ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

Διδακτορικη Διατριβη

του

ΓΕΩΡΓΙΟΥ ΑΥΛΟΓΙΑΡΗ

Τοπικές ϕ-Αποκλίσεις στη Στατιστική Θεωρία Πληροφοριών και Εφαρμογές σε Ελέγχους Στατιστικών Υποθέσεων και Επιλογής Μοντέλου

ΙΩΑΝΝΙΝΑ, 2017

Τριμελής Συμβουλευτική Επιτροπή

Καθηγητής Κωνσταντίνος Α. Ζωγράφος

Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων

(Επιβλέπων Καθηγητής)

Αναπληρωτής Καθηγητής Αθανάσιος Χ. Μιχέας

Τμήμα Στατιστικής, Πανεπιστήμιο Missouri-Columbia

Επίκουρος Καθηγητής Απόστολος Δ. Μπατσίδης

Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων

Η παρούσα Διδακτορική Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Διδακτορικού Διπλώματος που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 7/9/2017 από την επταμελή Εξεταστική Επιτροπή, η οποία αποτελούνταν από τους κ.κ

Κωνσταντίνο Ζωγράφο, Καθηγητή του Τμήματος Μαθηματικών, Πανεπιστήμιο Ιωαννίνων (Επιβλέποντα Καθηγητή),

Αθανάσιο Μιχέα, Αναπληρωτή Καθηγητή του Τμήματος Στατιστικής, Πανεπιστήμιο Missouri-Columbia (Μέλους Τριμελούς),

Απόστολο Μπατσίδη, Επίκουρο Καθηγητή του Τμήματος Μαθηματικών, Πανεπιστήμιο Ιωαννίνων (Μέλους Τριμελούς),

Σωτήριο Λουκά, Καθηγητή του Τμήματος Μαθηματικών, Πανεπιστήμιο Ιωαννίνων,

Κωνσταντίνα Σκούρη, Αναπληρώτρια Καθηγήτρια του Τμήματος Μαθηματικών, Πανεπιστήμιο Ιωαννίνων,

Κωνσταντίνο Αδαμίδη, Καθηγητή του Τμήματος Βιολογικών Εφαρμογών και Τεχνολογιών, Πανεπιστήμιο Ιωαννίνων,

Σταύρο Κουρούκλη, Καθηγητή του Τμήματος Μαθηματικών, Πανεπιστήμιο Πατρών.

Η έγκριση της Διδακτορικής Διατριβής από το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων δεν υποδηλοί αποδοχή των γνωμών του συγγραφέα (Άρθρο 202 §2, Ν. 5343/1932 και Άρθρο 50 §Β, Ν.1268/1982)

ΥΠΕΥΘΥΝΗ ΔΗΔΩΣΗ

«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 υποψηφίου διδάκτορα)

Αφιερώνεται στους γονείς μου

Θωμά και Στεργιανή

Ευχαριστίες

Η παρούσα Διδακτορική Διατριβή εκπονήθηκε στον Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του Τμήματος Μαθηματικών της Σχολής Θετικών Επιστημών του Πανεπιστημίου Ιωαννίνων, με Επιβλέποντα τον Καθηγητή κ. Κωνσταντίνο Ζωγράφο και μέλη της Τριμελούς Συμβουλευτικής Επιτροπής τον Αναπληρωτή Καθηγητή κ. Αθανάσιο Μιχέα του Τμήματος Στατιστικής του Πανεπιστημίου Missouri-Columbia και τον Επίκουρο Καθηγητή κ. Απόστολο Μπατσίδη του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Από τη θέση αυτή, θα ήθελα να εκφράσω την ευγνωμοσύνη και τις ευχαριστίες μου σε όσους συνέλαβαν στην πραγματοποίηση αυτής της διατριβής.

Πρωτίστως, στον δάσκαλο μου κ. Κωνσταντίνο Ζωγράφο που με προθυμία δέχτηκε να αναλάβει την επίβλεψη και εισηγήθηκε το θέμα της διατριβής. Τον ευχαριστώ θερμά, για τις πολύτιμες υποδείξεις του, τη συνεχή καθοδήγηση, την υπομονή του και τον πολύτιμο χρόνο που διέθεσε. Χωρίς τις συμβουλές του και την ηθική στήριξή του η παρούσα διατριβή θα αντιμετώπιζε σημαντικές δυσκολίες υλοποίησης.

Επίσης, ευχαριστώ θερμά τον Αναπληρωτή Καθηγητή κ. Αθανάσιο Μιχέα για τη συμμετοχή του στην Τριμελή Συμβουλευτική Επιτροπή, το συνεχές ενδιαφέρον του και τον πολύτιμο χρόνο που διέθεσε. Οι αμέτρητες ώρες τηλεδιάσκεψης μαζί του και με τον δάσκαλό μου κ. Κωνσταντίνο Ζωγράφο θα είναι κάτι που θα με ακολουθεί συνέχεια. Ακόμη, ευχαριστώ ιδιαίτερα τον Επίκουρο Καθηγητή του Τμήματος Μαθηματικών κ. Απόστολο Μπατσίδη για τις υποδείξεις του ως μέλους της Τριμελούς Συμβουλευτικής Επιτροπής.

Θα ήθελα επίσης να ευχαριστήσω τους κ.κ Σωτήριο Λουκά, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, Κωνσταντίνα Σκούρη, Αναπληρώτρια Καθηγήτρια του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, Κωνσταντίνο Αδαμίδη, Καθηγητή του Τμήματος Βιολογικών Εφαρμογών και Τεχνολογιών του Πανεπιστημίου Ιωαννίνων, Σταύρο Κουρούκλη, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Πατρών, για τη συμμετοχή τους στην Επταμελή Εξεταστική Επιτροπή και τις παρατηρήσεις τους.

Επίσης, ευχαριστώ το Πανεπιστήμιο Ιωαννίνων και τον Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του Τμήματος Μαθηματικών για την τεχνική και υλική βοήθεια που μου παρείχε.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την κάθε είδους συμπαράστασή τους.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Με τον όρο «πληροφορία» εννοούμε το ποσό κατά το οποίο ελαττώνεται η αβεβαιότητα σχετικά με την πραγματοποίηση ενός γεγονότος. Η έννοια του μέτρου ή του «ποσού» της πληροφορίας είναι θεμελιώδους σημασίας στη Στατιστική και έχει εισαχθεί από τον Fisher (1925). Από τότε έχουν προταθεί και μελετηθεί διάφορα μέτρα πληροφορίας που ταξινομούνται σε τρεις κατηγορίες: παραμετρικά ή τύπου Fisher, μέτρα τύπου απόκλισης (divergence) και μέτρα τύπου εντροπίας (βλέπε, Ferentinos and Papaioannou, 1981). Για τη θεμελίωση ενός μέτρου πληροφορίας άλλοι ερευνητές θέτουν αξιώματα που απορρέουν από τη φύση της πληροφορίας και το πλαίσιο στο οποίο αυτή αναφέρεται, ενώ άλλοι θεμελιώνουν τα προτεινόμενα μέτρα με βάση τις ιδιότητες και τις εφαρμογές τους. Ο κλάδος της Στατιστικής που επικεντρώνεται στην έννοια της «πληροφορίας», τη θεμελίωση μέτρων πληροφορίας, τις ιδιότητες αυτών και τις εφαρμογές τους σε διάφορα προβλήματα και επιστημονικά πεδία, παρέχοντας επιπρόσθετα εργαλεία για τη μέτρηση της απόκλισης μεταξύ δύο μέτρων πιθανότητας ή μεταξύ των αντίστοιχων πυκνοτήτων τους, ονομάζεται Στατιστική Θεωρία Πληροφοριών. Ως εκ τούτου, κατά τη διάρκεια των τελευταίων δεκαετιών, ο κλάδος αυτός, αποτελεί βασικό εργαλείο για την ανάπτυξη μεθοδολογιών στη Μαθηματική Στατιστική. Αντικείμενο της παρούσας διατριβής είναι η εισαγωγή μιας ευρείας κλάσης μέτρων τοπικής απόκλισης (local divergence measures) μεταξύ δύο μέτρων πιθανότητας ή μεταξύ των αντίστοιχων πυκνοτήτων τους, η θεμελίωση και η μελέτη των ιδιοτήτων της καθώς και η αξιοποίησή της στον έλεγχο στατιστικών υποθέσεων και την επιλογή κατάλληλου παραμετρικού μοντέλου. Στο πλαίσιο αυτό, στο εισαγωγικό αυτό κεφάλαιο δίνεται, αρχικά, μια γενική και ταυτόχρονα σύντομη ανασκόπηση στα μέτρα πληροφορίας τύπου απόκλισης, με το ενδιαφέρον να επικεντρώνεται ειδικότερα στις ϕ-αποκλίσεις. Παρατίθενται οι εκφράσεις των κυριότερων μέτρων και σκιαγραφούνται βασικές ιδιότητές τους. Τέλος, παρατίθεται σε περίληψη το περιεχόμενο των κεφαλαίων της διατριβής και επισημαίνεται η συμβολή του πρωτότυπου μέρους της.

1.1 Μέτρα πληροφορίας τύπου απόκλισης

Όπως έχει ήδη αναφερθεί η έννοια της πληροφορίας έχει εισαχθεί στη βιβλιογραφία από τον Fisher (1925) με τον ορισμό της δεύτερης τάξης ροπής της μερικής παραγώγου του λογάριθμου της πιθανοφάνειας, ως ένα μέτρο του ποσού της πληροφορίας που περιέχεται στα δεδομένα σχετικά με την άγνωστη παράμετρο του πληθυσμού από τον οποίο προέρχονται. Ειδικότερα, έστω ο μετρήσιμος χώρος $(\mathcal{X}, \mathcal{A})$, η παραμετρική οικογένεια μέτρων πιθανότητας $\{P_\theta : \theta \in \Theta\}$, με Θ να είναι ένα ανοιχτό υποσύνολο του R^M, $M \geq 1$, και το σ-πεπερασμένο μέτρο μ στο χώρο αυτό, έτσι ώστε $P_\theta \ll \mu$, $\theta \in \Theta$. Τότε το μέτρο πληροφορίας του Fisher (1925) ορίζεται από τη σχέση:

$$I_F(\theta) = \begin{cases} \int_{\mathcal{X}} \left(\frac{\partial \log f_\theta(x)}{\partial \theta}\right)^2 f_\theta(x) d\mu(x), & \text{αν } M = 1 \\ \left(\int_{\mathcal{X}} f_\theta(x) \frac{\partial \log f_\theta(x)}{\partial \theta_i} \frac{\partial \log f_\theta(x)}{\partial \theta_j} d\mu(x)\right)_{M \times M}, \ i, j = 1, ..., M, & \text{αν } M > 1, \end{cases}$$

όπου $f_\theta(x)$ είναι η Radon-Nikodym παράγωγος του P_θ ως προς μ, δηλαδή $f_\theta(x) = dP_\theta(x)/d\mu(x)$ και με $(a_{ij})_{M \times M}$ συμβολίζεται ο $M \times M$ πίνακας με a_{ij} να είναι το στοιχείο της i-οστής γραμμής και j-οστής στήλης, $i, j = 1, ..., M$.

Παρατήρηση 1.1. *Χάριν πληρότητας ορίζεται στη συνέχεια η έννοια της απόλυτης συνέχειας της Θεωρίας Μέτρου που χρησιμοποιήθηκε στον παραπάνω ορισμό του μέτρου πληροφορίας του Fisher και θα αναφερθεί και στη συνέχεια της διατριβής. Στο πλαίσιο αυτό, έστω ο μετρήσιμος χώρος $(\mathcal{X}, \mathcal{A})$ και μ, ν δύο προσημασμένα ή θετικά μέτρα στον $(\mathcal{X}, \mathcal{A})$. Το μέτρο ν θα λέγεται ότι είναι απόλυτα συνεχές ως προς το μέτρο μ αν για κάθε $A \in \mathcal{A}$ με $\mu(A) = 0$ ισχύει $\nu(A) = 0$ και γράφουμε $\nu \ll \mu$.*

Το μέτρο πληροφορίας του Fisher ορίζεται για κάθε παραμετρική οικογένεια κατανομών, παρέχει ενδιαφέρουσες θεωρητικές πληροφορίες και στατιστικές ιδιότητες υπό την προϋπόθεση ότι ικανοποιούνται συγκεκριμένες συνθήκες ομαλότητας της $f_\theta(x)$ (Ferentinos and Papaioannou, 1981, Papaioannou, 1985). Ένα σπουδαίο αποτέλεσμα επαγόμενο από την πληροφορία Fisher είναι η γνωστή ανισότητα Cramér-Rao, που δηλώνει πως το αντίστροφο της πληροφορίας Fisher είναι ένα κάτω φράγμα της διακύμανσης για κάθε αμερόληπτη εκτιμήτρια της άγνωστης παραμέτρου θ.

Έπειτα από τον ορισμό του μέτρου πληροφορίας του Fisher προτάθηκαν και μελετήθηκαν στη βιβλιογραφία διάφορα άλλα μέτρα πληροφορίας, τα οποία μπορούν να ταξινομηθούν σε τρεις κατηγορίες: παραμετρικά ή τύπου Fisher, μέτρα τύπου απόκλισης (divergence) και μέτρα τύπου εντροπίας. Καθώς στην παρούσα διατριβή θα οριστούν, μελετηθούν και θα εφαρμοστούν στον έλεγχο στατιστικών υποθέσεων, τοπικά μέτρα τύπου απόκλισης και μάλιστα ϕ-απόκλισης στην παράγραφο αυτή το ενδιαφέρον θα επικεντρωθεί στα μέτρα τύπου απόκλισης και ειδικότερα ϕ-απόκλισης που ορίζονται σε ολόκληρο το πεδίο ορισμού των κατανομών που στηρίζονται.

Η έννοια της απόκλισης είναι θεμελιώδους σημασίας στα μαθηματικά και σε όλα σχεδόν τα επιστημονικά πεδία. Ένα μέτρο απόκλισης χρησιμοποιείται ως ένας δείκτης εγγύτητας μεταξύ δύο πληθυσμών ή των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας που περιγράφουν τους εν' λόγω πληθυσμούς. Ως εκ τούτου, τα μέτρα τύπου απόκλισης έχουν σημαντικό ρόλο στη Θεωρία Πιθανοτήτων και στη Μαθηματική Στατιστική. Για παράδειγμα, ο έλεγχος Kolmogorov-Smirnov βασίζεται σε ένα μέτρο απόκλισης μεταξύ της εμπειρικής κατανομής και της αντίστοιχης συνάρτησης κατανομής που καθορίζεται από τη μηδενική υπόθεση. Ο κλασικός χ^2 έλεγχος καλής προσαρμογής (chi-square goodness-of-fit) βασίζεται σε ένα μέτρο απόκλισης μεταξύ των θεωρητικών πιθανοτήτων και των αντίστοιχων εμπειρικών. Πολλές άλλες στατιστικές διαδικασίες στηρίζονται σε ένα μέτρο απόκλισης μεταξύ δύο κατανομών πιθανότητας. Στη συνέχεια δίνονται οι συναρτησιακές εκφράσεις των κυριότερων μέτρων πληροφορίας τύπου απόκλισης για τη διευκόλυνση της μελέτης των υπολοίπων κεφαλαίων της διατριβής.

Το πιο γνωστό μέτρο απόκλισης είναι το μέτρο των Kullback-Leibler, το οποίο εισήχθη στη βιβλιογραφία από τους Kullback and Leibler (1951) και μελετήθηκε εκτενώς από τον Kullback (1959) στο σύγγραμμά του με τίτλο *Information Theory and Statistics.* Το προαναφερθέν σύγγραμμα μπορεί να θεωρηθεί ότι σηματοδοτεί το ξεκίνημα της Στατιστικής Θεωρίας Πληροφοριών, παρόλο που πέρασαν μερικά χρόνια ακόμη προκειμένου οι στατιστικοί να επιστρέψουν στο θέμα. Σε ένα πλήρες παραμετρικό στατιστικό πλαίσιο, το μέτρο πληροφορίας των Kullback-Leibler μεταξύ των μέτρων πιθανότητας P_{θ_1} και P_{θ_2} ή μεταξύ των αντίστοιχων Radon-Nikodym παραγώγων f_{θ_1} και f_{θ_2}, ορίζεται ως εξής:

Ορισμός 1.1. *Έστω ο μετρήσιμος χώρος* $(\mathcal{X}, \mathcal{A})$ *και* $\{P_\theta\}$, $\theta \in \Theta$, *μια παραμετρική οικογένεια μέτρων πιθανότητας, με* Θ *ένα ανοικτό υποσύνολο του* $\mathcal{R}^M$, $M \geq 1$. *Για* $\theta_1, \theta_2 \in \Theta$, *συμβολίζουμε με* f_{θ_i}, $i = 1, 2$, *τις Radon-Nikodym παραγώγους των μέτρων* P_{θ_i}, $i = 1, 2$, *ήτοι* $f_{\theta_i} = \frac{dP_{\theta_i}}{d\mu}$, $i = 1, 2$, *όπου* μ *είναι ένα* σ*-πεπερασμένο μέτρο στο χώρο* $(\mathcal{X}, \mathcal{A})$ *και* $P_{\theta_i} \ll \mu$, $i = 1, 2$. *Η ποσότητα*

$$D_0(f_{\theta_1}, f_{\theta_2}) = D_0(\theta_1, \theta_2) = \int_{\mathcal{X}} f_{\theta_1}(x) \log \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} d\mu(x), \tag{1.1}$$

ονομάζεται μέτρο πληροφορίας των Kullback-Leibler ή απόκλιση Kullback-Leibler.

Ο Ορισμός 1.1 καλύπτει και τη διακριτή περίπτωση, όπου το μέτρο μ είναι αριθμήσιμο, οι $f_{\theta_i}(x)$, $i = 1, 2$, διακριτές συναρτήσεις κατανομής και το σύμβολο της ολοκλήρωσης αντικαθίσταται από το σύμβολο της άθροισης. Άμεσα προκύπτει ότι το μέτρο των Kullback-Leibler είναι μη αρνητικό και παίρνει την ελάχιστη τιμή του, ίση με το 0, εάν-ν τα μέτρα πιθανότητας στα οποία στηρίζεται ταυτίζονται. Το μέτρο αυτό όμως δεν ικανοποιεί τη συμμετρική ιδιότητα, ήτοι $D_0(f_{\theta_1}, f_{\theta_2}) \neq D_0(f_{\theta_2}, f_{\theta_1})$. Ο Jeffreys (1946) όρισε την απόκλιση

$$J(f_{\theta_1}, f_{\theta_2}) = J(\theta_1, \theta_2) = D_0(\theta_1, \theta_2) + D_0(\theta_2, \theta_1),$$

η οποία, σε αντίθεση με την απόκλιση Kullback-Leibler, είναι συμμετρική και ονομάζεται J-απόκλιση. Η τριγωνική ιδιότητα δεν ικανοποιείται γενικά από το μέτρο των Kullback-Leibler και δεν μπορεί έτσι να θεωρηθεί ως μέτρο απόστασης μεταξύ κατανομών.

Διάφορες προσπάθειες γενίκευσης της απόκλισης Kullback-Leibler ή του ορισμού μιας ευρείας κλάσης μέτρων απόκλισης έχουν εμφανιστεί στη βιβλιογραφία. Η πιο σημαντική προσπάθεια να οριστεί μια ευρεία κλάση μέτρων απόκλισης μεταξύ δύο μέτρων πιθανότητας ή μεταξύ των αντίστοιχων Radon-Nikodym παραγώγων τους, έγινε από τον Csiszár (1963, 1967) και ανεξάρτητα από τους Ali and Silvey (1966). Η ϕ-απόκλιση Csiszár (ή απλά ϕ-απόκλιση) μεταξύ των μέτρων πιθανότητας P και Q ή μεταξύ των αντίστοιχων Radon-Nikodym παραγώγων p και q, ορίζεται ως εξής:

Ορισμός 1.2. *Έστω ο μετρήσιμος χώρος* $(\mathcal{X}, \mathcal{A})$, *τα μέτρα πιθανότητας* P, Q *και το* σ*-πεπερασμένο μέτρο* μ *στο χώρο αυτό, έτσι ώστε* $P \ll \mu$ *και* $Q \ll \mu$. *Συμβολίζονται με* p *και* q *οι Radon-Nikodym παράγωγοι των* P *και* Q *αντίστοιχα ως προς* μ, *ήτοι* $p = \frac{dP}{d\mu}$ *και*

$q = \frac{dQ}{d\mu}$. *Η ποσότητα*

$$D_\phi(P,Q) = D_\phi(p,q) = \int_{\mathcal{X}} \phi\left(\frac{dP}{dQ}\right) dQ = \int_{\mathcal{X}} q(x)\phi\left(\frac{p(x)}{q(x)}\right) d\mu(x), \tag{1.2}$$

ονομάζεται ϕ-απόκλιση Csiszár (ή απλά ϕ-απόκλιση) μεταξύ των μέτρων P και Q ή μεταξύ των Radon-Nikodym παραγώγων p και q.

Η συνάρτηση $\phi(\cdot)$ που εμφανίζεται στον παραπάνω ορισμό είναι συνεχής, κυρτή, πραγματική συνάρτηση, ορίζεται στο διάστημα $[0,\infty)$ και πληροί τις συνθήκες $\phi(1) = 0$, $0\phi\left(\frac{0}{0}\right) = 0$ και $0\phi\left(\frac{u}{0}\right) = u \lim\limits_{v\to\infty} \frac{\phi(v)}{v}$, με τις δύο τελευταίες να είναι αναγκαίες ώστε να αποφευχθούν εκφυλισμένες περιπτώσεις (βλέπε, Csiszár, 1967 και Pardo, 2006).

Άμεσα προκύπτει από τον παραπάνω ορισμό ότι σε ένα πλήρες παραμετρικό πλαίσιο η Csiszár ϕ-απόκλιση μεταξύ των πυκνοτήτων f_{θ_1} και f_{θ_2}, ορίζεται ως εξής:

Ορισμός 1.3. *Έστω ο μετρήσιμος χώρος $(\mathcal{X},\mathcal{A})$ και $\{P_\theta\}$, $\theta \in \Theta$, μια παραμετρική οικογένεια μέτρων πιθανότητας, με Θ ένα ανοικτό υποσύνολο του $\mathcal{R}^M$, $M \geq 1$. Για $\theta_1, \theta_2 \in \Theta$, συμβολίζουμε με f_{θ_i}, $i = 1,2$, τις Radon-Nikodym παραγώγους των μέτρων P_{θ_i}, $i = 1,2$, ήτοι $f_{\theta_i} = \frac{dP_{\theta_i}}{d\mu}$, $i = 1,2$, όπου μ είναι ένα σ-πεπερασμένο μέτρο στο χώρο $(\mathcal{X},\mathcal{A})$ και $P_{\theta_i} \ll \mu$, $i = 1,2$. Η ποσότητα*

$$D_\phi(f_{\theta_1}, f_{\theta_2}) = D_\phi(\theta_1,\theta_2) = \int_{\mathcal{X}} f_{\theta_2}(x)\phi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) d\mu(x), \tag{1.3}$$

ονομάζεται Csiszár ϕ-απόκλιση μεταξύ των συναρτήσεων πυκνότητας πιθανότητας f_{θ_1} και f_{θ_2}.

Μετά από τις πρωτοπόρες εργασίες του Csiszár (1963, 1967) και ανεξάρτητα των Ali and Silvey (1966) σχετικά με τον ορισμό και τη μελέτη της ϕ-απόκλισης, μια πληθώρα άρθρων και βιβλίων έχουν δημοσιευθεί. Μερικά από αυτά επικεντρώνονται στους χαρακτηρισμούς και τη μελέτη των ιδιοτήτων της ϕ-απόκλισης, ενώ άλλα στη γενίκευση της ϕ-απόκλισης. Ένα μεγάλο μέρος της βιβλιογραφίας ασχολείται με τη διαμόρφωση και λύση μιας μεγάλης ποικιλίας προβλημάτων στις Πιθανότητες, στη Στατιστική και σχεδόν σε όλους τους επιστημονικούς κλάδους. Τα βιβλία και οι μονογραφίες των Kullback (1959), Csiszár and Korner (1981), Liese and Vajda (1987), Vajda (1989, 1995), Pardo (2006) και Basu et al. (2011) και τα άρθρα επισκόπησης των Papaioannou (1985, 2001), Ullah (1996), Soofi (2000), Ebrahimi et al. (2010) και οι αναφορές που υπάρχουν σ' αυτά αποτελούν μια βάση της υπάρχουσας βιβλιογραφίας σχετικά με τα μέτρα ϕ-απόκλισης. Στη συνέχεια παρατίθενται κάποιες από τις ιδιότητες της ϕ-απόκλισης και αναφέρονται κάποιες ειδικές περιπτώσεις της.

Παρατήρηση 1.2. *Αν η συνάρτηση ϕ, πέραν των ιδιοτήτων που παρετέθηκαν μετά τον Ορισμό 1.2, είναι αυστηρά κυρτή στο 1, τότε*

$$D_\phi(p,q) \geq 0, \text{ με την ισότητα να επιτυγχάνεται αν και μόνο αν } p(x) = q(x), \text{ σ.π.}, x \in \mathcal{X} \tag{1.4}$$

(βλέπε, Pardo, 2006, σελ. 9), όπου η συντομογραφία σ.π. συμβολίζει τον όρο «σχεδόν παντού». Η παραπάνω σημαντική ιδιότητα είναι ο λόγος για τον οποίο η $D_\phi(P,Q)$ έχει καθιερωθεί στη βιβλιογραφία ως ένα μέτρο απόκλισης μεταξύ των μέτρων πιθανότητας P και Q ή μεταξύ των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας p και q.

Παρατήρηση 1.3. *Η $D_\phi(P,Q)$ άμεσα προκύπτει ότι δεν είναι συμμετρική. Επιπλέον, αποδεικνύεται ότι (βλέπε, Liese and Vajda, 1987, σελ. 14) μπορεί να εκφραστεί ως ένα συμμετρικό μέτρο με τη χρήση της κυρτής συνάρτησης*

$$\phi^*(u) = \phi(u) + u\phi\left(\frac{1}{u}\right), \; u > 0.$$

Τότε

$$D_{\phi^*}(P,Q) = D_{\phi^*}(Q,P) = D_\phi(P,Q) + D_\phi(Q,P).$$

Παρατήρηση 1.4. *Η $D_\phi(P,Q)$ δεν πληροί γενικά την τριγωνική ανισότητα, ήτοι δεν πληροί τη σχέση:*

$$D_\phi(P,Q) \leq D_\phi(P,R) + D_\phi(Q,R), \text{ για κάθε } P,Q,R \text{ μέτρα πιθανότητας στον } (\mathcal{X},\mathcal{A}).$$

Από την παραπάνω Παρατήρηση 1.4 προκύπτει άμεσα το συμπέρασμα ότι η $D_\phi(P,Q)$ δεν είναι μια απόσταση με τη συνήθη έννοια της μετρικής, δεδομένου ότι δεν πληροί γενικά την τριγωνική ανισότητα. Έτσι, θεωρούμε τα μέτρα απόκλισης ως στατιστικές αποστάσεις και τα χειριζόμαστε με τον ίδιο τρόπο με τον οποίο χειριζόμαστε μια συνάρτηση απώλειας σε προβλήματα της Θεωρίας Αποφάσεων. Πιο απλά, αν $D_\phi(P,Q) = 0$ τότε τα μέτρα πιθανότητας P και Q συμπίπτουν και επιπλέον όσο πιο «κοντά» είναι η τιμή της $D_\phi(P,Q)$ στο 0, τόσο πιο «κοντά» είναι τα μέτρα πιθανότητας P και Q. Για περισσότερες λεπτομέρειες παραπέμπουμε, μεταξύ άλλων, στους Liese and Vajda (1987, 2006), Vajda (1989) και Stummer and Vajda (2010), όπου η Csiszár ϕ-απόκλιση έχει μελετηθεί εκτενώς.

Μέτρα ϕ-απόκλισης	Κυρτή Συνάρτηση ϕ
$Kullback - Leibler$ (1951)	$\phi(x) = x \log x$ ή $\phi(x) = x \log x - x + 1,\ x > 0$
$J -$ Απόκλιση	$\phi(x) = (x-1) \log x,\ x > 0$
$Pearson$ (1900), $Kagan$ (1963)	$\phi(x) = (x-1)^2,\ x > 0$
$Matusita$(1964)	$\phi(x) = \lvert 1 - x^{\alpha} \rvert^{1/\alpha},\ \ 0 < \alpha < 1,\ \ x > 0$
$Affinity$ του $Matusita$ (1967)	$\phi(x) = sgn(\alpha - 1) x^{\alpha},\ \ x > 0,\ \ \alpha > 0$
$\chi -$ απόκλιση τάξης α $Vajda$(1973)	$\phi(x) = \lvert 1 - x \rvert^{\alpha},\ \alpha \geq 1,\ \ x > 0$
$Cressie\ and\ Read$ (1984)	$\phi_{\lambda}(x) = \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)},\ \ \lambda \neq 0, -1$

Πίνακας 1.1: Μέτρα πληροφορίας Csiszár ϕ-απόκλισης.

Όπως έχει ήδη αναφερθεί η ϕ-απόκλιση Csiszár εισήχθη στη βιβλιογραφία με αρχικό στόχο να αποτελέσει μια ευρεία κλάση μέτρων απόκλισης. Εύκολα διαπιστώνεται ότι για την ειδική περίπτωση που η κυρτή συνάρτηση ϕ της σχέσης (1.3) είναι της μορφής

$$\phi(x) = x \log x,\ \ x > 0$$

ή

$$\phi(x) = x \log x - x + 1,\ \ x > 0,$$

προκύπτει η σχέση (1.1), δηλαδή το μέτρο απόκλισης των Kullback-Leibler. Εκτός του μέτρου απόκλισης των Kullback-Leibler, η ϕ-απόκλιση Csiszár περιλαμβάνει ως ειδικές περιπτώσεις, για διάφορες επιλογές της κυρτής συνάρτησης ϕ, σημαντικά μέτρα απόκλισης (βλέπε, Pardo, 2006, σελ. 6), όπως είναι η λ-απόκλιση των Cressie-Read, που εισήχθη ανεξάρτητα από τους Cressie and Read (1984) και Liese and Vajda (1987), η απόκλιση του Kagan (1963), του Vajda (1973) κ.ά. Στον Πίνακα 1.1 παραθέτουμε τη μορφή της συνάρτησης $\phi(x)$ για τα πιο ευρέως χρησιμοποιούμενα μέτρα αποκλίσεων που ανήκουν στην ευρεία κλάση των ϕ-αποκλίσεων του Csiszár. Για περισσότερα παραδείγματα παραπέμπουμε στους Arndt (2001), Pardo (2006) και Vajda (1989, 1995).

Από στατιστική άποψη η πιο σημαντική, ίσως, οικογένεια ϕ-αποκλίσεων μεταξύ των πυκνοτήτων f_{θ_1} και f_{θ_2} είναι η απόκλιση Cressie and Read (1984), η οποία ορίζεται από τη σχέση

$$D_{\phi_\lambda}(\theta_1,\theta_2) = \frac{1}{\lambda(\lambda+1)}\left(\int_{\mathcal{X}} \frac{f_{\theta_1}^{\lambda+1}(x)}{f_{\theta_2}^{\lambda}(x)} d\mu(x) - 1\right), \quad \lambda \neq 0, \ \lambda \neq -1.$$

Για $\lambda = 0$ ή $\lambda \neq -1$ έχουμε αντίστοιχα,

$$\lim_{\lambda \to 0} D_{\phi_\lambda}(\theta_1,\theta_2) = D_0(\theta_1,\theta_2)$$

και

$$\lim_{\lambda \to -1} D_{\phi_\lambda}(\theta_1,\theta_2) = D_0(\theta_2,\theta_1).$$

Έτσι, η απόκλιση Cressie and Read ορίζεται για $\lambda \in R$ από τη συνάρτηση

$$\phi_\lambda(x) = \begin{cases} \frac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x - \lambda(x-1)), \quad \lambda \neq 0, \quad \lambda \neq -1 \\ \lim\limits_{\lambda \to 0}\left\{\frac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x - \lambda(x-1))\right\} = x\log x - x + 1 \\ \lim\limits_{\lambda \to -1}\left\{\frac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x - \lambda(x-1))\right\} = -x\log x + x - 1 \end{cases}$$

(βλέπε, Pardo, 2006, σελ. 7).

Ένα από τα σχετικά πρόσφατα προτεινόμενα μέτρα απόκλισης είναι το μέτρο των Basu - Harris - Hjort και Jones ή απόκλιση BHHJ μεταξύ των πυκνοτήτων g και f και προτάθηκε από τους Basu et al. (1998) κατά την ανάπτυξη μιας μεθόδου εκτίμησης παραμέτρων.

Ορισμός 1.4. *Η ποσότητα*

$$D_\alpha(g,f) = \int_{\mathcal{X}} \left\{ f^{1+\alpha}(x) - (1+\frac{1}{\alpha})g(x)f^{\alpha}(x) + \frac{1}{\alpha}g^{1+\alpha}(x)\right\} d\mu(x), \quad \alpha > 0, \tag{1.5}$$

ονομάζεται μέτρο των Basu - Harris - Hjort and Jones ή απόκλιση BHHJ μεταξύ των πυκνοτήτων g και f.

Η απόκλιση BHHJ εξαρτάται και επομένως ρυθμίζεται από μια θετική παράμετρο a. Το όριο της οικογένειας αποκλίσεων BHHJ για $a \to 0$ είναι η απόκλιση Kullback-Leibler, ενώ για $a = 1$ η απόκλιση (1.5) οδηγεί στο τετράγωνο της τυπικής L_2 απόστασης (ή Ευκλείδειας απόστασης) μεταξύ των κατανομών g και f. Το ενδιαφέρον εστιάζεται στην οικογένεια μέτρων BHHJ, διότι η συναρτησιακή της έκφραση είναι χρήσιμη για τη δημιουργία κριτήριων πληροφορίας και χρησιμοποιήθηκε από τους Mattheou et al. (2009) για τη δημιουργία του κριτηρίου επιλογής DIC. Αυτή η οικογένεια μέτρων προτάθηκε από τους Basu et al. (1998) για την ανάπτυξη μιας μεθόδου εκτίμησης παραμέτρων με βάση την ελάχιστη απόκλιση BHHJ. Ο δείκτης a ελέγχει την ασυμπτωτική ευρωστία και αποτελεσματικότητα των εκτιμητών των παραμέτρων που είναι οι ποσότητες που ελαχιστοποιούν τη σχέση (1.5).

Η γενίκευση της οικογένειας αποκλίσεων BHHJ (1.5) (βλέπε, Mattheou and Karagrigoriou, 2010) δίνεται από μια πραγματική κυρτή συνάρτηση ϕ_α στο $[0, +\infty)$ που πληροί τις συνθήκες $\phi_\alpha(1) = 0$, $\phi'_\alpha(1) = 0$, $\phi''_\alpha(1) \neq 0$ και $0\phi_\alpha(0/0) = 0$, $0\phi_\alpha(u/0) = u \lim\limits_{v\to\infty} \phi_\alpha(v)/v$, για $u > 0$ και δίνεται από την έκφραση

$$D_\alpha(g, f) = E_g\left(g^\alpha(X)\phi_\alpha\left(\frac{f(X)}{g(X)}\right)\right) = \int_{\mathcal{X}} g^{1+\alpha}(x)\phi_\alpha\left(\frac{f(x)}{g(x)}\right) d\mu, \quad \alpha > 0, \tag{1.6}$$

Το μέτρο BHHJ προκύπτει από την παραπάνω γενικευμένη οικογένεια αν η συνάρτηση ϕ_α έχει τη μορφή

$$\phi_\alpha(u) = u^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) u^\alpha + \frac{1}{\alpha}. \tag{1.7}$$

Σ' αυτό το σημείο θα ήταν παράλειψη να μην επισημανθεί ότι αν και αρκετά μέτρα ανήκουν στην οικογένεια των μέτρων Csiszár, εν τούτοις, υπάρχουν σημαντικά μέτρα απόκλισης που δεν ανήκουν σ' αυτήν την οικογένεια. Για παράδειγμα δεν ανήκει το μέτρο που ορίστηκε από τον Rényi (1961), ως μια πρώτη προσπάθεια γενίκευσης του μέτρου των Kullback-Leibler και ορίζεται ως εξής:

$$D_r(\theta_1, \theta_2) = \frac{1}{r-1} \log \int_{\mathcal{X}} f_{\theta_1}^r(x) f_{\theta_2}^{1-r}(x) d\mu(x), \ r > 0, \ r \neq 1. \tag{1.8}$$

Για $r = 1/2$ το μέτρο Rényi είναι το διπλάσιο του γνωστού μέτρου του Bhattacharyya (1943), που δίνεται από τη σχέση

$$D^{Bh}(f_{\theta_1}, f_{\theta_2}) = -\log \int_{\mathcal{X}} \sqrt{f_{\theta_1}(x) f_{\theta_2}(x)} d\mu(x). \tag{1.9}$$

Το μέτρο Rényi τάξης r ορίστηκε αργότερα από τους Liese and Vajda (1987), για όλες τις τιμές του $r \neq 0, 1$. Ειδικότερα, ορίζεται από τη σχέση

$$D_r(\theta_1, \theta_2) = \frac{1}{r(r-1)} \log \int_{\mathcal{X}} f_{\theta_1}^r(x) f_{\theta_2}^{1-r}(x) d\mu(x), \tag{1.10}$$

για $r \neq 0, 1$, το οποίο αναφέρεται ως μέτρο του Rényi. Οι ειδικές περιπτώσεις για $r = 0$ και $r = 1$ ορίζονται από τις σχέσεις

$$\lim_{r\to 1} D_r(\theta_1, \theta_2) = D_0(\theta_1, \theta_2)$$

και

$$\lim_{r\to 0} D_r(\theta_1,\theta_2) = D_0(\theta_2,\theta_1),$$

αντίστοιχα, όπου $D_0(\cdot,\cdot)$ το μέτρο πληροφορίας των Kullback-Leibler.

Παρατήρηση 1.5. *Η παραπάνω διαπίστωση της ύπαρξης μέτρων απόκλισης που δεν ανήκουν στην οικογένεια των μέτρων Csiszár οδήγησε τους Menéndez et al. (1995) στον ορισμό και τη μελέτη της (h,ϕ) οικογένειας αποκλίσεων, η οποία περιλαμβάνει μια επιπρόσθετη διαφορίσιμη και αύξουσα πραγματική συνάρτηση h, τέτοια ώστε $h(0)=0$ και $h'(0)>0$. Ειδικότερα η οικογένεια αυτή των μέτρων (h,ϕ)-απόκλισης ορίζεται (βλέπε, Pardo, 2006, σελ. 8) από τη σχέση:*

$$D^{h,\phi}(f_{\theta_1}, f_{\theta_2}) = D^{h,\phi}(\theta_1,\theta_2) = h(D_\phi(\theta_1,\theta_2)).$$

Η οικογένεια των (h,ϕ) μέτρων απόκλισης έχει ερευνηθεί διεξοδικά αν και η χρήση των δύο συναρτήσεων (ϕ και h) αυξάνει την πολυπλοκότητα και δυσκολεύει την εφαρμογή αυτών.

1.2 Περίληψη-Συμβολή Διατριβής

Στην παράγραφο αυτή παρουσιάζεται συνοπτικά το περιεχόμενο της διατριβής και επισημαίνεται η συμβολή του πρωτότυπου μέρους της. Η ϕ-απόκλιση όπως ορίστηκε στην (1.2) ποσοτικοποιεί τη διαφορά, την απόκλιση μεταξύ των μέτρων πιθανότητας P και Q ή των αντίστοιχων πυκνοτήτων p και q, σε ολόκληρο το πεδίο ορισμού $\mathcal{X}$. Ωστόσο, υπάρχουν στην πράξη περιπτώσεις όπου το ενδιαφέρον μας επικεντρώνεται στις διαφορές μεταξύ δύο πληθυσμών ή μεταξύ των αντίστοιχων κατανομών σε ένα υποσύνολο του κοινού πεδίου ορισμού $\mathcal{X}$. Για παράδειγμα, ένας ερευνητής επιθυμεί να διερευνήσει αν οι πληθυσμοί των ανδρών και των γυναικών παρουσιάζουν την ίδια συμπεριφορά για υψηλά ή χαμηλά επίπεδα χοληστερόλης στο αίμα. Συνεπώς, η εφαρμογή της σχέσης (1.2) παρά το γεγονός ότι δίνει μια εικόνα σχετική με την ομοιογένεια ανδρών και γυναικών για όλα τα επίπεδα χοληστερόλης, η σχέση (1.2) δε δίνει καμία πληροφορία σχετική με την ομοιογένεια ανδρών και γυναικών σε χαμηλά ή υψηλά ή φυσιολογικά (ενδιάμεσα) επίπεδα χοληστερόλης. Είναι φανερό ότι θα πρέπει να οριστεί ένα μέτρο απόκλισης κατά τέτοιο τρόπο ώστε να μας παρέχει μία ένδειξη για την ομοιότητα, την ομοιογένεια των δύο πληθυσμών ή των αντίστοιχων κατανομών σε ένα υποσύνολο του κοινού πεδίου ορισμού $\mathcal{X}$.

Παρακινούμενοι από την παρατήρηση αυτή, το **Κεφάλαιο 2** επικεντρώνεται στον ορισμό μιας ευρείας κλάσης μέτρων τοπικής απόκλισης (local divergence measures) μεταξύ δύο μέτρων πιθανότητας ή μεταξύ των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας. Οι εισαχθείσες τοπικές αποκλίσεις βασίζονται στην κλασική απόκλιση του Csiszár και αποδίδουν ένα μέτρο της ψευτο-απόστασης (ή στατιστικής απόστασης ή απόκλισης) μεταξύ δύο κατανομών σε μία συγκεκριμένη περιοχή του κοινού πεδίου ορισμού τους. Κατ' αυτόν τον τρόπο παρέχουν ένα χρήσιμο εργαλείο για την ποσοτικοποίηση της στατιστικής απόστασης μεταξύ δύο κατανομών, τοπικά, σε μια συγκεκριμένη δηλαδή περιοχή του κοινού πεδίου ορισμού τους, η οποία παρουσιάζει ενδιαφέρον τόσο από θεωρητική πλευρά όσο και στο επίπεδο των εφαρμογών.

Επιπλέον στο Κεφάλαιο 2 μελετάται το σύνολο τιμών των τοπικών αποκλίσεων που παρουσιάζονται και προσδιορίζονται, περαιτέρω, οι αναλυτικές εκφράσεις των προτεινόμενων τοπικών αποκλίσεων, όταν οι κατανομές στις οποίες στηρίζονται είναι μέλη της εκθετικής οικογένειας κατανομών και όταν ταυτίζονται με εκείνες της πολυδιάστατης κανονικής κατανομής. Το Κεφάλαιο 2 ολοκληρώνεται με μελέτες προσομοίωσης που αποτυπώνουν την εύρωστη συμπεριφορά των μέτρων τοπικής απόκλισης, υπό την έννοια ότι οι διαφορές μεταξύ δύο πληθυσμών που είτε δεν μπορούν να αναδειχθούν είτε είναι δυσδιάκριτες με τα καθολικά μέτρα, εντοπίζονται χρησιμοποιώντας το κατάλληλο μέτρο τοπικής απόκλισης. Ένα μέρος των αποτελεσμάτων αυτού του κεφαλαίου περιλαμβάνεται στην εργασία των Avlogiaris et al. (2016 a).

Όπως έχει ήδη αναφερθεί η Στατιστική Θεωρία Πληροφοριών μεταξύ άλλων ποσοτικοποιεί την απόκλιση μεταξύ δύο ή περισσότερων στατιστικών μοντέλων. Ως εκ τούτου, κατά τη διάρκεια των τελευταίων δεκαετιών αποτελεί βασικό εργαλείο για την ανάπτυξη μεθοδολογιών στη Μαθηματική Στατιστική. Οι μονογραφίες των Read and Cressie (1988) και Pardo (2006) και οι αναφορές που εκεί παρατίθενται παρέχουν μια εξαντλητική συζήτηση για τη χρήση των μέτρων απόκλισης στον έλεγχο στατιστικών υποθέσεων, μέχρι τη χρονολογία συγγραφής τους, ενώ νεότερα αποτελέσματα παρατίθενται στην εργασία των Basu et al. (2013). Έλεγχοι ομοιογένειας μεταξύ δύο ή περισσότερων κατανομών έχουν μελετηθεί στην εργασία του Zografos (1998).

Χάριν πληρότητας και για να περιγραφεί η χρήση μέτρων απόκλισης σε ελέγχους στατιστικών υποθέσεων σε παραμετρικό πλαίσιο, έστω δύο ανεξάρτητα τυχαία δείγματα $X_1, ..., X_n$ και $Y_1, ..., Y_m$ από πληθυσμούς που περιγράφονται από τις κατανομές f_{θ_1} και f_{θ_2}, $\theta_1, \theta_2 \in \Theta$. Ένα μέτρο απόκλισης $D(f_{\theta_1}, f_{\theta_2})$ ποσοτικοποιεί την ομοιότητα μεταξύ των κατανομών f_{θ_1} και f_{θ_2} και η απόκλιση $D(f_{\theta_1}, f_{\theta_2})$ πληροί τη βασική ιδιότητα $D(f_{\theta_1}, f_{\theta_2}) \geq 0$, με ισότητα εάν-ν τα μοντέλα f_{θ_1} και f_{θ_2} συμπίπτουν. Συνεπώς, μια εμπειρική εκδοχή ενός μέτρου απόκλισης $D(f_{\theta_1}, f_{\theta_2})$, έστω $D(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$, όπου $\widehat{\theta}_1$ και $\widehat{\theta}_2$ οι ΕΜΠ (Εκτιμητές Μέγιστης Πιθανοφάνειας) των θ_1 και θ_2 με βάση τα δύο τυχαία δείγματα $X_1, ..., X_n$ και $Y_1, ..., Y_m$ αντίστοιχα, μπορεί να χρησιμεύσει ως τη στατιστική συνάρτηση για τον έλεγχο της υπόθεσης ότι τα μοντέλα f_{θ_1} και f_{θ_2} συμπίπτουν (έλεγχος ομοιογένειας, βλέπε, Zografos, 1998). Μικρές τιμές της $D(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ είναι υπέρ της μηδενικής υπόθεσης $H_0 : \theta_1 = \theta_2$ της ομοιογένειας, έναντι της εναλλακτικής $H_a : \theta_1 \neq \theta_2$, ενώ μεγάλες τιμές της $D(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ είναι υπέρ της απόρριψης της H_0. Με παρόμοιο τρόπο μια εμπειρική εκδοχή ενός μέτρου απόκλισης $D(f_\theta, f_{\theta_0})$, θ_0 γνωστό, έστω $D(f_{\widehat{\theta}}, f_{\theta_0})$, χρησιμεύει ως μια στατιστική συνάρτηση για τον έλεγχο αν το πραγματικό αλλά άγνωστο μοντέλο f_θ μπορεί να προσεγγιστεί από το μοντέλο f_{θ_0}, το οποίο καθορίζεται από τη μηδενική υπόθεση $H_0 : \theta = \theta_0$ (έλεγχος καλής προσαρμογής). Ο έλεγχος της μηδενικής υπόθεσης $H_0 : \theta = \theta_0, \theta_0 \in \Theta$, θ_0 γνωστό, έναντι της εναλλακτικής $H_a : \theta \neq \theta_0$ έχει εκτενώς μελετηθεί μέσω μέτρων ϕ-απόκλισης (βλέπε, Pardo, 2006, Κεφάλαιο 9).

Τα διαθέσιμα στη βιβλιογραφία παραμετρικά στατιστικά τεστ που ορίζονται μέσω μέτρων ϕ-αποκλίσεων, διερευνούν τα προβλήματα σε ολόκληρο το πεδίο ορισμού των κατανομών που περιγράφουν ένα χαρακτηριστικό μέγεθος ή ιδιότητα των μελών του πληθυσμού. Παρά το γεγονός ότι ο ισχυρισμός της καλής προσαρμογής ή της ομοιογένειας μπορεί να γίνεται δεκτός ή να απορριφθεί σε ολόκληρο το πεδίο ορισμού των θεωρούμενων κατανομών, ανάλυση πραγματικών δεδομένων οδηγεί στο εντελώς αντίθετο συμπέρασμα όταν το ενδιαφέρον εστιάζεται σε ένα υποσύνολο του πεδίου ορισμού των κατανομών. Η διαπίστωση αυτή έδωσε το έναυσμα για την εισαγωγή, την κατασκευή και τη μελέτη παραμετρικών στατιστικών τεστ καλής προσαρμογής και ομοιογένειας, τοπικά, σε ένα υποσύνολο του πεδίου ορισμού των θεωρούμενων κατανομών. Η κατασκευή των παραπάνω στατιστικών τεστ για τον έλεγχο υποθέσεων, τοπικά, βασίζεται στην έννοια της τοπικής ϕ-απόκλισης (local ϕ-divergence) η οποία ορίστηκε και μελετήθηκε στο Κεφάλαιο 2.

Στο στατιστικό αυτό πλαίσιο, το **Κεφάλαιο 3** επικεντρώνεται στην κατασκευή και μελέτη παραμετρικών στατιστικών τεστ για τον έλεγχο καλής προσαρμογής (one sample problem) και τον έλεγχο ομοιογένειας (two samples problem), τοπικά, σε ένα υποσύνολο του πεδίου ορισμού των υπό θεώρηση κατανομών. Αναλυτικότερα, προσδιορίζονται οι στατιστικές συναρτήσεις, καθώς και οι ασυμπτωτικές τους κατανομές υπό τη μηδενική υπόθεση η οποία μπορεί να είναι απλή η σύνθετη. Επιπλέον, στο Κεφάλαιο 3 με μελέτες προσομοίωσης διερευνάται η αποτελεσματικότητα των εισαγόμενων τοπικών παραμετρικών ελέγχων έχοντας ως κριτήρια το σφάλμα τύπου I και την ισχύ του ελέγχου. Διερευνάται επίσης, η συμπεριφορά των προτεινόμενων στατιστικών τεστ σε πραγματικά δεδομένα. Τα αποτελέσματα αυτού του κεφαλαίου περιλαμβάνονται στην εργασία των Avlogiaris et al. (2016 b)

Ας στρέψουμε τώρα το ενδιαφέρον στην περίπτωση που διαθέτουμε ένα τυχαίο δείγμα από έναν άγνωστο πληθυσμό και δύο ή περισσότερα παραμετρικά μοντέλα, υποψήφια για τη μοντελοποίηση των δεδομένων αυτών. Τα κριτήρια επιλογής μοντέλων παρέχουν μια συστηματική και ενδελεχή μέθοδο και επιτρέπουν στους ερευνητές να επιλέξουν το καταλληλότερο μοντέλο, από μια συλλογή πιθανών κατάλληλων μοντέλων. Η δημιουργία τέτοιων κριτηρίων απαιτεί τη χρήση ενός μέτρου ομοιότητας μεταξύ δύο μοντέλων, τα οποία συνήθως περιγράφονται από τις κατανομές τους. Τα κριτήρια αυτά μπορεί να θεωρούνται ως αμερόληπτες εκτιμήτριες της αναμενόμενης συνολικής διαφοράς, μιας μη αρνητικής ποσότητας, η οποία εκφράζει την απόσταση μεταξύ του αληθινού, αλλά άγνωστου μοντέλου, και του υποψήφιου μοντέλου, που χρησιμοποιείται για να περιγράψει τα δεδομένα. Εάν η τιμή του κριτηρίου είναι μικρή τότε το υποψήφιο ή προσαρμοσμένο μοντέλο προτιμάται. Διάφορα μέτρα απόκλισης έχουν χρησιμοποιηθεί για τη δημιουργία κριτηρίων επιλογής μοντέλου. Το γνωστό Kullback-Leibler (1951) μέτρο απόκλισης χρησιμοποιήθηκε από τον Akaike (1973), προκειμένου να αναπτυχθεί το φημισμένο κριτήριο του Akaike (AIC). Μετά την πρωτοπόρα εργασία του Akaike, υπήρξε μια τεράστια βιβλιογραφία σχετική με τη κατασκευή κριτηρίων επιλογής μοντέλου. Παραπέμπουμε μεταξύ άλλων στους Schwarz (1978), Konishi and Kitagawa (1996), Spiegelhalter et al. (2002), Seghouane and Bekara (2004), Cavanaugh (2004), Bengtsson and Cavanaugh (2006), Shang and Cavanaugh (2008), Shang (2008), Mattheou et al. (2009), Toma and Broniatowski (2011), και Toma (2014). Μια διαφορετική προσέγγιση σε κριτήρια επιλογής μοντέλου παρουσιάζεται στους Claeskens and Hjort (2003). Τέλος, το βιβλίο των Claeskens and Hjort (2008) και οι αναφορές που εκεί παρατίθενται παρέχουν μια εξαντλητική συζήτηση σε κριτήρια επιλογής μοντέλου.

Παρακινούμενοι από τη συζήτηση που προηγήθηκε, φαίνεται ελκυστική η ιδέα της ανάπτυξης μιας μεθόδου που διερευνά ολόκληρο το πεδίο ορισμού $\mathcal{X}$ και επιλέγει το καλύτερο μοντέλο μεταξύ διάφορων διαθέσιμων υποψήφιων μοντέλων, σε ορισμένες περιοχές του πεδίου ορισμού $\mathcal{X}$. Η ιδέα αυτή αναπτύσσεται και μελετάται στο Κεφάλαιο 4. Ειδικότερα, στο **Κεφάλαιο 4** εισάγεται η έννοια της τοπικής επιλογής μοντέλου, και το τοπικό κριτήριο *LDiv.IC* (Local Divergence Information Criterion) χρησιμοποιώντας την τοπική εκδοχή της απόκλισης των Basu et al. (1998), που παρουσιάστηκε στον Ορισμό 1.4. Το παραπάνω μέτρο απόκλισης χρησιμοποιήθηκε γιατί πληροί κάποιες χαρακτηριστικές ιδιότητες, που διεξοδικά θα παρουσιασθούν στο Κεφάλαιο 4 και κρίθηκε κατ' αυτόν τον τρόπο κατάλληλο στη δημιουργία κριτηρίων πληροφορίας για την τοπική επιλογή μοντέλου (local model selection). Επιπρόσθετα στο Κεφάλαιο 4 παρατίθενται τα αποτελέσματα μελετών προσομοίωσης για τη διερεύνηση της συμπεριφοράς του τοπικού κριτηρίου επιλογής μοντέλου καθώς και αποτελέσματα εφαρμογής του σε πραγματικά δεδομένα.

Τέλος, στο **Κεφάλαιο 5** δίνονται προτάσεις για περαιτέρω έρευνα στην περιοχή της Στατιστικής Θεωρίας Πληροφοριών. Η παρούσα διατριβή ολοκληρώνεται με το **Κεφάλαιο 6** στο οποίο δίνεται μια περίληψη των κεφαλαίων που την απαρτίζουν, στην Αγγλική (*Abstract of Dissertation*) και τη Βιβλιογραφία.

Λέξεις Κλειδιά

ϕ-απόκλιση, απόκλιση Csiszár, απόκλιση Kullback-Leibler, απόκλιση Cressie and Read, α-πόκλιση BHHJ, τοπική απόκλιση, εκθετική οικογένεια κατανομών, τοπικός πίνακας πληροφορίας Fisher, τοπικό στατιστικό ελέγχου Wald, τοπικό στατιστικό ελέγχου Rao, τοπικός έλεγχος καλής προσαρμογής, τοπικός έλεγχος ομοιογένειας, τοπικό στατιστικό ελέγχου ϕ-απόκλισης, επιλογή μοντέλου, τοπική επιλογή μοντέλου, AIC, LDiv.IC, LAIC, κριτήριο πληροφορίας τοπικής απόκλισης, τοπικό κριτήριο επιλογής μοντέλου, τοπική αναμενόμενη συνολική διαφορά, τοπική απόκλιση BHHJ, μοντέλα μείξης, θεωρία point process.

Κεφάλαιο 2

Τοπικές ϕ-αποκλίσεις μεταξύ δύο μέτρων πιθανότητας

Όπως αναφέρθηκε στην Ενότητα 1.2, αντικείμενο μελέτης του Κεφαλαίου 2 είναι ο ορισμός μιας ευρείας κλάσης τοπικών αποκλίσεων (local divergence measures) μεταξύ δύο μέτρων πιθανότητας ή μεταξύ των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας. Οι εισαχθείσες τοπικές αποκλίσεις βασίζονται στην κλασική απόκλιση Csiszár (βλέπε, Ορισμό 1.3) και αποδίδουν ένα μέτρο της ψευτο-απόστασης (ή στατιστικής απόστασης ή απόκλισης) μεταξύ δύο κατανομών σε μία συγκεκριμένη περιοχή του κοινού πεδίου ορισμού τους.

Κατ' αυτόν τον τρόπο παρέχουν ένα χρήσιμο εργαλείο για την ποσοτικοποίηση της στατιστικής απόστασης μεταξύ δύο κατανομών, τοπικά, σε μια συγκεκριμένη δηλαδή περιοχή του κοινού πεδίου ορισμού τους, η οποία παρουσιάζει ενδιαφέρον τόσο από θεωρητικής πλευράς όσο και στο επίπεδο των εφαρμογών.

Σ' αυτό το πλαίσιο η διάθρωση του Κεφαλαίου 2 έχει ως εξής. Στην Ενότητα 2.1 ορίζεται η κλάση των τοπικών ϕ-αποκλίσεων και μελετώνται κάποιες ιδιότητές της. Στη συνέχεια, στην Ενότητα 2.2, προσδιορίζονται οι αναλυτικές εκφράσεις των προτεινόμενων τοπικών αποκλίσεων για ειδικές περιπτώσεις κυρτών συναρτήσεων ϕ, όταν οι κατανομές στις οποίες στηρίζονται είναι μέλη της εκθετικής οικογένειας κατανομών. Λόγω της σπουδαιότητας της πολυδιάστατης κανονικής κατανομής, παρατίθενται αναλυτικές εκφράσεις της τοπικής απόκλισης γι' αυτήν την περίπτωση. Στην Ενότητα 2.3 ορίζεται η τοπική εκδοχή της απόκλισης BHHJ, της απόκλισης που εισήχθη από τους Basu et al. (1998) και προσδιορίζεται η αναλυτική της έκφραση όταν οι κατανομές στις οποίες στηρίζεται είναι μέλη της εκθετικής οικογένειας κατανομών. Η τοπική αυτή απόκλιση μπορεί να θεωρηθεί ότι αποτελεί ένα είδος επέκτασης της τοπικής ϕ-απόκλισης. Η τοπική απόκλιση BHHJ διαδραματίζει σημαντικό ρόλο στη δημιουργία κριτήριων πληροφορίας για την τοπική επιλογή μοντέλου, αντικείμενο μελέτης του Κεφαλαίου 4. Τέλος στην Ενότητα 2.4, διερευνάται μέσω δυο παραδείγματων και μιας εφαρμογής σε ένα πραγματικό σύνολο δεδομένων, ο τρόπος με τον οποίο η τοπική ϕ-απόκλιση ποσοτικοποιεί τη στατιστική απόσταση μεταξύ δύο κατανομών σε ένα υποσύνολο του πεδίου ορισμού τους.

