



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ



Πασχάλης Χ. Μαυρίδης

ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ ΓΙΑ ΤΗ ΣΥΝΑΡΤΗΣΗ
ΡΥΘΜΟΥ ΕΠΙΚΙΝΔΥΝΟΤΗΤΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννινα, 2023



UNIVERSITY OF IOANNINA
Department of Mathematics



Paschalis C. Mavridis

GOODNESS-OF-FIT TESTING FOR THE HAZARD
RATE FUNCTION

Master's Thesis

Ioannina, 2023

*Αφιερώνεται στη μνήμη του πατέρα μου Χρυσολόγου και του παππού μου
Κωνσταντίνου.*

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

Εγκρίθηκε την 03/05/2023 από την εξεταστική επιτροπή:

Όνοματεπώνυμο	Βαθμίδα
Δημήτριος Μπάγκαβος	Επίκουρος Καθηγητής (Επιβλέπων)
Απόστολος Μπατσίδης	Αναπληρωτής Καθηγητής
Ιωάννης Δημητρίου	Αναπληρωτής Καθηγητής

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.”

Πασχάλης Χ. Μαυρίδης

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διατριβή εκπονήθηκε στα πλαίσια των υποχρεώσεων μου για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στον τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, με επιβλέποντα τον Επίκουρο Καθηγητή κ. Δημήτριο Μπάγκαβο και μέλη της Τριμελούς Εξεταστικής Επιτροπής τους κ.κ. Απόστολο Μπατσίδα και Ιωάννη Δημητρίου. Στην συνέχεια θα ήθελα να εκφράσω τις ευχαριστίες μου σε όσους συνέβαλαν στην ολοκλήρωση αυτής της διατριβής.

Αρχικά θα ήθελα να ευχαριστήσω τον κ. Δημήτριο Μπάγκαβο που με προθυμία δέχτηκε να αναλάβει την επίβλεψη αυτής της διατριβής. Τον ευχαριστώ βαθύτατα για την επιστημονική βοήθεια, την καθοδήγηση, την υποστήριξη, τον χρόνο και τις πολύτιμες συμβουλές που μου προσέφερε καθόλη τη διάρκεια της εκπόνησης της διατριβής. Επιπλέον θα ήθελα να τον ευχαριστήσω θερμά για την εμπιστοσύνη που έδειξε στη δουλειά μου καθώς επίσης και για την επιβράβευση της προσπάθειάς μου μέσα από τα λόγια του.

Ακόμη, θα ήθελα να εκφράσω τις ευχαριστίες μου προς τους κ.κ. Ιωάννη Δημητρίου, Αναπληρωτή Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων και Απόστολο Μπατσίδα, Αναπληρωτή Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για την συμμετοχή τους στην Τριμελή Εξεταστική Επιτροπή και για τις πολύτιμες παρατηρήσεις και υποδείξεις τους στην διατριβή μου με σκοπό την βελτίωσή της. Θα ήθελα επίσης να ευχαριστήσω και τους υπόλοιπους καθηγητές του Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας για τις πολύτιμες γνώσεις που μου παρείχαν καθ' όλη την διάρκεια των προπτυχιακών αλλά και των μεταπτυχιακών μου σπουδών.

Στη συνέχεια θα ήθελα να ευχαριστήσω από καρδιάς την οικογένειά μου για την απέραντη αγάπη, υποστήριξη και κατανόησή τους κατά τη διάρκεια της προπτυχιακής και της μεταπτυχιακής μου πορείας. Οι γονείς μου και τα μέλη της οικογένειάς μου ήταν πάντοτε εκεί για μένα, παρέχοντας μου συνεχή υποστήριξη και ενθάρρυνση. Η αφοσίωσή τους στην εκπαίδευσή μου και η ανεκτίμητη πίστη τους στις ικανότητές μου με ενέπνευσαν να προχωρήσω πέρα από τα όρια μου.

Επιπλέον, θα ήθελα να ευχαριστήσω τον παππού, τη γιαγιά και τα αδέρφια μου για τις θυσίες, την εμπιστοσύνη και την ενθάρρυνσή τους να πετύχω τους στόχους μου.

Επίσης, ένα μεγάλο ευχαριστώ στους φίλους και συμφοιτητές μου που πίστεψαν σε μένα.

Τέλος, θέλω να ευχαριστήσω την αγαπημένη μου σύντροφο για την αμέριστη στήριξη και κατανόησή της καθ' όλη τη διάρκεια της προπτυχιακής και μεταπτυχιακής μου πορείας.

