

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΟΜΕΑΣ ΠΙΘΑΝΟΤΗΤΩΝ, ΣΤΑΤΙΣΤΙΚΗΣ
& ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

ΑΡΓΥΡΙΟΣ ΤΖΗΜΟΠΟΥΛΟΣ
ΜΑΘΗΜΑΤΙΚΟΣ

ΕΛΕΓΧΟΙ ΜΟΝΟΔΙΑΣΤΑΤΗΣ ΣΥΜΜΕΤΡΙΑΣ
ΜΙΑ ΚΡΙΤΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ, 2014

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 19-06-2014 από την εξεταστική επιτροπή:

Ζωγράφος Κωνσταντίνος, Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Λουκάς Σωτήριος, Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Μπατσίδης Απόστολος, Λέκτορας του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων (Επιβλέπων).

Υπεύθυνη Δήλωση

«Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή»

Υπογραφή

Αργύριος Τζημόπουλος

Αφιερώνεται
στους γονείς μου
και στην αδερφή μου

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διατριβή πραγματοποιήθηκε κατά τη διάρκεια του τελευταίου έτους φοίτησής μου στο Τμήμα Μαθηματικών της Σχολής Θετικών Επιστημών του Πανεπιστημίου Ιωαννίνων για την απόκτηση μεταπτυχιακού διπλώματος ειδίκευσης στον κλάδο «Στατιστική και Επιχειρησιακή Έρευνα». Ως θέμα έχει την παρουσίαση των κυριότερων στατιστικών τεστ για τον έλεγχο της μονοδιάστατης συμμετρίας κατανομών, χωρίζοντας τους ελέγχους συμμετρίας σε δύο κατηγορίες ανάλογα με το αν το κέντρο συμμετρίας είναι γνωστό εκ των προτέρων ή αν είναι άγνωστο. Στην περάτωση της διατριβής αυτής συνέβαλαν σημαντικά ορισμένα άτομα, τα οποία σπεύδω να ευχαριστήσω και να εκφράσω ευγνωμοσύνη.

Αρχικά, ευχαριστώ θερμά τον κ. Απόστολο Μπατσίδα, ο οποίος με προθυμία ανέλαβε την επίβλεψη και πρότεινε το θέμα της διατριβής. Η καθοδήγηση, η υπομονή, το ενδιαφέρον που επέδειξε και η ηθική υποστήριξη που μου παρείχε καθ' όλη τη διάρκεια της προσπάθειάς μου, με βοήθησαν σημαντικά ώστε να καταφέρω να ολοκληρώσω την παρούσα μεταπτυχιακή διατριβή.

Ειλικρινείς είναι οι ευχαριστίες μου προς τους κ.κ. Κωνσταντίνο Ζωγράφο και Σωτήριο Λουκά, για τη συμμετοχή τους στην Τριμελή Επιτροπή Κρίσης και τις παρατηρήσεις τους καθώς επίσης και προς όλους τους διδάσκοντες του Τομέα «Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας» για την πολύτιμη βοήθεια τους όλα αυτά τα χρόνια.

Εκτιμώ επίσης τη βοήθεια όλων των καθηγητών του Τμήματος Μαθηματικών της Σχολής Θετικών Επιστημών του Πανεπιστημίου

Ιωαννίνων. Ακόμη, ευχαριστώ τους συναδέλφους μου του προγράμματος Μεταπτυχιακών Σπουδών του Τομέα «Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας» για την συνεργασία τους και εκτιμώ την ψυχολογική και συναισθηματική στήριξη των φίλων μου καθώς και την κατανόηση τους κατά τη διάρκεια των σπουδών μου.

Τέλος, επιθυμώ να εκφράσω την ευγνωμοσύνη μου στους γονείς μου Αντώνη, Ειρήνη και στην αδελφή μου Μελίνα που δεν έπαψαν να με στηρίζουν πνευματικά, συναισθηματικά, ηθικά και οικονομικά όλα αυτά τα χρόνια. Χωρίς τη βοήθεια τους, την προτροπή τους και την ενθάρρυνσή τους θα ήταν αδύνατη η ολοκλήρωση των προπτυχιακών και μεταπτυχιακών μου σπουδών.

Αργύριος Τζημόπουλος

Ιωάννινα, Ιούνιος 2014.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή.....	1
1.1 Χρήσιμες έννοιες.....	8
1.2 Σκοπός της μεταπτυχιακής διατριβής και η περίληψή της.....	21

ΚΕΦΑΛΑΙΟ 2

Έλεγχος μονοδιάστατης συμμετρίας γύρω από γνωστό σημείο.....	25
2.1 Τεστ που στηρίζονται στο προσημικό τεστ	26
2.1.1 Ο έλεγχος συμμετρίας με το προσημικό τεστ	26
2.1.2 Ο έλεγχος των Cheng and Balakrishnan (2004)	29
2.1.3 Έλεγχοι προσημασμένης διάταξης.....	32
2.2 Τεστ που στηρίζονται στο τεστ των ροών.....	39
2.2.1 Το τεστ του McWilliams (1990).....	39
2.2.2 Ο έλεγχος του Baklizi (2003)	45
2.2.3 Ο έλεγχος των Modarres and Gastwirth (1996)	50
2.2.4 Το τεστ των Corzo and Babativa (2013)	53
2.2.5 Το τεστ του Baklizi (2007).....	56
2.3 Τεστ με την Εμπειρική Αθροιστική Συνάρτηση Κατανομής (ε.α.σ.κ.)	62
2.3.1 Το τεστ των Smirnov (1947) και Butler (1969).....	63
2.3.2 Ο έλεγχος των Rothman and Woodroffe (1972)	67
2.3.3 Το τεστ των Srinivasan and Godio (1973)	71
2.3.4 Ο έλεγχος των Hill and Rao (1977)	75
2.3.5 Το τεστ των Baringhaus and Henze (1992)	80
2.4 Ο έλεγχος συμμετρίας του Tajuddin (1994) και μια τροποποίηση του..	85
2.4.1 Ο έλεγχος συμμετρίας του Tajuddin (1994)	85
2.4.2 Ο έλεγχος των Modarres and Gastwirth (1998) και η βελτίωση του.....	88

ΚΕΦΑΛΑΙΟ 3

Έλεγχος μονοδιάστατης συμμετρίας γύρω από άγνωστο σημείο.....	93
3.1 Επεκτάσεις των μεθόδων ελέγχου συμμετρίας γύρω από γνωστό σημείο.....	93
3.1.1 Τροποποίηση του τεστ των Smirnov (1947) και Butler (1969)	94
3.1.2 Ο τροποποιημένος έλεγχος των Rothman and Woodroffe (1972) ..	96
3.1.3 Ο τροποποιημένος έλεγχος του McWilliams (1990)	97
3.1.4 Τροποποιήσεις του προσημικού ελέγχου.....	99
3.1.5 Τροποποιήσεις του ελέγχου προσημασμένης διάταξης.....	103
3.1.6 Οι έλεγχοι των Antille and Kersting (1977) και Ekström and Jammalamadaka (2006)	107
3.2 Τεστ που στηρίζονται σε περιγραφικά μέτρα	115
3.2.1 Δειγματικός συντελεστής λοξότητας	116
3.2.2 Ο έλεγχος των Cabilio and Masaro (1996) και η τροποποίηση των Miao <i>et al.</i> (2006)	119
3.2.3 Ο έλεγχος του Ghosh (2011)	125
3.2.4 Ο έλεγχος των Hotelling and Solomons (1932)	129
3.2.5 Ο έλεγχος του Boos (1982) και οι έλεγχοι του Auda (2006)	132
3.2.6 Ο έλεγχος της Mira (1999)	139
3.2.7 Ο κανόνας των Mukherjee <i>et al.</i> (2013)	143
3.3 Τεστ που στηρίζονται στην εμπειρική χαρακτηριστική συνάρτηση	147
3.3.1 Ο έλεγχος των Csörgö and Heathcote (1987)	147
3.3.2 Ο έλεγχος του Koutrouvelis (1985)	153
3.4 Έλεγχοι με U-στατιστικές συναρτήσεις.....	156
3.4.1 Triple test	157
3.4.2 Ο έλεγχος του Gupta (1967) και μια επέκταση του	161
3.4.3 Ο έλεγχος του Finch (1977)	169

ΚΕΦΑΛΑΙΟ 4

Συγκριτική μελέτη173

4.1 Μέθοδοι ελέγχου συμμετρίας γύρω από γνωστό σημείο174

4.2 Μέθοδοι ελέγχου συμμετρίας γύρω από άγνωστο σημείο180

ΚΕΦΑΛΑΙΟ 5

Εφαρμογές.....195

5.1 Ελέγχων συμμετρίας γύρω από γνωστό σημείο195

5.2 Ελέγχων συμμετρίας γύρω από άγνωστο σημείο208

Αντί επιλόγου233

ΠΑΡΑΡΤΗΜΑ Α.....235

ΠΑΡΑΡΤΗΜΑ Β.....249

ABSTRACT OF MSc DISSERTATION281

ΒΙΒΛΙΟΓΡΑΦΙΑ283

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Ένα από τα προβλήματα που έχει απασχολήσει τη στατιστική βιβλιογραφία από τις αρχές της προηγούμενης εκατονταετίας (βλέπε, μεταξύ άλλων, Pearson (1905), Yule (1911) και Fisher (1929)) είναι αυτό του ελέγχου της συμμετρίας μίας μονοδιάστατης κατανομής γύρω από ένα σημείο, έστω μ . Ο ορισμός της συμμετρίας, διατυπώνεται στη συνέχεια.

Ορισμός 1.1 Λέμε ότι η τυχαία μεταβλητή (τ.μ.) X είναι συμμετρική γύρω από το σημείο μ αν και μόνο αν $X - \mu \stackrel{d}{=} \mu - X$, όπου με $\stackrel{d}{=}$ συμβολίζεται ότι οι τ.μ. έχουν την ίδια κατανομή.

Στην παρούσα διατριβή το ενδιαφέρον θα επικεντρωθεί στην περίπτωση συνεχών τ.μ., όπου προκύπτει ο ακόλουθος ισοδύναμος ορισμός.

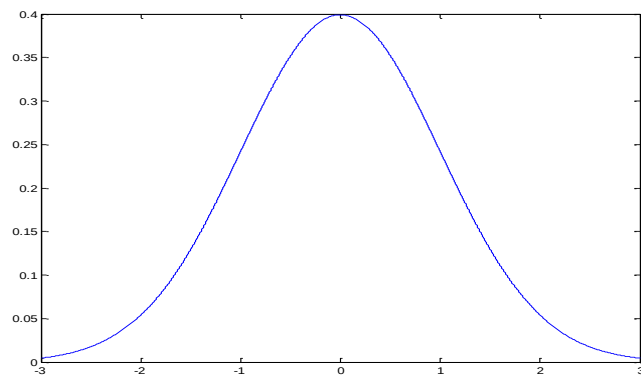
Ορισμός 1.2 Έστω X μια συνεχής τ.μ. με αθροιστική συνάρτηση κατανομής (α.σ.κ.) $F(x) = P(X \leq x)$ και συνάρτηση πυκνότητας πιθανότητας (σ.π.π) $f(\cdot)$, η συναρτησιακή μορφή των οποίων είναι άγνωστη. Η τ.μ. X είναι συμμετρική γύρω από το σημείο μ αν και μόνο αν:

$$f(x - \mu) = f(x + \mu), \text{ για κάθε } x \text{ στο πεδίο ορισμού της,} \quad (1.1)$$

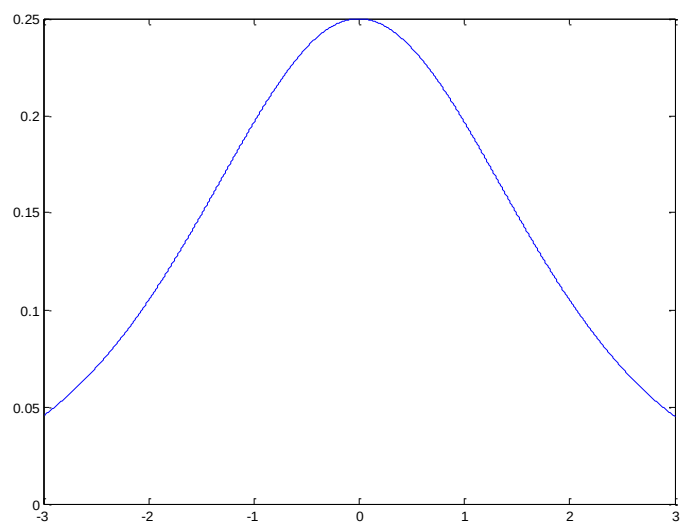
ή

$$F(\mu + x) = 1 - F(\mu - x), \text{ για κάθε } x. \quad (1.2)$$

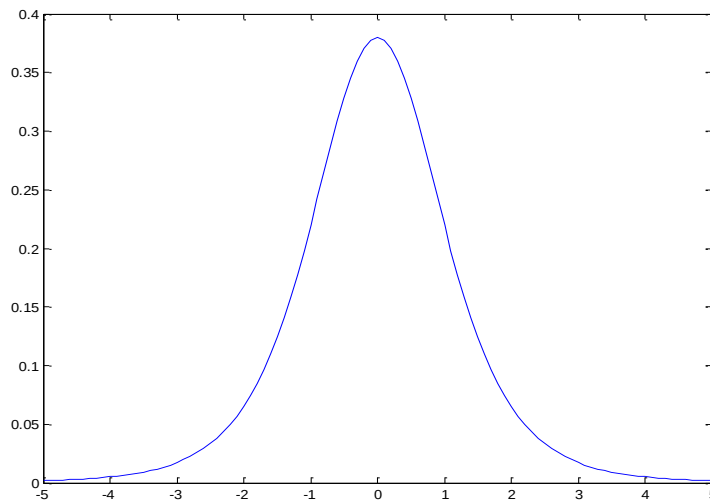
Χαρακτηριστικά παραδείγματα κατανομών οι οποίες είναι συμμετρικές γύρω από ένα σημείο, όπως φαίνεται στις γραφικές παραστάσεις 1-3, είναι η κανονική, η λογιστική και η t κατανομή.



Γράφημα 1- Σ.π.π. της κανονικής κατανομής όταν $\mu = 0$ και $\sigma = 1$, για $x \in [-3, 3]$.



Γράφημα 2- Σ.π.π. της λογιστικής κατανομής όταν $\mu = 0$ και $\sigma = 1$, για $x \in [-3, 3]$.



Γράφημα 3- Σ.π.π. της t κατανομής με 5 βαθμούς ελευθερίας όταν $\mu = 0$ και $\sigma = 1$, για $x \in [-5, 5]$.

Έστω $X_1, \dots, X_n$ ένα τυχαίο δείγμα (τ.δ.) από κάποιον πληθυσμό με σ.π.π. $f(\cdot)$ και α.σ.κ. $F(\cdot)$. Για λόγους που θα αναφερθούν στη συνέχεια, είναι σημαντικό να καθορίσουμε αν ο υποκείμενος πληθυσμός έχει συμμετρική κατανομή. Ο έλεγχος της συμμετρίας έγκειται στην ανάπτυξη στατιστικών μεθοδολογιών σχετικά με την απόφαση περί αποδοχής ή απόρριψης της μηδενικής υπόθεσης:

$$H_0 : F(\mu + x) = 1 - F(\mu - x), \text{ για κάθε } x. \quad (1.3)$$

Ο παραπάνω έλεγχος διεξάγεται έχοντας ως εναλλακτική ότι η κατανομή είναι είτε θετικά (αρνητικά) λοξή είτε γενικά ασύμμετρη, δηλαδή έχοντας ως εναλλακτική μια εκ των

$$H_1 : F(x + \mu) < 1 - F(\mu - x), \text{ για κάθε } x, \quad (1.4)$$

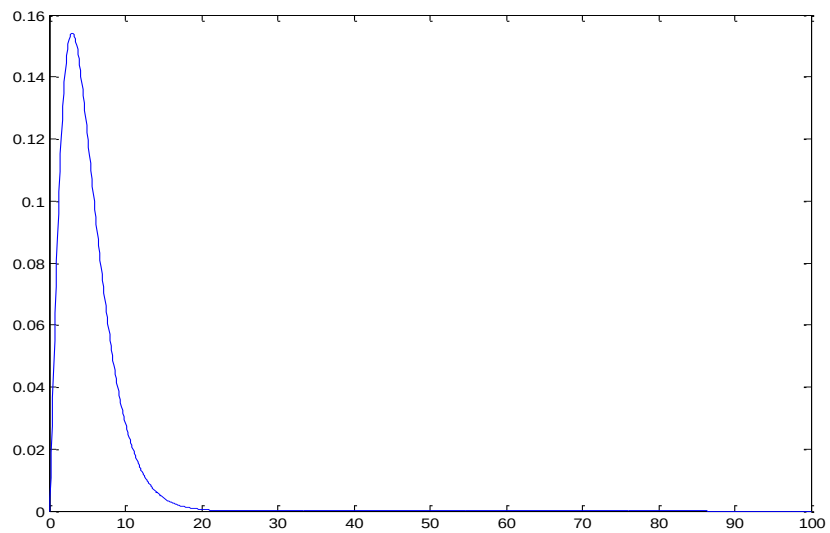
ή

$$H_2 : F(\mu + x) > 1 - F(\mu - x), \text{ για κάθε } x, \quad (1.5)$$

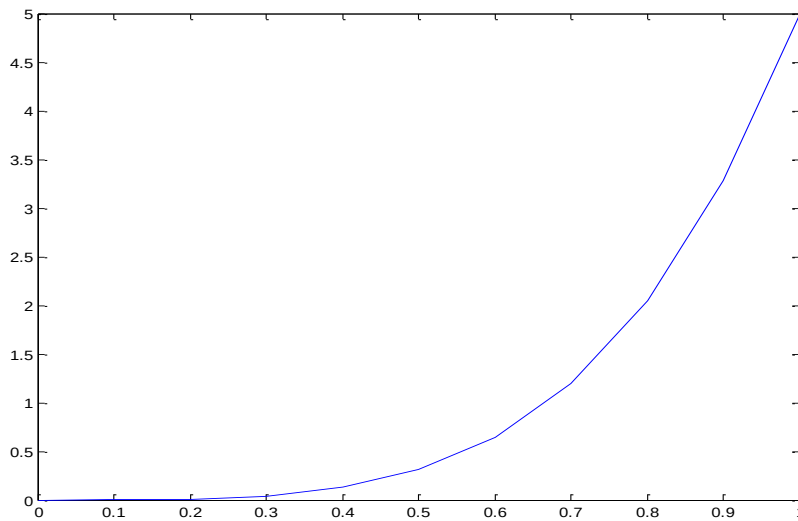
ή

$$H_3 : F(\mu + x) \neq 1 - F(\mu - x), \text{ για κάποιο } x, \quad (1.6)$$

αντίστοιχα. Προφανώς, οι δύο πρώτες από τις εναλλακτικές υποθέσεις είναι μονόπλευρες και η τρίτη είναι δίπλευρη. Στις γραφικές παραστάσεις 4 και 5 δίνονται παραδείγματα θετικά και αρνητικά λοξών κατανομών.



Γράφημα 4- Σ.π.π. της χι-τετράγωνο κατανομής με 5 βαθμούς ελευθερίας, για $x \in [0, 100]$.



Γράφημα 5- Σ.π.π. της Βήτα κατανομής με παραμέτρους 5 και 1.

Οι λόγοι της αναγκαιότητας της ανάπτυξης τρόπων ελέγχου της συμμετρίας είναι πολλοί. Στη συνέχεια παρατίθενται οι πιο σημαντικοί από αυτούς (βλέπε, μεταξύ άλλων, Ghosh (2011), Ekström and Jammalamadaka (2007), Hollander (1981)).

α) Είναι γνωστό ότι πολλές στατιστικές μεθοδολογίες ισχύουν υπό την υπόθεση ότι τα δεδομένα προέρχονται από πληθυσμό που περιγράφεται ικανοποιητικά από την κανονική κατανομή. Ωστόσο, η υπόθεση της κανονικότητας πολλές φορές δεν ικανοποιείται. Παρόλα αυτά έχει αποδειχθεί ότι κάποιες από τις προαναφερθείσες διαδικασίες είναι ανθεκτικές σε αποκλίσεις από την κανονικότητα και η θεώρηση ότι τα δεδομένα προέρχονται από συμμετρικό πληθυσμό είναι συχνά επαρκής για την εγκυρότητά τους.

β) Αρκετές μέθοδοι της Μη Παραμετρικής Στατιστικής στηρίζονται στην υπόθεση ότι τα δεδομένα προέρχονται από πληθυσμό με συμμετρική κατανομή.

γ) Σε πολλές πρακτικές εφαρμογές, είναι σημαντικό να καθορίσουμε αν ο υποκείμενος πληθυσμός έχει συμμετρική κατανομή. Για παράδειγμα, αν κάποιος ενδιαφέρεται να εκτιμήσει το μέτρο θέσης του πληθυσμού, η παρουσία λοξής κατανομής θα οδηγούσε στη θεώρηση περισσότερων του ενός μέτρων θέσης, αφού δεν θα υπήρχε ένα μοναδικό φυσικό μέτρο θέσης και έτσι θα έπρεπε να εξεταστεί με περισσότερη λεπτομέρεια ποια παράμετρος ουσιαστικά θα εκτιμηθεί (βλέπε Antille *et al.* (1982)).

δ) Έστω ότι έχουμε n το πλήθος ανεξάρτητα ζεύγη παρατηρήσεων $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Ένα μέλος κάθε ζεύγους επιλέγεται τυχαία να ανήκει στην ομάδα "ελέγχου", ενώ το άλλος μέλος στην ομάδα "πείραμα". Η υπόθεση ότι δεν υπάρχει καμία επίδραση του "πειράματος" είναι ισοδύναμη με το ότι η από κοινού κατανομή της απόκρισης (X, Y) είναι ανταλλάξιμη (exchangeable), δηλαδή ότι ισχύει η σχέση:

$$P(X \leq x; Y \leq y) = P(X \leq y; Y \leq x), \text{ για κάθε } (x, y). \quad (1.7)$$

Παίρνοντας τις διαφορές $Z = Y - X$, προκύπτει ότι υπό την (1.7) η κατανομή της τ.μ. Z είναι ίδια με την κατανομή της τ.μ. $-Z = X - Y$, ή ισοδύναμα ότι η κατανομή της τ.μ. Z είναι συμμετρική γύρω από το μηδέν. Άρα, με βάση τις παρατηρήσεις $Z_1, Z_2, \dots, Z_n$, όπου $Z_i = Y_i - X_i$, για $i = 1, 2, \dots, n$, αν απορρίπτεται η υπόθεση της συμμετρίας γύρω από το μηδέν, θα απορρίπτεται και η υπόθεση της ύπαρξης σημαντικής επίδρασης του "πειράματος", δηλαδή της ανταλλαξιμότητας. Είναι σαφές ότι σε περίπτωση μη απόρριψης της συμμετρίας γύρω από το μηδέν, δεν εξάγεται κάποιο συμπέρασμα για την ικανοποίηση ή όχι της (1.7), καθώς υπάρχουν από κοινού κατανομές $G(x, y) = P(X \leq x; Y \leq y)$ για τις οποίες η κατανομή του Z είναι συμμετρική γύρω από το μηδέν, αλλά η G δεν είναι ανταλλάξιμη (βλέπε Hollander (1981)).

ε) Τέλος, στην παλινδρόμηση η συμμετρία των υπολοίπων είναι η βάση για ανάπτυξη τεστ με στόχο την ορθότητα του μοντέλου.

Επομένως, γίνεται αντιληπτό ότι η υπόθεση της συμμετρίας συναντάται συχνά και είναι σημαντικό να προβούμε στον έλεγχο της υπόθεσης αυτής. Για το λόγο αυτό υπάρχουν διαθέσιμες στη βιβλιογραφία αρκετές μεθοδολογίες ελέγχου. Στόχος των ερευνητών ήταν να δημιουργήσουν στατιστικά τεστ τα οποία:

(α) στηρίζονται σε στατιστικές συναρτήσεις που είτε είναι ασυμπτωτικά ανεξάρτητες της αληθινής κατανομής του πληθυσμού είτε έχουν πιθανότητα σφάλματος τύπου-I, σχετικά σταθερή,

(β) διατηρούν το επίπεδο σημαντικότητας και έχουν καλή ισχύ,

(γ) ο έλεγχος είναι συνεπής, δηλαδή η ισχύς του ελέγχου για μια σταθερή και μη αληθή υπόθεση τείνει στο 1 καθώς αυξάνεται το μέγεθος του δείγματος.

(δ) Τέλος, δύο επιθυμητές ιδιότητες είναι οι σ.σ. να παραμένουν αναλλοίωτες στους μετασχηματισμούς $(X_1, \dots, X_n) \rightarrow (-X_1, \dots, -X_n)$ και $(X_1, \dots, X_n) \rightarrow (X_1^{-1}, \dots, X_n^{-1})$ καθώς αν η τ.μ. X είναι συμμετρική το ίδιο θα ισχύει και για τις τ.μ. $-X$ και X^{-1} .

Στο πλαίσιο αυτό, οι μεθοδολογίες που προτάθηκαν ανάλογα με τη διατύπωση της υπό έλεγχο υπόθεσης ταξινομούνται σε δύο μεγάλες κατηγορίες. Η πρώτη κατηγορία περιλαμβάνει τις μεθόδους για τον έλεγχο της συμμετρίας γύρω από ένα γνωστό κέντρο συμμετρίας μ . Οι περισσότερες μέθοδοι αυτής της κατηγορίας επικεντρώνονται σε ελέγχους συμμετρίας γύρω από το μηδέν. Αυτό συμβαίνει γιατί ο έλεγχος της συμμετρίας της τ.μ. X γύρω από κάποια γνωστή τιμή μ , είναι ισοδύναμος με τον έλεγχο της υπόθεσης ότι η κατανομή της τ.μ. $X - \mu$ είναι συμμετρική γύρω από το μηδέν. Η δεύτερη κατηγορία μεθοδολογιών περιλαμβάνει τους ελέγχους συμμετρίας γύρω από ένα

άγνωστο κέντρο συμμετρίας μ . Είναι προφανές ότι η υπόθεση αυτή είναι πιο ρεαλιστική και πιο δύσκολη αφού το κέντρο συμμετρίας πρέπει να εκτιμηθεί.

Σκοπός αυτής της μεταπτυχιακής διατριβής είναι η κριτική ανασκόπηση και παρουσίαση των κυριοτέρων και δημοφιλέστερων μεθόδων στατιστικού ελέγχου της μονοδιάστατης συμμετρίας. Στο πλαίσιο αυτό, στο εισαγωγικό αυτό κεφάλαιο για τη διευκόλυνση της μελέτης της διατριβής θα δοθούν κάποιες χρήσιμες έννοιες στις οποίες στηρίζονται κάποιες από τις μεθόδους που θα παρουσιασθούν στα επόμενα κεφάλαια. Τέλος, θα δοθεί ο σκοπός και η περίληψη της διατριβής.

Παρατήρηση 1.1 Στο σημείο αυτό θα ήταν παράλειψη να μην επισημανθεί ότι στη βιβλιογραφία έχουν εισαχθεί τρόποι ορισμού και ελέγχου τόσο της διδιάστατης όσο και της πολυδιάστατης συμμετρίας. Η κριτική ανασκόπηση και παρουσίαση αυτών δεν αποτελεί αντικείμενο μελέτης της παρούσας διατριβής. Τέλος, ο κατάλογος των προτεινόμενων μεθόδων ελέγχου της μονοδιάστατης συμμετρίας είναι εκτενής και στη μεταπτυχιακή αυτή διατριβή θα παρουσιαστούν οι πιο ευρέως χρησιμοποιούμενες στα πλαίσια του στατιστικού ελέγχου υποθέσεων.

1.1 Χρήσιμες έννοιες

Κάποιοι από τους ελέγχους που θα παρουσιαστούν στις επόμενες ενότητες στηρίζονται στο τεστ των ροών, στον προσημασμένης διάταξης έλεγχο του Wilcoxon (signed rank Wilcoxon test) και στην εμπειρική αθροιστική συνάρτηση κατανομής. Για το λόγο αυτό, στην ενότητα αυτή θα δοθούν οι ορισμοί αυτών καθώς και χρήσιμες ιδιότητές τους. Τέλος, η ενότητα αυτή θα ολοκληρωθεί δίνοντας τον ορισμό της διαδικασίας Wiener (γνωστής και ως κίνησης

Brown), των U-στατιστικών συναρτήσεων καθώς και συνοπτικές πληροφορίες για την kernel εκτίμηση και τους kernel εκτιμητές, έννοιες που θα χρησιμοποιηθούν συχνά στα υπόλοιπα κεφάλαια της διατριβής.

Εμπειρική αθροιστική συνάρτηση κατανομής

Ορισμός 1.1.1 Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από έναν πληθυσμό με α.σ.κ. $F(\cdot)$. Η εμπειρική αθροιστική συνάρτηση κατανομής (ε.α.σ.κ.), συμβολίζεται με $F_n(x)$ και ορίζεται ως εξής:

$$F_n(x) = \frac{\text{αριθμός } X_i \leq x}{n}, \quad x \in \mathbb{R}. \quad (1.1.1)$$

Για σταθερό x , η F_n είναι μια σ.σ. καθώς είναι συνάρτηση των $X_1, \dots, X_n$. Στις προτάσεις που ακολουθούν παρατίθενται κάποιες χρήσιμες ιδιότητες της ε.α.σ.κ..

Πρόταση 1.1.1 Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από έναν πληθυσμό με συνεχή α.σ.κ. $F(\cdot)$. Ισχύει ότι:

- α) για σταθερό x , η κατανομή της σ.σ. $nF_n(x)$, όπου $F_n(x)$ η ε.α.σ.κ., είναι διωνυμική με παραμέτρους n και $F(x)$,
- β) η ε.α.σ.κ. είναι συνεπής εκτιμητής της α.σ.κ. $F(x)$, και
- γ) η ε.α.σ.κ. ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή $F(x)$ και διακύμανση $\{F(x)(1-F(x))\}/n$.

Απόδειξη (βλέπε μεταξύ άλλων Kvam and Vidakovic (2007))

- α) Από τον ορισμό της εμπειρικής αθροιστικής συνάρτησης κατανομής εύκολα προκύπτει ότι οι δυνατές της τιμές ανήκουν στο σύνολο:

$\{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$. Για σταθερό x , η τιμή της εμπειρικής αθροιστικής συνάρτησης είναι ίση με k/n , $k = 0, \dots, n$, αν και μόνο αν k το πλήθος δειγματικές τιμές είναι μικρότερες ή ίσες του x . Επομένως, ο υπολογισμός της πιθανότητας $P\left(F_n(x) = \frac{k}{n}\right)$ ανάγεται στην εύρεση της πιθανότητας να έχουμε k το πλήθος επιτυχίες σε n δοκιμές ενός διωνυμικού πειράματος, όπου επιτυχία θεωρείται μία δειγματική τιμή X_i να είναι μικρότερη ή ίση από την σταθεροποιημένη τιμή x , με πιθανότητα επιτυχίας $p = P(X_i \leq x) = F(x)$. Άρα,

$$P(nF_n(x) = k) = \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}, \quad k = 0, \dots, n, \quad (1.1.2)$$

που ολοκληρώνει την απόδειξη.

β) Καθώς ισχύει ότι $nF_n(x) \sim B(n, F(x))$ άμεσα προκύπτει ότι:

$$E(F_n(x)) = F(x).$$

Επιπλέον, έχουμε ότι:

$$\text{Var}(nF_n(x)) = nF(x)(1 - F(x)) \Leftrightarrow \text{Var}(F_n(x)) = \frac{F(x)(1 - F(x))}{n},$$

και επομένως $\text{Var}(F_n(x)) \rightarrow 0$, καθώς το $n \rightarrow \infty$, και η απόδειξη του β) ολοκληρώθηκε.

γ) Είναι γνωστό από τον ορισμό της εμπειρικής αθροιστικής συνάρτησης κατανομής ότι:

$$F_n(x) = \frac{\sum_{i=1}^n I_{(-\infty, x]}(X_i)}{n}, \quad x \in R, \quad (1.1.3)$$

όπου

$$I_C(X_i) = \begin{cases} 1, & \text{αν } X_i \in C \\ 0, & \text{αν } X_i \notin C. \end{cases} \quad (1.1.4)$$

Επομένως, η σ.σ. $F_n(x)$ είναι ο δειγματικός μέσος των τ.μ. $I_{(-\infty, x]}(X_i)$, η κατανομή των οποίων είναι Bernoulli με πιθανότητα επιτυχίας $F(x)$, μέση τιμή $F(x)$ και διακύμανση $F(x)(1-F(x))$. Εφαρμόζοντας το Κεντρικό Οριακό Θεώρημα προκύπτει άμεσα ότι:

$$F_n(x) = \frac{\sum_{i=1}^n I_{(-\infty, x]}(X_i)}{n} \stackrel{\text{ασυμπτ.}}{\sim} N\left(F(x), \frac{F(x)(1-F(x))}{n}\right). \quad (1.1.5)$$

Από τα παραπάνω προκύπτει ότι αν κάποιος ενδιαφέρεται να εκτιμήσει την α.σ.κ. $F(x)$, για κάθε x , τότε θα ήθελε να διαπιστώσει πόσο κοντά είναι η ε.α.σ.κ. στην α.σ.κ.. Στο επόμενο θεώρημα, η απόδειξη του οποίου παραλείπεται, θεμελιώνεται ότι με πιθανότητα 1 η σύγκλιση της $F_n(x)$ στην $F(x)$ είναι ομοιόμορφη στο x .

Θεώρημα 1.1.1 (Glivenco-Cantelli)

Αν $F_n(x)$ είναι η εμπειρική αθροιστική συνάρτηση κατανομής ενός τυχαίου δείγματος $X_1, \dots, X_n$, τότε:

$$\sup_x |F_n(x) - F(x)| \xrightarrow{a.s.} 0, \quad (1.1.6)$$

δηλαδή $P\left(\sup_x |F_n(x) - F(x)| \rightarrow 0\right) = 1$.

Τάξεις

Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από έναν πληθυσμό με συνεχή κατανομή. Επιπλέον υποθέτουμε ότι στο τ.δ. $X_1, \dots, X_n$, δεν υπάρχουν δεσμοί, δηλαδή $X_i \neq X_j$, για $i \neq j$, $i, j = 1, \dots, n$. Οι τάξεις $R_1, R_2, \dots, R_n$ αναφέρονται στο πως διατάσσονται οι δειγματικές παρατηρήσεις και ορίζονται ως εξής:

$$R_i = \text{ο αριθμός των } X_j, \text{ με } X_j \leq X_i, \quad i = 1, 2, \dots, n, \quad (1.1.7)$$

ή ισοδύναμα

$$R_i = \sum_{j=1, j \neq i}^n I_j, \quad i = 1, 2, \dots, n, \quad (1.1.8)$$

όπου

$$I_j = \begin{cases} 1, & \text{αν } X_j \leq X_i \\ 0, & \text{διαφορετικά.} \end{cases} \quad (1.1.9)$$

Στη συνέχεια παρατίθενται κάποιες ιδιότητες των τάξεων χρήσιμες σε όσα έπονται στη διατριβή αυτή.

Πρόταση 1.1.2 Αν $R_1, \dots, R_n$ είναι οι τάξεις ενός τυχαίου δείγματος $X_1, \dots, X_n$ από μία συνεχή α.σ.κ. τότε:

α) $R_1, \dots, R_n$ είναι ένα τ.δ. από τη διακριτή ομοιόμορφη κατανομή, με τιμές στο σύνολο $\{1, \dots, n\}$,

$$\beta) E(R_i) = \frac{n+1}{2} \text{ και } Var(R_i) = \frac{n^2-1}{12},$$

$$\gamma) Cov(R_i, R_j) = -\frac{n+1}{12}, \text{ για } i \neq j, \quad i, j = 1, \dots, n.$$

Απόδειξη (βλέπε μεταξύ άλλων Kvam and Vidakovic (2007))

α) Από τον τρόπο ορισμού τους και λαμβάνοντας υπόψη ότι δεν υπάρχουν δεσμοί στο δείγμα των $X_1, \dots, X_n$, προκύπτει άμεσα ότι οι δυνατές τιμές των τάξεων R_i ανήκουν στο σύνολο $\{1, 2, \dots, n\}$ και επιπλέον $P(R_i = j) = \frac{1}{n}$, $j = 1, \dots, n$ και η απόδειξη ολοκληρώνεται.

β) Από τον ορισμό της μέσης τιμής και της διακύμανσης μίας διακριτής τ.μ. προκύπτει ότι:

$$E(R_i) = \sum_{j=1}^n jP(R_i = j) = \sum_{j=1}^n j \frac{1}{n} = \frac{n(n+1)}{2} \frac{1}{n} = \frac{n+1}{2},$$

και

$$Var(R_i) = E(R_i^2) - [E(R_i)]^2,$$

όπου

$$E(R_i^2) = \sum_{j=1}^n j^2 P(R_i = j) = \sum_{j=1}^n j^2 \frac{1}{n} = \frac{n(n+1)(2n+1)}{6} \frac{1}{n} = \frac{(n+1)(2n+1)}{6}.$$

Επομένως,

$$Var(R_i) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}.$$

γ) Είναι:

$$Cov(R_i, R_j) = \sum_{k=1}^n \sum_{s=1, k \neq s}^n (k - E(R_i))(s - E(R_j))P(R_i = k, R_j = s),$$

και λαμβάνοντας υπόψη το α) και ότι

$$P(R_i = k, R_j = s) = P(R_i = k)P(R_j = s / R_i = k) = \frac{1}{n} \frac{1}{n-1},$$

προκύπτει:

$$\begin{aligned}
\text{Cov}(R_i, R_j) &= \sum_{k=1}^n \sum_{s=1, k \neq s}^n \left(k - \frac{n+1}{2}\right) \left(s - \frac{n+1}{2}\right) \frac{1}{n(n-1)} \\
&= \frac{1}{n(n-1)} \sum_{k=1}^n \left(k - \frac{n+1}{2}\right) \sum_{s=1}^n \left(s - \frac{n+1}{2}\right) - \frac{1}{n(n-1)} \sum_{k=1}^n \left(k - \frac{n+1}{2}\right)^2.
\end{aligned}$$

Όμως,

$$\sum_{k=1}^n \left(k - \frac{n+1}{2}\right) = \sum_{s=1}^n \left(s - \frac{n+1}{2}\right) = \frac{n(n+1)}{2} - \frac{n(n+1)}{2} = 0,$$

οπότε

$$\begin{aligned}
\text{Cov}(R_i, R_j) &= -\frac{1}{n(n-1)} \sum_{k=1}^n \left(k - \frac{n+1}{2}\right)^2 \\
&= -\frac{1}{n(n-1)} \left\{ \sum_{k=1}^n k^2 - 2 \frac{n+1}{2} \sum_{k=1}^n k + \sum_{k=1}^n \left(\frac{n+1}{2}\right)^2 \right\} \\
&= -\frac{1}{n(n-1)} \left\{ \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2 n}{2} + \frac{(n+1)^2 n}{4} \right\} \\
&= -\frac{n+1}{12}.
\end{aligned}$$

Ροές

Στις περισσότερες από τις στατιστικές μεθοδολογίες υποθέτουμε ότι το δείγμα είναι τυχαίο. Υπάρχουν όμως περιπτώσεις που αμφιβάλλουμε για την τυχειότητα των δειγματικών παρατηρήσεων. Ένας τρόπος ελέγχου της τυχειότητας ή μη ενός δείγματος είναι το λεγόμενο τεστ των ροών (Wald and Wolfowitz (1940)). Καθώς διάφοροι τρόποι ελέγχου της συμμετρίας στηρίζονται στην έννοια της ροής, στη συνέχεια παρατίθενται συνοπτικά κάποιες χρήσιμες έννοιες και ιδιότητες.

Έστω ότι στο δείγμα των n το πλήθος παρατηρήσεων $X_1, X_2, \dots, X_n$, n_1 από αυτές ανήκουν στην Ομάδα Α και οι τιμές αυτών αντικαθίστανται με το σύμβολο $+$, ενώ οι $n_2 = n - n_1$ το πλήθος τιμές των υπολοίπων παρατηρήσεων που ανήκουν στην Ομάδα Β αντικαθίστανται με το σύμβολο $-$. Για παράδειγμα, τα υπόλοιπα ενός μοντέλου γραμμικής παλινδρόμησης ανήκουν στην Ομάδα Α όταν είναι θετικά, ενώ διαφορετικά ανήκουν στην Ομάδα Β. Επομένως, κάθε δειγματική παρατήρηση αντικαθίσταται από το σύμβολο $+$ ή το σύμβολο $-$, τα οποία σύμβολα χαρακτηρίζουν κατά αυτόν τον τρόπο σε ποια ομάδα από τις δύο ανήκει η δειγματική παρατήρηση. **Ροή**, ονομάζεται μια ακολουθία ομοίων συμβόλων στο δείγμα των $+$ και $-$. Το πλήθος των όρων που αποτελούν αυτήν την ακολουθία ονομάζεται **μήκος ροής**. Έτσι για παράδειγμα στην περίπτωση της ακολουθίας συμβόλων:

++---++---+++

έχουμε ότι $n = 12$, $n_1 = 7$, $n_2 = 5$ και το πλήθος των ροών, έστω R , είναι ίσο με 5.

Έστω R ο αριθμός των ακολουθιών ομοίων συμβόλων στην ακολουθία n συμβόλων, εκ των οποίων n_1 και n_2 είναι το πλήθος των $+$ και $-$, αντίστοιχα. Από τον τρόπο ορισμού της τ.μ. R προκύπτει ότι μπορεί να γραφεί ως το παρακάτω άθροισμα: $R = 1 + \sum_{i=2}^n I_i$, όπου I_i , $i = 2, \dots, n$, μία δείκτρια συνάρτηση που λαμβάνει την τιμή 1 αν το i -οστό σύμβολο είναι διαφορετικό από το $(i-1)$ -οστό σύμβολο, για $i = 2, \dots, n$.

Πρόταση 1.1.3 Δοθέντος των n_1 και n_2 είναι

$$E(R|n_1, n_2) = 1 + \frac{2n_1n_2}{n}, \quad (1.1.10)$$

και

$$Var(R|n_1, n_2) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}. \quad (1.1.11)$$

Απόδειξη (βλέπε μεταξύ άλλων Gibbons and Chakraborti (2003))

Καθεμία τ.μ. I_i , $i = 2, \dots, n$, είναι διακριτή και ακολουθεί Bernoulli

κατανομή με παράμετρο $p = \frac{n_1n_2}{\binom{n}{2}} = \frac{2n_1n_2}{n(n-1)}$. Επομένως:

$$E(R) = E\left(1 + \sum_{i=2}^n I_i\right) = 1 + \sum_{i=2}^n E(I_i) = 1 + \sum_{i=2}^n \frac{2n_1n_2}{n(n-1)} = 1 + \frac{2n_1n_2}{n}.$$

Επιπλέον ισχύει ότι:

$$\begin{aligned} Var(R) &= Var\left(1 + \sum_{i=2}^n I_i\right) = Var\left(\sum_{i=2}^n I_i\right) = \sum_{i=2}^n Var(I_i) + \sum_{2 \leq i \neq j \leq n} Cov(I_i, I_j) \\ &= \sum_{i=2}^n Var(I_i) + \sum_{2 \leq i \neq j \leq n} \{E(I_i I_j) - E(I_i)E(I_j)\} \\ &= \sum_{i=2}^n \{E(I_i^2) - (E(I_i))^2\} + \sum_{2 \leq i \neq j \leq n} E(I_i I_j) - (n-2)(n-1)(E(I_i))^2 \\ &= (n-1)E(I_i^2) - (n-1)(E(I_i))^2 - (n-2)(n-1)(E(I_i))^2 + \sum_{2 \leq i \neq j \leq n} E(I_i I_j), \end{aligned}$$

όπου $E(I_i^2) = Var I_i + (E I_i)^2 = p(1-p) + p^2 = p = \frac{2n_1n_2}{n(n-1)}$. Επομένως, για

την εύρεση της διακύμανσης απαιτείται η εύρεση των $(n-1)(n-2)$ το

πλήθος ροπών τύπου $E(I_i I_j)$. Για τον υπολογισμό αυτών λαμβάνουμε υπόψη τα ακόλουθα: για τις $2(n-2)$ το πλήθος περιπτώσεις που είναι τέτοιες ώστε είτε $i = j-1$ είτε $i = j+1$ ισχύει ότι

$$E(I_i I_j) = \frac{n_1 n_2 (n_1 - 1) + n_2 n_1 (n_2 - 1)}{n(n-1)(n-2)} = \frac{n_1 n_2}{n(n-1)}, \text{ ενώ για τις υπόλοιπες}$$

$(n-1)(n-2) - 2(n-2) = (n-2)(n-3)$ το πλήθος περιπτώσεις ισχύει ότι:

$$E(I_i I_j) = \frac{4n_1 n_2 (n_1 - 1)(n_2 - 1)}{n(n-1)(n-2)(n-3)}.$$

Επομένως, συνδυάζοντας τα ενδιάμεσα αποτελέσματα, ύστερα από λίγη άλγεβρα προκύπτει ότι:

$$\text{Var}(R) = \frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n-1)}.$$

Διαδικασία Wiener ή κίνηση Brown (Χρυσόφινου (2008))

Η διαδικασία Wiener (1923) είναι μια στοχαστική διαδικασία συνεχούς χρόνου, η οποία εισήχθη στη βιβλιογραφία θέλοντας να περιγράψει με ένα μαθηματικό μοντέλο την άτακτη κίνηση ενός μορίου διαμέτρου 10^{-4} εκατοστών όταν αυτό βυθίζεται σε ένα υγρό ή αέριο. Συχνά αναφέρεται και ως κίνηση Brown, καθώς ο Brown ήταν ο πρώτος που παρατήρησε την άτακτη κίνηση του μορίου. Πριν προχωρήσουμε στο μαθηματικό ορισμό της διαδικασίας Wiener, για λόγους πληρότητας, αναφέρουμε ότι με τον όρο στοχαστική διαδικασία που προαναφέρθηκε εννοούμε ένα σύνολο τυχαίων μεταβλητών ορισμένων πάνω σε έναν κοινό πιθανο-θεωρητικό χώρο.

Ορισμός 1.1.2. Μια στοχαστική διαδικασία $\{X(t), t \geq 0\}$ είναι στοχαστική διαδικασία Wiener αν:

- α) έχει ανεξάρτητες προσauξήσεις, δηλαδή για κάθε $v > 2$ και για όλα τα $t_1 < t_2 < \dots < t_v$ οι προσauξήσεις, δηλαδή οι τ.μ. $X(t_2) - X(t_1), \dots, X(t_v) - X(t_{v-1})$, είναι ανεξάρτητες,
- β) έχει ομογενείς προσauξήσεις, δηλαδή οι προσauξήσεις $X(t) - X(s)$ και $X(t+h) - X(s+h)$, για κάθε $h > 0$ και $0 \leq s < t$, είναι ισόνομες,
- γ) για κάθε $t \geq 0$, η $X(t)$ κατανέμεται κανονικά με $E(X(t)) = 0$, $\forall t > 0$,
- δ) $X(0) = 0$ και η $X(t)$ είναι συνεχής στο μηδέν.

U-στατιστικές συναρτήσεις

Έστω $X_1, X_2, \dots, X_n$ τυχαίο δείγμα από κάποιο πληθυσμό. Ο όρος U -στατιστική συνάρτηση οφείλεται στον Hoeffding (1948) και ο ορισμός τους είναι ο ακόλουθος.

Έστω $h: \mathbb{R}^m \rightarrow \mathbb{R}$ μια συνάρτηση m το πλήθος μεταβλητών, όπου m ένας θετικός ακέραιος με $m \leq n$ και $h(\cdot)$ μια συνάρτηση συμμετρική στο όρισμα της και τέτοια ώστε $E[h(X_1, X_2, \dots, X_m)] < \infty$. Λέμε ότι η σ.σ. U_n είναι U -στατιστική συνάρτηση (U -σ.σ.) με kernel function $h(\cdot)$, τάξης m , αν είναι της μορφής

$$U_n = \binom{n}{m}^{-1} \sum_{C_{m,n}} h(X_{i_1}, X_{i_2}, \dots, X_{i_m}), \quad (1.1.12)$$

όπου το παραπάνω άθροισμα λαμβάνεται στο σύνολο $C_{m,n}$ όλων των $\binom{n}{m}$ το πλήθος δυνατών συνδυασμών των m ακεραίων $\{i_1, i_2, \dots, i_m\}$ με $i_1 < \dots < i_m$ από το σύνολο $\{1, 2, \dots, n\}$.

Ειδικές περιπτώσεις U -σ.σ.:

α) Για $m=1$ και $h(x)=x$ προκύπτει η δειγματική μέση τιμή.

β) Για $m=1$ και $h(x_i)=x_i^k$ προκύπτει η δειγματική ροπή k -τάξης.

γ) Για $m=1$ και παίρνοντας ως kernel function h τη δείκτρια συνάρτηση της σχέσης (1.1.4) προκύπτει η ε.α.σ.κ..

δ) Για $m=2$ και $h(x_1, x_2)=(x_1 - x_2)^2/2$ προκύπτει η δειγματική διακύμανση

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \frac{(X_i - X_j)^2}{2} = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = S^2.$$

ε) Για $m=2$ και $h(x_1, x_2)=|x_1 - x_2|$ προκύπτει η U -σ.σ.

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} |X_i - X_j|,$$

η οποία είναι γνωστή στη βιβλιογραφία ως μέση διαφορά του Gini (Gini's mean difference).

Kernel εκτίμηση της σ.π.π.

Οι συναρτήσεις kernel βρίσκουν εφαρμογή, μεταξύ άλλων, στη Μη- Παραμετρική στατιστική, όταν η υποκείμενη σ.π.π. είναι άγνωστη και πρέπει να εκτιμηθεί. Έστω $X_1, X_2, \dots, X_n$ τυχαίο δείγμα από την άγνωστη σ.π.π. f . Η kernel εκτίμηση της σ.π.π. f δίνεται από τη σχέση:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

με $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$, όπου $h > 0$ είναι μια παράμετρος, η οποία καλείται εύρος (bandwidth) και επηρεάζει σημαντικά την εκτίμηση, ενώ $K(\cdot)$ είναι μια συνάρτηση kernel (πυρήνα). Με τον όρο kernel συνάρτηση εννοούμε μια μη-αρνητική πραγματική, συμμετρική και ολοκληρώσιμη συνάρτηση που είναι τέτοια ώστε:

$$\int_{-\infty}^{+\infty} K(u) du = 1 \text{ και } K(-u) = K(u), \text{ για όλες τις τιμές του } u.$$

Η πρώτη συνθήκη εξασφαλίζει ότι η μέθοδος kernel για εκτίμηση σ.π.π. καταλήγει σε εκτίμηση που είναι σ.π.π., ενώ η δεύτερη συνθήκη εξασφαλίζει ότι η συνάρτηση kernel είναι συμμετρική. Το εύρος επηρεάζει σημαντικά το «άνοιγμα» του πυρήνα, δηλαδή πόσο διάστημα θα χαραχτεί γύρω από το σημείο x που γίνεται η εκτίμηση. Οι παρατηρήσεις που βρίσκονται σε αυτό διάστημα θα χρησιμοποιηθούν για την εκτίμηση της σ.π.π. σύμφωνα με τον ορισμό της συνάρτησης πυρήνα.

Διαισθητικά, η παράμετρος h επιλέγεται σύμφωνα με κάποιο κανόνα που "ισορροπεί" τη μεροληψία και τη διακύμανση του εκτιμητή, ώστε κάποιο μέτρο (π.χ. μέσο τετραγωνικό σφάλμα) να ελαχιστοποιείται. Χαρακτηριστικά αναφέρουμε ότι μερικές φορές προτείνεται η χρήση της τιμής $h = 1.06\hat{\sigma}n^{-1/5}$, όπου $\hat{\sigma}$ η δειγματική τυπική απόκλιση (βλέπε Chapter 3, από Silverman (1986)). Στον παρακάτω πίνακα παρατίθενται ορισμένες συναρτήσεις kernel, με $I(\cdot)$ να είναι η συνήθης δείκτρια συνάρτηση.

Πίνακας 1. Παραδείγματα συναρτήσεων kernel

Ονομασία	$K(\cdot)$
Ομοιόμορφη	$K(u) = \frac{1}{2} I(u \leq 1)$
Τριγωνική (Triangular)	$K(u) = (1 - u) I(u \leq 1)$
Epanechnikov	$K(u) = \frac{3}{4} (1 - u^2) I(u \leq 1)$
Quartic	$K(u) = \frac{15}{16} (1 - u^2)^2 I(u \leq 1)$
Gaussian	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Logistic	$K(u) = \frac{1}{e^u + 2 + e^{-u}}$
Tricube	$K(u) = \frac{70}{81} (1 - u ^3)^3 I(u \leq 1)$

1.2 Σκοπός της μεταπτυχιακής διατριβής και περίληψή της

Στη βιβλιογραφία έχουν εμφανιστεί διάφοροι τρόποι ελέγχου της μονοδιάστατης συμμετρίας, οι οποίοι ταξινομούνται σε δύο μεγάλες κατηγορίες. Η πρώτη κατηγορία περιλαμβάνει στατιστικά τεστ για τον έλεγχο της συμμετρίας μιας μονοδιάστατης κατανομής

γύρω από ένα γνωστό σημείο, ενώ η δεύτερη κατηγορία περιλαμβάνει τρόπους ελέγχου της συμμετρίας ως προς άγνωστο σημείο. Αντικείμενο αυτής της μεταπτυχιακής διατριβής είναι η κριτική ανασκόπηση και παρουσίαση των κυριότερων μεθόδων στατιστικού ελέγχου καθεμίας εκ των δύο παραπάνω κατηγοριών. Στο πλαίσιο αυτό, θα δίνεται η κεντρική ιδέα στην οποία στηρίζεται ο κάθε τρόπος ελέγχου και έπειτα θα παρατίθεται, υπό τη μηδενική υπόθεση, είτε η ακριβής είτε η ασυμπτωτική κατανομή της ελεγχοσυνάρτησης. Τέλος, θα δίνονται κάποιες ιδιότητες των στατιστικών τεστ. Στη συνέχεια, θα δοθεί μια σύντομη περίληψη του περιεχομένου της μεταπτυχιακής διατριβής.

Αντικείμενο μελέτης στο **Κεφάλαιο 2** (Έλεγχος συμμετρίας γύρω από γνωστό σημείο) είναι η κριτική ανασκόπηση, η παρουσίαση και ταξινόμηση, αν είναι εφικτή, των κυριότερων μεθόδων ελέγχου της υπόθεσης της συμμετρίας γύρω από γνωστό σημείο. Οι τρόποι ελέγχου που θα παρουσιασθούν στηρίζονται κυρίως στο προσημικό τεστ (sign test), στο τεστ των ροών (runs test), στον έλεγχο προσημασμένης διάταξης του Wilcoxon (Wilcoxon signed-rank test) και στην ε.α.σ.κ.

Στο **Κεφάλαιο 3** (Έλεγχος συμμετρίας γύρω από άγνωστο σημείο) το ενδιαφέρον επικεντρώνεται στην κριτική ανασκόπηση, στην παρουσίαση και ταξινόμηση, αν είναι εφικτή, των κυριότερων μεθόδων ελέγχου της υπόθεσης της συμμετρίας γύρω από άγνωστο σημείο. Αρχικά, θα παρουσιαστούν οι τρόποι ελέγχου που έχουν εισαχθεί στη βιβλιογραφία και προέκυψαν τροποποιώντας κατάλληλα κάποιους από τους ελέγχους του προηγούμενου κεφαλαίου. Έπειτα, θα παρουσιαστούν έλεγχοι που στηρίζονται, μεταξύ άλλων, σε κατάλληλα ορισμένα περιγραφικά μέτρα και στην εμπειρική χαρακτηριστική συνάρτηση.

Στο **Κεφάλαιο 4** (Συγκριτική μελέτη) το ενδιαφέρον επικεντρώνεται στη σύγκριση των μεθόδων των προηγούμενων κεφαλαίων. Ειδικότερα, οι μεθοδολογίες που παρουσιάστηκαν θα

συγκριθούν τόσο ως προς την ικανοποίηση κάποιων ιδιοτήτων (σ.σ. αναλλοίωτες σε μετασχηματισμούς και ανεξάρτητες κατανομής υπό τη μηδενική υπόθεση), όσο και ως προς την απόδοση τους να διατηρούν το επίπεδο σημαντικότητας και να έχουν καλή ισχύ. Στο πλαίσιο αυτό, θα δοθεί μια ανασκόπηση των μελετών προσομοίωσης που έχουν εμφανιστεί στη βιβλιογραφία.

Στο **Κεφάλαιο 5** (Μια εφαρμογή) οι κυριότερες μέθοδοι ελέγχου συμμετρίας θα εφαρμοστούν σε ένα σύνολο δεδομένων.

Στο **Παράρτημα Α** παρατίθενται κάποιοι χρήσιμοι πίνακες, ενώ στο **Παράρτημα Β** δίνονται οι απαραίτητοι κώδικες που χρησιμοποιήθηκαν για την υλοποίηση των ελέγχων και την απόκτηση των αποτελεσμάτων που δίνονται στο Κεφάλαιο 5. Τέλος, η διατριβή ολοκληρώνεται με την περίληψή της στα Αγγλικά (**Abstract of MSc Dissertation**) και τη **Βιβλιογραφία**.

ΚΕΦΑΛΑΙΟ 2

Έλεγχος μονοδιάστατης συμμετρίας γύρω από γνωστό σημείο

Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από έναν πληθυσμό με συνεχή και διαφορίσιμη α.σ.κ., έστω F , και σ.π.π. f . Όπως έχει ήδη αναφερθεί στο εισαγωγικό κεφάλαιο, ο έλεγχος της συμμετρίας έγκειται στην ανάπτυξη στατιστικών μεθοδολογιών σχετικά με την απόφαση περί αποδοχής ή απόρριψη της μηδενικής υπόθεσης

$$H_0 : F(\mu + x) = 1 - F(\mu - x), \text{ για κάθε } x. \quad (2.1)$$

Στο κεφάλαιο αυτό, το ενδιαφέρον μας επικεντρώνεται στον έλεγχο της παραπάνω μηδενικής υπόθεσης στην ειδική περίπτωση που το κέντρο συμμετρίας μ είναι γνωστό. Τότε, χωρίς βλάβη της γενικότητας, εφαρμόζοντας τον μετασχηματισμό $X_i - \mu, i = 1, \dots, n$, ο παραπάνω έλεγχος ανάγεται στον έλεγχο της συμμετρίας γύρω από το μηδέν, δηλαδή της υπόθεσης

$$H_0 : F(x) = 1 - F(-x), \text{ για κάθε } x, \quad (2.2)$$

έναντι μίας εκ των τριών εναλλακτικών που προκύπτουν από τις σχέσεις (1.4) έως (1.6) θέτοντας $\mu = 0$.

Πολλοί συγγραφείς έχουν προτείνει τεστ για τον έλεγχο της παραπάνω υπόθεσης. Αυτοί οι τρόποι ελέγχου στηρίζονται κυρίως στο προσημικό τεστ (sign test), στο τεστ των ροών (runs test), στον έλεγχο προσημασμένης διάταξης του Wilcoxon (Wilcoxon signed-rank test) και στην ε.α.σ.κ.. Αντικείμενο μελέτης αυτού του κεφαλαίου είναι η κριτική ανασκόπηση, η παρουσίαση και ταξινόμηση, αν είναι εφικτή, των κυριότερων μεθόδων ελέγχου της παραπάνω υπόθεσης. Στο πλαίσιο αυτό, θα δίνεται η κεντρική ιδέα στην οποία στηρίζεται ο κάθε

τρόπος ελέγχου και έπειτα θα παρατίθεται είτε η ακριβής είτε η ασυμπτωτική κατανομή της ελεγχοσυνάρτησης υπό τη μηδενική υπόθεση. Τέλος, θα δίνονται κάποιες ιδιότητες των στατιστικών τεστ. Η συγκριτική μελέτη των μεθόδων που θα παρουσιαστούν σε αυτό το κεφάλαιο αποτελεί αντικείμενο μελέτης του 4^{ου} Κεφαλαίου.

2.1 Τεστ που στηρίζονται στο προσημικό τεστ

Στην ενότητα αυτή θα παρουσιασθούν τρόποι ελέγχου της μονοδιάστατης συμμετρίας μιας συνεχούς κατανομής γύρω από γνωστό κέντρο συμμετρίας που στηρίζονται στο προσημικό τεστ. Στο πλαίσιο αυτό, αρχικά, θα παρουσιαστεί ο έλεγχος του προσημικού τεστ. Έπειτα, θα παρουσιαστεί μια τροποποίηση του που εμφανίστηκε πρόσφατα στη βιβλιογραφία και τέλος θα παρουσιαστεί ο έλεγχος προσημασμένης διάταξης (signed rank test).

2.1.1 Ο έλεγχος συμμετρίας με το προσημικό τεστ

Στο εδάφιο αυτό θα παρουσιαστεί ο ευκολότερος τρόπος ελέγχου της υπόθεσης ότι n το πλήθος παρατηρήσεις προέρχονται από έναν πληθυσμό με συνεχή κατανομή, συμμετρική, χωρίς βλάβη της γενικότητας, γύρω από το μηδέν. Ο έλεγχος αυτός δεν είναι άλλος από το προσημικό τεστ (sign test).

Η μέθοδος του προσημικού ελέγχου στηρίζεται στο πρόσημο των δειγματικών τιμών $X_1, X_2, \dots, X_n$. Στο σημείο αυτό επισημαίνουμε ότι οι παρατηρήσεις είναι τέτοιες ώστε να μην υπάρχουν δεσμοί, δηλαδή $X_i \neq 0$, για $i = 1, \dots, n$. Έστω T_s το πλήθος των θετικών X_i , δηλαδή ο αριθμός των θετικών προσήμων, δηλαδή,

$$T_s = \sum_{i=1}^n I(X_i > 0), \quad (2.1.1.1)$$

όπου $I(\cdot)$ η συνήθης δείκτρια συνάρτηση. Υπό την υπόθεση της συμμετρίας, η στατιστική συνάρτηση (σ.σ.) T_S ουσιαστικά περιγράφει τον αριθμό των επιτυχιών σε n δοκιμές Bernoulli, όπου επιτυχία θεωρείται η εμφάνιση θετικού προσήμου, με πιθανότητα επιτυχίας υπό την υπόθεση της συμμετρίας $p=0.5$. Αυτό συμβαίνει επειδή υπό την H_0 , εξ ορισμού, $P(X > 0) = 0.5 = P(X < 0)$. Επιπλέον, η πιθανότητα αυτή παραμένει αμετάβλητη από δοκιμή σε δοκιμή. Επομένως, υπό την H_0 , η σ.σ. T_S ακολουθεί διωνυμική κατανομή με παραμέτρους n και $p=0.5$. Δηλαδή $T_S \stackrel{H_0}{\sim} B(n, p=0.5)$, με $P(T_S = t | H_0 \text{ αληθής}) = \binom{n}{t} 0.5^n$.

Γίνεται αντιληπτό ότι η υπόθεση της συμμετρίας που διατυπώθηκε στη σχέση (1.3), απορρίπτεται για επίπεδο σημαντικότητας α , έναντι της εναλλακτικής υπόθεσης της ύπαρξης θετικής (αρνητικής) λοξότητας, για μεγάλες (μικρές) τιμές του T_S . Ειδικότερα, στην πρώτη περίπτωση απορρίπτεται αν $T_S \geq t_{\alpha_1}$, όπου το t_{α_1} είναι ο μικρότερος ακέραιος για τον οποίο ισχύει $\sum_{t=t_{\alpha_1}}^n \binom{n}{t} 0.5^n \leq \alpha$, ενώ στη δεύτερη περίπτωση αν $T_S \leq t_{\alpha_2}$, όπου t_{α_2} είναι ο μεγαλύτερος ακέραιος για τον οποίο ισχύει $\sum_{t=0}^{t_{\alpha_2}} \binom{n}{t} 0.5^n \leq \alpha$. Τέλος, η H_0 απορρίπτεται για επίπεδο σημαντικότητας α , έναντι της εναλλακτικής υπόθεσης ότι η F είναι ασύμμετρη, αν $T_S \leq t_{\alpha_3}$ ή $T_S \geq t_{\alpha_4}$, όπου t_{α_3} είναι ο μεγαλύτερος ακέραιος τέτοιος ώστε $\sum_{t=0}^{t_{\alpha_3}} \binom{n}{t} 0.5^n \leq \alpha/2$ και t_{α_4} ο μικρότερος ακέραιος έτσι ώστε $\sum_{t=t_{\alpha_4}}^n \binom{n}{t} 0.5^n \leq \alpha/2$. Στον Πίνακα 1 του

Παραρτήματος Α δίνονται οι τιμές της α.σ.κ. της $B(n, p = 0.5)$ για τον ευκολότερο υπολογισμό των παραπάνω κρίσιμων τιμών.

Για να είναι πιο εύκολα εφαρμόσιμος στην πράξη ο παραπάνω έλεγχος και καθώς στηρίζεται σε μια διακριτή τ.μ., που έχει ως συνέπεια περιορισμένες επιλογές του επιπέδου σημαντικότητας α , μπορεί να χρησιμοποιηθεί η προσέγγιση της διωνυμικής κατανομής από την κανονική. Ειδικότερα, στην επόμενη πρόταση προσδιορίζεται η ασυμπτωτική κατανομή της σ.σ. T_S υπό τη μηδενική υπόθεση.

Πρόταση 2.1.1.1 Για μεγάλες τιμές του μεγέθους δείγματος n , υπό την υπόθεση της συμμετρίας γύρω από το 0, ισχύει ότι:

$$Z_s = \frac{T_s - 0.5n}{\sqrt{0.25n}} \stackrel{H_0}{\underset{\text{προσ.}}{\sim}} N(0,1). \quad (2.1.1.2)$$

Επομένως, η H_0 απορρίπτεται έναντι της θετικής λοξότητας αν $Z_s \geq z_\alpha$, έναντι της αρνητικής λοξότητας αν $Z_s \leq -z_\alpha$ και έναντι της ασυμμετρίας αν $|Z_s| \geq z_{\alpha/2}$, όπου z_γ είναι η τιμή εκείνη για την οποία $P(Z \geq z_\gamma) = \gamma$, με $Z \sim N(0,1)$.

Εναλλακτικά, μπορεί να χρησιμοποιηθεί και η σ.σ. που προκύπτει με διόρθωση συνεχείας και δίνεται από τη σχέση:

$$Z'_s = \begin{cases} \frac{T_s - 0.5 - 0.5n}{\sqrt{0.25n}}, & \text{αν } T_s > n/2 \\ \frac{T_s + 0.5 - 0.5n}{\sqrt{0.25n}}, & \text{αν } T_s < n/2. \end{cases} \quad (2.1.1.3)$$

2.1.2 Ο έλεγχος των Cheng and Balakrishnan (2004)

Στη βιβλιογραφία έχουν παρουσιασθεί διάφορες τροποποιήσεις του προσημικού τεστ για τον έλεγχο της υπόθεσης ότι μια συνάρτηση κατανομής είναι συμμετρική γύρω από το μηδέν. Μια από τις τροποποιήσεις αυτές είναι αυτή που προτάθηκε από τους Cheng and Balakrishnan (2004). Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n , από μια συνεχή αθροιστική συνάρτηση κατανομής F και ας είναι $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, οι διατεταγμένες τιμές με βάση την απόλυτη τιμή, διατηρώντας τα πρόσημα. Σε όσα ακολουθούν συμβολίζουμε με $S_1, S_2, \dots, S_n$, τις δείκτριες μεταβλητές οι οποίες προσδιορίζουν το πρόσημο των $X_{(i)}$, δηλαδή:

$$S_i = \begin{cases} 1, & \text{αν } X_{(i)} \geq 0 \\ 0, & \text{διαφορετικά.} \end{cases} \quad (2.1.2.1)$$

Με βάση αυτό τον συμβολισμό, προκύπτει ότι η σ.σ. του προσημικού τεστ που ορίστηκε στη σχέση (2.1.1.1), δεν είναι τίποτε άλλο παρά ο αριθμός των θετικών $X_{(i)}$.

Χρησιμοποιώντας τον παραπάνω συμβολισμό, ο έλεγχος που προτάθηκε από τους Cheng and Balakrishnan (2004) βασίζεται στο άθροισμα των k τελευταίων όρων της ακολουθίας $S_1, S_2, \dots, S_n$, δηλαδή στη σ.σ.

$$C_k = S_{n-k+1} + S_{n-k+2} + \dots + S_n, \quad (2.1.2.2)$$

όπου τα S_i δίνονται στη σχέση (2.1.2.1).

Σε αναλογία με το προσημικό τεστ, είναι προφανές ότι η μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν, απορρίπτεται έναντι της εναλλακτικής ότι η F είναι ασύμμετρη, για μικρές ή μεγάλες

τιμές της σ.σ. C_k . Για την εύρεση της κρίσιμης τιμής του ελέγχου είναι απαραίτητη η εύρεση της κατανομής της σ.σ. C_k , υπό την υπόθεση της συμμετρίας. Η κατανομή αυτή, με παρόμοιο σκεπτικό με αυτό για τον προσδιορισμό της κατανομής της σ.σ. T_s , προσδιορίστηκε από τους Cheng and Balakrishnan (2004) και παρατίθεται στο θεώρημα που ακολουθεί.

Θεώρημα 2.1.2.1 (Cheng and Balakrishnan (2004)) Υπό τη μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν, ισχύει ότι η σ.σ. C_k ακολουθεί διωνυμική κατανομή με παραμέτρους k και $1/2$. Δηλαδή,

$$C_k \stackrel{H_0}{\sim} B(k, 1/2). \quad (2.1.2.3)$$

Επομένως, για επίπεδο σημαντικότητας α , απορρίπτεται η H_0 έναντι της εναλλακτικής ότι η F είναι ασύμμετρη αν $C_k \leq t_{\alpha_1}$ ή $C_k \geq t_{\alpha_2}$, όπου t_{α_1} είναι ο μεγαλύτερος ακέραιος για τον οποίο ισχύει

$$\sum_{t=0}^{t_{\alpha_1}} \binom{k}{t} 0.5^k \leq \alpha/2, \text{ ενώ } t_{\alpha_2} \text{ είναι ο μικρότερος ακέραιος τέτοιος ώστε}$$

$$\sum_{t=t_{\alpha_2}}^k \binom{k}{t} 0.5^k \leq \alpha/2.$$

Γίνεται αντιληπτό από τα παραπάνω ότι για δοθέν επίπεδο σημαντικότητας α , οι κρίσιμες τιμές εξαρτώνται από την επιλογή του k . Με μελέτη προσομοίωσης οι Cheng and Balakrishnan (2004) κατέληξαν να προτείνουν την τιμή $k=6$, δηλαδή να υπολογίζεται το άθροισμα των 6 τελευταίων όρων της ακολουθίας $S_1, S_2, \dots, S_n$. Σε αυτήν την περίπτωση προκύπτει ότι η κρίσιμη περιοχή θα είναι: $C_6 \geq 6$ ή $C_6 \leq 0$, καθώς ισχύει ότι $P(C_6 \geq 6 \text{ ή } C_6 \leq 0) = 0.03125 < 0.05$.

Επομένως, το πραγματικό επίπεδο σημαντικότητας του ελέγχου είναι 3.125%.