2.1 Τοπική ϕ-απόκλιση: Ορισμός και ιδιότητες

Ας υποθέσουμε ότι ένας ερευνητής ενδιαφέρεται για την ομοιογένεια ή την ομοιότητα δύο πληθυσμών σχετικά με ένα κοινό χαρακτηριστικό γνώρισμα των μελών τους. Για παράδειγμα, ας θεωρήσουμε τους πληθυσμούς των ανδρών και γυναικών και έστω ότι το κοινό χαρακτηριστικό των μελών τους είναι το επίπεδο χοληστερόλης στο αίμα. Υιοθετούμε ότι το κοινό χαρακτηριστικό των δύο πληθυσμών περιγράφεται από δύο πυκνότητες, μία για κάθε πληθυσμό. Ένα μέτρο ϕ-απόκλισης που προκύπτει από τη σχέση (1.2) του Ορισμού 1.2, για μια συγκεκριμένη επιλογή της συνάρτησης ϕ, ποσοτικοποιεί τη διαφορά, την απόκλιση μεταξύ των μέτρων πιθανότητας P και Q ή των αντίστοιχων πυκνοτήτων p και q, σε ολόκληρο το πεδίο ορισμού $\mathcal{X}$. Ωστόσο, υπάρχουν στην πράξη περιπτώσεις, όπου το ενδιαφέρον μας επικεντρώνεται στις διαφορές μεταξύ δύο πληθυσμών ή μεταξύ των αντίστοιχων κατανομών σε ένα υποσύνολο του κοινού πεδίου ορισμού $\mathcal{X}$. Κάτι τέτοιο συμβαίνει όταν, στο πλαίσιο του παραδείγματος, ο ερευνητής εστιάζει το ενδιαφέρον του στη διερεύνηση, για παράδειγμα, της ομοιότητας ή της διαφορετικότητας ανδρών και γυναικών σε υψηλά ή χαμηλά επίπεδα χοληστερόλης. Είναι φανερό ότι θα πρέπει, στην περίπτωση αυτή, να οριστεί ένα μέτρο απόκλισης κατά τέτοιο τρόπο ώστε να παρέχει μία ένδειξη για την ομοιότητα, την ομοιογένεια των δύο πληθυσμών ή των αντίστοιχων κατανομών σε ένα υποσύνολο του κοινού πεδίου ορισμού $\mathcal{X}$.

Μία πρώτη προσέγγιση στη λύση του προβλήματος αυτού μπορεί να δοθεί αν η τοπική απόκλιση οριστεί ως εξής:

$$D_{\phi}^{A}(f_{\theta_1}, f_{\theta_2}) = \int\limits_{\mathcal{X}} I_A(x) f_{\theta_2}(x) \phi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) d\mu(x),$$

όπου $I_A(x) = 1$ αν $x \in A$ και $I_A(x) = 0$ αν $x \notin A$, με $A \subseteq \mathcal{X}$.

Ωστόσο, αυτή η επιλογή, με βάση το Λήμμα 1.1 του Csiszár (1967), οδηγεί σε μέτρα απόκλισης που παραβιάζουν τη συνθήκη (1.4) της μη αρνητικότητας και ταυτότητας των δυσδιάκριτων κατανομών (non-negativity and identity of indiscernibles) ενός μέτρου απόκλισης, που δόθηκε στην Παρατήρηση 1.2. Η ιδιότητα αυτή αποτελεί ουσιαστικά απαραίτητη συνθήκη η οποία πρέπει να πληροίτε από ένα μέτρο απόκλισης. Ένας δεύτερος τρόπος ορισμού μέτρου απόκλισης, σε τοπικό επίπεδο, θα μπορούσε να προκύψει με αντικατάσταση των πυκνοτήτων στον ορισμό της ϕ-απόκλισης από τις αντίστοιχες περικομμένες πυκνότητες (truncated densities), στο επιθυμητό υποσύνολο A του $\mathcal{X}$. Ωστόσο, αυτή η δεύτερη προσέγγιση οδηγεί σε ένα μέτρο απόκλισης μεταξύ μοντέλων τα οποία δεν είναι απαραίτητα τα κατάλληλα μοντέλα που πρέπει να χρησιμοποιηθούν για να περιγράψουν τα υπό εξέταση δεδομένα.

Στην υποενότητα που ακολουθεί θα οριστεί μια ευρεία κλάση μέτρων απόκλισης κατά τέτοιο τρόπο ώστε να ξεπερνιούνται τα προβλήματα που προαναφέρθηκαν και επιπλέον να παρέχει μια ένδειξη για την ομοιότητα των δύο πληθυσμών σε ένα υποσύνολο του κοινού πεδίου ορισμού τους. Μελετάται επίσης η ιδιότητα της μη αρνητικότητας και ταυτότητας των κατανομών της τοπικής ϕ-απόκλισης και διερευνάται το σύνολο τιμών της.

2.1.1 Τοπική ϕ-απόκλιση: Ορισμός

Στην υποενότητα αυτή ορίζουμε ένα μέτρο τοπικής απόκλισης μεταξύ δύο μέτρων πιθανότητας ή των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας. Το μέτρο αυτό, όπως αναφέρθηκε και στην Ενότητα 1.2, έχει τις ρίζες του στην Csiszár ϕ-απόκλιση η οποία ορίστηκε στη σχέση (1.2) και έχει πολυάριθμες εφαρμογές στη Στατιστική αλλά και σε πολλές άλλες επιστημονικές περιοχές.

Θεωρούμε το μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$ και την κλάση Φ^* των πραγματικών κυρτών συναρτήσεων ϕ που ορίζονται στο διάστημα $[0, \infty)$, έτσι ώστε $\phi(1) = 0$, $0\phi\left(\frac{0}{0}\right) = 0$ και $0\phi\left(\frac{u}{0}\right) = u \lim_{v\to\infty} \frac{\phi(v)}{v}$, με τις δύο τελευταίες συνθήκες να είναι αναγκαίες προκειμένου να αποφευχθούν απροσδιοριστίες (βλέπε, Csiszár, 1967, σελ. 299 και Pardo, 2006, σελ. 5). Επιπλέον, υποθέτουμε ότι η συνάρτηση ϕ είναι αυστηρά κυρτή στο 1. Επομένως, η κλάση Φ^* των κυρτών συναρτήσεων ϕ συνοπτικά γράφεται

$$\Phi^* = \left\{ \phi : \phi \text{ είναι αυστηρά κυρτή στο } 1,\ \phi(1) = 0, 0\phi\left(\frac{0}{0}\right) = 0, 0\phi\left(\frac{u}{0}\right) = u \lim_{v\to\infty} \frac{\phi(v)}{v} \right\}. \tag{2.1}$$

Παρατήρηση 2.1. *Στο σημείο αυτό, θα πρέπει να σημειωθεί ότι όλες οι κυρτές συναρτήσεις ϕ που οδηγούν σε σημαντικές περιπτώσεις Csiszár ϕ-αποκλίσεων, όπως των Kullback-Leibler (1951) ($\phi(u) = u \log u$ ή $\phi(u) = u \log u - u + 1,\ u > 0$), του Kagan (1963) ($\phi(u) = (u-1)^2,\ u > 0$), του Vajda (1973) ($\phi(u) = |1-u|^\alpha,\ u > 0,\ \alpha \geq 1$), η λ-απόκλιση των Cressie and Read (1984) ($\phi(u) = \dfrac{u^{\lambda+1} - u - \lambda(u-1)}{\lambda(\lambda+1)},\ \lambda \neq 0, -1$), κ.ά, ανήκουν στην κλάση Φ^* των κυρτών συναρτήσεων.*

Παρακινούμενοι από τη ϕ-απόκλιση του Csiszár η οποία ορίστηκε στο Κεφάλαιο 1 στη σχέση (1.2), ένα μέτρο τοπικής απόκλισης μεταξύ δύο μέτρων πιθανότητας P και Q ή μεταξύ των αντίστοιχων Radon-Nikodym παραγώγων p και q, μπορεί να οριστεί μέσω της σχέσης (1.2), αν μια πρόσθετη συνάρτηση, έστω $A(\cdot,\omega)$, εισαχθεί στην έκφραση της ϕ-απόκλισης, προκειμένου να μετατοπιστεί η μάζα του ολοκληρώματος (1.2) στο επιθυμητό υποσύνολο του κοινού πεδίου ορισμού $\mathcal{X}$. Η συνάρτηση $A(\cdot,\omega)$ έχει το ρόλο ενός πυρήνα και σε πλήρη αναλογία με το καθολικό μέτρο ϕ-απόκλισης της σχέσης (1.2), η συναρτησιακή έκφραση του μέτρου τοπικής απόκλισης μπορεί να αποδοθεί ως εξής:

$$\begin{aligned} D_\phi^A(P,Q) = D_\phi^A(p,q) &= \int_{\mathcal{X}} A(x,\omega)\phi\left(\tfrac{dP}{dQ}\right) dQ \\ &= \int_{\mathcal{X}} A(x,\omega) q(x) \phi\left(\tfrac{p(x)}{q(x)}\right) d\mu(x). \end{aligned} \tag{2.2}$$

Σημειώνουμε ότι, αν $A(x,\omega) = 1$, τότε $D_\phi^1(P,Q) = D_\phi(P,Q)$. Η προσθήκη του πυρήνα $A(x,\omega)$ στη συναρτησιακή έκφραση της ϕ-απόκλισης (1.2) σταθμίζει τη στατιστική απόσταση (ή απόκλιση) μεταξύ των μέτρων πιθανότητας P και Q ή μεταξύ των αντίστοιχων Radon-Nikodym παραγώγων p και q, δίνοντας τη δυνατότητα να επικεντρωθούμε σε συγκεκριμένες περιοχές του πεδίου ορισμού $\mathcal{X}$, που μπορεί να παρουσιάζουν ιδιαίτερο ενδιαφέρον. Στην πράξη, ο πυρήνας $A(\cdot,\cdot)$ μπορεί να θεωρηθεί ως ένα «παράθυρο» πάνω στο οποίο υπολογίζεται το ολοκλήρωμα (2.2) του οποίου μπορεί να ρυθμιστεί η διάσταση του (μέσω της παραμέτρου ω) για να αναδείξει τα ιδιαίτερα χαρακτηριστικά των μέτρων πιθανότητας P και Q, σε ορισμένες περιοχές του $\mathcal{X}$.

Σε ό,τι ακολουθεί και προκειμένου να αποφευχθούν προβλήματα που σχετίζονται με την ύπαρξη του ολοκληρώματος που εμφανίζεται στη σχέση (2.2), θα περιοριστούμε στις συναρτήσεις $A(x,\omega)$ που είναι σχετικές με ένα μέτρο πιθανότητας R στον ίδιο μετρήσιμο χώρο $(\mathcal{X},\mathcal{A})$. Ειδικότερα, η συνάρτηση $A(x,\omega)$ θα θεωρείται ότι είναι η Radon-Nikodym παράγωγος του R ως προς μ, με μ ένα σ-πεπερασμένο μέτρο στον μετρήσιμο χώρο $(\mathcal{X},\mathcal{A})$. Στο πλαίσιο αυτό, ο ορισμός της τοπικής ϕ-απόκλισης μεταξύ των μέτρων P και Q ή μεταξύ των Radon-Nikodym παραγώγων p και q έχει ως εξής:

Ορισμός 2.1. *Θεωρούμε το μετρήσιμο χώρο $(\mathcal{X},\mathcal{A})$, τα μέτρα πιθανότητας P, Q, R και το σ-πεπερασμένο μέτρο μ στο χώρο αυτό, έτσι ώστε $P \ll \mu$, $Q \ll \mu$ και $R \ll \mu$. Συμβολίζουμε με p, q και r τις Radon-Nikodym παραγώγους των P, Q και R αντίστοιχα ως προς μ, ήτοι $p = \frac{dP}{d\mu}$, $q = \frac{dQ}{d\mu}$ και $r = \frac{dR}{d\mu}$. Επιπρόσθετα, έστω μια συνάρτηση ϕ που ανήκει στην κλάση των κυρτών συναρτήσεων Φ^* που ορίστηκε παραπάνω. Η ποσότητα*

$$\begin{aligned} D_\phi^R(P,Q) = D_\phi^r(p,q) &= \int_{\mathcal{X}} \frac{dR}{d\mu}\phi\left(\frac{dP}{dQ}\right)dQ \\ &= \int_{\mathcal{X}} r(x)q(x)\phi\left(\frac{p(x)}{q(x)}\right)d\mu(x), \end{aligned} \tag{2.3}$$

ονομάζεται τοπική ϕ-απόκλιση Csiszár (ή απλά τοπική ϕ-απόκλιση) μεταξύ των μέτρων P και Q που οδηγείται από το μέτρο πυρήνα R ή μεταξύ των Radon-Nikodym παραγώγων p και q που οδηγείται από την πυκνότητα πυρήνα r.

Ο παραπάνω ορισμός μπορεί να τροποποιηθεί ανάλογα, ώστε να ισχύει και για μια παραμετρική οικογένεια μέτρων πιθανότητας. Στο παραμετρικό αυτό πλαίσιο, η τοπική ϕ-απόκλιση ορίζεται ως εξής:

Ορισμός 2.2. *Έστω ο μετρήσιμος χώρος $(\mathcal{X},\mathcal{A})$ και $\{P_\theta\}$, $\theta \in \Theta$, μια παραμετρική οικογένεια μέτρων πιθανότητας, με Θ ένα ανοικτό υποσύνολο του $\mathcal{R}^M$, $M \geq 1$. Για $\theta_1, \theta_2 \in \Theta$, συμβολίζουμε με f_{θ_i}, $i = 1,2$, τις Radon-Nikodym παραγώγους των μέτρων P_{θ_i}, $i = 1,2$, ήτοι $f_{\theta_i} = \frac{dP_{\theta_i}}{d\mu}$, $i = 1,2$, όπου μ ένα σ-πεπερασμένο μέτρο στο χώρο $(\mathcal{X},\mathcal{A})$ και $P_{\theta_i} \ll \mu$, $i = 1,2$. Έστω επίσης, ένα μέτρο πιθανότητας $H_\omega \ll \mu$, $\omega \in \Theta^* \subseteq R^K$, $K \geq 1$, στον ίδιο χώρο $(\mathcal{X},\mathcal{A})$, με Radon-Nikodym παράγωγο $h_\omega = \frac{dH_\omega}{d\mu}$ και μια συνάρτηση ϕ που ανήκει στην κλάση των κυρτών συναρτήσεων Φ^*. Η ποσότητα*

$$\begin{aligned} D_\phi^\omega(f_{\theta_1}, f_{\theta_2}) = D_\phi^\omega(\theta_1,\theta_2) &= \int_{\mathcal{X}} \frac{dH_\omega}{d\mu}\phi\left(\frac{dP_{\theta_1}}{dP_{\theta_2}}\right)dP_{\theta_2} \\ &= \int_{\mathcal{X}} h_\omega(x) f_{\theta_2}(x)\phi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right)d\mu(x), \end{aligned} \tag{2.4}$$

ονομάζεται τοπική ϕ-απόκλιση Csiszár μεταξύ δύο μελών P_{θ_1} και P_{θ_2} της οικογένειας $\{P_\theta : \theta \in \Theta \subseteq R^M\}$ που οδηγείται από το μέτρο πυρήνα H_ω ή μεταξύ των συναρτήσεων πυκνότητας πιθανότητας f_{θ_1} και f_{θ_2} που οδηγείται από την πυκνότητα πυρήνα h_ω.

Η τοπική ϕ-απόκλιση, όπως ορίστηκε παραπάνω, είναι ένα μέτρο απόκλισης μεταξύ δύο μελών της παραμετρικής οικογένειας $\{P_\theta : \theta \in \Theta \subseteq R^M\}$ και διέπεται από ένα άλλο μέτρο H_ω, με $\omega \in \Theta^* \subseteq R^K$, $K \geq 1$, που καθορίζει την περιοχή πάνω στην οποία υπολογίζεται η απόκλιση. Ο πυρήνας πυκνότητα h_ω εξαρτάται από την παράμετρο ω η οποία «καθοδηγεί» και ορίζει το «παράθυρο» πάνω στο οποίο υπολογίζεται το ολοκλήρωμα. Ο υπολογισμός του μέτρου (2.4) σε μια κλειστή μορφή καθίσταται ευκολότερος όταν το μέτρο H_ω ή η αντίστοιχη πυκνότητα h_ω, ανήκει στην παραμετρική οικογένεια $\{P_\theta\}$, $\theta \in \Theta$, ωστόσο αυτό δεν είναι απαραίτητο να ισχύει στην πράξη. Ο πυρήνας πυκνότητα h_ω μπορεί να επιλεγεί με τέτοιο τρόπο ώστε να αναδείξει συγκεκριμένα χαρακτηριστικά των μέτρων πιθανότητας στη περιοχή επί της οποίας υπολογίζεται το ολοκλήρωμα. Επιπρόσθετα, η τοπική ϕ-απόκλιση, όπως ορίστηκε στις (2.3) ή (2.4) παραπάνω, είναι αρκετά όμοια με αυτήν που ορίστηκε στις (1.2) ή (1.3), αντίστοιχα. Η μόνη διαφορά είναι η πυκνότητα r ή h_ω που εισέρχεται στην έκφραση της κλασικής ϕ-απόκλισης και πιο συγκεκριμένα η πρόσθετη παράμετρος $\omega \in \Theta^*$. Ο ρόλος της παραμέτρου ω είναι αποφασιστικής σημασίας στον παραπάνω ορισμό και ακριβώς αυτός ο ρόλος θα μελετηθεί και διερευνηθεί με τη βοήθεια κάποιων παραδειγμάτων στην Ενότητα 2.4.

Παρατήρηση 2.2. (α) *Αν το πεδίο ορισμού $\mathcal{X}$ είναι πεπερασμένο (ή μετρήσιμο) σύνολο, έστω $\mathcal{X}=\{1,2,...,n\}$ και τα μέτρα πιθανότητας P, Q και R αντιπροσωπεύονται από τις διακριτές κατανομές πιθανοτήτων $p=(p_1,...,p_n)$, $q=(q_1,...,q_n)$ και $r=(r_1,...,r_n)$ αντίστοιχα, τότε η τοπική ϕ-απόκλιση μεταξύ των p και q που οδηγείται από την κατανομή r, ορίζεται, με βάση τη σχέση (2.3), από το παρακάτω άθροισμα*

$$D_\phi^r(p,q) = \sum_{i=1}^{n} r_i q_i \phi\left(\frac{p_i}{q_i}\right).$$

Το τελευταίο αυτό μέτρο είναι γνωστό στη βιβλιογραφία ως η σταθμισμένη ϕ-απόκλιση (weighted ϕ-divergence) και έχει μελετηθεί στα άρθρα των Landaburu and Pardo (2000, 2003), Landaburu et al. (2005) καθώς και στις αναφορές αυτών.
(β) *Ειδική περίπτωση της τοπικής ϕ-απόκλισης μελετήθηκε από τους McElroy and Holan, (2009).*

Παρατήρηση 2.3. *Μια επέκταση της τοπικής ϕ-απόκλισης, η οποία ορίστηκε στη σχέση (2.3) ή τη σχέση (2.4), μπορεί να ληφθεί με παρόμοιο τρόπο με εκείνον της Παρατήρησης 1.4. Πιο συγκεκριμένα, αν h είναι μια πραγματική, διαφορίσιμη, αύξουσα συνάρτηση, τότε η τοπική (h,ϕ)-απόκλιση ορίζεται ως εξής:*

$$D_{h,\phi}^\omega(P,Q) = h\left(D_\phi^\omega(P,Q)\right). \tag{2.5}$$

Το τελευταίο αυτό μέτρο μας επιτρέπει να ορίσουμε γενικότερα μέτρα τοπικής απόκλισης, για διάφορες επιλογές των συναρτήσεων h και ϕ. Ωστόσο, ο κύριος λόγος για τον παραπάνω μετασχηματισμό της τοπικής ϕ-απόκλισης $D_\phi^\omega(P,Q)$ είναι ότι μας επιτρέπει να ορίσουμε το μέτρο τοπικής απόκλισης του Rényi με βάση την τοπική (h,ϕ)-απόκλιση. Η τοπική απόκλιση Rényi ορίζεται από τη σχέση (2.5) για

$$h(x) = \frac{1}{r(r-1)} \log(r(r-1)x+1)$$

και

$$\phi(x) = \frac{x^r - r(x-1) - 1}{r(r-1)},$$

με $r \neq 0, 1$.

Παρατήρηση 2.4. *Γενικά, η καθολική ϕ-απόκλιση $D_\phi(P,Q)$ δεν μπορεί να προκύψει από την τοπική ϕ-απόκλιση $D_\phi^R(P,Q)$, εκτός κι αν το μέτρο R είναι, για παράδειγμα, ένα ομοιόμορφο μέτρο επί του $\mathcal{X}$. Στην περίπτωση αυτή η $D_\phi^R(P,Q)$ είναι πολλαπλάσιο της $D_\phi(P,Q)$. Επιπλέον, $D_\phi^R(P,Q) = D_\phi(P,Q)$ όταν $E_q\left[(1-r(X))\phi\left(\frac{p(X)}{q(X)}\right)\right] = 0$.*

2.1.2 Σύνολο τιμών της τοπικής ϕ-απόκλισης

Υπάρχει ένας ευρύς κατάλογος ιδιοτήτων που ικανοποιούνται από την κλασική, καθολική ϕ-απόκλιση Csiszár. Μερικές από τις ιδιότητες αυτές είναι μαθηματικής ή στατιστικής φύσεως, ενώ άλλες υποκινούνται από τις ερευνητικές περιοχές στις οποίες εφαρμόζεται η κλασική ϕ-απόκλιση. Μελέτες σχετικές με τις ιδιότητες που ικανοποιούνται από την καθολική ϕ-απόκλιση Csiszár και τα παράγωγά της μπορούν να βρεθούν, μεταξύ άλλων, στα άρθρα των Papaioannou (1985, 2001), Liese and Vajda (2006) και στα συγγράμματα των Liese and Vajda (1987) και Vajda (1989).

Στο εδάφιο αυτό θα προσδιορισθεί το σύνολο τιμών της τοπικής ϕ-απόκλισης. Θα αποδειχθεί ότι η τοπική ϕ-απόκλιση είναι μη αρνητική και επιτυγχάνει το κάτω φράγμα εάν και μόνο εάν οι κατανομές στις οποίες στηρίζεται ταυτίζονται. Αποτελεί η ιδιότητα αυτή, σημαντικό χαρακτηρισμό της τοπικής ϕ-απόκλισης και επιτρέπει έτσι τη χρησιμοποίησή της ως ένα μέτρο απόκλισης η ψευτο-απόστασης μεταξύ κατανομών.

Χάριν μαθηματικής γενικότητας, θα προσδιορισθεί το σύνολο τιμών μιας γενικότερης μορφής της τοπικής ϕ-απόκλισης. Από αυτό θα προκύψει το αντίστοιχο σύνολο τιμών της τοπικής ϕ-απόκλισης, όπως αυτή ορίστηκε στη σχέση (2.3) ή στη σχέση (2.4), αν περιοριστούμε στις κυρτές συναρτήσεις $\phi \in \Phi^*$, που επιπλέον ικανοποιούν την ιδιότητα $\phi'(1) = 0$, με Φ^* να ορίζεται από τη σχέση (2.1). Η απαίτηση $\phi'(1) = 0$ δεν περιορίζει καθόλου τη γενικότητα. Αντίθετα, αποτελεί την αναγκαία συνθήκη για την εφαρμογή της καθολικής τοπικής ϕ-απόκλισης του Csiszár σε ελέγχους υποθέσεων (βλέπε, Pardo, 2006, σελ. 6). Η ίδια συνθήκη είναι αναγκαία για τη μελέτη των Κεφαλαίων 3 και 4.

Η θεωρητική διερεύνηση των θεμάτων του εδάφιου αυτού στηρίζεται στο ακόλουθο λήμμα.

Λήμμα 2.1 (Stummer and Vajda, 2010, σελ. 171). *Αν ϕ μια οποιαδήποτε κυρτή συνάρτηση τέτοια ώστε $\phi \in \Phi^*$, τότε η συνάρτηση*

$$\overline{\phi}(u) = \phi(u) - \phi'_+(1)(u-1), \tag{2.6}$$

ανήκει και αυτή στην κλάση Φ^*, $\overline{\phi} \in \Phi^*$ και επιπλέον πληροί τη σχέση

$$\overline{\phi}(1) = \overline{\phi}'(1) = 0, \tag{2.7}$$

όπου με $\phi'_+(1)$ συμβολίζεται η δεξιά παράγωγος της ϕ στο σημείο 1. Επιπρόσθετα $\overline{\phi}(u) > 0$, για $u \geq 0$, $u \neq 1$ και $\overline{\phi}(u) = 0$ για $u = 1$.

Χρησιμοποιώντας το παραπάνω Λήμμα 2.1 καταλήγουμε στην ακόλουθη πρόταση.

Πρόταση 2.1. *Έστω ο μετρήσιμος χώρος $(\mathcal{X}, \mathcal{A})$, τα μέτρα πιθανότητας P, Q, R και το σ-πεπερασμένο μέτρο μ στο χώρο αυτό, έτσι ώστε $P \ll \mu$, $Q \ll \mu$ και $R \ll \mu$. Επιπρόσθετα, έστω μια συνάρτηση ϕ που ανήκει στην κλάση των κυρτών συναρτήσεων Φ^* και $\overline{\phi}$ η συνάρτηση της σχέσης (2.6). Τότε,*

$$0 \leq D^R_{\overline{\phi}}(P,Q) = D^R_{\phi}(P,Q) - \phi'_+(1) \int_{\mathcal{X}} r(x)\,(p(x) - q(x))\,d\mu(x), \tag{2.8}$$

όπου με $\phi'_+(1)$ συμβολίζεται η δεξιά παράγωγος της ϕ στο σημείο 1.

Απόδειξη. Λαμβάνοντας υπόψη ότι $\overline{\phi}(u) \geq 0$, για $u \geq 0$ και με βάση τις σχέσεις (2.3) και (2.6) έχουμε ότι

$$\begin{aligned} 0 \leq D^R_{\overline{\phi}}(P,Q) &= \int_{\mathcal{X}} r(x)q(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right) d\mu(x) \\ &= \int_{\mathcal{X}} r(x)q(x)\left(\phi\left(\frac{p(x)}{q(x)}\right) - \phi'_+(1)\left(\frac{p(x)}{q(x)} - 1\right)\right) d\mu(x) \\ &= D^R_{\phi}(P,Q) - \phi'_+(1) \int_{\mathcal{X}} r(x)\,(p(x) - q(x))\,d\mu(x). \end{aligned} \tag{2.9}$$

∎

Είναι πλέον σαφές, με βάση τη σχέση (2.8) της Πρότασης 2.1, ότι η τοπική απόκλιση η οποία ορίζεται με τη βοήθεια της κυρτής συνάρτησης $\overline{\phi}$ της σχέσης (2.6), είναι πάντα μη αρνητική και ως εκ τούτου μπορεί να θεωρηθεί ως μέτρο τοπικής απόκλισης μεταξύ δύο κατανομών πιθανότητας. Γι' αυτό το λόγο στη συνέχεια, παρακινούμενοι από τους Stummer and Vajda (2010, σελ. 171), εστιάζουμε το ενδιαφέρον μας στις τοπικές $\overline{\phi}$-αποκλίσεις και επαναδιατυπώνουμε τον ορισμό της τοπικής ϕ-απόκλισης ως εξής:

Ορισμός 2.3. *Έστω ο μετρήσιμος χώρος $(\mathcal{X},\mathcal{A})$, τα μέτρα πιθανότητας P, Q, R και το σ-πεπερασμένο μέτρο μ στο χώρο αυτό, έτσι ώστε $P \ll \mu$, $Q \ll \mu$ και $R \ll \mu$. Συμβολίζουμε με p, q και r τις Radon-Nikodym παραγώγους των P, Q και R αντίστοιχα ως προς μ, ήτοι $p = \frac{dP}{d\mu}$, $q = \frac{dQ}{d\mu}$ και $r = \frac{dR}{d\mu}$. Επιπρόσθετα, έστω η συνάρτηση $\phi \in \Phi^*$, με Φ^* όπως ορίστηκε στη σχέση (2.1). Τότε η τοπική ϕ-απόκλιση μεταξύ των p και q που οδηγείται από το μέτρο πυρήνα R, ορίζεται ως εξής:*

$$\widetilde{D}_{\phi}^{R}(P,Q) = D_{\overline{\phi}}^{R}(P,Q) = D_{\phi}^{R}(P,Q) - \phi'_{+}(1)\int_{\mathcal{X}} r(x)\,(p(x)-q(x))\,d\mu(x), \tag{2.10}$$

όπου η τοπική απόκλιση $D_{\overline{\phi}}^{R}(P,Q)$ ορίστηκε στη σχέση (2.3).

Σύμφωνα με τη σχέση (2.8), είναι σαφές ότι οι δύο τοπικές αποκλίσεις $\widetilde{D}_{\phi}^{R}(P,Q)$ και $D_{\phi}^{R}(P,Q)$ συμπίπτουν αν συμπεριλάβουμε τη συνθήκη $\phi'(1) = 0$ στην κλάση Φ^*. Έτσι, αν θεωρήσουμε τις τοπικές αποκλίσεις που ορίστηκαν στις σχέσεις (2.3) και (2.4), στο σύνολο των κυρτών συναρτήσεων

$$\Phi = \Phi^* \cap \{\phi : \phi'(1) = 0\}, \tag{2.11}$$

τότε αυτές είναι πάντα μη αρνητικές (βλέπε επίσης, Pardo 2006, σελ. 6).

Παρατήρηση 2.5. *Στο σημείο αυτό, θα πρέπει να τονιστεί, ότι όλες οι κυρτές συναρτήσεις ϕ που οδηγούν σε σημαντικές περιπτώσεις ϕ-αποκλίσεων Csiszár, όπως η Kullback-Leibler (1951) ($\phi(u) = u\log u - u + 1, u > 0$), η Kagan (1963) ($\phi(u) = (u-1)^2, u > 0$), η λ-απόκλιση των Cressie and Read (1984) ($\phi(u) = \dfrac{u^{\lambda+1} - u - \lambda(u-1)}{\lambda(\lambda+1)}$, $\lambda \neq 0, -1$), κ.ά., ανήκουν στο σύνολο Φ, που ορίστηκε στη (2.11).*

Παρατήρηση 2.6. *Χάριν πληρότητας αναφέρουμε στον παρακάτω ορισμό κάποιες έννοιες της Θεωρίας Μέτρου που θα χρησιμοποιηθούν στην απόδειξη του θεωρήματος που ακολουθεί.*

Ορισμός 2.4 (Κουμουλλής και Νεγρεπόντης, 2005, σελ. 153). *Έστω ο μετρήσιμος χώρος $(\mathcal{X}, \mathcal{A})$ και μ ένα προσημασμένο ή θετικό μέτρο στον $(\mathcal{X}, \mathcal{A})$. Λέγεται ότι το μ είναι συγκεντρωμένο σ' ένα υποσύνολο $A \in \mathcal{A}$, αν $\mu(B) = 0$ για κάθε $B \in \mathcal{A}$, με $B \subset \mathcal{X} - A$. Φυσικά στην περίπτωση που το μ είναι θετικό μέτρο αυτό είναι ισοδύναμο με $\mu(\mathcal{X} - A) = 0$.*

Έστω τώρα ν ένα επίσης προσημασμένο ή θετικό μέτρο στον $(\mathcal{X}, \mathcal{A})$. Λέγεται ότι τα μ και ν είναι κάθετα (μεταξύ τους), αν υπάρχει $A \in \mathcal{A}$ ώστε το μ να είναι συγκεντρωμένο στο A και το ν να είναι συγκεντρωμένο στο $\mathcal{X} - A$ και γράφουμε $\mu \perp \nu$.

Στο θεώρημα που ακολουθεί μελετάται το σύνολο τιμών της τοπικής ϕ-απόκλισης, όπως ορίστηκε στη (2.10).

Θεώρημα 2.1. *(α) Για $\phi \in \Phi^*$, η τοπική ϕ-απόκλιση όπως ορίζεται στη σχέση (2.10), πληροί την ανισότητα*

$$0 \leq \widetilde{D}_{\phi}^{R}(P, Q) \leq \ \phi(0)\xi_0 + \phi^*(0)\xi_1 + \phi'_{+}(1)\,(\xi_0 - \xi_1)\,,$$

όπου $\xi_0 = \int\limits_{\mathcal{X}} r(x)q(x)d\mu(x)$, $\xi_1 = \int\limits_{\mathcal{X}} r(x)p(x)d\mu(x)$ και $\phi^* \in \Phi^*$, με ϕ^* η συνάρτηση που ορίζεται ως εξής: $\phi^*(u) = u\phi\left(\frac{1}{u}\right)$, $u > 0$.
(β) $\widetilde{D}_{\phi}^{R}(P, Q) = 0$ αν και μόνο αν $P = Q$.
(γ) $\widetilde{D}_{\phi}^{R}(P, Q) = \phi(0)\xi_0 + \phi^*(0)\xi_1 + \phi'_{+}(1)\,(\xi_0 - \xi_1)$ αν $P \perp Q$. Επιπλέον, αν $\phi(0) + \phi^*(0) < \infty$ και $\widetilde{D}_{\phi}^{R}(P, Q) = \phi(0)\xi_0 + \phi^*(0)\xi_1 + \phi'_{+}(1)\,(\xi_0 - \xi_1)$, τότε $P \perp Q$.

Απόδειξη. (α) Από τις σχέσεις (2.8) και (2.10) είναι σαφές ότι $0 \leq \widetilde{D}_{\phi}^{R}(P, Q)$. Συνεχίζουμε για το άνω φράγμα της $\widetilde{D}_{\phi}^{R}(P, Q)$. Είναι $\overline{\phi}(1) = 0$ και παρακινούμενοι από μια παρόμοια απόδειξη των Stummer and Vajda (2010, σελ. 174), έχουμε

$$\widetilde{D}_{\phi}^{R}(P, Q) = \int\limits_{\{p<q\}} r(x)q(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right) d\mu(x) + \int\limits_{\{q<p\}} r(x)q(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right) d\mu(x). \tag{2.12}$$

Ορίζουμε τη συνάρτηση

$$\overline{\phi^*}(u) = \phi^*(u) + \phi'_{+}(1)(u - 1). \tag{2.13}$$

Τότε, για $u = \frac{q(x)}{p(x)} = \frac{q}{p}$, $x \in \mathcal{X}$, προκύπτει ότι

$$rp\ \overline{\phi^*}\left(\frac{q}{p}\right) = rp\phi^*\left(\frac{q}{p}\right) + rp\phi'_{+}(1)\left(\frac{q}{p} - 1\right). \tag{2.14}$$

Συνεπώς,

$$\begin{aligned} \int\limits_{\{q<p\}} r(x)p(x)\overline{\phi^*}\left(\tfrac{q(x)}{p(x)}\right)d\mu(x) &= \int\limits_{\{q<p\}} r(x)p(x)\phi^*\left(\tfrac{q(x)}{p(x)}\right)d\mu(x) \\ &+ \phi'_+(1)\int\limits_{\{q<p\}} r(x)\left(q(x)-p(x)\right)d\mu(x) \end{aligned} \tag{2.15}$$

και λαμβάνοντας υπόψη ότι $\phi^*(u) = u\phi\left(1/u\right),\ u > 0$, συνάγεται ότι

$$\begin{aligned} \int\limits_{\{q<p\}} r(x)p(x)\overline{\phi^*}\left(\tfrac{q(x)}{p(x)}\right)d\mu(x) &= \int\limits_{\{q<p\}} r(x)q(x)\phi\left(\tfrac{p(x)}{q(x)}\right)d\mu(x) \\ &- \phi'_+(1)\int\limits_{\{q<p\}} r(x)\left(p(x)-q(x)\right)d\mu(x) \\ &= \int\limits_{\{q<p\}} r(x)q(x)\overline{\phi}\left(\tfrac{p(x)}{q(x)}\right)d\mu(x). \end{aligned} \tag{2.16}$$

Έτσι, από τις εξισώσεις (2.12) και (2.16) καταλήγουμε στη σχέση

$$\widetilde{D}^R_\phi(P,Q) = \int\limits_{\{p<q\}} r(x)q(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right)d\mu(x) + \int\limits_{\{q<p\}} r(x)p(x)\overline{\phi^*}\left(\frac{q(x)}{p(x)}\right)d\mu(x). \tag{2.17}$$

Από την άλλη πλευρά, σύμφωνα με τους Stummer and Vajda (2010, σελ. 174), για μια κυρτή συνάρτηση $\phi \in \Phi^*$ η οποία είναι επιπλέον αυστηρά κυρτή στο 1, με $\phi'_+(1) = 0$, ισχύει ότι

$$0 = \phi(1) \le \phi(t_2) \le \phi(t_1) \le \phi(0),\ \text{για κάθε } 0 \le t_1 \le t_2 \le 1. \tag{2.18}$$

Εφαρμόζοντας την ανισότητα (2.18) για $\phi = \overline{\phi}$, είναι φανερό ότι για το υποσύνολο $\{x \in \mathcal{X} : p(x) < q(x)\}$ του $\mathcal{X}$, ισχύει ότι $0 \le \overline{\phi}\left(\frac{p}{q}\right) \le \overline{\phi}(0)$ και συνεπώς

$$0 \le \int\limits_{\{p<q\}} r(x)q(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right)d\mu(x) \le \overline{\phi}(0)\int\limits_{\{p<q\}} r(x)q(x)d\mu(x).$$

Επιπλέον, με βάση τη μη-αρνητικότητα των r και q σε κάθε υποσύνολο του $\mathcal{X}$, ισχύει ότι

$$0 \le \int\limits_{\{p<q\}} r(x)q(x)d\mu(x) \le \int\limits_{\mathcal{X}} r(x)q(x)d\mu(x) = \xi_0.$$

Επομένως, οι δύο τελευταίες ανισότητες οδηγούν στην ανισότητα

$$0 \le \int\limits_{\{p<q\}} r(x)q(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right)d\mu(x) \le \overline{\phi}(0)\xi_0. \tag{2.19}$$

Με παρόμοιο τρόπο με τον παραπάνω, εφαρμόζοντας την ανισότητα (2.18) για $\phi = \overline{\phi^*}$, είναι φανερό ότι στο υποσύνολο $\{x \in \mathcal{X} : q(x) < p(x)\}$ του $\mathcal{X}$, ισχύει ότι $0 \leq \overline{\phi^*}\left(\frac{q}{p}\right) \leq \overline{\phi^*}(0)$, και συνεπώς

$$0 \leq \int\limits_{\{q<p\}} r(x)p(x)\overline{\phi^*}\left(\frac{q(x)}{p(x)}\right) d\mu(x) \leq \overline{\phi^*}(0) \int\limits_{\{q<p\}} r(x)p(x)d\mu(x).$$

Επομένως,

$$0 \leq \int\limits_{\{q<p\}} r(x)p(x)\overline{\phi^*}\left(\frac{q(x)}{p(x)}\right) d\mu(x) \leq \overline{\phi^*}(0) \int\limits_{\{q<p\}} r(x)p(x)d\mu(x) \leq \overline{\phi^*}(0)\xi_1. \tag{2.20}$$

Συνδυάζοντας τις (2.17), (2.19) και (2.20) προκύπτει ότι

$$\widetilde{D}_\phi^R(P,Q) \leq \overline{\phi}(0)\xi_0 + \overline{\phi^*}(0)\xi_1.$$

Με βάση τις σχέσεις (2.6) και (2.13) είναι $\overline{\phi}(0) = \phi(0) + \phi'_+(1)$ και $\overline{\phi^*}(0) = \phi^*(0) - \phi'_+(1)$. Αυτές οι ταυτότητες μαζί με την προηγούμενη ανισότητα ολοκληρώνουν την απόδειξη του μέρους (α) του θεωρήματος.

(β) Αρχικά υποθέτουμε ότι $P = Q$. Τότε, από τη σχέση (2.10) είναι φανερό ότι $\widetilde{D}_\phi^R(P,P) = D_\phi^R(P,P) = \phi(1) = 0$, αφού $\phi \in \Phi^*$.

Αντίστροφα, έστω ότι $\widetilde{D}_\phi^R(P,Q) = 0$. Λαμβάνοντας υπόψη τη σχέση (2.8) και το γεγονός ότι $\phi(1) = 0$, έχουμε

$$\phi\left(\frac{p(x)}{q(x)}\right) = \phi(1) + \phi'_+(1)\left(\frac{p(x)}{q(x)} - 1\right), \tag{2.21}$$

σ.π. ως προς το μέτρο μ, με Radon-Nikodym παράγωγο r θετική στο $\mathcal{X}$. Από την άλλη πλευρά, σύμφωνα με τον Vajda (1989, σελ. 58), έχουμε ότι

$$\phi\,(x) > \phi(1) + \phi'_+(1)\,(x-1)\,, \text{ για κάθε } x \neq 1,$$

διότι η ϕ είναι αυστηρά κυρτή στο 1. Συνεπώς, ο μόνος τρόπος για να ισχύει η ισότητα (2.21), λαμβάνοντας υπόψη την παραπάνω ανισότητα, είναι όταν $\frac{p(x)}{q(x)} = 1$ ή $P = Q$, το οποίο ολοκληρώνει την απόδειξη του μέρους (β) του θεωρήματος.

(γ) Υποθέτουμε ότι $P \perp Q$. Τότε, σύμφωνα με τον Vajda (1972, σελ. 227), $u = \frac{dQ}{dP+dQ} = 0$, σ.π. P και $u = \frac{dQ}{dP+dQ} = 1$, σ.π. Q. Λαμβάνοντας υπόψη ότι $u = \frac{q}{p+q}$, συμπεραίνουμε ότι αν $P \perp Q$ τότε $q(x) = 0$, σ.π. P, $x \in \mathcal{X}$ και $p(x) = 0$, σ.π. Q, $x \in \mathcal{X}$. Η σχέση (2.17) επαναδιατυπώνεται ως εξής:

$$\widetilde{D}^R_\phi(P,Q) = \int\limits_{\{p<q\}} r(x)\overline{\phi}\left(\frac{p(x)}{q(x)}\right) dQ(x) + \int\limits_{\{q<p\}} r(x)\overline{\phi^*}\left(\frac{q(x)}{p(x)}\right) dP(x),$$

και εφόσον $P \perp Q$, προκύπτει ότι

$$\begin{aligned} \widetilde{D}^R_\phi(P,Q) &= \int\limits_{\{p<q\}} r(x)\overline{\phi}\left(\tfrac{p(x)}{q(x)}\right) dQ(x) + \int\limits_{\{q<p\}} r(x)\overline{\phi^*}\left(\tfrac{q(x)}{p(x)}\right) dP(x) \\ &= \overline{\phi}\,(0) \int\limits_{\{p<q\}} r(x)dQ(x) + \overline{\phi^*}\,(0) \int\limits_{\{q<p\}} r(x)dP(x). \end{aligned} \tag{2.22}$$

Από την άλλη πλευρά, επειδή $p(x) = 0$, σ.π. Q, $x \in \mathcal{X}$, είναι σαφές ότι

$$Q(\{p \geq q\}) = Q(\{p > q\}) + Q(\{p = q\}) = Q(\{p > q\}) = Q(\{q < 0\}) = Q(\varnothing) = 0$$

αφού $Q(\{q = p\}) = 0$, λαμβάνοντας υπόψη ότι $P \perp Q$. Αυτή η τελευταία ισότητα οδηγεί στο $\int\limits_{\{p \geq q\}} r(x)dQ(x) = 0$, και συνεπώς

$$\begin{aligned} \xi_0 = \int\limits_{\mathcal{X}} r(x)q(x)d\mu(x) = \int\limits_{\mathcal{X}} r(x)dQ(x) &= \int\limits_{\{p<q\}} r(x)dQ(x) + \int\limits_{\{p\geq q\}} r(x)dQ(x) \\ &= \int\limits_{\{p<q\}} r(x)dQ(x). \end{aligned} \tag{2.23}$$

Ομοίως, μπορεί να αποδειχθεί ότι $P(\{q \geq p\}) = 0$ και συνεπώς

$$\xi_1 = \int\limits_{\mathcal{X}} r(x)q(x)d\mu(x) = \int\limits_{\{q<p\}} r(x)dP(x). \tag{2.24}$$

Σύμφωνα με τις σχέσεις (2.22), (2.23) και (2.24), αν $P \perp Q$ τότε $\widetilde{D}^R_\phi(P,Q) = \overline{\phi}\,(0)\,\xi_0 + \overline{\phi^*}\,(0)\,\xi_1$ και η απόδειξη αυτού του τμήματος του θεωρήματος ολοκληρώνεται μέσω των ισοτήτων $\overline{\phi}(0) = \phi(0) + \phi'_+(1)$ και $\overline{\phi^*}(0) = \phi^*(0) - \phi'_+(1)$.

Απομένει να αποδειχθεί ότι, εάν $\phi(0)+\phi^*(0) < \infty$ και $\widetilde{D}^R_\phi(P,Q) = \overline{\phi}\,(0)\,\xi_0 + \overline{\phi^*}\,(0)\,\xi_1$, τότε $P \perp Q$. Οι σχέσεις (2.12) και (2.18) οδηγούν άμεσα στη σχέση

$$\widetilde{D}^R_\phi(P,Q) \leq \overline{\phi}\,(0) \int\limits_{\{p<q\}} r(x)dQ(x) + \overline{\phi^*}\,(0) \int\limits_{\{q<p\}} r(x)dP(x). \tag{2.25}$$

Η τελευταία ανισότητα, με την παραδοχή ότι $\widetilde{D}_{\phi}^{R}(P,Q) = \overline{\phi}\,(0)\,\xi_0 + \overline{\phi^*}\,(0)\,\xi_1$ και $\phi(0) + \phi^*(0) < \infty$, δίνει

$$\int\limits_{\{p<q\}} r(x)dQ(x) = \xi_0 \quad \text{και} \quad \int\limits_{\{q<p\}} r(x)dP(x) = \xi_1,$$

και συνεπώς

$$\int\limits_{\{p\geq q\}} r(x)dQ(x) = \int\limits_{\{q\geq p\}} r(x)dP(x) = 0.$$

Η τελευταία αυτή ισότητα δίνει

$$Q(\{p \geq q\}) = 0 \quad \text{και} \quad P(\{q \geq p\}) = 0,$$

ή

$$Q(\{p \geq q\}) = 0 \quad \text{και} \quad P(\{q > p\}) = 0,$$

για Radon-Nikodym παράγωγο r θετική στο $\mathcal{X}$. Το τελευταίο αυτό συμπέρασμα αποδεικνύει ότι $P \perp Q$ και η απόδειξη του θεωρήματος έχει ολοκληρωθεί. ∎

2.2 Τοπική ϕ-απόκλιση για την εκθετική οικογένεια κατανομών

Όπως έχει ήδη αναφερθεί στο Κεφάλαιο 1 μέτρα τύπου απόκλισης (ή και τύπου εντροπίας) έχουν εφαρμοστεί ευρέως σε διάφορους κλάδους και πλαίσια, όχι μόνο στην κλασική και σύγχρονη στατιστική, αλλά σχεδόν σε κάθε επιστημονικό πεδίο. Συνεπώς, χρήζει μεγάλης σημασίας, ο προσδιορισμός των εκφράσεων των αποκλίσεων, σε κλειστούς τύπους, για ευρέως χρησιμοποιούμενες οικογένειες κατανομών. Για περισσότερες λεπτομέρειες παραπέμπουμε, μεταξύ άλλων, στους Soofi and Retzer (2002), Zografos and Nadarajah (2005), Zografos (2008) και Pardo (2006) καθώς και στις αναφορές που υπάρχουν στις εργασίες αυτές.

Σκοπός της ενότητας αυτής είναι να δοθούν αναλυτικές εκφράσεις για συγκεκριμένες επιλογές της τοπικής ϕ-απόκλισης μεταξύ δύο μελών της εκθετικής οικογένειας κατανομών. Ειδικότερα θα προσδιοριστούν οι αναλυτικές εκφράσεις της τοπικής απόκλισης Cressie and Read και της τοπικής απόκλισης Kullback-Leibler μεταξύ δύο μελών της εκθετικής οικογένειας κατανομών. Κίνητρο για τη μελέτη αυτή αποτελεί το γεγονός ότι από τη μια μεριά η εκθετική οικογένεια κατανομών είναι μια ευρεία οικογένεια η οποία περιλαμβάνει την πλειοψηφία των πιο γνωστών χρησιμοποιημένων, στην πράξη, στατιστικών κατανομών, ενώ από την άλλη μεριά οι αποκλίσεις Cressie and Read και Kullback-Leibler είναι οι πιο γνωστές περιπτώσεις ϕ-αποκλίσεων, ευρέως χρησιμοποιούμενες στη πράξη. Στη συνέχεια παρατίθεται ο ορισμός της εκθετικής οικογένειας για λόγους πληρότητας. Σε όσα ακολουθούν με A^t συμβολίζουμε τον ανάστροφο ενός διανύσματος ή ενός πίνακα, ανάλογα, με $|A|$ την ορίζουσα του πίνακα A και με A^{-1} τον αντίστροφο του τετραγωνικού πίνακα A. Τέλος, θα πρέπει να σημειωθεί ότι το εσωτερικό γινόμενο των $\alpha = (u, M)$ και $\beta = (v, N)$ τα οποία αποτελούνται από δύο συνιστώσες, μια διανυσματική u και v και μια πινάκων M και N, ορίζεται από τη σχέση $\alpha^t \beta = u^t v + trace(M^t N)$ (βλέπε, Nielsen and Nock, 2011, σελ. 6).

Ορισμός 2.5. *Μια κατανομή ανήκει στην εκθετική οικογένια κατανομών αν η συνάρτηση πιθανότητας ή η συνάρτηση πυκνότητας πιθανότητάς της μπορεί να γραφεί στη μορφή*

$$f_C(x, \theta) = \exp\left\{\theta^t T(x) - C(\theta) + w(x)\right\}, \ x \in \mathcal{X}, \tag{2.26}$$

όπου $\theta \in \Theta \subseteq R^M$ και $T(x) = (T_1(x), ..., T_M(x))^t$, $x \in \mathcal{X}$. Οι $w(x)$ και $T_1(x), ..., T_M(x)$ είναι πραγματικές συναρτήσεις οι οποίες είναι ανεξάρτητες της παράμετρου θ, ενώ η $C(\theta)$ είναι πραγματική συνάρτηση ανεξάρτητη του x.

Ειδική περίπτωση της εκθετικής οικογένειας κατανομών αποτελεί η πολυδιάστατη κανονική κατανομή, η πιο ευρέως χρησιμοποιούμενη κατανομή της Πολυμεταβλητής Στατιστικής. Ειδικότερα, η συνάρτηση πυκνότητας πιθανότητας ενός k-διάστατου κανονικού τυχαίου διανύσματος X δίνεται από τη σχέση

$$f_X(x) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right), x \in R^k$$

και λέγεται ότι ακολουθεί την κανονική κατανομή με μέσο διάνυσμα $\mu \in R^k$ και πίνακα διακυμάνσεων-συνδιακυμάνσεων Σ, με Σ θετικά ορισμένο. Σε μια τέτοια περίπτωση γράφουμε ότι $X \sim N_k(\mu, \Sigma)$. Εύκολα προκύπτει ότι η ανωτέρω k-διάστατη κανονική κατανομή ανήκει στην εκθετική οικογένεια κατανομών (2.26) με φυσικές παραμέτρους

$$\begin{aligned}
\theta = (\theta_1, \theta_2) = \left(\Sigma^{-1}\mu, \tfrac{1}{2}\Sigma^{-1}\right), T(x) &= (T_1(x), T_2(x)) = \left(x, xx^t\right), \\
C(\theta) &= \log\left((2\pi)^{k/2}|\Sigma|^{1/2}\right) + \tfrac{1}{2}\mu^t\Sigma^{-1}\mu \\
&= \log(2\pi)^{k/2} - \tfrac{1}{2}\log\left(|-2\theta_2|\right) - \tfrac{1}{4}\theta_1^t\theta_2^{-1}\theta_1, \\
w(x) &= 0.
\end{aligned} \tag{2.27}$$

2.2.1 Τοπική απόκλιση Cressie and Read για την εκθετική οικογένεια κατανομών

Στην υποενότητα αυτή θα προσδιοριστεί η έκφραση της τοπικής απόκλισης Cressie and Read μεταξύ δύο μελών της εκθετικής οικογένειας κατανομών του Ορισμού 2.5, δηλαδή δύο μελών με σ.π.π., $f_C(x,\theta_i)$, $\theta_i \in \Theta \subseteq R^M$, $i=1,2$ που προσδιορίζονται από τη σχέση (2.26). Τα γενικά αποτελέσματα που θα προκύψουν εφαρμόζονται στην ειδική περίπτωση της πολυδιάστατης κανονικής κατανομής.

Λαμβάνοντας υπόψη τη σχέση (2.4) και το γεγονός ότι σε αυτήν την ειδική περίπτωση (βλέπε και Πίνακα 1.1. του Κεφαλαίου 1) η κυρτή συνάρτηση είναι η

$$\phi(u) = \phi_\lambda(u) = \frac{u^{\lambda+1} - u - \lambda(u-1)}{\lambda(\lambda+1)}, \lambda \neq 0, -1,$$

η τοπική απόκλιση Cressie and Read δίνεται από την παρακάτω σχέση

$$D^{\omega}_{\phi_\lambda}(\theta_1,\theta_2) = \frac{1}{\lambda(\lambda+1)}\left[K_{\lambda,\omega}(\theta_1,\theta_2) - (\lambda+1)E_{\theta_1}\left(h_\omega(X)\right) + \lambda E_{\theta_2}\left(h_\omega(X)\right)\right], \quad (2.28)$$

με $\lambda \neq 0, -1$, όπου

$$K_{\lambda,\omega}(\theta_1,\theta_2) = \int_{\mathcal{X}} h_\omega(x)\frac{f_C^{\lambda+1}(x,\theta_1)}{f_C^{\lambda}(x,\theta_2)}d\mu(x) \quad (2.29)$$

και

$$E_{\boldsymbol{\theta}_i}\left(h_\omega(X)\right) = \int_{\mathcal{X}} h_\omega(x) f_C(x,\theta_i)d\mu(x), \quad (2.30)$$

$\omega \in \Theta^* \subseteq R^K$, $\theta_i \in \Theta \subseteq R^M$, $i=1,2$.

Στην επόμενη πρόταση παρουσιάζεται η αναλυτική έκφραση της $D^{\omega}_{\phi_\lambda}(\theta_1,\theta_2)$, όταν ο πυρήνας πυκνότητα h_ω ορίζεται στο $\mathcal{X}$ αλλά δεν ανήκει απαραίτητα στην οικογένεια πυκνοτήτων της σχέσης (2.26).

Πρόταση 2.2. *Έστω ο πυρήνας πυκνότητα h_ω που ορίζεται στο σύνολο $\mathcal{X}$ και ας θεωρήσουμε δύο μέλη $f_C(x,\theta_1)$ και $f_C(x,\theta_2)$ της (2.26). Αν $(\lambda+1)\theta_1 - \lambda\theta_2 \in \Theta$, με $\lambda \neq 0,-1$, τότε η τοπική απόκλιση Cressie and Read μεταξύ των πυκνοτήτων $f_C(x,\theta_1)$ και $f_C(x,\theta_2)$ που οδηγείται από την πυκνότητα h_ω, δίνεται από τη σχέση*

$$\begin{aligned} D^{\omega}_{\phi_\lambda}(\theta_1,\theta_2) &= \tfrac{1}{\lambda(\lambda+1)}\left\{\left(\exp\left[M^{(1)}_{C,\lambda}(\theta_1,\theta_2)\right]\right)E_{(\lambda+1)\theta_1-\lambda\theta_2}\left(h_\omega(X)\right)\right\} \\ &-\tfrac{1}{\lambda}E_{\theta_1}\left(h_\omega(X)\right) + \tfrac{1}{\lambda+1}E_{\theta_2}\left(h_\omega(X)\right) \end{aligned}, \quad (2.31)$$

όπου

$$M^{(1)}_{C,\lambda}(\theta_1,\theta_2) = \lambda C(\theta_2) - (\lambda+1)C(\theta_1) + C((\lambda+1)\theta_1 - \lambda\theta_2) \qquad (2.32)$$

και $E_{(\lambda+1)\theta_1-\lambda\theta_2}\left(h_\omega(X)\right)$, $E_{\theta_i}\left(h_\omega(X)\right)$, $i=1,2$, *όπως ορίζονται σύμφωνα με τη σχέση (2.30).*

Απόδειξη. Με βάση τη σχέση (2.29), απλή άλγεβρα δίνει

$$\begin{aligned} K_{\lambda,\omega}(\theta_1,\theta_2) &= \int_{\mathcal{X}} h_\omega(x)\exp\left([(\lambda+1)\theta_1^t - \lambda\theta_2^t]T(x)\right) \\ &\times \exp\left(\lambda C(\theta_2) - (\lambda+1)C(\theta_1) + w(x)\right) d\mu(x). \end{aligned}$$

Έτσι,

$$\begin{aligned} K_{\lambda,\omega}(\theta_1,\theta_2) &= \int_{\mathcal{X}} h_\omega(x)\exp\left([(\lambda+1)\theta_1^t - \lambda\theta_2^t]T(x) - C\left((\lambda+1)\theta_1 - \lambda\theta_2\right) + w(x)\right) \\ &\times \exp\left(\lambda C(\theta_2) - (\lambda+1)C(\theta_1)\right) \times \exp\left(C\left((\lambda+1)\theta_1 - \lambda\theta_2\right)\right) d\mu(x), \end{aligned}$$

η οποία λαμβάνοντας υπόψη τις σχέσεις (2.28) και (2.32) οδηγεί στο επιθυμητό αποτέλεσμα. ∎

Στην πρόταση που ακολουθεί προσδιορίζεται η αναλυτική έκφραση της $D^{\omega}_{\phi_\lambda}(\theta_1,\theta_2)$ όταν ο πυρήνας πυκνότητα h_ω ανήκει στην κλάση των πυκνοτήτων (2.26), όποτε $\Theta^* = \Theta$.