Ιωάννινα, Μάιος 2023
Πασχάλης Χ. Μαυρίδης

ΠΕΡΙΛΗΨΗ

Η συνάρτηση κινδύνου ή συνάρτηση ρυθμού επικινδυνότητας (Hazard Rate function) κατέχει εξέχουσα θέση στην ανάλυση επιβίωσης, καθώς εκφράζει τον στιγμιαίο ρυθμό με τον οποίο συμβαίνουν γεγονότα στους υπό μελέτη οργανισμούς ή στα υπό μελέτη αντικείμενα κατά τη διάρκεια μίας συγκεκριμένης χρονικής περιόδου. Τα τελευταία χρόνια έχει αυξηθεί το ενδιαφέρον για την ανάπτυξη ελέγχων καλής προσαρμογής ειδικά για τη συνάρτηση ρυθμού επικινδυνότητας. Τέτοιοι έλεγχοι είναι ιδιαίτερα σημαντικοί διότι επιτρέπουν στους ερευνητές να καθορίσουν αν η υποτιθέμενη κατανομή για τη συνάρτηση ρυθμού επικινδυνότητας, είναι κατάλληλη για τα παρατηρούμενα δεδομένα. Εάν η υποτιθέμενη κατανομή δεν περιγράφει ικανοποιητικά τα διαθέσιμα δεδομένα, οι ερευνητές μπορεί να οδηγηθούν σε μεροληπτικά αποτελέσματα και κατά συνέπεια σε λανθασμένα συμπεράσματα. Σκοπός αυτής της μεταπτυχιακής διατριβής είναι η πρόταση ενός ελέγχου καλής προσαρμογής για τη συνάρτηση ρυθμού επικινδυνότητας, ειδικά για δεδομένα χρόνων ζωής που είναι δεξιά λογοκριμένα. Συγκεκριμένα στο Κεφάλαιο 1 (Εισαγωγή) παρουσιάζεται η έννοια των ελέγχων καλής προσαρμογής και ο ρόλος τους στη στατιστική, όπως επίσης και κάποιοι χρήσιμοι ορισμοί που θα χρησιμοποιηθούν σε όλη την έκταση της παρούσας μεταπτυχιακής διατριβής. Το Κεφάλαιο 2 (Η συνάρτηση κινδύνου) εστιάζει στη θεωρία της συνάρτησης κινδύνου και στον ρόλο που διαδραματίζει στην ανάλυση επιβίωσης. Συγκεκριμένα, εξετάζονται οι μαθηματικές ιδιότητες της συνάρτησης κινδύνου, συμπεριλαμβανομένης της σχέσης της με άλλες έννοιες της ανάλυσης επιβίωσης, όπως η συνάρτηση επιβίωσης και η αθροιστική συνάρτηση κινδύνου. Επιπλέον, γίνεται ιδιαίτερη αναφορά στην μορφή την οποία μπορεί να πάρει μία συνάρτηση κινδύνου, μέσω της μελέτης της μονοτονίας της. Το κεφάλαιο ολοκληρώνεται παρουσιάζοντας την έννοια της λογοκρισίας και τα τρία είδη της που συναντώνται σε δεδομένα χρόνων ζωής. Το Κεφάλαιο 3 (Εκτίμηση της συνάρτησης κινδύνου) είναι αφιερωμένο στην εκτίμηση της συνάρτησης κινδύνου, κάνοντας μία αναφορά στην παραμετρική εκτίμησή της και στην συνέχεια στη μη παραμετρική εκτίμησή της. Συγκεκριμένα, παρουσιάζεται ένας εκτιμητής τύπου πυρήνα και μελετώνται αναλυτικά η μέση τιμή, η διακύμανση και η κατανομή του. Στο Κεφάλαιο 4 (Έλεγχος καλής προσαρμογής) εισάγεται ένας νέος έλεγχος καλής προσαρμογής για τη συ-

νάφτηση κινδύνου. Αρχικά, ορίζονται η μηδενική και η εναλλακτική υπόθεση, ενώ στη συνέχεια δημιουργείται η ελεγχοσυνάρτηση, η οποία βασίζεται στο ολοκληρωμένο τετραγωνικό σφάλμα. Σε επόμενο βήμα μελετάται η κατανομή της, με τη χρήση θεωρίας πιθανοτήτων, καθώς και σχέσεων για ροπές, παραγώγων, γινόμενων και συνελίξεων. Έπειτα, στο Κεφάλαιο 5 (Προσομοιώσεις) προσδιορίζεται η εμπειρική συνάρτηση ισχύος του ελέγχου, σε ειδικές περιπτώσεις εναλλακτικών υποθέσεων, οι παράμετροι των οποίων επιλέγονται με τέτοιο τρόπο ώστε η απόκλιση Kullback-Leibler από τη μηδενική υπόθεση να αυξάνεται συνεχώς. Επιπρόσθετα, προσδιορίζεται η κρίσιμη περιοχή όπως επίσης και άλλοι παράμετροι που απαιτούνται για να εφαρμοστεί το τεστ στην πράξη. Τα αποτελέσματα επιβεβαιώνονται στη συνέχεια μέσω προσομοιώσεων χρησιμοποιώντας τη γλώσσα προγραμματισμού της R. Στο Κεφάλαιο 6 (Επίλογος - Θέμα προς περαιτέρω έρευνα), παρουσιάζεται ένα θέμα για περαιτέρω έρευνα. Τέλος, στο Παράρτημα Α' (Πίνακες), περιέχονται οι πίνακες που προέκυψαν από την υλοποίηση του ελέγχου μέσω του λογισμικού της R. Συγκεκριμένα, παρατίθενται οι πίνακες των Kullback-Leibler αποκλίσεων των εναλλακτικών συναρτήσεων κινδύνου από την αντίστοιχη της μηδενικής υπόθεσης, οι πίνακες των σημείων αποκοπής και οι πίνακες της εμπειρικής ισχύος για κάθε έλεγχο που πραγματοποιείται. Η διατριβή ολοκληρώνεται με την παράθεση της βιβλιογραφίας που χρησιμοποιήθηκε.

ABSTRACT

The hazard function, or the hazard rate function, holds a prominent position in survival analysis as it expresses the instantaneous rate at which events occur in the organisms under study or in the objects under study during a specific time period. In recent years, interest has increased in developing goodness-of-fit tests specifically for the hazard function. Such tests are particularly important because they allow researchers to determine whether the assumed distribution for the hazard rate function is suitable for the observed data. If the assumed distribution does not adequately describe the available data, researchers may be led to biased results and, consequently, incorrect conclusions. The purpose of this master's thesis is to propose a goodness-of-fit test for the hazard rate function, specifically for right-censored lifetime data. Specifically, in Chapter 1 (Introduction), the concept of goodness-of-fit tests is presented, as well as why they are important in statistics. Some useful definitions that will be used throughout this entire master's thesis are also presented. In Chapter 2 (The hazard function) focuses on the theory of the hazard function and its role in survival analysis. Specifically, the mathematical properties of the hazard function are examined, including its relationship with other concepts in survival analysis such as the survival function and the cumulative hazard function. In addition, special attention is paid to the form that a hazard function can take, through the study of its monotonicity. The chapter concludes by presenting the concept of censoring and the three types of censoring that can be found in lifetime data. In Chapter 3 (Estimation of hazard rate function) is dedicated to the estimation of the hazard rate function, with a reference to parametric estimation while focusing on non-parametric estimation. Specifically, a kernel-based estimator is presented and the mean, variance, and distribution of the estimator are thoroughly studied. In Chapter 4 (Goodness-of-Fit test), a new goodness-of-fit test for the hazard rate function is introduced. Initially, the null and alternative hypotheses are defined, and then the test statistic based on the integrated squared error is constructed. In the next step, its distribution is studied using probability theory, as well as formulas for moments, derivatives, products, and convolutions. Then, in Chapter 5 (Simulations),