Στην πράξη είναι προτιμότερο να χρησιμοποιείται η προσέγγιση της διωνυμικής από την κανονική κατανομή. Δηλαδή, χρησιμοποιείται ότι:

$$Z_{CB} = \frac{C_k - 0.5k}{\sqrt{0.25k}} \stackrel{H_0}{\sim}_{\text{προσ.}} N(0,1), \quad (2.1.2.4)$$

με κρίσιμη περιοχή του ελέγχου την $|Z_{CB}| \geq z_{\alpha/2}$. Επιπλέον, προτείνεται να χρησιμοποιείται η σ.σ. που προκύπτει με διόρθωση συνεχείας, δηλαδή η:

$$Z'_{CB} = \begin{cases} \frac{C_k + 0.5 - 0.5k}{\sqrt{0.25}}, & \text{αν } C_k > \frac{k}{2} \\ \frac{C_k - 0.5 - 0.5k}{\sqrt{0.25k}}, & \text{αν } C_k < \frac{k}{2}. \end{cases} \quad (2.1.2.5)$$

Παρατήρηση 2.1.2.1 Οι Cheng and Balakrishnan (2004) έδωσαν έναν εναλλακτικό τρόπο έκφρασης της σ.π. της σ.σ. C_k . Ειδικότερα, απέδειξαν ότι οι πιθανότητες $P(C_k = l)$, για $l = 0, 1, \dots, k$, μπορούν να γραφούν σε όρους της F , καθώς

$$P(C_k = l) = \sum_{l=s_{n-k+1}+\dots+s_n} \dots \sum P(S_{n-k+1} = s_{n-k+1}, \dots, S_n = s_n),$$

όπου καθένα από τα s_i παίρνουν τιμή 0 ή 1. Παρότι κάθε τέτοιος όρος φαίνεται περίπλοκος, μπορεί να γραφεί ως ένα πολλαπλό ολοκλήρωμα το οποίο περιλαμβάνει την F . Ένα πλεονέκτημα αυτού του τρόπου έκφρασης είναι ότι αντικαθιστώντας την F με την α.σ.κ. μιας εναλλακτικής κατανομής, μπορεί να προσδιοριστεί η ισχύς του τεστ.

2.1.3 Έλεγχοι προσημασμένης διάταξης

Οι προσημασμένης διάταξης σ.σ. που υιοθετήθηκαν για τον έλεγχο της συμμετρίας γύρω από γνωστό σημείο, αρχικά από τους Hajek and Sidak (1967), ορίζονται ως εξής:

Ορισμός 2.1.3.1 (Hajek and Sidak (1967)) Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από έναν πληθυσμό με συνεχή α.σ.κ. F . Επιπλέον, έστω ότι R_i^+ είναι η τάξη του $|X_i|$ στην ακολουθία $|X_1|, |X_2|, \dots, |X_n|$. Λέμε ότι μία σ.σ. είναι προσημασμένης διάταξης αν είναι της μορφής:

$$T_n = \sum_{i=1}^n \text{sign}(X_i) a_n(R_i^+), \quad (2.1.3.1)$$

όπου

$$\text{sign}(X_i) = \begin{cases} -1, & \text{αν } X_i < 0 \\ 0, & \text{αν } X_i = 0 \\ 1, & \text{αν } X_i > 0 \end{cases}, \quad (2.1.3.2)$$

και $a_n(1), a_n(2), \dots, a_n(n)$ είναι τα λεγόμενα σκορ.

Από τον προηγούμενο ορισμό γίνεται σαφές ότι συγκεκριμένες επιλογές για τον καθορισμό των σκορ $a_n(1), a_n(2), \dots, a_n(n)$ οδηγούν σε ειδικές περιπτώσεις σ.σ.. Στη συνέχεια, παρατίθενται κάποιες ειδικές περιπτώσεις σ.σ. που έχουν μελετηθεί στη βιβλιογραφία και παρουσιάζουν ενδιαφέρον (βλέπε μεταξύ άλλων Sen (2006) και Prentice (2006)).

Προσημικό τεστ: Αν στη σχέση (2.1.3.1) θέσουμε $a_n(i) = 1$, για $i = 1, \dots, n$, τότε η σ.σ. T_n ανάγεται στη σ.σ. που παριστάνει τη διαφορά

του πλήθους των θετικών και αρνητικών X_i . Προφανώς, καθώς το πλήθος των παρατηρήσεων είναι δεδομένο, ο έλεγχος αυτός είναι ισοδύναμος με τον προσημικό έλεγχο της προηγούμενης ενότητας. Επιπρόσθετα, σε αυτήν την περίπτωση είναι σαφές ότι δεν χρησιμοποιείται καθόλου η πληροφορία που δίνεται από τις τάξεις R_i^+ .

Η προσημασμένης διάταξης σ.σ. του Wilcoxon: Αν στη σχέση (2.1.3.1) θέσουμε $a_n(i) = i$, για $i = 1, \dots, n$, τότε η σ.σ. T_n ανάγεται στη σ.σ.

$T_n' = \sum_{i=1}^n \text{sign}(X_i) R_i^+$, η οποία είναι η προσημασμένης διάταξης σ.σ. του Wilcoxon (1945).

Στατιστική συνάρτηση με κανονικά σκορ: Σε αυτήν την περίπτωση $a_n(i) = E(Y_{i:n})$, όπου $Y_{i:n}$ είναι η i -οστή διατεταγμένη παρατήρηση ενός τ.δ. μεγέθους n από την X -τετράγωνο κατανομή με ένα βαθμό ελευθερίας, X_1^2 (Fraser 1957). Μια παραλλαγή προκύπτει θεωρώντας ότι:

$$a_n(i) = \Phi^{-1}\left(\frac{1+i/(n+1)}{2}\right), \quad (2.1.3.3)$$

όπου Φ η α.σ.κ. της τυπικής κανονικής.

Όλες οι παραπάνω περιπτώσεις μπορούν να προκύψουν θεωρώντας ότι:

$$a_n(i) = \varphi^*\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n, \quad (2.1.3.4)$$

όπου

$$\varphi^*(u) = \varphi((1+u)/2), \quad 0 < u < 1, \quad (2.1.3.5)$$

με φ μια συνάρτηση τέτοια ώστε:

$$\varphi(u) + \varphi(1-u) = 0, \text{ για κάθε } u \in (0,1). \quad (2.1.3.6)$$

Ειδικότερα, το προσημικό τεστ προκύπτει όταν $\varphi^*(u) = \text{sign}(2u-1)$, η προσημασμένης διάταξης σ.σ. του Wilcoxon αν $\varphi^*(u) = 2u-1$, ενώ τέλος η σ.σ. με κανονικά σκορ όταν $\varphi^*(u) = \Phi^{-1}(u)$.

Άλλες τρεις ειδικές περιπτώσεις που παρουσιάζουν ενδιαφέρον είναι οι ακόλουθες:

Στατιστική συνάρτηση του Savage (1959): Σε αυτήν την περίπτωση

$$a_n(i) = n^{-1} + (n-1)^{-1} + \dots + (n-i+1)^{-1}, \quad i=1, \dots, n. \quad (2.1.3.7)$$

Στατιστική συνάρτηση του Van der Waerden (1953): Σε αυτήν την περίπτωση

$$a_n(i) = \Phi^{-1}\left(\frac{i}{n+1}\right), \quad i=1, \dots, n. \quad (2.1.3.8)$$

Στατιστική συνάρτηση του Koziol (1980): Η σ.σ. που προτάθηκε από τον Koziol (1980) προκύπτει ότι είναι ασυμπτωτικά ισοδύναμη με τη σ.σ. που προκύπτει ως ειδική περίπτωση της T_n όταν

$$a_n(i) = \varphi_1\left(\frac{1}{2} + i/2n\right), \quad i=1, 2, \dots, n, \quad (2.1.3.9)$$

όπου

$$\varphi_1(t) = \sqrt{2} \cos(\pi t), \quad 0 \leq t \leq 1. \quad (2.1.3.10)$$

Παρατήρηση 2.1.3.1 (Huskova (1984)) Τα σκορ $a_n(i)$, $i=1, \dots, n$, συνήθως συνδέονται με μια τετραγωνικά ολοκληρώσιμη συνάρτηση φ , ορισμένη στο διάστημα $(0,1)$, μέσω της σχέσης:

$$\lim_{n \rightarrow \infty} \int_0^1 \left(a_n([un] + 1) - \varphi(u) \right)^2 du = 0. \quad (2.1.3.11)$$

Τότε προκύπτει η σ.σ. $T_n(\varphi)$ αντί της T_n . Συνήθως,

$$a_n(i) = E\left(\varphi(U_{(i)})\right), \quad i = 1, 2, \dots, n, \quad (2.1.3.12)$$

ή

$$a_n(i) = \varphi\left(\frac{i}{n+1}\right), \quad i = 1, 2, \dots, n, \quad (2.1.3.13)$$

όπου $U_{(i)}$ η i -οστή διατεταγμένη παρατήρηση ενός δείγματος μεγέθους n από την ομοιόμορφη κατανομή στο διάστημα $(0,1)$. Η πρώτη επιλογή για τα $a_n(i)$ ικανοποιεί τη σχέση (2.1.3.11), ενώ η δεύτερη επιλογή την ικανοποιεί αν η συνάρτηση φ μπορεί να γραφεί ως διαφορά μονότονων συναρτήσεων. Επιπλέον, η Huskova (1984) παραθέτει την ακόλουθη ισοδύναμη μορφή της σ.σ. T_n :

$$T_n = \sum_{j=1}^n a_n(j) T_j - \sum_{j=1}^n a_j(j) / 2, \quad (2.1.3.14)$$

όπου

$$T_i = \begin{cases} 1, & \text{αν η } i\text{-οστή μικρότερη παρατήρηση κατά απόλυτη τιμή είναι θετική} \\ 0, & \text{αλλιώς.} \end{cases}$$

Οι σ.σ. T_n έχουν προταθεί για τον έλεγχο της συμμετρίας γύρω από το μηδέν, έναντι της εναλλακτικής υπόθεσης της ύπαρξης τόσο γενικής ασυμμετρίας, όσο και των εναλλακτικών υποθέσεων ότι το τ.δ. προέρχεται από την α.σ.κ $F(x-\theta)$ με $\theta > 0$ ή $\theta < 0$. Από τον τρόπο ορισμού της σ.σ. T_n προκύπτει άμεσα ότι η κρίσιμη περιοχή των παραπάνω ελέγχων, για επίπεδο σημαντικότητας α , είναι $|T_n| > c_{\alpha/2}$, $T_n > c_\alpha$ και $T_n < -c_\alpha$, αντίστοιχα, όπου c_α είναι το σημείο εκείνο τέτοιο

ώστε $P_{H_0}(T_n > c_\alpha) = \alpha$. Επομένως, για τον καθορισμό του απαιτείται ο προσδιορισμός της κατανομής της σ.σ. T_n υπό την υπόθεση της συμμετρίας. Έστω ότι $S = (\text{sign}(X_1), \text{sign}(X_2), \dots, \text{sign}(X_n))$ και $R^+ = (R_1^+, R_2^+, \dots, R_n^+)$. Υπό την υπόθεση της συμμετρίας γύρω από το μηδέν, ισχύει ότι (βλέπε Hajek και Sidak (1967)) τα διανύσματα S και R^+ είναι ανεξάρτητα. Επιπλέον, καθώς έχουμε αποκλείσει την περίπτωση ύπαρξης δεσμών εύκολα προκύπτει ότι:

$$P_{H_0}(\text{sign}(X_1) = s_1, \text{sign}(X_2) = s_2, \dots, \text{sign}(X_n) = s_n) = 2^{-n}, \quad (2.1.3.15)$$

με $s_i \in \{-1, 1\}$. Σε αυτήν την περίπτωση το διάνυσμα $R^+ = (R_1^+, R_2^+, \dots, R_n^+)$ παριστάνει μια τυχαία μετάθεση του $(1, 2, \dots, n)$. Οπότε:

$$P_{H_0}(R_1^+ = r_1, \dots, R_n^+ = r_n) = \frac{1}{n!}, \quad (2.1.3.16)$$

με $(r_1, r_2, \dots, r_n)$ να είναι οι δυνατές μεταθέσεις του $(1, \dots, n)$.

Είναι προφανές (βλέπε Sen (2006)) ότι καθώς η σ.σ. T_n εξαρτάται από τις δειγματικές παρατηρήσεις μόνο μέσω των τιμών των διανυσμάτων S και R^+ , υπό τη μηδενική υπόθεση, η κατανομή της είναι ανεξάρτητη της F . Επιπλέον, η κατανομή της T_n υπό τη μηδενική υπόθεση μπορεί να προσδιοριστεί λαμβάνοντας υπόψη τα $2^n n!$ ισοπίθανα ενδεχόμενα πραγματοποίησης των $s_i \in \{-1, 1\}$ και $(r_1, r_2, \dots, r_n)$. Είναι σαφές ότι ο παραπάνω προσδιορισμός είναι εφικτός για δοθείσες μορφές των $a_n(i)$, $i = 1, 2, \dots, n$ και για μικρά σε μέγεθος δείγματα και υπάρχουν διαθέσιμοι πίνακες στη βιβλιογραφία (βλέπε Owen (1962)). Για την ευκολότερη εφαρμογή του ελέγχου στην πράξη, το ενδιαφέρον επικεντρώθηκε στην εύρεση της ασυμπτωτικής κατανομής της σ.σ. T_n . Ο προσδιορισμός της ασυμπτωτικής κατανομής

της σ.σ. υπό την υπόθεση της συμμετρίας, περιορίζεται στην περίπτωση που η F ανήκει στην οικογένεια $\mathbb{F}_1^* \subset \mathbb{F}_1$, όπου $\mathbb{F}_1$ είναι η οικογένεια των συναρτήσεων κατανομής που είναι συνεχείς και συμμετρικές γύρω από το μηδέν, ενώ $\mathbb{F}_1^*$ είναι η οικογένεια των συνεχών συναρτήσεων κατανομής με απόλυτα συνεχείς σ.π.π. f και πεπερασμένο μέτρο πληροφορίας Fisher, δηλαδή η οικογένεια των κατανομών που ικανοποιούν τη σχέση:

$$0 < \int_{-\infty}^{+\infty} (f'(x))^2 (f(x))^{-1} dx < +\infty. \quad (2.1.3.17)$$

Τότε, καθώς υπό την μηδενική υπόθεση ισχύουν

$$E_{H_0} T_n = 0 \text{ και } \text{Var}_{H_0} T_n = \sum_{i=1}^n a_n^2(i),$$

με εφαρμογή του Κεντρικού Οριακού Θεωρήματος προκύπτει το ακόλουθο αποτέλεσμα.

Θεώρημα 2.1.3.1 (Hajek and Sidak (1967), Huskova (1984)) Αν

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} a_n^2(j) \left(\sum_{j=1}^n a_n^2(j) \right)^{-1} = 0, \quad (2.1.3.18)$$

τότε

$$\left(T_n \left(\sum_{j=1}^n a_n^2(j) \right)^{-1/2} \middle| H_0 \right) \xrightarrow[n \rightarrow \infty]{d} N(0,1). \quad (2.1.3.19)$$

Επιπλέον, αν τα σκορ $a_n(j)$, $j=1, 2, \dots, n$ ικανοποιούν τη σχέση

$$\lim_{n \rightarrow \infty} \int_0^1 (a_n([un]+1) - \varphi(u))^2 du = 0, \quad (2.1.3.20)$$

τότε η ποσότητα $\sum_{j=1}^n a_n^2(j)$ μπορεί να αντικατασταθεί από την $n \int_0^1 \varphi^2(u) du$.

Επομένως, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της γενικής ασυμμετρίας αν $\left| T_n \left(\sum_{j=1}^n a_n^2(j) \right)^{-1/2} \right| > z_{\alpha/2}$, απορρίπτεται έναντι της εναλλακτικής υπόθεσης ότι το τ.δ. προέρχεται από την α.σ.κ. $F(x-\theta)$ με $\theta > 0$ αν $T_n \left(\sum_{j=1}^n a_n^2(j) \right)^{-1/2} > z_\alpha$, ενώ τέλος απορρίπτεται έναντι της εναλλακτικής υπόθεσης ότι το τ.δ. προέρχεται από την α.σ.κ. $F(x-\theta)$ με $\theta < 0$ αν $T_n \left(\sum_{j=1}^n a_n^2(j) \right)^{-1/2} < -z_\alpha$.

Παρατήρηση 2.1.3.2 Θεωρώντας όπου $[\cdot]$ το ακέραιο μέρος, οι Hajek and Sidak (1967) υπό τη συνθήκη ότι τα $a_n(j)$, $j=1,2,\dots,n$, είναι τέτοια ώστε:

$$\lim_{n \rightarrow \infty} \int_0^1 \left(a_n([un]+1) - \varphi(u) \right)^2 du = 0, \quad (2.1.3.21)$$

και η φ είναι τετραγωνικά ολοκληρώσιμη, προσδιόρισαν την ασυμπτωτική κατανομή της σ.σ. T_n . Κατά αυτόν τον τρόπο, κατάφεραν να καθορίσουν τις βέλτιστες επιλογές για τα σκορ, που οδηγούν στη βέλτιστη ισχύ του ελέγχου για ειδικές περιπτώσεις εναλλακτικών. Ειδικότερα, απέδειξαν ότι τα κανονικά σκορ, τα Wilcoxon σκορ και τα κλασικά σκορ του προσημικού είναι βέλτιστα για την κανονική, τη λογιστική και τη διπλά εκθετική κατανομή αντίστοιχα.

Τέλος, αποδεικνύεται (βλέπε Sen (2006)) ότι η τιμή της σ.σ. T_n είναι αναλλοίωτη υπό τους μετασχηματισμούς των δεδομένων $X_i \rightarrow aX_i$ και $X_i \rightarrow g(X_i)$, όπου $g(\cdot)$ μια συνάρτηση μονότονη και τέτοια ώστε $g(-x) = -g(x)$, για κάθε x και όπου a οποιαδήποτε θετική σταθερά.

2.2 Τεστ που στηρίζονται στο τεστ των ροών

Στην ενότητα αυτή, θα παρουσιαστούν τρόποι ελέγχου της μονοδιάστατης συμμετρίας μιας συνεχούς κατανομής γύρω από γνωστό κέντρο συμμετρίας που στηρίζονται στο τεστ των ροών. Στο πλαίσιο αυτό, αρχικά θα παρουσιαστεί ο τρόπος ελέγχου που προτάθηκε από τον McWilliams (1990), ενώ έπειτα θα παρουσιασθούν οι κυριότερες τροποποιήσεις αυτού του ελέγχου που έχουν εμφανιστεί στη βιβλιογραφία.

2.2.1 Το τεστ του McWilliams (1990)

Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από έναν πληθυσμό με συνεχή και διαφορίσιμη α.σ.κ., έστω F , και σ.π.π. f . Για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από το μηδέν, έναντι της εναλλακτικής υπόθεσης της γενικής ασυμμετρίας, ο McWilliams (1990) πρότεινε τα ακόλουθα. Έστω $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, το διατεταγμένο τ.δ. που προκύπτει διατάσσοντας τις παρατηρήσεις $X_1, X_2, \dots, X_n$, κατά αύξουσα τάξη μεγέθους της απόλυτης τιμής, διατηρώντας τα πρόσημα, ενώ $S_1, S_2, \dots, S_n$, είναι οι δείκτριες μεταβλητές που ορίστηκαν στη σχέση (2.1.2.1) και οι οποίες προσδιορίζουν το πρόσημο των $X_{(i)}$. Στο

παραπάνω πλαίσιο, ο τρόπος ελέγχου που προτάθηκε από τον McWilliams (1990) βασίζεται στον αριθμό των ακολουθιών ομοίων συμβόλων (ροών), έστω R^* , της ακολουθίας $\{S_i\}$. Εύκολα προκύπτει ότι:

$$R^* = 1 + \sum_{k=2}^n I_k, \quad (2.2.1.1)$$

όπου για $k = 2, \dots, n$,

$$I_k = \begin{cases} 0, & \text{αν } S_k = S_{k-1} \\ 1, & \text{αν } S_k \neq S_{k-1}. \end{cases} \quad (2.2.1.2)$$

Η αποσαφήνιση των παραπάνω συμβολισμών, καθώς επίσης και του σκεπτικού για την υιοθέτηση της παραπάνω σ.σ. για τον έλεγχο της υπόθεσης της συμμετρίας, γίνεται στη συνέχεια μέσω ενός αριθμητικού παραδείγματος.

Παράδειγμα 2.2.1.1 (McWilliams (1990))

Έστω οι ακόλουθες $n=15$ το πλήθος δειγματικές παρατηρήσεις:

−1.8, −4.6, 6.4, 12.7, −1.1, −2.18, 1.7, 21.6, −2.16, 6.3, −4.0, 11.8, −0.1, 14.6, −4.63

Αρχικά, διατάσσονται αυτές οι παρατηρήσεις κατά αύξουσα τάξη μεγέθους ως προς την απόλυτη τιμή, διατηρώντας τα πρόσημα τους. Τότε προκύπτει εύκολα ότι:

$$X_{(i)} : -0.1, -1.1, 1.7, -1.8, -2.16, -2.18, -4.0, -4.6, \\ -4.63, 6.3, 6.4, 11.8, 12.7, 14.6, 21.6.$$

Για τον ευκολότερο προσδιορισμό της τιμής της σ.σ. οι παρατηρήσεις με θετικό πρόσημο αντικαθίστανται στη συνέχεια με το σύμβολο +, ενώ οι παρατηρήσεις με αρνητικό πρόσημο με το σύμβολο −. Επομένως, κάθε δειγματική παρατήρηση αντικαθίσταται από το σύμβολο + ή το σύμβολο −, τα οποία σύμβολα χαρακτηρίζουν κατά

αυτόν τον τρόπο σε ποια ομάδα από τις δύο ανήκει η δειγματική παρατήρηση. Έτσι, για το υπό θεώρηση παράδειγμα είναι: $-, -, +, -, -, -, -, -, +, +, +, +, +, +$.

Ο McWilliams (1990) όπως έχει ήδη αναφερθεί, πρότεινε να αποφασίζουμε για την αποδοχή ή όχι της υπόθεσης της συμμετρίας γύρω από το μηδέν, θεωρώντας τη σ.σ. R^* που ορίζεται ως ο αριθμός των ακολουθιών ομοίων συμβόλων στην ακολουθία των n το πλήθος συμβόλων $+$ και $-$. Στο συγκεκριμένο παράδειγμα προκύπτει ότι $R^* = 4$.

Αν ήταν $R^* = 2$, αυτό θα σήμαινε ότι είτε αρχικά όλες οι παρατηρήσεις θα ήταν $+$ και έπειτα $-$ είτε αντίστροφα. Επομένως, μέχρι κάποιο σημείο κάθε παρατήρηση της μιας ομάδας ακολουθείται από παρατήρηση που ανήκει στην ίδια ομάδα και από εκείνο το σημείο και μετά οι υπόλοιπες παρατηρήσεις ανήκουν στην άλλη ομάδα. Αυτό πρακτικά σημαίνει ότι η υπόθεση της συμμετρίας απορρίπτεται. Αν ήταν $R^* = 15$, που είναι ο μέγιστος αριθμός ροών, θα σήμαινε ότι μια δειγματική παρατήρηση της μιας ομάδας διαδέχεται παρατήρηση της άλλης. Με βάση τα όσα αναφέρθηκαν στην Πρόταση 1.1.3, ο αναμενόμενος αριθμός των ροών της ακολουθίας είναι $1 + \frac{2n_1(n - n_1)}{n}$ και υποθέτοντας ότι το n είναι σταθερό, εύκολα προκύπτει ότι η παραπάνω συνάρτηση είναι κοίλη και μεγιστοποιείται ως προς n_1 για $n_1 = n_2 = n/2$. Άρα, ο μέγιστος αριθμός των ροών αναμένεται όταν οι δύο ομάδες έχουν τον ίδιο αριθμό παρατηρήσεων, δηλαδή όταν οι παρατηρήσεις έχουν χωριστεί ομοιόμορφα στις δύο ομάδες. Αυτό συμβαίνει όταν δεν έχουμε αποκλίσεις από την υπόθεση της συμμετρίας.

Μέσω του προηγούμενου παραδείγματος έγινε αντιληπτό ότι η ποσότητα R^* μπορεί να χρησιμοποιηθεί για το ζητούμενο έλεγχο. Επιπλέον, είναι προφανές ότι πολύ μικρές τιμές της σ.σ. R^*

υποδεικνύουν αποκλίσεις από τη συμμετρία γύρω από το μηδέν. Για το λόγο αυτό απαιτείται η εύρεση της κατανομής της σ.σ. R^* , υπό τη μηδενική υπόθεση.

Στο πλαίσιο αυτό, έστω $D_1, D_2, \dots, D_n$ οι δείκτριες μεταβλητές οι οποίες προσδιορίζουν το πρόσημο των X_i . Δηλαδή:

$$D_i = \begin{cases} 1, & \text{αν } X_i \geq 0 \\ 0, & \text{διαφορετικά.} \end{cases} \quad (2.2.1.3)$$

Είναι προφανές ότι υπό την υπόθεση της συμμετρίας οι $D_1, D_2, \dots, D_n$ είναι ανεξάρτητες και ισόνομες τ.μ. που ακολουθούν κατανομή Bernoulli με πιθανότητα επιτυχίας $p = 0.5$. Επιπλέον, οι $D_1, D_2, \dots, D_n$ είναι ανεξάρτητες των $|X_1|, |X_2|, \dots, |X_n|$. Επομένως, προκύπτει ότι η ακολουθία των τ.μ. $S_1, S_2, \dots, S_n$, όπως αυτή ορίστηκε στη σχέση (2.1.2.1), δεν είναι τίποτε άλλο παρά μια αναδιάταξη, ως προς τις απόλυτες τιμές, της ακολουθίας των τ.μ. $D_1, D_2, \dots, D_n$. Χρησιμοποιώντας το παραπάνω σκεπτικό ο McWilliams (1990) κατέληξε στο ακόλουθο συμπέρασμα.

Λήμμα 2.2.1.1 (McWilliams 1990) Υπό τη μηδενική υπόθεση της συμμετρίας, η σ.σ. $R^* - 1$ ακολουθεί διωνυμική κατανομή με παραμέτρους $n - 1$ και $1/2$.

Παρατήρηση 2.2.1.1 Ο παραπάνω έλεγχος έχει καθιερωθεί στη βιβλιογραφία ως ο έλεγχος που προτάθηκε από τον McWilliams (1990). Θα ήταν παράλειψη να μην αναφερθεί ότι ο τρόπος αυτός προτάθηκε ανεξάρτητα και ταυτόχρονα και από τους Cohen and Menjoge (1988), ενώ ο Henze (1993) μελέτησε, μεταξύ άλλων, τη συνέπεια του ελέγχου.

Επομένως, με βάση το Λήμμα 2.2.1.1 και τη συζήτηση που προηγήθηκε, για επίπεδο σημαντικότητας α , απορρίπτεται η H_0 αν $R^* \leq k_\alpha$, όπου η τιμή k_α είναι τέτοια ώστε:

$$\sum_{r=2}^{k_\alpha} P(R^* = r) = \alpha, \quad (2.2.1.4)$$

και μπορεί να υπολογιστεί λαμβάνοντας υπόψη το Λήμμα 2.2.1.1.

Όμως ο υπολογισμός της τιμής k_α τις περισσότερες φορές είναι ιδιαίτερα δύσκολος και επιπλέον όπως συμβαίνει σε όλα τα στατιστικά τεστ που στηρίζονται σε διακριτές τ.μ. υπάρχει πεπερασμένο πλήθος επιλογών για το επίπεδο σημαντικότητας α . Για τους παραπάνω λόγους, χρησιμοποιείται η προσέγγιση της διωνυμικής κατανομής από την κανονική μέσω του Κεντρικού Οριακού Θεωρήματος. Ειδικότερα, εύκολα προκύπτει ότι:

$$Z_{McW} = \frac{(R^* - 1) - \frac{(n-1)}{2}}{\sqrt{\frac{(n-1)}{4}}} \stackrel{H_0}{\underset{\text{ασυμπτ.}}{\sim}} N(0,1). \quad (2.2.1.5)$$

Επομένως, για επίπεδο σημαντικότητας α , απορρίπτεται η μηδενική υπόθεση έναντι της εναλλακτικής ότι η F είναι ασύμμετρη, αν $Z_{McW} < -z_\alpha$ ή ισοδύναμα ύστερα από λίγη άλγεβρα αν $R^* \leq c_\alpha$, όπου

$$c_\alpha = 1 + \frac{n-1}{2} - \frac{\Phi^{-1}(1-\alpha)}{2} \sqrt{n-1}, \quad \text{με } \Phi^{-1} \text{ να είναι η αντίστροφη}$$

συνάρτηση της α.σ.κ. της τυπικής κανονικής κατανομής $N(0,1)$.

Παρατήρηση 2.2.1.2 Για μικρές τιμές του n συνιστάται η εφαρμογή της διόρθωσης συνεχείας:

$$Z'_{McW} = \begin{cases} \frac{(R^*-1) + 0.5 - \frac{(n-1)}{2}}{\sqrt{\frac{(n-1)}{4}}}, & \text{αν } R^*-1 > \frac{(n-1)}{2} \\ \frac{(R^*-1) - 0.5 - \frac{(n-1)}{2}}{\sqrt{\frac{(n-1)}{4}}}, & \text{αν } R^*-1 < \frac{(n-1)}{2}. \end{cases} \quad (2.2.1.6)$$

Η ενότητα αυτή ολοκληρώνεται παραθέτοντας κάποιες βασικές ιδιότητες της σ.σ. R^* .

Ιδιότητα 1 (McWilliams 1990) Η σ.σ. R^* είναι αναλλοίωτη υπό οποιοδήποτε μετασχηματισμό $h(X)$, όπου η h είναι περιττή συνάρτηση, δηλαδή τέτοια ώστε $h(-x) = -h(x)$, για κάθε x και επιπλέον είναι αυστηρά μονότονη συνάρτηση επί των θετικών πραγματικών αριθμών.

Παρατήρηση 2.2.1.3 Επομένως, σύμφωνα με την Ιδιότητα 1 η τιμή της σ.σ. παραμένει αναλλοίωτη για μια οικογένεια μετασχηματισμών πάνω στις αρχικές παρατηρήσεις. Μέλη αυτής της οικογένειας μετασχηματισμών είναι ο αντίστροφος μετασχηματισμός $h(X) = 1/X$ και ο μετασχηματισμός $h(X) = -X$. Η ιδιότητα αυτή είναι επιθυμητή καθώς αν μια συνεχής τ.μ. X είναι συμμετρική γύρω από το μηδέν τότε και οι τ.μ. $1/X$ και $-X$ είναι συμμετρικές γύρω από το μηδέν.

Καθώς η κατανομή της σ.σ. R^*-1 υπό την H_0 είναι διωνυμική, ανεξάρτητα από την αληθινή κατανομή $F(\cdot)$ του πληθυσμού, προκύπτει η ακόλουθη ιδιότητα.

Ιδιότητα 2 (McWilliams 1990) Το τεστ του McWilliams (1990) είναι ανεξάρτητο κατανομής.

Ιδιότητα 3 (Cohen and Menjoge (1988), Henze (1993)) Υπό την υπόθεση ότι η σ.π.π. $f(\cdot)$ είναι συνεχής σχεδόν παντού, αποδεικνύεται ότι για $n \rightarrow \infty$ η σ.σ. $\frac{1}{n}R^*$ συγκλίνει κατά πιθανότητα και σχεδόν

παντού στο $D(f)$, όπου $D(f) = \int_0^\infty 2 \left(\frac{1}{f(x)} + \frac{1}{f(-x)} \right)^{-1} dx$.

Ιδιότητα 4 (Cohen and Menjoge (1988), Henze (1993)) Ισχύει ότι $D(f) \leq 1/2$ με την ισότητα να ισχύει στην περίπτωση όπου η f είναι συμμετρική γύρω από το μηδέν.

Οι Ιδιότητες 3 και 4 θεμελιώνουν από θεωρητική σκοπιά ότι η μηδενική υπόθεση της συμμετρίας απορρίπτεται για μικρές τιμές του R^* . Τέλος, ισχύει ότι:

Ιδιότητα 5 (βλέπε Henze (1993)) Το τεστ απορρίπτει τη μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν για μικρές τιμές της σ.σ. R^* και ο έλεγχος είναι συνεπής έναντι γενικών εναλλακτικών.

2.2.2 Ο έλεγχος του Baklizi (2003)

Ο έλεγχος που προτάθηκε από τον Baklizi (2003) αποτελεί ουσιαστικά μια τροποποίηση του ελέγχου που προτάθηκε ανεξάρτητα από τον McWilliams (1990) και τους Cohen and Menjoge (1988). Προέκυψε στην προσπάθεια του Baklizi να δημιουργήσει ένα στατιστικό τεστ το οποίο θα είναι ανθεκτικότερο από το τεστ του McWilliams (1990) και ταυτόχρονα θα έχει μεγαλύτερη απόδοση ισχύος. Στο ίδιο πλαίσιο με τον McWilliams (1990), το τεστ που προτάθηκε από τον Baklizi (2003) βασίζεται στη δεσμευμένη

κατανομή της σ.σ. R^* , δοθέντος των n_1 και n_2 , που ορίζονται να είναι το πλήθος των μηδενικών και των μονάδων, αντίστοιχα, στην ακολουθία των $\{S_i\}$, όπου τα S_i δίνονται στη σχέση (2.1.2.1).

Για την εύρεση της κατανομής του R^* υπό τη μηδενική υπόθεση είναι χρήσιμο το ακόλουθο αποτέλεσμα. Ο αριθμός των διαφορετικών τρόπων τοποθέτησης n το πλήθος όμοιων στοιχείων σε r το πλήθος μη κενά κελιά είναι ίσος με $\binom{n-1}{r-1}$. Μία απόδειξη αυτού (βλέπε Gibbons and Chakraborti (2003)) έχει ως εξής: αρχικά τοποθετούνται τα n το πλήθος στοιχεία, έστω π.χ. άσπρες μπάλες σε μία σειρά. Έπειτα, τα r το πλήθος κελιά μπορούν να δημιουργηθούν παρεμβάλλοντας ανάμεσα σε οποιεσδήποτε δύο άσπρες μπάλες $r-1$ το πλήθος μαύρες μπάλες. Καθώς υπάρχουν $n-1$ το πλήθος δυνατές θέσεις τέτοιας τοποθέτησης ο συνολικός αριθμός αυτών των τρόπων είναι $\binom{n-1}{r-1}$.

Μετά και την παράθεση του προηγούμενου αποτελέσματος στην επόμενη πρόταση προσδιορίζεται η δεσμευμένη κατανομή του R^* .

Πρόταση 2.2.2.1 (βλέπε μεταξύ άλλων Gibbons και Chakraborti (2003))

Αν R^* είναι ο αριθμός των ακολουθιών ομοίων συμβόλων στην ακολουθία n συμβόλων εκ των οποίων n_1 το πλήθος είναι $+$ και $n_2 = n - n_1$ το πλήθος είναι $-$, τότε υπό την υπόθεση της συμμετρίας, για $r = 2, 3, \dots, n$, ισχύει ότι:

$$f_{R^*}(r|n_1, n_2) = \begin{cases} \frac{2 \binom{n_1-1}{r/2-1} \binom{n_2-1}{r/2-1}}{\binom{n}{n_1}}, & r \text{ άρτιος} \\ \frac{\binom{n_1-1}{(r-1)/2} \binom{n_2-1}{(r-3)/2} + \binom{n_1-1}{(r-3)/2} \binom{n_2-1}{(r-1)/2}}{\binom{n}{n_1}}, & r \text{ περιττός} \end{cases}$$

ενώ στην ειδική περίπτωση όπου είτε $n_1 = 0$ είτε $n_2 = 0$ ισχύει ότι $P(R^* = 1) = 1$.

Απόδειξη

Ο δυνατός αριθμός των διατάξεων των n_1 το πλήθος συμβόλων + είναι: $\binom{n}{n_1}$. Αν τώρα το πλήθος των ακολουθιών ομοίων συμβόλων

είναι ίσο με r , όπου r ζυγός, δηλαδή $r = 2, 4, \dots$, τότε υπό την υπόθεση της συμμετρίας, θα έχουμε $r/2$ το πλήθος ακολουθίες συμβόλων + και $r/2$ το πλήθος ακολουθίες συμβόλων -. Αφού έχουμε $r/2$ το πλήθος ακολουθίες συμβόλων + αυτό σημαίνει ότι τα n_1 το πλήθος θετικά σύμβολα χωρίζονται σε $r/2$ ομάδες. Αυτό σύμφωνα με όσα προηγήθηκαν επιτυγχάνεται με $\binom{n_1-1}{r/2-1}$ τρόπους. Όμοια, έχουμε

$\binom{n_2-1}{r/2-1}$ τρόπους κατασκευής των $r/2$ το πλήθος ρών συμβόλων -.

Οπότε από τον πολλαπλασιαστικό κανόνα και καθώς μπορεί να έχουμε είτε πρώτα μια ακολουθία συμβόλων + και μετά μια ακολουθία συμβόλων - είτε αντίστροφα προκύπτει άμεσα ότι:

$$f_{R^*}(r|n_1, n_2) = \frac{2 \binom{n_1-1}{r/2-1} \binom{n_2-1}{r/2-1}}{\binom{n}{n_1}}, \text{ όταν } r \text{ άρτιος.}$$

Στην περίπτωση που έχουμε r το πλήθος ακολουθίες όμοιων συμβόλων με r περιττό αριθμό, υπό τη μηδενική υπόθεση, θα έχουμε είτε $(r-1)/2$ το πλήθος ακολουθίες συμβόλων $+$ και $(r-1)/2+1$ το πλήθος ακολουθίες συμβόλων $-$, είτε $(r-1)/2+1$ το πλήθος ακολουθίες συμβόλων $+$ και $(r-1)/2$ το πλήθος ακολουθίες συμβόλων $-$. Με παρόμοιο τρόπο όπως πριν προκύπτει ότι:

$$f_{R^*}(r|n_1, n_2) = \frac{\binom{n_1-1}{(r-1)/2} \binom{n_2-1}{(r-3)/2} + \binom{n_1-1}{(r-3)/2} \binom{n_2-1}{(r-1)/2}}{\binom{n}{n_1}}, r \text{ περιττός.}$$

Από τη συζήτηση που προηγήθηκε στην προηγούμενη ενότητα, προκύπτει ότι η μηδενική υπόθεση απορρίπτεται όταν $R^* \leq r_\alpha$, όπου r_α

είναι ένας αριθμός τέτοιος ώστε: $\sum_{r=2}^{r_\alpha} P(R^* = r|n_1, n_2) = \alpha$, με τη

δεσμευμένη κατανομή της σ.σ. R^* να προσδιορίζεται στην Πρόταση 2.2.2.1. Οι κρίσιμες τιμές του τεστ για μικρές θετικές τιμές των n_1 και n_2 δίνονται από τους Gibbons and Chakraborti (2003), καθώς δεν είναι τίποτα άλλο παρά μια απλή τροποποίηση του τεστ των ροών. Για

παράδειγμα, στην ειδική περίπτωση που $n_1 = 5$ και $n_2 = 4$, έχοντας ως κρίσιμη περιοχή την $R^* \leq 2$ προκύπτει ότι $\alpha = 2/126$ κι αυτό διότι $f_{R^*}(R^* = 2 | n_1 = 5, n_2 = 4) = 2/126$, ενώ έχοντας ως κρίσιμη περιοχή την $R^* \leq 3$ προκύπτει ότι $\alpha = 7/126$, καθώς $f_{R^*}(R^* = 3 | n_1 = 5, n_2 = 4) = 7/126$ (βλέπε Gibbons and Chakraborti (2003)).

Είναι προφανές ότι ο υπολογισμός της τιμής r_α είναι ιδιαίτερα δύσκολος και οδηγεί σε πεπερασμένο πλήθος επιλογών για το επίπεδο σημαντικότητας α . Για τους παραπάνω λόγους οδηγούμαστε στην εύρεση ενός προσεγγιστικού στατιστικού τεστ. Καθώς η σ.σ. R^* μπορεί να γραφεί ως άθροισμα ανεξάρτητων τ.μ., αρκεί να υπολογιστούν οι ποσότητες $E(R^*)$ και $Var(R^*)$ για να χρησιμοποιηθεί το Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.). Οι ποσότητες αυτές έχουν προσδιοριστεί στην Πρόταση 1.1.2. Στο πλαίσιο αυτό, υποθέτοντας ότι το πλήθος των θετικών και αρνητικών προσήμων είναι τέτοια ώστε $n_1 / (n_1 + n_2) \rightarrow \gamma$, όπου γ ένας αριθμός με $0 < \gamma < 1$, αποδεικνύεται ότι (βλέπε Baklizi (2003)):

$$Z_B = \frac{R^* - 2(n_1 + n_2)\gamma(1-\gamma)}{2(n_1 + n_2)^{1/2}\gamma(1-\gamma)} \stackrel{H_0}{\underset{\text{ασυμπτ.}}{\sim}} N(0,1). \quad (2.2.2.1)$$

Επομένως, απορρίπτεται η υπόθεση της συμμετρίας έχοντας ως εναλλακτική την υπόθεση της ασυμμετρίας, αν $Z_B < -z_\alpha$.

Παρατήρηση 2.2.2.1 Η ανθεκτικότητα του τεστ του Baklizi (2003) έναντι του τεστ του McWilliams (1990) προκύπτει με βάση το ακόλουθο σκεπτικό (βλέπε Baklizi (2003) σελ. 714). Ας υποθέσουμε ότι η κατανομή είναι συμμετρική αλλά η προκαθορισμένη διάμεσος δεν είναι σωστή. Τότε υπάρχει μια ανισορροπία μεταξύ του αριθμού των 0 και 1 στην ακολουθία των $\{S_i\}$, $i=1,...,n$ και προκύπτει ότι ο πιθανός αριθμός των ροών θα μειωθεί με αποτέλεσμα να λαμβάνει μικρές τιμές η σ.σ. R^* . Το γεγονός αυτό οδηγεί σε αύξηση του αριθμού των φορών που απορρίπτεται η υπόθεσης της συμμετρίας χρησιμοποιώντας το τεστ του McWilliams (1990), το οποίο στηρίζεται στη μη δεσμευμένη κατανομή του R^* , ενώ η μηδενική υπόθεση της συμμετρίας είναι αληθής, παρότι το κέντρο συμμετρίας έχει καθοριστεί εσφαλμένα. Αντίθετα, χρησιμοποιώντας το τεστ του Baklizi (2003), η δυσαναλογία μεταξύ 0 και 1 υπολογίζεται και τροποποιούνται κατάλληλα οι κρίσιμες τιμές λαμβάνοντας υπόψη τις τιμές των n_1 και n_2 που παριστάνουν το πλήθος των 0 και 1 στην ακολουθία $\{S_i\}$. Επομένως, κατά αυτόν τον τρόπο ο αριθμός των απορρίψεων της υπόθεσης της συμμετρίας αναμένεται να είναι πλησιέστερα στο αληθινό επίπεδο σημαντικότητας. Η ιδιότητα αυτή έχει επαληθευτεί με μια εκτενή μελέτη προσομοίωσης από τον Baklizi (2003).

2.2.3 Ο έλεγχος των Modarres and Gastwirth (1996)

Στο εδάφιο αυτό θα παρουσιαστεί άλλη μια τροποποίηση του τεστ του McWilliams (1990) η οποία εισήχθη στη βιβλιογραφία από τους Modarres and Gastwirth (1996). Όπως έχει ήδη αναφερθεί, ο McWilliams (1990) χρησιμοποίησε για τον έλεγχο της συμμετρίας γύρω από το μηδέν ως σ.σ. τον αριθμό των ροών R^* στην ακολουθία $\{S_i\}$, που προσδιορίζεται από τη σχέση (2.1.2.1).

Οι Modarres and Gastwirth (1996) αρχικά επισημαίνουν ότι υπό τη μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν, αν I_k είναι οι δείκτριες τ.μ. που ορίστηκαν στη σχέση (2.2.1.2), τότε οι πιθανότητες $P(I_k = x)$, $x = 0, 1$, είναι σταθερές, ενώ υπό την εναλλακτική υπόθεση της ασυμμετρίας οι πιθανότητες αυτές δεν είναι ανεξάρτητες του k , $k = 2, 3, \dots, n$. Με βάση αυτήν την παρατήρηση οι Modarres and Gastwirth (1996) πρότειναν ένα στατιστικό τεστ αξιοποιώντας την τιμή του k για τις ροές, έχοντας ως στόχο να προκύψει ένας έλεγχος ισχυρότερος από αυτόν του McWilliams (1990). Παρατηρώντας ότι για λοξές εναλλακτικές οι ροές θα πρέπει να βρίσκονται στις ουρές της κατανομής και παρακινούμενοι από την ιδέα του Gastwirth (1965) της χρήσης του τεστ των ροών με βάρη, τροποποιούν το τεστ του McWilliams (1990) εισάγοντας συντελεστές βαρύτητας στις ροές.

Ειδικότερα, οι Modarres and Gastwirth (1996) προτείνουν την οικογένεια των σ.σ.:

$$M_p = \sum_{k=[np]+2}^n \varphi(k) I_k, \quad (2.2.3.1)$$

όπου

$$\varphi(k) = \begin{cases} k - np, & \text{αν } k > np \\ 0, & \text{διαφορετικά,} \end{cases} \quad (2.2.3.2)$$

με p να είναι το περικομμένο ποσοστό (trimming proportion) και με $[x]$ να συμβολίζεται το ακέραιο μέρος του x . Για $p = 0$ οι συντελεστές $\varphi(k)$ είναι τα λεγόμενα Wilcoxon scores, ενώ για $p = 0$ και $\varphi(k) = 1$ προκύπτει η σ.σ. $R^* - 1$ που προτάθηκε ανεξάρτητα από τους McWilliams (1990) και Cohen and Menjoge (1988).

Η ακριβής κατανομή της σ.σ. M_p υπό την μηδενική υπόθεση είναι εύκολα υπολογίσιμη για μικρά μεγέθη δείγματος, καθώς η σ.σ.

M_p προσδιορίζεται πλήρως από τις τιμές των I_k . Ειδικότερα, ο δειγματικός χώρος αποτελείται από 2^{n-1} το πλήθος στοιχεία που αντιστοιχούν στις διαφορετικές τιμές $(I_2, I_3, \dots, I_n)$. Καθώς καθένα από τα στοιχεία του δειγματικού χώρου είναι ισοπίθανα, η κατανομή του M_p υπό την H_0 προκύπτει εύκολα και προσδιορίζεται στην ακόλουθη πρόταση.

Πρόταση 2.2.3.1 (Modarres and Gastwirth (1996)) Υπό την υπόθεση της συμμετρίας ισχύει ότι:

$$P(M_p = m) = \frac{\eta(m)}{2^{n-1}}, \quad (2.2.3.3)$$

όπου $\eta(m)$ είναι ο αριθμός των τρόπων που αναθέτοντας μηδέν ή ένα στις τ.μ. I_k , $k = [np] + 2, \dots, n$, προκύπτει ότι η τιμή του αθροίσματος

$$\sum_{k=[np]+2}^n \varphi(k) I_k, \text{ είναι τέτοια ώστε } M_p = m.$$

Από την παραπάνω πρόταση είναι προφανές ότι η ακριβής κατανομή της σ.σ. M_p είναι δύσκολα εφαρμόσιμη. Για το λόγο αυτό οι Modarres and Gastwirth (1996) προσδιόρισαν την ασυμπτωτική της κατανομή, υπό τη μηδενική υπόθεση, η οποία παρατίθεται στο λήμμα που ακολουθεί.

Λήμμα 2.2.3.1 (Modarres and Gastwirth (1996)) Υπό τη μηδενική υπόθεση της συμμετρίας η σ.σ. M_p ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή και διακύμανση:

$$\mu(M_p) = \frac{1}{4}(n(1-p)-1)(n(1-p)+2), \quad (2.2.3.4)$$

και

$$\sigma^2(M_p) = \frac{1}{24}(n(1-p)-1)(2n^2(1-p)^2 + 5n(1-p) + 6), \quad (2.2.3.5)$$

αντίστοιχα.

Σκιαγράφηση Απόδειξης

Υπό τη μηδενική υπόθεση της συμμετρίας οι τ.μ. I_k , $k = 2, \dots, n$, είναι ανεξάρτητες και ισόνομες Bernoulli με παράμετρο $p = 1/2$. Η απόδειξη ολοκληρώνεται λαμβάνοντας υπόψη ότι η σ.σ. M_p γράφεται ως άθροισμα τ.μ. (βλέπε σχέση 2.2.3.1) και η ασυμπτωτική κανονικότητα προκύπτει καθώς πληρούνται οι προϋποθέσεις του Κεντρικού Οριακού Θεωρήματος.

Έτσι, με βάση το Λήμμα 2.2.3.1 προκύπτει ότι:

$$Z_{MG} = \frac{M_p - \mu(M_p)}{\sqrt{\sigma^2(M_p)}} \stackrel{H_0}{\underset{\text{ασυμπ.}}{\sim}} N(0,1). \quad (2.2.3.6)$$

Επομένως, έχοντας ως εναλλακτική υπόθεση ότι η F είναι ασύμμετρη, η H_0 απορρίπτεται αν $Z_{MG} < -z_\alpha$.

Παρατήρηση 2.2.3.1 Σχετικά με την επιλογή του ποσοστού περικοπής p , με βάση τα αποτελέσματα μιας εκτενούς μελέτης προσομοίωσης, οι Modarres and Gastwirth (1996) προτείνουν ανεξάρτητα από το μέγεθος του δείγματος n την τιμή $p = 0.2$. Στην ειδική περίπτωση όπου έχουμε δείγματα με μέγεθος μεγαλύτερο ή ίσο του 50 και λοξή εναλλακτική η επιλογή $p = 0.25$ είναι προτιμότερη. Στο σημείο αυτό πρέπει να αναφερθεί ότι τα παραπάνω αποτελέσματα προέκυψαν συγκρίνοντας τις τιμές $p = 0, 0.1, 0.2, 0.25$.

2.2.4 Το τεστ των Corzo and Babativa (2013)

Στην παράγραφο αυτή θα παρουσιαστεί μια ακόμα τροποποίηση του τεστ του McWilliams (1990), η οποία εισήχθη στη

βιβλιογραφία πρόσφατα από τους Corzo and Babativa (2013). Η σ.σ. των Corzo and Babativa (2013), όπως θα δούμε στη συνέχεια έχει τη μορφή της σ.σ. M_p των Modarres and Gastwirth (1996), που δόθηκε στη σχέση (2.2.3.1).

Ειδικότερα, η σ.σ. που προτάθηκε από τους Corzo and Babativa (2013) συμβολίζεται με J_k και ορίζεται να είναι ο αριθμός των ροών στους τελευταίους k όρους της ακολουθίας $I_2, \dots, I_n$, έχοντας ως συντελεστή βαρύτητας τη θέση του εκάστοτε όρου. Συνεπώς,

$$J_k = 1 + \sum_{i=1}^k (n-k+i) I_{n-k+i}, \text{ για } k = 2, \dots, n-1. \quad (2.2.4.1)$$

Παρατήρηση 2.2.4.1 Όταν $k = n-1$, τότε προκύπτει ότι:

$$J_{n-1} - 1 = \sum_{i=1}^{n-1} (i+1) I_{i+1} = \sum_{i=2}^n i I_i = M_0,$$

όπου M_0 είναι η σ.σ. των Modarres and Gastwirth (1996) που προκύπτει από τη σχέση (2.2.3.1) για $p = 0$.

Είναι προφανές ότι η ιδέα των Corzo and Babativa (2013) προέκυψε θέλοντας να αξιοποιηθεί το γεγονός ότι για λοξές εναλλακτικές οι ροές θα πρέπει να βρίσκονται στις ουρές της κατανομής και για αυτόν τον λόγο επιλέγονται οι k τελευταίοι όροι της ακολουθίας $I_2, \dots, I_n$.

Η μορφή της κρίσιμης περιοχής του ελέγχου συμμετρίας που στηρίζεται στη σ.σ. J_k προκύπτει με ανάλογο σκεπτικό με αυτό που αναπτύχθηκε για τη σ.σ. του McWilliams (1990). Έτσι, η μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται για μικρές τιμές της σ.σ. J_k . Το επόμενο βήμα για τον καθορισμό της κρίσιμης περιοχής είναι η εύρεση της κατανομής της σ.σ. J_k υπό τη μηδενική

υπόθεση. Καθώς η J_k είναι το άθροισμα των ανεξάρτητων τ.μ. I_j , αρκεί να υπολογιστεί η αναμενόμενη τιμή και η διακύμανση της, υπό τη μηδενική υπόθεση.

Θεώρημα 2.2.4.1 (Corzo and Babativa (2013)) Υπό τη μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν, ισχύει ότι:

$$E(J_k - 1) = \frac{1}{2} \sum_{i=1}^k (n - k + i) = \frac{1}{4} k (2n - k + 1), \quad (2.2.4.2)$$

και

$$Var(J_k - 1) = \frac{1}{24} k (6n^2 + 6n + 2k^2 - 3k - 6nk + 1), \quad (2.2.4.3)$$

αντίστοιχα.

Επομένως, προκύπτει ότι η μέση τιμή και η διακύμανση εξαρτώνται τόσο από το μέγεθος του δείγματος όσο και από το k . Για να ξεπεραστεί αυτό το πρόβλημα οι Corzo and Babativa (2013) θεώρησαν έναν αριθμό, έστω γ , μεταξύ μηδέν και ένα, δηλαδή $0 < \gamma < 1$. Θέτοντας τότε $k_n = [\gamma n]$ και αντικαθιστώντας στον ορισμό του J_k το k με το k_n προκύπτει μια σ.σ. η οποία εξαρτάται μόνο από το n και είναι η:

$$J_{k_n} = 1 + \sum_{i=1}^{k_n} (n - k_n + i) I_{n - k_n + i}. \quad (2.2.4.4)$$

Στο επόμενο θεώρημα, λαμβάνοντας υπόψη το Θεώρημα 2.2.4.1, προσδιορίζεται η ασυμπτωτική κατανομή της σ.σ. J_{k_n} υπό την H_0 .

Θεώρημα 2.2.4.2 (Corzo and Babativa (2013)) Έστω k_n τέτοιο ώστε $k_n \rightarrow \infty$. Τότε,

$$Z_{COB} = \frac{(J_{k_n} - 1) - E(J_{k_n} - 1)}{\sqrt{Var(J_{k_n} - 1)}} \stackrel{H_0}{\underset{\text{ασυμπτ.}}{\sim}} N(0,1), \quad (2.2.4.5)$$

όπου οι ποσότητες $E(J_k - 1)$ και $Var(J_k - 1)$ προκύπτουν από τις σχέσεις (2.2.4.2) και (2.2.4.3), θέτοντας όπου k το k_n .

Επομένως, χρησιμοποιώντας τα παραπάνω αποτελέσματα προκύπτει ότι έχοντας ως εναλλακτική την υπόθεση της ασυμμετρίας, η H_0 απορρίπτεται αν $Z_{COB} < -z_\alpha$.

Παρατήρηση 2.2.4.2 Η επιλογή της τιμής $k=6$ στη σ.σ. J_k , προτείνεται με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης που πραγματοποίησαν οι Corzo and Babativa (2013) μεταξύ των τιμών $4 \leq k \leq 20$.

2.2.5 Το τεστ του Baklizi (2007)

Στην ενότητα αυτή θα παρουσιαστεί το τεστ που προτάθηκε από τον Baklizi (2007) για τον έλεγχο της συμμετρίας γύρω από το μηδέν. Σε προηγούμενη ενότητα ορίσαμε τα S_i να δίνονται από τη σχέση (2.1.2.1) και όπως είναι γνωστό, μεγάλος αριθμός ροών συνηγορεί υπέρ της υπόθεσης της συμμετρίας γύρω από το μηδέν. Επιπλέον είναι προφανές ότι το μήκος κάθε ροής, δηλαδή το πλήθος των όμοιων συμβόλων σε κάθε ροή, αναμένεται να είναι μικρό, καθώς το πλήθος των ροών αναμένεται μεγάλο.

Επομένως, η ύπαρξη ροών είτε μηδενικών είτε μονάδων μεγάλου μήκους μπορεί να ληφθεί ως υποψία ασυμμετρίας. Χρησιμοποιώντας αυτό το σκεπτικό, η σ.σ. που χρησιμοποιήθηκε από τον Baklizi (2007) και η οποία συμβολίζεται με $L_{n,p}^*$ ορίζεται να είναι το

μήκος της μακρύτερης ροής στις τελευταίες $n - [np]$ παρατηρήσεις της ακολουθίας $S_1, S_2, \dots, S_n$, όπου με $[\cdot]$ συμβολίζεται το ακέραιο μέρος και p είναι το ποσοστό περικοπής. Προφανώς, αν το ποσοστό περικοπής p ισούται με το μηδέν, τότε $L_{n,0}^*$ είναι η σ.σ. που παριστάνει το μεγαλύτερο μήκος ροής (longest run). Η ιδέα του Baklizi (2007) για τη χρησιμοποίηση του ποσοστού περικοπής, βασίστηκε στην παρατήρηση ότι μικρές τιμές του ποσοστού περικοπής p , δίνουν μεγαλύτερη ισχύ για τον συγκεκριμένο έλεγχο ειδικά αν ο λόγος της απόκλισης από τη συμμετρία είναι η ύπαρξη λοξότητας. Αυτό συμβαίνει επειδή στη συγκεκριμένη περίπτωση, μεγάλη ακολουθία ομοίων συμβόλων στις ροές είναι πολύ πιθανό να υπάρχει στους τελευταίους όρους της ακολουθίας $S_1, S_2, \dots, S_n$. Έτσι το $L_{n,p}^*$ αφαιρεί τη συμβολή των πρώτων $[np]$ το πλήθος παρατηρήσεων. Άρα, η συνολική συμβολή στο μέγεθος του τεστ προέρχεται από τους τελευταίους $n - [np]$ όρους της ακολουθίας $S_1, S_2, \dots, S_n$.

Παρατήρηση 2.2.5.1 Η ιδέα της χρήσης του ποσοστού περικοπής έχει χρησιμοποιηθεί από πολλούς συγγραφείς (βλέπε Gastwirth (1965), Modarres and Gastwirth (1998)). Σχετικά με την επιλογή του ποσοστού περικοπής p , οι Modarres and Gastwirth (1998) αναφέρουν ότι η επιλογή αυτή εξαρτάται από το μέγεθος του δείγματος και την προς έλεγχο εναλλακτική υπόθεση. Με βάση τα αποτελέσματα μιας εκτενούς μελέτης προσομοίωσης ο Baklizi (2007) προτείνει να χρησιμοποιείται προσεγγιστικά η τιμή $p = 0.8$. Εναλλακτικά, ο Baklizi (2007) για την αποφυγή της επιλογή ενός συγκεκριμένου ποσοστού περικοπής p , που επιλέγεται με εκτενή μελέτη προσομοίωσης, προτείνει να χρησιμοποιείται η σ.σ. $L = \sup_p (L_{n,p}^* - \mu_p) / \sigma_p$, όπου μ_p και σ_p^2 είναι η μέση τιμή και η διακύμανση αντίστοιχα της L_{np}^* . Όμως, από τα αποτελέσματα μιας μελέτης προσομοίωσης που διεξήγαγε ο

Baklizi (2007) εξάγεται το συμπέρασμα ότι η χρησιμοποίηση της σ.σ. L δεν αυξάνει σημαντικά την ισχύ στο περιβάλλον της προσομοίωσης που χρησιμοποιήθηκε και για αυτό δεν την προτείνει.

Από τη συζήτηση που έχει προηγηθεί, γίνεται σαφές ότι απορρίπτεται η υπόθεση της συμμετρίας έναντι της ασυμμετρίας για μεγάλες τιμές της σ.σ. $L_{n,p}^*$. Ο προσδιορισμός της κρίσιμης περιοχής απαιτεί την εύρεση της κατανομής της σ.σ. $L_{n,p}^*$ υπό τη μηδενική υπόθεση.

Στη συνέχεια, χωρίς βλάβη της γενικότητας, θα προσδιοριστεί η κατανομή της τ.μ. L_n^* , που παριστάνει το μήκος της μακρύτερης ροής σε n το πλήθος δειγματικές παρατηρήσεις. Η ανάλυση που ακολουθεί προέρχεται από την εργασία του Baklizi (2007). Έστω ότι έχουμε μια ακολουθία από n δοκιμές Bernoulli και $N_{r,1}$ είναι η τ.μ. που παριστάνει τον αριθμό των δοκιμών Bernoulli μέχρι να εμφανιστεί η πρώτη ροή συμβόλων + μήκους ακριβώς r με $p_{k,r}^* = P(N_{r,1} = k)$, $k = r, r+1, \dots$. Ακόμη, έστω $N_{r,2}$ η τ.μ. που παριστάνει τον αριθμό των δοκιμών Bernoulli μέχρι να εμφανιστεί η πρώτη ροή συμβόλων - μήκους ακριβώς r . Από τους Blom *et al.* (1994) προκύπτει ότι η πιθανογεννήτρια συνάρτηση των $N_{r,i}$, $i = 1, 2$, δίνεται από τη σχέση:

$$G_{r,i}(t) = \frac{0.5^r t^r (1 - 0.5t)}{1 - t + 0.5^{r+1} t^{r+1}}, \quad i = 1, 2. \quad (2.2.5.1)$$

Έστω επιπλέον N_r η τ.μ. που παριστάνει τον αριθμό των δοκιμών Bernoulli μέχρι να εμφανιστεί η πρώτη ροή (οποιοδήποτε τύπου) μήκους ακριβώς r . Προκύπτει ότι η πιθανογεννήτρια συνάρτηση του N_r (βλέπε Feller (1968)) είναι:

$$G_r(t) = \frac{2(0.5^r t^r (1 - 0.5t)(1 - 0.5^r t^r))}{1 - t + 0.5^r t^{r+1} - 0.5^{2r} t^{2r}}. \quad (2.2.5.2)$$

Στο σημείο αυτό θα πρέπει να επισημανθεί ότι η γνώση της πιθανογεννήτριας μας δίνει τη δυνατότητα εύρεσης των πιθανοτήτων $p_{k,r} = P(N_r = k)$, καθώς οι πιθανότητες αυτές δεν είναι τίποτα άλλο παρά ο συντελεστής του όρου t^k . Η εύρεση αυτή είναι εφικτή αναπτύσσοντας την πιθανογεννήτρια $G_r(t)$ σε σειρά Taylor γύρω από το t .

Από τους Blom *et al.* (1973, 1994) προκύπτει ότι αν $B_{n,r}$ είναι το ενδεχόμενο να εμφανίζεται τουλάχιστον μια ροή μήκους r ή μεγαλύτερου σε n το πλήθος δοκιμές Bernoulli, τότε αυτό συμβαίνει αν και μόνο αν εμφανίζεται μια ροή μήκους r στη n -οστή ή σε προηγούμενη προσπάθεια. Προφανώς, η συντομότερη δυνατή προσπάθεια είναι η r . Έτσι,

$$P(B_{n,r}) = P(L_n^* \geq r) = P(N_r \leq n) = \sum_{k=r}^n P(N_r = k). \quad (2.2.5.3)$$

Για την σ.π. της τ.μ. L_n^* , δηλαδή του μήκους της μεγαλύτερης ροής οποιουδήποτε τύπου ισχύει ότι:

$$\begin{aligned} P(L_n^* = r) &= P(N_r \leq n) - P(N_{r+1} \leq n) \\ &= \sum_{k=r}^n P(N_r = k) - \sum_{k=r+1}^n P(N_{r+1} = k) \end{aligned} \quad (2.2.5.4)$$

με $r=1, 2, \dots, n$, ενώ η πιθανότητα $p_{k,r}$ είναι ο συντελεστής του όρου t^k στο ανάπτυγμα Taylor του $G_r(t)$.

Έχει ήδη αναφερθεί ότι η μηδενική υπόθεση της συμμετρίας απορρίπτεται για μεγάλες τιμές της σ.σ. L_n^* . Επομένως, το ενδιαφέρον μας επικεντρώνεται στην εύρεση των πιθανοτήτων της μορφής $q_{r,n} = P(N_r \leq n) = P(L_n^* \geq r)$, $r=1, \dots, n$. Έστω $Q_{r,n}(t) = q_{r,n}t^r + \dots + q_{n,n}t^n$ η πιθανογεννήτρια συνάρτηση των $q_{r,n}, \dots, q_{n,n}$. Τότε ο συντελεστής του

t^j , για $j = r, \dots, n$, στο ανάπτυγμα Taylor των $Q_{r,n}(t)$ και $G_r(t)/(1-t)$ είναι ο ίδιος. Αυτό ισχύει, καθώς ο συντελεστής του t^j στο ανάπτυγμα Taylor του $(1-t)Q_{r,n}(t)$ είναι $q_{r,j} - q_{r,j-1} = p_{j,r}$. Όμως, $p_{j,r}$ είναι ο συντελεστής του t^j στο ανάπτυγμα Taylor της πιθανογεννήτριας $G_r(t)$ και επομένως προκύπτει ότι οι συντελεστές αυτοί ταυτίζονται. Συνεπώς, οι πιθανότητες της μορφής $P(L_n^* \geq r)$, μπορούν να βρεθούν παραγωγίζοντας την πιθανογεννήτρια συνάρτηση $Q_{r,n}(t)$ ως προς t και θέτοντας $t=0$. Κάποιες πιθανότητες της μορφής $P(L_n^* \geq r)$ για ειδικές περιπτώσεις των n, r προσδιορίζονται στον Πίνακα 3 του Baklizi (2007). Ειδικότερα, παρατίθενται αποτελέσματα για $n=5, 6, \dots, 25$. Καθώς στην πράξη είναι πολύ πιθανό να μην περιέχεται ο συνδυασμός των (n, r) στον Πίνακα 3 του Baklizi (2007, σελ. 344) και κατά αυτόν τον τρόπο να μην μπορεί να υπολογιστεί η κρίσιμη περιοχή, προτάθηκε από τον Baklizi (2007) η ακόλουθη προσέγγιση για τον υπολογισμό των πιθανοτήτων $q_{r,n}$:

$$q_{r,n} \approx 1 - \frac{0.5^{r-1} (t^*)^r}{(t^*)^{n+1} \left(\left(1 - \sum_{k=1}^{r-1} 0.5^k (t^*)^k \right) + (1-t^*) \sum_{k=1}^{r-1} 0.5^k k (t^*)^{k-1} \right)}, \quad r=1, \dots, n,$$

όπου t^* είναι η μικρότερη, κατά απόλυτη τιμή, ρίζα του $t = 1 + \left(\frac{t}{2} \right)^r$. Η προσέγγιση αυτή δίνει σωστά αποτελέσματα με ακρίβεια τουλάχιστον τριών δεκαδικών ψηφίων ακόμα και για μικρά μεγέθη δείγματος, π.χ. για $n=7$ (βλέπε Baklizi (2007)).

Είναι φανερό ότι η παραπάνω προσέγγιση δεν είναι ιδιαίτερα εύχρηστη. Για το λόγο αυτό ο Baklizi (2007) πρότεινε μια εναλλακτική προσέγγιση, η οποία δίνεται στην πρόταση που ακολουθεί.

Πρόταση 2.2.5.1. (Baklizi (2007)) Ισχύει ότι υπό την υπόθεση της συμμετρίας γύρω από το μηδέν $P(L_n^* \geq r) \approx e^{-\frac{n-r}{2^r}}$.

Απόδειξη

Το αποτέλεσμα προκύπτει ως εφαρμογή του αποτελέσματος των Balakrishnan and Koutras (2002, σελ.197) σύμφωνα με το οποίο:

$$\left| P(L_n^* < r) - (1 - e^{-\lambda_0}) \right| \leq \frac{1}{2^{r-1}} + (1 - e^{-\lambda_0})(2r+1) \left(\frac{1}{2^{r-1}} \right) + \left(\frac{r}{2^{r-4}} \right)$$

όπου $\lambda_0 = (n-r)/2^r$. Επομένως, για μεγάλα n, r προκύπτει ότι

$$P(L_n^* \geq r) \approx e^{-\lambda_0} = e^{-\frac{n-r}{2^r}}.$$

Παρότι η πρώτη προσέγγιση είναι ποιο ακριβής, σε πρακτικές εφαρμογές συνίσταται η χρήση της προσέγγισης της Πρότασης 2.2.5.1. Σύμφωνα με αυτήν, η κρίσιμη περιοχή του ελέγχου με επίπεδο σημαντικότητας α είναι $L_n^* \geq r$, όπου το r είναι τέτοιο ώστε $e^{-\frac{n-r}{2^r}} = \alpha$.

Παρατήρηση 2.2.5.2 (Baklizi (2007)) Μια πιθανή τροποποίηση του τεστ του Baklizi (2007) προκύπτει συνδυάζοντας τον έλεγχο αυτό με το τεστ των Modarres and Gastwirth (1996). Ειδικότερα, η μηδενική υπόθεση της συμμετρίας απορρίπτεται αν είτε αυτή απορρίπτεται με τη σ.σ. M_p για επίπεδο σημαντικότητας α_1 είτε από τη σ.σ. $L_{n,p}^*$ για επίπεδο σημαντικότητας α_2 . Το συνολικό επίπεδο σημαντικότητας α προκύπτει μέσω της σχέσης:

$$\begin{aligned} \alpha &= P(L_{n,p}^* \text{ ή } M_p \text{ απορ. } H_0) = P(L_{n,p}^* \text{ απορ. } H_0) + P(M_p \text{ απορ. } H_0) \\ &\quad - P(L_{n,p}^* \text{ και } M_p \text{ απορ. } H_0) = \alpha_1 + \alpha_2 - \alpha_1 \alpha_2. \end{aligned}$$

Παρόμοια, το τεστ του McWilliams (1990) μπορεί να χρησιμοποιηθεί στη θέση της σ.σ. M_p των Modarres and Gastwirth (1996) ώστε να προκύψει μια ακόμη τροποποίηση του τεστ του Baklizi (2007).

Παρατήρηση 2.2.5.3 (Baklizi (2007)) Το τεστ του Baklizi (2007) μπορεί εύκολα να τροποποιηθεί ανάλογα με το αν η υπό έλεγχο εναλλακτική υπόθεση είναι αυτή της ύπαρξης θετικής ή αρνητικής λοξότητας. Για ελέγχους όπου έχουμε ως εναλλακτική υπόθεση την ύπαρξη θετικής λοξότητας, αν αυτή είναι αληθής αναμένονται μεγάλες ροές με 1 στους τελευταίους όρους της ακολουθίας $\{S_i\}$. Έτσι, μία αποτελεσματικά μεγάλη ροή 1 ($L_{n,p}^1$) είναι ένδειξη θετικής λοξότητας. Ανάλογα, όταν έχουμε ως εναλλακτική υπόθεση την ύπαρξη αρνητικής λοξότητας, αν αυτή είναι αληθής αναμένονται μεγάλες ροές με 0 στους τελευταίους όρους της ακολουθίας $\{S_i\}$. Έτσι, μία αποτελεσματικά μεγάλη ροή 0 ($L_{n,p}^0$) είναι ένδειξη υπέρ της υπόθεσης της αρνητικής λοξότητας. Οι κατανομές των σ.σ. $L_{n,p}^0$ και $L_{n,p}^1$ υπό την υπόθεση της συμμετρίας ταυτίζονται. Επιπλέον, υπό την υπόθεση της συμμετρίας οι κατανομές αυτές προκύπτουν από τις κατανομές των τ.μ. L_n^0 και L_n^1 , που παριστάνουν το μέγιστο μήκος ροής 0 ή 1 σε όλη την ακολουθία των $S_1, S_2, \dots, S_n$.