Πρόταση 2.3. *Θεωρούμε δύο μέλη $f_C(x,\theta_1)$ και $f_C(x,\theta_2)$ της (2.26) και ας υποθέσουμε ότι και ο πυρήνας πυκνότητα $h_\omega(x) = f_C(x,\omega)$ είναι μέλος της (2.26). Τότε, κάτω από τις υπόθεσεις ότι $\theta_i + \omega \in \Theta$, $i=1,2$ και $(\lambda+1)\theta_1 - \lambda\theta_2 + \omega \in \Theta$, με $\lambda \neq 0,-1$, η τοπική απόκλιση Cressie and Read μεταξύ των πυκνοτήτων $f_C(x,\theta_1)$ και $f_C(x,\theta_2)$ που οδηγείται από τον πυρήνα h_ω, δίνεται από τη σχέση*

$$\begin{aligned} D^{\omega}_{\phi_\lambda}(\theta_1,\theta_2) &= \tfrac{1}{\lambda(\lambda+1)}\left\{\left(\exp\left[M^{(2)}_{C,\lambda}(\theta_1,\theta_2,\omega)\right]\right) E_{(\lambda+1)\theta_1-\lambda\theta_2+\omega}\left(\exp\left(w(X)\right)\right)\right. \\ &-(\lambda+1)\exp[C(\theta_1+\omega) - C(\theta_1) - C(\omega)] \times E_{\theta_1+\omega}\left(\exp\left(w(X)\right)\right) \\ &\left.+\lambda\exp[C(\theta_2+\omega) - C(\theta_2) - C(\omega)] \times E_{\theta_2+\omega}\left(\exp\left(w(X)\right)\right)\right\}, \end{aligned} \qquad (2.33)$$

όπου

$$M^{(2)}_{C,\lambda}(\theta_1,\theta_2,\omega) = \lambda C(\theta_2) - (\lambda+1)C(\theta_1) - C(\omega) + C((\lambda+1)\theta_1 - \lambda\theta_2 + \omega) \qquad (2.34)$$

και

$$E_{\theta_i+\omega}\left(\exp\left(w(X)\right)\right) = \int_{\mathcal{X}} \left\{\exp\left(w(X)\right)\right\} f_C(x,\theta_i+\omega)d\mu(x),\ i=1,2, \qquad (2.35)$$

$$E_{(\lambda+1)\theta_1-\lambda\theta_2+\omega}\left(\exp\left(w(X)\right)\right)=\int\limits_{\mathcal{X}}\left\{\exp\left(w(X)\right)\right\}f_C(x,(\lambda+1)\theta_1-\lambda\theta_2+\omega)d\mu(x). \quad (2.36)$$

Απόδειξη. Σύμφωνα με τη (2.29), λίγη άλγεβρα δίνει

$$\begin{aligned} K_{\lambda,\omega}(\theta_1,\theta_2) &= \exp\{\lambda C(\theta_2)-(\lambda+1)C(\theta_1)-C(\omega)+C((\lambda+1)\theta_1-\lambda\theta_2+\omega)\} \\ &\times\int\limits_{\mathcal{X}}\exp\{w(x)\}\exp\left\{\left(\sum_{i=1}^{k}[(\lambda+1)\theta_{1i}-\lambda\theta_{2i}+\omega_i]T_i(x)\right)\right. \\ &\left.-C((\lambda+1)\theta_1-\lambda\theta_2+\omega)+w(x)\right\}d\mu(x). \end{aligned}$$

Λαμβάνοντας υπόψη τις (2.34) και (2.36), έχουμε

$$K_{\lambda,\omega}(\theta_1,\theta_2)=\exp\left\{M_{C,\lambda}^{(2)}(\theta_1,\theta_2,\omega)\right\}E_{(\lambda+1)\theta_1-\lambda\theta_2+\omega}\{\exp(w(x))\}. \quad (2.37)$$

Από την άλλη πλευρά, μπορεί εύκολα να αποδειχθεί ότι η μέση τιμή

$$E_{\theta_j}\left(h_\omega(X)\right)=\int\limits_{\mathcal{X}}h_\omega(x)f_C(x,\theta_j)d\mu(x),\quad j=1,2,$$

που ορίστηκε στη (2.30), δίνεται από τη σχέση

$$\begin{aligned} E_{\theta_j}\left(h_\omega(X)\right) &= \exp\{-C(\theta_j)-C(\omega)+C(\theta_j+\omega)\} \\ &\times\int\limits_{\mathcal{X}}\exp(w(x))\exp\left\{\left(\sum_{i=1}^{k}(\omega_i+\theta_{ji})T_i(x)-C(\theta_j+\omega)+w(x)\right)\right\}d\mu(x), \end{aligned}$$

$j=1,2$, ή

$$E_{\theta_j}\left(h_\omega(X)\right)=\exp\{-C(\theta_j)-C(\omega)+C(\theta_j+\omega)\}E_{\theta_j+\omega}\left(\exp\left(w(X)\right)\right),\ j=1,2. \quad (2.38)$$

Το αποτέλεσμα (2.33) έπεται ως μια εφαρμογή των (2.28), (2.37) και (2.38). ∎

Στην ακόλουθη πρόταση εφαρμόζεται το αποτέλεσμα της Πρότασης 2.3 στην ειδική περίπτωση της k-διάστατης κανονικής κατανομής. Ειδικότερα, υποθέτουμε ότι ο πυρήνας πυκνότητα $f_{N_k(\mu,\Sigma)}$, είναι η πολυδιάστατη κανονική κατανομή $N_k(\mu,\Sigma)$ με μέσο διάνυσμα $\mu\in R^k$ και πίνακα διακυμάνσεων-συνδιακυμάνσεων Σ. Επιπλέον, θεωρούμε δύο πυκνότητες $f_{N_k(\mu_1,\Sigma_1)}$ και $f_{N_k(\mu_2,\Sigma_2)}$ στο χώρο $\mathcal{X}=R^k$, που ακολουθούν τις k-διάστατες κανονικές κατανομές $N_k(\mu_1,\Sigma_1)$ και $N_k(\mu_2,\Sigma_2)$, με παραμέτρους (μ_1,Σ_1) και (μ_2,Σ_2), αντίστοιχα.

Πρόταση 2.4. *Η τοπική απόκλιση Cressie and Read* $D_{\phi_\lambda}^{(\mu,\Sigma)} = D_{\phi_\lambda}^{(\mu,\Sigma)}((\mu_1,\Sigma_1),(\mu_2,\Sigma_2))$ *μεταξύ δύο k-διάστατων κανονικών κατανομών* $N_k(\mu_1,\Sigma_1)$ *και* $N_k(\mu_2,\Sigma_2)$, *οδηγούμενη από μια k-διάστατη κανονική κατανομή* $N_k(\mu,\Sigma)$, *δίνεται από τη σχέση*

$$\begin{aligned}
D_{\phi_\lambda}^{(\mu,\Sigma)} &= \tfrac{1}{\lambda(\lambda+1)} \Big\{ (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} |\Sigma_1|^{-\frac{\lambda+1}{2}} |\Sigma_2|^{\frac{\lambda}{2}} \left|(\lambda+1)\Sigma_1^{-1} - \lambda\Sigma_2^{-1} + \Sigma^{-1}\right|^{-\frac{1}{2}} \\
&\times \exp\left(-\tfrac{1}{2}\left(\mu^t\Sigma^{-1}\mu + (\lambda+1)\mu_1^t\Sigma_1^{-1}\mu_1 - \lambda\mu_2^t\Sigma_2^{-1}\mu_2 - B_1^t B_2 B_1\right)\right) \\
&-(\lambda+1)(2\pi)^{-k/2}|\Sigma|^{-1/2}|\Sigma_1|^{-1/2}\left|\Sigma_1^{-1} + \Sigma^{-1}\right|^{-1/2} \\
&\times \exp\left(-\tfrac{1}{2}(\mu-\mu_1)^t(\Sigma+\Sigma_1)^{-1}(\mu-\mu_1)\right) \\
&+\lambda(2\pi)^{-k/2}|\Sigma|^{-1/2}|\Sigma_2|^{-1/2}\left|\Sigma_2^{-1} + \Sigma^{-1}\right|^{-1/2} \\
&\times \exp\left(-\tfrac{1}{2}(\mu-\mu_2)^t(\Sigma+\Sigma_2)^{-1}(\mu-\mu_2)\right)\Big\},
\end{aligned}$$

όπου

$$\begin{aligned}
B_1 &= (\lambda+1)\Sigma_1^{-1}\mu_1 - \lambda\Sigma_2^{-1}\mu_2 + \Sigma^{-1}\mu, \\
B_2 &= \left((\lambda+1)\Sigma_1^{-1} - \lambda\Sigma_2^{-1} + \Sigma^{-1}\right)^{-1},
\end{aligned}$$

εφόσον $(\lambda+1)\Sigma_1^{-1} - \lambda\Sigma_2^{-1} + \Sigma^{-1} > 0$, *για* $\lambda \neq 0, -1$.

Απόδειξη. Με βάση την Πρόταση 2.3, έχουμε

$$\begin{aligned}
D_{\phi_\lambda}^{(\mu,\Sigma)} &= \tfrac{1}{\lambda(\lambda+1)} \left\{ \left(\exp\left[M_{C,\lambda}^{(2)}(\theta_1,\theta_2,\omega)\right]\right) E_{(\lambda+1)\theta_1-\lambda\theta_2+\omega}\left(\exp\left(w(X)\right)\right) \right. \\
&-(\lambda+1)\exp[C(\theta_1+\omega) - C(\theta_1) - C(\omega)] \times E_{\theta_1+\omega}\left(\exp\left(w(X)\right)\right) \\
&\left. +\lambda\exp[C(\theta_2+\omega) - C(\theta_2) - C(\omega)] \times E_{\theta_2+\omega}\left(\exp\left(w(X)\right)\right) \right\},
\end{aligned} \tag{2.39}$$

με $\theta_1 = (\theta_{11},\theta_{12}) = (\Sigma_1^{-1}\mu_1, -\frac{1}{2}\Sigma_1^{-1})$, $\theta_2 = (\theta_{21},\theta_{22}) = (\Sigma_2^{-1}\mu_2, -\frac{1}{2}\Sigma_2^{-1})$, $\omega = (\omega_1,\omega_2) = (\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1})$ και

$$M_{C,\lambda}^{(2)}(\theta_1,\theta_2,\omega) = \lambda C(\theta_2) - (\lambda+1)C(\theta_1) - C(\omega) + C((\lambda+1)\theta_1 - \lambda\theta_2 + \omega). \tag{2.40}$$

Σύμφωνα με τις σχέσεις (2.27), έχουμε

$$\begin{aligned}
C(\theta_i) &= \log\left((2\pi)^{k/2}|\Sigma_i|^{1/2}\right) + \tfrac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i, \ i = 1,2 \\
C(\omega) &= \log\left((2\pi)^{k/2}|\Sigma|^{1/2}\right) + \tfrac{1}{2}\mu^t\Sigma^{-1}\mu.
\end{aligned} \tag{2.41}$$

Από την άλλη πλευρά,

$$\theta_1 + \omega = \left(\Sigma_1^{-1}\mu_1 + \Sigma^{-1}\mu, -\frac{1}{2}(\Sigma_1^{-1} + \Sigma^{-1})\right),$$

και άμεσα προκύπτει, μέσω της (2.27), ότι

$$\begin{aligned} C(\theta_1+\omega) &= \log\left((2\pi)^{k/2}|\Sigma^{-1}+\Sigma_1^{-1}|^{-1/2}\right) \\ &+\tfrac{1}{2}\left(\Sigma^{-1}\mu+\Sigma_1^{-1}\mu_1\right)^t\left(\Sigma^{-1}+\Sigma_1^{-1}\right)^{-1}\left(\Sigma^{-1}\mu+\Sigma_1^{-1}\mu_1\right). \end{aligned} \tag{2.42}$$

Λαμβάνοντας υπόψη την ταυτότητα

$$\begin{aligned} &\left(\Sigma^{-1}\mu+\Sigma_i^{-1}\mu_i\right)^t\left(\Sigma^{-1}+\Sigma_i^{-1}\right)^{-1}\left(\Sigma^{-1}\mu+\Sigma_i^{-1}\mu_i\right)-\mu^t\Sigma^{-1}\mu-\mu_i^t\Sigma_i^{-1}\mu_i \\ &\qquad = -(\mu-\mu_i)^t(\Sigma+\Sigma_i)^{-1}(\mu-\mu_i),\ i=1,2, \end{aligned}$$

(βλέπε, Pardo, 2006, σελ. 49), μετά από λίγη άλγεβρα συνεπάγεται ότι

$$\begin{aligned} C(\theta_i+\omega)-C(\theta_i)-C(\omega) &= \log\left((2\pi)^{-k/2}|\Sigma^{-1}+\Sigma_i^{-1}|^{-1/2}|\Sigma|^{-1/2}|\Sigma_i|^{-1/2}\right) \\ &-\tfrac{1}{2}(\mu-\mu_i)^t(\Sigma+\Sigma_i)^{-1}(\mu-\mu_i),\ i=1,2. \end{aligned} \tag{2.43}$$

Απομένει να αποδειχθεί ότι η ποσότητα $M_{C,\lambda}^{(2)}(\theta_1,\theta_2,\omega)$ δίνεται από τη σχέση (2.40). Είναι εύκολο να δούμε ότι

$$\begin{aligned} (\lambda+1)\theta_1-\lambda\theta_2+\omega &= \left((\lambda+1)\Sigma_1^{-1}\mu_1-\lambda\Sigma_2^{-1}\mu_2+\Sigma^{-1}\mu,\right. \\ &\quad \left.(-1/2)\left((\lambda+1)\Sigma_1^{-1}-\lambda\Sigma_2^{-1}+\Sigma^{-1}\right)\right), \end{aligned}$$

και συνεπώς

$$\begin{aligned} C((\lambda+1)\theta_1-\lambda\theta_2+\omega) &= \log\left((2\pi)^{k/2}|(\lambda+1)\Sigma_1^{-1}-\lambda\Sigma_2^{-1}+\Sigma^{-1}|^{-1/2}\right) \\ &+\tfrac{1}{2}\left((\lambda+1)\Sigma_1^{-1}\mu_1-\lambda\Sigma_2^{-1}\mu_2+\Sigma^{-1}\mu\right)^t \\ &\times\left((\lambda+1)\Sigma_1^{-1}-\lambda\Sigma_2^{-1}+\Sigma^{-1}\right)^{-1} \\ &\times\left((\lambda+1)\Sigma_1^{-1}\mu_1-\lambda\Sigma_2^{-1}\mu_2+\Sigma^{-1}\mu\right), \end{aligned} \tag{2.44}$$

με $(\lambda+1)\Sigma_1^{-1}-\lambda\Sigma_2^{-1}+\Sigma^{-1}>0$, για $\lambda\neq 0,-1$. Με βάση τις σχέσεις (2.40), (2.41) και (2.44), προκκύπτει ότι

$$\begin{aligned} M_{C,\lambda}^{(2)}(\theta_1,\theta_2,\omega) = \ & \log\left((2\pi)^{-k/2}\right)|\Sigma|^{-\frac{1}{2}}|\Sigma_1|^{-\frac{\lambda+1}{2}}|\Sigma_2|^{\frac{\lambda}{2}}\left|(\lambda+1)\Sigma_1^{-1}-\lambda\Sigma_2^{-1}+\Sigma^{-1}\right|^{-\frac{1}{2}} \\ & -\tfrac{1}{2}\left(\mu^t\Sigma^{-1}\mu+(\lambda+1)\mu_1^t\Sigma_1^{-1}\mu_1-\lambda\mu_2^t\Sigma_2^{-1}\mu_2-B_1^tB_2B_1\right), \end{aligned} \tag{2.45}$$

με

$$\begin{aligned} B_1 &= (\lambda+1)\Sigma_1^{-1}\mu_1-\lambda\Sigma_2^{-1}\mu_2+\Sigma^{-1}\mu, \\ B_2 &= \left((\lambda+1)\Sigma_1^{-1}-\lambda\Sigma_2^{-1}+\Sigma^{-1}\right)^{-1}. \end{aligned}$$

Λαμβάνοντας υπόψη ότι $w(X)=0$ (βλέπε 2.27), το αποτέλεσμα προκύπτει ως εφαρμογή των (2.39), (2.43) και (2.45). ∎

Παρατήρηση 2.7. *Η αναλυτική έκφραση για την τοπική απόκλιση Cressie and Read μεταξύ των μονοδιάστατων κανονικών κατανομών $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$ και πυρήνα $N(\mu, \sigma^2)$ μπορεί να προκύψει με άμεση εφαρμογή της παραπάνω πρότασης. Οι αντίστοιχοι τύποι έχουν ως εξής:*

$$D_{\phi_\lambda}^{\omega}(\theta_1, \theta_2) = \frac{1}{\lambda(\lambda+1)} \left\{ K_{\lambda,\omega}(\theta_1, \theta_2) - (\lambda+1) E_{\theta_1}\left[h_\omega(X)\right] + \lambda E_{\theta_2}\left[h_\omega(X)\right] \right\}, \quad (2.46)$$

όπου

$$\begin{aligned} K_{\lambda,\omega}(\theta_1, \theta_2) &= \int_{\mathcal{X}} h_\omega(x) f_{\theta_1}^{\lambda+1}(x) f_{\theta_2}^{-\lambda}(x) d\mu(x) \\ &= \frac{(2\pi)^{-1/2} \sigma_1^{-\lambda} \sigma_2^{\lambda+1}}{\left(\sigma_1^2\sigma_2^2 + (\lambda+1)\sigma^2\sigma_2^2 - \lambda\sigma^2\sigma_1^2\right)^{1/2}} \exp\left\{-\tfrac{1}{2}\left(B_1 + B_2\right)\right\}, \end{aligned}$$

με

$$B_1 = -\frac{\lambda(\lambda+1)\left(\mu_1 - \mu_2\right)^2}{(\lambda+1)\sigma_2^2 - \lambda\sigma_1^2}, \quad B_2 = (\mu - \widetilde{\mu})^2 \frac{(\lambda+1)\sigma_2^2 - \lambda\sigma_1^2}{\sigma_1^2\sigma_2^2 + (\lambda+1)\sigma^2\sigma_2^2 - \lambda\sigma^2\sigma_1^2},$$

και

$$\widetilde{\mu} = \frac{(\lambda+1)\mu_1\sigma_2^2 - \lambda\mu_2\sigma_1^2}{(\lambda+1)\sigma_2^2 - \lambda\sigma_1^2}.$$

Επιπλέον, είναι

$$E_{\theta_i}\left[h_\omega(X)\right] = \left(2\pi(\sigma^2 + \sigma_i^2)\right)^{-1/2} \exp\left\{-\frac{(\mu - \mu_i)^2}{2(\sigma^2 + \sigma_i^2)}\right\}, \quad i = 1, 2.$$

2.2.2 Τοπική απόκλιση Kullback-Leibler για την εκθετική οικογένεια κατανομών

Στην υποενότητα αυτή θα προσδιοριστεί η έκφραση της τοπικής απόκλισης Kullback-Leibler μεταξύ δύο μελών της εκθετικής οικογένειας κατανομών του Ορισμού 2.5, δηλαδή δύο μελών με σ.π.π., $f_C(x, \theta_i)$, $\theta_i \in \Theta \subseteq R^M$, $i = 1, 2$, που προσδιορίζονται από τη σχέση (2.26). Τα γενικά αποτελέσματα που θα προκύψουν εφαρμόζονται στην ειδική περίπτωση της πολυδιάστατης κανονικής κατανομής.

Λαμβάνοντας υπόψη τη σχέση (2.4) και το γεγονός ότι σε αυτήν την ειδική περίπτωση (βλέπε και Πίνακα 1.1 του Κεφαλαίου 1) η κυρτή συνάρτηση είναι η $\phi(u) = u \log u - u + 1$ προκύπτει ότι η τοπική απόκλιση Kullback-Leibler ορίζεται από τη σχέση

$$\begin{aligned} D_0^R(P, Q) &= \int_{\mathcal{X}} \frac{dR}{d\mu} \frac{dP}{dQ} \log\left(\frac{dP}{dQ}\right) dQ - \int_{\mathcal{X}} \frac{dR}{d\mu} dP + \int_{\mathcal{X}} \frac{dR}{d\mu} dQ \\ &= \int_{\mathcal{X}} r(x)p(x) \log\left(\frac{p(x)}{q(x)}\right) d\mu(x) - \int_{\mathcal{X}} r(x)p(x) d\mu(x) + \int_{\mathcal{X}} r(x)q(x) d\mu(x). \end{aligned} \quad (2.47)$$

Παρατήρηση 2.8. *Θα πρέπει να σημειωθεί ότι, η κλασική απόκλιση Kullback-Leibler δίνεται στη (1.1) και λαμβάνεται από την (1.3) για $\phi(u) = u \log u$ ή $\phi(u) = u \log u - u + 1$, $u > 0$. Αυτό δεν ισχύει για την τοπική απόκλιση Kullback-Leibler. Η τελευταία, ορίζεται από τη σχέση (2.47), ως μια συγκεκριμένη περίπτωση της (2.3) ή της (2.4), μόνο όταν $\phi(u) = u \log u - u + 1$.*

Η επόμενη πρόταση παρέχει τις αναλυτικές εκφράσεις της τοπικής Kullback-Leibler απόκλισης μεταξύ δύο μελών της εκθετικής οικογένειας, αλλά και μεταξύ δύο πολυδιάστατων κανονικών κατανομών.

Πρόταση 2.5. (α) *Η τοπική απόκλιση Kullback-Leibler (2.47) μεταξύ δύο μελών $f_C(x, \theta_1)$ και $f_C(x, \theta_2)$ της εκθετικής οικογένειας (2.26) οδηγούμενη από τον πυρήνα πυκνότητα $f_C(x, \omega)$, μέλος της (2.26), δίνεται από τη σχέση*

$$\begin{aligned} D_0^{\omega}(\theta_1, \theta_2) &= \left(\exp\left(C(\theta_1+\omega) - C(\theta_1) - C(\omega)\right)\right)\{(C(\theta_2) - C(\theta_1))\, E_{\theta_1+\omega}\left(\exp\left(w(X)\right)\right) \\ &+ (\theta_1 - \theta_2)^t E_{\theta_1+\omega}\left(T(X)\exp\left(w(X)\right)\right)\} \\ &- \left(\exp\left(C(\theta_1+\omega) - C(\theta_1) - C(\omega)\right)\right) E_{\theta_1+\omega}\left(\exp\left(w(X)\right)\right) \\ &+ \left(\exp\left(C(\theta_2+\omega) - C(\theta_2) - C(\omega)\right)\right) E_{\theta_2+\omega}\left(\exp\left(w(X)\right)\right), \end{aligned}$$

με $E_{\theta_i+\omega}\left(\exp\left(w(X)\right)\right)$, $i = 1, 2$, όπως ορίστηκε στη σχέση (2.35).
(β) *Η τοπική απόκλιση Kullback-Leibler (2.47) μεταξύ δύο πολυδιάστατων κανονικών κατανομών $f_{N(\mu_1,\Sigma_1)}$ και $f_{N(\mu_2,\Sigma_2)}$ στο σύνολο $\mathcal{X} = R^k$, οδηγούμενη από την πολυδιάστατη κανονική κατανομή $f_{N(\mu,\Sigma)}$, δίνεται από τη σχέση*

$$\begin{aligned} D_0^{(\mu,\Sigma)}((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)) &= \tfrac{1}{2}(2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}|\Sigma_1|^{-\frac{1}{2}}\left|\Sigma^{-1}+\Sigma_1^{-1}\right|^{-\frac{1}{2}} \\ &\times \exp\left(-\tfrac{1}{2}(\mu-\mu_1)^t(\Sigma+\Sigma_1)^{-1}(\mu-\mu_1)\right) \\ &\times \left\{\log\tfrac{|\Sigma_2|}{|\Sigma_1|} - trace\left((\Sigma^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_1^{-1}-\Sigma_2^{-1})\right)\right. \\ &-(\mu^*-\mu_1)^t\Sigma_1^{-1}(\mu^*-\mu_1) \\ &\left.+(\mu^*-\mu_2)^t\Sigma_2^{-1}(\mu^*-\mu_2)\right\} \\ &-E_{(\mu_1,\Sigma_1)}\left(f_{N(\mu,\Sigma)}(X)\right) + E_{(\mu_2,\Sigma_2)}\left(f_{N(\mu,\Sigma)}(X)\right), \end{aligned}$$

όπου

$$\begin{aligned} E_{(\mu_i,\Sigma_i)}\left(f_{N(\mu,\Sigma)}(X)\right) &= (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}|\Sigma_i|^{-\frac{1}{2}}\left|\Sigma^{-1}+\Sigma_i^{-1}\right|^{-\frac{1}{2}} \\ &\times \exp\left\{-\tfrac{1}{2}(\mu-\mu_i)^t(\Sigma+\Sigma_i)^{-1}(\mu-\mu_i)\right\}, \ i = 1, 2, \end{aligned}$$

και

$$\mu^* = (\Sigma^{-1}+\Sigma_1^{-1})^{-1}(\Sigma^{-1}\mu + \Sigma_1^{-1}\mu_1).$$

Απόδειξη. (α) Με βάση τη σχέση (2.47) και λαμβάνοντας υπόψη τη (2.26), απλή άλγεβρα οδηγεί στο επιθυμητό αποτέλεσμα.
(β) Με βάση το τμήμα (α) και τις εξισώσεις (2.27), έχουμε

$$\begin{aligned} D_0^{(\mu,\Sigma)}((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)) &= \exp\{C(\theta_1+\omega)-C(\theta_1)-C(\omega)\} \\ &\times \left(C(\theta_2)-C(\theta_1)+(\theta_1-\theta_2)^t E_{\theta_1+\omega}\left(T(X)\right)\right) \\ &- \exp\{C(\theta_1+\omega)-C(\theta_1)-C(\omega)\} \\ &+ \exp\{C(\theta_2+\omega)-C(\theta_2)-C(\omega)\}, \end{aligned} \tag{2.48}$$

με

$$\theta_i = \left(\Sigma_i^{-1}\mu_i, -\frac{1}{2}\Sigma_i^{-1}\right),\ i=1,2,\ \omega = \left(\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1}\right) \ \text{ και } \ T(X) = \left(X, XX^t\right). \tag{2.49}$$

Απλές αλγεβρικές πράξεις συνάγουν ότι

$$C(\theta_2)-C(\theta_1) = \frac{1}{2}\left(\log\frac{|\Sigma_2|}{|\Sigma_1|} + \mu_2^t\Sigma_2^{-1}\mu_2 - \mu_1^t\Sigma_1^{-1}\mu_1\right). \tag{2.50}$$

Από την άλλη πλευρά, λαμβάνοντας υπόψη ότι

$$E_{\theta_1+\omega}(X) = \int_{\mathcal{X}} x f_C(x,\theta_1+\omega)d\mu(x),$$

η εξίσωση (2.49) δίνει

$$E_{\theta_1+\omega}(X) = \left(\Sigma^{-1}+\Sigma_1^{-1}\right)^{-1}\left(\Sigma^{-1}\mu+\Sigma_1^{-1}\mu_1\right). \tag{2.51}$$

Τότε

$$\begin{aligned} E_{\theta_1+\omega}\left(XX^t\right) &= Var_{\theta_1+\omega}(X) + \left(E_{\theta_1+\omega}(X)\right)\left(E_{\theta_1+\omega}(X)\right)^t \\ &= (\Sigma^{-1}+\Sigma_1^{-1})^{-1} + \left\{(\Sigma^{-1}+\Sigma_1^{-1})^{-1}\left(\Sigma^{-1}\mu+\Sigma_1^{-1}\mu_1\right)\right. \\ &\times \left.\left(\Sigma^{-1}\mu+\Sigma_1^{-1}\mu_1\right)^t(\Sigma^{-1}+\Sigma_1^{-1})^{-1}\right\} \end{aligned} \tag{2.52}$$

και

$$\theta_1-\theta_2 = \left(\Sigma_1^{-1}\mu_1-\Sigma_2^{-1}\mu_2, -\frac{1}{2}\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)\right). \tag{2.53}$$

Με βάση τις (2.49) και (2.53), έχουμε

$$\begin{aligned} (\theta_1-\theta_2)^t E_{\theta_1+\omega}\left(T(X)\right) &= \left(\Sigma_1^{-1}\mu_1-\Sigma_2^{-1}\mu_2\right)^t E_{\theta_1+\omega}\left(X\right) \\ &+ trace\left\{-\tfrac{1}{2}\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)^t E_{\theta_1+\omega}\left(XX^t\right)\right\}. \end{aligned} \tag{2.54}$$

Ως εκ τούτου, λαμβάνοντας υπόψη τις (2.51), (2.52) και (2.54), είναι

$$\begin{aligned}(\theta_1-\theta_2)^t E_{\theta_1+\omega}\left(T(X)\right) &= \left(\Sigma_1^{-1}\mu_1-\Sigma_2^{-1}\mu_2\right)^t\mu^* \\ &-\tfrac{1}{2}\{trace\left(\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)\left(\Sigma^{-1}+\Sigma_1^{-1}\right)^{-1}\right) \\ &+(\mu^*)^t\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)\mu^*\},\end{aligned} \tag{2.55}$$

με $\mu^*=E_{\theta_1+\omega}(X)$. Επιπλέον, λαμβάνοντας υπόψη ότι, για $i=1,2$,

$$\begin{aligned}\exp\left(C(\theta_i+\omega)-C(\theta_i)-C(\omega)\right) = \;& (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}|\Sigma_i|^{-\frac{1}{2}}\left|\Sigma^{-1}+\Sigma_i^{-1}\right|^{-\frac{1}{2}} \\ &\times\exp\left\{-\tfrac{1}{2}(\mu-\mu_i)^t(\Sigma+\Sigma_i)^{-1}(\mu-\mu_i)\right\} \\ &= E_{(\mu_i,\Sigma_i)}\left(f_{N(\mu,\Sigma)}(X)\right),\end{aligned}$$

οι σχέσεις (2.43), (2.48), (2.50) και (2.55) ολοκληρώνουν την απόδειξη του τμήματος (β) της πρότασης. ∎

Παρατήρηση 2.9. (α) *Αναλυτικές εκφράσεις για την τοπική απόκλιση Kullback-Leibler μεταξύ μονοδιάστατων κανονικών κατανομών μπορούν να προκύψουν από την άμεση εφαρμογή του μέρους (β) της παραπάνω πρότασης. Οι αντίστοιχοι τύποι δίνονται από τις σχέσεις*

$$\begin{aligned}D_0^{(\mu,\sigma^2)}((\mu_1,\sigma_1^2),(\mu_2,\sigma_2^2)) &= \tfrac{1}{2}(2\pi(\sigma^2+\sigma_1^2))^{-\frac{1}{2}}\exp\left(-\tfrac{(\mu-\mu_1)^2}{2(\sigma^2+\sigma_1^2)}\right) \\ &\times(\log\tfrac{\sigma_2^2}{\sigma_1^2}-\tfrac{\sigma^2(\sigma_2^2-\sigma_1^2)}{\sigma_2^2(\sigma^2+\sigma_1^2)}-\tfrac{(\mu^*-\mu_1)^2}{\sigma_1^2}+\tfrac{(\mu^*-\mu_2)^2}{\sigma_2^2}) \\ &-E_{(\mu_1,\sigma_1^2)}\left(f_{N(\mu,\sigma^2)}(X)\right)+E_{(\mu_2,\sigma_2^2)}\left(f_{N(\mu,\sigma^2)}(X)\right),\end{aligned}$$

όπου

$$E_{(\mu_i,\sigma_i^2)}\left(f_{N(\mu,\sigma^2)}(X)\right) = \left(2\pi(\sigma^2+\sigma_i^2)\right)^{-1/2}\exp\left\{-\tfrac{(\mu-\mu_i)^2}{2(\sigma^2+\sigma_i^2)}\right\},\ i=1,2,$$

και

$$\mu^*=\frac{\mu\sigma_1^2+\mu_1\sigma^2}{\sigma^2+\sigma_1^2}.$$

(β) Η Πρόταση 2.5 (β) μπορεί να χρησιμοποιηθεί έτσι ώστε να ληφθεί η αναλυτική έκφραση για την τοπική απόκλιση Kullback-Leibler πολυδιάστατων κανονικών κατανομών με κοινό πίνακα διακυμάνσεων-συνδιακύμανσεων $\Sigma_1=\Sigma_2=\Sigma_*$. Αποκτάται εύκολα με απλή εφαρμογή της Πρότασης 2.4 (β), για $\Sigma_1=\Sigma_2=\Sigma_*$.

2.3 Τοπική απόκλιση BHHJ: Ορισμός, σύνολο τιμών και αναλυτική έκφραση για την εκθετική οικογένεια κατανομών

Στην ενότητα αυτή παρακινούμενοι από την καθολική απόκλιση BHHJ, που ορίστηκε από τους Basu et al. (1998) και δόθηκε στη σχέση (1.5), αρχικά ορίζεται η τοπική εκδοχή της. Έπειτα σε πλήρη αντιστοιχία με την προηγούμενη ενότητα μελετάται το σύνολο τιμών της και προσδιορίζονται οι αναλυτικές εκφράσεις για την περίπτωση της εκθετικής οικογένειας κατανομών.

Ορισμός 2.6. *Η ποσότητα*

$$D_a^{\omega}(g, f_\theta) = \int_{\mathcal{X}} h_\omega(x) \left(f_\theta^{1+a}(x) - (1 + \frac{1}{a}) g(x) f_\theta^a(x) + \frac{1}{a} g^{1+a}(x) \right) d\mu(x), \ a > 0. \tag{2.56}$$

ονομάζεται τοπική απόκλιση BHHJ μεταξύ των πυκνοτήτων g και f_θ οδηγούμενη από την πυκνότητα πυρήνα h_ω.

Η επόμενη πρόταση χαρακτηρίζει το κατώτερο όριο της τοπικής απόκλισης BHHJ και αποδεικνύεται εύκολα χρησιμοποιώντας το Θεώρημα 2.1 και πιο συγκεκριμένα τα τμήματα (α) και (β).

Πρόταση 2.6. *Η ποσότητα $D_a^{\omega}(g, f_\theta)$ είναι ένα μέτρο απόκλισης μεταξύ των πυκνοτήτων g και f_θ, υπό την έννοια ότι:*
(α) $D_a^{\omega}(g, f_\theta) \geq 0$, για κάθε g, f_θ, και
(β) $D_a^{\omega}(g, f_\theta) = 0$ αν και μόνο αν $g = f_\theta$ σ.π., ανεξάρτητα του πυρήνα πυκνότητα h_ω.

Απόδειξη. Η τοπική απόκλιση BHHJ μπορεί να δοθεί από τη σχέση

$$D_a^{\omega}(g, f_\theta) = \int_{\mathcal{X}} h_\omega(x) g^{1+a}(x) \phi_a \left(\frac{f_\theta(x)}{g(x)} \right) dx, \ a > 0, \tag{2.57}$$

(βλέπε, σχέση 1.6) με

$$\phi_a(u) = u^{1+a} - (1 + \frac{1}{a}) u^a + \frac{1}{a}, \ \phi_a \in \Phi, a > 0. \tag{2.58}$$

Λαμβάνοντας υπόψη το Θεώρημα 2.1, τμήματα (α) και (β), έχουμε το επιθυμητό αποτέλεσμα. ∎

Στην επόμενη πρόταση προσδιορίζεται η έκφραση της τοπικής απόκλισης BHHJ για δύο μέλη της εκθετικής οικογένειας, $f_C(x, \theta_i)$, $\theta_i \in \Theta \subseteq R^M$, $i = 1, 2$. Επιπλέον, θεωρούμε τον πυρήνα πυκνότητα h_ω ορισμένο στο μετρήσιμο χώρο $\mathcal{X}$ και χωρίς να ανήκει, κατ' ανάγκη, στην κλάση των πυκνοτήτων (2.26).

Πρόταση 2.7. *Έστω δύο μέλη της εκθετικής οικογένειας, $f_C(x, \theta_i)$, $\theta_i \in \Theta \subseteq R^M$, $i = 1, 2$ της οικογένειας (2.26). Επιπρόσθετα, έστω ο πυρήνας πυκνότητα h_ω ορισμένος στο μετρήσιμο χώρο $\mathcal{X}$. Αν $\theta_1 + a\theta_2 \in \Theta$, για $a > 0$, τότε η τοπική απόκλιση BHHJ μεταξύ $f_C(x, \theta_1)$ και $f_C(x, \theta_2)$, οδηγούμενη από την πυκνότητα h_ω, δίνεται από τη σχέση*

$$\begin{aligned} D_\alpha^\omega(\theta_1, \theta_2) &= E_{\theta_2}\left(h_\omega(X) f_C^a(X, \theta_2)\right) \\ &-(1 + a^{-1}) \exp\{M_{C,a}^{(1)}(\theta_1, \theta_2)\} E_{\theta_1 + a\theta_2}(h_\omega(X) \exp\{aw(X)\}) \\ &+a^{-1} E_{\theta_1}\left(h_\omega(X) f_C^a(X, \theta_1)\right), \end{aligned} \tag{2.59}$$

με

$$M_{C,a}^{(1)}(\theta_1, \theta_2) = C(\theta_1 + a\theta_2) - C(\theta_1) - aC(\theta_2) \tag{2.60}$$

και

$$E_{\boldsymbol{\theta}_i}\left(h_\omega(X) f_C^a(X, \theta_i)\right) = \int_{\mathcal{X}} h_\omega(x) f_C^{a+1}(x, \theta_i) d\mu(x), \ i = 1, 2, \tag{2.61}$$

$$E_{\theta_1 + a\theta_2}\left(h_\omega(X) \exp\{aw(X)\}\right) = \int_{\mathcal{X}} h_\omega(x) \exp\{aw(X)\} f_C(x, \theta_1 + a\theta_2) d\mu(x). \tag{2.62}$$

Απόδειξη. Λαμβάνοντας υπόψη τη σχέση (2.56), η τοπική απόκλιση BHHJ ορίζεται από τη σχέση

$$D_\alpha^\omega(\theta_1, \theta_2) = E_{\theta_2}\left(h_\omega(X) f_C^a(X, \theta_2)\right) - (1 + \frac{1}{a}) K_{a,\omega}(\theta_1, \theta_2) + \frac{1}{a} E_{\theta_1}\left(h_\omega(X) f_C^a(x, \theta_1)\right), \tag{2.63}$$

για $a > 0$, με

$$K_{a,\omega}(\theta_1, \theta_2) = \int_{\mathcal{X}} h_\omega(x) f_C(x, \theta_1) f_C^a(x, \theta_2) d\mu(x), \tag{2.64}$$

$$E_{\boldsymbol{\theta}_i}\left(h_\omega(X) f_C^a(X, \theta_i)\right) = \int_{\mathcal{X}} h_\omega(x) f_C^{a+1}(x, \theta_i) d\mu(x) \tag{2.65}$$

και $\theta_i \in \Theta \subseteq R^M$, $i = 1, 2$. Μετά από λίγη άλγεβρα έχουμε το επιθυμητό αποτέλεσμα.

∎

Η πρόταση που ακολουθεί δίνει την αναλυτική έκφραση για την $D_a^{\omega}(\theta_1,\theta_2)$, όταν ο πυρήνας πυκνότητα h_ω ανήκει στην κλάση (2.26). Η απόδειξη παραλείπεται καθώς είναι παρόμοια με την απόδειξη της Πρότασης 2.3.

Πρόταση 2.8. *Έστω δύο μέλη $f_C(x,\theta_1)$ και $f_C(x,\theta_2)$ της οικογένειας (2.26) και έστω ο πυρήνας πυκνότητα $h_\omega(x) = f_C(x,\omega)$ που ανήκει στην (2.26). Τότε, κάτω από τις υποθέσεις $a\theta_i + \omega \in \Theta$, $i = 1,2$, και $\theta_1 + a\theta_2 + \omega \in \Theta$, για $a > 0$, η τοπική απόκλιση BHHJ μεταξύ των πυκνοτήτων $f_C(x,\theta_1)$ και $f_C(x,\theta_2)$, οδηγούμενη από την πυκνότητα $h_\omega(x)$, δίνεται από τη σχέση*

$$\begin{aligned} D_a^{\omega}(\theta_1,\theta_2) &= \exp\{C(a\theta_2+\omega) - aC(\theta_2) - C(\omega)\}E_{a\theta_2+\omega}\left(\exp\{aw(X)\}\right) \\ &\quad -(1+a^{-1})\exp\{M_{C,a}^{(2)}(\theta_1,\theta_2,\omega)\}E_{\theta_1+a\theta_2+\omega}(\exp\{(a+1)w(X)\}) \\ &\quad +a^{-1}\exp\{C(a\theta_1+\omega) - aC(\theta_1) - C(\omega)\}E_{a\theta_1+\omega}\left(\exp\{aw(X)\}\right), \end{aligned} \tag{2.66}$$

με

$$M_{C,a}^{(2)}(\theta_1,\theta_2,\omega) = C(\theta_1 + a\theta_2 + \omega) - C(\theta_1) - aC(\theta_2) - C(\omega) \tag{2.67}$$

και

$$E_{a\theta_i+\omega}(\exp\{aw(X)\}) = \int_{\mathcal{X}} \exp\{aw(X)\} f_C(x, a\theta_i+\omega)d\mu(x),\ i = 1,2, \tag{2.68}$$

$$E_{\theta_1+a\theta_2+\omega}(\exp\{(a+1)w(X)\}) = \int_{\mathcal{X}} \exp\{(a+1)w(X) f_C(x,\theta_1+a\theta_2+\omega)d\mu(x). \tag{2.69}$$

Το ενδιαφέρον τώρα επικεντρώνεται στον προσδιορισμό της αναλυτικής μορφής της τοπικής απόκλισης BHHJ, μεταξύ δυο k-διάστατων κανονικών κατανομών, $N_k(\mu_1,\Sigma_1)$ και $N_k(\mu_2,\Sigma_2)$. Έστω ότι ο πυρήνας πυκνότητα h_ω είναι η πολυδιάστατη κανονική κατανομή $N_k(\mu,\Sigma)$, με μέσο διάνυσμα $\mu \in R^k$ και πίνακα διακυμάνσεων-συνδιακυμάνσεων Σ. Το αποτέλεσμα προκύπτει ως ειδική περίπτωση της προηγούμενης πρότασης.

Πρόταση 2.9. *Η τοπική απόκλιση BHHJ $D_a^{(\mu,\Sigma)} = D_a^{(\mu,\Sigma)}((\mu_1,\Sigma_1),(\mu_2,\Sigma_2))$, μεταξύ δυο k-διάστατων κανονικών κατανομών $N_k(\mu_1,\Sigma_1)$ και $N_k(\mu_2,\Sigma_2)$, οδηγούμενη από μια k-διάστατη κανονική κατανομή $N_k(\mu,\Sigma)$, δίνεται από*

$$\begin{aligned} D_a^{(\mu,\Sigma)} = \ & (2\pi)^{-ak/2}|\Sigma^{-1} + a\Sigma_2^{-1}|^{-1/2}|\Sigma|^{-1/2}|\Sigma_2|^{-a/2} \\ & \times \exp\{-\tfrac{1}{2}(\mu-\mu_2)^t(\Sigma + a\Sigma_2)^{-1}(\mu-\mu_2)\} \\ & -(1+a^{-1})(2\pi)^{-(a+1)k/2}|\Sigma|^{-\frac{1}{2}}|\Sigma_1|^{-\frac{1}{2}}|\Sigma_2|^{-\frac{a}{2}}\left|\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}\right|^{-\frac{1}{2}} \\ & \times \exp\{-\tfrac{1}{2}\left(\mu^t\Sigma^{-1}\mu + \mu_1^t\Sigma_1^{-1}\mu_1 + a\mu_2^t\Sigma_2^{-1}\mu_2 - B_1^tB_2B_1\right)\} \\ & +a^{-1}(2\pi)^{-ak/2}|\Sigma^{-1} + a\Sigma_1^{-1}|^{-1/2}|\Sigma|^{-1/2}|\Sigma_1|^{-a/2} \\ & \times \exp\{-\tfrac{1}{2}(\mu-\mu_1)^t(\Sigma + a\Sigma_1)^{-1}(\mu-\mu_1)\}, \end{aligned}$$

με

$$
\begin{aligned}
B_1 &= \Sigma_1^{-1}\mu_1 + a\Sigma_2^{-1}\mu_2 + \Sigma^{-1}\mu,\\
B_2 &= \left(\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}\right)^{-1},
\end{aligned}
$$

για $a > 0$.

Απόδειξη. Με βάση την Πρόταση 2.8 και λαμβάνοντας υπόψη ότι $w(X) = 0$, έχουμε

$$
\begin{aligned}
D_a^{(\mu,\Sigma)}((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)) &= \exp\{C(a\theta_2+\omega) - aC(\theta_2) - C(\omega)\}\\
&\quad -(1+a^{-1})\exp\{M_{C,a}^{(2)}(\theta_1,\theta_2,\omega)\}\\
&\quad +a^{-1}\exp\{C(a\theta_1+\omega) - aC(\theta_1) - C(\omega)\}
\end{aligned} \tag{2.70}
$$

με $\theta_1 = (\theta_{11},\theta_{12}) = (\Sigma_1^{-1}\mu_1, -\frac{1}{2}\Sigma_1^{-1})$, $\theta_2 = (\theta_{21},\theta_{22}) = (\Sigma_2^{-1}\mu_2, -\frac{1}{2}\Sigma_2^{-1})$, $\omega = (\omega_1,\omega_2) = (\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1})$ και

$$
M_{C,a}^{(2)}(\theta_1,\theta_2,\omega) = C(\theta_1 + a\theta_2 + \omega) - C(\theta_1) - aC(\theta_2) - C(\omega). \tag{2.71}
$$

Σύμφωνα με τις σχέσεις (2.27),

$$
\begin{aligned}
C(\theta_i) &= \log\left((2\pi)^{k/2}|\Sigma_i|^{1/2}\right) + \tfrac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i,\ i=1,2\\
C(\omega) &= \log\left((2\pi)^{k/2}|\Sigma|^{1/2}\right) + \tfrac{1}{2}\mu^t\Sigma^{-1}\mu.
\end{aligned} \tag{2.72}
$$

Από την άλλη πλευρά, είναι

$$
a\theta_i + \omega = \left(a\Sigma_i^{-1}\mu_i + \Sigma^{-1}\mu, -\frac{1}{2}(a\Sigma_i^{-1} + \Sigma^{-1})\right),\ i = 1,2.
$$

Άμεσα προκύπτει, μέσω της (2.27), ότι

$$
\begin{aligned}
C(a\theta_i+\omega) &= \log\left((2\pi)^{k/2}|\Sigma^{-1} + a\Sigma_i^{-1}|^{-1/2}\right)\\
&\quad +\tfrac{1}{2}\left(\Sigma^{-1}\mu + a\Sigma_i^{-1}\mu_i\right)^t\left(\Sigma^{-1} + a\Sigma_i^{-1}\right)^{-1}\left(\Sigma^{-1}\mu + a\Sigma_i^{-1}\mu_i\right),\ i=1,2.
\end{aligned} \tag{2.73}
$$

Λαμβάνοντας υπόψη την ισότητα (βλέπε Pardo, 2006, σελ. 49)

$$
\begin{aligned}
&\left(\Sigma^{-1}\mu + a\Sigma_i^{-1}\mu_i\right)^t\left(\Sigma^{-1} + a\Sigma_i^{-1}\right)^{-1}\left(\Sigma^{-1}\mu + a\Sigma_i^{-1}\mu_i\right) - \mu^t\Sigma^{-1}\mu - \mu_i^t a\Sigma_i^{-1}\mu_i\\
&\qquad = -(\mu-\mu_i)^t(\Sigma + a\Sigma_i)^{-1}(\mu-\mu_i),\ i=1,2,
\end{aligned}
$$

μετά από αλγεβρικές πράξεις, έχουμε

$$
\begin{aligned}
C(a\theta_i+\omega) - aC(\theta_i) - C(\omega) &= \log\left((2\pi)^{-ak/2}|\Sigma^{-1} + a\Sigma_i^{-1}|^{-1/2}|\Sigma|^{-1/2}|\Sigma_i|^{-a/2}\right)\\
&\quad -\tfrac{1}{2}(\mu-\mu_i)^t(\Sigma + a\Sigma_i)^{-1}(\mu-\mu_i),\ i=1,2.
\end{aligned} \tag{2.74}
$$

Απομένει να προσδιορίσουμε την ποσότητα $M_{C,a}^{(2)}(\theta_1,\theta_2,\omega)$, η οποία δίνεται στην (2.71). Είναι,

$$\begin{aligned}\theta_1 + a\theta_2 + \omega &= \big(\Sigma_1^{-1}\mu_1 + a\Sigma_2^{-1}\mu_2 + \Sigma^{-1}\mu, \\ &\quad (-1/2)\left(\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}\right)\big),\end{aligned}$$

συνεπώς

$$\begin{aligned}C(\theta_1 + a\theta_2 + \omega) &= \log\left((2\pi)^{k/2}|\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}|^{-1/2}\right) \\ &+ \tfrac{1}{2}\left(\Sigma_1^{-1}\mu_1 + a\Sigma_2^{-1}\mu_2 + \Sigma^{-1}\mu\right)^t \left(\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}\right)^{-1} \\ &\times \left(\Sigma_1^{-1}\mu_1 + a\Sigma_2^{-1}\mu_2 + \Sigma^{-1}\mu\right),\end{aligned} \tag{2.75}$$

$a > 0$. Σύμφωνα με τις σχέσεις (2.71), (2.72) και (2.75),

$$\begin{aligned}M_{C,a}^{(2)}(\theta_1,\theta_2,\omega) &= \log((2\pi)^{-(a+1)k/2}|\Sigma|^{-\frac{1}{2}}|\Sigma_1|^{-\frac{1}{2}}|\Sigma_2|^{-\frac{a}{2}}\left|\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}\right|^{-\frac{1}{2}}) \\ &-\tfrac{1}{2}\left(\mu^t\Sigma^{-1}\mu + \mu_1^t\Sigma_1^{-1}\mu_1 + a\mu_2^t\Sigma_2^{-1}\mu_2 - B_1^tB_2B_1\right),\end{aligned} \tag{2.76}$$

με

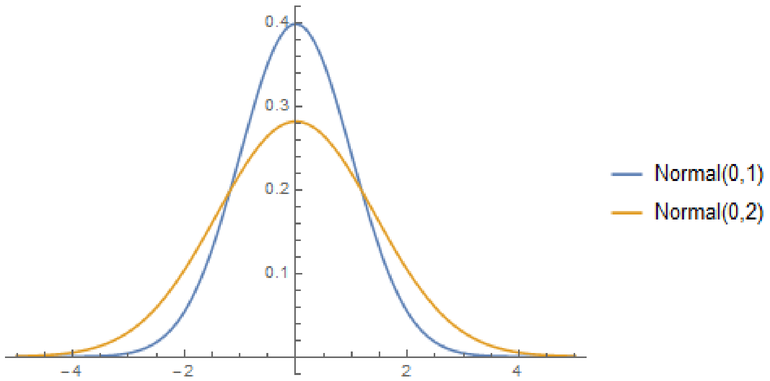
$$\begin{aligned}B_1 &= \Sigma_1^{-1}\mu_1 + a\Sigma_2^{-1}\mu_2 + \Sigma^{-1}\mu, \\ B_2 &= \left(\Sigma_1^{-1} + a\Sigma_2^{-1} + \Sigma^{-1}\right)^{-1}.\end{aligned}$$

Το αποτέλεσμα έπεται από τις (2.70), (2.74) και (2.76). ∎

2.4 Αριθμητικά παραδείγματα-Εφαρμογές

Η τοπική ϕ-απόκλιση που ορίστηκε στη σχέση (2.3) ή στη σχέση (2.4), είναι αρκετά όμοια με τη ϕ-απόκλιση που ορίζεται στη (1.2) ή στη (1.3), αντίστοιχα. Η μοναδική διαφορά έγκειται στη συνάρτηση κατανομής r ή h_ω που εισέρχεται στην έκφραση της κλασικής ϕ-απόκλισης Csiszár και πιο συγκεκριμένα στην επιπρόσθετη παράμετρο ω. Ο ρόλος της παραμέτρου ω είναι καθοριστικός στον ορισμό της τοπικής ϕ-απόκλισης, και αυτός ακριβώς ο ρόλος θα εξεταστεί στα παρακάτω παραδείγματα. Στο πρώτο παράδειγμα θα χρησιμοποιηθούν κανονικές κατανομές προκειμένου να διερευνηθεί ότι ο Ορισμός 2.2 στην ουσία ποσοτικοποιεί την απόκλιση μεταξύ δύο κανονικών μοντέλων σε ένα υποσύνολο του πεδίου ορισμού τους. Σε αυτό το παράδειγμα, αποσαφηνίζεται ο ρόλος της παραμέτρου ω στον ορισμό της D_ϕ^ω.

Παράδειγμα 2.1 (Κανονικές κατανομές). *Έστω P_θ, $\theta \in \Theta = \left\{(\mu,\sigma^2) : \mu,\sigma^2 \in R,\ \sigma^2 > 0\right\}$ είναι μια μονοδιάστατη οικογένεια κανονικών κατανομών. Για τρεις τιμές της*



Σχήμα 2.1: Γραφική παράσταση κανονικών κατανομών με παραμέτρους $\theta_1 = (0,1)$ και $\theta_2 = (0,2)$.

παραμέτρου θ, $\theta_1 = (\mu_1, \sigma_1^2)$, $\theta_2 = (\mu_2, \sigma_2^2)$ *και* $\omega = (\mu, \sigma^2)$, *συμβολίζουμε με* f_{θ_1}, f_{θ_2} *και* f_ω *τις αντίστοιχες μονοδιάστατες κανονικές πυκνότητες. Θεωρούμε την τοπική απόκλιση Cressie and Read, όπως αυτή λαμβάνεται από τη σχέση (2.4) για* $\phi(u) = \phi_\lambda(u) = \frac{u^{\lambda+1}-u-\lambda(u-1)}{\lambda(\lambda+1)}$, $\lambda \neq 0, -1$. *Η αναλυτική (κλειστή) μορφή της τοπικής απόκλισης* $D^\omega_{\phi_\lambda}(\theta_1, \theta_2)$ *μεταξύ* f_{θ_1} *και* f_{θ_2} *οδηγούμενη από την πυκνότητα* f_ω*, δόθηκε στην Παρατήρηση 2.7, στη σχέση (2.46).*

Στον Πίνακα 2.1, χρησιμοποιώντας τη σχέση (2.46), παρουσιάζονται οι τοπικές αποκλίσεις $D^\omega_{\phi_{2/3}}(\theta_1, \theta_2)$, $D^\omega_{\phi_{2/3}}(\theta_2, \theta_1)$ *καθώς και η συμμετρική εκδοχή τους* $D^\omega_{\phi_{2/3}}(\theta_1, \theta_2) + D^\omega_{\phi_{2/3}}(\theta_2, \theta_1)$, *για* $\theta_1 = (0,1)$, $\theta_2 = (0,2)$ *και διάφορες τιμές της παραμέτρου* $\omega = (\mu, \sigma^2)$, *Επικεντρωνόμαστε στην τιμή* $\lambda = 2/3$, *διότι η επιλογή αυτή για την παράμετρο* λ *θεωρείται ιδανική σε πολλές στατιστικές εφαρμογές της κλασικής απόκλισης Cressie and Read. Επιπλέον, ο Πίνακας 2.1 περιλαμβάνει την κλασική απόκλιση Cressie and Read* $D_{\phi_{2/3}}(\theta_1, \theta_2) = \int_{\mathcal{X}} f_{\theta_2}(x)\phi_{2/3}\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) d\mu(x)$ *και τιμές του ολοκληρώματος*

$$I_{i,j} = \int_{\mathcal{X}} I_A(x) f_{\theta_j}(x) \phi_{2/3}\left(\frac{f_{\theta_i}(x)}{f_{\theta_j}(x)}\right) d\mu(x), \; i,j = 1,2, \; i \neq j,$$

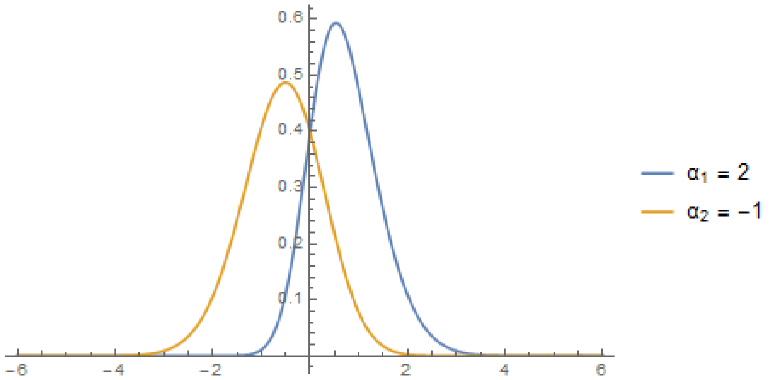
το οποίο στην ουσία είναι η κλασική ϕ*-απόκλιση Csiszár, περιορισμένη στο σύνολο* $A \subseteq \mathcal{X}$. *Σημειώνεται και πάλι ότι το* $I_{i,j}$ *δεν ικανοποιεί την ιδιότητα της μη αρνητικότητας και της ταυτότητας των δυσδιάκριτων κατανομών σύμφωνα με το Λήμμα 1.1 του Csiszár (1967). Σύμφωνα με τον Πίνακα 2.1, η κλασική απόκλιση Cressie and Read μεταξύ δύο μονοδι-*

$\omega=(\mu,\sigma^2)$	$D^{\omega}_{\phi_{2/3}}(\theta_1,\theta_2), D^{\omega}_{\phi_{2/3}}(\theta_2,\theta_1)$, $D^{\omega}_{\phi_{2/3}}(\theta_1,\theta_2)+D^{\omega}_{\phi_{2/3}}(\theta_2,\theta_1)$ $\theta_1=(0,1),\ \theta_2=(0,2)$	$A\subseteq X$ $A=(\alpha,\beta)$	$I_{1,2}, I_{2,1}, I_{1,2}+I_{2,1}$
$(0,0.1)$	$0.0195, 0.0151, 0.0346$	$(-0.5,0.5)$	$0.0199, 0.0154, 0.0353$
$(1,0.1)$	$0.0040, 0.0035, 0.0075$	$(0.5,1.5)$	$0.0037, 0.0032, 0.0069$
$(2,0.1)$	$0.0115, 0.0217, 0.0332$	$(1.5,2.5)$	$0.0116, 0.0216, 0.0332$
$(3,0.1)$	$0.0120, 0.0509, 0.0629$	$(2.5,3.5)$	$0.0121, 0.0513, 0.0634$
$(4,0.1)$	$0.0033, 0.0448, 0.0481$	$(3.5,4.5)$	$0.0032, 0.0450, 0.0482$
$(5,0.1)$	$0.0004, 0.0247, 0.0251$	$(4.5,5.5)$	$0.0004, 0.0246, 0.0250$
$(6,0.1)$	$0.0000, 0.0104, 0.0104$	$(5.5,6.5)$	$0.0000, 0.0103, 0.0103$
$D_{\phi_{2/3}}(\theta_1,\theta_2)=0.082, D_{\phi_{2/3}}(\theta_2,\theta_1)=0.3373$		$(-3\sigma_2,3\sigma_2)$	$0.0804, 0.2383$

Πίνακας 2.1: Τιμές των $D_{\phi_{2/3}}(\theta_1,\theta_2)$, $D_{\phi_{2/3}}(\theta_2,\theta_1)$, $D^{\omega}_{\phi_{2/3}}(\theta_1,\theta_2)$, $D^{\omega}_{\phi_{2/3}}(\theta_2,\theta_1)$, $D^{\omega}_{\phi_{2/3}}(\theta_1,\theta_2)+D^{\omega}_{\phi_{2/3}}(\theta_2,\theta_1)$, $I_{1,2}$, $I_{2,1}$ και $I_{1,2}+I_{2,1}$ για κανονικές κατανομές με παραμέτρους $\theta_1=(\mu_1,\sigma_1^2)=(0,1)$, $\theta_2=(\mu_2,\sigma_2^2)=(0,2)$ για διάφορες τιμές της παραμέτρου $\omega=(\mu,\sigma^2)$.

άστατων κανονικών μοντέλων $N(0,1)$ και $N(0,2)$, είναι ίση με $D_{\phi_{2/3}}(\theta_1,\theta_2)=0.082$. Η τιμή της μειώνεται σημαντικά εάν εστιάσουμε σε συγκεκριμένα υποσύνολα $A=(\alpha,\beta)$ του πεδίου ορισμού $\mathcal{X}$, όπως προκύπτει (μετριέται) από το ολοκλήρωμα $I_{i,j}$. Έτσι, εξηγείται ο ρόλος της πυκνότητας h_ω, στη (2.4), και πιο συγκεκριμένα ο ρόλος της παραμέτρου ω, η οποία ρυθμίζει το υποσύνολο του $\mathcal{X}$ στο οποίο θα υπολογισθεί η απόκλιση μεταξύ των κανονικών μοντέλων $N(0,1)$ και $N(0,2)$. Παρατηρούμε ότι όταν εστιάζουμε στις ουρές (εκτός του διαστήματος $[-5,5]$) των κατανομών, όπως φαίνεται και στο Σχήμα 2.1, τόσο πιο όμοιες γίνονται οι δύο πυκνότητες.

