



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ



Ευάγγελος Καρβούνης

ΜΙΑ ΑΝΑΣΚΟΠΗΣΗ ΣΤΙΣ ΣΤΑΤΙΣΤΙΚΕΣ ΑΠΟΚΛΙΣΕΙΣ
(DIVERGENCES) ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ ΣΤΗ
ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννινα, 2025



UNIVERSITY OF IOANNINA
Department of Mathematics



Evangelos Karvounis

A REVIEW OF STATISTICAL DIVERGENCES AND
THEIR APPLICATIONS IN STATISTICAL
INFERENCE

Master's Thesis

Ioannina, 2025

*Αφιερώνεται στους γονείς μου
Χρήστο και Σοφία*

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

Εγκρίθηκε την 01/07/2025 από την εξεταστική επιτροπή:

Ονοματεπώνυμο	Βαθμίδα
Κωνσταντίνο Ζωγράφο	Καθηγητή
Απόστολο Μπατσίδα	Αναπληρωτή Καθηγητή
Γεώργιου Αυλογίαρη	Επίκουρο Καθηγητή

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

“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.”

Ευάγγελος Καρβούνης

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα Μεταπτυχιακή 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 τα μέλη της Τριμελούς Επιτροπής τον Αναπληρωτή Καθηγητή κ. Απόστολο Μπατσίδα του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων και τον Επίκουρο Καθηγητή κ. Γεώργιο Αυλογίαρη του Τμήματος Στατιστικής & Ασφαλιστικής Επιστήμης του Πανεπιστημίου Δυτικής Μακεδονίας.

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

Πρωτίστως, στον καθηγητή 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 ιδιαίτερα τον Αναπληρωτή Καθηγητή κ. Απόστολο Μπατσίδα του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων και τον Επίκουρο Καθηγητή κ. Γεώργιο Αυλογίαρη του Πανεπιστημίου Δυτικής Μακεδονίας για τις υποδείξεις τους ως μέλη της Τριμελούς Επιτροπής.

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

ΠΕΡΙΛΗΨΗ

Το ενδιαφέρον της παρούσας μεταπτυχιακής διατριβής επικεντρώνεται στην έννοια της στατιστικής απόκλισης (divergence) και των εφαρμογών της στη στατιστική συμπερασματολογία, εκτίμηση και έλεγχο υποθέσεων. Οι στατιστικές αποκλίσεις αποτελούν μέτρα ψευδο-απόστασης μεταξύ κατανομών πιθανότητας, τα οποία έχουν σχεδιαστεί και προσαρμοστεί ώστε να χρησιμοποιούνται στη μοντελοποίηση, την ανάλυση και τη μελέτη προβλημάτων που σχετίζονται με τη θεωρία πιθανοτήτων, τη μαθηματική στατιστική και άλλα συναφή επιστημονικά πεδία.

Στο **Κεφάλαιο 1** της μεταπτυχιακής διατριβής, πραγματοποιείται μια ανασκόπηση των βασικών στατιστικών αποκλίσεων που έχουν προταθεί στη βιβλιογραφία. Αρχικά, ορίστηκε το μέτρο απόκλισης Kullback and Leibler (1951), Kullback (1959), το οποίο αποτελεί ένα από τα βασικότερα μέτρα απόκλισης στη Θεωρία Πληροφορίας. Στη συνέχεια, πραγματοποιείται μια αναφορά σε άλλα μέτρα απόκλισης (divergence) που υπάρχουν, τα οποία αποτελούν ειδικές περιπτώσεις μιας ευρείας κλάσης αποκλίσεων, γνωστή και ως οικογένεια φ - αποκλίσεων (φ - divergence) που εισήχθη από τον Csiszár (1963, 1967) και ανεξάρτητα από τους Ali and Silvey (1966), για την οποία δίνεται ο ορισμός και βασικές ιδιότητες. Επιπλέον, παρουσιάζεται η Power Divergence Family των Cressie and Read (1984), η οποία είναι μία από τις σημαντικότερες αποκλίσεις, καθώς ορίστηκε για να μελετήσει υπάρχοντες χ^2 πολυωνυμικούς ελέγχους καλής προσαρμογής. Το Κεφάλαιο 1 ολοκληρώνεται με τον ορισμό της οικογένειας Density Power Divergence των Basu et al. (1998), που αποτελεί κίνητρο και βασίζονται σε αυτή τα επόμενα κεφάλαια.

Στο **Κεφάλαιο 2** της μεταπτυχιακής διατριβής, δίνεται ανασκόπηση μεθόδων στατιστικής συμπερασματολογίας που βασίζονται σε στατιστικές αποκλίσεις μεταξύ του θεωρητικού αλλά άγνωστου πιθανοθεωρητικού μοντέλου που κυβερνά τα δεδομένα και ενός εμπειρικού εκτιμητή του. Στο πλαίσιο αυτό, πραγματοποιείται μια ανασκόπηση μεθόδων εκτίμησης τύπου ελάχιστης απόκλισης που γενικεύουν και επεκτείνουν την κλασική μέθοδο μέγιστης πιθανοφάνειας. Αναλυτικότερα, χρησιμοποιείται η μέθοδος εκτίμησης που εισάγεται στους Basu et al.

(1998), οι οποίοι ελαχιστοποιώντας την οικογένεια αποκλίσεων Density Power Divergence, εισάγουν μια μέθοδο εκτίμησης τύπου ελάχιστης απόστασης, η οποία οδηγεί κατά κανόνα σε εύρωστους (robust) εκτιμητές. Επιπλέον, μελετώνται οι ιδιότητες των εκτιμητών αυτών, όπως η συνέπεια και η ασυμπτωτική κατανομή τους, οι οποίες και αποδεικνύονται λεπτομερώς στο Κεφάλαιο 5, Παράρτημα.

Το **Κεφάλαιο 3** της μεταπτυχιακής διατριβής επικεντρώνεται στην κατασκευή και μελέτη στατιστικών τεστ για τον έλεγχο της απλής μηδενικής υπόθεσης ($H_0 : \theta = \theta_0$ έναντι $H_a : \theta \neq \theta_0$) χρησιμοποιώντας την οικογένεια Density Power Divergence. Συγκεκριμένα, ορίζεται η στατιστική συνάρτηση του τεστ και προσδιορίζεται η ασυμπτωτική κατανομή της για τον έλεγχο της απλής μηδενικής υπόθεσης. Το κεφάλαιο αυτό ολοκληρώνεται, μελετώντας τη Συνάρτηση Ισχύος, που είναι ένα μέτρο αξιολόγησης της απόδοσης του προτεινόμενου ελέγχου, καθώς και με τη μελέτη της ασυμπτωτικής κατανομής του στατιστικού ελέγχου της απόκλισης Density Power Divergence, υπό μία ακολουθία τοπικών υποθέσεων.

Τέλος στον επίλογο, **Κεφάλαιο 4** πραγματοποιείται μία σύντομη περιγραφή ερευνητικών εργασιών, που αφορούν εφαρμογές της οικογένειας Density Power Divergence σε προβλήματα παραμετρικής Συμπερασματολογίας, παραπέμποντας στη σχετική βιβλιογραφία για περαιτέρω διεξοδική συζήτηση και μελέτη.

ABSTRACT

This Master Thesis is focused on the concept of statistical divergence and its applications in statistical inference estimation and testing statistical hypotheses. Statistical divergences are pseudo-distance measures between probability distributions, designed and adapted for use in modeling, analysis and the study of problems which appear in probability theory, mathematical statistics, and other related scientific fields.

Chapter 1 of the thesis is devoted in a review of the main statistical divergences proposed in the literature. This chapter begins with the definition of the pioneer divergence measure introduced by Kullback and Leibler (cf. Kullback and Leibler (1951), Kullback (1959)), which is one of the most fundamental divergence measures in Information Theory. In the sequel, some other divergence measures are discussed, which are special cases of a broad class of divergences, namely the φ -divergence family introduced by Csiszár (1963, 1967) and independently by Ali and Silvey (1966). Some fundamental properties of Csiszár's divergence are also discussed. Moreover, the Power Divergence Family introduced by Cressie and Read (1984) is presented. This is one of the most important divergence families, as it was defined in order to study the existing polynomial chi-square (χ^2) goodness-of-fit tests. Chapter 1 concludes with the introduction of the Density Power Divergence family introduced by Basu et al. (1998), which consists the basis for the subsequent chapters.

Chapter 2, is reviewing methods of statistical inference which are based on statistical divergences between the theoretical, but unknown, probabilistic model which governs the data and its empirical estimator. In this context, a review of minimum divergence-type estimation methods is presented, which generalizes and extends the classical maximum likelihood method. Specifically, the estimation method introduced by Basu et al. (1998) is employed, which minimizes the Density Power Divergence family and results in a robust minimum distance estimation method. Furthermore, the properties of these estimators—such as consistency and their asymptotic distribution—are

studied and their detailed proof is presented in Chapter 5, Appendix.

Chapter 3 focuses on the construction and analysis of statistical tests for testing the simple null hypothesis ($H_0 : \theta = \theta_0$ versus $H_\alpha : \theta \neq \theta_0$) by using the Density Power Divergence family. In particular, the test statistic is defined, and its asymptotic distribution is derived for testing the simple null hypothesis. The chapter concludes with an analysis of the Power Function, which is a measure of the effectiveness of the proposed test, along with the study of the asymptotic distribution of the Density Power Divergence test statistic under a sequence of local alternatives.

Finally, in the epilogue, Chapter 4, a brief overview of several recent research articles is provided, which concerns applications of the Density Power Divergence family to problems in parametric inference, citing the relevant literature for a more detailed discussion and study.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη

Abstract

1	Μέτρα Απόκλισης	3
1.1	Εισαγωγή	3
1.2	Kullback Leibler Απόκλιση	4
1.2.1	Kullback Leibler Απόκλιση: Ορισμοί και Ιδιότητες	4
1.2.2	Ειδικές Περιπτώσεις Αποκλίσεων	7
1.3	Csiszár Φ-Απόκλιση	9
1.3.1	Φ-απόκλιση Csiszár: Ορισμός	10
1.3.2	Φ-απόκλιση Csiszár: Ιδιότητες	12
1.3.3	Cressie and Read	19
1.4	Basu-Harris-Hjort-Jones Απόκλιση	20
2	Εκτίμηση τύπου Ελάχιστης Απόστασης μέσω της Density Power Divergence	23
2.1	Density Power Divergence: Ορισμός και Ιδιότητες	24
2.1.1	Ελάχιστη Απόκλιση Kullback-Leibler (KLD) και Εκτιμητές Μέγιστης Πιθανοφάνειας (MLE)	26
2.2	Εκτιμητής Ελάχιστης Density Power Divergence	29
2.2.1	Εκτιμητές τύπου Ελάχιστης Απόστασης μέσω της Density Power Divergence	29

2.3	Ασυμπτωτική κατανομή των MDPDE	33
3	Έλεγχος Στατιστικών Υποθέσεων Μέσω της Density Power Divergence	35
3.1	Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD	36
3.1.1	Συνάρτηση Ισχύος	46
3.2	Τοπική Υπόθεση (local alternative)	49
4	Εφαρμογές της Density Power Divergence	51
5	Παράρτημα	55
	Βιβλιογραφία	71

ΚΕΦΑΛΑΙΟ 1

ΜΕΤΡΑ ΑΠΟΚΛΙΣΗΣ

1.1 Εισαγωγή

Στη Μαθηματική Στατιστική σημαντικό ρόλο παίζει η ποσότητα κατά την οποία μειώνεται η αβεβαιότητα πραγματοποίησης ενός γεγονότος. Η ποσότητα αυτή ορίζεται ως «Πληροφορία» και έχει εισαχθεί στη βιβλιογραφία από τον Fisher (1925). Στη στατιστική υπάρχουν διάφορα μέτρα πληροφορίας, τα οποία ομαδοποιούνται στις εξής τρεις κατηγορίες: Μέτρα Απόκλισης (divergences), μέτρα τύπου Fisher και μέτρα τύπου εντροπίας (βλέπε Ferentinos and Papaioannou (1981)). Τα μέτρα απόκλισης (divergences) θα αποτελέσουν κύριο αντικείμενο αυτής της διατριβής, καθώς είναι σημαντικά στη Θεωρία Πιθανοτήτων και στη Μαθηματική Στατιστική. Ως μέτρο απόκλισης (divergence) ορίζεται η διαφορά που υπάρχει μεταξύ δύο κατανομών ή πληθυσμών στη στατιστική. Για παράδειγμα, ο έλεγχος Kolmogorov-Smirnov βασίζεται σε ένα μέτρο απόκλισης μεταξύ της εμπειρικής κατανομής και της αντίστοιχης συνάρτησης κατανομής. Επιπλέον, θεωρούνται ένας τύπος στατιστικής απόστασης μεταξύ δύο κατανομών πιθανότητας και όχι ως μέτρο απόστασης με τη συνήθη μετρική έννοια, καθώς δεν ικανοποιούν απαραίτητα όλες τις ιδιότητες μιας ευκλείδειας μετρικής.

Ειδικότερα, μια συνάρτηση $d : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ ορίζεται ως ευκλείδεια μετρική απόσταση μεταξύ των σημείων $x = (x_1, \dots, x_p)^t$ και $y = (y_1, \dots, y_p)^t$ του p -διαστάτου Ευκλείδειου χώρου $\mathbb{R}^p$ ¹, αν ικανοποιεί τις εξής τρεις ιδιότητες

1. Μη αρνητικότητα

$$d(x, y) \geq 0, \quad \forall x, y \in \mathbb{R}^p, \text{ με ισότητα αν και μόνο αν } x = y.$$

2. Συμμετρία

¹Βικιπαίδεια, *Ευκλείδεια απόσταση*, https://el.wikipedia.org/wiki/μμ_Ε, τελευταία πρόσβαση: 26 Μαΐου 2025.

$$d(x, y) = d(y, x), \quad \forall x, y \in \mathbb{R}^p.$$

3. Τριγωνική Ανισότητα

$$d(x, z) \leq d(x, y) + d(y, z), \quad \forall x, y, z \in \mathbb{R}^p.$$

Ωστόσο, αρκετά από τα μέτρα απόκλισης δεν ικανοποιούν την ιδιότητα της συμμετρίας, ενώ κανένα από αυτά δεν ικανοποιεί την ιδιότητα της τριγωνικής ανισότητας. Παρόλα αυτά, τα μέτρα απόκλισης αποτελούν εξαιρετικά χρήσιμα εργαλεία για τη μέτρηση της διαφοράς μεταξύ δύο κατανομών πιθανότητας, ιδίως στη Θεωρία Πληροφορίας, όπου και θα πραγματοποιηθεί στη συνέχεια, μια ανασκόπηση των κυριότερων μέτρων απόκλισης (divergence).

1.2 Kullback Leibler Απόκλιση

Στην ενότητα αυτή θα ασχοληθούμε με το πιο ευρέως γνωστό μέτρο απόκλισης (divergence), που είναι το μέτρο Kullback-Leibler. Η απόκλιση Kullback-Leibler εμφανίστηκε πρώτη φορά από τους Kullback and Leibler (1951) και μελετήθηκε αναλυτικότερα στο σύγγραμμα του Kullback (1959). Το μέτρο απόκλισης Kullback-Leibler είναι ένας τύπος στατιστικής απόστασης, η οποία υποδηλώνει το κατά πόσο διαφέρουν δύο πληθυσμοί ή οι αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας.

1.2.1 Kullback Leibler Απόκλιση: Ορισμοί και Ιδιότητες

Στην υποενότητα αυτή θα παρουσιαστεί το μέτρο απόκλισης Kullback-Leibler. Αρχικά θα διατυπωθεί ο ορισμός της απόκλισης Kullback-Leibler σε έναν μετρήσιμο χώρο, καθώς στη βιβλιογραφία ο ορισμός της βασίζεται στη θεωρία μέτρου. Χάριν πληρότητας δίνουμε τον ορισμό της έννοιας της απόλυτης συνέχειας της θεωρίας μέτρου (βλέπε Billingsley (1995), σελ. 422).

Παρατήρηση 1.1. Έστω ο μετρήσιμος χώρος $(\mathcal{X}, \mathcal{A})$ και μ, ν δύο προσημασμένα ή θετικά μέτρα στον χώρο $(\mathcal{X}, \mathcal{A})$. Το μέτρο ν 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 συμβολίζεται με $\nu \ll \mu$.

Διατυπώνουμε επίσης το θεώρημα των Radon-Nikodym παραγώγων στην παρακάτω παρατήρηση (βλέπε Billingsley (1995), σελ. 419).

Παρατήρηση 1.2. Έστω ο μετρήσιμος χώρος $(\mathcal{X}, \mathcal{A})$ και $P_\theta \ll \mu$, $\forall \theta$, όπου $\{P_\theta\}_{\theta \in \Theta}$ μια παραμετρική οικογένεια μέτρων πιθανότητας, με Θ ένα υποσύνολο του $\mathbb{R}^p$, $p \geq 1$. Για θ που ανήκει στον χώρο Θ , το θεώρημα των Radon-Nikodym διατυπώνει ότι για ένα σ-πεπερασμένο μέτρο μ , υπάρχει μια $f : \mathcal{X} \rightarrow [0, \infty)$ τέτοια ώστε

$$P_\theta(A) = \int f_\theta d\mu.$$

Η συνάρτηση $f_\theta = \frac{dP_\theta(x)}{d\mu}$ ονομάζεται Radon-Nikodym παράγωγος.

Βάσει των δύο αυτών παρατηρήσεων, μπορεί να οριστεί η στατιστική απόκλιση των Kullback-Leibler των μέτρων πιθανότητας P_{θ_1} και P_{θ_2} ή των αντίστοιχων Radon-Nikodym παραγώγων f_{θ_1} και f_{θ_2} , βασισμένη σε ένα πλήρως παραμετρικό πλαίσιο (βλέπε Kullback (1959)) 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 να είναι ανοικτό υποσύνολο του $\mathbb{R}^p$, $p \geq 1$. Για $\theta_1, \theta_2 \in \Theta$, συμβολίζουμε με f_{θ_i} , όπου $i = 1, 2$ τις Radon-Nikodym παραγώγους των μέτρων P_{θ_i} , $i = 1, 2$, δηλαδή $f_{\theta_i} = \frac{dP_{\theta_i}(x)}{d\mu}$, όπου μ είναι ένα σ-πεπερασμένο μέτρο στον χώρο $(\mathcal{X}, \mathcal{A})$ και $P_{\theta_i} \ll \mu$, $i = 1, 2$. Η ποσότητα

$$D_{Kull}(f_{\theta_1}, f_{\theta_2}) = D_0(\theta_1, \theta_2) = \int_{\mathcal{X}} f_{\theta_1}(x) \ln \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} d\mu(x), \quad (1.1)$$

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

Όμως χωρίς βλάβη της γενικότητας και για λόγους απλότητας, στην παρούσα διατριβή δεν θα επεκταθούμε στον μετρήσιμο χώρο $(\mathcal{X}, \mathcal{A})$ που αναφέρεται στον παραπάνω ορισμό, αλλά τα ορίσματα και τα αποτελέσματα θα διατυπωθούν στην πραγματική ευθεία ή τον Ευκλείδειο χώρο. Επιπλέον, επειδή στα επόμενα κεφάλαια θα αναπτυχθεί συμπερασματολογία με βάση στατιστικές αποκλίσεις, αυτές θα διατυπωθούν με βάση το πραγματικό αλλά άγνωστο μοντέλο που περιγράφει τα δεδομένα και την παραμετρική οικογένεια κατανομών που υιοθετείται για να τα μοντελοποιήσει.

Επομένως, πριν δοθεί ο ορισμός της απόκλισης Kullback-Leibler στην πραγματική ευθεία, θα διατυπωθεί στην παρακάτω παρατήρηση η έννοια της ταυτοποιήσιμης ή αναγνωρίσιμης (identifiable) παραμετρικής οικογένειας κατανομών (βλέπε Lehmann and Casella (1998), σελ. 24).

Παρατήρηση 1.3. *Μια παραμετρική οικογένεια πυκνοτήτων $\{f_\theta\}$, με $\theta \in \Theta$ λέγεται ότι είναι identifiable αν και μόνο αν για $\theta_1, \theta_2 \in \Theta$ με $\theta_1 \neq \theta_2$ συνεπάγεται $f_{\theta_1}(x) \neq f_{\theta_2}(x)$.*

Ορισμός 1.2. *Έστω μια τυχαία ποσότητα, η οποία περιγράφεται από την πραγματική αλλά άγνωστη πυκνότητα g . Λαμβάνοντας υπόψη ότι το πραγματικό μοντέλο είναι άγνωστο, υιοθετείται μια identifiable παραμετρική οικογένεια κατανομών f_θ για να μοντελοποιήσει τα δεδομένα, όπου θ είναι μια παράμετρος, $\theta \in \Theta$, με Θ να είναι ανοιχτό υποσύνολο του $\mathbb{R}^p$, με $p \geq 1$. Τότε ονομάζουμε απόκλιση Kullback-Leibler μεταξύ των πυκνοτήτων g και f_θ την ποσότητα*

$$D_{Kull}(g, f_\theta) = D_0(g, f_\theta) = \int_{\mathbb{R}} g(x) \ln \left(\frac{g(x)}{f_\theta(x)} \right) dx. \quad (1.2)$$

Παρατήρηση 1.4. *Ο ορισμός 1.2. εκτός από τη συνεχή περίπτωση καλύπτει και τη διακριτή με τη μοναδική διαφορά, ότι το σύμβολο του ολοκληρώματος αντικαθίσταται από εκείνο του αθροίσματος.*

Μια πρώτη ιδιότητα της απόκλισης Kullback-Leibler, είναι αυτή της μη αρνητικότητας με την ελάχιστη τιμή, να είναι η τιμή μηδέν και να επιτυγχάνεται όταν ταυτίζονται οι δύο πυκνότητες g και f_θ . Η μη αρνητικότητα της απόκλισης Kullback-Leibler αποτελεί την πρώτη από τις τρεις ιδιότητες που χαρακτηρίζουν την έννοια της «Απόστασης». Η δεύτερη ιδιότητα της Απόστασης είναι η συμμετρία, όμως στην περίπτωση της απόκλισης Kullback-Leibler, η συμμετρία δεν ικανοποιείται καθώς $D_0(g, f_\theta) \neq D_0(f_\theta, g)$. Μια πρώτη ιδέα για την αντιμετώπιση αυτού του προβλήματος προέρχεται από τον Jeffreys (1946), που όρισε την απόκλιση

$$\mathbf{J}(g, f_\theta) = D_0(g, f_\theta) + D_0(f_\theta, g), \quad (1.3)$$

η οποία είναι συμμετρική ως προς τα ορίσματά της.

Η προηγούμενη απόκλιση, αντίθετα από την Kullback-Leibler, ικανοποιεί την ιδιότητα της συμμετρίας και ορίζεται ως **J-Απόκλιση**. Όμως η τρίτη ιδιότητα της Απόστασης, αυτή της Τριγωνικής Ανισότητας, δεν ικανοποιείται γενικά, ενώ μια συζήτηση για την ικανοποίηση της τριγωνικής ιδιότητας παρατίθεται στους Liese and Vajda (1987) και Vajda (2009) (βλέπε Zografos (2023), σελ. 297). Έτσι λοιπόν, οι αποκλίσεις δεν θεωρούνται Αποστάσεις με την έννοια της «Απόστασης» στη μαθηματική βιβλιογραφία, αλλά θεωρούνται ως Στατιστικές Αποστάσεις.

1.2.2 Ειδικές Περιπτώσεις Αποκλίσεων

Προηγουμένως, αναφερθήκαμε στην κυριότερη απόκλιση Kullback-Leibler. Ωστόσο, στη βιβλιογραφία έχουν μελετηθεί και άλλα σημαντικά μέτρα απόκλισης, τα οποία αποτελούν ειδικές περιπτώσεις της φ -απόκλισης, που θα παρουσιαστεί αναλυτικότερα στην επόμενη ενότητα. Στην παρούσα υποενότητα, επιχειρείται μια σύντομη ανασκόπηση κάποιων μέτρων αποκλίσεων (divergences). Η συμβολή του Papaiοannου στην Encyclopedia of Statistics, Papaiοannου (1985), είναι ιδιαίτερα σημαντική, καθώς παρέχει μια συνεκτική παρουσίαση και ανάλυση αυτών των αποκλίσεων (divergences). Ακολουθεί, λοιπόν, μια συνοπτική ανασκόπηση ορισμένων αποκλίσεων (divergences).

Ένα πρώτο μέτρο απόκλισης δημιουργείται από τον κλασικό χ^2 έλεγχο καλής προσαρμογής ή αλλιώς χ^2 του Pearson (1900). Ο έλεγχος αυτός χρησιμοποιείται για κατηγορικά δεδομένα, είναι ένα κριτήριο που ποσοτικοποιεί τη συνολική διαφορά μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων. Η απόκλιση chi-square, που εισήχθη από τον Pearson, δίνεται από τη σχέση

$$\chi^2 = \sum_{i=1}^n \frac{(n_i - e_i)^2}{e_i}, \text{ για } e_i = np_i,$$

όπου p_i είναι η πιθανότητα μία παρατήρηση να ανήκει στην i -οστή κατηγορία δεδομένης μιας κατανομής, και n_i είναι ο αριθμός των παρατηρήσεων που ανήκουν στην i -οστή κατηγορία, $i = 1, \dots, n$. Ο έλεγχος αυτός παρουσιάζεται από τον Karl Pearson (Pearson (1900)). Ο Kagan (1963), γενικεύοντας το χ^2 του Pearson, ορίζει ως μέτρο απόκλισης στη συνεχή περίπτωση την ποσότητα

$$D_{\chi^2}(g, f_\theta) = \int_{\mathbb{R}} f_\theta(x) \left(1 - \frac{g(x)}{f_\theta(x)}\right)^2 dx, \quad (1.4)$$

η οποία ονομάζεται chi-square divergence μεταξύ της g και f_θ .

Επιπλέον, από τη συνεισφορά του Vajda (1973), ορίζεται το μέτρο απόκλισης

$$D_{\chi^\alpha}(g, f_\theta) = \int_{\mathbb{R}} f_\theta(x) \left|1 - \frac{g(x)}{f_\theta(x)}\right|^\alpha dx, \quad (1.5)$$

όπου $\alpha \geq 1$.

Το μέτρο αυτό αποτελεί γενίκευση της παραπάνω απόκλισης και ονομάζεται Vajda divergence ή χ^α -απόκλιση μεταξύ της g και f_θ . Η ιδιαιτερότητα της απόκλισης αυτής είναι η παράμετρος α , η οποία καθορίζει την ευαισθησία της

απόκλισης. Συγκεκριμένα για $\alpha = 2$ λαμβάνεται η απόκλιση του Kagan, ενώ για $\alpha = 1$ το total variation Saks (1937).