the empirical power function of the test is determined in special cases of alternative hypotheses, whose parameters are chosen so that the Kullback-Leibler divergence from the null hypothesis is increasingly deviant. Subsequently, the critical region is determined, as well as other parameters that are required to apply the test in practice. The results are then confirmed via simulation study using the R programming language. In Chapter 6 (Conclusion - Topic for Further Research), a topic is presented for further research. Finally, in Appendix A' (Tables), the tables resulting from the implementation of test using the R software are included. Specifically, the tables of Kullback-Leibler divergences of the alternative hazard functions from the corresponding hazard function under the null hypothesis, the tables with the cut-off points as well as those with the empirical power of the test. The thesis is concluded by providing the bibliography that was used.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη	i
Abstract	iii
1 Εισαγωγή	5
2 Η συνάρτηση κινδύνου	11
2.1 Συνεχής μονοδιάστατη περίπτωση	11
2.1.1 Η συνάρτηση κινδύνου εκφραζόμενη μέσω των $f(x)$, $F(x)$ και $S(x)$	13
2.1.2 Οι $f(x)$, $F(x)$ και $S(x)$ εκφραζόμενες μέσω της συνάρτη- σης κινδύνου	14
2.1.3 Αθροιστική συνάρτηση κινδύνου	16
2.2 Μορφές της συνάρτησης κινδύνου	18
2.2.1 Κατανομές μονότονων συναρτήσεων κινδύνου	18
2.2.2 Κατανομές μη μονότονων συναρτήσεων κινδύνου	23
2.3 Λογοκριμένες και περικομμένες παρατηρήσεις	29
2.3.1 Λογοκρισία (Censoring)	29
2.3.2 Περικομμένες παρατηρήσεις (Truncated observations)	33
3 Εκτίμηση της συνάρτησης κινδύνου	35
3.1 Παραμετρική εκτίμηση της συνάρτησης κινδύνου	35
3.1.1 Συνάρτηση πιθανοφάνειας	36
3.1.2 Εκτίμηση μέγιστης πιθανοφάνειας για τη συνάρτηση κινδύνου	38

3.2	Μη παραμετρική εκτίμηση της συνάρτησης κινδύνου	39
3.2.1	Εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο των πυρήνων	39
3.2.2	Μέση τιμή και διακύμανση του εκτιμητή $\hat{\lambda}(x)$	45
3.2.3	Ασυμπτωτική κατανομή του εκτιμητή $\hat{\lambda}(x)$	59
4	Έλεγχος καλής προσαρμογής	85
4.1	Διατύπωση των υποθέσεων και της στατιστικής συνάρτησης . . .	86
4.1.1	Διατύπωση των υποθέσεων	86
4.1.2	Διατύπωση της στατιστικής συνάρτησης	87
4.2	Μελέτη της κατανομής της στατιστικής συνάρτησης	88
5	Προσομοιώσεις	127
5.1	Υλοποίηση του ελέγχου	127
5.2	Αποτελέσματα των προσομοιώσεων	130
5.2.1	Αποτελέσματα του ελέγχου για εναλλακτικές συναρτήσεις κινδύνου από την Εκθετική κατανομή	132
5.2.2	Αποτελέσματα του ελέγχου για εναλλακτικές συναρτήσεις κινδύνου από την Weibull κατανομή	132
5.2.3	Αποτελέσματα του ελέγχου για εναλλακτικές συναρτήσεις κινδύνου από την Γάμμα κατανομή	134
6	Επίλογος - Θέμα προς περαιτέρω έρευνα	137
A'	Πίνακες	141
A'.1	Αποκλίσεις Kullback-Leibler	141
A'.2	Σημεία αποκοπής	142
A'.3	Ισχύς για εναλλακτικές από την Εκθετική κατανομή	144
A'.4	Ισχύς για εναλλακτικές από την Weibull κατανομή	148
A'.5	Ισχύς για εναλλακτικές από την Γάμμα κατανομή	152

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Η μαθηματική στατιστική και συγκεκριμένα η περιοχή ανάπτυξης ελέγχου υποθέσεων καλής προσαρμογής (Goodness-of-Fit testing) είναι ένας ιδιαίτερος κλάδος της στατιστικής με κύριο αντικείμενο έρευνας την καταλληλότητα της προσαρμογής εκτιμώμενων συναρτήσεων (όπως για παράδειγμα πυκνότητας πιθανότητας, κινδύνου και επιβίωσης) με βάση ένα δείγμα από τον υπάρχοντα πληθυσμό, για την περιγραφή σημαντικών ιδιοτήτων του πληθυσμού, όπως για παράδειγμα την κατανομή του. Μαζί με την ανάπτυξη της κατάλληλης στατιστικής θεωρίας, η οποία θεμελιώνει αναλυτικά τις στατιστικές ιδιότητες των αναπτυχθέντων τεστ, η απόδοση τους, συνήθως, επιβεβαιώνεται και αριθμητικά με δεδομένα από γνωστές κατανομές ή μέσω της εφαρμογής τους σε πραγματικά δεδομένα. Το τελικό αποτέλεσμα είναι ο ερευνητής, με χρήση του τεστ, να μπορεί να αποφανθεί με ικανοποιητικό βαθμό σιγουριάς για την καταλληλότητα της υιοθέτησης της συγκεκριμένης κατανομής.