Παράδειγμα 2.2 (Λοξές κανονικές κατανομές). *Ας θεωρήσουμε την τυπική λοξή κανονική κατανομή με παράμετρο λοξότητας α και πυκνότητα $2\phi(x)\Phi(\alpha x)$, όπου ϕ και Φ είναι οι σ.π.π. και η α.σ.κ. της τυπικής κανονικής κατανομής. Ο Πίνακας 2.2 παρουσιάζει τις τιμές των Cressie and Read αποκλίσεων $D^{\omega}_{\phi_{2/3}}(\alpha_1,\alpha_2)$, $D^{\omega}_{\phi_{2/3}}(a_2,a_1)$ και τη συμμετρική εκδοχή τους $D^{\omega}_{\phi_{2/3}}(a_1,a_2)+D^{\omega}_{\phi_{2/3}}(a_2,a_1)$, μεταξύ δύο λοξών κανονικών μοντέλων με παραμέτρους $\alpha_1=2$ και $\alpha_2=-1$. Η πυκνότητα h_ω της τοπικής απόκλισης Cressie and Read είναι εκείνη της μονοδιάστατης κανονικής κατανομής με παραμέτρο $\omega=(\mu,\sigma^2)$.*



Σχήμα 2.2: Γραφική παράσταση τυπικών λοξών κανονικών κατανομών με παραμέτρους λοξότητας $\alpha_1 = 2$ και $\alpha_2 = -1$.

$\omega = (\mu, \sigma^2)$	$D^{\omega}_{\phi_{2/3}}(\alpha_1,\alpha_2), D^{\omega}_{\phi_{2/3}}(\alpha_2,\alpha_1)$ $D^{\omega}_{\phi_{2/3}}(\alpha_1,\alpha_2) + D^{\omega}_{\phi_{2/3}}(\alpha_2,\alpha_1)$ $\alpha_1 = 2,\ \alpha_2 = -1$	$A \subseteq X$ $A = (\alpha, \beta)$	$I_{1,2}, I_{2,1}, I_{1,2} + I_{2,1}$
$(0, 0.1)$	$0.0878, 0.1314, 0.2192$	$(-0.5, 0.5)$	$0.0819, 0.1015, 0.1834$
$(1, 0.1)$	$0.7249, 0.1665, 0.8914$	$(0.5, 1.5)$	$0.7284, 0.1693, 0.8977$
$(2, 0.1)$	$1.0133, 0.0671, 1.0804$	$(1.5, 2.5)$	$1.0191, 0.0666, 1.0857$
$(3, 0.1)$	$0.6426, 0.0075, 0.6501$	$(2.5, 3.5)$	$0.6425, 0.0071, 0.6496$
$(4, 0.1)$	$0.2521, 0.0003, 0.2524$	$(3.5, 4.5)$	$0.2502, 0.0003, 0.2505$
$(5, 0.1)$	$0.0678, 5.29 \times 10^{-6}, 0.0678$	$(4.5, 5.5)$	$0.0668, 4.05 \times 10^{-6}, 0.0668$
$(6, 0.1)$	$0.0129, 3.57 \times 10^{-8}, 0.0129$	$(5.5, 6.5)$	$0.0126, 2.27 \times 10^{-8}, 0.0126$
$D_{\phi_{2/3}}(\alpha_1,\alpha_2) = 3.0839$ $D_{\phi_{2/3}}(\alpha_2,\alpha_1) = 2.32617 \times 10^6$		$(-4, 4)$	$3.0786,\ 503072.48$

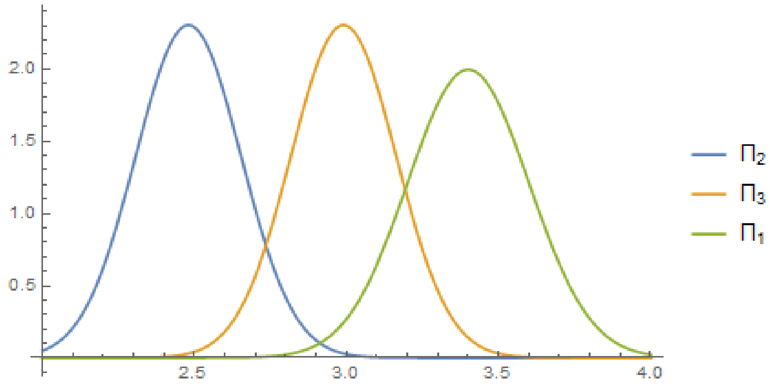
Πίνακας 2.2: Τιμές των $D_{\phi_{2/3}}(\alpha_1,\alpha_2)$, $D_{\phi_{2/3}}(\alpha_2,\alpha_1)$, $D^{\omega}_{\phi_{2/3}}(\alpha_1,\alpha_2)$, $D^{\omega}_{\phi_{2/3}}(\alpha_2,\alpha_1)$, $D^{\omega}_{\phi_{2/3}}(\alpha_1,\alpha_2) + D^{\omega}_{\phi_{2/3}}(\alpha_2,\alpha_1)$, $I_{1,2}$, $I_{2,1}$, και $I_{1,2} + I_{2,1}$ για τυπικές λοξές κανονικές κατανομές με παραμέτρους $\alpha_1 = 2$ και $\alpha_2 = -1$ για διάφορες τιμές της παραμέτρου $\omega = (\mu, \sigma^2)$.

Ο Πίνακας 2.2 οδηγεί σε αρκετά παρόμοια συμπεράσματα με αυτά του προηγούμενου παραδείγματος. Απεικονίζει ότι η απόκλιση μεταξύ δύο κατανομών πιθανότητας σε ολόκληρο το πεδίο ορισμού $\mathcal{X}$ διαφέρει σημαντικά από την τοπική απόκλιση σε κάποια υποσύνολα του $\mathcal{X}$ που καθορίζει ο πυρήνας h_ω. Το Σχήμα 2.2 μας βοηθά να οπτικοποιήσουμε τις δύο λοξές κανονικές κατανομές και το πώς οι τιμές του Πίνακα 2.2 υποδεικνύουν τη διαφορά των δύο κατανομών στα διάφορα υποδιαστήματα του πεδίου ορισμού τους. Παρατηρούμε ότι το κλασικό καθολικό μέτρο προτείνει απόκλιση, ενώ τοπικά, και πιο συγκεκριμένα κοντά στις ουρές οι κατανομές είναι παρόμοιες.

Το κεφάλαιο αυτό ολοκληρώνεται με μια εφαρμογή στην οποία μελετάται η συμπεριφορά των μέτρων τοπικής απόκλισης σε ένα παράδειγμα πραγματικών δεδομένων της βιβλιογραφίας.

Εφαρμογή 2.1. *Στην εφαρμογή αυτή, το ενδιαφέρον εστιάζεται στους μέσους όρους βαθμολογίας (GPA) μαθητών που ενδιαφέρονται για εισαγωγή σε σχολή διοίκησης επιχειρήσεων (βλέπε, Johnson and Wichern, 1992, σελ. 532, Παράδειγμα 11.11). Υπάρχουν τρεις ομάδες αιτούντων, οι οποίες έχουν κατηγοριοποιηθεί ως εξής:* Π_1*: εισαχθέντες,* Π_2*: μη εισαχθέντες και* Π_3*: οριακά εισαχθέντες με βάση τις βαθμολογίες τους. Ενδιαφερόμαστε να μελετήσουμε τυχόν διαφορές μεταξύ των τριών πληθυσμών των μαθητών, είτε γενικά, δηλαδή πάνω σε ολόκληρο το πεδίο ορισμού των κατανομών που περιγράφουν κάθε πληθυσμό, είτε τοπικά, εστιάζοντας σε μια συγκεκριμένη περιοχή του πεδίου ορισμού, όπου δύο πληθυσμοί μπορεί να διαφέρουν. Θεωρούμε ότι το κέντρο του πυρήνα πυκνότητα είναι ένας κυρτός συνδυασμός των μέσων όρων των δύο υπό εξέταση πληθυσμών. Με αυτόν τον τρόπο, ο πυρήνας λειτουργεί ως ένα «παράθυρο» που μπορεί να μετακινηθεί σε όλο το πεδίο ορισμού και να επικεντρωθεί σε μια μικρή περιοχή κάθε φορά, που εξαρτάται από τη μεταβλητότητα του πυρήνα πυκνότητα.*

Στον Πίνακα 2.3 παρουσιάζονται οι έλεγχοι κανονικότητας μαζί με τα βασικά περιγραφικά μέτρα για τους τρεις πληθυσμούς, τις δειγματικές μέσες τιμές και διακυμάνσεις. Σημειώνεται ότι η υπόθεση της κανονικότητας μπορεί να υιοθετηθεί για τις τρεις ομάδες μαθητών, και γι' αυτό μπορούμε να προσαρμόσουμε τρεις μονοδιάστατες κανονικές κατανομές προκειμένου να περιγράψουμε τα δεδομένα. Με βάση την κανονικότητα και θεωρώντας τις δειγματικές μέσες τιμές και διακυμάνσεις καθορίζονται πλήρως οι κατανομές των τριών πληθυσμών. Το Σχήμα 2.3 απεικονίζει τις τρεις πυκνότητες για τους πληθυσμούς Π_1, Π_2 *και* Π_3 *χρησιμοποιώντας τις εκτιμώμενες μέσες τιμές και διακυμάνσεις από τον Πίνακα 2.3.*



Σχήμα 2.3: Γραφική παράσταση των τριών πυκνοτήτων που περιγράφουν τους πληθυσμούς Π_1, Π_2 και Π_3, βάσει των δειγματικών μέσων τιμών και διακυμάνσεων.

Χρησιμοποιείται η απόκλιση Cressie-Read προκειμένου να συγκρίθουν οι πληθυσμοί Π_1 *με* Π_2, Π_1 *με* Π_3, *και* Π_2 *με* Π_3. *Τα αποτελέσματα των συγκρίσεων δίνονται στους Πίνακες 2.4, 2.5 και 2.6, αντίστοιχα. Σε όλους τους πίνακες, παρουσιάζεται η τοπική Cressie-Read απόκλιση* $D^{\omega}_{\phi_\lambda}(\widehat{\theta}_1, \widehat{\theta}_2)$ *για διάφορες τιμές της παραμέτρου* λ, *πιο συγκεκριμένα, για* $\lambda = -2, -0.5, \frac{2}{3}, 1$ *και* 2. *Η τελευταία γραμμή κάθε πίνακα δείχνει τις τιμές του καθολικού μέτρου* $D_{\phi_\lambda}(\widehat{\theta}_1, \widehat{\theta}_2)$. *Ο πυρήνας και τα πληθυσμιακά μοντέλα είναι μονοδιάστατες κανονικές κατανομές, με τις εκτιμώμενες παραμέτρους των* Π_1, Π_2 *και* Π_3 *να είναι* $\widehat{\theta}_1 = (\widehat{\mu}_1, \widehat{\sigma}^2_1) = (3.40, 0.04)$, $\widehat{\theta}_2 = (\widehat{\mu}_2, \widehat{\sigma}^2_2) = (2.48, 0.03)$, *και* $\widehat{\theta}_3 = (\widehat{\mu}_3, \widehat{\sigma}^2_3) = (2.99, 0.03)$, *αντίστοιχα. Χρησιμοποιούνται κυρτοί συνδυασμοί των εκτιμώμενων μέσων τιμών για διαφορετικές τιμές του* k, *και το αποτέλεσμα θεωρείται ως μέση τιμή του πυρήνα. Οι παράμετροι του πυρήνα βρίσκονται στη δεύτερη στήλη των Πινάκων 2.4, 2.5 και 2.6.*

Η διακύμανση του πυρήνα είναι 0.1 *σε όλες τις περιπτώσεις, μια μικρή τιμή που δίνει περισσότερο βάρος στις τιμές της μέσης τιμής* μ, *αναδεικνύοντας έτσι τις διαφορές των δύο πληθυσμών σε μία περιοχή κοντά στη μέση τιμή του πυρήνα. Οι τιμές του* k *που μελετώνται είναι* $k = 0, 0.1, 0.3, 0.5, 0.7, 0.9$ *και* 1, *και οδηγούν το «παράθυρο» πάνω στο πεδίο ορισμού και κινείται από τον ένα μέσο όρο του πληθυσμού προς τον άλλο. Όταν δύο πληθυσμοί βρίσκονται κοντά ο ένας με τον άλλον σε ένα συγκεκριμένο «παράθυρο», προσδοκούμε το τοπικό μέτρο να λάβει μικρότερες τιμές, εκτός εάν οι δύο πληθυσμοί είναι τελείως διαφορετικοί μεταξύ τους. Αυτός ο ισχυρισμός υποστηρίζεται από τα αποτελέσματα των Πινάκων 2.4, 2.5 και 2.6, με όλες τις τιμές του μέτρου «μακριά» από το μηδέν, υποδεικνύοντας ότι* *όλοι οι πληθυσμοί είναι διαφορετικοί μεταξύ τους. Για παράδειγμα, όταν συγκρίνουμε τον πληθυσμό* Π_2 *με τον* Π_3, *με* $\lambda = \frac{2}{3}$, *το καθολικό μέτρο είναι* $D_{\phi_{2/3}}(\widehat{\theta}_2, \widehat{\theta}_3) = 110.3$, *ενώ η τιμή* $k = 0$, *δίδει στο τοπικό μέτρο τιμή* $D^{\omega}_{\phi_{2/3}}(\widehat{\theta}_2, \widehat{\theta}_3) = 7.7$. *Εστιάζουμε στην περιοχή που περιγράφεται από τον πυρήνα με μέση τιμή ίδια με τη μέση τιμή του* Π_3, *αλλά η διακύμανση του πυρήνα* $\sigma^2 = 0.1$ *είναι πολύ μεγαλύτερη από αυτή του πληθυσμού* Π_3 $(\widehat{\sigma}_3^2 = 0.03)$.

Στον Πίνακα 2.7, διερευνάται η συμπεριφορά του τοπικού μέτρου Kullback-Leibler και τα αποτελέσματα είναι παρόμοια με εκείνα της τυπικής απόκλισης Cressie and Read. Όλες οι τιμές δείχνουν ότι οι τρεις πληθυσμοί είναι σαφώς διαφορετικοί σύμφωνα με το καθολικό μέτρο, όμως για ορισμένες τιμές της παραμέτρου λ *οι πληθυσμοί* Π_2 *και* Π_3 *δεν είναι τόσο διαφορετικοί, τοπικά, όταν ο πυρήνας εστιάζει στη μέση τιμή του πληθυσμού* Π_3.

Πληθυσμός	Μέγεθος δείγματος	Τεστ κανονικότητας (p-τιμές)		Δειγματικοί μέσοι
	n	$Shapiro - Wilk$	$Kolmogorov - Smirnov$	και διακυμάνσεις
Π_1	31	0.86	0.2 (Κάτω φράγμα)	$(\overline{X}_1 = 3.40, S_1^2 = 0.04)$
Π_2	28	0.85	0.2 (Κάτω φράγμα)	$(\overline{X}_2 = 2.48, S_2^2 = 0.03)$
Π_3	26	0.11	0.2 (Κάτω φράγμα)	$(\overline{X}_3 = 2.99, S_3^2 = 0.03)$

Πίνακας 2.3: Έλεγχοι κανονικότητας και δειγματικοί μέσοι και διακυμάνσεις για την μεταβλητή GPA στους τρεις πληθυσμούς.

k	Π_1-Π_2					
	$\omega=(\mu,\sigma^2)$	$\lambda=-2$	$\lambda=-0.5$	$\lambda=\frac{2}{3}$	$\lambda=1$	$\lambda=2$
0	$(2.48,0.1)$	3.84881×10^{6}	2.20	1435.39	5.81244×10^{7}	7.44435×10^{41}
0.1	$(2.57,0.1)$	2.49542×10^{6}	2.18	4509.07	2.67675×10^{8}	2.16225×10^{43}
0.3	$(2.76,0.1)$	806689.19	1.93	42412.55	5.69638×10^{9}	2.35027×10^{46}
0.5	$(2.94,0.1)$	211568.98	1.77	284582.94	8.38124×10^{10}	1.51982×10^{49}
0.7	$(3.13,0.1)$	38797.18	1.91	1.68285×10^{6}	1.14943×10^{12}	1.20013×10^{52}
0.9	$(3.31,0.1)$	5947.27	2.12	7.27367×10^{6}	1.11536×10^{13}	5.73363×10^{54}
1	$(3.40,0.1)$	2111.43	2.14	1.39560×10^{7}	3.22035×10^{13}	1.18589×10^{56}
Καθολικό	ϕ_λ-απόκλιση	1.16070×10^{7}	3.81	5.23885×10^{8}	1.27025×10^{18}	4.08814×10^{109}

Πίνακας 2.4: Τιμές της τοπικής απόκλισης Cressie-Read $D^{\omega}_{\phi_\lambda}(\widehat{\theta}_1,\widehat{\theta}_2)$ για διάφορες τιμές της παραμέτρου λ, προκειμένου να συγκριθούν οι πληθυσμοί Π_1 και Π_2. Στην τελευταία γραμμή φαίνεται η τιμή του καθολικού μέτρου $D_{\phi_\lambda}(\widehat{\theta}_1,\widehat{\theta}_2)$. Ο πυρήνας και τα πληθυσμιακά μοντέλα είναι μονοδιάστατες κανονικές κατανομές με τις εκτιμώμενες παραμέτρους για τους πληθυσμούς Π_1 και Π_2 να είναι $\widehat{\theta}_1=(\widehat{\mu}_1,\widehat{\sigma}^2_1)=(3.40,0.04)$ και $\widehat{\theta}_2=(\widehat{\mu}_2,\widehat{\sigma}^2_2)=(2.48,0.03)$, αντίστοιχα. Βάσει των εκτιμητών αυτών, θεωρούμε κυρτούς συνδυασμούς των εκτιμώμενων μέσων τιμών για διαφορετικές τιμές του k και το αποτέλεσμα λαμβάνεται ως μέση τιμή του πυρήνα, δηλαδή $\theta=(\mu,\sigma^2)=(k\widehat{\mu}_1+(1-k)\widehat{\mu}_2,0.1)$. Οι παράμετροι του πυρήνα παρουσιάζονται στη δεύτερη στήλη. Όλες οι τιμές δείχνουν απόκλιση μεταξύ των πληθυσμών Π_1 και Π_2.

k	Π_1-Π_3					
	$\omega = (\mu, \sigma^2)$	$\lambda = -2$	$\lambda = -0.5$	$\lambda = \frac{2}{3}$	$\lambda = 1$	$\lambda = 2$
0	$(2.99, 0.1)$	12.41	1.26	3.85	20.88	5.33459×10^7
0.1	$(3.03, 0.1)$	11.36	1.29	4.78	28.19	1.03967×10^8
0.3	$(3.12, 0.1)$	8.91	1.33	7.59	53.60	4.54339×10^8
0.5	$(3.20, 0.1)$	6.83	1.35	11.06	91.09	1.63409×10^9
0.7	$(3.28, 0.1)$	5.00	1.36	15.53	148.82	5.70873×10^9
0.9	$(3.36, 0.1)$	3.53	1.36	20.95	233.62	1.93717×10^{10}
1	$(3.40, 0.1)$	2.93	1.35	23.95	288.35	3.52976×10^{10}
Καθολικό	ϕ_λ-απόκλιση	14.40	1.82	49.84	2369.73	1.7256×10^{21}

Πίνακας 2.5: Τιμές της τοπικής απόκλισης Cressie-Read $D^{\omega}_{\phi_\lambda}(\widehat{\theta}_1, \widehat{\theta}_3)$ για διάφορες τιμές της παραμέτρου λ, προκειμένου να συγκριθούν οι πληθυσμοί Π_1 και Π_3. Στην τελευταία γραμμή φαίνεται η τιμή του καθολικού μέτρου $D_{\phi_\lambda}(\widehat{\theta}_1, \widehat{\theta}_3)$. Ο πυρήνας και τα πληθυσμιακά μοντέλα είναι μονοδιάστατες κανονικές κατανομές με τις εκτιμώμενες παραμέτρους για τους πληθυσμούς Π_1 και Π_3 να είναι $\widehat{\theta}_1 = (\widehat{\mu}_1, \widehat{\sigma}^2_1) = (3.40, 0.04)$ και $\widehat{\theta}_3 = (\widehat{\mu}_3, \widehat{\sigma}^2_3) = (2.99, 0.03)$, αντίστοιχα. Βάσει των εκτιμητών αυτών, θεωρούμε κυρτούς συνδυασμούς των εκτιμώμενων μέσων τιμών για διαφορετικές τιμές του k και το αποτέλεσμα λαμβάνεται ως μέση τιμή του πυρήνα, δηλαδή $\theta = (\mu, \sigma^2) = (k\widehat{\mu}_1 + (1-k)\widehat{\mu}_3, 0.1)$. Οι παράμετροι του πυρήνα παρουσιάζονταιι στη δεύτερη στήλη. Όλες οι τιμές δείχουν απόκλιση μεταξύ των πληθυσμών Π_1 και Π_3.

k	Π_2-Π_3					
	$\omega = (\mu, \sigma^2)$	$\lambda = -2$	$\lambda = -0.5$	$\lambda = \frac{2}{3}$	$\lambda = 1$	$\lambda = 2$
0	$(2.99, 0.1)$	1184.27	1.86	7.70	59.08	4.48366×10^6
0.1	$(2.94, 0.1)$	963.90	1.90	10.42	86.47	7.99876×10^6
0.3	$(2.84, 0.1)$	602.76	1.94	18.29	175.20	2.40296×10^7
0.5	$(2.74, 0.1)$	349.05	1.95	30.05	329.10	6.68441×10^7
0.7	$(2.64, 0.1)$	187.24	1.94	45.95	572.69	1.72176×10^8
0.9	$(2.53, 0.1)$	86.47	1.90	67.19	963.90	4.46054×10^8
1	$(2.48, 0.1)$	59.08	1.86	77.46	1184.27	6.66715×10^8
Καθολικό	ϕ_λ-απόκλιση	2912.25	2.65	110.30	2912.25	3.29495×10^{10}

Πίνακας 2.6: Τιμές της τοπικής απόκλισης Cressie-Read $D^{\omega}_{\phi_\lambda}(\widehat{\theta}_2, \widehat{\theta}_3)$ για διάφορες τιμές της παραμέτρου λ, προκειμένου να συγκριθούν οι πληθυσμοί Π_2 και Π_3. Στην τελευταία γραμμή φαίνεται η τιμή του καθολικού μέτρου $D_{\phi_\lambda}(\widehat{\theta}_2, \widehat{\theta}_3)$. Ο πυρήνας και τα πληθυσμιακά μοντέλα είναι μονοδιάστατες κανονικές κατανομές με τις εκτιμώμενες παραμέτρους για τους πληθυσμούς Π_2 και Π_3 να είναι $\widehat{\theta}_2 = (\widehat{\mu}_2, \widehat{\sigma}^2_2) = (2.48, 0.03)$ και $\widehat{\theta}_3 = (\widehat{\mu}_3, \widehat{\sigma}^2_3) = (2.99, 0.03)$, αντίστοιχα. Βάσει των εκτιμητών αυτών, θεωρούμε κυρτούς συνδυασμούς των εκτιμώμενων μέσων τιμών για διαφορετικές τιμές του k, και το αποτέλεσμα λαμβάνεται ως μέση τιμή του πυρήνα, δηλαδή $\theta = (\mu, \sigma^2) = (k\widehat{\mu}_2 + (1-k)\widehat{\mu}_3, 0.1)$. Οι παράμετροι του πυρήνα παρουσιάζονται στη δεύτερη στήλη. Όλες οι τιμές δείχουν απόκλιση μεταξύ των πληθυσμών Π_2 και Π_3.

k	Π_1-Π_2		Π_1-Π_3		Π_2-Π_3	
	$\omega = (\mu, \sigma^2)$	$D_0^\omega(\theta_1, \theta_2)$	$\omega = (\mu, \sigma^2)$	$D_0^\omega(\theta_1, \theta_3)$	$\omega = (\mu, \sigma^2)$	$D_0^\omega(\theta_2, \theta_3)$
0	$(2.48, 0.1)$	1.38	$(2.99, 0.1)$	1.24	$(2.99, 0.1)$	1.65
0.1	$(2.57, 0.1)$	1.62	$(3.03, 0.1)$	1.35	$(2.94, 0.1)$	1.85
0.3	$(2.76, 0.1)$	2.70	$(3.12, 0.1)$	1.61	$(2.84, 0.1)$	2.31
0.5	$(2.94, 0.1)$	5.06	$(3.20, 0.1)$	1.87	$(2.74, 0.1)$	2.85
0.7	$(3.13, 0.1)$	9.04	$(3.28, 0.1)$	2.13	$(2.64, 0.1)$	3.41
0.9	$(3.31, 0.1)$	12.81	$(3.36, 0.1)$	2.37	$(2.53, 0.1)$	3.93
1	$(3.40, 0.1)$	13.99	$(3.40, 0.1)$	2.47	$(2.48, 0.1)$	4.10
Καθολικό	$D_0(\theta_1, \theta_2) = 14.13$		$D_0(\theta_1, \theta_3) = 2.82$		$D_0(\theta_2, \theta_3) = 4.34$	

Πίνακας 2.7: Τιμές της τοπικής απόκλισης Kullback-Leibler $D_0^\omega(\cdot, \cdot)$, προκειμένου να συγκριθούν ανά δύο οι τρεις πληθυσμοί. Στην τελευταία γραμμή φαίνεται η τιμή του καθολικού μέτρου $D_0(\cdot, \cdot)$. Ο πυρήνας και τα πληθυσμιακά μοντέλα είναι μονοδιάστατες κανονικές κατανομές, με τις εκτιμώμενες παραμέτρους για τους πληθυσμούς Π_1, Π_2 και Π_3 να είναι $\widehat{\theta}_1 = (\widehat{\mu}_1, \widehat{\sigma}_1^2) = (3.40, 0.04)$, $\widehat{\theta}_2 = (\widehat{\mu}_2, \widehat{\sigma}_2^2) = (2.48, 0.03)$ και $\widehat{\theta}_3 = (\widehat{\mu}_3, \widehat{\sigma}_3^2) = (2.99, 0.03)$, αντίστοιχα. Βάσει των εκτιμητών αυτών, θεωρούμε κυρτούς συνδυασμούς των εκτιμώμενων μέσων τιμών για διάφορες τιμές του k και το αποτέλεσμα λαμβάνεται ως μέση τιμή του πυρήνα. Όλες οι τιμές δείχουν απόκλιση μεταξύ των πληθυσμών Π_1-Π_2, Π_1-Π_3 και Π_2-Π_3.

Κεφάλαιο 3

Έλεγχος στατιστικών υποθέσεων μέσω μέτρων τοπικής απόκλισης

Κατά τη διάρκεια των τελευταίων δεκαετιών η Στατιστική Θεωρία Πληροφοριών αποτελεί βασικό εργαλείο για την ανάπτυξη μεθοδολογιών στη Μαθηματική Στατιστική και ειδικότερα στους ελέγχους στατιστικών υποθέσεων που αποτελούν και το αντικείμενο του κεφαλαίου αυτού. Η εισαγωγή μέτρων απόκλισης σε ελέγχους υποθέσεων σηματοδοτείται από τις εργασίες των Cressie and Read (1984) και Zografos et al. (1990), μεταξύ άλλων, στις οποίες με τη χρήση της λ-απόκλισης και της ϕ-απόκλισης, αντίστοιχα, κατασκευάζονται στατιστικά τεστ για πολυωνυμικά δεδομένα. Οι μονογραφίες των Read and Cressie (1988) και Pardo (2006) και οι αναφορές που εκεί παρατίθενται παρέχουν μια εξαντλητική συζήτηση για τη χρήση των μέτρων απόκλισης στον έλεγχο στατιστικών υποθέσεων, μέχρι τη χρονολογία συγγραφής τους, ενώ νεότερα αποτελέσματα παρατίθενται στην εργασία των Basu et al. (2013). Έλεγχοι ομοιογένειας μεταξύ δύο ή περισσότερων κατανομών είχαν προγενέστερα μελετηθεί στην εργασία του Zografos (1998). Τα διαθέσιμα στη βιβλιογραφία παραμετρικά στατιστικά τεστ που ορίζονται μέσω ϕ-αποκλίσεων, διερευνούν τα προβλήματα σε ολόκληρο το πεδίο ορισμού των κατανομών στις οποίες στηρίζονται.

Όμως, όπως έχει ήδη αναφερθεί στην Ενότητα 1.2, είναι πιθανό ο ισχυρισμός της καλής προσαρμογής ή της ομοιογένειας ή οποιασδήποτε άλλης στατιστικής υπόθεσης να απορρίπτεται ή να μην μπορεί να απορριφθεί σε ολόκληρο το πεδίο ορισμού των θεωρούμενων κατανομών, αλλά να προκύπτει διαφορετικό συμπέρασμα όταν το ενδιαφέρον επικεντρώνεται σ' ένα υποσύνολο του πεδίου ορισμού των κατανομών. Η διαπίστωση αυτή, η οποία προέκυψε μέσω αριθμητικών παραδειγμάτων, έδωσε το έναυσμα για την εισαγωγή, την κατασκευή και τη μελέτη παραμετρικών στατιστικών τεστ καλής προσαρμογής και ομοιογένειας, τοπικά, σ' ένα υποσύνολο του πεδίου ορισμού των θεωρούμενων κατανομών.

Στο στατιστικό αυτό πλαίσιο, το Κεφάλαιο 3 επικεντρώνεται στην κατασκευή και μελέτη παραμετρικών στατιστικών τεστ για τον έλεγχο καλής προσαρμογής (one sample problem) και τον έλεγχο ομοιογένειας (two samples problem), τοπικά, σε ένα υποσύνολο του πεδίου ορισμού των υπό θεώρηση κατανομών. Αναλυτικότερα, η διάρθρωση αυτού του κεφαλαίου έχει ως εξής. Στην Ενότητα 3.1 μελετάται το πρόβλημα του ελέγχου καλής προσαρμογής και προσδιορίζονται οι ασυμπτωτικές κατανομές τριών στατιστικών συναρτήσεων για τον έλεγχο της απλής μηδενικής υπόθεσης, η οποία καθορίζει πλήρως την κατανομή του πληθυσμού από την οποία προέρχεται το τυχαίο δείγμα. Αυτό σημαίνει ότι υπό τη μηδενική υπόθεση καθορίζεται τόσο η συναρτησιακή μορφή της κατανομής όσο και η τιμή της παραμέτρου που υπεισέρχεται σ' αυτή. Η ενότητα ολοκληρώνεται με τον προσδιορισμό της ασυμπτωτικής κατανομής της στατιστικής συνάρτησης με βάση την τοπική ϕ-απόκλιση υπό τοπικές εναλλακτικές (local alternatives) στατιστικές υποθέσεις. Στην Ενότητα 3.2 το τοπικό πρόβλημα ελέγχου καλής προσαρμογής μελετάται για σύνθετη μηδενική υπόθεση, δηλαδή για την περίπτωση που υπό τη μηδενική υπόθεση καθορίζεται μόνο η συναρτησιακή μορφή της κατανομής και παραμένει άγνωστη η τιμή της παραμέτρου της. Στην Ενότητα 3.3 το ενδιαφέρον επικεντρώνεται στον τοπικό έλεγχο ομοιογένειας χρησιμοποιώντας την τοπική ϕ-απόκλιση. Η αξιολόγηση της απόδοσης των προτεινόμενων τοπικών κριτηρίων των Ενοτήτων 3.1, 3.2 και 3.3 ως προς το επίπεδο σημαντικότητας και την ισχύ διενεργείται μέσω προσομοιώσεων. Τα αποτελέσματα αυτών των προσομοιώσεων παρουσιάζονται στην Ενότητα 3.4. Τέλος, το κεφάλαιο αυτό ολοκληρώνεται με την εφαρμογή των προτεινόμενων τοπικών κριτηρίων σε δύο πραγματικά σύνολα δεδομένων. Στόχος των εφαρμογών αυτών είναι η ανάδειξη της χρησιμότητας των προτεινόμενων τοπικών στατιστικών τεστ στην πράξη.

Έστω $\theta = (\theta_1, ..., \theta_M)$ ένα M-διάστατο διάνυσμα του χώρου R^M. Σε όσα ακολουθούν με $\nabla_\theta l_\theta(\cdot)$ συμβολίζεται το M-διάστατο ανάδελτα διάνυσμα της $l_\theta(\cdot)$ ως προς θ, ήτοι

$$\nabla_\theta l_\theta(\cdot) = \left[\frac{\partial}{\partial\theta_1}l_\theta(\cdot), \frac{\partial}{\partial\theta_2}l_\theta(\cdot), ..., \frac{\partial}{\partial\theta_M}l_\theta(\cdot)\right]^t$$

και με $\nabla^2_\theta l_\theta(\cdot)$ ο αντίστοιχος $M \times M$ Εσσιανός πίνακας (Hessian matrix), ήτοι

$$\nabla^2_\theta l_\theta(\cdot) = \begin{pmatrix} \frac{\partial^2}{\partial\theta_1^2}l_\theta(\cdot) & \frac{\partial^2}{\partial\theta_1\partial\theta_2}l_\theta(\cdot) & \cdots & \frac{\partial^2}{\partial\theta_1\partial\theta_M}l_\theta(\cdot) \\ \frac{\partial^2}{\partial\theta_2\partial\theta_1}l_\theta(\cdot) & \frac{\partial^2}{\partial\theta_2^2}l_\theta(\cdot) & \cdots & \frac{\partial^2}{\partial\theta_2\partial\theta_M}l_\theta(\cdot) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2}{\partial\theta_M\partial\theta_1}l_\theta(\cdot) & \frac{\partial^2}{\partial\theta_M\partial\theta_2}l_\theta(\cdot) & \cdots & \frac{\partial^2}{\partial\theta_M^2}l_\theta(\cdot) \end{pmatrix},$$

όπου $l_\theta(\cdot)$ είναι μια πραγματική συνάρτηση.

3.1 Τοπικός έλεγχος καλής προσαρμογής: Απλή μηδενική υπόθεση

Στην ενότητα αυτή θα προταθούν και θα μελετηθούν τρεις στατιστικές συναρτήσεις για τον τοπικό έλεγχο καλής προσαρμογής, στην περίπτωση της απλής μηδενικής υπόθεσης.

Έστω $(\mathcal{X}, \mathcal{A}, P_\theta)$, $\theta \in \Theta \subseteq \mathcal{R}^M$, με $M \geq 1$, είναι ένας χώρος πιθανότητας και P_θ μια οικογένεια μέτρων πιθανότητας και ένα σ-πεπερασμένο μέτρο μ στον ίδιο μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$, με $P_\theta \ll \mu$, $\theta \in \Theta$. Έστω επίσης στον ίδιο μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$, ένα μέτρο πιθανότητας H_ω, $\omega \in \Theta^*$, $\Theta^* \subseteq \mathcal{R}^K$, με $K \geq 1$ και $H_\omega \ll \mu$. Για γνωστό $\omega \in \Theta^*$, θεωρούμε τις αντίστοιχες Radon-Nikodym παραγώγους $f_\theta = \frac{dP_\theta}{d\mu}$ και $f_\omega = \frac{dP_\omega}{d\mu}$. Ας υποθέσουμε επίσης ότι η οικογένεια πυκνοτήτων $\{f_\theta : \theta \in \Theta \subseteq \mathcal{R}^M\}$ είναι ταυτοποιήσιμη (identifiable), δηλαδή ότι ισχύει: Αν $f_{\theta_1} = f_{\theta_2}$, τότε $\theta_1 = \theta_2$ για κάθε $\theta_1, \theta_2 \in \Theta$.

Στο παραπάνω στατιστικό πλαίσιο και με βάση ένα τυχαίο δείγμα $X_1, ..., X_n$ από την f_θ, είναι ενδιαφέρον να προσδιοριστεί μια στατιστική συνάρτηση για τον έλεγχο της μηδενικής υπόθεσης $H_0 : f_\theta = f_{\theta_0}$, η οποία είναι ταυτόσημη με την υπόθεση

$$H_0 : \theta = \theta_0,\ \theta_0 \in \Theta \text{ και } \theta_0 \text{ γνωστό.}$$

Στη βιβλιογραφία έχουν εμφανιστεί τρόποι ελέγχου της παραπάνω υπόθεσης που στηρίζονται σε καθολικά μέτρα τύπου ϕ-απόκλισης. Η μονογραφία του Pardo (2006, Chapter 9) και οι αναφορές που εκεί παρατίθενται παρέχουν μια εξαντλητική συζήτηση για το θέμα αυτό. Ωστόσο, όπως προηγουμένως αναφέρθηκε, οι έλεγχοι αυτοί στηρίζονται σε καθολικά μέτρα και επομένως το όποιο συμπέρασμα, σχετικά με την ταύτιση της f_θ με την f_{θ_0} ή ισοδύναμα της θ με την θ_0, θα αναφέρεται στο χώρο $\mathcal{X}$. Ενδέχεται όμως σε κάποια σημεία του $\mathcal{X}$ να υπάρχει απόκλιση της f_θ από την f_{θ_0}, την οποία ένα καθολικό μέτρο ϕ-απόκλισης δεν μπορεί να εντοπίσει και αναγνωρίσει σε ένα υποσύνολο του πεδίου ορισμού. Στην ενότητα αυτή προτείνονται τρεις στατιστικές συναρτήσεις για τον τοπικό έλεγχο καλής προσαρμογής. Για κάθε στατιστική συνάρτηση που προτείνεται μελετάται η ασυμπτωτική της κατανομή υπό τη μηδενική υπόθεση. Σε όσα ακολουθούν, χρησιμοποιούνται οι κλασικές τεχνικές της ασυμπτωτικής στατιστικής. Ενδεικτικά παραπέμπουμε στο σύγγραμμα των Sen and Singer (1993, Κεφάλαιο 5) για θέματα που άπτονται της θεωρίας των μεγάλων δειγμάτων.

3.1.1 Έλεγχος καλής προσαρμογής μέσω της τοπικής ϕ-απόκλισης

Οι στατιστικές συναρτήσεις που έχουν προταθεί στη βιβλιογραφία για τον έλεγχο της απλής μηδενικής υπόθεσης $H_0 : \theta = \theta_0$ χρησιμοποιώντας καθολικά μέτρα τύπου ϕ-απόκλισης διατυπώνονται μέσω του καθολικού μέτρου απόκλισης $D_\phi(f_\theta, f_{\theta_0})$ και βασίζονται στη στατιστική συνάρτηση $D_\phi(f_{\widehat{\theta}}, f_{\theta_0})$, όπου $\widehat{\theta}$ ένας εκτιμητής της παραμέτρου θ. Παρακινούμενοι από αυτήν την παρατήρηση και έχοντας ως στόχο την ανάπτυξη μιας μεθοδολογίας, η οποία θα επιτρέπει να εστιάζουμε σε ένα υποσύνολο του πεδίου ορισμού $\mathcal{X}$ και να ελέγχουμε σε αυτό την εγκυρότητα της μηδενικής υπόθεσης, θα χρησιμοποιήσουμε για τη διατύπωσή της την τοπική ϕ-απόκλιση.

Ειδικότερα, λαμβάνοντας υπόψη τη σχέση (2.4) του Ορισμού 2.2, προκύπτει ότι για την ποσότητα,

$$D_\phi^\omega(f_\theta, f_{\theta_0}) = D_\phi^\omega(\theta, \theta_0) = \int\limits_{\mathcal{X}} h_\omega(x) f_{\theta_0}(x) \phi\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right) d\mu(x), \tag{3.1}$$

όπου ϕ κυρτή συνάρτηση, $\phi \in \Phi$, με Φ όπως ορίστηκε στη (2.11), ισχύει, από τις ιδιότητες της τοπικής ϕ απόκλισης, ότι $D_\phi^\omega(f_\theta, f_{\theta_0}) = 0$ αν και μόνο αν $f_\theta = f_{\theta_0}$. Από τη συζήτηση που προηγήθηκε και καθώς η ποσότητα $D_\phi^\omega(f_\theta, f_{\theta_0})$ εξαρτάται από την άγνωστη παράμετρο θ είναι προφανές ότι ο έλεγχος της μηδενικής υπόθεσης, τοπικά, βασίζεται στην εκτιμώμενη τοπική ϕ-απόκλιση

$$D_\phi^\omega(f_{\widehat{\theta}}, f_{\theta_0}) = D_\phi^\omega(\hat{\theta}, \theta_0) = \int_{\mathcal{X}} f_\omega(x) f_{\theta_0}(x) \phi\left(\frac{f_{\widehat{\theta}}(x)}{f_{\theta_0}(x)}\right) d\mu(x), \text{ δοθέντος } \omega \in \Theta^*, \quad (3.2)$$

όπου $\widehat{\theta}$ ο ΕΜΠ της παραμέτρου $\theta \in \Theta$. Είναι σαφές ότι μικρές τιμές της απόκλισης $D_\phi^\omega(f_{\widehat{\theta}}, f_{\theta_0})$ είναι υπέρ της μηδενικής υπόθεσης $H_0 : \theta = \theta_0$, ενώ μεγάλες τιμές της $D_\phi^\omega(f_{\widehat{\theta}}, f_{\theta_0})$ είναι υπέρ της απόρριψης της μηδενικής υπόθεσης $H_0 : \theta = \theta_0$ στο υποσύνολο του $\mathcal{X}$ που καθορίζεται από την παράμετρο ω της πυκνότητας πυρήνα h_ω, η οποία οδηγεί την κύρια μάζα του ολοκληρώματος (3.1) στο υποσύνολο του $\mathcal{X}$ που υπολογίζεται η τοπική ϕ-απόκλιση. Η τοπική μηδενική υπόθεση θα γράφεται ως

$$H_0 : f_\theta(x) = f_{\theta_0}(x), \text{ δοθέντος } h_\omega, \ \omega \in \Theta^* \quad (3.3)$$

ή

$$H_0 : f_\theta(x) = f_{\theta_0}(x), \ x \in A \text{ σ.π., δοθέντος A}, \quad (3.4)$$

όπου ω η παράμετρος του πυρήνα που καθορίζει το υποσύνολο A του $\mathcal{X}$ στο οποίο υπολογίζεται η απόκλιση της f_θ από την f_{θ_0} και στο οποίο ελέγχεται ουσιαστικά η εγκυρότητα της H_0. Σε όσα ακολουθούν η τοπική μηδενική υπόθεση σύντομα θα γράφεται

$$H_0^\omega : \theta = \theta_0, \quad (3.5)$$

αντί για $H_0 : f_\theta(x) = f_{\theta_0}(x)$, δοθέντος h_ω, $\omega \in \Theta^*$.

Για τον προσδιορισμό των τιμών εκείνων που οδηγούν σε απόρριψη της μηδενικής υπόθεσης, δηλαδή της κρίσιμης περιοχής, είναι απαραίτητος ο προσδιορισμός της ασυμπτωτικής κατανομής της $D_\phi^\omega(f_{\widehat{\theta}}, f_{\theta_0})$, υπό τη μηδενική υπόθεση. Για τον προσδιορισμό αυτό υποθέτουμε ότι ικανοποιούνται οι συνήθεις συνθήκες ομαλότητας της ασυμπτωτικής στατιστικής (βλέπε, Serfling, 1980, σελ. 144, Pardo, 2006, σελ. 58). Για χάριν πληρότητας οι συνθήκες αυτές παρατίθενται στη συνέχεια.

(R_1) Η πρώτη, δεύτερη και τρίτη μερικές παράγωγοι

$$\frac{\partial f_\theta(x)}{\partial \theta_i}, \frac{\partial^2 f_\theta(x)}{\partial \theta_i \partial \theta_j}, \frac{\partial^3 f_\theta(x)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \ i, j, k = 1, ..., M,$$

υπάρχουν για όλους τους δείκτες $i, j, k = 1, ..., M$.

(R_2) Η πρώτη, δεύτερη και τρίτη μερικές παράγωγοι της $f_\theta(x)$ ως προς θ_i, $i = 1, ..., M$, είναι απόλυτα φραγμένες από τις συναρτήσεις α, β και γ, αντίστοιχα, για τις οποίες ισχύει

$$\int_{\mathcal{X}} \alpha(x)d\mu(x) < \infty, \quad \int_{\mathcal{X}} \beta(x)d\mu(x) < \infty, \quad \int_{\mathcal{X}} \gamma(x)f_\theta(x)d\mu(x) < \infty, \tag{3.6}$$

(R_3) Για κάθε $\theta \in \Theta$ ο πίνακας πληροφορίας Fisher

$$I_F(\theta) = \left(\int_{\mathcal{X}} f_\theta(x) \frac{\partial \log f_\theta(x)}{\partial \theta_i} \frac{\partial \log f_\theta(x)}{\partial \theta_j} d\mu(x) \right)_{i,j=1,\dots,M}, \tag{3.7}$$

υπάρχει και είναι θετικά ορισμένος, με στοιχεία συνεχείς συναρτήσεις της παραμέτρου θ.

Η τοπική επέκταση του κλασικού πίνακα πληροφορίας Fisher $I_F(\theta)$ διαδραματίζει βασικό ρόλο στην ανάπτυξη της θεωρίας των τοπικών στατιστικών συναρτήσεων για τον τοπικό έλεγχο υποθέσεων.

Για δοθέν $\omega \in \Theta^*$, ορίζεται ο σταθμισμένος ή τοπικός πίνακας πληροφορίας $I_F^\omega(\theta)$ Fisher ως εξής:

Ορισμός 3.1. *Για δοθέν $\omega \in \Theta^*$, ο πίνακας*

$$I_F^\omega(\theta) = \left(\int_{\mathcal{X}} h_\omega(x) f_\theta(x) \frac{\partial \log f_\theta(x)}{\partial \theta_i} \frac{\partial \log f_\theta(x)}{\partial \theta_j} d\mu(x) \right)_{i,j=1,\dots,M}. \tag{3.8}$$

ονομάζεται σταθμισμένος ή τοπικός πίνακας πληροφορίας Fisher.

Σε όσα ακολουθούν υποθέτουμε ότι ο $I_F^\omega(\theta)$ υπάρχει και είναι θετικά ορισμένος με στοιχεία συνεχείς συναρτήσεις της παραμέτρου θ. Οι υποθέσεις αυτές είναι παρόμοιες με εκείνες για τον κλασικό πίνακα πληροφορίας Fisher. Επιπρόσθετα, με βάση τον Pardo (2006, σελ. 408), η μελέτη θα περιοριστεί στις κυρτές συναρτήσεις ϕ που ανήκουν στην κλάση των κυρτών συναρτήσεων Φ, όπως αυτή ορίζεται στη σχέση (2.11) και οι οποίες επιπλέον πληρούν τις ακόλουθες προϋποθέσεις.

(Φ_1) Η συνάρτηση $\phi \in \Phi$ είναι δύο φορές διαφορίσιμη με $\phi''(1) > 0$.

(Φ_2) Για κάθε $\theta_0 \in \Theta$ υπάρχει μια ανοιχτή περιοχή $N(\theta_0)$ έτσι ώστε για κάθε $\theta \in N(\theta_0)$ και $1 \le i, j \le M$, να ισχύει:

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} h_\omega(x) f_{\theta_0}(x) \phi\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right) d\mu(x) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} \left(h_\omega(x) f_{\theta_0}(x) \phi\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right)\right) d\mu(x),$$

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\int\limits_{\mathcal{X}} h_\omega(x)f_{\theta_0}(x)\phi\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right)d\mu(x) = \int\limits_{\mathcal{X}} \frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(h_\omega(x)f_{\theta_0}(x)\phi\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right)\right)d\mu(x)$$

και αυτές οι συναρτήσεις να είναι συνεχείς στην περιοχή $N(\theta_0)$.

Στο επόμενο θεώρημα προσδιορίζεται η ασυμπωτική κατανομή της εκτιμώμενης τοπικής ϕ-απόκλισης $D^\omega_\phi(\widehat{\theta},\theta_0)$ που ορίστηκε στη σχέση (3.2), υπό τη μηδενική υπόθεση.

Θεώρημα 3.1. *Έστω ότι η κυρτή συνάρτηση $\phi \in \Phi$ ικανοποιεί τις υποθέσεις Φ_1, Φ_2 και επιπλέον ισχύουν οι συνθήκες ομαλότητας R_1, R_2, R_3. Υπό τη μηδενική υπόθεση $H^\omega_0 : \theta = \theta_0$, η ασυμπτωτική κατανομή της ποσότητας $T^\omega_{\phi,n}(\widehat{\theta},\theta_0)$ είναι η*

$$T^\omega_{\phi,n}(\widehat{\theta},\theta_0) = \frac{2nD^\omega_\phi(\widehat{\theta},\theta_0)}{\phi''(1)} \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r}\beta_i Z_i^2, \tag{3.9}$$

όπου $Z_1, ...Z_r$, είναι r το πλήθος ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές, με

$$r = rank(I_F^{-1}(\theta_0)I_F^\omega(\theta_0)I_F^{-1}(\theta_0))$$

και $\beta_1, \beta_2, ..., \beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα

$$I_F^\omega(\theta_0)I_F^{-1}(\theta_0),$$

όπου $I_F(\theta)$ ο πίνακας πληροφορίας Fisher και $I_F^\omega(\theta)$ ο τοπικός πίνακας πληροφορίας Fisher, που ορίστηκαν στις σχέσεις (3.7) και (3.8), αντίστοιχα.

Απόδειξη. Το ανάπτυγμα Taylor της τοπικής ϕ-απόκλισης $D^\omega_\phi(\theta,\theta_0)$ γύρω από τη θ_0, για $\theta = \hat{\theta}$, δίνεται από τη σχέση

$$\begin{aligned} D^\omega_\phi(\hat{\theta},\theta_0) &= D^\omega_\phi(\theta_0,\theta_0) + \nabla_\theta\left(D^\omega_\phi(\theta,\theta_0)\right)_{\theta=\theta_0}(\hat{\theta}-\theta_0) \\ &+ \frac{1}{2}(\hat{\theta}-\theta_0)^t\nabla^2_\theta\left(D^\omega_\phi(\theta,\theta_0)\right)_{\theta=\theta_0}(\hat{\theta}-\theta_0) + o\left(\|\hat{\theta}-\theta_0\|^2\right). \end{aligned} \tag{3.10}$$

Σύμφωνα με τον Pardo (2006, σελ. 411-412) προκύπτει ότι $D^\omega_\phi(\theta_0,\theta_0) = 0$ και

$$\frac{\partial D^\omega_\phi(\theta,\theta_0)}{\partial\theta_i} = \int_{\mathcal{X}} h_\omega(x)\phi'(\frac{f_\theta(x)}{f_{\theta_0}(x)})\frac{\partial f_\theta(x)}{\partial\theta_i}d\mu(x).$$

Τότε, για $\theta = \theta_0$ έχουμε

$$\left(\frac{\partial D^\omega_\phi(\theta,\theta_0)}{\partial\theta_i}\right)_{\theta=\theta_0} = \phi'(1)\int_{\mathcal{X}} h_\omega(x)\left(\frac{\partial f_\theta(x)}{\partial\theta_i}\right)_{\theta=\theta_0}d\mu(x) = 0$$

εφόσον $\phi'(1) = 0$. Συνεπώς $\nabla_\theta \left(D^\omega_\phi(\theta, \theta_0) \right)_{\theta=\theta_0} = 0$. Επιπλέον, έχουμε

$$\begin{aligned} \frac{\partial^2 D^\omega_\phi(\theta,\theta_0)}{\partial\theta_i \partial\theta_j} &= \int_{\mathcal{X}} h_\omega(x) \phi''(\frac{f_\theta(x)}{f_{\theta_0}(x)}) \frac{1}{f_{\theta_0}(x)} \frac{\partial f_\theta(x)}{\partial \theta_i} \frac{\partial f_\theta(x)}{\partial \theta_j} d\mu(x) \\ &+ \int_{\mathcal{X}} h_\omega(x)) \phi'(\frac{f_\theta(x)}{f_{\theta_0}(x)}) \frac{\partial^2 f_\theta(x)}{\partial\theta_i \partial\theta_j} d\mu(x), \end{aligned}$$

και για $\theta = \theta_0$, ισχύει

$$\begin{aligned} \left(\frac{\partial^2 D^\omega_\phi(\theta,\theta_0)}{\partial\theta_i \partial\theta_j} \right)_{\theta=\theta_0} &= \phi''(1) \int_{\mathcal{X}} h_\omega(x) f_{\theta_0}(x) \frac{\partial \log f_{\theta_0}(x)}{\partial \theta_i} \frac{\partial \log f_{\theta_0}(x)}{\partial \theta_j} d\mu(x) \\ &= \phi''(1) \left(I^\omega_F(\theta_0) \right)_{i,j} \end{aligned}$$

ή

$$\nabla^2_\theta \left(D^\omega_\phi(\theta, \theta_0) \right)_{\theta=\theta_0} = \phi''(1) I^\omega_F(\theta_0).$$

Χρησιμοποιώντας τη σχέση (3.10) και λαμβάνοντας υπόψη τις παραπάνω ισότητες, προκύπτει ότι

$$D^\omega_\phi(\hat\theta, \theta_0) = \frac{1}{2} (\hat\theta - \theta_0)^t \phi''(1) I^\omega_F(\theta_0)(\hat\theta - \theta_0) + o\left(\|\hat\theta - \theta_0\|^2 \right),$$

ή

$$\frac{2D^\omega_\phi(\hat\theta, \theta_0)}{\phi''(1)} = (\hat\theta - \theta_0)^t I^\omega_F(\theta_0)(\hat\theta - \theta_0) + o\left(\|\hat\theta - \theta_0\|^2 \right).$$

Επίσης, $o\left(\|\hat\theta - \theta_0\|^2 \right) = o_p(n^{-1})$ (βλέπε, Pardo, 2006, σελ. 411), υπό την μηδενική υπόθεση H_0. Συνεπώς, για δοθέν $\omega \in \Theta^*$ είναι

$$T^\omega_{\phi,n}(\hat\theta, \theta_0) = \frac{2nD^\omega_\phi(\hat\theta, \theta_0)}{\phi''(1)} = n(\hat\theta - \theta_0)^t I^\omega_F(\theta_0)(\hat\theta - \theta_0) + o_p(1).$$

Επίσης, υπό τη μηδενική υπόθεση H_0, ισχύει ότι

$$\sqrt{n}(\hat\theta - \theta_0) \xrightarrow[n\to\infty]{L} N(0, I_F^{-1}(\theta_0)).$$

Το αποτέλεσμα εξάγεται χρησιμοποιώντας το Πόρισμα 2.1 των Dik and de Gunst (1985). ∎

Στη συνέχεια, και χάριν πληρότητας διατυπώνονται τα Πορίσματα 2.1 και 2.2 των Dik and de Gunst (1985), που θα χρησιμοποιηθούν σε αποδείξεις διαφόρων αποτελεσμάτων της διατριβής.

Πόρισμα 3.1 (Πόρισμα 2.1 Dik and de Gunst, 1985). *Έστω X μια q-διάστατη τυχαία μεταβλητή που ακολουθεί πολυδιάστατη κανονική κατανομή με μέση τιμή το μηδενικό διάνυσμα και πίνακα διακυμάνσεων-συνδιακυμάνσεων Σ, ήτοι $X \sim N_q(0, \Sigma)$. Έστω A ένας πραγματικός συμμετρικός πίνακας τάξης q. Έστω επίσης $r = rank(\Sigma A \Sigma)$, $r \geq 1$ και $\beta_1, \beta_2, ..., \beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα $A\Sigma$. Τότε η κατανομή της τετραγωνικής μορφής $X^t AX$ συμπίπτει με την κατανομή της τυχαίας μεταβλητής $\sum_{i=1}^{r} \beta_i Z_i^2$, όπου $Z_1, ...Z_r$ είναι ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές.*

Πόρισμα 3.2 (Πόρισμα 2.2 Dik and de Gunst, 1985). *Έστω X μια q-διάστατη τυχαία μεταβλητή που ακολουθεί πολυδιάστατη κανονική κατανομή με μέση τιμή το q-διάστατο διάνυσμα d και πίνακα διακυμάνσεων-συνδιακυμάνσεων Σ, ήτοι $X \sim N_q(d, \Sigma)$. Έστω A ένας πραγματικός, συμμετρικός και θετικά ορισμένος πίνακας τάξης q. Έστω επίσης $r = rank(\Sigma A \Sigma)$, $r \geq 1$ και $\beta_1, \beta_2, ..., \beta_r$ οι θετικές ιδιοτιμές του πίνακα $A\Sigma$. Τότε η κατανομή της τετραγωνικής μορφής $X^t AX$ συμπίπτει με την κατανομή της τυχαίας μεταβλητής $\sum_{i=1}^{r} \beta_i (Z_i + w_i)^2 + \xi$, όπου $Z_1, ...Z_r$ είναι ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές. Επίσης, είναι*

$$w = \Lambda^{-1} R^t S^t Ad$$

$$\xi = d^t Ad - w^t \Lambda w$$

με $\Lambda = diag(\beta_1, \beta_2, ..., \beta_r)$, S είναι μια αυθαίρετη ρίζα του πίνακα Σ και R είναι ο πίνακας των ιδιοδιανυσμάτων του $S^t AS$.

Παρατήρηση 3.1. *(1) Από το Θεώρημα 3.1, η τοπική μηδενική υπόθεση $H_0^\omega : \theta = \theta_0$ έναντι της εναλλακτικής $H_a^\omega : \theta \neq \theta_0$ απορρίπτεται σε επίπεδο σημαντικότητας α $(0 < \alpha < 1)$, όταν $T_{\phi,n}^\omega(\widehat{\theta}, \theta_0) > c_\alpha$, όπου το κρίσιμο σημείο c_α προσδιορίζεται έτσι ώστε ο έλεγχος να έχει επίπεδο σημαντικότητας α. Επομένως, προσδιορίζεται από την πιθανότητα σφάλματος τύπου Ι, η οποία ορίζεται ως εξής:*

$$\begin{aligned} \alpha &= P(\textit{Απόρριψη της } H_0^\omega \mid H_0^\omega \textit{ Αληθής}) \\ &= P(T_{\phi,n}^\omega(\widehat{\theta}, \theta) > c_\alpha \mid \theta = \theta_0 \textit{ και } h_\omega,\ \omega \in \Theta^*). \end{aligned} \tag{3.11}$$

Επιπλέον, η πιθανότητα σφάλματος τύπου ΙΙ ορίζεται από τη σχέση:

$$\begin{aligned} \beta &= P(\textit{Αποδοχή της } H_0^\omega \mid H_a^\omega \textit{ Αληθής}) \\ &= P(T_{\phi,n}^\omega(\widehat{\theta}, \theta) \leq c_\alpha \mid \theta = \theta^*(\theta^* \neq \theta_0) \textit{ και } h_\omega,\ \omega \in \Theta^*). \end{aligned} \tag{3.12}$$

Ο προσδιορισμός του κρίσιμου σημείου c_α δεν είναι εφικτός από τη σχέση (3.11), καθώς απαιτείται ο προσδιορισμός της κατανομής του γραμμικού συνδυασμού χ^2 τυχαίων μεταβλητών με ένα βαθμό ελευθερίας. Για να ξεπεραστεί το πρόβλημα αυτό, το κρίσιμο σημείο c_α προσεγγίζεται χρησιμοποιώντας μία εκ των διαδικασιών που χρησιμοποιούνται στη σχετική βιβλιογραφία και περιγράφονται στη συνέχεια.