Η Hellinger distance εισήχθη ως έννοια το 1909 από τον Γερμανό μαθηματικό Ernst Hellinger Hellinger (1909), είναι ένα στατιστικό μέτρο που χρησιμοποιείται για να ποσοτικοποιήσει την ομοιότητα ή τη διαφορά μεταξύ δύο κατανομών πιθανότητας ². Η Hellinger distance θεωρείται ειδική περίπτωση της φ -απόκλισης και το μέτρο

$$D_{He}(g, f_\theta) = \frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{g(x)} - \sqrt{f_\theta(x)} \right)^2 dx \quad (1.6)$$

ή ισοδύναμα

$$D_{He}(g, f_\theta) = 1 - \int_{\mathbb{R}} \sqrt{g(x)f_\theta(x)} dx, \quad (1.7)$$

ονομάζεται Hellinger divergence μεταξύ της g και f_θ .

Μια απόκλιση που μπορεί να θεωρηθεί ως γενίκευση της Hellinger εμφανίζεται από τη συνεισφορά του Matusita (1964). Η Matusita απόκλιση έχει εισαχθεί ως ένας έλεγχος αξιολόγησης της ομοιότητας ή της διαφοράς μεταξύ δύο διαφορετικών κατανομών. Η μορφή της είναι η εξής:

$$D_{M,\alpha}(g, f_\theta) = \int_{\mathbb{R}} (g^\alpha(x) - f_\theta^\alpha(x))^{\frac{1}{\alpha}} dx, 0 < \alpha < 1 \quad (1.8)$$

και ονομάζεται Matusita divergence μεταξύ της g και f_θ .

Η απόκλιση αυτή συνδέεται με την Hellinger, όταν η παράμετρος $\alpha = 1/2$. Επίσης, στην εργασία του Matusita (1964) εισάγεται και το affinity δύο συναρτήσεων πυκνότητας πιθανότητας με την απόκλιση

$$\rho_\alpha(g, f_\theta) = \int_{\mathbb{R}} (g^\alpha(x) - f_\theta^{1-\alpha}(x)) dx. \quad (1.9)$$

Μια άλλη σχετική απόσταση είναι η Bhattacharyya distance (βλέπε Bhattacharyya (1943)), που ποσοτικοποιεί και αυτή την «ομοιότητα» δύο συναρτήσεων πυκνότητας πιθανότητας. Σε αντίθεση με την Kullback-Leibler που δεν ικανοποιεί τη συμμετρία, η απόσταση αυτή είναι συμμετρική, αλλά ούτε και αυτή υπακούει στην τρίτη ιδιότητα των αποστάσεων, αυτής της τριγωνικής ανισότητας. Έτσι λοιπόν ορίζεται βάσει αυτής της απόστασης μια νέα απόκλιση ως εξής

$$D_{Ba}(g, f_\theta) = -\ln \left(\int_{\mathbb{R}} \sqrt{g(x)f_\theta(x)} dx \right), \quad (1.10)$$

²Βικιπαίδεια, *Hellinger distance*, https://en.wikipedia.org/wiki/Hellinger_distance, τελευταία πρόσβαση: 26 Μαΐου 2025.

που ονομάζεται Bhattacharyya divergence μεταξύ της g και f_θ .

Ακόμη ένα μέτρο απόκλισης, που εισήχθη από τον Rényi (1961) είναι η Rényi απόκλιση (Rényi divergence). Είναι ένα μέτρο που αποτυπώνει την απόσταση μεταξύ κατανομών. Η Rényi απόκλιση ορίζεται από τη σχέση

$$D_{R,\alpha}(g, f_\theta) = \frac{1}{\alpha - 1} \ln \left(\int_{\mathbb{R}} g(x)^\alpha f_\theta(x)^{1-\alpha} dx \right), \quad (1.11)$$

όπου $\alpha > 0$ και $\alpha \neq 1$.

Το μέτρο αυτό σχετίζεται άμεσα με την Kullback-Leibler, καθώς στην περίπτωση που $\alpha \rightarrow 1$ έχουμε

$$\lim_{\alpha \rightarrow 1} D_{R,\alpha}(g, f_\theta) = D_{Kull}(g, f_\theta).$$

Η Rényi απόκλιση αποτελεί γενίκευση της Kullback-Leibler, προσφέροντας μεγαλύτερη ευελιξία λόγω της παραμέτρου α , η οποία επιτρέπει την προσαρμογή της ευαισθησίας σε συγκεκριμένες διαφορές μεταξύ των κατανομών, δηλαδή δίνει μεγαλύτερη ή μικρότερη βαρύτητα σε διαφορετικές περιοχές των κατανομών.

1.3 Csiszár Φ-Απόκλιση

Η απόκλιση Kullback-Leibler που επινοήθηκε από τους Kullback and Leibler (1951), αποτέλεσε ένα από τα σημαντικότερα μέτρα για τη θεωρία πληροφοριών. Όπως αναφέρθηκε και προηγουμένως, η απόκλιση Kullback-Leibler ορίζεται ως ένας τύπος απόστασης στη στατιστική, που μετρά τη διαφορά μεταξύ δύο κατανομών ή δύο συναρτήσεων πυκνότητας πιθανότητας. Για τον λόγο αυτόν έχουν πραγματοποιηθεί διάφορες προσπάθειες γενίκευσης της απόκλισης Kullback-Leibler ή του ορισμού μιας ευρείας κλάσης μέτρων απόκλισης. Μια από τις σημαντικότερες ενέργειες εύρεσης μιας ευρείας κλάσης μέτρων απόκλισης εμφανίστηκε μετά την απόκλιση Rényi από τον Csiszár (1963, 1967) και ανεξάρτητα από τους Ali and Silvey (1966). Η απόκλιση που θα μας απασχολήσει στην ενότητα αυτή θα είναι η φ -απόκλιση Csiszár. Συγκεκριμένα, ο Csiszár όρισε μία γενική κλάση αποκλίσεων που βασίζεται σε μία κυρτή συνάρτηση φ . Ανάλογα με την επιλογή της συνάρτησης φ , προκύπτουν διαφορετικά μέτρα απόκλισης, συμπεριλαμβανομένων και αυτών που συζητήθηκαν στην προηγούμενη ενότητα. Στις υποενότητες που ακολουθούν θα δοθεί ο ορισμός της απόκλισης και κάποιες από τις ιδιότητές της.

1.3.1 Φ-απόκλιση Csiszár: Ορισμός

Η φ -απόκλιση Csiszár ορίζει μια ευρεία κλάση μέτρων απόκλισης μεταξύ της πραγματικής αλλά άγνωστης πυκνότητας και του παραμετρικού μοντέλου που χρησιμοποιείται για την περιγραφή των δεδομένων. Ο ορισμός της διατυπώνεται ως εξής:

Ορισμός 1.3. Έστω μια τυχαία ποσότητα, η οποία περιγράφεται από την πραγματική αλλά άγνωστη πυκνότητα g . Λαμβάνοντας υπόψη ότι το πραγματικό μοντέλο είναι άγνωστο, υιοθετείται μια *identifiable* παραμετρική οικογένεια κατανομών f_θ για να μοντελοποιήσει τα δεδομένα, όπου θ είναι μια παράμετρος, $\theta \in \Theta$, με Θ να είναι ανοιχτό υποσύνολο του $\mathbb{R}^p$, με $p \geq 1$. Τότε ονομάζουμε φ -απόκλιση Csiszár μεταξύ των πυκνοτήτων g και f_θ την ποσότητα

$$D_\varphi(g, f_\theta) = \int_{\mathbb{R}} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx, \quad \varphi \in \Phi^*, \quad (1.12)$$

όπου Φ^* είναι η κλάση όλων των κυρτών συναρτήσεων $\varphi(x)$, $x \geq 0$, τέτοιων ώστε στο $x = 1$, $\varphi(1) = 0$, στο $x = 0$, $0\varphi(0/0) = 0$ και $0\varphi(\rho/0) = \rho \lim_{r \rightarrow \infty} \varphi(r)/r$ (βλέπε Pardo (2006), σελ. 5).

Παρατήρηση 1.5. Έστω $\varphi \in \Phi^*$ είναι διαφορίσιμη στο 1 τότε η συνάρτηση

$$\psi(x) \equiv \varphi(x) - \varphi'(1)(x - 1),$$

επίσης ανήκει στην Φ^* και έχει επιπλέον την ιδιότητα ότι $\psi'(1) = 0$. Αυτή η ιδιότητα, μαζί με την κυρτότητα, συνεπάγεται ότι $\psi(x) \geq 0$ για κάθε $x \geq 0$. Τότε,

$$\begin{aligned} D_\psi(g, f_\theta) &= \int_{\mathbb{R}} f_\theta(x) \left(\varphi\left(\frac{g(x)}{f_\theta(x)}\right) - \varphi'(1)\left(\frac{g(x)}{f_\theta(x)} - 1\right) \right) dx \\ &= \int_{\mathbb{R}} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &= D_\varphi(g, f_\theta). \end{aligned}$$

Δεδομένου ότι τα δύο μέτρα απόκλισης ταυτίζονται, μπορούμε να θεωρήσουμε ότι το σύνολο Φ^* είναι ισοδύναμο με το σύνολο

$$\Phi = \Phi^* \cap \{\varphi : \varphi'(1) = 0\}.$$

Για να είναι χρήσιμη σε στατιστικές εφαρμογές, η Φ^* εμπλουτίζεται με την πρόσθετη υπόθεση $\varphi'(1) = 0$. Η ιδιότητα $\varphi'(1) = 0$ διασφαλίζει ότι το

μέτρο απόκλισης διατηρεί τις επιθυμητές ιδιότητες, όπως μη αρνητικότητα και κυρτότητα, ενώ ταυτόχρονα απλοποιεί τον υπολογισμό και τη θεωρητική ανάλυση.

Από την παραπάνω παρατήρηση μπορούμε να δούμε ότι η Kullback-Leibler απόκλιση μπορεί να ληφθεί για $\psi(x) = x \ln(x) - x + 1$ ή για $\varphi(x) = x \ln(x)$ τότε $\psi(x) = x \ln(x) - \varphi'(1)(x - 1)$.

Παρατήρηση 1.6. Η $D_\varphi(g, f_\theta)$ ως απόκλιση δεν είναι συμμετρική. Οι Liese and Vajda (1987), σελ. 14, όρισαν μια νέα συνάρτηση η οποία είναι κυρτή. Η συνάρτηση αυτή είναι της μορφής

$$\psi(r) = \varphi(r) + r\varphi\left(\frac{1}{r}\right), \text{ για } r > 0,$$

τότε η $D_\psi(g, f_\theta)$ μπορεί να αποδοθεί ως συμμετρική και έχουμε ότι

$$D_\psi(g, f_\theta) = D_\psi(f_\theta, g) = D_\varphi(g, f_\theta) + D_\varphi(f_\theta, g).$$

Παρατήρηση 1.7. Για την απόκλιση $D_\varphi(g, f_\theta)$ η τριγωνική ανισότητα δεν ισχύει και άρα η $D_\varphi(g, f_\theta)$ δεν μπορεί να θεωρηθεί απόσταση με τη συνήθη έννοια της μετρικής (Ευκλείδεια απόσταση).

Στον παρακάτω πίνακα παρουσιάζονται οι βασικές αποκλίσεις που αναφέρθηκαν στην Ενότητα 1.2. Ο πίνακας προέρχεται από το βιβλίο του Pardo (2006), βλέπε σελ. 6, όπου καταγράφονται εκεί και άλλες σχετικές αποκλίσεις.

Φ-Συναρτήσεις	Αποκλίσεις
$x \ln(x) - x + 1$	Kullback-Leibler
$(x - 1) \ln(x), x > 0$	j-Απόκλιση
$\frac{1}{2}(x - 1)^2$	Pearson, Kagan
$ 1 - x ^\alpha, \alpha \geq 1$	χ -απόκλιση τάξης α Vajda
$\frac{1}{2}(1 - x^{\frac{1}{2}})^2$	Hellinger
$ 1 - x^\alpha ^{\frac{1}{\alpha}}, 0 < \alpha < 1$	Matusita
$\text{sgn}(\alpha - 1)x^\alpha, \alpha > 0$	Affinity Matusita

Θα ήταν παράλειψη να μην αναφερθούμε σε μέτρα που δεν ανήκουν στις φ -αποκλίσεις. Τέτοια μέτρα είναι οι αποκλίσεις Rényi και Bhattacharyya, τα οποία δεν ανήκουν στις φ -αποκλίσεις. Ωστόσο, τα μέτρα αυτά μπορούν να εκφραστούν μέσω της (h, φ) -απόκλισης

$$D_\varphi^h(g, f_\theta) = h(D_\varphi(g, f_\theta)), \quad (1.13)$$

όπου h είναι μια διαφορίσιμη αύξουσα πραγματική συνάρτηση στο $[0, \varphi(0) + \lim_{t \rightarrow \infty} f(t)/t]$ του $[0, \infty]$ και παρατίθενται στον παρακάτω πίνακα (βλέπε Pardo (2006), σελ. 8).

Αποκλίσεις	$h(x)$	$\varphi(x)$
Rényi	$\frac{1}{r(r-1)} \ln(r(r-1)x + 1), r \neq 0, 1$	$\frac{x^r - r(x-1) - 1}{r(r-1)}$
Bhattacharay	$-\ln(1-x)$	$-x^{\frac{1}{2}} + \frac{1}{2}(x+1)$

Αυτές οι αποκλίσεις ονομάζονται (h, φ) -Divergences και έχουν εισαχθεί στη βιβλιογραφία από την Menéndez et al. (1995).

1.3.2 Φ-απόκλιση Csiszár: Ιδιότητες

Σε αυτήν την υποενότητα θα επεκταθούμε σε κάποιες από τις σημαντικότερες ιδιότητες των φ -αποκλίσεων. Το εύρος των ιδιοτήτων είναι τεράστιο, μια αναλυτική μελέτη αυτών παρουσιάζεται από τον Vajda (1987, 1995). Όπως αναφέρει και ο Pardo (2006), είναι λογικό να απαιτούμε από μια απόκλιση την ιδιότητα να αυξάνεται όταν οι δύο συναρτήσεις πυκνότητας αποκλίνουν μεταξύ τους. Η ιδέα αυτή θα αποτελέσει την πρώτη πρόταση που αναφέρεται παρακάτω (Pardo (2006), Proposition 1.1, σελ. 8).

Πρόταση 1.1. Έστω ότι g και f_θ είναι δύο συναρτήσεις πυκνότητας πιθανότητας και έστω $\varphi \in \Phi^*$ διαφορίσιμη συνάρτηση στο $t = 1$. Τότε

$$0 \leq D_\varphi(g, f_\theta) \leq \varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \quad (1.14)$$

όπου

$$D_\varphi(g, f_\theta) = 0, \text{ αν } g(x) \equiv f_\theta(x) \quad (1.15)$$

και

$$D_\varphi(g, f_\theta) = \varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \text{ αν } S_1 \cap S_2 = \emptyset. \quad (1.16)$$

Αν φ είναι επίσης αυστηρά κυρτή συνάρτηση στο $t = 1$, τότε η (1.15) ισχύει αν και μόνο αν $g(x) \equiv f_\theta(x)$. Αν επιπλέον

$$\varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} < \infty,$$

τότε η (1.16) ισχύει αν και μόνο αν $S_1 \cap S_2 = \emptyset$, όπου S_i , για, $i = 1, 2$, είναι το σύνολο θετικότητας των συναρτήσεων πυκνότητας πιθανότητας g και f_θ , αντίστοιχα.

Απόδειξη. Η απόδειξη ακολουθεί πιστά την απόδειξη της Πρότασης 1.1 στη μονογραφία Pardo (2006), σελ. 8. Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης και παραπέμπουμε επίσης σε ανάλογη απόδειξη στον Αυλογιάρης (2017), σελ. 36, σε περιβάλλον τοπικών αποκλίσεων.

Χρησιμοποιώντας τη μη αρνητικότητα της συνάρτησης ψ από την Παρατήρηση 1.5.

$$\psi(x) \equiv \varphi(x) - \varphi'(1)(x - 1), \quad (1.17)$$

έχουμε $D_\psi(g, f_\theta) \geq 0$, επίσης από την Παρατήρηση 1.5. έχουμε ότι

$$D_\psi(g, f_\theta) = D_\varphi(g, f_\theta).$$

Συνεπώς, $D_\varphi(g, f_\theta) \geq 0$.

Είναι γνωστό ότι για κάθε κυρτή συνάρτηση φ , ισχύει (βλέπε Liese and Vajda, 1987, σελ. 210)

$$\varphi(t) \leq \varphi(0) + t \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \quad t \geq 0. \quad (1.18)$$

Αν η φ είναι αυστηρά κυρτή συνάρτηση στο $t_0 \in (0, \infty)$ τότε η (1.18) είναι αυστηρά κυρτή συνάρτηση για όλα τα $t > 0$ και άρα χρησιμοποιώντας την (1.18) και αντικαθιστώντας το t με $g(x)/f_\theta(x)$ έχουμε

$$\begin{aligned} D_\varphi(g, f_\theta) &= \int_{\mathbb{R}} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &\leq \int_{\mathbb{R}} f_\theta(x) \left(\varphi(0) + \frac{g(x)}{f_\theta(x)} \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} \right) dx \\ &= \int_{\mathbb{R}} \left(f_\theta(x) \varphi(0) + f_\theta(x) \frac{g(x)}{f_\theta(x)} \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} \right) dx \\ &= \varphi(0) \int_{\mathbb{R}} f_\theta(x) dx + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} \int_{\mathbb{R}} g(x) dx \\ &= \varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}. \end{aligned}$$

Καθώς τα ολοκληρώματα $\int_{\mathbb{R}} f_\theta(x) dx$ και $\int_{\mathbb{R}} g(x) dx$ είναι ίσα με τη μονάδα από τον ορισμό των συναρτήσεων πυκνότητας πιθανότητας, το πεδίο τιμών της $D_\varphi(g, f_\theta)$ είναι το

$$0 \leq D_\varphi(g, f_\theta) \leq \varphi(0) + \lim_{r \rightarrow \infty} \varphi(r)/r.$$

Είναι σαφές ότι αν $g(x) \equiv f_\theta(x) \implies D_\varphi(g, f_\theta) = 0$, από ιδιότητα της $\varphi(\cdot)$ συνάρτησης όπου $\varphi(g(x)/f_\theta(x)) = \varphi(1) = 0$.

Στη συνέχεια, το σύνολο θετικότητας που αναφέρεται πιο πάνω είναι το σύνολο των σημείων του πεδίου ορισμού όπου η αντίστοιχη συνάρτηση πυκνότητας είναι διάφορη του μηδενός. Συγκεκριμένα με $S_1^c \cap S_2$ θα συμβολίζουμε το σύνολο για το οποίο η $g(x) = 0$ ενώ $f_\theta(x) \neq 0$ και με $S_2^c \cap S_1$ θα συμβολίζουμε το σύνολο για το οποίο η $g(x) \neq 0$ ενώ $f_\theta(x) = 0$.

Έτσι, αν $S_1 \cap S_2 = \emptyset$ έχουμε

$$\begin{aligned} D_\varphi(g, f_\theta) &= \int_{\mathbb{R}} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &= \int_{S_1^c \cap S_2} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx + \int_{S_1 \cap S_2^c} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx. \end{aligned}$$

Για $x \in S_1^c \cap S_2 \implies g(x) = 0$, το δεξί μέρος της παραπάνω εξίσωσης γίνεται

$$\begin{aligned} \int_{S_1^c \cap S_2} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx &= \int_{S_1^c \cap S_2} f_\theta(x) \varphi\left(\frac{0}{f_\theta(x)}\right) dx \\ &= \int_{S_1^c \cap S_2} f_\theta(x) \varphi(0) dx \\ &= \varphi(0) \int_{S_1^c \cap S_2} f_\theta(x) dx \\ &= \varphi(0), \end{aligned}$$

όπου $\int_{S_1^c \cap S_2} f_\theta(x) dx = 1$.

Ενώ για $x \in S_1 \cap S_2^c \implies f_\theta(x) = 0$, το δεξί μέρος της παραπάνω εξίσωσης γίνεται

$$\begin{aligned} \int_{S_1 \cap S_2^c} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx &= \int_{S_1 \cap S_2^c} 0 \varphi\left(\frac{g(x)}{0}\right) dx \\ &\stackrel{0\varphi(\frac{r}{0}) = \rho \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}}{=} \int_{S_1 \cap S_2^c} g(x) \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} dx \\ &= \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} \int_{S_1 \cap S_2^c} g(x) dx \\ &= \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \end{aligned}$$

όπου $\int_{S_1 \cap S_2^c} g(x) dx = 1$.

Συνεπώς, συνδυάζοντας τις παραπάνω εξισώσεις, αν $S_1 \cap S_2 = \emptyset$ έχουμε

$$\begin{aligned} D_\varphi(g, f_\theta) &= \int_{\mathbb{R}} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &= \int_{S_1^c \cap S_2} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx + \int_{S_1 \cap S_2^c} f_\theta(x) \varphi\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &= \varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}. \end{aligned}$$

Τώρα στην περίπτωση που η φ είναι αυστηρά κυρτή συνάρτηση στο $t = 1$, θα δείξουμε ότι ισχύει η συνεπαγωγή $D_\varphi(g, f_\theta) = 0 \implies g(x) \equiv f_\theta(x)$.

Δεδομένου ότι, η συνάρτηση φ είναι αυστηρά κυρτή στο $t = 1$ από τη σχέση (1.17) η συνάρτηση $\psi(\cdot)$ εκφράζει την απόκλιση της $\varphi(\cdot)$ από την εφαπτομένη της στο σημείο 1. Η $\varphi(\cdot)$ θα είναι αυστηρά κυρτή συνάρτηση πάνω από την εφαπτομένη της σε κάθε σημείο $t \neq 1$ και συνεπώς $\psi(\cdot) > 0$. Άρα έχουμε

$$\psi\left(\frac{g(x)}{f_\theta(x)}\right) > 0,$$

για $g(x)/f_\theta(x) > 1$ και για $g(x)/f_\theta(x) < 1$, όπου ψ ορίζεται από τη σχέση (1.17).

Αν $D_\varphi(g, f_\theta) = D_\psi(g, f_\theta) = 0$ τότε $g(x)/f_\theta(x) \leq 1$ ή $g(x)/f_\theta(x) \geq 1$. Ας υποθέσουμε πρώτα ότι $g(x)/f_\theta(x) \leq 1$, είναι γνωστό ότι $D_\psi(g, f_\theta) = 0$

$$\begin{aligned} D_\psi(g, f_\theta) &= \int_{\mathbb{R}} f_\theta(x) \psi\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &= \int_{\mathbb{R}} f_\theta(x) \left(\varphi\left(\frac{g(x)}{f_\theta(x)}\right) - \varphi'(1) \left(\frac{g(x)}{f_\theta(x)} - 1 \right) \right) dx \\ &= D_\varphi(g, f_\theta) - \varphi'(1) \int_{\mathbb{R}} f_\theta(x) \left(\frac{g(x)}{f_\theta(x)} - 1 \right) dx \\ &= 0 - \varphi'(1) \int_{\mathbb{R}} f_\theta(x) \left(\frac{g(x)}{f_\theta(x)} - 1 \right) dx \\ &= -\varphi'(1) \int_{\mathbb{R}} \left(\frac{g(x)}{f_\theta(x)} - 1 \right) dF_\theta(x) \end{aligned}$$

όπου $\frac{dF_\theta(x)}{dx} = f_\theta(x)$ ισοδυναμεί με $f_\theta(x)dx = dF_\theta(x)$ με $F_\theta(\cdot)$ να είναι η αθροιστική συνάρτηση κατανομής της $f_\theta(\cdot)$.

Εφόσον η φ είναι αυστηρά κυρτή στο $t = 1$, πρέπει να ισχύει $g(x) \equiv f_\theta(x)$ για να ισχύει η ισότητα $D_\psi(g, f_\theta) = 0$. Όμοια αποτελέσματα μπορούν να προσδιοριστούν για την περίπτωση $g(x)/f_\theta(x) \geq 1$.

Στη συνέχεια, η αυστηρή κυρτότητα της φ στο $t = 1$ έχει ως συνέπεια την αυστηρή ανισότητα στη σχέση (1.18), δηλαδή ότι

$$\varphi(t) < \varphi(0) + t \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \quad \forall t > 0,$$

Θεωρούμε τη συνάρτηση

$$l(t) = \varphi(0) - \varphi(t) + t \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \quad \forall t > 0,$$

όπου $l(t)$ είναι θετική $\forall t > 0$.

Για $x \in S_1$, συνεπάγεται ότι x είναι τέτοιο ώστε $g(x) > 0$, τότε $t = \frac{g(x)}{f_\theta(x)} > 0$ και $l\left(\frac{g(x)}{f_\theta(x)}\right) > 0$.

Επομένως,

$$\begin{aligned} D_l(g, f_\theta) &= \int_{\mathbb{R}} f_\theta(x) l\left(\frac{g(x)}{f_\theta(x)}\right) dx \\ &= \int_{\mathbb{R}} f_\theta(x) \left(\varphi(0) - \varphi\left(\frac{g(x)}{f_\theta(x)}\right) + \frac{g(x)}{f_\theta(x)} \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} \right) dx \\ &= -D_\varphi(g, f_\theta) + \varphi(0) \int_{\mathbb{R}} f_\theta(x) dx + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r} \int_{\mathbb{R}} g(x) dx \\ &= -D_\varphi(g, f_\theta) + \varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r}, \end{aligned}$$

επειδή $\int_{\mathbb{R}} f_\theta(x) dx = 1 = \int_{\mathbb{R}} g(x) dx$.

Όμως από τη σχέση (1.16) έχουμε

$$D_\varphi(g, f_\theta) = \varphi(0) + \lim_{r \rightarrow \infty} \frac{\varphi(r)}{r},$$

επομένως,

$$D_l(g, f_\theta) = 0 \quad \mu\epsilon \quad l\left(\frac{g(x)}{f_\theta(x)}\right) > 0.$$

Τότε, $f_\theta(x) = 0$ επειδή $D_l(g, f_\theta) = 0$ και $l\left(\frac{g(x)}{f_\theta(x)}\right) > 0$, συνεπάγεται $x \notin S_2$, που ολοκληρώνει την απόδειξη. □

Για την καλύτερη κατανόηση της πρότασης που ακολουθεί, θα ορίσουμε για αυτήν το εξής: Με f_{θ_1} θα συμβολίζουμε την αληθινή αλλά άγνωστη συνάρτηση πυκνότητας και με f_{θ_2} την πυκνότητα που χρησιμοποιούμε για να περιγράψουμε το μοντέλο.