2.3 Τεστ με την Εμπειρική Αθροιστική Συνάρτηση Κατανομής (ε.α.σ.κ.)

Στις προηγούμενες ενότητες παρουσιάστηκαν έλεγχοι μονοδιάστατης συμμετρίας γύρω από γνωστό κέντρο συμμετρίας που βασίζονται στο προσημικό τεστ και στις ροές. Στην ενότητα αυτή θα παρουσιασθούν έλεγχοι που στηρίζονται στην εμπειρική α.σ.κ (ε.α.σ.κ.).

Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από έναν πληθυσμό με συνεχή α.σ.κ. F και ας είναι F_n η ε.α.σ.κ. Σε όσα ακολουθούν σε αυτήν την ενότητα θα συμβολίζεται με $\mathbb{F}$ η κλάση όλων των συνεχών συναρτήσεων κατανομής και με $\mathbb{F}_1$ η κλάση των συναρτήσεων κατανομής που ανήκουν στην $\mathbb{F}$ και οι οποίες είναι συμμετρικές γύρω από το μηδέν. Οι έλεγχοι που θα παρουσιαστούν σε αυτήν την ενότητα διατυπώθηκαν τόσο για τον έλεγχο της συμμετρίας έχοντας ως εναλλακτική υπόθεση την ύπαρξη γενικής ασυμμετρίας όσο και την ύπαρξη αρνητικής (θετικής) λοξότητας. Η κεντρική ιδέα των μεθοδολογιών αυτής της ενότητας στηρίζεται στην ιδιότητα ότι υπό την υπόθεση της συμμετρίας γύρω από το μηδέν ισχύει ότι $F(x) + F(-x) = 1$. Επομένως, είναι λογικό να αναζητηθεί ένας έλεγχος που θα στηρίζεται στην εκτίμηση της ποσότητας $F(x) + F(-x)$ από τις διαθέσιμες δειγματικές παρατηρήσεις και την απόκλιση αυτής της εκτιμώμενης ποσότητας από την τιμή 1. Διάφοροι τρόποι μέτρησης αποκλίσεων έχουν εισαχθεί στη βιβλιογραφία. Ένας τέτοιος τρόπος οδηγεί στους λεγόμενους Kolmogorov-Smirnov τύπους ελέγχου, όπου χρησιμοποιείται είτε το supremum είτε το infimum, ενώ όταν για παράδειγμα χρησιμοποιείται ένα κατάλληλα ορισμένο ολοκλήρωμα έχουμε τους λεγόμενους Cramer-Von Mises τύπους ελέγχου. Στο πλαίσιο αυτό, ο έλεγχος των Smirnov (1947), του Butler (1967), των Srinivasan and Godio (1974) και των Hill and Rao (1977) που θα παρουσιαστούν στη συνέχεια ανήκουν στην πρώτη κατηγορία, ενώ ο έλεγχος των Rothman and Woodroffe (1972) και των Baringhaus and Henze (1992) στη δεύτερη.

2.3.1 Το τεστ των Smirnov (1947) και Butler (1969)

Ο τρόπος ελέγχου που θα παρουσιαστεί σε αυτήν την παράγραφο πρωτοεμφανίστηκε στη βιβλιογραφία σε μία εργασία του

Smirnov (1947). Η πρόσβαση σε αυτήν ήταν περιορισμένη και αυτό οδήγησε σε μια ανεξάρτητη μελέτη που διεξήχθη από τον Butler (1969). Καθώς η εργασία του Butler (1969) ήταν πολύ συνοπτική, οι Chatterjee and Sen (1973) παραθέτουν μια πιο λεπτομερή ανάλυση και μελέτη, ενώ μια απλουστευμένη παρουσίαση δόθηκε από τον Niederhausen (1979). Αρχικά, θα παρουσιαστεί η κεντρική ιδέα αυτού του τρόπου ελέγχου. Έπειτα, θα προσδιοριστεί η μορφή της περιοχής απόρριψης για τις διάφορες περιπτώσεις εναλλακτικών υποθέσεων και θα καθοριστούν οι κρίσιμες τιμές με τη βοήθεια τόσο της ακριβής κατανομής της ελεγχουσυνάρτησης υπό τη μηδενική υπόθεση όσο και με τη χρήση της ασυμπτωτικής κατανομής της.

Όπως έχουμε αναφέρει η F είναι συμμετρική γύρω από το μηδέν αν $F(x) = 1 - F(-x)$, για κάθε $x \in \mathbb{R}$. Αν $X_1, X_2, \dots, X_n$ είναι οι διαθέσιμες δειγματικές παρατηρήσεις, τότε είναι λογικό κάποιος να χρησιμοποιήσει για τον έλεγχο της υπόθεσης της συμμετρίας έναντι των εναλλακτικών της ύπαρξης θετικής, αρνητικής λοξότητας ή ασυμμετρίας τις σ.σ.:

$$B_n^+ = \sup_{x \geq 0} \{1 - F_n(x) - F_n(-x-)\}, \quad (2.3.1.1)$$

$$B_n^- = \inf_{x \geq 0} \{1 - F_n(x) - F_n(-x-)\}, \quad (2.3.1.2)$$

και

$$B_n = \max \{B_n^+, -B_n^-\} = \sup_{x \geq 0} |F_n(x) + F_n(-x-) - 1|, \quad (2.3.1.3)$$

αντίστοιχα, όπου επειδή η ε.α.σ.κ. είναι βηματική συνάρτηση χρησιμοποιείται η $F_n(-x-)$ αντί της $F_n(-x)$. Προφανώς, μεγάλες τιμές των B_n^+ , B_n και B_n^- συνηγορούν στην απόρριψη της υπόθεσης της συμμετρίας γύρω από το μηδέν.

Ο προσδιορισμός της κατανομής των σ.σ. που ορίστηκαν στις σχέσεις (2.2.1.1.)-(2.2.1.3), υπό την H_0 δίνεται συνοπτικά από τον Butler (1969), ενώ αναλυτικά από τους Chatterjee and Sen (1973).

Πρόταση 2.3.1.1 (Butler (1969), Chatterjee and Sen (1973)) Υπό την υπόθεση της συμμετρίας γύρω από το μηδέν και για $k=1, \dots, n$, ισχύει ότι:

$$P(nB_n^+ \geq k) = P(nB_n^- \geq k) = \begin{cases} \sum_{j=k}^n (k/j) 2^{-j} \binom{j}{r}, & \text{αν } j-k=2r, r=0,1,2,\dots \\ 0 & , \text{αν } j-k=2r+1, j \geq k \geq 1, \end{cases}$$

ενώ για κάθε $k \geq 1$

$$P(nB_n \geq k) = 2 \sum_{j=0}^u (-1)^j P(nB_n^+ \geq (2j+1)k),$$

όπου $u = [n/(2k)] - 1$, $k \geq 1$, ενώ για $k=1$ είναι $P(nB_n \geq k) = 1$.

Από την προηγούμενη πρόταση εξάγεται το ακόλουθο πόρισμα.

Πόρισμα 2.3.1.1 (Butler (1969)) Υπό την υπόθεση της συμμετρίας γύρω από το μηδέν οι σ.σ. B_n^+ , B_n και B_n^- είναι ανεξάρτητες κατανομής.

Καθώς $B_{n,0.5}(x) = 2^{-n} \sum_{i=0}^{[x]} \binom{n}{i}$ είναι η α.σ.κ της διωνυμικής

κατανομής με πιθανότητα επιτυχίας $p = 0.5$, όπου $[\cdot]$ το ακέραιο μέρος, γίνεται αντιληπτό από τα αποτελέσματα της Πρότασης 2.3.1.1 ότι ο υπολογισμός των πιθανοτήτων γίνεται άμεσα χρησιμοποιώντας τους πίνακες της α.σ.κ της διωνυμικής κατανομής με πιθανότητα επιτυχίας 0.5 (βλέπε Πίνακα 1 Παράρτημα Α). Στην περίπτωση που το μέγεθος δείγματος είναι $n > 10$, προτείνεται να χρησιμοποιείται η ασυμπτωτική

κατανομή των σ.σ. υπό την υπόθεση της συμμετρίας γύρω από το μηδέν. Η κατανομή αυτή προσδιορίστηκε αναλυτικά από τους Chatterjee and Sen (1973) και το αποτέλεσμα δίνεται στην επόμενη πρόταση και αποδείχθηκε, με προσομοιώσεις, ότι είναι ικανοποιητική για μεγέθη δείγματος $n > 40$.

Πρόταση 2.3.1.2 (Butler (1969), Chatterjee and Sen (1973)) Υπό την υπόθεση της συμμετρίας γύρω από το μηδέν και για σταθεροποιημένο $y (0 \leq y < \infty)$ ισχύει ότι:

$$\lim_{n \rightarrow \infty} P(n^{1/2} B_n^+ \geq y) = \lim_{n \rightarrow \infty} P(n^{1/2} B_n^- \geq y) = 2\Phi(-y), \quad (2.3.1.4)$$

ενώ

$$\lim_{n \rightarrow \infty} P(n^{1/2} B_n \geq y) = 4 \sum_{j=1}^{\infty} (-1)^{j+1} \Phi(-(2j-1)y), \quad (2.3.1.5)$$

όπου $\Phi(\cdot)$ η α.σ.κ. της τυπικής κανονικής κατανομής $N(0,1)$.

Είναι προφανές ότι με βάση την Πρόταση 2.3.1.2 μπορεί να καθοριστεί η κρίσιμη τιμή του ελέγχου. Ειδικότερα, για τους μονόπλευρους ελέγχους προκύπτει ότι η κρίσιμη περιοχή για επίπεδο σημαντικότητας α είναι η: $n^{1/2} B_n^+ (n^{1/2} B_n^-) \geq -\Phi^{-1}(\alpha/2)$.

Παρατήρηση 2.3.1.1 (Niederhausen (1979)) Μια ισοδύναμη μορφή των σ.σ. B_n^+ και B_n^- , χρήσιμη σε υπολογισμούς, είναι η ακόλουθη. Αρχικά διατάσσουμε τις παρατηρήσεις κατά απόλυτη τιμή και ορίζουμε τις τυχαίες μεταβλητές

$$T_i = \begin{cases} -1, & \text{αν η } i\text{-οστή μικρότερη παρατήρηση είναι αρνητική} \\ 1, & \text{αν η } i\text{-οστή μικρότερη παρατήρηση είναι θετική.} \end{cases} \quad (2.3.1.6)$$

Τότε, $B_n^+ = \max_{1 \leq j \leq n+1} \sum_{i=j}^{n+1} T_i$ και $B_n^- = \min_{1 \leq j \leq n+1} \sum_{i=j}^{n+1} T_i$, όπου $T_{n+1} = 0$.

2.3.2 Ο έλεγχος των Rothman and Woodroffe (1972)

Στο εδάφιο αυτό θα παρουσιαστεί ο τρόπος ελέγχου της μονοδιάστατης συμμετρίας γύρω από γνωστό σημείο ο οποίος προτάθηκε από τους Rothman and Woodroffe (1972). Όπως έχει ήδη αναφερθεί, πρόκειται για έναν τρόπο ελέγχου που βασίζεται στην ε.α.σ.κ. και είναι ένας Cramer-Von Mises τύπος ελέγχου. Με ανάλογο σκεπτικό με εκείνο των Smirnov (1947) και Butler (1969), οι Rothman and Woodroffe (1972) προτείνουν για τον έλεγχο της συμμετρίας γύρω από το μηδέν, τη χρήση της σ.σ.:

$$R_n = n \int_{-\infty}^{+\infty} [F'_n(x) + F'_n(-x) - 1]^2 dF_n(x), \quad (2.3.2.1)$$

όπου $F'_n(\cdot)$ η σ.σ. που προσδιορίζεται από τη σχέση:

$$2F'_n(x) = F_n(x+0) + F_n(x-0), \text{ για } x \in \mathbb{R}. \quad (2.3.2.2)$$

Παρατηρούμε ότι οι Rothman and Woodroffe (1972) χρησιμοποιούν τη σ.σ. F'_n αντί της F_n . Η ιδέα αυτή στηρίζεται κυρίως στην παρατήρηση ότι υιοθετώντας την F'_n , η σ.σ. R_n παραμένει αναλλοίωτη όταν τα δεδομένα πολλαπλασιάζονται με -1 , η οποία είναι μια επιθυμητή ιδιότητα (βλέπε Παρατήρηση 2.2.1.2). Στο σημείο αυτό πρέπει να αναφερθεί ότι ο υπολογισμός της τιμής της σ.σ. R_n είναι πιο εύκολος χρησιμοποιώντας τη σχέση:

$$R_n = \sum_{i=1}^n \left[F'_n(-X_{(i)}) - \frac{2n-2i+1}{2n} \right]^2, \quad (2.3.2.3)$$

όπου $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ είναι οι διατεταγμένες τιμές των $X_1, X_2, \dots, X_n$.

Είναι προφανές ότι η μηδενική υπόθεση της συμμετρίας απορρίπτεται για μεγάλες τιμές της σ.σ. R_n . Ο καθορισμός της κρίσιμης περιοχής απαιτεί την εύρεση της κατανομής της σ.σ. R_n υπό την H_0 .

Στο παραπάνω πλαίσιο, στη συνέχεια παρατίθεται η ασυμπτωτική κατανομή της σ.σ. R_n υπό την H_0 .

Θεώρημα 2.3.2.1 (Rothman and Woodroffe (1972)) Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από μια συνεχή και συμμετρική συνάρτηση κατανομής F . Τότε, υπό την υπόθεση της συμμετρίας,

$$R_n \xrightarrow{d} R = \int_0^1 (W(t))^2 dt, \text{ καθώς } n \rightarrow \infty, \quad (2.3.2.4)$$

όπου $W(\cdot)$ είναι μία τυπική διαδικασία Wiener στο διάστημα $[0,1]$.

Η παραπάνω οριακή κατανομή του R_n είναι δύσκολο να εφαρμοστεί στην πράξη. Οι Rothman και Woodroffe (1972) θέλοντας να προσδιορίσουν την κρίσιμη τιμή του ελέγχου απέδειξαν ότι η οριακή κατανομή της σ.σ. R_n μπορεί να γραφεί ως ένα σταθμισμένο άθροισμα μη πεπερασμένων το πλήθος χ^2 – τυχαίων μεταβλητών.

Πόρισμα 2.3.2.1 (Rothman and Woodroffe (1972), Shepp (1967))

Για την τ.μ. R του Θεωρήματος 2.3.2.1 ισχύει ότι:

$$R = \sum_{i=1}^{\infty} \frac{4Z_i^2}{(2i-1)^2 \pi^2}, \quad (2.3.2.5)$$

όπου Z_i ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, με $Z_i \sim N(0,1)$.

Επιπρόσθετα, χρησιμοποιώντας το αποτέλεσμα του Πορίσματος 2.3.2.1 οι Rothman and Woodroffe (1972) προσδιόρισαν την ασυμπτωτική κατανομή της σ.σ. R , υπό την μηδενική υπόθεση, ως ανάπτυγμα σειράς.

Θεώρημα 2.3.2.2 (Rothman and Woodroffe (1972)) Η συνάρτηση κατανομής της σ.σ. R υπό τη μηδενική υπόθεση δίνεται από τη σχέση:

$$H(x) = 2^{3/2} \sum_{i=0}^{\infty} \binom{-\frac{1}{2}}{i} \left(1 - \Phi \left(\frac{4i+1}{2x^{1/2}} \right) \right), \quad x > 0, \quad (2.3.2.6)$$

όπου Φ η α.σ.κ της $N(0,1)$.

Αξιοποιώντας το παραπάνω αποτέλεσμα και με τη βοήθεια ηλεκτρονικού υπολογιστή οι Rothman and Woodroffe (1972) προσδιόρισαν για επιλεγμένες τιμές του επιπέδου σημαντικότητας α , τις κρίσιμες τιμές του ελέγχου, έστω x_α , οι οποίες παρατίθενται στον πίνακα που ακολουθεί.

Επίπεδο Σημαντικότητας α	0.10	0.05	0.025	0.01
Κρίσιμη τιμή x_α	1.196	1.656	2.135	2.78

Οι Rothman and Woodroffe (1972) με μελέτη προσομοίωσης συμπέραναν ότι οι παραπάνω κρίσιμες τιμές είναι ικανοποιητικές ακόμα και για δείγματα μεγέθους $n=20$. Συνοψίζοντας, για επίπεδο σημαντικότητας α απορρίπτουμε την H_0 αν: $R_n \geq x_\alpha$.

Τέλος, καθώς η συνάρτηση κατανομής F είναι σταθερή και μη-συμμετρική, ισχύει ότι $R_n \rightarrow \infty$ με πιθανότητα 1 για $n \rightarrow \infty$ και επομένως προκύπτει η ακόλουθη ιδιότητα.

Πόρισμα 2.3.2.2 Τα τεστ που στηρίζονται στη σ.σ. R_n και απορρίπτουν τη μηδενική υπόθεση της συμμετρίας γύρω από το μηδέν για μεγάλες τιμές της σ.σ. R_n , είναι συνεπή έναντι όλων των μη-συμμετρικών εναλλακτικών.

Θα ήταν παράλειψη στο σημείο αυτό να μην αναφερθεί ότι προτάθηκε από τον Orlov (1972), ανεξάρτητα από τους Rothman and Woodroffe (1972), ένας τρόπος ελέγχου της συμμετρίας, ο οποίος είναι ελάχιστα διαφορετικός και ασυμπτωτικά ισοδύναμος με αυτόν των Rothman and Woodroffe (1972). Ειδικότερα, ο Orlov (1972) πρότεινε τη σ.σ.

$$V_n^2 = \int_{-\infty}^0 Q_n^2(x) dF_n(x), \quad (2.3.2.7)$$

όπου $Q_n(x) = (n/2)^{1/2} [F_n(x) + F_n(-x) - 1]$ και η οποία αποδεικνύεται

ότι είναι ασυμπτωτικά ισοδύναμη με τη σ.σ. $\tilde{V}_n^2 = \int_{-\infty}^0 Q_n^2(x) dF(x)$.

Επιπρόσθετα, ο Orlov (1972) απέδειξε ότι:

$$\lim_{n \rightarrow \infty} P(V_n^2 \leq x) = S(x) = P\left(\int_0^{1/2} W^2(t) dt \leq x\right), \quad (2.3.2.8)$$

όπου $W(\cdot)$ είναι μία τυπική διαδικασία Wiener. Στον Πίνακα 2 του Παραρτήματος Α παρατίθενται οι τιμές της $S(x)$ όπως αυτές δόθηκαν από τον Orlov (1972) και διορθώθηκαν από τον Koziol (1980). Ενδεικτικά αναφέρουμε ότι $S(0.4) = 0.9457124$ και $S(0.7) = 0.9901722$. Δηλαδή, με επίπεδο σημαντικότητας 5% απορρίπτεται η μηδενική υπόθεση αν $V_n^2 \geq 0.4$.

Ο Koziol (1980) εκτός του ότι διόρθωσε δύο τυπογραφικά λάθη που υπήρχαν στον πίνακα που δόθηκε από τον Orlov (1972), μελέτησε τις σ.σ. που ανεξάρτητα προτάθηκαν από τους Orlov (1972) και Rothman and Woodroffe (1972). Παρακινούμενος από τους Durbin and Knott (1972), ο Koziol (1980) θεώρησε την ορθογώνια αναπαράσταση της σ.σ. του Orlov (1972) και χρησιμοποίησε τις

ορθογώνιες συνιστώσες της για τον έλεγχο συμμετρίας. Ειδικότερα, για $j \geq 1$, η σ.σ. που προτάθηκε από τον Kozioł (1980) είναι:

$$\tilde{Z}_{nj} = (2/n)^{1/2} \left\{ \sum_{X_i < 0} \cos[(2j-1)\pi F_n(X_i)] - \sum_{X_i > 0} \cos[(2j-1)\pi F_n(-X_i)] \right\}.$$

Η οικογένεια αυτών των σ.σ. αποτελεί ειδική περίπτωση των σ.σ. προσημασμένης διάταξης και για το λόγο αυτό δεν χρήζει περαιτέρω μελέτης σε αυτήν την ενότητα. Αξίζει μόνο να αναφερθεί ότι μεταξύ των μελών της, έχει μέγιστη ισχύ και επομένως προτείνεται η περίπτωση $\tilde{Z}_{n1}$.

2.3.3 Το τεστ των Srinivasan and Godio (1973)

Στην ενότητα αυτή θα παρουσιαστεί ο τρόπος ελέγχου της υπόθεσης της συμμετρίας μιας συνεχούς συνάρτησης κατανομής γύρω από το μηδέν έναντι γενικών εναλλακτικών που προτάθηκε από τους Srinivasan and Godio (1973). Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. μεγέθους n από έναν πληθυσμό με άγνωστη, απόλυτα συνεχή συνάρτηση κατανομής F . Ας είναι $Z_1 < Z_2 < \dots < Z_n$ οι απόλυτες τιμές των $X_1, X_2, \dots, X_n$, διατεταγμένες σε αύξουσα σειρά. Για $j = 1, 2, \dots, n$, έστω N_j η τ.μ. που παριστάνει τον αριθμό των αρνητικών X_i που είναι τέτοια ώστε $|X_i| \leq Z_j$, ενώ P_j η τ.μ. που περιγράφει τον αριθμό των θετικών X_i για τα οποία ισχύει ότι $|X_i| \leq Z_j$. Στο πλαίσιο αυτό, η σ.σ. που προτάθηκε από τους Srinivasan and Godio (1973), έστω SG_n , είναι η:

$$SG_n = \sum_{j=1}^n \left(N_j - \frac{1}{2} j \right)^2. \quad (2.3.3.1)$$

Η διαισθητική έννοια της σ.σ. SG_n είναι σαφής, αφού, υπό την υπόθεση της συμμετρίας, η αναμενόμενη τιμή του N_j είναι $\frac{1}{2}j$, για κάθε j . Αυτό προκύπτει εύκολα καθώς υπό την υπόθεση της συμμετρίας, η τ.μ. N_j ακολουθεί διωνυμική κατανομή με παραμέτρους j και $1/2$, δηλαδή $B(j, 1/2)$. Επομένως, η σ.σ. SG_n δεν είναι τίποτε άλλο παρά ένα μέτρο της απόκλισης των τ.μ. N_j από την αναμενόμενη τιμή τους, υπό τη μηδενική υπόθεση.

Παρατήρηση 2.3.3.1 (Srinivasan and Godio (1973)) Εξ ορισμού, οι τυχαίες μεταβλητές N_j και P_j , $j=1,2,\dots,n$ είναι τέτοιες ώστε $N_j + P_j = j$. Χρησιμοποιώντας τη σχέση αυτή προκύπτει ότι:

$$SG_n = \sum_{j=1}^n \left(N_j - \frac{1}{2}(N_j + P_j) \right)^2 = \frac{1}{4} \sum_{j=1}^n (N_j - P_j)^2. \quad (2.3.3.2)$$

Επιπλέον, ισχύει ότι (βλέπε Hill and Rao (1977)):

$$S_n = \frac{1}{n^2} \sum_{j=1}^n (N_j - P_j)^2 = n \int_{-\infty}^{+\infty} \left\{ F_n^{(0)}(x) + F_n^{(0)}(-x) - 2F_n(0) \right\}^2 dF_n(x), \quad (2.3.3.3)$$

όπου

$$F_n^{(0)}(x) = \begin{cases} F_n(x-), & \text{αν } x \leq 0 \\ F_n(x+), & \text{αν } x > 0. \end{cases} \quad (2.3.3.4)$$

Η ισοδύναμη αυτή μορφή δικαιολογεί την ταξινόμηση αυτού του τρόπου ελέγχου σε αυτήν την κατηγορία.

Είναι προφανές από τον τρόπο κατασκευής της σ.σ. SG_n ότι η υπόθεση της συμμετρίας απορρίπτεται για μεγάλες τιμές της. Για τον ακριβή καθορισμό της κρίσιμης περιοχής απαιτείται η εύρεση της κατανομής της SG_n υπό την υπόθεση της συμμετρίας. Η κατανομή αυτή προσδιορίστηκε από τους Srinivasan and Godio (1973) και

παρατίθεται στο επόμενο Θεώρημα. Σε όσα ακολουθούν με $P_j(i, n)$ συμβολίζεται η πιθανότητα σε ένα σύνολο j παρατηρήσεων από την F να υπάρχουν i αρνητικές παρατηρήσεις τέτοιες ώστε $S_j = u$.

Θεώρημα 2.3.3.1 (Srinivasan and Godio (1973)) Υπό την υπόθεση της συμμετρίας γύρω από το μηδέν ισχύει ότι:

$$P(SG_n = s) = \sum_{i=0}^n P_n(i, s), \quad (2.3.3.5)$$

όπου για κάθε $j \geq 1$ και $i < j$ είναι:

$$P_j(i, u) = \frac{1}{2} \left[P_{j-1} \left(i, u - \left(i - \frac{1}{2} j \right)^2 \right) + P_{j-1} \left(i-1, u - \left(i - \frac{1}{2} j \right)^2 \right) \right], \quad (2.3.3.6)$$

και

$$P_j(0, u) = P_j(j, u) = \begin{cases} \left(\frac{1}{2} \right)^j, & \text{αν } u = j(j+1)(2j+1)/24 \\ 0, & \text{αλλιώς.} \end{cases} \quad (2.3.3.7)$$

Απόδειξη (Srinivasan and Godio (1973))

Η σχέση (2.3.3.5) και οι πιθανότητες της σχέσης (2.3.3.7) προκύπτουν άμεσα από τον τρόπο ορισμού των $P_j(i, u)$. Η σχέση (2.3.3.6) προκύπτει μέσω του διαχωρισμού των ισοπίθανων ενδεχομένων ότι η τ.μ. Z_j είναι η απόλυτη τιμή της μικρότερης αρνητικής παρατήρησης ή ότι η τ.μ. Z_j είναι η απόλυτη τιμή της μεγαλύτερης θετικής παρατήρησης.

Επομένως, από τη συζήτηση που προηγήθηκε, προκύπτει ότι για επίπεδο σημαντικότητας α , απορρίπτεται η υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της γενικής ασυμμετρίας αν $SG_n \geq c_\alpha$, όπου η τιμή c_α είναι τέτοια ώστε $P(SG_n \geq c_\alpha) = \alpha$.

Για τον καθορισμό της τιμής c_α , οι Srinivasan and Godio (1973) χρησιμοποιώντας το προηγούμενο αποτέλεσμα, παραθέτουν στον Πίνακα 1 της εργασίας (βλέπε Πίνακα 3 Παραρτήματος Α) τους τα αποτελέσματα για τον υπολογισμό των πιθανοτήτων της μορφής $P(SG_n \leq x)$, για δείγματα μεγέθους $n=10(1)20$. Ενδεικτικά, προκύπτει ότι για $n=10$ έχουμε $P(SG_n \leq 43.25) = 0.9473$, ενώ για $n=20$ έχουμε $P(SG_n \leq 170.50) = 0.9503$.

Για τους λόγους που αναφέρθηκαν πρωτίτερα όταν παρουσιάστηκαν και άλλοι τρόποι ελέγχου, είναι απαραίτητη η εύρεση της ασυμπτωτικής κατανομής της σ.σ. SG_n υπό την H_0 . Στο επόμενο θεώρημα δίνεται η ασυμπτωτική κατανομή του $S_n = \frac{4SG_n}{n^2}$ υπό την H_0 και το αποτέλεσμα αυτό μπορεί να χρησιμοποιηθεί για τον προσδιορισμό των κρίσιμων τιμών για δείγματα μεγέθους $n > 20$.

Θεώρημα 2.3.3.2 (Srinivasan and Godio (1973)) Υπό την υπόθεση της συμμετρίας της F γύρω από το μηδέν ισχύει ότι:

$$\lim_{n \rightarrow \infty} P(S_n \leq x) = S(x) = P\left(\int_0^{1/2} W^2(t) dt \leq x\right), \quad (2.3.3.8)$$

με $W(t)$ να είναι μια διαδικασία Wiener στο διάστημα $\left[0, \frac{1}{2}\right]$.

Από το Θεώρημα 2.3.3.2 προκύπτει ότι η σ.σ. S_n των Srinivasan and Godio (1973) είναι ασυμπτωτικά ισοδύναμη με τη σ.σ. του Orlov (1972), καθώς ακολουθούν την ίδια ασυμπτωτική κατανομή. Ως συνέπεια της ισοδυναμίας των σ.σ. SG_n και R_n και λαμβάνοντας υπόψη το Πόρισμα 2.3.2.2, προκύπτει ότι ο έλεγχος αυτός είναι συνεπής έναντι όλων των μη-συμμετρικών εναλλακτικών της H_0 .

Συνοψίζοντας, αν χρησιμοποιηθεί η σ.σ. S_n της σχέσης (2.3.3.3), η υπόθεση της συμμετρίας απορρίπτεται αν $S_n \geq c_\alpha$ και η κρίσιμη τιμή c_α προσδιορίζεται από τη σχέση :

$$S(c_\alpha) = 1 - \alpha. \quad (2.3.3.9)$$

Στον Πίνακα 2 του Παραρτήματος Α παρατίθενται οι τιμές της $S(x)$, για διάφορες επιλογές του x , όπως αυτές δόθηκαν από τον Orlov (1972) και διορθώθηκαν από τον Koziol (1980).

2.3.4 Ο έλεγχος των Hill and Rao (1977)

Στις προηγούμενες ενότητες παρουσιάστηκαν οι έλεγχοι συμμετρίας που προτάθηκαν από τους Rothman and Woodrooffe (1972) καθώς και από τους Srinivasan and Godio (1973). Οι Hill and Rao (1977) αρχικά γενίκευσαν τις σ.σ. R_n και S_n σε δύο οικογένειες σ.σ. και μελέτησαν τις ιδιότητες τους. Έπειτα, συνδυάζοντας αυτές τις δύο οικογένειες πρότειναν μια νέα για τον έλεγχο της συμμετρίας μιας συνεχούς κατανομής γύρω από μια γνωστή διάμεσο. Τα μέλη αυτής της νέας οικογένειας προκύπτει ότι παραμένουν αναλλοίωτα κάτω από συγκεκριμένους μετασχηματισμούς δεδομένων, ιδιότητα που δεν την έχουν οι σ.σ. S_n και R_n και η οποία είναι επιθυμητή για τον υπό μελέτη έλεγχο (βλέπε Παρατήρηση 2.2.1.2).

Ειδικότερα, στο πλαίσιο που παρουσιάστηκε στα προηγούμενα εδάφια, οι Hill and Rao (1977) εισήγαγαν τις σ.σ.

$$R_n^{(a)} = n \int_{-\infty}^{+\infty} \{F_n^{(a)}(x) + F_n^{(a)}(-x) - 1\}^2 dF_n(x), \quad (2.3.4.1)$$

όπου για $0 \leq a \leq 1$

$$F_n^{(a)}(x) = \begin{cases} aF_n(x+) + (1-a)F_n(x-), & \text{αν } x \leq 0 \\ (1-a)F_n(x+) + aF_n(x-), & \text{αλλιώς.} \end{cases} \quad (2.3.4.2)$$

Έτσι, η πρώτη οικογένεια σ.σ. είναι η $\{R_n^{(a)}, \text{ για } 0 \leq a \leq 1\}$ και προκύπτει αντικαθιστώντας στη σ.σ. των Rothman and Woodroffe (1972) την F_n' με την $F_n^{(a)}$. Άμεσα συμπεραίνουμε ότι η σ.σ. R_n είναι μέλος αυτής της οικογένειας αφού για $a = \frac{1}{2}$ έχουμε ότι $R_n^{(1/2)} = R_n$.

Η δεύτερη οικογένεια σ.σ. η οποία εισήχθη από τους Hill and Rao (1977) είναι η $\{S_n^{(a)}, \text{ για } 0 \leq a \leq 1\}$ και προέκυψε αντικαθιστώντας στη σ.σ. S_n των Srinivasan and Godio (1974) την $F_n^{(0)}$ με την $F_n^{(a)}$. Επομένως,

$$S_n^{(a)} = n \int_{-\infty}^{+\infty} \{F_n^{(a)}(x) + F_n^{(a)}(-x) - 2F_n(0)\}^2 dF_n(x), \quad (2.3.4.3)$$

με την $F_n^{(a)}(\cdot)$ να δίνεται στη σχέση (2.3.4.2). Η σ.σ. S_n είναι μέλος της οικογένειας αυτής αφού για $a = 0$ προκύπτει ότι $S_n^{(0)} = S_n$.

Οι παραπάνω μορφές των $S_n^{(a)}$ και $R_n^{(a)}$ είναι δύσκολα εφαρμόσιμες στην πράξη. Για το λόγο αυτό οι Hill and Rao (1977) παραθέτουν κάποιες ισοδύναμες μορφές αυτών, που είναι χρήσιμες τόσο για υπολογιστικούς σκοπούς όσο και για την μελέτη των ιδιοτήτων τους. Ειδικότερα, έστω $Z_1 < Z_2 < \dots < Z_n$ οι απόλυτες τιμές των $X_1, X_2, \dots, X_n$ διατεταγμένες σε αύξουσα σειρά και για $k = 1, 2, \dots, n$, $\delta_k = -1$ αν το Z_k αντιστοιχεί σε αρνητική παρατήρηση και $\delta_k = 1$ αν το Z_k αντιστοιχεί σε θετική παρατήρηση. Επιπρόσθετα, αν

$$P_k = \frac{1}{2} \sum_{j=1}^k (1 + \delta_j), \quad (2.3.4.4)$$

και

$$N_k = \frac{1}{2} \sum_{j=1}^k (1 - \delta_j), \quad (2.3.4.5)$$

τότε

$$R_n^{(a)} = n^{-2} \sum_{k=1}^n \{N'_k - P'_k + (1-a)\delta'_k\}^2, \quad (2.3.4.6)$$

και

$$S_n^{(a)} = n^{-2} \sum_{k=1}^n \{N_k - P_k + a\delta_k\}^2, \quad (2.3.4.7)$$

όπου N'_k , P'_k και δ'_k είναι οι αντίστοιχες τιμές των N_k , P_k και δ_k υπολογισμένες για τις δειγματικές παρατηρήσεις $X_1^{-1}, X_2^{-1}, \dots, X_n^{-1}$.

Στη συνέχεια παρατίθενται κάποιες ιδιότητες για τις οικογένειες σ.σ. των Hill and Rao (1977).

Ιδιότητα 1 (Hill and Rao (1977)) Οι σ.σ. $R_n^{(a)}$ και $S_n^{(a)}$ παραμένουν αναλλοίωτες κάτω από το μετασχηματισμό $(X_1, X_2, \dots, X_n) \rightarrow (-X_1, -X_2, \dots, -X_n)$.

Το αποτέλεσμα προκύπτει καθώς υπό τον μετασχηματισμό $(X_1, X_2, \dots, X_n) \rightarrow (-X_1, -X_2, \dots, -X_n)$, $N_k - P_k = -\{N'_k - P'_k\}$ και $\delta_k = -\delta'_k$. Επιπλέον, καθώς ισχύει ότι $\delta_k = \delta'_{n-k+1}$, με βάση τις σχέσεις (2.3.4.4)-(2.3.4.7) έχουμε την ακόλουθη ιδιότητα.

Ιδιότητα 2 (Hill and Rao (1977)) Η τιμή της σ.σ. $S_n^{(a)}$ υπολογισμένη στο τ.δ. $X_1, X_2, \dots, X_n$, ισούται με την τιμή της σ.σ. $R_n^{(1-a)}$ υπολογισμένη στο δείγμα των αντιστρόφων του αρχικού δείγματος, δηλαδή στις παρατηρήσεις $X_1^{-1}, X_2^{-1}, \dots, X_n^{-1}$.

Από τους Hill and Rao (1977) προκύπτει ότι για κάθε $0 \leq a \leq 1$, οι κατανομές των $S_n^{(a)}$ και $R_n^{(1-a)}$ ταυτίζονται υπό την υπόθεση της συμμετρίας γύρω από το μηδέν. Έτσι έχουμε την ακόλουθη ιδιότητα.

Ιδιότητα 3 (Hill and Rao (1977)) Οι κρίσιμες τιμές της σ.σ. S_n που προσδιορίστηκαν από τους Srinivasan and Godio (1973) είναι και κρίσιμες τιμές της σ.σ. $R_n^{(1)}$. Επιπλέον, οι κρίσιμες τιμές της σ.σ. R_n είναι και κρίσιμες τιμές της σ.σ. $S_n^{(1/2)}$, οι οποίες μπορούν να υπολογιστούν ακολουθώντας ανάλογη πορεία με αυτήν των Srinivasan and Godio (1973).

Υπό την υπόθεση της συμμετρίας, οι Hill and Rao (1977) απέδειξαν ότι οι ασυμπτωτικές κατανομές των $R_n^{(a)}$ και $S_n^{(a)}$ δεν εξαρτώνται από την τιμή του a . Με βάση αυτήν την πρόταση έχουμε το ακόλουθο αποτέλεσμα.

Ιδιότητα 4 (Hill and Rao (1977)) Υπό την υπόθεση της συμμετρίας, οι ασυμπτωτικές κατανομές των $R_n^{(a)}$ και $S_n^{(a)}$ είναι ίδιες με τις ασυμπτωτικές κατανομές των R_n και S_n , αντίστοιχα, που προσδιορίστηκαν σε προηγούμενες ενότητες.

Τέλος, οι Hill and Rao (1977) απέδειξαν ότι υπό την υπόθεση της ασυμμετρίας, $R_n^{(a)} \rightarrow \infty$ με πιθανότητα 1, για $0 \leq a \leq 1$. Συνδυάζοντας αυτό το αποτέλεσμα με την Ιδιότητα 2 προκύπτει ότι:

Ιδιότητα 5 (Hill and Rao (1977)) Οι σ.σ. $R_n^{(a)}$ και $S_n^{(a)}$ παρέχουν για κάθε $0 \leq a \leq 1$ συνεπή τεστ για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από το μηδέν.

Έχοντας ορίσει τις δύο οικογένειες σ.σ. προκύπτει το ερώτημα ποιες σ.σ. από τις κλάσεις $\{R_n^{(a)}, \text{για } 0 \leq a \leq 1\}$ και $\{S_n^{(a)}, \text{για } 0 \leq a \leq 1\}$ είναι καταλληλότερες. Οι Hill and Rao (1977) προτείνουν για την S_n

οικογένεια να χρησιμοποιείται η τιμή $a=0$, ενώ για την R_n οικογένεια να χρησιμοποιείται η τιμή $a=1$. Πλεονέκτημα των επιλογών αυτών είναι ότι οι σ.σ. υπολογίζονται ευκολότερα και επιπλέον ότι οι σ.σ. $S_n^{(0)}$ και $R_n^{(1)}$ έχουν την ιδιότητα να παραμένουν αναλλοίωτες υπό τον μετασχηματισμό $(X_1, X_2, \dots, X_n) \rightarrow (-X_1, -X_2, \dots, -X_n)$, μια ιδιότητα επιθυμητή στους ελέγχους συμμετρίας. Άλλη μια επιθυμητή ιδιότητα σε τέτοιους ελέγχους είναι το να παραμένουν αναλλοίωτες οι σ.σ. υπό τον μετασχηματισμό $(X_1, X_2, \dots, X_n) \rightarrow (X_1^{-1}, X_2^{-1}, \dots, X_n^{-1})$. Ωστόσο, καμία από τις σ.σ. $S_n^{(0)}$ και $R_n^{(1)}$ δεν είναι αναλλοίωτη υπό το μετασχηματισμό $(X_1, X_2, \dots, X_n) \rightarrow (X_1^{-1}, X_2^{-1}, \dots, X_n^{-1})$. Αυτό το μειονέκτημα, έδωσε την ιδέα στους Hill and Rao (1977) να δημιουργήσουν μία νέα οικογένεια σ.σ., συνδυάζοντας τις σ.σ. $R_n^{(a)}$ και $S_n^{(a)}$. Σκοπός τους, ήταν η δημιουργία ενός ελέγχου ο οποίος θα είναι συνεπής και θα στηρίζεται σε μια σ.σ. η οποία θα παραμένει αναλλοίωτη στους προαναφερθέντες μετασχηματισμούς. Με αυτό ως στόχο, πρότειναν την οικογένεια των σ.σ. $\{T_n^{(a)}, 0 \leq a \leq 1\}$, όπου

$$T_n^{(a)} = \frac{1}{2} (S_n^{(a)} + R_n^{(1-a)}). \quad (2.3.4.8)$$

Για τη νέα αυτή οικογένεια σ.σ. οι Hill and Rao (1977) απέδειξαν ότι ικανοποιεί τις προαναφερθείσες επιθυμητές ιδιότητες και επιπλέον ότι μας εφοδιάζει με ένα συνεπές τεστ για τον έλεγχο της συμμετρίας έναντι γενικών εναλλακτικών.

Για τον έλεγχο της H_0 μια λογική επιλογή στατιστικού είναι το $T_n^{(0)}$, αφού για $a=0$ εμφανίζονται οι σ.σ. $S_n^{(0)}$ και $R_n^{(1)}$ που προτάθηκαν προηγουμένως. Κρίσιμες τιμές του στατιστικού $T_n^{(0)}$ δίνονται από τους Hill and Rao (1977) για μεγέθη δείγματος από $n=10$ μέχρι $n=24$ και για διάφορα επίπεδα σημαντικότητας α . Ο Hill (1976) στην διδακτορική διατριβή του επισημαίνει ότι οι κρίσιμες τιμές που

αντιστοιχούν σε μέγεθος δείγματος $n=24$ μπορούν να χρησιμοποιηθούν δίνοντας ικανοποιητικά αποτελέσματα και για μεγέθη δείγματος μεγαλύτερα από 24. Αυτό γίνεται διότι η ασυμπτωτική κατανομή του $T_n^{(0)}$ υπό την H_0 είναι ιδιαίτερα δύσχρηστη. Στον Πίνακα 4 του Παραρτήματος Α παρατίθεται ο Πίνακας των κρίσιμων τιμών που δόθηκε από τον Hill (1976).

2.3.5 Το τεστ των Baringhaus and Henze (1992)

Στην παράγραφο αυτή θα παρουσιαστούν οι σ.σ. που προτάθηκαν από τους Baringhaus and Henze (1992), για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από το μηδέν. Αρχικά, θα παρουσιαστεί ο χαρακτηρισμός της συμμετρίας που χρησιμοποίησαν και έπειτα θα προσδιοριστεί η ασυμπτωτική κατανομή των σ.σ. υπό την H_0 .

Οι σ.σ. των Baringhaus and Henze (1992) στηρίζονται στο χαρακτηρισμό που ακολουθεί.

Λήμμα 2.3.5.1 (Baringhaus and Henze (1992)) Έστω X και Y ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, με συνεχή συνάρτηση κατανομής F . Η F είναι συμμετρική γύρω από το μηδέν αν και μόνο αν οι τυχαίες μεταβλητές $\max(|X, Y|)$ και $|X|$ έχουν την ίδια κατανομή.

Απόδειξη

Η απόδειξη προκύπτει εύκολα καθώς η α.σ.κ. της τ.μ. $|X|$ είναι $G(x) = F(x) - F(-x)$, $\forall x \geq 0$ και της τ.μ. $\max(|X, Y|)$ είναι $G(x) [F(x) + F(-x)]$, $\forall x \geq 0$.

Με βάση αυτόν το χαρακτηρισμό, ένας έλεγχος της υπόθεσης της συμμετρίας μπορεί να στηρίζεται στην απόκλιση των εκτιμητών των α.σ.κ. των τ.μ. $|X|$ και $\max(X, Y)$.

Ειδικότερα, αν $G_n(x) = n^{-1} \sum_{j=1}^n I(|X_j| \leq x)$ είναι η εμπειρική συνάρτηση κατανομής των $|X_i|$ και $H_n(x) = \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq n} I(|\max(X_j, X_k)| \leq x)$ η εμπειρική συνάρτηση κατανομής των $\binom{n}{2}$ το πλήθος τυχαίων μεταβλητών $|\max(X_j, X_k)|$, για $1 \leq j < k \leq n$, οι Baringhaus and Henze (1992) ορίζουν τις σ.σ.:

$$KS_n = \sqrt{n} \sup_{x \geq 0} |H_n(x) - G_n(x)|, \quad (2.3.5.1)$$

και

$$CM_n = n \int_0^\infty [H_n(x) - G_n(x)]^2 dG_n(x). \quad (2.3.5.2)$$

Οι έλεγχοι με τις σ.σ. KS_n και CM_n δεν είναι παρά τρόποι ελέγχου τύπου Kolmogorov-Smirnov και Cramer-Von Mises, αντίστοιχα. Είναι προφανές ότι όταν η H_0 είναι αληθής αναμένουμε μικρές τιμές των KS_n και CM_n .

Καθώς οι σχέσεις (2.3.5.1) και (2.3.5.2) δεν είναι εύχρηστες στην πράξη για τον υπολογισμό της τιμής των σ.σ. KS_n και CM_n , οι Baringhaus and Henze (1992) παρέθεσαν εναλλακτικές μορφές για την αντιμετώπιση αυτού του μειονεκτήματος. Ειδικότερα, έστω $0 < Z_1 < Z_2 < \dots < Z_n$ τα διατεταγμένα στατιστικά των $|X_1|, |X_2|, \dots, |X_n|$

και έστω D_k ο δείκτης εκείνος για τον οποίο ισχύει ότι $|X_{D_k}| = Z_k$. Έστω S_k η δείκτρια τ.μ. που μας δίνει το πρόσημο του X_{D_k} . Δηλαδή,

$$S_k = \text{sign}(X_{D_k}) = \begin{cases} 1, & \text{αν } X_{D_k} > 0 \\ -1, & \text{αν } X_{D_k} < 0. \end{cases} \quad (2.3.5.3)$$

Υπό τη μηδενική υπόθεση της συμμετρίας τα $S_1, S_2, \dots, S_n$ είναι ανεξάρτητες και ισόνομες τ.μ. με $P(S_i = 1) = \frac{1}{2} = P(S_i = -1)$. Στο πλαίσιο αυτό, οι Baringhaus and Henze (1992) απέδειξαν ότι:

$$KS_n(t) = \frac{1}{(n-1)\sqrt{n}} \max_{1 \leq k \leq n-1} \left| k \sum_{i=k+1}^n S_i \right|, \quad (2.3.5.4)$$

και

$$CM_n(t) = \frac{1}{(n-1)^2} \sum_{k=1}^{n-1} \left(\frac{k}{n} \right)^2 \left(\sum_{i=k+1}^n S_i \right)^2. \quad (2.3.5.5)$$

Επιπλέον, οι Baringhaus and Henze (1992) παραθέτουν εναλλακτικές μορφές υπολογισμού της σ.σ. B_n του Butler (1969) της σχέσης (2.3.1.3), της σ.σ. SG_n των Srinivasan and Godio (1974) της σχέσης (2.3.3.1), της σ.σ. T_s του προσημικού ελέγχου της σχέσης (2.1.1.1) και της σ.σ. T_n' του ελέγχου προσημασμένης διάταξης του Wilcoxon (Wilcoxon signed rank test). Ειδικότερα, είναι:

$$B_n(t) = \frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k S_i \right|, \quad (2.3.5.6)$$

$$SG_n(t) = \frac{1}{4n^2} \sum_{k=1}^n \left(\sum_{i=1}^k S_i \right)^2, \quad (2.3.5.7)$$

$$SI_n(t) = \frac{1}{2\sqrt{n}} \left| \sum_{i=1}^n S_i \right|, \quad (2.3.5.8)$$

και τέλος

$$T'_n = \frac{1}{\sqrt{n(n+1)}} \left| \sum_{k=1}^n k S_k \right|. \quad (2.3.5.9)$$

Παρατήρηση 2.3.5.1 Οι παραπάνω σ.σ. εξαρτώνται μόνο από τα $S_1, S_2, \dots, S_n$ και συνεπώς είναι ανεξάρτητες κατανομής υπό την H_0 .

Όπως έχουμε ήδη αναφέρει η υπόθεση της συμμετρίας απορρίπτεται για μεγάλες τιμές των KS_n και CM_n . Για την εύρεση των κρίσιμων τιμών απαιτείται η εύρεση της κατανομής των KS_n και CM_n υπό την H_0 . Οι Baringhaus and Henze (1992) κατέληξαν στα αποτελέσματα που δίνονται στο ακόλουθο θεώρημα.

Θεώρημα 2.3.5.1 (Baringhaus and Henze (1992)) Υπό την H_0 , η κατανομή των σ.σ. KS_n και CM_n είναι ίδια με την κατανομή των

$$\max_{0 \leq t \leq 1} |W(1) - W(t)|, \quad (2.3.5.10)$$

και

$$\int_0^1 t^2 (W(1) - W(t))^2 dt, \quad (2.3.5.11)$$

αντίστοιχα, όπου $\{W(t), 0 \leq t \leq 1\}$ είναι μια συνήθης διαδικασία Wiener στο διάστημα $[0,1]$.

Ο προσδιορισμός της κατανομής της $\int_0^1 t^2 (W(1) - W(t))^2 dt$ δεν ήταν εφικτός, ενώ οι Baringhaus and Henze (1992) απέδειξαν ότι η κατανομή του $\max_{0 \leq t \leq 1} |W(1) - W(t)|$ είναι ισοδύναμη με αυτή της τ.μ.

$$C = \sum_{k \geq 1} \lambda_k N_k^2, \quad (2.3.5.12)$$

όπου $N_1, N_2, \dots$ είναι ανεξάρτητες $N(0,1)$ τυχαίες μεταβλητές. Καθώς οι Baringhaus and Henze (1992) κατάφεραν να προσδιορίσουν τις ροπές $E(C)$, $Var(C)$, $E[C - E(C)]^3$ και $E[C - E(C)]^4$, ήταν εφικτή η προσέγγιση της κατανομής της τ.μ. C από ένα μέλος της οικογένειας κατανομών του Johnson (Johnson (1949)). Αυτό επιτυγχάνεται καθώς υπάρχει μοναδικό μέλος αυτής της οικογένειας που έχει δοθείσες τιμές για τις 4 προαναφερθείσες ροπές. Ειδικότερα, προέκυψε ότι έχει σ.π.π.:

$$f_T(t) = \gamma + \delta \log \left(\frac{T - \xi}{\xi + \lambda - T} \right), \quad \xi < T < \xi + \lambda, \quad (2.3.5.13)$$

με $(\gamma, \delta, \lambda, \xi) = (3.2009489, 0.9952496, 1.5033677, -0.00324769)$. Έτσι προκύπτει ο ακόλουθος πίνακας κρίσιμων τιμών.

Επίπεδο σημαντικότητας α	0.2	0.1	0.05	0.025	0.01
Κρίσιμη Τιμή KS_n ελέγχου	0.125	0.188	0.257	0.332	0.438

Θεώρημα 2.3.5.2 (Baringhaus and Henze 1992) Για οποιαδήποτε σταθερή εναλλακτική υπόθεση έχουμε ότι ο έλεγχος των σ.σ. KS_n και CM_n είναι συνεπής.

Παρατήρηση 2.3.5.2 Σύμφωνα με τα αποτελέσματα μιας μελέτης προσομοίωσης που διεξήχθη από τους Baringhaus and Henze (1992), προέκυψε ότι οι έλεγχοι που στηρίζονται στις σ.σ. KS_n και CM_n διατηρούν το επίπεδο σημαντικότητας 0.05 και είναι ανταγωνιστικοί από άποψη ισχύος με τους ελέγχους του Butler (1969), των Srinivasan

and Godio (1974), τον προσημικό έλεγχο και τέλος τον έλεγχο προσημασμένης διάταξης Wilcoxon.

2.4 Ο έλεγχος συμμετρίας του Tajuddin (1994) και μια τροποποίηση του

Στην ενότητα αυτή, αρχικά, θα παρουσιασθεί ο τρόπος ελέγχου που προτάθηκε από τον Tajuddin (1994) για τον έλεγχο της υπόθεσης ότι μια α.σ.κ. F είναι συμμετρική γύρω από το μηδέν, έναντι της εναλλακτικής υπόθεσης ότι η F είναι ασύμμετρη. Ο έλεγχος αυτός κατά μια έννοια βασίζεται στον έλεγχο του Wilcoxon για δύο δείγματα. Έπειτα, θα παρουσιασθεί μια τροποποίηση αυτού του ελέγχου που προτάθηκε από τους Modarres and Gastwirth (1998).

2.4.1 Ο έλεγχος συμμετρίας του Tajuddin (1994)

Η ιδέα του Tajuddin (1994) οφείλεται σε μια αδυναμία των ελέγχων που προτάθηκαν σε προηγούμενες ενότητες και περιγράφηκε από τον Tajuddin (1994) ως εξής. Έστω n_1 ο αριθμός των θετικών παρατηρήσεων και $n_2 = n - n_1$ ο αριθμός των αρνητικών παρατηρήσεων σε ένα δείγμα n παρατηρήσεων από έναν πληθυσμό με διάμεσο ίση με μηδέν. Τότε η πιθανότητα $P(|n_1 - n_2| > 0)$ είναι μεγαλύτερη στις λοξές κατανομές σε σχέση με τις συμμετρικές κατανομές. Ο Tajuddin (1994) επισημαίνει ότι το τεστ που προτάθηκε από τον McWilliams (1990), επηρεάζεται από τη διαφορά $|n_1 - n_2|$ και συγκεκριμένα, μεγάλη τιμή αυτής τείνει να μειώσει τον αριθμό των ροών. Ωστόσο, για παράδειγμα, το προσημασμένης διάταξης τεστ του Wilcoxon (Wilcoxon Signed-rank test) λαμβάνει υπόψη το πλήθος των

n_1 ή n_2 αντί της απόλυτης διαφοράς $|n_1 - n_2|$ και έτσι η μηδενική υπόθεση δεν απορρίπτεται πολύ συχνά όταν η αληθής κατανομή είναι ασύμμετρη. Το πρόβλημα της μικρής απόδοσης ισχύος σύμφωνα με τον Tajuddin (1994), μπορεί να αντιμετωπιστεί διαχωρίζοντας το τ.δ. όπως παραπάνω και χρησιμοποιώντας τον έλεγχο του Wilcoxon για δύο δείγματα, για τις δοθείσες τιμές των n_1, n_2 .

Ειδικότερα, παρακινούμενος από τον McWilliams (1990), ο Tajuddin (1994) θεώρησε ως δύο δείγματα τα σύνολα των n_1 θετικών και n_2 αρνητικών παρατηρήσεων και υπέθεσε ότι αυτά προέρχονται από τις δεσμευμένες κατανομές G και H , αντίστοιχα, οι οποίες ορίζονται ως:

$$G(x) = P(X \leq x \mid X < 0), \quad (2.4.1.1)$$

και

$$H(x) = P(X \leq x \mid X > 0), \quad (2.4.1.2)$$

αντίστοιχα.

Υπό την υπόθεση της συμμετρίας γύρω από το μηδέν, ισχύει ότι οι κατανομές των απολύτων τιμών των παρατηρήσεων από τους πληθυσμούς G ή H ταυτίζονται, δηλαδή $G_{-x}(x) = H_x(x)$, για κάθε x . Έτσι, ο έλεγχος της υπόθεσης ότι η F είναι συμμετρική γύρω από το μηδέν ανάγεται στον έλεγχο της υπόθεσης

$$H_0 : G_{-x}(x) = H_x(x), \text{ για κάθε } x, \quad (2.4.1.3)$$

έναντι της εναλλακτικής

$$H_1 : G_{-x}(x) \neq H_x(x), \text{ για κάποια } x. \quad (2.4.1.4)$$

Σε ένα δείγμα μεγέθους n με n_2 αρνητικές παρατηρήσεις, θεωρούμε τις απόλυτες τιμές των n_2 το πλήθος αρνητικών

παρατηρήσεων που προέρχονται από την κατανομή G_{-X} να είναι το ένα δείγμα, ενώ τις υπόλοιπες n_1 το πλήθος θετικές παρατηρήσεις να είναι το άλλο δείγμα μεγέθους n_1 από την κατανομή H . Υπό αυτές τις υποθέσεις, η σ.σ. που προτάθηκε από τον Tajuddin (1994), συμβολίζεται με S και ορίζεται να είναι η σ.σ. του τεστ του Wilcoxon για δύο δείγματα μείον το άθροισμα των τάξεων των παρατηρήσεων που προέρχονται από τους δύο πληθυσμούς G και H . Ορίζοντας για $k = 1, 2, \dots, n$, τις τ.μ. V_k ως

$$V_k = \begin{cases} 1, & \text{αν η } X_{(k)} \text{ προέρχεται από την } H \\ 0, & \text{αλλιώς,} \end{cases} \quad (2.4.1.5)$$

ισχύει ότι

$$S = \sum_{k=1}^n kV_k. \quad (2.4.1.6)$$

Από τον τρόπο ορισμού της σ.σ. και για τον έλεγχο της H_0' έναντι της H_1' που διατυπώθηκαν προωύτερα, προκύπτει ότι για επίπεδο σημαντικότητας α , απορρίπτεται η H_0' , αν $S > S_{1-\alpha/2}$ ή $S < S_{\alpha/2}$, με S_α να είναι το α -εκατοστιαίο σημείο της κατανομής της σ.σ. του Wilcoxon για δύο δείγματα.

Υπό την υπόθεση της συμμετρίας, η μέση τιμή και η διακύμανση του S είναι (βλέπε Gibbons (1971)):

$$E(S) = n_1(n+1)/2, \quad (2.4.1.7)$$

και

$$Var(S) = n_1n_2(n+1)/12, \quad (2.4.1.8)$$

αντίστοιχα. Συνεπώς, για μεγάλα μεγέθη δείγματος μέσω του Κεντρικού Οριακού Θεωρήματος προκύπτει ότι:

$$Z_{TAJ} = \frac{S - E(S)}{\sqrt{\text{Var}(S)}} \stackrel{\text{ασυμπ.}}{\underset{H_0}{\sim}} N(0,1). \quad (2.4.1.9)$$

Επομένως, έχοντας ως εναλλακτική υπόθεση ότι η F είναι ασύμμετρη, απορρίπτουμε την H_0 αν $|Z_{TAJ}| \geq z_{\alpha/2}$.

Παρατήρηση 2.4.1.1 (Gibbons (1971)) Η παραπάνω προσέγγιση μέσω του Κ.Ο.Θ. προκύπτει ότι είναι ικανοποιητική ακόμα και για δείγματα μεγέθους $n=12$.

2.4.2 Ο έλεγχος των Modarres and Gastwirth (1998) και η βελτίωση του

Οι Modarres and Gastwirth (1998) παρακινούμενοι από τον Tajuddin (1994), πρότειναν έναν "υβριδικό" τρόπο ελέγχου, επιπέδου σημαντικότητας α , της συμμετρίας μιας συνεχούς συνάρτησης κατανομής F γύρω από το μηδέν. Η μέθοδος που προτάθηκε από τους Modarres and Gastwirth (1998) αποτελείται από δύο βήματα και βελτιώθηκε, διορθώθηκε από τον Baklizi (2008). Στη συνέχεια θα περιγραφεί η μέθοδος των δύο βημάτων που προτάθηκε από τους Modarres and Gastwirth (1998).

Ειδικότερα, στο πρώτο βήμα, για $X_1, X_2, \dots, X_n$ τ.δ. από την α.σ.κ. F και με επίπεδο σημαντικότητας $\alpha_1 < \alpha$, προτείνουν τη χρησιμοποίηση του προσημικού τεστ για την αποδοχή ή όχι της υπό έλεγχο υπόθεσης. Στο πλαίσιο αυτό και λαμβάνοντας υπόψη τη σχέση (2.1.1.2), η H_0 απορρίπτεται αν:

$$|Z_s| = \left| \frac{T_s - n/2}{\sqrt{n/4}} \right| \geq z_{\alpha_1/2}, \quad (2.4.2.1)$$

όπου n το μέγεθος του δείγματος.

Αν στο πρώτο βήμα απορρίπτεται η μηδενική υπόθεση, τότε η διαδικασία σταματά και αποδεχόμαστε την υπόθεση ότι η F δεν είναι συμμετρική γύρω από το μηδέν. Σε περίπτωση αποδοχής της μηδενικής υπόθεσης, προχωράμε στο βήμα 2, όπου οι Modarres and Gastwirth (1998) προτείνουν να διεξάγεται ένας ακόμα έλεγχος τροποποιώντας τη μέθοδο του Tajuddin (1994). Ειδικότερα, καθώς είναι επιθυμητό η όλη διαδικασία να έχει επίπεδο σημαντικότητας α και στο πρώτο βήμα χρησιμοποιήθηκε επίπεδο σημαντικότητας α_1 , το επίπεδο σημαντικότητας που θα χρησιμοποιείται στον τροποποιημένο έλεγχο του Tajuddin (1994) είναι α_2 , τέτοιο ώστε να ισχύει $\alpha = \alpha_1 + (1 - \alpha_1)\alpha_2$. Τυπικά, το α_1 προτείνεται να είναι μικρότερο από το α_2 . Για παράδειγμα, αν $\alpha_1 = 0.01$ τότε για $\alpha_2 = 0.0404$ προκύπτει ότι $\alpha = 0.05$.

Στο παραπάνω πλαίσιο, παρακινούμενοι από τον Tajuddin (1994) και με παρόμοιο σκεπτικό με αυτό που αναπτύχθηκε στο Εδάφιο 2.1.3 όταν παρουσιάστηκε ο έλεγχος των Modarres and Gastwirth (1996), οι Modarres and Gastwirth (1998) προτείνουν τη σ.σ.:

$$W_p = \sum_{k=np+1}^n \varphi(k) V_k, \quad (2.4.2.2)$$

όπου p είναι το ποσοστό περικοπής, V_k οι τ.μ. που ορίστηκαν στη σχέση (2.4.1.5) και $\varphi(k)$ οι συντελεστές βαρών που ορίστηκαν στη σχέση (2.2.3.2).

Από τους Modarres and Gastwirth (1998) προκύπτει ότι:

$$E(W_p) = \frac{1}{2} n_1 (1 - p) [n(1 - p) + 1], \quad (2.4.2.3)$$

και

$$\begin{aligned} \text{Var}(W_p) = & \frac{n_1 n_2}{12(n-1)} (1-p) [n(1-p)+1] * \\ & * [n(1-p)(3p+1)+3p-1], \end{aligned} \quad (2.4.2.4)$$

με n_1 και n_2 να είναι το πλήθος των θετικών και των αρνητικών παρατηρήσεων του δείγματος μεγέθους n . Για $n \rightarrow \infty$, υπό την H_0 και για κάθε n_1 τέτοιο ώστε $\min(n_1, n-n_1) \rightarrow \infty$, μέσω του Κεντρικού Οριακού Θεωρήματος προκύπτει ότι:

$$Z_{\text{MOG}} = \frac{W_p - E(W_p)}{(\text{Var}(W_p))^{1/2}} \stackrel{H_0}{\underset{\text{ασυμπ.}}{\sim}} N(0,1). \quad (2.4.2.5)$$

Επομένως, στο δεύτερο βήμα με επίπεδο σημαντικότητας α_2 , απορρίπτεται η υπόθεση της συμμετρίας γύρω από το μηδέν αν $|Z_{\text{MOG}}| > z_{\alpha_2/2}$.

Παρατήρηση 2.4.2.1 (Modarres and Gastwirth (1998), Baklizi (2008))

Οι Modarres and Gastwirth (1998) ισχυρίστηκαν ότι η διαδικασία που εισήγαγαν συνδυάζει την απλότητα του προσημικού τεστ με την υψηλή αποτελεσματικότητα του τεστ του Wilcoxon. Όμως, ο Baklizi (2008) έδειξε ότι το πρώτο βήμα του παραπάνω ελέγχου δεν είναι αποτελεσματικό για την ανίχνευση αποκλίσεων από τη συμμετρία. Αυτό συμβαίνει καθώς η κατανομή του T_s , τόσο υπό τη μηδενική υπόθεση όσο και υπό την εναλλακτική, είναι διωνυμική με παράμετρο $p=1/2$, διότι η διάμεσος είναι γνωστή και αμετάβλητη, ανεξάρτητα από το αν η κατανομή είναι συμμετρική ή όχι. Με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης που έγινε από τον Baklizi (2008), προκύπτει ότι το τεστ που περιλαμβάνει μόνο το δεύτερο βήμα του παραπάνω ελέγχου είναι πιο ισχυρό από το τεστ των δύο βημάτων των Modarres and Gastwirth (1998). Αυτό συμβαίνει καθώς το προσημικό τεστ δεν είναι ευαίσθητο σε αποκλίσεις από τη συμμετρία, με αποτέλεσμα το μέγεθος του προσημικού τεστ του πρώτου βήματος

να μη συμβάλει, παρά ελάχιστα, στη συνολική ισχύ. Επομένως, ο Baklizi (2008) προτείνει μόνο το δεύτερο βήμα, δηλαδή να χρησιμοποιείται μόνο ο έλεγχος της σ.σ. W_p .

Παρατήρηση 2.4.2.2 Οι Modarres and Gastwirth (1998) προτείνουν η παράμετρος p να επιλέγεται ανάλογα με το μέγεθος του δείγματος. Συγκεκριμένα, για $n < 50$ προτείνουν την τιμή $p = 0.7$, ενώ για $n \geq 50$ προτείνουν την τιμή $p = 0.8$.

Παρατήρηση 2.4.2.3 (Koopman (1979)) Η ιδέα της σύνθεσης δύο διαδικασιών για τον έλεγχο της συμμετρίας, πρωτοπραγματοποιήθηκε, όσο είμαστε σε θέση να γνωρίζουμε, από τον Koopman (1979). Ειδικότερα, ο Koopman (1979) πρότεινε μια μεθοδολογία ελέγχου της συμμετρίας γύρω από γνωστό σημείο συνδυάζοντας τον προσημικό έλεγχο και τον προσημασμένης διάταξης έλεγχο.

ΚΕΦΑΛΑΙΟ 3

Έλεγχος μονοδιάστατης συμμετρίας γύρω από άγνωστο σημείο

Στο προηγούμενο κεφάλαιο δόθηκε μια ανασκόπηση των κυριότερων μεθοδολογιών ελέγχου της υπόθεσης της συμμετρίας μιας τ.μ. γύρω από ένα γνωστό σημείο, το οποίο χωρίς βλάβη της γενικότητας θεωρήθηκε ότι είναι το μηδέν. Καθώς πολλές φορές η επιθυμητή ιδιότητα είναι αυτή της συμμετρίας γύρω από οποιοδήποτε σημείο, η ανάπτυξη μεθοδολογιών για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από ένα άγνωστο σημείο προσέλκυσε το ενδιαφέρον των ερευνητών. Στο πλαίσιο αυτό, αντικείμενο μελέτης αυτού του κεφαλαίου είναι η κριτική ανασκόπηση, η παρουσίαση και ταξινόμηση, αν είναι εφικτή, των κυριότερων και δημοφιλέστερων μεθόδων ελέγχου της παραπάνω υπόθεσης. Ειδικότερα, θα δίνεται η κεντρική ιδέα στην οποία στηρίζεται ο κάθε τρόπος ελέγχου και έπειτα θα παρατίθεται όπου είναι εφικτό είτε η ακριβής είτε η ασυμπτωτική κατανομή της ελεγχουσυνάρτησης υπό τη μηδενική υπόθεση. Τέλος, θα δίνονται κάποιες ιδιότητες των στατιστικών τεστ. Η συγκριτική μελέτη των μεθόδων που θα παρουσιασθούν αποτελεί αντικείμενο μελέτης του επόμενου κεφαλαίου.

3.1 Επεκτάσεις των μεθόδων ελέγχου συμμετρίας γύρω από γνωστό σημείο

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν μέθοδοι ελέγχου της συμμετρίας γύρω από ένα γνωστό σημείο. Κάποιες από αυτές τις μεθόδους έχουν τροποποιηθεί κατάλληλα έτσι ώστε να μπορούν να

χρησιμοποιηθούν για τον έλεγχο της συμμετρίας γύρω από άγνωστο σημείο. Στην ενότητα αυτή θα παρουσιασθούν εκείνες οι επεκτάσεις, τροποποιήσεις που έχει αποδειχθεί ότι έχουν καλή απόδοση ισχύος. Οι περισσότερες από αυτές προτάθηκαν και μελετήθηκαν πρόσφατα από τους Mukherjee *et al.* (2013).

3.1.1 Τροποποίηση του τεστ των Smirnov (1947) και Butler (1969)

Στην ενότητα 2.3.1 παρουσιάστηκε ο τρόπος ελέγχου της υπόθεσης της συμμετρίας γύρω από γνωστό σημείο που αρχικά προτάθηκε από τον Smirnov (1947) και στη συνέχεια από τον Butler (1969). Πρόσφατα, οι Mukherjee *et al.* (2013) πρότειναν και μελέτησαν μια τροποποίηση του, ώστε να μπορεί να χρησιμοποιηθεί για τον έλεγχο της συμμετρίας γύρω από ένα άγνωστο σημείο, έστω μ .

Ειδικότερα, έστω $X_1, \dots, X_n$ ένα τ.δ. από κάποιον πληθυσμό με σ.π.π. $f(\cdot)$ και α.σ.κ. $F(\cdot)$. Θέλοντας να ελεγχθεί η

$$H_0 : F(\mu + x) = 1 - F(\mu - x), \text{ για κάθε } x, \quad (3.1.1.1)$$

έναντι της εναλλακτικής υπόθεσης της ύπαρξης θετικής λοξότητας, δηλαδή έχοντας ως εναλλακτική την

$$H_1 : F(x + \mu) < 1 - F(\mu - x), \text{ για κάθε } x, \quad (3.1.1.2)$$

οι Mukherjee *et al.* (2013) προτείνουν τη σ.σ.

$$D_n = \max_{i: X_i < \tilde{\mu}} [1 - F_n(X_i) - F_n(2\tilde{\mu} - X_i)], \quad (3.1.1.3)$$

όπου $\tilde{\mu}$ ένας εκτιμητής της άγνωστης πληθυσμιακής διαμέσου και F_n η ε.α.σ.κ.. Η ιδέα τους στηρίζεται στην ιδιότητα ότι υπό την υπόθεση της συμμετρίας γύρω από το μ ισχύει ότι $X \stackrel{d}{=} 2\mu - X$ και αντίστροφα. Με

παρόμοιο σκεπτικό με αυτό του Butler (1969), είναι προφανές ότι με επίπεδο σημαντικότητας α και έχοντας ως εναλλακτική υπόθεση ότι η α.σ.κ. F είναι θετικά λοξή, η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται αν $D_n > c_1$, όπου η τιμή c_1 είναι τέτοια ώστε $P(D_n > c_1 | H_0) = \alpha$.

