο

$$\mathbb{P}(II) = \Phi\left(\frac{\sqrt{n}(\theta_0 - \theta_1)}{\sigma} + z_\alpha\right)$$

από όπου μπορούμε να πάρουμε την αντίστοιχη τιμή του β .

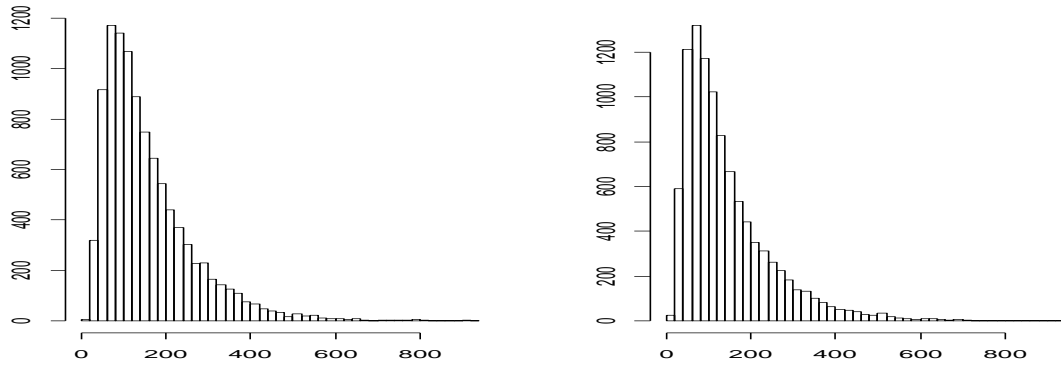
Κατά την υλοποίηση των παραπάνω (βλ. Παράρτημα IV, Πρόγραμμα 5) εφαρμόζουμε τα εξής: Για την εκτίμηση των ποσοτήτων $\mathbb{E}_{\theta_0}(T)$, $\mathbb{P}(II)$ και $\mathbb{E}_{\theta_1}(T)$, $\mathbb{P}(I)$ του SPRT χρησιμοποιούμε δύο βρόχους επανάληψης (μέσω της εντολής while) όπου εξετάζουμε εάν το άθροισμα S_n ξεπεράσει τις τιμές των a και b (βλ. Πίνακα 3.6.1), παράγοντας τυχαίους αριθμούς από την $N(\theta_0, \sigma^2)$ και $N(\theta_1, \sigma^2)$ αντίστοιχα. Έξω από τους δύο αυτούς βρόχους επανάληψης ελέγχουμε εάν έχουμε λανθασμένη απόρριψη ή λανθασμένη αποδοχή της μηδενικής υπόθεσης, αυξάνοντας τις τιμές μετρητών για την εκτίμηση των πιθανοτήτων. Με τον τερματισμό των βρόχων while παίρνουμε μία τιμή για τα T_{θ_0} και T_{θ_1} . Εσωκλείοντας τα παραπάνω σε έναν βρόχο $N = 10^4$ επαναλήψεων αποθηκεύουμε τα αποτελέσματα σε μία γραμμή του πίνακα matrixSPRT. Ωστόσο, αυτά θα ισχύουν για συγκεκριμένες τιμές των α και β . Όπως αναφέραμε παραπάνω, για δοσμένο α και n (το n θα πάρει 7 τιμές) υπολογίζουμε το $\beta = \mathbb{P}(II)$, και με αυτά τα α και β υπολογίζουμε κάθε φορά τα φράγματα a και b για τον SPRT.

ΠΙΝΑΚΑΣ 3.6.1

Σύγκριση στατιστικών ελέγχων Neyman-Pearson και SPRT

Neyman-Pearson				SPRT						
n	α	β	$\alpha + \beta$	\bar{T}_{θ_0}	$SE(\bar{T}_{\theta_0})$	\bar{T}_{θ_1}	$SE(\bar{T}_{\theta_1})$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\beta}$
30	0,05	0,709	0,759	14,04	0,0019	22,74	0,0024	0,0565	0,625	0,681
50	0,05	0,591	0,641	22,58	0,0029	37,97	0,0036	0,0511	0,520	0,571
100	0,05	0,361	0,411	45,03	0,0046	72,92	0,0056	0,0479	0,319	0,367
150	0,05	0,211	0,261	71,05	0,0064	101,80	0,0073	0,0475	0,192	0,239
200	0,05	0,118	0,168	99,07	0,0081	122,36	0,0085	0,0479	0,106	0,154
250	0,05	0,065	0,115	127,82	0,0094	134,10	0,0094	0,0445	0,056	0,100
300	0,05	0,034	0,084	157,82	0,0107	144,75	0,0107	0,0469	0,029	0,076

Παρατηρούμε ότι για όλα τα n οι ποσότητες \bar{T}_{θ_0} και \bar{T}_{θ_1} είναι μικρότερες του n . Ακόμα, σε όλες τις περιπτώσεις το άθροισμα $\hat{\alpha} + \hat{\beta}$ είναι μικρότερο του $\alpha + \beta$ γεγονός που επιβεβαιώνει τα θεωρητικά αποτελέσματα της Παραγράφου 3.3. Τέλος από την υλοποίηση στο πρόγραμμα R πήραμε τα δύο επόμενα ιστογράμματα για $\alpha = 0.05$, $\beta = 0.034$ και $n = 10^4$



Σχήμα 3.6.1 Ιστογράμματα της κατανομής της τ.μ. T υπό την \mathcal{H}_0 και \mathcal{H}_1 αντίστοιχα

Το σχήμα των ιστογραμμάτων συμφωνεί με την ιδιότητα του Θεωρήματος 3.2.1.

Για να συγκρίνουμε τις συναρτήσεις ASN και OC, δίνουμε δύο χαρακτηριστικά παραδείγματα που εμφανίζονται στη βιβλιογραφία.

3.6.2 Περίπτωση ανεξάρτητων μεταβλητών

Έστω οι i.i.d. τ.μ. $(X_n)_{n \geq 1}$ με $X_i \sim N(\theta, \sigma^2)$. Θα βρούμε τις συναρτήσεις OC και ASN για τον έλεγχο της υπόθεσης (3.1), για δοσμένες πιθανότητες σφαλμάτων α και β . Από το Παράδειγμα (3.3.1), θα είναι

$$Z_i = \log \frac{f(X_i | \theta_1)}{f(X_i | \theta_0)} = -\frac{(X_i - \theta_1)^2}{2\sigma^2} + \frac{(X_i - \theta_0)^2}{2\sigma^2} = \frac{\theta_0 - \theta_1}{\sigma^2} \left(X_i - \frac{\theta_0 + \theta_1}{2} \right)$$

και παίρνοντας την μέση τιμή θα είναι

$$\mathbb{E}_\theta(Z) = \frac{\theta_1 - \theta_0}{\sigma^2} \left(\theta - \frac{\theta_0 + \theta_1}{2} \right) \quad (3.11)$$

$$\mathbb{E}_\theta(Z^2) = k^2(\sigma^2 + \theta^2) + k^2\tilde{\theta}(\tilde{\theta} - 2\theta)$$

όπου $k = (\theta_1 - \theta_0)/\sigma^2$ και $\tilde{\theta} = (\theta_1 + \theta_0)/2$. Ισχύει ότι $0 < F_X(\tilde{\theta}) < 1$, όπου F_X η σ.κ. της X και επομένως $\mathbb{P}(Z < 0) > 0$ και $\mathbb{P}(Z > 0) > 0$. Προφανώς, $\theta_0 < \tilde{\theta} < \theta_1$, που επιβεβαιώνεται και από τις ιδιότητες της απόκλισης K-L. Η ροπογεννήτρια $M_Z(t)$ είναι πεπερασμένη αφού

$$M_Z(t) = e^{-k\tilde{\theta}t} M_X(kt) < \infty.$$

Αποδείξαμε ως τώρα ότι η τ.μ. Z ικανοποιεί τις προϋποθέσεις της Πρότασης (1.5.1) όταν η παράμετρος θ της κανονικής κατανομής είναι διάφορη του $\tilde{\theta}$. Δηλαδή υπάρχει μοναδικό $t_0^* \neq 0$ τέτοιο ώστε $M_Z(t_0^*) = 1$ ή $M_Z(-t_0^*) = 1$, όπου $t_0 = -t_0^*$. Θέλοντας να εφαρμόσουμε τη

θεμελιώδη εξίσωση, αρκεί να δειχθεί ότι η Z ικανοποιεί την υπόθεση (ii) του Θεωρήματος (1.5.1). Σύμφωνα με τον ορισμό της τ.μ. T

$$\forall n < T: a < S_n < b \Rightarrow |S_n| \leq \max\{|a|, |b|\}$$

και επομένως

$$\mathbb{E}(e^{-t_0 Z}) = 1 \Rightarrow t_0(\theta) = \frac{2}{\theta_1 - \theta_0} \left(\theta - \frac{\theta_0 + \theta_1}{2} \right) \quad (3.12)$$

όπου χρησιμοποιήθηκε ότι η ροπογεννήτρια της X είναι $M_X(t) = \exp\{\theta t + \sigma^2 t^2/2\}$. Όταν $\theta = \tilde{\theta}$ θέτουμε $t_0(\tilde{\theta}) = 0$.

Σημειώνουμε ότι για να λύσουμε την εξίσωση $\mathbb{E}(e^{-t_0 Z}) = 1$ στη γενική περίπτωση που η σχέση μεταξύ των t_0 και θ είναι μη γραμμική, δεν χρειάζεται απαραίτητα να χρησιμοποιήσουμε αριθμητικές μεθόδους προσέγγισης, λόγω της μοναδικότητας του t_0 . Παράγουμε τιμές του t_0 σε σύνολο της μορφής $[-l_1, 0) \cup (0, l_2]$, όπου $l_1, l_2 > 0$, παίρνουμε αντίστοιχες τιμές για το θ , αντικαθιστούμε στην σχέση (3.8), βρίσκουμε τιμές της συνάρτησης $Q(\theta)$ και έπειτα της $\mathbb{E}_\theta(T)$ από την (3.9).

3.6.3 Περίπτωση συσχετισμένων μεταβλητών

Αυτή η περίπτωση εμφανίζεται κατά την μελέτη διαγραμμάτων ελέγχου συσχετισμένων τ.μ. Ωστόσο, θα δούμε πως στη συγκεκριμένη περίπτωση (AR(1) μοντέλο) ο λογάριθμος της πιθανοφάνειας μπορεί να εκφραστεί με απλό τρόπο. Έστω ότι $(X_n)_{n \geq 1}$ είναι μία Γκαουσιανή AR(1) ανέλιξη με $\mathbb{E}(X_n) = \theta$ και $\text{var}(X_n) = \sigma^2$, δηλαδή η X_n δίνεται από τον τύπο $X_{n+1} = \rho X_n + \gamma + \varepsilon_n$, όπου $X_0 = 0$, $|\rho| < 1$, $\gamma \in \mathbb{R}$, $\varepsilon_n \sim N(0, \sigma_\varepsilon^2)$ και η από κοινού κατανομή των X_1, X_2, \dots, X_n είναι η πολυμεταβλητή κανονική κατανομή

$$f(\mathbf{x}|\boldsymbol{\theta}, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\theta})^t \Sigma^{-1}(\mathbf{x} - \boldsymbol{\theta})\right\}$$

όπου συμβολίσαμε $\mathbf{x} = (x_1, \dots, x_n)^t$ και $\boldsymbol{\theta} = (\theta, \dots, \theta)^t$. Αποδεικνύεται ότι $\gamma = (1 - \rho)\theta$, $\sigma_\varepsilon^2 = (1 - \rho^2)\sigma^2$ και ότι ο πίνακας διασπορών-συνδιασπορών είναι

$$\Sigma = \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$

Θα κάνουμε τον έλεγχο της υπόθεσης (3.1). Για να εφαρμόσουμε τον SPRT υπολογίζουμε το άθροισμα

$$\begin{aligned}
S_n = \log(R_n) &= \log\left(\frac{f(X_1, \dots, X_n|\theta_1)}{f(X_1, \dots, X_n|\theta_0)}\right) \\
&= \log\left(\frac{f(X_1, \dots, X_{n-1}|\theta_1)f(X_n|X_1, \dots, X_{n-1}; \theta_1)}{f(X_1, \dots, X_{n-1}|\theta_0)f(X_n|X_1, \dots, X_{n-1}; \theta_0)}\right) \\
&= \log\left(\frac{f(X_1, \dots, X_{n-1}|\theta_1)}{f(X_1, \dots, X_{n-1}|\theta_0)}\right) + \log\left(\frac{f(X_n|X_1, \dots, X_{n-1}; \theta_1)}{f(X_n|X_1, \dots, X_{n-1}; \theta_0)}\right) \\
&= S_{n-1} + \lambda_n
\end{aligned} \tag{3.13}$$

Αρκεί να προσδιορίσουμε το λ_n . Εάν $\mathbf{X}_1 = (X_1, \dots, X_{n-1})^t$ και $\mathbf{X}_2 = X_n$ τότε για την δεσμευμένη κατανομή $\mathbf{X}_2|\mathbf{X}_1$ γνωρίζουμε ότι (π.χ. βλ. Johnson, 2007, σελ. 156-163)

$$\mathbf{X}_2|\mathbf{X}_1 \sim N(\theta + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\theta}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

όπου τα Σ_{ij} ορίζονται ως $\Sigma_{12} = \sigma^2[\rho^{n-1} \ \rho^{n-2} \ \dots \ \rho]^t$, $\Sigma_{21} = \sigma^2[\rho^{n-1} \ \rho^{n-2} \ \dots \ \rho]$,

$$\Sigma_{11} = \begin{bmatrix} 1 & \rho & \dots & \rho^{n-2} \\ \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \quad \Sigma_{11}^{-1} = \frac{1}{(1-\rho^2)\sigma^2} \begin{bmatrix} 1 & -\rho & \dots & 0 \\ -\rho & 1+\rho^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

$\boldsymbol{\theta}_1$ είναι το διάνυσμα της μέσης τιμής του \mathbf{X}_1 και \mathbf{x}_1 είναι μία πραγματοποίηση του. Με πράξεις προκύπτει ότι

$$\Sigma_{21}\Sigma_{11}^{-1} = [0, 0, \dots, \rho]$$

και

$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = [0, 0, \dots, \rho] \cdot (\sigma^2[\rho^{n-1}, \rho^{n-2}, \dots, \rho]^t) = \sigma^2\rho^2$$

από όπου έχουμε ότι

$$X_n|X_1, \dots, X_{n-1} \sim N(\theta + \rho(X_{n-1} - \theta), \sigma^2(1 - \rho^2))$$

Από την (3.13):

$$\begin{aligned}
\lambda_n &= -\frac{1}{2(1-\rho^2)\sigma^2} [(X_n - \theta_1 - \rho(X_{n-1} - \theta_1))^2 - (X_n - \theta_0 - \rho(X_{n-1} - \theta_0))^2] \\
&= \frac{(\theta_1 - \theta_0)}{(1+\rho)\sigma^2} \left[((1-\rho)\theta + \varepsilon_n) - (1-\rho)\frac{(\theta_0 + \theta_1)}{2} \right]
\end{aligned} \tag{3.14}$$

3.6.4 Σύγκριση των συναρτήσεων ASN και OC

Θα συγκρίνουμε την προσέγγιση του Wald για τον υπολογισμό των συναρτήσεων ASN και OC με τις συναρτήσεις που προκύπτουν από προσομοίωση. Επίσης, θα υπολογίσουμε τις τιμές των συναρτήσεων για το AR(1) μοντέλο για τρεις τιμές του ρ (βλ. Παράρτημα IV).

Για την κατασκευή των αλγορίθμων (βλ. Παράρτημα IV, Πρόγραμμα 6 και 7) ακολουθούμε τα εξής βήματα: Όσον αφορά τον υπολογισμό των συναρτήσεων ASN και OC με τη προσέγγιση του Wald, δημιουργούμε έναν βρόχο 11 επαναλήψεων για τιμές του θ , όταν $\alpha = 0.05$, $\beta = 0.1$, $\theta_0 = 1$, $\theta_1 = 1.4$, $\sigma = 2$, $a = \log\left(\frac{\beta}{1-\alpha}\right)$, $b = \log\left(\frac{1-\beta}{\alpha}\right)$. Σε κάθε επανάληψη υπολογίζουμε το $t_0(\theta)$ από τον τύπο (3.12), μετά το $Q(\theta)$ από τη σχέση (3.8) και τέλος τον ASN από τους τύπους (3.9) και (3.10), εξετάζοντας εάν $\theta = \frac{\theta_0 + \theta_1}{2}$. Οι τιμές αυτές θα αποθηκευτούν σε έναν πίνακα A.wald. Τώρα για τον υπολογισμό των προσομοιωμένων τιμών, σε ένα βρόχο επανάληψης while εξετάζουμε εάν το άθροισμα S_n ξεπεράσει τα άκρα a και b (βλ. Πίνακα 3.1.1). Ο μετρητής count θα χρησιμοποιηθεί για την εκτίμηση του ASN ενώ ο μετρητής q για την εκτίμηση του OC. Οι επαναλήψεις αυτές θα πραγματοποιηθούν 11 φορές αποθηκεύοντας όλες τιμές σε έναν πίνακα A.sim 11 γραμμών τα αποτελέσματα.