Στην ανάλυση επιβίωσης μία από τις σημαντικότερες συναρτήσεις προς μελέτη, μαζί με τη συνάρτηση επιβίωσης, είναι η συνάρτηση κινδύνου ή συνάρτηση ρυθμού επικινδυνότητας. Αυτό διότι η συνάρτηση κινδύνου εκφράζει τον στιγμιαίο κίνδυνο αποτυχίας ενός οργανισμού ή ενός αντικειμένου και επομένως περιέχει πληροφορίες οι οποίες είναι πολύ χρήσιμες στις μελέτες που αφορούν την επιβίωση (survival studies). Επιπλέον, στις μελέτες αυτές, οι ερευνητές ενδιαφέρονται να ελέγξουν αν μία συγκεκριμένη συνάρτηση κινδύνου μπορεί να εξηγήσει τον ρυθμό αποτυχίας ενός συγκεκριμένου συνόλου δεδομένων με την πάροδο του χρόνου ή να ελέγξουν αν δύο ή περισσότερα υποσύνολα ενός πληθυσμού έχουν την ίδια συνάρτηση ρυθμού επικινδυνότητας. Η μέθοδος ελέγχου της καταλληλότητας της συνάρτησης κινδύνου είναι ένα πολύ ισχυρό εργαλείο της μαθηματικής στατιστικής προκειμένου να βεβαιώσει την εγχυρότητα του εκτιμώμενου μηχανισμού που εκφράζει τον ρυθμό αποτυχίας (χρόνου μέχρι το συμβάν) για τα διαθέσιμα δεδομένα. Χρησιμοποιείται ευρύτατα στην καθημερινότητα σε πλειάδα εφαρμογών, για παράδειγμα στην φαρμακοκινητική και φαρμακοδυναμική (PKPD) για

Κεφάλαιο 1

τη μοντελοποίηση της υποκείμενης σχέσης μεταξύ του χρόνου μέχρι το συμβάν και της έκθεσης στο φάρμακο, καθώς επίσης και σε άλλους ερευνητικούς τομείς, όπως στην ιατρική, στη μηχανική και στα χρηματοοικονομικά.

Συχνά, τα υπό μελέτη δεδομένα είναι λογοκριμένα, δηλαδή οι παρατηρήσεις δεν εκφράζουν όλες την ολική - πλήρη διάρκεια των γεγονότων των οποίων έχουν μετρήσει. Καθώς μία μελέτη επιβίωσης αποτελείται από αρχή και τέλος, οι ερευνητές μπορούν να γνωρίζουν μόνο εντός αυτών των χρονικών ορίων εάν το γεγονός συνέβη ή όχι στους υπό μελέτη οργανισμούς. Εκτός αυτών των ορίων, οι ερευνητές δεν μπορούν να αποφανθούν για την κατάσταση των οργανισμών και επομένως θεωρούνται λογοκριμένοι. Στο πλαίσιο αυτό, υπάρχουν τρία είδη λογοκρισίας. Η δεξιά λογοκρισία κατά την οποία ο οργανισμός θεωρείται ότι λογοκρίνεται αν επιβιώσει καθ' όλη τη διάρκεια της έρευνας, η αριστερή λογοκρισία κατά την οποία το υπό μελέτη γεγονός έχει συμβεί στον οργανισμό πριν αυτός εισέλθει στην έρευνα και τέλος η λογοκρισία σε διαστήματα κατά την οποία ο οργανισμός εξετάζεται σε τακτά χρονικά διαστήματα και θεωρείται λογοκριμένος εάν το υπό μελέτη γεγονός του συμβεί μεταξύ δύο χρονικών σημείων παρακολούθησης.

Στην παρούσα μεταπτυχιακή διατριβή σκοπός μας είναι να προτείνουμε και να επικεντρωθούμε κυρίως στην ακριβή ποσοτικοποίηση των στατιστικών ιδιοτήτων ενός ελέγχου καλής προσαρμογής για τη συνάρτηση κινδύνου, μελετώντας την κατανομή της στατιστικής συνάρτησης. Η στατιστική συνάρτηση προκύπτει μέσω του ολοκληρωμένου τετραγωνικού σφάλματος (Integrated Squared Error) και αποτελεί το ολοκλήρωμα της τετραγωνικής διαφοράς μεταξύ της πραγματικής συνάρτησης κινδύνου και διάφορων γνωστών συναρτήσεων κινδύνου υπό τη μηδενική υπόθεση. Στη συνέχεια θα μελετήσουμε θεωρητικά την κατανομή της στατιστικής συνάρτησης ελέγχου που έχουμε προτείνει, δίνοντάς μας τη δυνατότητα να προσεγγίσουμε την πιθανότητα σωστής διάγνωσης της απόκλισης της πραγματικής κατανομής από την υποτιθέμενη, όσο η απόσταση μεταξύ τους αυξάνεται. Επιπλέον, σε συνέχεια των προηγούμενων αποτελεσμάτων, θα προσδιοριστεί η συνάρτηση ισχύος του παραγόμενου τεστ, η οποία θα δώσει αυξημένη διαγνωστική ισχύ σε μία πλειάδα διαφορετικών τύπων εναλλακτικών υποθέσεων

Η πραγματική συνάρτηση κινδύνου που αναφέρθηκε στην προηγούμενη παράγραφο, καθώς άγνωστη, εκτιμάται με τη μέθοδο των πυρήνων. Ένας εκτιμητής πυρήνα αποτελεί τη συνέλιξη μίας συνάρτησης εξομάλυνσης, του λεγόμενου πυρήνα (Kernel), και ενός εμπειρικού εκτιμητή που επιλέγονται με τέτοιο τρόπο ώστε να παράγουν έναν ομαλό εκτιμητή μίας συνάρτησης. Η συνάρτηση αυτή μπορεί, για παράδειγμα, να είναι μία συνάρτηση πυκνότητας πιθανότητας ή μία συνάρτηση κινδύνου.