Για τον προσδιορισμό της τιμής c_1 απαιτείται η εύρεση της κατανομής της σ.σ. D_n υπό την υπόθεση της συμμετρίας. Οι Mukherjee *et al.* (2013) επισημαίνουν ότι για μικρά μεγέθη δείγματος δεν διατηρείται η επιθυμητή ιδιότητα να είναι η κατανομή της σ.σ. ανεξάρτητη της κατανομής $F(\cdot)$, ιδιότητα που ικανοποιούν οι σ.σ. B_n^+ , B_n^- και B_n . Επιπλέον, με τη βοήθεια προσομοιώσεων, οι Mukherjee *et al.* (2013) συμπεραίνουν ότι για μικρά μεγέθη δείγματος ο παραπάνω έλεγχος έχει μικρή ισχύ και ότι η τιμή του κρίσιμου σημείου c_1 διαφοροποιείται ανάλογα με την υποκείμενη συνάρτηση κατανομής. Επομένως, ανάλογα με την αληθινή, αλλά άγνωστη, συνάρτηση κατανομής, διαφοροποιείται η τιμή c_1 . Καθώς κάτι τέτοιο δεν είναι επιθυμητό σε ελέγχους συμμετρίας, οι Mukherjee *et al.* (2013) προτείνουν ως διορθωτικό μέτρο να επιλέγεται ως κρίσιμη τιμή η:

$$c_1^*(\alpha) = \max_{F \in \mathbb{F}} c_1(\alpha), \quad (3.1.1.4)$$

όπου $\mathbb{F}$ είναι η οικογένεια των διαθέσιμων συμμετρικών συναρτήσεων κατανομής. Με πλήθος προσομοιώσεων από διάφορες συμμετρικές κατανομές, οι Mukherjee *et al.* (2013) παρατήρησαν ότι η κρίσιμη τιμή $c_1^*(0.05)$ αντιστοιχεί στην κρίσιμη τιμή $c_1(0.05)$ που προκύπτει από την Βήτα(2,2) κατανομή. Για επίπεδο σημαντικότητας 5% και για διάφορες τιμές του n οι κρίσιμες αυτές τιμές παρατίθενται στον πίνακα που ακολουθεί.

Μέγεθος δείγματος	20	30	50	100	200
Κρίσιμη τιμή $c_1^* (0.05)$	0.200	0.200	0.180	0.130	0.095

3.1.2 Ο τροποποιημένος έλεγχος των Rothman and Woodrooffe (1972)

Οι Rothman and Woodrooffe (1972) με ανάλογο τρόπο με το σκεπτικό των Smirnov (1947) και Butler (1969) πρότειναν έναν τρόπο ελέγχου της υπόθεσης της συμμετρίας γύρω από γνωστό σημείο που βασίζεται στην εμπειρική α.σ.κ.. Οι Mukherjee *et al.* (2013) τροποποίησαν τον έλεγχο των Rothman and Woodrooffe (1972) έτσι ώστε να χρησιμοποιείται για τον έλεγχο της συμμετρίας γύρω από άγνωστο σημείο.

Συγκεκριμένα, για τον έλεγχο της υπόθεσης (3.1.1.1) έναντι της εναλλακτικής που δίνεται στη σχέση (3.1.1.2), οι Mukherjee *et al.* (2013) προτείνουν την ακόλουθη σ.σ. τροποποιώντας τη σ.σ. των Rothman and Woodrooffe (1972) (βλέπε σχέση 2.3.2.3)

$$R_n = \sum_{i=1}^n a_i \left[F_n \left(2\tilde{\mu} - X_{(i)} \right) - \frac{2n-2i+1}{2n} \right]^2, \quad (3.1.2.1)$$

όπου τα βάρη a_i είναι τέτοια ώστε $a_1 > a_2 > \dots > a_n > 0$, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές του τ.δ. $X_1, X_2, \dots, X_n$, $\tilde{\mu}$ ο εκτιμητής της άγνωστης πληθυσμιακής διαμέσου και F_n η ε.α.σ.κ.. Οι Mukherjee *et al.* (2013) για τον παραπάνω έλεγχο προτείνουν να χρησιμοποιούνται ως βάρη τα $a_i = 1/i$, $i = 1, 2, \dots, n$.

Με παρόμοιο σκεπτικό με αυτό των Rothman and Woodrooffe (1972), είναι προφανές ότι με επίπεδο σημαντικότητας α και έχοντας ως εναλλακτική υπόθεση ότι η α.σ.κ. F είναι θετικά λοξή, η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται αν $R_n > c_2$, όπου η τιμή c_2 είναι τέτοια ώστε $P(R_n > c_2 | H_0) \leq \alpha$. Για τον προσδιορισμό της τιμής c_2 απαιτείται η εύρεση της κατανομής της σ.σ. D_n υπό την υπόθεση της συμμετρίας. Όμοια με τον τροποποιημένο έλεγχο του Butler (1969), οι Mukherjee *et al.* (2013) επισημαίνουν ότι για μικρά μεγέθη δείγματος δεν διατηρείται η επιθυμητή ιδιότητα να είναι η κατανομή της σ.σ. ανεξάρτητη της κατανομής $F(\cdot)$ και προτείνουν διορθωτικό μέτρο παρόμοιο με εκείνο της σχέσης (3.1.1.4) για την επιλογή της κρίσιμης τιμής. Με βάση εκτενείς μελέτες προσομοίωσης, συμπεραίνουν ότι για επίπεδο σημαντικότητας $\alpha = 5\%$, πρέπει να χρησιμοποιούνται οι κρίσιμες τιμές που προκύπτουν όταν η αληθινή συνάρτηση κατανομής είναι και πάλι η Βήτα(2,2). Για επίπεδο σημαντικότητας 5% και για διάφορες τιμές του n οι κρίσιμες αυτές τιμές παρατίθενται στον πίνακα που ακολουθεί.

Μέγεθος δείγματος	20	30	50	100	200
Κρίσιμη τιμή	0.110	0.087	0.062	0.034	0.018

3.1.3 Ο τροποποιημένος έλεγχος του McWilliams (1990)

Ο έλεγχος που προτάθηκε από τον McWilliams (1990), ο οποίος παρουσιάστηκε στην ενότητα 2.2.1, ήταν ουσιαστικά μια τροποποίηση του τεστ των ροών και αφορούσε τον έλεγχο της συμμετρίας γύρω από το μηδέν. Οι Mukherjee *et al.* (2013) τροποποίησαν τον έλεγχο του McWilliams (1990) ώστε να χρησιμοποιείται για τον έλεγχο της

υπόθεσης που δίνεται στη σχέση (3.1.1.1) έναντι της εναλλακτικής υπόθεσης που διατυπώνεται στη σχέση (3.1.1.2).

Ειδικότερα, έστω $X_1, \dots, X_n$ τ.δ. από κάποιον πληθυσμό με σ.π.π. $f(\cdot)$ και α.σ.κ. $F(\cdot)$ και $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές αυτού. Σε όσα ακολουθούν συμβολίζουμε με $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ τις τιμές αυτού του τ.δ., διατεταγμένες από τη μικρότερη στη μεγαλύτερη με βάση την απόλυτη τιμή της διαφοράς $X_{(i)} - \tilde{\mu}$, όπου $\tilde{\mu}$ ένας εκτιμητής της άγνωστης τιμής μ . Έστω ακόμη ότι $S_1, S_2, \dots, S_n$ είναι οι δείκτριες μεταβλητές που ορίζονται από τη σχέση:

$$S_i = \begin{cases} 1, & \text{αν } Y_{(i)} > 0 \\ 0, & \text{αλλιώς} \end{cases}, \quad i = 1, 2, \dots, n. \quad (3.1.3.1)$$

Η τροποποίηση της σ.σ. του McWilliams (1990) που προτάθηκε από τους Mukherjee *et al.* (2013), βασίζεται στον αριθμό των ροών του 1 στην ακολουθία $\{S_i\}$ και δίνεται σε αναλογία με τον έλεγχο του McWilliams (1990) από τη σχέση:

$$R^* = I_2 + I_3 + \dots + I_n, \quad (3.1.3.2)$$

όπου

$$I_k = \begin{cases} 1, & \text{αν } S_k = S_{k-1} = 1 \\ 0, & \text{αλλιώς.} \end{cases} \quad (3.1.3.3)$$

Είναι προφανές ότι με επίπεδο σημαντικότητας α και έχοντας ως εναλλακτική υπόθεση ότι η α.σ.κ. F είναι θετικά λοξή, η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται αν $R^* > c_3$, όπου η τιμή c_3 είναι τέτοια ώστε $P(R^* > c_3 | H_0) \leq \alpha$. Σε αναλογία με τις δύο προηγούμενες υποενότητες, οι Mukherjee *et al.* (2013) παρατήρησαν ότι η τιμή του κρίσιμου σημείου c_3 διαφοροποιείται ανάλογα με την υποκείμενη συνάρτηση κατανομής και πρότειναν παρόμοιο διορθωτικό

μέτρο με εκείνο της σχέσης (3.1.1.4). Με βάση εκτενείς μελέτες προσομοίωσης συμπεραίνουν ότι για επίπεδο σημαντικότητας $\alpha = 5\%$, πρέπει να χρησιμοποιούνται οι κρίσιμες τιμές που προκύπτουν όταν η αληθινή συνάρτηση κατανομής είναι και πάλι η Βήτα(2,2). Για επίπεδο σημαντικότητας 5% και για διάφορες τιμές του n οι κρίσιμες αυτές τιμές παρατίθενται στον πίνακα που ακολουθεί.

Μέγεθος δείγματος	20	30	50	100	200
Κρίσιμη τιμή	6	9	15	29	56

3.1.4 Τροποποιήσεις του προσημικού έλεγχου

Στην ενότητα αυτή θα παρουσιαστεί αρχικά η επέκταση του προσημικού ελέγχου που προτάθηκε από τον Gastwirth (1971) για τον έλεγχο της συμμετρίας γύρω από άγνωστο σημείο. Καθώς η επέκταση αυτή έχει το μειονέκτημα να βασίζεται σε σ.σ. που δεν είναι ανεξάρτητη κατανομής, τροποποιήθηκε πρόσφατα από τους Mukherjee *et al.* (2013).

Στην ενότητα 2.1.1 παρουσιάσθηκε ο έλεγχος της υπόθεσης της συμμετρίας γύρω από ένα γνωστό σημείο, έστω μ , με το προσημικό τεστ και ο οποίος στηρίζεται στον αριθμό των παρατηρήσεων που είναι μεγαλύτερες του μ . Ο Gastwirth (1971) τροποποίησε τον προσημικό έλεγχο με σκοπό να μπορεί να χρησιμοποιηθεί για τον έλεγχο της υπόθεσης ότι μια συνεχής α.σ.κ. F είναι συμμετρική γύρω από κάποιο άγνωστο κέντρο συμμετρίας, έστω μ .

Σε αναλογία με τον προσημικό έλεγχο, η σ.σ. T^* της τροποποίησης του Gastwirth (1971) στηρίζεται στον αριθμό των παρατηρήσεων που είναι μικρότερες από την εκτιμώμενη τιμή της άγνωστης ποσότητας μ . Ο Gastwirth (1971) πρότεινε την εκτίμηση

του μ από την δειγματική μέση τιμή $\bar{X}$ και κατά συνέπεια η σ.σ. στηρίζεται στο πρόσημο των $X_i - \bar{X}$, $i = 1, \dots, n$. Ορίζοντας τις δείκτριες τυχαίες μεταβλητές I_i^* ως:

$$I_i^* = \begin{cases} 1, & \text{αν } X_i < \bar{X} \\ 0, & \text{αλλιώς,} \end{cases} \quad (3.1.4.1)$$

εύκολα προκύπτει ότι:

$$T^* = \sum_{i=1}^n I_i^*. \quad (3.1.4.2)$$

Παρατήρηση 3.1.4.1 (Gastwirth (1971)) Έστω R η τ.μ. που παριστάνει τον αριθμό των παρατηρήσεων που είναι μεταξύ $\bar{X}$ και μ και έστω $R^* = R$, αν $\bar{X} > \mu$ και $R^* = -R$, αν $\bar{X} < \mu$. Τότε, η σ.σ. T_s της σχέσης (2.1.1.1) και η σ.σ. T^* διαφέρουν κατά R^* , δηλαδή $T^* = T_s + R^*$.

Χρησιμοποιώντας την προηγούμενη παρατήρηση ο Gastwirth (1971) κατάφερε να προσδιορίσει την ασυμπτωτική κατανομή της σ.σ. T^* και το αποτέλεσμα παρατίθεται στο θεώρημα που ακολουθεί.

Θεώρημα 3.1.4.1 (Gastwirth (1971)) Έστω $X_1, X_2, \dots, X_n$ τ.δ. από μια απόλυτα συνεχή α.σ.κ $F(\cdot)$ με μέση τιμή μ και διακύμανση σ^2 , της οποίας η σ.π.π. έστω $f(\cdot)$ είναι συνεχής στο μ . Τότε, η σ.σ. $n^{-1/2}(T^* - np)$ ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή 0 και διακύμανση

$$pq + \sigma^2 f^2(\mu) + 2f(\mu) \left[\int_{-\infty}^{\mu} x dF(x) - p\mu \right], \quad (3.1.4.3)$$

όπου $p = F(\mu)$ και $q = 1 - p$. Υπό την υπόθεση της συμμετρίας γύρω από το μ ισχύει ότι $p = 0.5$.

Παρατήρηση 3.1.4.2 (Gastwirth (1971)) Ο Gastwirth (1971) απέδειξε ότι για μικρά σε μέγεθος δείγματα, η αναμενόμενη τιμή της σ.σ. T^* δεν ισούται με np . Για παράδειγμα, για δείγματα μεγέθους δύο από οποιαδήποτε συνεχή σ.π.π. η αναμενόμενη τιμή του T^* αποδεικνύεται εύκολα ότι ισούται με 1.

Από το Θεώρημα 3.1.4.1 προκύπτει ότι η σ.σ. T^* ακολουθεί ασυμπτωτικά κανονική κατανομή ανεξάρτητα από την τιμή του p . Συνεπώς, η στατιστική συνάρτηση T^* είναι ασυμπτωτικά κανονική τόσο υπό την υπόθεση της συμμετρίας γύρω από το μ , όπου $p=0.5$, όσο και υπό τις αντίστοιχες εναλλακτικές υποθέσεις. Έτσι μπορεί να υπολογιστεί η ασυμπτωτική ισχύ του τεστ έναντι συγκεκριμένων ασύμμετρων εναλλακτικών. Τέλος, από μελέτες προσομοίωσης που έγιναν από τον Gastwirth (1971) προκύπτει ότι το επίπεδο του τροποποιημένου προσημικού τεστ εξαρτάται από την υποκείμενη σ.π.π. και πολλές φορές η ισχύς του ελέγχου δεν είναι ικανοποιητική. Έτσι, ως συμπέρασμα προκύπτει ότι η εκτίμηση του κέντρου συμμετρίας μ από τα δεδομένα έχει ως αποτέλεσμα το τροποποιημένο προσημικό τεστ να μην είναι ισχυρό και να μην είναι ανεξάρτητο κατανομής.

Αυτό το μειονέκτημα, ώθησε τους Mukherjee *et al.* (2013) να προτείνουν, πρόσφατα, μια τροποποίηση του τεστ του Gastwirth (1971), για τον έλεγχο της υπόθεσης της σχέσης (3.1.1.1) έναντι της εναλλακτικής υπόθεσης της ύπαρξης θετικής λοξότητας.

Ειδικότερα, για $i=1,2,\dots,n$, οι Mukherjee *et al.* (2013) προτείνουν τη σ.σ.

$$T = \sum_{i=1}^{[n/2]} I_i, \quad (3.1.4.4)$$

όπου

$$I_i = \begin{cases} 1, & \text{αν } X_{(n+1-i)} > 2\tilde{\mu} - X_{(i)} \\ 0, & \text{αν } X_{(n+1-i)} \leq 2\tilde{\mu} - X_{(i)}, \end{cases} \quad (3.1.4.5)$$

$\tilde{\mu}$ ο εκτιμητής του μ , $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές των $X_1, X_2, \dots, X_n$ και $[\cdot]$ το ακέραιο μέρος.

Με παρόμοιο σκεπτικό με αυτό του προσημικού ελέγχου, είναι προφανές ότι με επίπεδο σημαντικότητας α και έχοντας ως εναλλακτική υπόθεση ότι η α.σ.κ. F είναι θετικά λοξή, η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται αν $T > c_4$, όπου η τιμή c_4 είναι τέτοια ώστε $P(T > c_4 | H_0) \leq \alpha$. Όπως συνέβη στον έλεγχο του Gastwirth (1971), έτσι και στην τροποποίηση των Mukherjee *et al.* (2013) η σ.σ. δεν είναι ανεξάρτητη κατανομής όταν το κέντρο συμμετρίας είναι άγνωστο και εκτιμάται από τις παρατηρούμενες τιμές.

Σε αναλογία με τις προηγούμενες υποενότητες, οι Mukherjee *et al.* (2013) παρατήρησαν ότι η τιμή του κρίσιμου σημείου c_4 διαφοροποιείται ανάλογα με την υποκείμενη συνάρτηση κατανομής και πρότειναν παρόμοιο διορθωτικό μέτρο με εκείνο της σχέσης (3.1.1.4). Με εκτενείς μελέτες προσομοίωσης συμπεραίνουν ότι για επίπεδο σημαντικότητας $\alpha = 5\%$, πρέπει να χρησιμοποιούνται οι κρίσιμες τιμές που προκύπτουν όταν η αληθινή συνάρτηση κατανομής είναι και πάλι η Βήτα(2,2). Οι κρίσιμες αυτές τιμές για επίπεδο σημαντικότητας 5% και για διάφορα μεγέθη δείγματος παρατίθενται στον πίνακα που ακολουθεί.

Μέγεθος δείγματος	20	30	50	100	200
Κρίσιμη τιμή	9	14	24	49	98

3.1.5 Τροποποιήσεις του ελέγχου προσημασμένης διάταξης

Στην ενότητα αυτή θα παρουσιαστούν δύο τροποποιήσεις του ελέγχου προσημασμένης διάταξης του Wilcoxon που προτάθηκαν από τους Bhattacharya *et al.* (1982), με σκοπό να χρησιμοποιούνται για τον έλεγχο συμμετρίας γύρω από άγνωστο σημείο, έστω μ . Οι Bhattacharya *et al.* (1982) επικέντρωσαν τον ενδιαφέρον τους στον έλεγχο της υπόθεσης της συμμετρίας που διατυπώθηκε στη σχέση (3.1.1.1) έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας και ειδικότερα της υπόθεσης ότι η σ.π.π. είναι της μορφής

$$f_{\varepsilon} = (1-\varepsilon)f_0 + \varepsilon f_1, \text{ με } 0 < \varepsilon < 0.5, \quad (3.1.5.1)$$

όπου f_0 μία σ.π.π. συμμετρική γύρω από το σημείο μ και f_1 μια σ.π.π. με παράμετρο θέσης $\nu > \mu$. Επιπρόσθετα, οι Bhattacharya *et al.* (1982) θεώρησαν μια σταθερά p , με $0 < p < 1$ και $q = 1 - p$ και τον ακόλουθο συμβολισμό:

$$F(\mu) = \frac{1}{2}, F(\eta) = \frac{1}{2}q, F(\zeta) = 1 - \frac{1}{2}q.$$

Έστω τώρα $X_1, X_2, \dots, X_N$ τ.δ. μεγέθους $N = 2n + 1$ από έναν πληθυσμό με α.σ.κ. $F(\cdot)$ και $X_{(1)}, X_{(2)}, \dots, X_{(N)}$ οι διατεταγμένες τιμές του δείγματος. Σε όσα ακολουθούν, χωρίς βλάβη της γενικότητας και για αποφυγή περαιτέρω συμβολισμών οι Bhattacharya *et al.* (1982) υποθέτουν ότι το nq είναι ακέραιος. Τότε, εύκολα προκύπτει ότι οι δειγματικές τιμές των μ , η , και ζ είναι $X_{(n+1)}$, $X_{(nq+1)}$ και $X_{(N-nq)}$, αντίστοιχα.

Στο παραπάνω πλαίσιο, η πρώτη σ.σ. που προτάθηκε από τους Bhattacharya *et al.* (1982), βασίζεται στην ιδέα τους να συγκρίνουν τις

$\frac{1}{2}q$ το πλήθος άνω παρατηρήσεις με τις $\frac{1}{2}q$ το πλήθος κάτω παρατηρήσεις. Ειδικότερα, προτείνουν τη σ.σ.

$$T_N = (nq)^{-2} \sum_{i=1}^{nq} \sum_{j=1}^{nq} \chi(Y_{(i)} - Z_{(j)}), \quad (3.1.5.2)$$

όπου $Y_{(i)} = X_{(nq+1)} - X_{(nq+1-i)}$ και $Z_{(i)} = X_{(N-nq+i)} - X_{(N-nq)}$, για $i=1, \dots, nq$, ενώ

$$\chi(u) = \begin{cases} 1, & \text{αν } u < 0 \\ 0.5, & \text{αν } u = 0 \\ 0, & \text{αν } u > 0. \end{cases} \quad (3.1.5.3)$$

Η δεύτερη σ.σ. που προτάθηκε από τους Bhattacharya *et al.* (1982) προέκυψε από την ιδέα τους να χρησιμοποιήσουν μια σ.σ. της μορφής των σ.σ. που χρησιμοποιήθηκαν από τους Hajek and Sidak (1967) (βλέπε ενότητα 2.1.3). Ειδικότερα, προτείνουν τη σ.σ.

$$S_N = N^{-1/2} \sum I_i a_N(R_i^+), \quad (3.1.5.4)$$

όπου

$$I_i = \begin{cases} 1, & \text{αν } X_i \geq M_N = X_{(n+1)} \\ -1, & \text{αν } X_i < M_N = X_{(n+1)}, \end{cases} \quad (3.1.5.5)$$

R_i^+ η τάξη του $|X_i - X_{(n+1)}|$, $a_N(i) = \varphi\{i/(N+1)\}$ και $M_N = X_{(n+1)}$ η δειγματική διάμεσος.

Έχοντας ορίσει τις σ.σ. που προτάθηκαν από τους Bhattacharya *et al.* (1982) για τον έλεγχο της υπόθεσης (3.1.1.1) έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, είναι προφανές ότι η υπόθεση της συμμετρίας γύρω από το μ θα απορρίπτεται για μεγάλες τιμές των σ.σ.. Για τον καθορισμό της κρίσιμης περιοχής του ελέγχου, είναι απαραίτητη η εύρεση είτε της ακριβής κατανομής είτε της

ασυμπτωτικής κατανομής των T_N και S_N , υπό την υπόθεση της συμμετρίας γύρω από το σημείο μ . Χωρίς βλάβη της γενικότητας οι Bhattacharya *et al.* (1982) υποθέτουν ότι το σημείο αυτό είναι το μηδέν, καθώς οι σ.σ. T_N και S_N παραμένουν αναλλοίωτες σε μετασχηματισμούς της μορφής $aX+b$. Στο παραπάνω πλαίσιο προκύπτει το ακόλουθο θεώρημα.

Θεώρημα 3.1.5.1 (Bhattacharya *et al.* (1982)) Αν η α.σ.κ. F έχει σ.π.π. f η οποία είναι παραγωγίσιμη στα σημεία η και ζ , τότε υπό τη μηδενική υπόθεση της συμμετρίας ισχύει ότι:

$$\frac{(T_N - 0.5)\sqrt{N}}{\sigma_1} \xrightarrow[H_0]{d} N(0,1), \quad (3.1.5.6)$$

με

$$\sigma_1^2 = (3q)^{-1} + q^{-1} \left[1 - 4 \{qf(\zeta)\}^{-1} \int_{\zeta}^{\infty} f^2(x) dx \right]^2. \quad (3.1.5.7)$$

Επιπρόσθετα, αν η f είναι μια συμμετρική σ.π.π. παραγωγίσιμη στη διάμεσο της, η σ.σ. S_N ακολουθεί, υπό τη μηδενική υπόθεση της συμμετρίας, ασυμπτωτικά κανονική κατανομή με μέση τιμή 0 και διακύμανση

$$\frac{1}{3}q^3 - \frac{1}{4}q^4 + \left\{ 2f(0)^{-1} \int_{\zeta}^{\infty} f^2(x) dx - \frac{1}{2}q \right\}^2, \quad (3.1.5.8)$$

όπου η τιμή του ζ προσδιορίζεται από τη σχέση $\int_{\zeta}^{\infty} f(x) dx = \frac{1}{2}q$.

Είναι προφανές ότι η ασυμπτωτική διακύμανση των σ.σ. εξαρτάται από την αληθινή αλλά άγνωστη σ.π.π. f και επομένως η

ασυμπτωτική κατανομή των σ.σ. T_N και S_N δεν είναι ανεξάρτητη κατανομής. Οι Bhattacharya *et al.* (1982) υπολόγισαν την τιμή των ασυμπτωτικών διακυμάνσεων για ειδικές περιπτώσεις συμμετρικών σ.π.π. και ειδικές επιλογές της ποσότητας q . Τα αποτελέσματα αυτά παρατίθενται στον πίνακα που ακολουθεί.

Σ.Σ.	Q	Κανονική	Λογιστική	Διπλά Εκθετική	t_7	t_3	Cauchy
T_N	1.0	0.505	0.444	0.333	0.449	0.396	0.333
	0.5	0.754	0.691	0.667	0.693	0.667	0.816
	0.25	1.41	1.340	1.33	1.34	1.35	1.74
S_N	1.0	0.126	0.111	0.0833	0.112	0.099	0.0833
	0.5	0.0394	0.33	0.026	0.0334	0.0286	0.0274
	0.25	0.00600	0.00491	0.00423	0.00494	0.00434	0.00458

Σε αναλογία με τις προηγούμενες υποενότητες, μπορούμε να παρατηρήσουμε ότι η τιμή του ελέγχου διαφοροποιείται ανάλογα με την υποκείμενη συνάρτηση κατανομής, καθώς διαφοροποιείται η ασυμπτωτική διακύμανση. Οι Bhattacharya *et al.* (1982, σελ. 381) συνιστούν σε πρακτικές εφαρμογές την αντικατάσταση της άγνωστης διακύμανσης από την εκτιμώμενη που προκύπτει εκτιμώντας την σ.π.π. f και τα ολοκληρώματα των σχέσεων (3.1.5.7) και (3.1.5.8) με τους λεγόμενους uniform kernel εκτιμητές (βλέπε ενότητα 1.1).

Οι τροποποιήσεις που προτάθηκαν από τους Bhattacharya *et al.* (1982) παρότι αποτέλεσαν μια σημαντική θεωρητική συμβολή, με μελέτες προσομοίωσης, έχει διαπιστωθεί ότι είναι ισχυρότερα από τον έλεγχο που προκύπτει αν στο τεστ προσημασμένης διάταξης αντικατασταθεί η τιμή μ από τη δειγματική διάμεσο, αλλά δεν έχουν καλύτερη ισχύ από το κλασικό τεστ με το συντελεστή λοξότητας που

θα παρουσιαστεί στην ενότητα 3.2.1. Λόγω αυτής της συμπεριφοράς δεν παρατίθενται περισσότερες λεπτομέρειες.

Παρατήρηση 3.1.5.1 Μειονέκτημα των ελέγχων των σ.σ. T_N και S_N είναι ότι προϋποθέτουν δείγμα περιττού μεγέθους και όσο είμαστε σε θέση να γνωρίζουμε δεν υπάρχει τροποποίηση για άρτιου μεγέθους δείγματα.

3.1.6 Οι έλεγχοι των Antille and Kersting (1977) και Ekström and Jammalamadaka (2006)

Στην παράγραφο αυτή αρχικά θα παρουσιασθούν οι τρόποι ελέγχου της συμμετρίας γύρω από άγνωστο σημείο που μελετήθηκαν από τους Antille and Kersting (1977) και η επέκταση, τροποποίηση ενός εξ αυτών που προτάθηκε από τους Ekstrom and Jammalamadaka (2006).

Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από έναν πληθυσμό με α.σ.κ. F και σ.π.π. f . Υποθέτουμε ότι $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ είναι οι διατεταγμένες τιμές του τ.δ., $Y_i = X_{(i)} - X_{(i-1)}$, $i = 2, \dots, n$, ενώ $\bar{X}_n$ και M_n είναι η δειγματική μέση τιμή και διάμεσος, αντίστοιχα. Ακόμη, σε όσα ακολουθούν με ψ συμβολίζεται η F^{-1} . Οι σ.σ. που μελετήθηκαν από τους Antille and Kersting (1977) για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από το άγνωστο σημείο μ , είναι οι:

$$S_1 = n^{-1/2} \sum_{1 \leq i \leq n} \left(I(X_i - \bar{X}_n \leq 0) - \frac{1}{2} \right), \quad (3.1.6.1)$$

$$S_2 = n^{-3/2} \sum_{1 \leq i < j \leq n} \left(I(X_i + X_j \leq 2\bar{X}_n) - \frac{1}{2} \right), \quad (3.1.6.2)$$

$$S_3 = n^{-3/2} \sum_{1 \leq i < j \leq n} \left(I(X_i + X_j \leq 2M_n) - \frac{1}{2} \right), \quad (3.1.6.3)$$

και

$$S_{4,\varepsilon} = n^{-1/2} \sum_{[\varepsilon n] + 1 \leq i \leq [n/2]} \left(I(Y_i - Y_{n-i+2} \leq 0) - \frac{1}{2} \right), \quad (3.1.6.4)$$

όπου $0 < \varepsilon < 1/2$ και όπου $[\cdot]$ το ακέραιο μέρος.

Παρατήρηση 3.1.6.1 Έστω $X_{(1)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές των $X_1, \dots, X_n$ και $Y_i = X_{(i+1)} - X_{(i)}$, $i = 1, \dots, n-1$. Έστω $J(t)$ μια συνάρτηση η οποία παίρνει τιμές στο $\left(0, \frac{1}{2}\right)$. Μια εναλλακτική μορφή της σ.σ. $S_{4,\varepsilon}$ προσδιορίστηκε από τους Antille *et al.* (1982) και δίνεται από τη σχέση:

$$T_2(J) = n^{-1/2} \sum_{i=1}^{[(n-1)/2]} J\left(\frac{i}{n}\right) \left[I(Y_i - Y_{n-i} \leq 0) - \frac{1}{2} \right]. \quad (3.1.6.5)$$

Με μελέτη προσομοίωσης προτείνεται η συνάρτηση

$$J(t) = \begin{cases} 1, & \text{αν } \alpha \leq t \leq 1/2 \\ 0, & \text{αλλιώς,} \end{cases} \quad (3.1.6.6)$$

όπου $0 \leq \alpha \leq \frac{1}{2}$. Επιπλέον, οι Antille *et al.* (1982) απέδειξαν ότι η ασυμπτωτική κατανομή της σ.σ. $T_2(0.05)$ μπορεί να προσεγγιστεί ικανοποιητικά από την κανονική κατανομή $N(0, 0.1008)$. Στο πλαίσιο αυτό, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|T_2(0.05)| \geq 0.3174z_{\alpha/2}$. Για περισσότερη ακρίβεια προτείνεται η χρήση των πιθανοτήτων $P(T_2(0.05) \leq x) = \alpha \approx \bar{\alpha}$, όπου για περιττό μέγεθος

δείγματος n συνίσταται $\bar{\alpha} = \Phi\left(\frac{x+1/(2n^{1/2})}{0.3175}\right)$, ενώ για άρτιο μέγεθος

δείγματος n με εφαρμογή κάποιας επιπλέον διόρθωσης συνίσταται $\bar{\alpha} = \Phi\left(\frac{x+1/2(n-1)^{1/2}}{0.3175((n-1)/n)^{1/2}}\right)$.

Παρατηρούμε ότι οι σ.σ. που μελετήθηκαν από τους Antille and Kersting (1977) μπορούν να χωριστούν σε δύο κατηγορίες. Η πρώτη κατηγορία περιλαμβάνει τροποποιήσεις των σ.σ. του προσημικού ελέγχου και του ελέγχου Wilcoxon, όπου το άγνωστο κέντρο συμμετρίας εκτιμάται από τη μέση τιμή ή τη διάμεσο του τυχαίου δείγματος, ενώ η δεύτερη κατηγορία περιλαμβάνει ένα περικομμένο προσημικό έλεγχο. Πιο συγκεκριμένα, η σ.σ. S_1 είναι παραλλαγή της σ.σ. του προσημικού ελέγχου με το άγνωστο κέντρο συμμετρίας μ να εκτιμάται από τη δειγματική μέση τιμή $\bar{X}_n$, ενώ οι σ.σ. S_2 και S_3 είναι παραλλαγές της σ.σ. του ελέγχου Wilcoxon με αντικατάσταση του άγνωστου κέντρου συμμετρίας από τη δειγματική διάμεσο M_n και τη δειγματική μέση τιμή $\bar{X}_n$, αντίστοιχα. Τέλος, η σ.σ. $S_{4,\varepsilon}$ είναι μια περικομμένη σ.σ. προσημικού ελέγχου που βασίζεται στις διαφορές των συμμετρικά τοποθετημένων διαστημάτων των διατεταγμένων στατιστικών $Y_i - Y_{n-i+2}$, με το ποσοστό περικοπής να καθορίζεται από το $[\varepsilon n]$. Οι Antille and Kersting (1977) αναφέρουν ότι στη δεύτερη κατηγορία θα μπορούσαν να συμπεριληφθούν και περικομμένες σ.σ. τύπου Wilcoxon, ωστόσο αυτές οι σ.σ. δεν προτείνονται καθώς από μια αρχική μελέτη προέκυψε ότι δεν έχουν καλή συμπεριφορά.

Στα θεωρήματα που ακολουθούν παρατίθεται η ασυμπτωτική κατανομή των παραπάνω σ.σ. υπό κάποιες συνθήκες. Σε όσα

ακολουθούν $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$.

Θεώρημα 3.1.6.1 (Antille and Kersting (1977)) Έστω ότι η σ.π.π. f είναι συνεχής στο άγνωστο κέντρο συμμετρίας μ . Αν $f(\mu) \neq 0$ και $\int_{-\infty}^{\infty} x^2 f(x) dx < \infty$, τότε υπό την υπόθεση της συμμετρίας η σ.σ. S_1 ακολουθεί ασυμπτωτικά κανονική κατανομή $N(0, \sigma_1^2)$, όπου

$$\sigma_1^2 = \frac{1}{4} + f(\mu) [\sigma^2 f(\mu) - \mu_0], \quad (3.1.6.7)$$

$$\text{με } \mu_0 = 2 \int_{\mu}^{\infty} (x - \mu) f(x) dx.$$

Θεώρημα 3.1.6.2 (Antille and Kersting (1977)) Έστω ότι η σ.π.π. f είναι απόλυτα συνεχής με $f' \in L_2(-\infty, +\infty)$. Αν $\int_{-\infty}^{\infty} x^2 f(x) dx < \infty$, τότε υπό την υπόθεση της συμμετρίας η σ.σ. S_2 ακολουθεί ασυμπτωτικά κανονική κατανομή $N(0, \sigma_2^2)$, με

$$\begin{aligned} \sigma_2^2 = & \frac{1}{12} + \sigma^2 \left(\int_{-\infty}^{\infty} f^2(x) dx \right)^2 \\ & + 2 \left(\int_{-\infty}^{\infty} f^2(x) dx \right) \left(\int_{-\infty}^{\infty} x \left(\frac{1}{2} - F(x) \right) f(x) dx \right). \end{aligned} \quad (3.1.6.8)$$

Επιπλέον, αν $f(\mu) \neq 0$, η σ.σ. S_3 ακολουθεί ασυμπτωτικά κανονική κατανομή $N(0, \sigma_3^2)$, με

$$\sigma_3^2 = \frac{1}{48} + \frac{1}{4} \left(\frac{1}{f(\mu)} \int_{-\infty}^{\infty} f^2(x) dx - \frac{1}{2} \right)^2. \quad (3.1.6.9)$$

Θεώρημα 3.1.6.3 (Antille and Kersting (1977)) Υπό την προϋπόθεση ότι το πεδίο ορισμού της τ.μ. X είναι διάστημα και η σ.π.π. f είναι αυστηρώς θετική στο διάστημα αυτό και δύο φορές παραγωγίσιμη με συνεχείς παραγώγους, η σ.σ. $S_{4,\varepsilon}$ ακολουθεί υπό την υπόθεση της συμμετρίας ασυμπτωτικά κανονική κατανομή $N(0, \sigma_{4,\varepsilon}^2)$, όπου

$$\begin{aligned}\sigma_{4,\varepsilon}^2 &= \frac{1}{8} + \frac{1}{4} \int_{\varepsilon}^{1/2} (\ln(\psi'(1/2)) - \ln(\psi'(t))) dt \\ &\quad + \frac{1}{8} \int_{\varepsilon}^{1/2} (\ln(\psi'(1/2)) - \ln(\psi'(t)))^2 dt \\ &\quad + \frac{\varepsilon}{16} (\ln(\psi'(1/2)) - \ln(\psi'(\varepsilon)))^2 - \frac{\varepsilon}{4}.\end{aligned}\quad (3.1.6.10)$$

Επιπλέον, το όριο $\lim_{\varepsilon \rightarrow 0} \sigma_{4,\varepsilon}^2$ υπάρχει και είναι ίσο με

$$\sigma_{4,0}^2 = \frac{1}{16} + \frac{1}{16} \int_{-\infty}^{\infty} \left(1 + \frac{\ln f(t)}{f(\mu)} \right)^2 f(t) dt, \quad (3.1.6.11)$$

υπό την προϋπόθεση ότι το παραπάνω ολοκλήρωμα υπάρχει.

Θεώρημα 3.1.6.4 (Antille *et al.* (1982)) Αποδεικνύεται ότι υπό την υπόθεση της συμμετρίας, η σ.σ. $T_2(J)$ που αποτελεί μια εναλλακτική μορφή της σ.σ. $S_{4,\varepsilon}$, ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή 0 και διακύμανση

$$\sigma_2^2(J) = \frac{1}{8} \int_0^{1/2} J^2(t) dt + \frac{1}{8} \int_0^{1/2} \left[J(t) + \int_t^{1/2} J(s) \frac{\psi''(s)}{\psi'(s)} ds \right]^2 dt. \quad (3.1.6.12)$$

Για την συνάρτηση $J(t)$ της σχέσης (3.1.6.6) προκύπτει ότι:

$$\frac{1}{8} \left(\frac{1}{2} - \alpha \right) \leq \sigma_2^2(\alpha) < \infty. \quad (3.1.6.13)$$

Από τα παραπάνω θεωρήματα προκύπτει ότι οι ασυμπτωτικές διακυμάνσεις των σ.σ. S_i , $i=1,2,3$, και $S_{4,\varepsilon}$ εξαρτώνται από την υποκείμενη σ.π.π. και αυτό έχει ως συνέπεια να μην είναι ανεξάρτητες της κατανομής. Στον Πίνακα 1 της εργασίας των Antille and Kersting (1977) υπολογίστηκαν οι ασυμπτωτικές διακυμάνσεις σ_i^2 , $i=1,2,3,4$ έχοντας ως f διάφορες ειδικές περιπτώσεις συμμετρικών κατανομών και δίνονται στον Πίνακα που ακολουθεί.

Κατανομή	Κανονική	Λογιστική	Διπλά Εκθετική	Cauchy
σ_1^2	0.09085	0.10904	0.25000	-
σ_2^2	0.00376	0.00805	0.02083	-
σ_3^2	0.03155	0.02778	0.02083	0.02083
σ_4^2	0.10938	0.11621	0.12500	0.27744

Με βάση τον παραπάνω Πίνακα, χρησιμοποιώντας τον έλεγχο της σ.σ. S_1 , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $|S_1| \geq z_{\alpha/2} \max(\sqrt{\sigma_1^2})$ και επομένως για επίπεδο σημαντικότητας $\alpha=0.05$ απορρίπτεται αν $|S_1| \geq 0.98$. Ανάλογα, εφαρμόζοντας τον έλεγχο της σ.σ. S_2 , η υπόθεση της συμμετρίας απορρίπτεται αν $|S_2| \geq 0.2828$. Εφαρμόζοντας τον έλεγχο της σ.σ. S_3 , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $|S_3| \geq 0.3481$ και τέλος εφαρμόζοντας τον έλεγχο της σ.σ. $S_{4,\varepsilon}$, η υπόθεση της συμμετρίας απορρίπτεται αν $|S_{4,\varepsilon}| \geq 1.0323$.

Επιπλέον, εύκολα μπορεί να παρατηρηθεί ότι σε κάποιες περιπτώσεις η ασυμπτωτική διακύμανση δεν μεταβάλλεται αρκετά. Στην πράξη για να χρησιμοποιηθούν οι σ.σ. των Antille and Kersting (1977) χωρίς τον παραπάνω κανόνα είτε πρέπει

α) να υπολογιστούν οι κρίσιμες τιμές του ελέγχου για διάφορες κατανομές και να ακολουθηθεί παρόμοια μέθοδος με αυτή που προτάθηκε από τους Mukherjee *et al.* (2013), είτε

β) να αντικατασταθούν οι ασυμπτωτικές διακυμάνσεις από συνεπείς εκτιμητές τους.

Το δεύτερο βήμα, αλλά μόνο για την περίπτωση της σ.σ. $S_{4,\varepsilon}$ ήταν αυτό που ακολούθησαν οι Ekström and Jammalamadaka (2006). Ειδικότερα, οι Ekström and Jammalamadaka (2006) αποδεικνύουν ότι:

Θεώρημα 3.1.6.5 (Ekström and Jammalamadaka (2006)) Έστω $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές του τ.δ. $X_1, X_2, \dots, X_n$ και

$$D_{i,m} = X_{(i+m)} - X_{(i-m+1)}, \text{ για } i = m, \dots, n-m. \quad (3.1.6.14)$$

Έστω ότι $m/n \rightarrow 0$ και $m/\log n \rightarrow \infty$ καθώς $n \rightarrow \infty$. Τότε, οι εκτιμητές

$$\begin{aligned} \hat{\sigma}_{4,\varepsilon}^2 = & \frac{1}{16} + \frac{1}{16(n-2(m+\lceil \varepsilon n \rceil)+1)} \sum_{i=m+\lceil \varepsilon n \rceil}^{n-m-\lceil \varepsilon n \rceil} \left(1 - \log \frac{D_{i,m}}{D_{k,m}}\right)^2 + \\ & + \frac{\varepsilon}{16} \left(\log \frac{D_{\lceil \varepsilon n \rceil, m}}{D_{k,m}}\right)^2 - \frac{\varepsilon}{8}, \end{aligned} \quad (3.1.6.15)$$

και

$$\tilde{\sigma}_{\delta}^2 = \frac{1}{16} + \frac{1}{16(n-2(m+\lceil \delta n \rceil)+1)} \sum_{i=m+\lceil \delta n \rceil}^{n-m-\lceil \delta n \rceil} \left(1 - \log \frac{D_{i,m}}{D_{k,m}}\right)^2, \quad (3.1.6.16)$$

με δ ένα μικρό θετικό αριθμό, είναι συνεπείς εκτιμητές των $\sigma_{4,\varepsilon}^2$ και

$$\sigma_{\delta}^2 = \frac{1}{16} + \frac{1}{16} \int_{\psi(\delta)}^{\psi(1-\delta)} \left(1 + \log \frac{f(t)}{f(\mu)} \right)^2 dF(t), \quad (3.1.6.17)$$

δηλαδή $\hat{\sigma}_{4,\varepsilon}^2 \xrightarrow{a.s.} \sigma_{4,\varepsilon}^2$ και $\tilde{\sigma}_{\delta}^2 \xrightarrow{a.s.} \sigma_{\delta}^2$ καθώς $n \rightarrow \infty$, όπου με $\xrightarrow{a.s.}$ συμβολίζεται η σχεδόν παντού σύγκλιση.

Έχοντας ως στόχο τη δημιουργία μιας σ.σ. η οποία θα στηρίζεται στο $S_{4,\varepsilon}$ και θα είναι ανεξάρτητη κατανομής για τον ζητούμενο έλεγχο, οι Ekström and Jammalamadaka (2006) προτείνουν να χρησιμοποιούνται οι σ.σ.

$$T_{4,0} = \frac{S_{4,0}}{\hat{\sigma}_{4,0}}, \quad (3.1.6.18)$$

και

$$T_{4,\varepsilon} = \frac{S_{4,\varepsilon}}{\hat{\sigma}_{4,\varepsilon}}. \quad (3.1.6.19)$$

Στο Θεώρημα που ακολουθεί διατυπώνεται η ασυμπτωτική κανονικότητα της στατιστικής συνάρτησης $T_{4,\varepsilon}$.

Θεώρημα 3.1.6.6 (Ekström and Jammalamadaka (2006)) Έστω $X_1, X_2, \dots, X_n$ τ.δ. από την α.σ.κ. F , με την F να είναι συμμετρική. Έστω ακόμα ότι το πεδίο ορισμού της F είναι διάστημα και ότι η σ.π.π. f είναι αυστηρώς θετική και δύο φορές παραγωγίσιμη στο διάστημα αυτό με συνεχείς παραγώγους. Τότε, η σ.σ. $T_{4,\varepsilon}$ ακολουθεί ασυμπτωτικά τυπική κανονική κατανομή $N(0,1)$.

Άρα, με βάση τα παραπάνω, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας της F γύρω από το άγνωστο σημείο μ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|T_{4,\varepsilon}| > z_{\alpha/2}$.

Παρατήρηση 3.1.6.2 Με βάση τα αποτελέσματα μια εκτενούς μελέτης προσομοίωσης που έγινε από τους Ekström and Jammalamadaka (2006), προκύπτει ότι ο έλεγχος που προτάθηκε με χρήση της σ.σ. $T_{4,\varepsilon}$, διατηρεί το επίπεδο σημαντικότητας α με μεγάλη ακρίβεια ακόμη και για δείγματα μεγέθους $n = 20$. Τέλος, με βάση τα αποτελέσματα αυτής της προσομοίωσης, οι Ekström and Jammalamadaka (2006) εικάζουν ότι η ασυμπτωτική κανονικότητα ισχύει και για την σ.σ. $T_{4,0}$.

Παρατήρηση 3.1.6.3 Διεξάγοντας μια εκτενή μελέτη προσομοίωσης οι Ekström and Jammalamadaka (2006), συμπεραίνουν ότι η τιμή της παραμέτρου m πρέπει να είναι μικρή και να αυξάνεται καθώς αυξάνεται το μέγεθος του δείγματος n , έτσι ώστε να διατηρείται το επίπεδο σημαντικότητας. Ειδικότερα, για τιμές του n από 20 μέχρι 100 προτείνουν την τιμή $m = 3$, ενώ για $n = 500$ την τιμή $m = 4$ ή 5. Επιπλέον, εξάγουν το συμπέρασμα ότι ο έλεγχος που στηρίζεται στη σ.σ. $T_{4,0}$ είναι ανταγωνιστικός του ελέγχου του δειγματικού συντελεστή λοξότητας b_1 και μάλιστα πιο γενικός αφού μπορεί να εφαρμοστεί για κατανομές με βαριές ουρές.

3.2 Τεστ που στηρίζονται σε περιγραφικά μέτρα

Στην ενότητα αυτή θα παρουσιαστούν οι μέθοδοι που προτάθηκαν για τον έλεγχο της συμμετρίας μιας κατανομής γύρω από κάποιο άγνωστο σημείο και στηρίζονται σε περιγραφικά μέτρα ή/και χαρακτηρισμούς μέσω των δειγματικών τιμών. Η παρακίνηση για τη δημιουργία ελέγχων που βασίζονται σε περιγραφικά μέτρα προέκυψε από το γεγονός ότι έχοντας ένα περιγραφικό μέτρο (συνήθως ένα μέτρο λοξότητας), μπορεί να προκύψει ένας έλεγχος συμμετρίας, χρησιμοποιώντας τη δειγματική τιμή του μέτρου και βρίσκοντας την κατανομή του. Συγκεκριμένα, θα παρουσιαστεί ο έλεγχος που

στηρίζεται στο δειγματικό συντελεστή λοξότητας b_1 , ο έλεγχος των Hotelling and Solomons (1932), ο έλεγχος του Boos (1982), ο έλεγχος των Cabilio and Masaro (1996) με την τροποποίηση των Miao *et al.* (2006), ο έλεγχος της Mira (1999), ο έλεγχος του Ghosh (2011) και τέλος ο κανόνας των Mukherjee *et al.* (2013).

3.2.1 Δειγματικός συντελεστής λοξότητας

Στην ενότητα αυτή θα παρουσιαστεί η πιο γενική μέθοδος που έχει εμφανιστεί στην βιβλιογραφία για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από άγνωστο σημείο. Ο έλεγχος αυτός βασίζεται στο δειγματικό συντελεστή λοξότητας που δίνεται από τη σχέση:

$$b_1 = \frac{m_3}{m_2^{3/2}}, \quad (3.2.1.1)$$

όπου

$$m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k. \quad (3.2.1.2)$$

Ο έλεγχος αυτός χρησιμοποιείται για οποιαδήποτε υποκείμενη α.σ.κ. F , υπό την προϋπόθεση ότι οι μέχρι 6^{ης} τάξης ροπές της F είναι πεπερασμένες. Στο ακόλουθο θεώρημα προσδιορίζεται η ασυμπτωτική κατανομή της σ.σ. b_1 .

Θεώρημα 3.2.1.1 (Gupta (1967)) Έστω $X_1, X_2, \dots, X_n$ τ.δ. από τη μη-εκφυλισμένη α.σ.κ. $F(x - \mu)$. Έστω $\mu_k = E(X^k)$ οι k -τάξης πληθυσμιακές κεντρικές ροπές της F και έστω ότι η F έχει πεπερασμένες πληθυσμιακές ροπές μέχρι 6^{ης} τάξης. Τότε υπό την υπόθεση της συμμετρίας και για $n \rightarrow \infty$ ισχύει ότι:

$$n^{1/2} \frac{b_1}{\sigma_0(b_1)} \xrightarrow{d} N(0,1), \quad (3.2.1.3)$$

όπου

$$\sigma_0^2(b_1) = \frac{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3}{n\mu_2^3}. \quad (3.2.1.4)$$

Έχοντας προσδιορίσει την ασυμπτωτική κατανομή της σ.σ. b_1 υπό την H_0 , το επόμενο βήμα είναι ο προσδιορισμός της κρίσιμης περιοχής του ελέγχου. Συγκεκριμένα, η υπόθεση της συμμετρίας γύρω από το μ , απορρίπτεται έναντι της εναλλακτικής υπόθεσης ότι η α.σ.κ. F είναι θετικά λοξή, για μεγάλες τιμές της σ.σ. b_1 . Δηλαδή, η H_0 απορρίπτεται αν $b_1 > c_\alpha$, με το σημείο c_α να είναι τέτοιο ώστε $\alpha = P(b_1 > c_\alpha | H_0 : \text{αληθής})$. Επομένως, ο παραπάνω έλεγχος δεν είναι ασυμπτωτικά ανεξάρτητος κατανομής καθώς η ασυμπτωτική διακύμανση του ελέγχου $\sigma_0^2(b_1)$, υπό την H_0 , εξαρτάται από την υποκείμενη συνάρτηση κατανομής μέσω των κεντρικών ροπών. Ωστόσο, μπορεί να τροποποιηθεί ώστε να γίνει ασυμπτωτικά ανεξάρτητος κατανομής, εκτιμώντας το $n\sigma_0^2(b_1)$ από την συνάρτηση που προκύπτει αντικαθιστώντας τις κεντρικές ροπές με τις δειγματικές ροπές. Έτσι, ένας συνεπής εκτιμητής του $n\sigma_0^2(b_1)$ είναι:

$$n\hat{\sigma}_0^2(b_1) = \frac{m_6 - 6m_2m_4 + 9m_2^3}{m_2^3}, \quad (3.2.1.5)$$

όπου $m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$. Με βάση τα παραπάνω, ο έλεγχος γίνεται ασυμπτωτικά κανονικός και για $n \rightarrow \infty$ προκύπτει ότι:

$$Z_{b_1} = \frac{n^{1/2}b_1}{\hat{\sigma}_0(b_1)} \xrightarrow{d} N(0,1). \quad (3.2.1.6)$$

Συνεπώς, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της θετικής λοξότητας, αν $Z_{b_1} > z_\alpha$.

Παρατήρηση 3.2.1.1 (Gupta (1967)) Ο παραπάνω έλεγχος μπορεί να τροποποιηθεί ώστε να χρησιμοποιηθεί ως έλεγχος συμμετρίας γύρω από γνωστό σημείο, όταν το κέντρο συμμετρίας μ είναι γνωστό. Αρκεί να επαναπροσδιοριστούν τα m_k ώστε να είναι οι k -τάξης δειγματικές ροπές ως προς μ , δηλαδή $m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^k$ και στη συνέχεια να γίνει η ίδια διαδικασία όπως στην περίπτωση της σ.σ. b_1 .

Είναι γνωστό ότι αν η α.σ.κ. $F(x - \mu)$ είναι συμμετρική, τότε $\beta_1 = 0$ (καθώς $\mu_3 = 0$). Όμως, υπάρχουν ασύμμετρες κατανομές για τις οποίες $\mu_3 = 0$ και επομένως ο έλεγχος που στηρίζεται στη σ.σ. b_1 δεν είναι ικανός να ανιχνεύσει τέτοιους ασύμμετρους πληθυσμούς. Το γεγονός αυτό αποτελεί ένα μειονέκτημα του ελέγχου.

Παρατήρηση 3.2.1.2 Είναι προφανές ότι ο έλεγχος με το συντελεστή λοξότητας b_1 δεν μπορεί να εφαρμοστεί για κατανομές με βαριές ουρές για τις οποίες δεν υπάρχουν οι ροπές, π.χ. για την κατανομή Cauchy.

Παρατήρηση 3.2.1.3 Ο Gupta (1967) προτείνει μια εναλλακτική μορφή του ελέγχου του δειγματικού συντελεστή λοξότητας b_1 . Συγκεκριμένα, στο Θεώρημα 2.1 της εργασίας του (βλέπε Theorem 2.1, σελ. 850, Gupta (1967)), προτείνει μία εναλλακτική μορφή της τυπικής απόκλισης του b_1 , έστω τ και έπειτα έναν συνεπή εκτιμητή αυτής $\hat{\tau}$.

Στην περίπτωση αυτή, η σ.σ. του ελέγχου είναι η $Z_{b_1}^* = \frac{n^{1/2}b_1}{\hat{\tau}}$ και η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της θετικής λοξότητας, αν $Z_{b_1}^* > z_\alpha$.

3.2.2 Ο έλεγχος των Cabilio and Masaro (1996) και η τροποποίηση των Miao *et al.* (2006)

Στην παράγραφο αυτή θα παρουσιαστεί ο έλεγχος των Cabilio and Masaro (1996) για την υπόθεσης της συμμετρίας μιας συνάρτησης κατανομής γύρω από κάποιο άγνωστο σημείο, έστω θ και μια τροποποίηση του ελέγχου που δημιουργήθηκε από τους Miao *et al.* (2006).

Έστω $X_1, X_2, \dots, X_n$ τ.δ. από έναν συνεχή πληθυσμό με α.σ.κ. $F_X(x) = F(x - \theta)$ και σ.π.π. $f_X(x) = f(x - \theta)$, με $F_X(0) = 0.5$ και $F'_X(0) > 0$. Οι παραπάνω υποθέσεις γίνονται έτσι ώστε η διάμεσος θ να υπάρχει και να είναι μοναδική. Σε όσα ακολουθούν η μέση τιμή συμβολίζεται με μ και η διακύμανση με σ^2 , ενώ F_n είναι η ε.α.σ.κ.. Στο παραπάνω πλαίσιο, από τους Cabilio and Masaro (1996) δίνεται το ακόλουθο θεώρημα.

Θεώρημα 3.2.2.1 (Cabilio and Masaro (1996)) Έστω ότι η α.σ.κ. F_X ικανοποιεί τη σχέση $\int F_X(x)(1 - F_X(x))^{1/2} dx < \infty$ και έχει θετική παράγωγο στο θ . Επιπρόσθετα, έστω

$$\sigma^2(T, F) = \text{Var}[IC(X; T, F)], \quad (3.2.2.1)$$

όπου

$$IC(x; T, F) = - \int_{-\infty}^{\infty} \{I(x \leq y) - F_X(y)\} dy + \frac{I(x \leq \theta) - 1/2}{f_X(\theta)}, \quad (3.2.2.2)$$

με $I(\cdot)$ τη συνήθη δείκτρια συνάρτηση και έστω $0 < \sigma^2(T, F) < \infty$. Τότε για $n \rightarrow \infty$ όταν $\mu = \theta$ ισχύει ότι:

$$S_K = \frac{\sqrt{n}(\bar{X} - m)}{S} \xrightarrow{d} N(0, \sigma_0^2(F)), \quad (3.2.2.3)$$

όπου $\bar{X}$, m και S^2 η δειγματική μέση τιμή, διάμεσος και διακύμανση αντίστοιχα και

$$\sigma_0^2(F) = 1 + \frac{1}{e_{m,\bar{X}}(F)} - \frac{2}{\sqrt{e_{m,\bar{X}}(F)}} E \left| \frac{X - \mu}{\sigma} \right|, \quad (3.2.2.4)$$

με

$$e_{m,\bar{X}}(F) = 4\sigma^2 f^2(0). \quad (3.2.2.5)$$

Με σκοπό τη χρησιμοποίηση του Θεωρήματος 3.2.2.1 για την προς έλεγχο υπόθεση, οι Cabilio and Masaro (1996) υπολόγισαν για μια κλάση συμμετρικών κατανομών με πεπερασμένη πρώτη και δεύτερης τάξης ροπή, την ασυμπτωτική διακύμανση $\sigma_0^2(F)$ και τα ασυμπτωτικά 95 και 97.5 εκατοστιαία σημεία, που προσδιορίζονται από τις σχέσεις:

$$P_{95} = 1.644854\sqrt{\sigma_0^2(F)}, \quad (3.2.2.6)$$

και

$$P_{97.5} = 1.959964\sqrt{\sigma_0^2(F)}, \quad (3.2.2.7)$$

αντίστοιχα. Τα αποτελέσματα αυτά δίνονται στον Πίνακα 1 της εργασίας των Cabilio and Masaro (1996)) και παρατίθενται στο Πίνακα 5 του Παραρτήματος Α. Παρατηρώντας τα αποτελέσματα αυτού του πίνακα, οι Cabilio and Masaro (1996) εξαγουν το συμπέρασμα ότι για μια μεγάλη κλάση συμμετρικών κατανομών, τα ασυμπτωτικά εκατοστιαία σημεία P_{95} και $P_{97.5}$ έχουν τιμή η οποία διαφέρει ελάχιστα από τα αντίστοιχα P_{95} και $P_{97.5}$ της κανονικής κατανομής και προτείνουν να χρησιμοποιούνται γενικά οι κρίσιμες τιμές της κανονικής κατανομής, οι οποίες δίνονται στον Πίνακα που ακολουθεί. Επιπλέον, καθώς το αποτέλεσμα του Θεωρήματος 3.2.2.1

είναι ασυμπτωτικό, οι Cabilio and Masaro (1996) θέλησαν να εξετάσουν την ταχύτητα σύγκλιση της σ.σ. S_K στην κανονική κατανομή προσομοιώνοντας 10.000 δείγματα μεγέθους $n = 5(1)30$, από την κανονική κατανομή. Τα προσομοιωμένα ποσοστιαία σημεία μαζί με τα ασυμπτωτικά δίνονται στον πίνακα που ακολουθεί.

n	P_{90}	P_{95}	$P_{97.5}$	n	P_{90}	P_{95}	$P_{97.5}$
5	0.88	1.07	1.21	19	0.94	1.20	1.42
6	0.71	0.88	1.02	20	0.88	1.12	1.33
7	0.91	1.13	1.30	21	0.95	1.20	1.43
8	0.77	0.96	1.13	22	0.89	1.13	1.35
9	0.92	1.15	1.35	23	0.95	1.21	1.43
10	0.81	1.01	1.19	24	0.89	1.14	1.36
11	0.93	1.17	1.37	25	0.95	1.21	1.44
12	0.83	1.05	1.24	26	0.90	1.15	1.37
13	0.93	1.18	1.39	27	0.95	1.21	1.44
14	0.85	1.08	1.27	28	0.90	1.15	1.37
15	0.94	1.19	1.41	29	0.95	1.21	1.44
16	0.86	1.10	1.30	30	0.91	1.16	1.38
17	0.94	1.19	1.41	∞	0.97	1.24	1.48
18	0.87	1.11	1.32	-	-	-	-

Από τον παραπάνω πίνακα προκύπτει ότι τα εκατοστιαία σημεία διαφέρουν ανάλογα με το αν το μέγεθος του δείγματος n είναι άρτιος ή περιττός αριθμός και σε όλες τις περιπτώσεις έχουν τιμή μικρότερη από τα ασυμπτωτικά εκατοστιαία σημεία, με την τιμή τους

να αυξάνεται, καθώς αυξάνεται το μέγεθος του δείγματος n . Για n περιττό, παρατηρείται ότι η σύγκλιση στην οριακή τιμή γίνεται πιο γρήγορα.

Συνοψίζοντας, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το θ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|S_K| \geq P_{1-\alpha/2}$, με τα $P_{1-\alpha/2}$ να προκύπτουν είτε ασυμπτωτικά (βλέπε Πίνακα 5 Παραρτήματος Α) είτε από τον παραπάνω Πίνακα.

Όταν είναι επιθυμητό να μην θεωρείται η κανονική κατανομή για τον υπολογισμό των κρίσιμων τιμών του S_K , χρησιμοποιείται μια τροποποίηση του παραπάνω ελέγχου. Συγκεκριμένα, οι Cabilio and Masaro (1996) προτείνουν την αντικατάσταση της ασυμπτωτικής διακύμανσης $\sigma_0^2(T, F)$, που δόθηκε στη σχέση (3.2.2.4), από έναν συνεπή εκτιμητή της. Προκύπτει ότι ένας συνεπής εκτιμητής της $\sigma_0(T, F)$ είναι ο:

$$\hat{\sigma}_0(T, F) = \left(S^2 + \frac{1}{4f_n^2(m)} - \frac{d_{\bar{X}}}{f_n(m)} \right)^{1/2}, \quad (3.2.2.8)$$

όπου

$$d_{\bar{X}} = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|, \quad (3.2.2.9)$$

και η kernel εκτίμηση είναι

$$f_n(m) = \frac{1}{2nh_n} \max \left\{ 1, \sum_{i=1}^n I(m - h_n \leq X_i \leq m + h_n) \right\}, \quad (3.2.2.10)$$

με $h_n \rightarrow 0$ και $nh_n \rightarrow \infty$ καθώς $n \rightarrow \infty$.

Συνεπώς, στην περίπτωση αυτή, η υπόθεση της συμμετρίας απορρίπτεται έναντι της γενικής εναλλακτικής της ασυμμετρίας αν

$$\left| S_K^* = \frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}_0(T, F)} \right| \geq z_{\alpha/2}. \quad (3.2.2.11)$$

Η τροποποίηση των Miao *et al.* (2006)

Οι Miao *et al.* (2006) ασχολήθηκαν με τον έλεγχο των Cabilio and Masaro (1996) και πρότειναν μια τροποποίηση του. Ειδικότερα, πρότειναν την αντικατάσταση της τυπικής απόκλισης S της σ.σ. S_k της σχέσης (3.2.2.3) από έναν εκτιμητή της, ο οποίος είναι ο:

$$J = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^n |X_i - M|, \quad (3.2.2.12)$$

όπου M η δειγματική διάμεσος. Η σ.σ. του ελέγχου των Miao *et al.* (2006) είναι η

$$T_{Mi} = \frac{\bar{X} - M}{J}. \quad (3.2.2.13)$$

Στο επόμενο Θεώρημα προσδιορίζεται η ασυμπτωτική κατανομή της σ.σ. T_{Mi} .

Θεώρημα 3.2.2.2 (Miao *et al.* (2006)) Αν η α.σ.κ. F είναι συμμετρική γύρω από το μ , η σ.σ. $\sqrt{n}T_{Mi}$ ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή μηδέν και διακύμανση

$$\sigma_{T_{Mi}}^2 = \frac{2}{\pi\tau^2} \left[\sigma^2 + \frac{1}{4f^2(\mu)} - \frac{\tau}{f(\mu)} \right], \quad (3.2.2.14)$$

όπου σ^2 η πληθυσμιακή διακύμανση και $\tau = E(X) - 2\gamma$, με

$$\gamma = \int_{-\infty}^{\mu} xf(x)dx.$$

Επομένως, χρησιμοποιώντας τον έλεγχο των Miao *et al.* (2006), για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|T_{Mi}| \geq \frac{z_{\alpha/2} \sigma_{T_{Mi}}}{\sqrt{n}}$.

Γίνεται αντιληπτό ότι ασυμπτωτική διακύμανση της σ.σ. T_{Mi} εξαρτάται από την υποκείμενη συνάρτηση κατανομής η οποία συνήθως είναι άγνωστη. Έτσι, ο έλεγχος δεν είναι ασυμπτωτικά ανεξάρτητος κατανομής. Οι Miao *et al.* (2006) προτείνουν την εφαρμογή της μεθόδου kernel εκτίμησης για τη δημιουργία ενός ελέγχου ο οποίος θα είναι ασυμπτωτικά ανεξάρτητος κατανομής, όταν αυτό είναι απαραίτητο. Στον ακόλουθο Πίνακα δίνονται οι τιμές της ασυμπτωτικής διακύμανσης $\sigma_{T_{Mi}}^2$ για δείγματα από ορισμένες συμμετρικές κατανομές.

Κατανομή	Διακύμανση	Κατανομή	Διακύμανση
Κανονική	0.5708	t_3	0.9689
Λογιστική	0.5799	Ομοιόμορφη	0.8488
Διπλά Εκθετική	0.6366	Βήτα(2,2)	0.6539
t_5	0.6373	$0.9N(0,1) + 0.1N(0,9)$	0.7165

Παρακινούμενοι από τους Cabilio and Masaro (1996), προτείνουν να χρησιμοποιείται η κρίσιμη περιοχή $|T_{Mi}| \geq \frac{z_{\alpha/2} \sqrt{0.5708}}{\sqrt{n}}$, δηλαδή να υιοθετείται η ασυμπτωτική διακύμανση υπό την υπόθεση της κανονικότητας.

Τέλος, οι Miao *et al.* (2006) ολοκληρώνουν την εργασία τους διεξάγοντας μια εκτενή μελέτη προσομοίωσης. Αντικείμενο της μελέτης ήταν η σύγκριση του πραγματικού επιπέδου σημαντικότητας του ελέγχου T_{Mi} όταν εφαρμόζεται η προσεγγιστική τιμή της ασυμπτωτικής διακύμανσης και όταν εφαρμόζεται η πραγματική. Από την μελέτη αυτή προκύπτει ότι για $\alpha = 0.05$ και συμμετρικές κατανομές που βρίσκονται κοντά στην κανονική κατανομή, όπως η t_5 , η διπλά εκθετική κατανομή και η λογιστική κατανομή, διατηρείται το ονομαστικό επίπεδο σημαντικότητας $\alpha = 0.05$. Ωστόσο, για συμμετρικές κατανομές οι οποίες διαφέρουν αποτελεσματικά από την κανονική κατανομή, όπως η t_3 και η ομοιόμορφη κατανομή, οι οποίες είναι κατανομές με κοντές και βαριές ουρές, το πραγματικό επίπεδο σημαντικότητας ξεπερνά το ονομαστικό. Έτσι, σε τέτοιες περιπτώσεις ο παραπάνω έλεγχος δεν πρέπει να εφαρμόζεται.

3.2.3 Ο έλεγχος του Ghosh (2011)

Στην ενότητα αυτή θα παρουσιαστεί ο τρόπος ελέγχου της συμμετρίας γύρω από κάποιο άγνωστο σημείο μ που προτάθηκε από τον Ghosh (2011). Ας είναι $X_1, X_2, \dots, X_n$ ένα σύνολο παρατηρήσεων (όχι απαραίτητα ανεξάρτητων και ισόνομων), από μια συνεχή κατανομή με α.σ.κ. F και σ.π.π. f . Για $p \in (0,1)$ ο Ghosh (2011) αρχικά θεώρησε τη συνάρτηση:

$$g(p) = g(1-p) = F\left(\frac{F^{-1}(p) + F^{-1}(1-p)}{2}\right). \quad (3.2.3.1)$$

Έπειτα, παρατήρησε ότι $g(p) < 1/2$ αν η f είναι αρνητικά λοξή, $g(p) > 1/2$ αν η f είναι θετικά λοξή και $g(p) = 1/2$ αν η f είναι συμμετρική. Με βάση αυτήν την παρατήρηση εύκολα προκύπτει ότι αν

$$\tau = \tau(F) = \int_0^{1/2} g(p) dp, \quad (3.2.3.2)$$

τότε $\tau = 1/4$ όταν υπάρχει συμμετρία, $\tau < 1/4$ αν υπάρχει αρνητική λοξότητα, ενώ $\tau > 1/4$ αν υπάρχει θετική λοξότητα.

Επομένως, είναι λογικό ένας τρόπος ελέγχου της υπόθεσης της συμμετρίας να στηρίζεται σε έναν εκτιμητή του τ και στην απόκλιση του από το 0.25. Στο παραπάνω πλαίσιο, ο Ghosh (2011) πρότεινε τον εκτιμητή

$$\hat{\tau} = \int_0^{1/2} \hat{F} \left(\frac{\hat{F}^{-1}(p) + \hat{F}^{-1}(1-p)}{2} \right) dp, \quad (3.2.3.3)$$

όπου $\hat{F}$ είναι ένας εκτιμητής της άγνωστης α.σ.κ. F , τέτοιος ώστε

$$\sqrt{n}(\hat{F} - F) \xrightarrow{d} W_F, \quad (3.2.3.4)$$

με W_F να είναι μια σ.σ. κανονικών κατανομών με πίνακα διακυμάνσεων συνδιακυμάνσεων $K_F(\cdot, \cdot)$. Στο Πόρισμα που ακολουθεί προσδιορίζεται η ασυμπτωτική κατανομή του εμπειρικού εκτιμητή $\hat{\tau}$.

Πόρισμα 3.2.3.1 (Ghosh (2011)) Έστω $X_1, X_2, \dots, X_n$ τ.δ. από την F , η οποία είναι συμμετρική γύρω από το μ και έστω ότι η F έχει πεπερασμένη ροπή δεύτερης τάξης. Τότε ισχύει ότι:

$$\sqrt{n} \left(\hat{\tau} - \frac{1}{4} \right) \xrightarrow{d} N(0, \omega^2), \quad (3.2.3.5)$$

όπου

$$\omega^2 = \frac{1}{16} - \frac{1}{4} f(\mu) E(|X - \mu|) + \frac{1}{4} f^2(\mu) E(X - \mu)^2. \quad (3.2.3.6)$$

Είναι προφανές ότι η ασυμπτωτική διακύμανση εξαρτάται από την αληθινή και άγνωστη σ.π.π. f καθώς και από το σημείο μ . Επειδή

η σ.σ. $\hat{\tau}$ είναι αναλλοίωτη σε μετασχηματισμούς της μορφής $aX + b$, ο Ghosh (2011) υπολόγισε την ασυμπτωτική διακύμανση για ένα σύνολο συμμετρικών κατανομών γύρω από το μηδέν, με τα αποτελέσματα να δίνονται στον Πίνακα που ακολουθεί.