Στο AR(1) μοντέλο (βλ. Παράρτημα IV, Πρόγραμμα 7), η μεθοδολογία είναι παρόμοια. Εξετάζουμε πάλι σε ένα βρόχο επανάληψης while τότε το άθροισμα S_n θα εξέλθει από το διάστημα (a, b) . Οι τιμές αυτού του αθροίσματος προκύπτουν παράγοντας τυχαίο αριθμό από την $N(0, \sigma_\varepsilon^2)$ και αντικαθιστώντας στην σχέση (3.14). Τέλος τα παραπάνω εσωκλείονται σε ένα βρόχο επανάληψης για τρεις τιμές του ρ .

ΠΙΝΑΚΑΣ 3.6.2
Σύγκριση τιμών OC και ASN ανάμεσα στην προσέγγιση του Wald
και σε Monte Carlo προσομοίωση

		Ανεξάρτητες τυχαίες μεταβλητές				Μοντέλο AR(1)					
		Προσέγγιση Wald		Προσομοίωση		$\rho = 0.1$		$\rho = 0.5$		$\rho = 0.9$	
t_0	θ	OC	ASN	OC	ASN	OC	ASN	OC	ASN	OC	ASN
-1,0	1,00	0,950	99,71	0,963	105,59	0,957	131,018	0,957	314,391	0,945	1961,19
-0,8	1,04	0,916	113,69	0,911	121,19	0,911	145,756	0,930	346,230	0,912	2236,76
-0,6	1,08	0,863	128,87	0,877	141,02	0,872	171,881	0,841	401,662	0,877	2438,38
-0,4	1,12	0,786	143,74	0,784	154,93	0,806	186,359	0,789	466,843	0,799	2707,46
-0,2	1,16	0,683	155,88	0,694	170,70	0,689	205,640	0,700	484,150	0,688	2943,03
0,0	1,20	0,562	162,68	0,567	171,77	0,530	213,436	0,547	529,469	0,549	3154,75
0,2	1,24	0,436	162,60	0,397	177,08	0,422	218,780	0,413	528,497	0,436	3093,65
0,4	1,28	0,319	156,07	0,303	184,22	0,326	206,288	0,319	493,541	0,323	3089,11
0,6	1,32	0,224	145,09	0,220	169,72	0,233	189,738	0,210	450,108	0,211	2715,06
0,8	1,36	0,151	132,04	0,149	154,95	0,131	168,969	0,159	401,801	0,142	2524,26
1,0	1,40	0,100	118,81	0,104	126,14	0,091	154,751	0,074	372,405	0,095	2220,86

Σύμφωνα με τις τιμές για τις ανεξάρτητες τ.μ., η ύπαρξη διαφορών, κυρίως για την συνάρτηση ASN, οφείλεται στο γεγονός ότι υπάρχει σφάλμα στην προσέγγιση του Wald για την τιμή της τ.μ. S_T ($S_T \approx a$ ή $S_T \approx b$). Ουσιαστικά η προσέγγιση του Wald δεν λαμβάνει υπ'όψιν τις διαφορές $|S_T - a|$ και $|S_T - b|$ αλλά θεωρεί τα φράγματα σαν φράγματα απορρόφησης.

Παρατηρούμε ότι όταν το ρ απομακρύνεται από το 0.1 τόσο περισσότερες παρατηρήσεις θα χρειαστούμε για να αποφασίσουμε υπέρ ή κατά της \mathcal{H}_0 . Δηλαδή το είδος της συσχέτισης μεταξύ των τ.μ. παίζει καθοριστικό ρόλο στο υπολογισμό του αριθμού ASN. Αυτό αποδεικνύεται και αναλυτικά παίρνοντας την μέση τιμή στην (3.14)

$$\mathbb{E}(\lambda_n) = \frac{(1 - \rho)(\theta_1 - \theta_0)}{(1 + \rho)\sigma^2} \left(\theta - \frac{\theta_0 + \theta_1}{2} \right) \quad (3.15)$$

και αφήνοντας το ρ να τείνει στο 1 και στο -1. Τότε η προσαύξηση λ_n θα γίνεται είτε πολύ μικρή και θα χρειαζόμαστε πολλές παρατηρήσεις, είτε πολύ μεγάλη και θα χρειαζόμαστε λίγες. Επίσης, αποδεικνύεται (Baseville, 1993, σελ 142-143) ότι η συνάρτηση ASN στο AR(1) μοντέλο θα είναι το γινόμενο της ποσότητας $(1 + \rho)/(1 - \rho)$ με την αντίστοιχη συνάρτηση ASN στην i.i.d. περίπτωση.

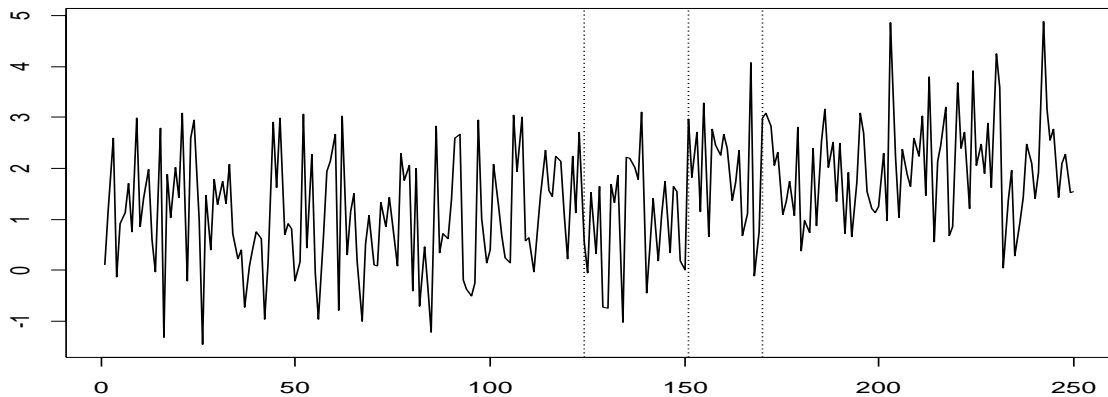
Αλγόριθμος Σωρευτικών Αθροισμάτων

4.1 Γενικά περί στατιστικών αλγορίθμων ανίχνευσης

Οι αλγόριθμοι ανίχνευσης σημείου αλλαγής κατανομής χρησιμοποιήθηκαν αρχικά στον στατιστικό έλεγχο ποιότητας. Κατά τον Ghosh (1991) ο πρώτος ακολουθιακός αλγόριθμος δημιουργήθηκε από τον Shewhart (1931), γνωστός και ως *διάγραμμα ελέγχου του Shewhart*. Σε γενικές γραμμές τα διαγράμματα ελέγχου χρησιμοποιούνται για την επιτήρηση διεργασιών, κυρίως ως προς ένα συνεχές χαρακτηριστικό. Συνήθως εξετάζεται η μέση τιμή της κατανομής του χαρακτηριστικού. Σε κάθε διάγραμμα ελέγχου απεικονίζεται μία στατιστική συνάρτηση και τις περισσότερες φορές, δύο παράλληλες γραμμές, υποδεικνύοντας ότι εάν η τιμή της συνάρτησης βρεθεί πέρα από το διάστημα που ορίζεται από αυτές τις γραμμές, θα υπάρξει ένδειξη μετατόπισης του μέσου. Πολλοί άλλοι αλγόριθμοι κατασκευάστηκαν με σκοπό την ανίχνευση μετατόπισης μέσου ή διασποράς ή και των δύο, οι οποίοι επεκτάθηκαν και σε εφαρμογές πέρα από τον έλεγχο ποιότητας, όπως για παράδειγμα στην στατιστική επεξεργασία σήματος, στην επιτήρηση ιατρικών διαδικασιών, στα συστήματα πλοήγησης κ.α. Εδώ χρειάζεται να επισημάνουμε κάποιες σημαντικές διαφορές. Υπάρχουν δύο είδη στατιστικής ανίχνευσης. Κατά το πρώτο έχουμε συλλέξει όλα τα δεδομένα και μετά αναζητούμε πότε και αν συνέβη κάποια αλλαγή στην κατανομή του χαρακτηριστικού που εξετάζουμε. Η διαδικασία αυτή ονομάζεται *offline*. Στο δεύτερο είδος, συλλέγουμε μία-μία τις παρατηρήσεις και η ανίχνευση γίνεται εκείνη την χρονική στιγμή. Η διαδικασία αυτή ονομάζεται *online* και αποτελεί μία από τις δημοφιλέστερες εφαρμογές της ακολουθιακής στατιστικής. Ενεργό ερευνητικό πεδίο είναι η αντιμετώπιση προβλημάτων *ταχύτατης ανίχνευσης (quickest detection)*, όπου εξετάζεται η ύπαρξη και η μορφή των στατιστικών συναρτήσεων και των κανόνων απόφασης που θα ικανοποιούν ορισμένα κριτήρια βελτιστοποίησης, για τον κατά το δυνατόν ταχύτερο εντοπισμό μικρής μετατόπισης του μέσου.

Για να περιγράψουμε την όλη διαδικασία, θα δανειστούμε κάποιες βασικές έννοιες από το στατιστικό έλεγχο ποιότητας. Υποθέτουμε αρχικά ότι εξετάζουμε εάν υπάρχει μετατόπιση του μέσου της κατανομής. Θέτουμε ένα κατώφλι (ή δύο ανάλογα με το αν θέλουμε να ελέγξουμε και μείωση του μέσου) και όταν η τιμή της στατιστικής συνάρτησης περάσει αυτήν την τιμή, τότε έχουμε ένδειξη «συναγερμού» (*alarm*) ότι ο μέσος της κατανομής άλλαξε, σύμφωνα με την κλασσική ορολογία (π.χ. βλ. Αντζουλάκος, 2012). Διακρίνουμε ότι

υπάρχει περίπτωση να υπάρξει συναγερμός ενώ δεν έχει αλλάξει η μέση τιμή της κατανομής, δηλαδή να έχουμε ένδειξη εσφαλμένου συναγερμού (*false alarm* ή σφάλμα τύπου I), και να μην υπάρξει συναγερμός όταν στην πραγματικότητα υπάρχει μετατόπιση του μέσου (σφάλμα τύπου II). Στην δεύτερη περίπτωση, ενδεχομένως ο αλγόριθμος να εντοπίσει την αλλαγή αλλά σε προχωρημένο στάδιο. Άρα τότε θα υπάρξει κάποια καθυστέρηση στον εντοπισμό του σημείου αλλαγής. Για παράδειγμα ας δούμε το παρακάτω σχήμα.



Σχήμα 4.1 Μετατόπιση μέσου κατανομής

Τα πρώτα 150 σημεία προέρχονται από την κατανομή $N(1,1)$ ενώ τα υπόλοιπα 100 από την κατανομή $N(2,1)$. Εάν οι παράμετροι του αλγορίθμου και κυρίως το κατώφλι, ήταν τέτοια ώστε ο αλγόριθμος να ανιχνεύει όσο το δυνατόν γρηγορότερα κάποια αλλαγή, τότε θα μπορούσαμε να είχαμε ένδειξη συναγερμού για αύξηση της μέσης τιμής, στο σημείο από το που διέρχεται η πρώτη κάθετος αντί για την δεύτερη (δηλαδή αφού δεν έχει αλλάξει η κατανομή, πρόκειται ουσιαστικά για εσφαλμένο συναγερμό). Εάν πάλι, ο αλγόριθμος ήταν τροποποιημένος ώστε να έχει μικρό αριθμό εσφαλμένων συναγερμών, τότε θα μπορούσε να μην εντόπιζε την αλλαγή κατανομής ικανοποιητικά γρήγορα, με αποτέλεσμα να υπήρχε καθυστέρηση αρκετών σημείων, δηλαδή να εντοπιζόταν η αλλαγή κατανομής στο σημείο που διέρχεται η τρίτη κάθετος. Σκοπός των μεθόδων της ακολουθιακής ανάλυσης είναι να ελαχιστοποιήσουμε την μέση καθυστέρηση ανίχνευσης (*mean delay*) όταν ο αναμενόμενος χρόνος μεταξύ εσφαλμένων συναγερμών (*mean time between false alarms*) είναι σταθερός. Θα δούμε στις επόμενες παραγράφους πως ολόκληρη η θεωρία των martingales και του ελέγχου SPRT μπορεί να χρησιμοποιηθεί για την κατασκευή του αλγορίθμου CUSUM, ο οποίος έχει ευρεία χρήση σε διάφορες επιστήμες και είναι ενσωματωμένος σε στατιστικά πακέτα όπως Minitab, Statgraphics κ.α.

4.2 Αλγόριθμος CUSUM

4.2.1 Διαισθητική ερμηνεία

Ο αλγόριθμος *σωρευτικών αθροισμάτων* (*cumulative sum-CUSUM*) προτάθηκε από τον Page (1954), ανακοινώθηκε δηλαδή επτά χρόνια μετά τη δημοσίευση του Wald (1947) περί του SPRT. Ο CUSUM σε αντίθεση με τα κλασσικά διαγράμματα Shewhart λαμβάνει υπ' όψιν του τις προηγούμενες παρατηρήσεις μέσω ενός αθροίσματος, όπως θα δούμε παρακάτω.

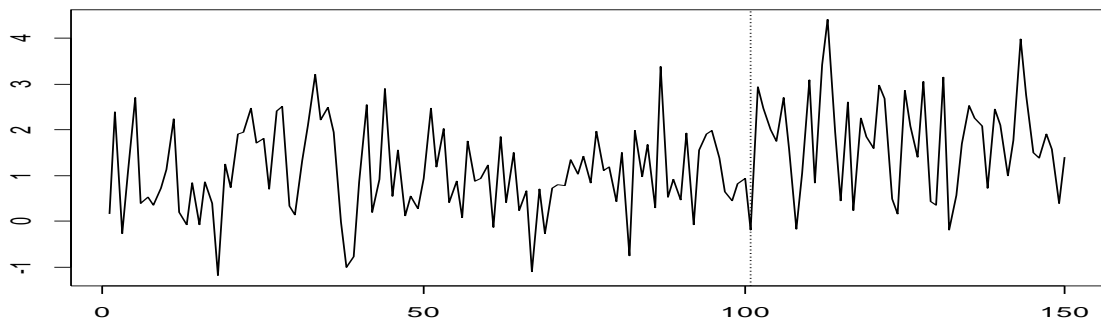
Το πείραμα που θα εξετάσουμε τώρα έχει την εξής μορφή: Λαμβάνουμε δείγμα μεγέθους ένα κάθε φορά, από την κανονική κατανομή $N(\theta, \sigma^2)$ και κάνουμε τον έλεγχο της υπόθεσης (3.1). Υποθέτουμε αρχικά ότι τα θ_0 και θ_1 καθώς και η διακύμανση σ^2 είναι γνωστά. Θεωρούμε τα αθροίσματα

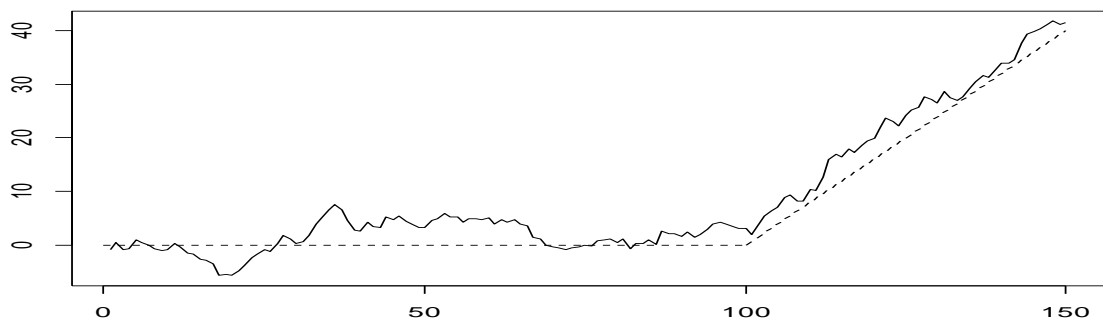
$$S_n = \sum_{i=1}^n (X_i - \theta_0).$$

Όσο ισχύει η \mathcal{H}_0 περιμένουμε τα σημεία (n, S_n) να κινούνται γύρω από τον άξονα xx' γιατί $\mathbb{E}(S_n) = 0$. Εάν στο σημείο με τετμημένη τ_0 η μέση τιμή γίνει $\theta_1 > \theta_0$ τότε

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^{\tau_0-1} (X_i - \theta_0) + \sum_{i=\tau_0}^n (X_i - \theta_0)\right) = (\theta_1 - \theta_0)(n - \tau_0)$$

Άρα τα σημεία (n, S_n) για $n > \tau_0$ θα βρίσκονται γύρω από ευθεία με κλίση $\lambda = (\theta_1 - \theta_0)$. Σημειώνουμε ότι το τ_0 ονομάζεται *σημείο αλλαγής κατανομής* (*change point*).