Έστω $X_1, X_2, \dots, X_n$ να είναι ένα δείγμα από την αληθή αλλά άγνωστη αθροιστική συνάρτηση κατανομής F_{θ_1} με $\theta_1 \in \Theta$. Έστω F_{θ_2} το μοντέλο που υιοθετείται για να μοντελοποιήσει τα δεδομένα και $f_{\theta_i} = \frac{dF_{\theta_i}}{dx}(x)$ είναι η αντίστοιχη συνάρτηση πυκνότητας πιθανότητας, $i = 1, 2$. Υποθέτουμε ότι υπάρχει ένας μετασχηματισμός $T: \mathbb{R} \rightarrow \mathbb{R}$ τέτοιος ώστε

$$G_{\theta_i}(A) = F_{\theta_i}(T^{-1}(A)), \quad (1.19)$$

με $A \subseteq \mathbb{R}$, και έστω

$$g_{\theta_i}(t) = \frac{dG_{\theta_i}}{dt}(t), f_{\theta_i}(x/t) = \frac{dF_{\theta_i}}{dG_{\theta_i}} \quad (1.20)$$

όπου t συμβολίζονται οι τιμές του T .

Η επόμενη πρόταση εξετάζει τη συμπεριφορά των μέτρων απόκλισης υπό μετασχηματισμούς δεδομένων. Συγκεκριμένα, όταν εφαρμόζουμε έναν μετασχηματισμό που απλοποιεί τα δεδομένα, η απόκλιση μεταξύ των κατανομών μειώνεται, καθώς μέρος της πληροφορίας μπορεί να χαθεί. Εντούτοις, αν ο μετασχηματισμός T είναι επαρκής για τις κατανομές f_{θ_1} και f_{θ_2} , τότε διατηρείται η πλήρης πληροφορία σχετικά με την απόσταση των κατανομών, και η απόκλιση παραμένει ίση με την αρχική (Pardo (2006), Proposition 1.2, σελ. 11).

Πρόταση 1.2. Έστω $\varphi \in \Phi^*$ και $F_{\theta_i}, G_{\theta_i}$ είναι δύο άγνωστες αθροιστικές συναρτήσεις που ορίζονται από τις σχέσεις (1.19) και (1.20), τότε έχουμε

$$D_{\varphi}(g_{\theta_1}, g_{\theta_2}) \leq D_{\varphi}(f_{\theta_1}, f_{\theta_2}).$$

Η ισότητα ισχύει αν T είναι επαρκής στατιστική συνάρτηση για τις συναρτήσεις πυκνότητας πιθανότητας $f_{\theta_1}(x)$ και $f_{\theta_2}(x)$.

Απόδειξη. Η απόδειξη ακολουθεί πιστά την απόδειξη της Πρότασης 1.2 στη μονογραφία Pardo (2006), σελ. 12. Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης.

Για την απόκλιση $D_\varphi(f_{\theta_1}, f_{\theta_2})$ έχουμε

$$\begin{aligned} D_\varphi(f_{\theta_1}, f_{\theta_2}) &= \int_{\mathbb{R}} f_{\theta_2}(x) \varphi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_{\theta_2}(x/t) g_{\theta_2}(t) \varphi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) dt dx \\ &= \int_{\mathbb{R}} g_{\theta_2}(t) \left(\int_{\mathbb{R}} f_{\theta_2}(x/t) \varphi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) dx \right) dt. \end{aligned}$$

Εφαρμόζοντας την Jensen's ανισότητα για την κυρτή $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, προκύπτει

$$D_\varphi(f_{\theta_1}, f_{\theta_2}) \geq \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\int_{\mathbb{R}} f_{\theta_2}(x/t) \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} dx \right) \right) dt. \quad (1.21)$$

Παρατηρείτε ότι

$$\begin{aligned} D_\varphi(f_{\theta_1}, f_{\theta_2}) &\geq \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\int_{\mathbb{R}} f_{\theta_2}(x/t) \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} dx \right) \right) dt \\ &= \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\int_{\mathbb{R}} f_{\theta_2}(x/t) \frac{\frac{dF_{\theta_1}(x)}{dx}}{\frac{dF_{\theta_2}(x)}{dx}} dx \right) \right) dt \\ &= \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\int_{\mathbb{R}} f_{\theta_2}(x/t) \frac{\frac{dF_{\theta_1}(x) dG_{\theta_1}(t)}{dG_{\theta_1}(t) dt}}{\frac{dF_{\theta_2}(x) dG_{\theta_2}(t)}{dG_{\theta_2}(t) dt}} dx \right) \right) dt \\ &= \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\int_{\mathbb{R}} f_{\theta_2}(x/t) \frac{f_{\theta_1}(x/t) g_{\theta_1}(t)}{f_{\theta_2}(x/t) g_{\theta_2}(t)} dx \right) \right) dt \\ &= \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\int_{\mathbb{R}} f_{\theta_1}(x/t) \frac{g_{\theta_1}(t)}{g_{\theta_2}(t)} dx \right) \right) dt \\ &= \int_{\mathbb{R}} g_{\theta_2}(t) \left(\varphi \left(\frac{g_{\theta_1}(t)}{g_{\theta_2}(t)} \int_{\mathbb{R}} f_{\theta_1}(x/t) dx \right) \right) dt \\ &= D_\varphi(g_{\theta_1}, g_{\theta_2}). \end{aligned}$$

Συνεπώς, προκύπτει η σχέση

$$D_\varphi(f_{\theta_1}, f_{\theta_2}) \geq \int_{\mathbb{R}} g_{\theta_2}(t) \varphi\left(\frac{g_{\theta_1}(t)}{g_{\theta_2}(t)}\right) dt = D_\varphi(g_{\theta_1}, g_{\theta_2}). \quad (1.22)$$

Από την απόδειξη της ανισότητας $D_\varphi(f_{\theta_1}, f_{\theta_2}) \geq D_\varphi(g_{\theta_1}, g_{\theta_2})$ προκύπτει ότι

$$\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} = \frac{f_{\theta_1}(x/t) g_{\theta_1}(t)}{f_{\theta_2}(x/t) g_{\theta_2}(t)}.$$

Επιπλέον, αν φ είναι αυστηρά κυρτή, η ισότητα ισχύει αν μόνο αν

$$\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} = \int_{\mathbb{R}} f_{\theta_2}(x/t) \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} dx, \text{ για όλα τα } x.$$

Τότε χρησιμοποιώντας το παραγοντικό θεώρημα, η ισότητα ισχύει αν το T είναι επαρκές για κάθε συνάρτηση πυκνότητας πιθανότητας $f_{\theta_1}(x)$ και $f_{\theta_2}(x)$. $\square$

Παρατήρηση 1.8. Είναι προφανές ότι αν η h των (h, φ) -Divergences είναι μια διαφορίσιμη αύξουσα συνάρτηση τότε οι Προτάσεις 1.1 και 1.2, ικανοποιούνται και για τις (h, φ) -Divergences (βλέπε Pardo (2006)).

Το μέτρο απόκλισης του Csiszár, που παρουσιάστηκε, ορίζει μια ευρεία κλάση αποκλίσεων η οποία περιλαμβάνει ως ειδικές περιπτώσεις ποικιλία αποκλίσεων που είχαν εμφανιστεί στη βιβλιογραφία. Στη συνέχεια θα αναφερθούμε σε μία από τις σημαντικότερες οικογένειες αποκλίσεων που ανήκει στις φ -αποκλίσεις, την Power Divergence Family των Cressie and Read (1984).

1.3.3 Cressie and Read

Ακόμη μια απόκλιση που θεωρείται στη στατιστική από τις σημαντικότερες οικογένειες αποκλίσεων, είναι η απόκλιση των Cressie and Read (1984). Η οικογένεια αυτή είναι γνωστή ως Power Divergence Family και δίνεται από τη σχέση

$$D_{\varphi_{(\lambda)}}(g, f_{\theta}) = \frac{1}{\lambda(\lambda+1)} \left(\int_{\mathbb{R}} \frac{g^{1+\lambda}(x)}{f_{\theta}^{\lambda}(x)} dx - 1 \right), -\infty < \lambda < \infty, \lambda \neq -1, 0, \quad (1.23)$$

και προκύπτει από την Csiszár φ -απόκλιση για

$$\varphi_{(\lambda)}(x) = \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}, \lambda \neq -1, 0, x > 0.$$

Για $\lambda = 0$ η Power Divergence Family ταυτίζεται με την Kullback-Leibler απόκλιση $D(g, f_{\theta})$, ενώ για $\lambda = -1$ η Power Divergence Family ταυτίζεται με την Kullback-Leibler απόκλιση $D(f_{\theta}, g)$. Η $D_{\varphi_{(\lambda)}}$ είναι επίσης συνεχής ως προς λ .

Η οικογένεια Power Divergence Family, ορίστηκε για να μελετήσει τους υπάρχοντες πολυωνυμικούς ελέγχους καλής προσαρμογής τύπου χ^2 . Αυτή η οικογένεια περιλαμβάνει ως ειδικές περιπτώσεις τον έλεγχο του Pearson χ^2 για

($\lambda \rightarrow 1$), τον λόγο πιθανοφανειών (log likelihood ratio test) για ($\lambda \rightarrow 0$), τροποποιημένες μορφές αυτών και τον έλεγχο Freeman-Tukey για ($\lambda \rightarrow -1/2$). Τα τελευταία χρόνια, έχουν δημοσιευθεί πολλά άρθρα στη στατιστική βιβλιογραφία που χρησιμοποιούν την οικογένεια Power Divergence Family, για την εξαγωγή ανταγωνιστικών εκτιμητών καθώς και στατιστικών τεστ. Για περισσότερες πληροφορίες παραπέμπουμε στο βιβλίο του Pardo (2006).

1.4 Basu-Harris-Hjort-Jones Απόκλιση

Στην ενότητα αυτή θα αναφερθούμε σε ένα από τα πιο πρόσφατα μέτρα απόκλισης, που προτάθηκε από μια διαδικασία εύρεσης μιας μεθοδολογίας εκτίμησης παραμέτρων. Το μέτρο απόκλισης δημοσιεύτηκε στην εργασία των Basu-Harris-Hjort-Jones, Basu et al. (1998), και αναφέρεται στην density power divergence μεταξύ των g και f_θ συναρτήσεων πυκνότητας. Αυτή η απόκλιση θα αναλυθεί εκτενώς και θα αποτελέσει το κεντρικό αντικείμενο συζήτησης στο Κεφάλαιο 2.

Ορισμός 1.4. *Η ποσότητα*

$$d_\alpha(g, f_\theta) = \int_{\mathbb{R}} \left\{ f_\theta^{\alpha+1}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f_\theta^\alpha(x) + \frac{1}{\alpha} g^{\alpha+1}(x) \right\} dx, \alpha > 0, \quad (1.24)$$

ονομάζεται *Basu-Harris-Hjort-Jones απόκλιση μεταξύ των g και f_θ ή BHHJ απόκλιση*.

Η BHHJ απόκλιση εξαρτάται και ρυθμίζεται από την παράμετρο α . Η παράμετρος α είναι θετική και όπως θα δούμε και στο Κεφάλαιο 2 το όριο της απόκλισης BHHJ, για $\alpha \rightarrow 0$ ταυτίζεται με την Kullback-Leibler, ενώ για $\alpha \rightarrow 1$ ταυτίζεται με την L_2 απόσταση (ή Ευκλείδεια απόσταση) μεταξύ των g και f .

Παρατήρηση 1.9. *Η οικογένεια αποκλίσεων BHHJ από τη συνεισφορά των Mattheou and Karagrigoriou (2010), μπορεί να γενικευτεί μέσω μιας κυρτής συνάρτησης φ_α στο $[0, \infty)$. Ειδικότερα είναι:*

$$\begin{aligned} d_\alpha(g, f_\theta) &= \int_{\mathbb{R}} \left\{ f_\theta^{\alpha+1}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f_\theta^\alpha(x) + \frac{1}{\alpha} g^{\alpha+1}(x) \right\} dx \\ &= E_g \left(g^\alpha(x) \varphi_\alpha \left(\frac{f_\theta(x)}{g(x)} \right) \right) = \int_{\mathbb{R}} g^{\alpha+1}(x) \varphi_\alpha \left(\frac{f_\theta(x)}{g(x)} \right) dx, \alpha > 0, \end{aligned}$$

όπου η φ_α πληροί τις συνθήκες $\varphi_\alpha(1) = 0$, $\varphi'_\alpha(1) = 0$, $\varphi''_\alpha(1) \neq 0$, $0\varphi_\alpha\left(\frac{0}{0}\right) = 0$ και $0\varphi_\alpha\left(\frac{u}{0}\right) = u \lim_{r \rightarrow \infty} \frac{\varphi_\alpha(r)}{r}$ για $u > 0$. Η γενικευμένη απόκλιση BHHJ δίνεται

για φ_α ίση με

$$\varphi_\alpha(u) = u^{(\alpha+1)} - \left(1 + \frac{1}{\alpha}\right) u^\alpha + \frac{1}{\alpha}. \quad (1.25)$$

Επιπλέον, η απόκλιση BHHJ αποτελεί ειδική περίπτωση της Bregman Divergence που έχει τύπο

$$D_\alpha(g, f_\theta) = \int_{\mathbb{R}} \{T(f_\theta(x)) - T(g(x)) - [f_\theta(x) - g(x)]T'(g(x))\} dx, \quad \alpha > 0,$$

όπου T κυρτή συνάρτηση και για $T(u) = u^{\alpha+1}$ προκύπτει η $d_\alpha(g, f)$ (βλέπε Basu et al. (2011), σελ 303).

ΚΕΦΑΛΑΙΟ 2

ΕΚΤΙΜΗΣΗ ΤΥΠΟΥ ΕΛΑΧΙΣΤΗΣ ΑΠΟΣΤΑΣΗΣ ΜΕΣΩ ΤΗΣ DENSITY POWER DIVERGENCE

Κατά τη διάρκεια των τελευταίων δεκαετιών, η εκτίμηση έχει αποτελέσει έναν από τους κυριότερους στόχους στη Στατιστική. Στόχος της εκτίμησης είναι η προσέγγιση των άγνωστων παραμέτρων ενός πληθυσμού, μιας κατανομής, βασισζόμενη στα δεδομένα που συλλέγονται από την κατανομή αυτή. Συγκεκριμένα, με την εισαγωγή των μέτρων αποκλίσεων ή στατιστικών αποστάσεων, έχουν προταθεί διάφορες μεθοδολογίες εκτίμησης. Αρκετές εργασίες χρησιμοποιούν την οικογένεια των φ -αποκλίσεων, για να εισαχθούν μέθοδοι εκτίμησης τύπου ελάχιστης απόστασης. Ο λόγος αξιοποίησης αυτών των αποκλίσεων είναι ότι με την ελαχιστοποίηση της απόστασης παράγονται εκτιμητές ασυμπτωτικά αποδοτικοί και αρκετοί από αυτούς παρουσιάζουν και καλές ιδιότητες ευρωστίας (robustness). Μια οργανωμένη προσπάθεια παρουσίασης αυτών των εκτιμητών επιχειρείται στη μονογραφία του Pardo (2006), Κεφάλαιο 5, όμως η συγκεκριμένη ανάλυση επικεντρώνεται κυρίως σε διακριτά μοντέλα.

Η παρούσα διατριβή δεν θα εστιάσει το ενδιαφέρον της σε μεθόδους εκτίμησης τύπου ελάχιστης απόστασης, χρησιμοποιώντας την οικογένεια των φ -αποκλίσεων, αλλά χρησιμοποιώντας μια άλλη οικογένεια αποκλίσεων. Η οικογένεια αποκλίσεων αυτή ονομάζεται Density Power Divergence για την οποία πραγματοποιήθηκε μια αρχική αναφορά στο Κεφάλαιο 1 αυτής της διατριβής. Βάσει αυτής της απόκλισης θα παρουσιαστούν μέθοδοι στατιστικής συμπερασματολογίας που βασίζονται σε στατιστικές αποκλίσεις μεταξύ του άγνωστου πιθανοθεωρητικού μοντέλου που κυβερνά τα δεδομένα και της παραμετρικής οικογένειας κατανομών που υιοθετείται για να τα μοντελοποιήσει. Στο πλαίσιο αυτό, θα γίνει μια ανασκόπηση μεθόδων εκτίμησης τύπου ελάχιστης απόκλισης που γενικεύουν και επεκτείνουν την κλασική μέθοδο μέγιστης πιθανοφάνειας.

2.1 Density Power Divergence: Ορισμός και Ιδιότητες

Σε αυτήν την ενότητα θα αναφερθούμε στην οικογένεια αποκλίσεων Density Power Divergence, η οποία εισήχθη στην εργασία των Basu et al. (1998). Για τη διατύπωση του ορισμού της Density Power Divergence, θεωρούμε μια παραμετρική οικογένεια κατανομών $\{F_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$ που εξαρτάται από μία παράμετρο $\theta \in \Theta \subseteq \mathbb{R}^p$, με αντίστοιχη συνάρτηση πυκνότητας πιθανότητας $\{f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Υποθέτουμε ακόμη ότι g είναι μία άλλη συνάρτηση πυκνότητας πιθανότητας.

Ας θεωρήσουμε ότι υπάρχει ένα τυχαίο δείγμα $X_1, \dots, X_n$ της X , το οποίο περιγράφεται από την πραγματική αλλά άγνωστη πυκνότητα g . Λαμβάνοντας υπόψη ότι το πραγματικό μοντέλο g είναι άγνωστο, υποθέτουμε ότι $\{f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$ είναι μία ταυτοποιήσιμη (identifiable) παραμετρική οικογένεια κατανομών, η οποία χρησιμοποιείται για την περιγραφή των παρατηρήσεων $\{x_1, \dots, x_n\}$. Στο πλαίσιο αυτό, ο ορισμός της Density Power Divergence διατυπώνεται ως εξής:

Ορισμός 2.1. Η ποσότητα $d_\alpha(g, f_\theta)$

$$d_\alpha(g, f_\theta) = \int_{\mathbb{R}} \{f_\theta^{1+\alpha}(x) - (1 + \frac{1}{\alpha})g(x)f_\theta^\alpha(x) + \frac{1}{\alpha}g^{1+\alpha}(x)\}dx, \quad \alpha > 0, \quad (2.1)$$

ονομάζεται *Density Power Divergence* ή *DPD* απόκλιση μεταξύ των g και f_θ πυκνοτήτων.

Παρατηρούμε ότι η Density Power Divergence είναι άμεσα συνδεδεμένη με μια θετική παράμετρο α . Οι Basu et al. (1998) κατασκεύασαν την οικογένεια αυτή με τέτοιο τρόπο ώστε όταν η παράμετρος είναι μη αρνητική ($\alpha > 0$), κατά την ελαχιστοποίηση της απόκλισης αυτής, να παράγονται εκτιμητές που είναι εύρωστοι (robust), ενώ όταν η παράμετρος α τείνει στο μηδέν, η απόκλιση $d_\alpha(g, f_\theta)$ να συμπίπτει με την αντίστοιχη Kullback-Leibler, όπως θα δειχτεί αναλυτικά στις επόμενες ενότητες.

Γενικά θα μας απασχολήσουν τιμές της παραμέτρου α που είναι μεταξύ του μηδενός και του ένα, όπου για $\alpha \rightarrow 1$ συμπίπτει με την L_2 απόσταση, δηλαδή την Ευκλείδεια απόσταση $d_1(g, f_\theta) = \int_{\mathcal{X}} (g(x) - f_\theta(x))^2 dx$, καθώς η διαδικασία για τιμές μεγαλύτερες της μονάδας γίνεται λιγότερο αποτελεσματική.

Μια πρώτη προσπάθεια των Basu et al. (1998) ήταν να δείξουν ότι η νέα οικογένεια αποκλίσεων αποτελεί έναν τύπο στατιστικής απόστασης. Για να επιτευχθεί αυτό, αρκεί να δειχθεί ότι η $d_\alpha(g, f_\theta)$ της σχέσης (2.1) είναι μη αρνητική

με την ελάχιστη τιμή να επιτυγχάνεται, όταν οι δύο πυκνότητες ταυτίζονται, $g \equiv f$. Η ιδιότητα αυτή παρουσιάζεται στο επόμενο θεώρημα, το οποίο προέρχεται από το βιβλίο των Basu et al. (2011), Theorem 9.1, σελ. 301.

Θεώρημα 2.1. Για δύο συναρτήσεις πυκνότητας f και g , η ποσότητα $d_\alpha(g, f)$ είναι μια απόκλιση, μη αρνητική για όλα τα f, g και είναι μηδέν αν και μόνο αν $f \equiv g$ σχεδόν παντού.

Απόδειξη. Η απόδειξη ακολουθεί πιστά την απόδειξη του Θεωρήματος 9.1 του βιβλίου των Basu et al. (2011), σελ. 301. Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης.

Για $\alpha \rightarrow 0$, είναι γνωστό ότι η απόκλιση που ορίζεται από τη σχέση (2.1) ταυτίζεται με την Kullback-Leibler, η οποία είναι μη αρνητική και ισούται με το μηδέν εάν και μόνο αν $g \equiv f$.

Για $\alpha \in (0, 1]$ μπορούμε εύκολα να δείξουμε ότι η απόκλιση που ορίζεται από τη σχέση (2.1) είναι μη αρνητική. Εφόσον, οι ποσότητες g και f είναι συναρτήσεις πυκνότητας πιθανότητας, ισχύει $g \geq 0$ και $f \geq 0$ σχεδόν παντού. Έστω χωρίς βλάβη της γενικότητας ότι $g > 0$.

Βγάζοντας κοινό παράγοντα την ποσότητα $g^{1+\alpha}(x)/a$ από τη σχέση (2.1) έχουμε

$$d_\alpha(g, f) = \int_{\mathbb{R}} \frac{1}{\alpha} g^{1+\alpha}(x) \left\{ \alpha \frac{f^{1+\alpha}(x)}{g^{1+\alpha}(x)} - (\alpha + 1) \frac{f^\alpha(x)}{g^\alpha(x)} + 1 \right\} dx.$$

Αν θέσουμε $y = f(x)/g(x)$, όπου $y \geq 0$, προκύπτει

$$d_\alpha(g, f) = \int_{\mathbb{R}} \frac{1}{\alpha} g^{1+\alpha}(x) \left\{ \alpha y^{1+\alpha} - (\alpha + 1) y^\alpha + 1 \right\} dx.$$

Για να δείξουμε ότι η απόκλιση $d_\alpha(g, f_\theta)$ είναι μη αρνητική, θεωρούμε τη συνάρτηση

$$I_\alpha(y) = \alpha y^{1+\alpha} - (1 + \alpha) y^\alpha + 1, \text{ με } y \geq 0$$

και αρκεί να δείξουμε ότι η $I_\alpha(y)$ είναι μη αρνητική.

Παρατηρούμε ότι για $y = 1$ έχουμε $I_\alpha(1) = \alpha - (\alpha + 1) + 1 = 0$ και είναι το μοναδικό ελάχιστο, καθώς η I_α είναι αυστηρά μη αρνητική,

Πράγματι, παραγωγίζοντας τη I_α προκύπτει η σχέση

$$\frac{d}{dy} I_\alpha(y) = \alpha(\alpha + 1) y^\alpha - (\alpha + 1) \alpha y^{\alpha-1},$$

όπου

- Για $y < 1$ ισχύει $\frac{d}{dy}I_\alpha(y) < 0 \implies I_\alpha(y)$ φθίνουσα
- Για $y = 1$ ισχύει $I_\alpha(y) = 0$
- Για $y > 1$ ισχύει $\frac{d}{dy}I_\alpha(y) > 0 \implies I_\alpha(y)$ αυξουσα

Άρα η συνάρτηση I_α είναι μη αρνητική, καθώς αρχικά είναι φθίνουσα και στη συνέχεια αύξουσα, με ελάχιστη τιμή το μηδέν.

Επιπλέον, η δεύτερη παράγωγος της $I_\alpha(y)$ είναι

$$\frac{d^2}{(dy)^2}I_\alpha(y) = \alpha(\alpha + 1) \left[y^{\alpha-2}(\alpha y - (\alpha - 1)) \right],$$

η οποία για $\alpha \in (0, 1]$ και $y \geq 0$ είναι μη αρνητική, με αποτέλεσμα στο $y = 1$ να επιτυγχάνει το μοναδικό ελάχιστό της, καθώς η $I_\alpha(y)$ είναι κυρτή συνάρτηση.

Δεδομένου ότι $I_\alpha \geq 0$ και $\frac{1}{\alpha}g^{1+\alpha}(x) > 0$ προκύπτει ότι και η απόκλιση $d_\alpha(g, f)$ είναι επίσης μη αρνητική, καθώς το ολοκλήρωμα μιας μη αρνητικής συνάρτησης είναι μη αρνητικό. Η ελάχιστη τιμή της απόκλισης, δηλαδή το μηδέν, επιτυγχάνεται αν και μόνο αν οι δύο συναρτήσεις είναι ίσες, δηλαδή όταν $f \equiv g$.

□

2.1.1 Ελάχιστη Απόκλιση Kullback-Leibler (KLD) και Εκτιμητές Μέγιστης Πιθανοφάνειας (MLE)

Όπως αναφέρθηκε πιο πάνω, ένας από τους σημαντικότερους στόχους στη Στατιστική Συμπερασματολογία είναι η εκτίμηση μιας ή περισσότερων άγνωστων παραμέτρων. Μια από τις δημοφιλέστερες και σημαντικότερες προσεγγίσεις, στην κατεύθυνση αυτή, είναι οι Εκτιμητές Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimator). Οι Εκτιμητές Μέγιστης Πιθανοφάνειας προκύπτουν μεγιστοποιώντας τη συνάρτηση πιθανοφάνειας ή τον λογάριθμό της ως προς την άγνωστη παράμετρο. Ένα πλεονέκτημα αυτών των εκτιμητών είναι ότι διαδραματίζουν σημαντικό ρόλο στην ανάπτυξη του log-likelihood ratio test. Μολονότι οι Εκτιμητές Μέγιστης Πιθανοφάνειας μπορούν εύκολα να προσδιοριστούν, υπάρχουν περιπτώσεις που τα δεδομένα περιέχουν ακραίες τιμές (outliers), με αποτέλεσμα οι εκτιμητές αυτοί να χάνουν την αποτελεσματικότητά τους. Ένα βασικό αποτέλεσμα, το οποίο θα αποτελέσει το κίνητρο, στην επόμενη ενότητα, για την ανάπτυξη μεθόδου εκτίμησης τύπου ελάχιστης απόστασης, είναι ότι οι

εκτιμητές που προκύπτουν από την ελαχιστοποίηση της Kullback-Leibler απόκλισης ταυτίζονται με τους εκτιμητές που προκύπτουν από τη μεγιστοποίηση της συνάρτησης πιθανοφάνειας.