Ορισμός 1.0.1. Ως πυρήνας (Kernel) ορίζεται μια συνάρτηση $K(z) : \mathbb{R} \rightarrow \mathbb{R}$

Κεφάλαιο 1

η οποία ικανοποιεί τις παρακάτω συνθήκες

1. $\int K(z) dz = 1$.
2. $K(z) \geq 0$ για κάθε z , το οποίο σε συνδυασμό με την πρώτη συνθήκη υποδηλώνει ότι η $K(z)$ είναι συνάρτηση πυκνότητας πιθανότητας.
3. Είναι συμμετρική συνάρτηση και κατά συνέπεια ισχύει ότι $K(z) = K(-z)$. Στην περίπτωση αυτή όλες οι ροπές της ($\mu_i(K) = \int z^i K(z) dz, i = 1, \dots, n$) με i - περιττό είναι ίσες με μηδέν.
4. Η τάξη της ορίζεται από την πρώτη μη μηδενική ροπή. Συμμετρικοί πυρήνες έχουν ζυγό αριθμό τάξης.
5. Είναι πυρήνας “ανώτερης τάξης” εάν η τάξη του είναι μεγαλύτερη του 2. Τέτοιου είδους πυρήνες δεν αποτελούν συναρτήσεις πυκνότητας πιθανότητας, καθώς θα δέχονται και αρνητικές τιμές.

Στον Πίνακα 1.1 παρατίθενται μερικές συναρτήσεις πυρήνων και στο Σχήμα 1.1 η σχηματική απεικόνισή τους.

Για την υλοποίηση ενός εκτιμητή πυρήνα μίας συνάρτησης απαιτείται ο προσδιορισμός του κατάλληλου εύρους ζώνης h . Το εύρος ζώνης, στην πλειονότητα των περιπτώσεων, είναι άγνωστο και για τον λόγο αυτόν η επιλογή του πραγματοποιείται μέσα από μία πληθώρα μεθόδων που έχουν προταθεί στη βιβλιογραφία. Ο ορισμός του δίνεται ως εξής.

Ορισμός 1.0.2. Ως εύρος ζώνης (*Bandwidth*) ορίζεται μια παράμετρος η οποία προσδιορίζει το ποσοστό της εξομάλυνσης που εφαρμόζεται στα δεδομένα και καθορίζει το εύρος της συνάρτησης πυρήνα. Συμβολίζεται με h .

Μία πιθανή διαδικασία εύρεσης του εύρους ζώνης, και λιγότερο προτιμητέα, είναι η “οπτική” επιλογή του. Η διαδικασία αυτή πραγματοποιείται ελέγχοντας οπτικά μέσω γραφικών παραστάσεων τον αντίστοιχο εκτιμητή για διαφορετικές επιλογές του εύρους ζώνης ξεκινώντας από μία μεγάλη τιμή και συνεχίζοντας με μικρότερες τιμές, έτσι ώστε το τελικό αποτέλεσμα να είναι και το πλησιέστερο γραφικά στην πραγματική συνάρτηση.

Η διαδικασία επιλογής η οποία όμως προτιμάται περισσότερο είναι εκείνη κατά την οποία το εύρος ζώνης επιλέγεται αυτόματα από τα δεδομένα. Ένας από τους λόγους είναι ότι η οπτική επιλογή του εύρους ζώνης μπορεί να αποβεί εξαιρετικά χρονοβόρα. Άλλος λόγος είναι το ότι σε πολλές περιπτώσεις, ο ερευνητής δεν

Κεφάλαιο 1

Όνομασία πυρήνα	Μαθηματική έκφραση
Epanechnikov Kernel	$K(z) = \frac{3}{4} (1 - z^2), \quad z \leq 1$
Gaussian Kernel	$K(z) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R}$
Triangular Kernel	$K(z) = (1 - z), \quad z \leq 1$
Biweight Kernel	$K(z) = \frac{15}{16} (1 - z^2)^2, \quad z \leq 1$

Πίνακας 1.1: Συναρτήσεις πυρήνων, $K(z)$

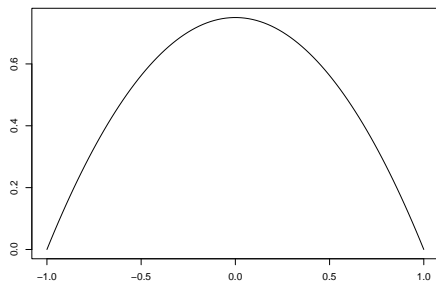
έχει την εκ των προτέρων γνώση για τη δομή των δεδομένων, γεγονός το οποίο καθιστά εξαιρετικά δύσκολο την επιλογή εύρους ζώνης που να δίνει εκτίμηση πλησιέστερη στην πραγματική συνάρτηση.

Όσο μεγαλύτερη είναι η τιμή του εύρους ζώνης τόσο περισσότερη θα είναι και η εξομάλυνση (Smoothing) της καμπύλης, γεγονός το οποίο εγκυμονεί τον κίνδυνο της υπέρ-εξομάλυνσης (Oversmoothing). Ο κίνδυνος αυτός ελαχιστοποιείται μειώνοντας το εύρος ζώνης. Αντίθετα, πολύ μικρή τιμή του εύρους ζώνης καθιστά πιο “μυτερές” τις κορυφές της καμπύλης του εκτιμητή, γεγονός το οποίο ερμηνεύεται με τον όρο της υπό-εξομάλυνσης (Undersmoothing).