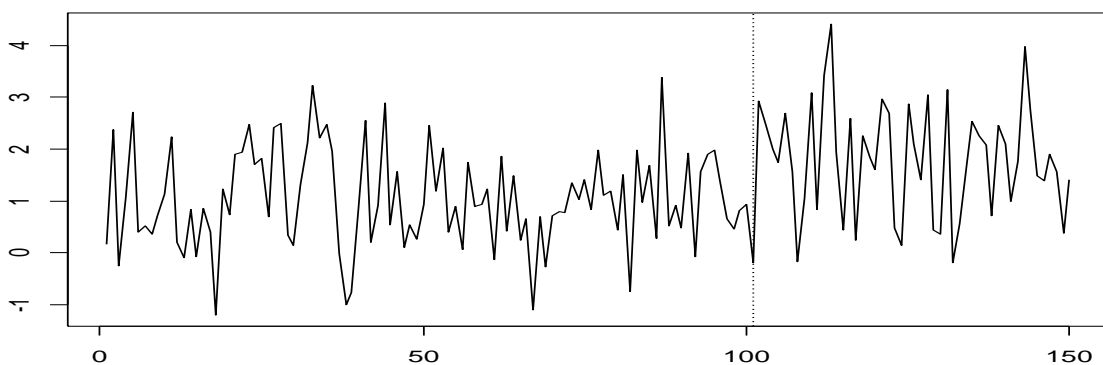
Σχήμα 4.2.1 Συμπεριφορά των σημείων (n, S_n) κατά την μετατόπιση μέσου με σημείο αλλαγής $\tau_0 = 100$.

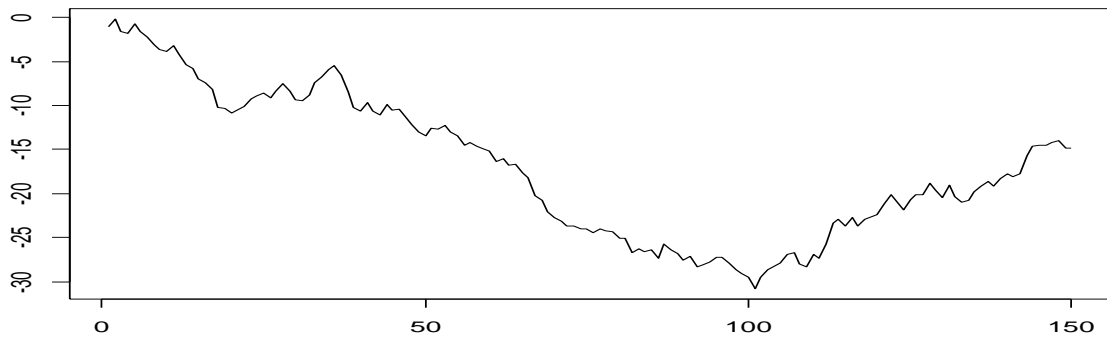
Ας δούμε τώρα πως μπορούμε να εφαρμόσουμε ανάλογα παραδείγματα με χρήση του λόγου των πιθανοφανειών R_n . Συμβολίζουμε όπως και στο τέταρτο κεφάλαιο με $Z = \log\left(\frac{f(X|\theta_1)}{f(X|\theta_0)}\right)$.

Σύμφωνα με τις ιδιότητες τις απόκλισης K-L (Παράρτημα (I)) επειδή

$$\mathbb{E}_{\theta_0}(Z) < 0 \text{ και } \mathbb{E}_{\theta_1}(Z) > 0$$

θα περιμέναμε ότι, μετά από μία μετατόπιση της μέσης τιμής της κατανομής f , το άθροισμα $S_n = \log(R_n) = \sum_{i=1}^n Z_i$ θα αρχίσει να αυξάνεται. Ας είναι τ_0 το σημείο μετατόπισης του μέσου που εντοπίζουμε με χρήση των στατιστικών συναρτήσεων. Από το Παράδειγμα (3.3.1) παίρνουμε ότι $Z_i = \frac{\theta_1 - \theta_0}{\sigma^2} \left(X_i - \frac{\theta_1 + \theta_0}{2} \right)$.





Σχήμα 4.2.2 Συμπεριφορά των σημείων (n, S_n) κατά την μετατόπιση μέσου με τη χρήση του λόγου πιθανοφανειών

Ο Page (1954) πρότεινε τη χρήση ενός κατωφλιού $h > 0$, που θα ορίζεται από τον χρήστη, τέτοιο ώστε κάθε φορά που η ποσότητα

$$G_n = S_n - \min_{1 \leq k < n} S_k$$

γίνεται μεγαλύτερη ή ίση του h , θα έχουμε ένδειξη μετατόπισης του μέσου. Ορίζουμε το χρόνο στάσης

$$T = \inf \{n \in \mathbb{N} : G_n \geq h\}, \quad \inf \emptyset = \infty.$$

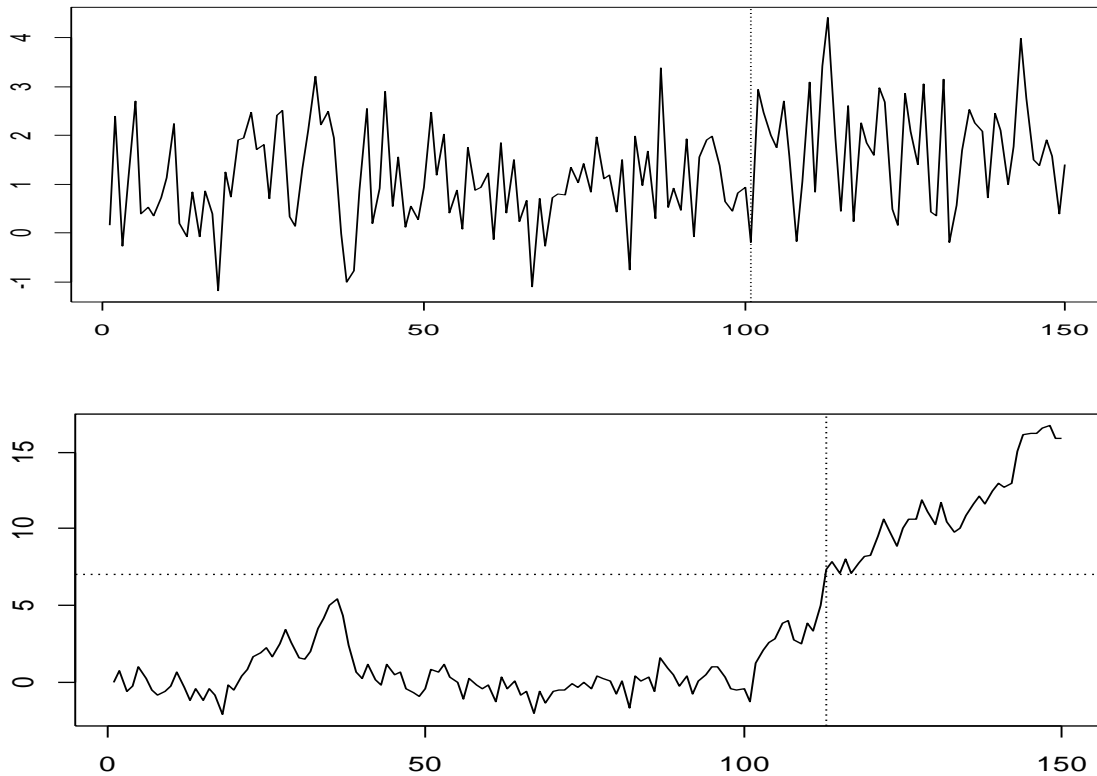
Για την υλοποίηση του αλγορίθμου στην R θέτουμε ότι $h = 7$ (ο τρόπος με τον οποίο θα υπολογίζουμε τη σταθερά h , θα αναφερθεί στην Παράγραφο 4.3.3) τα πρώτα 100 σημεία προέρχονται από την κατανομή $N(\theta_0, \sigma^2)$, όπου $\theta_0 = 1$ και $\sigma = 1$, ενώ τα επόμενα 50 από την $N(\theta_1, \sigma^2)$ όπου $\theta_1 = 1.8$. Υπολογίζουμε το άθροισμα $S_n = \sum_{i=1}^n Z_i$, αποθηκεύουμε τις τιμές σε ένα πίνακα S και τις τιμές $S_n - \min_{1 \leq k < n} S_k$ σε έναν πίνακα G . Προφανώς η εκτίμηση του T θα γίνει όταν κάποιο στοιχείο του G ξεπεράσει για πρώτη φορά την τιμή $h = 7$. Σε έναν πίνακα m αποθηκεύουμε όλες τις τιμές i για τις οποίες $G[i] > h$ και βρίσκουμε το ελάχιστο του, που θα είναι το T .

```
> set.seed(5)
> theta0 <-1 ; theta1 <-1.8;sigma<-1
> S<-matrix();X1<-rnorm(100,mean=theta0,sd=1);X2<-rnorm(50,mean=theta1,sd=1)
> X<-c(X1,X2)
> S[1] <- (theta1-theta0)/sigma^2*(X[1]-(theta1+theta0)/2)
> for(i in 1:99) {
+ S[i+1]<-S[i]+(theta1-theta0)/sigma^2*(X[i+1]-(theta1+theta0)/2)
+ }
> for(i in 100:149) {
+ S[i+1] <- S[i]+(theta1-theta0)/sigma^2*(X[i+1]-(theta1+theta0)/2)
+ }
> par(mar=c(3,3,2,2));index<-1:150
> plot(index,X, type="l", lwd=1, pch=16,ylab="",xlab="")
> abline(v=101,lty=3)
> h<-7;k<-1;m<-c();G<-c();G[1]<-0
```

```

> for(i in 2:length(S)) {
+   G[i]<-S[i]-min(S[seq(1,i-1,1)])
+   if(G[i]>h) {
+     m[k]<-i
+     k<-k+1
+   }
+ }
> plot(G,type="l");abline(h=h,lty=3);abline(v=min(m),lty=3)
> T<-min(m)
> for(i in 1:length(S)){
+   if( S[i]==min(S[seq(1,T-1,1)]) ) {
+     t0hat<-i+1
+   }
+ }
> print(t0hat)
[1]102

```



Σχήμα 4.2.3 Συμπεριφορά των σημείων (n, G_n) κατά την μετατόπιση μέσου. Η οριζόντια γραμμή δηλώνει το κατώφλι h

Για να εκτιμήσουμε το κατεξοχήν σημείο αλλαγής τ_0 , σύμφωνα με το σχήμα 4.2.3 υπολογίζουμε την ποσότητα

$$\hat{\tau}_0 = \underset{1 \leq k < T}{\operatorname{argmin}} S_k + 1$$

Στη συγκεκριμένη περίπτωση εκτιμήσαμε ότι ο χρόνος αλλαγής κατανομής είναι $\hat{\tau}_0 = 102$, αφού η ελάχιστη τιμή του αθροίσματος S_k επιτυγχάνεται όταν $k = 101$. Συνεπώς, έχουμε καθυστερήσει για σήμανση συναγεραμού μόνο $\hat{\tau}_0 - \tau_0 = 102 - 101 = 1$ παρατήρηση.

4.2.2 Ο αλγόριθμος CUSUM ως επαναλαμβανόμενος έλεγχος SPRT

Ένας δεύτερος τρόπος που όρισε ο Page (1954) τον αλγόριθμο CUSUM, είναι με χρήση του SPRT. Ο CUSUM, για ανίχνευση μετατόπισης (αύξησης) του μέσου θ , μπορεί να υλοποιηθεί ως επαναλαμβανόμενος έλεγχος SPRT με κάτω φράγμα το μηδέν και άνω φράγμα το h , υπό την εξής έννοια: κάθε φορά που το σωρευτικό άθροισμα S_n γίνεται μικρότερο του μηδέν τότε ξεκινάει ο αλγόριθμος από την αρχή μηδενίζοντας το S_n . Ωστε ο CUSUM θα περιγράφεται από τη σχέση

$$S_n = (S_{n-1} + Z_n)^+$$

όπου $S_0 = 0$, $(x)^+ = \max(x, 0)$, και τα Z_i όπως ορίστηκαν στην Παράγραφο 3.3. Ο χ .σ. τώρα θα είναι

$$T = \inf \{n \in \mathbb{N} : S_n \geq h\}, \quad \inf \emptyset = \infty. \quad (4.1)$$

Όπως και στο προηγούμενο πρόγραμμα τα πρώτα 100 σημεία προέρχονται από την κατανομή $N(\theta_0, \sigma^2)$, όπου $\theta_0 = 1$ και $\sigma = 1$, ενώ τα επόμενα 50 από την $N(\theta_1, \sigma^2)$ όπου $\theta_1 = 1.8$. Χρησιμοποιούμε ως κατώφλι $h = 7$. Υπολογίζουμε το άθροισμα S_n όσο βρίσκεται εντός του διαστήματος $(0, h)$. Εάν γίνει μικρότερο ή ίσο του μηδέν, τότε θέτουμε $S_n = 0$ και αποθηκεύουμε το σημείο n σε ένα πίνακα L (το πλήθος των σημείων του πίνακα L θα είναι και ο αριθμός των φορών που επανεκκίνησε ο αλγόριθμος). Εάν $S_n \geq h$ τοποθετούμε το n σε ένα πίνακα k .

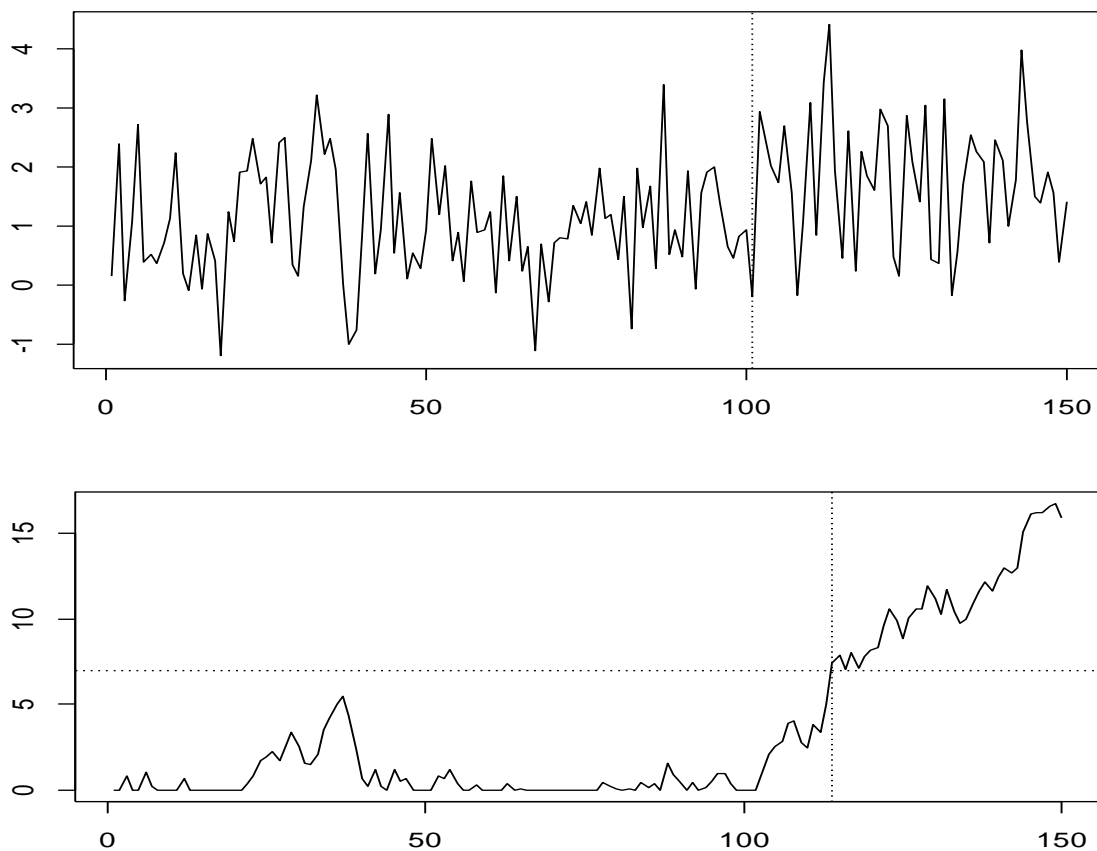
```
> set.seed(5)
> theta0 <-1;thetal<-1.8;sigma<-1
> S <-c();Z<-c();L<-c();k<-c();j<-1;l<-1;Z[1]<-0;h <-7
> X1<-rnorm(100,mean=theta0,sd=1);X2<-rnorm(50,mean=thetal,sd=1);X<-c(X1,X2)
> S[1]<-0;
> Z[1]<-(thetal-theta0)/sigma^2*(X[1]-(thetal+theta0)/2)
> S[2]<-S[1]+Z[1]
> for(i in 2:149) {
+   if((S[i]>0) & (S[i]<h)) {
+     Z[i]<-(thetal-theta0)/sigma^2*(X[i]-(thetal+theta0)/2)
+     S[i+1]<-S[i]+Z[i]
+   }
+   if(S[i]<=0) {
+     S[i]<-0
+     Z[i]<-(thetal-theta0)/sigma^2*(X[i]-(thetal+theta0)/2)
+     S[i+1]<-S[i]+Z[i]
+     L[j] <- i
+     j <-j+1
+   }
+   if(S[i]>=h) {
+     k[l]<-i
```



```

+ l<-l+1
+ Z[i]<-(theta1-theta0)/sigma^2*(X[i]-(theta1+theta0)/2)
+ S[i+1]<-S[i]+Z[i]
+ }
+ }
> par(mar=c(3,3,2,2)); plot(S,type="l",ylab="",xlab="")
> abline(v=k[1],lty=3);abline(h=h,lty=3)
> for(i in 1:length(S)){
+ if( S[i]==min(S[seq(1,T-1,1)]) ) {
+ t0hat<-i+1
+ }
+ }
> print(t0hat);print(length(L)) # σημείο αλλαγής και αριθμός φορών που ο SPRT
επανεκκίνησε
[1] 103
[1]48

```



Σχήμα 4.2.4 Συμπεριφορά των σημείων (n, S_n) κατά την μετατόπιση του μέσου, με χρήση του SPRT

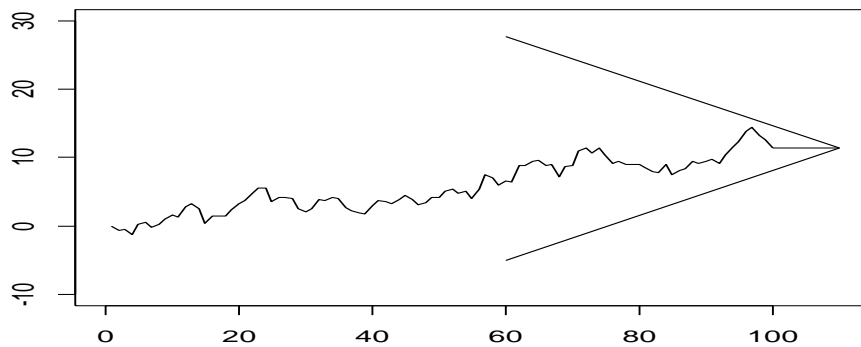
Σύμφωνα με τα αποτελέσματα από το πρόγραμμα R, πήραμε ότι το σημείο αλλαγής \hat{t}_0 θα είναι ίσο με 103 καθώς και ότι ο αριθμός των φορών που επανεκκίνησε ο αλγόριθμος CUSUM είναι 48.