(α) Προσέγγιση Monte Carlo: Σύμφωνα με την προσέγγιση αυτή, παράγεται ένα τυχαίο δείγμα $T_1, ..., T_n$ μεγέθους n από την κατανομή $T = \sum_{i=1}^{r} \beta_i Z_i^2$. Το κρίσιμο σημείο c_α επιλέγεται κατά τέτοιο τρόπο, έτσι ώστε

$$\frac{1}{n}\sum_{i=1}^{n} I(T_i > c_a) \simeq a, \tag{3.13}$$

για δεδομένο επίπεδο σημαντικότητας α, όπου $I(\cdot)$ η δείκτρια συνάρτηση.

(β) Σύμφωνα με τους Landaburu and Pardo (2000, σελ. 594) και τη βιβλιογραφία στην εργασία αυτή, προτείνεται η χρήση του στατιστικού ελέγχου $S_{\phi,n}^1 = \frac{T_{\phi,n}^\omega(\widehat{\theta},\theta_0)}{\beta_*}$, με $\beta_* = max\{\beta_1, \beta_2, ..., \beta_r\}$ και η μηδενική υπόθεση H_0^ω απορρίπτεται αν $S_{\phi,n}^1 > \chi_{r,\alpha}^2$ σε επίπεδο σημαντικότητας α. Επιπλέον, μπορεί να χρησιμοποιηθεί το στατιστικό ελέγχου $S_{\phi,n}^2 = \frac{T_{\phi,n}^\omega(\widehat{\theta},\theta_0)}{\bar{\beta}}$, με $\bar{\beta} = \frac{1}{r}\sum_{i=1}^{r}\beta_i$ και η μηδενική υπόθεση H_0^ω απορρίπτεται αν $S_{\phi,n}^2 > \chi_{r,\alpha}^2$ σε επίπεδο σημαντικότητας α.

(2) Όσον αφορά την ερμηνεία ενός τοπικού ελέγχου, θα πρέπει να σημειωθεί ότι η μόνη διαφορά μεταξύ του τοπικού ελέγχου και του συνηθισμένου (καθολικού) ελέγχου υποθέσεων, είναι ότι η υπόθεση αντιμετωπίζεται τοπικά, σε ένα συγκεκριμένο υποσύνολο του πεδίου ορισμού.

Στο επόμενο θεώρημα προσδιορίζεται η ασυμπτωτική κατανομή του εκτιμητή $D_\phi^\omega(\widehat{\theta}, \theta_0)$ της τοπικής απόκλισης $D_\phi^\omega(\theta, \theta_0)$. Η απόδειξη παραλείπεται καθώς ακολουθεί παρόμοια βήματα με την απόδειξη του Θεωρήματος 9.2 του Pardo (2006, σελ. 413).

Θεώρημα 3.2. *Έστω ότι η κυρτή συνάρτηση $\phi \in \Phi$ ικανοποιεί τις υποθέσεις Φ_1, Φ_2 και επιπλέον πληρούνται οι συνθήκες ομαλότητας R_1, R_2, R_3. Αν θ^* είναι η πραγματική τιμή της παραμέτρου θ, με $\theta^* \neq \theta_0$, τότε*

$$\sqrt{n}(D_\phi^\omega(\widehat{\theta}, \theta_0) - D_\phi^\omega(\theta^*, \theta_0)) \xrightarrow[n\to\infty]{L} N(0, \sigma_\phi^2(\theta^*)),$$

όπου $\sigma_\phi^2(\theta^*) = T^t I_F^{-1}(\theta^*) T$, $T = (t_1, ..., t_M)^T$, με στοιχεία $t_j = \left.\frac{\partial D_\phi^\omega(\theta,\theta_0)}{\partial \theta_j}\right|_{\theta=\theta^*}$, $j = 1, ..., M$.

Παρατήρηση 3.2. *Το Θεώρημα 3.2 επιτρέπει τον προσδιορισμό της συνάρτησης ισχύος σε κλειστή μορφή. Πιο συγκεκριμένα, η συνάρτηση ισχύος στο $\theta^* \neq \theta_0$, δίνεται από τη σχέση*

$$\beta_{n,\phi}(\theta^*) = 1 - \Phi\left(\frac{\sqrt{n}}{\sigma_{\phi(\theta^*)}}\left(\frac{\phi''(1)c_\alpha}{2n} - D_\phi^\omega(\theta^*, \theta_0)\right)\right),$$

όπου $\Phi(\cdot)$ η αθροιστική συνάρτηση κατανομής της τυπικής κανονικής κατανομής και c_α το κρίσιμο σημείο που προσδιορίζεται από την πιθανότητα σφάλματος τύπου I, ήτοι

$$P(T_{\phi,n}^\omega(\hat{\theta}, \theta_0) > c_\alpha) = \alpha.$$

Σύμφωνα με τον Pardo (2006, σελ. 414), αν $\theta^* \neq \theta_0$ είναι η πραγματική τιμή της παραμέτρου τότε η πιθανότητα απόρριψης της τιμής της παραμέτρου θ_0 με τον κανόνα απόρριψης $T_{\phi,n}^\omega(\widehat{\theta}, \theta_0) > c_\alpha$, για ένα σταθερό επίπεδο σημαντικότητας α, τείνει στο 1 όταν $n \to \infty$. Από το αποτέλεσμα αυτό εξάγεται το συμπέρασμα ότι ο έλεγχος είναι συνεπής (consistent).

Το τελευταίο αποτέλεσμα της ενότητας αυτής μελετά την ασυμπτωτική κατανομή του στατιστικού έλεγχου τοπικής ϕ-απόκλισης υπό μια τοπική εναλλακτική υπόθεση, που ουσιαστικά είναι μια ακολουθία εναλλακτικών υποθέσεων. Η απόδειξη του θεωρήματος περιγράφεται συνοπτικά χάριν πληρότητας, καθώς ακολουθεί παρόμοια βήματα με την απόδειξη του Θεωρήματος 9.3 του Pardo (2006, σελ. 414) και με το Θεώρημα 4 των Landaburu and Pardo (2000, σελ. 597).

Θεώρημα 3.3. *Υποθέτουμε ότι η κυρτή συνάρτηση $\phi \in \Phi$ ικανοποιεί τις υποθέσεις Φ_1, Φ_2 και ότι ισχύουν οι συνθήκες ομαλότητας R_1, R_2, R_3. Υπό την εναλλακτική υπόθεση $H_{\alpha,n}^\omega : \theta_n = \theta_0 + \frac{1}{\sqrt{n}}d$, όπου d είναι ένα σταθερό διάνυσμα στον $\mathcal{R}^M$, τέτοιο ώστε $\theta_n \in \Theta \subseteq \mathcal{R}^M$, ισχύει ότι*

$$T_{\phi,n}^\omega(\widehat{\theta}, \theta_0) = \frac{2nD_\phi^\omega(\widehat{\theta}, \theta_0)}{\phi''(1)} \xrightarrow[n\to\infty]{L} \sum_{i=1}^r \beta_i(Z_i + w_i)^2 + \xi,$$

όπου $Z_1, ...Z_r$ είναι r ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές, με

$$r = rank\left(I_F^{-1}(\theta_0)I_F^\omega(\theta_0)I_F^{-1}(\theta_0)\right)$$

και $\beta_1, \beta_2, ..., \beta_r$ οι θετικές ιδιοτιμές του πίνακα

$$I_F^\omega(\theta_0)I_F^{-1}(\theta_0),$$

όπου $I_F(\theta)$ ο πίνακας πληροφορίας Fisher και $I_F^\omega(\theta)$ ο τοπικός πίνακας πληροφορίας Fisher, αντίστοιχα. Επιπλέον

$$w = \ \Lambda^{-1} R^t \ S^t I_F^\omega(\theta_0) \ d,$$

και

$$\xi = d^t I_F^\omega(\theta_0) d - w^t \Lambda w,$$

με $\Lambda = diag(\beta_1, \beta_2, ..., \beta_r)$, S μια αυθαίρετη ρίζα του $I_F^{-1}(\theta_0)$ και R ο πίνακας των ιδιοδιανυσμάτων $S^t I_F^\omega(\theta_0) S$.

Απόδειξη. Σύμφωνα με τον Pardo (2006, σελ. 414),

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \sqrt{n}(\widehat{\theta} - \theta_n) + \sqrt{n}(\theta_n - \theta_0) = \sqrt{n}(\widehat{\theta} - \theta_n) + d.$$

Υπό την εναλλακτική υπόθεση $H_{\alpha,n}$, έχουμε

$$\sqrt{n}(\widehat{\theta} - \theta_n) \xrightarrow[n\to\infty]{L} N(0, I_F^{-1}(\theta_0)),$$

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow[n\to\infty]{L} N(d, I_F^{-1}(\theta_0)).$$

Το Θεώρημα 3.1, για δοθέν $\omega \in \Theta^*$, δίνει

$$T_{\phi,n}^\omega(\widehat{\theta}, \theta_0) = \frac{2nD_\phi^w(\widehat{\theta}, \theta_0)}{\phi''(1)} = n(\widehat{\theta} - \theta_0)^T I_F^\omega(\theta_0)(\widehat{\theta} - \theta_0) + o_p(1),$$

το οποίο οδηγεί στο επιθυμητό αποτέλεσμα, λαμβάνοντας υπόψη το Πόρισμα 2.2 των Dik and de Gunst, 1985. ∎

Παρατήρηση 3.3. *Το παρατάνω θεώρημα θα μπορούσε να συνδυαστεί με την αποδοτικότητα Pitman (Pitman efficiency) και να χρησιμοποιηθεί για τη διερεύνηση και την πιθανή επιλογή πυρήνα ή τη σύγκριση δύο πυρήνων ή ακόμη και την επιλογή της κυρτής συνάρτησης ϕ. Πιο συγκεκριμένα, αν η αποδοτικότητα Pitman μπορεί να προσδιορισθεί και αυτή εξαρτάται από τον πυρήνα h_ω ή/και την κυρτή συνάρτηση ϕ, τότε η αποδοτικότητα Pitman μπορεί να χρησιμοποιηθεί ως ένα κριτήριο επιλογής των συναρτήσεων h_ω και ϕ, που αποτελούν αναπόσπαστα στοιχεία της στατιστικής συνάρτησης $T_{\phi,n}^\omega(\widehat{\theta}, \theta_0)$. Η διερεύνηση αυτή θα μπορούσε να αποτελέσει αντικείμενο μελλοντικής έρευνας.*

3.1.2 Έλεγχος καλής προσαρμογής μέσω των τοπικών στατιστικών Wald και Rao

Στην ενότητα αυτή προτείνονται άλλοι δύο τρόποι τοπικού ελέγχου καλής προσαρμογής, στην περίπτωση της απλής μηδενικής υπόθεσης. Ειδικότερα, θα επεκταθούν σε τοπικό περιβάλλον το στατιστικό ελέγχο τύπου Wald (βλέπε, Pardo, 2006, σελ. 408-409 για τον καθολικό έλεγχο) και το στατιστικό έλεγχου Rao. Στο στατιστικό πλαίσιο της προηγούμενης υποενότητας, προτείνονται οι στατιστικές συναρτήσεις

$$W_n^\omega = n(\widehat{\theta}-\theta_0)^t I_F^\omega(\widehat{\theta})(\widehat{\theta}-\theta_0),$$

και

$$R_n^\omega = \frac{1}{n} U_n^t(\theta_0) I_F^\omega(\theta_0)^{-1} U_n(\theta_0),$$

με

$$U_n(\theta_0) = \left(\sum_{i=1}^{n} \frac{\partial \log f_\theta(X_i)}{\partial \theta_1}, \ldots, \sum_{i=1}^{n} \frac{\partial \log f_\theta(X_i)}{\partial \theta_M} \right)^t_{\theta=\theta_0}.$$

Στο θεώρημα που ακολουθεί προσδιορίζεται η ασυμπτωτική κατανομή των τοπικών εκδοχών των στατιστικών συναρτήσεων Wald και Rao για τον έλεγχο της τοπικής μηδενικής υπόθεσης $H_0^\omega : \theta = \theta_0$.

Θεώρημα 3.4. *Υποθέτουμε ότι οι συνθήκες ομαλότητας R_1, R_2, R_3 ικανοποιούνται. Υπό τη μηδενική υπόθεση $H_0^\omega : \theta = \theta_0$, για την τοπική εκδοχή του στατιστικού Wald, W_n^ω, ισχύει ότι*

$$W_n^\omega = n(\widehat{\theta}-\theta_0)^T I_F^\omega(\widehat{\theta})(\widehat{\theta}-\theta_0) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

ενώ για την τοπική εκδοχή του στατιστικού Rao, R_n^ω, ισχύει ότι

$$R_n^\omega \xrightarrow[n\to\infty]{L} \sum_{i=1}^{s} a_i Z_i^2,$$

όπου $Z_1, ..., Z_r, ..., Z_s$ είναι ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές, με

$$r = rank(I_F^{-1}(\theta_0) I_F^\omega(\theta_0) I_F^{-1}(\theta_0)),$$

και

$$s = rank(I_F(\theta_0) I_F^\omega(\theta_0)^{-1} I_F(\theta_0)).$$

Οι $\beta_1, \beta_2, ..., \beta_r$ και $a_1, a_2, ..., a_s$ είναι οι μη μηδενικές ιδιοτιμές του πίνακα

$$I_F^\omega(\theta_0) I_F^{-1}(\theta_0)$$

και του πίνακα

$$I_F^\omega(\theta_0)^{-1} I_F(\theta_0),$$

αντίστοιχα, με $I_F(\theta)$ να είναι ο πίνακας πληροφορίας του Fisher και $I_F^\omega(\theta)$ ο τοπικός πίνακας πληροφορίας του Fisher, όπως ορίστηκαν στις (3.7) και (3.8), αντίστοιχα.

Απόδειξη. Υπό τη μηδενική υπόθεση $H_0 : \theta = \theta_0$ και με βάση τους Pardo (2006, σελ. 410) και Sen and Singer (1993, σελ. 236),

$$I_F(\widehat{\theta}) \xrightarrow[n\to\infty]{P} I_F(\theta_0).$$

Συνεπώς για δοθέν $\omega \in \Theta^*$, προκύπτει ότι

$$I_F^\omega(\widehat{\theta}) \xrightarrow[n\to\infty]{P} I_F^\omega(\theta_0), \tag{3.14}$$

όπου $\widehat{\theta}$ είναι ο ΕΜΠ της παραμέτρου θ. Επιπλέον, ισχύει ότι,

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow[n\to\infty]{L} N(0, I_F^{-1}(\theta_0)). \tag{3.15}$$

Κατά συνέπεια, η ασυμπτωτική κατανομή του τοπικού στατιστικού έλεγχου του Wald λαμβάνεται από τις (3.14), (3.15) και το Πόρισμα 2.1 των Dik and de Gunst (1985). Σύμφωνα με τους Pardo (2006, σελ. 410) και Sen and Singer (1993, σελ. 236), υπό την μηδενική υπόθεση $H_0 : \theta = \theta_0$, ισχύει ότι

$$\frac{1}{\sqrt{n}} U_n(\theta_0) \xrightarrow[n\to\infty]{L} N(0, I_F(\theta_0)).$$

Η ασυμπτωτική κατανομή του τοπικού στατιστικού έλεγχου του Rao R_n^ω, προσδιορίζεται χρησιμοποιώντας τον ορισμό της στατιστικής συνάρτησης R_n^ω και το Πόρισμα 2.1 των Dik and de Gunst (1985). ∎

Για τον προσδιορισμό των κρίσιμων σημείων για τη διενέργεια του ελέγχου εφαρμόζεται η προσέγγιση Monte Carlo της Παρατήρησης 3.1.

3.2 Τοπικός έλεγχος καλής προσαρμογής: Σύνθετη μηδενική υπόθεση

Σκοπός αυτής της ενότητας είναι η ανάπτυξη στατιστικού τεστ για τον έλεγχο σύνθετης μηδενικής υπόθεσης, χρησιμοποιώντας την τοπική ϕ-απόκλιση.

Μια σύνθετη μηδενική υπόθεση (βλέπε, Sen and Singer, 1993, σελ. 239) είναι μια υπόθεση της μορφής

$$H_0 : h(\theta) = 0, \tag{3.16}$$

όπου h είναι μια συνάρτηση τέτοια ώστε $h : \Theta \subseteq \mathcal{R}^M \longrightarrow \mathcal{R}^d$, με $d < M$. Στο πλαίσιο αυτό, ενδιαφέρει ο έλεγχος της μηδενικής υπόθεσης H_0 έναντι της εναλλακτικής

$$H_\alpha : h(\theta) \neq 0.$$

Είναι σαφές ότι η συνάρτηση h περιορίζει τον αρχικό παραμετρικό χώρο Θ σε έναν άλλο χώρο, έστω $\Theta_0 \subseteq \Theta$ και ο χώρος Θ_0 διέπεται από τους περιορισμούς που επιβάλλονται στις παραμέτρους από τη συνάρτηση h.

Προκειμένου να μετατραπεί η σύνθετη υπόθεση H_0 σε απλή, θεωρούμε μια συνάρτηση $g : B \subseteq \mathcal{R}^{M-d} \longrightarrow \Theta \subseteq \mathcal{R}^M$ τέτοια ώστε οι H_0 και H_α να είναι ισοδύναμες με τις υποθέσεις

$$H_0 : \theta = g(\beta) \text{ έναντι της } H_\alpha : \theta \neq g(\beta),$$

για κάποιο $\beta \in B$.

Το παρακάτω παράδειγμα δίνει μια εικόνα του τρόπου μετατροπής μιας σύνθετης μηδενικής υπόθεσης σε απλή, έτσι ώστε να μπορεί να αξιοποιηθεί η μεθοδολογία που έχει αναπτυχθεί για την απλή μηδενική υπόθεση. Για μια ενδελεχή μελέτη των θεμάτων αυτών παραπέμπουμε στο βιβλίο των Sen and Singer (1993).

Παράδειγμα 3.1. *Έστω ένα τυχαίο δείγμα από πληθυσμό με συνάρτηση πυκνότητας πιθανότητας f_θ, όπου $\theta = (\theta_1, \theta_2)^t$ με $\theta_1, \theta_2 \in \mathcal{R}$ και $\theta_2 > 0$. Τότε ο παραμετρικός χώρος είναι $\Theta = \{(\theta_1, \theta_2) \in \mathcal{R}^2 : \theta_1 \in \mathcal{R}, \theta_2 > 0\}$. Έστω ότι ενδιαφέρει ο έλεγχος,*

$$H_0 : \theta_1 = \theta_2 \quad \textit{έναντι της} \quad H_1 : \theta_1 \neq \theta_2.$$

Οι υποθέσεις μπορούν να γράφουν ισοδύναμα στην μορφή

$$H_0 : h(\theta) = 0 \quad \textit{έναντι της} \quad H_1 : h(\theta) \neq 0,$$

όπου $h : \Theta \subseteq \mathcal{R}^2 \to B \subseteq \mathcal{R}$, *με* $h(\theta_1, \theta_2) = \theta_1 - \theta_2$.

Ο παραμετρικός χώρος Θ *είναι το ημιεπίπεδο πάνω από τον άξονα* x, *ενώ η* H_0 *περιορίζει τον αρχικό παραμετρικό χώρο* Θ *στα σημεία της ευθείας* $\theta_1 - \theta_2 = 0$. *Έτσι, υπό την* $H_0 : \theta_1 = \theta_2$, *ο παραμετρικός χώρος* Θ *περιορίζεται στο σύνολο* $B \subseteq \mathcal{R}$ *με*

$$B = \{(\theta_1, \theta_2) \in \mathcal{R}^2 : \theta_1 \in \mathcal{R},\ \theta_2 > 0 \text{ και } \theta_1 = \theta_2\}.$$

Για να εκφραστεί η σύνθετη μηδενική υπόθεση $H_0 : h(\theta) = 0$ *ως απλή μηδενική υπόθεση ακολουθούμε τις ισοδυναμίες*

$$H_0 : h(\theta) = 0 \Leftrightarrow H_0 : \theta_1 = \theta_2 \Leftrightarrow H_0 : \theta = (\theta_1, \theta_2) = (\beta, \beta)$$

για $\beta \in B \subseteq \mathcal{R}$. *Η τελευταία αυτή ισοδυναμία οδηγεί στον ορισμό της συνάρτησης* $g : B \subseteq \mathcal{R} \longrightarrow \Theta \subseteq \mathcal{R}^2$ *με* $g(\beta) = (\beta, \beta)$ *για* $\beta \in B$ *και* $(\theta_1, \theta_2) \in \Theta \subseteq \mathcal{R}^2$, *οπότε η αρχική σύνθετη υπόθεση* $H_0 : h(\theta) = 0$ *μπορεί να γραφεί ισοδύναμα ως μια απλή μηδενική υπόθεση της μορφής* $H_0 : \theta = g(\beta)$.

Προκειμένου να διατυπωθεί το στατιστικό τεστ τοπικής ϕ-απόκλισης, απαιτούνται κάποιες τεχνικές προϋποθέσεις, παρόμοιες με εκείνες του καθολικού ελέγχου (βλέπε, για παράδειγμα, Pardo, 2006, σελ. 420). Πιο συγκεκριμένα, υποθέτουμε ότι,

(A1) Ο $M \times d$ πίνακας $H(\theta) = \nabla_\theta h(\theta)$ υπάρχει και τα στοιχεία του είναι συνεχείς συναρτήσεις του θ, με $rank(H(\theta)) = d$.

(A2) Ο $M \times (M - d)$ πίνακας $G(\beta) = \nabla_\beta g(\beta)$ υπάρχει, με $rank(G(\beta)) = M - d$ και $\beta \in B \subseteq \mathcal{R}^{M-d}$.

Στο παραπάνω στατιστικό πλαίσιο, έστω $\widehat{\theta}$ ο ΕΜΠ της θ στον παραμετρικό χώρο Θ και έστω επίσης $\widetilde{\theta}$ ο ΕΜΠ του θ υπό τη μηδενική υπόθεση $H_0 : h(\theta) = 0$, δηλαδή ο ΕΜΠ $\widetilde{\theta}$ ικανοποιεί τον περιορισμό $h(\widetilde{\theta}) = 0$. Το ακόλουθο λήμμα εξετάζει, προσδιορίζει τη σχέση μεταξύ των εκτιμητών μέγιστης πιθανοφάνειας της παράμετρου θ, $\widetilde{\theta}$ και $\widehat{\theta}$, αντίστοιχα.

Λήμμα 3.1. *Ισχύει η ακόλουθη ισότητα*

$$\sqrt{n}(\widetilde{\theta} - \theta) = \left(I_M - I_F^{-1}(\theta) B(\theta)\right) \sqrt{n}(\widehat{\theta} - \theta) + o_p(1),$$

όπου

$$B(\theta) = H(\theta) \left[H^t(\theta) I_F^{-1}(\theta) H(\theta)\right]^{-1} H^t(\theta).$$

Απόδειξη. Με βάση τις εξισώσεις (5.6.10), (5.6.23), και (5.6.24) των Sen and Singer (1993), έχουμε ότι

$$\sqrt{n}(\tilde{\theta} - \theta) = P(\theta) n^{-1/2} U_n(\theta), \tag{3.17}$$

όπου

$$P(\theta) = I_F^{-1}(\theta) \left(I_M - B(\theta) I_F^{-1}(\theta) \right)$$

και

$$U_n(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(x_i, \theta)$$

ή

$$\frac{1}{\sqrt{n}} U_n(\theta) = I_F(\theta) \sqrt{n}(\hat{\theta} - \theta) + o_p(1). \tag{3.18}$$

Συνδυάζοντας τις σχέσεις (3.17) και (3.18) προκύπτει το επιθυμητό αποτέλεσμα. ∎

Το επόμενο λήμμα παίζει σημαντικό ρόλο στον προσδιορισμό της ασυμπτωτικής κατανομής του στατιστικού έλεγχου τοπικής ϕ-απόκλισης για τον έλεγχο της σύνθετης μηδενικής υπόθεσης.

Λήμμα 3.2. *Ισχύει η ακόλουθη ισότητα*

$$\sqrt{n}(\widehat{\theta} - \widetilde{\theta})^t I_F^{\omega}(\widetilde{\theta}) \sqrt{n}(\widehat{\theta} - \widetilde{\theta}) = \sqrt{n}(\widehat{\theta} - \theta)^t A(\theta) \sqrt{n}(\widehat{\theta} - \theta) + o_p(1),$$

όπου

$$A(\theta) = B(\theta) I_F^{-1}(\theta) I_F^{\omega}(\theta) I_F^{-1}(\theta) B(\theta),$$

και

$$B(\theta) = H(\theta) \left[H^t(\theta) I_F^{-1}(\theta) H(\theta) \right]^{-1} H^t(\theta).$$

Απόδειξη. Από το Λήμμα 3.1, προκύπτει ότι

$$\sqrt{n}(\tilde{\theta} - \theta) - \sqrt{n}(\hat{\theta} - \theta) = -I_F^{-1}(\theta) B(\theta) \sqrt{n}(\hat{\theta} - \theta) + o_p(1),$$

ή ισοδύναμα ότι

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) = I_F^{-1}(\theta) B(\theta) \sqrt{n}(\hat{\theta} - \theta) + o_p(1).$$

Με βάση την παραπάνω σχέση, έχουμε

$$\sqrt{n}(\hat{\theta} - \tilde{\theta})^t I_F^{\omega}(\tilde{\theta}) \sqrt{n}(\hat{\theta} - \tilde{\theta}) = n(\hat{\theta} - \theta)^t A^*(\theta)(\hat{\theta} - \theta) + o_p(1),$$

με
$$A^*(\theta) = B(\theta) I_F^{-1}(\theta) I_F^w(\tilde{\theta}) I_F^{-1}(\theta) B(\theta).$$

Σύμφωνα με τους Morales et al. (1997, σελ. 150), για κάθε θ της μηδενικής υπόθεσης H_0,
$$I_F(\tilde{\theta}) \xrightarrow[n\to\infty]{P} I_F(\theta),$$
και ως εκ τούτου, για δοθέν $\omega \in \Theta^*$,
$$I_F^\omega(\tilde{\theta}) \xrightarrow[n\to\infty]{P} I_F^\omega(\theta).$$

Άρα
$$\sqrt{n}(\hat{\theta} - \tilde{\theta})^t I_F^\omega(\tilde{\theta}) \sqrt{n}(\hat{\theta} - \tilde{\theta}) = n(\hat{\theta} - \theta)^t A(\theta)(\hat{\theta} - \theta) + o_p(1),$$
με
$$A(\theta) = B(\theta) I_F^{-1}(\theta) I_F^w(\theta) I_F^{-1}(\theta) B(\theta),$$
όπου
$$B(\theta) = H(\theta) \left[H^t(\theta) I_F^{-1}(\theta) H(\theta) \right]^{-1} H^t(\theta),$$
υπό τη μηδενική υπόθεση H_0^ω. ∎

Στο επόμενο θεώρημα μελετάται η ασυμπωτική κατανομή του στατιστικού ελέγχου με βάση την τοπική ϕ-απόκλιση για τον έλεγχο σύνθετης μηδενικής υπόθεσης.

Θεώρημα 3.5. *Ας υποθέσουμε ότι η κυρτή συνάρτηση $\phi \in \Phi$ ικανοποιεί τις υποθέσεις Φ_1, Φ_2 και ότι πληρούνται οι συνθήκες ομαλότητας R_1, R_2, R_3. Τότε υπό τη σύνθετη μηδενική υπόθεση $H_0^\omega : h(\theta) = 0$, η οποία ικανοποιεί τις υποθέσεις (A1) και (A2), ισχύει ότι*
$$T_{\phi,n}^\omega(\widehat{\theta}, \widetilde{\theta}) = \frac{2n D_\phi^\omega(\widehat{\theta}, \widetilde{\theta})}{\phi''(1)} \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$
όπου $Z_1, ...Z_r$ είναι ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές, με
$$r = rank\left(I_F^{-1}(\theta) A(\theta) I_F^{-1}(\theta) \right),\ r \geq 1,$$
$$A(\theta) = B(\theta) I_F^{-1}(\theta) I_F^\omega(\theta) I_F^{-1}(\theta) B(\theta),$$
και $\beta_1, \beta_2, ..., \beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα
$$A(\theta) I_F^{-1}(\theta).$$

Απόδειξη. Το ανάπτυγμα Taylor της τοπικής ϕ-απόκλισης $D_\phi^\omega(\theta, \widetilde{\theta})$ γύρω από το $\widetilde{\theta}$, για $\theta = \hat{\theta}$ δίνει

$$\begin{aligned} D_\phi^\omega(\widehat{\theta}, \widetilde{\theta}) &= D_\phi^\omega(\widetilde{\theta}, \widetilde{\theta}) + \nabla_\theta^t \left(D_\phi^\omega(\theta, \widetilde{\theta}) \right)_{\theta=\widetilde{\theta}} (\widehat{\theta} - \widetilde{\theta}) \\ &+ \tfrac{1}{2}(\widehat{\theta} - \widetilde{\theta})^t \nabla_\theta^2 \left(D_\phi^\omega(\theta, \widetilde{\theta}) \right)_{\theta=\widetilde{\theta}} (\widehat{\theta} - \widetilde{\theta}) + o\left(\|\widehat{\theta} - \widetilde{\theta}\|^2 \right). \end{aligned} \tag{3.19}$$

Ακολουθώντας τα βήματα της απόδειξης του Θεωρήματος 3.1, έχουμε ότι

$$D_\phi^\omega(\widehat{\theta}, \widetilde{\theta}) = \frac{1}{2}(\widehat{\theta} - \widetilde{\theta})^t \phi''(1) I_F^\omega(\widetilde{\theta})(\widehat{\theta} - \widetilde{\theta}) + o\left(\|\widehat{\theta} - \widetilde{\theta}\|^2 \right),$$

ή ισοδύναμα

$$T_{\phi,n}^\omega(\widehat{\theta}, \widetilde{\theta}) = n(\widehat{\theta} - \widetilde{\theta})^t I_F^\omega(\widetilde{\theta})(\widehat{\theta} - \widetilde{\theta}) + no\left(\|\widehat{\theta} - \widetilde{\theta}\|^2 \right).$$

Από την άλλη πλευρά (βλέπε, Pardo, 2006, σελ. 63)

$$no\left(||\widehat{\theta} - \widetilde{\theta}|^2 \right) \leq no\left(||\widehat{\theta} - \theta||^2 \right) + no\left(||\widetilde{\theta} - \theta||^2 \right)$$

και συνεπώς

$$no\left(||\widehat{\theta} - \theta|| \right) = o_P(1) \text{ και } no\left(||\widetilde{\theta} - \theta|| \right) = o_P(1).$$

Άρα, $o\left(||\widehat{\theta} - \widetilde{\theta}||^2 \right) = o_P(1)$, δηλαδή

$$\widehat{\theta} - \widetilde{\theta} \xrightarrow[n\to\infty]{P} 0.$$

Επομένως, αρκεί να προσδιοριστεί η κατανομή της τετραγωνικής μορφής

$$\sqrt{n}(\widehat{\theta} - \widetilde{\theta})^t I_F^\omega(\widetilde{\theta}) \sqrt{n}(\widehat{\theta} - \widetilde{\theta}).$$

Από το Λήμμα 3.2, προκύπτει ότι

$$\sqrt{n}(\widehat{\theta} - \widetilde{\theta})^t I_F^\omega(\widetilde{\theta}) \sqrt{n}(\widehat{\theta} - \widetilde{\theta}) = \sqrt{n}(\widehat{\theta} - \theta)^t A(\theta) \sqrt{n}(\widehat{\theta} - \theta) + o_p(1), \tag{3.20}$$

με

$$A(\theta) = B(\theta) I_F^{-1}(\theta) I_F^\omega(\theta) I_F^{-1}(\theta) B(\theta),$$

όπου

$$B(\theta) = H(\theta) \left[H^t(\theta) I_F^{-1}(\theta) H(\theta) \right]^{-1} H^t(\theta).$$

Επιπλέον,

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow[n\to\infty]{L} N(0, I_F^{-1}(\theta)). \tag{3.21}$$

Το επιθυμητό αποτέλεσμα προκύπτει συνδυάζοντας τις σχέσεις (3.20) και (3.21) και λαμβάνοντας υπόψη το Πόρισμα 2.1 των Dik and de Gunst (1985). ∎

Ακολουθώντας τους Sen and Singer (1993, σελ. 239) το τοπικό στατιστικό έλεγχου Wald στην περίπτωση σύνθετης μηδενικής υπόθεσης διαμορφώνεται ως εξής,

$$W_n^{\omega,c} = nh(\widehat{\theta})^t \left(H(\widehat{\theta})^t I_F^{\omega}(\widehat{\theta})^{-1} H(\widehat{\theta}) \right)^{-1} h(\widehat{\theta}),$$

όπου $\widehat{\theta}$ είναι ο ΕΜΠ της παραμέτρου $\theta \in \Theta$ και $I_F^{\omega}(\widehat{\theta})^{-1}$ είναι ο αντίστροφος του τοπικού πίνακα πληροφορίας Fisher.

Στο θεώρημα που ακολουθεί προσδιορίζεται η ασυμπτωτική κατανομή της στατιστικής συνάρτησης $W_n^{\omega,c}$ υπό τη σύνθετη μηδενική υπόθεση $H_0^{\omega} : h(\theta) = 0$.

Θεώρημα 3.6. *Έστω ότι πληρούνται οι συνθήκες ομαλότητας R_1, R_2, R_3. Τότε υπό τη σύνθετη μηδενική υπόθεση $H_0^{\omega} : h(\theta) = 0$, για την οποία υποθέτουμε ότι ικανοποιούνται οι υποθέσεις (A1) και (A2), ισχύει ότι*

$$W_n^{\omega,c} \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

όπου $Z_1, ..., Z_r$ είναι ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές,

$$r = rank(H(\theta)^t I_F^{-1}(\theta) H(\theta) \left(H(\theta)^t I_F^{\omega}(\theta)^{-1} H(\theta) \right)^{-1} H(\theta)^t I_F^{-1}(\theta) H(\theta)), \quad r \geq 1,$$

και $\beta_1, \beta_2, ..., \beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα

$$\left(H(\theta)^T I_F^{\omega}(\theta)^{-1} H(\theta) \right)^{-1} H(\theta)^T I_F(\theta)^{-1} H(\theta).$$

Απόδειξη. Με βάση τα Θεωρήματα 5.4.1 (σελ. 220) και 5.6.3 (σελ. 240) των Sen and Singer (1993), είναι

$$\sqrt{n}(h(\widehat{\theta}) - h(\theta)) \xrightarrow[n\to\infty]{L} N(0, H(\theta)^t I_F^{-1}(\theta) H(\theta)).$$

Υπό τη μηδενική υπόθεση $H_0 : h(\theta) = 0$, έχουμε

$$\sqrt{n} h(\widehat{\theta}) \xrightarrow[n\to\infty]{L} N(0, H(\theta)^t I_F^{-1}(\theta) H(\theta))$$

και ως εκ τούτου, για δοθέν $\omega \in \Theta^*$, προκύπτει ότι

$$\left(H(\widehat{\theta})^t I_F^{\omega}(\widehat{\theta})^{-1} H(\widehat{\theta}) \right)^{-1} \xrightarrow[n\to\infty]{P} \left(H(\theta)^t I_F^{\omega}(\theta)^{-1} H(\theta) \right)^{-1}.$$

Κατά συνέπεια, η ασυμπτωτική κατανομή του στατιστικού $W_n^{\omega,c}$ συμπίπτει με την κατανομή της τετραγωνικής μορφής

$$nh(\widehat{\theta})^t \left(H(\theta)^t I_F^{\omega}(\theta)^{-1} H(\theta)\right)^{-1} h(\widehat{\theta}).$$

Λαμβάνοντας υπόψη το Πόρισμα 2.1 των Dik and de Gunst (1985), συνάγεται το επιθυμητό αποτέλεσμα, δηλαδή ότι

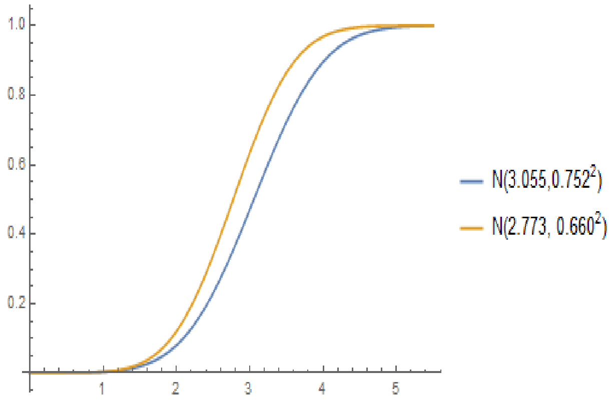
$$W_n^{\omega,c} \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2.$$

∎

3.3 Έλεγχος ομοιογένειας με βάση την τοπική ϕ-απόκλιση

Οι έλεγχοι ομοιογένειας είναι ιδιαίτερα σημαντικοί στη Στατιστική και το στατιστικό τους πλαίσιο περιγράφεται στη συνέχεια. Έστω δύο ανεξάρτητα τυχαία δείγματα, $X_1, ..., X_n$ και $Y_1, ..., Y_m$, από πληθυσμούς που περιγράφονται από τα μοντέλα f_{θ_1} και f_{θ_2}, αντίστοιχα. Ενδιαφέρει στην περίπτωση αυτή να ελεγχθεί η υπόθεση ότι τα δείγματα προέρχονται από τον ίδιο πληθυσμό, δηλαδή, η υπόθεση της ομοιογένειας που διατυπώνεται από τη μηδενική υπόθεση $H_0 : \theta_1 = \theta_2$. Όπως έχει ήδη αναφερθεί ένα μέτρο απόκλισης $D(f_{\theta_1}, f_{\theta_2})$, για το οποίο πληροίται η βασική ιδιότητα ότι $D(f_{\theta_1}, f_{\theta_2}) \geq 0$ και $D(f_{\theta_1}, f_{\theta_2}) = 0$ αν και μόνο αν τα μοντέλα f_{θ_1} και f_{θ_2} ταυτίζονται, ποσοτικοποιεί την απόκλιση μεταξύ των μοντέλων f_{θ_1} και f_{θ_2}. Κατά συνέπεια, μια εμπειρική εκδοχή του μπορεί να χρησιμεύσει ως το στατιστικό έλεγχου για τον έλεγχο της ομοιογένειας των f_{θ_1} και f_{θ_2}, σε ολόκληρο το πεδίο ορισμού των f_{θ_1} και f_{θ_2}. Έλεγχοι αυτής της φύσεως που λειτουργούν σε ολόκληρο το πεδίο ορισμού των κατανομών και βασίζονται σε αποκλίσεις έχουν καθιερωθεί στη βιβλιογραφία (βλέπε, Pardo, 2006, Κεφάλαιο 9).

Ωστόσο, οι έλεγχοι αυτοί λειτουργούν σε ολόκληρο το πεδίο ορισμού και προφανώς δεν μπορούν να ελέγξουν την ομοιογένεια των f_{θ_1} και f_{θ_2}, σε κάποια περιοχή του κοινού πεδίου ορισμού τους. Υπάρχουν παραδείγματα πραγματικών συνόλων δεδομένων, όπου μια στατιστική περιγραφική ανάλυση των δεδομένων δείχνει ότι μπορεί να υπάρχουν περιοχές του πεδίου ορισμού $\mathcal{X}$ των δεδομένων όπου παραβιάζεται η υπόθεση της ομοιογένειας, ακόμη και αν η ομοιογένεια είναι αποδεκτή σε ολόκληρο το πεδίο ορισμού $\mathcal{X}$, χρησιμοποιώντας



Σχήμα 3.1: Γραφική παράσταση των αθροιστικών συναρτήσεων κατανομών για τους κανονικούς πληθυσμούς $N(3.055, 0.752^2)$ και $N(2.773, 0.660^2)$.

οποιονδήποτε έλεγχο ομοιογένειας. Ένα τέτοιο παράδειγμα θα παρουσιαστεί στο τέλος του παρόντος κεφαλαίου.

Στο σημείο αυτό όμως, θα προσπαθήσουμε να αναδείξουμε την αναγκαιότητα εισαγωγής μιας μεθόδου ελέγχου της ομοιογένειας τοπικά μέσω ενός παραδείγματος και του αντίστοιχου σχήματος. Στο Σχήμα 3.1 παρατίθενται οι γραφικές παραστάσεις των αθροιστικών συναρτήσεων κατανομών για τους κανονικούς πληθυσμούς $N(3.055, 0.752^2)$ και $N(2.773, 0.660^2)$. Είναι σαφές από το Σχήμα 3.1 ότι οι δύο κατανομές είναι ανόμοιες σε μια περιοχή του κοινού πεδίου ορισμού τους, παρόλο που συμπίπτουν στις ουρές του κοινού πεδίου ορισμού τους. Ως εκ τούτου, ένας έλεγχος καθολικής ομοιογένειας, βασισμένος στην κλασική ϕ-απόκλιση $D_\phi(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ όπως δόθηκε στην (1.3), θα απορρίψει τη μηδενική υπόθεση της ομοιογένειας $H_0 : \theta_1 = \theta_2$ σε όλο το πεδίο ορισμού ενώ αυτή ισχύει, για παράδειγμα, στις ουρές των κατανομών. Ως εκ τούτου, φαίνεται ενδιαφέρον να αναπτυχθεί μια μέθοδος που θα εξετάζει όλο το πεδίο ορισμού $\mathcal{X}$ και θα ερευνά την εγκυρότητα της $H_0 : \theta_1 = \theta_2$, σε ορισμένες περιοχές της $\mathcal{X}$. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας ένα μέτρο απόκλισης μεταξύ των εμπειρικών μοντέλων $f_{\widehat{\theta}_1}$ και $f_{\widehat{\theta}_2}$, το οποίο επικεντρώνεται σε ένα συγκεκριμένο υποσύνολο του $\mathcal{X}$. Ένα κατάλληλο μέτρο απόκλισης που πληροί αυτές τις απαιτήσεις είναι η τοπική ϕ-απόκλιση που ορίστηκε στη (2.4) και η εμπειρική εκδοχή της δίνεται από τη σχέση,

$$D_\phi^\omega(\hat{\theta}_1, \hat{\theta}_2) = D_\phi^\omega(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2}), \tag{3.22}$$

όπου $\widehat{\theta}_1$ και $\widehat{\theta}_2$ είναι οι ΕΜΠ των θ_1 και θ_2, βάσει των τυχαίων δειγμάτων $X_1, ..., X_n$ και $Y_1, ..., Y_m$, αντίστοιχα. Είναι σαφές ότι μικρές τιμές της $D^{\omega}_{\phi}(\hat{\theta}_1, \hat{\theta}_2)$ είναι υπέρ της τοπικής ομοιογένειας των f_{θ_1} και f_{θ_2}, ενώ μεγάλες τιμές της $D^{\omega}_{\phi}(\hat{\theta}_1, \hat{\theta}_2)$ οδηγούν σε απόρριψη της μηδενικής υπόθεσης $H^{\omega}_0 : \theta_1 = \theta_2$, τοπικά. Για τον προσδιορισμό των τιμών εκείνων που οδηγούν σε απόρριψη της μηδενικής υπόθεσης είναι απαραίτητος ο προσδιορισμός της ασυμπτωτικής κατανομής της $D^{\omega}_{\phi}(\hat{\theta}_1, \hat{\theta}_2)$, υπό τη μηδενική υπόθεση.

Για τη μαθηματική διατύπωση του προβλήματος, θεωρούμε τον χώρο πιθανότητας $(\mathcal{X}, \mathcal{A}, P_\theta)$, $\theta \in \Theta \subseteq \mathcal{R}^M$, με $M \geq 1$, με P_θ μια οικογένεια μέτρων πιθανότητας και ένα σ-πεπερασμένο μέτρο μ στο χώρο $(\mathcal{X}, \mathcal{A})$, με $P_\theta \ll \mu$, $\theta \in \Theta$. Έστω, επιπλέον H_ω, $\omega \in \Theta^*$, ένα μέτρο πιθανότητας στον ίδιο χώρο $(\mathcal{X}, \mathcal{A})$, με $H_\omega \ll \mu$ και $f_{\theta_i} = \frac{dP_{\theta_i}}{d\mu}$, $f_\omega = \frac{dH_\omega}{d\mu}$, $i = 1, 2$ οι αντίστοιχες Radon-Nikodym παράγωγοι. Ας υποθέσουμε επίσης ότι η οικογένεια πυκνοτήτων $\{f_\theta : \theta \in \Theta \subseteq \mathcal{R}^M\}$ είναι ταυτοποιήσιμη (identifiable), δηλαδή ότι ισχύει: Αν $f_{\theta_1} = f_{\theta_2}$ τότε $\theta_1 = \theta_2$, για κάθε $\theta_1, \theta_2 \in \Theta$. Στο στατιστικό αυτό πλαίσιο και με βάση δύο ανεξάρτητα τυχαία δείγματα $X_1, ..., X_n$ από την f_{θ_1} και $Y_1, ..., Y_m$ από την f_{θ_2}, κατασκευάζεται ένας στατιστικός έλεγχος για την τοπική μηδενική υπόθεση:

$$H^{\omega}_0 : \theta_1 = \theta_2 \text{ έναντι της } H^{\omega}_\alpha : \theta_1 \neq \theta_2.$$

Η ασυμπτωτική κατανομή του στατιστικού ελέγχου $D^{\omega}_{\phi}(\hat{\theta}_1, \hat{\theta}_2)$ προσδιορίζεται στο ακόλουθο θεώρημα.

Θεώρημα 3.7. *Ας υποθέσουμε ότι η κυρτή συνάρτηση $\phi \in \Phi$ ικανοποιεί τις υποθέσεις Φ_1, Φ_2 και επιπλέον ότι πληρούνται οι συνθήκες ομαλότητας R_1, R_2, R_3. Υπό τη μηδενική υπόθεση $H^{\omega}_0 : \theta_1 = \theta_2$, ισχύει ότι*

$$T^{\omega}_{\phi,n,m}(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{2nm D^{\omega}_{\phi}(\widehat{\theta}_1, \widehat{\theta}_2)}{(m+n)\phi''(1)} \xrightarrow[n,m\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

όπου $Z_1, ...Z_r$, r ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές, με

$$r = rank(I_F^{-1}(\theta_1) I_F^{\omega}(\theta_1) I_F^{-1}(\theta_1))$$

και $\beta_1, \beta_2, ..., \beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα

$$I_F^{\omega}(\theta_1) I_F^{-1}(\theta_1),$$

όπου $I_F(\theta)$ συμβολίζει τον πίνακα πληροφορίας Fisher και $I_F^{\omega}(\theta)$ συμβολίζει τον τοπικό πίνακα πληροφορίας Fisher.

Απόδειξη. Ακολουθώντας τα βήματα παρόμοιας αποδείξης του Pardo (2006, σελ. 442-443), και λαμβάνοντας υπόψη ότι $\phi(1) = 0$ και $\phi'(1) = 0$, προκύπτει ότι

$$\begin{aligned} D_\phi^\omega(\hat\theta_1, \hat\theta_2) &= \frac{1}{2}(\hat\theta_1 - \theta_1)^t \nabla^2_{\theta_1} \left(D_\phi^\omega(\theta_1, \theta_2) \right)_{\theta_1=\theta_2} (\hat\theta_1 - \theta_1) \\ &+ (\hat\theta_1 - \theta_1)^t \nabla^2_{\theta_2\theta_1} \left(D_\phi^\omega(\theta_1, \theta_2) \right)_{\theta_1=\theta_2} (\hat\theta_2 - \theta_2) \\ &+ \frac{1}{2}(\hat\theta_2 - \theta_2)^t \nabla^2_{\theta_2} \left(D_\phi^\omega(\theta_1, \theta_2) \right)_{\theta_1=\theta_2} (\hat\theta_2 - \theta_2) \\ &+ o\left(\|\hat\theta_1 - \theta_1\|^2 \right) + o\left(\|\hat\theta_2 - \theta_2\|^2 \right). \end{aligned}$$

Είναι

$$\frac{\partial D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{1i}} = \int_{\mathcal{X}} h_\omega(x) \phi'(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}) \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} d\mu(x), \ i = 1, ..., M,$$

και

$$\begin{aligned} \frac{\partial^2 D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} &= \int_{\mathcal{X}} h_\omega(x) \phi''(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}) \frac{1}{f_{\theta_2}(x)} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1j}} d\mu(x) \\ &+ \int_{\mathcal{X}} h_\omega(x) \phi'(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}) \frac{\partial^2 f_{\theta_1}(x)}{\partial \theta_{1i} \partial \theta_{1j}} d\mu(x), \ i, j = 1, ..., M. \end{aligned}$$

Τότε

$$\begin{aligned} \left(\frac{\partial^2 D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} \right)_{\theta_1=\theta_2} &= \phi''(1) \int_{\mathcal{X}} h_\omega(x) f_{\theta_1}(x) \frac{\partial \log f_{\theta_1}(x)}{\partial \theta_{1i}} \frac{\partial \log f_{\theta_1}(x)}{\partial \theta_{1j}} d\mu(x) \\ &= \phi''(1) \left(I_F^\omega(\theta_1) \right)_{i,j} \end{aligned}$$

και

$$\left(\frac{\partial^2 D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{2i} \partial \theta_{2j}} \right)_{\theta_1=\theta_2} = \left(\frac{\partial^2 D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} \right)_{\theta_1=\theta_2},$$

$$\left(\frac{\partial^2 D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{2j}} \right)_{\theta_1=\theta_2} = -\left(\frac{\partial^2 D_\phi^\omega(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} \right)_{\theta_1=\theta_2}.$$

Έτσι,

$$\begin{aligned}
\frac{2D_\phi^\omega(\hat\theta_1,\hat\theta_2)}{\phi''(1)} &= (\hat\theta_1-\theta_1)^t I_F^\omega(\theta_1)(\hat\theta_1-\theta_1)\\
&- (\hat\theta_1-\theta_1)^t I_F^\omega(\theta_1)(\hat\theta_2-\theta_2)\\
&+ (\hat\theta_2-\theta_2)^t I_F^\omega(\theta_1)(\hat\theta_2-\theta_2)\\
&+ o\left(\|\hat\theta_1-\theta_1\|^2\right)+o\left(\|\hat\theta_2-\theta_2\|^2\right)\\
&= (\hat\theta_1-\hat\theta_2)^t I_F^\omega(\theta_1)(\hat\theta_1-\hat\theta_2)\\
&+ o\left(\|\hat\theta_1-\theta_1\|^2\right)+o\left(\|\hat\theta_2-\theta_2\|^2\right).
\end{aligned}$$

Από την άλλη πλευρά, για $\lambda = \lim\limits_{n,m\to\infty}\frac{m}{m+n}$, είναι

$$\sqrt{n}(\hat\theta_1-\theta_1)\xrightarrow[n\to\infty]{L} N(0,I_F^{-1}(\theta_1))$$

και

$$\sqrt{m}(\hat\theta_2-\theta_2)\xrightarrow[n\to\infty]{L} N(0,I_F^{-1}(\theta_2)).$$

Τότε

$$\sqrt{\frac{mn}{m+n}}(\hat\theta_1-\theta_1)\xrightarrow[n\to\infty]{L} N(0,\lambda I_F^{-1}(\theta_1))$$

και

$$\sqrt{\frac{mn}{m+n}}(\hat\theta_2-\theta_2)\xrightarrow[n\to\infty]{L} N(0,(1-\lambda) I_F^{-1}(\theta_2)).$$

Έτσι, υπό τη μηδενική υπόθεση $H_0:\theta_1=\theta_2$, παίρνουμε

$$\sqrt{\frac{mn}{m+n}}(\hat\theta_1-\hat\theta_2)\xrightarrow[n\to\infty]{L} N(0,I_F^{-1}(\theta_1)).$$

Επιπλέον, είναι

$$o\left(\|\hat{\theta_1}-\theta_1\|^2\right)=o_p(n^{-1}) \text{ και } \left(\|\hat{\theta_2}-\theta_2\|^2\right)=o_p(m^{-1}).$$

Το συμπέρασμα του θεωρήματος προκύπτει άμεσα από το Πόρισμα 2.1 των Dik and Gunst (1985), δηλαδή η στατιστική συνάρτηση

$$\frac{2mnD_\phi^\omega(\hat\theta_1,\hat\theta_2)}{(m+n)\phi''(1)},$$

έχει την ίδια κατανομή με εκείνη της ποσότητας $\sum_{i=1}^r \beta_i Z_i^2$, όπου $Z_1,...Z_r$ είναι ανεξάρτητες τυπικές τυχαίες κανονικές μεταβλητές με $r = rank(I_F^{-1}(\theta_1)I_F^\omega(\theta_1)I_F^{-1}(\theta_1))$ και $\beta_1,\beta_2,...,\beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα $I_F^\omega(\theta_1)I_F^{-1}(\theta_1)$. ∎

3.4 Μελέτη προσομοίωσης

Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα μιας μελέτης προσομοίωσης Monte Carlo, μέσω της οποίας θα αξιολογηθεί η απόδοση των τοπικών κριτηρίων ελέγχου καλής προσαρμογής και ομοιογένειας των προηγούμενων εδαφίων. Η απόδοση ενός τεστ αξιολογείται από τη μελέτη του σφαλμάτος τύπου I και της ισχύος του. Σ' αυτό το πλαίσιο, αρχικά θα μελετηθούν τα ποσοστά σφάλματος τύπου I για τον τοπικό έλεγχο καλής προσαρμογής και για τον τοπικό έλεγχο ομοιογένειας και στη συνέχεια η εμπειρική ισχύς αυτών των ελέγχων. Επισημαίνεται ότι για τον προσδιορισμό των κρίσιμων σημείων χρησιμοποιείται η προσέγγιση Monte Carlo της Παρατήρησης 3.1. Τέλος, τα αποτελέσματα της προσομοίωσης θα βασιστούν στην αξιοποίηση της τοπικής απόκλισης των Cressie and Read (1984). Σημειώνεται ότι αναλυτικές εκφράσεις της τοπικής απόκλισης Cressie and Read (1984) έχουν προσδιοριστεί στο Κεφάλαιο 2, στην Πρόταση 2.4, όταν οι κατανομές στις οποίες στηρίζεται η τοπική ϕ-απόκλιση είναι διδιάστατες κανονικές κατανομές και ο πυρήνας είναι επίσης διδιάστατη κανονική κατανομή.

3.4.1 Μελέτη προσομοίωσης Monte Carlo για το σφάλμα τύπου I

Στο εδάφιο αυτό, διενεργείται μια μελέτη προσομοίωσης Monte Carlo και παρουσιάζονται τα εμπειρικά αποτελέσματα που αφορούν το ποσοστό σφάλματος τύπου I για τους τοπικούς ελέγχους καλής προσαρμογής και ομοιογένειας.

Ειδικότερα, η Monte Carlo μελέτη για τα ποσοστά σφάλματος τύπου I του ελέγχου που βασίζεται στη στατιστική συνάρτηση $T^{\omega}_{\phi_\lambda}$, που ορίστηκε στο Θεώρημα 3.1 στη σχέση (3.9), ακολουθεί τα εξής βήματα:

- Παράγονται $j = 1, ..., L$ δείγματα μεγέθους n, p-διάστατων παρατηρήσεων $X_1^{(j)}, ..., X_n^{(j)}$ για διάφορες τιμές του n (όπως περιγράφεται λεπτομερώς παρακάτω), όπου L είναι ο προκαθορισμένος αριθμός επαναλήψεων. Τα παραχθέντα δείγματα είναι τέτοια ώστε να ικανοποιείται η μηδενική υπόθεση H_0^{ω}.
- Υπολογίζονται οι τιμές του δειγματικού μέσου διανύσματος και του δειγματικού πίνακα διακυμάνσεων-συνδιακυμάνσεων, από τις σχέσεις

$$\bar{X}_{(j)} = \frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}$$

και

$$S_j = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^{(j)} - \bar{X}_{(j)})(X_i^{(j)} - \bar{X}_{(j)})^t,$$

$j = 1, ..., L$, αντίστοιχα.

- Υπολογίζεται η τιμή του τοπικού στατιστικού ελέγχου $T_{\phi_\lambda}^{\omega,j}$, $j = 1, ..., L$, για διάφορες τιμές της παραμέτρου λ.
- Απορρίπτεται η μηδενική υπόθεση H_0^ω με βάση το j-οστό δείγμα αν $T_{\phi_\lambda}^{\omega,j} > c_a$, $j = 1, ..., L$, όπου α είναι το προκαθορισμένο επίπεδο σημαντικότητας.
- Μια προσέγγιση για το ποσοστό σφάλματος τύπου I προκύπτει από τον υπολογισμό του κλάσματος: αριθμός φορών που η μηδενική υπόθεση απορρίπτεται προς το πλήθος L των επαναλήψεων.

Θεωρούμε τις ακόλουθες προσομοιώσεις:

1. Σύμφωνα με τα θεωρητικά αποτελέσματα της Ενότητας 3.1, θεωρούμε το στατιστικό ελέγχου $T_{\phi_\lambda,n}^{\omega}(\hat{\theta}, \theta_0)$, για $\lambda = 2/3, 1, -0.5, 2, \ -2$. Στη μελέτη προσομοίωσης, $L = 10000$ σύνολα δεδομένων υπό τη μηδενική υπόθεση H_0^ω: $\theta = \theta_0$ με μέγεθος δειγμάτων $n = 25$, 50, 100, 200, 500, 1000 παράγονται. Το πληθυσμιακό μοντέλο και ο πυρήνας είναι διδιάστατες κανονικές κατανομές με παραμέτρους

$$\theta_0 = (\mu_0, \Sigma_0) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \right),$$

$$\omega = (\mu_\omega, \Sigma_\omega) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

   αντίστοιχα. Τα ποσοστά σφάλματος τύπου I παρουσιάζονται στον Πίνακα 3.1. Το κρίσιμο σημείο ισούται με $c_\alpha = 0.2182$ για προκαθορισμένο επίπεδο σημαντικότητας $\alpha = 0.05$.

2. Σύμφωνα με τα θεωρητικά αποτελέσματα της Ενότητας 3.3, θεωρούμε το στατιστικό ελέγχου $T_{\phi_\lambda,n,m}^{\omega}(\hat{\theta}_1, \hat{\theta}_2)$, για $\lambda = 2/3, 1, -0.5, 2, \ -2$. Στη μελέτη αυτή, παράγονται $L = 10000$ σύνολα δεδομένων υπό τη μηδενική υπόθεση H_0^ω: $\theta_1 = \theta_2 = \theta_0$, με μέγεθος δειγμάτων $(n, m) = (25, 25)$, $(50, 50)$, $(100, 100)$, $(200, 200)$, $(500, 500)$,

Τοπικός έλεγχος καλής προσαρμογής: Ποσοστά σφάλματος τύπου I

n	$\lambda = 2/3$	1	-0.5	2	-2
25	0.1141	0.1215	0.1245	0.1453	0.1924
50	0.0799	0.0843	0.0826	0.0977	0.1250
100	0.0621	0.0675	0.0705	0.0766	0.0802
200	0.0606	0.0602	0.0608	0.0590	0.0677
500	0.0509	0.0534	0.0479	0.0553	0.0573
1000	0.0488	0.0507	0.0496	0.0518	0.0569

Πίνακας 3.1: Ποσοστά σφάλματος τύπου I για τον τοπικό έλεγχο καλής προσαρμογής με βάση την τοπική απόκλιση Cressie and Read και κρίσιμο σημείο $c_\alpha = 0.2182$.