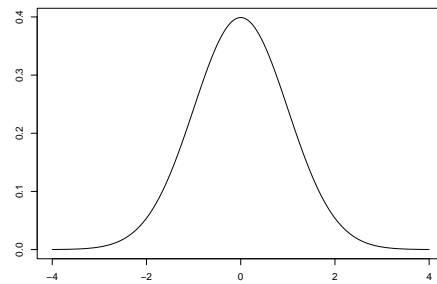
Στα δύο πρώτα κεφάλαια που ακολουθούν παρουσιάζονται εκτενώς, μέσω βιβλιογραφικής ανασκόπησης, θεωρητικά στοιχεία που αφορούν τη συνάρτηση κινδύνου, τη μη παραμετρική εκτίμησή της και τη λογοκρισία. Επιπλέον, σε επόμενα κεφάλαια, ορίζεται αναλυτικά ένας νέος έλεγχος καλής προσαρμογής για τη συνάρτηση του ρυθμού επικινδυνότητας και υλοποιείται με τη χρήση κατάλληλων

Κεφάλαιο 1

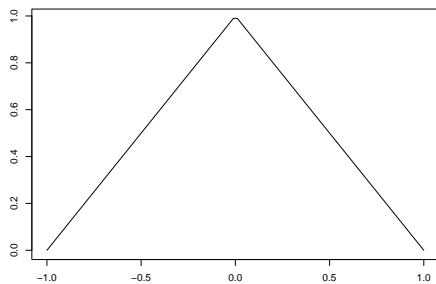
μεθόδων της στατιστικής συμπερασματολογίας. Τέλος μέσω προσομοιώσεων, χρησιμοποιώντας το στατιστικό πακέτο της R, επιβεβαιώνονται τα θεωρητικά αποτελέσματα που προέκυψαν για την κατανομή της ελεγχοσυνάρτησης, ενώ παρατίθενται και τα παραγόμενα αποτελέσματα από την υλοποίηση της συνάρτησης ισχύος προκειμένου να εξετασθεί η αξιοπιστία του ελέγχου.



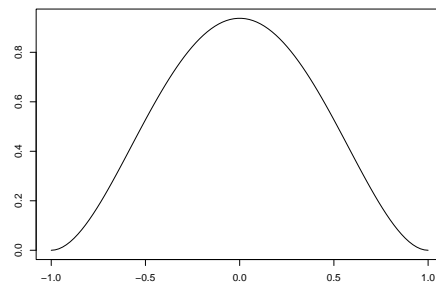
(α') Epanechnikov Kernel



(β') Gaussian Kernel



(γ') Triangular Kernel



(δ') Biweight Kernel

Σχήμα 1.1: Σχηματική απεικόνιση των πυρήνων του Πίνακα 1.1 .

Κεφάλαιο 1

ΚΕΦΑΛΑΙΟ 2

Η ΣΥΝΑΡΤΗΣΗ ΚΙΝΔΥΝΟΥ

2.1 Συνεχής μονοδιάστατη περίπτωση

Στην ανάλυση επιβίωσης, μείζον θέμα προς μελέτη αποτελεί η στιγμιαία πιθανότητα να συμβεί ένα γεγονός στο χρονικό διάστημα $(x, x + \Delta x)$. Δηλαδή το υπό μελέτη γεγονός να συμβεί την επόμενη χρονική στιγμή δεδομένου ότι δεν έχει συμβεί μέχρι τη χρονική στιγμή x . Μαθηματικά η στιγμιαία πιθανότητα αυτή μπορεί να γραφεί ως:

$$\lambda(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x < X \leq x + \Delta x | X > x)}{\Delta x} \quad \text{για } x \geq 0, \quad (2.1)$$

όπου X είναι μία τυχαία μεταβλητή (τ.μ.) που παριστάνει τον χρόνο επιβίωσης ενός ατόμου ή αντικειμένου που συμμετέχει στη μελέτη. Η συνάρτηση $\lambda(x)$ ονομάζεται συνάρτηση κινδύνου ή συνάρτηση ρυθμού επικινδυνότητας.

Έστω $f(x)$ η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.), $F(x)$ η αθροιστική συνάρτηση κατανομής (α.σ.κ.) και $S(x)$ η συνάρτηση επιβίωσης (σ.ε.) της τ.μ. X .

Στο χρονικό διάστημα $(x, x + \Delta x)$, όπου $x \geq 0$, ισχύει ότι:

$$P(x < X \leq x + \Delta x | X > x) \approx f(\Delta x | X > x) \Delta x. \quad (2.2)$$

Για την $f(\Delta x | X > x)$ προκύπτει ότι:

$$\begin{aligned} f(\Delta x | X > x) &= \frac{\partial F(\Delta x | X > x)}{\partial \Delta x} \\ &= \frac{\partial}{\partial \Delta x} \left(\frac{F(x + \Delta x) - F(x)}{1 - F(x)} \right) \\ &= \frac{f(x + \Delta x)}{1 - F(x)}, \end{aligned}$$

ή ισοδύναμα ότι:

$$f(\Delta x|X > x) = \frac{f(x + \Delta x)}{S(x)}. \quad (2.3)$$

Αντικαθιστώντας την (2.3) στην (2.2) λαμβάνουμε ότι:

$$P(x < X \leq x + \Delta x|X > x) \approx \frac{f(x + \Delta x)}{S(x)}. \quad (2.4)$$

Στη συνέχεια λόγω του αποτελέσματος (2.4), η (2.1) γίνεται:

$$\begin{aligned} \lambda(x) &= \lim_{\Delta x \rightarrow 0} \frac{P(x < X \leq x + \Delta x|X > x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\frac{f(x + \Delta x)}{S(x)} \Delta x}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x)}{S(x)}, \end{aligned}$$