4.2.3 Μέθοδος V-mask

Έστω τώρα ότι θέλουμε να εξετάσουμε τον έλεγχο της υπόθεσης H_0 έναντι δύο εναλλακτικών

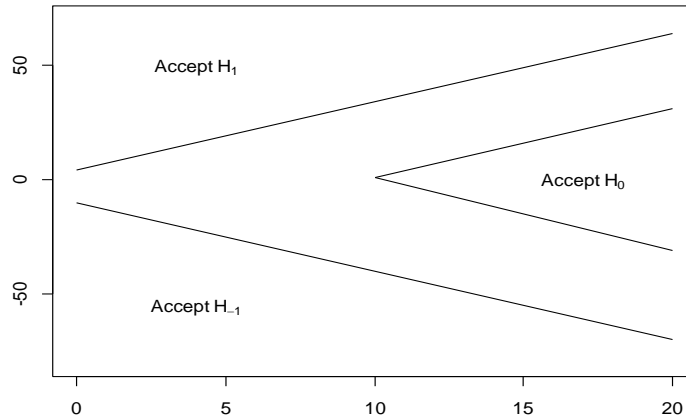
$$H_{-1}: \theta = \theta_0 - \delta\sigma - H_0: \theta = \theta_0 - H_1: \theta = \theta_0 + \delta\sigma \quad (4.2)$$

με $\delta > 0$. Ο Barnard (1959) είχε υποδείξει έναν γεωμετρικό τρόπο για τον έλεγχο μετατόπισης του μέσου ως εξής: απεικονίζουμε στο διάγραμμα τα σωρευτικά αθροίσματα $S_n = \sum_{i=1}^n (X_i - \theta)$, $1 \leq n \leq m$. Από το m -στο σημείο που είναι και το τελευταίο, τοποθετούμε σε απόσταση d ένα σημείο O , και από το σημείο O φέρνουμε δύο ευθείες, όπου η κάθε μία από αυτές σχηματίζει γωνία φ με τον άξονα $x'x$. Εάν κάποιο από τα σημεία βρεθεί πάνω από την άνω ευθεία τότε θα έχουμε ένδειξη ότι η μέση τιμή της αυξήθηκε σε σχέση με την αρχική τιμή της $\mu = \mu_0$ ενώ αν βρεθεί πέρα από την κάτω ευθεία θα έχουμε ένδειξη ότι μειώθηκε. Με αυτόν τον τρόπο μπορούμε να ελέγξουμε όλα τα σημεία, ξεκινώντας από το τελευταίο.



Σχήμα 4.2.5 Περιγραφή του CUSUM ως V-mask.

Φαίνεται λογικό ότι οι παράμετροι φ και d θα σχετίζονται με τις πιθανότητες σφαλμάτων I και II αλλά και με την παράμετρο δ . Ο Johnson (1962) πρότεινε μία παραλλαγή του ελέγχου του Armitage (1950). Ο έλεγχος αυτός αποτελεί την ταυτόχρονη εφαρμογή δύο ελέγχων SPRT για τον έλεγχο της υπόθεσης H_0 έναντι της H_{-1} και της H_0 έναντι της H_1 .



Σχήμα 4.2.6 Έλεγχος του Armitage

Εάν λοιπόν δούμε το σχήμα 4.2.5 αντίστροφα, παρατηρούμε ότι μπορεί να ερμηνευθεί όπως ο δίπλευρος έλεγχος του Armitage, με την προϋπόθεση ότι δεν θα αποδεχθούμε ποτέ την H_0 (δηλαδή για $\beta = 0$) και παίρνοντας τις παρατηρήσεις ξεκινώντας από την τελευταία

$$Y_1 = \sigma^{-1}(X_n - \theta_0), Y_2 = \sigma^{-1}(X_{n-1} - \theta_0), \dots, Y_n = \sigma^{-1}(X_1 - \theta_0)$$

Τότε ο λόγος πιθανοφανειών, για τον πρώτο έλεγχο θα είναι ίσος με

$$R_n = \frac{\left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n (Y_i + \delta)^2\right)}{\left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n Y_i^2\right)} = \exp\left(\delta \sum_{i=1}^n Y_i + \frac{n\delta^2}{2}\right)$$

Εφαρμόζοντας λοιπόν την μεθοδολογία που αναπτύχθηκε στο Κεφάλαιο 3 παίρνουμε μετά από πράξεις ότι θα αποδεχόμαστε την H_{-1} όταν $\sum_{i=1}^n Y_i \leq -\frac{n\delta}{2} - \frac{1}{\delta} \log\left(\frac{1-\beta}{\alpha}\right)$, ενώ όταν $\beta = 0$ θα έχουμε $\sum_{i=1}^n Y_i \leq -\frac{n\delta}{2} + \frac{1}{\delta} \log(\alpha)$. Όμοια θα αποδεχόμαστε την H_1 όταν $\sum_{i=1}^n Y_i \geq \frac{n\delta}{2} - \frac{1}{\delta} \log(\alpha)$. Οι ευθείες $\varepsilon_1: y_1 = -\frac{x\delta}{2} + \frac{1}{\delta} \log(\alpha)$ και $\varepsilon_2: y_2 = \frac{x\delta}{2} - \frac{1}{\delta} \log(\alpha)$ θα τέμνονται στο σημείο $\left(\frac{2}{\delta^2} \log(\alpha), 0\right)$, άρα εάν συμβολίσουμε με d την απόσταση από την αρχή $(0,0)$ και φ την κλίση της ευθείας ε_1 με τον x' έχουμε ότι $\tan\varphi = \frac{\delta}{2}$. Καταλήξαμε ότι η μέθοδος που πρότεινε ο Barnard (1959), μπορεί να υλοποιηθεί με χρήση μεθόδων της ακολουθιακής ανάλυσης. Πολλά στατιστικά πακέτα περιέχουν ενσωματωμένες ρουτίνες για την υλοποίηση του αλγορίθμου CUSUM και της V-mask, για δοσμένες παραμέτρους.

4.2.4 H-K μέθοδος

Ο Page (1954) είχε κάνει χρήση μίας ακόμα μορφής του αλγορίθμου CUSUM για τον δίπλευρο έλεγχο υπόθεσης (4.2). Ορίζουμε δύο σωρευτικά αθροίσματα

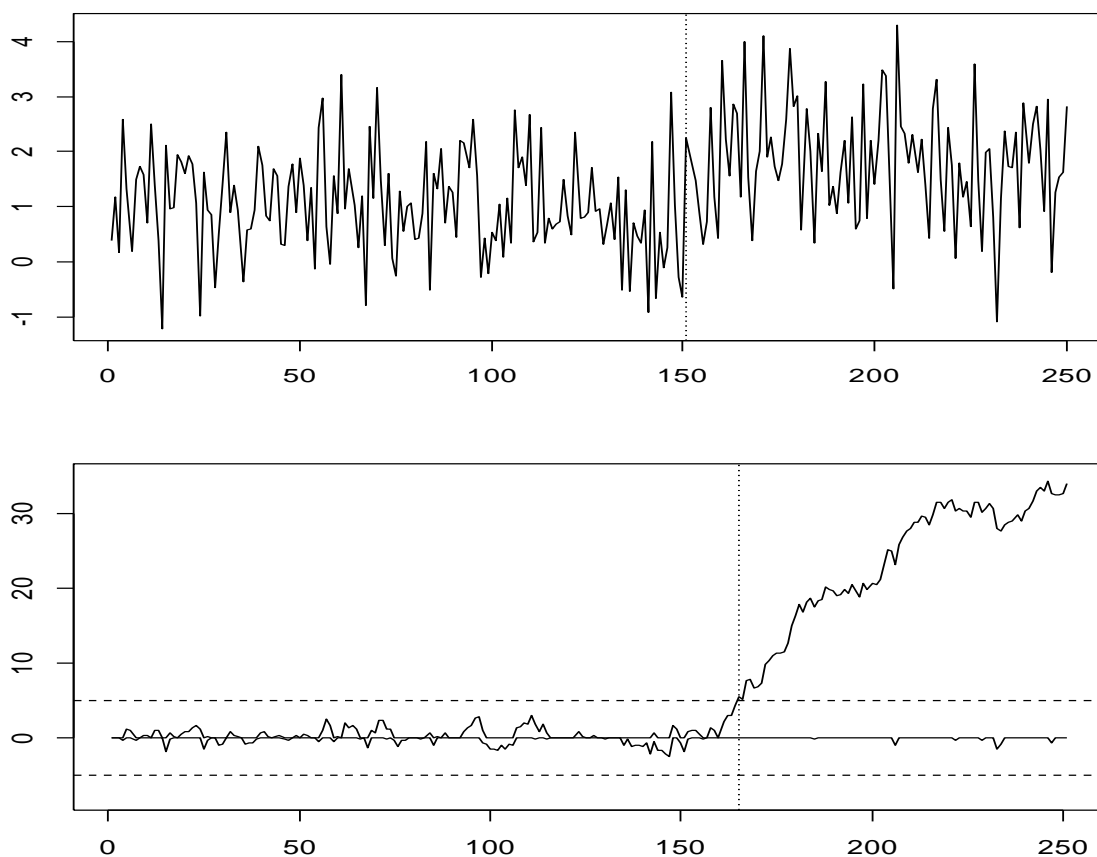
$$C_n^+ = \max\{0, X_n - (\theta_0 + K) + C_{n-1}^+\}, \quad C_0^+ = 0$$

$$C_n^- = \min\{0, X_n - (\theta_0 - K) + C_{n-1}^-\}, \quad C_0^- = 0$$

όπου $K = k\sigma$ και $k = \frac{\delta}{2}$. Έστω ότι $H = h\sigma$. Θα αποδεχόμαστε την \mathcal{H}_1 όταν $C_n^+ > H$ και θα αποδεχόμαστε την \mathcal{H}_{-1} όταν $C_n^- < -H$. Σύμφωνα με τη βιβλιογραφία οι παράμετροι K και H θα ονομάζονται *τιμή αναφοράς (reference value)* και *διάστημα απόφασης (decision interval)* αντίστοιχα. Στο διάγραμμα παρουσιάζονται και τα δύο σωρευτικά αθροίσματα C_n^+ , C_n^- .

Τα πρώτα 150 σημεία προέρχονται από την κατανομή $N(\theta_0, \sigma^2)$, όπου $\theta_0 = 1$ και $\sigma = 1$, ενώ τα επόμενα 100 από την $N(\theta_1, \sigma^2)$ όπου $\theta_1 = 1.8$. Χρησιμοποιούμε ως $H = 5$ και ως τιμή αναφοράς $K = 0.5$. Υπολογίζουμε τα αθροίσματα C_n^+ και C_n^- σύμφωνα με τους παραπάνω τύπους. Εάν $C_n^- < -H$ ή $C_n^+ > H$ αποθηκεύουμε το n σε ένα πίνακα f (Προφανώς το πρώτο στοιχείο του πίνακα που είναι και το ελάχιστο, αποτελεί το πρώτο σημείο εκτός του διαστήματος $(-H, H)$). Εκτιμήσαμε ότι $n = 165$.

```
> set.seed(1)
> theta0<-1;theta1<-1.8;sigma<-1;K<-k*sigma;delta<-1;k<-delta/2;h<-5;H<-h*sigma
> Cnplus<-c();Cnminus<-c();X<-c()
> i<-1;d<-0
> X0<-rnorm(150,mean=theta0,sd=sigma); X1<-rnorm(100,mean=theta1,sd=sigma)
> X<-c(X0,X1);plot(X,type="l");abline(v=151,lty=3)
> Cnplus[1]<-0;Cnminus[1]<-0;j<-1;f<-c()
> for(i in 1:length(X)) {
+ Cnplus[i+1]<-max(0,X[i]-(theta0+K)+Cnplus[i])
+ Cnminus[i+1]<-min(0,X[i]-(theta0-K)+Cnminus[i])
+ if((Cnplus[i+1]>H) | (Cnminus[i+1]<(-H))) {
+ f[j]<-i+1
+ j<-j+1
+ }
+ }
> par(mar=c(3,3,2,2))
> plot(Cnplus,type="l",main="",ylim=c(-8,35),xlab="",ylab="")
> lines(Cnminus,type="l");
> abline(h=H,lty=2);abline(h=-H,lty=2);abline(v=min(f),lty=3)
> print(min(f)) # πρώτο σημείο έξω από το διάστημα (-H,H)
[1]165
```



Σχήμα 4.2.7 Η-Κ μέθοδος CUSUM για την ανίχνευση μετατόπισης της μέσης τιμής όταν $k = 0.5$ και $h = 5$. Οι οριζόντιες γραμμές δείχνουν τα άνω και κάτω κατώφλια H και $-H$ αντίστοιχα

Αυτή η μέθοδος ενσωματώνεται στα στατιστικά πακέτα για την υλοποίηση του αλγορίθμου CUSUM και αποτελεί έναν ανάλογο δίπλευρο διάγραμμα ελέγχου όπως το κλασικό διάγραμμα του Shewhart.

4.3 Συνάρτηση ARL

Θα χρειαστούμε κάποιους βασικούς ορισμούς αντίστοιχους με την έννοια του μέσου αριθμού δείγματος από το Κεφάλαιο 3 οι οποίοι χρησιμοποιούνται για την αξιολόγηση των στατιστικών αλγορίθμων ανίχνευσης. Έστω τ η τ.μ. που δείχνει τον αριθμό των σημείων που απεικονίζονται σε ένα διάγραμμα ελέγχου μέχρι και την ένδειξη συναγερμού.

Ορισμός 4.3.1 Θα ονομάζουμε μέσο μήκος ροής (average run length-ARL) ενός στατιστικού αλγορίθμου ανίχνευσης, την αναμενόμενη τιμή της τ.μ. τ :

$$ARL = \mathbb{E}_{\theta}(\tau).$$

Σημειώνουμε στα πλαίσια του ελέγχου (3.1) ότι διακρίνουμε δύο περιπτώσεις:

(i) Όταν $\theta = \theta_0$ τότε η ποσότητα $ARL_0 = \mathbb{E}_{\theta_0}(\tau)$ δηλώνει τον αναμενόμενο αριθμό σημείων πριν από μία εσφαλμένη σήμανση συναγερμού (*false alarm*). Ισοδύναμα μπορούμε να πούμε ότι είναι ο αναμενόμενος χρόνος μεταξύ εσφαλμένων συναγερμών.

(ii) Όταν $\theta = \theta_1$ τότε η ποσότητα $ARL_1 = \mathbb{E}_{\theta_1}(\tau)$ δηλώνει τον αναμενόμενο αριθμό σημείων μέχρι την ανίχνευση της μετατόπισης. Προφανώς αυτή η ποσότητα ισοδυναμεί με τη μέση καθυστέρηση ανίχνευσης (*mean detection delay*).

Θα χρησιμοποιήσουμε αυτές τις δύο έννοιες στην Παράγραφο 4.3.3 για να επιλέξουμε μία ικανοποιητική τιμή του h . Σημειώνουμε ότι γενικά για έναν αλγόριθμο στατιστικής ανίχνευσης επιθυμούμε να έχουμε μεγάλο ARL_0 και μικρό ARL_1 . Αρχικά όμως θα βρούμε αναλυτικές μεθόδους για τον υπολογισμό του ARL_θ για οποιαδήποτε τιμή του θ . Θα θεωρούμε τον CUSUM με την μορφή του επαναλαμβανόμενου ελέγχου SPRT (βλ. Παράγραφο 4.2.2).