$(1000, 1000)$. Το πληθυσμιακό μοντέλο και ο πυρήνας είναι διδιάστατες κανονικές κατανομές με παραμέτρους

$$\theta_1 = (\mu_1, \Sigma_1) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix} \right),$$

$$\theta_2 = (\mu_2, \Sigma_2) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix} \right)$$

και

$$\omega = (\mu_\omega, \Sigma_\omega) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

αντίστοιχα. Τα ποσοστά σφάλματος τύπου I παρουσιάζονται στον Πίνακα 3.2. Το κρίσιμο σημείο ισούται με $c_\alpha = 0.1326$ για προκαθορισμένο επίπεδο σημαντικότητας $\alpha = 0.05$.

Στο σημείο αυτό επισημαίνεται ότι η επιλογή του πυρήνα και των παραμέτρων $\omega = (\mu_\omega, \Sigma_\omega)$, έγινε με γνώμονα να αξιολογηθεί η συμπεριφορά του ελέγχου σε μια περιοχή κοντά στο μοντέλο υπό τη μηδενική υπόθεση. Παρατηρούμε ότι όσο το μέγεθος του δείγματος αυξάνει και με προκαθορισμένη την τιμή του κρίσιμου σημείου c_α, το ποσοστό σφάλματος τύπου I προσεγγίζει το 0.05, όπως προσδοκάται και στις δύο περιπτώσεις, ανεξάρτητα από την επιλογή της παραμέτρου λ για τον υπολογισμό της τοπικής απόκλισης των Cressie and Read.

Τοπικός έλεγχος ομοιογένειας: Ποσοστά σφάλματος τύπου I					
n, m	$\lambda = 2/3$	1	-0.5	2	-2
$25, 25$	0.1374	0.1553	0.1041	0.2071	0.1530
$50, 50$	0.0934	0.0983	0.0720	0.1347	0.0990
$100, 100$	0.0631	0.0683	0.0616	0.0882	0.0725
$200, 200$	0.0569	0.0579	0.0554	0.0695	0.0582
$500, 500$	0.0564	0.0549	0.0561	0.0627	0.0546
$1000, 1000$	0.0485	0.0517	0.0536	0.0546	0.0513

Πίνακας 3.2: Ποσοστά σφάλματος τύπου I για τον τοπικό έλεγχο ομογένειας με βάση την τοπική απόκλιση Cressie and Read και κρίσιμο σημείο $c_\alpha = 0.1326$.

3.4.2 Μελέτη προσομοίωσης Monte Carlo για την ισχύ

Η ισχύς ενός ελέγχου στατιστικών υποθέσεων είναι ένα ακόμη χαρακτηριστικό, το σημαντικότερο ίσως, για την αξιολόγηση της απόδοσης του προτεινόμενου ελέγχου. Σε αυτήν την υποενότητα παρουσιάζονται τα αποτελέσματα μιας μελέτης προσομοίωσης Monte Carlo της εμπειρικής ισχύος των τοπικών ελέγχων καλής προσαρμογής και των τοπικών ελέγχων ομοιογένειας. Παρόμοια βήματα με αυτά της υποενότητας 3.4.1 ακολουθούνται με τη μοναδική διαφορά να εστιάζεται στο πρώτο βήμα της διαδικασίας, όπου τα δείγματα που παράγονται είναι υπό την εναλλακτική υπόθεση H_a^ω. Οι λεπτομέρειες της μελέτης προσομοίωσης είναι οι ακόλουθες:

1. Σύμφωνα με τα θεωρητικά αποτελέσματα της Ενότητας 3.1, η στατιστική συνάρτηση ελέγχου είναι η $T^\omega_{\phi_\lambda,n}(\hat\theta,\theta_0)$, για $\lambda = 2/3, 1, -0.5, 2, -2$. Στη μελέτη προσομοίωσης παράγονται 10000 σύνολα δεδομένων υπό την H_a^ω: $\theta = \theta^*$ $(\theta^* \neq \theta_0)$, με δειγματικά μεγέθη $n = 25$, 50, 100, 200, 500, 1000. Τα πληθυσμιακά μοντέλα και ο πυρήνας είναι διδιάστατες κανονικές κατανομές με παραμέτρους

$$\theta_0 = (\mu_0, \Sigma_0) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \right),$$

$$\theta = \theta^* = (\mu^*, \Sigma^*) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & -0.5 \\ -0.5 & 4 \end{pmatrix} \right)$$

Τοπικός έλεγχος καλής προσαρμογής: Ισχύς

n	$\lambda = 2/3$	1	-0.5	2	-2
25	0.1179	0.1432	0.0972	0.2566	0.1438
50	0.1526	0.1784	0.1130	0.3048	0.1331
100	0.3401	0.3698	0.2600	0.4838	0.2370
200	0.7292	0.7511	0.6594	0.8066	0.5944
500	0.9977	0.9981	0.9962	0.9986	0.9933
1000	1	1	1	1	1

Πίνακας 3.3: Η ισχύς του τοπικού ελέγχου καλής προσαρμογής με βάση την τοπική απόκλιση Cressie and Read και κρίσιμο σημείο $c_\alpha = 0.2182$.

και

$$\omega = (\mu_\omega, \Sigma_\omega) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

αντίστοιχα. Στον Πίνακα 3.3 φαίνονται οι τιμές της ισχύος του ελέγχου. Το κρίσιμο σημείο ισούται με $c_\alpha = 0.2182$ για προκαθορισμένο επίπεδο σημαντικότητας $\alpha = 0.05$.

2. Σύμφωνα με τα θεωρητικά αποτελέσματα της Ενότητας 3.3, η στατιστική συνάρτηση ελέγχου είναι η $T^{\omega}_{\phi_\lambda,n,m}(\hat{\theta}_1, \hat{\theta}_2)$, για $\lambda = 2/3, 1, -0.5, 2, -2$. Παράγονται 10000 σύνολα δεδομένων υπό την εναλλακτική H^{ω}_a: $\theta_1 \neq \theta_2$ με δειγματικά μεγέθη $(n, m) = (25, 25)$, $(50, 50)$, $(100, 100)$, $(200, 200)$, $(500, 500), (1000, 1000)$. Τα πληθυσμιακά μοντέλα και ο πυρήνας είναι διδιάστατες κανονικές κατανομές με παραμέτρους

$$\theta_1 = (\mu_1, \Sigma_1) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix} \right),$$

$$\theta_2 = (\mu_2, \Sigma_2) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix} \right)$$

και

$$\omega = (\mu_\omega, \Sigma_\omega) = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

Τοπικός έλεγχος ομοιογένειας: Ισχύς

n, m	$\lambda = 2/3$	1	-0.5	2	-2
$25, 25$	0.4207	0.4533	0.3769	0.5290	0.5409
$50, 50$	0.6870	0.6999	0.6844	0.7486	0.8317
$100, 100$	0.9613	0.9599	0.9717	0.9657	0.9926
$200, 200$	1	1	1	1	1
$500, 500$	1	1	1	1	1
$1000, 1000$	1	1	1	1	1

Πίνακας 3.4: Η ισχύς του τοπικού ελέγχου ομοιογένειας με βάση την τοπική απόκλιση Cressie and Read και κρίσιμο σημείο $c_\alpha = 0.1326$.

αντίστοιχα. Στον Πίνακα 3.4 παρουσιάζονται οι τιμές της ισχύος του ελέγχου. Το κρίσιμο σημείο ισούται με $c_\alpha = 0.1326$ για προκαθορισμένο επίπεδο σημαντικότητας $\alpha = 0.05$.

Παρατηρούμε ότι όσο αυξάνεται το μέγεθος του δείγματος, η ισχύς αυξάνεται μέχρι την τιμή ένα, όπως προσδοκάται, και για τους δύο ελέγχους, καλής προσαρμογής και ομοιογένειας.

3.5 Εφαρμογές

Στην ενότητα αυτή αναλύονται δύο σύνολα πραγματικών δεδομένων της βιβλιογραφίας, αξιοποιώντας τα θεωρητικά αποτελέσματα του κεφαλαίου αυτού. Διευκρινίζεται έτσι η εφαρμογή των θεωρητικών αποτελεσμάτων στην πράξη και επεξηγείται η χρησιμότητα των στατιστικών τεστ που παρουσιάστηκαν, σε τοπικό επίπεδο.

3.5.1 Τοπικός έλεγχος καλής προσαρμογής: Δεδομένα Kalbfleisch and Prentice (1980)

Αρχικά, εφαρμόζεται ο τοπικός έλεγχος καλής προσαρμογής σε ένα πραγματικό σύνολο δεδομένων, που εμφανίστηκε στους Kalbfleisch and Prentice (1980). Στο σύνολο αυτό

Πληθυσμός	n	Τεστ ($p-$τιμές)	Δειγματικός Μέσος
		Kolmogorov-Smirnov	
Π	100	0.076	$\overline{X} = 1.638$

Πίνακας 3.5: Τεστ καλής προσαρμογής για εκθετική κατανομή και η δειγματική μέση τιμή του χρόνου επιβίωσης των 100 ασθενών.

καταγράφονται οι χρόνοι επιβίωσης σε έτη, 100 ασθενών, οι οποίοι έχουν εκτεθεί σε ακτινοβολία από τη στιγμή της διάγνωσης καρκίνου του φάρυγγα. Ο Πίνακας 3.5 παρουσιάζει το κλασικό τεστ καλής προσαρμογής για μια εκθετική κατανομή για να επιβεβαιωθεί ότι τα δεδομένα αυτά μπορεί να θεωρηθεί ότι προέρχονται από εκθετική κατανομή.

$\omega = (\mu_\omega, \sigma_\omega^2)$	R_n^ω	$c_{0.05}$	W_n^ω	$c_{0.05}$	$T^\omega_{\phi_{2/3},n}(\hat{\theta}, \theta_0)$	$c_{0.05}$	Απόφαση
$(0, 0.1^2)$	0.0548	7.3738	0.0151	2.0012	0.0153	2.0012	Αποδ. H_0
$(0.5, 0.1^2)$	0.1286	17.300	0.0063	0.8530	0.0065	0.8530	Αποδ. H_0
$(1, 0.1^2)$	0.5201	69.943	0.0015	0.2110	0.0016	0.2110	Αποδ. H_0
$(2, 0.1^2)$	3.7412	503.13	0.0003	0.0293	0.0002	0.0293	Αποδ. H_0
Καθολικό τεστ	0.0285	3.841	0.0285	3.841	0.0286	3.841	Αποδ. H_0

Πίνακας 3.6: Οι τιμές των τοπικών στατιστικών ελέγχων Rao, Wald και ϕ-απόκλισης για μέγεθος δείγματος $n = 100$ και $\theta_0 = 0.6$. Κάθε γραμμή δείχνει έναν διαφορετικό τοπικό έλεγχο καλής προσαρμογής για διάφορες τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$ του πυρήνα. Η τελευταία γραμμή περιλαμβάνει την τιμή των καθολικών στατιστικών ελέγχων Rao, Wald και ϕ-απόκλισης. Το πληθυσμιακό μοντέλο είναι μια μονοδιάστατη εκθετική κατανομή με την εκτιμώμενη παράμετρο θ να ισούται με $\hat{\theta} = 0.61$ έτη και με $c_{0.05}$ τα αντίστοιχα κρίσιμα σημεία.

Εφαρμόζονται οι τοπικοί έλεγχοι Rao, Wald και ϕ-απόκλισης για να ελεγχθεί η καλή προσαρμογή των δεδομένων στο εκθετικό μοντέλο με παράμετρο $\theta_0 = 0.6$, μια τιμή πολύ κοντινή στην αντίστροφη τιμή της δειγματικής μέσης τιμής. Στον Πίνακα 3.6 παρουσιάζονται οι τιμές των τοπικών στατιστικών ελέγχων R_n^ω, W_n^ω και $T^\omega_{\phi_{2/3},n}(\widehat{\theta}, \theta_0)$, όπου σε κάθε γραμμή παρουσιάζεται ένας διαφορετικός τοπικός έλεγχος για ποικίλες τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$ του περικομμένου στον θετικό ημιάξονα κανονικού πυρήνα. Οι τιμές των αντίστοιχων καθολικών στατιστικών είναι $R_n = 0.0285$, $W_n = 0.0285$ και

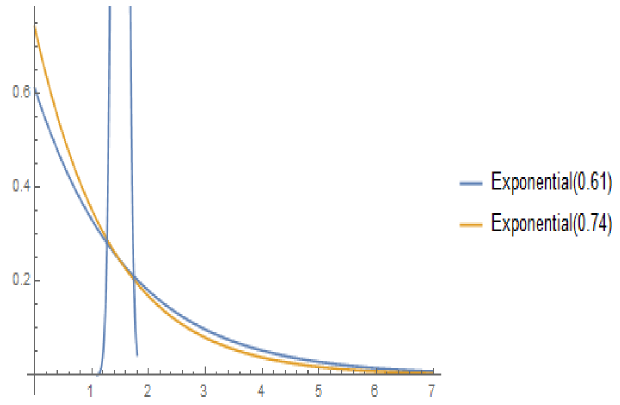
$\omega = (\mu_\omega, \sigma_\omega^2)$	Στατιστικό $Wald$ W_n^ω	Κρίσιμα σημεία c_α για α=0.05	Απόφαση
$(0.5, 0.1^2)$	1.0005	0.8123	Απόρριψη H_0
$(1.5, 0.1^2)$	0.0127	0.0156	Αποδοχή H_0
$(3, 0.1^2)$	0.3044	0.4634	Αποδοχή H_0
$(4, 0.1^2)$	0.4975	0.5694	Αποδοχή H_0
$W_n = 4.5151$		3.841	Απόρριψη H_0

Πίνακας 3.7: Οι τιμές του τοπικού στατιστικού ελέγχου Wald W_n^ω για μέγεθος δείγματος $n = 100$ και $\theta_0 = 0.74$. Κάθε γραμμή δείχνει έναν διαφορετικό τοπικό έλεγχο καλής προσαρμογής για διάφορες τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$ του πυρήνα. Η τελευταία γραμμή περιλαμβάνει την τιμή του καθολικού στατιστικού ελέγχου Wald W_n. Το πληθυσμιακό μοντέλο είναι μια μονοδιάστατη εκθετική κατανομή με εκτιμώμενη παράμετρο θ να ισούται με $\hat{\theta} = 0.61$ έτη (βλέπε, Σχήμα 3.2).

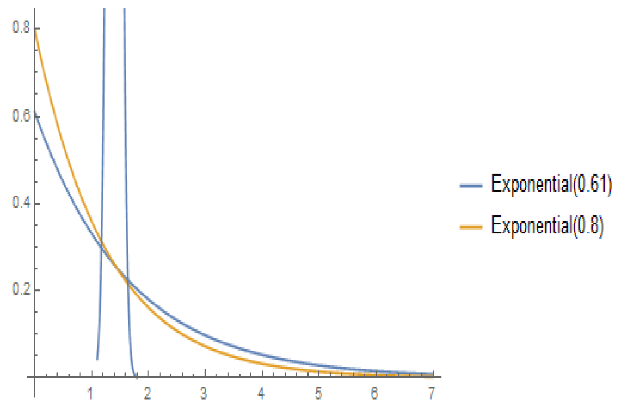
$T_{\phi_{2/3},n}(\widehat{\theta}, \theta_0) = 0.0286$ και γι' αυτό γίνεται γενικά δεκτή η υπόθεση της καλής προσαρμογής σε επίπεδο $\alpha = 0.05$. Οι τοπικοί έλεγχοι δεν μπορούν επίσης να απορρίψουν τη μηδενική υπόθεση $H_0^\omega : \theta = 0.6$.

Στη συνέχεια, εφαρμόζεται ο τοπικός έλεγχος Wald για να ελεγχθεί η καλή προσαρμογή των δεδομένων στο εκθετικό μοντέλο με παράμετρο $\theta_0 = 0.74$, μια τιμή που δεν μπορεί να θεωρηθεί κοντινή στην αντίστροφη της δειγματικής μέσης τιμής. Στον Πίνακα 3.7 δίνονται οι τιμές του τοπικού στατιστικού ελέγχου W_n^ω, όπου σε κάθε γραμμή παρουσιάζεται το αποτέλεσμα ενός διαφορετικού τοπικού ελέγχου, που προκύπτει για διαφορετικές τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$ του περικομμένου στον θετικό ημιάξονα κανονικού πυρήνα. Η τιμή του καθολικού στατιστικού είναι $W_n = 4.5151$, και γι' αυτό η υπόθεση της καλής προσαρμογής απορρίπτεται γενικά σε επίπεδο $\alpha = 0.05$. Ο τοπικός έλεγχος με την παράμετρο $\omega = (0.5, 0.1^2)$ απορρίπτει τη μηδενική υπόθεση $H_0^\omega : \theta = 0.74$, ενώ για τις άλλες τιμές της παραμέτρου ω, οι τοπικοί έλεγχοι δείχνουν καλή προσαρμογή (δεν μπορεί να απορριφθεί η $H_0^\omega : \theta = 0.74$). Από το Σχήμα 3.2 παίρνουμε μία ένδειξη της συμπεριφοράς των δύο εκθετικών μοντέλων σε τοπικό επίπεδο.

Το Σχήμα 3.3 απεικονίζει την εκθετική κατανομή υπό τη μηδενική υπόθεση με παράμετρο $\theta_0 = 0.8$ και την εκθετική κατανομή χρησιμοποιώντας τον ΕΜΠ $\widehat{\theta} = 1/\overline{X} = 0.61$. Απεικονίζεται επίσης ένας περικομμένος κανονικός πυρήνας με παράμετρο $\omega = (1.4, 0.1^2)$.



Σχήμα 3.2: Γράφημα των εκθετικών κατανομών με παραμέτρους $\theta_0 = 0.74$, $\widehat{\theta} = 0.61$ και του περικομμένου κανονικού πυρήνα με παράμετρο $\omega = (1.5, 0.1^2)$.



Σχήμα 3.3: Γράφημα των εκθετικών κατανομών με παραμέτρους $\theta_0 = 0.8$, $\widehat{\theta} = 0.61$ και του περικομμένου κανονικού πυρήνα με παράμετρο $\omega = (1.4, 0.1^2)$.

$\omega = (\mu_\omega, \sigma_\omega^2)$	Στατιστικό ελέγχου $T^{\omega}_{\phi_{2/3},n}(\hat{\theta}, \theta_0)$	Κρίσιμα σημεία :c_α για α=0.05	Απόφαση
$(0, 0.1^2)$	3.9428	2.5484	Απόρριψη H_0
$(0.5, 0.1^2)$	1.4591	0.7732	Απόρριψη H_0
$(1.4, 0.1^2)$	0.0104	0.0194	Αποδοχή H_0
$(4, 0.1^2)$	1.2139	0.6055	Απόρριψη H_0
$T_{\phi_{2/3},n}(\hat{\theta}, \theta_0) = 9.5619$		3.841	Απόρριψη H_0

Πίνακας 3.8: Οι τιμές του τοπικού στατιστικού ελέγχου $T^{\omega}_{\phi_{2/3},n}(\hat{\theta}, \theta_0)$, για μέγεθος δείγματος $n = 100$ και $\theta_0 = 0.8$. Κάθε γραμμή δείχνει έναν διαφορετικό τοπικό έλεγχο καλής προσαρμογής για διάφορες τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$ του πυρήνα. Η τελευταία γραμμή περιλαμβάνει την τιμή του καθολικού στατιστικού ελέγχου ϕ-απόκλισης $T_{\phi_{2/3},n}(\hat{\theta}, \theta_0)$. Το πληθυσμιακό μοντέλο είναι μια μονοδιάστατη εκθετική κατανομή με την εκτιμώμενη παράμετρο θ να ισούται με $\hat{\theta} = 0.61$ (βλέπε, Σχήμα 3.3).

Η τιμή της παραμέτρου θ_0 επιλέχθηκε, διότι είναι κάπως μακριά από την τιμή του ΕΜΠ και προκειμένου να μελετηθεί η επίδραση μιας τέτοιας επιλογής στη διαμόρφωση των δεδομένων.

Το Σχήμα 3.3 υποδεικνύει ότι τα δύο εκθετικά μοντέλα είναι παρόμοια στην περιοχή του πυρήνα και γι' αυτό ισχυριζόμαστε ότι οι δύο πυκνότητες δε διαφέρουν σημαντικά στατιστικά σε τοπικό επίπεδο. Προκειμένου να αξιολογηθεί περαιτέρω αυτός ο αρχικός ισχυρισμός, εφαρμόζουμε τον τοπικό έλεγχο με βάση την τοπική Cressie and Read απόκλιση του Θεωρήματος 3.1 για να αξιολογήσουμε την καλή προσαρμογή των δεδομένων στο εκθετικό μοντέλο με παράμετρο $\theta_0 = 0.8$. Ο Πίνακας 3.8 παρουσιάζει τις τιμές του τοπικού στατιστικού ελέγχου $T^{\omega}_{\phi_{2/3},n}(\widehat{\theta}, \theta_0)$, όπου σε κάθε γραμμή δίνεται ένας διαφορετικός τοπικός στατιστικός έλεγχος για διαφορετικές τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$, του περικομμένου κανονικού πυρήνα. Η τιμή του καθολικού στατιστικού είναι $T_{\phi_{2/3},n}(\widehat{\theta}, \theta_0) = 9.5619$ και γι' αυτό η υπόθεση ότι τα δεδομένα προέρχονται από μια εκθετική κατανομή με $\theta_0 = 0.8$ απορρίπτεται σε επίπεδο $\alpha = 0.05$, ενώ αν θέσουμε την παράμετρο του πυρήνα ίση με $\omega = (1.4, 0.1^2)$, ο τοπικός έλεγχος δεν μπορεί να απορρίψει τη $H_0^{\omega} : \theta = 0.8$.

Συνεπώς, για τους ασθενείς καρκίνου με διάρκεια ζωής περίπου 1.1 έως 1.7 χρόνια, κάποιος μπορεί ακόμα και να θεωρήσει και συνεπώς να υιοθετήσει ένα εκθετικό μοντέλο με παράμετρο 0.8 (ή μέσο όρο ζωής 1.25 = 1 / 0.8 έτη), αντί για το εκθετικό μοντέλο με παράμετρο $\widehat{\theta} = 0.61$. Σημειώνεται ότι σε όλο το πεδίο ορισμού (καθολικά) το εκθετικό μοντέλο με παράμετρο 0.8 αποτυγχάνει στο να περιγράψει τα δεδομένα.

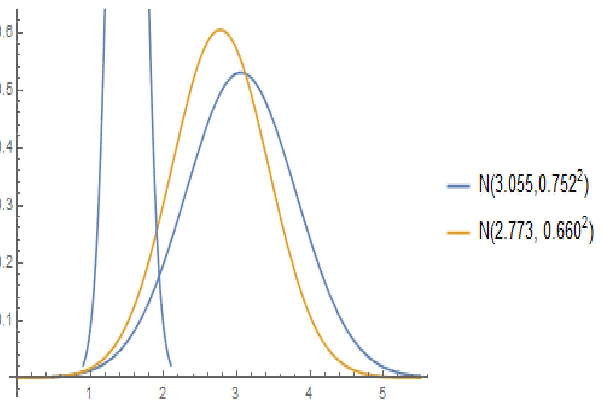
3.5.2 Τοπικός έλεγχος ομοιογένειας: Δεδομένα Hosmer and Lemeshow (1989)

Το ενδιαφέρον μας τώρα στρέφεται σε μια εφαρμογή του ελέγχου τοπικής ομοιογένειας, χρησιμοποιώντας ένα πραγματικό σύνολο δεδομένων των Hosmer and Lemeshow (1989, Παράρτημα 1, σελ. 247). Το αρχείο δεδομένων αναφέρεται στο βάρος γέννησης (BWT) ενός βρέφους, που εξαρτάται από το αν καπνίζει η μητέρα του κατά τη διάρκεια της εγκυμοσύνης. Τα δεδομένα σχετικά με τη μεταβλητή BWT θεωρούνται ότι προέρχονται από δύο πληθυσμούς, τον πληθυσμό, έστω Π_1, των βρεφών που οι μητέρες τους δεν κάπνιζαν κατά τη διάρκεια της εγκυμοσύνης και τον πληθυσμό, έστω Π_2, των βρεφών που οι μητέρες τους κάπνιζαν κατά τη διάρκεια της εγκυμοσύνης. Μας ενδιαφέρει να ερευνήσουμε τυχόν διαφορές μεταξύ των δύο πληθυσμών, είτε γενικά, δηλαδή σε ολόκληρο το πεδίο ορισμού των κατανομών που περιγράφουν κάθε πληθυσμό, είτε τοπικά, επικεντρώνοντας σε μια συγκεκριμένη περιοχή του πεδίου ορισμού παρατήρησης, όπου μπορεί να διαφέρουν οι δύο πληθυσμοί. Ο Πίνακας 3.9 παρουσιάζει τους ελέγχους κανονικότητας μαζί με βασικά περιγραφικά στατιστικά για τα δεδομένα από τους δύο πληθυσμούς, συμπεριλαμβανομένων των δειγματικών μέσων και διακυμάνσεων. Σημειώνεται ότι η υπόθεση κανονικότητας μπορεί να υιοθετηθεί για τους δύο πληθυσμούς και γι' αυτό οι δύο μονοδιάστατες κανονικές κατανομές μπορούν να χρησιμοποιηθούν για να περιγράψουν τα δύο σύνολα δεδομένων.

Πληθυσμός	n	Έλεγχοι κανονικότητας (p-τιμές)		Δειγματικός μέσος
		Shapiro-Wilk	Kolmogorov-Smirnov	και διακύμανση
Π_1	115	0.344	0.2 (Κάτω φράγμα)	$\overline{X}_1 = 3.055, S_1^2 = 0.752^2$
Π_2	74	0.410	0.2 (Κάτω φράγμα)	$\overline{X}_2 = 2.773, S_2^2 = 0.660^2$

Πίνακας 3.9: Έλεγχοι κανονικότητας και περιγραφικά στατιστικά για τις δύο ομάδες βρεφών.

Στο Σχήμα 3.4 απεικονίζονται οι πυκνότητες των πληθυσμών Π_1 και Π_2, χρησιμοποιώντας τις εκτιμώμενες μέσες τιμές και διακυμάνσεις του Πίνακα 3.9, μαζί με ένα κανονικό πυρήνα με παράμετρο $\omega = (1.5, 0.2^2)$. Πιο συγκεκριμένα, οι πληθυσμοί Π_1 και Π_2 πε-



Σχήμα 3.4: Γράφημα των κανονικών κατανομών με παραμέτρους $\theta_1 = (3.055, 0.752^2)$, $\theta_2 = (2.773, 0.660^2)$ και του κανονικού πυρήνα με παράμετρο $\omega = (1.5, 0.2^2)$.

ριγράφονται από μονοδιάστατες κανονικές κατανομές με παραμέτρους $\widehat{\theta}_1 = (\widehat{\mu}_1, \widehat{\sigma}_1^2) = (3.055, 0.752^2)$ και $\widehat{\theta}_2 = (\widehat{\mu}_2, \widehat{\sigma}_2^2) = (2.773, 0.660^2)$, αντίστοιχα.

Για την σύγκριση των πληθυσμών Π_1 και Π_2, χρησιμοποιείται το στατιστικό ελέγχου που βασίζεται στην τοπική λ-απόκλιση των Cressie and Read. Στον Πίνακα 3.10 παρουσιάζονται οι τιμές του τοπικού στατιστικού ελέγχου της Cressie and Read απόκλισης $T^{\omega}_{\phi_{2/3},n,m}(\widehat{\theta}_1, \widehat{\theta}_2)$, με τιμή $\lambda = 2/3$, η οποία θεωρείται κατάλληλη επιλογή σε ανάλυση δεδομένων με την απόκλιση αυτή. Σε κάθε γραμμή του Πίνακα 3.10 δίνεται το αποτέλεσμα ενός διαφορετικού τοπικού ελέγχου ομοιογένειας, με βάση το Θεώρημα 3.7, για διάφορες τιμές της παραμέτρου $\omega = (\mu_\omega, \sigma_\omega^2)$ της πυκνότητα πυρήνα h_ω. Επιλέγεται η τυπική απόκλιση του πυρήνα να είναι 0.2, μια μικρή τιμή που αποδίδει περισσότερο βάρος σε τιμές κοντά στη μέση τιμή του πυρήνα και αναδεικνύει έτσι τις διαφορές των δύο πληθυσμών σε μια περιοχή πολύ κοντά στη μέση τιμή του πυρήνα. Στην προτελευταία γραμμή παρατίθεται το αποτέλεσμα του καθολικού ελέγχου ομοιογένειας, με $\phi_\lambda(u) = \frac{u^{\lambda+1} - u - \lambda(u-1)}{\lambda(\lambda+1)}$ και $\lambda = 2/3$. Η τιμή του καθολικού στατιστικού είναι $T_{\phi_{2/3},n,m}(\widehat{\theta}_1, \widehat{\theta}_2) = 13.158$, ενώ η τιμή του κλασικού t-τεστ για τον έλεγχο ισότητας των μέσων τιμών δύο ανεξάρτητων κανονικών πληθυσμών με άγνωστες διακυμάνσεις είναι $t = 2.634$. Σύμφωνα και με τις δύο αυτές τιμές η υπόθεση της ομοιογένειας για τους πληθυσμούς Π_1 και Π_2 απορρίπτεται γενικά σε επίπεδο σημαντικότητας $\alpha = 0.05$. Αντίθετα, η ίδια υπόθεση δεν μπορεί να απορριφθεί σε επίπεδο $\alpha = 0.05$, εάν η παράμετρος του πυρήνα έχει επιλεγεί να τονίσει τις αριστερές ουρές των πυκνοτήτων, δηλαδή για $\omega = (1.5, 0.2^2)$. Αυτό σημαίνει ότι εστιάζοντας στη συγκεκριμένη

$\omega = (\mu_\omega, \sigma^2_\omega)$	Στατιστικό ελέγχου $T^{\omega}_{\phi_{2/3},n,m}(\hat{\theta}_1, \hat{\theta}_2)$	Κρίσιμα σημεία c_α για επίπεδο σημαντικότητας $a = 0.05$	Απόφαση
$(1, 0.2^2)$	0.0691	1.3843	Αποδοχή H_0
$(1.5, 0.2^2)$	0.5646	2.1745	Αποδοχή H_0
$(2.7, 0.2^2)$	1.1810	0.8869	Απόρριψη H_0
$(4.5, 0.2^2)$	7.2151	2.1979	Απόρριψη H_0
$T_{\phi_{2/3},n,m}(\hat{\theta}_1, \hat{\theta}_2) = 13.158$		5.991	Απόρριψη H_0
$t = 2.634$		$t_{0.025,187} = 1.973$	Απόρριψη H_0

Πίνακας 3.10: Οι τιμές του τοπικού στατιστικού έλεγχου ϕ-απόκλισης $T^{\omega}_{\phi_{2/3},n,m}(\hat{\theta}_1, \hat{\theta}_2)$ για δύο δείγματα μεγέθους $n = 115$ και $m = 74$, αντίστοιχα. Κάθε γραμμή δείχνει ένα διαφορετικό έλεγχο τοπικής ομοιογένειας για διάφορες τιμές του $\omega = (\mu_\omega, \sigma^2_\omega)$. Η προτελευταία γραμμή δείχνει την τιμή του καθολικού στατιστικού ελέγχου ϕ-απόκλισης $T_{\phi_{2/3},n,m}(\hat{\theta}_1, \hat{\theta}_2)$. Η τελευταία γραμμή δείχνει την τιμή του κλασσικού t-ελέγχου για τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_1 = \mu_2$. Τα πληθυσμιακά μοντέλα είναι μονοδιάστατες, κανονικές κατανομές με εκτιμώμενες παραμέτρους των θ_1 και θ_2 να είναι $\hat{\theta}_1 = (\hat{\mu}_1, \hat{\sigma}^2_2) = (3.055, 0.752^2)$ και $\hat{\theta}_2 = (\hat{\mu}_2, \hat{\sigma}^2_2) = (2.773, 0.660^2)$, αντίστοιχα.

περιοχή του πεδίου ορισμού (δηλαδή για βρέφη που γεννιούνται με βάρος μικρότερο από 2.25 κιλά), οι δύο πληθυσμοί δε διαφέρουν στατιστικά σημαντικά και γι' αυτό σ' αυτήν την περίπτωση η κατάσταση καπνίσματος μιας μητέρας δεν έχει καμία επίδραση στο βάρος του νεογέννητου. Τα αποτελέσματα αυτά φαίνονται αρκετά λογικά διαισθητικά, δεδομένου ότι το κάπνισμα της μητέρας κατά την εγκυμοσύνη είναι μόνο ένας από τους πιθανούς παράγοντες που συντελούν και επηρεάζουν το βάρος του νεογέννητου, με το τελευταίο να επηρεάζεται από πολλούς άλλους παράγοντες, όπως οι διατροφικές συνήθειες της μητέρας ή η λήψη φαρμακευτικής αγωγής κατά τη διάρκεια της εγκυμοσύνης.

Κεφάλαιο 4

Κριτήριο πληροφορίας για την τοπική επιλογή μοντέλου

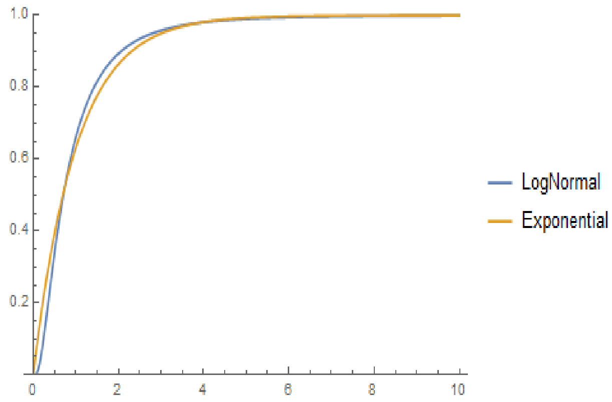
Έστω ένα τυχαίο δείγμα από ένα πληθυσμό που περιγράφεται από μια άγνωστη κατανομή στον χώρο $\mathcal{X}$ και έστω δύο ή περισσότερα παραμετρικά μοντέλα, με κοινό πεδίο ορισμού $\mathcal{X}$, τα οποία είναι υποψήφια για τη μοντελοποίηση των δεδομένων αυτών. Όπως αναφέρθηκε στην Ενότητα 1.2, είναι ελκυστική η ιδέα της ανάπτυξης μιας μεθόδου που θα διερευνά ολόκληρο το πεδίο ορισμού $\mathcal{X}$, με σκοπό την επιλογή του καλύτερου μοντέλου, μεταξύ των διαθέσιμων υποψήφιων μοντέλων, σε ορισμένες περιοχές του πεδίου ορισμού $\mathcal{X}$. Η ιδέα αυτή αποτελεί αντικείμενο μελέτης αυτού του κεφαλαίου. Ειδικότερα, στο πλαίσιο αυτό, η διάθρωση του Κεφαλαίου 4 έχει ως εξής. Στην Ενότητα 4.1 εισάγεται η έννοια της τοπικής επιλογής μοντέλου (local model selection) και το τοπικό κριτήριο $LDiv.IC$ (Local Divergence Information Criterion). Το κριτήριο αυτό βασίζεται στην τοπική εκδοχή της απόκλισης BHHJ των Basu et al. (1998), η οποία ορίστηκε στην Ενότητα 2.3. Στην Ενότητα 4.2 αξιολογείται η απόδοση του προτεινόμενου κριτηρίου τοπικής επιλογής μοντέλου μέσω προσομοιώσεων. Τέλος, στην Ενότητα 4.3 εφαρμόζεται το προτεινόμενο κριτήριο επιλογής μοντέλου σε τρία κλασικά σύνολα δεδομένων της βιβλιογραφίας.

4.1 Κριτήρια για την τοπική επιλογή μοντέλου

Τα κριτήρια επιλογής μοντέλων παρέχουν στους στατιστικούς μια συστηματική και ενδελεχή μέθοδο που επιτρέπει στους ερευνητές να επιλέξουν το καταλληλότερο μοντέλο από μια συλλογή πιθανών κατάλληλων μοντέλων. Η δημιουργία τέτοιων κριτηρίων βασίζεται στη δημιουργία ενός μέτρου ομοιότητας ή απόκλισης μεταξύ του άγνωστου μοντέλου και καθενός εκ των δύο ή περισσότερων υποψήφιων μοντέλων, τα οποία συνήθως περιγράφονται από τις συναρτήσεις κατανομών πιθανότητάς τους. Το πιο γνωστό κριτήριο επιλογής μοντέλου είναι το φημισμένο Akaike Information Criterion (AIC), που προτάθηκε από τον Akaike (1973) και βασίζεται στο μέτρο απόκλισης των Kullback-Leibler. Μετά την πρωτοπόρα εργασία του Akaike, πλήθος ερευνητικών εργασιών έχει δημοσιευτεί στη βιβλιογραφία που σχετίζεται με την κατασκευή κριτηρίων επιλογής μοντέλου σε όλο το πεδίο ορισμού (βλέπε, Ενότητα 1.2 και αναφορές που παρατίθενται εκεί). Στην ενότητα αυτή εισάγεται ένα κριτήριο πληροφορίας για την επιλογή μοντέλου, τοπικά, χρησιμοποιώντας την τοπική απόκλιση BHHJ, που ορίστηκε στη σχέση (2.56) του Ορισμού 2.6. Για την εισαγωγή του κριτηρίου ακολουθείται παρόμοια μεθοδολογία με αυτήν που χρησιμοποιήθηκε για την κατασκευή του κριτηρίου AIC (βλέπε, για παράδειγμα, Burham and Anderson, 2003) ή του κριτηρίου DIC (βλέπε, Mattheou et al., 2009).

Πριν προχωρήσουμε στην εισαγωγή του κριτηρίου, μέσω ενός κλασικού παραδείγματος θα προσπαθήσουμε να τονίσουμε την αναγκαιότητα εισαγωγής μιας μεθόδου που σαρώνει, εξετάζει, διερευνά ολόκληρο το πεδίο ορισμού $\mathcal{X}$ και στο τέλος επιλέγει το καλύτερο μοντέλο, μεταξύ των διαθέσιμων υποψήφιων μοντέλων, σε ορισμένες περιοχές του πεδίου ορισμού $\mathcal{X}$. Το εισαγωγικό παράδειγμα που θα δοθεί έχει μακρά ιστορία στη στατιστική βιβλιογραφία καθώς έχει απασχολήσει πλήθος ερευνητών. Μεταξύ άλλων παραπέμπουμε στους Vuong and Wang (1993), Jimenez-Gamero et al. (2011) και στις αναφορές που δίνονται στις εργασίες αυτές. Ειδικότερα, θεωρούμε ότι τα δύο ανταγωνιστικά μοντέλα είναι αυτά της εκθετικής κατανομής με παράμετρο $\lambda = 1$ και της λογαριθμο-κανονικής κατανομής με παραμέτρους $\mu = -0.347$ και $\eta = 0.833$.

Στο Σχήμα 4.1 παρουσιάζεται η αθροιστική συνάρτηση κατανομής των δύο αυτών κατανομών. Είναι σαφές, από το Σχήμα 4.1, ότι οι δύο κατανομές είναι ανόμοιες μεταξύ τους σε ένα διάστημα του κοινού πεδίου ορισμού τους, είναι πολύ κοντά (σχεδόν όμοιες) σε κάποιο άλλο διάστημα, ενώ στις ουρές τους συμπίπτουν. Τυπικά, με τις κλασικές μεθόδους επιλογής μοντέλου η σύγκριση πραγματοποιείται στο κοινό πεδίο ορισμού των δύο



Σχήμα 4.1: Αθροιστικές συναρτήσεις κατανομών της λογαριθμο-κανονικής κατανομής (log-normal) με παραμέτρους $\mu = -0.347$ και $\eta = 0.833$ και της εκθετικής κατανομής (exponential) με παράμετρο $\lambda = 1$.

αυτών μοντέλων. Έτσι, ο ερευνητής θα επιλέξει ένα εκ' των δύο αυτών μοντέλων παρά το γεγονός ότι αυτά συμπίπτουν σε κάποιες περιοχές του κοινού πεδίου ορισμού τους και επομένως θα μπορούσε, τουλάχιστον για τις περιοχές αυτές, να επιλεγεί το ανταγωνιστικό μοντέλο. Προκύπτει, με βάση το παράδειγμα αυτό, η ανάγκη εισαγωγής ενός κριτηρίου το οποίο θα σαρώνει το κοινό πεδίο ορισμού των υποψήφιων μοντέλων και θα επιλέγει το καταλληλότερο από αυτά στις διάφορες περιοχές του κοινού πεδίου ορισμού.

Η σύγκριση των υποψήφιων μοντέλων σε ορισμένη περιοχή του πεδίου ορισμού μπορεί να επιτευχθεί με τη χρήση ενός μέτρου απόκλισης που εστιάζει σε συγκεκριμένο υποσύνολο του κοινού πεδίου ορισμού $\mathcal{X}$. Ενός μέτρου τοπικής απόκλισης μεταξύ του πραγματικού, αλλά άγνωστου μοντέλου, έστω g, και ενός υποψηφίου μοντέλου, έστω f_θ από μια παραμετρική οικογένεια κατανομών $\mathcal{F}$. Ένα τέτοιο μέτρο απόκλισης είναι η τοπική εκδοχή της απόκλισης BHHJ που ορίστηκε στην Ενότητα 2.3, η οποία εφοδιασμένη με την κατάλληλη επιλογή της πυκνότητας h_ω, μπορεί να οδηγήσει στο επιθυμητό υποσύνολο του $\mathcal{X}$. Η εισαγωγή αυτού του κριτηρίου της τοπικής επιλογής μοντέλου αποτελεί αντικείμενο μελέτης της υποενότητας που ακολουθεί.

Σε όσα ακολουθούν, θεωρούμε το μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$ και έστω $\mathcal{F} = \{F_\theta\}$ μια παραμετρική οικογένεια μέτρων πιθανότητας σε αυτόν, με $\theta \in \Theta \subseteq R^M$, $M \geq 1$. Έστω επίσης $\mathcal{G}$ η κλάση των μέτρων πιθανότητας G στο χώρο $(\mathcal{X}, \mathcal{A})$ και ένα σ-πεπερασμένο μέτρο μ στον ίδιο μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$, έτσι ώστε $G \ll \mu$ και $g = \frac{dG}{d\mu}$ η Radon-Nikodym παράγωγος της G ως προς μ. Για $\theta \in \Theta$, υποθέτουμε ότι $F_\theta \ll \mu$ και με $f_\theta = \frac{dF_\theta}{d\mu}$ συμβολίζεται η αντίστοιχη Radon-Nikodym παράγωγος.

4.1.1 Κριτήριο τοπικής επιλογής μοντέλου

Έστω $X_1, ..., X_n$ ένα τυχαίο δείγμα από έναν πληθυσμό με κατανομή G ορισμένη στον μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$ και αντίστοιχη πραγματική, αλλά άγνωστη, συνάρτηση πυκνότητας πιθανότητας g. Θεωρούμε ένα υποψήφιο παραμετρικό μοντέλο F_θ με συνάρτηση πυκνότητας πιθανότητας f_θ, από την παραμετρική οικογένεια κατανομών πιθανότητας $\mathcal{F} = \{F_\theta\}$ (ή πυκνοτήτων $\{f_\theta\}$) στο μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$, με άγνωστη M-διάστατη παράμετρο $\theta \in \Theta \subseteq R^M$, $M \geq 1$.

Το πρώτο βήμα για την εισαγωγή ενός κριτηρίου επιλογής μοντέλου τοπικά, όπως είναι προφανές από τη συζήτηση που προηγήθηκε, είναι η υιοθέτηση ενός τοπικού μέτρου απόκλισης μεταξύ του αληθινού, αλλά άγνωστου, μοντέλου και του υποψήφιου μοντέλου. Η εισαγωγή ενός τοπικού μέτρου απόκλισης υπήρξε το αντικείμενο μελέτης του δεύτερου κεφαλαίου αυτής της διατριβής. Για τη χρησιμοποίηση ενός τέτοιου μέτρου απαιτείται επιπλέον η εισαγωγή, η χρησιμοποίηση μιας παραμετρικής οικογένειας κατανομών H_ω στο μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$, με $H_\omega \ll \mu$. Για μια γνωστή τιμή της παραμέτρου ω, $\omega \in \Theta^* \subseteq R^K$, $K \geq 1$ η πυκνότητα πυρήνας $h_\omega = \frac{dH_\omega}{d\mu}$ είναι η Radon-Nikodym παράγωγος.

Στο δεύτερο κεφάλαιο παρουσιάστηκε μια πλειάδα μέτρων τοπικής απόκλισης. Το τοπικό μέτρο απόκλισης που θα χρησιμοποιηθεί για την επιλογή μοντέλου τοπικά είναι η τοπική απόκλιση BHHJ, όπως δικαιολογείται στην εργασία των Basu et al. (1998). Ειδικότερα, η επιλογή μοντέλου, τοπικά, βασίζεται στην ποσότητα

$$W_\alpha^\omega(\theta) = \int\limits_{\mathcal{X}} h_\omega(x) f_\theta^{1+a}(x)dx - (1+\alpha^{-1}) \int\limits_{\mathcal{X}} h_\omega(x) g(x) f_\theta^\alpha(x) d\mu(x), \; a > 0, \tag{4.1}$$

η οποία ισοδύναμα μπορεί να γραφεί, λαμβάνοντας υπόψη τη σχέση (2.56), στη μορφή

$$W_\alpha^\omega(\theta) = D_a^\omega(g, f_\theta) - \frac{1}{a} \int\limits_{\mathcal{X}} h_\omega(x) g^{1+a}(x) d\mu(x), \; a > 0, \tag{4.2}$$

όπου $D_a^\omega(g, f_\theta)$ είναι η τοπική απόκλιση BHHJ που ορίστηκε στη σχέση (2.56). Σημειώνεται ότι η ποσότητα $\frac{1}{a} \int\limits_{\mathcal{X}} h_\omega(x) g^{1+a}(x) d\mu(x)$ παραμένει σταθερή ανεξάρτητα του μοντέλου

f_θ που χρησιμοποιείται. Ως εκ τούτου, λαμβάνοντας υπόψη τους Burham and Anderson, 2003, σελ. 364, η ποσότητα $W_\alpha^\omega(\theta)$ μπορεί να θεωρηθεί ως μέτρο της τοπικής απόκλισης μεταξύ g και f_θ η οποία διαφέρει από την τοπική απόκλιση BHHJ κατά μια σταθερά. Σημειώνεται επίσης ότι, η (4.1) μπορεί να γραφεί ως

$$W_\alpha^\omega(\theta) = E_{f_\theta}(h_\omega(X) f_\theta^\alpha(X)) - (1+\alpha^{-1}) E_g(h_\omega(X) f_\theta^\alpha(X)), \ a > 0. \tag{4.3}$$

Όμως, όπως είναι φανερό από τη σχέση (4.2), η ποσότητα $W_\alpha^\omega(\theta)$ δεν μπορεί να χρησιμοποιηθεί άμεσα για την τοπική επιλογή μοντέλου καθώς, μεταξύ άλλων, στον υπολογισμό της υπεισέρχεται η άγνωστη παράμετρος θ. Ένας τρόπος επίλυσης αυτού του προβλήματος είναι η αντικαστάστατη της άγνωστης παραμέτρου θ από ένα συνεπή εκτιμητή της, ο οποίος θα ήταν επιθυμητό να πληροί και κάποιες επιπλέον ιδιότητες, όπως για παράδειγμα να ακολουθεί ασυμπτωτικά κανονική κατανομή. Σε αυτό το πλαίσιο, αν η πραγματική κατανομή g ανήκει στην παραμετρική οικογένεια $\mathcal{F} = \{F_\theta\}$ ένας τέτοιος εκτιμητής μπορεί να προκύψει είτε με τη μεγιστοποίηση της συνάρτησης πιθανοφάνειας ή την ελαχιστοποίηση της απόκλισης BHHJ (βλέπε, Basu et al., 2011). Στην τελευταία περίπτωση, η ιδιότητα της συνέπειας, καθώς και η ασυμπτωτική κανονικότητα του εκτιμητή εξασφαλίζεται από το Θεώρημα 4.2 των Basu et al. (1998).

Με σκοπό την εισαγωγή του τοπικού κριτηρίου επιλογής μοντέλου, έστω ότι το πραγματικό μοντέλο g ανήκει στην παραμετρική οικογένεια κατανομών $\{f_\theta\}$ και έστω $\hat{\theta}$ ένας συνεπής και ασυμπτωτικά κανονικός εκτιμητής της θ με βάση το τυχαίο δείγμα που είναι διαθέσιμο από την κατανομή g. Ο εκτιμητής $\hat{\theta}$ μπορεί να είναι εκείνος που προκύπτει από την ελαχιστοποίηση της απόκλισης BHHJ, όπως προαναφέρθηκε. Ακολουθώντας την κλασική μεθοδολογία για την κατασκευή του AIC (βλέπε, Burham and Anderson, 2003) ή την κατασκευή του DIC (βλέπε, Mattheou et al., 2009) το ενδιαφέρον εντοπίζεται στη σταθμισμένη ή τοπική αναμενόμενη συνολική διαφορά

$$\begin{aligned} E_g(W_\alpha^\omega(\hat{\theta})) &= E_g(W_\alpha^\omega(\theta) | \theta = \hat{\theta}) \\ &= E_g\left(E_{f_{\hat{\theta}}}(h_\omega(X) f_{\hat{\theta}}^\alpha(X))\right) - (1+\alpha^{-1}) E_g\left(E_g(h_\omega(X) f_{\hat{\theta}}^\alpha(X))\right), \ a > 0, \end{aligned} \tag{4.4}$$

με $\hat{\theta}$ ένας συνεπής και ασυμπτωτικά κανονικός εκτιμητής της παραμέτρου θ.

Στη συνέχεια, παρακινούμενοι, μεταξύ άλλων, από τους Burham and Anderson (2003, σελ. 240), υιοθετείται το ακόλουθο κριτήριο:

«επιλέγουμε το μοντέλο $f_{\hat{\theta}}$ που ελαχιστοποιεί την ποσότητα $E_g(W_\alpha^\omega(\hat{\theta}))$», δηλαδή επιλέγουμε το μοντέλο $f_{\hat{\theta}}$ που ελαχιστοποιεί τη σταθμισμένη ή τοπική αναμενόμενη συνολική διαφορά μεταξύ των g και $f_{\hat{\theta}}$.

Όμως, η ποσότητα $E_g(W_\alpha^\omega(\hat{\theta}))$ εξακολουθεί να εξαρτάται από την άγνωστη παράμετρο θ. Για να επιλυθεί αυτό το πρόβλημα, θα πρέπει να προσδιοριστεί ένας ασυμπωτικά αμερόληπτος εκτιμητής της τοπικής αναμενόμενης συνολικής διαφοράς $E_g(W_\alpha^\omega(\hat{\theta}))$, με $g \in \{f_\theta\}$.

Για το σκοπό αυτό, αρχικά θα παρουσιαστούν, σε μορφή λήμματος, ορισμένα ενδιάμεσα αναγκαία αποτελέσματα για την απόκτηση της βασικής πρότασης της ενότητας αυτής. Υπενθυμίζεται ότι με $\nabla_\theta l_\theta(\cdot)$ συμβολίζεται το M-διάστατο ανάδελτα διάνυσμα της $l_\theta(\cdot)$ ως προς θ και με $\nabla_\theta^2 l_\theta(\cdot)$ ο αντίστοιχος $M \times M$ Εσσιανός πίνακας (Hessian matrix), όπου $l_\theta(\cdot)$ είναι μια πραγματική συνάρτηση. Τέλος, προκειμένου να διερευνηθεί η ασυμπτωτική συμπεριφορά των εκτιμητών, θεωρούμε ότι οι βασικές συνθήκες ομαλότητας της ασυμπτωτικής στατιστικής, που αναφέρθηκαν στο Κεφάλαιο 3, ικανοποιούνται.

Λήμμα 4.1. *Ισχύει ότι*

$$\nabla_\theta W_\alpha^\omega(\theta) = (a+1)\left(\int_{\mathcal{X}} h_\omega(x) u_\theta(x) f_\theta^{1+a}(x) d\mu(x) - E_g(h_\omega(X) u_\theta(X) f_\theta^\alpha(X))\right) \tag{4.5}$$

και

$$\begin{aligned}\nabla_\theta^2 W_\alpha^\omega(\theta) \quad &= (a+1)\Big\{(a+1)\int_{\mathcal{X}} h_\omega(x) u_\theta(x) u_\theta^t(x) f_\theta^{1+a}(x) d\mu(x) \\ &-\int_{\mathcal{X}} h_\omega(x)\Xi_\theta(x) f_\theta^{1+a}(x) d\mu(x) \\ &+E_g(h_\omega(X)\Xi_\theta(X) f_\theta^a(X)) - E_g(a h_\omega(X) u_\theta(X) u_\theta^t(X) f_\theta^\alpha(X))\Big\},\end{aligned} \tag{4.6}$$

όπου $u_\theta(x) = \nabla_\theta \log(f_\theta(x))$ *και* $\Xi_\theta(x) = -\nabla_\theta^2 \log(f_\theta(x))$.

Απόδειξη. Για το ανάδελτα της ποσότητας $W_\alpha^\omega(\theta)$, έχουμε

$$\begin{aligned}\nabla_\theta W_\alpha^\omega(\theta) \quad &= (a+1)\int_{\mathcal{X}} h_\omega(x) f_\theta^a(x) \nabla_\theta f_\theta(x) d\mu(x) \\ &-\tfrac{a+1}{a} E_g(a h_\omega(X) f_\theta^{\alpha-1}(X) \nabla_\theta f_\theta(x)) \\ &= (a+1)\Big(\int_{\mathcal{X}} h_\omega(x) f_\theta^{1+a}(x) \nabla_\theta \log(f_\theta(x)) d\mu(x) \\ &-E_g(h_\omega(X) f_\theta^\alpha(X)) \nabla_\theta \log(f_\theta(x))\Big).\end{aligned} \tag{4.7}$$

Επιπλέον, για τον Εσσιανό πίνακα της ποσότητας $W_\alpha^\omega(\theta)$, είναι

$$
\begin{aligned}
\nabla_\theta^2 W_\alpha^\omega(\theta) &= (a+1)\Big\{\int_{\mathcal{X}} h_\omega(x)\left[-\Xi_\theta(x)+u_\theta(x)(a+1)f_\theta^a(x)\nabla_\theta^t f_\theta(x)d\mu(x)\right] \\
&+E_g(h_\omega(x)\Xi_\theta(x)f_\theta^a(x)-ah_\omega(x)u_\theta(x)f_\theta^{a-1}(x)\nabla_\theta^t f_\theta(x))\Big\} \\
&= (a+1)\Big\{(a+1)\int_{\mathcal{X}} h_\omega(x)u_\theta(x)u_\theta^t(x)f_\theta^{1+a}(x)d\mu(x) \\
&-\int_{\mathcal{X}} h_\omega(x)\Xi_\theta(x)f_\theta^{1+a}(x)d\mu(x) \\
&+E_g(h_\omega(X)\Xi_\theta(X)f_\theta^a(X))-E_g(ah_\omega(X)u_\theta(X)u_\theta^t(X)f_\theta^\alpha(X))\Big\}.
\end{aligned}
\tag{4.8}
$$

∎

Μια άμεση συνέπεια του Λήμματος 4.1, είναι η ακόλουθη.

Πόρισμα 4.1. *Αν η πραγματική κατανομή g ανήκει στην παραμετρική οικογένεια κατανομών $\{f_\theta\}$ και θ_0 είναι η αληθινή τιμή της άγνωστης παραμέτρου θ, τότε το ανάδελτα και ο Εσσιανός πίνακας της ποσότητας $W_\alpha^\omega(\theta)$ δίνονται από τις σχέσεις*

$$[\nabla_\theta W_\alpha^\omega(\theta)]_{\theta=\theta_0}=0, \tag{4.9}$$

και

$$[\nabla_\theta^2 W_\alpha^\omega(\theta)]_{\theta=\theta_0}=(a+1)\int_{\mathcal{X}} h_\omega(x)u_{\theta_0}(x)u_{\theta_0}^t(x)f_{\theta_0}^{1+a}(x)d\mu(x)=(a+1)J^\omega(\theta_0), \tag{4.10}$$

με

$$J^\omega(\theta)=\left(\int_{\mathcal{X}} h_\omega(x)f_\theta^{1+\alpha}(x)\frac{\partial\log f_\theta(x)}{\partial\theta_i}\frac{\partial\log f_\theta(x)}{\partial\theta_j}d\mu(x)\right)_{i,j=1,\dots,M}, \tag{4.11}$$

όπου $u_\theta(x)=\nabla_\theta\log(f_\theta(x))$.

Το επόμενο λήμμα περιγράφει τον τρόπο με τον οποίο η τοπική αναμενόμενη συνολική διαφορά $E_g(W_\alpha^\omega(\hat\theta))$ μπορεί να γραφεί μέσω της επέκτασης Taylor.

Λήμμα 4.2. *Αν η πραγματική κατανομή g ανήκει στην παραμετρική οικογένεια κατανομών $\{f_\theta\}$ και θ_0 είναι η πραγματική τιμή της παραμέτρου θ, τότε η τοπική αναμενόμενη συνολική διαφορά $E_g(W_\alpha^\omega(\hat\theta))$ μεταξύ των g και $f_{\hat\theta}$, δίνεται από τη σχέση*

$$E_g(W_\alpha^\omega(\hat\theta))=W_\alpha^\omega(\theta_0)+\frac{\alpha+1}{2}E_g((\hat\theta-\theta_0)^tJ^\omega(\theta_0)(\hat\theta-\theta_0))+E_g(R_n),\ a>0, \tag{4.12}$$

όπου $R_n=o\left(\|\hat\theta-\theta_0\|^2\right)$ και $J^\omega(\theta_0)$ όπως ορίστηκε στη (4.11).

Απόδειξη. Αναπτύσσοντας τη συνάρτηση $W_\alpha^\omega(\theta)$ σε σειρά Taylor γύρω από την πραγματική παράμετρο θ_0, για $\theta = \hat{\theta}$ και λαμβάνοντας υπόψη τις σχέσεις (4.9) και (4.10), προκύπτει,

$$W_\alpha^\omega(\hat{\theta}) = W_\alpha^\omega(\theta_0) + \frac{\alpha+1}{2}(\hat{\theta}-\theta_0)^t J^\omega(\theta_0)(\hat{\theta}-\theta_0)) + o\left(\|\hat{\theta}-\theta_0\|^2\right), \ a > 0, \quad (4.13)$$

από την οποία άμεσα προκύπτει η (4.12). ∎

Στη συνέχεια, για $\omega \in \Theta^*$ και $\theta \in \Theta$, ένας εκτιμητής $Q_\alpha^\omega(\theta)$ της ποσότητας $W_\alpha^\omega(\theta)$ λαμβάνεται αντικαθιστώντας την ποσότητα $E_g(h_\omega(X) f_\theta^\alpha(X))$ στην έκφραση της $W_\alpha^\omega(\theta)$, που ορίστηκε στη (4.3), με τον εκτιμητή $\frac{1}{n}\sum\limits_{i=1}^{n} h_\omega(X_i) f_\theta^\alpha(X_i)$, όπου $X_1, ..., X_n$ είναι ένα τυχαίο δείγμα από την πραγματική συνάρτηση κατανομής g. Τότε, ο εκτιμητής $Q_\alpha^\omega(\theta)$ δίνεται από τη σχέση

$$Q_\alpha^\omega(\theta) = \int\limits_{\mathcal{X}} h_\omega(x) f_\theta^{1+\alpha}(x) d\mu(x) - (1+\alpha^{-1})\frac{1}{n}\sum_{i=1}^{n} h_\omega(X_i) f_\theta^\alpha(X_i), \ a > 0, \quad (4.14)$$

Το ανάδελτα και ο Εσσιανός πίνακας της $Q_\alpha^\omega(\theta)$, ως προς θ, δίνονται στο ακόλουθο Λήμμα.