όπου τελικά προκύπτει ότι:

$$\lambda(x) = \frac{f(x)}{S(x)}, \quad S(x) > 0. \quad (2.5)$$

Μία άλλη πιθανή ερμηνεία της συνάρτησης κινδύνου είναι ότι παριστάνει τον ρυθμό με τον οποίο συμβαίνουν γεγονότα (αποτυχίες) στη μονάδα του χρόνου σε σχέση με το τμήμα του πληθυσμού που δεν έχει ακόμα αποτύχει. Γενικά όσο μεγαλύτερη είναι η τιμή της $\lambda(x)$, τόσο πιο πιθανό είναι να συμβεί το γεγονός που εκφράζεται από την X . Όπως θα δούμε και παρακάτω, η συνάρτηση κινδύνου δίνει μη αρνητικές τιμές. Οι τιμές αυτές μπορούν να είναι μεγαλύτερες της μονάδας καθώς, δοθείσης της φύσης της (2.1), η $\lambda(x)$ μπορεί μόνο προσεγγιστικά να συνδεθεί με πιθανότητα.

Κάθε συνάρτηση κινδύνου ικανοποιεί τις παρακάτω συνθήκες:

$$\begin{aligned} \lambda(x) &\geq 0, \\ \int_0^\infty \lambda(x) dx &= \infty. \end{aligned}$$

Η πρώτη συνθήκη είναι προφανής καθώς $f(x) \geq 0$ και $S(x) > 0$, ενώ η δεύτερη συνθήκη, δοθέντος ότι $\lambda(x) = -\partial \ln S(x)/\partial x$, προκύπτει ως εξής:

$$\begin{aligned}
\int_0^{\infty} \lambda(x) dx &= \int_0^{\infty} \left(-\frac{\partial \ln S(x)}{\partial x} \right) dx \\
&= - \int_0^{\infty} \partial \ln S(x) dx \\
&= - [\ln S(x)]_0^{\infty} \\
&= - \ln S(\infty) + \ln S(0) \\
&= \infty,
\end{aligned}$$

καθώς $S(0) = 1$.

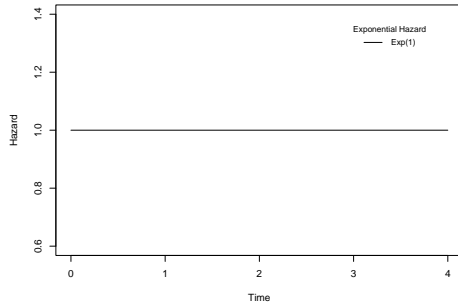
Η συνάρτηση ρυθμού επικινδυνότητας μπορεί να μοντελοποιηθεί χρησιμοποιώντας μια ποικιλία στατιστικών κατανομών, όπως για παράδειγμα είναι η Εκθετική, η Weibull και η Γάμμα. Καθεμία από αυτές τις κατανομές διαθέτει συνάρτηση κινδύνου η οποία αποτυπώνει διαφορετικά μοτίβα στα δεδομένα επιβίωσης, όπως φαίνεται και στο Σχήμα 2.1. Η εκθετική κατανομή έχει σταθερή συνάρτηση κινδύνου η οποία δεν εξαρτάται από τον χρόνο, ενώ οι κατανομές Weibull και Γάμμα διαθέτουν συναρτήσεις κινδύνου οι οποίες μπορεί είτε να αυξάνουν, είτε να φθίνουν, είτε να παραμένουν σταθερές, ανάλογα με την τιμή της παραμέτρου μορφής (shape parameter). Η επιλογή της κατανομής βασίζεται, συνήθως, στις υποκείμενες υποθέσεις σχετικά με τη φύση των χρόνων επιβίωσης και μπορεί να έχει σημαντικό αντίκτυπο στα αποτελέσματα της εκάστοτε ανάλυσης.

2.1.1 Η συνάρτηση κινδύνου εκφραζόμενη μέσω των $f(x)$, $F(x)$ και $S(x)$

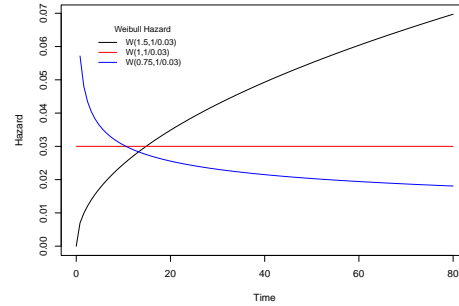
Η συνάρτηση κινδύνου που δίνεται στη σχέση (2.5), εκφραζόμενη μέσω της σ.π.π. $f(x)$, της α.σ.κ. $F(x)$ και της σ.ε. $S(x)$ αντίστοιχα, δίνεται από τις παρακάτω σχέσεις:

$$\begin{aligned}
\lambda(x) &= \frac{f(x)}{\int_x^{\infty} f(y) dy} \\
&= \frac{F'(x)}{1 - F(x)} \\
&= -\frac{S'(x)}{S(x)} = -\frac{\partial \ln S(x)}{\partial x}.
\end{aligned}$$

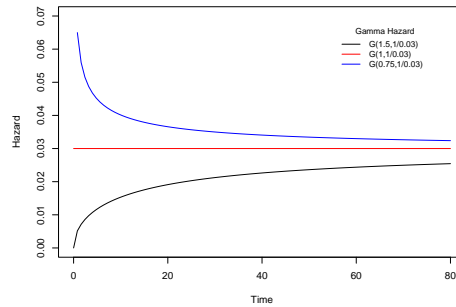
Παρατηρούμε ότι ενώ η σ.ε. δίνει πιθανότητα, η συνάρτηση κινδύνου δίνει τον ρυθμό με τον οποίο συμβαίνει το γεγονός που μελετάται, συναρτήσει του χρόνου.