Συμβολίζουμε με:

z : την αρχική τιμή της στατιστικής συνάρτησης του αλγορίθμου CUSUM

$P(z)$: την πιθανότητα μία ροή που ξεκινά από το z , να τερματίσει στο κάτω φράγμα

$N(z)$: το μέσο μήκος μίας ροής που ξεκινά με αρχική τιμή z

$N_1(z)$: το μέσο μήκος ροής που τερματίζει στο 0

$N_2(z)$: το μέσο μήκος ροής που τερματίζει στο h

$L(z)$: το μέσο μήκος ροής του αλγορίθμου CUSUM

τ_k : την τ.μ. που δηλώνει τον αριθμό των ελέγχων SPRT (βλ. Παράγραφο 4.2.2, όπου $\tau_k = 48$) όπου το άθροισμα S_n κατέληξε στο κάτω φράγμα 0

$T_{\varepsilon,h}$: τον χρόνο στάσης (3.4) με κάτω φράγμα το ε και άνω το h

Παρατηρούμε ότι η τ_k ακολουθεί γεωμετρική κατανομή με παράμετρο $P(0)$ και επομένως

$$\mathbb{E}(\tau_k) = \frac{P(0)}{1 - P(0)},$$

και

$$ARL = L(0) = \frac{P(0)}{1 - P(0)} \cdot N_1(0) + N_2(0)$$

$$= \frac{P(0)N_1(0) + (1 - P(0))N_2(0)}{1 - P(0)} = \frac{N(0)}{1 - P(0)} \quad (4.3)$$

Για τον υπολογισμό των ποσοτήτων $N(0)$ και $P(0)$ έχουν προταθεί διάφοροι τρόποι. Οι συναρτήσεις $P(z)$ και $N(z)$ ικανοποιούν τις εξισώσεις του Fredholm δευτέρου είδους (Page, 1954)

$$P(z) = \int_{-\infty}^{-z} f(x)dx + \int_0^h P(x)f(x-z)dx$$

και

$$N(z) = 1 + \int_0^h N(x)f(x-z)dx$$

όπου η f είναι η σ.π.π. της τ.μ. $Z = \log\left(\frac{f(X|\theta_1)}{f(X|\theta_0)}\right)$.

Οι εξισώσεις αυτές αν και δεν λύνονται αναλυτικά μπορούν λυθούν με αριθμητική προσέγγιση σύμφωνα με τη μέθοδο του Γκαουσιανού τετραγωνισμού (*Gaussian quadrature*), υπό την προϋπόθεση ότι η f ικανοποιεί κάποιες συνθήκες ομαλότητας (βλ. Baseville, 1993, σελ. 169). Έχει διαπιστωθεί ότι η μέθοδος αυτή δίνει αρκετά ικανοποιητική προσέγγιση του ARL (βλ. Granjon, 2012, σελ. 13) αλλά απαιτεί σύνθετους υπολογισμούς.

4.3.1 Προσεγγίσεις των Wald και Siegmund

Θα χρησιμοποιήσουμε τώρα τα αποτελέσματα από τη θεωρία του Wald, στο Κεφαλαίο 3. Σύμφωνα με τους συμβολισμούς,

$$L(0) = \frac{\mathbb{E}(T_{0,h})}{1 - P(0)} \quad (4.4)$$

και επιπλέον όταν $\theta \neq \tilde{\theta}$ (το $\tilde{\theta}$ είναι τέτοιο ώστε $\mathbb{E}_{\tilde{\theta}}(Z) = 0$), θα ισχύει

$$\mathbb{E}_{\theta}(T_{\varepsilon,h}) \approx \frac{\varepsilon\mathbb{P}(S_T \leq \varepsilon) + h(1 - \mathbb{P}(S_T \leq \varepsilon))}{\mathbb{E}_{\theta}(Z)}$$

$$\mathbb{P}(S_T \leq \varepsilon) \approx \frac{e^{-t_0(\theta)h} - 1}{e^{-t_0(\theta)h} - e^{-t_0(\theta)\varepsilon}} \quad (4.5)$$

όπου το t_0 είναι λύση της εξίσωσης

$$\mathbb{E}_{\theta}(e^{-t_0(\theta)Z}) = 1.$$

Επειδή η αντικατάσταση του τύπου (4.5) για $\varepsilon = 0$ στον (4.4) θα δώσει απροσδιόριστη ποσότητα ($P(0) = \mathbb{P}(S_T \leq 0) = 1$), θα πάρουμε το όριο καθώς το ε τείνει στο μηδέν. Αντικαθιστώντας την σχέση (4.5) στον τύπο $\mathbb{E}_\theta(T_{\varepsilon,h})$ παίρνουμε ότι

$$\mathbb{E}_\theta(T_{\varepsilon,h}) \approx \frac{1}{\mathbb{E}_\theta(Z)} \left(h + \varepsilon \frac{1 - e^{-t_0(\theta)h}}{1 - e^{t_0(\theta)\varepsilon}} \right)$$

και όταν $\varepsilon \rightarrow 0$,

$$\mathbb{E}_\theta(T_{0,h}) \approx \frac{1}{\mathbb{E}_\theta(Z)} \left(h + \frac{e^{-t_0(\theta)h}}{t_0(\theta)} - \frac{1}{t_0(\theta)} \right).$$

Εάν τώρα $\theta = \tilde{\theta}$,

$$\mathbb{E}_{\tilde{\theta}}(T_{\varepsilon,h}) \approx \frac{\varepsilon^2 \mathbb{P}(S_T \leq \varepsilon) + h^2 (1 - \mathbb{P}(S_T \leq \varepsilon))}{\mathbb{E}_{\tilde{\theta}}(Z^2)}$$

και

$$\mathbb{P}(S_T \leq \varepsilon) \approx \frac{h}{h + \varepsilon}.$$

Με αντικατάσταση του $\mathbb{P}(S_T \leq \varepsilon)$ στον τύπο του $\mathbb{E}_{\tilde{\theta}}(T_{\varepsilon,h})$ και όταν $\varepsilon \rightarrow 0$,

$$\mathbb{E}_{\tilde{\theta}}(T_{0,h}) = \frac{h^2}{\mathbb{E}_{\tilde{\theta}}(Z^2)}$$

Άρα καταλήγουμε ότι

$$\text{ARL}_\theta \approx \begin{cases} \frac{1}{\mathbb{E}_\theta(Z)} \left(h + \frac{e^{-t_0(\theta)h}}{t_0(\theta)} - \frac{1}{t_0(\theta)} \right), & \theta \neq \tilde{\theta} \\ \frac{h^2}{\mathbb{E}_\theta(Z^2)}, & \theta = \tilde{\theta}. \end{cases}$$

Ο Siegmund (1985) είχε προτείνει μια παραλλαγή του παραπάνω τύπου. Ουσιαστικά, ο Wald (1947) δεν λάμβανε υπ' όψιν του τις διαφορές $|S_T - h|$, $|S_T - 0|$ ενώ ο Siegmund, κάνοντας χρήση της θεωρίας προσεγγίσεων διαχύσεων (*diffusion approximation theory*) εκτίμησε αυτήν την διαφορά.

Αν συμβολίσουμε

$$\rho_+ = \mathbb{E}_\theta(S_T - h | S_T - h \geq 0)$$

$$\rho_- = \mathbb{E}_\theta(S_T | S_T \leq 0)$$

τότε αποδεικνύεται ότι (βλ. Baseville, 1993, σελ. 174-175)

$$\text{ARL}_\theta \approx \begin{cases} \frac{1}{\mathbb{E}_\theta(Z)} \left(h + \rho_+ - \rho_- + \frac{e^{-t_0(\theta)(h+\rho_+-\rho_-)}}{t_0(\theta)} - \frac{1}{t_0(\theta)} \right), & \theta \neq \tilde{\theta} \\ \frac{(h + \rho_+ - \rho_-)^2}{\mathbb{E}_\theta(Z^2)}, & \theta = \tilde{\theta} \end{cases}$$

Πρόκειται λοιπόν για την εκτίμηση του Wald όπου στη θέση του h είναι τώρα το $h + \rho_+ - \rho_-$. Στην περίπτωση της κανονικής κατανομής ισχύει ότι

$$\rho_+ - \rho_- = -\frac{2}{\pi} \int_0^{+\infty} x^{-2} \log \left[\frac{2}{x^2} \left(1 - e^{-\frac{1-x^2}{2}} \right) \right] dx \approx 2 \cdot 0.583 = 1.166$$

4.3.2 Σύγκριση συναρτήσεων ARL

Θα εξετάσουμε τώρα την αποδοτικότητα των τύπων του ARL για διάφορες τιμές του θ . Εφαρμόζουμε τον αλγόριθμο CUSUM, ως επαναλαμβανόμενο έλεγχο SPRT, για $\theta_0 = 1$, $\theta_1 = \theta_0 + \delta\sigma$, $\sigma = 1$, $\delta = 1$ και $h = 4$. Σύμφωνα με το Κεφάλαιο 3 υπολογίζουμε ότι

$$\mathbb{E}_\theta(Z) = \theta - 1.5, \quad \mathbb{E}_{1.5}(Z^2) = 1, \quad t_0(\theta) = 2\theta - 3.$$

Επομένως η συνάρτηση ARL θα δίνεται από τις σχέσεις

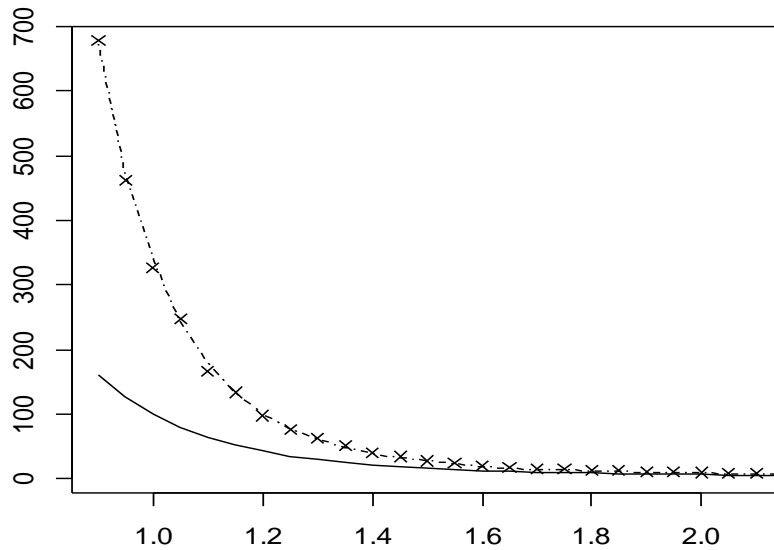
$$\text{ARL}_\theta \approx \begin{cases} \frac{1}{(\theta - 1.5)} \left(h + \frac{e^{-(2\theta-3)h}}{2\theta - 3} - \frac{1}{2\theta - 3} \right), & \theta \neq 1.5 \\ h^2, & \theta = 1.5 \end{cases}$$

$$\text{ARL}_\theta \approx \begin{cases} \frac{1}{(\theta - 1.5)} \left(h + 1.166 + \frac{e^{-(2\theta-3)(h+1.66)}}{2\theta - 3} - \frac{1}{2\theta - 3} \right), & \theta \neq 1.5 \\ (h + 1.166)^2, & \theta = 1.5 \end{cases}$$

σύμφωνα με τις προσεγγίσεις των Wald και Siegmund αντίστοιχα.

Υλοποιούμε τους παραπάνω τύπους στην R (βλ. Παράρτημα IV, Πρόγραμμα 8), για να δούμε τις διαφορές μεταξύ των δύο προσεγγίσεων των συναρτήσεων ARL, ως προς θ , αλλά και μέσω προσομοίωσης. Δημιουργούμε έναν βρόχο επανάληψης για διάφορες τιμές του θ . Οι τιμές των συναρτήσεων προκύπτουν απλά, εξετάζοντας μόνο την περίπτωση εάν $\theta = 1.5$ ή $\theta \neq 1.5$.

Για τις προσομοιωμένες τιμές, χρησιμοποιούμε έναν βρόχο επανάληψης while, εξετάζοντας εάν το άθροισμα $S_n = \sum_{i=1}^n Z_i$ είναι μικρότερο από το $h = 4$. Εάν για κάποια από αυτές τις τιμές του S_n ισχύει ότι $S_n \leq 0$, τότε θέτουμε $S_n = 0$, σύμφωνα με τον ορισμό του αλγορίθμου CUSUM.



Σχήμα 4.3.1 Συνάρτηση ARL για τις διάφορες τιμές του θ με προσομοίωση (x), με την προσέγγιση του Siegmund (διακεκομμένη γραμμή) και με την προσέγγιση του Wald (συνεχής γραμμή)

Παρατηρούμε ότι η προσέγγιση του Siegmund είναι σε γενικές γραμμές καλύτερη από αυτήν του Wald και ειδικά για τιμές του θ μικρότερες του $\tilde{\theta}$.

Θα συγκρίνουμε τώρα τον αλγόριθμο CUSUM και το διάγραμμα Shewhart για μεμονωμένες παρατηρήσεις όταν η διασπορά είναι γνωστή.

Αποδεικνύεται ότι η συνάρτηση ARL για το δίπλευρο διάγραμμα Shewhart, συναρτήσει της μετατόπισης δ δίνεται από τον τύπο (π.χ. βλ. Αντζουλάκος, 2010, σελ. 36)

$$ARL(\delta) = \frac{1}{1 - \Phi(\delta + 3) + \Phi(\delta - 3)}$$

ενώ για το δίπλευρο διάγραμμα CUSUM (βλ. Granjon, 2012, σελ. 15)

$$\frac{1}{ARL(\delta)} = \frac{1}{ARL_+(\delta)} + \frac{1}{ARL_-(\delta)}$$

όπου $ARL_+(\delta)$ είναι το μέσο μήκος ροής για ανίχνευση αύξησης του μέσου κατά $|\delta|\sigma$ και $ARL_-(\delta)$ για ανίχνευση μείωσης του μέσου κατά $|\delta|\sigma$. Τύπος για την συνάρτηση $ARL_-(\delta)$ ανάλογος με αυτόν του $ARL_+(\delta)$ που υπολογίστηκε αναλυτικά, μπορεί να βρεθεί χρησιμοποιώντας το αντίστοιχο $t_0 = t_0(\theta)$.

Σημειώνουμε ότι οι παράμετροι για τον δίπλευρο αλγόριθμο CUSUM ώστε το ARL_0 να είναι 370.4, υπολογίζονται παρακάτω με χρήση της προσέγγισης του Siegmund και είναι ίσες με $h = 4.76713$ και $\delta = 1$. Αρχικά βρίσκουμε την τιμή ARL_0 (προσέγγιση Siegmund) για τον δίπλευρο έλεγχο CUSUM, με χρήση μίας συνάρτησης. Η συνάρτηση έχει παράμετρο το h

και χρησιμοποιείται σε ένα βρόχο για τον εντοπισμό εκείνων των h για τα οποία $ARL_0 > 370.3$ και $ARL_0 < 370.5$. Έχοντας πάρει δύο ενδεικτικές τιμές, μπορούμε με δοκιμές να διαπιστώσουμε ότι το ζητούμενο h είναι 4.76713.

```
> # Εύρεση του h για ARL0~370.4 και δ=1
> theta0<-1;theta<-theta0
> ARLf<-function(h) {
+ ARLplus<-1/(theta-1.5)*(h+1.166+exp(-(2*theta-3)*(h+1.166)))/(2*theta-3)-
1/(2*theta-3))
+ ARLminus<-1/(0.5-theta)*(h+1.166+exp(-(1-2*theta)*(h+1.166)))/(1-2*theta)-
1/(1-2*theta))
+ w <- 1/ARLplus + 1/ARLminus
+ return(1/w)
+ }
> for (i in seq(1,5,0.00025)) {
+   if (ARLf(i)>370.3 & ARLf(i)<370.5) {
+     print(ARLf(i));print(i) }
+ }
[1] 370.3517
[1] 4.767
[1] 370.4458
[1] 4.76725
```

Τώρα για τον υπολογισμό του ARL για διάφορες τιμές του δ , χρησιμοποιούμε τους παραπάνω τύπους, σε δύο βρόχους επανάληψης έναν για τον αλγόριθμο CUSUM και έναν για τον Shewhart, αντίστοιχα.

```
# Υπολογισμός του ARL για διάφορες τιμές του δ
> h<-4.76713;sigma<-1;theta0<-1
> ARLcsum<-c();ARLplus<-c();ARLminus<-c();j<-1
> for(delta in seq(0,4,0.25)){
+ theta<-theta0+delta*sigma
+ ARLplus[j]<-ifelse(theta==1.5, (h+1.166)^2,1/(theta-1.5)*(h+1.166+exp(-
(2*theta-3)*(h+1.166)))/(2*theta-3)-1/(2*theta-3)))
+ ARLminus[j]<-ifelse(theta==0.5, ((h+1.166)^2)/2,1/(0.5-
theta)*(h+1.166+exp(-(1-2*theta)*(h+1.166)))/(1-2*theta)-1/(1-2*theta))
+ ARLcsum[j]<-1/(1/ARLplus[j]+1/ARLminus[j])
+ j<-j+1
+ }
> ARLshew<-c();k<-1 # ARL shew
> for(delta in seq(0,4,0.25)){
+ ARLshew[k]<-1/(1-pnorm(delta+3,mean=0,sd=1)+pnorm(delta-3,mean=0,sd=1))
+ k<-k+1
+ }
> cbind(seq(0,4,0.25),ARLcsum,ARLshew)
```