Έστω ότι $X_1, \dots, X_n$ είναι ένα τυχαίο δείγμα της X , το οποίο περιγράφεται από την πραγματική αλλά άγνωστη πυκνότητα g . Λαμβάνοντας υπόψη ότι το πραγματικό μοντέλο g είναι άγνωστο, υποθέτουμε ότι $\{f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$ είναι μία ταυτοποιήσιμη (identifiable) παραμετρική οικογένεια κατανομών, η οποία χρησιμοποιείται για την περιγραφή των παρατηρήσεων $\{x_1, \dots, x_n\}$. Σε αυτήν την περίπτωση, η συνάρτηση πιθανοφάνειας ορίζεται να είναι η από κοινού συνάρτηση πυκνότητας πιθανότητας που βασίζεται στις περιθώριες κατανομές και έχει τη μορφή

$$L(\theta|x) = f_\theta(x) = \prod_{i=1}^n f_\theta(x_i). \quad (2.2)$$

Η $L(\theta|x)$ ονομάζεται συνάρτηση πιθανοφάνειας (likelihood function), και ο λογάριθμος της συνάρτησης πιθανοφάνειας (log-likelihood function) ορίζεται

$$\ln(L(\theta|x)) = \ln \left(\prod_{i=1}^n f_\theta(x_i) \right) = \sum_{i=1}^n \ln(f_\theta(x_i)). \quad (2.3)$$

Οι Εκτιμητές Μέγιστης Πιθανοφάνειας ορίζονται

$$\hat{\theta}_{MLE,n} = \arg \max_{\theta \in \Theta} \{L(\theta|x)\} \quad (2.4)$$

ή

$$\hat{\theta}_{MLE,n} = \arg \max_{\theta \in \Theta} \{\ln(L(\theta|x))\} \quad (2.5)$$

και μπορούν να ληφθούν και από τη λύση της εξίσωσης

$$u_\theta(x) = \frac{d}{d\theta} \{\ln(L(\theta|x))\} = 0. \quad (2.6)$$

Στη συνέχεια, θα δειχθεί ότι οι Εκτιμητές Μέγιστης Πιθανοφάνειας μπορούν να προκύψουν μέσω της ελαχιστοποίησης του μέτρου απόκλισης Kullback-Leibler, για μεγάλο μέγεθος δείγματος n .

Η Kullback-Leibler απόκλιση ανάμεσα στη συνάρτηση πυκνότητας $g(x)$ και την $f_\theta(x)$ δίνεται από τη σχέση

$$\begin{aligned} D_{Kul}(g, f_\theta) &= \int_{\mathbb{R}} g(x) \ln \left(\frac{g(x)}{f_\theta(x)} \right) dx \\ &= \int_{\mathbb{R}} g(x) \ln(g(x)) dx - \int_{\mathbb{R}} g(x) \ln(f_\theta(x)) dx. \end{aligned}$$

Το ολοκλήρωμα $\int_{\mathbb{R}} g(x) \ln(g(x)) dx$ μπορεί να αγνοηθεί, καθώς δεν εξαρτάται από την παράμετρο θ . Άρα, μπορούμε να ορίσουμε τον εκτιμητή βασισμένο στην ελαχιστοποίηση της απόκλισης Kullback-Leibler ως εξής

$$\hat{\theta}_{Kull,n} = \arg \min_{\theta \in \Theta} D_{Kull}(g, f_{\theta}), \quad (2.7)$$

$$\begin{aligned} \hat{\theta}_{Kull,n} &= \arg \min_{\theta \in \Theta} \left\{ - \int_{\mathbb{R}} g(x) \ln(f_{\theta}(x)) dx \right\}, \\ \hat{\theta}_{Kull,n} &= \arg \min_{\theta \in \Theta} \left\{ - \int_{\mathbb{R}} \ln(f_{\theta}(x)) dG(x) \right\}, \end{aligned} \quad (2.8)$$

όπου G είναι η συνάρτηση κατανομής που αντιστοιχεί στην g . Αν η G στη σχέση (2.8) που είναι άγνωστη αντικατασταθεί από την εμπειρική συνάρτηση κατανομής G_n , για μεγάλο n , έχουμε

$$\hat{\theta}_{Kull,n} = \arg \min_{\theta \in \Theta} \left\{ - \int_{\mathbb{R}} \ln(f_{\theta}(x)) dG_n(x) \right\}, \quad (2.9)$$

ή

$$\hat{\theta}_{Kull,n} = \arg \min_{\theta \in \Theta} \left\{ - \frac{1}{n} \sum_{i=1}^n \ln(f_{\theta}(x_i)) \right\}. \quad (2.10)$$

Επειδή ο όρος $1/n$ δεν επηρεάζει τη διαδικασία ελαχιστοποίησης, μπορεί να παραληφθεί και έτσι

$$\hat{\theta}_{Kull,n} = \arg \max_{\theta \in \Theta} \left\{ \sum_{i=1}^n \ln(f_{\theta}(x_i)) \right\}. \quad (2.11)$$

Επομένως, παρατηρούμε ότι ο εκτιμητής μέγιστης πιθανοφάνειας δίνεται από

$$\hat{\theta}_{MLE,n} = \arg \max_{\theta \in \Theta} \left\{ \ln(L(\theta|x)) \right\} \stackrel{\text{Large } n}{\approx} \arg \min_{\theta \in \Theta} \left\{ D_{Kull}(g, f_{\theta}) \right\} = \hat{\theta}_{Kull,n}. \quad (2.12)$$

Η ανάλυση αυτή, που διατυπώνεται και στους Castilla et al. (2018), επιβεβαιώνει ότι ο εκτιμητής μέγιστης πιθανοφάνειας είναι μαθηματικά ισοδύναμος με τον εκτιμητή που προκύπτει από την ελαχιστοποίηση της απόκλισης Kullback-Leibler, όταν το δείγμα είναι αρκετά μεγάλο, $n \rightarrow \infty$.

2.2 Εκτιμητής Ελάχιστης Density Power Divergence

Η μέθοδος εκτίμησης τύπου ελάχιστης απόστασης εκτιμά την παράμετρο που μας ενδιαφέρει, ελαχιστοποιώντας την απόκλιση μεταξύ του θεωρητικού αλλά άγνωστου μοντέλου και ενός παραμετρικού μοντέλου που χρησιμοποιείται για να περιγράψει τα δεδομένα και σκιαγραφήθηκε πιο πάνω μέσω της απόκλισης Kullback-Leibler. Μεταξύ των πρώτων εργασιών που ανέπτυξαν τη μέθοδο εκτίμησης τύπου ελάχιστης απόστασης ήταν η εργασία του Beran (1977). Ο Beran χρησιμοποίησε την Hellinger απόσταση και βασιζόμενος στην πυκνότητα, ανέπτυξε μια μέθοδο παραγωγής εκτιμητών για συνεχή μοντέλα με καλές ιδιότητες ευρωστίας (robustness). Με κατάλληλες απλοποιήσεις, τα αποτελέσματα είναι αποδεκτά και για διακριτά μοντέλα. Ωστόσο, τα αποτελέσματα αυτά συχνά χρειάζονται χρήση μη παραμετρικών μεθόδων, όπως η εκτίμηση πυκνότητας με χρήση πυρήνων (kernel density estimation). Στο βιβλίο των Basu et al. (2011), σελ. 74, αναφέρεται ότι η εργασία του Beran (1977) αποτέλεσε κίνητρο και για πολλούς συγγραφείς, όπως οι Tamura and Boos (1986), Simpson (1987), Basu and Lindsay (1994), Cao and Fraiman (1995) and Toma (2008). Επίσης, οι Wu and Karunamuni (2009) και οι Karunamuni and Wu (2009) επέκτειναν τη εργασία του Beran (1977) σε ημιπαραμετρικά μοντέλα.

Στο πλαίσιο λοιπόν της εκτίμησης παραμέτρων, στην ενότητα που ακολουθεί θα αναφερθούμε στην ανάπτυξη εκτιμητών ελάχιστης Density Power Divergence. Το βασικό πλεονέκτημα της συγκεκριμένης μεθόδου κατασκευής εκτιμητών έγκειται στο γεγονός ότι δεν απαιτεί μη παραμετρική εξομάλυνση για οποιαδήποτε τιμή του α και μπορεί να χρησιμοποιηθεί τόσο σε διακριτά όσο και σε συνεχή μοντέλα. Η ιδέα παρουσιάζεται στη δημοσίευση των Basu et al. (1998), οι οποίοι εισήγαγαν την οικογένεια πυκνοτήτων Density Power Divergence. Στην υποενότητα που ακολουθεί θα περιγραφεί η ιδέα των Basu et al. (1998) για την ανάπτυξη μεθοδολογίας εύρεσης εκτιμητών μέσω αυτής της οικογένειας αποκλίσεων.

2.2.1 Εκτιμητές τύπου Ελάχιστης Απόστασης μέσω της Density Power Divergence

Η ιδέα της εκτίμησης που παρουσιάστηκε από τους Basu et al. (1998), ξεκινά από το γεγονός ότι οι εκτιμητές μέγιστης πιθανοφάνειας, συμπίπτουν με τον εκτιμητή της παραμέτρου θ που ελαχιστοποιεί την Kullback-Leibler απόκλιση μεταξύ του πραγματικού αλλά άγνωστου μοντέλου και της παραμετρικής

οικογένειας κατανομών που υιοθετείται για να μοντελοποιήσει τα δεδομένα. Με βάση αυτήν την ιδιότητα, οι Basu et al. (1998) ορίζουν μια οικογένεια αποκλίσεων που περιέχει, ως ειδική περίπτωση, την Kullback-Leibler και επιπλέον, με την ελαχιστοποίηση της Density Power Divergence να προκύπτουν εκτιμητές που χαρακτηρίζονται από αποτελεσματικότητα ακόμη και στην περίπτωση ακραίων παρατηρήσεων.

Η διαδικασία εκτίμησης που θα παρουσιαστεί αφορά τιμές της παραμέτρου α στο διάστημα $(0, 1]$, καθώς εκεί αρκετοί από τους εκτιμητές έχουν ισχυρές ιδιότητες ευρωστίας (robustness) με μικρές απώλειες ασυμπτωτικής αποδοτικότητας (asymptotic efficiency), συγκριτικά με τη μέθοδο μέγιστης πιθανοφάνειας. Από την άλλη, τιμές μεγαλύτερες της μονάδας καθιστούν τη μέθοδο εξαιρετικά αναποτελεσματική. Βάσει αυτών οι Basu et al. (1998) παρουσιάζουν μια μέθοδο εκτίμησης τύπου ελάχιστης απόστασης.

Ας θεωρήσουμε λοιπόν μία παραμετρική οικογένεια κατανομών $\{f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$, όπου μας ενδιαφέρει η εκτίμηση της θ δεδομένου ενός τυχαίου δείγματος $X_1, \dots, X_n$, από την κατανομή G , η οποία είναι η πραγματική αλλά άγνωστη κατανομή με αντίστοιχη συνάρτηση πυκνότητας g . Η παράμετρος θ εκτιμάται ελαχιστοποιώντας το μέτρο απόκλισης (Density Power Divergence) στον παραμετρικό χώρο Θ . Ο εκτιμητής αυτός είναι γνωστός ως Minimum Density Power Divergence Estimator (MDPDE) και προκύπτει από τη σχέση

$$\hat{\theta}_n^\alpha = \arg \min_{\theta \in \Theta} d_\alpha(g, f_\theta). \quad (2.13)$$

Είναι Fisher-συνεπής εκτιμητής, βλέπε Rao (1965), δηλαδή η μέθοδος εκτίμησης για μεγάλο δείγμα θα δίνει την αληθινή τιμή της παραμέτρου.

Στη συνέχεια θα περιγραφεί η διαδικασία κατασκευής των MDPDE εκτιμητών που ορίζονται στη σχέση (2.13).

Γνωρίζουμε από τη σχέση (2.1) ότι η Density Power Divergence είναι

$$\begin{aligned} d_\alpha(g, f_\theta) &= \int_{\mathbb{R}} \left\{ f_\theta^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f_\theta^\alpha(x) + \frac{1}{\alpha} g^{1+\alpha}(x) \right\} dx \\ &= \int_{\mathbb{R}} f_\theta^{1+\alpha}(x) dx - \int_{\mathbb{R}} \left(1 + \frac{1}{\alpha}\right) g(x) f_\theta^\alpha(x) dx + \int_{\mathbb{R}} \frac{1}{\alpha} g^{1+\alpha}(x) dx \end{aligned}$$

Το τελευταίο ολοκλήρωμα $\int_{\mathbb{R}} \frac{1}{\alpha} g^{1+\alpha}(x) dx$ είναι ανεξάρτητο της παραμέτρου θ . Εάν το αγνοήσουμε και αντικαταστήσουμε την άγνωστη κατανομή G με την εμπειρική συνάρτηση κατανομής G_n , ο εκτιμητής MDPDE προκύπτει για μεγάλο

μέγεθος δείγματος n να είναι

$$\begin{aligned}\hat{\theta}_n^\alpha &= \arg \min_{\theta \in \Theta} d_\alpha(g, f_\theta) = \\ &= \arg \min_{\theta \in \Theta} \left\{ \int_{\mathbb{R}} f_\theta^{1+\alpha}(x) dx - \int_{\mathbb{R}} \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(x) dG_n(x) \right\} \\ &= \arg \min_{\theta \in \Theta} \left\{ \int_{\mathbb{R}} f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i) \right\}.\end{aligned}$$

Θεωρούμε τώρα την αντικειμενική συνάρτηση

$$H_n^\alpha(\theta) = \int_{\mathbb{R}} f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i). \quad (2.14)$$

Βάσει αυτής, ο εκτιμητής MDPDE μπορεί να ληφθεί από την ελαχιστοποίηση ως προς θ της αντικειμενικής συνάρτησης

$$\theta_n^\alpha = \arg \min_{\theta \in \Theta} H_n^\alpha(\theta).$$

Φαίνεται λοιπόν ότι μπορούμε να ελαχιστοποιήσουμε την DPD ως προς την παράμετρο θ , μέσω της G_n που είναι η εμπειρική συνάρτηση κατανομής.

Γενικά, όπως θα δούμε στην επόμενη ενότητα, υπό την υπόθεση ότι το μοντέλο πληροί κατάλληλες συνθήκες διαφορισιμότητας και κανονικότητας (regularity assumptions), ο εκτιμητής MDPDE που ορίζεται από την ελαχιστοποίηση της αντικειμενικής συνάρτησης της (2.14), μπορεί να υπολογιστεί λύνοντας την εξίσωση εκτίμησης

$$\int_{\mathbb{R}} u_\theta(x) f_\theta^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_\theta(X_i) f_\theta^\alpha(X_i) = 0, \quad (2.15)$$

που είναι η παράγωγος της σχέσης (2.14), όπου $u_\theta(x) = \nabla \ln f_\theta(x)$ είναι η συνάρτηση σκορ (score function).

Αναφέρθηκε ότι για $\alpha \rightarrow 0$ η Density Power Divergence ταυτίζεται με την απόκλιση Kullback-Leibler, δηλαδή ισχύει ότι:

$$d_0(g, f_\theta) = \lim_{\alpha \rightarrow 0} d_\alpha(g, f) = \int_{\mathbb{R}} g(x) \ln \left(\frac{g(x)}{f_\theta(x)} \right) dx.$$

Δεδομένης μιας οικογένειας πυκνοτήτων $\{f_\theta\}$ αποδείχθηκε προηγουμένως ότι η $d_0(g, f_\theta)$ ελαχιστοποιείται στον χώρο των παραμέτρων από τον Εκτιμητή Μέγιστης Πιθανοφάνειας.

Στη συνέχεια διατυπώνουμε ένα παράδειγμα, το οποίο δείχνει ότι για τιμές της παραμέτρου α της Density Power Divergence κοντά στο μηδέν, ο εκτιμητής MDPDE ταυτίζεται με τον Εκτιμητή Μέγιστης Πιθανοφάνειας.

Παράδειγμα 2.1. Έστω ότι $X_1, \dots, X_n$ είναι ένα τυχαίο δείγμα από την $Ek\theta(\theta)$, με άγνωστη παράμετρο $\theta > 0$. Έστω $f_\theta(x)$ παριστά τη συνάρτηση πυκνότητας της $Ek\theta(\theta)$. Ο εκτιμητής MDPDE προκύπτει από τη λύση της εξίσωσης (2.15), δηλαδή από την εξίσωση

$$\int_{\mathbb{R}} u_\theta(x) f_\theta^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_\theta(X_i) f_\theta^\alpha(X_i) = 0,$$

που για $\alpha \rightarrow 0$ γίνεται

$$\int_{\mathbb{R}} u_\theta(x) f_\theta(x) dx - \frac{1}{n} \sum_{i=1}^n u_\theta(X_i) = 0.$$

Λαμβάνοντας υπόψη ότι $u_\theta(x) = \frac{d}{d\theta} \ln f_\theta(x)$ και $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$

$$\begin{aligned} 0 &= \int \frac{d}{d\theta} \ln f_\theta(x) f_\theta(x) dx - \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \ln f_\theta(X_i) \\ &= \int_0^\infty \left(\frac{d}{d\theta} \ln(\theta e^{-\theta x}) \right) \theta e^{-\theta x} dx - \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \ln(\theta e^{-\theta X_i}) \\ &= \int_0^\infty \left(\frac{1}{\theta} - x \right) \theta e^{-\theta x} dx - \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\theta} - X_i \right) \\ &= \int_0^\infty \left(\frac{1}{\theta} - x \right) f_\theta(x) dx - \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\theta} - X_i \right) \\ &= \frac{1}{\theta} - E(X) - \frac{1}{\theta} + \frac{\sum_{i=1}^n X_i}{n} \\ &= \frac{1}{\theta} - \frac{1}{\theta} - \frac{1}{\theta} + \frac{\sum_{i=1}^n X_i}{n} \end{aligned}$$

και επομένως

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n X_i} \quad \text{ή} \quad \hat{\theta} = \frac{1}{\bar{X}}.$$

Παρατηρούμε λοιπόν ότι ο εκτιμητής $\hat{\theta} = \frac{1}{\bar{X}}$ που προκύπτει από τη λύση της σχέσης (2.15), για τιμή του $\alpha \rightarrow 0$ ταυτίζεται με τον αντίστοιχο Εκτιμητή Μέγιστης Πιθανοφάνειας.

2.3 Ασυμπτωτική κατανομή των MDPDE

Είναι λογικό, μετά τον ορισμό των εκτιμητών, να διερευνηθούν οι ιδιότητές τους. Σε αυτήν την ενότητα, θα μελετηθεί η συνέπεια και η ασυμπτωτική κατανομή του εκτιμητή MDPDE που προκύπτει από την ελαχιστοποίηση της Density Power Divergence.

Ας υποθέσουμε ότι η άγνωστη παράμετρος θ είναι p -διάστατη και έστω ότι το παραμετρικό μοντέλο που μοντελοποιεί τα δεδομένα είναι $\{f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$. Ο εκτιμητής τύπου ελάχιστης απόστασης, μέσω της Density Power Divergence, προκύπτει από τη λύση της εξίσωσης (2.15) μέσω ελαχιστοποίησης της αντικειμενικής συνάρτησης (2.14).

Έστω $i_\theta(x) = \nabla(u_\theta(x))$, όπου $u_\theta(x) = \nabla \ln f_\theta(x)$, η συνάρτηση πληροφορίας του μοντέλου. Αυτή η συνάρτηση σχετίζεται με τον πίνακα πληροφορίας Fisher, ο οποίος, υπό κατάλληλες συνθήκες κανονικότητας, ορίζεται από την αρνητική αναμενόμενη τιμή αυτής της ποσότητας και εκφράζει τη συνολική πληροφορία που περιέχεται στα δεδομένα για τις άγνωστες παραμέτρους του μοντέλου. Επίσης, έστω οι πίνακες $J = J(\theta_0)$ και $K = K(\theta_0)$ που ορίζονται

$$J(\theta_0) = \int u_{\theta_0}(x) u_{\theta_0}(x)^t f_{\theta_0}^{1+\alpha}(x) dx + \int \{i_{\theta_0}(x) - \alpha u_{\theta_0}(x) u_{\theta_0}(x)^t\} \{g(x) - f_{\theta_0}(x)\} f_{\theta_0}^\alpha(x) dx \quad (2.16)$$

και

$$K(\theta_0) = \int u_{\theta_0}(x) u_{\theta_0}^t(x) f_{\theta_0}^{2\alpha}(x) g(x) dx - \xi \xi^t, \quad (2.17)$$

με $\xi = \int u_{\theta_0}(x) f_{\theta_0}^\alpha(x) g(x) dx$.

Στη συνέχεια θεωρούμε ότι ικανοποιούνται οι ακόλουθες συνθήκες κανονικότητας, όπως αυτές διατυπώνονται στη μονογραφία των Basu et al. (2011), σελίδα 304.

(Σ_1): Οι κατανομές f_θ έχουν κοινό σύνολο τιμών $\mathcal{X}$, τέτοιο ώστε το $\mathcal{X} = \{x | f_\theta(x) > 0\}$ να είναι ανεξάρτητο του θ . Η άγνωστη κατανομή G έχει επίσης κοινό σύνολο το $\mathcal{X}$, στο οποίο η αντίστοιχη πυκνότητα g είναι μεγαλύτερη του μηδενός.

(Σ_2): Υπάρχει ένα ανοιχτό υποσύνολο Ω του παραμετρικού χώρου Θ , το οποίο περιέχει την παράμετρο θ_0 , που είναι η αληθής αλλά άγνωστη τιμή της θ , τέτοια ώστε για σχεδόν όλα τα $x \in \mathcal{X}$ και όλα τα $\theta \in \Omega$ η πυκνότητα $f_\theta(x)$ να είναι

τρεις φορές παραγωγίσιμη ως προς θ και οι τρίτης τάξης μερικοί παράγωγοι να είναι συνεχείς ως προς θ .

(Σ_3): Τα ολοκληρώματα $\int f_\theta^{a+1}(x)dx$ και $\int f_\theta^a(x)g(x)dx$ μπορούν να παραγωγισθούν τρεις φορές ως προς θ και οι παράγωγοι μπορούν να υπολογιστούν εντός του ολοκληρώματος.

(Σ_4): Ο πίνακας $J(\theta)_{p \times p}$, με στοιχεία:

$$J_{kl}(\theta) = \frac{1}{1+\alpha} \left(E_g \left[\nabla_{kl} \left(\int f_\theta^{1+\alpha}(x)dx - \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(X) \right) \right] \right) \quad (2.18)$$

είναι θετικά ορισμένος, όπου E_g συμβολίζει την αναμενόμενη τιμή υπό την g .

(Σ_5): Υπάρχει μια συνάρτηση $M_{jkl}(x)$ τέτοια ώστε

$$\left| \nabla_{jkl} \left(\int f_\theta^{1+\alpha}(x)dx + \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(x)dx \right) \right| \leq M_{jkl}(x) \quad (2.19)$$

για όλα τα $\theta \in \Omega$ όπου $E_g[M_{jkl}(X)] = m_{jkl} < \infty, \forall j, k, l$.

Το επόμενο θεώρημα πηγάζει από τους Basu et al. (2011), Theorem 9.2, σελ. 304, μεταφέρεται πιστά από εκεί και αφορά τη συνέπεια και την ασυμπτωτική κανονικότητα του MDPDE.

Θεώρημα 2.2. Έστω ότι ισχύουν οι συνθήκες κανονικότητας (Σ_1) – (Σ_5). Τότε

(i) Ο εκτιμητής ελάχιστης απόστασης MDPDE $\hat{\theta}_n^\alpha = (\theta_{1n}, \dots, \theta_{pn})^t$ είναι συνεπής εκτιμητής του $\theta_0 = (\theta_{01}, \dots, \theta_{0p})^t$, δηλαδή, για κάθε $\zeta > 0$, ισχύει

$$P(\|\theta_n - \theta_0\| < \zeta) \xrightarrow{n \rightarrow \infty} 1,$$

όπου $\|\cdot\|$ είναι μία νόρμα στον χώρο Θ .

(ii) Το τυχαίο διάνυσμα $\sqrt{n}(\hat{\theta}_n^\alpha - \theta_0)$ έχει ασυμπτωτική πολυδιάστατη κανονική κατανομή, με μέση τιμή το μηδενικό διάνυσμα και πίνακα διακυμάνσεων-συνδιακυμάνσεων $J^{-1}KJ^{-1}$, όπου J και K ορίζονται από τις σχέσεις (2.16) και (2.17) αντίστοιχα.

Η απόδειξη ακολουθεί πιστά την απόδειξη του Θεωρήματος 9.2 του βιβλίου των Basu et al. (2011), σελ. 304 και τις αποδείξεις των Theorem 1 και Theorem 2 των Leroy et al. (2016). Μεταφέρεται εδώ, λεπτομερώς παρουσιασμένη, χάριν πληρότητας της παρουσίασης και βρίσκεται στο Κεφάλαιο 5 (Παράρτημα).

ΚΕΦΑΛΑΙΟ 3

ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ ΜΕΣΩ ΤΗΣ DENSITY POWER DIVERGENCE

Τα τελευταία χρόνια, η Στατιστική Θεωρία Πληροφοριών αποτελεί θεμελιώδες εργαλείο για την ανάπτυξη μεθοδολογιών στη Μαθηματική Στατιστική και ειδικότερα στους ελέγχους Στατιστικών Υποθέσεων, οι οποίοι θα αποτελέσουν κύριο αντικείμενο αυτού του κεφαλαίου (βλέπε Αυλογιάρης (2017)). Οι έλεγχοι υποθέσεων στα μέτρα απόκλισης έκαναν την εμφάνισή τους, μεταξύ άλλων, στις εργασίες των Cressie and Read (1984) και Zografos et al. (1990), με τη χρήση των λ και φ -αποκλίσεων, αντίστοιχα, κατασκευάζοντας στατιστικά τεστ για πολυωνυμικά δεδομένα. Επιπλέον, στη μονογραφία του Pardo (2006) παρατίθενται αναλυτικά αποτελέσματα για τους ελέγχους στατιστικών υποθέσεων, που βασίζονται στην οικογένεια των φ -αποκλίσεων.