Κατανομή	ω^2
Ομοιόμορφη	0.0208
Κανονική	0.0227
Laplace	0.0625
$t_\nu, (\nu > 2)$	$\frac{1}{16} - \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right]^2 \frac{\nu-3}{4\pi(\nu-1)(\nu-2)}$
Βήτα(m, m)	$\frac{1}{16} - \frac{m+1}{16^m m(2m+1)} \left[\frac{\Gamma(2m)}{\Gamma^2(m)} \right]^2$
Λογιστική	0.0273

Στην πράξη, ο Ghosh (2011) προτείνει την αντικατάσταση της ασυμπτωτικής διακύμανσης ω^2 από έναν συνεπή εκτιμητή της. Στο πλαίσιο αυτό, παρακινούμενος από τους Cabilio and Masaro (1996), προτείνει την εκτίμηση του ω^2 από το $\hat{\omega}^2$ με

$$\hat{\omega}^2 = \frac{1}{16} - \frac{1}{4} \hat{f}(\hat{\mu}) \frac{1}{n} \sum_{i=1}^n |X_i - \hat{\mu}| + \frac{1}{4} \hat{f}^2(\hat{\mu}) S^2, \quad (3.2.3.7)$$

όπου S^2 και $\hat{\mu}$ η δειγματική διακύμανση και διάμεσος, αντίστοιχα, ενώ $\hat{f}(\hat{\mu})$ είναι ο kernel εκτιμητής της $f(\mu)$ (βλέπε ενότητα 1.1) που δίνεται από τη σχέση:

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad (3.2.3.8)$$

με

$$h_n = 1.06 \min(S, \text{IQR}/1.34) n^{-1/5}, \quad (3.2.3.9)$$

όπου IQR το ενδοτεταρτομοριακό εύρος, δηλαδή

$$\text{IQR} = q_3 - q_1, \quad (3.2.3.10)$$

με q_3 και q_1 το 75^ο και 25^ο εκατοστιαίο σημείο αντίστοιχα και

$$K(t) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}t^2\right), & \text{αν } |t| < \sqrt{5} \\ 0, & \text{αλλιώς,} \end{cases} \quad (3.2.3.11)$$

ο πυρήνας Epanechnikov, ο οποίος είναι ο πιο αποτελεσματικός πυρήνας σύμφωνα με τους Wald and Jones (1995).

Από τον Ghosh (2011) δίνεται μια εναλλακτική μορφή του $\hat{\tau}$, εύχρηστη σε υπολογισμούς με

$$\hat{\tau} = \begin{cases} \frac{1}{n^2} \left[\# \left(X \leq \frac{X_{(1)} + X_{(n)}}{2} \right) + \# \left(X \leq \frac{X_{(2)} + X_{(n-1)}}{2} \right) + \dots + \# \left(X \leq \frac{X_{(n/2)} + X_{(n/2)+1}}{2} \right) \right], & \text{αν } n \text{ άρτιος} \\ \frac{1}{n^2} \left[\# \left(X \leq \frac{X_{(1)} + X_{(n)}}{2} \right) + \# \left(X \leq \frac{X_{(2)} + X_{(n-1)}}{2} \right) + \dots + \# \left(X \leq X_{((n+1)/2)} \right) \right], & \text{αν } n \text{ περιττός,} \end{cases}$$

όπου με $\#(X \leq y)$ συμβολίζεται το πλήθος των παρατηρήσεων του τυχαίου δείγματος $X_1, X_2, \dots, X_n$ με τιμή μικρότερη ή ίση του y .

Επομένως, με βάση τα παραπάνω, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν

$$Z_G = \sqrt{n} \left| \frac{\hat{\tau} - \frac{1}{4}}{\hat{\omega}} \right| > z_{\alpha/2}. \quad (3.2.3.12)$$

Παρατήρηση 3.2.3.1 Από μελέτες προσομοίωσης που έγιναν από τον Ghosh (2011) προκύπτει ότι η ασυμπτωτικά ανεξάρτητης κατανομής μορφή του ελέγχου έχει μικρή ισχύ, ιδιαίτερα για μικρά δείγματα και πιστεύεται ότι αυτό συμβαίνει λόγω της αργής σύγκλισης στην οριακή κατανομή. Επιπλέον, από τον Πίνακα 5.12 της εργασίας του Ghosh (2011) προκύπτει ότι ο έλεγχος αυτός δεν διατηρεί το επίπεδο σημαντικότητας ακόμα και για μέγεθος δείγματος $n = 500$ και για αυτό το λόγο δεν μπορεί να προταθεί.

3.2.4 Ο έλεγχος των Hotelling and Solomons (1932)

Στην παράγραφο αυτή θα παρουσιαστεί ένας από τους αρχαιότερους ελέγχους που εμφανίστηκαν στη βιβλιογραφία για την υπόθεση της συμμετρίας μιας συνάρτησης κατανομής γύρω από κάποιο άγνωστο σημείο, ο έλεγχος των Hotelling and Solomons (1932). Οι Hotelling and Solomons (1932) αρχικά πρότειναν ένα μέτρο λοξότητας για κάποιο πληθυσμό και στη συνέχεια χρησιμοποιώντας τη δειγματική τιμή του μέτρου δημιούργησαν έναν έλεγχο συμμετρίας.

Έστω $X_1, X_2, \dots, X_n$ τ.δ. από ένα συνεχή πληθυσμό με α.σ.κ. $F(x - \theta)$ και σ.π.π. $f(x - \theta)$, όπου η διάμεσος θ είναι το άγνωστο κέντρο συμμετρίας. Η ιδέα των Hotelling and Solomons (1932) για τη δημιουργία της σ.σ. τους, προέκυψε από τον ορισμό της συμμετρίας ενός πληθυσμού. Συγκεκριμένα, όταν κάποιος πληθυσμός είναι συμμετρικός, τότε η διάμεσος, η κορυφή και η μέση τιμή ταυτίζονται. Επομένως, είναι προφανές ότι η διαφορά μεταξύ της διαμέσου

(median) και της μέσης τιμής (mean) ή η διαφορά μεταξύ της κορυφής (mode) και της μέσης τιμής μπορούν να ληφθούν ως μέτρα λοξότητας.

Έτσι, το πληθυσμιακό μέτρο λοξότητας που προτάθηκε από τους Hotelling and Solomons (1932) δίνεται από τη σχέση $T = \frac{\mu - \theta}{\sigma}$,

όπου μ η πληθυσμιακή μέση τιμή, θ η πληθυσμιακή διάμεσος και σ η τυπική απόκλιση. Οι Hotelling and Solomons (1932) απέδειξαν ότι η τιμή του μέτρου T κυμαίνεται από -1 μέχρι 1. Όταν για κάποιο πληθυσμό το παραπάνω μέτρο έχει τιμή μηδέν, τότε ο πληθυσμός είναι συμμετρικά κατανομημένος ως προς θ . Αντίθετα, όταν το παραπάνω μέτρο έχει θετική τιμή (ή αρνητική) τότε υπάρχει θετική λοξότητα (ή αρνητική λοξότητα). Η αντίστοιχη δειγματική τιμή του μέτρου T είναι

$$D = \frac{n(\bar{X} - m)}{S}, \quad (3.2.4.1)$$

όπου $\bar{X}$, m και S η δειγματική μέση τιμή, διάμεσος και τυπική απόκλιση, αντίστοιχα, με $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. Το επόμενο βήμα για τον

ακριβή προσδιορισμό της κρίσιμης περιοχής είναι η εύρεση της ασυμπτωτικής κατανομής της σ.σ. D , υπό την υπόθεση της συμμετρίας. Από τον Gastwirth (1971) προκύπτει ότι:

Πόρισμα 3.2.4.1 (Gastwirth (1971)) Αν η σ.π.π. f είναι συμμετρική γύρω από το μηδέν, τότε:

$$n^{-1/2} D \overset{\text{ασυμπτ.}}{\underset{H_0}{\sim}} N(0, \tau^2), \quad (3.2.4.2)$$

όπου

$$\tau^2 = 1 - \frac{4f(0)\mu_1 - 1}{4f^2(0)\sigma^2}, \quad (3.2.4.3)$$

με $\mu_1 = \int_{-\infty}^{\infty} |x| dF(x) = 2 \int_0^{\infty} x dF(x)$. Επιπλέον, προκύπτει ότι αν η σ.π.π. f είναι μονοκόρυφη, τότε $\frac{1}{4} \leq \tau^2 \leq 1$, με το πάνω όριο της ανισότητας να πετυχαίνεται όταν η f είναι η σ.π.π. της ομοιόμορφης κατανομής και το κάτω όριο να προσεγγίζεται αυθαίρετα όταν

$$f(x) = \begin{cases} 2, & \text{αν } |x| < \frac{a}{4} \\ 1-a, & \text{αν } \frac{a}{4} < |x| < \frac{a}{4} + \frac{1}{2} \\ 0, & \text{αλλιώς,} \end{cases} \quad (3.2.4.4)$$

και $a \rightarrow 0$.

Η παραπάνω σ.π.π. οφείλεται στον Hoeffding (1955) και είναι γνωστή στη βιβλιογραφία ως σ.π.π. του Hoeffding. Ο Gastwirth (1971) με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης επισήμανε ότι για $\tau^2 = 1$ και παρατηρήσεις από τη σ.π.π. Hoeffding, το πραγματικό επίπεδο σημαντικότητας είναι 0.0005, αρκετά μικρότερο του ονομαστικού επιπέδου σημαντικότητας 0.05.

Για να γίνει ο παραπάνω έλεγχος ασυμπτωτικά ανεξάρτητος κατανομής, αρκεί να βρεθεί ένας συνεπής εκτιμητής του τ^2 . Τότε από τη σχέση (3.2.4.3), με αντικατάσταση του τ^2 από τον αντίστοιχο συνεπή εκτιμητή, προκύπτει η ασυμπτωτικά ανεξάρτητης κατανομής μορφή του ελέγχου, με την κρίσιμη περιοχή και τα κρίσιμα σημεία να προσδιορίζονται με τη βοήθεια των ποσοστιαίων σημείων της τυπικής κανονικής κατανομής.

Τέλος, αξίζει να αναφερθεί ότι ο παραπάνω έλεγχος συμμετρίας δεν ενδείκνυται να εφαρμόζεται σε όλες τις περιπτώσεις. Είναι προφανές ότι υπάρχουν ασύμμετρες κατανομές για τις οποίες η

διάμεσος και η μέση τιμή ταυτίζονται. Έτσι, ο έλεγχος που προτάθηκε από τους Hotelling and Solomons (1932) δεν είναι ικανός να ανιχνεύσει τέτοιου είδους ασύμμετρους πληθυσμούς.

Παρατήρηση 3.2.4.1 Η σ.σ. του ελέγχου των Cabilio and Masaro (1996) συμπίπτει με την σ.σ. του ελέγχου των Hotelling and Solomons (1932).

3.2.5 Ο έλεγχος του Boos (1982) και οι έλεγχοι του Auda (2006)

Έστω $X_1, X_2, \dots, X_n$ τ.δ. από έναν πληθυσμό με συνεχή α.σ.κ. $F = F_0\left(\frac{x-\mu}{\sigma}\right)$. Στην ενότητα αυτή θα παρουσιαστεί ο τρόπος ελέγχου που προτάθηκε από τον Boos (1982) για τον έλεγχο της υπόθεσης της συμμετρίας της F_0 γύρω από το άγνωστο σημείο μ . Από τα αποτελέσματα που δόθηκαν στο πρώτο κεφάλαιο της διατριβής είναι γνωστό ότι η ε.α.σ.κ. F_n είναι συνεπής εκτιμητής της α.σ.κ. και έτσι δικαιολογείται να χρησιμοποιήσει κανείς ως ένα μέτρο της απόκλισης από την υπόθεση της συμμετρίας την ποσότητα (βλέπε Boos (1982))

$$d_n(\mu) = \int_{-\infty}^{+\infty} [F_n(\mu+x) + F_n(\mu-x) - 1]^2 dx, \quad (3.2.5.1)$$

ή την ισοδύναμη έκφραση της

$$d_n(\mu) = -\frac{2}{n^2} \sum_{i < j} |X_i - X_j| + \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left| \frac{X_i + X_j}{2} - \mu \right|. \quad (3.2.5.2)$$

Είναι φανερό ότι η ποσότητα αυτή εξαρτάται από το άγνωστο κέντρο συμμετρίας μ και επομένως σε πρακτικές εφαρμογές θα πρέπει να αντικατασταθεί από κατάλληλα ορισμένη ποσότητα. Στο πλαίσιο αυτό, ο Boos (1982) πρότεινε να αντικατασταθεί η άγνωστη

παράμετρος θέσης από τον λεγόμενο Hodges-Lehmann εκτιμητή ή από οποιοδήποτε άλλο εκτιμητή που είναι ασυμπτωτικά ισοδύναμος με αυτόν. Ο Hodges-Lehmann εκτιμητής είναι ένας ανθεκτικός (robust), μη παραμετρικός και εύκολα υπολογίσιμος εκτιμητής της παραμέτρου θέσης. Ειδικότερα, αν $X_1, \dots, X_n$ είναι οι διαθέσιμες παρατηρήσεις, για κάθε ζεύγος παρατηρήσεων της μορφής X_i, X_j , με $i, j = 1, \dots, n$, υπολογίζεται ο μέσος όρος τους και η διάμεσος αυτών των $n(n+1)/2$ το πλήθος τιμών ορίζεται να είναι ο Hodges-Lehmann εκτιμητής, δηλαδή

$$\hat{\mu} = \text{median}\{(X_i + X_j)/2, 1 \leq i, j \leq n\}. \quad (3.2.5.3)$$

Η σ.σ. που θεώρησε ο Boos (1982) προκύπτει ως ένα πολλαπλάσιο της σ.σ. $nd_n(\hat{\mu})$ διαιρεμένο με την

$$\hat{\sigma}_G = \binom{n}{2}^{-1} \sum_{i < j} |X_i - X_j|, \quad (3.2.5.4)$$

και δίνεται από τη σχέση:

$$T_{B_n} = n \left\{ -1 + \frac{\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n |X_i + X_j - 2\hat{\mu}|}{\sum_{i < j} |X_i - X_j|} \right\}. \quad (3.2.5.5)$$

Παρατήρηση 3.2.5.1 Η ποσότητα $\hat{\sigma}_G = \binom{n}{2}^{-1} \sum_{i < j} |X_i - X_j|$ είναι ένα μέτρο

της μεταβλητότητας που είναι γνωστό ως η Gini's Mean Difference (Gini, 1912). Η χρησιμοποίηση της στη σχέση (3.5.1.2) δίνει μια μορφή της σ.σ. που είναι εύχρηστη σε υπολογισμούς καθώς και έναν λόγο των απόλυτων διαφορών.

Καθώς αν οι τυχαίες μεταβλητές X_i και X_j είναι συμμετρικά κατανομημένες τότε $X_i + X_j - 2\mu \stackrel{d}{=} X_i + X_j$, είναι προφανές ότι η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται για μεγάλες τιμές της σ.σ. T_{B_n} . Το επόμενο βήμα για τον καθορισμό της κρίσιμης τιμής του ελέγχου είναι να προσδιοριστεί υπό την υπόθεση της συμμετρίας είτε η ακριβής είτε η ασυμπτωτική κατανομή της σ.σ. T_{B_n} .

Θεώρημα 3.2.5.1 (Boos (1982)) Έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες τ.μ. με α.σ.κ $F(x) = F_0((x - \mu)/\sigma)$. Υποθέτουμε ότι για κάθε $\varepsilon > 0$ υπάρχει η $(2 + \varepsilon)$ -τάξης ροπή της F_0 . Έστω ακόμα ότι F_0 έχει σ.π.π. f_0 η οποία είναι φραγμένη, συμμετρική γύρω από το μηδέν και συνεχής σχεδόν παντού. Τότε ισχύει ότι:

$$T_{B_n} \xrightarrow{d} (\sigma_G(F_0))^{-1} \frac{\int_0^1 [U(t) + U(1-t)]^2 dF_0^{-1}(t) - \left(\int_0^1 [U(t) + U(1-t)] dt \right)^2}{\int_0^1 f_0(F_0^{-1}(t)) dt},$$

όπου

$$\sigma_G(F_0) = \int |x_1 - x_2| dF_0(x_1) dF_0(x_2), \quad (3.2.5.6)$$

η πληθυσμιακή τιμή του μέτρου του Gini.

Από τη συζήτηση που προηγήθηκε προκύπτει ότι για επίπεδο σημαντικότητας α , απορρίπτεται η υπόθεσης της συμμετρίας γύρω από το μ έναντι της εναλλακτικής υπόθεσης της γενικής ασυμμετρίας, αν $T_{B_n} \geq c_\alpha$, όπου η τιμή c_α είναι τέτοια ώστε υπό την υπόθεση της συμμετρίας να ισχύει $P(T_{B_n} \geq c_\alpha) = \alpha$. Όμως, η τιμή αυτή δεν είναι ίδια για κάθε συμμετρική κατανομή καθώς από το Θεώρημα 3.2.5.1

παρατηρούμε ότι η ασυμπτωτική κατανομή της σ.σ. T_{B_n} εξαρτάται από την υποκείμενη α.σ.κ. F_0 . Για παράδειγμα, στην ειδική περίπτωση της Λογιστικής κατανομής προκύπτει το ακόλουθο Πόρισμα.

Πόρισμα 3.2.5.1 (Boos (1982)): Έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες τυχαίες μεταβλητές οι οποίες έχουν α.σ.κ. $F(x) = [1 + \exp(-(x - \mu)/\sigma)]^{-1}$. Τότε:

$$T_{B_n} \xrightarrow{d} \sum_{j=2}^{\infty} \frac{Z_{2j-1}^2}{j(2j-1)} = \sum_{i=1}^{\infty} \frac{Z_i^2}{(2i+1)(i+1)}, \quad (3.2.5.7)$$

όπου $Z_1, Z_2, \dots$, ανεξάρτητες τ.μ. που ακολουθούν τυπική κανονική κατανομή $N(0,1)$. Επομένως, η ασυμπτωτική κατανομή του T_{B_n} είναι ένα σταθμισμένο άθροισμα χ^2 -τυχαίων μεταβλητών.

Χρησιμοποιώντας αυτό το αποτέλεσμα ο Boos (1982) υπολόγισε τα ασυμπτωτικά ποσοστιαία σημεία όταν η αληθινή κατανομή είναι η λογιστική και αυτά δίνονται στον πίνακα που ακολουθεί.

$P(T_{B_n} \leq \text{τιμή πίνακα})$					
	$\alpha = 0.75$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$
$n \rightarrow \infty$	0.480	0.712	0.899	1.093	1.358

Με μελέτη προσομοίωσης ο Boos (1982) θέλησε να απαντήσει στα ακόλουθα ερωτήματα:

α) πόσο γρήγορα συγκλίνουν σε κάποια σταθερή τιμή τα άνω ποσοστιαία σημεία,

β) πόσο επηρεάζεται η τιμή των άνω ποσοστιαίων σημείων από την αληθινή συμμετρική κατανομή, και τέλος

γ) αν ο έλεγχος που προκύπτει είναι όντως ισχυρός.

Στον πίνακα που ακολουθεί δίνονται τα εκτιμώμενα άνω ποσοστιαία σημεία της σ.σ. T_{B_n} για την ειδική περίπτωση της δειγματοληψίας από την λογιστική κατανομή.

$P(T_{B_n} \leq \text{τιμή πίνακα})$					
	$\alpha = 0.75$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$
$n = 10$	0.490	0.710	0.880	1.070	1.350
$n = 20$	0.480	0.710	0.890	1.090	1.320
$n = 30$	0.480	0.710	0.900	1.090	1.360
$n \rightarrow \infty$	0.480	0.712	0.899	1.093	1.358

Ενδεικτικά, προκύπτει ότι για $n=10$ έχουμε $P(T_{B_n} \leq 0.88) = 0.95$, ενώ για $n=20$ έχουμε $P(T_{B_n} \leq 0.89) = 0.95$. Από τους δύο προηγούμενους πίνακες, προκύπτει ότι τα ασυμπτωτικά εκατοστιαία σημεία του T_{B_n} μπορούν να χρησιμοποιηθούν για όλα τα δείγματα μεγέθους $n \geq 10$ υπό τη λογιστική κατανομή. Έπειτα, ο Boos (1982) θέλοντας να απαντήσει στο δεύτερο ερώτημα, με επίπεδο σημαντικότητα 5% εφαρμόζει τον προτεινόμενο έλεγχο με κρίσιμη περιοχή $T_{B_n} \geq 0.899$ (δηλαδή με κρίσιμη περιοχή αυτή που προέκυψε με αληθή κατανομή τη λογιστική), σε 10.000 προσομοιωμένα δείγματα

μεγέθους $n = 20,50$, από διάφορες συμμετρικές κατανομές, πλην της λογιστικής κατανομής. Από τα αποτελέσματα του Πίνακα 2 του Boos (1982) μπορεί κάποιος να ισχυριστεί ότι δεν επηρεάζεται σημαντικά η τιμή των άνω ποσοστιαίων σημείων από την αληθινή συμμετρική κατανομή και επομένως η κρίσιμη περιοχή $T_{B_n} \geq 0.899$ είναι αυτή που προτείνεται.

Τέλος, με εκτενή μελέτη προσομοίωσης ο Boos (1982) κατέληξε στο συμπέρασμα ότι η σ.σ. T_{B_n} , για μια μεγάλη κλάση εναλλακτικών, δίνει καλά αποτελέσματα ανιχνεύοντας σε μεγάλο ποσοστό αποκλίσεις από την συμμετρία, με τα αποτελέσματα αυτά να δίνονται στον Πίνακα 3 του Boos (1982).

Δύο νέοι έλεγχοι συμμετρίας που στηρίζονται στο Gini's Mean Difference

Πρόσφατα στη βιβλιογραφία εμφανίστηκαν δύο νέοι έλεγχοι συμμετρίας που στηρίζονται στο Gini's Mean Difference και προτάθηκαν από τους Wang *et al.* (2013), ενώ περισσότερες λεπτομέρειες είναι διαθέσιμες στη διδακτορική διατριβή του Auda (2006). Η ιδέα του Auda (2006) για τη δημιουργία αυτών των ελέγχων, προέκυψε από τη σύγκριση των κατανομών των δειγμάτων που βρίσκονται στις δύο πλευρές της διαμέσου.

Έστω $X_1, X_2, \dots, X_n$ τ.δ. από μια άγνωστη α.σ.κ. F με άγνωστη πληθυσμιακή διάμεσο και έστω $X_{(1)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές αυτών. Ο Auda (2006) θεώρησε ως εκτιμητή της πληθυσμιακής διαμέσου τη δειγματική διάμεσο M και έπειτα για $j = 1, \dots, k$, όπου $k = \lfloor n/2 \rfloor$, όρισε τις ποσότητες $Y_j^+ = X_{(n+1-j)} - M$ και $Y_j^- = M - X_{(j)}$. Για

$\hat{\sigma}_G^+ = \binom{k}{2}^{-1} \sum_{i < j}^k |Y_i^+ - Y_j^+|$ και $\hat{\sigma}_G^- = \binom{k}{2}^{-1} \sum_{i < j}^k |Y_i^- - Y_j^-|$, προκύπτει η σ.σ. του πρώτου ελέγχου η οποία είναι η

$$S_{GMD} = \log \frac{\hat{\sigma}_G^+}{\hat{\sigma}_G^-}. \quad (3.2.5.8)$$

Στην ουσία η σ.σ. S_{GMD} συγκρίνει τις τιμές του Gini's Mean Difference για τις παρατηρήσεις που βρίσκονται πάνω και κάτω από τη δειγματική διάμεσο.

Η σ.σ. του δεύτερου ελέγχου προκύπτει από την παρατήρηση ότι μια σ.σ. ανάλογη με αυτή της σχέσης (3.2.5.8) μπορεί να προκύψει αν στη θέση των παρατηρήσεων χρησιμοποιηθούν οι τάξεις. Έστω $R^{(1)}(i)$ η τάξη του Y_{k-i+1}^+ ως προς τα $Y_1^-, \dots, Y_k^-, Y_1^+, \dots, Y_k^+$ και $R^{(2)}(i)$ η τάξη του Y_{k-i+1}^- ως προς τα $Y_1^-, \dots, Y_k^-, Y_1^+, \dots, Y_k^+$. Ο Auda (2006) όρισε τις ποσότητες

$$R_{GMD+} = \frac{2 \sum_{i=1}^k i R^{(1)}(i) - (k+1) \sum_{i=1}^k R^{(1)}(i)}{k(k-1)/2}, \quad (3.2.5.9)$$

και

$$R_{GMD-} = \frac{2 \sum_{i=1}^k i R^{(2)}(i) - (k+1) \sum_{i=1}^k R^{(2)}(i)}{k(k-1)/2}, \quad (3.2.5.10)$$

και η σ.σ. του δεύτερου ελέγχου του Auda (2006) είναι η

$$RS_{GMD} = \log \frac{R_{GMD+}}{R_{GMD-}}. \quad (3.2.5.11)$$

Ο Auda (2006) απέδειξε ότι οι σ.σ. S_{GMD} και RS_{GMD} ακολουθούν ασυμπτωτικά κανονική κατανομή, με τη μέση τιμή και τη διακύμανση να εξαρτώνται από την υποκείμενη συνάρτηση κατανομής από την

οποία προέρχεται το αρχικό συμμετρικό δείγμα. Επιπλέον, με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης συμπέρανε ότι η σύγκλιση στην κανονική κατανομή είναι αργή, ενώ προτείνει τον καθορισμό των κρίσιμων σημείων προσομοιώνοντας δεδομένα από την τυπική κανονική κατανομή.

Επιπρόσθετα, προτείνει να χρησιμοποιείται μια τυχαιοποίηση των ελέγχων των σ.σ. S_{GMD} και RS_{GMD} . Στον Πίνακα 7 του Παραρτήματος Α δίνονται οι κρίσιμες τιμές c_α του ελέγχου της σ.σ. $\exp(RS_{GMD})$, όπως αυτές υπολογίστηκαν από τον Auda (2006) και δόθηκαν στον Πίνακα 5.1 της εργασίας του, για μεγέθη δείγματος $n=8,9,...,19$, για διάφορες τιμές του επιπέδου σημαντικότητας και για όλους τους συνδυασμούς των n_1 και n_2 , όπου n_1 το πλήθος των παρατηρήσεων που είναι μεγαλύτερες ή ίσες της δειγματικής διαμέσου και n_2 το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της δειγματικής διαμέσου. Στο πλαίσιο αυτό, χρησιμοποιώντας τον έλεγχο της σ.σ. RS_{GMD} , για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $RS_{GMD} < \log(c_\alpha)$ ή $RS_{GMD} > \log\left(\frac{1}{c_\alpha}\right)$.

3.2.6 Ο έλεγχος της Mira (1999)

Στο εδάφιο αυτό θα παρουσιαστεί ο τρόπος ελέγχου της υπόθεσης συμμετρίας γύρω από άγνωστο σημείο που πρόσφατα προτάθηκε από την Mira (1999). Η ιδέα της Mira (1999) για τη δημιουργία αυτού του ελέγχου, προέκυψε από την παρατήρηση ότι έχοντας ένα μέτρο λοξότητας μπορεί να δημιουργηθεί ένας έλεγχος συμμετρίας, βρίσκοντας την κατανομή του μέτρου είτε για πεπερασμένου μεγέθους δείγματα είτε για απείρου μεγέθους δείγματα

(ασυμπτωτική κατανομή). Επειδή όμως ο προσδιορισμός της κατανομής τέτοιου είδους σ.σ. για δείγματα πεπερασμένου μεγέθους είναι αρκετά δύσκολος, η Mira (1999) επικέντρωσε το ενδιαφέρον της στα ασυμπτωτικά αποτελέσματα.

Η Mira (1999) στηρίχτηκε στο μέτρο λοξότητας που προτάθηκε από τον Bonferroni (1933). Συγκεκριμένα, έστω F μια α.σ.κ. με άγνωστη μέση τιμή μ_F και διάμεσο $\tilde{\mu}_F$. Ο Bonferroni (1933) εισήγαγε στη βιβλιογραφία την ποσότητα γ_p ως μέτρο λοξότητας της α.σ.κ. F με:

$$\gamma_p(F) = F^{-1}(p) + F^{-1}(1-p) - 2\tilde{\mu}_F, \quad \forall p \in \left(0, \frac{1}{2}\right). \quad (3.2.6.1)$$

Ειδικότερα, πρότεινε τη χρήση του πληθυσμιακού μέτρου

$$\gamma_1(F) = \int_0^1 \gamma_p(F) dp = 2 \int_0^1 F^{-1}(p) dp - 2\tilde{\mu}_F = 2(\mu_F - \tilde{\mu}_F). \quad (3.2.6.2)$$

Αν $X_1, X_2, \dots, X_n$ τ.δ. από την α.σ.κ. F και F_n η ε.α.σ.κ., τότε η δειγματική τιμή του μέτρου Bonferroni (1933) είναι $\gamma_1(F_n) = \int_0^1 \gamma_p(F_n) dp$. Η Mira (1999) προσδιόρισε την ασυμπτωτική κατανομή της σ.σ. $\gamma_1(F_n)$, με το αποτέλεσμα αυτό να δίνεται στο παρακάτω Θεώρημα.

Θεώρημα 3.2.6.1 (Mira (1999)) Έστω F μια α.σ.κ. με τυπική απόκλιση $\sigma_F < \infty$ και σ.π.π. f , τέτοια ώστε $f(\tilde{\mu}_F) > 0$. Τότε, υπό την H_0 ισχύει ότι:

$$n^{1/2} \gamma_1(F_n) \xrightarrow{d} N(0, \sigma^2(\gamma_1, F)), \quad (3.2.6.3)$$

όπου

$$\sigma^2(\gamma_1, F) = 4\sigma_F^2 + \left[\frac{1}{f(\tilde{\mu}_F)} \right]^2 - 4 \frac{1}{f(\tilde{\mu}_F)} S_{\tilde{\mu}_F}. \quad (3.2.6.4)$$

Στο παραπάνω πλαίσιο, μεγάλες τιμές της $|\gamma_1(F_n)|$ συνηγορούν στην απόρριψη της υπόθεσης της συμμετρίας γύρω από το $\tilde{\mu}_F$, έναντι της γενικής εναλλακτικής υπόθεσης της ασυμμετρίας.

Από το Θεώρημα 3.2.6.1 προκύπτει ότι η σ.σ. $\gamma_1(F_n)$ δεν είναι ασυμπτωτικά ανεξάρτητη κατανομής, αφού εξαρτάται από τη υποκείμενη συνάρτηση κατανομής μέσω της ασυμπτωτικής διακύμανσης. Η Mira (1999) έχοντας ως στόχο τη δημιουργία ενός ελέγχου συμμετρίας ο οποίος θα είναι ασυμπτωτικά ανεξάρτητος κατανομής, πρότεινε έναν συνεπή εκτιμητή της ασυμπτωτικής διακύμανσης $\sigma^2(\gamma_1, F)$ της σχέσης (3.2.6.4).

Στο πλαίσιο αυτό, έστω $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές του τυχαίου δείγματος $X_1, X_2, \dots, X_n$ και έστω $X_{(s)}$ η δειγματική διάμεσος η οποία δίνεται από τον τύπο

$$X_{(s)} = \begin{cases} X_{((n+1)/2)}, & \text{αν } n \text{ άρτιος} \\ \frac{1}{2} (X_{(n/2)} + X_{((n/2)+1)}), & \text{αν } n \text{ περιττός.} \end{cases} \quad (3.2.6.5)$$

Με βάση τον παραπάνω συμβολισμό, η δειγματική τιμή του μέτρου λοξότητας Bonferroni μπορεί να γραφεί ως $\gamma_1(F_n) = 2(\bar{X}_n - X_{(s)})$. Στο ακόλουθο θεώρημα προσδιορίζεται ο συνεπής εκτιμητής της $\sigma^2(\gamma_1, F)$ που προτάθηκε από την Mira (1999).

Θεώρημα 3.2.6.2 (Mira (1999)) Ένας συνεπής εκτιμητής της ασυμπτωτικής διακύμανσης $\sigma^2(\gamma_1, F)$ δίνεται από τη σχέση:

$$S_c^2(\gamma_1, F) = 4\hat{\sigma}^2 + \{D_{n,c}(\tilde{\mu}_F)\}^2 - 4D_{n,c}(\tilde{\mu}_F)\hat{S}_{\tilde{\mu}_F}, \quad (3.2.6.6)$$

όπου

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (3.2.6.7)$$

$$\hat{S}_{\tilde{\mu}_F} = \bar{X}_n - \frac{2}{n} \sum_{i=1}^n X_i I(X_i \leq X_{(s)}), \quad (3.2.6.8)$$

και

$$D_{n,c}(\tilde{\mu}_F) = \frac{n^{1/5}}{2c} \left(X_{\left(\left\lfloor \frac{n}{2} + cn^{4/5} \right\rfloor\right)} - X_{\left(\left\lfloor \frac{n}{2} - cn^{4/5} + 1 \right\rfloor\right)} \right), \quad (3.2.6.9)$$

με $[\cdot]$ να είναι το ακέραιο μέρος.

Από το Θεώρημα 3.2.6.2 προκύπτει ότι ο εκτιμητής $S_c^2(\gamma_1, F_n)$ εξαρτάται από την τιμή της παραμέτρου c . Η Mira (1999) πρότεινε η παράμετρος c να επιλέγεται ως η τιμή που ελαχιστοποιεί το ασυμπτωτικά αναμενόμενο μέσο τετραγωνικό σφάλμα του $D_{n, cn^{4/5}}(u_p)$, όπου u_p είναι το p -τάξης εκατοστιαίο σημείο. Έπειτα από λίγη άλγεβρα κατέληξε ότι για $p = 1/2$, η βέλτιστη επιλογή είναι $c = 0.5, 0.4$ και 0.58 για την κανονική, την Cauchy και τη λογιστική κατανομή, αντίστοιχα.

Στη συνέχεια, πραγματοποίησε μελέτες προσομοίωσης χρησιμοποιώντας διάφορες τιμές του c , ώστε να εξετάσει την επίδραση της μη-βέλτιστης επιλογής του c στην ακρίβεια της εκτίμησης. Ως αποτέλεσμα αυτής της μελέτης προτείνεται η εφαρμογή της τιμής $c = 0.5$, καθώς με αυτήν δεν υπάρχει σημαντική απώλεια, υπό την έννοια του μέσου τετραγωνικού σφάλματος, στην εκτίμηση της ασυμπτωτικής διακύμανσης, ανεξάρτητα από την κατανομή που χρησιμοποιείται.

Στο παραπάνω πλαίσιο, υπό τη μηδενική υπόθεση και με εφαρμογή του θεωρήματος 3.2.6.1 προκύπτει ότι:

$$Z_M = \frac{n^{1/2} \gamma_1(F_n)}{S_c(\gamma_1, F_n)} \xrightarrow{d} N(0,1). \quad (3.2.6.10)$$

Επομένως, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|Z_M| \geq z_{\alpha/2}$, για $n \rightarrow \infty$. Αν η εναλλακτική υπόθεση είναι μονόπλευρη, δηλαδή αν είναι η ύπαρξη θετικής ή αρνητικής λοξότητας, τότε η κρίσιμη περιοχή τροποποιείται ανάλογα.

Τέλος, με βάση τα αποτελέσματα μιας εκτενούς μελέτης προσομοίωσης, η Mira (1999) προτείνει την εφαρμογή της ασυμπτωτικά κανονικής κατανομής για δείγματα μεγέθους $n \geq 50$, ενώ για δείγματα με μικρότερο μέγεθος προτείνει την εκτίμηση των κρίσιμων τιμών διεξάγοντας μελέτη προσομοίωσης. Ενδεικτικά, παρατίθενται τα ποσοστιαία σημεία στον Πίνακα που ακολουθεί (βλέπε Table 6, σελ. 968, Mira (1999)).

	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.975$
$n = 20$	1.213	1.541	1.821
$n = 30$	1.092	1.392	1.645
$n = 40$	1.118	1.118	1.696

3.2.7 Ο κανόνας των Mukherjee *et al.* (2013)

Στην παράγραφο αυτή θα παρουσιασθεί ο κανόνας που προτάθηκε από τους Mukherjee *et al.* (2013) για τον έλεγχο της υπόθεσης της συμμετρίας έναντι της εναλλακτικής υπόθεσης της

θετικής λοξότητας. Στις παραγράφους 3.1.1-3.1.4 είδαμε τις τροποποιήσεις των ελέγχων του 2^{ου} Κεφαλαίου που προτάθηκαν από τους Mukherjee *et al.* (2013) για τον έλεγχο της ζητούμενης υπόθεσης. Οι Mukherjee *et al.* (2013) παρατήρησαν ότι ο καθορισμός των κρίσιμων σημείων τις περισσότερες φορές είναι δύσκολος και έτσι θεωρείται η ασυμπτωτική προσέγγιση. Επιπλέον, κατέληξαν ότι δεν υπάρχει μοναδικός βέλτιστος έλεγχος για την υπόθεση της συμμετρίας, έναντι της θετικής λοξότητας, ειδικά όταν το μέγεθος του δείγματος είναι μικρό και το κέντρο συμμετρίας άγνωστο. Έτσι πρότειναν τον παρακάτω κανόνα.

Έστω $X_1, X_2, \dots, X_n$ τυχαίο δείγμα από μια α.σ.κ. F . Ας είναι $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ τα διατεταγμένα στατιστικά και έστω $\tilde{\mu}$ η δειγματική διάμεσος. Οι Mukherjee *et al.* (2013) παρατήρησαν ότι αν η υπόθεση της συμμετρίας είναι αληθής, τότε τα $\tilde{\mu} - X_{(1)}$ και $X_{(n)} - \tilde{\mu}$ είναι σχεδόν ίσα. Με άλλα λόγια, υπό την H_0 , το $X_{(n)}$ θα βρίσκεται κοντά στο $2\tilde{\mu} - X_{(1)}$. Αντιθέτως, στην περίπτωση ύπαρξης θετικής λοξότητας, το $X_{(n)}$ μαζί με κάποιες ακόμα διατεταγμένες τιμές ανωτέρας τάξης θα υπερβαίνουν το $2\tilde{\mu} - X_{(1)}$.

Στο παραπάνω πλαίσιο, αν $x_{(j)}$ είναι η παρατηρούμενη τιμή του $X_{(j)}$, ο κανόνας των Mukherjee *et al.* (2013) για επίπεδο σημαντικότητας α , απορρίπτει την υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της θετικής λοξότητας, αν $x_{(t)} > 2\tilde{\mu} - x_{(1)}$, με το t να προσδιορίζεται από την σχέση

$$\max_{[(n+1)/2] \leq t \leq n} P(X_{(t)} > 2\tilde{\mu} - X_{(1)} | H_0 : \text{αληθής}) \leq \alpha, \quad (3.2.7.1)$$

όπου $[\cdot]$ το ακέραιο μέρος.

Έστω $\mathfrak{F}$ η κλάση των απόλυτα συνεχών α.σ.κ. F , με συνεχή πρώτη παράγωγο $dF(x) = f(x)dx$, όπου f η σ.π.π.. Επειδή τις περισσότερες φορές είναι δύσκολος ο προσδιορισμός του t μέσω της σχέσης (3.2.7.1), οι Mukherjee *et al.* (2013) πρότειναν μια προσέγγιση του t για μεγάλα μεγέθη δείγματος. Το αποτέλεσμα αυτό δίνεται στο Πόρισμα που ακολουθεί.

Πόρισμα 3.2.7.1 (Mukherjee *et al.* (2013)) Για οποιαδήποτε απόλυτα συνεχή α.σ.κ. F από την κλάση $\mathfrak{F}$ και για δείγμα μετρίου μεγέθους, υποθέτουμε ότι το στατιστικό $X_{(t)}$ εξασφαλίζει το ονομαστικό επίπεδο σημαντικότητας $\alpha \in [0,1)$. Τότε, το t μπορεί να προσεγγιστεί από το ακέραιο μέρος του $n\alpha^{1/n}$, δηλαδή $t \approx \left[n(\alpha^{1/n}) \right]$.

Με βάση τα παραπάνω και με περαιτέρω πράξεις, όταν $\alpha = 0.05$ και $n > 3$ οι Mukherjee *et al.* (2013) προτείνουν να επιλέγεται $t = n - 3$. Επιπλέον, διεξάγοντας μελέτες προσομοίωσης κατέληξαν στο συμπέρασμα ότι η προσέγγιση αυτή είναι αρκετά ακριβής για $n = 4, 5, \dots, 10^6$. Έτσι, για τα μεγέθη δείγματος που αναφέρθηκαν η υπόθεση της συμμετρίας απορρίπτεται αν $X_{(n-3)} > 2\tilde{\mu} - X_{(1)}$.

Παρόμοια, για επίπεδο σημαντικότητας $\alpha = 0.01$ και $n = 5, 6, \dots, 15$ προτείνεται να επιλέγεται $t = n - 4$, ενώ για $n > 15$ προτείνεται το $t = n - 5$.

Οι Mukherjee *et al.* (2013) αναφέρουν τον παραπάνω έλεγχο ως κανόνα, αφού δεν μπορεί να εφαρμοστεί για έλεγχο συμμετρίας έναντι της γενικής εναλλακτικής της ασυμμετρίας και εφαρμόζεται μόνο στην ειδική περίπτωση του ελέγχου της υπόθεσης της συμμετρίας έχοντας ως εναλλακτική υπόθεση τη θετική λοξότητα. Δύο πλεονεκτήματα του κανόνα είναι ότι είναι εύκολα εφαρμόσιμος και ότι για μικρά μεγέθη δείγματος οι κρίσιμες τιμές του ελέγχου καθορίζονται πλήρως.

Παρατήρηση 3.2.7.1 Με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης που έγινε από τους Mukherjee *et al.* (2013), προκύπτει ότι ο κανόνας διατηρεί το επίπεδο σημαντικότητας $\alpha = 0.05$ για μια μεγάλη κλάση συμμετρικών κατανομών. Επιπλέον, θεωρώντας μια κλάση από θετικά λοξές κατανομές ως εναλλακτικές και υπολογίζοντας την ισχύ του ελέγχου, προκύπτει ότι ο κανόνας έχει αρκετά καλή ισχύ ακόμα και για μικρά μεγέθη δείγματος όπως $n = 20$. Για δείγματα μεγέθους $n = 200$ και 500 η ισχύς παίρνει τιμή 1 σε όλες τις περιπτώσεις που θεωρήθηκαν και έτσι ο κανόνας είναι αποτελεσματικό εργαλείο για την ανίχνευση αποκλίσεων από τη συμμετρία που οφείλεται στην ύπαρξη θετικής λοξότητας. Ειδικότερα, στη μελέτη προσομοίωσης που διεξήγαγαν, προσομοιώνοντας 50.000 δείγματα μεγέθους $n = 20, 50$ και 100 από μια μεγάλη κλάση θετικά λοξών κατανομών, υπολόγισαν την ισχύ του κανόνα, των τροποποιήσεων των Ενοτήτων 3.1.1-3.1.4 και του Triple test που θα παρουσιασθεί στην Ενότητα 3.4.1. Από τη μελέτη αυτή οι Mukherjee *et al.* (2013) συμπεραίνουν ότι:

- Για μεγάλα μεγέθη δείγματος, όπως $n = 100$, ο κανόνας που πρότειναν έχει μεγαλύτερη ισχύ από τους υπόλοιπους ελέγχους, ενώ την μικρότερη ισχύ παρουσιάζουν οι έλεγχοι των σ.σ. R^* και T των σχέσεων (3.1.3.2) και (3.1.4.5), αντίστοιχα.
- Για $n = 20$, η ισχύς του κανόνα είναι μεγαλύτερη από αυτήν των υπολοίπων ελέγχων, εκτός από τον έλεγχο T_{R1} που παρουσιάζει μεγαλύτερη ισχύ σε δειγματοληψία από τις κατανομές Lognormal(0,2) και Weibull(0.5,1).
- Σε περίπτωση δειγματοληψίας από την κατανομή Lognormal(0,1), για $n \leq 20$ ο έλεγχος T_{R1} έχει τη μεγαλύτερη ισχύ, ενώ για $n > 20$ ο έλεγχος T_{R1} και ο κανόνας κυριαρχούν από άποψη ισχύος.

3.3 Τεστ που στηρίζονται στην εμπειρική χαρακτηριστική συνάρτηση

Στην ενότητα αυτή θα παρουσιαστούν τα κυριότερα τεστ που έχουν εμφανιστεί στη βιβλιογραφία, για τον έλεγχο της υπόθεσης της συμμετρίας μιας συνάρτησης κατανομής γύρω από άγνωστο σημείο και στηρίζονται στην εμπειρική χαρακτηριστική συνάρτηση. Συγκεκριμένα, θα παρουσιαστεί ο έλεγχος των Csörgö and Heathcote (1987), ο οποίος στηρίζεται στην εμπειρική χαρακτηριστική συνάρτηση μέσω της εμπειρικής χαρακτηριστικής συμμετρικής συνάρτησης και ο έλεγχος που προτάθηκε από τον Koutrounelis (1985).

3.3.1 Ο έλεγχος των Csörgö and Heathcote (1987)

Στην παράγραφο αυτή θα παρουσιαστεί ο τρόπος έλεγχου της συμμετρίας γύρω από ένα άγνωστο σημείο, έστω μ , που προτάθηκε από τους Csörgö and Heathcote (1987) και ο οποίος στηρίζεται στην χαρακτηριστική συμμετρική συνάρτηση (χ.σ.σ.). Για το λόγο αυτό στη συνέχεια δίνεται ο ορισμός της χ.σ.σ..

Ορισμός 3.3.1.1 Αν X είναι μία τ.μ. τότε η χαρακτηριστική της συνάρτηση ορίζεται από τη σχέση:

$$C(t) = E(e^{itX}) = U(t) + iV(t) = E(\cos tX) + iE(\sin tX), \quad (3.3.1.1)$$

ενώ η χαρακτηριστική συμμετρική συνάρτηση της τ.μ. X ορίζεται να είναι η:

$$\mu(t) = t^{-1} \arctan \left\{ \frac{V(t)}{U(t)} \right\}, \quad \text{για } 0 < t < r_0, \quad (3.3.1.2)$$

όπου r_0 , με $0 < r_0 \leq \infty$ η πρώτη θετική ρίζα του $U(t)$.

Η κεντρική ιδέα της μεθοδολογίας αυτής της ενότητας στηρίζεται στο μονοσήμαντο της χ.σ. και στον ακόλουθο χαρακτηρισμό.

Θεώρημα 3.3.1.1 Η χ.σ.σ. $\mu(t)$ είναι σταθερή αν και μόνο αν ικανοποιείται η υπόθεση της συμμετρίας.

Απόδειξη (Csörgö and Heathcote (1987))

Υπό την υπόθεση της συμμετρίας γύρω από το άγνωστο σημείο μ ισχύει ότι (βλέπε Csörgö and Heathcote (1987)):

$$U(t) = \cos(t\mu) E[\cos\{t(X - \mu)\}], \quad (3.3.1.3)$$

και

$$V(t) = \sin(t\mu) E[\sin\{t(X - \mu)\}]. \quad (3.3.1.4)$$

Είναι προφανές ότι υπό την υπόθεση της συμμετρίας γύρω από το άγνωστο σημείο μ προκύπτει ότι η χ.σ.σ. $\mu(t)$ είναι σταθερή.

Αντίστροφα, αν η συνάρτηση $\mu(t)$ είναι σταθερή για $0 < t < r_0$, τότε $E(\sin t(X - \mu)) = 0$ για $t \in (-r_0, r_0)$ και για $t \rightarrow 0$ είναι:

$$\mu(t) = \mu - \frac{t^2}{3!} E(X - \mu)^3 + \frac{t^4}{5!} [E(X - \mu)^5 - 10\{E(X - \mu)^3\}\{E(X - \mu)^2\}],$$

με

$$E(X) = \mu = \lim_{t \rightarrow 0} \mu(t).$$

Επομένως, προκύπτει ότι αν η χ.σ.σ. $\mu(t)$ είναι σταθερή τότε αυτό αποτελεί ένδειξη συμμετρίας και είναι λογική η αναζήτηση ενός ελέγχου που θα στηρίζεται στην εκτίμηση της $\mu(t)$ από τις διαθέσιμες

δειγματικές παρατηρήσεις και θα εξετάζει αν αυτή είναι σταθερή συνάρτηση.

Στο παραπάνω πλαίσιο, αν $X_1, X_2, \dots, X_n$ είναι οι διαθέσιμες δειγματικές παρατηρήσεις, ένας λογικός εκτιμητής της ποσότητας $\mu(t)$ είναι ο:

$$\mu_n(t) = t^{-1} \arctan \{V_n(t)/U_n(t)\}, \quad t \in L = [a, b], \quad (3.3.1.5)$$

όπου

$$U_n(t) = n^{-1} \sum_{j=1}^n \cos(tX_j), \quad (3.3.1.6)$$

και

$$V_n(t) = n^{-1} \sum_{j=1}^n \sin(tX_j), \quad (3.3.1.7)$$

είναι το πραγματικό και το φανταστικό μέρος, αντίστοιχα, της εμπειρικής χαρακτηριστικής συνάρτησης $C_n(t) = n^{-1} \sum_{j=1}^n \exp(itX_j)$, με το διάστημα L να επιλέγεται έτσι ώστε να ισχύει $0 < a < b < r_0$.

Παρατήρηση 3.3.1.1 Η χρήση της σ.σ. $\mu_n(t)$ εκτός από λογική, δικαιολογείται καθώς αποδεικνύεται ότι (βλέπε μεταξύ άλλων Csörgö and Heathcote (1987)) συγκλίνει ομοιόμορφα με πιθανότητα 1 στη $\mu(t)$, αν ικανοποιείται η συνθήκη:

$$E(\log|X|)^{1+\delta} < \infty, \quad \text{για } \delta > 0. \quad (3.3.1.8)$$

Επιπρόσθετα, υπό αυτή τη συνθήκη αποδεικνύεται ότι η σ.σ.

$$\tilde{T}_n(t) = n^{1/2} \{\mu_n(t) - \mu(t)\}, \quad t \in L, \quad (3.3.1.9)$$

που δεν είναι τίποτε άλλο παρά ένα μέτρο απόκλισης της $\mu_n(t)$ από την $\mu(t)$, συγκλίνει κατά κατανομή σε μια στοχαστική διαδικασία (σ.δ.), έστω T , κανονικών τυχαίων μεταβλητών με μέση τιμή μηδέν και πίνακα διακυμάνσεων συνδιακυμάνσεων με στοιχεία

$$\sigma(t,s) = E\{T(t)T(s)\} = \frac{h(t,s)}{2ts\{U^2(t)+V^2(t)\}\{U^2(s)+V^2(s)\}}, \quad (3.3.1.10)$$

όπου

$$\begin{aligned} h(t,s) = & U(t-s)\{U(t)U(s)+V(t)V(s)\} + U(t+s)\{V(t)V(s)-U(t)U(s)\} \\ & + V(t-s)\{V(t)U(s)-U(t)V(s)\} - V(t+s)\{U(t)V(s)+V(t)U(s)\}. \end{aligned}$$

Από τη συζήτηση που προηγήθηκε είναι φανερό ότι ένας τρόπος ελέγχου της υπόθεσης της συμμετρίας θα μπορούσε να στηρίζεται σε ένα μέτρο της απόκλισης της $\mu_n(t)$ από τη σταθερή συνάρτηση. Οι Csörgö and Heathcote (1987) προτείνουν ως μέτρο αυτής της απόκλισης την ποσότητα:

$$\xi_n = n^{1/2} \sup_{s,t \in L} |\mu_n(s) - \mu_n(t)|, \quad (3.3.1.11)$$

ή ισοδύναμα, την

$$\xi_n = \sup_{s,t \in L} |\tilde{T}_n(s) - \tilde{T}_n(t)|. \quad (3.3.1.12)$$

Υπό τον περιορισμό (3.3.1.8) οι Csörgö and Heathcote (1987) απέδειξαν ότι η σ.σ. ξ_n συγκλίνει κατά κατανομή στην τ.μ.

$$\xi = \sup_{s,t \in L} |T(s) - T(t)| = \left| \sup_{s \in L} T(s) - \inf_{t \in L} T(t) \right|. \quad (3.3.1.13)$$

Όμως, ακόμη και αν είναι γνωστή η κατανομή της τ.μ. X , δεν προκύπτει κάποια πληροφορία σχετικά με την κατανομή της τ.μ. ξ . Ωστόσο, η μορφή του ξ έδωσε την ιδέα στους Csörgö and Heathcote

(1987) να ορίσουν ένα άλλο μέτρο απόκλισης της $\mu_n(t)$ από κάποια σταθερά. Συγκεκριμένα, παρακινούμενοι από τους Fernique (1971) και Marcus and Shepp (1971), προτείνουν τη χρησιμοποίηση μιας σ.σ. η οποία θα στηρίζεται στην ποσότητα $|\mu_n(s_0) - \mu_n(t_0)|$, όπου s_0 και t_0 είναι τέτοια ώστε:

$$\sigma^2(s_0) = \max_{s \in L} \sigma^2(s), \quad \sigma^2(t_0) = \min_{t \in L} \sigma^2(t), \quad (3.3.1.14)$$

με (βλέπε σχέση (3.3.1.10))

$$\sigma^2(t) = \sigma(t, t). \quad (3.3.1.15)$$

Καθώς η συνάρτηση $\sigma^2(t)$ είναι συνεχής στο κλειστό διάστημα L , οι ποσότητες $\sigma^2(s_0)$ και $\sigma^2(t_0)$ είναι καλά ορισμένες και μοναδικές. Συγκεκριμένα, τα s_0 και t_0 είναι μοναδικά και τέτοια ώστε:

$$\sigma^2(s_0) > \sigma^2(t), \quad \sigma^2(t_0) < \sigma^2(t), \quad \text{για κάθε } t \in L. \quad (3.3.1.16)$$

Είναι προφανές ότι σε πρακτικές εφαρμογές τα σημεία s_0 και t_0 είναι άγνωστα και πρέπει να εκτιμηθούν. Στο πλαίσιο αυτό, έστω $\sigma_n(t, s)$ η εκτίμηση της συνδιακύμανσης $\sigma(t, s)$ της $T(t)$, που προκύπτει αντικαθιστώντας στην $\sigma(t, s)$ τις ποσότητες $U(t)$, $V(t)$ με τις αντίστοιχες εμπειρικές ποσότητες $U_n(t)$, $V_n(t)$ και έστω $\sigma_n^2(t) = \sigma_n(t, t)$. Τότε προκύπτει ότι:

$$s_n = \min \left\{ s : \sigma_n^2(s) = \sup_{t \in L} \sigma_n^2(t) \right\} \quad \text{και} \quad t_n = \max \left\{ t : \sigma_n^2(t) = \inf_{s \in L} \sigma_n^2(s) \right\}.$$

Η χρήση των s_n και t_n εκτός από λογική, δικαιολογείται καθώς αποδεικνύεται ότι $s_n \xrightarrow{a.s.} s$ και $t_n \xrightarrow{a.s.} t$ καθώς $n \rightarrow \infty$, όπου με $\xrightarrow{a.s.}$ συμβολίζεται η σχεδόν σίγουρη σύγκλιση.

Υπό την υπόθεση της συμμετρίας, αποδεικνύεται (βλέπε Csörgö and Heathcote (1982) και Csörgö and Heathcote (1983)) ότι η ποσότητα

$$n^{1/2} |\mu_n(s_n) - \mu_n(t_n)| = |\tilde{T}_n(s_n) - \tilde{T}_n(t_n)|, \quad (3.3.1.17)$$

είναι ανεξάρτητη από οποιοδήποτε εκτιμητή του κέντρου συμμετρίας μ και συγκλίνει κατά κατανομή στην τ.μ. $|T(s_0) - T(t_0)|$, δηλαδή στην απόλυτη τιμή μιας κανονικά κατανομημένης τ.μ. με μέση τιμή μηδέν και διακύμανση $\sigma_0^2(s_0) + \sigma_0^2(t_0) - 2\sigma(s_0, t_0)$.

Αυτό το αποτέλεσμα οδήγησε τους Csörgö and Heathcote (1987) στον ορισμό της ακόλουθης σ.σ. για το ζητούμενο έλεγχο:

$$T_n = n^{1/2} \frac{|\mu_n(s_n) - \mu_n(t_n)|}{\{\sigma_n^2(s_n) + \sigma_n^2(t_n) - 2\sigma_n(s_n, t_n)\}^{1/2}}. \quad (3.3.1.18)$$

Από τον ορισμό της σ.σ. T_n προκύπτει ότι μεγάλες τιμές αυτής συνηγορούν στην απόρριψη της υπόθεσης της συμμετρίας. Με βάση όσα προηγήθηκαν, στο ακόλουθο Θεώρημα δίνεται η ασυμπτωτική κατανομή της σ.σ. T_n υπό την H_0 , θεωρώντας ότι ικανοποιούνται οι σχέσεις (3.3.1.8) και (3.3.1.16).

Θεώρημα 3.3.1.2 (Csörgö and Heathcote (1987)) Αν η υποκείμενη α.σ.κ. είναι συμμετρική ως προς κάποιο κέντρο συμμετρίας, τότε:

$$\lim_{n \rightarrow \infty} P(T_n \leq x) = 2\Phi(x) - 1, \text{ για κάθε } x \geq 0, \quad (3.3.1.19)$$

όπου $\Phi(\cdot)$ η α.σ.κ. της τυπικής κανονικής κατανομής $N(0,1)$.

Με βάση το παραπάνω θεώρημα και για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $T_n > x_\alpha$, με το σημείο x_α να προσδιορίζεται από τη σχέση $2\Phi(x_\alpha) - 1 = 1 - \alpha$.

Μελέτες προσομοίωσης που έγιναν από τους Csörgö and Heathcote (1987) έδειξαν ότι ο προτεινόμενος έλεγχος είναι αποδεκτός για δείγματα μεγέθους $n \geq 40$. Ακόμη, οι Csörgö and Heathcote (1987) απέδειξαν ότι ο έλεγχος είναι συνεπής έναντι οποιασδήποτε λοξής εναλλακτικής για την οποία ικανοποιείται η σχέση $\mu(s_0) \neq \mu(t_0)$.

Παρατήρηση 3.3.1.2 Σε εφαρμογές μπορεί να χρησιμοποιηθεί οποιοδήποτε διάστημα $L = [a, b]$, $0 < a < b < r_n$, όπου το a παίρνει τιμή κοντά στο μηδέν και το r_n είναι η πρώτη θετική ρίζα του U_n . Τέλος, οι Csörgö and Heathcote (1987) προτείνουν μαζί με τον έλεγχο να χρησιμοποιείται και ένα γράφημα των $\sigma_n^2(t)$. Κατά αυτόν τον τρόπο βρίσκονται ταυτόχρονα οι τιμές $\sigma_n^2(s_n)$, $\sigma_n^2(t_n)$ και τα σημεία s_n και t_n , με αποτέλεσμα οι ποσότητες $\sigma_n(s_n, t_n)$, $\mu_n(s_n)$ και $\mu_n(t_n)$ να μπορούν να υπολογιστούν άμεσα.

3.3.2 Ο έλεγχος του Koutrouvelis (1985)

Στην παράγραφο αυτή θα παρουσιασθεί ο έλεγχος που προτάθηκε από τον Koutrouvelis (1985). Πρόκειται για μια μέθοδο που στηρίζεται στην εμπειρική χαρακτηριστική συνάρτηση για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από κάποιο άγνωστο σημείο και η σ.σ. του ελέγχου είναι μια τετραγωνική μορφή.

Έστω $X_1, X_2, \dots, X_n$ τ.δ. από μια α.σ.κ. $F(\cdot)$, με χαρακτηριστική συνάρτηση $\varphi(u) = C(u) + iS(u)$ και έστω η εμπειρική χαρακτηριστική συνάρτηση $\varphi_n(u) = \frac{1}{n} \sum_{j=1}^n \exp(iuX_j)$. Αρχικά, ο Koutrounelis (1985) επικέντρωσε το ενδιαφέρον του στην εύρεση ενός εκτιμητή της άγνωστης διαμέσου. Για το σκοπό αυτό θεώρησε $F_0(y)$ μια συνάρτηση κατανομής συμμετρική στο όρισμα της, με τη χαρακτηριστική της συνάρτηση να είναι η πραγματική και άρτια συνάρτηση $\varphi_0(u) = \int_{-\infty}^{+\infty} \cos(uy) dF_0(y)$. Στο πλαίσιο αυτό, αν $F(x) = F_0(x + \delta)$, δηλαδή αν η α.σ.κ. F είναι συμμετρική γύρω από το δ , τότε το πραγματικό και το φανταστικό μέρος της $\varphi(u)$ είναι:

$$C(u) = \cos(\delta u) \varphi_0(u) \text{ και } S(u) = \sin(\delta u) \varphi_0(u),$$

αντίστοιχα. Τότε, για κάθε πραγματικό αριθμό u τέτοιο ώστε $C(u) \neq 0$, υπάρχει ακέραιος $k(u)$ ο οποίος ικανοποιεί τη σχέση:

$$\arctan \left\{ \frac{S(u)}{C(u)} \right\} + \pi k(u) = \delta u. \quad (3.3.2.1)$$

Με βάση τη σχέση (3.3.2.1), ο Koutrounelis (1985) πρότεινε την εκτίμηση της άγνωστης διαμέσου δ από ένα γραμμικό μοντέλο όπου η εξαρτημένη μεταβλητή παίρνει τιμές $z_{nl} + \pi k_l$, $l = 1, \dots, r$, με

$$z_{nl} = \arctan \left\{ \frac{S_n(t_l)}{C_n(t_l)} \right\}, \quad k_l = k(t_l) \text{ και } t_1, \dots, t_r \text{ πραγματικούς αριθμούς.}$$

Στη συνέχεια, ο Koutrounelis (1985) όρισε το διάνυσμα $\underline{t}_n' = (t_{1n}, \dots, t_{rn})$ με $t_{kn} = k\tau / c_n$, όπου τ είναι μια κατάλληλα επιλεγμένη απόσταση μεταξύ των t -τιμών και c_n είναι το ημί-διατεταρτομοριακό

εύρος. Ο έλεγχος του Koutrouvelis (1985) στηρίζεται στην τετραγωνική μορφή:

$$Q_n(\underline{t}_n) = 2n(\underline{h}_n(\underline{t}_n) - \tilde{d}_n(\underline{t}_n)\underline{t}_n)' D_n^{-1}(\underline{t}_n)(\underline{h}_n(\underline{t}_n) - \tilde{d}_n(\underline{t}_n)\underline{t}_n), \quad (3.3.2.2)$$

με $\tilde{d}_n(\underline{t}_n) = (\underline{t}_n' D_n^{-1}(\underline{t}_n) \underline{t}_n)^{-1} \underline{t}_n' D_n^{-1}(\underline{t}_n) \underline{h}_n(\underline{t}_n)$ και $\underline{h}_n(\underline{t}_n) = (h_{n1}, \dots, h_{nr})$, όπου τα h_{nl} προκύπτουν με τον ίδιο τρόπο που βρέθηκαν τα z_{nl} αλλά χρησιμοποιώντας το τ.δ. $X'_1, \dots, X'_n$, με $X'_i = X_i - \delta_{0n} - c\delta_{cn}$, c μια γνωστή σταθερά, δ_{cn} κατάλληλη σταθερά και δ_{0n} να είναι η 25% περικομμένη μέση τιμή του αρχικού δείγματος. Ο πίνακας D_n είναι ένας $r \times r$ πίνακας με στοιχεία:

$$\begin{aligned} d_{kl} = & \left| \varphi(t_k) \right|^{-2} \left| \varphi(t_l) \right|^{-2} \left[-C'_{n0}(t_k + t_l) \{ C'_{n0}(t_k) C'_{n0}(t_l) - S'_{n0}(t_k) S'_{n0}(t_l) \} \right. \\ & + C'_{n0}(t_k - t_l) \{ C'_{n0}(t_k) C'_{n0}(t_l) + S'_{n0}(t_k) S'_{n0}(t_l) \} \\ & - S'_{n0}(t_k + t_l) \{ S'_{n0}(t_k) S'_{n0}(t_l) + C'_{n0}(t_k) C'_{n0}(t_l) \} \\ & \left. + S'_{n0}(t_k - t_l) \{ S'_{n0}(t_k) C'_{n0}(t_l) - C'_{n0}(t_k) S'_{n0}(t_l) \} \right], \end{aligned}$$

$$\text{όπου } C'_{n0}(u) = \frac{1}{n} \sum_{j=1}^n \cos(uX'_j) \text{ και } S'_{n0}(u) = \frac{1}{n} \sum_{j=1}^n \sin(uX'_j).$$

Οι τετραγωνικές μορφές $Q_n(\underline{t}_n)$ ορίζουν μια οικογένεια σ.σ. τα μέλη της οποίας καθορίζονται από τις επιλογές των σταθερών τ και r . Στο παρακάτω θεώρημα δίνεται η ασυμπτωτική κατανομή της σ.σ. $Q_n(\underline{t}_n)$.

Θεώρημα 3.3.2.1 (Koutrouvelis (1985)) Αν $E|X|^2 < \infty$ και $r \geq 2$, τότε υπό την υπόθεση της συμμετρίας γύρω από το δ , η τετραγωνική μορφή $Q_n(\underline{t}_n)$ ακολουθεί ασυμπτωτικά την χ -τετράγωνο κατανομή με $r-1$ βαθμούς ελευθερίας, χ^2_{r-1} .

Όταν η υπόθεση της συμμετρίας δεν είναι αληθής, τότε η τετραγωνική μορφή $Q_n(t_n)$ παίρνει μεγάλες τιμές. Επομένως, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το δ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $Q_n(t_n) \geq \chi_{\alpha, r-1}^2$, όπου το $\chi_{\alpha, r-1}^2$ είναι το $100(1-\alpha)$ εκατοστιαίο σημείο της χ_{r-1}^2 κατανομής.

Παρατήρηση 3.3.2.1 Με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης που έγινε από τον Koutrounelis (1985), προκύπτει ότι ο έλεγχος της τετραγωνικής μορφής $Q_n(t_n)$ υποεκτιμά το επίπεδο σημαντικότητας. Ειδικότερα, το πρόβλημα χειροτερεύει όταν το r είναι μεγάλο, καθώς σε τέτοιες περιπτώσεις η κατανομή της $Q_n(t_n)$ αποκλίνει σε μεγαλύτερο βαθμό από την ασυμπτωτική χ -τετράγωνο κατανομή. Επιπλέον, θεωρώντας ως εναλλακτικές κατανομές ορισμένες χ -τετράγωνο κατανομές, προκύπτει ότι η ισχύς του ελέγχου αυξάνεται προς τη μονάδα καθώς αυξάνεται το μέγεθος του δείγματος. Άρα, ο έλεγχος είναι συνεπής έναντι των ασύμμετρων χ -τετράγωνο κατανομών. Τέλος, η ισχύς τείνει να μικραίνει όσο η κατανομή γίνεται λιγότερο ασύμμετρη.

3.4 Έλεγχοι με U-στατιστικές συναρτήσεις

Στην ενότητα αυτή θα παρουσιαστούν οι κυριότερες μεθοδολογίες που αναπτύχθηκαν για τον έλεγχο της υπόθεσης της συμμετρίας μιας κατανομής γύρω από κάποιο άγνωστο σημείο και στηρίζονται στις U -στατιστικές συναρτήσεις (η μορφή των οποίων παρουσιάσθηκε στην ενότητα 1.1). Στο πλαίσιο αυτό, θα παρουσιαστεί το triple test που προτάθηκε ανεξάρτητα από τους Randles *et al.* (1980) και τους Davis and Quade (1978), ο έλεγχος του Gupta (1969)

με τη γενίκευση των Antille *et al.* (1982) και τέλος ο έλεγχος του Finch (1977).

3.4.1 Triple τεστ

Στην ενότητα αυτή θα παρουσιαστεί ο έλεγχος που προτάθηκε ανεξάρτητα από τους Davis and Quade (1978) και από τους Randles *et al.* (1980), για την υπόθεση της συμμετρίας ενός πληθυσμού γύρω από κάποιο άγνωστο σημείο, έστω μ . Ο έλεγχος αυτός είναι γνωστός ως triple τεστ καθώς στηρίζεται σε τριάδες παρατηρήσεων (triples observations) και ανήκει στην οικογένεια των U -σ.σ. (U -statistics). Η παρουσίαση που ακολουθεί βασίζεται στην εργασία των Randles *et al.* (1980).

Συγκεκριμένα, έστω $X_1, X_2, \dots, X_n$ τυχαίο δείγμα από ένα συνεχή πληθυσμό με α.σ.κ. $F(x - \mu)$. Αν i, j, k , είναι τρεις διαφορετικοί μεταξύ τους δείκτες τέτοιοι ώστε $1 < i, j, k < n$ τότε η τριάδα παρατηρήσεων (X_i, X_j, X_k) λέγεται δεξιά (right triple), αν η μεσαία παρατήρηση βρίσκεται πιο κοντά στη μικρότερη από ότι στη μεγαλύτερη παρατήρηση. Ανάλογα, αν η μεσαία παρατήρηση μιας τριάδας (X_i, X_j, X_k) βρίσκεται πιο κοντά στη μεγαλύτερη παρατήρηση από ότι στην μικρότερη, τότε η τριάδα αυτή ονομάζεται αριστερή τριάδα (left triple).

Οι Randles *et al.* (1980) παρατήρησαν ότι αν η υποκείμενη συνάρτηση κατανομής είναι συμμετρική, τότε η κατανομή πιθανότητας του $X_1 + X_2 - 2X_3$ θα ταυτίζεται με την κατανομή πιθανότητας του $-X_1 - X_2 + 2X_3$ και επομένως

$$\eta = P(X_1 + X_2 - 2X_3 > 0) - P(X_1 + X_2 - 2X_3 < 0) = 0. \quad (3.4.1.1)$$

Επιπλέον, από τους Randles *et al.* (1980) προκύπτει ότι η κλάση των ασύμμετρων κατανομών για τις οποίες ισχύει $\eta=0$ είναι μικρή. Η παρατήρηση αυτή έδωσε την ιδέα στους Randles *et al.* (1980) να στηρίζουν τον έλεγχο τους σε έναν εκτιμητή του η .

Ειδικότερα, η σ.σ. που προτάθηκε από τους Randles *et al.* (1980) για τον έλεγχο της υπόθεσης που διατυπώνεται στη σχέση (3.1.1.1), είναι της μορφής των U -στατιστικών συναρτήσεων και δίνεται από τον τύπο:

$$\hat{\eta} = \binom{n}{3}^{-1} \sum_{i < j < k} f^*(X_i, X_j, X_k), \quad (3.4.1.2)$$

όπου

$$f^*(X_i, X_j, X_k) = \frac{1}{3} \left[\text{sign}(X_i + X_j - 2X_k) + \right. \\ \left. + \text{sign}(X_i + X_k - 2X_j) + \text{sign}(X_j + X_k - 2X_i) \right], \quad (3.4.1.3)$$

με

$$\text{sign}(u) = \begin{cases} 1, & \text{αν } u > 0 \\ 0, & \text{αν } u = 0 \\ -1, & \text{αν } u < 0. \end{cases} \quad (3.4.1.4)$$

Εύκολα προκύπτει ότι αν η τριάδα (X_i, X_j, X_k) είναι δεξιά τότε $f^*(X_i, X_j, X_k) = \frac{1}{3}$, αν είναι αριστερή τότε $f^*(X_i, X_j, X_k) = -\frac{1}{3}$, ενώ η πιθανότητα να ισχύει $f^*(X_i, X_j, X_k) = 0$ είναι μηδέν για δείγματα από συνεχή πληθυσμό. Επομένως, προκύπτει ότι:

$$\hat{\eta} = \frac{[(\text{πλήθος δεξιών τριάδων}) - (\text{πλήθος αριστερών τριάδων})]}{\left[3 \binom{n}{3} \right]}.$$

Συνεπώς, ο έλεγχος της υπόθεσης της συμμετρίας ανάγεται στον έλεγχο της $H_0: \eta = 0$ έναντι της εναλλακτικής $H_1: \eta \neq 0$, μέσω της σ.σ. $\hat{\eta}$. Η κατανομή της σ.σ. $\hat{\eta}$ προσδιορίζεται στο θεώρημα που ακολουθεί.