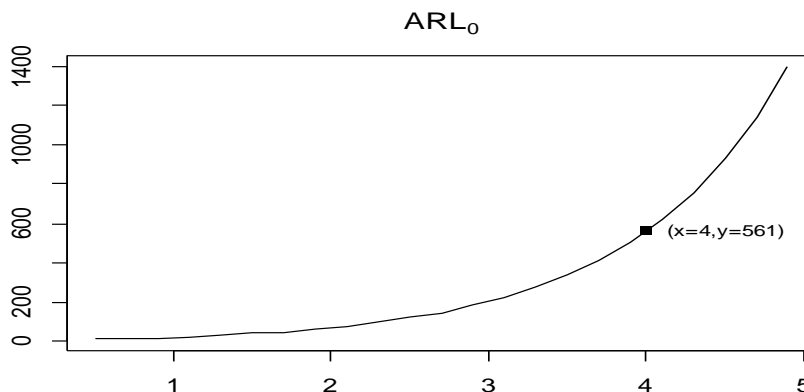
ΠΙΝΑΚΑΣ 4.3.1
*Σύγκριση τιμών συναρτήσεων ARL μεταξύ των
 στατιστικών αλγορίθμων CUSUM και Shewhart*

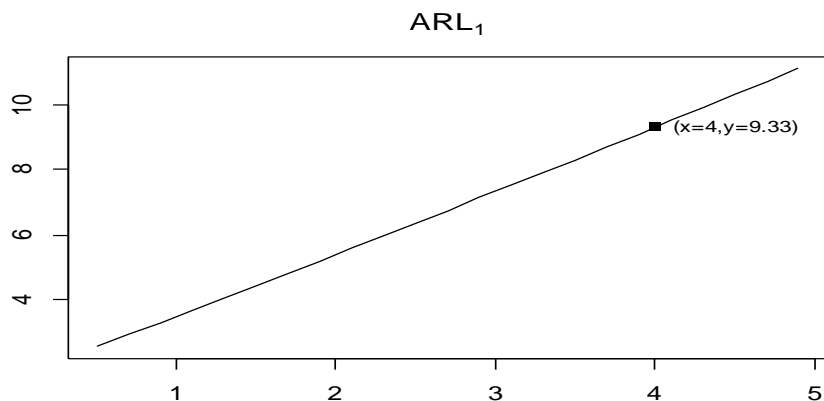
δ	CUSUM	Shewhart	δ	CUSUM	Shewhart
0,00	370,40	370,40	2,00	3,73	6,30
0,25	121,36	281,15	2,25	3,23	4,41
0,50	35,18	155,22	2,50	2,84	3,24
0,75	16,14	81,22	2,75	2,54	2,49
1,00	9,87	43,89	3,00	2,29	2,00
1,25	7,02	24,96	3,25	2,09	1,67
1,50	5,43	14,97	3,50	1,92	1,45
1,75	4,43	9,47	3,75	1,78	1,29

Παρατηρούμε ότι για μικρές του δ , ο αλγόριθμος CUSUM έχει μικρότερο μέσο μήκος ροής σε σχέση με το διάγραμμα Shewhart, που σημαίνει ότι ανιχνεύει γρηγορότερα την μετατόπιση του μέσου. Το αντίθετο συμβαίνει όταν το δ παίρνει τιμές μεγαλύτερες του 2.75, αλλά τότε η διαφορά είναι αμελητέα. Καταλήγουμε ότι αποδοτικότερος αλγόριθμος του μέσου είναι ο CUSUM είναι ενώ για πολύ ακραίες μετατοπίσεις (μεγαλύτερες ή ίσες του 3σ) το διάγραμμα Shewhart είναι ελάχιστα καλύτερο.

4.3.3 Υπολογισμός Παραμέτρων

Όπως έχουμε δει μέχρι τώρα, για την χρήση του αλγορίθμου CUSUM απαιτούνται κάποιες παράμετροι όπως τα θ_0 , θ_1 , σ^2 και h . Όταν το θ_0 και το σ^2 είναι άγνωστα, κάνουμε χρήση ενός αρχικού, αρκούντως μεγάλου μεγέθους δείγματος και υπολογίζουμε το δειγματικό μέσο και τη δειγματική διασπορά. Όταν το θ_1 είναι άγνωστο, η επιλογή του δ εξαρτάται από την *εκ των προτέρων* γνώση μας για τα δεδομένα. Δηλαδή, εξαρτάται από τον χρήστη για το τι τάξης μετατόπιση θα εξετάσει σύμφωνα με τις προηγούμενες παρατηρήσεις.





Σχήμα 4.3.2 Συναρτήσεις ARL για θ_0 και θ_1 , σύμφωνα με την προσέγγιση του Siegmund, ως προς τις διάφορες τιμές του h

Τα παραπάνω σχήματα μας δίνουν σημαντικές πληροφορίες για τον αλγόριθμο CUSUM (βλ. Granjon, 2012, σελ. 14). Παρατηρούμε ότι ο CUSUM με παράμετρο $h = 4$, κατά μέσο όρο, θα σημάνει εσφαλμένο συναγερμό κάθε 561 παρατηρήσεις. Ενώ, εντοπίζει μετατόπιση του μέσου της τάξεως 1σ με μέση καθυστέρηση περίπου 9 σημεία. Από τα δύο αυτά διαγράμματα συμπεραίνουμε επίσης ότι όσο πιο μικρό είναι το h τόσο πιο ευαίσθητος γίνεται ο αλγόριθμος στην ανίχνευση μετατόπισης, δίνοντας όμως μικρές τιμές του ARL_0 άρα πολλούς εσφαλμένους συναγερμούς. Εάν πάλι έχουμε μεγάλο h , ο αλγόριθμος γίνεται πιο «επιφυλακτικός» δίνοντας μικρές τιμές στη σήμανση συναγερμού, αλλά δημιουργείται παράλληλα καθυστέρηση στην πραγματική ανίχνευση μετατόπισης. Αυτή η αναλογία, καθορίζεται κάθε φορά αποκλειστικά από τον χρήστη του αλγορίθμου.

4.4 Βέλτιστη ιδιότητα του CUSUM

Σε αυτήν την παράγραφο θα διατυπώσουμε κάποια προβλήματα από την θ.β.σ. Σημειώνουμε ότι όταν τ_0 είναι το σημείο αλλαγής, θα συμβολίζουμε με \mathbb{P}_n το μέτρο πιθανότητας με σ.π.π. που δίνεται από την σχέση

$$f_{\theta_1}(x_1, \dots, x_n) = \prod_{k=1}^{\tau_0-1} f_{\theta_0}(x_k) \cdot \prod_{k=\tau_0}^n f_{\theta_1}(x_k)$$

ενώ με \mathbb{P}_∞ το μέτρο πιθανότητας εάν δεν έχει συμβεί κάποια αλλαγή στην κατανομή.

Ένα εύλογο ερώτημα που θα μπορούσε να τεθεί είναι το εξής: Αφού ο αλγόριθμος CUSUM χρησιμοποιεί τον SPRT και ο SPRT έχει βέλτιστη ιδιότητα υπό την έννοια ότι απαιτεί το ελάχιστο μέγεθος δείγματος, μήπως και ο CUSUM έχει κάποια ιδιότητα που να τον χαρακτηρίζει, όπως για παράδειγμα κάποια ιδιότητα που αφορά την ταχύτητα ανίχνευσης; Η απάντηση σε αυτό το ερώτημα είναι θετική και δόθηκε αρχικά από τον Lorden (1971) για την ασυμπτωτική περίπτωση όπως θα δούμε παρακάτω. Ενώ στον έλεγχο SPRT

είχαμε τις ποσότητες a και β που ήταν τα σφάλματα τύπου I και II και χρησιμοποιούνταν ως αρχικές συνθήκες, τώρα έχουμε τα ARL_0 και ARL_1 . Για την αξιολόγηση του CUSUM ενδιαφερόμαστε για τον μέσο χρόνο μεταξύ δύο εσφαλμένων συναγεργμών και για τον μέσο χρόνο καθυστέρησης όπως αναφέραμε στην Παράγραφο 4.1. Αυτές οι δύο ποσότητες είναι ουσιαστικά τα ARL_0 και ARL_1 αντίστοιχα.

Συμβολίζουμε με τ_0 το πραγματικό σημείο αλλαγής κατανομής, με T τον χ.σ. ενός στατιστικού αλγορίθμου και ως $\Delta_\gamma = \{T: \mathbb{E}_\infty(T) \geq \gamma\}$ όπου $\gamma > 1$.

Ο Lorden (1971) είχε διατυπώσει ένα μέτρο αξιολόγησης της καθυστέρησης ανίχνευσης μετατόπισης των στατιστικών αλγορίθμων ως το *ουσιώδες άνω φράγμα μέσης καθυστέρησης ανίχνευσης* (*essential supremum average detection delay-ESADD*) που ορίζεται ως

$$ESADD(T) = \sup_{\tau_0 \geq 1} \text{ess sup } \mathbb{E}_{\theta_1}[(T - \tau_0 + 1)^+ | \mathcal{F}_{\tau_0-1}]$$

όπου \mathcal{F}_{τ_0-1} είναι η συνήθης διήθηση και η μέση τιμή ως προς την σ.π.π f_{θ_1} . Σκοπός τώρα είναι να βρεθεί ένας χ.σ. $T \in \Delta_\gamma$ ώστε να ελαχιστοποιεί την $ESADD(T)$. Αποδείχθηκε αρχικά (Lorden, 1971) ότι ο χ.σ. (4.1) δηλαδή ο αλγόριθμος CUSUM, ικανοποιεί αυτή την συνθήκη ασυμπτωτικά, δηλαδή όταν $\gamma \rightarrow \infty$ και ότι η σχέση που συνδέει τη μέση καθυστέρηση ανίχνευσης με τον αναμενόμενο αριθμό σημείων μέχρι την ένδειξη συναγεργμού είναι, ασυμπτωτικά:

$$ESADD(T) \approx \frac{\log(\gamma)}{D(f_1 || f_0)}$$

Αργότερα, (Moustakides, 1986) αποδείχθηκε ότι ο αλγόριθμος CUSUM ελαχιστοποιεί την ποσότητα $ESADD(T)$ και για οποιοδήποτε $\gamma > 1$.

Αναφέρουμε ακόμα δύο περιπτώσεις προβλημάτων εύρεσης μέτρων αξιολόγησης που παρουσιάζονται συχνά στη βιβλιογραφία της ακολουθιακής ανάλυσης, των οποίων όμως οι λύσεις είναι διαφορετικοί αλγόριθμοι από τον CUSUM.

(i) Ο Pollack (1985) είχε ορίσει ένα διαφορετικό μέτρο καθυστέρησης ανίχνευσης μετατόπισης ως το *άνω πέρασ των μέσων χρόνων καθυστέρησης* (*supremum average detection delay-SADD*)

$$SADD(\tau) = \sup_{\tau_0 \geq 1} \mathbb{E}_{\tau_0}(T - \tau_0 | T \geq \tau_0)$$

Πάλι αναζητούμε τον χρόνο στάσης $T \in \Delta_\gamma$ που να ελαχιστοποιεί το SADD. Η τ.μ. T που επιλύει αυτό το πρόβλημα είναι ο χ.σ. των Shiryaev-Roberts (βλ. Veeravalli, 2012, σελ. 26).

(ii) Όταν το σημείο αλλαγής κατανομής τ_0 είναι τ.μ. με μη αρνητικές τιμές, τότε μπορούμε να ορίσουμε τον *μέσο χρόνο καθυστέρησης* (*average detection delay-ADD*) ως

$$\text{ADD}(\tau) = \mathbb{E}((T - \tau_0)^+) = \sum_{n=1}^{\infty} \pi_n \mathbb{E}_n(T - \tau_0)^+$$

όπου $\pi_n = \mathbb{P}(\tau_0 = n)$. Συμβολίζουμε με PFA την πιθανότητα εσφαλμένου συναγερμού (*probability of false alarm-PFA*)

$$\text{PFA}(T) = \mathbb{P}(T < \tau_0) = \sum_{n=1}^{\infty} \pi_n \mathbb{P}_n(T < \tau_0)$$

Έστω $\Delta_a = \{T: \text{PFA}(T) \leq a\}$, $a \in (0,1)$. Αποδεικνύεται ότι όταν η τ.μ. τ_0 ακολουθεί γεωμετρική κατανομή, ο χ.σ. $T \in \Delta_a$ που ελαχιστοποιεί την ποσότητα είναι ο αλγόριθμος του Shiryaev (βλ. Veeravalli, 2012, σελ. 11-14)

Αυτά τα προβλήματα που αναφέρθηκαν είναι ουσιαστικά προβλήματα εύρεσης χρόνου στάσης που ελαχιστοποιεί κάποια ποσότητα, ικανοποιώντας παράλληλα αρχικές συνθήκες. Ο κλάδος της θ.β.σ. αποτελεί πλέον το καταλληλότερο εργαλείο για την κατασκευή χρόνων στάσης που εμφανίζονται στη θεωρία αλγορίθμων στατιστικής ανίχνευσης.

Σύνοψη

Είδαμε πως η θεωρία των martingales και των χρόνων στάσης συμβάλλει στην απόδειξη των εξισώσεων της ακολουθιακής ανάλυσης. Οι εξισώσεις του Wald αποτελούν τα μαθηματικά εργαλεία για την εκτίμηση διαφόρων προβλημάτων στη στατιστική όπως η εύρεση άνω φράγματος για τον απαιτούμενο μέσο αριθμό δείγματος στην κατασκευή δ.ε. με καθορισμένο εύρος, η δημιουργία της συνάρτησης μέσου αριθμού δείγματος των στατιστικών ελέγχων, η εκτίμηση της συνάρτησης μέσου μήκος ροής του αλγορίθμου CUSUM, αλλά και σε διάφορα προβλήματα όπως το πρόβλημα χρωκοπίας του παίκτη (Ghosh, 1991).

Εξετάζοντας τα αριθμητικά αποτελέσματα, είδαμε ότι εν γένει οι στατιστικές τεχνικές της ακολουθιακής ανάλυσης απαιτούν μικρό μέγεθος δείγματος, όπως η πλήρως ακολουθιακή μέθοδος κατασκευής δ.ε., ο SPRT σε σχέση με τον έλεγχο Neyman-Pearson και ο αλγόριθμος CUSUM σε σχέση με τον αντίστοιχο του Shewhart.

Από τα σημαντικότερα επιτεύγματα της ακολουθιακής ανάλυσης ήταν η διατύπωση της βέλτιστης ιδιότητας του ελέγχου SPRT καθώς έδωσε κίνητρα για την δημιουργία της θ.β.σ. Αυτή εν συνεχεία, επιλύει θεμελιώδη προβλήματα στον κλάδο της στατιστικής θεωρίας ανίχνευσης κατηγοριοποιώντας του αλγορίθμους ανάλογα με τις βέλτιστες ιδιότητες τους.

Οι σύγχρονες εφαρμογές της ακολουθιακής στατιστικής αφορούν προβλήματα κατασκευής στατιστικών αλγορίθμων ταχύτατης ανίχνευσης αλλαγής κατανομής και τον ορισμό των ιδιοτήτων τους αλλά για συσχετισμένες τ.μ. Επίσης όταν οι ιδιότητες δεν κληρονομούνται γίνεται προσπάθεια να εφαρμοστούν ασυμπτωτικά. Ενδεικτικά (Baseville, 1993) αναφέρονται κάποιες παραλλαγές του αλγορίθμου CUSUM, αλγόριθμοι τύπου Bayes και οι επεκτάσεις τους. Μια πλήρης περιγραφή των ιδιοτήτων των στατιστικών αλγορίθμων στην γενική περίπτωση των συσχετισμένων τ.μ., δίνεται από τον Veeravalli (2012).

Παράρτημα

(I) Απόκλιση Kullback-Leibler

Ορισμός. Έστω δύο συνεχείς τ.μ. με σ.π.π. f_1 και f_2 αντίστοιχα. Ορίζουμε ως απόκλιση K-L την ποσότητα

$$D_{KL}(f_1||f_2) = \int_{-\infty}^{\infty} f_1(x) \log\left(\frac{f_1(x)}{f_2(x)}\right) dx$$

Πρόταση. Για την απόκλιση K-L ισχύει ότι $D_{KL}(f_1||f_2) \geq 0$.

Απόδειξη. Ισχύει ότι

$$\begin{aligned} D_{KL}(f_1||f_2) &= \int_{-\infty}^{\infty} f_1(x) \log\left(\frac{f_1(x)}{f_2(x)}\right) dx \\ &= \mathbb{E}_{f_1} \left[-\log\left(\frac{f_2(x)}{f_1(x)}\right) \right] \geq -\log \left[\mathbb{E}_{f_1} \left(\frac{f_2(x)}{f_1(x)} \right) \right] \\ &= -\log \left[\int_{-\infty}^{\infty} \frac{f_2(x)}{f_1(x)} f_1(x) dx \right] = -\log(1) = 0, \end{aligned}$$

όπου στη δεύτερη σχέση χρησιμοποιήσαμε την ανισότητα του Jensen. \square

Έστω $Z = \log\left(\frac{f(x|\theta_1)}{f(x|\theta_0)}\right)$, με $\theta_1 > \theta_0$. Τότε παρατηρούμε ότι

$$\begin{aligned} \mathbb{E}_{\theta_0}(Z) &= \int_{-\infty}^{\infty} f(x|\theta_0) \log\left(\frac{f(x|\theta_1)}{f(x|\theta_0)}\right) dx \\ &= -\int_{-\infty}^{\infty} f(x|\theta_0) \log\left(\frac{f(x|\theta_0)}{f(x|\theta_1)}\right) dx = -D_{KL}(f_0||f_1) < 0. \end{aligned}$$

Με όμοιο τρόπο αποδεικνύεται ότι $\mathbb{E}_{\theta_1}(Z) > 0$.