Το Κεφάλαιο 3 της παρούσας διατριβής, επικεντρώνεται στην κατασκευή και μελέτη ενός παραμετρικού στατιστικού τεστ (one sample problem). Αναλυτικότερα, ορίζεται η στατιστική συνάρτηση του τεστ και προσδιορίζεται η ασυμπτωτική κατανομή της για τον έλεγχο της απλής μηδενικής υπόθεσης, χρησιμοποιώντας την οικογένεια Density Power Divergence, η οποία αναφέρθηκε λεπτομερώς στο προηγούμενο Κεφάλαιο. Η εργασία των Basu et al. (2013) αποτελεί μία από τις πρώτες προσεγγίσεις στην ανάπτυξη ελέγχων υποθέσεων που βασίζονται στην Density Power Divergence.

Συνεπώς, στο κεφάλαιο αυτό θα πραγματοποιηθεί μία ανασκόπηση του ελέγχου της απλής μηδενικής υπόθεσης, χρησιμοποιώντας την οικογένεια αποκλίσεων (DPD).

3.1 Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

Στην ενότητα αυτή, θα βασιστούμε στην εργασία των Basu et al. (2013), η οποία αναφέρεται στον έλεγχο παραμετρικών υποθέσεων μέσω της Density Power Divergence. Έστω το τυχαίο δείγμα $X_1, \dots, X_n$, μεγέθους n , από την ταυτοποιήσιμη (identifiable) παραμετρική οικογένεια κατανομών $\{f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$. Ορίζουμε ως θ_0 μία γνωστή τιμή της παραμέτρου και το ενδιαφέρον μας να επικεντρωθεί στον έλεγχο $H_0 : \theta = \theta_0$ έναντι $H_a : \theta \neq \theta_0$, με το $\theta_0 \in \Theta$ γνωστό. Η Density Power Divergence, μεταξύ των f_θ και f_{θ_0} κατανομών, ορίζεται ως εξής

$$d_\alpha(f_\theta, f_{\theta_0}) = \int \left\{ f_{\theta_0}^{\alpha+1}(x) - \left(1 + \frac{1}{\alpha}\right) f_{\theta_0}^\alpha(x) f_\theta(x) + \frac{1}{\alpha} f_\theta^{\alpha+1}(x) \right\} dx. \quad (3.1)$$

Κύρια ιδιότητα της απόκλισης αυτής είναι η μη αρνητικότητα της, $d_\alpha(f_\theta, f_{\theta_0}) \geq 0$, με την ισότητα να ισχύει αν και μόνο αν $f_\theta = f_{\theta_0}$. Η απόκλιση $d_\alpha(f_\theta, f_{\theta_0})$ εξαρτάται από μια άγνωστη παράμετρο θ , και συνεπώς, για τον έλεγχο της μηδενικής υπόθεσης, είναι λογικό να βασιστούμε στην εκτιμώμενη απόκλιση $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$, που ορίζεται από τη σχέση

$$\begin{cases} d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) = \int \{ f_{\theta_0}^{\beta+1}(x) - (1 + \frac{1}{\beta}) f_{\theta_0}^\beta(x) f_{\hat{\theta}_\alpha}(x) + \frac{1}{\beta} f_{\hat{\theta}_\alpha}^{\beta+1}(x) \} dx, & \alpha > 0, \\ d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) = \int f_{\hat{\theta}_\alpha}(x) \ln \left(\frac{f_{\hat{\theta}_\alpha}(x)}{f_{\theta_0}(x)} \right) dx, & \alpha \rightarrow 0, \end{cases} \quad (3.2)$$

όπου $\hat{\theta}_\alpha$ είναι ο εκτιμητής τύπου ελάχιστης απόκλισης με βάση την Density Power Divergence (MDPDE) της παραμέτρου $\theta \in \Theta$. Επιπρόσθετα, παρατηρείται ότι η εκτιμώμενη απόκλιση $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$ εξαρτάται από δύο διαφορετικές παραμέτρους, τις ονομάσουμε, ρύθμισης, την α και την β . Η παράμετρος α είναι αυτή που σχετίζεται με τη διαδικασία εκτίμησης της άγνωστης παραμέτρου θ με βάση την DPD, ενώ η β είναι αυτή που χρησιμοποιείται για την κατασκευή του στατιστικού ελέγχου, προκειμένου να ελεγχθεί η υπόθεση που μας ενδιαφέρει.

Ακόμη, είναι προφανές ότι μικρές τιμές της εκτιμώμενης απόκλισης $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$ υποστηρίζουν την αποδοχή της μηδενικής υπόθεσης, σε αντίθεση με μεγάλες τιμές της εκτιμώμενης απόκλισης $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$ που σηματοδοτούν την απόρριψη της μηδενικής υπόθεσης. Όταν η μηδενική υπόθεση δεν μπορεί να απορριφθεί, τότε η f_{θ_0} θα είναι η συνάρτηση πυκνότητας που διέπει, που κυβερνά τα δεδομένα.

Για να καθοριστούν οι τιμές εκείνες που οδηγούν στην απόρριψη της μηδενικής υπόθεσης, δηλαδή η κρίσιμη περιοχή, θα χρειαστεί να προσδιοριστεί η

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

ασυμπτωτική κατανομή της $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$ υπό τη μηδενική υπόθεση. Για λόγους πληρότητας, για να προσδιοριστεί αυτή, υποθέτουμε ότι ικανοποιούνται οι συνθήκες κανονικότητας, όπως αναφέρθηκαν στο Κεφάλαιο 2 της παρούσας διατριβής.

Το επόμενο Θεώρημα προέρχεται από την εργασία των Basu et al. (2013), Theorem 1, βλέπε σελ. 325, μεταφέρεται πιστά από εκεί και αφορά τον προσδιορισμό της ασυμπτωτικής κατανομής της απόκλισης $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$, που ορίστηκε στη σχέση (3.2), υπό την μηδενική υπόθεση, όπως επίσης και την ασυμπτωτική κατανομή του στατιστικού

$$T_\beta(\hat{\theta}_\alpha, \theta_0) = 2nd_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}), \quad (3.3)$$

υπό την μηδενική υπόθεση, όπου $\hat{\theta}_\alpha$ ο εκτιμητής MDPDE της θ .

Θεώρημα 3.1. Έστω ότι ισχύουν οι συνθήκες κανονικότητας $(\Sigma_1) - (\Sigma_5)$. Τότε, υπό τη μηδενική υπόθεση $H_0 : \theta = \theta_0$, η ασυμπτωτική κατανομή της $T_\beta(\hat{\theta}_\alpha, \theta_0)$ συμπίπτει με την κατανομή του

$$\sum_{i=1}^r \lambda_i^{\alpha, \beta}(\theta_0) Z_i^2, \quad (3.4)$$

όπου Z_i είναι ανεξάρτητες τυχαίες μεταβλητές με τυπική κανονική κατανομή, με Z_i^2 να έχουν χ_1^2 κατανομή και $\lambda_i^{\alpha, \beta}$ να είναι οι μη μηδενικές ιδιοτιμές του πίνακα $A_\beta(\theta_0)J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0)$, όπου $A_\beta(\theta_0) = \nabla_\theta^2 \left(d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) \right)_{\theta=\theta_0}$, για $\theta = \hat{\theta}_\alpha$ και

$$\begin{aligned} r &= \text{rank} \left(J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0)A_\beta(\theta_0)J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0) \right) \\ &= \text{rank} \left(K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0)A_\beta(\theta_0)J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0) \right), \end{aligned}$$

Απόδειξη. Η απόδειξη ακολουθεί πιστά την απόδειξη του Theorem 1 στην εργασία των Basu et al. (2013), σελ. 325. Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης.

Θεωρούμε το ανάπτυγμα Taylor δεύτερης τάξης, της απόκλισης $d_\beta(f_\theta, f_{\theta_0})$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

γύρω από το θ_0 ως συνάρτηση του $\hat{\theta}_\alpha$, το οποίο δίνεται από τη σχέση

$$\begin{aligned} d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) &= d_\beta(f_{\theta_0}, f_{\theta_0}) + \nabla_\theta \left(d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0} (\hat{\theta}_\alpha - \theta_0) + \\ &+ \frac{1}{2} (\hat{\theta}_\alpha - \theta_0)^t \nabla_\theta^2 \left(d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0} (\hat{\theta}_\alpha - \theta_0) + o(\|\theta - \theta_0\|^2). \end{aligned} \quad (3.5)$$

Σύμφωνα με την ιδιότητα των αποκλίσεων, η οποία αναφέρεται στο Θεώρημα 2.1 αυτής της διατριβής, προκύπτει ότι η απόκλιση $d_\beta(f_{\theta_0}, f_{\theta_0}) = 0$.

Επιπλέον,

$$\begin{aligned} &\left(\frac{\partial}{\partial \theta_j} d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0} = \\ &= \left(\frac{\partial}{\partial \theta_j} \left\{ \int f_{\theta_0}^{\beta+1}(x) + \left(1 + \frac{1}{\beta} \right) f_{\theta_0}^\beta(x) f_\theta(x) + \frac{1}{\beta} f_\theta^{\beta+1}(x) \right\} dx \right)_{\theta=\theta_0} \\ &\stackrel{\Sigma_3}{=} \left(\int \left\{ \frac{\partial}{\partial \theta_j} f_{\theta_0}^{\beta+1}(x) - \left(1 + \frac{1}{\beta} \right) \frac{\partial}{\partial \theta_j} f_{\theta_0}^\beta(x) f_\theta(x) + \frac{1}{\beta} \frac{\partial}{\partial \theta_j} f_\theta^{\beta+1}(x) \right\} dx \right)_{\theta=\theta_0} \\ &= \left(\int \left\{ 0 - \left(1 + \frac{1}{\beta} \right) f_{\theta_0}^\beta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) + \frac{1}{\beta} \frac{\partial}{\partial \theta_j} f_\theta^{\beta+1}(x) \right\} dx \right)_{\theta=\theta_0} \\ &= \left(\int \left\{ - \left(\frac{\beta+1}{\beta} \right) f_{\theta_0}^\beta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) + \frac{1}{\beta} (\beta+1) f_\theta^\beta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) \right\} dx \right)_{\theta=\theta_0} \\ &= 0. \end{aligned}$$

Συνεπώς,

$$\nabla_\theta \left(d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0} = 0. \quad (3.6)$$

Ακόμη, έχουμε ότι

$$\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} d_\beta(f_\theta, f_{\theta_0}) =$$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

$$\begin{aligned}
&= \frac{\partial}{\partial \theta_k} \int \left\{ -\left(\frac{\beta+1}{\beta}\right) f_{\theta_0}^\beta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) + \frac{\beta+1}{\beta} f_\theta^\beta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) \right\} dx = \\
&\stackrel{\Sigma_3}{=} \int -\left(\frac{\beta+1}{\beta}\right) f_{\theta_0}^\beta(x) \frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} f_\theta(x) + \frac{\beta+1}{\beta} \beta f_\theta^{\beta-1}(x) \frac{\partial}{\partial \theta_k} f_\theta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) \\
&\quad + \frac{\beta+1}{\beta} f_\theta^\beta(x) \frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} f_\theta(x) \} dx,
\end{aligned}$$

και για $\theta = \theta_0$ προκύπτει

$$\left(\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0} = \int (\beta+1) f_{\theta_0}^{\beta-1}(x) \frac{\partial}{\partial \theta_k} f_\theta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) dx.$$

Ορίζουμε ότι

$$\alpha_{jk}^\beta(\theta_0) = \left(\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0},$$

οπότε ισχύει

$$\alpha_{jk}^\beta(\theta_0) = (\beta+1) \int f_{\theta_0}^{\beta-1}(x) \left(\frac{\partial}{\partial \theta_k} f_\theta(x) \frac{\partial}{\partial \theta_j} f_\theta(x) \right)_{\theta=\theta_0} dx,$$

όπου $\alpha_{jk}^\beta(\theta_0)$ είναι στοιχεία του πίνακα $A_\beta(\theta_0) = \left(\alpha_{jk}^\beta(\theta_0) \right)_{p \times p}$.

Επομένως, προκύπτει ότι

$$A_\beta(\theta_0) = \nabla_\theta^2 \left(d_\beta(f_\theta, f_{\theta_0}) \right)_{\theta=\theta_0}. \quad (3.7)$$

Χρησιμοποιώντας τη σχέση (3.5) και λαμβάνοντας υπόψιν τις σχέσεις (3.6) και (3.7) προκύπτει

$$d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) = \frac{1}{2} (\hat{\theta}_\alpha - \theta_0)^t A_\beta(\theta_0) (\hat{\theta}_\alpha - \theta_0) + o\left(\|\hat{\theta} - \theta_0\|^2\right), \quad (3.8)$$

ή

$$2nd_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) = \sqrt{n}(\hat{\theta}_\alpha - \theta_0)^t A_\beta(\theta_0) \sqrt{n}(\hat{\theta}_\alpha - \theta_0) + o\left(\|\hat{\theta} - \theta_0\|^2\right). \quad (3.9)$$

Επίσης, $o\left(\|\hat{\theta} - \theta_0\|^2\right) = o_p(n^{-1})$, βλέπε Basu et al. (2013), σελ. 325, υπό την μηδενική υπόθεση H_0 .

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

Συνεπώς,

$$T_\beta(\hat{\theta}_\alpha, \theta_0) = 2nd_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) = \sqrt{n}(\hat{\theta}_\alpha - \theta_0)^t A_\beta(\theta_0) \sqrt{n}(\hat{\theta}_\alpha - \theta_0) + o_p(1).$$

Επίσης, με βάση το Θεώρημα 2.2 σχετικό με την ασυμπτωτική κανονικότητα του MDPDE $\hat{\theta}_\alpha$, υπό την μηδενική υπόθεση H_0 , ισχύει ότι

$$\sqrt{n}(\hat{\theta}_\alpha - \theta_0) \xrightarrow{L} N(0, J_\beta(\theta_0)^{-1} K_\beta(\theta_0) J_\beta(\theta_0)^{-1}).$$

Έτσι, η κατανομή της στατιστικής συνάρτησης $T_\beta(\hat{\theta}_\alpha, \theta_0)$ είναι η κατανομή της τετραγωνικής μορφής του ασυμπτωτικά κανονικού τυχαίου διανύσματος $\sqrt{n}(\hat{\theta}_\alpha - \theta_0)$. Η κατανομή αυτή, με βάση το Πόρισμα 2.1 των Dik and de Gunst (1985) ταυτίζεται με την κατανομή του $\sum_{j=1}^r \lambda_j^{\alpha, \beta} Z_j$, όπου $\lambda_1^{\alpha, \beta}, \dots, \lambda_r^{\alpha, \beta}$ οι μη μηδενικές ιδιοτιμές του πίνακα $A_\beta(\theta_0) J_\alpha^{-1}(\theta_0) K_\alpha(\theta_0) J_\alpha^{-1}(\theta_0)$. Επιπλέον, το πλήθος των μη μηδενικών ιδιοτιμών δίνεται από την εξίσωση

$$\begin{aligned} r &= \text{rank} \left(J_\alpha^{-1}(\theta_0) K_\alpha(\theta_0) J_\alpha^{-1}(\theta_0) A_\beta(\theta_0) J_\alpha^{-1}(\theta_0) K_\alpha(\theta_0) J_\alpha^{-1}(\theta_0) \right) \\ &= \text{rank} \left(K_\alpha(\theta_0) J_\alpha^{-1}(\theta_0) A_\beta(\theta_0) J_\alpha^{-1}(\theta_0) K_\alpha(\theta_0) \right), \end{aligned}$$

η οποία ισχύει από το Πόρισμα 8.3.3 του συγράμματος του Harville (2008). $\square$

Για λόγους πληρότητας διατυπώνεται το Πόρισμα 2.1 των Dik and de Gunst (1985), στην επόμενη παρατήρηση.

Παρατήρηση 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 , δηλαδή $X \sim N_q(0, \Sigma)$. Έστω A ένας πραγματικός συμμετρικός πίνακας τάξης q . Έστω επίσης $r = \text{rank}(\Sigma A \Sigma)$ και $\lambda_1, \dots, \lambda_r$ οι μη αρνητικές ιδιοτιμές του πίνακα $A \Sigma$. Τότε, η κατανομή της τετραγωνικής μορφής $X^t A X$ ταυτίζεται με την κατανομή του $\sum_{j=1}^r \lambda_j Z_j^2$ όπου $Z_1, \dots, Z_r$ ανεξάρτητες τυπικές κανονικές κατανομές.

Μετά τον προσδιορισμό της ασυμπτωτικής κατανομής της στατιστικής συνάρτησης $T_\beta(\hat{\theta}_\alpha, \theta_0) = 2nd_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$, υπό τη μηδενική υπόθεση H_0 , ακολουθεί ο προσδιορισμός της κρίσιμης περιοχής. Η μηδενική υπόθεση $H_0 : \theta = \theta_0$ έναντι της εναλλακτικής $H_a : \theta \neq \theta_0$ απορρίπτεται, σε επίπεδο σημαντικότητας α

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

($0 < a < 1$), όταν $T_\beta(\hat{\theta}_\alpha, \theta_0) > c$, όπου c το κρίσιμο σημείο που προσδιορίζεται έτσι ώστε ο έλεγχος να έχει επίπεδο σημαντικότητας a . Η περιοχή απόρριψης καθορίζεται από την πιθανότητα σφάλματος Τύπου I, η οποία ορίζεται ως εξής

$$a = P(\text{Απόρριψη της } H_0 | H_0 \text{ Αληθής}) = P(T_\beta(\hat{\theta}_\alpha, \theta_0) > c | H_0).$$

Επιπλέον, η πιθανότητα σφάλματος Τύπου II ορίζεται από τη σχέση

$$b = P(\text{Αποδοχή της } H_0 | H_a \text{ Αληθής}) = P(T_\beta(\hat{\theta}_\alpha, \theta_0) \leq c | H_a).$$

Με βάση τα παραπάνω δεν είναι εφικτός ο προσδιορισμός της κρίσιμης περιοχής, καθώς προαπαιτείται ο προσδιορισμός της κατανομής του γραμμικού συνδυασμού χ^2 τυχαίων μεταβλητών με έναν βαθμό ελευθερίας.

Για να ξεπεραστεί το πρόβλημα αυτό, θα βασιστούμε στο Remark 3 της δημοσίευσης των Basu et al. (2013), σελ. 326, και στη σχετική βιβλιογραφία, όπου και προτείνονται εναλλακτικές στατιστικές συναρτήσεις, των οποίων η κατανομή μπορεί ασυμπτωτικά να περιγραφεί ως γραμμικός συνδυασμός ανεξάρτητων χ^2 μεταβλητών. Συγκεκριμένα, προτείνονται τέσσερις διαδικασίες για την προσέγγιση του κρίσιμου σημείου c , οι οποίες περιγράφονται στη συνέχεια.

(α). Αρχικά, οι Basu et al. (2013) προτείνουν τη στατιστική συνάρτηση

$$T_\beta^{(1)}(\hat{\theta}_\alpha, \theta_0) = \frac{T_\beta(\hat{\theta}_\alpha, \theta_0)}{\lambda_{\max}^{\alpha, \beta}} \equiv \frac{\sum_{i=1}^r \lambda_i^{\alpha, \beta} Z_i^2}{\lambda_{\max}^{\alpha, \beta}},$$

όπου $\lambda_{\max}^{\alpha, \beta} = \max\{\lambda_1^{\alpha, \beta}, \dots, \lambda_r^{\alpha, \beta}\}$ (βλέπε Rao and Scott (1981)).

Υποθέτουμε, χωρίς βλάβη της γενικότητας ότι η μεγαλύτερη ιδιοτιμή είναι η $\lambda_1^{\alpha, \beta}$ και ότι οι ιδιοτιμές διατάσσονται ως εξής, $\lambda_1^{\alpha, \beta} > \lambda_2^{\alpha, \beta} > \dots > \lambda_r^{\alpha, \beta}$. Τότε, προκύπτει ότι

$$\frac{\sum_{i=1}^r \lambda_i^{\alpha, \beta} Z_i^2}{\lambda_{\max}^{\alpha, \beta}} = \frac{\lambda_1^{\alpha, \beta} Z_1^2 + \dots + \lambda_r^{\alpha, \beta} Z_r^2}{\lambda_1^{\alpha, \beta}} \leq \frac{\lambda_1^{\alpha, \beta} (Z_1^2 + \dots + Z_r^2)}{\lambda_1^{\alpha, \beta}} \leq \sum_{i=1}^r Z_i^2,$$

δεδομένου ότι τα Z_i είναι τυπικές κανονικές κατανομές, συμπεραίνεται ότι

$$\sum_{i=1}^r Z_i^2 \sim \chi_r^2.$$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

Επομένως, η μηδενική υπόθεση απορρίπτεται σε προσεγγιστικό επίπεδο σημαντικότητας a , με βάση το στατιστικό $T_\beta^{(1)}$, για $T_\beta^{(1)}(\hat{\theta}_\alpha, \theta_0) > \chi_{r,a}^2$, με $\chi_{r,a}^2$ το εκατοστιαίο σημείο της χ_r^2 κατανομής, το οποίο ορίζεται από τη σχέση

$$P(\chi_r^2 \geq \chi_{r,a}^2) = a.$$

(β). Μια άλλη διαφορετική προσέγγιση, εκείνη των ασυμπτωτικών tail probabilities, διατυπώνεται με βάση τη στατιστική συνάρτηση

$$T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0) = \frac{T_\beta(\hat{\theta}_\alpha, \theta_0)}{\bar{\lambda}^{\alpha,\beta}},$$

όπου $\bar{\lambda}^{\alpha,\beta} = \frac{1}{r} \sum_{i=1}^r \lambda_i^{\alpha,\beta}$ (βλέπε Satterthwaite (1946)).

Το στατιστικό $T_\beta^{(2)}$ προσεγγίζεται από κατανομή χ^2 με r βαθμούς ελευθερίας, καθώς όπως φαίνεται η αναμενόμενη τιμή του είναι

$$E\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right) = E\left(\frac{T_\beta(\hat{\theta}_\alpha, \theta_0)}{\bar{\lambda}^{\alpha,\beta}}\right) = \frac{E\left(T_\beta(\hat{\theta}_\alpha, \theta_0)\right)}{\bar{\lambda}^{\alpha,\beta}}, \quad (3.10)$$

όπου

$$E\left(T_\beta(\hat{\theta}_\alpha, \theta_0)\right) \simeq E\left(\sum_{i=1}^r \lambda_i^{\alpha,\beta} Z_i^2\right) = \sum_{i=1}^r \lambda_i^{\alpha,\beta} E\left(Z_i^2\right),$$

και δεδομένου ότι Z_i^2 ακολουθεί χ_1^2 , δηλαδή $E\left(Z_i^2\right) = 1$, προκύπτει ότι η αναμενόμενη τιμή του $T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)$ ισούται κατά προσέγγιση με

$$E\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right) = \frac{E\left(T_\beta(\hat{\theta}_\alpha, \theta_0)\right)}{\bar{\lambda}^{\alpha,\beta}} \simeq \frac{\sum_{i=1}^r \lambda_i^{\alpha,\beta}}{\frac{1}{r} \sum_{i=1}^r \lambda_i^{\alpha,\beta}} = r = E(\chi_r^2). \quad (3.11)$$

Επίσης, παρατηρείται ότι η διακύμανση του είναι

$$Var\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right) = Var\left(\frac{T_\beta(\hat{\theta}_\alpha, \theta_0)}{\bar{\lambda}^{\alpha,\beta}}\right) = \frac{1}{(\bar{\lambda}^{\alpha,\beta})^2} Var\left(T_\beta(\hat{\theta}_\alpha, \theta_0)\right), \quad (3.12)$$

όπου

$$Var\left(T_\beta(\hat{\theta}_\alpha, \theta_0)\right) \simeq Var\left(\sum_{i=1}^r \lambda_i^{\alpha,\beta} Z_i^2\right) = \sum_{i=1}^r (\lambda_i^{\alpha,\beta})^2 Var\left(Z_i^2\right),$$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

και δεδομένου ότι Z_i^2 ακολουθεί χ_1^2 , δηλαδή $Var(Z_i^2) = 2$, προκύπτει ότι

$$Var\left(T_\beta(\hat{\theta}_\alpha, \theta_0)\right) \simeq 2 \sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2.$$

Επομένως, η διακύμανση του $T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)$ ισούται κατά προσέγγιση με

$$Var\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right) = Var\left(\frac{T_\beta(\hat{\theta}_\alpha, \theta_0)}{\bar{\lambda}^{\alpha, \beta}}\right) \simeq \frac{2 \sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2},$$

όπου αν προσθεθεί και αφαιρεθεί στον αριθμητή η ποσότητα $2r(\bar{\lambda}^{\alpha, \beta})^2$, προκύπτει

$$\begin{aligned} \frac{\sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2}{2 \frac{\sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2}} &= 2 \frac{\sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2 + r(\bar{\lambda}^{\alpha, \beta})^2 - r(\bar{\lambda}^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2} \\ &= 2r + 2 \frac{\sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2 - r(\bar{\lambda}^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2} \\ &= 2r + 2 \frac{\sum_{i=1}^r (\lambda_i^{\alpha, \beta} - \bar{\lambda}^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2} > 2r = Var[\chi_r^2]. \end{aligned}$$

Συνεπώς, η διακύμανση του στατιστικού $T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)$ είναι

$$Var\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right) = 2r + 2 \frac{\sum_{i=1}^r (\lambda_i^{\alpha, \beta} - \bar{\lambda}^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2} > 2r = Var[\chi_r^2]. \quad (3.13)$$

Βάσει των σχέσεων (3.11) και (3.13), το στατιστικό προσεγγίζεται από κατανομή χ^2 με r βαθμούς ελευθερίας και η μηδενική υπόθεση απορρίπτεται σε προσεγγιστικό επίπεδο σημαντικότητας a , για $T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0) > \chi_{r,a}^2$, με $\chi_{r,a}^2$ το εκατοστιαίο σημείο της χ_r^2 κατανομής, το οποίο ορίζεται

$$P(\chi_r^2 \geq \chi_{r,a}^2) = a.$$

(γ). Επιπλέον, οι Basu et al. (2013) αναφέρουν το στατιστικό

$$T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0) = \frac{T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)}{\nu^{\alpha, \beta}} = \frac{T_\beta(\hat{\theta}_\alpha, \theta_0)}{\nu^{\alpha, \beta} \bar{\lambda}^{\alpha, \beta}},$$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

όπου η τιμή $\nu^{\alpha,\beta}$ μπορεί να υπολογιστεί, επιβάλλοντας την προϋπόθεση $Var[T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)] = 2E[T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)]$, όπως ισχύει στην χ^2 κατανομή.