Λήμμα 4.3. *Το ανάδελτα και ο Εσσιανός πίνακας της ποσότητας $Q_\alpha^\omega(\theta)$ που ορίστηκε στη σχέση (4.14), δίνονται από τις σχέσεις*

$$\nabla_\theta Q_\alpha^\omega(\theta) = (a+1)\left(\int\limits_{\mathcal{X}} h_\omega(x) u_\theta(x) f_\theta^{1+a}(x) d\mu(x) - \frac{1}{n}\sum_{i=1}^{n} h_\omega(X_i) u_\theta(X_i) f_\theta^\alpha(X_i)\right) \quad (4.15)$$

και

$$\begin{aligned}\nabla_\theta^2 Q_\alpha^\omega(\theta) &= (a+1)\Big\{(a+1)\int\limits_{\mathcal{X}} h_\omega(x) u_\theta(x) u_\theta^t(x) f_\theta^{1+a}(x) d\mu(x) \\ &- \int\limits_{\mathcal{X}} h_\omega(x) \Xi_\theta(x) f_\theta^{1+a}(x) d\mu(x) \\ &+ \frac{1}{n}\sum_{i=1}^{n} h_\omega(X_i) \Xi_\theta(X_i) f_\theta^a(X_i) - \frac{1}{n}\sum_{i=1}^{n} a h_\omega(X_i) u_\theta(X_i) u_\theta^t(X_i) f_\theta^\alpha(X_i)\Big\},\end{aligned} \quad (4.16)$$

όπου $u_\theta(x) = \nabla_\theta \log(f_\theta(x))$ και $\Xi_\theta(x) = -\nabla_\theta^2 \log(f_\theta(x))$.

Η απόδειξη ακολουθεί παρόμοια βήματα με την απόδειξη του Λήμματος 4.1 και παραλείπεται.

Το ακόλουθο λήμμα διατυπώνει το τελευταίο ενδιάμεσο αποτέλεσμα που απαιτείται προκειμένου να αποδειχθεί η κύρια πρόταση αυτού του κεφαλαίου.

Λήμμα 4.4. *Αν το πραγματικό μοντέλο g ανήκει στην παραμετρική οικογένεια κατανομών $\{f_\theta\}$, η τοπική αναμενόμενη συνολική διαφορά $E_g(W_\alpha^\omega(\hat{\theta}))$ δίνεται από τη σχέση*

$$E_g(W_\alpha^\omega(\hat{\theta})) = E_g\left(Q_\alpha^\omega(\hat{\theta}) + (a+1)(\hat{\theta}-\theta_0)^t J^\omega(\theta_0)(\hat{\theta}-\theta_0) + R_n\right), \tag{4.17}$$

όπου $R_n = o\left(\|\hat{\theta}-\theta_0\|^2\right)$, θ_0 είναι η πραγματική παράμετρος και $J^\omega(\theta_0)$ ο πίνακας που ορίστηκε στη (4.11).

Απόδειξη. Με βάση τη σχέση (9) των Mattheou et al. (2009), το ανάπτυγμα Taylor της ποσότητας $Q_\alpha^\omega(\theta)$, που ορίστηκε στη σχέση (4.14), γύρω από τον εκτιμητή $\hat{\theta}$ δίνει την προσέγγιση,

$$\begin{aligned} Q_\alpha^\omega(\theta) &= Q_\alpha^\omega(\hat{\theta}) + (\theta-\hat{\theta})^t[\nabla_\theta Q_\alpha^\omega(\theta)]_{\theta=\hat{\theta}} \\ &+\tfrac{1}{2}(\theta-\hat{\theta})^t[\nabla_\theta^2 Q_\alpha^\omega(\theta)]_{\theta=\hat{\theta}}(\theta-\hat{\theta}) + o\left(\|\hat{\theta}-\theta\|^2\right), \end{aligned} \tag{4.18}$$

$a > 0$. Επιπλέον, με βάση τις σχέσεις (12) των Mattheou et al. (2009) και τα Λήμματα 4.1 και 4.3, έχουμε

$$[\nabla_\theta Q_\alpha^\omega(\theta)]_{\theta=\theta_0} \xrightarrow[n\to\infty]{P} [\nabla_\theta W_\alpha^\omega(\theta)]_{\theta=\theta_0} \text{ και } [\nabla_\theta^2 Q_\alpha^\omega(\theta)]_{\theta=\theta_0} \xrightarrow[n\to\infty]{P} [\nabla_\theta^2 W_\alpha^\omega(\theta)]_{\theta=\theta_0}, \tag{4.19}$$

με $n \to \infty$. Εφόσον $\hat{\theta} \xrightarrow[n\to\infty]{P} \theta_0$, από το Πόρισμα 4.1, έχουμε ότι ισχύει

$$[\nabla_\theta Q_\alpha^\omega(\theta)]_{\theta=\hat{\theta}} \xrightarrow[n\to\infty]{P} 0 \text{ και } [\nabla_\theta^2 Q_\alpha^\omega(\theta)]_{\theta=\hat{\theta}} \xrightarrow[n\to\infty]{P} (a+1)J^\omega(\theta_0). \tag{4.20}$$

Στη συνέχεια, λαμβάνοντας υπόψη την (4.20), η σχέση (4.18) γράφεται

$$Q_\alpha^\omega(\theta) = Q_\alpha^\omega(\hat{\theta}) + \frac{a+1}{2}(\theta-\hat{\theta})^t J^\omega(\theta_0)(\theta-\hat{\theta}) + o\left(\|\hat{\theta}-\theta\|^2\right). \tag{4.21}$$

Θέτοντας $\theta = \theta_0$ στην παραπάνω ισότητα, έχουμε

$$Q_\alpha^\omega(\theta_0) = Q_\alpha^\omega(\hat{\theta}) + \frac{a+1}{2}(\theta_0-\hat{\theta})^t J^\omega(\theta_0)(\theta_0-\hat{\theta}) + o\left(\|\hat{\theta}-\theta_0\|^2\right), \tag{4.22}$$

συνεπώς

$$E_g(Q_\alpha^\omega(\theta_0)) = E_g(Q_\alpha^\omega(\hat{\theta})) + \frac{\alpha+1}{2}E_g((\theta_0-\hat{\theta})^t J^\omega(\theta_0)(\theta_0-\hat{\theta})) + E_g(R_n), \tag{4.23}$$

όπου $R_n = (\|\hat{\theta} - \theta_0\|^2)$. Επιπλέον, χρησιμοποιώντας το Λήμμα 4.2 παίρνουμε

$$E_g(W_\alpha^\omega(\hat{\theta})) = W_\alpha^\omega(\theta_0) + \frac{\alpha+1}{2} E_g((\hat{\theta} - \theta_0)^t J^\omega(\theta_0)(\hat{\theta} - \theta_0)) + E_g(R_n). \tag{4.24}$$

Από την άλλη πλευρά, είναι $E_g(Q_\alpha^\omega(\theta_0)) = W_\alpha^\omega(\theta_0)$. Πράγματι,

$$\begin{aligned}
E_g(Q_\alpha^\omega(\theta_0)) &= E_g\left(E_{f_{\theta_0}}(h_\omega(X) f_{\theta_0}^{1+\alpha}(X))\right) - (1+\alpha^{-1}) E_g\left(\tfrac{1}{n}\sum_{i=1}^n h_\omega(X_i) f_{\theta_0}^\alpha(X_i)\right) \\
&= E_{f_{\theta_0}}(h_\omega(X) f_{\theta_0}^{1+\alpha}(X)) - (1+\alpha^{-1}) \tfrac{1}{n}\sum_{i=1}^n E_g\left(h_\omega(X_i) f_{\theta_0}^\alpha(X_i)\right) \\
&= E_{f_{\theta_0}}(h_\omega(X) f_{\theta_0}^{1+\alpha}(X)) - (1+\alpha^{-1}) \tfrac{1}{n}\sum_{i=1}^n E_g\left(h_\omega(X) f_{\theta_0}^\alpha(X)\right) \\
&= E_{f_{\theta_0}}(h_\omega(X) f_{\theta_0}^{1+\alpha}(X)) - (1+\alpha^{-1}) E_g\left(h_\omega(X) f_{\theta_0}^\alpha(X)\right) \\
&= W_\alpha^\omega(\theta_0).
\end{aligned}$$

Συνεπώς, συνδυάζοντας τις σχέσεις (4.23) και (4.24) έχουμε το επιθυμητό αποτέλεσμα. ∎

4.1.2 Κατασκευή τοπικού κριτήριου επιλογής μοντέλου: LDiv.IC

Στην υποενότητα αυτή κατασκευάζεται το τοπικό κριτήριο πληροφορίας για την επιλογή μοντέλου LDiv.IC, στηριζόμενοι στο μαθηματικό υπόβαθρο που αναπτύχθηκε στην προηγούμενη υποενότητα 4.1.1. Το βασικό αποτέλεσμα αυτού του κεφαλαίου περιγράφεται παρακάτω.

Πρόταση 4.1. *Αν το πραγματικό μοντέλο g ανήκει στην παραμετρική οικογένεια κατανομών $\{f_\theta\}$, τότε ισχύουν τα ακόλουθα:*
(α) Ένας ασυμπτωτικά αμερόληπτος εκτιμητής της τοπικής αναμενόμενης συνολικής διαφοράς $E_g(W_\alpha^\omega(\hat{\theta}))$, που ορίστηκε στη (4.4), πολλαπλασιασμένος με n δίνεται από τη σχέση

$$\widehat{nE_g(W_\alpha^\omega(\hat{\theta}))} = nQ_\alpha^\omega(\hat{\theta}) + n(a+1)(\hat{\theta} - \theta_0)^t J^\omega(\theta_0)(\hat{\theta} - \theta_0), \tag{4.25}$$

όπου $\hat{\theta}$ είναι ένας συνεπής και ασυμπτωτικά κανονικός εκτιμητής και $Q_\alpha^\omega(\hat{\theta})$ η ποσότητα που ορίστηκε στη σχέση (4.14).
(β) Η αναμενόμενη τιμή της τετραγωνικής μορφής $n(\hat{\theta} - \theta_0)^t J^\omega(\theta_0)(\hat{\theta} - \theta_0)$ δίνεται από τη σχέση

$$E_g[n(\hat{\theta} - \theta_0)^t J^\omega(\theta_0)(\hat{\theta} - \theta_0)] = \sum_{i=1}^r \beta_i, \tag{4.26}$$

όπου $\hat{\theta}$ *είναι ένας εκτιμητής που ικανοποιεί τις ιδιότητες του Θεώρηματος 4.2 των Basu et al. (1998) και* $\beta_1, \beta_2, ..., \beta_r$ *οι μη μηδενικές ιδιοτιμές του πίνακα*

$$J^{\omega}(\theta_0)AVar(\theta_0),$$

με θ_0 *να είναι η πραγματική τιμή της παραμέτρου* θ *και* r *η βαθμίδα του πίνακα*

$$AVar(\theta_0)J^{\omega}(\theta_0)AVar(\theta_0),$$

ήτοι

$$r = rank(AVar(\theta_0)J^{\omega}(\theta_0)AVar(\theta_0)).$$

Ο πίνακας $J^{\omega}(\theta_0)$ *ορίστηκε στην (4.11) και*

$$AVar(\theta_0) = J^{-1}(\theta_0)K(\theta_0)J^{-1}(\theta_0),$$

είναι ο ασυμτωτικός πίνακας διακυμάνσεων-συνδιακυμάνσεων του $\sqrt{n}(\hat{\theta} - \theta_0)$, *όπου οι* $J(\theta_0)$ *και* $K(\theta_0)$ *προσδιορίζονται από τις ακόλουθες σχέσεις,*

$$J(\theta_0) = \int u_{\theta_0}(x)u^t_{\theta_0}(x)f^{1+a}_{\theta_0}(x)dx$$

και

$$K(\theta_0) = \int u_{\theta_0}(x)u^t_{\theta_0}(x)f^{1+2a}_{\theta_0}(x)dx - \int u_{\theta_0}(x)f^{1+a}_{\theta_0}(x)dx \int u^t_{\theta_0}(x)f^{1+a}_{\theta_0}(x)dx.$$

Απόδειξη. (α) Το Λήμμα 4.4 και η ισότητα $no\left(\|\hat{\theta} - \theta_0\|^2\right) = no_p(n^{-1}) = o_p(1)$ (βλέπε, Pardo, 2006, σελ. 411-412), οδηγούν στο επιθυμητό αποτέλεσμα.
(β) Από το Θεώρημα 4.2 των Basu et al. (1998), έχουμε ότι $\hat{\theta}$ είναι ένας συνεπής εκτιμητής της παραμέτρου θ, με

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow[n\to\infty]{L} N(0, AVar(\theta_0))$$

και ασυμπτωτική διακύμανση

$$AVar(\theta_0) = J^{-1}(\theta_0)K(\theta_0)J^{-1}(\theta_0),$$

όπου $J(\theta_0)$ και $K(\theta_0)$ δίνονται από τις σχέσεις

$$J(\theta_0) = \int u_{\theta_0}(x)u^t_{\theta_0}(x)f^{1+a}_{\theta_0}(x)d\mu(x)$$

και

$$K(\theta_0) = \int u_{\theta_0}(x) u^t_{\theta_0}(x) f^{1+2a}_{\theta_0}(x) d\mu(x) - \int u_{\theta_0}(x) f^{1+a}_{\theta_0}(x) d\mu(x) \int u^t_{\theta_0}(x) f^{1+a}_{\theta_0}(x) d\mu(x),$$

αντίστοιχα. Σύμφωνα με το Πόρισμα 2.1 των Dik and de Gunst (1985), έχουμε

$$n(\hat{\theta} - \theta_0)^T J^\omega(\theta_0)(\hat{\theta} - \theta_0) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

όπου $Z_1, ...Z_r$, είναι ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές με

$$r = rank(AVar(\theta_0) J^\omega(\theta_0) AVar(\theta_0))$$

και $\beta_1, \beta_2, ..., \beta_r$ οι μη μηδενικές ιδιοτιμές του πίνακα

$$J^\omega(\theta_0) AVar(\theta_0).$$

Σύμφωνα με τα παραπάνω, τελικά έχουμε

$$E_g[n(\hat{\theta} - \theta_0)^t J^\omega(\theta_0)(\hat{\theta} - \theta_0)] = \sum_{i=1}^{r} \beta_i. \tag{4.27}$$

∎

Παρατήρηση 4.1 (Διαδικασία τοπικής επιλογής μοντέλου). *Στη συνέχεια περιγράφεται η γενική προσέγγιση της επιλογής μοντέλου, τοπικά, η οποία θα εφαρμοστεί στις Ενότητες 4.2 και 4.3. Συνοψίζοντας τη συζήτηση που προηγήθηκε, το κριτήριο πληροφορίας τοπικής απόκλισης* ($LDiv.IC$) *ορίζεται από τη σχέση*

$$L^\omega_{\alpha,n}(\hat{\theta}, \beta_1...\beta_r) = nQ^\omega_a(\hat{\theta}) + (\alpha + 1)\sum_{i=1}^{r} \beta_i. \tag{4.28}$$

Αρχικά υποθέτουμε ότι η πραγματική κατανομή g ανήκει στην παραμετρική οικογένεια $\{f_\theta\}$. Επιπλέον, έστω $\hat{\theta}$ ένας συνεπής και ασυμπτωτικά κανονικός εκτιμητής της παραμέτρου θ που προκύπτει χρησιμοποιώντας το τυχαίο δείγμα $X_1, ..., X_n$ από τον πληθυσμό με άγνωστη κατανομή $g \in \{f_\theta\}$. Τότε, για $\omega \in \Theta^*$ και $\alpha > 0$, η τιμή της $L^\omega_{\alpha,n}(\hat{\theta}, \beta_1...\beta_r)$, όπου η ποσότητα $Q^\omega_a(\hat{\theta})$ ορίστηκε στη σχέση (4.14) και $\beta_1, \beta_2, ..., \beta_r$ είναι οι μη μηδενικές ιδιοτιμές ενός συγκεκριμένου πίνακα που προσδιορίστηκε στην Πρόταση 3.1, μας παρέχει ένα κριτήριο για να επιλέξουμε το πιο κατάλληλο μοντέλο από μια συλλογή $\{f_\theta\}$ πιθανών κατάλληλων μοντέλων, σε κάποια περιοχή του $\mathcal{X}$, που καθορίζεται από την πυκνότητα

h_ω. Ειδικότερα, υπολογίζουμε την τιμή του κριτηρίου LDiv.IC μεταξύ δύο υποψήφιων μοντέλων $f^1_{\theta_1}$ και $f^2_{\theta_2}$ από το πραγματικό g, τοπικά, δηλαδή, υπολογίζουμε τις τιμές των $L^\omega_{\alpha,n}(\hat{\theta}_1, \beta_1...\beta_r)$ και $L^\omega_{\alpha,n}(\hat{\theta}_2, a_1...a_s)$ με βάση το τυχαίο δείγμα $X_1, ..., X_n$ και αν

$$L^\omega_{\alpha,n}(\hat{\theta}_1, \beta_1...\beta_r) < L^\omega_{\alpha,n}(\hat{\theta}_2, a_1...a_s), \tag{4.29}$$

τότε το $f_{\hat{\theta}_1}$ θεωρείται πιο κατάλληλο μοντέλο από το $f_{\hat{\theta}_2}$. Η διαδικασία συνεχίζεται παρόμοια για όλα τα υποψήφια μοντέλα.

Παρατήρηση 4.2. *Οι πίνακες $J^\omega(\theta_0)$ και $I_F(\theta_0)$ μπορούν να εκτιμηθούν αντικαθιστώντας την πραγματική αλλά άγνωστη παράμετρο θ_0 με έναν εκτιμητή $\hat{\theta}$. Έτσι*

$$I_F(\hat{\theta}) = \left(\int_{\mathcal{X}} f_\theta(x) \frac{\partial \log f_\theta(x_i)}{\partial \theta_i} \frac{\partial \log f_\theta(x_i)}{\partial \theta_j} d\mu(x) \bigg|_{\theta=\hat{\theta}} \right)_{i,j=1,\dots,M}, \tag{4.30}$$

και

$$J^\omega(\hat{\theta}) = \left(\int_{\mathcal{X}} h_\omega(x) f_\theta^{1+\alpha}(x) \frac{\partial \log f_\theta(x)}{\partial \theta_i} \frac{\partial \log f_\theta(x)}{\partial \theta_j} d\mu(x) \bigg|_{\theta=\hat{\theta}} \right)_{i,j=1,\dots,M}. \tag{4.31}$$

Παρατήρηση 4.3. *Όταν ο εκτιμητής $\hat{\theta}$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας (ΕΜΠ), ο πίνακας διακυμάνσεων-συνδιακυμάνσεων του $\sqrt{n}(\hat{\theta} - \theta_0)$ είναι ο αντίστροφος του πίνακα πληροφορίας Fisher $I_F^{-1}(\theta_0)$, με $\beta_1, \beta_2, ..., \beta_r$ να είναι οι μη μηδενικές ιδιοτιμές του πίνακα*

$$J^\omega(\theta_0) I_F^{-1}(\theta_0),$$

όπου

$$r = rank(I_F^{-1}(\theta_0) J^\omega(\theta_0) I_F^{-1}(\theta_0)).$$

Αυτό που διακρίνει τον ΕΜΠ σε σχέση με τον BHHJ εκτιμητή είναι ότι υπολογίζεται πολύ πιο γρήγορα και με μεγαλύτερη ακρίβεια (βλέπε, Mattheou et al., 2009, σελ. 233). Όπως προαναφέρθηκε, η μέθοδος εύρεσης του εκτιμητή ΕΜΠ είναι υπολογιστικά πιο γρήγορη από την μέθοδο των Basu et al. (1998). Αυτό οφείλεται στο γεγονός ότι ο ΕΜΠ δίνεται σε κλειστό τύπο στις περισσότερες των περιπτώσεων των γνωστών κατανομών σε αντίθεση με τη μέθοδο εύρεσης εκτιμητή που προτάθηκε από τους Basu et al. (1998), η οποία δεν παρέχει σε κλειστό τύπο εκτιμητές με αποτέλεσμα να χρησιμοποιούνται αριθμητικές μέθοδοι για την εύρεση των εκτιμητών αυτών. Τέτοιες αριθμητικές μέθοδοι συνήθως *σχετίζονται με σφάλματα που μπορεί να μην είναι ελέγξιμα. Κατά συνέπεια, η μέθοδος εύρεσης ΕΜΠ είναι περισσότερο ακριβής από τη μέθοδο Basu et al. (1998) και ταυτόχρονα ικανοποιεί τις ιδιότητες που απαιτούνται από την εν λόγω διαδικασία, δηλαδή τη συνέπεια και την ασυμπτωτική κανονικότητα. Οι πρακτικές συνέπειες αυτών των δύο μορφών του κριτήριου LDiv.IC θα αναδειχθούν στην Ενότητα 4.2 της προσομοίωσης.*

Παρατήρηση 4.4. *Για $\alpha \to 0$, το όριο της τοπικής απόκλισης BHHJ που ορίστηκε στην (2.56) είναι η τοπική Kullback-Leibler απόκλιση. Συνεπώς, μπορούμε να γράψουμε*

$$D_0^{\omega}(g, f_\theta) = \lim_{a \to 0} D_\alpha^{\omega}(g, f_\theta) = E_{f_\theta}(h_\omega(X)) - E_g(h_\omega(X)) + \int_{\mathcal{X}} h_\omega(x) g(x) \log \frac{g(x)}{f_\theta(x)} d\mu(x). \tag{4.32}$$

Έτσι, όταν $\alpha \to 0$, προκύπτει

$$W_0^{\omega}(\theta) = \lim_{\alpha \to 0} W_\alpha^{\omega}(\theta) = E_{f_\theta}(h_\omega(X)) - E_g(h_\omega(X) \log(f_\theta(X))), \tag{4.33}$$

και

$$Q_0^{\omega}(\theta) = \lim_{\alpha \to 0} Q_a^{\omega}(\theta) = E_{f_\theta}(h_\omega(X) - \frac{1}{n} \sum_{i=1}^{n} h_\omega(X_i) \log(f_\theta(X_i)). \tag{4.34}$$

Στην περίπτωση αυτή, το κριτήριο LDiv.IC γίνεται το αντίστοιχο του κλασικού κριτηρίου του Akaike σε τοπικό περιβάλλον. Θα ονομάζεται τοπικό Akaike κριτήριο (LAIC) και δίνεται από τη σχέση

$$LAIC = nQ_0^{\omega}(\hat{\theta}) + \sum_{i=1}^{r} \beta_i. \tag{4.35}$$

Χάριν πληρότητας, η παρατήρηση αυτή ολοκληρώνεται με την απόδειξη της σχέσης (4.32). Είναι (βλέπε, Basu et al., 1998),

$$
\begin{aligned}
D_0^{\omega}(g, f_\theta) &= \lim_{a\to 0} D_\alpha^{\omega}(g, f_\theta) \\
&= \lim_{a\to 0} \left\{ \int_{\mathcal{X}} h_\omega(x) \left(f_\theta^{1+a}(x) - (1+\tfrac{1}{a}) g(x) f_\theta^a(x) + \tfrac{1}{a} g^{1+a}(x) \right) dx \right\} \\
&= \lim_{a\to 0} \int_{\mathcal{X}} h_\omega(x) f_\theta^{1+a}(x) dx - \lim_{a\to 0} \int_{\mathcal{X}} h_\omega(x) g(x) f_\theta^a(x) dx \\
&+ \lim_{a\to 0} \int_{\mathcal{X}} h_\omega(x) \tfrac{g(x)(g^a(x) - f_\theta^a(x))}{\alpha} dx \\
&= \int_{\mathcal{X}} h_\omega(x) f_\theta(x) dx - \int_{\mathcal{X}} h_\omega(x) g(x) dx + \int_{\mathcal{X}} h_\omega(x) g(x) \lim_{a\to 0} \tfrac{(g^a(x) - f_\theta^a(x))}{\alpha} dx \\
&= E_{f_\theta}(h_\omega(X)) - E_g(h_\omega(X)) + \int_{\mathcal{X}} h_\omega(x) g(x) \lim_{a\to 0} \{ g^a(x) \log g(x) \\
&- f_\theta^a(x) \log f_\theta(x) \} dx \\
&= E_{f_\theta}(h_\omega(X)) - E_g(h_\omega(X)) + \int_{\mathcal{X}} h_\omega(x) g(x) \log \tfrac{g(x)}{f_\theta(x)} dx.
\end{aligned}
$$

4.2 Μελέτη προσομοίωσης

Στην ενότητα αυτή θα επιχειρηθεί, μέσω παραδειγμάτων που αφορούν τη μονοδιάστατη και πολυδιάστατη περίπτωση, να διερευνηθεί η αποτελεσματικότητα του προτεινόμενου τοπικού κριτηρίου πληροφορίας για την επιλογή μοντέλου. Ειδικότερα, θα παρουσιαστούν τα αποτελέσματα μιας μελέτης προσομοίωσης Monte Carlo με την αξιοποίηση του κριτηρίου $LDiv.IC$ για την επιλογή μοντέλων.

4.2.1 Μονοδιάστατη περίπτωση

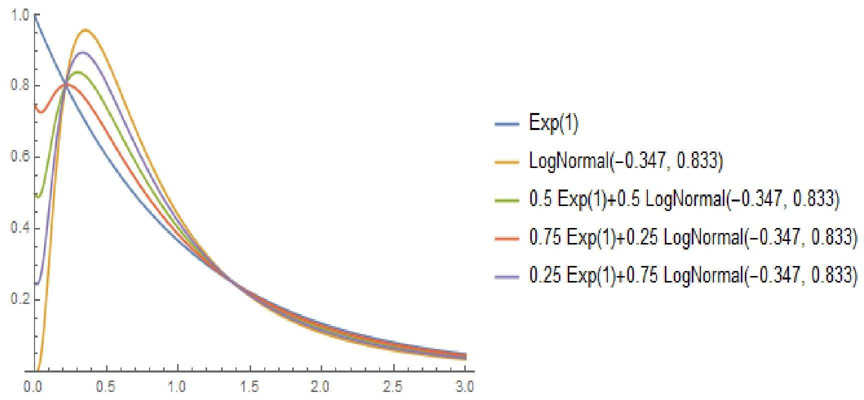
Στην υποενότητα αυτή το ενδιαφέρον επικεντρώνεται στη διερεύνηση της αποδοτικότητας του κριτηρίου $LDiv.IC$ για το πρόβλημα της επιλογής μεταξύ ενός εκθετικού μοντέλου, με πυκνότητα

$$f(x;\lambda) = f_\lambda(x) = \lambda \exp(-\lambda x),\ x > 0,\ \lambda > 0,$$

και ενός λογαριθμο-κανονικού μοντέλου, με πυκνότητα

$$g(x;\mu,\eta) = g_{\mu,\eta}(x) = \frac{1}{x\eta\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\eta^2} \right),\ x > 0,\ \mu \in \mathcal{R},\ \eta > 0,$$

τοπικά. Το πρόβλημα αυτό, όπως έχει αναφερθεί στην Ενότητα 4.1 (βλέπε και Σχήμα 4.1), έχει μακρά ιστορία στη στατιστική βιβλιογραφία, όταν η σύγκριση γίνεται σε όλο το πεδίο ορισμού των εν λόγω κατανομών.



Σχήμα 4.2: Γραφική απεικόνιση των πυκνοτήτων που χρησιμοποιούνται στη μελέτη προσομοίωσης (βλέπε, Πίνακες 4.1 και 4.2).

Με σκοπό να αξιολογηθεί η απόδοση του προτεινόμενου τοπικού κριτηρίου επιλογής μοντέλου, προσομοιώνουμε 1000 δείγματα (simulations runs=1000), μεγέθους $n = 100, 250, 500, 1000$ από τις μίξεις εκθετικής και λογαριθμο-κανονικής κατανομής, ήτοι από την

$$h(x;t) = tf(x;1) + (1-t)g(x;-0.347, 0.833),$$

για $t = 0.5, 1, 0, 0.75, 0.25$, με την τιμή $t = 1$ να αντιστοιχεί σε δειγματοληψία από την εκθετική με παράμετρο 1 και την τιμή $t = 0$ να αντιστοιχεί σε δειγματοληψία από την λογαριθμο-κανονική με παραμέτρους $(-0.347, 0.833)$. Η γραφική απεικόνιση των πυκνοτήτων που χρησιμοποιούνται στη μελέτη προσομοίωσης δίνεται στο Σχήμα 4.2.

Όπως έχει αναφερθεί στην Παρατήρηση 4.1 η εφαρμογή του προτεινόμενου τοπικού κριτηρίου επιλογής μοντέλου προϋποθέτει:
(α) την επιλογή μιας τιμής για την παράμετρο α, της τοπικής απόκλισης BHHJ.
(β) τον υπολογισμό ενός συνεπή και ασυμπτωτικά κανονικού εκτιμητή των παραμέτρων λ, μ και η της εκθετικής κατανομής και της λογαριθμικο-κανονικής κατανομής, και
(γ) την επιλογή πυκνότητα πυρήνα h_ω.

Όσον αφορά το (α) επιλέγεται η τιμή $\alpha = 0.1$. Η επιλογή του $\alpha = 0.1$ είναι αυθαίρετη. Ωστόσο, οποιαδήποτε από τις τυπικές μεθόδους μπορούν να χρησιμοποιηθούν για την κατάλληλη επιλογή του α (βλέπε για παράδειγμα, Diggle, 2013, σελ. 132). Όσον αφορά το (β) χρησιμοποιούνται στο παράδειγμα αυτό δύο μέθοδοι εκτίμησης των άγνωστων παραμέτρων. Ειδικότερα, οι εκτιμητές των παραμέτρων προκύπτουν είτε από την ελαχιστοποίηση του μέτρου των BHHJ (Πίνακας 4.1 και σ' αυτή την περίπτωση η παράμετρος η θεωρείται γνωστή και τίθεται εκ των προτέρων ίση με 0.833) είτε προκύπτουν με τη μέθοδο εκτίμησης της μέγιστης πιθανοφάνειας (Πίνακας 4.2). Στο σημείο αυτό, πρέπει να αναφερθεί ότι με τη μέθοδο BHHJ οι εκτιμητές των παραμέτρων δε δίνονται σε κλειστή, αναλυτική μορφή και υπολογίζονται επίλυοντας αριθμητικά την εξίσωση

$$\frac{1}{n}\sum_{i=1}^{n} u_\theta(X_i) f_\theta^\alpha(X_i) - \int_{\mathcal{X}} u_\theta(x) f_\theta^{1+a}(x)dx = 0, \tag{4.36}$$

(βλέπε, Basu et al., 1998), όπου $u_\theta(x) = \nabla_\theta \log(f_\theta(x))$ και $\alpha = 0.1$. Τέλος, όσον αφορά το (γ), την επιλογή του πυρήνα, στο παράδειγμα αυτό θεωρήθηκε ότι ο πυρήνας h_ω είναι η σ.π.π. της περικομμένης (truncated) κανονικής κατανομής με παράμετρο $\omega = (\mu, \sigma^2 = 0.1^2)$, για διάφορες επιλογές της παραμέτρου μ. Ειδικότερα, παρατίθενται τα αποτελέσματα για $\mu = 0.6, 1, 1.5, 2, 3$.

Στο πλαίσιο αυτό, για κάθε συνδυασμό του δειγματοληπτικού πλαισίου, του μεγέθους δείγματος, της επιλογής πυρήνα και της μεθόδου εκτίμησης (ήτοι συνολικά για $5\times4\times5\times2 = 200$ συνδυασμούς) καταγράφεται το ποσοστό επιλογής καθενός εκ των δύο ανταγωνιστικών μοντέλων, στη βάση 1000 προσομοιώσεων. Τα αποτελέσματα παρατίθενται στους Πίνακες 4.1 και 4.2. Από τα αποτελέσματα αυτά εξάγονται τα ακόλουθα συμπεράσματα:

(α) Καθώς το μέγεθος του δείγματος αυξάνεται το τοπικό κριτήριο $LDiv.IC$ επιλέγει αποτελεσματικά το πραγματικό μοντέλο, τοπικά (βλέπε, Πίνακες 4.1 και 4.2, περιπτώσεις b και c). Αυτό επιτυγχάνεται εύκολα σε περιοχές του πεδίου ορισμού (όπως καθορίζονται από τον πυρήνα), όπου τα υποψήφια μοντέλα είναι σαφώς διαφορετικά (περιπτώσεις b και c των Πινάκων 4.1 και 4.2, για $\mu = 0.6, 1, 1.5, 2$ και 3). Για παράδειγμα, για πυρήνα που εστιάζει στο $\mu = 0.6$ το πραγματικό μοντέλο επιλέγεται 99,9% ή 100% των φορών (βλέπε, Πίνακα 4.1 για $n = 1000$).

(β) Στην περίπτωση που το πραγματικό μοντέλο είναι μίξη των μοντέλων $Exp(1)$ και $LogNormal(-0.347,\ 0.833)$ (βλέπε, Πίνακες 4.1 και 4.2, περιπτώσεις a, d και e), εστιάζοντας σε συγκεκριμένες περιοχές, το τοπικό κριτήριο $LDiv.IC$ μπορεί να καθορίσει το πραγματικό μοντέλο ακόμη και αν τα δεδομένα προκύπτουν από μοντέλο μείξη (π.χ. στις περιπτώσεις d και e του Πίνακα 4.1, $\mu = 0.6$, για $n = 1000$, ή στην περίπτωση e στον Πίνακα 4.2, $\mu = 0.6$, $n = 1000$).

Πίνακας 4.1: Ποσοστά επιλογής μεταξύ ενός εκθετικού ή λογαριθμο-κανονικού υποψήφιου μοντέλου για διάφορα μεγέθη δείγματος και πυρήνες χρησιμοποιώντας τους εκτιμητές BHHJ για λ και μ. Η παράμετρος η θεωρείται γνωστή και ίση με 0.833. Ο πυρήνας είναι μια περικομμένη (truncated) κανονική κατανομή για διάφορους μέσους μ και σταθερή τυπική απόκλιση $\sigma = 0.1$. Το πραγματικό μοντέλο το οποίο παράγει τα δεδομένα φαίνεται σε κάθε τμήμα (περιπτώσεις (a)-(e)).

n	100		250		500		1000	
Υποψήφια μοντέλα	*Exp*	*LogN*	*Exp*	*LogN*	*Exp*	*LogN*	*Exp*	*LogN*
Πυρήνας: Περικομμένη Κανονική ($\mu, \sigma = 0.1$)								
(a)	Πραγματικό μοντέλο: $0.5Exp(1) + 0.5LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	53.1	46.9	48.5	51.5	50.2	49.8	50.3	49.7
1	49.5	50.5	48.3	51.7	39	61	40	60
1.5	58.8	41.2	58.8	41.2	64.5	35.5	67.2	32.8
2	53.4	46.6	56.1	43.9	58.7	41.3	66.4	33.6
3	48	52	51.5	48.5	54.7	45.3	60.7	39.3
(b)	Πραγματικό μοντέλο: *Exp(1)*							
$\mu = 0.6$	84.2	15.8	95.9	4.1	99.1	0.9	99.9	0.1
1	57	43	66.4	33.6	72.4	27.6	80	20
1.5	64.6	35.4	72.7	27.3	78.	21.1	87.2	12.8
2	67.3	32.7	78.3	21.7	85.1	14.9	94.3	5.7
3	61.4	38.6	66.8	33.2	79.9	20.1	89.5	10.5
(c)	Πραγματικό μοντέλο: $LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	16.4	83.6	6	94	1	99	0	100
1	36.6	63.4	28.3	71.7	19.9	80.1	7.7	92.3
1.5	55	45	49.5	50.5	44	56	42.2	57.8
2	42.6	57.4	34.8	65.2	28.3	71.7	21.2	78.8
3	28.4	71.6	27.7	72.3	26.6	73.4	22.4	77.6
(d)	Πραγματικό μοντέλο: $0.75Exp(1) + 0.25LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	68.1	31.9	80.6	19.4	90.6	9.4	95.5	4.5
1	54.3	45.7	53.5	46.5	60.6	39.4	59.9	40.1
1.5	60.8	39.2	66.7	33.3	73.6	26.4	80.8	19.2
2	58.2	41.8	67.4	32.6	74.7	25.3	86.5	13.5
3	52.2	47.8	60.7	39.3	69.2	30.8	20.7	79.3
(e)	Πραγματικό μοντέλο: *0.25Exp(1) + 0.75LogNormal(−0.347, 0.833)*							
$\mu = 0.6$	28.7	71.3	21.8	78.2	13.8	86.2	4.5	95.5
1	44.3	55.7	36.4	63.6	30.1	69.9	22.5	77.5
1.5	56.1	43.9	54.2	45.8	54.1	45.9	51.5	48.5
2	43.1	56.9	47	53	46.1	53.9	40.3	59.7
3	37.1	62.9	41.8	58.2	41	59	41.3	58.7

Πίνακας 4.2: Ποσοστά επιλογής μεταξύ ενός εκθετικού ή λογαριθμο-κανονικού υποψήφιου μοντέλου για διάφορα μεγέθη δείγματος και πυρήνες χρησιμοποιώντας τους εκτιμητές μέγιστης πιθανοφάνειας (ΕΜΠ) για όλες τις παραμέτρους των υποψήφιων μοντέλων. Ο πυρήνας είναι μια περικομμένη (truncated) κανονική κατανομή για διάφορες μέσες τιμές μ και σταθερή τυπική απόκλιση $\sigma = 0.1$. Το πραγματικό μοντέλο το οποίο παράγει τα δεδομένα φαίνεται σε κάθε τμήμα (περιπτώσεις (a)-(e)).

n	100		250		500		1000	
Υποψήφια μοντέλα	*Exp*	*LogN*	*Exp*	*LogN*	*Exp*	*LogN*	*Exp*	*LogN*
Πυρήνας: Περικομμένη Κανονική ($\mu, \sigma = 0.1$)								
(a)	Πραγματικό μοντέλο: $0.5Exp(1) + 0.5LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	39.7	60.3	24.4	75.6	12.1	87.9	4.2	95.8
1	67.1	32.9	74	26	77.8	22.2	87	13
1.5	64.2	35.8	65.4	34.6	70.6	29.4	75.5	24.5
2	52.1	47.9	54.1	45.9	55.1	44.9	57.4	42.6
3	47.1	52.9	43.9	56.1	43.8	56.2	37.9	62.1
(b)	Πραγματικό μοντέλο: $Exp(1)$							
$\mu = 0.6$	73.4	26.6	79.6	20.4	79.6	20.4	87.6	12.4
1	65.7	34.3	78.7	21.3	78.7	21.3	85.1	14.9
1.5	68.3	31.7	83.6	16.4	83.6	16.4	90.9	9.1
2	65.2	34.8	82.1	17.9	82.1	17.9	89.5	10.5
3	56.1	43.9	61.5	38.5	61.5	38.5	69	31
(c)	Πραγματικό μοντέλο: $LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	11.9	88.1	2.4	97.6	0.4	99.6	0	100
1	37.1	62.9	22	78	11.5	88.5	3.6	96.4
1.5	61.8	38.2	57	43	48.9	51.1	44.2	55.8
2	45	55	33.6	66.4	32.4	67.6	19.3	80.7
3	39.9	60.1	32.4	67.6	23.5	76.5	14.2	85.8
(d)	Πραγματικό μοντέλο: $0.75Exp(1) + 0.25LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	44.5	55.5	50.1	49.9	61.1	38.9	41.1	58.9
1	83.2	16.8	74.1	25.9	72.2	27.8	89.5	10.5
1.5	78.6	21.4	71	29	66.5	33.5	85	15
2	70.1	29.9	59.9	40.1	58.5	41.5	78.2	21.8
3	50.9	49.1	51.7	48.3	49.2	50.8	55.6	44.4
(e)	Πραγματικόό μοντέλο: $0.25Exp(1) + 0.75LogNormal(-0.347, 0.833)$							
$\mu = 0.6$	25.4	74.6	8.3	91.7	2.2	97.8	0.1	99.9
1	56.5	43.5	46	54	35.6	64.4	26.3	73.7
1.5	57.7	42.3	58.7	41.3	63.3	36.7	67.5	32.5
2	46.9	53.1	43.1	56.9	42.7	57.3	37.9	62.1
3	44.6	55.4	37.8	62.2	36.2	63.8	24.4	75.6

4.2.2 Πολυδιάστατη περίπτωση: Παράδειγμα point process

Θεωρούμε μια περιοχή $\mathcal{W} \subset R^2$ και ας υποθέσουμε ότι παρατηρούμε n σημεία $\{x_i\}_{i=1}^n$. Για τη μοντελοποίηση αυτής της συλλογής σημείων χρησιμοποιώντας point process μεθοδολογία, θεωρούμε ότι τα παρατηρούμενα σημεία $\{x_i\}_{i=1}^n$ προκύπτουν από ένα μοντέλο με δύο πηγές τυχαιότητας: 1) τον τυχαίο αριθμό των σημείων n και 2) τα εν' λόγω σημεία παράγονται τυχαία στην περιοχή $\mathcal{W}$, δοθέντος n. Τα παραγόμενα σημεία (γεγονότα) ονομάζονται μοτίβο σημείων (point pattern) και αντιμετωπίζονται ως μια point process N στο παράθυρο $\mathcal{W}$.

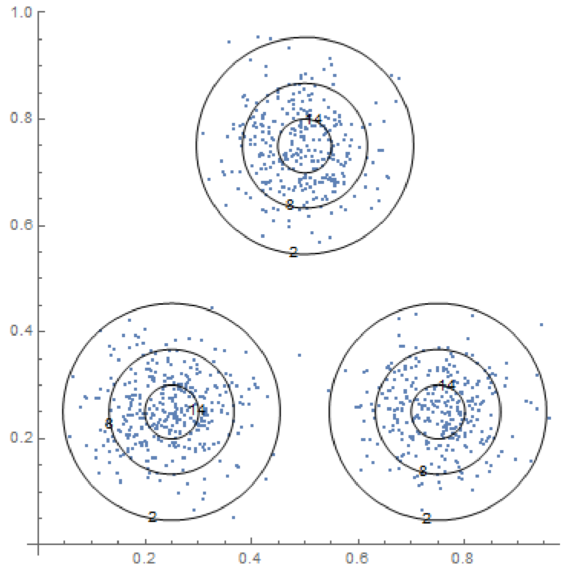
Ομοίως, με την πρώτη ροπή για τυχαίες μεταβλητές οι point processes χαρακτηρίζονται από την αντίστοιχη συνάρτηση έντασης $\lambda(x)$ (intensity function), $x \in \mathcal{W}$, όπου το $\lambda(x)dx$ είναι η πιθανότητα παρατήρησης ενός σημείου σε έναν απειροελάχιστο δίσκο (χώρο R^d) με κέντρο x και όγκο dx. Φυσικά η πρώτη ροπή δεν χαρακτηρίζει μοναδικά την point process, εκτός και εάν πληρούνται ορισμένες προϋποθέσεις (βλέπε, Gressie, 1993). Πιο συγκεκριμένα, έστω ότι $N(\mathcal{W})$ συμβολίζει τον αριθμό των σημείων στο παράθυρο $\mathcal{W}$, και υποθέτουμε ότι $N(\mathcal{W}) \sim$ Poisson $(\Lambda(\mathcal{W}))$, όπου $\Lambda(\mathcal{W})$ συμβολίζει τη μέση τιμή της point process, δηλαδή $E(N(\mathcal{W})) = \Lambda(\mathcal{W})$. Επιπλέον υποθέτουμε ότι οι μετρήσεις σε μη επικαλυπτόμενες περιοχές είναι ανεξάρτητες τυχαίες μεταβλητές. Αυτές οι δύο παραδοχές οδηγούν στο πιο σημαντικό, ίσως, μοντέλο point process, που είναι γνωστό ως η μη ομογενής (inhomogeneous) Poisson point process N $(IPPP)$ με μέτρο έντασης Λ. Αν επιπλέον το μέτρο Λ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue, τότε υπάρχει μια μετρήσιμη συνάρτηση $\lambda(x|\theta)$ τέτοια ώστε

$$\Lambda_\theta(\mathcal{W}) = \int_{\mathcal{W}} \lambda(x|\theta)dx, \tag{4.37}$$

για κάποια παράμετρο θ. Στην περίπτωση αυτή, η από κοινού κατανομή των σημείων που παρατηρήθηκαν σε κάποιο παράθυρο $\mathcal{W}$, δοθέντος $N(\mathcal{W}) = n$, δίνεται από τη σχέση

$$f(x_1, ..., x_n|\theta, N(\mathcal{W}) = n) = \prod_{i=1}^{n} \frac{\lambda(x_i|\theta)}{\Lambda_\theta(\mathcal{W})}, \text{ για } x_i \in \mathcal{W}, i = 1, ..., n. \tag{4.38}$$

Η συνάρτηση $\lambda(x|\theta)$ ονομάζεται συνάρτηση έντασης (intensity function) και προσδιορίζει μοναδικά τη συνάρτηση κατανομής της point process. Σκοπός μας είναι να εκτιμήσουμε την παράμετρο θ με ένα ισχυρό-εύρωστο τρόπο (robust way), ενώ αλλάζουμε την περιοχή εστίασης στο παράθυρο παρατηρήσεων $\mathcal{W}$. Επιλέγουμε να μοντελοποιηθεί η



Σχήμα 4.3: Μη ομογενής Poisson point process με $\xi = 1025$ και ισοϋψείς από τρεις διδιάστατους κανονικούς πυρήνες. Υπάρχουν $n = 1000$ σημεία (παρατηρηθέντα γεγονότα) στο παράθυρο $(0, 1) \times (0, 1)$.

συνάρτηση έντασης με ένα πολλαπλάσιο μιας πυκνότητας $\lambda_\theta(x)$ (βλέπε, Micheas, 2014), δηλαδή,

$$\lambda(x|\theta, \xi) = \xi\lambda_\theta(x),\ \xi > 0, \tag{4.39}$$

η οποία δεν είναι μια πυκνότητα γενικά.

Προκειμένου να αξιολογηθεί το προτεινόμενο τοπικό κριτήριο πληροφορίας, πραγματοποιήθηκε Monte Carlo μελέτη προσομοίωσης με τη χρήση του κριτηρίου $LDiv.IC$. Το Σχήμα 4.3 απεικονίζει $n = 1000$ γεγονότα που παρατηρήθηκαν στο παράθυρο $\mathcal{W} = (0, 1) \times (0, 1)$, τα οποία προκύπτουν ως υλοποίηση μιας μη ομογενούς Poisson point process με $\xi = 1025$ και $\lambda_\theta(x)$ ένα μοντέλο μίξης τριών κανονικών κατανομών. Συγκεκριμένα,

$$\lambda_\theta(x) = 0.4f(x; (0.25, 0.25), \Sigma) + 0.3f(x; (0.75, 0.25), \Sigma) + 0.3f(x; (0.5, 0.75), \Sigma),$$

με $\Sigma = \begin{pmatrix} 0.005 & 0 \\ 0 & 0.005 \end{pmatrix}$ και $f(x; \mu, \Sigma)$ η διδιάστατη κανονική κατανομή με μέσο διάνυσμα μ και πίνακα διακύμανσης-συνδιακύμανσης Σ.

Θεωρούμε το πρόβλημα επιλογής μεταξύ τριών υποψήφιων μοντέλων, τοπικά. Ως υποψήφια μοντέλα θεωρούμε τις διδιάστατες κανονικές συνιστώσες της κατανομής $\lambda_\theta(x)$, δηλαδή τις διδιάστατες κανονικές κατανομές

$$f(x;\mu_i,\Sigma_i) = (2\pi)^{-1}|\Sigma_i|^{-1/2}\exp\left(-\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)\right),$$

με μέσα διάνυσματα μ_i, $i = 1,2,3$ και πίνακες διακυμάνσεων-συνδιακυμάνσων Σ_i, $i = 1,2,3$. Χρησιμοποιώντας την κλασική μέθοδο προσομοίωσης για *IPPP* των Lewis and Shedler (1979), η μελέτη προσομοίωσης με βάση το κριτήριο *LDiv.IC* έχει τα ακόλουθα χαρακτηριστικά:

- Παράγεται ο αριθμός των σημείων $N(\mathcal{W}) \sim$ Poisson $(\Lambda(\mathcal{W}))$, όπου $\Lambda(\mathcal{W}) = \int_{\mathcal{W}} \xi\lambda_\theta(x)dx$. Έστω $N(\mathcal{W}) = n$.
- Υπολογίζουμε την τιμή $\lambda^* = \max\limits_{x\in\mathcal{W}}\lambda_\theta(x)dx$.
- Παράγεται ένα σημείο $x_0 \in \mathcal{W}$ ομοιόμορφα στο παράθυρο $\mathcal{W}$ και έστω $U \sim \mathcal{U}(0,1)$. Ελέγχουμε αν ισχύει $U < \frac{\lambda_\theta(x_0)}{\lambda^*}$, αν ναι, κρατάμε το σημείο x_0. Επαναλαμβάνουμε την διαδικασία μέχρι να κρατήσουμε n-σημεία. Τα σημεία αυτά σχηματίζουν ένα μοτίβο σημείων από την υποψήφια Poisson point process.
- Υπολογίζουμε τις τιμές των *AIC* και *DIC* για $\alpha = 0.1$. Σημειώνεται ότι, $AIC = -2L(\hat\theta) + 2M$, όπου $L(\theta)$ η συνάρτηση πιθανοφάνειας και $DIC_{\text{ΕΜΠ}} = nQ_{\hat\theta} + (2\pi)^{-\alpha/2}(1+\alpha)^{-M/2}M$, με $Q_\theta = \int_{\mathcal{X}} f_\theta^{1+\alpha}(x)d\mu(x) - (1+a^{-1})\frac{1}{n}\sum\limits_{i=1}^{n} f_\theta^\alpha(X_i)$ (βλέπε, Mattheou et al., 1998, σελ. 232).
- Υπολογίζουμε το κριτήριο *LDiv.IC* για διάφορους πυρήνες από τη διδιάστατη κανονική κατανομή. Πιο συγκεκριμένα, επιλέγουμε το μέσο διάνυσμα μ_0 να είναι ένα από τα διανύσματα $(0.25, 0.25)$, $(0.75, 0.75)$ και $(0.5, 0.75)$, και για πίνακα διακύμανσης-συνδιακύμανσης Σ_0 επιλέγουμε τον πίνακα $\Sigma_0 = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$.

Τα αποτελέσματα παρουσιάζονται στον Πίνακα 4.3, όπου οι τιμές του τοπικού κριτήριου υπολογίζονται για τρεις πυρήνες στο παράθυρο $\mathcal{W} = (0,1)\times(0,1)$. Παρατηρούμε ότι στην περιοχή με κέντρο $(0.25, 0.25)$ επιλέγεται η πρώτη κανονική συνιστώσα, στην περιοχή με κέντρο $(0.75, 0.25)$ επιλέγεται η δεύτερη κανονική συνιστώσα, ενώ η τρίτη κανονική

Πίνακας 4.3: Τιμές του κριτηρίου πληροφορίας $LDiv.IC$ για τρία υποψήφια μοντέλα και για διάφορους πυρήνες. Το Μοντέλο 1 είναι μια διδιάστατη κανονική κατανομή που αντιστοιχεί στη πρώτη συνιστώσα του μοντέλου μείξης, το Μοντέλο 2 αντιστοιχεί στη δεύτερη και το Μοντέλο 3 αντιστοιχεί στην τρίτη συνιστώσα. Ο πυρήνας είναι διδιάστατη κανονική κατανομή με διάφορες μέσες τιμές μ_0 και σταθερό Σ_0. Στην τελευταία στήλη φαίνεται το επιλεγμένο μοντέλο.

	Μοντέλο 1	Μοντέλο 2	Μοντέλο 3	Επιλογή
Πυρήνας: Διδιάστη Κανονική ($\mu_0, \Sigma_0 = [(0.01, 0), (0, 0.01)]$)				
	$LDiv.IC$	$LDiv.IC$	$LDiv.IC$	Επιλεγόμενο Μοντέλο
$(0.25, 0.25)$	-42416.8	-5340.27	-2933.20	1
$(0.75, 0.25)$	-4832.02	-36591.1	-2632.57	2
$(0.5, 0.75)$	-2805.56	-2749.72	-34964.1	3
AIC	30594.2	33136.6	37800.5	1
DIC	-7508.30	-6865.79	-6500.77	1

συνιστώσα επιλέγεται στην περιοχή με κέντρο $(0.5, 0.75)$. Αποτέλεσμα της εφαρμογής του κριτηρίου $LDiv.IC$, στο παράδειγμα αυτό, είναι η επιλογή του πραγματικού μοντέλου σε κάθε μια περιοχή που εστιάζουμε.

Επιπλέον, παρουσιάζονται οι τιμές των κριτηρίων AIC και DIC. Θα πρέπει να σημειωθεί ότι και τα δύο κλασικά κριτήρια (AIC και DIC) επιλέγουν τη συνιστώσα με το μεγαλύτερο βάρος στο πραγματικό μοντέλο. Ως εκ τούτου, τα κλασικά κριτήρια επιλογής μοντέλου παρέχουν λανθασμένη πληροφορία σε αυτήν την περίπτωση, δεδομένου ότι χάνονται τα τοπικά χαρακτηριστικά, και έτσι επιβεβαιώνεται, μέσω του παραδείγματος αυτού, ότι τα εν' λόγω κριτήρια δεν περιλαμβάνουν έναν μηχανισμό που να μπορεί να συλλάβει, να μετουσιώσει και να εκφράσει αυτήν την τοπική συμπεριφορά.

Σε αντίθεση, το τοπικό κριτήριο $LDiv.IC$ είναι σε θέση, από το τρόπο κατασκευής του, να εντοπίσει και να συμπεριλάβει τοπικά χαρακτηριστικά για τη λήψη τεκμηριωμένης απόφασης σχετικά με τα ανταγωνιστικά μοντέλα. Τέλος, η επιλογή του «καλύτερου» μοντέλου σε κάθε περιοχή εστίασης μας επιτρέπει να εκτιμήσουμε τις παραμέτρους του μοντέλου με έναν «εύρωστο-ισχυρό» τρόπο καθώς η περιοχή εστίασης αλλάζει, δεδομένου ότι είμαστε σε θέση να ανακαλύψουμε τις παραμέτρους της πραγματικής συνιστώσας που γεννά τα δεδομένα.

4.3 Εφαρμογές

Στην ενότητα αυτή το προτεινόμενο τοπικό κριτήριο πληροφορίας για την επιλογή μοντέλου εφαρμόζεται σε τρία γνωστά στη βιβλιογραφία σύνολα δεδομένων, προκειμένου και πάλι να διερευνηθεί η εφαρμογή του σε πραγματικά προβλήματα και να αξιολογηθεί η συμπεριφορά του.

4.3.1 Μονοδιάστατη περίπτωση: Γαλαξιακά δεδομένα (galaxy data)

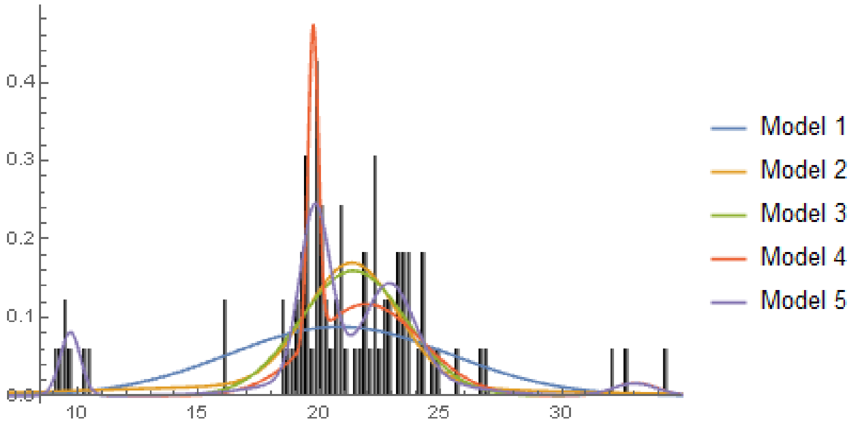
Το πρώτο σύνολο δεδομένων μελετήθηκε για πρώτη φορά από τους Postman et al. (1986). Στο σύνολο αυτό των δεδομένων καταγράφεται η ταχύτητα μετρημένη, σε km^3/sec, 82 γαλαξίων από τις έξι κωνικές περιοχές του αστερισμού Borealist. Ο Roeder (1992) ήταν ο πρώτος που εφάρμοσε ένα μοντέλο μείξης κατανομών σε αυτά τα δεδομένα και από τότε αυτό το σύνολο δεδομένων αποτέλεσε σημείο αναφοράς πολλών ερευνητικών εργασιών. Για μια συγκριτική παρουσίαση των διαφορετικών προσεγγίσεων που χρησιμοποιούνται για την ανάλυση αυτών των δεδομένων παραπέμπουμε στον Aitkin (2001).

Οι περισσότεροι από τους ερευνητές πιστεύουν ότι οι παρατηρήσεις του προαναφερθέντος συνόλου δεδομένων προκύπτουν από ένα μεικτό μοντέλο με κανονικές συνιστώσες. Στο Σχήμα 4.4 παρουσιάζεται το ιστόγραμμα των δεδομένων. Στο ιστόγραμμα παρουσιάζονται κενά μέσα στον κύριο όγκο των παρατηρήσεων, καθώς και ορισμένες ακραίες τιμές που συγκεντρώνονται κοντά στις ταχύτητες 10 και 32. Ως εκ τούτου, είναι λογικό να υποθέσουμε ότι τα δεδομένα προέρχονται από μια μείξη από τουλάχιστον τρεις συνιστώσες και η υπόθεση αυτή είναι η μόνη στην οποία αρκετοί ερευνητές φαίνεται να συμφωνούν. Ειδικότερα, οι περισσότερες εργασίες στην βιβλιογραφία εκτιμούν ότι ο αριθμός των συνιστωσών είναι μεταξύ 3 και 8. Ακολούθως, θεωρούμε για τα δεδομένα αυτά τα εξής πέντε υποψήφια μοντέλα: κανονική κατανομή, μείξη δύο κανονικών κατανομών, μείξη τριών κανονικών κατανομών, μείξη τεσσάρων κανονικών κατανομών και μείξη πέντε κανονικών κατανομών. Οι εκτιμητές, με τη μέθοδο της μέγιστης πιθανοφάνειας, των παραμέτρων κάθε συνιστώσας της μείξης και των αντίστοιχων συντελεστών βάρους, δίνονται παρακάτω:

Μοντέλο 1: $\hat{\theta}_1 = (\hat{\mu_1}, \hat{\sigma_1}^2) = (20.83, 4.54^2)$,

Μοντέλο 2: $\hat{\theta}_2 = (\hat{\mu_2}, \hat{\mu_3}, \hat{\sigma_2}^2, \hat{\sigma_3}^2, \hat{w_2}, \hat{w_3}) = (21.35, 19.36, 1.88^2, 8.15^2, 0.74, 0.26)$,

Μοντέλο 3: $\hat{\theta}_3 = (\hat{\mu_4}, \hat{\mu_5}, \hat{\mu_6}, \hat{\sigma_4}^2, \hat{\sigma_5}^2, \hat{\sigma_6}^2, \hat{w_4}, \hat{w_5}, \hat{w_6}) = (33.04, 21.40, 9.71, 0.92^2,$



Σχήμα 4.4: Ιστόγραμμα των γαλαξιακών δεδομένων και οι πυκνότητες πέντε υποψήφιων μοντέλων, όπου το μοντέλο i αποτελείται από ένα μείγμα με i μονοδιάστατες κανονικές συνιστώσες, $i = 1, ..., 5$.