(II) Θεωρία πινάκων

Για να βρούμε τον αντίστροφο του πίνακα Σ_{11} θα υπολογίσουμε πρώτα την ορίζουσα του πίνακα Σ

$$\Sigma = \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$

Πρόταση Ισχύει ότι $\det \Sigma = (1 - \rho^2)^n$

Απόδειξη. Για $n = 2$ προκύπτει ότι $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ και $\det \Sigma = 1 - \rho^2$. Έστω ότι η ζητούμενη σχέση ισχύει για $n = k$. Τότε εάν στον πίνακα

$$\Sigma^k = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{k-1} & \rho^k \\ \rho & 1 & \rho^3 & \dots & \rho^{k-2} & \rho^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \rho^{k-3} & \dots & 1 & \rho \\ \rho^k & \rho^{k-1} & \rho^{k-2} & \dots & \rho & 1 \end{bmatrix}$$

πολλαπλασιάσουμε με $-\rho$ την προτελευταία γραμμή και την προσθέσουμε στην τελευταία θα προκύψει ο πίνακας

$$\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{k-1} & \rho^k \\ \rho & 1 & \rho^3 & \dots & \rho^{k-2} & \rho^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \rho^{k-3} & \dots & 1 & \rho \\ 0 & 0 & 0 & \dots & 0 & 1 - \rho^2 \end{bmatrix}$$

και αναπτύσσοντας την ορίζουσα του παραπάνω πίνακα ως προς την τελευταία γραμμή θα πάρουμε

$$(1 - \rho^2) \det \Sigma^{k-1} = (1 - \rho^2)(1 - \rho^2)^k = (1 - \rho^2)^{k+1}$$

που είναι το ζητούμενο λόγω επαγωγής. \square

Τώρα για να βρούμε τον αντίστροφο του πίνακα Σ^k θα εργαζόμαστε πρώτα στις διαστάσεις δύο, τρία και τέσσερα. Οι αντίστροφοι πίνακες που προκύπτουν μετά από απλές πράξεις είναι

$$\Sigma_2^{-k} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}, \Sigma_3^{-k} = \frac{1}{(1-\rho^2)^2} \begin{bmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{bmatrix}$$

$$\Sigma_4^{-k} = \frac{1}{(1-\rho^2)^3} \begin{bmatrix} 1 & -\rho & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & -\rho & 1 \end{bmatrix}$$

Επομένως για τις k διαστάσεις ορίζουμε

$$\Sigma_k^{-k} = \frac{1}{(1-\rho^2)^{k-1}} \begin{bmatrix} 1 & -\rho & \dots & 0 \\ -\rho & 1+\rho^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & -\rho \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Προκύπτει ότι $\Sigma^k \Sigma_k^{-k} = I$ και λόγω μοναδικότητας του αντιστρόφου πίνακα $(\Sigma^k)^{-1} = \Sigma_k^{-k}$. Συνεπώς από όλα τα παραπάνω καταλήγουμε ότι ο τύπος για το Σ_{11}^{-1} στην παράγραφο 3.6.3 ισχύει.

(III) Κώδικας R για σχήματα

Σχήμα 1.5.1

```
> library(plotrix)
> par(mar=c(4,4,2,2))
> par(mfrow=c(1,2))
> x <- seq(-0.5,2,0.2)
> y <- x*(x-1)+1
> pl.1 <- plot(x,y,bty="n",type="l", col="black", bty="n",lwd=2,
ylab="",cex=0.9, xlab="",main="", ylim=c(0,2.5),xlim=c(-0.5,1.7))
> ablineclip(a=1,b=0, lty=1, lwd=2, col="black", x1=-0.5, x2=1.5)
> x1 <- seq(-1.8,0.5,0.2)
> y1 <- x1*(x1+1)+1
> pl.2 <- plot(x1,y1,bty="n",type="l", col="black", bty="n",lwd=2,
ylab="",cex=0.9, xlab="",main="", ylim=c(0,2.5), xlim=c(-2, 0.5))
> ablineclip(a=1,b=0, lty=1, lwd=2, col="black", x1=-1.5, x2=0.5)
```

Σχήμα 2.3.1

```
> set.seed(1)
> n <- 10000;n0 <- 10;mu <- 1;sigma <- 3;a <- 0.05;
> t <- qt(1-a/2,df=n0-1) # ποσοστιαίο σημείο της κατανομής Student
> Ts <- matrix() # πίνακας με τα μεγέθη δειγμάτων T
> j <- 1;d <- 0.5
> k <- qnorm(1-a/2,mean=0,sd=1)^2*sigma^2/d^2
> for(i in 1:n) {
+ X1 <- rnorm(n0, mean=mu, sd=sigma)
+ Ts[i] <- max(n0, floor(t^2*sd(X1)^2/d^2)+1)
+ }
> Prob <- matrix();f <-1;l <-1
> for(j in n0:600) {
+ if(j==n0) {
+ Prob[l] <- pchisq(n0*(n0-1)*d^2/(sigma^2*t^2), df=n0-1)
+ } else {
+ Prob[l] <- pchisq((n0+f)*(n0-1)*d^2/(sigma^2*t^2),df=n0-1) - pchisq(
(n0+f-1)*(n0-1)*d^2/(sigma^2*t^2),df=n0-1)
```

```

+ f <- f+1
+ }
+ l<-l+1
+ }
> G <- n0:600;par(mar=c(3,4,3,2));par(mfrow=c(1,1))
> Histo <- hist(Ts,
main="",prob=T,xlab="",ylab="",ylim=c(0,0.005),xlim=c(0,600),breaks=30)
> lines(x=G,y=Prob,col="gray21",lty=1, lwd=3))

```

Σχήμα 3.3.1

```

> library(MASS) ;library(plotrix)
> set.seed(2)
> a=0.05;b=0.1;sigma=2;theta0=1;theta1=1.4;A <- b/(1-a); B <- (1-b)/a
> C1 <- sigma^2*log(A)/(theta1-theta0);C2 <- sigma^2*log(B)/(theta1-theta0)
> D <- (theta1+theta0)/2
> Mat.H0 <- matrix();x0 <-rnorm(1,mean=theta0, sd=sigma)
> S <- x0;i <- 1;
> Mat.H0[i] <- S
> while(S>(C1+i*D) & S<(C2+i*D)) {
+ x <- rnorm(1, mean=theta0, sd=sigma);
+ S <- S+x;
+ i <- i+1
+ Mat.H0[i] <- S
+ }
> par(mar=c(3,5,4,2))
> plot(Mat.H0,type="l", col="black", bty="n",lwd=2,ylab="",cex=0.9,
xlab="",main="", ylim=c(-20,120),xlim=c(-20, 110))
> ablineclip(C1,D, lty=1, lwd=2, col="black", x1=-5 ,x2=100, y1=-20)
> ablineclip(C2,D, lty=1, lwd=2, col="black", x1=-10, x2=80)

```

Σχήμα 4.1.1

```

> set.seed(2)
> X<-rnorm(150,1,1);Y<-rnorm(100,2,1)
> Z <-c(X,Y);par(mar=c(3,3,2,2))
> index=1:250
> plot(index,Z, type="l", lwd=1,ylab="",xlab="")
> abline(v=151,lty=3);abline(v=124,lty=3)
> abline(v=170,lty=3)

```

Σχήμα 4.2.1

```

> set.seed(5)
> theta0 <-1;theta1 <-1.8;sigma<-1
> S<-matrix();X1<-rnorm(100,mean=theta0, sd=1);X2<-
rnorm(50,mean=theta1, sd=1);X<-c(X1,X2)
> S[1] <- X[1]-theta0
> for(i in 1:99) {
+ S[i+1]<-S[i]+(X[i+1]-theta0)
+ }
> for(i in 100:149) {
+ S[i+1] <- S[i]+(X[i+1]-theta0)
+ }
> par(mar=c(3,3,2,2));par(mfrow=c(1,1))
> plot(X,type="l",lwd=1,ylab="",xlab="")
> abline(v=101,lty=3)

```

```

> plot(S, type="l", lwd=1, ylab="", xlab="")
> segments(x0=0, y0=0, x1=100, y1=0, lty=2, lwd=1)
> segments(x0=100, y0=0, x1=150, y1=(theta1-theta0)*(150-100), lty=2, lwd=1)

```

Σχήμα 4.2.2

```

> set.seed(5)
> theta0 <-1 ; theta1 <-1.8;sigma<-1
> S<-matrix();X1<-rnorm(100,mean=theta0,sd=1);X2<-
rnorm(50,mean=theta1,sd=1);X<-c(X1,X2)
> S[1] <- (theta1-theta0)/sigma^2*(X[1]-(theta1+theta0)/2)
> for(i in 1:99) {
+ S[i+1]<-S[i]+(theta1-theta0)/sigma^2*(X[i+1]-(theta1+theta0)/2)
+ }
> for(i in 100:149) {
+ S[i+1] <- S[i]+(theta1-theta0)/sigma^2*(X[i+1]-(theta1+theta0)/2)
+ }
> par(mar=c(3,3,2,2))
> index<-1:150
> plot(index,X, type="l", lwd=1, pch=16, ylab="", xlab="")
> abline(v=101, lty=3)
> plot(S, type="l", lwd=1, ylab="", xlab="")

```

Σχήμα 4.2.4

```

> set.seed(1)
> theta0 <-1;sigma<-1
> X<-rnorm(100,theta0,sigma);S<-matrix();S[1]<-0
> for(i in 1:99) {
+ S[i+1] <-S[i]+(X[i]-theta0)
+ }
> par(mar=c(3,3,2,2))
> plot(S,type="l",xlim=c(0,110),ylim=c(-10,30),xlab="",ylab="")
> segments(x0=100,y0=S[100],x1=110,y1=S[100])
> segments(x0=110,y0=S[100],x1=60,y1=-5)
> segments(x0=110,y0=S[100],x1=60,y1=S[100]+abs(S[100]+5))

```

Σχήμα 4.2.5

```

> x <-0;y <-0;par(mar=c(3,3,2,2))
> plot(x,y,type="l",xlim=c(0,20),ylim=c(-80,70),ylab="",xlab="")
> segments(x0=0,y0=4,x1=20,y1=64);segments(x0=10,y0=1,x1=20,y1=31)
> segments(x0=10,y0=1,x1=20,y1=-31);segments(x0=0,y0=-10,x1=20,y1=-70)
> text(x=4,y=50,expression(paste("Accept", " ",H[1])))
> text(x=4,y=-55,expression(paste("Accept", " ",H[-1])))
> text(x=17,y=0,expression(paste("Accept", " ",H[0])))

```

Σχήμα 4.3.2

```

> # ARL0
> theta0<-1;theta<-theta0;ARL<-matrix();h<-seq(0.5,5,0.2)
> par(mar=c(3,3,3,2))
> ARLf<-function(h) {
+ w<-1/(theta-1.5)*(h+1.66+exp(-(2*theta-3)*(h+1.66)))/(2*theta-3)-
1/(2*theta-3)}

```

```

+ return(w)
+ }
> ARL<-1/(theta-1.5)*(h+1.66+exp(-(2*theta-3)*(h+1.66)))/(2*theta-3)-
1/(2*theta-3))
> plot(h,ARL,type="l",main=expression(paste(ARL[0])))
> points(x=4,ARLf(4),pch=15);print(ARLf(4))
[1] 560.9773
> text(x=4.5,y=ARLf(4),cex=0.8,expression(paste("(x=4,y=561)")))
> # ARL1
> theta1<-2;theta<-theta1;ARL<-matrix();h<-seq(0.5,5,0.2)
> par(mar=c(3,3,3,2))
> ARLf<-function(h) {
+ w<-1/(theta-1.5)*(h+1.66+exp(-(2*theta-3)*(h+1.66)))/(2*theta-3)-
1/(2*theta-3))
+ return(w)
+ }
> ARL<-1/(theta-1.5)*(h+1.66+exp(-(2*theta-3)*(h+1.66)))/(2*theta-3)-
1/(2*theta-3))
> plot(h,ARL,type="l",main=expression(paste(ARL[1])))
> points(x=4,ARLf(4),pch=15);print(ARLf(4))
[1] 9.326965
> text(x=4.5,y=ARLf(4),cex=0.8,expression(paste("(x=4,y=9.33)")))

```

(IV) Κώδικας R για προγράμματα

Πρόγραμμα 1

```

> # Μέθοδος του Stein
> set.seed(1)
> n <- 1000;n0 <- 10;mu <- 1;sigma <- 3;a <- 0.05;j <- 1
> t <- qt(1-a/2,df=n0-1);mat.Stein <- matrix(,6,5)
> Ts <- matrix() # πίνακας με τα μεγέθη δειγμάτων T

> for(d in c(1, 0.75 ,0.5, 0.25, 0.15, 0.05)) { # ακτίνα του δ.ε.
+ s<-0;k <- qnorm(1-a/2,mean=0,sd=1)^2*sigma^2/d^2

+ for(i in 1:n) {
+   X1 <- rnorm(n0, mean=mu, sd=sigma)
+   Ts[i] <- max(n0, floor(t^2*sd(X1)^2/d^2)+1)
+   if(Ts[i]>n0) {
+     X2 <- rnorm(Ts[i]-n0,mean=mu,sd=sigma) # επιλογή επιπλέον παρατηρήσεων
+     X3 <- c(X1,X2) # πίνακας με το τελικό δείγμα
+   }
+   else {
+     X3 <- X1
+   }
+   if(mu>=mean(X3)-d & mu<=mean(X3)+d) {
+     s <- s+1
+   }
+ }
+ mat.Stein [j,] <- c(k, mean(Ts), sd(Ts), mean(Ts)/k, s/n)
+ j <- j+1
+ }
> print(mat.Stein)

```

Πρόγραμμα 2

```
> # Πλήρως Ακολουθιακή Μέθοδος
> set.seed(1)
> n <- 1000;n0 <- 10;mu <- 1;sigma <- 3;a <- 0.05;l <-1
> Ta <- matrix() # πίνακας με τα μεγέθη δειγμάτων T
> mat.Seq <- matrix(,6,5); # πίνακας εξόδου του αλγορίθμου
> X <- matrix() # πίνακας με το τελικό δείγμα

> for(d in c(1, 0.75 , 0.5, 0.25, 0.15, 0.05)) { # ακτίνα του δ.ε.
+   k <- qnorm(1-a/2,mean=0,sd=1)^2*sigma^2/d^2
+   s <- 0
+   for(j in 1:n) {
+     i <- n0
+     X <- rnorm(n0, mean=mu, sd=sigma)
+     k0 <- qnorm(1-a/2, mean=0, sd=1)^2*sd(X)^2/d^2
+     while(i<k0) { # βρόχος για εκτίμηση του Ta
+       i <- i+1
+       X[i] <- rnorm(1, mean=mu, sd=sigma)
+       k0 <- qnorm(1-a/2, mean=0, sd=1)^2*sd(X)^2/d^2
+     }
+     Ta[j] <- i
+     if(mu>=mean(X)-d & mu<=mean(X)+d) {
+       s <- s+1
+     }
+   }
+   mat.Seq[l,] <- c(k, mean(Ta), sd(Ta), mean(Ta)/k, s/n)
+   l <- l+1
+ }
> print(mat.Seq)
```

Πρόγραμμα 3

```
# Μέθοδος του Healy
> set.seed(1);library(MASS)
> p=2;m=10;Mean=c(1,2);S=matrix(c(1,0.5,0.5,2),nrow=2,ncol=2)
;alpha=0.05;n=1000
> u<-p*(m-1)/(m-p)*qf(1-alpha, df1=p, df2=m-p);
mat.Healy <- matrix(,ncol=5,nrow=6);j<-1
> lambda_S <- max(eigen(S)$values) # ιδιοτιμή του πίνακα διασπορών Σ
>
> for(d in c(1,0.75, 0.5, 0.25, 0.15, 0.05)) { # ακτίνα περιοχής
+   C <- qchisq(1-alpha, df=p)*lambda_S/d^2
+   sum=0;XTbar <- c();T <- c();X3<-c()
+   for(i in 1:n) {
+
+     X1 <- mvrnorm(m, mu=Mean, Sigma=S)
+     Sn <- cov(X1) # sample covariance matrix
+     lambda <- max(eigen(Sn)$values)
+     T[i] <- max(m,floor(u*lambda/d^2)+1)
+
+     if(T[i]>m) {
+       X2 <- mvrnorm(T[i]-m, mu=Mean, Sigma=S)
+       X3 <- rbind(X1,X2)
+     } else {
```

```

+       X3 <- X1
+     }
+     XTbar <- c(mean(X3[,1]),mean(X3[,2])) # Διάνυσμα μέσης τιμής
+     if (t(XTbar-Mean)%*%(XTbar-Mean)<=d^2){ # Περιοχή εμπιστοσύνης
+       sum=sum+1
+     }
+   }
+ }
+ mat.Healy[j,] <- c(C, mean(T), sd(T), mean(T)/C, sum/n)
+ j <- j+1
+ }
+ }
+ print(mat.Healy)