Η αναμενόμενη τιμή του στατιστικού $T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)$ είναι

$$E\left(T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)\right) = E\left(\frac{T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)}{\nu^{\alpha,\beta}}\right) = \frac{1}{\nu^{\alpha,\beta}} E\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right) \simeq \frac{r}{\nu^{\alpha,\beta}}$$

και η διακύμανση του είναι

$$Var\left(T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)\right) = Var\left(\frac{T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)}{\nu^{\alpha,\beta}}\right) = \frac{1}{(\nu^{\alpha,\beta})^2} Var\left(T_\beta^{(2)}(\hat{\theta}_\alpha, \theta_0)\right).$$

όπου με αντικατάσταση από τη σχέση (3.13), γράφεται ως εξής

$$Var\left(T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)\right) \simeq \frac{1}{(\nu^{\alpha,\beta})^2} \left(2r + 2 \frac{\sum_{i=1}^r (\lambda_i^{\alpha,\beta} - \bar{\lambda}^{\alpha,\beta})^2}{(\bar{\lambda}^{\alpha,\beta})^2} \right),$$

ή ισοδύναμα

$$Var\left(T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)\right) \simeq \frac{2r}{(\nu^{\alpha,\beta})^2} \left(1 + \frac{\sum_{i=1}^r (\lambda_i^{\alpha,\beta} - \bar{\lambda}^{\alpha,\beta})^2}{r(\bar{\lambda}^{\alpha,\beta})^2} \right).$$

Για να ισχύει η προϋπόθεση $Var[T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)] = 2E[T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)]$, η τιμή $\nu^{\alpha,\beta}$ θα πρέπει να ισούται με

$$\nu^{\alpha,\beta} = 1 + \sum_{i=1}^r \frac{(\lambda_i^{\alpha,\beta} - \bar{\lambda}^{\alpha,\beta})^2}{r(\bar{\lambda}^{\alpha,\beta})^2} = 1 + CV^2(\{\lambda_i^{\alpha,\beta}\}_{i=1}^r),$$

όπου CV είναι ο Συντελεστής Μεταβλητότητας (coefficient of variation). Τότε, για μεγάλο μέγεθος δείγματος n , η ασυμπτωτική κατανομή του στατιστικού $T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)$ προσεγγίζεται από την κατανομή χ^2 με $r/\nu^{\alpha,\beta}$ βαθμούς ελευθερίας.

Παρατηρείται ότι οι βαθμοί ελευθερίας του στατιστικού $T_\beta^{(3)}(\hat{\theta}_\alpha, \theta_0)$ ενδέχεται να μην είναι ακέραιος αριθμός. Για να αποφευχθεί αυτή η δυσκολία, μπορεί να τροποποιηθεί το στατιστικό έτσι ώστε οι δύο πρώτες ροπές του να ταιριάζουν

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

συγκεκριμένα με την κατανομή χ_r^2 αντί για οποιαδήποτε άλλη χ^2 κατανομή. Συγκεκριμένα, έστω $X = T_{\beta}^{(2)}(\hat{\theta}_{\alpha}, \theta_0)$, με μέση τιμή

$$E[X] = r = E[\chi_r^2]$$

και διακύμανση

$$Var[X] = \frac{2 \sum_{i=1}^r (\lambda_i^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})} = 2r + 2 \sum_{i=1}^r \frac{(\lambda_i^{\alpha, \beta} - \bar{\lambda}^{\alpha, \beta})^2}{(\bar{\lambda}^{\alpha, \beta})^2} = 2r + c.$$

Έστω $Y = \frac{X-A}{B}$ όπου A, B σταθερές, τέτοιες ώστε να ισχύουν

$$\begin{aligned} E[Y] &= r, \\ Var[Y] &= 2r. \end{aligned}$$

Επομένως, έχουμε ότι

$$E(Y) = E\left(\frac{X-A}{B}\right) = \frac{E(X) - A}{B} = \frac{r - A}{B}$$

και

$$Var(Y) = Var\left(\frac{X-A}{B}\right) = \frac{1}{B^2} Var(X) = \frac{2r + c}{B^2}$$

και άρα προκύπτει το σύστημα

$$\frac{r - A}{B} = r \tag{3.14}$$

$$\frac{2r + c}{B^2} = 2r, \tag{3.15}$$

με λύσεις

$$\begin{aligned} A &= r(1 - B) \\ B &= \sqrt{1 + \frac{c}{2r}}. \end{aligned}$$

Έτσι, οι Basu et al. (2013) αναφέρουν μία τροποποίηση του στατιστικού $T_{\beta}^{(3)}(\hat{\theta}_{\alpha}, \theta_0)$, που δίνεται από

$$T_{\beta}^{(4)}(\hat{\theta}_{\alpha}, \theta_0) = \frac{T_{\beta}^{(2)}(\hat{\theta}_{\alpha}, \theta_0) - A}{B},$$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

η οποία για μεγάλο μέγεθος δείγματος μπορεί να προσεγγιστεί από την κατανομή χ_r^2 .

Εκτός από τις παραπάνω προσεγγίσεις, είναι δυνατόν να εξεταστούν πίνακες της αθροιστικής κατανομής $\sum_{i=1}^r a_i Z_i^2$ στην περίπτωση μικρού r (βλέπε Solomon (1960), Johnson and Kotz (1968), Eckler (1969) και Gupta (1963)). Οι προσεγγίσεις αυτές στοχεύουν στη διευκόλυνση του υπολογισμού των ουρών πιθανότητας (tail probabilities) γραμμικών συνδυασμών των μεταβλητών, που ακολουθούν την κατανομή χ^2 . Στην επόμενη ενότητα θα αναφερθούμε σε μία προσέγγιση της συνάρτησης ισχύος της $T_\beta(\hat{\theta}_\alpha, \theta_0)$ για τον έλεγχο της απλής μηδενικής υπόθεσης.

3.1.1 Συνάρτηση Ισχύος

Η ισχύς ενός ελέγχου στατιστικών υποθέσεων αποτελεί ένα κρίσιμο μέτρο για την αξιολόγηση της απόδοσης του προτεινόμενου ελέγχου, καθώς εκφράζει την πιθανότητα σωστής απόρριψης της μηδενικής υπόθεσης, δηλαδή την πιθανότητα

$$\gamma_n(\theta) = P_\theta(T_n \geq C_n), \text{ για } \theta \neq \theta_0,$$

όπου T_n είναι το στατιστικό και C_n το κρίσιμο σημείο.

Η ισχύς του στατιστικού ελέγχου, εξαρτάται από το μέγεθος της διαφοράς μεταξύ της πραγματικής τιμής της παραμέτρου και της τιμής που ορίζεται από τη μηδενική υπόθεση. Όσο μεγαλύτερη είναι αυτή η διαφορά, τόσο υψηλότερη είναι η ισχύς του ελέγχου. Είναι επίσης σημαντικό να αναφέρουμε ότι η ισχύς ισούται με $1 - b$, όπου b είναι η πιθανότητα σφάλματος Τύπου II.

Το επόμενο θεώρημα προέρχεται από τους Basu et al. (2013), Theorem 5, σελ. 329, μεταφέρεται πιστά και αναφέρεται σε μια προσέγγιση της συνάρτησης ισχύος, βάση της $T_\beta(\hat{\theta}_\alpha, \theta_0)$, για τον έλεγχο της απλής μηδενικής υπόθεσης $H_0 : \theta = \theta_0$.

Θεώρημα 3.2. Έστω ότι ισχύουν οι συνθήκες κανονικότητας $(\Sigma_1) - (\Sigma_5)$. Τότε, μια προσέγγιση της συνάρτησης ισχύος του στατιστικού $T_\beta(\hat{\theta}_\alpha, \theta_0)$ για την $H_0 : \theta = \theta_0$ έναντι της $H_a : \theta \neq \theta_0$ δίνεται από

$$\gamma_{n,a}^{\alpha,\beta}(\theta^*) \simeq 1 - \Phi \left(\frac{\sqrt{n}}{\sigma_{\alpha,\beta}(\theta^*)} \left(\frac{t_a^{\alpha,\beta}}{2n} - d_\beta(f_{\theta^*}, f_{\theta_0}) \right) \right), \quad (3.16)$$

για $\theta^* \neq \theta_0$, όπου $t_a^{\alpha,\beta}$ είναι το εκατοστιαίο σημείο της ασυμπτωτικής κατανομής της $T_\beta(\hat{\theta}_\alpha, \theta_0)$ και $\sigma_{\alpha,\beta}(\theta^*)$ ορίζεται από τη σχέση (3.18), πιο κάτω.

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

Απόδειξη. Η απόδειξη ακολουθεί πιστά την απόδειξη του Theorem 5 στην εργασία των Basu et al. (2013), σελ. 330. Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης.

Αρχικά θα προσδιοριστεί η ασυμπτωτική κατανομή της εκτιμώμενης DPD απόκλισης $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$ της απόκλισης $d_\beta(f_\theta, f_{\theta_0})$. Για τον προσδιορισμό αυτής, θεωρούμε το ανάπτυγμα Taylor πρώτης τάξης της $d_\beta(\hat{\theta}_\alpha, \theta_0)$ γύρω από το $\theta^* = \hat{\theta}_\alpha$, για $\theta^* \neq \theta_0$, το οποίο δίνεται από την σχέση

$$d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) = d_\beta(f_{\theta^*}, f_{\theta_0}) + B_\beta^t(\hat{\theta}_\alpha - \theta^*) + o(\|\hat{\theta}_\alpha - \theta^*\|), \quad (3.17)$$

όπου $B_\beta = (B_1^\beta, \dots, B_p^\beta)^t$ και $B_j^\beta = (\frac{\partial}{\partial \theta_j} d_\beta(f_\theta, f_{\theta_0}))_{\theta=\theta^*}$.

Γνωρίζουμε ότι από το Θεώρημα 2.2

$$\sqrt{n}(\hat{\theta}_\alpha - \theta^*) \xrightarrow{L} N(0, J_\alpha^{-1}(\theta^*) K_\alpha(\theta^*) J_\alpha^{-1}(\theta^*))$$

και

$$\sqrt{n}o(\|\hat{\theta}_\alpha - \theta^*\|) = o_p(1).$$

Τότε είναι προφανές ότι οι τυχαίες μεταβλητές $\sqrt{n}(d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) - d_\beta(f_{\theta^*}, f_{\theta_0}))$ και $B_\beta^t \sqrt{n}(\hat{\theta}_\alpha - \theta^*)$ έχουν ίδιες ασυμπτωτικές κατανομές

$$\sqrt{n}(d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) - d_\beta(f_{\theta^*}, f_{\theta_0})) \xrightarrow{L} N(0, \sigma_{\alpha,\beta}^2(\theta^*)),$$

όπου

$$\sigma_{\alpha,\beta}^2(\theta^*) = B_\beta^t(\theta^*) J_\alpha^{-1}(\theta^*) K_\alpha(\theta^*) J_\alpha^{-1}(\theta^*) B_\beta(\theta^*). \quad (3.18)$$

Εφόσον προσδιορίστηκε η ασυμπτωτική κατανομή του εκτιμητή $d_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0})$ της απόκλισης $d_\beta(f_\theta, f_{\theta_0})$, μπορεί να προσδιοριστεί μια προσέγγιση της συνάρτησης ισχύος, για $\theta^* \neq \theta_0$.

Η συνάρτηση ισχύος προκύπτει από

$$\gamma_{n,a}^{\alpha,\beta}(\theta) = P(T_\beta(\hat{\theta}_\alpha, \theta_0) > t_a^{\alpha,\beta}) = P(2nd_\beta(f_{\hat{\theta}_\alpha}, f_{\theta_0}) > t_a^{\alpha,\beta}). \quad (3.19)$$

ΚΕΦΑΛΑΙΟ 3 3.1. Έλεγχος της απλής στατιστικής υπόθεσης μέσω της DPD

Βάσει της σχέσης (3.17) έχουμε

$$\begin{aligned}
& P\left(2nd_{\beta}(f_{\hat{\theta}_{\alpha}}, f_{\theta_0}) - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0}) > t_a^{\alpha, \beta} - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})\right) \\
&= P\left(\frac{2nd_{\beta}(f_{\hat{\theta}_{\alpha}}, f_{\theta_0}) - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})}{\sigma_{\alpha, \beta}(\theta^*)} > \frac{t_a^{\alpha, \beta} - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})}{\sigma_{\alpha, \beta}(\theta^*)}\right) \\
&= P\left(\frac{d_{\beta}(f_{\hat{\theta}_{\alpha}}, f_{\theta_0}) - d_{\beta}(f_{\theta^*}, f_{\theta_0})}{\sigma_{\alpha, \beta}(\theta^*)} > \frac{t_a^{\alpha, \beta} - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})}{2n\sigma_{\alpha, \beta}(\theta^*)}\right) \\
&= P\left(\sqrt{n} \frac{d_{\beta}(f_{\hat{\theta}_{\alpha}}, f_{\theta_0}) - d_{\beta}(f_{\theta^*}, f_{\theta_0})}{\sigma_{\alpha, \beta}(\theta^*)} > \sqrt{n} \frac{t_a^{\alpha, \beta} - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})}{2n\sigma_{\alpha, \beta}(\theta^*)}\right) \\
&= 1 - P\left(\sqrt{n} \frac{d_{\beta}(f_{\hat{\theta}_{\alpha}}, f_{\theta_0}) - d_{\beta}(f_{\theta^*}, f_{\theta_0})}{\sigma_{\alpha, \beta}(\theta^*)} < \sqrt{n} \frac{t_a^{\alpha, \beta} - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})}{2n\sigma_{\alpha, \beta}(\theta^*)}\right) \\
&= 1 - P\left(Z < \sqrt{n} \frac{t_a^{\alpha, \beta} - 2nd_{\beta}(f_{\theta^*}, f_{\theta_0})}{2n\sigma_{\alpha, \beta}(\theta^*)}\right),
\end{aligned}$$

όπου $Z \xrightarrow{L} N(0, 1)$.

Συνεπώς, η συνάρτηση ισχύος προκύπτει από τη σχέση

$$\gamma_{n, a}^{\alpha, \beta}(\theta^*) \simeq 1 - \Phi\left(\frac{\sqrt{n}}{\sigma_{\alpha, \beta}(\theta^*)} \left(\frac{t_a^{\alpha, \beta}}{2n} - d_{\beta}(f_{\theta^*}, f_{\theta_0})\right)\right), \quad (3.20)$$

όπου $\Phi(x)$ η αθροιστική συνάρτηση κατανομής της τυπικής κανονικής κατανομής και $t_a^{\alpha, \beta}$ το κρίσιμο σημείο που προσδιορίζεται από την πιθανότητα σφάλματος Τύπου I

$$P(T_{\beta}(\hat{\theta}_{\alpha}, \theta_0) > t_a^{\alpha, \beta}) = a.$$

□

Παρατήρηση 3.2. Σύμφωνα με τον Basu et al. (2013), σελ. 330, αν $\theta^* \neq \theta_0$, ο έλεγχος θεωρείται συνεπής αν η ισχύς του τείνει στο 1, καθώς το μέγεθος του δείγματος αυξάνεται, δηλαδή αν

$$\gamma_n(\theta) \rightarrow 1, \text{ για κάθε } \theta \neq \theta_0, \text{ καθώς } n \rightarrow \infty.$$

3.2 Τοπική Υπόθεση (local alternative)

Ένα τελευταίο συμπέρασμα που μελετάται στην εργασία των Basu et al. (2013), είναι η ασυμπτωτική κατανομή του στατιστικού ελέγχου της Density Power Divergence απόκλισης, υπό μια ακολουθία τοπικών υποθέσεων. Το θεώρημα που δίνει την ασυμπτωτική κατανομή του $T_\beta(\hat{\theta}_\alpha, \theta_0)$, προέρχεται από την εργασία των Basu et al. (2013), Theorem 4, σελ. 329, και μεταφέρεται πιστά από εκεί.

Θεώρημα 3.3. Έστω ότι ισχύουν οι συνθήκες κανονικότητας $(\Sigma_1) - (\Sigma_5)$. Τότε, υπό τη μηδενική τοπική υπόθεση $H_{1,n} : \theta_n = \theta_0 + \frac{1}{\sqrt{n}}d$, όπου d είναι ένα σταθερό διάνυσμα στον χώρο $\mathbb{R}^p$, τέτοιο ώστε $\theta_n \in \Theta \subseteq \mathbb{R}^p$, η ασυμπτωτική κατανομή της $T_\beta(\hat{\theta}_\alpha, \theta_0)$ συμπίπτει με την κατανομή της

$$\sum_{i=1}^r \lambda_i^{\alpha, \beta}(\theta_0)(Z_i + w_i)^2 + \xi,$$

όπου Z_i είναι ανεξάρτητες τυχαίες μεταβλητές από τυπική κανονική κατανομή με Z_i^2 να είναι χ_1^2 και $\lambda_i^{\alpha, \beta}(\theta_0)$ να είναι οι θετικές ιδιοτιμές του πίνακα $A_\beta(\theta_0)J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0)$, $w = (w_1, \dots, w_r)$ και ξ να δίνονται από τις σχέσεις

$$\begin{aligned} w &= \Lambda_r^{-1} R^t S^t A_\beta(\theta_0) d, \\ \xi &= d^t A_\beta(\theta_0) d - w^t \Lambda_r w, \end{aligned}$$

όπου S είναι μια οποιαδήποτε ρίζα του πίνακα $J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0)$, $\Lambda_r = \text{diag}(\lambda_1^{\alpha, \beta}(\theta_0), \dots, \lambda_r^{\alpha, \beta}(\theta_0))$ και R είναι ο πίνακας των ορθοκανονικών ιδιοδιανυσμάτων του.

Απόδειξη. Η απόδειξη ακολουθεί πιστά την απόδειξη του Theorem 4 στην εργασία των Basu et al. (2013), σελ. 329. Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης.

Μπορούμε να γράψουμε

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

Κάτω από την τοπική υπόθεση $H_{1,n}$ έχουμε

$$\sqrt{n}(\hat{\theta}_\alpha - \theta_n) \xrightarrow[n \rightarrow \infty]{L} N(0, J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0))$$

και

$$\sqrt{n}(\hat{\theta}_\alpha - \theta_0) \xrightarrow[n \rightarrow \infty]{L} N(d, J_\alpha^{-1}(\theta_0)K_\alpha(\theta_0)J_\alpha^{-1}(\theta_0)).$$

Επιπλέον, ξέρουμε ότι

$$T_\beta(\hat{\theta}_\alpha, \theta_0) = \sqrt{n}(\hat{\theta}_\alpha - \theta_0)^t A_\beta(\theta_0) \sqrt{n}(\hat{\theta}_\alpha - \theta_0) + n o(\|\hat{\theta}_\alpha - \theta_0\|^2).$$

Τότε η στατιστική συνάρτηση $T_\beta(\hat{\theta}_\alpha, \theta_0)$ έχει ίδια ασυμπτωτική κατανομή με την τετραγωνική μορφή $\sqrt{n}(\hat{\theta}_\alpha - \theta_0)^t A_\beta(\theta_0) \sqrt{n}(\hat{\theta}_\alpha - \theta_0)$ και από το Πόρισμα 2.2. της εργασίας των Dik and de Gunst (1985) προκύπτει η ασυμπτωτική κατανομή. $\square$

Για λόγους πληρότητας διατυπώνεται το Πόρισμα 2.2 των Dik and de Gunst (1985).

Παρατήρηση 3.3. (Πόρισμα 2.2 Dik and de Gunst (1985)). Έστω X μια q -διάστατη τυχαία μεταβλητή που ακολουθεί πολυδιάστατη κανονική κατανομή, με μέση τιμή το διάνυσμα d και πίνακα διακυμάνσεων-συνδιακυμάνσεων Σ , δηλαδή $X \sim N_q(d, \Sigma)$. Έστω A ένας πραγματικός θετικά ορισμένος συμμετρικός πίνακας, τάξης q . Έστω επίσης $r = \text{rank}(\Sigma A \Sigma)$ και $\lambda_1, \dots, \lambda_r$ οι μη αρνητικές ιδιοτιμές του πίνακα $A \Sigma$. Τότε η κατανομή της τετραγωνικής μορφής $X^t A X$ ταυτίζεται με την κατανομή του $\sum_{j=1}^r \lambda_j (Z_j + w_j)^2 + \xi$, όπου $Z_1, \dots, Z_r$ ανεξάρτητες τυπικές κανονικές κατανομές. Επίσης είναι

$$\begin{aligned} w &= \Lambda^{-1} R^t S^t A d, \\ \xi &= d^t A d - w^t \Lambda w, \end{aligned}$$

όπου $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$, S είναι μια αυθαίρετη ρίζα του πίνακα Σ και R είναι ο πίνακας των ιδιοδιανυσμάτων του $S^t A S$.

ΚΕΦΑΛΑΙΟ 4

ΕΦΑΡΜΟΓΕΣ ΤΗΣ DENSITY POWER DIVERGENCE

Στο κεφάλαιο αυτό, πραγματοποιείται μία σύντομη περιγραφή ορισμένων σημαντικών ερευνητικών εργασιών, που αφορούν εφαρμογές της οικογένειας Density Power Divergence σε προβλήματα παραμετρικής Συμπερασματολογίας. Η συζήτηση αυτών θα είναι συνοπτική, παραπέμποντας στη σχετική βιβλιογραφία για περαιτέρω διεξοδική συζήτηση και μελέτη.

Παλινδρόμηση και Γενικευμένα Γραμμικά Μοντέλα

Στην Εφαρμοσμένη Στατιστική, η Παλινδρόμηση αποτελεί ένα από τα πιο διαδεδομένα εργαλεία για τη διερεύνηση της σχέσης μεταξύ ενός συνόλου επεξηγηματικών μεταβλητών και μιας εξαρτημένης μεταβλητής. Στην υπάρχουσα βιβλιογραφία, οι εκτιμήσεις των παραμέτρων του μοντέλου Παλινδρόμησης βασίζονται στους εκτιμητές Μέγιστης Πιθανοφάνειας, όμως σε περιπτώσεις μελέτης μεγάλων συνόλων δεδομένων, η μέθοδος αυτή είναι ιδιαίτερα ασταθής σύμφωνα με τους Durio and Isaia (2011), καθώς οι ακραίες τιμές μπορούν να προκαλέσουν σημαντικά προβλήματα στις κλασικές στατιστικές μεθόδους. Συγκεκριμένα, οι Durio and Isaia (2011) εξετάζουν τη χρήση του κριτηρίου Minimum Density Power Divergence (MDPD) ως ένα πρακτικό εργαλείο για την κατασκευή παραμετρικών μοντέλων παλινδρόμησης για μεγάλα σύνολα δεδομένων, τα οποία περιέχουν σημαντικό αριθμό ακραίων τιμών. Η εργασία τους ολοκληρώνεται με την πραγματοποίηση προσομοιώσεων, οι οποίες επιβεβαιώνουν τα αποτελέσματά τους.

Επίσης, μια άλλη εφαρμογή βασισμένη στο κριτήριο Minimum Density Power

ΚΕΦΑΛΑΙΟ 4

Divergence, η οποία προέρχεται από τους Ghosh and Basu (2015), αφορά τα γενικευμένα γραμμικά μοντέλα. Τα γενικευμένα γραμμικά μοντέλα αποτελούν ένα ιδιαίτερα σημαντικό εργαλείο στην ανάλυση πραγματικών δεδομένων, όπου η σχέση μεταξύ ενός συνόλου επεξηγηματικών μεταβλητών και μιας εξαρτημένης μεταβλητής ενδέχεται να μην είναι γραμμική ή οι κατανομές να μην είναι πάντοτε κανονικές. Ειδικότερα, οι Ghosh and Basu (2015) ανέπτυξαν μια εύρωστη διαδικασία εκτίμησης για τα γενικευμένα γραμμικά μοντέλα, η οποία βασίζεται στους εκτιμητές MDPD και μπορεί να παράγει εύρωστους (robust) εκτιμητές με ελάχιστη απώλεια αποδοτικότητας. Η μέθοδος αυτή έχει χρησιμοποιηθεί επιτυχώς σε εφαρμογές λογιστικής και Poisson παλινδρόμησης.

Ακόμη μία εφαρμογή της Density Power Divergence, η οποία προέρχεται από την εργασία των Ghosh and Basu (2018), που ασχολήθηκαν με το μοντέλο Παλινδρόμησης Αναλογικού Κινδύνου του Cox. Το μοντέλο του Cox είναι ένα δημοφιλές εργαλείο για την ανάλυση της σχέσης μεταξύ μιας λογοκριμένης μεταβλητής (διάρκεια ζωής) και άλλων σχετικών παραγόντων. Το ημιπαραμετρικό μοντέλο του Cox χρησιμοποιείται ευρέως για τη μελέτη διάφορων τύπων δεδομένων, που προκύπτουν από εφαρμοσμένα πεδία, όπως η ιατρική και η βιολογία. Ωστόσο, οι Ghosh and Basu (2018) επισημαίνουν ότι μια πλήρως παραμετρική εκδοχή του μοντέλου Cox, εφόσον διατυπωθεί σωστά, μπορεί να προσφέρει βελτιωμένη ερμηνεία και μεγαλύτερη ακρίβεια στην εκτίμηση των κινδύνων. Η πλήρως παραμετρική εκδοχή του μοντέλου Cox βασίζεται στη μέθοδο Μέγιστης Πιθανοφάνειας, η οποία παρουσιάζει υψηλή ευαισθησία σε προβλήματα νοθευμένων δεδομένων (data contamination). Οι Ghosh and Basu (2018) ανέπτυξαν μια εύρωστη διαδικασία για το παραμετρικό μοντέλο Cox, βασισμένη στους εκτιμητές Minimum Density Power Divergence. Η διαδικασία που προτείνουν παράγει ιδιαίτερα εύρωστες εκτιμήσεις σε δεδομένα τύπου (data contamination), με ελάχιστη απώλεια αποδοτικότητας. Επιπλέον, οι εκτιμήσεις αυτές παρέχουν πιο ακριβή αποτελέσματα σε σύγκριση με τις εκτιμήσεις που βασίζονται στη μέθοδο πιθανοφάνειας, είτε στο ημιπαραμετρικό μοντέλο Cox, είτε σε υπάρχουσες εύρωστες εκδοχές του.