Θεώρημα 3.4.1.1 (Hoeffding (1948)) Η αναμενόμενη τιμή της U -σ.σ. $\hat{\eta}$ είναι:

$$E(\hat{\eta}) = \eta = P(X_1 + X_2 - 2X_3 > 0) - P(X_1 + X_2 - 2X_3 < 0), \quad (3.4.1.5)$$

και η διακύμανση της είναι:

$$Var(\hat{\eta}) = \sigma_n^2 / n = \binom{n}{3}^{-1} \sum_{i=1}^3 \binom{3}{i} \binom{n-3}{3-i} \zeta_i, \quad (3.4.1.6)$$

όπου

$$\zeta_i = Var(f_i^*(X_1, \dots, X_i)) \text{ και } f_i^*(x_1, \dots, x_i) = E[f^*(x_1, \dots, x_i, X_{i+1}, \dots, X_3)].$$

Ακόμη, από τον Hoeffding (1948) προκύπτει ότι για $n \rightarrow \infty$ ισχύει:

$$\frac{n^{1/2}(\hat{\eta} - \eta)}{\sigma_n} \xrightarrow{d} N(0,1), \quad (3.4.1.7)$$

ενώ ορίζοντας ως $\sigma_A^2 = 9\zeta_1$ αποδεικνύεται ότι:

$$\frac{n^{1/2}(\hat{\eta} - \eta)}{\sigma_A} \xrightarrow{d} N(0,1). \quad (3.4.1.8)$$

Επομένως, προκύπτει ένας έλεγχος που δεν είναι ανεξάρτητος κατανομής. Οι Randles *et al.* (1980) τροποποίησαν τις σ.σ. των σχέσεων (3.1.4.7) και (3.1.4.8) με στόχο τη δημιουργία σ.σ. οι οποίες θα είναι ασυμπτωτικά ανεξάρτητες κατανομής. Έτσι, πρότειναν την αντικατάσταση των ασυμπτωτικών διακυμάνσεων σ_n και σ_A από τους συνεπείς εκτιμητές $\hat{\sigma}_n$ και $\hat{\sigma}_A$, αντίστοιχα. Οι εκτιμητές αυτοί

προκύπτουν με αντικατάσταση των ζ_i , $i=1,2,3$, από τις εκτιμώμενες τιμές $\hat{\zeta}_i$ που προσδιορίζονται στις ακόλουθες σχέσεις:

$$\hat{\zeta}_1 = \frac{1}{n} \sum_{i=1}^n \left(\hat{f}_1^*(X_i) - \hat{\eta} \right)^2, \quad (3.4.1.9)$$

όπου

$$\hat{f}_1^*(X_i) = \frac{1}{\binom{n-1}{2}} \sum_{\substack{j < k \\ j \neq i \neq k}} f^*(X_i, X_j, X_k), \quad (3.4.1.10)$$

$$\hat{\zeta}_2 = \frac{1}{\binom{n}{2}} \sum_{j < k} \left(\hat{f}_2^*(X_j, X_k) - \hat{\eta} \right)^2, \quad (3.4.1.11)$$

όπου

$$\hat{f}_2^*(X_j, X_k) = \frac{1}{n-2} \sum_{\substack{i=1 \\ i \neq j \neq k \\ i \neq k}}^n f^*(X_i, X_j, X_k), \quad (3.4.1.12)$$

και τέλος

$$\hat{\zeta}_3 = \frac{1}{9} - \hat{\eta}^2. \quad (3.4.1.13)$$

Επομένως, για τον έλεγχο της υπόθεσης H_0 οι Randles *et al.* (1980) προτείνουν τη χρήση των σ.σ.

$$T_{R1} = \frac{n^{1/2} \hat{\eta}}{\hat{\sigma}_n}, \quad (3.4.1.14)$$

και

$$T_{R2} = \frac{n^{1/2} \hat{\eta}}{\hat{\sigma}_A}, \quad (3.4.1.15)$$

οι οποίες ακολουθούν ασυμπτωτικά τυπική κανονική κατανομή $N(0,1)$. Στο παραπάνω πλαίσιο, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μ απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|T_{Ri}| > z_{\alpha/2}$, $i=1,2$. Οι έλεγχοι αυτοί είναι συνεπείς αν και μόνο αν $\eta \neq 0$.

Από μελέτες προσομοίωσης που έγιναν από τους Randles *et al.* (1980), προκύπτει ότι ο έλεγχος που στηρίζεται στη σ.σ. T_{R1} είναι αποτελεσματικός για να ανιχνεύει αποκλίσεις από τη συμμετρία και η προσαρμογή του είναι καλύτερη από τον αντίστοιχο έλεγχο της σ.σ. T_{R2} .

Παρατήρηση 3.1.4.1 Στη μελέτη προσομοίωσης των Randles *et al.* (1980) συμπεριλήφθηκε ο έλεγχος με κρίσιμη περιοχή $|T_{R1}| \geq t_{n,\alpha/2}$, όπου $t_{n,\alpha}$ οι κρίσιμες τιμές της t-κατανομής με n βαθμούς ελευθερίας και επίπεδο σημαντικότητας α . Αυτός ο έλεγχος διατηρεί καλύτερα το επίπεδο σημαντικότητας, προκαλώντας μικρή μείωση στην ισχύ. Τελικά, από τους Randles *et al.* (1980) προτείνεται η εφαρμογή του ελέγχου με κρίσιμη περιοχή $|T_{R1}| \geq t_{n,\alpha/2}$.

3.4.2 Ο έλεγχος του Gupta (1967) και μια επέκταση του

Στο εδάφιο αυτό αρχικά θα παρουσιαστεί ο έλεγχος που προτάθηκε από τον Gupta (1967) για την υπόθεση της συμμετρίας μιας συνάρτησης κατανομής γύρω από κάποιο άγνωστο σημείο μ , έχοντας ως εναλλακτική την υπόθεση ότι η κατανομή των θετικών παρατηρήσεων είναι στοχαστικά μεγαλύτερη από την κατανομή των απολύτων τιμών των αρνητικών παρατηρήσεων, δηλαδή την

$$H_1: P(X \leq x | X > 0) \leq P(-X \leq x | X < 0), \text{ για κάθε } x > 0. \quad (3.4.2.1)$$

Έπειτα, θα παρουσιασθεί ένας τρόπος ελέγχου που προτάθηκε από τους Antille *et al.* (1982) και περικλείει τη σ.σ. του Gupta (1967) ως ειδική περίπτωση.

Η σ.σ. που προτάθηκε από τον Gupta (1967) ανήκει στην οικογένεια των U -στατιστικών συναρτήσεων καθώς είναι της μορφής:

$$\hat{J} = \hat{J}_n = \binom{n}{2}^{-1} \sum_{i < j} \varphi(X_i - \hat{\mu}, X_j - \hat{\mu}), \quad (3.4.2.2)$$

όπου

$$\varphi(X_i - \hat{\mu}, X_j - \hat{\mu}) = \begin{cases} 1, & \text{αν είτε } X_i < \hat{\mu} < X_j \text{ και } \hat{\mu} - X_i < X_j - \hat{\mu}, \\ & \text{είτε } X_j < \hat{\mu} < X_i \text{ και } \hat{\mu} - X_j < X_i - \hat{\mu} \\ 0, & \text{αλλιώς,} \end{cases} \quad (3.4.2.3)$$

και $\hat{\mu}$ η δειγματική διάμεσος. Μια ισοδύναμη έκφραση αυτής της σ.σ. είναι η:

$$\hat{J}_n = \binom{n}{2}^{-1} \sum_{i < j} [I(X_i + X_j > 2\hat{\mu}) - I(X_i > \hat{\mu}, X_j > \hat{\mu})], \quad (3.4.2.4)$$

όπου

$$I(X_i + X_j > 2\hat{\mu}) = \begin{cases} 1, & \text{αν } X_i + X_j > 2\hat{\mu} \\ 0, & \text{διαφορετικά,} \end{cases} \quad (3.4.2.5)$$

και

$$I(X_i > \hat{\mu}, X_j > \hat{\mu}) = \begin{cases} 1, & \text{αν } X_i > \hat{\mu} \text{ και } X_j > \hat{\mu} \\ 0, & \text{διαφορετικά.} \end{cases} \quad (3.4.2.6)$$

Η σ.σ. $\hat{J}_n$ βασίζεται στο πλήθος των θετικών αποκλίσεων από τη δειγματική διάμεσο $\hat{\mu}$ οι οποίες υπερβαίνουν τις απόλυτες τιμές των αρνητικών αποκλίσεων από τη δειγματική διάμεσο $\hat{\mu}$. Αν η H_1

είναι αληθής, τότε αναμένεται μεγάλη τιμή της σ.σ. $\hat{J}_n$. Επομένως, η H_0 απορρίπτεται έναντι της H_1 για μεγάλες τιμές της σ.σ. $\hat{J}_n$.

Το επόμενο βήμα για τον ακριβή προσδιορισμό της κρίσιμης περιοχής είναι η εύρεση της ασυμπτωτικής κατανομής της σ.σ. $\hat{J}_n$ υπό την H_0 .

Θεώρημα 3.4.2.1 (Gupta (1967)) Έστω $X_1, X_2, \dots, X_n$ τυχαίο δείγμα από την α.σ.κ. $F(x - \mu)$ και έστω $A(t) = E[\varphi(X_1 - t, X_2 - t) | \mu = 0]$ με συνεχή παράγωγο σε μια περιοχή του ορίσματος του. Έστω ότι η σ.π.π. f είναι συνεχής κοντά στο όρισμα με $f(0) \neq 0$ και έστω ότι η σ.π.π. f είναι φραγμένη. Τότε για $n \rightarrow \infty$, υπό την H_0 , ισχύει ότι:

$$n^{1/2}(\hat{J}_n - E(\hat{J}_n)) \xrightarrow{d} N(0, \sigma_0^2), \quad (3.4.2.7)$$

$$\text{με } E(\hat{J}_n) = \frac{1}{4} \text{ και } \sigma_0^2 = \frac{1}{12} + \frac{1}{4} \left[1 - \frac{2 \int f^2(x) dx}{f(0)} \right]^2.$$

Έτσι, ο έλεγχος που στηρίζεται στη σ.σ. $\hat{J}_n$ δεν είναι ασυμπτωτικά ανεξάρτητος κατανομής, αφού εξαρτάται από την υποκείμενη κατανομή μέσω της ασυμπτωτικής διακύμανσης σ_0^2 . Ωστόσο, μπορεί να τροποποιηθεί ώστε να γίνει ασυμπτωτικά ανεξάρτητος κατανομής, βρίσκοντας συνεπείς εκτιμητές των $f(0)$ και $\int f^2(x) dx$ και στη συνέχεια συνεπή εκτιμητή του σ_0^2 . Τέτοιου είδους εκτιμητές προσδιορίστηκαν από τον Gupta (1967).

Αρχικά, για το $\int f^2(x) dx$, έστω $n'(\mu)$ το πλήθος των $(X_i + X_j)/2$ με $i < j$ που υπερβαίνουν το μ . Έστω ότι οι $\binom{n}{2}$ το

πλήθος μέσες τιμές $(X_i + X_j)/2$, $i < j$ διατάσσονται ως $A^{(1)} < A^{(2)} < \dots < A^{(M)}$, όπου $M = \binom{n}{2}$. Ο Lehmann (1963b) έδειξε ότι το διάστημα εμπιστοσύνης επιπέδου σημαντικότητας α , για τον δίπλευρο έλεγχο της υπόθεσης $H_0: \mu = \mu_0$, είναι

$$\underline{\mu} < \mu < \bar{\mu}, \quad (3.4.2.8)$$

όπου $\underline{\mu} = A^{(c_\alpha)}$ και $\bar{\mu} = A^{(M+1-c_\alpha)}$, με την τιμή c_α να προσδιορίζεται από τη σχέση $P\left(c_\alpha \leq n'(\mu_0) \leq \binom{n}{2} - c_\alpha \mid H_0\right) = 1 - \alpha$. Στο παραπάνω πλαίσιο, ο συνεπής εκτιμητής του $\int f^2(x)dx$ προκύπτει από το ακόλουθο Λήμμα.

Λήμμα 3.4.2.1 (Lehmann (1963a)) Αν $L = \bar{\mu} - \underline{\mu}$ είναι το μήκος του διαστήματος εμπιστοσύνης της σχέσης (3.4.2.8), τότε:

$$\frac{(3n)^{1/2} L}{k_{\alpha/2}} \xrightarrow{p} \frac{1}{\int f^2(x)dx}, \quad (3.4.2.9)$$

όπου $k_{\alpha/2}$ είναι το $100\alpha/2$ άνω εκατοστιαίο σημείο της τυπικής κανονικής κατανομής και όπου $\xrightarrow{p}$ το κατά πιθανότητα όριο.

Για τον συνεπή εκτιμητή του $f(0)$, ο Gupta (1967) όρισε τη συνάρτηση

$$f_n = \frac{\sum_{i=1}^n I(\hat{\mu} - h \leq X_i \leq \hat{\mu} + h)}{2nh}, \quad (3.4.2.10)$$

όπου h μια συνάρτηση του τυχαίου δείγματος τέτοια ώστε $\lim_{n \rightarrow \infty} h = 0$ και $\lim_{n \rightarrow \infty} nh = \infty$ και απέδειξε ότι είναι συνεπής εκτιμητής του $f(0)$.

Συνδυάζοντας τα παραπάνω αποτελέσματα, προκύπτει ένας συνεπής εκτιμητής έστω S_n^2 της σ_0^2 ο οποίος προσδιορίζεται στο παρακάτω Λήμμα.

Λήμμα 3.4.2.2 (Gupta (1967)) Ισχύει ότι:

$$S_n^2 = \frac{1}{12} + \frac{1}{4} \left[1 - \frac{2r_n}{f_n} \right]^2 \xrightarrow{p} \frac{1}{12} + \frac{1}{4} \left[1 - \frac{2 \int f^2(x) dx}{f(0)} \right]^2,$$

$$\text{όπου } r_n = \frac{k_{\alpha/2}}{(3n)^{1/2} L}.$$

Τελικά, υπό την H_0 , η ασυμπτωτικά ανεξάρτητης κατανομής σ.σ. που προτάθηκε από τον Gupta (1967) είναι η

$$Z_{GJ} = \frac{n^{1/2} \left(\hat{J} - E(\hat{J}_n) \right)}{S_n} \xrightarrow{d} N(0,1), \quad (3.4.2.11)$$

ενώ για επίπεδο σημαντικότητας α , η H_0 απορρίπτεται έναντι της H_1' , αν $Z_{GJ} \geq z_\alpha$.

Η γενίκευση των Antille *et al.* (1982)

Στη συνέχεια θα παρουσιαστούν οι έλεγχοι που προτάθηκαν από τους Antille *et al.* (1982) για την υπόθεση της συμμετρίας μιας συνάρτησης κατανομής γύρω από κάποιο άγνωστο σημείο, έστω μ , οι οποίοι περικλείουν τη σ.σ. του Gupta (1969) ως ειδική περίπτωση.

Έστω $X_1, X_2, \dots, X_n$ τ.δ. από κάποιο πληθυσμό με α.σ.κ. F και σ.π.π. f . Η ιδέα των Antille *et al.* (1982) στηρίχθηκε στην παρατήρηση τους ότι ένας τρόπος για να γίνει έλεγχος συμμετρίας γύρω από κάποιο άγνωστο σημείο, είναι η εκτίμηση του κέντρου συμμετρίας και στη συνέχεια η εφαρμογή ελέγχου συμμετρίας γύρω από το μηδέν. Έτσι, πρότειναν την εκτίμηση του κέντρου συμμετρίας μ από την εμπειρική διάμεσο M_n και όρισαν $R_i(M_n)$ την τάξη του $|X_i - M_n|$ ως προς τα $|X_1 - M_n|, |X_2 - M_n|, \dots, |X_n - M_n|$. Η σ.σ. που προτάθηκε από τους Antille *et al.* (1982) είναι η:

$$T_1(\alpha) = n^{-1/2} \sum_{i=1}^n G_\alpha \left(\frac{R_i(M_n)}{2(n+1)} \right) \text{sign}(X_i - M_n), \quad (3.4.2.12)$$

όπου $G_\alpha(x) = \min\left(x, \frac{1}{2} - \alpha\right)$, με $0 \leq x \leq \frac{1}{2}$, $0 \leq \alpha \leq \frac{1}{2}$ και το $\text{sign}(\cdot)$ να είναι όπως ορίστηκε στη σχέση (3.4.1.4). Η υπόθεση της συμμετρίας απορρίπτεται για αρκετά μεγάλες τιμές της σ.σ. $T_1(\alpha)$.

Παρατήρηση 3.4.2.1 Για $\alpha = 0$ η προκύπτει η σ.σ. $T_1(0)$, η οποία είναι η σ.σ. $\hat{J}_n$ της σχέσης (3.4.2.2) που προτάθηκε από τον Gupta (1967).

Στο θεώρημα που ακολουθεί προσδιορίζεται η ασυμπτωτική κατανομή της σ.σ. $T_1(\alpha)$.

Θεώρημα 3.4.2.2 (Antille *et al.* (1982)) Υπό την υπόθεση της συμμετρίας, η σ.σ. $T_1(\alpha)$ ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή μηδέν και διακύμανση

$$\sigma_1^2(\alpha) = \frac{1}{4}(1-2\alpha)^2 \left(\frac{1}{12} + \frac{\alpha}{3} - \alpha^2 \right) + \frac{1}{4} \left[A - \left(\frac{1}{2} - 2\alpha^2 \right) \right]^2,$$

όπου $A = \frac{1}{f(\mu)} \int_{\alpha}^{1-\alpha} f(F^{-1}(u)) du$. Επιπλέον, όταν η υποκείμενη σ.π.π. είναι μονοκόρυφη ισχύει:

$$\frac{1}{4}(1-2\alpha)^2 \left(\frac{1}{12} + \frac{\alpha}{3} - \alpha^2 \right) \leq \sigma_1^2(\alpha) \leq \frac{1}{4} \left[(1-2\alpha)^2 \left(\frac{1}{12} + \frac{\alpha}{3} - \alpha^2 \right) + \left(\frac{1}{2} - 2\alpha^2 \right)^2 \right].$$

Είναι γνωστό ότι ένας έλεγχος διατηρεί το επίπεδο σημαντικότητας αν η διακύμανση της σ.σ. του ελέγχου είναι φραγμένη από πάνω. Επιπλέον, η ισχύς του ελέγχου διατηρείται σταθερή αν η διακύμανση της σ.σ. του ελέγχου είναι φραγμένη από κάτω. Με το παραπάνω σκεπτικό, προκύπτει ότι η σ.σ. $T_1(\alpha)$ έχει καλύτερη προσαρμογή συγκριτικά με τους ελέγχους $T_2(J)$ και $T_3(J)$ που δόθηκαν στις σχέσεις (3.1.6.5) και (3.4.3.7), αντίστοιχα. Επιπλέον, καθώς η σ.σ. $T_1(0)$ έχει ασυμπτωτική διακύμανση

$$0.028 \leq \sigma_1^2(0) \leq 0.083, \text{ ενώ η σ.σ. } T_1\left(\frac{1}{6}\right) \text{ έχει ασυμπτωτική διακύμανση}$$

$$0.012 \leq \sigma_1^2\left(\frac{1}{6}\right) \leq 0.061, \text{ προκύπτει ότι προτιμάται η χρήση της σ.σ.}$$

$T_1(0)$ έναντι της $T_1\left(\frac{1}{6}\right)$. Ωστόσο, με βάση τα αποτελέσματα μιας μελέτης προσομοίωσης των Antille *et al.* (1982), οι πιθανότητες απόρριψης της συμμετρίας με εφαρμογή του ελέγχου της σ.σ. $T_1\left(\frac{1}{6}\right)$,

ποικίλλουν για μια κλάση συμμετρικών κατανομών από 0.02 μέχρι 0.08 (βλέπε Table 2, σελ. 641, Antille *et al.* (1982)). Επειδή αυτά τα αποτελέσματα είναι αποδεκτά, στην πράξη δεν μπορεί να αποκλειστεί η χρήση της σ.σ. $T_1\left(\frac{1}{6}\right)$.

Παρατήρηση 3.4.2.2 Οι Antille *et al.* (1982) επικέντρωσαν το ενδιαφέρον τους στη συγκριτική μελέτη των σ.σ. $T_1(0)$, $T_1\left(\frac{1}{6}\right)$, $T_2(0.05)$ και $T_3(0.05)$, με τη σ.σ. $T_3(0.05)$ να προσδιορίζεται στην Παράγραφο 3.4.3.

Ως αποτέλεσμα αυτής της μελέτης προσομοίωσης προκύπτει ότι αν η υποκείμενη κατανομή δεν έχει μεγάλες ουρές, τότε προτείνεται η χρήση των σ.σ. $T_1(0)$ και $T_1\left(\frac{1}{6}\right)$. Σε αντίθετη περίπτωση προτείνεται η εφαρμογή του ελέγχου της σ.σ. $T_3(0.05)$. Ο έλεγχος της σ.σ. $T_2(0.05)$ δεν ενδείκνυται για χρήση καθώς δεν διατηρεί το επίπεδο σημαντικότητας.

Παρατήρηση 3.4.2.3 Οι Antille *et al.* (1982) κατέληξαν στο συμπέρασμα ότι η ασυμπτωτική κατανομή της σ.σ. $T_1(0)$ προσεγγίζεται ικανοποιητικά από την κανονική κατανομή $N(0, 0.03156)$. Στο πλαίσιο αυτό, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $|T_1(0)| \geq 0.1776z_{\alpha/2}$. Επιπλέον, η ασυμπτωτική κατανομή της σ.σ. $T_1\left(\frac{1}{6}\right)$ προσεγγίζεται ικανοποιητικά από την κανονική κατανομή $N(0, 0.01736)$. Ανάλογα με πριν, η υπόθεση της συμμετρίας απορρίπτεται αν $\left|T_1\left(\frac{1}{6}\right)\right| \geq 0.1317z_{\alpha/2}$.

3.4.3 Ο έλεγχος του Finch (1977)

Στο εδάφιο αυτό θα παρουσιαστεί ο έλεγχος που προτάθηκε από τον Finch (1977) για την υπόθεση της συμμετρίας, που διατυπώθηκε στη σχέση (3.1.1.1).

Έστω $X_1, X_2, \dots, X_n$ το διαθέσιμο διατεταγμένο δείγμα, $R_i = X_{n-k+i} - X_{n-1-k+i}$ και $L_i = X_{k+2-i} - X_{k+1-i}$, $i = 1, \dots, k$ όπου $k = \lfloor n/2 \rfloor$. Ο Finch (1977) παρατήρησε ότι αν η κατανομή είναι συμμετρική, τότε $L_i = R_i$, για κάθε $i = 1, \dots, k$, ενώ αν η κατανομή δειγματοληψίας είναι λοξή, τότε τα $\{R_i\}$ είναι είτε μεγαλύτερα είτε μικρότερα από τα αντίστοιχα $\{L_i\}$. Η σ.σ. που προτάθηκε από τον Finch (1977) είναι η:

$$V = \sum_{i=1}^k w_i V_i, \quad (3.4.3.1)$$

όπου w_i είναι κατάλληλα επιλεγμένα βάρη και $V_i = (R_i - L_i) / (R_i + L_i)$. Η προσέγγιση της δειγματικής κατανομής της σ.σ. V προκύπτει με βάση το Λήμμα που ακολουθεί.

Λήμμα 3.4.3.1 (Finch (1977)) Έστω R και L ανεξάρτητες τ.μ. από την εκθετική κατανομή με μέση τιμή p και λ , αντίστοιχα. Αν $\theta = \lambda / p$ και $U = (R - L) / (R + L)$, τότε η σ.π.π. $f_U(u)$ της τ.μ. U είναι:

$$f_U(u) = \frac{2\theta}{(1 + \theta - (1 - \theta)u)^2}, \quad -1 < u < 1, \quad (3.4.3.2)$$

με

$$E(U) = \begin{cases} 0, & \text{αν } \theta = 1 \\ 2\theta \{ \ln \theta + (1 - \theta^2) / (2\theta) \} / (1 - \theta)^2, & \theta \neq 1. \end{cases} \quad (3.4.3.3)$$

Η προσέγγιση που χρησιμοποιήθηκε από τον Finch (1977) στο Λήμμα 3.4.3.1, είναι ότι οι ποσότητες R_i και L_i είναι ανεξάρτητες εκθετικά κατανεμημένες τ.μ. και έτσι η (3.4.3.3) είναι μια αποτελεσματική προσέγγιση της $E(U)$. Από τον Finch (1977) προτείνεται η εφαρμογή των βαρών

$$w_i = c \sum_{j=k-(i-1)}^{k+(i-1)} (j)^{-1}, \quad (3.4.3.4)$$

με το c να είναι τέτοιο ώστε $\sum_{i=1}^k w_i^2 = 1$ και για το μέγεθος του δείγματος να ισχύει ότι $n = 2k + 1$.

Ακριβή αποτελέσματα σχετικά με τη δειγματική κατανομή των V_i είναι δύσκολο να βρεθούν. Αν το τ.δ. είναι μικρού μεγέθους και προέρχεται από την κανονική κατανομή, μπορεί να προσδιοριστεί η από κοινού σ.π.π. των V_i , χωρίς ωστόσο να προσφέρει κάποια πληροφορία.

Ο Finch (1977) απέδειξε ότι η μοναδική περίπτωση που μπορεί να προσδιοριστεί η ακριβής από κοινού κατανομή των V_i , είναι όταν το τυχαίο δείγμα προέρχεται από την ομοιόμορφη κατανομή. Για το λόγο αυτό, από τον Finch (1977) προτάθηκε η σ.σ.:

$$V_A = \varphi_n(T^*) V \stackrel{\text{ασυμπ.}}{\underset{H_0}{\sim}} N(0,1), \quad (3.4.3.5)$$

όπου

$$\varphi_n(T^*) = 3c_n^2 + \frac{1 - c_n^2}{a_n(T^*)^2 + b_n T^* + 1/3}, \quad (3.4.3.6)$$

$$\text{με } 0.8 < T^* < 1.5, \quad a_n = 0.041 + \frac{23.4}{n^2}, \quad m_n = 1.47 - \frac{8}{n} = -\frac{b_n}{2a_n} \quad \text{και}$$

$$c_n = 0.278 + \frac{1.8}{n}.$$

Επομένως, προκύπτει ότι για επίπεδο σημαντικότητας α , απορρίπτεται η υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|V_A| \geq z_{\alpha/2}$.

Παρατήρηση 3.4.3.1 (Antille *et al.* (1982)) Οι Antille *et al.* (1982) δίνουν μια εναλλακτική μορφή της σ.σ. V , η οποία είναι η:

$$T_3(J) = n^{-1/2} \sum_{i=1}^{[(n-1)/2]} J\left(\frac{i}{n+1}\right) V_i, \quad (3.4.3.7)$$

όπου

$$V_i = \frac{Y_i - Y_{n-i}}{Y_i + Y_{n-i}}, \quad \text{για } i = 1, 2, \dots, \left[\frac{n-1}{2}\right], \quad (3.4.3.8)$$

με $Y_i = X_{(i+1)} - X_{(i)}$ για $i = 1, \dots, n-1$, $X_{(1)}, \dots, X_{(n)}$ οι διατεταγμένες τιμές των $X_1, \dots, X_n$ και όπου $J(t)$ μια συνάρτηση η οποία παίρνει τιμές στο $\left(0, \frac{1}{2}\right)$. Όμοια με πριν, μεγάλες κατά απόλυτη τιμή, τιμές της σ.σ. $T_3(J)$ συνηγορούν στην απόρριψη της H_0 . Επιπλέον, απέδειξαν ότι υπό την υπόθεση της συμμετρίας, η σ.σ. $T_3(J)$ ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή μηδέν και διακύμανση

$$\sigma_3^2(J) = \frac{1}{9} \int_0^{1/2} J^2(t) dt + \frac{2}{9} \int_0^{1/2} \left[J(t) + \int_t^{1/2} J(s) \frac{\psi''(s)}{\psi'(s)} ds \right]^2 dt, \quad (3.4.3.9)$$

όπου με ψ συμβολίζεται η F^{-1} . Ακόμη, απέδειξαν ότι για την διακύμανση $\sigma_3^2(J)$ ισχύει $\frac{1}{9} \left(\frac{1}{2} - \alpha \right) \leq \sigma_3^2 < \infty$. Τέλος, για μικρές τιμές

του α προτείνουν τη χρήση της συνάρτησης $J(t) = I\left(\left[\alpha, \frac{1}{2}\right]\right)$ ώστε να σταθεροποιείται η διακύμανση.

Παρατήρηση 3.4.3.2 (Antille *et al.* (1982)) Οι Antille *et al.* (1982) διεξάγοντας μια εκτενή μελέτη προσομοίωσης κατέληξαν στο συμπέρασμα ότι η ασυμπτωτική κατανομή της σ.σ. $T_3(0.05)$ προσεγγίζεται ικανοποιητικά από την κανονική κατανομή $N(0, 0.1293)$. Έτσι, για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $T_3(0.05) \geq 0.3595z_{\alpha/2}$. Για επιπλέον ακρίβεια, προτείνεται η χρήση των πιθανοτήτων $P(T_3(0.05) \leq x) = \alpha \approx \bar{\alpha}$, όπου για περιττό μέγεθος δείγματος $\bar{\alpha} = \Phi\left(\frac{x}{0.3596}\right)$, ενώ για άρτιο μέγεθος δείγματος $\bar{\alpha} = \Phi\left(\frac{x}{0.3596((n-1)/n)^{1/2}}\right)$.

ΚΕΦΑΛΑΙΟ 4

Συγκριτική μελέτη

Όπως είδαμε στα προηγούμενα κεφάλαια της παρούσας διατριβής, υπάρχει μια ποικιλία από ελέγχους για να κρίνουμε αν ισχύει η υπόθεση της συμμετρίας γύρω από ένα γνωστό ή άγνωστο σημείο. Στη λήψη απόφασης για το ποιες από τις μεθόδους που παρουσιάστηκαν μπορούν να προταθούν για να χρησιμοποιούνται στην πράξη κυρίαρχο ρόλο διαδραματίζουν οι ακόλουθοι παράγοντες:

- α) ο στατιστικός έλεγχος να είναι ανεξάρτητος κατανομής,
- β) η σ.σ. να παραμένει αναλλοίωτη σε μετασχηματισμούς των δεδομένων της μορφής $X \rightarrow X^{-1}$ και $X \rightarrow -X$,
- γ) η σ.σ. και η κρίσιμη περιοχή της να υπολογίζονται εύκολα,
- δ) να διατηρείται το επίπεδο σημαντικότητας του ελέγχου, και τέλος
- ε) ο έλεγχος να έχει καλή απόδοση ισχύος ως προς ένα ευρύ φάσμα εναλλακτικών.

Αν δύο ή περισσότερες στατιστικές μεθοδολογίες πληρούν τις επιθυμητές ιδιότητες, α)-ε), τότε ιδανικά καλύτερη είναι εκείνη που έχει την υψηλότερη ισχύ για όλες τις πιθανές μορφές εναλλακτικών κατανομών. Όπως είναι προφανές, η αποτελεσματικότητα αυτή εξαρτάται τόσο από την εναλλακτική κατανομή όσο και από το μέγεθος του δείγματος. Η διαδικασία που υιοθετείται για τη σύγκριση ως προς τα σημεία δ) και ε) που αναφέρθηκαν είναι η διεξαγωγή μιας Monte Carlo μελέτης. Ειδικότερα, η απόδοση ως προς το επίπεδο σημαντικότητας αξιολογείται προσομοιώνοντας συμμετρικά δεδομένα, ενώ η απόδοση ισχύος αξιολογείται προσομοιώνοντας δεδομένα που δεν είναι συμμετρικά. Στο πλαίσιο αυτό, πολλοί συγγραφείς

παρακινούμενοι από τον McWilliams (1990) θεωρούν δεδομένα που προέρχονται από τη Γενικευμένη Λάμδα Οικογένεια Κατανομών (Generalized Lambda Family Distribution). Τα μέλη αυτής της οικογένειας κατανομών έχουν αντίστροφη α.σ.κ. που δίνεται από τη σχέση:

$$F^{-1}(u) = \lambda_1 + \frac{[u^{\lambda_3} - (1-u)^{\lambda_4}]}{\lambda_2}, \quad 0 < u < 1.$$

Στην οικογένεια αυτή η παράμετρος λ_1 είναι παράμετρος θέσης, η λ_2 είναι παράμετρος κλίμακας, ενώ οι παράμετροι λ_3 και λ_4 καθορίζουν τη λοξότητα της κατανομής. Συγκεκριμένα, όταν $\lambda_3 = \lambda_4$ τότε η κατανομή που προκύπτει είναι συμμετρική.

Στο πλαίσιο αυτού του κεφαλαίου της διατριβής θα παρατεθούν τα συμπεράσματα των σημαντικότερων συγκριτικών μελετών των διάφορων τρόπων ελέγχου της μονοδιάστατης συμμετρίας γύρω από γνωστό και άγνωστο σημείο. Ειδικότερα, αρχικά, θα παρουσιαστούν τα συμπεράσματα προσομοιώσεων για τους ελέγχους του 2^{ου} Κεφαλαίου όπου το κέντρο συμμετρίας είναι γνωστό και έπειτα, τα αντίστοιχα συμπεράσματα για τους ελέγχους του 3^{ου} Κεφαλαίου όπου το κέντρο συμμετρίας είναι άγνωστο.

4.1 Μέθοδοι ελέγχου συμμετρίας γύρω από γνωστό σημείο

Ένας μεγάλος αριθμός συγκριτικών μελετών έχουν παρουσιασθεί στη βιβλιογραφία, καθώς κάθε νέος έλεγχος που παρουσιάζεται τις περισσότερες φορές συγκρίνεται με τους υπάρχοντες. Στη συνέχεια, θα παρουσιαστούν συνοπτικά τα αποτελέσματα των μελετών που έχουν γίνει από τους McWilliams (1990), Modarres and Gastwirth ((1996), (1998)), Cheng and

Balakrishnan (2004), Baklizi ((2007), (2008)) και τέλος Corzo and Babativa (2011). Για την καλύτερη μελέτη των αποτελεσμάτων στον πίνακα που ακολουθεί δίνεται ο συμβολισμός που θα χρησιμοποιηθεί, η ονομασία κάθε ελέγχου και η ενότητα που παρουσιάστηκε στο Κεφάλαιο 2.

Συμβολισμός	Ονομασία	Ενότητα
T_S	Προσημικός έλεγχος	2.1.1
C_k	Cheng and Balakrishnan (2004)	2.1.2
T'_n	Προσημασμένης διάταξης του Wilcoxon	2.1.3
R^*	McWilliams (1990)	2.2.1
$R^* n_1, n_2$	Baklizi (2003)	2.2.2
M_p	Modarres and Gastwirth (1996)	2.2.3
J_k	Corzo and Babativa (2013)	2.2.4
$L^*_{n,p}, L$	Baklizi (2007)	2.2.5
B_n	Butler (1969)	2.3.1
R_n	Rothman and Woodroffe (1972)	2.3.2
SG_n, S_n	Srinivasan and Godio (1973)	2.3.3
$T_n^{(a)}$	Hill and Rao (1977)	2.3.4
S	Tajuddin (1994)	2.4.1
Z_S, W_p	Modarres and Gastwirth (1998)	2.4.2

Ο McWilliams (1990) σύγκρινε την απόδοση του έλεγχου που πρότεινε με άλλες διαδικασίες ελέγχου που είχαν την ιδιότητα να είναι ανεξάρτητες κατανομής. Ειδικότερα, στη μελέτη προσομοίωσης που διεξήγαγε προσομοιώνοντας 1.000 δείγματα μεγέθους $n = 20, 30, 50$ και 100 από 9 κατανομές της Λάμδα οικογένειας σύγκρινε τις μεθοδολογίες που στηρίζονται στις σ.σ. R^* , B_n , R_n , T_n' και $T_n^{(a)}$. Οι επιλογές των παραμέτρων της Λάμδα οικογένειας που αρχικά υιοθετήθηκαν από τον McWilliams (1990) και μετέπειτα από τους υπόλοιπους συγγραφείς παρατίθενται στον πίνακα που ακολουθεί, με την πρώτη περίπτωση να αντιστοιχεί σε συμμετρική κατανομή και μάλιστα είναι εκείνη που πλησιάζει την κανονική κατανομή. Επιπλέον, για κάθε κατανομή δίνεται ο συντελεστής κύρτωσης (K) και λοξότητας (S).

Πίνακας 4.1.1 Μέλη της Λάμδα οικογένειας και αντίστοιχοι συντελεστές λοξότητας και κύρτωσης που θεωρήθηκαν από τον McWilliams (1990).

Περίπτωση	λ_1	λ_2	λ_3	λ_4	S	K
1	0.000000	0.197454	0.134915	0.134915	0	3
2	-0.116734	-0.351663	-0.130000	-0.160000	0.8	11.4
3	0.000000	-1.000000	-0.100000	-0.180000	2	21.2
4	3.586508	0.043060	0.025213	0.094029	0.9	4.2
5	0.000000	-1.000000	-0.007500	-0.300000	1.5	7.5
6	0.000000	1.000000	1.400000	0.250000	0.5	2.2
7	0.000000	1.000000	0.000070	0.100000	1.5	5.8
8	0.000000	-1.000000	-0.001000	-0.130000	3.16	23.8
9	0.000000	-1.000000	-0.000100	-0.170000	3.88	40.7

Από τη μελέτη αυτή, ο McWilliams (1990) συμπεραίνει ότι:

- όλοι οι έλεγχοι που συμμετείχαν στη συγκεκριμένη μελέτη προσομοίωσης διατηρούν το επίπεδο σημαντικότητας και ότι
- η μέθοδος που ο ίδιος πρότεινε είναι πιο ισχυρή στις περισσότερες από τις περιπτώσεις που θεωρήθηκαν και προτείνεται για τις συγκεκριμένες εναλλακτικές κατανομές.

Οι Modarres and Gastwirth (1996) συγκρίνουν την απόδοση του έλεγχου που πρότειναν με άλλες διαδικασίες ελέγχου που είχαν την ιδιότητα να είναι ανεξάρτητες κατανομής. Ειδικότερα, στη μελέτη προσομοίωσης που διεξήγαγαν προσομοιώνοντας 10.000 δείγματα μεγέθους $n = 20, 30, 50$ και 100 από τις 9 κατανομές του Πίνακα 4.1.1, συγκρίνουν τις μεθοδολογίες που στηρίζονται στις σ.σ. R^* , T_n' και M_p , για $p = 0, 0.1, 0.2, 0.25$. Στο σημείο αυτό πρέπει να επισημανθεί ότι στις συγκρίσεις δε συμμετείχε η μέθοδος της σ.σ. T_s , καθώς από τον Tajuddin (1994) αποδείχθηκε ότι έχει παρόμοια ισχύ με αυτή του έλεγχου του McWilliams (1990). Από τη μελέτη αυτή οι Modarres and Gastwirth (1996) συμπεραίνουν ότι:

- όλοι οι έλεγχοι που συμμετείχαν στη συγκεκριμένη μελέτη προσομοίωσης διατηρούν το επίπεδο σημαντικότητας,
- η μέθοδος που στηρίζεται στη σ.σ. $M_{0.2}$ είναι πιο ισχυρή στις περισσότερες από τις περιπτώσεις που θεωρήθηκαν και προτείνεται για τις συγκεκριμένες εναλλακτικές κατανομές και ότι
- όταν η ασυμμετρία εντοπίζεται σε περιοχές κοντά στη διάμεσο ο έλεγχος των Modarres and Gastwirth (1996) δεν έχει ικανοποιητική ισχύ και προτείνεται ο έλεγχος προσημασμένης διάταξης του Wilcoxon.

Λαμβάνοντας υπόψη τα προηγούμενα αποτελέσματα, οι Cheng and Balakrishnan (2004) συγκρίνουν την απόδοση της μεθοδολογίας που πρότειναν με αυτήν των Modarres and Gastwirth (1996). Για το σκοπό αυτό, προσομοιώνουν 10.000 δείγματα μεγέθους $n = 20, 30, 50$ και 100 από τις 9 κατανομές του Πίνακα 4.1.1 και συγκρίνουν την απόδοση των C_6 και M_p , για $p = 0, 0.1, 0.2, 0.25$. Από τη μελέτη αυτή οι Cheng and Balakrishnan (2004) συμπεραίνουν ότι:

- οι έλεγχοι που συμμετείχαν στη συγκεκριμένη μελέτη προσομοίωσης διατηρούν το επίπεδο σημαντικότητας,
- η μέθοδος που στηρίζεται στη σ.σ. C_6 είναι πιο ισχυρή στις περισσότερες περιπτώσεις που θεωρήθηκαν και προτείνεται για τις συγκεκριμένες εναλλακτικές κατανομές. Ειδικότερα, είναι πάντοτε ισχυρότερη για $n \geq 50$ και σε αρκετές περιπτώσεις η διαφορά στην ισχύ είναι μεγάλη.

Ο Baklizi (2007) συγκρίνει τον έλεγχο που πρότεινε και στηρίζεται στη σ.σ. $L_{n,p}^*$, καθώς και τον τροποποιημένο έλεγχο με τη σ.σ. L , με τους ελέγχους που στηρίζονται στις σ.σ. R^* , $M_{0.6}$ και $W_{0.7}$, με τον τελευταίο έλεγχο να είναι αυτός των δύο βημάτων που προτάθηκε από τους Modarres and Gastwirth (1998) με επίπεδο σημαντικότητας του πρώτου βήματος $\alpha_1 = \alpha/5$. Για το σκοπό αυτό, προσομοίωσε 5.000 δείγματα μεγέθους $n = 20, 30, 50$ και 100 από τις προαναφερθείσες 9 κατανομές της Λάμδα οικογένειας και οδηγήθηκε στα ακόλουθα συμπεράσματα:

- οι έλεγχοι που συμμετείχαν στη συγκεκριμένη μελέτη προσομοίωσης διατηρούν το επίπεδο σημαντικότητας,
- όλοι οι έλεγχοι έχουν αρκετά μεγαλύτερη ισχύ από τον έλεγχο του McWilliams (1990),

- η μέθοδος που στηρίζεται στη σ.σ. $L_{n,0.8}^*$ κυριαρχεί της μεθόδου που στηρίζεται στη σ.σ. L σε όλες τις περιπτώσεις,
- η μέθοδος $L_{n,0.8}^*$ είναι ανταγωνιστική ως προς τις $W_{0.7}$ και $M_{0.6}$, χωρίς ωστόσο κάποια να μπορεί να θεωρηθεί καλύτερη των άλλων.

Όπως είδαμε στο Κεφάλαιο 2, πρόσφατα οι Corzo and Babativa (2013) πρότειναν έναν νέο τρόπο ελέγχου της συμμετρίας γύρω από γνωστό σημείο. Στο πλαίσιο αυτό, διεξήγαγαν μια εκτενή μελέτη προσομοίωσης για να επιχειρηματολογήσουν ότι η μέθοδος που προτείνουν μπορεί να προστεθεί στον κατάλογο των ήδη υπαρχουσών. Έτσι, συγκρίνουν τους ελέγχους που στηρίζονται στις σ.σ.: J_6 , R^* , $R^*|n_1, n_2, T_n', S, C_6, M_p$ (για $p=0.2, 0.25$ και 0.6), $L_{n,p}^*$ (για $p=0.8$), L και W_p (για $p=0.7$), με το W_p να είναι η τροποποίηση του Baklizi (2008) που περιλαμβάνει μόνο το δεύτερο βήμα του ελέγχου των δύο βημάτων των Modarres and Gastwirth (1998). Για το σκοπό αυτό, προσομοίωσαν 10.000 δείγματα μεγέθους $n=20, 30, 50$ και 100 από τις προαναφερθείσες 9 κατανομές της Λάμδα οικογένειας και οδηγήθηκαν στα ακόλουθα συμπεράσματα:

- οι έλεγχοι που συμμετείχαν στη συγκεκριμένη μελέτη προσομοίωσης διατηρούν το επίπεδο σημαντικότητας,
- για μέγεθος δείγματος $n=20$, όταν τα δεδομένα προσομοιώνονται από τις περιπτώσεις 2-5 οι έλεγχοι W_p και J_6 είναι οι καλύτεροι, ενώ όταν η δειγματοληψία γίνεται από τις κατανομές των περιπτώσεων 6-9 ο έλεγχος της σ.σ. J_6 κυριαρχεί των υπολοίπων,
- για μέγεθος δείγματος $n=30$, όλοι οι έλεγχοι μπορούν να θεωρηθούν ισοδύναμοι,
- όταν τα δεδομένα προσομοιώνονται από την περίπτωση 3, ο έλεγχος W_p έχει τη μεγαλύτερη ισχύ, με τους ελέγχους $L_{n,0.8}^*$ και J_6 να ακολουθούν,

- όταν τα δεδομένα προσομοιώνονται από τις περιπτώσεις 4 και 5 οι έλεγχοι W_p και J_6 έχουν τη μεγαλύτερη ισχύ,
- για μέγεθος δείγματος $n=50$, όταν τα δεδομένα προσομοιώνονται από τις περιπτώσεις 2-5 οι έλεγχοι W_p , J_6 και C_6 έχουν τη μεγαλύτερη ισχύ, ενώ στις περιπτώσεις 6-9 κυριαρχούν οι έλεγχοι J_6 και C_6 ,
- για $n=100$ τη μεγαλύτερη ισχύ την έχουν οι έλεγχοι W_p , J_6 και C_6 .

Συνοψίζοντας,

α) για μικρά μεγέθη δείγματος, προτείνονται να εφαρμόζονται οι έλεγχοι των σ.σ. $W_{0.7}$, J_6 , $L_{n,0.8}^*$, και M_p (για $p=0.2$ ή 0.6).

β) Για μεγαλύτερα σε μέγεθος σύνολα δεδομένων, προτείνεται επιπλέον ο έλεγχος της σ.σ. C_6 , ενώ δεν προτείνεται ο έλεγχος της σ.σ. M_p .

γ) Δεν συνιστώνται οι έλεγχοι R^* , $R^*|n_1, n_2$, B_n , T_n' , R_n , $T_n^{(a)}$ και S .

4.2 Μέθοδοι ελέγχου συμμετρίας γύρω από άγνωστο σημείο

Ένας μεγάλος αριθμός συγκριτικών μελετών έχουν παρουσιασθεί στη βιβλιογραφία, καθώς κάθε νέος έλεγχος συμμετρίας γύρω από άγνωστο σημείο που παρουσιάζεται τις περισσότερες φορές συγκρίνεται με τους υπάρχοντες μέσω προσομοιώσεων. Στη συνέχεια, θα παρουσιαστούν συνοπτικά τα αποτελέσματα των μελετών που έχουν γίνει από τους Randles *et al.* (1980), Cabilio and Masaro (1996), Mira (1999), Ghosh (2011) και τέλος τους Wang *et al.* (2013). Αξίζει να αναφερθεί ότι η τελευταία εργασία είναι η πιο πλήρης καθώς συγκρίνει

την απόδοση των περισσότερων μεθοδολογιών, προσομοιώνοντας δεδομένα από ευρύτερο φάσμα κατανομών. Στον πίνακα που ακολουθεί δίνεται ο συμβολισμός κάθε ελέγχου και η ενότητα που παρουσιάστηκε στο Κεφάλαιο 3.

Συμβολισμός	Ονομασία	Ενότητα
T^*, T	Τροποποίηση προσημικού ελέγχου	3.1.4
T_N, S_N	Bhattacharya <i>et al.</i> (1982)	3.1.5
$S_{4,\varepsilon}, T_2(J)$	Antille and Kersting (1977)	3.1.6
$T_{4,0}, T_{4,\varepsilon}$	Ekström and Jammalamadaka (2006)	3.1.6
$b_1, Z_{b_1}^*$	Δειγματικός συντελεστής λοξότητας	3.2.1
S_K, S_K^*	Cabilio and Masaro (1996)	3.2.2
T_{Mi}	Miao <i>et al.</i> (2006)	3.2.2
$\hat{\tau}$	Ghosh (2011)	3.2.3
T_{B_n}	Boos (1982)	3.2.5
S_{GMD}, RS_{GMD}	Wang <i>et al.</i> (2013)	3.2.5
$\gamma_1(F)$	Mira (1999)	3.2.6
T_{R1}, T_{R2}	Randles <i>et al.</i> (1980)	3.4.1
$\hat{J}_n, T_1(\alpha)$	Gupta (1967), Antille <i>et al.</i> (1982)	3.4.2
$V, V_A, T_3(J)$	Finch (1977), Antille <i>et al.</i> (1982)	3.4.3

Οι Randles *et al.* (1980) θέλησαν να συγκρίνουν το triple test των σ.σ. T_{R1} και T_{R2} με τις μεθοδολογίες που στηρίζονται στις σ.σ. b_1 , $Z_{b_1}^*$ και $\hat{J}_n$. Στον έλεγχο $\hat{J}_n$, καθώς η εκτίμηση της ασυμπτωτικής διακύμανσης που προτάθηκε από τον Gupta (1967) χρησιμοποιεί το μήκος ενός διαστήματος εμπιστοσύνης, οι Randles *et al.* (1980) εφάρμοσαν τον έλεγχο χρησιμοποιώντας για μεγέθη δείγματος $n=20$, 80% και 95% διαστήματα εμπιστοσύνης. Έπειτα, παρατηρώντας ότι δεν υπάρχουν σημαντικές διαφορές στην ισχύ, για $n=30$ χρησιμοποίησαν μόνο τον έλεγχο που προκύπτει από το 95% διάστημα εμπιστοσύνης. Επιπλέον, στη συγκριτική μελέτη περικλείουν τον έλεγχο T_{R1} χρησιμοποιώντας τις κρίσιμες τιμές της t -κατανομής με n βαθμούς ελευθερίας, με τον έλεγχο αυτό σε όσα ακολουθούν να αναφέρεται ως T_{R1}^* .

Οι Randles *et al.* (1980) για τη σύγκριση των παραπάνω μεθοδολογιών ως προς τη διατήρηση του επιπέδου σημαντικότητας και την απόδοση ισχύος, διεξάγουν μια Monte Carlo μελέτη εφαρμόζοντας τις μεθόδους σε ένα ευρύ φάσμα συμμετρικών και μη συμμετρικών δεδομένων για διάφορα μεγέθη δείγματος. Ειδικότερα, προσομοιώνουν δεδομένα από 20 κατανομές, εκ των οποίων οι 14 είναι μέλη της Γενικευμένης Λάμδα οικογένειας κατανομών, με τιμές των παραμέτρων που παρατίθενται στον Πίνακα 4.2.1 που ακολουθεί, ενώ οι υπόλοιπες 6 προκύπτουν από τρία μέλη της Γενικευμένης Λάμδα οικογένειας κατανομών θέτοντας

$$Y = \begin{cases} X, & \text{αν } X \leq 0 \\ RX, & \text{αν } X > 0, \end{cases}$$

όπου X μια τ.μ. από την Γενικευμένη Λάμδα οικογένεια κατανομών συμμετρική γύρω από το μηδέν, R μια σταθερά με $R > 1$ και Y η τ.μ. που προκύπτει και έχει την επιθυμητή κατανομή. Οι παράμετροι αυτών των κατανομών δίνονται στον Πίνακα 4.2.2.

Παρατήρηση 4.2.1 Οι Randles *et al.* (1980) επισημαίνουν ότι:

α) οι κατανομές 1-6 είναι συμμετρικές (αφού $\lambda_3 = \lambda_4$), με την τιμή του συντελεστή κύρτωση να αυξάνεται, β) οι κατανομές 7-14 είναι λοξές, με διαφορετικό συντελεστή λοξότητας και αύξουσα τιμή του συντελεστή κύρτωσης, ενώ γ) οι κατανομές 15-20 είναι κατανομές για τις οποίες το δεξί μέρος της κατανομής είναι στοχαστικά μεγαλύτερο από το αριστερό.

Πίνακας 4.2.1 Μέλη της Λάμδα οικογένειας και αντίστοιχοι συντελεστές λοξότητας S και κύρτωσης K που θεωρήθηκαν από τους Randles *et al.* (1980).

Περίπτωση	λ_1	λ_2	λ_3	λ_4	S	K
1	0	2.000000	1.000000	1.000000	0	1.8
2	0	0.197454	0.134915	0.134915	0	3.0
3	0	-1.000000	-0.080000	-0.080000	0	6.0
4	0	-0.397912	-0.160000	-0.160000	0	11.6
5	0	-1.000000	-0.240000	-0.240000	0	126.0
6	0	-5.112560	-1.000000	-1.000000	-	-
7	0	1.000000	1.400000	0.250000	0.5	2.2
8	0	1.000000	0.000070	0,100000	1.5	5.8
9	3.586508	0.043060	0.025213	0.094029	0.9	4.2
10	0	-1.000000	-0.007500	-0.030000	1.5	7.5
11	-0.116734	-0.351663	-0.130000	-0.160000	0.8	11.4
12	0	-1.000000	-0.100000	-0.180000	2.0	21.2
13	0	-1.000000	-0.001000	-0.130000	3.16	23.8
14	0	-1.000000	-0.000100	-0.170000	3.88	40.7

Πίνακας 4.2.2 Ασύμμετρες κατανομές που θεωρήθηκαν από τους Randles *et al.* (1980).

Περίπτωση	λ_1	λ_2	λ_3	λ_4	R
15	0	2.000000	1.000000	1.000000	2
16	0	2.000000	1.000000	1.000000	4
17	0	0.197454	0.134915	0.134915	2
18	0	0.197454	0.134915	0.134915	4
19	0	-5.112560	-1.000000	-1.000000	2
20	0	-5.112560	-1.000000	-1.000000	4

Οι Randles *et al.* (1980) θέλοντας να συγκρίνουν την απόδοση των μεθοδολογιών που αναφέρθηκαν ως προς το επίπεδο σημαντικότητας και την ισχύ, προσομοιώνουν 1.000.000 δείγματα μεγέθους $n = 20$ και 30 από τις προαναφερθείσες 20 κατανομές. Από τη μελέτη αυτή οι Randles *et al.* (1980) συμπεραίνουν ότι:

- Οι έλεγχοι που στηρίζονται στις σ.σ. $Z_{b_1}^*$, T_{R2} και $\hat{J}$ (με το 95% διάστημα εμπιστοσύνης), δεν διατηρούν το επίπεδο σημαντικότητας, με τον τελευταίο σε αντίθεση με τους δύο πρώτους να είναι συντηρητικός. Επομένως, αυτοί οι τρεις τρόποι ελέγχου δεν θα συμπεριληφθούν στη μελέτη της απόδοσης ισχύος και δεν προτείνονται για χρήση. Αντιθέτως, οι έλεγχοι που στηρίζονται στις σ.σ. T_{R1} , T_{R1}^* και Z_{b_1} διατηρούν αποτελεσματικά το επίπεδο σημαντικότητας 0.05 για όλες τις συμμετρικές κατανομές που θεωρήθηκαν.
- Ο έλεγχος του Triple test T_{R1} είναι αποτελεσματικό εργαλείο για την ανίχνευση αποκλίσεων από τη συμμετρία, με την προσαρμογή του

να είναι καλύτερη σε σχέση με τους υπόλοιπους ελέγχους που θεωρήθηκαν.

- Η εφαρμογή του ελέγχου T_{R1}^* παρέχει επιπλέον προστασία του επιπέδου σημαντικότητας α , προκαλώντας μικρή μείωση στην ισχύ του ελέγχου της σ.σ. T_{R1} . Για τον παραπάνω λόγο προτείνεται η χρήση του.

Παρακινούμενοι από τα αποτελέσματα της μελέτης προσομοίωσης των Randles *et al.* (1980), οι Cabilio and Masaro (1996) αρχικά σύγκριναν τον έλεγχο που στηρίζεται στη σ.σ. S_K με τον έλεγχο της σ.σ. T_{R1} και την τροποποίηση αυτού T_{R1}^* . Για το σκοπό αυτό, προσομοιώνουν 1.000 δείγματα μεγέθους $n=20$ και 30 από τις προαναφερθείσες 20 κατανομές. Από τη μελέτη αυτή οι Cabilio and Masaro (1996) συμπεραίνουν ότι:

- Ο έλεγχος της σ.σ. S_K διατηρεί το επίπεδο σημαντικότητας $\alpha=0.05$, πλην της περίπτωσης δειγματοληψίας από την ομοιόμορφη κατανομή (Περίπτωση 1).
- Στην περίπτωση δειγματοληψίας από τις ασύμμετρες κατανομές των Περιπτώσεων 7-14, ο έλεγχος T_{R1}^* φαίνεται να αποδίδει καλύτερα σε σχέση με τους άλλους δύο.
- Στην περίπτωση δειγματοληψίας από τις ασύμμετρες κατανομές των Περιπτώσεων 15-20, φαίνεται να αποδίδει καλύτερα ο έλεγχος που στηρίζεται στη σ.σ. S_K .

Έπειτα, οι Cabilio and Masaro (1996) αφού απέδειξαν ότι ο έλεγχος τους είναι ανταγωνιστικός με τον έλεγχο των Randles *et al.* (1980), διεξάγουν μελέτη προσομοίωσης χρησιμοποιώντας τα ίδια μεγέθη δείγματος και τις ίδιες κατανομές με πριν, για τη σύγκριση του ελέγχου S_K με τους ελέγχους T_{B_n} , S_K^* , $T_1(0)$, $T_1\left(\frac{1}{6}\right)$, $T_2(0.05)$ και $T_3(0.05)$, με

την τιμή εντός της παρένθεσης να δηλώνει την τιμή της παραμέτρου. Από τη μελέτη αυτή οι Cabilio and Masaro (1996) συμπεραίνουν ότι:

- Στην περίπτωση δειγματοληψίας από την ομοιόμορφη Κατανομή (Περίπτωση 1), ο έλεγχος $T_2(0.05)$ διατηρεί καλύτερα από τους υπόλοιπους ελέγχους το επίπεδο σημαντικότητας $\alpha = 0.05$. Όμως, ο έλεγχος αυτός έχει την ιδιότητα να έχει τη μικρότερη ισχύ όταν τα προσομοιωμένα δεδομένα προέρχονται από ασύμμετρες εναλλακτικές κατανομές.
- Στην περίπτωση δειγματοληψίας από την κατανομή Cauchy (Περίπτωση 6), προκύπτει ότι για δείγματα μεγέθους $n=20$, οι εμπειρικές πιθανότητες σφάλματος τύπου I του ελέγχου $T_2(0.05)$ είναι μεγαλύτερες από τις αντίστοιχες του ελέγχου S_K , ενώ για $n=30$ είναι μικρότερες.
- Στην περίπτωση όπου τα δεδομένα προέρχονται από την ομοιόμορφη κατανομή, ο έλεγχος $T_3(0.05)$ έχει μεγαλύτερη ισχύ από τους ελέγχους T_{B_n} και S_K και διατηρεί καλύτερα το επίπεδο σημαντικότητας. Στην περίπτωση αυτή το επίπεδο σημαντικότητας του ελέγχου T_{B_n} είναι μικρότερο από αυτό του ελέγχου S_K .
- Οι έλεγχοι T_{B_n} και T_{R1}^* έχουν παρόμοια συμπεριφορά στις Περιπτώσεις 7-14, αλλά ο έλεγχος T_{B_n} έχει μεγαλύτερη ισχύ στις Περιπτώσεις 15-20.
- Στις περιπτώσεις δειγματοληψίας από τις κατανομές 7-14, ο έλεγχος T_{B_n} έχει ελαφρώς μεγαλύτερη ισχύ από τον έλεγχο S_K , με τους ελέγχους να έχουν παρόμοια προσαρμογή όταν τα δεδομένα προέρχονται από τις κατανομές 15-18. Ο έλεγχος T_{B_n} έχει αρκετά μεγαλύτερη ισχύ στις Περιπτώσεις 19 και 20.

- Η ισχύς των ελέγχων S_K και $T_3(0.05)$ δεν διαφέρει αρκετά, εκτός από τις Περιπτώσεις 19 και 20 όπου η ισχύς του ελέγχου S_K είναι μεγαλύτερη.

Η Mira (1999) σύγκρινε την απόδοση του ελέγχου της σ.σ. $\gamma_1(F)$ με τον έλεγχο της σ.σ. S_K . Ειδικότερα, στη μελέτη προσομοίωσης που διεξήγαγε προσομοιώνοντας 100.000 δείγματα μεγέθους $n=20$ και 30 από τις κατανομές του Πίνακα 4.2.1, σύγκρινε τις μεθοδολογίες των σ.σ. $\gamma_1(F)$ και S_K και κατέληξε στα ακόλουθα συμπεράσματα:

- Για δείγματα μεγέθους $n=20$, στις Περιπτώσεις 1,2,3 ο έλεγχος $\gamma_1(F)$ διατηρεί το επίπεδο σημαντικότητας εξίσου καλά με τον έλεγχο της σ.σ. S_K και στις Περιπτώσεις 4,5,6 τον ξεπερνά.
- Για $n=30$, οι δύο έλεγχοι έχουν παρόμοια προσαρμογή εκτός από τις Περιπτώσεις 1 και 6 όπου ο έλεγχος $\gamma_1(F)$ διατηρεί καλύτερα το επίπεδο σημαντικότητας.
- Για δειγματοληψία από τις κατανομές των Περιπτώσεων 7-14 και $n \leq 20$ ο έλεγχος S_K έχει μεγαλύτερη ισχύ. Ωστόσο, η αποτελεσματικότητα του ελέγχου $\gamma_1(F)$ αυξάνεται γοργά με την αύξηση του μεγέθους δείγματος και για $n \geq 30$ ο έλεγχος $\gamma_1(F)$ έχει συγκρίσιμη ή μεγαλύτερη ισχύ από τον έλεγχο S_K σε όλες τις εναλλακτικές κατανομές, εκτός της Περίπτωσης 7.

Ο Ghosh (2011) σύγκρινε την απόδοση του ελέγχου $\hat{\tau}$ που πρότεινε με αυτήν των ελέγχων των σ.σ. $\gamma_1(F)$, S_K και T_{Mi} . Ειδικότερα, παρακινούμενος από τους Randles *et al.* (1980) υπολόγισε

τις κρίσιμες τιμές των ελέγχων εμπειρικά, με 100.000 προσομοιώσεις από την κανονική κατανομή και έπειτα τις χρησιμοποίησε για τις συγκρίσεις ισχύος. Στη μελέτη προσομοίωσης που διεξήγαγε, προσομοιώνοντας 100.000 δείγματα μεγέθους $n=30,50,100$ και 500 από τις κατανομές του Πίνακα 4.2.1, σύγκρινε τους παραπάνω ελέγχους. Από τη μελέτη αυτή ο Ghosh (2011) συμπεραίνει ότι:

- Στην περίπτωση δειγματοληψίας από την Ομοιόμορφη Κατανομή (Περίπτωση 1), προκύπτει ότι ο έλεγχος $\hat{\tau}$ είναι ο μόνος που διατηρεί το επίπεδο σημαντικότητας $\alpha=0.05$.
- Για δειγματοληψίες από τις κατανομές των Περιπτώσεων 3-5, ο μόνος έλεγχος που διατηρεί το επίπεδο σημαντικότητας είναι ο έλεγχος $\gamma_1(F)$.
- Για δειγματοληψίες από την Περίπτωση 6, όλοι οι έλεγχοι αποτυγχάνουν να διατηρήσουν το επίπεδο σημαντικότητας.
- Για δειγματοληψίες από την Περίπτωση 11, όλοι οι έλεγχοι έχουν αρκετά μικρή ισχύ, ακόμη και για μεγάλα μεγέθη δείγματος.
- Στις Περιπτώσεις 9-12, ο έλεγχος $\hat{\tau}$ έχει τη μεγαλύτερη ισχύ.
- Στις Περιπτώσεις 7-14, ο έλεγχος $\gamma_1(F)$ έχει τη μικρότερη ισχύ για όλα τα μεγέθη δείγματος.

Στη συνέχεια θα παρουσιασθεί η μελέτη προσομοίωσης των Wang *et al.* (2013). Πρόκειται για μια από τις πιο πρόσφατες μελέτες που έχουν εμφανιστεί στη βιβλιογραφία αλλά και μια από τις πιο γενικές αφού συγκρίνει τις περισσότερες από τις μεθόδους που έχουν παρουσιασθεί στη βιβλιογραφία, προσομοιώνοντας δεδομένα από ευρύτερο φάσμα κατανομών. Η μελέτη αυτή διεξήχθη με σκοπό να συγκριθούν οι έλεγχοι των σ.σ. S_{GMD} , RS_{GMD} , T_{R1} , $\hat{J}_n$, $T_1\left(\frac{1}{6}\right)$, $T_2(J)$, $T_{4,0}$, $T_3(J)$, b_1 , T_{B_n} , T , S_K και τέλος των T_N και S_N . Οι δύο τελευταίες σ.σ. χρησιμοποιούνται για τις ειδικές περιπτώσεις που $q=1$ και $q=0.25$.

Στο σημείο αυτό επισημαίνεται ότι για $q=1$ οι σ.σ. συμβολίζονται με T_{N2} και S_{N2} . Οι κρίσιμες τιμές των ελέγχων υπολογίστηκαν εμπειρικά με 20.000 προσομοιώσεις από την κανονική κατανομή.

Οι Wang *et al.* (2013) θέλοντας να συγκρίνουν την απόδοση των παραπάνω ελέγχων, αρχικά, ως προς το επίπεδο σημαντικότητας, προσομοίωσαν 10.000 δείγματα μεγέθους $n=20, 50$ και 100 από μια μεγάλη κλάση συμμετρικών κατανομών. Ειδικότερα, θεώρησαν τα μέλη της Γενικευμένης Λάμδα οικογένειας που αντιστοιχούν στις Περιπτώσεις 1-6 του Πίνακα 4.2.1 καθώς και τις κατανομές που παρατίθενται στον Πίνακα 4.2.3.

Από τη μελέτη αυτή οι Wang *et al.* (2013) συμπεραίνουν ότι:

- Το επίπεδο σημαντικότητας για τους ελέγχους των σ.σ. S_{GMD} , T_N , S_N , S_{N2} , T , W , $T_1\left(\frac{1}{6}\right)$, S_K , T_{B_n} , $T_{4,0}$, b_1 και T_{R1} εξαρτάται από το είδος της κατανομής.
- Ο έλεγχος T_{N2} διατηρεί το επίπεδο σημαντικότητας για μέγεθος δείγματος τουλάχιστον $n=50$, ενώ για $n=100$ αυτό συμβαίνει για τον έλεγχο $T_2(J)$.
- Οι έλεγχοι των σ.σ. RS_{GMD} , $T_2(J)$, $T_3(J)$, S_K , T_{N2} και $T_{4,0}$ έχουν επίπεδο σημαντικότητας με τιμή αρκετά κοντά στο 0.05, ανεξάρτητα από την κατανομή και το μέγεθος δείγματος.

Πίνακας 4.2.3 Συμμετρικές κατανομές που θεωρήθηκαν από τους Wang *et al.* (2013).

Κατανομή-Συμβολισμός	Σ.Π.Π.
Τυπική κανονική κατανομή $N(0,1)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in \mathbb{R}$
Ομοιόμορφη κατανομή $U(0,1)$	$\frac{1}{b-a}, a < x < b$
Βήτα(a, a), a = 3, 5.	$\frac{x^{a-1}(1-x)^{a-1}}{B(a,a)}, 0 < x < 1$
Λογιστική κατανομή $Logistic(\mu, \sigma)$	$\frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma \left(1 + e^{-\frac{x-\mu}{\sigma}}\right)}, -\infty < x < +\infty$
Laplace κατανομή $L(\mu, \lambda)$	$\frac{\lambda}{2} e^{-\lambda x-\mu }, -\infty < x < +\infty$
t – κατανομές $t_v, v = 3, 6$	$\frac{N(0,1)}{\sqrt{\chi_v^2 / v}}, -\infty < x < +\infty$
Cauchy $C(0,1)$	$\frac{1}{\pi(1+x^2)}, -\infty < x < +\infty$
Slash Distribution	$\frac{\varphi(0) - \varphi(x)}{x^2}, \text{ όπου } \varphi(\cdot) \text{ η σ.π.π. της } N(0,1)$
Μίξη κανονικών	$0.5N(-\mu, 1) + 0.5N(\mu, 1)$

Από τα παραπάνω προκύπτει το συμπέρασμα ότι μόνο οι έλεγχοι των σ.σ. RS_{GMD} , $T_2(J)$, $T_3(J)$, S_K , T_{N2} και $T_{4,0}$ πρέπει να συμπεριληφθούν στη μελέτη προσομοίωσης και σύγκρισης ως προς την απόδοση ισχύος. Στο πλαίσιο αυτό, οι Wang *et al.* (2013) προσομοίωσαν 10.000 δείγματα μεγέθους $n=20,50$ και 100 από τις ασύμμετρες κατανομές των Πινάκων 4.2.1 και 4.2.2 καθώς επίσης και από τις κατανομές του πίνακα που ακολουθεί. Οι συντελεστές κύρτωσης και λοξότητας των κατανομών του Πίνακα 4.2.4 δίνονται στον Πίνακα 4.2.5.

Πίνακας 4.2.4 Επιπρόσθετες ασύμμετρες κατανομές που θεωρήθηκαν από τους Wang *et al.* (2013).

Κατανομή- Συμβολισμός	Σ.Π.Π.
Γάμμα($a,1$)	$\frac{x^{a-1}e^{-x}}{\Gamma(a)}, x \geq 0$
Βήτα(a,b)	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, 0 < x < 1$
$SN(\lambda)$	$2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Phi(\lambda x), x \in \mathbb{R}$ όπου $\Phi(\cdot)$ η α.σ.κ. της $N(0,1)$

Πίνακας 4.2.5 Συντελεστές κύρτωσης και λοξότητας των κατανομών των Wang *et al.* (2013).

Κατανομή	Συντελεστής Λοξότητας	Συντελεστής Κύρτωσης
Γάμμα(16,1)	0.5	3.375
Γάμμα(4,1)	1	4.5
Γάμμα(1,1)	2	9
Γάμμα(0.25,1)	4	27
$SN(2.1738)$	0.99	3.347
Βήτα(1,1.837)	0.5	2.283
Βήτα(1,3.698)	1	3.576
Βήτα(0.5,5.552)	2	7.674
Βήτα(1,2.575)	4	21.714

Από τη μελέτη αυτή Wang *et al.* (2013) συμπεραίνουν ότι:

- Για μικρά μεγέθη δείγματος, όπως $n=20$, οι έλεγχοι T_{N_2} και $T_{4,0}$ έχουν τη μεγαλύτερη ισχύ αλλά δεν διατηρούν το επίπεδο σημαντικότητας. Επιπλέον, καθώς το μέγεθος του δείγματος αυξάνεται, η ισχύς του ελέγχου T_{N_2} μειώνεται.
- Για δείγματα μεγέθους $n=50$, οι έλεγχοι RS_{GMD} , T_{N_2} και $T_2(J)$ έχουν τη μεγαλύτερη ισχύ.

- Για δείγματα μεγέθους $n=100$, ο καλύτερος έλεγχος από άποψη ισχύος είναι ο RS_{GMD} με τους ελέγχους T_{N2} , $T_2(J)$ και $T_3(J)$ να ακολουθούν.
- Σε όλες τις περιπτώσεις που θεωρήθηκαν η ισχύς του ελέγχου RS_{GMD} είναι ελαφρώς μεγαλύτερη από αυτή του ελέγχου $T_3(J)$.
- Ο έλεγχος S_K δεν προτείνεται για χρήση καθώς εξαρτάται από την δειγματική μέση τιμή $\bar{X}$ και την τυπική απόκλιση S και δεν είναι ανθεκτικός στην ύπαρξη ακραίων τιμών.