```

Πρόγραμμα 4

```

> # Μέθοδος του Srivastava
> set.seed(1)
> library(MASS)
> p=2;m=10;Mean=c(1,2);S=matrix(c(1,0.5,0.5,2),nrow=2,ncol=2);
alpha=0.05;n=1000
> u<-p*(m-1)/(m-p)*qf(1-alpha, df1=p, df2=m-p);mat.Sriva <-
matrix(,ncol=5,nrow=6);j<-1
> a<- qchisq(1-alpha, df=p);
> lambda_S <- max(eigen(S)$values)
>
> for(d in c(1,0.75, 0.5, 0.25, 0.15, 0.05)) { # ακτίνα περιοχής
+   C <- qchisq(1-alpha, df=p)*lambda_S/d^2 # βέλτιστο μέγεθος δειγμ.
+   sum=0;XTbar <- c();T <- c();X<-c()
+   for(i in 1:n) {
+     X <- mvrnorm(m, mu=Mean, Sigma=S)
+     Sn <- cov(X) # sample covariance matrix
+     lambda_n <- max(eigen(Sn)$values)
+     N <- m # Αρχική τιμή του N
+     while(N<a/d^2*lambda_n){
+       N <- N+1
+       Xnew <- mvrnorm(1, mu=Mean, Sigma=S)
+       X <- rbind(X,Xnew)
+       Sn <- cov(X)
+       lambda_n <- max(eigen(Sn)$values)
+     }
+     T[i] <- N # estimation of sample size
+     XTbar <- c(mean(X[,1]),mean(X[,2])) # διάνυσμα μέσης τιμής
+     if (t(XTbar-Mean)%*%(XTbar-Mean)<=d^2){ # περιοχή εμπιστοσύνης
+       sum=sum+1
+     }
+   }
+   mat.Sriva[j,] <- c(C, mean(T), sd(T), mean(T)/C, sum/n)
+   j <- j+1
+ }
> print(mat.Sriva)

```


Πρόγραμμα 5

```

> set.seed(1)
> sigma<-2;theta0<-1;thetal<-1.4;alpha<-0.05;n<-200;beta<-matrix();j<-1;N<-10000;
> for(n in c(30,50,100,150,200,250,300)) {
+ beta[j]<-pnorm((theta0-thetal)/sqrt(sigma^2/n))+qnorm(1-
alpha,mean=0,sd=1),mean=0,sd=1)
+ j<-j+1
+ }
> matrixNP<-matrix(ncol=3,nrow=length(beta)) # πίνακας με στοιχεία N-P
> matrixNP[,1]<-c(30,50,100,150,200,250,300)
> matrixNP[,2]<-rep(alpha,length(beta));matrixNP[,3]<-beta
> matrixSPRT<-matrix(nrow=length(beta),ncol=8)
> for(j in 1:length(beta)) {
+ A <- beta[j]/(1-alpha); B <- (1-beta[j])/alpha
+ a<-log(A);b<-log(B)
+ mat.H0<-matrix();mat.H1<-matrix() # πίνακες με τιμές των T0 και T1
+ k<-0;l<-0 # μετρητές για τα Pr(I) και Pr(II)
+ for(i in 1:N) {
+ X0<-rnorm(1,mean=theta0,sd=sigma)
+ Z0<-(thetal-theta0)/sigma^2*(X0-(theta0+thetal)/2)
+ Sn.0<-Z0 # αρχική τιμή του αθροίσματος Sn υπό την H0
+ countH0<-1 # μετρητής για την μέση τιμή E[T0]
+ X1<-rnorm(1,mean=thetal,sd=sigma)
+ Z1<-(thetal-theta0)/sigma^2*(X1-(theta0+thetal)/2)
+ Sn.1<-Z1 # αρχική τιμή του αθροίσματος Sn υπό την H1
+ countH1<-1 # μετρητής για την μέση τιμή E[T1]
+ while((a<Sn.0) & (Sn.0<b)) { # βρόχος για εκτίμηση των Pr[I] και E[T0]
+ X0<-rnorm(1,mean=theta0,sd=sigma)
+ Z0<-(thetal-theta0)/sigma^2*(X0-(theta0+thetal)/2)
+ Sn.0<-Sn.0+Z0
+ countH0<-countH0+1
+ }
+ while((a<Sn.1) & (Sn.1<b)) { # βρόχος για εκτίμηση των Pr[II] και E[T1]
+ X1<-rnorm(1,mean=thetal,sd=sigma)
+ Z1<-(thetal-theta0)/sigma^2*(X1-(theta0+thetal)/2)
+ Sn.1<-Sn.1+Z1
+ countH1<-countH1+1
+ }
+ if(Sn.0>=b) {
+ k<-k+1
+ }
+ if(Sn.1<=a) {
+ l<-l+1
+ }
+ mat.H0[i]<-countH0
+ mat.H1[i]<-countH1
+ }
+ matrixSPRT[j,]<-c(mean(mat.H0),sd(mat.H0)/length(mat.H0)
+ ,mean(mat.H1),sd(mat.H1)/length(mat.H1),k/N,l/N,k/N+1/N,alpha+beta[j])
+ }
> totalmatrix<-cbind(matrixNP,matrixSPRT) # πίνακας εξόδου
> colnames(totalmatrix)<-c("n","alpha","beta_NP","Mean[T;H0]",
"s.e(T0_bar)","Mean[T;H1]", "s.e(T1_bar)","Pr[I]","Pr[II]", "Pr[I]+Pr[II]","a+b")
> print(totalmatrix)
> par(mfrow=c(1,2));par(mar=c(3,3,2,2))
> hist(mat.H0,breaks=40,xlab="",ylab="",main="");hist(mat.H1,breaks=40,xlab="",
ylab="",main="")

```

Πρόγραμμα 6

```
> set.seed(1) # Προσέγγιση Wald
> alpha <- 0.05;beta <- 0.1;a <- log(beta/(1-alpha));b <- log((1-beta)/alpha)
> theta0 <- 1;theta1 <- 1.4;sigma <- 2;k <- (theta1-theta0)/sigma^2;theta.hat <-
(theta1+theta0)/2
> i <- 1;A.wald <- matrix(nrow=length(seq(1,1.4,0.04)), ncol=4, byrow=TRUE)
> Q <- c();ASN <-c();EZ1 <-c();t0<-c();EZ1.sq <- c()
> for(theta in seq(1,1.4,0.04)) {
+   if (theta!=theta.hat) {
+     t0[i] <- 2/(theta1-theta0)*(theta-theta.hat)
+     Q[i] <- (exp(1)^(-t0[i]*b)-1)/(exp(1)^(-t0[i]*b) - exp(1)^(-t0[i]*a))
+     EZ1[i] <- (theta1-theta0)/sigma^2*(theta-theta.hat)
+     ASN[i] <- (a*Q[i]+b*(1-Q[i]))/EZ1[i]
+   } else {
+     t0[i]<-0
+     Q[i] <- b/(b-a)
+     EZ1.sq[i] <- k^2*(sigma^2+theta^2)+k^2*theta.hat*(theta.hat-2*theta)
+     ASN[i] <- (a^2*Q[i]+b^2*(1-Q[i]))/EZ1.sq[i]
+   }
+   A.wald[i,] <- c(t0[i],theta,Q[i], ASN[i])
+   i <- i+1
+ }
> j<-1
> ASN.sim <- c();oc.sim <-c(); # ASN και OC με προσομοίωση
>
> for(theta.sim in seq(1,1.4,0.04)) {
+   count <- c();q <- c();n <- 1000
+   for(i in 1:n) {
+     count[i] <- 1;q[i] <- 0
+     X <- rnorm(1,mean=theta.sim, sd=sigma)
+     Sx <- log(dnorm(X,mean=theta1,sd=sigma)/dnorm(X, mean=theta0, sd=sigma))
+     while(Sx>a & Sx<b) {
+       X1 <- rnorm(1, mean=theta.sim, sd=sigma)
+       Sx <- Sx +log(dnorm(X1, mean=theta1, sd=sigma)/dnorm(X1, mean=theta0,
sd=sigma))
+       count[i] <- count[i]+1
+     }
+     if(Sx<=a) {
+       q[i] <- q[i]+1
+     }
+   }
+   ASN.sim[j] <- mean(count)
+   oc.sim[j] <- mean(q)
+   j <- j+1
+ }
> A.sim<-cbind(oc.sim,ASN.sim); A<-cbind(A.wald,A.sim); print(A) # Πίνακας εξόδου
```

Πρόγραμμα 7

```
> # ASN για το μοντέλο AR(1) για τρεις τιμές του ρ
> set.seed(1);alpha <- 0.05;beta <- 0.1;theta0 <- 1;theta1 <- 1.4;sigma <-2
> a<- log(beta/(1-alpha));b<- log((1-beta)/alpha);n <- 1000
> mat.ASN.p <- matrix(,nrow=length(seq(1,1.4,0.04)),ncol=4,byrow=TRUE)
> mat.ASN.p[,1] <-seq(1,1.4,0.04)
> mat.OC.p <- matrix(,nrow=length(seq(1,1.4,0.04)),ncol=3,byrow=TRUE);k <-2
> for(p in c(0.1,0.5,0.9)) {
+   j <-1;ASN <- c();OC <- c()
+   for(theta in seq(1,1.4,0.04)) {
+     sigmaE <- sqrt((1-p^2))*sigma; oc <- c();count <- c()
+     for(i in 1:n) {
+       oc[i] <- 0
```

```

+   En <- rnorm(1,0,sigmaE)
+   count[i] <- 2
+   S <- (theta1-theta0)/((1+p)*sigma^2)*((1-p)*theta+En) - (1-
p)*(theta1+theta0)/2 )
+   while (S>a & S<b) {
+     En <- rnorm(1,0,sigmaE)
+     S <- S+(theta1-theta0)/((1+p)*sigma^2)*((1-p)*theta+En) - (1-
p)*(theta1+theta0)/2 )
+     count[i] <- count[i]+1
+   }
+   if(S<a) {
+     oc[i] <- oc[i]+1
+   }
+ }
+ ASN[j] <- mean(count)
+ OC[j] <- mean(oc)
+ j <- j+1
+ }
+ mat.ASN.p[,k] <- ASN;mat.OC.p[,k-1] <- OC
+ k <- k+1
+ }
> totalmatrix<-cbind(mat.ASN.p,mat.OC.p);print(totalmatrix) # Πίνακας εξόδου

```

Πρόγραμμα 8

```

> set.seed(1)
> theta0 <-1;theta1<-2;sigma<-1;n<-1000
> h<-4;l<-1;N<-seq(0.9,2.5,0.05)
> ARLsim<-matrix();ARLwald<-matrix()
> ARLsieg<-matrix();ARLmat<-matrix(nrow=length(N),ncol=4);meanEZ<-matrix()
> for(theta in N) {
+   k<-matrix()
+   for(j in 1:n) {
+     S<-matrix();Z<-matrix()
+     X<-rnorm(1,mean=theta,sd=sigma)
+     S[1]<-0;Z[1]<-(theta1-theta0)/sigma^2*(X-(theta1+theta0)/2)
+     S[2]<-S[1]+Z[1]
+     i<-2
+     while(S[i]<h) {
+       if(S[i]<=0) {
+         S[i]<-0;
+       }
+       X<-rnorm(1,mean=theta,sd=sigma)
+       Z[i]<-(theta1-theta0)/sigma^2*(X-(theta1+theta0)/2)
+       S[i+1]<-S[i]+Z[i]
+       i<-i+1
+     }
+     k[j]<-i
+   }
+   ARLsim[1] <- mean(k)
+   ARLwald[1]<-ifelse(theta==1.5,h^2,1/(theta-1.5)*(h+exp(-(2*theta-3)*h)/(2*theta-
3)-1/(2*theta-3)))
+   ARLsieg[1]<-ifelse(theta==1.5,(h+1.166)^2,1/(theta-1.5)*(h+1.166+exp(-(2*theta-
3)*(h+1.166))/(2*theta-3)-1/(2*theta-3)))
+   meanEZ[1] <- (theta1-theta0)/sigma^2*(theta-(theta1+theta0)/2)
+   l<-l+1
+ }
> ARLmat[,1]<-meanEZ;ARLmat[,2]<-ARLsim;ARLmat[,3]<-ARLwald;ARLmat[,4]<-ARLsieg;
> par(mar=c(3,3,2,2))
> plot(N,ARLsieg,type="l",lty=4,xlab="",xlim=c(0.9,2.1),ylab="",main="")
> points(N,ARLsim,pch=4)
> lines(N,ARLwald,type="l",lty=1)

```

Βιβλιογραφία

Ελληνική

Δ.Αντζουλάκος (2010) *Στατιστικός Έλεγχος Ποιότητας*, Πανεπιστήμιο Πειραιώς, Σημειώσεις παραδόσεων.

Δ. Χελιώτης (2014) *Εισαγωγή στο στοχαστικό λογισμό*, Πανεπιστήμιο Αθηνών, Σημειώσεις παραδόσεων.

Ξένα

F. Anscombe (1952) Large sample theory of sequential estimation, *Proc. Camb. Phil. Soc.*, Vol.49, pp 600-607.

P. Armitage (1950) Sequential analysis with more than two alternative hypothesis and its relation to discriminant function analysis, *J.Roy.Statistist.Soc. Ser. B*, Vol 12, No 1, pp 137-144.

G.A. Barnard (1959) Control charts and stochastic processes, *J. Roy. Statistist. Soc.*, Ser. B, Vol 21, pp 239-271.

M. Baseville & I.Nikiforov (1993), *Detection of abrupt changes:theory and applications*, Prentice Hall.

Y.Chow & H. Robbins (1965), On the asymptotic theory of fixed width sequential confidence intervals for the mean, *Ann. Math. Statist.*, Vol 36, pp 457-462.

H.Dodge & H. Romig (1929), A method of sampling inspection, *Bell Syst. Tech. J.* Vol 8, pp 613-631.

T. Ferguson (2000) Optimal stopping and applications, unpublished manuscript <http://www.math.ucla.edu/~tom/Stopping/Contents.html>.

Z. Govindarajulu (1974) *Sequential Statistical Procedures*, Academic Press Inc.

M. Ghosh (1991) *Handbook of Sequential Analysis*, CRC Press.

M.Ghosh & N.Mukhopadhyay (1997) *Sequential Estimation*, Wiley

- P. Granjon (2012) The CUSUM algorithm, a small review, unpublished manuscript.
- W.C. Healy (1956) Two sample procedures in simultaneous estimation, *Ann. Math. Statist.*, Vol 27, pp 687-702.
- R. Hogg & A. Craig (1970) *Introduction to mathematical statistics*, Macmillan Co.
- R.A. Johnson (2007) *Applied Multivariate Analysis*, Pearson
- N. Johnson & F. Leone (1962) Cumulative sum control charts: mathematical principles applied to their construction and use, *Industrial Quality Control*, Vol 18, pp 15-21.
- N. Johnson & S. Kotz (1970) *Continuous univariate distributions-2*, Wiley: New York, pp 102.
- T.L. Lai (2001) Sequential analysis: some classical problems and new challenges, *Statistica Sinica*, 11(2):303-408.
- E. Lehman (1951) Notes on the theory of estimation, *University of California Press: Berkeley*.
- G. Lorden (1971) Procedures for reacting to a change in distribution, *Ann. Math. Statist.*, Vol 42, pp 1897-1908.
- G. Moustakides (1986) Optimal procedures for detecting changes in distribution, *Annals Statistics*, Vol 14, pp 1379-1397.
- N. Mukhopadhyay (2009) *Sequential methods and their applications*, Chapman and Hall/CRC..
- J. Neyman & E. Pearson (1933) On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society of London*, Ser.A, pp 289-337.
- R. Nowak (2010) Statistical Signal Processing, Lecture notes, ECE 830, Lecture 7
http://nowak.ece.wisc.edu/ece830/ece830_lecture7.pdf
- E.S. Page (1954) Continuous inspection schemes, *Biometrika* Vol 41, No 1, pp 100-115.
- M. Pollack (1985) Optimal detection of a change in distribution, *Annals Statistics*, Vol 13, pp 206-227.
- W.A. Shewhart (1931) *Economic control of manufactured product*, Van Nostrand, New York.
- A.N. Shiryaev (2007) *Optimal stopping rules*, Springer-Verlag, New York, NY.

- A.N. Shiryaev & M.V. Zhitlukhin (2013) Optimal stopping problems for Brownian motion with drift and disorder: application to mathematical finance and engineering, Steklov Mathematical Institute, Moscow.
- D. Siegmund (1985) Corrected diffusion approximation and their applications, Proceedings of the Berkeley Conference in honor of Jerzy Neyman and Jack Kiefer, Vol II, pp 599-617.
- M.S. Srivastava (1967) On fixed width confidence bounds for regression parameters and mean vector, *J. Roy. Statist. Soc.*, Series B, Vol 29, pp 132-140.
- C. Stein (1949) Some problems in sequential estimation. *Econometrica*, 17, pp 77-78.
- V.V. Veeravalli & T. Banerjee (2012) *Quickest change detection*, ECE Department and Coordinated Science Laboratory.
- A. Wald & J. Wolfowitz (1948) Optimum character of the sequential probability ratio test, *Ann. Math. Statist.* Vol 19, pp 326-339.
- A. Wald (1947) *Sequential analysis*, Wiley, New York.
- J. Walsh (2014) Elementary introduction to martingales, unpublished manuscript.
<https://www.math.ubc.ca/~walsh/marts.pdf>
- M. Woodroffe (1977) Second order approximation for sequential point and interval estimation, *Ann. Statist.*, Vol 5, pp 984-995.