2.20^2, 0.42^2, 0.037, 0.878, 0.085),

Μοντέλο 4: $\hat{\theta}_4 = (\hat{\mu_7}, \hat{\mu_8}, \hat{\mu_9}, \hat{\mu_{10}}, \hat{\sigma_7}^2, \hat{\sigma_8}^2, \hat{\sigma_9}^2, \hat{\sigma_{10}}^2, \hat{w_7}, \hat{w_8}, \hat{w_9}, \hat{w_{10}}) = (33.05$, 21.94, 19.75, 9.71, 0.92^2, 2.27^2, 0.45^2, 0.42^2, 0.037, 0.665, 0.213, 0.085), και

Μοντέλο 5: $\hat{\theta}_5 = (\hat{\mu_{11}}, \hat{\mu_{12}}, \hat{\mu_{13}}, \hat{\mu_{14}}, \hat{\mu_{15}}, \hat{\sigma_{11}}^2, \hat{\sigma_{12}}^2, \hat{\sigma_{13}}^2, \hat{\sigma_{14}}^2, \hat{\sigma_{15}}^2, \hat{w_{11}}, \hat{w_{12}}, \hat{w_{13}}, \hat{w_{14}}, \hat{w_{15}}) = (33.04$, 22.92, 21.85, 19.82, 9.71, 0.92^2, 1.02^2, 3.05^2, 0.63^2, 0.42^2, 0.036, 0.289, 0.245, 0.344, 0.085).

Τέλος, στον Πίνακα 4.4 παρατίθενται οι τιμές του κριτηρίου πληροφορίας $LDiv.IC$ για διάφορες επιλογές του κανονικού πυρήνα, καθώς και η τιμή του καθολικού κριτηρίου AIC. Από το καθολικό κριτήριο AIC εξάγεται το συμπέρασμα ότι το πιο κατάλληλο μοντέλο μεταξύ των πέντε υποψήφιων είναι η μείξη τεσσάρων κανονικών κατανομών (τιμή $AIC = 418.916$).

Όσον αφορά την εφαρμογή του τοπικού κριτηρίου επιλογής μοντέλου αρχικά επιλέγουμε να διαιρέσουμε το εύρος των παρατηρήσεων με τρεις κανονικούς πυρήνες με μέσες τιμές 10, 20 και 30 και σταθερή τυπική απόκλιση $\sigma = 1.67$. Στην περίπτωση αυτή και με τη χρήση του τοπικού κριτήριου διαφορετικές μείξης κανονικών κατανομών επιλέγονται σε κάθε διάστημα, ως το καλύτερο μοντέλο, με τα διαστήματα να καθορίζονται από κανονικούς πυρήνες (βλέπε, τμήμα (a) του Πίνακα 4.4). Στη συνέχεια, διερευνούμε λεπτότερες

Πίνακας 4.4: Τιμές του κριτήριου πληροφορίας $LDiv.IC$ για τα γαλαξιακά δεδομένα και την επιλογή μεταξύ πέντε υποψήφιων μοντέλων για διάφορους πυρήνες χρησιμοποιώντας ΕΜΠ. Ο πυρήνας είναι κανονικός για διάφορες μέσες τιμές μ_0 και σταθερή σ_0^2. Τα υποψήφια μοντέλα είναι μείξεις μονοδιάστατων κανονικών συνιστωσών από 1-5 συνιστώσες. Στην τελευταία στήλη παρουσιάζεται το μοντέλο που επιλέγεται, από τα υποψήφια πέντε μοντέλα.

	Μοντέλο 1	Μοντέλο 2	Μοντέλο 3	Μοντέλο 4	Μοντέλο 5	
(a) Πυρήνας Κανονικός ($\mu_0 = 10, 20, 30,\ \sigma_0^2 = 1.67^2$)						
$(10, 1.67^2)$	-9.8125	-9.9070	-11.3280	-11.3286	-11.3127	4
$(20, 1.67^2)$	-78.9007	-79.3333	-79.7111	-80.2637	-80.3412	5
$(30, 1.67^2)$	-1.0762	-1.5513	-1.6083	-1.5919	-1.51024	3
(b) Πυρήνας Κανονικός ($\mu_0 = 10, 16, 22, 28,\ \sigma_0^2 = 1^2$)						
$(10, 1^2)$	-15.4213	-15.3812	-17.4966	-17.4969	-17.4888	4
$(16, 1^2)$	-4.00205	-5.30619	-5.28678	-5.21688	-5.27787	2
$(22, 1^2)$	-77.1162	-76.9099	-77.0492	-77.3723	-77.2399	4
$(28, 1^2)$	-1.93588	-2.63204	-2.76677	-2.75332	-2.61609	3
(c) Πυρήνας Κανονικός ($\mu_0 = 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30,\ \sigma_0^2 = 0.4175^2$)						
$(10, 0.4175^2)$	-23.2203	-22.9663	-25.6755	-25.6759	-25.6697	4
$(12.5, 0.4175^2)$	0.9708	0.2343	0.0054	0.0035	0.0755	4
$(15, 0.4175^2)$	1.9243	-0.0748	-0.1721	-0.2238	-0.0871	4
$(17.5, 0.4175^2)$	2.8029	0.1507	0.8093	-0.0772	-0.2833	5
$(20, 0.4175^2)$	-142.619	-144.392	-145.232	-146.537	-146.858	5
$(22.5, 0.4175^2)$	-85.8838	-85.9618	-86.5534	-86.402	-85.971	3
$(25, 0.4175^2)$	-20.4891	-20.9452	-20.8027	-20.701	-20.657	2
$(27.5, 0.4175^2)$	-2.9717	-3.7274	-3.7838	-3.7863	-3.6396	4
$(30, 0.4175^2)$	0.6793	0.2212	0.0676	0.0718	0.1642	3
AIC	484.676	472.553	424.360	418.916	422.132	4

περιοχές του πεδίου ορισμού επιλέγοντας μικρότερη τυπική απόκλιση στους κανονικούς πυρήνες ($\sigma = 1$ και $\sigma = 0.4175$) και θεωρούμε ανάλογες μέσες τιμές. Στα τμήματα (b) και (c) του Πίνακα 4.4 παρατίθενται τα αποτελέσματα από όπου προκύπτει ότι διαφορετικά μοντέλα επιλέγονται ανάλογα με την περιοχή εστίασης (βλέπε, στήλη 5 τμήματα (b) και (c) του Πίνακα 4.4). Πιο συγκεκριμένα, για όλα τα τμήματα (a), (b) και (c) του Πίνακα 4.4, η μείξη με τέσσερις κανονικές συνιστώσες αναδεικνύεται ως το μοντέλο που επί το πλείστον επιλέγεται (7/16 φορές) ανεξάρτητα από την περιοχή που εστιάζουμε να υπολογίσουμε το κριτήριο τοπικής επιλογής. Αυτό το συμπέρασμα ταυτίζεται με το αντίστοιχο που διατυπώνεται με χρήση του καθολικού κριτηρίου AIC.

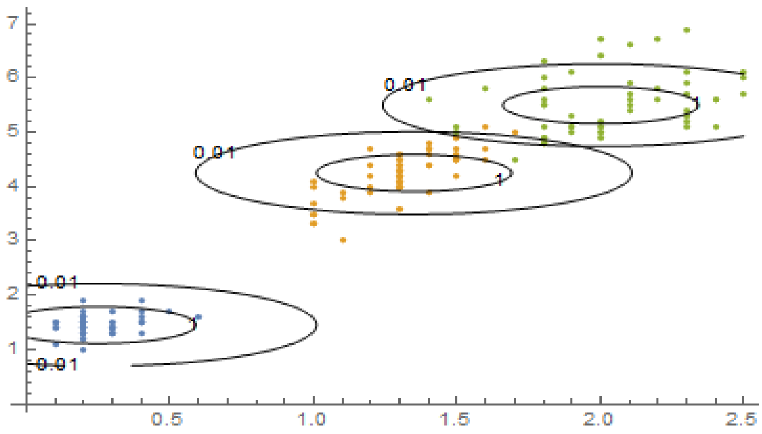
4.3.2 Πολυδιάστατη περίπτωση: Δεδομένα iris

Το δεύτερο σύνολο δεδομένων είναι ίσως το πιο γνωστό και ευρέως χρησιμοποιούμενο σύνολο δεδομένων της Πολυμεταβλητής Ανάλυσης. Πρόκειται για το λεγόμενο σύνολο δεδομένων Fisher iris data, τα οποία πρωτοεμφανίστηκαν στη βιβλιογραφία από τον R. A. Fisher (1936), στην προσπάθειά του να εισάγει τη χρήση της γραμμικής διαχωριστικής συνάρτησης που πρότεινε. Πρόκειται για 50 μετρήσεις σε φυτά από κάθενα εκ των τριών τύπων αγριόκρινων (iris): Setosa, Versicolor και Virginical, για τις ακόλουθες τέσσερις μεταβλητές: μήκος σέπαλου (sepal length), πλάτος σέπαλου (sepal width), μήκος πετάλου (petal length) και πλάτος πετάλου (petal width), με μονάδα μέτρησης τα εκατοστά. Στην παρούσα εφαρμογή θέλοντας να αξιολογήσουμε το προτεινόμενο τοπικό κριτήριο επιλογής μοντέλου, ήτοι το κριτήριο $LDiv.IC$, το ενδιαφέρον επικεντρώνεται στις μεταβλητές πλάτος πετάλου και μήκος πετάλου.

Στο πλαίσιο αυτό, στο Σχήμα 4.5 απεικονίζονται οι $n = 150$ παρατηρήσεις (50 για κάθε ποικιλία αγριόκρινων) για τις υπό μελέτη μεταβλητές πλάτος πετάλου (x-άξονας) και μήκος πετάλου (y-άξονας). Σε όσα ακολουθούν υποθέτουμε ότι το πραγματικό μοντέλο περιγράφεται από μία μείξη τριών κανονικών διδιάστατων κανονικών συνιστωσών κατανομών, με τις παραμέτρους της κάθε συνιστώσας να προκύπτουν με εκτίμηση χρησιμοποιώντας την προσέγγιση προσαύξησης δεδομένων (data augmentation) (βλέπε, Dempster et al., 1977). Έτσι θεωρούμε ότι το πραγματικό μοντέλο περιγράφεται από την κατανομή

$$f_\theta(x) = \frac{1}{3}N((0.246, 1.462), \Sigma_1) + \frac{1}{3}N((1.326, 4.260), \Sigma_2) + \frac{1}{3}N((2.026, 5.552), \Sigma_3),$$

με $\Sigma_1 = \begin{pmatrix} 0.0109 & 0.0059 \\ 0.0059 & 0.0296 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 0.0383 & 0.0716 \\ 0.0716 & 0.2164 \end{pmatrix}$ και $\Sigma_3 = \begin{pmatrix} 0.0739 & 0.0478 \\ 0.0478 & 0.2985 \end{pmatrix}$.



Σχήμα 4.5: Γράφημα των μεταβλητών πλάτος πέταλου (x-άξονας), μήκος πέταλου (y-άξονας) και οι ισοϋψείς από τρεις πυρήνες. Τα δεδομένα περιέχουν 50 παρατηρήσεις από κάθε είδος iris, setosa (μπλε), virginica (κίτρινο) και versicolor (πράσινο).

Στη συνέχεια εφαρμόζεται το τοπικό κριτήριο επιλογής μοντέλου, έχοντας ως τρία υποψήφια μοντέλα τις διδιάστατες κανονικές συνιστώσες της κατανομής $f_\theta(x)$, δηλαδή τις κατανομές με σ.π.π.

$$f(x;\mu_i,\Sigma_i) = (2\pi)^{-1}|\Sigma_i|^{-1/2}\exp\left(-\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)\right),$$

όπου τα μέσα διανύσματα μ_i και οι πίνακες διακυμάνσεων-συνδιακυμάνσεων Σ_i, $i = 1, 2, 3$, προσδιορίστηκαν πρωτύτερα.

Στο πλαίσιο αυτό, υπολογίζονται οι τιμές του τοπικού κριτηρίου $LDiv.IC$ για διάφορους πυρήνες από την διδιάστατη κανονική κατανομή. Πιο συγκεκριμένα, επιλέγουμε μ_0 να είναι ένα από τα διανύσματα $(0.25, 1.45)$, $(1.35, 4.25)$ και $(2, 5.5)$, ενώ επιλέγεται ως πίνακας διακύμανσεων-συνδιακύμανσεων ο $\Sigma_0 = \begin{pmatrix} 0.005 & 0 \\ 0 & 0.005 \end{pmatrix}$. Τέλος, υπολογίζεται το καθολικό κριτήριο AIC καθώς και το καθολικό κριτήριο DIC για $\alpha = 0.1$.

Τα αποτελέσματα που αφορούν τις τιμές του τοπικού κριτηρίου καθώς και των καθολικών κριτηρίων παρατίθενται στον Πίνακα 4.5. Παρατηρούμε ότι στην περιοχή με κέντρο $(0.25, 1.45)$ επιλέγεται η πρώτη κανονική συνιστώσα, στη περιοχή με κέντρο το $(1.35, 4.25)$ επιλέγεται η δεύτερη κανονική συνιστώσα, ενώ η τρίτη κανονική συνιστώσα επιλέγεται στην περιοχή με κέντρο $(2, 5.5)$. Επιπλέον, παρατηρούμε ότι και τα δύο καθολικά κριτήρια επι-

Πίνακας 4.5: Τιμές του κριτήριου πληροφορίας $LDiv.IC$ για τρία υποψήφια μοντέλα για διάφορους πυρήνες. Το Μοντέλο 1 είναι μια διδιάστατη κανονική κατανομή που αντιστοιχεί στη πρώτη συνιστώσα του μοντέλου μείξης, το Μοντέλο 2 αντιστοιχεί στη δεύτερη και το Μοντέλο 3 στην τρίτη συνιστώσα. Ο πυρήνας είναι μια διδιάστατη κανονική κατανομή με διάφορες μέσες τιμές μ_0 και σταθερό Σ_0. Στη τελευταία στήλη φαίνεται το μοντέλο που επιλέγεται από το τοπικό κριτήριο $LDiv.IC$.

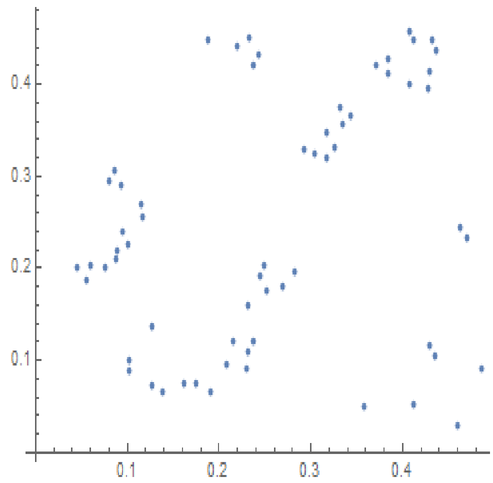
Δεδομένα iris	Μοντέλο 1	Μοντέλο 2	Μοντέλο 3	Επιλογή
Πυρήνας: Διδιάστατη κανονική ($\mu_0, \Sigma_0 = [(0.01, 0), (0, 0.005)]$)				
μ_0	$LDiv.IC$	$LDiv.IC$	$LDiv.IC$	Επιλεγόμενο Μοντέλο
$(0.25, 1.45)$	-1114.76	0.00001	0	1
$(1.35, 4.25)$	-0.00152	-805.079	-6.74515	2
$(2, 5.5)$	0	-3.80021	-272.447	3
AIC	48426.18	2692.16	4358.32	2
DIC	-709.715	-917.613	-792.693	2

λέγουν τη δεύτερη συνιστώσα. Η συμπεριφορά αυτή εξηγείται καθώς τα κλασικά κριτήρια τείνουν να επιλέγουν το μοντέλο που αντιστοιχεί στο κέντρο της περιοχής των δεδομένων. Αυτό συμβαίνει καθώς οι καθολικές διαδικασίες δε λαμβάνουν υπόψη τα τοπικά χαρακτηριστικά των δεδομένων.

4.3.3 Τοπική επιλογή μοντέλου σε point process: Δεδομένα redwoodfull

Το τρίτο σύνολο δεδομένων (αρχείο redwoodfull) περιγράφηκε και αναλύθηκε για πρώτη φορά από τον Strauss (1975). Πρόκειται για δεδομένα που αντιπροσωπεύουν τις θέσεις 95 φυντανιών και δενδρυλλίων σε μια τετράγωνη περιοχή δειγματοληψίας της Καλιφόρνιας. Στην παρούσα εφαρμογή θέλοντας να αξιολογήσουμε το προτεινόμενο τοπικό κριτήριο επιλογής μοντέλου, ήτοι το κριτήριο $LDiv.IC$, περιοριζόμαστε στις 63 παρατηρήσεις στο μειωμένο παράθυρο $\mathcal{W} = (0, 0.5) \times (0, 0.5)$, που φαίνεται στο Σχήμα 4.6.

Στο παραπάνω πλαίσιο, ως υποψήφια μοντέλα για τη συνάρτηση έντασης (intesity function) της Poisson point process, θεωρούμε έξι μοντέλα μείξης διδιάστατων κανονικών συνιστωσών, όπου το $i-$οστό μοντέλο αποτελείται από μια μείξη από i το πλήθος διδιάστα-



Σχήμα 4.6: Οι θέσεις των δένδρων redwood στην Καλιφόρνια. Υπάρχουν $n = 63$ θέσεις.

τες κανονικές συνιστώσες, για $i = 1, ..., 6$. Οι άγνωστες παράμετροι των έξι υποψήφιων μοντέλων προκύπτουν χρησιμοποιώντας την προσέγγιση της τεχνικής προσαύξησης δεδομένων (data augmentation). Οι εκτιμήσεις αυτές για κάθε μοντέλο παρατίθενται στο τέλος αυτής της εφαρμογής. Στη συνέχεια, υπολογίζεται το τοπικό κριτήριο $LDiv.IC$ για διάφορους διδιάστατους κανονικούς πυρήνες. Πιο συγκεκριμένα, επιλέγουμε παράμετρο θέσης $\mu_0 = (0.08, 0.25)$, $(0.20, 0.45)$, $(0.20, 0.15)$, $(0.35, 0.35)$, $(0.4, 0.4)$, ή $(0.45, 0.10)$ και σταθερό πίνακα διακυμάνσεων-συνδιακυμάνσεων τον $\Sigma_0 = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$.

Τέλος, υπολογίζεται η τιμή του καθολικού κριτηρίου AIC. Τα αποτελέσματα παρατίθενται στον Πίνακα 4.6, όπου οι τιμές του τοπικού κριτηρίου υπολογίζονται για κάθε πυρήνα. Παρατηρούμε ότι στις περιοχές εστίασης με βάση τους πυρήνες με μέσες τιμές $\mu_0 = (0.08, 0.25)$, $(0.20, 0.45)$, $(0.20, 0.15)$, $(0.35, 0.35)$, $(0.4, 0.4)$), και σταθερό Σ_0, επιλέγεται το Μοντέλο 4, ενώ στην περιοχή εστίασης με μέση τιμή $\mu_0 = (0.45, 0.10)$ επιλέγεται το Μοντέλο 2. Επιπλέον, το καθολικό κριτήριο AIC προτείνει το Μοντέλο 4 με τιμή κριτηρίου $AIC = -173.835$. Ως αποτέλεσμα, το μοντέλο μείξης τεσσάρων κανονικών συνιστωσών αναδύεται ως η πλέον κατάλληλη επιλογή τόσο σε όλο το πεδίου ορισμού όσο και τοπικά.

Πίνακας 4.6: Τιμές του τοπικού κριτηρίου $LDiv.IC$ για έξι υποψήφια μοντέλα για διάφορους πυρήνες. Ο πυρήνας είναι μια διδιάστατη κανονική κατανομή με μέσο διάνυσμα μ_0, όπως φαίνεται στην πρώτη στήλη, και σταθερό πίνακα διακύμανσης-συνδιακύμανσης Σ_0. Στη τελευταία στήλη φαίνεται το μοντέλο που επιλέγεται με βάση το κριτήριο $LDiv.IC$.

Redwood	Μοντέλο 1	Μοντέλο 2	Μοντέλο 3	Μοντέλο 4	Μοντέλο 5	Μοντέλο 6	Επιλογή
Πυρήνας: Διδιάστατη Κανονική $(\mu_0, \Sigma_0 = [(0.01, 0), (0, 0.01)])$							
μ_0	$LDiv.IC$	$LDiv.IC$	$LDiv.IC$	$LDiv.IC$	$LDiv.IC$	$LDiv.IC$	Επιλεγόμενο Μοντέλο
$(0.08, 0.25)$	-2941.13	-2958.20	-3008.57	-3050.93	-3039.29	-3010.76	4
$(0.2, 0.45)$	-1645.45	-1647.45	-1612.76	-1659.03	-1640.06	-1612.99	4
$(0.2, 0.15)$	-3689.42	-3690.03	-3686.59	-3738.56	-3703.70	-3670.25	4
$(0.35, 0.35)$	-3170.21	-3200.66	-3111.32	-3260.05	-3237.78	-3209.04	4
$(0.40, 0.40)$	-2958.16	-2994.83	-2915.98	-3078.63	-3061.47	-3032.35	4
$(0.45, 0.10)$	-1177.27	-1197.46	-1142.21	-1183.27	-1188.05	-1122.75	2
AIC	-147.732	-158.662	-137.785	-173.835	-162.56	-149.903	4

Παράρτημα 4.1 (Εφαρμογή: Δεδομένα redwoodfull)**.** *Οι εκτιμώμενες τιμές των παραμέτρων των έξι υποψήφιων μοντέλων, ήτοι οι εκτιμώμενες τιμές των παραμέτρων θέσης και κλίμακας των κανονικών συνιστωσών κατανομών καθώς και αυτές των αντίστοιχων βαρών, συντελεστών μείξης, είναι:*

Μοντέλο 1: *διδιάστατη κανονική με εκτιμώμενες παραμέτρους* $\hat{\mu}_1 = (0.26, 0.25)$ *και* $\hat{\Sigma}_1 = \begin{pmatrix} 0.0170 & 0.0049 \\ 0.0049 & 0.0179 \end{pmatrix}$.

Μοντέλο 2: *μείξη δυο διδιάστατων κανονικών συνιστωσών με εκτιμώμενες παραμέτρους* $\hat{w}_2 = 0.598$, $\hat{\mu}_2 = (0.21, 0.17)$, $\hat{\Sigma}_2 = \begin{pmatrix} 0.0156 & -0.0032 \\ -0.0032 & 0.0068 \end{pmatrix}$, $\hat{w}_3 = 0.402$, $\hat{\mu}_3 = (0.32, 0.36)$ *και* $\hat{\Sigma}_3 = \begin{pmatrix} 0.0087 & 0.0007 \\ 0.0007 & 0.0056 \end{pmatrix}$.

Μοντέλο 3: *μείξη τριών διδιάστατων κανονικών συνιστωσών με εκτιμώμενες παραμέτρους* $\hat{w}_4 = 0.2764$, $\hat{\mu}_4 = (0.12, 0.27)$, $\hat{\Sigma}_5 = \begin{pmatrix} 0.0036 & 0.0040 \\ 0.0040 & 0.0084 \end{pmatrix}$, $\hat{w}_6 = 0.2506$, $\hat{\mu}_6 = (0.34, 0.36)$, $\hat{\Sigma}_6 = \begin{pmatrix} 0.0083 & -0.0016 \\ -0.0016 & 0.0123 \end{pmatrix}$, $\hat{w}_7 = 0.4730$, $\hat{\mu}_7 = (0.28, 0.22)$ *και* $\hat{\Sigma}_7 = \begin{pmatrix} 0.0105 & 0.0079 \\ 0.0079 & 0.0142 \end{pmatrix}$.

Μοντέλο 4: *μείξη τεσσάρων διδιάστατων κανονικών συνιστωσών με εκτιμώμενες παραμέτρους* $\hat{w}_8 = 0.1565$, $\hat{\mu}_8 = (0.38, 0.18)$, $\hat{\Sigma}_8 = \begin{pmatrix} 0.0055 & 0.0000 \\ 0.0000 & 0.0105 \end{pmatrix}$, $\hat{w}_9 = 0.3004$, $\hat{\mu}_9 = (0.34, 0.35)$, $\hat{\Sigma}_9 = \begin{pmatrix} 0.0050 & 0.0047 \\ 0.0047 & 0.0071 \end{pmatrix}$, $\hat{w}_{10} = 0.2715$, $\hat{\mu}_{10} = (0.12, 0.29)$, $\hat{\Sigma}_{10} = \begin{pmatrix} 0.0040 & 0.0049 \\ 0.0049 & 0.0083 \end{pmatrix}$, $\hat{w}_{11} = 0.2715$, $\hat{\mu}_{11} = (0.22, 0.15)$ *και* $\hat{\Sigma}_{11} = \begin{pmatrix} 0.0053 & 0.0029 \\ 0.0029 & 0.0049 \end{pmatrix}$.

Μοντέλο 5: *μείξη πέντε διδιάστατων κανονικών συνιστωσών με εκτιμώμενες παραμέτρους* $\hat{w}_{12} = 0.2588$, $\hat{\mu}_{12} = (0.12, 0.28)$, $\hat{\Sigma}_{12} = \begin{pmatrix} 0.0037 & 0.0046 \\ 0.0046 & 0.0077 \end{pmatrix}$, $\hat{w}_{13} = 0.1435$,

$\hat{\mu}_{13} = (0.40, 0.17)$, $\hat{\Sigma}_{13} = \begin{pmatrix} 0.0045 & 0.0010 \\ 0.0010 & 0.0094 \end{pmatrix}$, $\hat{w}_{14} = 0.2776$, $\hat{\mu}_{14} = (0.33, 0.34)$, $\hat{\Sigma}_{14} = \begin{pmatrix} 0.0050 & 0.0052 \\ 0.0052 & 0.0073 \end{pmatrix}$, $\hat{w}_{15} = 0.0916$, $\hat{\mu}_{15} = (0.24, 0.25)$, $\hat{\Sigma}_{15} = \begin{pmatrix} 0.0057 & 0.0057 \\ 0.0057 & 0.0095 \end{pmatrix}$, $\hat{w}_{16} = 0.2284$, $\hat{\mu}_{16} = (0.23, 0.15)$ *και* $\hat{\Sigma}_{16} = \begin{pmatrix} 0.0049 & 0.0034 \\ 0.0034 & 0.0055 \end{pmatrix}$.

Μοντέλο 6: *μείξη έξι διδιάστατων κανονικών συνιστωσών με εκτιμώμενες παραμέτρους* $\hat{w}_{17} = 0.2478$, $\hat{\mu}_{17} = (0.12, 0.28)$, $\hat{\Sigma}_{17} = \begin{pmatrix} 0.0036 & 0.0044 \\ 0.0044 & 0.0074 \end{pmatrix}$, $\hat{w}_{18} = 0.0883$, $\hat{\mu}_{18} = (0.25, 0.29)$, $\hat{\Sigma}_{18} = \begin{pmatrix} 0.0061 & 0.0067 \\ 0.0067 & 0.0108 \end{pmatrix}$, $\hat{w}_{19} = 0.2527$, $\hat{\mu}_{19} = (0.33, 0.34)$, $\hat{\Sigma}_{19} = \begin{pmatrix} 0.0054 & 0.0059 \\ 0.0059 & 0.0083 \end{pmatrix}$, $\hat{w}_{20} = 0.1372$, $\hat{\mu}_{20} = (0.41, 0.16)$, $\hat{\Sigma}_{20} = \begin{pmatrix} 0.0039 & 0.0016 \\ 0.0016 & 0.0092 \end{pmatrix}$, $\hat{w}_{21} = 0.1943$, $\hat{\mu}_{21} = (0.23, 0.16)$, $\hat{\Sigma}_{21} = \begin{pmatrix} 0.0046 & 0.0042 \\ 0.0042 & 0.0067 \end{pmatrix}$, $\hat{w}_{22} = 0.0797$, $\hat{\mu}_{22} = (0.24, 0.20)$ *και* $\hat{\Sigma}_{22} = \begin{pmatrix} 0.0055 & 0.0059 \\ 0.0059 & 0.0090 \end{pmatrix}$.

Κεφάλαιο 5

Προτάσεις για περαιτέρω έρευνα

Στο κεφάλαιο παρουσιάζονται ερευνητικά ερωτήματα για περαιτέρω έρευνα σε θέματα σχετικά με το αντικείμενο της διατριβής. Τα ερωτήματα αυτά γεννήθηκαν τόσο κατά το αρχικό στάδιο της βιβλιογραφικής ενημέρωσης όσο και κατά την επεξεργασία των Κεφαλαίων 2-4.

Το Κεφάλαιο 2 επικεντρώνεται στον ορισμό μιας ευρείας κλάσης τοπικών αποκλίσεων μεταξύ δύο μέτρων πιθανότητας ή μεταξύ των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας. Οι εισαχθείσες τοπικές αποκλίσεις βασίζονται στην κλασική απόκλιση του Csiszár και αποδίδουν ένα μέτρο της στατιστικής απόστασης μεταξύ δύο κατανομών σε μία συγκεκριμένη περιοχή του κοινού πεδίου ορισμού τους. Κατ' αυτόν τον τρόπο παρέχουν ένα χρήσιμο εργαλείο για την ποσοτικοποίηση της στατιστικής απόστασης μεταξύ δύο κατανομών, τοπικά, σε μια συγκεκριμένη δηλαδή περιοχή του κοινού πεδίου ορισμού τους. Επιπλέον, όπως έχει επισημανθεί στο Κεφάλαιο 2, μελετάται το σύνολο τιμών των τοπικών αποκλίσεων που παρουσιάζονται και προσδιορίζονται, περαιτέρω, οι αναλυτικές εκφράσεις των προτεινόμενων τοπικών αποκλίσεων όταν οι κατανομές στις οποίες στηρίζονται είναι μέλη της εκθετικής οικογένειας κατανομών και όταν ταυτίζονται με εκείνες της πολυδιάστατης κανονικής κατανομής

Υπάρχει ένας ευρύς κατάλογος ιδιοτήτων που ικανοποιούνται από την κλασική, καθολική ϕ-απόκλιση Csiszár. Μελέτες σχετικές με τις ιδιότητες που ικανοποιούνται από την καθολική ϕ-απόκλιση Csiszár και τα παράγωγά της μπορούν να βρεθούν, μεταξύ άλλων, στα άρθρα των Papaioannou (1985, 2001), Liese and Vajda (2006) και στα συγγράμματα των Liese and Vajda (1987) και Vajda (1989). Ανοικτό παραμένει το πρόβλημα της μελέτης αυτών των ιδιοτήτων για την τοπική ϕ-απόκλιση. Ειδικότερα, οι ιδιότητες της καθολικής ϕ-απόκλισης που περιγράφονται στα Θεωρήματα 9.4 (Theorem of uniqueness), 9.6 (Theorem of symmetry), 9.9 (Theorem of monotonicity), 9.12 (Theorem of finite approximation), 9.19 (Theorem of isomorphy) και 9.23 (Theorem of monotone convergence) του βιβλίου του Vajda (1989), παρουσιάζει ενδιαφέρον να μελετηθούν και να επεκταθούν για την τοπική ϕ-απόκλιση που ορίστηκε στο Κεφάλαιο 2 της διατριβής.

'Οπως, αναφέρθηκε, είναι σημαντική για τις εφαρμογές η ύπαρξη αναλυτικών εκφράσεων των μέτρων απόκλισης για ευρέως χρησιμοποιούμενες οικογένειες κατανομών. Ενδεικτικά αναφέρουμε τις εργασίες των Soofi and Retzer (2002), Zografos and Nadarajah (2005), Zografos (2008) και Pardo (2006), καθώς και στις αναφορές αυτών. Ένα θέμα που προκύπτει για περαιτέρω διερεύνηση είναι η εύρεση κλειστών τύπων της τοπικής ϕ-απόκλισης για κατανομές που ανήκουν, για παράδειγμα, στη σφαιρική ή και ευρύτερα στην ελλειπτική οικογένεια κατανομών. Ιδιαίτερο ενδιαφέρον παρουσιάζει η εύρεση κλειστού τύπου στην περίπτωση που η τοπική ϕ-απόκλιση στηρίζεται σε μείξη συνιστωσών κατανομών. Στο πλαίσιο αυτό, παρουσιάζει ενδιαφέρον μια επέκταση του ορισμού της τοπικής ϕ-απόκλισης ώστε να συμπεριλάβει μείξης κατανομών με συνιστώσες από την εκθετική οικογένεια κατανομών. Στην περίπτωση αυτή προηγείται ο ορισμός της καθολικής ϕ-απόκλισης σε μείξεις κατανομών μιας και δεν είναι διαθέσιμος ακόμη στη βιβλιογραφία. Στο πλαίσιο αυτό, ο ορισμός της καθολικής ϕ-απόκλισης και της τοπικής ϕ-απόκλισης μπορεί να πραγματοποιηθεί ακολουθώντας, για παράδειγμα, μια από τις τρεις παρακάτω προσεγγίσεις: (α) ακριβής υπολογισμός (β) με βάση τον αλγόριθμο Monte Carlo, και (γ) σύμφωνα με τη διαδικασία προσαύξησης δεδομένων (data augmentation).

Στο Κεφάλαιο 3 αναπτύχθηκαν τοπικοί έλεγχοι με βάση την κλάση μέτρων τοπικής ϕ-απόκλισης που ορίστηκαν στο Κεφάλαιο 2. Μελετήθηκε η ασυμπτωτική κατανομή των στατιστικών συναρτήσεων που δημιουργήθηκαν υπό τη μηδενική υπόθεση, η οποία μπορεί να είναι απλή ή σύνθετη καθώς και η ασυμπτωτική κατανομή της εκτιμώμενης τοπικής απόκλισης υπό την εναλλακτική υπόθεση. Ειδικότερα, κατασκευάστηκε ο έλεγχος καλής προσαρμογής, τοπικά και ο έλεγχος ομοιογένειας, τοπικά. Ορίστηκε περαιτέρω η τοπική εκδοχή του πίνακα πληροφορίας Fisher, και οι τοπικές στατιστικές συναρτήσεις Wald και Rao. Τα κρίσιμα σημεία των ελέγχων προκύπτουν προσεγγιστικά με τη βοήθεια του αλγορίθμου Monte Carlo. Υπάρχουν πολλοί παράγοντες που μπορούν πιθανώς να επηρεάσουν την αποτελεσματικότητα των μεθοδολογιών που αναπτύχθηκαν. Για παράδειγμα, η επιλογή της κυρτής συνάρτησης ϕ ή η επιλογή του πυρήνα. Σχετικά με την επιλογή της κυρτής συνάρτησης ϕ και ειδικότερα την επιλογή της παραμέτρου λ στην ειδική περίπτωση της λ-απόκλισης των Cressie and Read (βλέπε Πίνακα 1.1) υπολογιστικές μελέτες δείχνουν ότι η επιλογή $\lambda = 2/3$ είναι ιδανική σε κάποιες περιπτώσεις (βλέπε, Read and Cressie, 1988). Στο ίδιο πλαίσιο μια μελέτη σχετική με την επιλογή του μέτρου πυρήνα, του Ορισμού 2.1, παρουσιάζει, ίσως, μεγάλο ενδιαφέρον.

Στο Κεφάλαιο 4 εισάγεται η έννοια της τοπικής επιλογής μοντέλου, και το τοπικό κριτήριο $LDiv.IC$. Μια επέκταση των αποτελεσμάτων του Κεφάλαιο 4 μπορεί να προκύψει υιοθετώντας για την κατασκευή τοπικού κριτηρίου πληροφορίας επιλογής μοντέλου, ίσως, ένα άλλο τοπικό μέτρο ομοιότητας πιο άμεσα συνδεμένο με την τοπική ϕ-απόκλιση του Ορισμού 2.1, αντί της τοπικής απόκλισης BHHJ. Σε μια τέτοια περίπτωση θα ήταν ενδιαφέρουσα μια μελέτη σύγκρισης των διαφορετικών τοπικών κριτήριων.

Η χρήση μέτρων απόκλισης κατανομών για την εκτίμηση παραμέτρων μιας κατανομής αποτελεί ξεχωριστό κομμάτι της στατιστικής βιβλιογραφίας. Εξέχουσα θέση στο πλαίσιο αυτό έχουν οι εκτιμητές ελάχιστης απόστασης Hellinger (minimum Hellinger distance estimators) οι οποίοι χαρακτηρίζονται από ευρωστία (robustness). Η πρόσφατη μονογραφία των Basu et al. (2011) επικεντρώνεται στην αξιοποίηση των αποκλίσεων BHHJ για την ανάπτυξη μεθοδολογίας και την κατασκευή εκτιμητών τύπου ελάχιστης απόκλισης. Οι εκτιμητές αυτοί δημιουργούνται από την ελαχιστοποίηση, ως προς την άγνωστη παράμετρο, της απόκλισης BHHJ μεταξύ του παραμετρικού μοντέλου και του εμπειρικού ανάλογου. Η μεθοδολογία αυτή μπορεί να επεκταθεί για την κατασκευή εκτιμητών με τη χρήση ενός μέτρου τοπικής απόκλισης. Στη περίπτωση αυτή, μια επέκταση της εξίσωσης (4.36) είναι η ακόλουθη

$$\frac{1}{n}\sum_{i=1}^{n} h_\omega(x)u_\theta(X_i)f_\theta^\alpha(X_i) - \int\limits_{\mathcal{X}} h_\omega(x)u_\theta(x)f_\theta^{1+a}(x)dx = 0,$$

με την εισαγωγή του πυρήνα πυκνότητα h_ω, με $\theta \in \Theta \subseteq R^M$, $u_\theta(x) = \nabla_\theta \log(f_\theta(x))$ και $X_1, ..., X_n$ τυχαίο δείγμα από την κατανομή $f_\theta(x)$. Η αριθμητική επίλυση της παραπάνω εξίσωσης θα οδηγήσει στους εκτιμητές ελάχιστης τοπικής απόκλισης, η μελέτη των οποίων παρουσιάζει ενδιαφέρον.

Η μελέτη της οριακής συμπεριφοράς των τοπικών μέτρων απόκλισης σε σχέση με την οριακή συμπεριφορά των ακολουθιών των κατανομών στις οποίες στηρίζονται είναι αντικείμενο περεταίρω μελέτης στη βάση του Κεφαλαίου 1 της Διδακτορικής Διατριβής του Zografos (1987) και της εργασίας Zografos et al. (1989). Για παράδειγμα, αν το πρώτο όρισμα του τοπικού μέτρου απόκλισης είναι μια ακολουθία κατανομών που συγκλίνει στο δεύτερο όρισμα, τότε η ακολουθία των τιμών του μέτρου συγκλίνει στην ελάχιστη τιμή που μπορεί να πάρει το μέτρο; Επιπρόσθετα, και η διερεύνηση του αντίστροφου ερωτήματος παρουσιάζει ενδιαφέρον. Για παράδειγμα, η οριακή συμπεριφορά ενός τοπικού μέτρου απόκλισης που στηρίζεται σε μια ακολουθία κατανομών έχει σαν αποτέλεσμα κάποιο τύπο σύγκλισης για την ακολουθία των κατανομών στις οποίες στηρίζεται; Μια θετική απάντηση στο ερώτημα αυτό, ανάγει τα μέτρα τοπικής ϕ-απόκλισης σε εργαλεία διερεύνησης σύγκλισης ακολουθιών κατανομών πιθανότητας σε τοπικό περιβάλλον. Τα ερευνητικά αυτά ερωτήματα μελετήθηκαν στην καθολική περίπτωση στα Θεωρήματα 2.3 και 2.4 της εργασίας των Zografos et al. (1989).

Τέλος, η μελέτη κλασικών θεμάτων στατιστικής συμπερασματολογίας, όπως εκτιμητές ΑΟΕΔ, αποδεκτοί (admissible), minimax και Bayes, κάτω φράγμα Cramér-Rao κ.ά., μπορεί να είναι θέματα μελλοντικής έρευνας στο πλαίσιο του ορισμού και της ερμηνείας των παραπάνω εννοιών σε τοπικό περιβάλλον.

Κεφάλαιο 6

Abstract of the Dissertation

This dissertation utilizes the concept of ϕ-divergence in order to define and study local statistical methods. In this section, we briefly discuss the contents of this dissertation and highlight the original contributions we achieved.

The ϕ-divergence, as defined by (1.2), quantifies the difference between probability measures P and Q or the corresponding densities p and q over a domain $\mathcal{X}$. However, in practice there are cases in which our interest is focused on the discrepancies between two populations or between their corresponding distributions in a subset of the joint domain $\mathcal{X}$. For example, a researcher is focused in the investigation of whether the populations of men and women exhibit the same behavior for high or low levels of blood cholesterol. Therefore, even though the implementation of relation (1.2) gives a picture regarding the homogeneity of men and women for all levels of cholesterol, (1.2) does not give any information regarding the homogeneity of men and women for low, high or normal levels of blood cholesterol. It is clear that a divergence measure will have to be defined in such a way that it can give us an indication of the similarity or homogeneity of the two populations or their corresponding distributions in a subset of the joint domain $\mathcal{X}$.

Prompted by this research question, **Chapter 2** focuses on the definition of a broad class of local divergence measures between two probability measures or between the corresponding densities. The local divergences developed are based on the classical

Csiszár divergence and provide a measure of the pseudo-distance (or statistical distance or divergence) between two distributions in a specific area of their joint domain. In this way they provide a useful tool for the quantification of the statistical distance between two distributions, locally, in a specific area of their joint domain.

In addition, in Chapter 2 the range of values of the local divergences are studied and the analytical expressions of the proposed local divergences are further determined when the distributions are members of the exponential family and the case of the multivariate normal distribution is also considered. Chapter 2 exemplifies the methodology via a simulation study which illustrates the robust behavior of the proposed local divergence measures in identifying the discrepancies between two populations which cannot be captured using classical global measures. Part of the results in this chapter is included in Avlogiaris et al. (2016 a).

Statistical Information Theory, among other things, quantifies the divergence between two or more statistical models. Therefore, over the last years it has served as a basic tool for the development of methodologies in Mathematical Statistics. The monographs of Read and Cressie (1988) and Pardo (2006) and the references therein, provide an exhaustive discussion for the use of the divergence measures in testing statistical hypotheses, with more recent results cited in the work of Basu et al. (2013). Homogeneity tests between two or more distributions have previously been studied in Zografos (1998).

In a fully parametric framework, consider two independent random samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ from the populations described by models f_{θ_1} and f_{θ_2}, $\theta_1, \theta_2 \in \Theta$, respectively. A divergence measure $D(f_{\theta_1}, f_{\theta_2})$ quantifies the similarity between the models f_{θ_1} and f_{θ_2} if the divergence $D(f_{\theta_1}, f_{\theta_2})$ satisfies the key property $D(f_{\theta_1}, f_{\theta_2}) \geq 0$, with equality if and only if the models f_{θ_1} and f_{θ_2} coincide. As a result, an empirical version of a divergence measure $D(f_{\theta_1}, f_{\theta_2})$, say $D(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$, where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are the MLE (Maximum Likelihood Estimators) of θ_1 and θ_2 based on the two random samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$, respectively, can be used as the test statistic in order to assess if models f_{θ_1} and f_{θ_2} coincide (homogeneity test, see Zografos, 1998). Low values of $D(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ are in favor of the null hypothesis $H_0 : \theta_1 = \theta_2$ against the alternative $H_a : \theta_1 \neq \theta_2$, while large values of $D(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ suggest rejection of H_0. In a similar manner, an empirical version of a divergence measure $D(f_\theta, f_{\theta_0})$, say $D(f_{\widehat{\theta}}, f_{\theta_0})$, serves as a test statistic for the hypothesis that the true, but unknown model f_θ can be described by the model f_{θ_0}, which is determined by the null hypothesis $H_0 : \theta = \theta_0$ (goodness of fit test). The test of the null hypothesis $H_0 : \theta = \theta_0,\ \theta_0 \in \Theta,\ \theta_0$ known, against the alternative $H_a : \theta \neq \theta_0$ has extensively been studied by means of ϕ-divergence measures (see, Pardo, 2006, Chapter 9).

The available parametric statistical tests in the bibliography which are created via ϕ-divergence measures, investigate the problems in the whole domain of the distributions, which describe a characteristic or property of the members of the population. Despite the fact that the claim of goodness of fit or homogeneity can be accepted or rejected in the whole domain of the distributions under consideration, analysis of real data indicates the opposite result when our interest is focused on a subset of the domain of the distributions. This finding has sparked the introduction, the construction and the study of parametric statistical tests of goodness of fit and homogeneity, locally, in a subset of the domain of the distributions under consideration. The construction of the above mentioned statistical tests for the test of hypotheses, locally, is based on the concept of the local ϕ-divergence which was defined and studied in Chapter 2.

As a result, **Chapter 3** focuses on the construction and study of parametric statistical tests for goodness of fit (one sample problem) and tests of homogeneity (two sample problem), locally, in a subset of the domain of the distributions under consideration. More precisely, test statistics are defined and their asymptotic distributions are obtained under the null hypothesis. In addition, in Chapter 3 the effectiveness of the developed local parametric tests are investigated by means of simulation studies based on their type I error and power. The behavior of the proposed statistical tests has also been investigated via real data. The results of this chapter have appeared in Avlogiaris et al. (2016 b).

In the sequel, consider a random sample from a population and two or more parametric models which are candidate models for the data under consideration. Model selection criteria provide a systematic and rigorous method that allows statisticians to choose the most appropriate model from a collection of possible models used to describe the data. The construction of such criteria requires the creation of a measure of similarity between two entertained models, which are typically described in terms of their distributions. This can be achieved if an unbiased estimator of the expected overall discrepancy is found, which measures the statistical distance between the true, but unknown model, and the entertained model. Therefore, the smaller the value of the criterion is the more preferable the model is.

Several cases of these measures have been utilized in the creation of model selection criteria. In particular, the well-known Kullback-Leibler (1951) measure of divergence was used by Akaike (1973) in order to develop the Akaike information criterion (AIC). Since Akaike's pioneering work, there has been a vast literature on the construction of model selection criteria. We refer to Schwarz (1978), Konishi and Kitagawa (1996), Spiegelhalter et al. (2002), Seghouane and Bekara (2004), Cavanaugh (2004), Bengtsson and Cavanaugh (2006), Shang and Cavanaugh (2008), Shang (2008), Mattheou et al. (2009), Toma and Broniatowski (2011) and Toma (2014) and the references therein for the development and illustration of many classic model selection criteria. An alternative approach to these classic methods was presented in Claeskens and Hjort (2003), where the authors allowed different methods to be selected for different parameters of interest. Finally, the book by Claeskens and Hjort (2008) and the references therein provides an exhaustive discussion of model selection criteria.

Based on the above discussion, it is appealing to develop a method that selects the best model, among various available candidate models, in some areas of $\mathcal{X}$. In **Chapter 4,** a local model selection criterion (LDiv.IC) is developed and studied using the local BHHJ power divergence, as defined in Definition 1.4. We focus on the BHHJ measures of divergence because their functional expression is particularly useful in the construction of the criterion. Moreover in Chapter 4, simulations are presented in order to evaluate the performance of the proposed model selection criterion in a local setting and three applications of the proposed methodology are given by analyzing real datasets.

In **Chapter 5,** we provide suggestions for further research and possible extensions of this work are indicated, along with and some open problems left to consider in this context.

Keywords

ϕ-divergence, Csiszár divergence, Kullback-Leibler divergence, Cressie and Read power divergence, BHHJ power divergence, Local divergence, Exponential family, Local Fisher information matrix, Local Rao and Wald tests, Local goodness of fit test, Local homogeneity test, Local ϕ-divergence test, Model selection, Local Model selection, AIC, LDiv.IC, LAIC, Local divergence information criterion, Local model selection criterion, Local expected overall discrepancy, Local BHHJ power divergence, Mixture models, Point process theory.

Βιβλιογραφία

[1] Aitkin, M. (2001). Likelihood and Bayesian analysis of mixtures. *Statistical Modeling,* **1**, 287-304.

[2] Akaike, H., (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceeding of the second International Symposium on Information Theory. Akademiai Kaido, Budapest.

[3] Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc.* Ser. B, **28**, 131-142.

[4] Arndt, C. (2001). *Information Measures.* Springer-Verlang, Berlin.

[5] Avlogiaris, G., Micheas A. and Zografos, K. (2016a). On local divergences between two probability measures. *Metrika,* **79**, 303-333.

[6] Avlogiaris, G., Micheas A. and Zografos, K. (2016b). On testing local hypotheses via local divergence. *Statistical Methodology,* **31**, 20-42.

[7] Basu, A., Harris, I.R., Hjort, N.L. and Jones, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika,* **85**, 549-559.

[8] Basu, A., Mandal, A., Martin, N. and Pardo, L. (2013). Testing statistical hypotheses based on the density power divergence. *Annals of the Institute of Statistical Mathematics*, **65**, 319-348.

[9] Basu, A., Shioya, H. and Park, C. (2011). *Statistical inference. The minimum distance approach.* Chapman & Hall/CRC.

[10] Bengtsson, T. and Cavanaugh, J.E. (2006). An improved Akaike information criterion for state-space model selection. *Comput. Statist. Data Anal.* **50**, 2635–2654.

[11] Bhattacharyya, A. (1943). On a measure of divergence between two statistical population defined by their probability distributions. *Bulletin Calcutta Mathematical Society,* **35**, 99-109.

[12] Burham, P. K. and Anderson, R. D. (2003). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*, Second Edition, Springer.

[13] Cavanaugh, J.E., (2004). Criteria for linear model selection based on Kullback's symmetric divergence. *Austral. New Zealand J. Statist.* **46**, 257–274.

[14] Claeskens, G. and Hjort, N. L. (2003). The Focused Information Criterion. *Journal of the American Statistical Association.* **98**:464, 900-916.

[15] Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging.* Cambridge University Press.

[16] Cressie, N., (1993). *Statistics for Spatial Data.* 2nd Edition. Wiley.

[17] Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc.* Ser. B **46**, 440-464.

[18] Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, **8**, 85-108.

[19] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, **2**, 299-318.

[20] Csiszár, I. and Korner, J. (1981). *Information theory. Coding theorems for discrete memoryless systems.* Akademiai Kiado, Budapest.

[21] Dempster, P. A., Laird, M. N. and Rubin B. D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society.* Series B, **39**, No. 1, 1-38.

[22] Dic, J. J. and Gunst, M. C. M. (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica*, **39**, 14-26.

[23] Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatial-Temporal Point Patterns.* 3rd edition. CRC press.

[24] Ebrahimi, N., Soofi, S. and Soyer, R. (2010). Information measures in perspective. *International Statistical Review*, **78**, 383-412.

[25] Ferentinos, K. and Papaioannou, T. (1981). New parametric measures of information. *Information and control*, **51**, 193-208.

[26] Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Hafner press, New York.

[27] Fisher, R. A. (1936). The use of multiple measurements Axonomic Problems. *Annals of Eugenics*, **7**, 179-188.

[28] Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression.* Wiley.

[29] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problem. *Proceedings of the Royal Society*, Series A, **186**, 453-561.

[30] Jiménez-Gamero, M.D., Pino-Mejias, R., Alba-Fernández, V. and Moreno Rebollo, JL. (2011). Minimum φ-divergence estimation in misspecified multinomial models. *Computational Statistics & Data Analysis,* **55**, 3365-3378.

[31] Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis.* Third Edition. Prentice Hall International Editions.

[32] Kagan, A. M. (1963). On the theory of Fisher's information quantity. *Dokl. Akad. Nauk SSSR*, **151**, 277-278.

[33] Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons.

[34] Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika,* **83**, 875–890.

[35] Κουμουλλής, Γ. και Νεγρεπόντης, Σ. (2005). *Θεωρία Μέτρου.* Εκδόσεις Συμμετρία, Αθήνα.

[36] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statistics*, **22**, 79-86.

[37] Kullback, S. (1959). *Information theory and statistics.* Wiley.

[38] Landaburu, E. and Pardo, L. (2000). Goodness of fit tests with weights in the classes based on $(h, \phi)-$divergences. *Kybernetika*, **36**, 589-602.

[39] Landaburu, E. and Pardo, L. (2003). Minimum (h, ϕ) -divergences estimators with weights. *Appl. Math. Comput.*, **140**, 15-28.

[40] Landaburu, E., Morales, D. and Pardo, L. (2005). Divergence-based estimation and testing with misclassified data. *Statist. Papers*, **46**, 397-409.

[41] Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics.* **26**:3, 403-413.

[42] Liese, F. and Vajda, I. (1987). *Convex statistical distances.* Teubner Texts in Mathematics, Leipzig.

[43] Liese, F. and Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory*, **52**, 4394-4412.

[44] Mattheou, K., Lee, S. and Karagrigoriou, A. (2009). A model selection criterion based on the BHHJ measure of divergence. *Journal of Statistical Planning and Inference,* **139,** 228-235.

[45] Mattheou, K. and Karagrigoriou, A. (2010). A new family of divergence measures for tests of fit. *Australian and New Zealand Journal of Statistics,* **52**(2), 187-200.

[46] Matusita, K. (1964). Distance and decision rules. *Annals of the Institute of Statistical Mathematics*, **16**, 305-320.

[47] Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications. *Annals of the Institute of Statistical Mathematics*, **19**, 181-192.

[48] McElroy, T. and Holan, S. (2009). A local spectral approach for assessing time series model misspecification. *Journal of Multivariate Analysis*, **100**, 604-621.

[49] Menéndez, M. L., Morales, D., Pardo, L. and Salicrú, M. (1995). Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: A unified study. *Statistical Papers*, **36**, 1-29.

[50] Micheas, A. (2014). Hierarchical Bayesian modeling of marked non-homogeneous Poisson processes with finite mixtures and inclusion of covariate information. *Journal of Applied Statistics*. **41**:12, 2596-2615.

[51] Morales, D., Pardo, L. and Vajda, I. (1997). Some new statistics for testing hypothesses in parametric models. *Journal of Multivariate Analysis*, **62**, 137-168.

[52] Nielsen, F. and Nock, R. (2011). On Rényi and Tsallis entropies and divergences for exponential families. arXiv:1105.3259v1 [cs.IT] 17 May 2011.

[53] Papaioannou T. (1985). Measures of information. In: *Encyclopedia of Statistical Sciences*, Vol. 5, Kotz S, Johnson NL (eds). Wiley, New York, 391-397.

[54] Papaioannou T. (2001). On distances and measures of information: case of diversity. In: *Probability and Statistical Models with Applications*, Charalambides CA, Koutras MV, Balakrishnan N (eds). Chapman & Hall/CRC, 503-515.

[55] Pardo, L. (2006). *Statistical inference based on divergence measures*. Chapman & Hall/CRC.

[56] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine*, **50**, 157-172.

[57] Postman, M., Geller, M. and Huchra, J. (1986). The cluster-cluster correlation function. *The Astronomical Journal*, **91**, 1267-1273.

[58] Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *J. Statist. Plann. Inference*, **49**, 137-162.

[59] Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness of Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, New York.

[60] Rényi, A. (1961). On measures of entropy and information. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **I**., Berkeley, 547-561.

[61] Roeder, K. (1992). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *Journal of Statistical Association*, **85**, 617-624.

[62] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461-464.

[63] Seghouane, A.K., Bekara, M. (2004). A small sample model selection based on the Kullback symmetric divergence. I*EEE Trans. Signal Process*, **52**, 3314–3323.

[64] Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics.* Chapman & Hall/CRC.

[65] Serfling, R. J. (1980). *Aproximation Theorems of Mathematical Statistics.* Wiley.

[66] Shang, J. (2008). Selection criteria based on Monte Carlo simulation and cross validation in mixed models. *Far East J Theor. Statist.*, **25**, 51–72.

[67] Shang, J. and Cavanaugh, J.E. (2008). Bootstrap variants of the Akaike information criterion for mixed model selection. *Comput. Statist. Data Anal.*, **52**, 2004–2021.

[68] Soofi, E. (2000). Principal information theoretic approaches. *J. Amer. Statist. Assoc.*, **95**, 1349-1353.

[69] Soofi, E. and Retzer, J. J. (2002). Information indices: unification and applications. and entropy econometrics. *J. Econometrics*, **107**, 17-40.

[70] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Roy. Statist. Soc.* B **64**, 583–639.

[71] Strauss, D.J. (1975). A model for clustering. *Biometrika*, **62**, 467-475.

[72] Stummer, W. and Vajda, I. (2010). On divergences of finite measures and their applicability in statistics and information theory. *Statistics*, **44**, 169-187.

[73] Toma, A. (2014). Model selection criteria using divergences. *Entropy.* **16**:5, 2686-2698.

[74] Toma, A. and Broniatowski, M. (2011). Dual divergence estimators and tests: robustness results. *Journal of Multivariate Analysis.* **102**:1, 20-36.

[75] Vajda, I. (1972). On the f-divergence and singularity of probability measures. *Periodica Methematica Hungarica*, **2**(1-4), 223-234.

[76] Vajda, I. (1973). χ^α-divergence and generalized Fisher's information. *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pp. 873-886.

[77] Vajda, I. (1989). *Theory of statistical inference and information.* Kluwer Academic Publishers.

[78] Vajda, I. (1995). Information theoretic methods in statistics. Research Report No 1834, *Academy of Sciences of the Czech Republic*. Institute of Information Theory and Automation. Prague.

[79] Vuong, QH. and Wang, W. (1993). Minimum chi-square estimation and tests for model selection. *Journal of Econometrics*, **56**, 141-168.

[80] Ζωγράφος, Κ. (1987). *Συμβολή στη Στατιστική Θεωρία Πληροφοριών.* Διδακτορική Διατριβή. Πανεπιστήμιο Ιωαννίνων.

[81] Zografos, K. (1998). f-dissimilarity of several distributions in testing statistical hypotheses. *Annals of the Institute of Statistical Mathematics*, **50**, 295-310.

[82] Zografos, K. (2008). On Mardia's and Song's measures of kurtosis in elliptical distributions. *J. Multivariate Anal.*, **99**, 858-879.

[83] Zografos, K., Ferentinos, K. and Papaioannou, T. (1989). Limiting Properties of Some Measures of Information. *Ann. Inst. Statist. Math.*, **41**, No. 3, 451-460.

[84] Zografos, K., Ferentinos, K. and Papaioannou, T. (1990). ϕ-divergence statistics: sampling properties and multinomial goodness of fit and divergence tests. *Communications in Statistics-Theory and Methods*, **19**(5), 1785-1802.

[85] Zografos, K. and Nadarajah, S. (2005). Expressions for Rényi and Shannon entropies for multivariate distributions. *Statist. Probab. Lett.*, **71**, 71-84.