Συνοψίζοντας,

α) για μικρά μεγέθη δείγματος, προτείνονται να εφαρμόζονται οι έλεγχοι των σ.σ. RS_{GMD} , $T_3(J)$, $\hat{\tau}$ και $T_{4,0}$.

β) Για μεγαλύτερα σε μέγεθος σύνολα δεδομένων, προτείνεται η εφαρμογή των ελέγχων των σ.σ. RS_{GMD} , $T_2(J)$, $T_3(J)$, $\hat{\tau}$ και T_{N2} .

γ) Δεν συνιστώνται οι έλεγχοι των σ.σ. S_{GMD} , T_N , S_N , S_{N2} , T , T_{R1} , T_{R2} ,

$\hat{J}_n$, b_1 , T_{B_n} , S_K , S_K^* , $T_2(J)$, $\gamma_1(F)$ και $T_1\left(\frac{1}{6}\right)$.

ΚΕΦΑΛΑΙΟ 5

Εφαρμογές

Αντικείμενο αυτού του κεφαλαίου της διατριβής είναι η επεξήγηση και η εφαρμογή των πιο ευρέως χρησιμοποιούμενων και αποτελεσματικότερων μεθόδων για τον έλεγχο της συμμετρίας γύρω από γνωστό ή άγνωστο σημείο που παρουσιάστηκαν στα Κεφάλαια 2 και 3, αντίστοιχα. Καθώς στα διάφορα στατιστικά πακέτα και προγράμματα υπάρχουν διαθέσιμες ελάχιστες μεθοδολογίες ελέγχου της συμμετρίας, για την εφαρμογή των υπολοίπων δημιουργήθηκαν κώδικες στη Matlab και χρησιμοποιήθηκαν κώδικες της R, οι οποίοι δίνονται στο Παράρτημα Β.

5.1 Ελέγχων συμμετρίας γύρω από γνωστό σημείο

Στο πλαίσιο της παρούσας ενότητας θα υλοποιηθούν οι συνηθέστερα χρησιμοποιούμενες τεχνικές ελέγχου της συμμετρίας γύρω από γνωστό σημείο. Όπως έχουμε δει, ο έλεγχος αυτός μπορεί να αναχθεί σε έλεγχο συμμετρίας γύρω από το μηδέν έχοντας ως εναλλακτική την υπόθεση της ασυμμετρίας. Τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από την εργασία των Ury and Forrester (1970) και είναι οι διαφορές τις αντοχής σε κλειστά τραύματα και σε τραύματα με ράμματα για 10 τρωκτικά. Οι μετρήσεις αυτές είναι οι ακόλουθες:

-63, 64, 70, 96, 202, 207, -213, 245, 397, 442.

Επισημαίνεται ότι σκοπός μας στην ενότητα αυτή είναι η επεξήγηση και υλοποίηση των μεθοδολογιών.

Ενότητα 2.1.1

Αρχικά θα πρέπει να υπολογιστεί η τιμή της σ.σ. T_S της σχέσης (2.1.1.1). Παρατηρούμε ότι στο δείγμα υπάρχουν 8 θετικές παρατηρήσεις και επομένως $T_S = 8$. Η H_0 απορρίπτεται σε επίπεδο σημαντικότητας 0.022 έναντι της εναλλακτικής ότι η F είναι ασύμμετρη, αν $T_S \leq 1$ ή $T_S \geq 9$, καθώς $\sum_{t=0}^1 \binom{10}{t} 0.5^{10} = 0.011$ και $\sum_{t=9}^{10} \binom{10}{t} 0.5^{10} = 0.011$. Για την εφαρμογή του προσεγγιστικού ελέγχου θα πρέπει να υπολογιστεί η τιμή της σ.σ. Z_S της σχέσης (2.1.1.2). Για $n=10$ προκύπτει $Z_S = 0.5270$ και έτσι για επίπεδο σημαντικότητας $\alpha = 0.05$, η υπόθεση της συμμετρίας γύρω από το μηδέν δεν απορρίπτεται με το συγκεκριμένο έλεγχο.

Ενότητα 2.1.2

Για την εφαρμογή του ελέγχου των Cheng and Balakrishnan (2004) απαιτείται η εύρεση της τιμής της σ.σ. C_k της σχέσης (2.1.2.2), για $k=6$. Στον Πίνακα 5.1.1 δίνονται τα $X_{(1)}, X_{(2)}, \dots, X_{(10)}$ που είναι οι διατεταγμένες τιμές με βάση την απόλυτη τιμή, διατηρώντας τα πρόσημα και οι δείκτριες μεταβλητές $S_1, S_2, \dots, S_{10}$, οι οποίες προσδιορίζουν το πρόσημο των $X_{(i)}$.

Πίνακας 5.1.1

$X_{(i)}$	-63	64	70	96	202	207	-213	245	397	442
S_i	0	1	1	1	1	1	0	1	1	1

Από το άθροισμα των 6 τελευταίων στοιχείων της δεύτερης γραμμής του παραπάνω Πίνακα προκύπτει $C_6 = 5$. Επομένως, για επίπεδο σημαντικότητας 0.032, απορρίπτεται η H_0 έναντι της εναλλακτικής ότι η F είναι ασύμμετρη, αν $C_6 = 0$ ή $C_k = 6$, καθώς $\sum_{t=0}^0 \binom{6}{t} 0.5^6 = 0.016$ και $\sum_{t=6}^6 \binom{6}{t} 0.5^6 = 0.016$. Επιπλέον, από τις σχέσεις (2.1.2.4) και (2.1.2.5) προκύπτουν $Z_{CB} = 0.2449$ και $Z'_{CB} = 0.8165$, αντίστοιχα. Άρα, ο συγκεκριμένος έλεγχος δεν απορρίπτει την υπόθεση της συμμετρίας για $\alpha = 0.05$, καθώς $|Z_{CB}| < z_{\alpha/2} = 1.96$.

Ενότητα 2.1.3

Θα υπολογιστεί η τιμή της σ.σ. T_n' του προσημασμένης διάταξης ελέγχου Wilcoxon. Στον Πίνακα 5.1.2 δίνονται η τάξη R_i^+ του $|X_i|$ στην ακολουθία $|X_1|, |X_2|, \dots, |X_{10}|$ καθώς και το $\text{sign}(X_i)$.

Πίνακας 5.1.2

$ X_i $	63	64	70	96	202	207	213	245	397	442
R_i^+	1	2	3	4	5	6	7	8	9	10
$\text{sign}(X_i)$	-1	1	1	1	1	1	-1	1	1	1

Πολλαπλασιάζοντας τα στοιχεία των δύο τελευταίων γραμμών του παραπάνω Πίνακα και αθροίζοντας προκύπτει $T'_n = 39$. Για επίπεδο σημαντικότητας α , η υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της γενικής ασυμμετρίας, αν $\left| T'_n \left(\sum_{j=1}^n j^2 \right)^{-1/2} \right| > z_{\alpha/2}$. Καθώς $\left| 39(385)^{-1/2} \right| = 1.9876 > 1.96$, η υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται.

Ενότητα 2.2.1

Ο έλεγχος του McWilliams (1990) στηρίζεται στη σ.σ. R^* της σχέσης (2.2.1.1), η τιμή της οποίας θα υπολογιστεί για το συγκεκριμένο σύνολο δεδομένων. Στον Πίνακα 5.1.2 υπάρχουν τα $X_{(1)}, X_{(2)}, \dots, X_{(10)}$ και τα $S_1, S_2, \dots, S_{10}$. Οι τιμές των $I_2, \dots, I_{10}$ της σχέσης (2.2.1.2) είναι οι:

I_i	-	1	0	0	0	0	1	1	0	0
-------	---	---	---	---	---	---	---	---	---	---

Ακόμη, $\sum_{k=2}^{10} I_k = 3$ και $R^* = 4$. Έτσι, για επίπεδο σημαντικότητας 0.036, η

υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται αν $R^* \leq 2$,

καθώς $\sum_{r=2}^2 P(R^* = 2) = P(X = 1 | X \sim B(9, 0.5)) = 0.018$. Από τη σχέση

(2.2.1.5) έχουμε ότι $Z_{McW} = -1$. Γνωρίζοντας ότι ο έλεγχος αυτός απορρίπτει την υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $Z_{McW} < -z_{\alpha}$ και καθώς για επίπεδο

σημαντικότητας $\alpha = 0.05$ ισχύει $Z_{McW} > -z_\alpha = -1.64$, προκύπτει ότι η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 2.2.2

Για την εφαρμογή του ελέγχου του Baklizi (2008) πρέπει να βρεθεί η τιμή της σ.σ. Z_B της σχέσης (2.2.2.1). Από τον Πίνακα 5.1.1 προκύπτει ότι στην ακολουθία S_i , $i=1,2,\dots,10$ υπάρχουν 2 παρατηρήσεις με $S_i = 0$ και οι υπόλοιπες 8 παρατηρήσεις έχουν $S_i = 1$.

Επομένως $n_1 = 2$ και $n_2 = 8$. Ακόμη, $\gamma = \frac{n_1}{n_1 + n_2} = \frac{1}{5}$ και όμοια με τον

έλεγχο της προηγούμενης ενότητας προκύπτει $R^* = 4$. Για την ειδική περίπτωση όπου $n_1 = 2$ και $n_2 = 8$, έχοντας ως κρίσιμη περιοχή την $R^* \leq 2$, προκύπτει $\alpha = 2/45 = 0.0444$ κι αυτό διότι

$$f_{R^*}(R^* = 2 | n_1 = 2, n_2 = 8) = \frac{2 \binom{1}{0} \binom{7}{0}}{\binom{10}{2}} = \frac{2}{45}.$$

Επιπλέον, από τη σχέση (2.2.2.1) προκύπτει $Z_B = 0.7906$. Ο έλεγχος αυτός απορρίπτει την υπόθεση της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $Z_B < -z_\alpha = -1.64$ και έτσι προκύπτει ότι η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 2.2.3

Για την εφαρμογή του ελέγχου των Modarres and Gastwirth (1996), πρέπει να καθοριστεί η τιμή της σ.σ. M_p της σχέσης (2.2.3.1), για $p=0.2$. Για $n=10$ και $p=0.2$ προκύπτει $k=[np]+2=4$. Για $i=4,5,\dots,10$ οι τιμές των $\varphi(i)$ της σχέσης (2.2.3.2) είναι:

$$2, 3, 4, 5, 6, 7, 8.$$

Πολλαπλασιάζοντας για $i=4,5,\dots,10$ τις ποσότητες $\varphi(i)$ και I_i και αθροίζοντας τα αποτελέσματα αυτά προκύπτει $M_p=11$. Επιπλέον, από τις σχέσεις (2.2.3.4) και (2.2.3.5) για $n=10$ και $p=0.2$ προκύπτουν $\mu(M_p)=17.50$ και $\sigma^2(M_p)=50.75$, αντίστοιχα. Από τη σχέση (2.2.3.6) με αντικατάσταση των παραπάνω αποτελεσμάτων προκύπτει $Z_{MG}=-0.9124$.

Ο έλεγχος αυτός απορρίπτει την υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $Z_{MG} < -z_\alpha$. Επειδή για $\alpha=0.05$ ισχύει $Z_{MG} > -z_\alpha = -1.64$, προκύπτει ότι η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 2.2.4

Αρχικά θα βρεθεί η τιμή της σ.σ. J_k της σχέσης (2.2.4.1) για $k=6$. Για το σκοπό αυτό, αρκεί να βρεθεί ο αριθμός των ροών στους τελευταίους 6 όρους της ακολουθίας $I_2,\dots,I_{10}$ και να πολλαπλασιαστεί

με τον συντελεστή βαρύτητας που είναι η θέση του κάθε όρου. Από την εφαρμογή της μεθοδολογίας της Ενότητας 2.2.1, όπου δίνονται οι τιμές των $I_5, \dots, I_{10}$ και τον Πίνακα 5.1.2 προκύπτει $J_6 = 16$. Επιπλέον, οι σχέσεις (2.2.4.2) και (2.2.4.3) δίνουν $E(J_k - 1) = 3.75$ και $Var(J_k - 1) = 59.1667$. Αντικαθιστώντας τα αποτελέσματα αυτά στη σχέση (2.2.4.5) προκύπτει $Z_{COB} = 1.4626$.

Γνωρίζουμε ότι ο έλεγχος αυτός απορρίπτει την υπόθεση της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $Z_{COB} < -z_\alpha$. Επειδή $Z_{COB} > -z_\alpha = -1.64$, η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 2.2.5

Η σ.σ. $L_{n,p}^*$ του ελέγχου είναι το μήκος της μακρύτερης ροής στις τελευταίες $n - [np]$ παρατηρήσεις της ακολουθίας $S_1, S_2, \dots, S_{10}$. Για $n = 10$ και $p = 0.8$ προκύπτει $n - [np] = 2$. Από τον Πίνακα 5.1.1 προκύπτει ότι το μήκος της μακρύτερης ροής στις τελευταίες 2 παρατηρήσεις της ακολουθίας S_i είναι 2 και έτσι $L_{n,p}^* = 2$.

Στο συγκεκριμένο έλεγχο η απόφαση περί αποδοχής ή απόρριψης της υπόθεσης της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας λαμβάνεται αξιοποιώντας την p -τιμή, ο υπολογισμός της οποία γίνεται με βάση

την Πρόταση 2.2.5.1. Επειδή p -τιμή=1, η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 2.3.1

Για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από το μηδέν με τον έλεγχο του Butler (1969) απαιτείται ο προσδιορισμός των τιμών των σ.σ. B_n^+ , B_n^- και B_n των σχέσεων (2.3.1.1)-(2.3.1.3), αντίστοιχα. Οι τιμές αυτών των σ.σ. θα προσδιοριστούν με βάση την Παρατήρηση 2.3.1.1. Στον Πίνακα 5.1.3 δίνονται οι τιμές των T_j , για $j=1,2,...,11$,

καθώς και των $\sum_{i=j}^{n+1} T_i$, με το T_{11} να είναι εξ ορισμού ίσο με το μηδέν.

Πίνακας 5.1.3

j	1	2	3	4	5	6	7	8	9	10	11
T_j	-1	1	1	1	1	1	-1	1	1	1	0
$\sum_{i=j}^{n+1} T_i$	6	7	6	5	4	3	2	3	2	1	0

Με βάση τον Πίνακα αυτό προκύπτουν $B_n^+ = \max_{1 \leq j \leq 10} \sum_{i=j}^{11} T_i = 7$,

$B_n^- = \min_{1 \leq j \leq 11} \sum_{i=j}^{11} T_i = 0$ και $B_n = \max\{B_n^+, B_n^-\} = 7$. Η απόφαση περί αποδοχής

ή απόρριψης της υπόθεσης της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της θετικής λοξότητας λαμβάνεται αξιοποιώντας τη σχέση (2.3.1.4) και υπολογίζοντας την p -τιμή του

ελέγχου. Η p -τιμή είναι 0.0267 και συνεπώς σε επίπεδο σημαντικότητας $\alpha = 0.05$, η υπόθεση της συμμετρίας απορρίπτεται.

Ενότητες 2.3.2-2.3.4

Στη συνέχεια θα εφαρμοστούν οι έλεγχοι των σ.σ. R_n , SG_n και $T_n^{(0)}$ των σχέσεων (2.3.2.1), (2.3.3.1) και (2.3.4.8), αντίστοιχα. Για τις δύο πρώτες οι τιμές των σ.σ. θα υπολογιστούν με τη βοήθεια των εναλλακτικών μορφών που προτάθηκαν από τους Hill and Rao (1977) και δίνονται στις σχέσεις (2.3.4.6) και (2.3.4.7). Στον επόμενο Πίνακα δίνονται οι τιμές των ποσοτήτων που συμμετέχουν στους υπολογισμούς.

Πίνακας 5.1.4

X_i	-63	64	70	96	202	207	-213	245	397	442
Z_i	63	64	70	96	202	207	213	245	397	442
δ_k	-1	1	1	1	1	1	-1	1	1	1
P_k	0	1	2	3	4	5	5	6	7	8
N_k	1	1	1	1	1	1	2	2	2	2
δ'_k	1	1	1	-1	1	1	1	1	1	-1
P'_k	1	2	3	3	4	5	6	7	8	8
N'_k	0	0	0	1	1	1	1	1	1	2

Από τη σχέση (2.3.4.7) και για $a = 0.5$ προκύπτει $S_n^{(0.5)} = 0.9650$. Τότε, με βάση την Ιδιότητα 2, προκύπτει $S_n^{(0.5)} = R_n^{(1-0.5)} = R_n^{(0.5)} = R_n$. Ο έλεγχος των Rothman and Woodroffe (1972) απορρίπτει την υπόθεση της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, με επίπεδο σημαντικότητας $\alpha = 0.05$, αν $R_n \geq 1.656$. Επομένως, με τον έλεγχο της σ.σ. R_n η υπόθεση της συμμετρίας δεν απορρίπτεται.

Από τη σχέση (2.3.4.7) και για $a = 0$ προκύπτει $S_n^{(0)} = S_n = 1.17$. Επομένως, από τη σχέση $S_n = \frac{4SG_n}{n^2}$ προκύπτει ότι $SG_n = 29.25$. Με βάση τον Πίνακα 3 του Παραρτήματος Α προκύπτει ότι:

$$P(SG_n \geq 29.25) = 1 - P(SG_n \leq 29.25) = 1 - 0.8555 = 0.1445 > 0.05,$$

και η υπόθεση της συμμετρίας δεν απορρίπτεται. Εναλλακτικά, με τον έλεγχο της σ.σ. S_n , η υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται αν $S_n \geq k_\alpha$, με τα k_α να προσδιορίζονται μέσω της σχέσης $S(k_\alpha) = 1 - \alpha$ από τον Πίνακα 2 του Παραρτήματος Α. Επειδή $0.25S_n < 0.4$ προκύπτει ότι και με τον έλεγχο της σ.σ. S_n η υπόθεση της συμμετρίας δεν απορρίπτεται.

Για την εφαρμογή του ελέγχου των Hill and Rao (1977), πρέπει να προσδιοριστεί για $a = 0$ η τιμή της σ.σ. $T_n^{(a)}$ της σχέσης (2.3.4.8). Από τη σχέση (2.3.4.6), για $a = 1$ και με βάση τον Πίνακα 5.2.4

προκύπτει $R_n^{(1)} = 1.89$. Έτσι, έχοντας βρει ότι $S_n^{(0)} = 1.17$ η σχέση (2.3.4.8) δίνει $T_n^{(0)} = 1.53$. Για επίπεδο σημαντικότητας $\alpha = 0.05$, με βάση τον Πίνακα 4 του Παραρτήματος Α και καθώς $\frac{1}{4}n^2T_n^{(0)} < 223.25$, προκύπτει ότι ο έλεγχος της σ.σ. $T_n^{(0)}$ δεν απορρίπτει την υπόθεση της συμμετρίας γύρω από το μηδέν.

Ενότητα 2.3.5

Στον Πίνακα 5.1.5 δίνονται τα αποτελέσματα των υπολογισμών που έγιναν με σκοπό την εύρεση της τιμής των σ.σ. KS_n και CM_n των σχέσεων (2.3.5.4) και (2.3.5.5). Σύμφωνα με τα αποτελέσματα της δεύτερης στήλης του Πίνακα 5.1.5 προκύπτει ότι $\max \left| k \sum_{i=k+1}^n S_i \right| = 21$. Έτσι, η σχέση (2.3.5.4) δίνει $KS_n = 0.7379$. Επιπλέον, αθροίζοντας τα στοιχεία της τελευταίας στήλης του Πίνακα 5.1.5 και εφαρμόζοντας τη σχέση (2.3.5.5) προκύπτει $CM_n = 0.0794$.

Πίνακας 5.1.5

S_k	$\left k \sum_{i=k+1}^n S_i \right $	$\sum_{i=k+1}^n S_i$	$\left(\sum_{i=k+1}^n S_i \right)^2$	$\left(\frac{k}{n} \right)^2$	$\left(\frac{k}{n} \right)^2 \left(\sum_{i=k+1}^n S_i \right)^2$
-1	7	7	49	0.01	0.49
1	12	6	36	0.04	1.44
1	15	5	25	0.09	2.25
1	16	4	16	0.16	2.56
1	15	3	9	0.25	2.25
1	12	2	4	0.36	1.44
-1	21	3	9	0.49	4.41
1	16	2	4	0.64	2.56
1	9	1	1	0.81	0.81
1	-	-	-	-	-

Καθώς η κατανομή της σ.σ. CM_n δεν έχει βρεθεί, θα ελέγξουμε την υπόθεση της συμμετρίας γύρω από το μηδέν με τη σ.σ. KS_n . Σύμφωνα με τους Baringhaus and Henze (1992), για επίπεδο σημαντικότητας $\alpha = 0.05$, η υπόθεση της συμμετρίας απορρίπτεται αν $KS_n \geq 0.257$. Επειδή αυτό ισχύει για το συγκεκριμένο σύνολο δεδομένων, προκύπτει ότι η υπόθεση της συμμετρίας γύρω από το μηδέν απορρίπτεται.

Ενότητα 2.4.1

Για τον έλεγχο της υπόθεσης της συμμετρίας γύρω από το μηδέν με τη μέθοδο του Tajuddin (1994), πρέπει να υπολογιστεί η τιμή της σ.σ. S της σχέσης (2.4.1.6). Αρχικά, στο δείγμα υπάρχουν 8 θετικές παρατηρήσεις και 2 αρνητικές και έτσι $n_1 = 8$ και $n_2 = 2$. Τα V_k , $k = 1, 2, \dots, 10$ προκύπτουν με τον ίδιο τρόπο που προέκυψαν τα S_k , $k = 1, 2, \dots, 10$ της Ενότητας 2.1.2 και δίνονται στον Πίνακα 5.1.1. Πολλαπλασιάζοντας κάθε V_k με τον δείκτη k και αθροίζοντας τα αποτελέσματα προκύπτει $S = 47$. Με αντικατάσταση των n_1 , n_2 και n στις σχέσεις (2.4.1.7) και (2.4.1.8) προκύπτει η αναμενόμενη τιμή και η διακύμανση της σ.σ. S , οι οποίες είναι $E(S) = 44$ και $Var(S) = 14.6667$, αντίστοιχα. Τότε, η σχέση (2.4.1.9) δίνει $Z_{TAJ} = 0.7833$. Ο έλεγχος του Tajuddin (1994) απορρίπτει την υπόθεση της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|Z_{TAJ}| \geq z_{\alpha/2}$. Άρα, για επίπεδο σημαντικότητας $\alpha = 0.05$ η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 2.4.2

Θα εφαρμοστεί η βελτίωση του ελέγχου των Modarres and Gastwirth (1998) που προτάθηκε από τον Baklizi (2008) και περιλαμβάνει μόνο το δεύτερο βήμα του ελέγχου των δύο βημάτων. Για το σκοπό αυτό, θα υπολογιστεί η τιμή της σ.σ. W_p της σχέσης (2.4.2.2),

για $p = 0.7$ όπως προτείνεται από τους Modarres and Gastwirth (1998).

Αρχικά, πρέπει να βρεθεί η τιμή της παραμέτρου k η οποία είναι $k = np + 1 = 10 \cdot 0.7 + 1 = 8$. Έπειτα, για $k = 8, 9, 10$ θα υπολογιστούν τα $\varphi(k)$ της σχέσης (2.2.3.2), τα οποία είναι $\varphi(8) = 1$, $\varphi(9) = 2$ και $\varphi(10) = 3$. Τα V_k , για $k = 8, 9, 10$ προκύπτουν όπως και στην προηγούμενη παράγραφο και είναι ίσα με 1. Επομένως, η σχέση (2.4.2.2) δίνει $W_p = 6$.

Η αναμενόμενη τιμή και η διακύμανση της σ.σ. W_p διατυπώθηκαν στις σχέσεις (2.4.2.3) και (2.4.2.4), αντίστοιχα και προκύπτουν με αντικατάσταση των n_1 , n_2 και n που προσδιορίστηκαν στην προηγούμενη ενότητα. Έτσι, $E(W_p) = 4.8$ και $Var(W_p) = 1.8489$. Συνεπώς, από τη σχέση (2.4.2.5) προκύπτει $Z_{MOG} = 0.8825$. Ο συγκεκριμένος έλεγχος απορρίπτει την υπόθεση της συμμετρίας γύρω από το μηδέν έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας, αν $|Z_{MOG}| > z_{\alpha/2}$ και επομένως η υπόθεση της συμμετρίας δεν απορρίπτεται.

5.2 Ελέγχων συμμετρίας γύρω από άγνωστο σημείο

Στην ενότητα αυτή θα εφαρμοστούν οι κυριότεροι μέθοδοι ελέγχου συμμετρίας ως προς άγνωστο σημείο σε ένα γνωστό σύνολο δεδομένων (βλέπε μεταξύ άλλων Table 3.9 Hollander and Wolfe

(1973)). Τα δεδομένα αυτά αποτελούν ένα υποσύνολο των δεδομένων που πρωτοπαρουσιάστηκαν από τον Ijzermans (1970) και αφορούν το ποσοστό του χρωμίου σε δείγματα ανοξείδωτου χάλυβα. Οι μετρήσεις αυτές είναι οι ακόλουθες:

17.4, 17.9, 17.6, 18.1, 17.6, 18.9, 16.9, 17.5, 17.8, 17.4, 24.6, 26.

Στον Πίνακα 5.2.1 δίνονται διάφορα περιγραφικά μέτρα για το παραπάνω σύνολο δεδομένων.

Πίνακας 5.2.1

Περιγραφικό Μέτρο	Τιμή
Δειγματική Μέση Τιμή	18.975
Δειγματική Διάμεσος	17.700
Δειγματική Διακύμανση	9.0475
Hodges-Lehmann εκτιμητής	17.850

Ενότητα 3.1.1

Θα πρέπει να υπολογιστεί η τιμή της σ.σ. D_n της σχέσης (3.1.1.3). Λαμβάνοντας υπόψη ότι η δειγματική διάμεσος είναι 17.7, προκύπτουν τα ακόλουθα αποτελέσματα για τις 6 δειγματικές παρατηρήσεις με τιμή μικρότερη από 17.7.

Πίνακας 5.2.2

$F_n(X_i)$	3	6	6	1	4	3
$F_n(2\tilde{\mu} - X_i)$	8	7	7	9	8	8
$1 - F_n(X_i) - F_n(2\tilde{\mu} - X_i)$	0.833	-0.833	-0.833	0.1667	0.0000	0.0833

Με βάση τα αποτελέσματα του Πίνακα 5.2.2 προκύπτει $D_n = 0.1667$. Για επίπεδο σημαντικότητας $\alpha = 0.05$ και για μέγεθος δείγματος 12, η κρίσιμη τιμή είναι 0.200. Άρα, επειδή $D_n < 0.200$ προκύπτει ότι δεν απορρίπτεται η υπόθεση της συμμετρίας έναντι της θετικής λοξότητας.

Ενότητα 3.1.2

Θα πρέπει να υπολογιστεί η τιμή της σ.σ. R_n της σχέσης (3.1.2.1). Οι παρατηρήσεις σε αύξουσα τάξη μεγέθους είναι:

16.9 17.4 17.4 17.5 17.6 17.6 17.8 17.9 18.1 18.9 24.6 26.0

Στον Πίνακα 5.2.3 δίνονται οι παράγοντες που συντελούν στον υπολογισμό της σ.σ. του ελέγχου.

Πίνακας 5.2.3

a_i	$F_n(2\tilde{\mu} - X_{(i)})$	$\left[F_n(2\tilde{\mu} - X_{(i)}) - \frac{2n-2i+1}{2n} \right]^2$	$a_i \left[F_n(2\tilde{\mu} - X_{(i)}) - \frac{2n-2i+1}{2n} \right]^2$
1.0000	9	0.9583	0.0434
0.0500	8	0.8750	0.0217
0.3300	8	0.7917	0.0052
0.2500	8	0.7083	0.0004
0.2000	7	0.6250	0.0003
0.1667	7	0.5417	0.0003
0.1429	6	0.4583	0.0002
0.1250	4	0.3750	0.0002
0.1111	1	0.2917	0.0048
0.1000	0	0.2083	0.0043
0.0909	0	0.1250	0.0014
0.0833	0	0.0417	0.0001

Έτσι, αθροίζοντας τα στοιχεία της τελευταίας στήλης του παραπάνω πίνακα προκύπτει η τιμή της σ.σ. η οποία είναι $R_n = 0.0826$. Επομένως, καθώς για επίπεδο σημαντικότητας $\alpha = 0.05$ και για μέγεθος δείγματος 12 η κρίσιμη τιμή είναι 0.110, προκύπτει ότι δεν απορρίπτεται η υπόθεση της συμμετρίας έναντι της θετικής λοξότητας.

Ενότητα 3.1.3

Θα πρέπει να υπολογιστεί η τιμή της σ.σ. R^* της σχέσης (3.1.3.2). Στον Πίνακα 5.2.4 δίνονται οι τιμές των ποσοτήτων που χρησιμοποιούνται στους υπολογισμούς.

Πίνακας 5.2.4

$X_{(i)} - \bar{X}_n$	$Y_{(i)}$	S_i	I_i
-0.8000	0.1000	1	-
-0.3000	0.1000	0	0
-0.3000	0.1000	1	0
-0.2000	0.2000	1	1
-0.1000	0.2000	0	0
-0.1000	0.3000	0	0
0.1000	0.3000	1	0
0.2000	0.4000	1	1
0.4000	0.8000	0	0
1.2000	1.2000	1	0
6.9000	6.9000	1	1
8.3000	8.3000	1	1

Με βάση τα παραπάνω στοιχεία, η τιμή της σ.σ. του ελέγχου προκύπτει αθροίζοντας τα στοιχεία της τελευταίας στήλης του Πίνακα

5.2.4 και είναι $R^* = 4$. Για επίπεδο σημαντικότητας $\alpha = 0.05$ και για μέγεθος δείγματος 12 η κρίσιμη τιμή είναι 6. Άρα, δεν απορρίπτεται η υπόθεση της συμμετρίας έναντι της θετικής λοξότητας.

Ενότητα 3.1.4

Αρχικά, θα πρέπει να υπολογιστούν οι τιμές των σ.σ. T^* και T των σχέσεων (3.1.4.2) και (3.1.4.4), αντίστοιχα.

Πίνακας 5.2.5

$X_i - \bar{X}_n$	I_i^*	$X_{(n+1-i)}$	$2\tilde{\mu} - X_{(i)}$	I_i
-1.5770	1	26.0000	18.5000	1
-1.0750	1	24.6000	18.0000	1
-1.3750	1	18.9000	18.0000	1
-0.8750	1	18.1000	17.9000	1
-1.3750	1	17.9000	17.8000	1
-0.0750	1	17.8000	17.8000	0
-2.0705	1	-	-	-
-1.5750	1	-	-	-
-1.1750	1	-	-	-
-1.5750	1	-	-	-
5.6250	0	-	-	-
7.0250	0	-	-	-

Στον Πίνακα 5.2.5 δίνονται οι ποσότητες που χρησιμοποιούνται στους υπολογισμούς. Αθροίζοντας τα στοιχεία των στηλών I_i^* και I_i προκύπτουν τα $T^*=10$ και $T=5$, αντίστοιχα. Για επίπεδο σημαντικότητας $\alpha=0.05$, δεν απορρίπτεται η υπόθεση της συμμετρίας έναντι της θετικής λοξότητας, καθώς η κρίσιμη τιμή του ελέγχου είναι 9.

Ενότητα 3.1.5

Οι έλεγχοι της Ενότητας 3.1.5 εφαρμόζονται για περιττό μέγεθος δείγματος και επομένως δεν υλοποιούνται στο συγκεκριμένο παράδειγμα. Επιπλέον, έχει αποδειχθεί ότι δεν είναι ιδιαίτερα αποτελεσματικοί.

Ενότητα 3.1.6

Θα πρέπει να υπολογιστούν οι τιμές των σ.σ. $S_1, S_2, S_3, S_{4,0}$ και $T_2(0.05)$ των σχέσεων (3.1.6.1)-(3.1.6.5) και της σ.σ. $T_{4,0}$ της σχέσης (3.1.6.18). Για την σ.σ. S_1 , με χρήση των δύο πρώτων στηλών του Πίνακα 5.2.5 προκύπτει $S_1=0.0278$. Άρα, για επίπεδο σημαντικότητας $\alpha=0.05$, γνωρίζοντας ότι η κρίσιμη περιοχή του ελέγχου είναι $|S_1| \geq 0.98$, προκύπτει ότι η υπόθεση της συμμετρίας δεν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Για την σ.σ. S_2 , αρχικά θα πρέπει για $i=1,2,...,12$ και $j=1,2,...,12$, με $i < j$, να υπολογιστούν οι διαφορές $X_i - X_j - 2\bar{X}_n$,

όπου $\bar{X}_n$ η δειγματική μέση τιμή. Υπάρχουν 66 συνδυασμοί που ικανοποιούν τον παραπάνω περιορισμό και οι διαφορές είναι:

-2.6500, -2.9500, -2.4500, -2.9500, -1.6500, -3.6500, -3.0500, -2.7500, -3.1500, 4.0500, 5.4500, -2.4500, -1.9500, -2.4500, -1.1500, -3.1500, -2.5500, -2.2500, -2.6500, 4.5500, 5.9500, -2.2500, -2.7500, -1.4500, -3.4500, -2.8500, -2.5500, -2.9500, 4.2500, 5.6500, -2.2500, -0.9500, -2.9500, -2.3500, -2.0500, -2.4500, 4.7500, 6.1500, -1.4500, -3.4500, -2.8500, -2.5500, -2.9500, 4.2500, 5.6500, -2.1500, -1.5500, -1.2500, -1.6500, 5.5500, 6.9500, -3.5500, -3.2500, -3.6500, 3.5500, 4.9500, -2.6500, -3.0500, 4.1500, 5.5500, -2.7500, 4.4500, 5.8500, 4.0500, 12.6500.

Από τις παραπάνω διαφορές, 45 έχουν αρνητική τιμή και $S_2 = 0.2887$. Από τη θεωρία γνωρίζουμε ότι η κρίσιμη περιοχή του ελέγχου είναι $|S_2| \geq 0.2828$. Άρα, η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Για τον υπολογισμό της τιμής της σ.σ. S_3 , βρέθηκαν για $i=1,2,\dots,12$ και $j=1,2,\dots,12$, με $i < j$, οι τιμές των διαφορών $X_i + X_j - 2M_n$, όπου M_n η δειγματική διάμεσος. Υπάρχουν 66 συνδυασμοί που ικανοποιούν αυτόν τον περιορισμό και οι τιμές των διαφορών είναι:

-0.1000, -0.4000, 0.1000, -0.4000, 0.9000, -1.1000, -0.5000, -0.2000, -0.6000, 6.6000, 8.0000, 0.1000, 0.6000, 0.1000, 1.4000, -0.6000, -

7.1054e-15, 0.3000, -0.1000, 7.1000, 8.5000, 0.3000, -0.2000, 1.1000, -0.9000, -0.3000, 0.0000, -0.4000, 6.8000, 8.2000, 0.3000, 1.6000, -0.4000, 0.2000, 0.5000, 0.1000, 7.3000, 8.7000, 1.1000, -0.9000, -0.3000, 0.0000, -0.4000, 6.8000, 8.2000, 0.4000, 1.0000, 1.3000, 0.9000, 8.1000, 9.5000, -1.0000, -0.7000, -1.1000, 6.1000, 7.5000, -0.1000, -0.5000, 6.7000, 8.1000, -0.2000, 7.0000, 8.4000, 6.6000, 8.0000, 15.2000.

Από τις παραπάνω τιμές, 48 είναι αρνητικές και προκύπτει $S_3 = -0.2165$. Από τη θεωρία έχουμε ότι η κρίσιμη περιοχή είναι $|S_3| \geq 0.3481$ και επομένως η υπόθεση της συμμετρίας δεν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Η τιμή της σ.σ. $S_{4,\varepsilon}$ εξαρτάται από την τιμή της παραμέτρου ε . Από τους περισσότερους μελετητές προτείνεται να χρησιμοποιείται η σ.σ. $S_{4,0}$, δηλαδή να επιλέγεται τιμή του ε η οποία να είναι αρκετά κοντά στο μηδέν. Στον Πίνακα 5.2.6 δίνονται τα αποτελέσματα των υπολογισμών που έγιναν. Από τον Πίνακα 5.2.6 προκύπτει ότι για τις διαφορές $Y_i - Y_{n-i+2}$ υπάρχουν 5 αρνητικές παρατηρήσεις και έτσι $S_{4,0} = 0.7217$. Καθώς για επίπεδο σημαντικότητας $\alpha = 0.05$ η κρίσιμη περιοχή του ελέγχου είναι $|S_{4,0}| \geq 1.0323$, προκύπτει ότι η υπόθεση της συμμετρίας δεν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Πίνακας 5.2.6

$X_{(i)}$	$Y_{(i)}$	$Y_i - Y_{n-i+2}$
16.9000	-	-
17.4000	0.5000	-0.9000
17.4000	0.0000	-5.7000
17.5000	0.1000	-0.7000
17.6000	0.1000	-0.1000
17.6000	0.0000	-0.1000
17.8000	0.2000	-
17.9000	0.1000	-
18.1000	0.2000	-
18.9000	0.8000	-
24.6000	5.7000	-
26.0000	1.4000	-

Στη συνέχεια, θα βρεθεί η τιμή της σ.σ. $T_2(0.05)$ που δόθηκε στη σχέση (3.1.6.5) και αποτελεί μια εναλλακτική μορφή της σ.σ. $S_{4,0}$. Ο Πίνακας 5.2.7 περιέχει τους υπολογισμούς που έγιναν για την σ.σ. $T_2(0.05)$.

Πίνακας 5.2.7

$\frac{i}{n}$	$J\left(\frac{i}{n}\right)$	$Y_i - Y_{n-i}$	$I(Y_i - Y_{n-i}) - 0.5$
0.0833	1	-0.9000	0.5000
0.1667	1	-5.7000	0.5000
0.2500	1	-0.7000	0.5000
0.3333	1	-0.1000	0.5000
0.4167	1	-0.1000	0.5000

Με βάση τα παραπάνω, για το συγκεκριμένο σύνολο δεδομένων προκύπτει $T_2(0.05) = 0.7217$. Η υπόθεση της συμμετρίας απορρίπτεται

έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $\left| \frac{T_2(0.05)}{\sqrt{0.1008}} \right| \geq z_{\alpha/2}$.

Για επίπεδο σημαντικότητας $\alpha = 0.05$ προκύπτουν $z_{\alpha/2} = 1.96$ και

$\frac{T_2(0.05)}{\sqrt{0.1008}} = 2.2737$. Επομένως, η υπόθεση της συμμετρίας απορρίπτεται.

Για τη σ.σ. $T_{4,0}$ απαιτείται ο προσδιορισμός των $D_{k,m}$ και $D_{i,m} = X_{(i+m)} - X_{(i-m+1)}$, με $k = [12/2] = 6$. Καθώς το δείγμα έχει 12 παρατηρήσεις από τη θεωρία προτείνεται η χρησιμοποίηση της τιμής $m = 3$ και έτσι $D_{k,m} = 0.6000$. Στον Πίνακα 5.2.8 δίνονται τα αποτελέσματα των υπολογισμών των μεγεθών που συμμετέχουν στην εύρεση της τιμής της σ.σ. $T_{4,0}$.

Πίνακας 5.2.8

$D_{i,m}$	$1 - \log \frac{D_{i,m}}{D_{k,m}}$	$\left(1 - \log \frac{D_{i,m}}{D_{k,m}}\right)^2$
0.7000	0.8458	0.7155
0.4000	1.4055	1.9753
0.5000	1.1823	1.3979
0.6000	1.0000	1.0000
1.3000	0.2268	0.0514
7.0000	-1.4567	2.1221
8.2000	-1.6150	2.6081

Με βάση τα παραπάνω και τη σχέση (3.1.6.15) προκύπτει

$$\hat{\sigma}_{4,0}^2 = 0.1506 \text{ και έτσι } T_{4,0} = \frac{S_{4,0}}{\hat{\sigma}_{4,0}} = \frac{0.7217}{0.3880} = 1.8600. \text{ Η κρίσιμη περιοχή}$$

του ελέγχου είναι $|T_{4,0}| \geq z_{\alpha/2}$. Επομένως, η υπόθεση της συμμετρίας δεν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Ενότητα 3.2.1

Η τιμή του δειγματικού συντελεστή λοξότητας b_1 της σχέσης (3.2.1.1) είναι $b_1 = 1.7305$. Για να βρεθεί ο εκτιμητής $\hat{\sigma}_0^2(b_1)$ της ασυμπτωτικής διακύμανσης της σ.σ. b_1 , θα πρέπει να υπολογιστούν οι δειγματικές ροπές m_i , $i = 2, 4, 6$. Αυτές είναι $m_2 = 8.2935$,

$m_4 = 290.2641$, $m_6 = 1.2667e+04$ και έτσι $\hat{\sigma}_0^2(b_1) = 5.8857$. Η σχέση (3.2.1.6) δίνει $Z_{b_1} = 2.4709$. Επομένως, γνωρίζοντας ότι η κρίσιμη περιοχή του ελέγχου είναι $|Z_{b_1}| \geq z_{\alpha/2}$, προκύπτει ότι για επίπεδο σημαντικότητας $\alpha = 0.05$, με τον έλεγχο του δειγματικού συντελεστή λοξότητας, η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Ενότητα 3.2.2

Για τον υπολογισμό της σ.σ. S_K της σχέσης (3.2.2.3) χρειάζονται η δειγματική μέση τιμή, διάμεσος και διακύμανση. Από τον Πίνακα 5.2.1 έχουμε $\bar{X} = 18.975$, $m = 17.700$, $S^2 = 9.0475$ και έτσι προκύπτει $S_K = 1.4684$. Με τον έλεγχο αυτό, η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $|S_K| \geq P_{1-\alpha/2}$, με τα $P_{1-\alpha/2}$ να δίνονται στον Πίνακα της Ενότητας 3.3.2. Στο συγκεκριμένο σύνολο δεδομένων ισχύει $|S_K| \geq P_{1-0.05/2} = 1.2400$ και επομένως η υπόθεση της συμμετρίας απορρίπτεται. Τέλος, για την τροποποίηση του παραπάνω ελέγχου η οποία είναι ο έλεγχος με τη σ.σ. S_K^* της σχέσης (3.2.2.11), προκύπτει $S_K^* = 1.798$ και η υπόθεση της συμμετρίας δεν απορρίπτεται.

Έλεγχος Miao et al. (2006)

Αρχικά, θα πρέπει να υπολογιστεί η τιμή της σ.σ. J της σχέσης (3.2.2.12). Για το σκοπό αυτό απαιτούνται οι διαφορές $X_i - M$, όπου M η δειγματική διάμεσος, οι οποίες είναι :

-0.3 0.2 -0.1 0.4 -0.1 1.2 -0.8 -0.2 0.1 -0.3 6.9 8.3.

Έπειτα, αθροίζοντας τις απόλυτες τιμές αυτών και πολλαπλασιάζοντας με τον συντελεστή $\sqrt{\frac{\pi}{2}} \frac{1}{n}$ προκύπτει $J = 1.9740$.

Στη συνέχεια, θα πρέπει να υπολογιστεί η τιμή της σ.σ. T_{Mi} της σχέσης (3.2.2.13). Καθώς η δειγματική μέση τιμή και διάμεσος είναι γνωστές, εύκολα προκύπτει ότι $T_{Mi} = 0.6459$. Η κρίσιμη τιμή είναι 0.4275 και επειδή $|T_{Mi}| > 0.4275$, προκύπτει ότι η υπόθεση της συμμετρίας απορρίπτεται.

Ενότητα 3.2.3

Για τον έλεγχο της υπόθεσης της συμμετρίας με τη μέθοδο του Ghosh (2011) απαιτείται η εύρεση της τιμής της σ.σ. Z_G της σχέσης (3.2.3.12). Για το σκοπό αυτό, αρχικά, θα πρέπει να βρεθεί η τιμή του $\hat{\tau}$ είτε μέσω της σχέσης (3.2.3.3) είτε με τη βοήθεια της εναλλακτικής μορφής του $\hat{\tau}$ που δόθηκε από τον Ghosh (2011). Από τον Πίνακα 5.2.9, προκύπτει $\hat{\tau} = \frac{1}{12^2} 48 = 0.3333$. Από τον Πίνακα 5.2.1, η τυπική

απόκλιση είναι $S = 3.0079$ και έτσι η σχέση (3.2.3.9) δίνει $h_n = 0.5053$, ενώ από τη σχέση (3.2.3.10) προκύπτει $IQR = 1.0500$. Για τον προσδιορισμό του Z_G απαιτείται η εύρεση του εκτιμητή $\hat{\omega}^2$ της σχέσης (3.2.3.6). Ο Πίνακας 5.2.10 περιέχει τα αποτελέσματα των υπολογισμών που έγιναν.

Πίνακας 5.2.9

$\frac{X_{(k)} + X_{(n-k)}}{2}$	Πλήθος X_i , με $X_i < \frac{X_{(k)} + X_{(n-k)}}{2}$
$\frac{X_{(1)} + X_{(12)}}{2} = 21.45$	10
$\frac{X_{(2)} + X_{(11)}}{2} = 21.00$	10
$\frac{X_{(3)} + X_{(10)}}{2} = 18.15$	9
$\frac{X_{(4)} + X_{(9)}}{2} = 17.80$	7
$\frac{X_{(5)} + X_{(8)}}{2} = 17.75$	6
$\frac{X_{(6)} + X_{(7)}}{2} = 17.70$	6

Πίνακας 5.2.10

$\frac{\hat{\mu} - X_i}{h_n}$	$K\left(\frac{\hat{\mu} - X_i}{h_n}\right)$	$ X_i - \hat{\mu} $
0.5937	0.3118	1.5750
-0.3958	0.3249	1.0750
0.1979	0.3328	1.3750
-0.7916	0.2934	0.8750
0.1979	0.3328	1.3750
-2.3748	0	0.0750
1.5832	0.1673	2.0750
0.3958	0.3249	1.4750
-0.1979	0.3328	1.1750
0.5937	0.3118	1.5750
-13.6551	0	5.6250
-16.4257	0	7.0250

Αθροίζοντας τα στοιχεία της δεύτερης στήλης του Πίνακα 5.2.10

προκύπτει $\sum_{i=1}^n K\left(\frac{\hat{\mu} - X_i}{h_n}\right) = 2.7323$. Επομένως, η σχέση (3.2.3.8) δίνει

$\hat{f}(\hat{\mu}) = 0.4506$. Επιπλέον, αθροίζοντας τα στοιχεία της τελευταίας

στήλης του παραπάνω πίνακα προκύπτει $\sum_{i=1}^n |X_i - \hat{\mu}| = 25.3000$. Έτσι, με αντικατάσταση των παραπάνω τιμών στη σχέση (3.2.3.7) προκύπτει $\hat{\omega}^2 = 0.2843$.

Άρα, $Z_G = 0.5414$ και καθώς από τη θεωρία γνωρίζουμε ότι η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $Z_G > z_{\alpha/2}$ και για $\alpha = 0.05$ είναι $Z_G < 1.96$, προκύπτει ότι η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 3.2.4

Ο έλεγχος της Ενότητας 3.2.4 έχει τη μικρότερη ισχύ μεταξύ των ελέγχων του 3^{ου} Κεφαλαίου και έχει προταθεί να αποφεύγεται. Συμπεριλήφθηκε στη διατριβή για ιστορικούς λόγους και δεν υλοποιήθηκε.

Ενότητα 3.2.5

Ο Hodges-Lehmann εκτιμητής για το συγκεκριμένο σύνολο δεδομένων είναι $\hat{\mu} = 17.85$. Για την εφαρμογή του ελέγχου του Boos (1980), θα πρέπει να υπολογιστεί η τιμή της συνάρτησης T_{B_n} της σχέσης (3.2.5.5).

Για $i = 1, 2, \dots, 12$ και $j = 1, 2, \dots, 12$, βρέθηκε το άθροισμα των απόλυτων τιμών των διαφορών $X_i + X_j - 2\hat{\mu}$, το οποίο είναι 415.4. Επιπλέον, υπολογίστηκε το άθροισμα των απόλυτων τιμών των

διαφορών $X_i - X_j$ για τους δείκτες με $i < j$ και είναι 179.5. Έτσι, $T_{B_n} = 1.8852$ και επειδή $P(T_{B_n} \geq 1.8852) < 0.01$, προκύπτει ότι δεν απορρίπτεται η υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Έλεγχοι συμμετρίας με το Gini's Mean Difference

Για την εφαρμογή των ελέγχων που προτάθηκαν από τους Wang *et al.* (2013), απαιτείται η εύρεση της τιμής των σ.σ. S_{GMD} και RS_{GMD} των σχέσεων (3.2.5.8) και (3.2.5.11), αντίστοιχα. Στον Πίνακα 5.2.11 δίνονται οι τιμές των ποσοτήτων που συμμετέχουν στους υπολογισμούς. Έτσι, για $k = [n/2] = [12/2] = 6$ προκύπτουν

$$\hat{\sigma}_G^+ = \binom{k}{2}^{-1} \sum_{i < j}^k |Y_i^+ - Y_j^+| = 61.9 \text{ και } \hat{\sigma}_G^- = \binom{k}{2}^{-1} \sum_{i < j}^k |Y_i^- - Y_j^-| = 4.2. \text{ Συνεπώς,}$$

η σχέση (3.2.5.8) δίνει $S_{GMD} = \log \frac{\hat{\sigma}_G^+}{\hat{\sigma}_G^-} = 2.6904$. Ακόμη, από τις σχέσεις

(3.2.5.9) και (3.2.5.10) προκύπτουν:

$$R_{GMD+} = \frac{2 \sum_{i=1}^k iR^{(1)}(i) - (k+1) \sum_{i=1}^k R^{(1)}(i)}{k(k-1)/2} = \frac{2*202 - 7*47.5}{6*5/2} = 4.7667,$$

και

$$R_{GMD-} = \frac{2 \sum_{i=1}^k iR^{(2)}(i) - (k+1) \sum_{i=1}^k R^{(2)}(i)}{k(k-1)/2} = \frac{2*132 - 7*30.5}{6*5/2} = 3.3667,$$

και με αντικατάσταση στη σχέση (3.2.5.11) προκύπτει

$$RS_{GMD} = \log \frac{R_{GMD+}}{R_{GMD-}} = 0.3477.$$

Πίνακας 5.2.11

$X_{(i)}$	Y_j^+	Y_j^-	$R^1(i)$	$R^2(i)$
16.9000	8.3	0.8	12	9
17.4000	6.9	0.3	11	6.5
17.4000	1.2	0.3	10	6.5
17.5000	0.4	0.2	8	4.5
17.6000	0.2	0.1	4.5	2
17.6000	0.1	0.1	2	2
17.8000	-	-	-	-
17.9000	-	-	-	-
18.1000	-	-	-	-
18.9000	-	-	-	-
24.6000	-	-	-	-
26.0000	-	-	-	-

Με βάση την διατριβή του Auda (2006, Table 5.1) και τον Πίνακα 7 του Παραρτήματος Α, για $n_1 = n_2 = 6$ και επίπεδο σημαντικότητας $\alpha = 0.05$, η κρίσιμη περιοχή του RS_{GMD} ελέγχου είναι:

$$RS_{GMD} < \log(0.595) = -0.5192 \text{ ή } RS_{GMD} > \log(1.651) = 0.5014.$$

Άρα, η υπόθεση της συμμετρίας δεν απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Ενότητα 3.2.6

Για την εφαρμογή του ελέγχου που προτάθηκε από την Mira (1999), απαιτείται ο προσδιορισμός της τιμής της σ.σ. Z_M της σχέσης (3.2.6.10), η οποία βρέθηκε ότι είναι $Z_M = 1.6057$, ενώ η p-τιμή του ελέγχου είναι 0.1083. Επομένως, για επίπεδο σημαντικότητας $\alpha = 0.05$, γνωρίζοντας ότι η κρίσιμη περιοχή του ελέγχου είναι $|Z_M| \geq z_{\alpha/2} = 1.96$, προκύπτει ότι έλεγχος της σ.σ. Z_M δεν απορρίπτει την υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας.

Ενότητα 3.2.7

Η εφαρμογή του κανόνα των Mukherjee *et al.* (2013) για τον έλεγχο της υπόθεσης της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ύπαρξης θετικής λοξότητας, απαιτεί τον προσδιορισμό της παραμέτρου t της σχέσης (3.2.7.1).

Οι Mukherjee *et al.* (2013) προτείνουν για δείγματα μεγέθους $n > 3$ να επιλέγεται $t = n - 3$. Άρα, στο συγκεκριμένο σύνολο δεδομένων με $n = 12$ προκύπτει $t = 9$. Η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ύπαρξης θετικής λοξότητας αν $X_{(t)} > 2\tilde{\mu} - X_{(1)}$, δηλαδή αν $X_{(9)} > 2\tilde{\mu} - X_{(1)}$, όπου $\tilde{\mu}$ η δειγματική

διάμεσος, η τιμή της οποίας από τον Πίνακα 5.2.1 είναι 17.7. Οι παρατηρήσεις σε αύξουσα τάξη μεγέθους είναι:

16.9 17.4 17.4 17.5 17.6 17.6 17.8 17.9 18.1 18.9 24.6 26.0

Επομένως, $X_{(1)} = 16.1$, $X_{(9)} = 18.1$ και $2\tilde{\mu} - X_{(1)} = 19.3$. Άρα, $X_{(9)} < 2\tilde{\mu} - X_{(1)}$ και έτσι ο κανόνας δεν απορρίπτει την υπόθεση της συμμετρίας.

Ενότητα 3.4.1

Στην Παράγραφο 3.4.1 παρουσιάστηκαν οι έλεγχοι των σ.σ. T_{R1} και T_{R2} των σχέσεων (3.4.1.14) και (3.4.1.15), αντίστοιχα. Στην Ενότητα 4.2 ο έλεγχος της σ.σ. T_{R2} κρίθηκε αναποτελεσματικός καθώς δεν διατηρεί το επίπεδο σημαντικότητας. Έτσι, στο συγκεκριμένο παράδειγμα εφαρμόστηκε μόνο ο έλεγχος της σ.σ. T_{R1} .

Αρχικά, υπολογίστηκε η τιμή του αθροίσματος $\sum_{i < j < k} f^*(X_i, X_j, X_k)$ η οποία είναι 205.3333. Έπειτα, διαιρώντας το προηγούμενο άθροισμα με $\binom{n}{3}$ βρέθηκε η τιμή $\hat{\eta} = 0.9333$. Στη συνέχεια το ενδιαφέρον επικεντρώθηκε στην εύρεση του εκτιμητή $\hat{\sigma}_n$, για τον οποίο απαιτούνται τα $\hat{\zeta}_i$, $i = 1, 2, 3$. Από τη σχέση (3.4.1.13), προκύπτει $\hat{\zeta}_3 = -0.7600$. Ακόμη, $\hat{\zeta}_2 = 0.6106$ και $\hat{\zeta}_1 = 0.5946$. Τότε, με αντικατάσταση των $\hat{\zeta}_i$ στην σχέση (3.1.4.6) προκύπτει $\hat{\sigma}_n^2 = 4.3604$ και

έτσι $T_{R1} = 1.5483$. Γνωρίζοντας ότι με τον έλεγχο της σ.σ. T_{R1} η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $|T_{R1}| > z_{\alpha/2}$, προκύπτει ότι για επίπεδο σημαντικότητας $\alpha = 0.05$ η υπόθεση της συμμετρίας δεν απορρίπτεται.

Ενότητα 3.4.2

Στη σχέση (3.4.2.12) δόθηκε η σ.σ. $T_1(\alpha)$ από την οποία για $\alpha = 0$ προκύπτει η σ.σ. του ελέγχου του Gupta (1967). Επιπλέον, από τους Antille *et al.* (1982) προτάθηκε η εφαρμογή του ελέγχου της σ.σ. $T_1\left(\frac{1}{6}\right)$. Για τον έλεγχο της υπόθεσης της συμμετρίας με αυτές τις δύο σ.σ. απαιτείται ο προσδιορισμός των τιμών των $T_1(0)$ και $T_1\left(\frac{1}{6}\right)$ για το συγκεκριμένο σύνολο δεδομένων. Στον Πίνακα 5.2.12 δίνονται τα αποτελέσματα των υπολογισμών που έγιναν.

Πολλαπλασιάζοντας τα στοιχεία της τρίτης και της τέταρτης στήλης και αθροίζοντας τα αποτελέσματα προκύπτει

$$\sum_{i=1}^n G_0 \left(\frac{R_i(M_n)}{2(n+1)} \right) \text{sign}(X_i - M_n) = 0.77 \quad \text{και} \quad \text{επομένως} \quad T_1(0) = 0.1963.$$

Ακόμη, πολλαπλασιάζοντας τα στοιχεία της τρίτης και της πέμπτης στήλης και αθροίζοντας τα αποτελέσματα προκύπτει

$$\sum_{i=1}^n G_{1/6} \left(\frac{R_i(M_n)}{2(n+1)} \right) \text{sign}(X_i - M_n) = 0.39 \quad \text{και} \quad \text{έτσι} \quad T_1\left(\frac{1}{6}\right) = 0.1125.$$

Από τη θεωρία γνωρίζουμε ότι για επίπεδο σημαντικότητας $\alpha = 0.05$, η υπόθεση της συμμετρίας απορρίπτεται έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $|T_1(0)| \geq 0.3480$ και αν $\left|T_1\left(\frac{1}{6}\right)\right| \geq 0.2581$. Επομένως, η υπόθεση της συμμετρίας δεν απορρίπτεται.

Πίνακας 5.2.12

$ X_i - M_n $	$R_i(M_n)$	$sign(X_i - M_n)$	$G_0\left(\frac{R_i}{2(n+1)}\right)$	$G_{1/6}\left(\frac{R_i}{2(n+1)}\right)$
0.3	6.5	-1	0.25	0.25
0.2	4.5	1	0.17	0.17
0.1	2	-1	0.07	0.07
0.4	8	1	0.30	0.30
0.1	2	-1	0.07	0.07
1.2	10	1	0.5	0.33
0.8	9	-1	0.34	0.33
0.5	4.5	-1	0.17	0.17
0.1	2	1	0.07	0.07
0.3	6.5	-1	0.25	0.25
6.9	11	1	0.42	0.33
8.3	12	1	0.46	0.33

Ενότητα 3.4.3

Θα εφαρμοστεί ο έλεγχος της σ.σ. $T_3(0.05)$ της σχέσης (3.4.3.7).

Οι τιμές $Y_{(i)}$ των διαφορών $X_{(i+1)} - X_{(i)}$ δίνονται στον Πίνακα 5.2.6 και χρησιμοποιούνται στη εύρεση των V_i . Στον Πίνακα 5.2.13 δίνονται οι τιμές των ποσοτήτων που συνθέτουν την σ.σ. $T_3(0.05)$.

Πίνακας 5.2.13

$\frac{i}{n+1}$	$J\left(\frac{i}{n+1}\right)$	V_i	$J\left(\frac{i}{n+1}\right)V_i$
0.0769	1	-0.4737	-0.4737
0.1538	1	-1.0000	-1.0000
0.2308	1	-0.7778	-0.7778
0.3077	1	-0.3333	-0.3333
0.3846	1	-1.0000	-1.0000

Αθροίζοντας τα στοιχεία της τελευταίας στήλης του Πίνακα 5.2.13 και πολλαπλασιάζοντας το αποτέλεσμα με $n^{-1/2}$ προκύπτει $T_3(0.05) = -1.0348$. Ο έλεγχος αυτός για επίπεδο σημαντικότητας $\alpha = 0.05$, απορρίπτει την υπόθεση της συμμετρίας έναντι της εναλλακτικής υπόθεσης της ασυμμετρίας αν $T_3(0.05) \geq z_{\alpha/2} \sqrt{0.1293} = 0.5903$. Στο συγκεκριμένο παράδειγμα αυτό δεν ισχύει και έτσι η υπόθεση της συμμετρίας δεν απορρίπτεται.

Αντί επιλόγου

Στη μεταπτυχιακή αυτή διατριβή το ενδιαφέρον επικεντρώθηκε στο πρόβλημα του ελέγχου της μονοδιάστατης συμμετρίας γύρω από ένα γνωστό ή ένα άγνωστο σημείο. Πλήθος ερευνητικών εργασιών έχουν εμφανιστεί στη βιβλιογραφία, λόγω της σημαντικότητας της υπόθεσης της συμμετρίας. Στο πλαίσιο αυτής της διατριβής το ενδιαφέρον εστιάστηκε στην κριτική παρουσίαση και ανασκόπηση των κυριότερων, δημοφιλέστερων και πιο εύκολα εφαρμόσιμων σε πρακτικές εφαρμογές μεθόδων. Η συγκριτική μελέτη των μεθοδολογιών αυτών έγινε μέσω των ιδιοτήτων τους καθώς και της ανασκόπησης των διαθέσιμων στη βιβλιογραφία συγκριτικών προσομοιώσεων. Εκτός από τις μεθοδολογίες που παρουσιάστηκαν στα Κεφάλαια 2 και 3, έχουν αναπτυχθεί κατά τη διάρκεια των ετών επιπλέον μεθοδολογίες για τον έλεγχο της μονοδιάστατης συμμετρίας. Ενδεικτικά, αναφέρονται οι εργασίες των Ahmad and Li (1997), Aki (1987), Hill and Rao (1981), Franch and Contreras (2002) και Reynolds (1975). Επιπλέον, υπάρχει εκτενής βιβλιογραφία μεταξύ άλλων για τον έλεγχο της πολυδιάστατης συμμετρίας και της ελλειπτικής συμμετρίας. Ενδεικτικά, για πολυδιάστατη συμμετρία αναφέρονται οι εργασίες των Diks and Tong (1999), Heathcote *et al.* (1995) και Blough (1989), ενώ για ελλειπτική συμμετρία αναφέρονται οι εργασίες των Li *et al.* (1997), Zhu and Neuhauss (2003) και Huffer and Park (2007).

Θεωρούμε ότι το θέμα παραμένει ανοικτό για περαιτέρω έρευνα τόσο ως προς την προσθήκη νέων μεθόδων όσο και ως προς την καλύτερη και εκτενέστερη συγκριτική μελέτη των υπαρχουσών.

ΠΑΡΑΡΤΗΜΑ Α

Στο Παράρτημα Α παρατίθενται χρήσιμοι πίνακες για την εφαρμογή των μεθοδολογιών ελέγχου της συμμετρίας που παρουσιάστηκαν στην παρούσα διατριβή. Ειδικότερα, δίνονται οι ακόλουθοι πίνακες:

Πίνακας 1. Πίνακας υπολογισμού των $P(X \leq x) * 1000$, όταν $X \sim B(n, 0.5)$. Πηγή: Παπαϊωάννου, Τ. και Λουκάς, Σ..

Πίνακας 2. Κρίσιμες τιμές της σ.σ. S_n . Πηγή: Orlov (1972).

Πίνακας 3. Κρίσιμες τιμές της σ.σ. SG_n . Πηγή: Srinivasan and Godio (1974).

Πίνακας 4. Πίνακας υπολογισμού των πιθανοτήτων $P = P(0.25n^2 T_n^{(0)} \leq x)$. Πηγή: Hill and Rao (1977).

Πίνακας 5. Κρίσιμες τιμές του ελέγχου των Cabilio and Masaro (1996). Πηγή: Cabilio and Masaro (1996).

Πίνακας 6. Πίνακας τυπικής κανονικής κατανομής.

Πίνακας 7. Κρίσιμες τιμές της σ.σ. $\exp(R_{GMD})$. Πηγή: Auda (2006).

Πίνακας 1. Πίνακας υπολογισμού των $P(X \leq x) * 1000$, όταν $X \sim B(n, 0.5)$.

Σημείωση: Με * έχουμε συμβολίσει τις τιμές που είναι ίσες ή περίπου ίσες με 1000

N	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	031	188	500	812	969	*										
6	016	109	344	656	891	984	*									
7	008	062	227	500	773	938	992	*								
8	004	035	145	363	637	855	965	996	*							
9	002	020	090	254	500	746	910	980	998	*						
10	001	011	055	172	377	623	828	945	989	999	*					
11		006	033	113	247	500	726	887	967	994	*	*				
12		003	019	073	194	387	613	806	927	981	997	*	*			
13		002	011	046	133	291	500	709	867	954	989	998	*	*		
14		001	006	029	090	212	395	605	788	910	971	994	999	*	*	
15			004	018	059	151	304	500	696	849	947	982	996	*	*	*
16			002	011	038	105	227	402	598	773	895	962	989	998	*	*
17			001	006	025	072	166	315	500	685	834	928	975	994	999	*
18			001	004	015	048	119	240	407	593	760	881	952	985	996	999
19				002	010	032	084	180	324	500	676	820	916	968	990	998
20				001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987
22					002	008	026	067	143	262	416	584	738	857	933	974
23					001	005	017	047	105	202	339	500	661	798	895	953
24					001	003	011	032	076	154	271	419	581	729	846	924
25						002	007	022	054	115	212	345	500	655	788	885

Πίνακας 2. Κρίσιμες τιμές της σ.σ. S_n .

x	$S(x)$	x	$S(x)$	x	$S(x)$
0.000	0.0000000	0.500	0.9696850	1.000	0.9980828
0.020	0.1090357	0.520	0.9729687	1.040	0.9984532
0.040	0.2988227	0.540	0.9758841	1.080	0.9987512
0.060	0.4347773	0.560	0.9784745	1.120	0.9989913
0.080	0.5328109	0.580	0.9807779	1.160	0.9991848
0.100	0.6069192	0.600	0.9828274	1.200	0.9993409
0.120	0.6651505	0.620	0.9846522	1.240	0.9994669
0.140	0.7122236	0.640	0.9862778	1.280	0.9995686
0.160	0.7510634	0.660	0.9877268	1.320	0.9996507
0.180	0.7535899	0.680	0.9890191	1.360	0.9997171
0.200	0.8111307	0.700	0.9901722	1.400	0.9997708
0.220	0.8346457	0.720	0.9912015	1.480	0.9998494
0.240	0.8548546	0.740	0.9921207	1.560	0.9999009
0.260	0.8723127	0.760	0.9929419	1.640	0.9999348
0.280	0.8874587	0.780	0.9936759	1.720	0.9999570
0.300	0.9006453	0.800	0.9943321	1.800	0.9999716
0.320	0.9121694	0.820	0.9949191	1.880	0.9999813
0.340	0.9222416	0.840	0.9954442	1.960	0.9999876

Πίνακας 2 (Συνέχεια)

x	$S(x)$	x	$S(x)$	x	$S(x)$
0.360	0.9319872	0.860	0.9959141	2.040	0.9999918
0.380	0.9388638	0.880	0.9963348	2.120	0.9999946
0.400	0.9457124	0.900	0.9967115	2.200	0.9999964
0.420	0.9517529	0.920	0.9970490	2.400	0.9999987
0.440	0.9570881	0.940	0.9973513	2.600	0.9999995
0.460	0.9618062	0.960	0.9976222	2.800	0.9999998
0.480	0.9659832	0.980	0.9978650	3.000	0.9999999

Πίνακας 3. Κρίσιμες τιμές της σ.σ. SG_n .

n=10		n=11		n=12		n=13	
x	$P(S_n \leq x)$	x	$P(S_n \leq x)$	x	$P(S_n \leq x)$	x	$P(S_n \leq x)$
22.25	0.8496	30.50	0.8486	35.50	0.8442	41.75	0.8472
26.25	0.8555	31.50	0.8535	36.50	0.8501	42.75	0.8535
32.25	0.8984	38.50	0.8994	45.50	0.8989	53.75	0.8997
33.25	0.8984	39.50	0.9053	46.50	0.9019	54.75	0.9033
43.25	0.9063	50.50	0.9463	61.50	0.9497	71.75	0.9480
44.25	0.9473	51.50	0.9502	62.50	0.9512	72.75	0.9507
53.25	0.9531	65.50	0.9746	78.50	0.9746	91.75	0.9746
54.25	0.9727	66.50	0.9766	79.50	0.9756	92.75	0.9766
65.25	0.9883	78.50	0.9893	97.50	0.9897	114.75	0.9895
66.25	0.9902	81.50	0.9902	99.50	0.9902	116.75	0.9900
n=14		n=15		n=16		n=17	
x	$P(S_n \leq x)$	x	$P(S_n \leq x)$	x	$P(S_n \leq x)$	x	$P(S_n \leq x)$
48.75	0.8470	56.00	0.8486	63.00	0.8476	71.25	0.8473
49.75	0.8521	57.00	0.8533	64.00	0.8513	72.25	0.8511
60.75	0.8978	71.00	0.8992	80.00	0.8990	90.25	0.8992
61.75	0.9006	72.00	0.9023	81.00	0.9014	91.25	0.9015
83.75	0.9487	96.00	0.9493	109.00	0.9495	123.25	0.9496
84.75	0.9506	97.00	0.9511	110.00	0.9509	124.25	0.9508
105.75	0.9745	121.00	0.9745	138.00	0.9746	156.25	0.9749
106.75	0.9751	122.00	0.9750	139.00	0.9754	157.25	0.9758
134.75	0.9899	154.00	0.9898	175.00	0.9989	199.25	0.9898
135.75	0.9902	155.00	0.9901	176.00	0.9900	200.25	0.9902

Πίνακας 3 (Συνέχεια)

n=18		n=19		n=20	
x	$P(S_n \leq x)$	x	$P(S_n \leq x)$	x	$P(S_n \leq x)$
79.25	0.8466	88.50	0.8487	98.50	0.8496
80.25	0.8502	89.50	0.8513	99.50	0.8519
100.25	0.8977	112.50	0.8997	124.50	0.8999
101.25	0.9000	113.50	0.9015	125.50	0.9015
137.25	0.9491	153.50	0.9499	169.50	0.9494
138.25	0.9501	154.50	0.9506	170.50	0.9503
174.25	0.9746	194.50	0.9748	215.50	0.9747
175.25	0.9751	195.50	0.9751	216.50	0.9751
223.25	0.9898	248.50	0.9899	275.50	0.9899
224.25	0.9901	249.50	0.9901	276.50	0.9900

Πίνακας 4. Πίνακας υπολογισμού των πιθανοτήτων

$$P = P(0.25n^2 T_n^{(0)} \leq x).$$

n=10		n=11		n=12		n=13	
x	P	x	P	x	P	x	P
32.25	0.8945	33.00	0.8955	43.00	0.8970	49.25	0.8982
33.25	0.9102	34.50	0.9072	44.00	0.9087	51.75	0.9009
38.25	0.9492	49.50	0.9443	53.00	0.9497	64.75	0.9490
39.25	0.9570	49.50	0.9502	54.00	0.9517	65.75	0.9504
44.15	0.9746	50.00	0.9736	71.50	0.7961	77.75	0.9736
48.25	0.9785	57.50	0.9775	73.50	0.7961	78.75	0.9751
60.25	0.9863	81.50	0.9893	80.50	0.9883	101.75	0.9883
62.25	0.9902	82.00	0.9912	83.50	0.9912	104.75	0.9902
n=14		n=15		n=16		n=17	
x	P	x	P	x	P	x	P
54.75	0.8979	68.00	0.8984	71.50	0.8988	83.25	0.8989
55.75	0.9000	68.50	0.9001	72.50	0.9013	83.75	0.9009
79.75	0.9497	84.00	0.9492	99.50	0.9487	115.25	0.9499
80.75	0.9512	84.50	0.9501	100.00	0.9509	115.75	0.9502
91.75	0.9736	112.50	0.9743	120.00	0.9748	137.75	0.9749
92.75	0.9751	113.00	0.9750	121.00	0.9752	138.25	0.9753
125.75	0.9899	132.50	0.9899	155.50	0.9897	184.75	0.9900
127.75	0.9907	133.00	0.9902	156.50	0.9903	185.25	0.9903

Πίνακας 4 (Συνέχεια)

n=18		n=19		n=20		n=21	
x	P	x	P	x	P	x	P
96.25	0.8989	101.50	0.8990	117.00	0.8981	125.25	0.8998
97.25	0.9008	102.00	0.9004	117.50	0.9003	126.25	0.9001
121.25	0.9493	141.50	0.9495	152.00	0.9496	167.25	0.9496
122.25	0.9506	142.00	0.9505	152.50	0.9503	168.25	0.9504
161.25	0.9746	169.50	0.9747	192.50	0.9748	218.25	0.9749
162.25	0.9752	170.00	0.9750	193.00	0.9753	218.75	0.9751
192.25	0.9898	220.00	0.9897	253.00	0.9900	263.75	0.9899
193.25	0.9907	220.50	0.9900	253.50	0.9902	264.25	0.9900
n=22		n=23		n=24			
x	P	x	P	x	P		
137.75	0.8997	155.75	0.8998	163.25	0.8996		
138.75	0.9006	156.25	0.9003	163.75	0.9007		
189.75	0.9499	199.25	0.9497	222.75	0.9498		
190.75	0.9506	199.75	0.9501	223.75	0.9505		
227.75	0.9747	257.25	0.9748	280.75	0.9750		
228.75	0.9751	257.75	0.9750	281.25	0.9751		
296.75	0.9900	332.25	0.9900	348.25	0.9900		
297.75	0.9901	332.75	0.9901	348.75	0.9901		

Πίνακας 5. Κρίσιμες τιμές του ελέγχου των Cabilio and Masaro (1996).