Ανάλυση Κύριων Συνιστωσών

Η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis - PCA) αποτελεί βασικό εργαλείο στην Πολυδιάστατη Ανάλυση, το οποίο χρησιμοποιείται κυρίως για τη μείωση της διάστασης των δεδομένων. Ωστόσο, η PCA είναι γνωστό ότι επηρεάζεται αρνητικά από την παρουσία ακραίων τιμών. Οι ευρωστοί μέθοδοι PCA που χρησιμοποιούν την M-εκτίμηση, έχουν θεωρητικά οφέλη, αλλά

ΚΕΦΑΛΑΙΟ 4

η ευρωστία τους μειώνεται σημαντικά σε δεδομένα με μεγάλη διάσταση. Σε πρόσφατη εργασία των Roy et al. (2024), προτείνεται μια νέα εύρωστη μέθοδος Principal Component Analysis, βασισμένη στους εκτιμητές Minimum Density Power Divergence Estimators. Η μέθοδος αυτή, γνωστή ως rPCAdpd, συνδυάζει τα θεωρητικά πλεονεκτήματα των M-εκτιμητών, με τα οφέλη των εκτιμητών τύπου ελαχίστης απόστασης της Density Power Divergence, προσφέροντας μια ιδιαίτερα ευέλικτη και αποδοτική προσέγγιση. Το βασικό χαρακτηριστικό της είναι η δυνατότητα της επιλογής της παραμέτρου α (robustness tuning parameter), η οποία επιτρέπει την ισορροπία μεταξύ ευρωστίας και αποδοτικότητας, ανεξαρτήτως της διάστασης των δεδομένων. Η προτεινόμενη μέθοδος, όχι μόνο διατηρεί υψηλό βαθμό ευρωστίας σε δεδομένα υψηλής διάστασης, αλλά και εξασφαλίζει αυξημένη σταθερότητα και αξιοπιστία στις εκτιμήσεις των κύριων συνιστωσών, ακόμη και σε περιβάλλον με σημαντική παρουσία ακραίων παρατηρήσεων.

Κριτήριο Επιλογής Μοντέλων

Τα κριτήρια επιλογής μοντέλων παρέχουν στους στατιστικούς μία μέθοδο που επιτρέπει την επιλογή του καταλληλότερου μοντέλου μέσα από μια ευρεία συλλογή πιθανών κατάλληλων μοντέλων. Η δημιουργία τέτοιων κριτηρίων βασίζεται σε μέτρα απόκλισης μεταξύ του αγνώστου μοντέλου και καθενός εκ των δύο ή περισσότερων υποψήφιων μοντέλων, τα οποία συνήθως περιγράφονται από τις συναρτήσεις κατανομών πιθανότητας. 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 έχουν δημοσιευθεί στη βιβλιογραφία που σχετίζεται με την κατασκευή κριτηρίων επιλογής μοντέλου. Η εργασία των Mattheou et al. (2009), θεωρείται ως μία εφαρμογή του μέτρου απόκλισης Density Power Divergence, καθώς αξιοποιεί το μέτρο απόκλισης Density Power Divergence για την κατασκευή ενός νέου κριτηρίου επιλογής μοντέλου, το κριτήριο Divergence Information Criterion (DIC).

Μοντέλα χρονοσειρών

Στην εργασία των Xiong and Zhu (2022), μελετάται μια εύρωστη μέθοδος εκτίμησης για μοντέλα χρονοσειρών με ακραίες τιμές. Ειδικότερα, οι Xiong and Zhu (2022) στην ανάλυσή τους, αποδεικνύουν θεωρητικά ότι ο εκτιμητής MDPD για τα NB-INGARCH μοντέλα είναι ισχυρά συνεπής (consistent) και η ασυμπτωτική

ΚΕΦΑΛΑΙΟ 4

κατανομή του είναι η κανονική, υπό τις κατάλληλες συνθήκες κανονικότητας. Παράλληλα, μέσω εκτεταμένων προσομοιώσεων αποδεικνύεται η ευρωστία του εκτιμητή σε σχέση με την κλασική μέθοδο Μέγιστης Πιθανοφάνειας, σε περιπτώσεις που τα δεδομένα περιέχουν ακραίες τιμές. Μια άμεση εφαρμογή της μεθόδου σε πραγματικά δεδομένα αποτελεί η μελέτη λοιμώξεων από καμπυλοβακτηρίωση (campylobacteriosis) (βλέπε, Xiong and Zhu (2022)). Η ανάλυση δείχνει ότι η MDPDE εκτίμηση προσφέρει σταθερότερα και πιο αξιόπιστα αποτελέσματα σε σχέση με τις κλασικές μεθόδους, καθιστώντας την κατάλληλη επιλογή για την ανάλυση χρονοσειρών καταμετρήσεων σε συνθήκες όπου υπάρχει πιθανότητα ακραίων παρατηρήσεων.

Bayesian εκτίμηση

Η κλασική Bayesian εκτίμηση βασίζεται στην εκ των υστέρων (posterior) κατανομή, η οποία προκύπτει από τον συνδυασμό της εκ των προτέρων (prior) γνώσης με την πληροφορία που παρέχουν τα δεδομένα, σύμφωνα με τον τύπο του Bayes. Παρόλο που αυτή η προσέγγιση θεωρείται ιδανική για την ενσωμάτωση σε στατιστικά μοντέλα της υποκειμενικής γνώσης, ο κλασικός Bayesian εκτιμητής μπορεί να αποδειχθεί εξαιρετικά ευαίσθητος στην παρουσία ακραίων τιμών (outliers) στα δεδομένα. Αυτή η ευαισθησία, περιορίζει τη χρησιμότητα των παραδοσιακών Bayesian μεθόδων σε εφαρμογές όπου τα δεδομένα περιέχουν ακραίες τιμές, όπως αναφέρουν οι Ghosh and Basu (2016).

Για την αντιμετώπιση αυτής της αδυναμίας, οι Ghosh and Basu (2016) πρότειναν έναν εναλλακτικό εκτιμητή Bayes, ο οποίος βασίζεται στη μέθοδο Minimum Density Power Divergence. Συγκεκριμένα, ανέπτυξαν μια εκ των υστέρων πυκνότητα $R(\alpha)$ ($R(\alpha)$ -posterior density), η οποία αποτελεί μια τροποποίηση της κλασικής εκ των υστέρων κατανομής. Η κατασκευή αυτή διατηρεί τη φιλοσοφία του Bayesian υποδείγματος ενσωματώνοντας την πληροφορία από τις εκ των προτέρων κατανομές, αλλά ταυτόχρονα ενισχύει σημαντικά την ευρωστία των εκτιμητών. Ο εκτιμητής που βασίζεται στην $R(\alpha)$ -posterior κατανομή, εξισορροπεί την αποδοτικότητα και την ευρωστία, ελέγχοντας την επιρροή των ακραίων τιμών μέσω της παραμέτρου α . Για μικρές τιμές του α , ο εκτιμητής πλησιάζει τον κλασικό Bayes εκτιμητή. Αντίθετα, για μεγαλύτερες τιμές του α , ενισχύεται η ευρωστία του εκτιμητή, μειώνοντας την ευαισθησία σε ακραίες παρατηρήσεις.

ΚΕΦΑΛΑΙΟ 5

ΠΑΡΑΡΤΗΜΑ

Απόδειξη Θεωρήματος 2.2. Η απόδειξη ακολουθεί πιστά την απόδειξη του Θεωρήματος 9.2 του βιβλίου των Basu et al. (2011), σελ. 304 και τις αποδείξεις των Theorem 1 και Theorem 2 των Leroy et al. (2016). Μεταφέρεται εδώ χάριν πληρότητας της παρουσίασης.

(i) Απόδειξη Συνέπειας του εκτιμητή MDPDE $\hat{\theta}_n^\alpha$

Σκοπός είναι να δειχθεί ότι υπάρχει εκτιμητής, που προκύπτει από τη λύση της εξίσωσης (2.15) και είναι (ασθενώς) συνεπής, δηλαδή ότι ισχύει

$$P(\|\hat{\theta}_n^\alpha - \theta_0\| < \varepsilon) \rightarrow 1,$$

όπου $\|\cdot\|$ είναι νόρμα στον Θ και θ_0 είναι η αληθής αλλά άγνωστη τιμή της θ .

Για να αποδείξουμε την ύπαρξη μιας ακολουθίας λύσεων κατά πιθανότητα, θα εξετάσουμε τη συμπεριφορά της αντικειμενικής συνάρτησης $H_n^\alpha(\theta)$ που δίνεται στην (2.14), πρώτα σε μία σφαίρα $Q_\zeta = \{\theta \in \Theta \subseteq \mathbb{R}^p : \|\theta - \theta_0\| \leq \zeta\}$ με κέντρο θ_0 και ακτίνα ζ . Θα δείξουμε ότι για κάθε επαρκώς μικρό ζ , κατά πιθανότητα, ισχύει ότι $H_n^\alpha(\theta) > H_n^\alpha(\theta_0)$ για όλα τα θ στην επιφάνεια Q_ζ , δηλαδή το $H_n^\alpha(\theta)$ έχει τοπικό ελάχιστο στο εσωτερικό της επιφάνειας Q_ζ . Δεδομένου ότι η αντικειμενική συνάρτηση $H_n^\alpha(\theta)$ είναι διαφορίσιμη, σε ένα τοπικό ελάχιστο οι εξισώσεις (2.15) της Density Power Divergence οφείλουν να ικανοποιούνται. Συνεπώς, θα προκύψει ότι για οποιοδήποτε $\zeta > 0$, καθώς το $n \rightarrow \infty$, η εξίσωση εκτίμησης έχει λύση $\hat{\theta}_{(\zeta)}$ εντός της Q_ζ .

Για να μελετήσουμε τη συμπεριφορά της $H_n^\alpha(\theta)$ στην επιφάνεια της σφαίρας Q_ζ , έστω το ανάπτυγμα σειράς Taylor της $H_n^\alpha(\theta)$ γύρω από το θ_0 . Από τη συνθήκη κανονικότητας (Σ_2), για $i = 1, \dots, n$ και $\forall \theta \in \Theta$, έχουμε

$$H_n^\alpha(\theta) = H_n^\alpha(\theta_0) + \sum_{j=1}^p \nabla_j H_n^\alpha(\theta)_{\theta=\theta_0} (\theta_j - \theta_{0j})$$

ΚΕΦΑΛΑΙΟ 5

$$\begin{aligned}
& + \frac{1}{2!} \sum_{j=1}^p \sum_{k=1}^p \nabla_{jk} H_n^\alpha(\theta)_{\theta=\theta_0} (\theta_j - \theta_{0j}) (\theta_k - \theta_{0k}) \\
& + \frac{1}{3!} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \nabla_{jkl} H_n^\alpha(\theta^*) (\theta_j - \theta_{0j}) (\theta_k - \theta_{0k}) (\theta_l - \theta_{0l}),
\end{aligned}$$

όπου θ^* ανήκει στο εσωτερικό της σφαίρας Q_ζ .

Θα διαιρέσουμε με $(1 + \alpha)$, καθώς παραγωγίζοντας την $H_n^\alpha(\theta)$ προκύπτει η ποσότητα $(1 + \alpha)$, την οποία θέλουμε να απαλείψουμε. Οπότε έχουμε

$$\begin{aligned}
\frac{1}{(1 + \alpha)} \left[H_n^\alpha(\theta_0) - H_n^\alpha(\theta) \right] &= \sum_{j=1}^p \frac{1}{(1 + \alpha)} \left(- \nabla_j H_n^\alpha(\theta) \right)_{\theta=\theta_0} (\theta_j - \theta_{0j}) \\
&+ \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{1}{(1 + \alpha)} \left(- \nabla_{jk} H_n^\alpha(\theta) \right)_{\theta=\theta_0} (\theta_j - \theta_{0j}) (\theta_k - \theta_{0k}) \quad (5.1) \\
&+ \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \frac{1}{(1 + \alpha)} \left(- \nabla_{jkl} H_n^\alpha(\theta^*) \right) (\theta_j - \theta_{0j}) (\theta_k - \theta_{0k}) (\theta_l - \theta_{0l}) \\
&= S_1 + S_2 + S_3.
\end{aligned}$$

Αναπτύσσοντας το άθροισμα S_1 , προκύπτει

$$S_1 = \sum_{j=1}^p \frac{(\theta_j - \theta_{0j})}{(1 + \alpha)} \left(- \frac{d}{d\theta_j} \left(\int f_\theta^{1+\alpha}(x) dx - \left(\frac{1 + \alpha}{\alpha} \right) \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i) \right) \right)_{\theta=\theta_0},$$

όπου από τη συνθήκη κανονικότητας (Σ_3) έχουμε ότι το ολοκλήρωμα $\int f_\theta^{a+1}(x) dx$ μπορεί να παραγωγιστεί ως προς θ και οι παράγωγοι μπορούν να υπολογιστούν εντός του ολοκληρώματος.

Έτσι,

$$S_1 = \sum_{j=1}^p \frac{(\theta_j - \theta_{0j})}{(1 + \alpha)} \left\{ - \left(\int \frac{d}{d\theta_j} f_\theta^{1+\alpha}(x) dx - \left(\frac{1 + \alpha}{\alpha} \right) \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta_j} f_\theta^\alpha(X_i) \right)_{\theta=\theta_0} \right\},$$

και παραγωγίζοντας έχουμε

$$\begin{aligned}
S_1 = \sum_{j=1}^p (\theta_j - \theta_{0j}) \frac{1}{(1 + \alpha)} \left\{ - \left(\int (1 + \alpha) f_\theta^\alpha(x) \frac{d}{d\theta_j} f_\theta(x) dx - \right. \right. \\
\left. \left. - \left(\frac{1 + \alpha}{\alpha} \right) \frac{1}{n} \sum_{i=1}^n f_\theta^{\alpha-1}(X_i) \frac{d}{d\theta_j} f_\theta(X_i) \right)_{\theta=\theta_0} \right\}.
\end{aligned}$$

ΚΕΦΑΛΑΙΟ 5

Αν οι δύο όροι πολλαπλασιαστούν και διαιρεθούν με την $f_\theta(x)$ και την $f_\theta(X_i)$ αντίστοιχα, έχουμε

$$S_1 = \sum_{j=1}^p (\theta_j - \theta_{0j}) \frac{1}{(1+\alpha)} \left\{ - \left(\int (1+\alpha) f_\theta^{1+\alpha}(x) \frac{\frac{d}{d\theta_j} f_\theta(x)}{f_\theta(x)} dx - \right. \right. \\ \left. \left. - (1+\alpha) \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i) \frac{\frac{d}{d\theta_j} f_\theta(X_i)}{f_\theta(X_i)} \right)_{\theta=\theta_0} \right\},$$

όπου $\frac{d}{d\theta_j} f_\theta(x) / f_\theta(x) = \frac{d}{d\theta_j} \ln f_\theta(x) = u_{j\theta}(x)$ και άρα το άθροισμα S_1 γράφεται

$$S_1 = \sum_{j=1}^p (\theta_j - \theta_{0j}) \left\{ - \left(\int u_{j\theta}(x) f_\theta^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_{j\theta}(X_i) f_\theta^\alpha(X_i) \right)_{\theta=\theta_0} \right\}. \quad (5.2)$$

Αναπτύσσοντας το άθροισμα S_2 , προκύπτει

$$S_2 = \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) \frac{1}{(1+\alpha)} \left\{ - \nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \right. \right. \\ \left. \left. - \frac{1+\alpha}{\alpha} \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i) \right)_{\theta=\theta_0} \right\} = \\ = \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) \frac{1}{(1+\alpha)} \left\{ - \frac{1}{n} \sum_{i=1}^n \nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \right. \right. \\ \left. \left. - \frac{1+\alpha}{\alpha} f_\theta^\alpha(X_i) \right)_{\theta=\theta_0} \right\},$$

που ισοδύναμα μπορεί να εκφραστεί από τη σχέση

$$S_2 = \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})}{(1+\alpha)} \left\{ - \frac{1}{n} \sum_{i=1}^n \nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \frac{1+\alpha}{\alpha} f_\theta^\alpha(X_i) \right) \right. \\ \left. + E_g \left[\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right] \right\}_{\theta=\theta_0} \\ + \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) \frac{1}{(1+\alpha)} \left\{ - E_g \left[\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) - \right. \right. \right. \\ \left. \left. \left. - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right]_{\theta=\theta_0} \right\}. \quad (5.3)$$

ΚΕΦΑΛΑΙΟ 5

Όμοια, αναπτύσσοντας το άθροισμα S_3 , προκύπτει η σχέση

$$\begin{aligned}
 S_3 &= \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p (\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_l - \theta_{0l}) \frac{1}{(1+\alpha)} \\
 &\quad \times \left\{ -\nabla_{jkl} \left(\int f_\theta^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i) \right) \right\}_{\theta=\theta^*} = \\
 &= \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_l - \theta_{0l})}{(1+\alpha)} \\
 &\quad \times \left(-\frac{1}{n} \right) \sum_{i=1}^n \nabla_{jkl} \left(\int f_\theta^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_\theta^\alpha(X_i) \right)_{\theta=\theta^*}, \tag{5.4}
 \end{aligned}$$

με θ^* στο εσωτερικό της σφαίρας $Q_\zeta(\theta_0)$.

Βάσει των σχέσεων (5.2), (5.3) και (5.4) η αρχική σχέση (5.1) αναπτύσσεται ως εξής:

$$\begin{aligned}
 \frac{1}{(1+\alpha)} [H_n^\alpha(\theta_0) - H_n^\alpha(\theta)] &= \sum_{j=1}^p (\theta_j - \theta_{0j}) \left\{ - \left(\int u_{j\theta}(x) f_\theta^{1+\alpha}(x) dx \right. \right. \\
 &\quad \left. \left. - \frac{1}{n} \sum_{i=1}^n u_{j\theta}(X_i) f_\theta^\alpha(X_i) \right)_{\theta=\theta_0} \right\} \\
 &+ \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})}{(1+\alpha)} \left\{ - \frac{1}{n} \sum_{i=1}^n \nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx \right. \right. \\
 &\quad \left. \left. - \frac{1+\alpha}{\alpha} f_\theta^\alpha(X_i) \right) + E_g \left[\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx \right. \right. \right. \\
 &\quad \left. \left. \left. - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right] \right\}_{\theta=\theta_0} \tag{5.5} \\
 &+ \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})}{(1+\alpha)} \left\{ - E_g \left[\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx \right. \right. \right. \right. \\
 &\quad \left. \left. \left. - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right] \right\}_{\theta=\theta_0}
 \end{aligned}$$

ΚΕΦΑΛΑΙΟ 5

$$+ \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_l - \theta_{0l})}{(1 + \alpha)} \\ \times \left(-\frac{1}{n} \right) \sum_{i=1}^n \nabla_{jkl} \left(\int f_{\theta}^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_{\theta}^{\alpha}(X_i) \right)_{\theta=\theta^*}.$$

$$= L_1 + L_2 + L_3 + L_4.$$

Θα εξεταστεί ξεχωριστά η οριακή σύγκλιση του κάθε όρου L_i , $i = 1, \dots, 4$ του δεξιού μέρους της εξίσωσης (5.5).

Για τον όρο L_1 , για κάθε $j = 1, \dots, p$, έχουμε από τον Νόμο Μεγάλων Αριθμών, για $n \rightarrow \infty$, ότι

$$\left(\frac{1}{n} \sum_{i=1}^n u_{j\theta}(X_i) f_{\theta}^{\alpha}(X_i) \right)_{\theta=\theta_0} \xrightarrow{P} E_g[u_{j\theta_0}(X) f_{\theta_0}^{\alpha}(X)].$$

και επομένως

$$\begin{aligned} & - \left\{ \int u_{j\theta}(x) f_{\theta}^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_{j\theta}(X_i) f_{\theta}^{\alpha}(X_i) \right\}_{\theta=\theta_0} \xrightarrow{P} \\ & \xrightarrow{P} - \left\{ \int u_{j\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx - \int u_{j\theta_0}(x) f_{\theta_0}^{\alpha}(x) g(x) dx \right\} \\ & = - \frac{1}{1+\alpha} \nabla_j d_{\alpha}(g(x), f_{\theta}(x)) \Big|_{\theta=\theta_0} = 0, \end{aligned} \quad (5.6)$$

με πιθανότητα να τείνει στο 1, αφού η αληθινή τιμή της παραμέτρου θ_0 οδηγεί στο μοντέλο g .

Επομένως,

$$\begin{aligned} |L_1| &= \left| \sum_{j=1}^p (\theta_j - \theta_{0j}) \left\{ - \left(\int u_{j\theta}(x) f_{\theta}^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_{j\theta}(X_i) f_{\theta}^{\alpha}(X_i) \right)_{\theta=\theta_0} \right\} \right| \\ &\leq \sum_{j=1}^p |\theta_j - \theta_{0j}| \left| - \left(\int u_{j\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_{j\theta_0}(X_i) f_{\theta_0}^{\alpha}(X_i) \right) \right| \\ &< \sum_{j=1}^p |\theta_j - \theta_{0j}| \zeta^2, \end{aligned}$$

αφού λόγω της κατά πιθανότητα σύγκλισης (5.6) έχουμε ότι

$$\left| - \left(\int u_{j\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_{j\theta_0}(X_i) f_{\theta_0}^{\alpha}(X_i) \right) \right| < \zeta^2,$$

ΚΕΦΑΛΑΙΟ 5

με πιθανότητα να τείνει στη μονάδα. Επιπλέον, στη σφαίρα Q_ζ με κέντρο θ_0 και ακτίνα ζ ισχύει $|\theta_j - \theta_{0j}| < \zeta$ και επομένως

$$|S_1| = |L_1| < \sum_{i=1}^n \zeta \zeta^2 = p\zeta^3, \quad (5.7)$$

με πιθανότητα να τείνει στο 1.

Σχετικά με τον όρο L_2 , για κάθε $(j, k) \in \{1, \dots, p\}^2$, από τον Νόμο των Μεγάλων Αριθμών έχουμε

$$\begin{aligned} & - \left\{ \frac{1}{(1+\alpha)n} \sum_{i=1}^n \left(\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_\theta^\alpha(X_i) \right)_{\theta=\theta_0} \right) \right\} \xrightarrow{P} \\ & \xrightarrow{P} - \left\{ \frac{1}{(1+\alpha)} E_g \left(\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_\theta^\alpha(X) \right)_{\theta=\theta_0} \right) \right\} = \\ & = -\frac{1}{1+\alpha} \left(E_g \left[\nabla_{kl} \left(\int f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right]_{\theta=\theta_0} \right) = -J_{jk}(\theta_0), \end{aligned} \quad (5.8)$$

όπου $J_{jk}(\theta_0)$ ορίζεται στη σχέση (2.18).

Έτσι,

$$\begin{aligned} |L_2| &= \left| \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})}{(1+\alpha)} \left\{ \right. \right. \\ & \quad \times \left(-\frac{1}{n} \right) \sum_{i=1}^n \nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \frac{1+\alpha}{\alpha} f_\theta^\alpha(X_i) \right) \\ & \quad \left. \left. + E_g \left[\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right] \right\}_{\theta=\theta_0} \right| \\ &\leq \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p |(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})| \\ & \quad \times \left| \frac{1}{(1+\alpha)} \left\{ -\frac{1}{n} \sum_{i=1}^n \nabla_{jk} \left(\int f_{\theta_0}^{1+\alpha}(x) dx - \frac{1+\alpha}{\alpha} f_{\theta_0}^\alpha(X_i) \right) \right. \right. \\ & \quad \left. \left. + E_g \left[\nabla_{jk} \left(\int f_{\theta_0}^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha} \right) f_{\theta_0}^\alpha(X) \right) \right] \right\} \right| \end{aligned}$$

ΚΕΦΑΛΑΙΟ 5

και επειδή η δεύτερη απόλυτη τιμή συγκλίνει κατά πιθανότητα στο 0, λόγω της κατά πιθανότητα σύγκλισης (5.8) έχουμε ότι

$$\left| \frac{1}{(1+\alpha)} \left\{ -\frac{1}{n} \sum_{i=1}^n \nabla_{jk} \left(\int f_{\theta_0}^{1+\alpha}(x) dx - \frac{1+\alpha}{\alpha} f_{\theta_0}^{\alpha}(X_i) \right) + E_g \left[\nabla_{jk} \left(\int f_{\theta_0}^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha} \right) f_{\theta_0}^{\alpha}(X) \right) \right] \right\} \right| < \zeta,$$

με πιθανότητα να τείνει στο 1. Επιπλέον, στη σφαίρα Q_ζ με κέντρο θ_0 και ακτίνα ζ έτσι ώστε $|(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})| < \zeta^2$. Έτσι,

$$|L_2| \leq \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \zeta^2 \zeta = \frac{1}{2} p^2 \zeta^3.$$

Για τον όρο L_3 , για κάθε $(j, k) \in \{1, \dots, p\}^2$, έχουμε

$$L_3 = \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})}{(1+\alpha)} \left\{ -E_g \left[\nabla_{jk} \left(\int f_{\theta}^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha} \right) f_{\theta}^{\alpha}(X) \right) \right]_{\theta=\theta_0} \right\},$$

$$L_3 = \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (-J_{jk})(\theta_j - \theta_{0j})(\theta_k - \theta_{0k}),$$

και επομένως η L_3 μπορεί να εκφραστεί ως διανυσματική μορφή

$$L_3 = \frac{1}{2} (\theta - \theta_0)^t (-J) (\theta - \theta_0),$$

όπου J είναι ο πίνακας με στοιχεία J_{jk} που ορίζεται στη σχέση (2.18).

Ακολουθώντας τώρα πιστά την απόδειξη στη σελ. 305 των Basu et al. (2011), επειδή ο πίνακας $-J$ είναι συμμετρικός και αρνητικά ορισμένος λόγω της συνθήκης (Σ_4) , υπάρχει ορθογώνιος πίνακας Q τέτοιος ώστε

$$Q^t (-J) Q = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix},$$

ΚΕΦΑΛΑΙΟ 5

όπου Λ διαγώνιος πίνακας με διαγώνια στοιχεία τις ιδιοτιμές του πίνακα $-J$ οι οποίες είναι αρνητικές και οι οποίες διατάσσονται ως εξής, $\lambda_p \leq \lambda_{p-1} \leq \dots \leq \lambda_1 < 0$. Επομένως, καθώς ο πίνακας μιας τετραγωνικής μορφής είναι συμμετρικός και συνεπώς διαγωνιοποιήσιμος σε μία ορθοκανονική βάση, έχουμε

$$\frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (-J_{jk})(\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) = \sum_{j=1}^p \lambda_j \xi_j^2,$$

$$\text{με } \sum_{j=1}^p \xi_j^2 = \sum_{j=1}^p (\theta_j - \theta_{0j})^2 = \zeta^2.$$

Ισχύει ότι

$$\sum_{j=1}^p \lambda_j \xi_j^2 = \lambda_1 \xi_1 + \dots + \lambda_p \xi_p \leq \lambda_1 \sum_{j=1}^p \xi_j^2 = \max(\lambda_j) \zeta^2.$$

Έτσι,

$$|L_3| = \left| \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})}{(1 + \alpha)} \left\{ -E_g \left[\nabla_{jk} \left(\int f_\theta^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha} \right) f_\theta^\alpha(X) \right) \right] \right\} \right| \leq \frac{1}{2} \max(\lambda_j) \zeta^2 < 0.$$

Συνεπώς, στη σφαίρα Q_ζ με κέντρο θ_0 και ακτίνα ζ έτσι ώστε $|(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})| < \zeta^2$, ισχύει ότι

$$|L_3| < \frac{1}{2} \max(\lambda_j) \zeta^2$$

με πιθανότητα να τείνει στο 1.