Κατανομή F	$e_{m,\bar{x}}(F)$	$E\left \frac{X-\mu}{\sigma}\right $	$\sigma_0^2(F)$	P_{95}	$P_{97.5}$
Κανονική	0.636620	0.797885	0.570796	1.2427	1.4808
T(0.01,3)	0.678413	0.783120	0.572463	1.2445	1.4829
T(0.05,3)	0.832840	0.741769	0.575095	1.2474	1.4863
T(0.1,3)	0.998220	0.713650	0.573212	1.2453	1.4839
t_3	1.621139	0.636620	0.616850	1.2919	.5394
t_5	0.960675	0.735105	0.540935	1.2098	1.4415
t_8	0.797526	0.765466	0.539592	1.2083	1.4397
Λογιστική	0.822467	0.764304	0.530322	1.1978	1.4273
Διπλά Εκθετική	2.000000	0.707107	0.500000	1.1631	1.3859
Βήτα(1,1)	0.333333	0.866025	1.000000	1.6449	1.9600
Βήτα(2,2)	0.450000	0.838525	0.722222	1.3979	1.6656
Βήτα(3,3)	0.502232	0.826797	0.657777	1.3340	1.5896
Βήτα(4,4)	0.531684	0.820313	-0.630816	1.3064	1.5567

Πίνακας 6. Πίνακας υπολογισμού των $P(0 \leq Z \leq z)$, όταν $Z \sim N(0,1)$.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890

2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981

Πίνακας 7. Κρίσιμες τιμές της σ.σ. $\exp(R_{GMD})$.

	0.995	0.990	0.950	0.900	0.200	0.100	0.050	0.01	0.005
n=8									
2,6			0.283	0.294	1.364	1.744			
3,5			0.351	0.417	1.333	1.667			
4,4		0.385	0.455	0.565	1.353	1.714	1.923	2.200	
n=9									
2,7			0.236	0.276	1.458	1.853			
3,6		0.323	0.400	0.492	1.429	1.556	1.86		
4,5		0.379	0.492	0.570	1.354	1.528	1.795		
4,4		0.467	0.520	0.615	1.300	1.556			
n=10									
2,8				0.246	1.400	2.000			
3,7			0.304	0.378	0.477	1.361	1.556		
4,6	0.352	0.373	0.507	0.580	1.279	1.531	1.761	2.000	2.063
5,5		0.400	0.565	0.652	1.333	1.533	1.750		
n=11									
2,9			0.225	0.231	1.479	1.826	2.149		
3,8		0.282	0.355	0.444	1.357	1.584	1.846	2.051	
4,7	0.336		0.490	0.595	1.323	1.555	1.701	1.909	
5,6		0.400	0.565	0.642	1.301	1.500	1.674	1.953	
5,5		0.464	0.609	0.667	1.277	1.444	1.588	1.833	
n=12									
2,10			0.211	0.217	1.599	1.895			
3,9		0.267	0.400	0.461	1.402	1.621	1.764	2.062	
4,8		0.324	0.511	0.583	1.339	1.500	1.644	1.904	
5,7		0.396	0.434	0.646	1.292	1.453	1.602	1.914	2.032
6,6	0.454	0.494	0.595	0.666	1.296	1.472	1.651	1.976	2.100
n=13									
2,11				0.205	1.492	1.996		2.630	
3,10	0.25	0.253	0.376	0.452	1.442	1.658	1.736	1.988	
4,9		0.340	0.494	0.587	1.317	1.500	1.621	1.852	
5,8	0.412	0.430	0.567	0.652	1.285	1.451	1.585	1.866	1.983
6,7	0.444	0.485	0.603	0.676	1.263	1.448	1.592	1.873	1.987
6,6	0.505	0.541	0.638	0.700	1.258	1.406	1.549	1.823	1.916
n=14									
2,12		0.184	0.188	0.202	1.585	1.867	2.357	2.667	
3,11	0.236	0.241	0.359	0.476	1.368	1.613	1.774	2.002	2.072
4,10	0.301	0.369	0.488	0.579	1.321	1.492	1.618	1.825	1.967
5,9	0.396	0.437	0.568	0.650	1.278	1.430	1.563	1.800	1.905
6,8	0.456	0.494	0.611	0.684	1.260	1.420	1.552	1.813	1.960
7,7	0.485	0.521	0.629	0.700	1.256	1.418	1.567	1.882	2.029
n=15									
2,13			0.175	0.191	1.536	1.921	2.412	2.696	
3,12		0.230	0.342	0.452	1.390	1.609	1.765	1.968	
4,11	0.290	0.352	0.478	0.576	1.314	1.472	1.619	1.811	1.909
5,10	0.395	0.437	0.572	0.647	1.278	1.427	1.545	1.790	1.894
6,9	0.451	0.488	0.612	0.685	1.252	1.400	1.521	1.783	1.888
7,8	0.497	0.527	0.639	0.704	1.248	1.395	1.530	1.794	1.920
7,7	0.527	0.560	0.661	0.727	1.227	1.372	1.500	1.761	1.842

Πίνακας 7 (Συνέχεια)

	0.995	0.990	0.950	0.900	0.200	0.100	0.050	0.01	0.005
n=16									
2,14		0.164	0.167	0.181	1.578	1.998	2.237	2.722	
3,13	0.212	0.215	0.339	0.442	1.430	1.617	1.767	1.979	
4,12	0.300	0.338	0.486	0.576	1.315	1.481	1.599	1.822	1.900
5,11	0.386	0.425	0.562	0.647	1.269	1.416	1.535	1.762	1.833
6,10	0.451	0.488	0.612	0.685	1.246	1.388	1.507	1.746	1.831
7,9	0.494	0.526	0.641	0.711	1.236	1.375	1.497	1.758	1.865
8,8	0.521	0.556	0.658	0.730	1.241	1.387	1.520	1.742	1.838
n=17									
2,15			0.159	0.171	1.534	2.058	2.275		
3,14	0.202	0.205	0.334	0.439	1.402	1.616	1.773	1.961	2.026
4,13	0.315	0.330	0.480	0.573	1.317	1.477	1.587	1.793	1.885
5,12	0.382	0.427	0.557	0.646	1.267	1.412	1.520	1.743	1.832
6,11	0.453	0.491	0.611	0.686	1.242	1.372	1.490	1.712	1.803
7,10		0.433	0.497	0.531	1.231	1.363	1.478	1.714	1.797
8,8	0.528	0.558	0.663	0.727	1.227	1.365	1.483	1.737	1.844
n=18									
2,16		0.148	0.162	0.169	1.572	1.938	2.324	2.542	
3,15	0.193	0.202	0.335	0.440	1.391	1.635	1.783	1.951	2.030
4,14	0.303	0.327	0.473	0.571	1.317	1.467	1.582	1.801	1.876
5,13	0.381	0.419	0.557	0.644	1.263	1.406	1.516	1.721	1.797
6,12	0.443	0.483	0.609	0.686	1.237	1.367	1.478	1.691	1.770
7,11	0.493	0.532	0.644	0.713	1.227	1.351	1.462	1.682	1.765
8,10	0.527	0.560	0.666	0.730	1.220	1.349	1.459	1.679	1.762
9,9	0.547	0.580	0.678	0.739	1.218	1.351	1.473	1.726	1.833
n=19									
2,17		0.142	0.145	0.160	1.531	1.991	2.366	2.572	2.778
3,16	0.186	0.194	0.338	0.461	1.414	1.626	1.792	1.959	2.033
4,15	0.291	0.314	0.472	0.569	1.315	1.467	1.585	1.784	1.887
5,14	0.376	0.421	0.557	0.645	1.264	1.397	1.511	1.719	1.798
6,13	0.445	0.485	0.609	0.686	1.237	1.361	1.469	1.671	1.737
7,12	0.493	0.528	0.645	0.715	1.221	1.344	1.448	1.655	1.729
8,11	0.528	0.561	0.668	0.733	1.213	1.335	1.442	1.659	1.741
9,10	0.551	0.582	0.682	0.744	1.210	1.336	1.447	1.681	1.771
9,9	0.577	0.608	0.700	0.757	1.177	1.278	1.361	1.469	1.485

ΠΑΡΑΡΤΗΜΑ Β

Στο Παράρτημα Β παρατίθενται οι πηγαίοι κώδικες (m_files) της MatLab που δημιουργήθηκαν για τις εφαρμογές του Κεφαλαίου 5.

ΜΕΡΟΣ Α

Στην ενότητα αυτή, θα δοθεί ο πηγαίος κώδικας που δημιουργήθηκε για την εφαρμογή των ελέγχων συμμετρίας γύρω από το μηδέν, με υλοποίηση του οποίου προέκυψαν τα αποτελέσματα της Ενότητας 5.1.

```
data=[-63;64;70;96;202;207; -213;245;397;442];
```

```
%Υπολογισμός βοηθητικών μεταβλητών-ποσοτήτων.
```

```
n=size(data);  
x=find(data<0);  
c=size(x);
```

```
% Εύρεση της ακολουθίας  $S_i$  (διάνυσμα  $s$ ) της σχέσης (2.1.2.1).
```

```
newabs=abs(data);  
y=sort(newabs);  
for i=1:c(1,1)  
for j=1:n(1,1)  
if y(j)==abs(data(x(i)))  
y(j)=-y(j);  
end  
end  
end  
y  
s=zeros(1,n(1,1));
```

```

for i=1:n(1,1)
    if y(i)>0 s(i)=1;
    end
end
s

```

*% Εύρεση των ποσοτήτων δ_k (dk) που χρησιμοποιούνται στην
Ενότητα 2.3.4 (βλέπε σχέση 2.3.4.7).*

```

t1=zeros(1,n(1,1)+1);
for i=1:n(1,1)
    if y(i)>0 t1(i)=1;
    else t1(i)=-1;
    end
end
t1
dk=t1(1:n(1,1))

```

*% Εύρεση των ποσοτήτων δ'_k ($dk1$) που χρησιμοποιούνται στην
Ενότητα 2.3.4 (βλέπε σχέση 2.3.4.6).*

```

antistr=1./data
newabs1=abs(1./data)
y11=sort(newabs1);
for i=1:c(1,1)
    for j=1:n(1,1)
        if y11(j)==abs(1./data(x(i)))
            y11(j)=-y11(j);
        end
    end
end
y11
t11=zeros(1,n(1,1)+1);
for i=1:n(1,1)
    if y11(i)>0 t11(i)=1;
    else t11(i)=-1;
    end
end

```

```
t11
```

```
dk1=t11(1:n(1,1))
```

% Υλοποίηση ελέγχων της Ενότητας 2.1.1 (Προσημικός έλεγχος)

% Υπολογισμός της σ.σ. T_S (ts), της αναμενόμενης τιμής της (ets) και της διακύμανσης της (varts). Προσέγγιση διωνυμικής κατανομής από κανονική (βλέπε σχέση 2.1.1.2) με τη σ.σ. Z_S (zs).

```
ts=n(1,1)-c(1,1)
```

```
ets=0.5*n(1,1);
```

```
varts=0.25*n(1,1);
```

```
zs=(ts-ets)\sqrt(varts)
```

% Υπολογισμός της σ.σ. Z'_S (zstonos) που προκύπτει με εφαρμογή της διόρθωσης συνεχείας (σχέση 2.1.1.3).

```
if ts>n(1,1)/2 zstonos=(ts-ets-0.5)\sqrt(varts);
```

```
else zstonos=(ts-ets+0.5)\sqrt(varts);
```

```
end
```

```
zstonos
```

% Κανόνας απόφασης με τις σ.σ. Z_S (zs) και Z'_S (zstonos). Αν zsapor=1 (zstonosapor=1, αντίστοιχα) τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(zs)>norminv(1-0.05/2,0,1) zsapor=1;
```

```
else zsapor=0;
```

```
end
```

```
zsapor
```

```
if abs(zstonos)>norminv(1-0.05/2,0,1) zstonosapor=1;
```

```
else zstonosapor=0;
```

```
end
```

```
zstonosapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.1.2 (Cheng and Balakrishnan (2004))

% Υπολογισμός της σ.σ. C_6 (c_6), της αναμενόμενης τιμής της (ec_6) και της διακύμανσης της ($varc_6$). Προσέγγιση διωνυμικής κατανομής από κανονική (βλέπε σχέση 2.1.2.4) με τη σ.σ. Z_{CB} (zcb).

```
c6=sum(s((n(1,1)-5):n(1,1)))
ec6=0.5*6;
varc6=0.25*6;
zcb=(ts-ec6)\sqrt(varc6)
```

% Υπολογισμός της σ.σ. Z'_{CB} ($zcbtonos$) που προκύπτει με εφαρμογή της διόρθωσης συνεχείας (σχέση 2.1.2.5).

```
if c6>3 zcbtonos=(c6-ec6-0.5)\sqrt(varc6);
else zcbtonos=(c6-ec6+0.5)\sqrt(varc6);
end
zcbtonos
```

% Κανόνας απόφασης με τις σ.σ. Z_{CB} (zcb) και Z'_{CB} ($zcbtonos$). Αν $zcbapor=1$ ($zcbtonosapor=1$, αντίστοιχα) τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(zcb)>norminv(1-0.05/2,0,1) zcbapor=1;
else zcbapor=0;
end
zcbapor
if abs(zcbtonos)>norminv(1-0.05/2,0,1) zcbtonosapor=1;
else zcbtonosapor=0;
end
zcbtonosapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.1.3 (Έλεγχος προσημασμένης διάταξης του Wilcoxon)

% Υπολογισμός των ποσοτήτων R_i^+ (riplus) και των ποσοτήτων $\text{sign}(X_i)$ (sign) της σχέσης (2.1.3.2).

```
riplus=tiedrank(newabs)
sign=ones(1,n(1,1));
for i=1:n(1,1)
    if data(i)<0 sign(i)=-1;
end
end
sign;
```

%Υπολογισμός της σ.σ. προσημασμένης διάταξης του Wilcoxon T_n' (tntonos) και κανόνας απόφασης.

```
tntonos=sign*riplus
if abs(tntonos/sqrt(sum((1:n(1,1)).^2)))>norminv(1-0.05/2,0,1)
tntonosapor=1;
else tntonosapor=0;
end
tntonosapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.2.1 (McWilliams (1990))

%Υπολογισμός των ποσοτήτων I_k (dianitel) της σχέσης (2.2.1.2).

```
diani=ones(1,n(1,1));
for i=2:n if s(i-1)==s(i)
    diani(i)=0;
end
end
diani;
dianitel=diani(2:n(1,1))
```

%Υπολογισμός της σ.σ. R^ (rasteri) της σχέσης 2.2.1.1.*

```
rasteri=1+sum(dianitel)
```

%Προσέγγιση της σ.σ. R^ μέσω Κ.Ο.Θ. (βλέπε σχέση 2.2.1.5) και σ.σ. Z_{McW} ($zmcw$).*

```
zmcw=((rasteri-1)-((n(1,1)-1)/2))/sqrt((n(1,1)-1)*0.25)
```

% Διόρθωση συνεχείας (βλέπε παρατήρηση 2.1.1.2). Προκύπτει η σ.σ. Z'_{McW} της σχέσης 2.2.1.6 ($zmcwtonos$).

```
if (rasteri-1)>((n(1,1)-1)/2) zmcwtonos=((rasteri-0.5)-((n(1,1)-1)/2))/sqrt((n(1,1)-1)*0.25);
else zmcwtonos=((rasteri-1.5)-((n(1,1)-1)/2))/sqrt((n(1,1)-1)*0.25);
end
zmcwtonos
```

% Κανόνας απόφασης με τις σ.σ. Z_{McW} ($zmcw$) και Z'_{McW} ($zmcwtonos$). Αν $zmcwapor=1$ ($zmcwtonosapor=1$, αντίστοιχα) τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if zmcw<-norminv(1-0.05,0,1) zmcwapor=1;
else zmcwapor=0;
end
zmcwapor
if zmcwtonos<-norminv(1-0.05,0,1) zmcwtonosapor=1;
else zmcwtonosapor=0;
end
zmcwtonosapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.2.2 (Baklizi (2003))

% Υπολογισμός των n_1 ($n1$) και n_2 ($n2$) που ορίζονται να είναι το πλήθος των μηδενικών και των μονάδων αντίστοιχα στην ακολουθία των S_i (βλέπε διάνυσμα s παραπάνω).

```
plithosmonadon=size(find(s>0));
n2=plithosmonadon(2)
n1=n(1,1)-n2
```

% Υπολογισμός της σ.σ. Z_B (zb) της σχέσης (2.2.2.1).

```
g=(n1/(n1+n2));
zb=(rasteri-2*(n1+n2)*g*(1-g))/(2*sqrt(n1+n2)*g*(1-g))
```

% Κανόνας απόφασης με τη σ.σ. Z_B (zb). Αν zbapor =1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if zb<-norminv(1-0.05,0,1) zbapor=1;
else zbapor=0;
end
zbapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.2.3 (Modarres and Gastwirth (1996))

% Ορίζεται το περικομμένο ποσοστό p και υπολογίζονται οι τιμές των $\varphi(k)$ (διάνυσμα ft) της σχέσης (2.2.3.2).

```
p=0.2;
k=fix(n(1,1)*p);
f=zeros(1,n(1,1));
for i=1:n(1,1) if i>n(1,1)*p f(i)=i-n(1,1)*p
end
end
f;
ft=f(2:n(1,1));
```

% Υπολογισμός της σ.σ. M_p (mp) της σχέσης (2.2.3.1), της αναμενόμενης τιμής της (emp) και της διακύμανσης της (varmp). Προσέγγιση με Κ.Ο.Θ. και σ.σ. Z_{MG} (zmg) της σχέσης (2.2.3.6).

```
mpolo=f.*diani
mp=sum(mpolo((k+2):n(1,1)))
emp=0.25*(n(1,1)*(1-p)-1)*(n(1,1)*(1-p)+2)
varmp=(1/24)*(n(1,1)*(1-p)-1)*(2*(n(1,1)^2)*((1-p)^2)+5*n(1,1)*(1-p)+6)
zmg=(mp-emp)/sqrt(varmp)
```

% Κανόνας απόφασης με τη σ.σ. Z_{MG} (zmg). Αν zmgapor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if zmg<-norminv(1-0.05,0,1) zmgapor=1;
else zmgapor=0;
end
zmgapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.2.4 (Corzo and Babativa (2013))

% Υπολογισμός της σ.σ. J_6 (j6) δηλαδή του αριθμού των ροών στους 6 τελευταίους όρους της ακολουθίας των S_i με συντελεστή βαρύτητας τη θέση του εκάστοτε όρου (βλέπε σχέση (2.2.4.1)).

```
j=[0 (2:n(1,1)).*dianitel];
j6=1+sum(j((n(1,1)-5):n(1,1)))
```

% Υπολογισμός των ποσοτήτων $E(J_6 - 1)$ (ej6minus1) και $Var(J_6 - 1)$ (varj6minus1) των σχέσεων (2.2.4.2) και (2.2.4.3), αντίστοιχα.

```
ej6minus1=0.25*(2*n(1,1)-5)
varj6minus1=(1/6)*(6*n(1,1)^2-30*n(1,1)+55)
```

% Υπολογισμός της σ.σ. Z_{COB} (zcob) της σχέσης (2.2.4.5).

```
zcob=((j6-1)-ej6minus1)/sqrt(varj6minus1)
```

% Κανόνας απόφασης με τη σ.σ. Z_{COB} (zcob). Αν zcobapor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if zcob<-norminv(1-0.05,0,1) zcobapor=1;
else zcobapor=0;
end
zcobapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.2.5 (Baklizi (2007))

% Ορίζεται το περικομμένο ποσοστό (p1) και υπολογίζεται το μεγαλύτερο μήκος ροής $L_{n,p}^$ (lnp) στους τελευταίους $n-[np]$ όρους της ακολουθίας S_i (διάνυσμα s) και υπολογίζεται η p-τιμή του ελέγχου (βλέπε Πρόταση 2.2.5.1).*

```
p1=0.8;
k1=fix(n(1,1)*p1)
sipo=s((k1+1):n(1,1))
findseq(sipo)
b=max(findseq(sipo))
if numel(b)==1 lnp=b;
else lnp=b(4);
end
lnp
pvaluelnp=exp(-((n(1,1)-k1)-lnp)/(2^lnp))
```

% Υλοποίηση ελέγχων της Ενότητας 2.3.1 (Smirnov (1947) και Butler (1969))

% Υπολογισμός των τιμών των σ.σ. B_n^+ (bnplus), B_n^- (bnminus) και B_n (bn) με βάση την παρατήρηση 2.3.1.1 (Niederhausen 1979). Υπολογισμός της p-τιμής (pvalubnplus) για μονόπλευρο έλεγχο με βάση τη σχέση (2.3.1.4).

```
wn=zeros(1,n(1,1)+1);
for j=1:n(1,1)+1 wn(j)=sum(t1(j:(n(1,1)+1)));
end
wn;
bnplus=max(wn)
bnminus=min(wn)
bn=max(bnplus,bnminus)
pvalubnplus=2*normcdf(-bnplus/sqrt(n(1,1)))
```

% Υλοποίηση ελέγχων των Ενοτήτων 2.3.2-2.3.4

% Υπολογισμός των P_k (pk) της σχέσης (2.3.4.4).

```
pk=zeros(1,n(1,1));
```

```

voithia1=1+dk;
voithia2=1-dk;
for j=1
    pk(j)=0.5*voithia1(j);
end
for j=2:n(1,1) pk(j)=pk(j-1)+0.5*voithia1(j);
end
pk

% Υπολογισμός των  $N_k$  ( $nk$ ) της σχέσης (2.3.4.5).

for j=1
    nk(j)=0.5*voithia2(j);
end
for j=2:n(1,1) nk(j)=nk(j-1)+0.5*voithia2(j);
end
nk

% Υπολογισμός των  $P_k'$  ( $pkn$ ) που είναι οι αντίστοιχες τιμές της  $P_k$ 
υπολογισμένες στα  $\delta_k'$ .

pkn=zeros(1,n(1,1));
voithia11=1+dk1;
voithia21=1-dk1;
for j=1
    pkn(j)=0.5*voithia11(j);
end
for j=2:n(1,1) pkn(j)=pkn(j-1)+0.5*voithia11(j);
end
pkn

% Υπολογισμός των  $N_k'$  ( $nkn$ ) που είναι οι αντίστοιχες τιμές της  $N_k$ 
υπολογισμένες στα  $\delta_k'$ .

for j=1
    nkn(j)=0.5*voithia21(j);
end

```

```

    for j=2:n(1,1) nkn(j)=nkn(j-1)+0.5*voithia21(j);
    end
nkn

% Υπολογισμός της σ.σ.  $R_n(rn)$  ως ειδική περίπτωση της σ.σ.  $S_n^{(a)}$  της
σχέσης (2.3.4.7) για  $\alpha = 0.05$ .

sn05=(1/(n(1,1)^2))*sum((nk-pk+0.5*dk).^2)
rn=sn05

% Κανόνας απόφασης με τη σ.σ.  $R_n(rn)$ . Η κρίσιμη τιμή προκύπτει
από την Ενότητα 2.3.2. Αν rnapor=1 τότε η μηδενική υπόθεση της
συμμετρίας απορρίπτεται.

if rn>1.656 rnapor=1;
else rnapor=0;
end
rnapor

% Υπολογισμός της σ.σ.  $T_n^{(0)}(tn0)$  ως ειδική περίπτωση της σ.σ.  $T_n^{(a)}$ 
της σχέσης (2.3.4.8) για  $\alpha=0$ . Υπενθύμιση: η σ.σ.  $S_n^{(0)}(sn0)$  είναι η σ.σ.
των Srinivasan and Godio (1973).

sn0=(1/(n(1,1)^2))*sum((nk-pk).^2)
rn1=(1/(n(1,1)^2))*sum((nkn-pkn).^2)
tn0=0.5*(sn0+rn1)

% Κανόνας απόφασης με τη σ.σ.  $S_n(sn0)$ . Η κρίσιμη τιμή προκύπτει
από τον Πίνακα 2 του Παραρτήματος Α. Αν sn0apor=1 τότε η μηδενική
υπόθεση της συμμετρίας απορρίπτεται.

if 0.25*sn0>0.4 sn0apor=1;
else sn0apor=0;
end
sn0apor

```

% Κανόνας απόφασης με τη σ.σ. $T_n^{(0)}$ ($tn0$). Η κρίσιμη τιμή προκύπτει από τον Πίνακα 4 του Παραρτήματος Α. Αν $tn0apor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if 0.25*n(1,1)^2*tn0>223.25 tn0apor=1;
else tn0apor=0;
end
tn0apor
```

% Υλοποίηση ελέγχων της Ενότητας 2.3.5 (Baringhaus and Henze (1992))

% Υπολογισμός της σ.σ. KS_n (ksn) από τη σχέση (2.3.5.4).

```
sgn=(1/(4*n(1,1)^2))*sum(wn.^2)
sin=abs(sum(t1))/2*sqrt(n(1,1))
win=abs(sum((1:n(1,1)+1).*t1))/(sqrt(n(1,1))*(n(1,1)+1))
for i=1:n(1,1)-1
    anew(i)=i*sum(t1(i+1:n(1,1)));
end
anew
athr=sum(t1(i+1:n(1,1)))
ksn=(1/((n(1,1)-1)*sqrt(n(1,1))))*max(anew)
```

% Κανόνας απόφασης με τη σ.σ. KS_n (ksn). Η κρίσιμη τιμή προκύπτει από την Ενότητα 2.3.5. Αν $ksnapor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if ksn>0.257 ksnapor=1;
else ksnapor=0;
end
ksnapor
```

% Υπολογισμός της σ.σ. CM_n (cmn) από τη σχέση (2.3.5.5).

```
for i=1:n(1,1)-1 bnew(i)=sum(t1(i+1:n(1,1)));
end
bnew
```



```
cmn=(sum(((1:n(1,1)-1).^2)./(n(1,1)^2).*bnew))/((n(1,1)-1)^2)
```

% Εναλλακτικός προσδιορισμός της σ.σ. S_n (sn) των Srinivasan and Godio (βλέπε και σχέση (2.3.5.7)) και κανόνας απόφασης. Αν snapor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
sn=4*sgn/(n(1,1)^2)
```

```
if sn>0.4 snapor=1;
```

```
else snapor=0;
```

```
end
```

```
snapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.4.1 (Tajuddin (1994))

% Υπολογισμός της σ.σ. S (tajud94) από τη σχέση (2.4.1.6). Επιπρόσθετα υπολογισμός της αναμενόμενης τιμής της (etaj) και της διακύμανσης της (vartaj). Υπολογισμός της σ.σ. Z_S (ztaj) της σχέσης (2.4.1.9).

```
tajud94=sum((1:n(1,1)).*s)
```

```
plithosthetikon=size(find(data>0))
```

```
n1new=plithosthetikon(1)
```

```
n2new=n(1,1)-n1new
```

```
etaj=n1new*(n(1,1)+1)/2
```

```
vartaj=n1new*n2new*(n(1,1)+1)/12
```

```
ztaj=(tajud94-etaj)/sqrt(vartaj)
```

% Κανόνας απόφασης με τη σ.σ. Z_S (ztaj). Αν ztajapor =1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(ztaj)>norminv(1-0.05/2,0,1) ztajapor=1;
```

```
else ztajapor=0;
```

```
end
```

```
ztajapor
```

% Υλοποίηση ελέγχων της Ενότητας 2.4.2 (Modarres and Gastwirth (1998) και βελτίωση αυτού)

% Ορίζεται το περικομμένο ποσοστό p ($p1$) και υπολογίζονται οι συντελεστές βαρύτητας $\varphi(k)$ ($f2$) που ορίστηκαν στη σχέση (2.2.3.2).

```
p2=0.7;
k2=fix(n(1,1)*p2);
f2=zeros(1,n(1,1));
for i=1:n(1,1) if i>n*p2 f2(i)=i-n(1,1)*p2;
end
end
f2
```

% Υπολογισμός της σ.σ. W_p (wp) της σχέσης (2.4.2.2), της αναμενόμενης τιμής της (ewp) και της διακύμανσης της ($varwp$). Υπολογισμός της σ.σ. Z_{MOG} ($zmog$) της σχέσης (2.4.2.5).

```
wpolo=f2.*s;
wp=sum(wpolo((k2+1):n(1,1)))
ewp=0.5*n1new*(1-p2)*(n(1,1)*(1-p2)+1)
varwp=(n1new*n2new)/(12*(n(1,1)-1))*(1-p2)*(n(1,1)*(1-p2)+1)*(n(1,1)*(1-p2)*(3*p2+1)+3*p2-1)
zmog=(wp-ewp)/sqrt(varwp)
```

% Κανόνας απόφασης με τη σ.σ. Z_{MOG} ($zmog$). Αν $zmogapor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(zmog)>norminv(1-0.05/2,0,1) zmogapor=1;
else zmogapor=0;
end
zmogapor
```

ΜΕΡΟΣ Β

Στην ενότητα αυτή, θα δοθεί ο πηγαίος κώδικας που δημιουργήθηκε για την εφαρμογή των ελέγχων συμμετρίας γύρω από άγνωστο κέντρο συμμετρίας, με υλοποίηση του οποίου προέκυψαν τα αποτελέσματα της Ενότητας 5.2.

```
data=[17.4;17.9;17.6;18.1;17.6; 18.9; 16.9; 17.5; 17.8; 17.4; 24.6; 26]
```

```
%Υπολογισμός βοηθητικών μεταβλητών-ποσοτήτων.
```

```
n=size(data);
m=median(data);
m1=mean(data)
x=find(data<m);
c=size(x);
```

```
% Υλοποίηση ελέγχων της Ενότητας 3.1.1 (Τροποποίηση του τεστ Smirnov (1947) και Butler (1969) των Mukherjee et al. (2013))
```

```
% Υπολογισμός της σ.σ.  $D_n$  (dnm) της σχέσης (3.1.1.3).
```

```
for i=1:c(1) a(i,:)=size(find(data<=data(x(i))));
    b(i,:)=size(find(data<=2*m-data(x(i))));
end
a
b
dn=1-a(:,1)'/n(1,1)-b(:,1)'/n(1,1);
dnm=max(dn)
```

% Κανόνας απόφασης με τη σ.σ. D_n (dnm). Για επίπεδο σημαντικότητας $\alpha = 0.05$ και μέγεθος δείγματος 12 η κρίσιμη τιμή είναι 0.2. Αν $dn_{maror} = 1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```

if dnm>0.2 dnmapor=1;
else dnmapor=0;
end
dnmapor

```

% Υλοποίηση ελέγχων της Ενότητας 3.1.2 (Τροποποίηση του τεστ των Rothman and Woodroffe (1972) των Mukherjee *et al.* (2013))

% Υπολογισμός της σ.σ. R_n (rn) της σχέσης (3.1.2.1).

```

y1=sort(data)
for i=1:n(1,1) nb(i,:)=size(find(data<=2*m-y1(i)));
    voith1(i)=(2*n(1,1)-2*i+1)/(2*n(1,1));
    voith2(i)=1/(i);
end
rn=sum(voith2.*((nb(:,1)'/n(1,1))-voith1).^2)

```

% Κανόνας απόφασης με τη σ.σ. R_n (rn). Για επίπεδο σημαντικότητας $\alpha = 0.05$ και μέγεθος δείγματος 12 η κρίσιμη τιμή είναι 0.11. Αν $rnapor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```

if rn>0.11 rnapor=1;
else rnapor=0;
end
rnapor

```

% Υλοποίηση ελέγχων της Ενότητας 3.1.3 (Τροποποίηση του τεστ του McWilliams (1990) των Mukherjee *et al.* (2013))

% Υπολογισμός της σ.σ. R^ ($rasteri$) της σχέσης (3.1.3.2). Προϋποθέτει υπολογισμό της ακολουθίας S_i (διάνυσμα s) της σχέσης (3.1.3.1) και των ποσοτήτων I_k (διάνυσμα $dianitel$) της σχέσης (3.1.3.3).*

```

z=y1-m;
xn=find(z<0);
cn=size(xn);
y=sort(abs(z));

```

```
for i=1:cn(1,1)
for j=1:n(1,1)
    if y(j)==abs(z(xn(i)));
        y(j)=-y(j);
    end
end
end
for j=1:n(1,1)-1 if y(j+1)==y(j);
    y(j+1)=-y(j+1);
end
end
y;
s=zeros(1,n(1,1));
for i=1:n(1,1)
    if y(i)>0 s(i)=1;
    end
end
s;
diani=zeros(1,n(1,1));
for i=2:n if s(i-1)==s(i) & s(i)==1;
    diani(i)=1;
end
end
diani;
dianitel=diani(2:n(1,1));
rasteri=sum(dianitel)
```

% Κανόνας απόφασης με τη σ.σ. R^ (rasteri). Για επίπεδο σημαντικότητας $\alpha = 0.05$ και μέγεθος δείγματος 12 η κρίσιμη τιμή είναι 6. Αν rasteriapor = 1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.*

```
if rasteri>6 rasteriapor =1;
else rasteriapor =0;
end
rasteriapor
```

% Υλοποίηση ελέγχων της Ενότητας 3.1.4 (Τροποποιήσεις του προσημικού τεστ των Mukherjee *et al.* (2013))

% Υπολογισμός της σ.σ. T^ (tasteri) της σχέσης (3.1.4.2). Προϋποθέτει υπολογισμό της ακολουθίας I_i^* (διάνυσμα iasteri) της σχέσης (3.1.4.1).*

```
iasteri=find(data<m1);
tasteriolo=size(iasteri);
tasteri=tasteriolo(1)
```

% Υπολογισμός της σ.σ. T (tneo) της σχέσης (3.1.4.4). Προϋποθέτει υπολογισμό της ακολουθίας I_i (διάνυσμα ideuteri) της σχέσης (3.1.4.5).

```
kn=fix(n(1,1)/2);
ideuteri=zeros(1,kn);
for i=1:kn if y1(n(1,1)+1-i)>2*m-y1(i);
    ideuteri(i)=1;
end
end
tneo=sum(ideuteri)
```

% Κανόνας απόφασης με τη σ.σ. T (tneo). Για επίπεδο σημαντικότητας $\alpha = 0.05$ και μέγεθος δείγματος 12 η κρίσιμη τιμή είναι 9. Αν tneoapor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if tneo>9 tneoapor =1;
else tneoapor =0;
end
tneoapor
```

% Υλοποίηση ελέγχων της Ενότητας 3.1.6 (Antille and Kersting (1977) και Ekstrom and Jammalamadaka (2006))

% Υπολογισμός της σ.σ. S_1 (s1) της σχέσης (3.1.6.1).

```
x11=find(data<m1);
vc=size(x11);
```

```
s1=(1/(n(1,1)^2))*(vc(1)-(n(1,1)/2))
```

% Κανόνας απόφασης με τη σ.σ. S_1 ($s1$). Για επίπεδο σημαντικότητας $\alpha = 0.05$, αν $s1apor = 1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(s1)>=0.98 s1apor =1;
else s1apor =0;
end
s1apor
```

% Υπολογισμός της σ.σ. S_2 ($s2$) της σχέσης (3.1.6.2).

```
count=0;
for i=1:n(1,1)
for j=1:n(1,1)
if i<j w=data(i)+data(j)-2*mean(data);
    if w<0 count=count+1;
    end
end
end
end
s2=(n(1,1)^(-3/2))*(count-sum(1:(n(1,1)-1))*0.5)
```

% Κανόνας απόφασης με τη σ.σ. S_2 ($s2$). Για επίπεδο σημαντικότητας $\alpha = 0.05$, η κρίσιμη περιοχή είναι $|S_2| \geq 0.2828$. Αν $s2apor = 1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(s2)>=0.2828 s2apor =1;
else s2apor =0;
end
s2apor
```

% Υπολογισμός της σ.σ. S_3 ($s3$) της σχέσης (3.1.6.3).

```
count1=0;
for i=1:n(1,1)
```

```

for j=1:n(1,1)
if i<j w1=data(i)+data(j)-2*median(data);
    if w1<0 count1=count1+1;
    end
end
end
end
s3=(n(1,1)^(-3/2))*(count1-sum(1:(n(1,1)-1))*0.5)

```

% Κανόνας απόφασης με τη σ.σ. S_3 ($s3$). Για επίπεδο σημαντικότητας $\alpha = 0.05$, η κρίσιμη περιοχή είναι $|S_3| \geq 0.3481$. Αν $s3apor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```

if abs(s3)>=0.3481 s3apor=1;
else s3apor=0;
end
s3apor

```

% Υπολογισμός της σ.σ. $S_{4,0}$ ($s40$) της σχέσης (3.1.6.4).

```

for i=2:n(1,1) yneo(i)=y1(i)-y1(i-1);
end
yneo;
ynew=yneo(2:n(1,1));
count3=0;
for i=2:kn
    yas(i)=yneo(i)-yneo(n(1,1)-i+2);
    if yas(i)<=0 count3=count3+1;
    end
end
yasz=yas(2:kn)
s40=n(1,1)^(-0.5)*(count3-(kn-1)*0.5)

```

% Κανόνας απόφασης με τη σ.σ. $S_{4,0}$ ($s40$). Για επίπεδο σημαντικότητας $\alpha = 0.05$, η κρίσιμη περιοχή είναι $|S_{4,0}| \geq 1.0323$. Αν $s40apor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.


```

if abs(s40)>=1.0323 s40apor =1;
else s40apor =0;
end
s40apor

```

% Υπολογισμός της σ.σ. $T_2(0.05)$ (t_{2005}) της σχέσης (3.1.6.5), ως ειδική περίπτωση για $\alpha = 0.05$.

```

kn1=fix((n(1,1)-1)/2);
for i=1:kn1
    voith3(i)= i/n(1,1);
    yas1(i)=ynew(i)-ynew(n(1,1)-i);
    if yas1(i)<=0 inea1(i)=0.5;
    else inea1(i)=-0.5;
    end
    if 0.05<=voith3(i) & voith3(i)<=0.5 j(i)=1;
    else j(i)=0;
    end
end
t2005=(n(1,1)^(-0.5))*sum(j.*inea1)

```

% Κανόνας απόφασης με τη σ.σ. $T_2(0.05)$ (t_{2005}). Αν $t_{2005apor}=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```

if abs(t2005)>norminv(1-0.05/2,0,1)*sqrt(0.1008) t2005apor=1;
else t2005apor=0;
end
t2005apor

```

% Υπολογισμός της σ.σ. $T_{4,0}$ (sekstrom) της σχέσης (3.1.6.18). Τιμή $m=3$ δηλώνεται (βλέπε Παρατήρηση 3.1.6.3).

```

n2fix=fix((n(1,1)/2));
mpar=3;

```

% Υπολογισμός των $D_{k,m}$ (dkm) της σχέσης (3.1.6.14) για $k=m$.

```
dkm=y1(n2fix+mpar)-y1(n2fix-mpar+1);
```

% Υπολογισμός της ποσότητας $\hat{\sigma}_{4,\varepsilon}^2$ (vars) της σχέσης (3.1.6.15) για $\varepsilon=0$.

```
for i=mpar:(n(1,1)-mpar) stoixia1(i)=(1-log((y1(i+mpar)-y1(i-mpar+1))/dkm))^2;
```

```
end
```

```
stoixia=stoixia1(mpar:(n(1,1)-mpar));
```

```
vars=(1/16)+(1/(16*(n(1,1)-2*mpar+1)))*sum(stoixia);
```

% Υπολογισμός της ποσότητας $S_{4,\varepsilon}$ (sekstrom) της σχέσης (3.1.6.4) για $\varepsilon=0$ και της σ.σ. $T_{4,0}$ (sekstrom) της σχέσης (3.1.6.18).

```
for i=1:n(1,1)-1
```

```
    dinea(i)=y1(i+1)-y1(i);
```

```
end
```

```
for i=1:kn1
```

```
    yas2(i)=dinea(i)-dinea(n(1,1)-i);
```

```
    if yas2(i)<=0 inea2(i)=0.5;
```

```
    else inea2(i)=-0.5;
```

```
    end
```

```
end
```

```
sekstrom=sum(ineas2)*(n(1,1)^(-0.5))
```

```
t40=sekstrom/sqrt(vars)
```

% Κανόνας απόφασης με τη σ.σ. $T_{4,0}$ (t40). Αν $t40_{apor}=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(t40)>norminv(1-0.05/2,0,1) t40apor=1;
```

```
else t40apor=0;
```

```
end
```

```
t40apor
```

% Υλοποίηση ελέγχων της Ενότητας 3.2.1 (Δειγματικός συντελεστής λοξότητας)

% Υπολογισμός της ποσότητας b_1 ($b1$) της σχέσης (3.2.1.1) και των χρήσιμων ποσοτήτων για τον υπολογισμό της σ.σ. Z_{b_1} ($zb1$) της σχέσης (3.2.1.6).

```
b1=skewness(data);
m2=(1/n(1,1))*sum((data-mean(data)).^2);
m4=(1/n(1,1))*sum((data-mean(data)).^4);
m6=(1/n(1,1))*sum((data-mean(data)).^6);
varso=(m6-6*m2*m4+9*(m2^3))/(m2^3);
zb1=sqrt(n(1,1))*b1/(sqrt(varso))
```

% Κανόνας απόφασης με τη σ.σ. Z_{b_1} ($zb1$). Αν $zb1apor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(zb1)>norminv(1-0.05/2,0,1) zb1apor=1;
else zb1apor=0;
end
zb1apor
```

% Υλοποίηση ελέγχων της Ενότητας 3.2.2 (Cabilio and Masaro (1996) και Miao *et al.* (2006))

% Υπολογισμός της σ.σ. S_k (sk) της σχέσης (3.2.2.3) και κανόνας απόφασης. Για την κρίσιμη τιμή βλέπε Ενότητα 3.2.2. Αν $skapor=1$ τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
sk=sqrt(n(1,1))*(mean(data)-median(data))*(1/sqrt(var(data)))
if abs(sk)>=1.24 skapor=1;
else skapor=0;
end
skapor
```

% Υπολογισμός της σ.σ. T_{Mi} (tmiao) της σχέσης (3.2.2.13) και κανόνας απόφασης. Για την κρίσιμη τιμή βλέπε Ενότητα 3.2.2. Αν tmiaoapor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
jmiao=sqrt(pi/2)*(1/n(1,1))*sum(abs(data-median(data)))
tmiao=(mean(data)-median(data))/jmiao
```

```
if      abs(tmiao)>=norminv(1-0.05/2,0,1)*sqrt(0.5708/n(1,1))
tmiaoapor=1;
else tmiaoapor=0;
end
tmiaoapor
```

% Υλοποίηση ελέγχων της Ενότητας 3.2.3 (Ghosh (2011))

```
for i=1:kn zitoumeno(i)=(y1(i)+y1(n(1,1)+1-i))/2;
end
zitoumeno
for i=1:kn number(i,:)=size(find(data<=zitoumeno(i)));
end
number
test=(1/(n(1,1)^2))*sum(number(:,1))
```

% Υπολογισμός της ποσότητας h_n (hn) της σχέσης (3.2.3.9), των kernel $K(t)$ της σχέσης (3.2.3.11) και της ποσότητας $\hat{f}$ (festim) της σχέσης (3.2.3.8).

```
hn=1.06*min(sqrt(var(data)),iqr(data)/1.34)*n(1,1)^(-1/5)
for i=1:n(1,1) orisma(i)=(median(data)-data(i))/hn;
end
orisma
for i=1:n(1,1) if abs(orisma(i))<sqrt(5) kernel(i)=(3/(4*sqrt(5)))*(1-
0.2*orisma(i)^2);
else kernel(i)=0;
end
end
kernel
festim=(1/(n(1,1)*hn))*sum(kernel)
```

% Υπολογισμός της ποσότητας $\hat{\omega}^2$ ($w2$) της σχέσης (3.2.3.7).

```
w2=1/16-0.25*festim*(1/n(1,1))*sum(abs(data-
mean(data)))+0.25*festim^2*var(data)
```

% Υπολογισμός της σ.σ. Z_G (zghosh) της σχέσης (3.2.3.12).

```
zghosh=sqrt(n(1,1))*abs((test-0.25)/sqrt(w2))
```

% Κανόνας απόφασης με τη σ.σ. Z_G (zghosh). Αν zghoshapor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(zghosh)>norminv(1-0.05/2,0,1) zghoshapor=1;
else zghoshapor=0;
end
zghoshapor
```

*% Υλοποίηση ελέγχων της Ενότητας 3.2.5 (Boos (1982) και Wang *et al.* (2013))*

% Υλοποίηση του ελέγχου του Boos (1982). Υπολογισμός της σ.σ. T_{B_n} (tbos) της σχέσης (3.2.5.5), αφού αρχικά δηλώσουμε την τιμή του Hodges-Lehmann εκτιμητή (hles) που βρέθηκε χρησιμοποιώντας τη συνάρτηση hl.loc(x) της γλώσσας προγραμματισμού R, όπου x το διάνυμα τιμών.

```
hles=17.85
bo1=0;
bo2=0;
for i=1:n(1,1)
    for j=1:n(1,1)
        bo1=bo1+0.5*abs((data(i)+data(j)-2*hles));
    end
end
for i=1:n(1,1)
    for j=1:n(1,1) if i<j bo2=bo2+abs(data(i)-data(j));
    end
end
```

```

end
bo1;
bo2;
tbos=n(1,1)*(-1+bo1/bo2)

% Από τον Table 1 του Boos (1982) προκύπτει ότι
 $P(T_{B_n} \geq 1.8852) < 0.01$  και έτσι η υπόθεση της συμμετρίας δεν
απορρίπτεται.

% Υλοποίηση του ελέγχου των Wang et al. (2013). Υπολογισμός των
σ.σ.  $S_{GMD}$  (sgmd) και  $RS_{GMD}$  (rs) των σχέσεων (3.2.5.8) και
(3.2.5.11), αντίστοιχα.

for j=1:kn
    yplus(j)=y1(n(1,1)+1-j)-median(data);
    yminus(j)=median(data)-y1(j);
end
wneo1=0;
wneo2=0;
for i=1:kn
    for j=1:kn
        if i<j wneo1=wneo1+abs(yplus(i)-yplus(j));
            wneo2=wneo2+abs(yminus(i)-yminus(j));
        end
    end
end
end

sgplus=(2/(kn*(kn-1)))*wneo1;
sgminus=(2/(kn*(kn-1)))*wneo2;
sgmd=log(sgplus/sgminus)
mix=[yplus yminus]
ranksmix=[12 11 10 8 4.5 2 9 6.5 6.5 4.5 2 2]
ranksyplus=ranksmix(1:kn)
ranksyminus=ranksmix(kn+1:(2*kn))
for i=1:kn r1(i)=ranksyplus(kn-i+1);
    r2(i)=ranksyminus(kn-i+1);
end
end

```

```

r1
r2
rgmdplus=(2*sum((1:kn).*r1)-(kn+1)*sum(r1))/(0.5*kn*(kn-1))
rgmdminus=(2*sum((1:kn).*r2)-(kn+1)*sum(r2))/(0.5*kn*(kn-1))
rs=log(rgmdplus/rgmdminus)

% Οι κρίσιμες τιμές του ελέγχου με τη σ.σ.  $RS_{GMD}$  ( $rs$ ) υπολογίζονται
% με βάση τον Πίνακα 7 του Παραρτήματος Α.

% Υλοποίηση ελέγχων της Ενότητας 3.4.1 (Triple test, Randles et al.
% (1980))

% Υπολογισμός της σ.σ.  $\hat{\eta}$  ( $hestim$ ) της σχέσης (3.4.1.2).

h=0;
for i=1:n(1,1)
    for j=1:n(1,1)
        for k=1:n(1,1)
            if i<j<k h=h+1/3*(proximo(data(i)+data(j)-
2*data(k))+proximo(data(i)+data(k)-
2*data(j))+proximo(data(k)+data(j)-2*data(i)));
            end
        end
    end
end
h
hestim=(nchoosek(n(1,1),3))^(-1)*h

% Υπολογισμός της ποσότητας  $\hat{f}_1$  ( $f1es$ ) της σχέσης (3.4.1.10) και της
% ποσότητας  $\hat{\zeta}_1$  ( $z1est$ ) της σχέσης (3.4.1.9).

f1sum=zeros(1,n(1,1));
for i=1:n(1,1)
    for j=1:n(1,1)
        for k=1:n(1,1)
            if j<k & j~=i & i~=k & j~=k
f1sum(i)=f1sum(i)+1/3*(proximo(data(i)+data(j)-

```

```

2*data(k))+proximo(data(i)+data(k)-
2*data(j))+proximo(data(k)+data(j)-2*data(i)));
    end
    end
end
f1sum
f1es=f1sum*2*factorial(n(1,1)-3)/factorial(n(1,1)-1)
z1est=sum((f1es-hestim).^2)/(n(1,1))

```

% Υπολογισμός της ποσότητας $\hat{f}_2$ (f2es) της σχέσης (3.4.1.10) και της ποσότητας $\hat{\xi}_2$ (z2est) της σχέσης (3.4.1.11).

```

f2sum=zeros(n(1,1),n(1,1));
for j=1:n(1,1)
    for k=1:n(1,1)
        for i=1:n(1,1)
            if j~=i & i~=k & j~=k

                f2sum(j,k)=f2sum(j,k)+1/3*(proximo(data(i)+data(j)-
2*data(k))+proximo(data(i)+data(k)-
2*data(j))+proximo(data(k)+data(j)-2*data(i)));
                    end
                end
            end
        end
    end
f2sum
f2sum1=f2sum/(n(1,1)-2)
zivoi=0;
for j=1:n(1,1)
    for k=1:n(1,1)
        if j<k zivoi=zivoi+(f2sum1(j,k)-hestim)^2;
        end
    end
end
end
z2est=zivoi*(nchoosek(n(1,1),2))^(-1)

```


% Υπολογισμός της ποσότητας $\hat{\zeta}_3$ (z3est) της σχέσης (3.4.1.13).

z3est=1/9-(hestim^2)

% Υπολογισμός της ποσότητας σ_n^2 (sn2) της σχέσης (3.4.1.6).

```
sumvar=3*nchoosek(n(1,1)-3,2)*z1est+3*nchoosek(n(1,1)-
3,1)*z2est+z3est;
sn2=(n(1,1)*factorial(3)*factorial(n(1,1)-
3)/factorial(n(1,1)))*sumvar
```

% Υπολογισμός της σ.σ. T_{R1} (tr1) της σχέσης (3.4.1.14).

tr1=sqrt(n(1,1))*hestim/sqrt(sn2)

% Κανόνας απόφασης με τη σ.σ. T_{R1} (tr1). Αν tr1apor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(tr1)>norminv(1-0.05/2,0,1) tr1apor=1;
else tr1apor=0;
end
tr1apor
```

*% Υλοποίηση ελέγχων της Ενότητας 3.4.2 (Gupta (1967), Antille *et al.* (1982))*

% Υπολογισμός της σ.σ. $T_1(0)$ (t10) της σχέσης (3.4.2.12) για την ειδική περίπτωση που $\alpha = 0$.

```
diaf=abs(data-m);
r=[6.5 4.5 2.0 8.0 2.0 10.0 9.0 4.5 2.0 6.5 11.0 12.0]';
t10=(n(1,1)^(-0.5))*ga(r/(2*n(1,1)+1),0)*proximo(data-m)
```

% Κανόνας απόφασης με τη σ.σ. $T_1(0)$ (t10). Αν t10apor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```
if abs(t10)>=0.3480 t10apor=1;
```

```

else t10apor=0;
end
t10apor

```

% Υπολογισμός της σ.σ. $T_1\left(\frac{1}{6}\right)$ (t116) της σχέσης (3.4.2.12) για την ειδική περίπτωση που $\alpha = \frac{1}{6}$.

```
t116=(n(1,1)^(-0.5))*(ga(r/(2*n(1,1)+1),1/6)*prosimo(data-m))
```

% Κανόνας απόφασης με τη σ.σ. $T_1\left(\frac{1}{6}\right)$ (t116). Αν t116apor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```

if abs(t116)>=0.2581 t116apor=1;
else t116apor=0;
end
t116apor

```

% Υλοποίηση ελέγχων της Ενότητας 3.4.3 (Finch (1982))

% Υπολογισμός της σ.σ. $T_3(0.05)$ (t3005) της σχέσης (3.4.3.7) για την ειδική περίπτωση που $\alpha = 0.05$ και κανόνας απόφασης. Προϋποθέτει τον υπολογισμό των ποσοτήτων V_i της σχέσης (3.4.3.8). Για την κρίσιμη τιμή βλέπε Παρατήρηση 3.4.3.2. Αν t3005apor=1 τότε η μηδενική υπόθεση της συμμετρίας απορρίπτεται.

```

for i=1:kn1
    voith31(i)=i/(n(1,1)+1);
    v(i)=(ynew(i)-ynew(n(1,1)-i))/(ynew(i)+ynew(n(1,1)-i));

    if 0.05<=voith31(i) & voith31(i)<=0.5 j1(i)=1;
    else j1(i)=0;
    end
end
t3005=(n(1,1)^(-0.5))*sum(j1.*v)

```

```
if t3005>=norminv(1-0.05,0,1)*sqrt(0.1293) t3005apor=1;
else t3005apor=0;
end
t3005apor
```

Οι συναρτήσεις "prosimo" και "ga" που χρησιμοποιούνται σε κάποιους από τους παραπάνω κώδικες είναι:

```
function y = prosimo(x)
x(x<0)=-1;
x(x==0)=0;
x(x>0)=1;
y=x;
```

και

```
function y = ga(x,a)
y=min(x,0.5-a);
```

Χρησιμοποιώντας την R και το πακέτο lawstat (Joseph L. Gastwirth; Yulia R. Gel, W. L. Wallace Hui, Vyacheslav Lyubchich; Weiwen Miao; Kimihiro Noguchi) μέσω της συνάρτησης symmetry.test μας δίνεται η δυνατότητα να επιλέξουμε ένα από τα τεστ Mira (1999), Cabilio and Masaro (1996) και Miao, Gel and Gastwirth (2006). Για παράδειγμα:

Test of Symmetry - Mira Test

```
>symmetry.test(x, option = "mira.test")
```

Test Statistic = 1.6057, p-value = 0.1083

Test of Symmetry - Cabilio-Masaro Test

```
> symmetry.test(x, option ="cabilio.masaro.test")
```

Test Statistic = 1.9436, p-value = 0.05195

Test of Symmetry - MGG Test

```
>symmetry.test(x)
```

Test Statistic = 2.9616, p-value = 0.003061.

Τέλος, χρησιμοποιώντας την R και τη συνάρτηση `wilcox.test` μπορεί να υλοποιηθεί ο έλεγχος του Wilcoxon για δοθείσα διάμεσο, ενώ χρησιμοποιώντας τη συνάρτηση `simple.median.test` μπορεί να υλοποιηθεί το προσημικό τεστ για δοθείσα διάμεσο.

ABSTRACT OF THE MSc DISSERTATION

Title: Testing Symmetry of univariate distributions. A critical review.

Symmetry is an important assumption in statistical methodology, since several statistical procedures require symmetry of the underlying distributions, while the presence or absence of symmetry plays an important role in terms of deciding what parameter to estimate. Simultaneously, testing for symmetry is one of the oldest classical problems and there exists a vast literature on tests of univariate symmetry. The work in this area can be broadly classified into two groups: a) procedures for testing symmetry around a known median, and b) approaches for testing symmetry around an unknown median. The goal of the researchers was and continue to be the design of tests that have good power for interesting alternatives to symmetry and are either asymptotically distribution-free or at least have a type I error rate that stays relatively constant for distributions satisfying the null hypothesis. Our main aim in this MSc Thesis dissertation is to review the most popular methods of testing univariate symmetry. In this frame, the present dissertation consists of five chapters, an appendix and the bibliography. We briefly present the subject-matter of each one in the sequel.

Chapter 1 (Introduction) explains the importance of testing univariate symmetry and presents some useful definitions. Also, a short

summary of this dissertation is presented. In **Chapter 2 (Testing symmetry around a known value)**, a critical review of the most popular existing methods of testing univariate symmetry around a known value is given, while in Chapter 3 (**Testing symmetry around an unknown value**) a critical review of the approaches for testing symmetry around an unknown value is given. The most frequent measures of the performance of a test for testing symmetry is its ability to control the type I error rate and its power. In **Chapter 4 (Comparison study)** a review of the simulation studies related to the performance of the tests is given. Based on this review, some recommendations are given. In **Chapter 5 (Application)** the methods presented in the previous chapters are illustrated via two numerical examples. For this purpose, we created some m_files in Matlab as well as we used some codes in R. These codes are given in **Appendix B**, while in **Appendix A** some useful tables are given. Finally, the **Bibliography** on which the dissertation was based is provided.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ahmad, I.A. and Li, Q. (1997). Testing symmetry of an unknown density function by kernel method. *J. Nonparametric Statistic.*, **7**, 279-293.

Aki, S. (1987). On nonparametric tests for symmetry. *Ann. Inst. Stat. Math.*, **39**, 457-472.

Antille , A. and Kersting, G. (1977). Test for Symmetry. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **39**, 235-255.

Antille, A., Kersting, G. and Zucchini, W. (1982). Testing symmetry. *J am Stat Assoc*, **77**(379): 639-646.

Auda, H. (2006). New Tests of Univariate Symmetry Based on the Gini Mean Difference. Ph.D. thesis, Kalamazoo, MI: Western Michigan University.

Baklizi, A. (2003). A conditional distribution runs test for symmetry. *J. Nonparametric Statistic.*, **15**, 713-718.

Baklizi, A. (2007). Testing symmetry using a trimmed longest run statistic. Austral. New Zealand. *J. Statist.*, **49**, 339-347.

Baklizi, A. (2008). Improving the power of the hybrid test. *Int. J. Contemp. Math. Sci.*, **3**, 497-499.

Balakrishnan, N. and Koutras, M. V. (2002). Runs and Scans with Applications. New York: *Wiley*.

Baringhaus, L. and Henze, N. (1992). A Characterization of and new Consistency Tests for Symmetry. *Commun. Statist.*, **21**, 1555-1566.

- Bhattacharya, P. K., Gastwirth, J. L. and Wright, A. L.** (1982). Two modified Wilcoxon tests for symmetry about an unknown location parameter. *Biometrika*, **69**, 377-382.
- Blough, D.K.** (1989). Multivariate symmetry via projection pursuit. *Ann. Inst. Stat. Math.*, **41**, 461-475.
- Blom, G., Holst, L. and Sandell, D.** (1994). Problems and Snapshots from the World Probability. Berlin: *Spriengel*.
- Boos, D.** (1982). A Test for Asymmetry Associated with the Hodges-Lehmann Estimator. *Journal of the American Statistical Association*, **41**, 647-651.
- Butler, C.** (1969). A Test for Symmetry Using the Sample Distribution Function. *The Annals of Mathematical Statistics*, **40**, 2211-2214.
- Cabilio, P. and Masaro, J.** (1996). A simple test of symmetry about an unknown median. *Can. J. Statist*, **24**, 349-361.
- Chatterjee, S. K. and Sen, P. K.** (1973). On Kolmogorov-Smirnov type tests for symmetry. *Amm. Inst. Stat. Math.*, **25**, 288-300.
- Cheng W. and Balakrishnan, N.** (2004). A modified sign test for symmetry. *Commun. Statist. Simul. Comput.*, **33**, 703-709.
- Cohen, J. P. and Menjoge, S. S.** (1988). One-sample runs test of symmetry. *Journal of Statistical Planning and Inference*, **18**, 93-100.
- Corzo, J. and Babativa, G.** (2011). A modified runs test for symmetry. *Journal of Statistical Computation and Simulation*, **83**, 984-991.
- Csörgö, S. and Heathcote, C. R.** (1982). Some results concerning symmetric distributions. *Bull. Aust. Math. Soc.*, **25**, 327-335.
- Csörgö, S. and Heathcote, C. R.** (1983). The theory of functional least squares. *J. Aust. Math. Soc. A*, **34**, 336-355.

Csörgö, S. and Heathcote, C. R. (1987). Testing for Symmetry. *Biometrika*, 177-184.

Davis, C. E. and Quade, D. (1978). U-Statistics for Skewness or Symmetry. *Communications in Statistics*, 7, 413-418.

Diks, C. and Tong, H. (1999). A test for symmetries of multivariate probability distributions. *Biometrika*, 86, 605-614.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 20, 393-403.

Durbin, J. and Knott, M. (1972). Components of Cramer-von Mises Statistics. I. *Journal of the Royal Statistical Society, Ser. B*, 34, 290-307.

Ekström, M. and Jammalamadaka, S. R. (2006). An asymptotically distribution-free test of symmetry. . *Journal of Statistical Planning and Inference*, 137, 799-810.

Feller, W. (1968). An Introduction to Probability Theory and its Applications: Volume 1. New York: *Wiley*.

Fernique, X. (1971). Règularité de processus gaussiens. *Invent. Math.*, 12, 304-320.

Finch, S. J. (1977). Robust univariate test of symmetry. *Journal of the American Statistical Association*, 72, 387-392.

Fisher, R. A. (1929). Moments and product moments of sampling distributions. *Proceedings of the London mathematical society (Series 2)*, 3, 199-238.

Franch, J. B. and Contreras, D. (2002). A Pearsons test for symmetry with an application to the Spanish business cycle. *Spanish Econ. Rev.*, 4, 221-238.

- Fraser, D. A. S.** (1957). Most powerfull rank-type tests. *The Annals of Mathematical Statistics*, **28**, 1040-1043.
- Gastwirth, J. L.** (1965). Percentille modification of two sample rank tests. *Journal of the American Statistical Association*, **15**, 713-718.
- Gastwirth, J. L.** (1971). On the sign test for symmetry. *Journal of the American Statistical Association*, **66**, 821-823.
- Ghosh, K.** (2011). Advances in Directional and Linear Statistics, A New Nonparametric Test of Symmetry. Springer *Heidelberg Dordrecht London New York*, 69-83.
- Gibbons, J.** (1971). *Nonparametric Statistical Inference*. New York: McGraw-Hill.
- Gibbons, J. and Chakraborti S.** (1992). *Nonparametric Statistical Inference*, ed. by Marcel Dekker, New York.
- Gupta, M. K.** (1967). An asymptotically nonparametric test of symmetry. *The Annals of Mathematical Statistics*, **38**, 849-866.
- Hajek, J. and Sidak, Z.** (1967). Theory of Rank Tests. New York: Academic Press.
- Heathcote, C.R., Rachev, S.T., and Cheng, B.** (1995). Testing multivariate symmetry. *J. Multivariate Anal.*, **54**, 91-112.
- Henze, N.** (1993). On the consistency of a test for symmetry based on a runs statistic. *Nonparametric Statistics*, **3**, 195-199.
- Hill, D. and Rao, P. V.** (1977). Tests of Symmetry Based on Cramer-von Mises Statistics. *Biometrika*, **64**, 489-494.
- Hill, D. and Rao, P. V.** (1981). Tests of symmetry based on Watson statistics. *Commun. Stat. Theory Methods*, **10**, 1111-1125.

Hoeffding, W. (1948). A class of statistics with asymptotic normal distribution. *The Annals of Mathematical Statistics*, **19**, 293-325.

Hollander, M. (1981). Testing for Symmetry. Technical Report. Florida State University.

Hollander, M. and Wolfe, D. A. (1973). Nonparametric Statistical Methods. New York: *Wiley*.

Hotelling, H. and Solomons, L. M. (1932). The limits of a measure of skewness. *The Annals of Mathematical Statistics*, **3**, 141-142.

Huffer, F. and Park, C. (2007). A test for elliptical symmetry. *J. Multivariate Anal.*, **98**, 256-281.

Huskova, M. (1984). Hypothesis of Symmetry. *Elsevier Science Publishers*, **4**, 63-78.

Johnson, N. L. (1949). Systems of frequency curves generated by the methods of translation. *Biometrika*, **36**, 149-176.

Koopman, P. A. R. (1979). Testing symmetry with a procedure combining the sign test and the sign rank test. *Stat. Neerl.*, **33**, 137-141.

Koutrouvelis, I. A. (1985). Distribution-free procedures for location and symmetry inference problems based on the empirical characteristic function. *Scand. J. Stat.*, **12**, 257-269.

Koziol, J. (1980). On a Cramer-von Mises type statistic for symmetry. *Journal of the American Statistical Association*, **75**, 161-167.

Kvam, P. L. and Vidakovic, B. (2007). Nonparametric statistics with applications to science and engineering. *Wiley Series in Probability and Statistics*.

Lehmann, E. L. (1963a). Robust estimation in analysis of variance. *The Annals of Mathematical Statistics*, **34**, 957-966.

Lehmann, E. L. (1963b). Nonparametric confidence intervals for a shift parameter. *The Annals of Mathematical Statistics*, **34**, 1507-1512.

Li, R. Z., Fang, K. T. and Zhu, L. X. (1997). Some Q-Q probability plots to test spherical and elliptical symmetry. *J. Comput. Graph. Statist.*, **6**, 435-450.

Markus, M. B. and Shepp, L. A. (1971). Sample behavior of Gaussian process. *Proc. 6th Berkeley Symp.*, **2**, 423-442.

McWilliams, T. P. (1990). A distribution-free test for symmetry based on a runs statistic. *Journal of the American Statistical Association*, **85**, 1130-1133.

Miao, W., Gel, Y. R., and Gastwirth J. L. (2006). A new test of symmetry about an unknown median. New Jersey: *World Scientific*, 199-214.

Mira, A. (1999). Distribution free-test for symmetry based on Bonferroni's measure. *Journal of Applied Statistics*, **26**, 959-972.

Modarres, R. and Gastwirth, J. L. (1996). A modified runs test for symmetry. *Statistics and Probability Letters*, **31**, 107-112.

Modarres, R. and Gastwirth, J. L. (1998). Hybrid test for the hypothesis of symmetry. *J Appl. Statist.*, **25**, 777-783.

Mukherjee, A., Abd-Elfattah, A. M. and Pukait, B. (2013). A rule of thumb for testing symmetry about an unknown median against a long right tail. *Journal of Statistical Computation and Simulation*, 1-18.

Niederhausen, H. (1979). Symmetry, Butler-Smirnov test of. *Encyclopedia of Statistics*, 9499-9502.

Orlov, A. I. (1972). On Testing the Symmetry of Distribution. *Theory of Probability and Its Applications*, **17**, 357-361.

Owen, B. (1962). Handbook of Statistical Hypothesis. *Wiley*, New York.

Pearson, K. (1905). Das fehlergesetz und seine verallgemeinerungen durch Fechner und Pearson. A rejoinder. *Biometrika*, **4**, 169-212.

Randles, R. H., Fligner, M. A., Policello, G. E., and Wolfe, D. A. (1980). An Asymptotically Distribution-Free Test for Symmetry Versus Asymmetry. *Journal of the American Statistical Association*, **75**, 168-172.

Reynolds, M. R. (1975). A sequential signed-rank test for symmetry. *The Annals of Mathematical Statistics*, **75**, 168-172.

Rothman, E. D. and Woodroffe, M. (1972). A Cramer-von Mises Type Statistic for Testing Symmetry. *The Annals of Mathematical Statistics*, **43**, 2035-2038.

Savage, J. R. (1959). Contribution to the theory of rank order statistics; one sample case. *The Annals of Mathematical Statistics*, **30**, 1018-1023.

Shepp, L. (1967). Radon-Nikodym derivatives of Gaussian measures. *The Annals of Mathematical Statistics*, **37**, 321-354.

Silverman, B. W. (1986). Density Estimation for Statistical and Data Analysis. *Chapman and Hall, Monographs on Statistics and Applied Probabilities*.

Smirnov, N. V. (1947). On Criteria for the Symmetry of Distribution Laws of Random Variables. *Doklady Akademii Nauk SSSR*, 11-14.

Srinivasan, R. and Godio, L. B. (1974). A Cramer-von Mises Type Statistic for Testing Symmetry. *Biometrika*, **61**, 196-198.

Tajuddin, I. H. (1994). Distribution-free test for symmetry based on Wilcoxon two-sample test. *Journal of Applied Statistics*, **21**, 2035-2038.

Ury, H. K. and Forrester, J. C. (1970). *Amer. Statisti.*, **24**, 25-26.

- Van der Waerden, B. L. and Nievergelt, E.** (1953). *Tfeln zum Vergleich zweier Stichproben mittels χ -Tests und Zeichentest*, Springer, Berlin.
- Wald, A. and Wolfowitz, J.** (1940). On a test whether two samples are from the same population. *Ann. Math. Statistics*, **11**, 147-162.
- Wand, M. P. and Jones, M. C.** (1995). Kernel Smoothing. *Chapman and Hall, Monographs on Statistics and Applied Probabilities*. Taylor and Francis Group.
- Wank, J. C., Auda, H. and Niewiadomska-Bugaj, M.** (2013). Comparing a New Gini Test with Other Symmetry Tests When Median is Unknown. *Communications in Statistics-Simulation and Computation*, Taylor and Francis Group, **42**, 1616-1627.
- Wilcoxon, F.** (1945). Individual comparisons by ranking methods. *Biometrics*, **1**, 80-83.
- Χρυσαφίνου, Ο.** (2008). Εισαγωγή στις στοχαστικές ανελίξεις. *Εκδόσεις Σοφία*.
- Yule, G. U.** (1911). Introduction to the Theory of Statistics. *Charles Griffin and Co, London*.
- Zhu, L.-X. and Neuhauss, G.** (2003). Conditional tests for elliptical symmetry. *J. Multivariate Anal.*, **84**, 284-298.