Συνδυάζοντας τις L_2 και L_3 , προκύπτει ότι $|S_2| = |L_2 + L_3| < (1/2)p^2\zeta^3 + \frac{1}{2}\max(\lambda_j)\zeta^2$ και αποδεικνύεται ότι για ένα ζ_0 και μία c θετική σταθερά, τέτοια ώστε για κάθε $\zeta < \zeta_0$, να έχουμε

$$|S_2| < \frac{1}{2}p^2\zeta^3 + \frac{1}{2}\max(\lambda_j)\zeta^2 = -c\zeta^2 \quad (5.9)$$

με πιθανότητα να τείνει στο 1.

Τέλος για τον όρο L_4 , ακολουθώντας τους Basu et al. (2011), σελ. 306 και Leroy et al. (2016), σελ. 102, για κάθε $(j, k, l) \in \{1, \dots, p\}^3$, και από την

ΚΕΦΑΛΑΙΟ 5

συνθήκη κανονικότητας (Σ_5) έχουμε ότι

$$-\frac{1}{n} \sum_{i=1}^n \nabla_{jkl} \left(\int f_{\theta}^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_{\theta}^{\alpha}(X_i) \right) \leq -\frac{1}{n} \sum_{i=1}^n M_{jkl},$$

όπου από Νόμο Μεγάλων Αριθμών έχουμε

$$\frac{1}{n} \sum_{i=1}^n M_{jkl} \xrightarrow{P} E_g(M_{jkl}) = m_{jkl} < \infty. \quad (5.10)$$

Επομένως,

$$\begin{aligned} |L_4| &= \left| \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \frac{(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_l - \theta_{0l})}{(1+\alpha)} \right. \\ &\quad \times \left(-\frac{1}{n} \right) \sum_{i=1}^n \nabla_{jkl} \left(\int f_{\theta}^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_{\theta}^{\alpha}(X_i) \right)_{\theta=\theta^*} \Bigg| \leq \\ &\leq \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \left| (\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_l - \theta_{0l}) \right| \\ &\quad \left| -\frac{1}{(1+\alpha)} \frac{1}{n} \sum_{i=1}^n \nabla_{jkl} \left(\int f_{\theta}^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_{\theta}^{\alpha}(X_i) \right)_{\theta=\theta^*} \right|. \end{aligned}$$

Επειδή, η δεύτερη απόλυτη τιμή συγκλίνει κατά πιθανότητα στο m_{jkl} , λόγω της σχέσης (5.10), έχουμε ότι

$$\left| -\frac{1}{(1+\alpha)} \frac{1}{n} \sum_{i=1}^n \nabla_{jkl} \left(\int f_{\theta}^{1+\alpha}(x) dx - \left(\frac{1+\alpha}{\alpha} \right) f_{\theta}^{\alpha}(X_i) \right)_{\theta=\theta^*} \right| < 2M,$$

με πιθανότητα να τείνει στο 1.

Επιπλέον, στη σφαίρα Q_{ζ} με κέντρο το θ_0 και ακτίνα ζ έτσι ώστε $|(\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_l - \theta_{0l})| < \zeta^3$.

Έτσι,

$$|S_3| = |L_4| < \frac{2}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \zeta^3 M = \frac{2}{6} p^3 \zeta^3 M = b \zeta^3, \quad (5.11)$$

όπου $b = p^3 M/3$.

ΚΕΦΑΛΑΙΟ 5

Συνδυάζοντας τις σχέσεις (5.7), (5.9) και (5.11) προκύπτει ότι

$$\frac{1}{(1+\alpha)} \left[H_n^\alpha(\theta_0) - H_n^\alpha(\theta) \right] < -c\zeta^2 + (b+s)\zeta^3,$$

όπου το $-c\zeta^2 + (b+s)\zeta^3 < 0$ αν και μόνο αν $\zeta < \frac{c}{b+s}$.

Επομένως αν υποθέσουμε ότι $\zeta < \frac{c}{b+s}$, προκύπτει ότι

$$\frac{1}{(1+\alpha)} \left[H_n^\alpha(\theta_0) - H_n^\alpha(\theta) \right] < 0, \quad \forall \theta \in Q_\zeta.$$

Ακολουθώντας πιστά την τεκμηρίωση στη μονογραφία των Basu et al. (2011), σελ. 306, αν επιλεγεί τιμή ζ μικρότερη από το $\min(\zeta_0, \frac{c}{b+s})$, η πιθανότητα

$$P(\forall \theta \in Q_\zeta, H_n^\alpha(\theta) - H_n^\alpha(\theta_0) > 0)$$

τείνει στο 1, όταν το $n \rightarrow \infty$. Επομένως υπάρχει $\hat{\theta}_n^\alpha$ το οποίο ανήκει στο εσωτερικό της σφαίρας Q_ζ τέτοιο ώστε $\|\hat{\theta}_n^\alpha - \theta\| < \zeta$ και η $H_n^\alpha(\theta)$ έχει τοπικό ελάχιστο το $\hat{\theta}_n^\alpha$, που ολοκληρώνει την απόδειξη.

Επομένως για επαρκώς μικρό ζ υπάρχει μια ακολουθία λύσεων $\hat{\theta}_n^\alpha = \hat{\theta}_{(\zeta)}^\alpha$ τέτοια ώστε: $P(\|\hat{\theta}_n^\alpha - \theta\| < \zeta) \rightarrow 1$, όπου $\|\cdot\|$ αναπαριστά τη νόρμα του χώρου Θ . Απομένει να δείξουμε ότι μπορούμε να προσδιορίσουμε μια τέτοια ακολουθία ανεξάρτητα από το ζ . Έστω θ^* είναι μια ρίζα κοντά στο θ . Αυτή υπάρχει, επειδή το όριο της ακολουθίας ριζών είναι επίσης ρίζα, λόγω της συνέχειας της $H_n^\alpha(\theta)$ ως συνάρτησης του θ . Τότε προφανώς, $P(\|\theta^* - \theta\| < \zeta) \rightarrow 1$ για όλα τα $\zeta > 0$. Έτσι ολοκληρώνεται η απόδειξη της ύπαρξης μιας ακολουθίας συνεπών λύσεων της Density Power Divergence εξίσωσης εκτίμησης κατά πιθανότητα.

(ii) Ασυμπτωτική κανονικότητα του MDPDE $\hat{\theta}_n^\alpha$

Τώρα θα παρουσιαστεί η απόδειξη της ασυμπτωτικής κανονικότητας του εκτιμητή τύπου ελάχιστης απόστασης με βάση την Density Power Divergence (MDPDE), που συμβολίζεται με $\hat{\theta}_n^\alpha$, η οποία ακολουθεί πιστά τα βήματα της απόδειξης των Basu et al. (2011), σελ. 306-307. Για να προσδιορισθεί η ασυμπτωτική κατανομή του εκτιμητή θεωρούμε το ανάπτυγμα Taylor, γύρω από το θ_0 , της $H_n^{\alpha(j)}(\theta)$, με $H_n^{\alpha(j)}(\theta) = \frac{\partial}{\partial \theta_j} H_n^\alpha(\theta)$ και $H_n^\alpha(\theta) = \int f_\theta^{1+\alpha}(x) dx - (1 + \frac{1}{\alpha}) \frac{1}{n} \sum_{i=1}^n f_\theta^\alpha(X_i)$ να δίνεται από τη σχέση (2.14).

ΚΕΦΑΛΑΙΟ 5

Τότε

$$\begin{aligned} H_n^{\alpha(j)}(\theta) &= H_n^{\alpha(j)}(\theta_0) + \sum_{k=1}^p (\theta_k - \theta_{0k}) H_n^{\alpha(jk)}(\theta)_{\theta=\theta_0} + \\ &+ \frac{1}{2} \sum_{k=1}^p \sum_{l=1}^p (\theta_k - \theta_{0k}) (\theta_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*), \end{aligned}$$

όπου θ^* ανήκει στο εσωτερικό της σφαίρας $Q_\zeta = \{\theta \in \Theta \subseteq \mathbb{R}^p : \|\theta - \theta_0\| \leq \zeta\}$ με κέντρο θ_0 και ακτίνα ζ . Επίσης $H_n^{\alpha(jk)}$, $H_n^{\alpha(jkl)}$ είναι οι μερικοί παράγωγοι δεύτερης και τρίτης τάξης της H_n^α .

Αντικαθιστώντας $\theta = \hat{\theta}_n^\alpha$ ισχύει ότι

$$H_n^{\alpha(j)}(\hat{\theta}_n^\alpha) = 0$$

και επομένως

$$H_n^{\alpha(j)}(\hat{\theta}_n^\alpha) = H_n^{\alpha(j)}(\theta_0) + \sum_{k=1}^p (\hat{\theta}_k - \theta_{0k}) \left\{ H_n^{\alpha(jk)}(\theta_0) + \frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*) \right\}$$

ή

$$H_n^{\alpha(j)}(\theta_0) + \sum_{k=1}^p (\hat{\theta}_k - \theta_{0k}) \left\{ H_n^{\alpha(jk)}(\theta_0) + \frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*) \right\} = 0$$

ή

$$-H_n^{\alpha(j)}(\theta_0) = \sum_{k=1}^p (\hat{\theta}_k - \theta_{0k}) \left\{ H_n^{\alpha(jk)}(\theta_0) + \frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*) \right\}$$

και τελικά,

$$-\sqrt{n} H_n^{\alpha(j)}(\theta_0) = \sqrt{n} \sum_{k=1}^p (\hat{\theta}_k - \theta_{0k}) \left\{ H_n^{\alpha(jk)}(\theta_0) + \frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*) \right\}. \quad (5.12)$$

Η σχέση (5.12) γράφεται σε διανυσματική μορφή

$$T_n = A_n Y_n \quad (5.13)$$

ΚΕΦΑΛΑΙΟ 5

ή ισοδύναμα

$$\sqrt{n} \begin{bmatrix} \left. \frac{\partial H_n^\alpha(\theta)}{\partial \theta_1} \right|_{\theta=\theta_0} \\ \vdots \\ \left. \frac{\partial H_n^\alpha(\theta)}{\partial \theta_p} \right|_{\theta=\theta_0} \end{bmatrix} = \begin{bmatrix} a_{11,n} & a_{12,n} & \cdots & a_{1p,n} \\ \vdots & \ddots & \cdots & \vdots \\ a_{p1,n} & a_{p2,n} & \cdots & a_{pp,n} \end{bmatrix} \sqrt{n} \begin{bmatrix} (\hat{\theta}_1 - \theta_{01}) \\ \vdots \\ (\hat{\theta}_p - \theta_{0p}) \end{bmatrix}$$

όπου

$$T_n = -\sqrt{n} H_n^{\alpha(j)}(\theta_0) = -\sqrt{n} \left(\frac{\partial H_n^\alpha(\theta_0)}{\partial \theta_1}, \dots, \frac{\partial H_n^\alpha(\theta_0)}{\partial \theta_p} \right)^t,$$

και $A_n \in \mathbb{R}^{p \times p}$ να είναι ο πίνακας με στοιχεία $a_{jk,n}$, με

$$a_{jk,n} = H_n^{\alpha(jk)}(\theta) \Big|_{\theta=\theta_0} + \frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta) \Big|_{\theta=\theta^*}.$$

Στόχος είναι να βρεθεί η ασυμπτωτική κατανομή του τυχαίου διανύσματος

$$Y_n = \sqrt{n}(\hat{\theta}_n^\alpha - \theta_0).$$

Γνωρίζουμε ότι από τη συνθήκη κανονικότητας (Σ_5), έχουμε ότι $H_n^{\alpha(jkl)}(\theta^*)$ είναι φραγμένο και εφόσον ισχύει $P(|\hat{\theta}_l - \theta_{0l}| < \alpha) \rightarrow 0$, συνεπάγεται ότι

$$\frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*) \xrightarrow{P} 0.$$

Ακόμη έχουμε δείξει στο πρώτο σκέλος της συνέπειας, με τη σύγκλιση της σχέσης (5.8), ότι

$$H_n^{\alpha(jk)}(\theta_0) \xrightarrow{P} (1 + \alpha) J_{jk}.$$

Επομένως

$$a_{jk,n} = H_n^{\alpha(jk)}(\theta_0) + \frac{1}{2} \sum_{l=1}^p (\hat{\theta}_l - \theta_{0l}) H_n^{\alpha(jkl)}(\theta^*) \xrightarrow{P} (1 + \alpha) J_{jk}$$

δηλαδή ο πίνακας $A_n \xrightarrow{P} A = (1 + \alpha)J$ με την έννοια της κατά στοιχείο σύγκλισης και με πιθανότητα να τείνει στο 1.

Επιπλέον, από τη συνθήκη κανονικότητας (Σ_4) γνωρίζουμε ότι ο πίνακας J είναι θετικά ορισμένος και συνεπώς ο πίνακας J είναι αντιστρέψιμος. Εφόσον

ΚΕΦΑΛΑΙΟ 5

ο πίνακας J είναι αντιστρέψιμος, η λύση του συστήματος της σχέσης (5.13) ως προς το διάνυσμα Y_n οδηγεί στη σχέση

$$Y_n = A_n^{-1}T_n. \quad (5.14)$$

Στην συνέχεια θα δειχθεί ότι το τυχαίο διάνυσμα $T_n = -\sqrt{n} \left(\frac{\partial H_n^\alpha(\theta_0)}{\partial \theta_1}, \dots, \frac{\partial H_n^\alpha(\theta_0)}{\partial \theta_p} \right)^t$ έχει πολυδιάστατη ασυμπτωτική κανονική κατανομή με μέσο διάνυσμα το μηδενικό διάνυσμα και πίνακα διακυμάνσεων-συνδιακυμάνσεων $(1 + \alpha)^2 K$, όπου K δίνεται από τη σχέση (2.16). Το αποτέλεσμα αυτό θα προκύψει κάνοντας χρήση του Κεντρικού Οριακού Θεωρήματος. Από τη σχέση (2.13) και από το γεγονός ότι ο όρος $\frac{1}{\alpha} g^{1+\alpha}(x)$ είναι ανεξάρτητος της θ , οι εκτιμητές MDPDE $\hat{\theta}_n^\alpha$ μπορούν να ληφθούν από την παράγωγο της $H(\theta)$, εξισώνοντάς την με το μηδέν, όπου

$$H(\theta) = \int f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha}\right) \int f_\theta^\alpha(x) g(x) dx. \quad (5.15)$$

Παραγωγίζοντας τη σχέση (5.15) προκύπτει

$$H^j(\theta) = \nabla_j H(\theta) = \frac{d}{d\theta_j} \left\{ \int f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha}\right) \int f_\theta^\alpha(x) g(x) dx \right\},$$

όπου από την (Σ_3) τα ολοκληρώματα μπορούν να παραγωγηθούν και έχουμε

$$\begin{aligned} H^j(\theta) &= \int \frac{d}{d\theta_j} f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha}\right) \int \frac{d}{d\theta_j} f_\theta^\alpha(x) g(x) dx \\ &= \int (1 + \alpha) f_\theta^\alpha(x) \frac{d}{d\theta_j} f_\theta(x) dx - \left(\frac{1 + \alpha}{\alpha}\right) \int \alpha f_\theta^{\alpha-1}(x) g(x) \frac{d}{d\theta_j} f_\theta(x) dx \\ &= (1 + \alpha) \left\{ \int f_\theta^\alpha(x) \frac{d}{d\theta_j} f_\theta(x) dx - \int f_\theta^{\alpha-1}(x) g(x) \frac{d}{d\theta_j} f_\theta(x) dx \right\}, \end{aligned}$$

που αν πολλαπλασιαστεί και διαιρεθεί η ποσότητα $f_\theta(x)$ στους δύο όρους προκύπτει

$$H^j(\theta) = (1 + \alpha) \left\{ \int f_\theta^{1+\alpha}(x) \frac{df_\theta(x)}{d\theta_j} \frac{1}{f_\theta(x)} dx - \int f_\theta^\alpha(x) g(x) \frac{df_\theta(x)}{d\theta_j} \frac{1}{f_\theta(x)} dx \right\},$$

με $\frac{d}{d\theta_j} f_\theta(x) / f_\theta(x) = \nabla_j \ln f_\theta(x) = u_{j\theta}(x)$.

Επομένως, προκύπτει

$$H^j(\theta) = (1 + \alpha) \left\{ \int u_{j\theta}(x) f_\theta^{1+\alpha}(x) dx - \int u_{j\theta}(x) f_\theta^\alpha(x) g(x) dx \right\}$$

ΚΕΦΑΛΑΙΟ 5

και οι εκτιμητές MDPDE $\hat{\theta}_n^\alpha$ προκύπτουν από τη λύση των εξισώσεων

$$\int u_{j\theta}(x) f_\theta^{1+\alpha}(x) dx - \int u_{j\theta}(x) f_\theta^\alpha(x) g(x) dx = 0, \text{ για } j = 1, \dots, p. \quad (5.16)$$

Όμως

$$E(T_n) = E_g(\sqrt{n} H_n^{\alpha(j)}(\theta_0)) = 0$$

αφού λόγω της (5.16), για μεγάλη n ,

$$\begin{aligned} E_g[H_n^{\alpha(j)}(\theta_0)] &= \frac{1}{n} \sum_{i=1}^n \left\{ \int u_{j\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx - \right. \\ &\quad \left. E_g \left[u_{j\theta_0}(X_i) f_{\theta_0}^\alpha(X_i) dx \right] \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \int u_{j\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx - \int u_{j\theta_0}(x) f_{\theta_0}^\alpha(x) g(x) dx \right\} \\ &= 0, \end{aligned}$$

δεδομένου ότι η αληθής τιμή της παραμέτρου θ_0 οδηγεί στο μοντέλο g .

Ο πίνακας διακυμάνσεων-συνδιακυμάνσεων του τυχαίου διανύσματος $T_n = -\sqrt{n} H_n^{\alpha(j)}(\theta_0)$ προκύπτει από τη σχέση

$$K(\theta) = Var_g \left[-\sqrt{n} H_n^{\alpha(j)}(\theta_0) \right] \quad (5.18)$$

$$\text{όπου } H_n^{\alpha(j)}(\theta_0) = (1 + \alpha) \left\{ \int u_{\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx - \frac{1}{n} \sum_{i=1}^n u_{\theta_0}(X_i) f_{\theta_0}^\alpha(X_i) \right\}.$$

Ο όρος $\int u_{\theta_0}(x) f_{\theta_0}^{1+\alpha}(x) dx$ είναι σταθερός ως προς g και μπορεί να αγνοηθεί. Τότε, ο πίνακας διακυμάνσεων-συνδιακυμάνσεων, του T_n , σύμφωνα με τους Basu et al. (2011), σελ. 307 είναι

$$\begin{aligned} K(\theta_0) &= (1 + \alpha)^2 \left\{ \int u_{\theta_0}(x) u_{\theta_0}^t(x) f_{\theta_0}^{2\alpha}(x) g(x) dx - \right. \\ &\quad \left. - \left(\int u_{\theta_0}(x) f_{\theta_0}^\alpha(x) g(x) \right) \left(\int u_{\theta_0}(x) f_{\theta_0}^\alpha(x) g(x) \right)^t \right\}. \end{aligned}$$

Επομένως, η ασυμπτωτική κατανομή του τυχαίου διανύσματος T_n είναι πολυδιάστατη κανονική

$$T_n = -\sqrt{n} H_n^{\alpha(j)}(\theta_0) \xrightarrow{L} N_p(0, K(\theta_0)).$$

Τέλος, από την (5.14) έχουμε ότι

$$Y_n = A_n^{-1}T_n,$$

όπου

$$A_n \xrightarrow{P} A = (1 + \alpha)J,$$

$$T_n \xrightarrow{L} N_p(0, K(\theta_0))$$

και άρα το τυχαίο διάνυσμα

$$Y_n = \sqrt{n}(\hat{\theta}_n^\alpha - \theta_0) \xrightarrow{L} N_p(0, J^{-1}(\theta_0)K(\theta_0)J^{-1}(\theta_0))$$

εφόσον,

$$A^{-1}(1 + \alpha)^2 K A^{-1} = ((1 + \alpha)J)^{-1}(1 + \alpha)^2 K((1 + \alpha)J)^{-1} = J^{-1}KJ^{-1},$$

που ολοκληρώνει την απόδειξη. $\square$

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In : Petrov, B. N. & Csaki, F., editor, *Proceeding of the second International Symposium on Information Theory*. Akademiai Kiado, Budapest.
- Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **28**:131–142.
- 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.
- Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: Efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, **46**:683–705.
- 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.
- Basu, A., Shioya, H., and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. Chapman & Hall/CRC.
- Beran, R. (1977). Minimum hellinger distance estimates for parametric models. *Annals of Statistics*, **5**:445–463.
- Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin Calcutta Mathematical Society*, **35**:99–109.

- Billingsley, P. (1995). *Probability and Measure*. Wiley, New York, 3rd edition.
- Cao, R. Cuevas, A. and Fraiman, R. (1995). Minimum distance density based estimation. *Computational Statistics and Data Analysis*, **20**:611–631.
- Castilla, E., Martín, N., Pardo, L., and Zografos, K. (2018). Composite likelihood methods based on minimum density power divergence estimator. *Entropy*, **20**(1):18.
- Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **46**:440–464.
- 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.
- Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, **2**:299–318.
- Dik, J. J. and de Gunst, M. C. M. (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica*, **39**:14–26.
- Durio, A. and Isaia, E. D. (2011). The minimum density power divergence approach in building robust regression models. *Informatica*, **22**(1):43–56.
- Eckler, A. R. (1969). A survey of coverage problems associated with point and area targets. *Technometrics*, **11**:561–589.
- Ferentinos, K. and Papaioannou, T. (1981). New parametric measures of information. *Information and Control*, **51**:193–208.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Hafner Press, New York.
- Ghosh, A. and Basu, A. (2015). Robust estimation in generalized linear models: the density power divergence approach. *Test*.
- Ghosh, A. and Basu, A. (2016). Robust bayes estimation using the density power divergence. *Annals of the Institute of Statistical Mathematics*, **68**(2):413–437.
- Ghosh, A. and Basu, A. (2018). Robust and efficient estimation in the parametric cox regression model under random censoring. *Statistics in Medicine*, **38**(27):1–17.

- Gupta, S. S. (1963). Bibliography on the multivariate normal integrals and related topics. *The Annals of Mathematical Statistics*, **34**:829–838.
- Harville, D. A. (2008). *Matrix Algebra from a statistician's perspective*. Springer, New York.
- Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik*, **136**:210–271.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A*, **186**:453–561.
- Johnson, N. L. and Kotz, S. (1968). Tables of distributions of positive definite quadratic forms in central normal variables. *Sankhya B—The Indian Journal of Statistics*, **30**:303–314.
- Kagan, A. M. (1963). On the theory of fisher's information quantity. *Dokl. Akad. Nauk SSSR*, **151**:277–278.
- Karunamuni, R. J. and Wu, J. (2009). Minimum hellinger distance estimation in a nonparametric mixture model. *Journal of Statistical Planning and Inference*, **139** :1118–1133.
- Kullback, S. (1959). *Information Theory and Statistics*. John Wiley & Sons, New York.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**(1):79–86.
- Lehmann, E. L. and Casella, N. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York, NY, USA.
- Leroy, F., Dauxois, J. Y., and Tubert-Bitter, P. (2016). On the parametric maximum likelihood estimator for independent but non-identically distributed observations with application to truncated data. *Journal of Statistical Theory and Applications*, **15**:96–107.
- Liese, F. and Vajda, I. (1987). *Convex Statistical Distances*. Teubner Texts in Mathematics. Teubner, Leipzig.
- 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.

- 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.
- Matusita, K. (1964). Distance and decision rules. *Annals of the Institute of Statistical Mathematics*, **16**:305–320.
- 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.
- Papaioannou, T. (1985). Measures of information. In *Encyclopedia of Statistical Sciences Vol.5* Kotz, S., Johnson, N. L. John Wiley & Sons, New York, 391–397.
- Pardo, L. (2006). *Statistical inference based on divergence measures*. Chapman & Hall/CRC.
- 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.
- Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.
- Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, **76**:221–230.
- Rényi, A. (1961). On measures of entropy and information. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **I**:547–561.
- Roy, S., Basu, A., and Ghosh, A. (2024). Robust principal component analysis using density power divergence. *Journal of Machine Learning Research*, **25**(324):1–40.
- Saks, S. (1937). *Theory of the Integral*. PWN, Lwów, Warszawa.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**:110–114.

- Simpson, D. G. (1987). Minimum hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**:802–807.
- Solomon, H. (1960). Distribution of quadratic forms: tables and applications. Technical Report 45, Applied Mathematics and Statistics Laboratories, Stanford University, Stanford.
- Tamura, R. N. and Boos, D. D. (1986). Minimum hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, **81**:223–229.
- Toma, A. (2008). Minimum hellinger distance estimators for multivariate distributions from the johnson system. *Journal of Statistical Planning and Inference*, **138**:803–816.
- Vajda, I. (1987). *Theory of Statistical Inference and Information*. Kluwer Academic Publishers.
- Vajda, I. (1995). Information theoretic methods in statistics. research report. *Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation*, Research Report No 1834.
- Vajda, I. (2009). On metric divergences of probability measures. *Kybernetika*, **45**(6):885–900.
- Wu, J. and Karunamuni, R. J. (2009). On minimum hellinger distance estimation. *Canadian Journal of Statistics*, **37**:514–533.
- Xiong, L. and Zhu, F. (2022). Minimum density power divergence estimator for negative binomial integer-valued garch models. *Communications in Mathematics and Statistics*, **10**:233–261.
- Zografos, K. (2023). On reconsidering entropies and divergences and their cumulative counterparts: Csiszár’s, dpd’s and fisher’s type cumulative and survival measures. *Probability in the Engineering and Informational Sciences*, **37**:294–321.
- 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.
- Αυλογιάρης, (2017). *Τοπικές ϕ -Αποκλίσεις στη Στατιστική Θεωρία Πληροφοριών και Εφαρμογές σε Ελέγχους Στατιστικών Υποθέσεων και Επιλογής Μοντέλου*. Διδακτορική Διατριβή, University of Ioannina.