(α) Εκθετική



(β') Weibull



(γ') Γάμμα

Σχήμα 2.1: Σχηματική απεικόνιση συναρτήσεων κινδύνου από την Εκθετική, τη Weibull και τη Γάμμα κατανομή.

2.1.2 Οι $f(x)$, $F(x)$ και $S(x)$ εκφραζόμενες μέσω της συνάρτησης κινδύνου

Αρχικά αποδεικνύεται ότι:

$$\int_0^x \lambda(u) du = [-\ln S(u)]_0^x = -\ln S(x) + \ln S(0) = -\ln S(x) + \ln 1 = -\ln S(x).$$

Επομένως, η σ.π.π. $f(x)$, η α.σ.κ. $F(x)$ και η σ.ε. $S(x)$ εκφραζόμενες μέσω της συνάρτησης κινδύνου $\lambda(x)$, δίνονται από τις παρακάτω σχέσεις:

$$S(x) = \exp \left\{ - \int_0^x \lambda(u) du \right\},$$

$$F(x) = 1 - S(x) = 1 - \exp \left\{ - \int_0^x \lambda(u) du \right\}, \quad (2.6)$$

και

$$\begin{aligned} f(x) &= \frac{\partial F(x)}{\partial x} = \frac{\partial [1 - \exp \{ - \int_0^x \lambda(u) du \}]}{\partial x} \\ &= \frac{\partial [- \exp \{ - \int_0^x \lambda(u) du \}]}{\partial x} \\ &= - \exp \left\{ - \int_0^x \lambda(u) du \right\} \frac{\partial [- \int_0^x \lambda(u) du]}{\partial x} \\ &= \lambda(x) \exp \left\{ - \int_0^x \lambda(u) du \right\}. \end{aligned} \quad (2.7)$$

Στη συνέχεια, δοθέντος τριών μοντέλων συναρτήσεων κινδύνου, εξετάζεται η αντίστοιχη μορφή της σ.π.π. για καθένα από αυτά με σκοπό την εύρεση της κατανομής από την οποία προέρχονται τα δεδομένα χρόνων ζωής. Ειδικότερα, τα μοντέλα τα οποία θεωρούνται είναι τα εξής:

1. Μοντέλο με σταθερή συνάρτηση κινδύνου (Constant Hazard Rate Model).
2. Μοντέλο με εκθετική συνάρτηση κινδύνου (Exponential Hazard Rate Model).
3. Μοντέλο με γραμμική συνάρτηση κινδύνου (Linear Hazard Rate Model).

Στο πλαίσιο αυτό, χρησιμοποιώντας τη σχέση (2.7), τα αποτελέσματα που προκύπτουν είναι τα ακόλουθα, ενώ η σχηματική απεικόνιση των τριών μοντέλων που αναφέρθηκαν παραπάνω δίνεται στο Σχήμα 2.2.

1. Για το μοντέλο με σταθερή συνάρτηση κινδύνου δίνεται ότι:

$$\lambda(x) = \lambda, \quad \text{για κάθε } x \geq 0, \lambda > 0.$$

Κατά συνέπεια, η σ.π.π. γίνεται:

$$f(x) = \lambda \exp \left\{ - \int_0^x \lambda du \right\} = \lambda e^{-\lambda x}.$$

Επομένως προκύπτει ότι αντιστοιχεί στη σ.π.π. της Εκθετικής κατανομής.

2. Για το μοντέλο με εκθετική συνάρτηση κινδύνου δίνεται ότι:

$$\lambda(x) = e^x, \quad x \geq 0.$$

Κατά συνέπεια, η σ.π.π. γίνεται:

$$f(x) = e^x \exp \left\{ - \int_0^x e^u du \right\} = e^x \exp \{ e^{-x} + 1 \}.$$

Επομένως προκύπτει ότι αντιστοιχεί στη σ.π.π. της κατανομής Gompertz.

3. Για το μοντέλο με γραμμική συνάρτηση κινδύνου δίνεται ότι:

$$\lambda(x) = a + bx, \quad \text{για κάθε } x \geq 0, a \geq 0, b \geq 0.$$

Κατά συνέπεια, η σ.π.π. γίνεται:

$$f(x) = (a + bx) \exp \left\{ - \int_0^x (a + bu) du \right\} = (a + bx) \exp \left\{ -ax - \frac{b}{2}x^2 \right\}.$$

Επομένως προκύπτει ότι αντιστοιχεί στη σ.π.π. της κατανομής Rayleigh.

2.1.3 Αθροιστική συνάρτηση κινδύνου

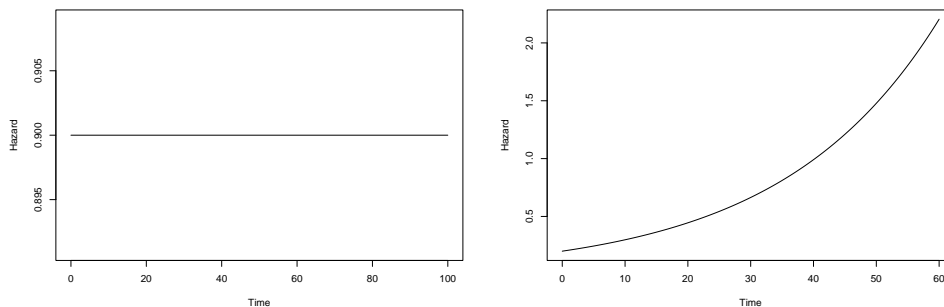
Η αθροιστική συνάρτηση κινδύνου μπορεί να θεωρηθεί ότι παρέχει τον συνολικό (συσσωρευμένο) κίνδυνο εμφάνισης του υπό μελέτη γεγονότος, ο οποίος κίνδυνος έχει αποκτηθεί καθ' όλη τη διάρκεια έως τη χρονική στιγμή x . Παρατηρήστε ότι, ενώ η συνάρτηση κινδύνου $\lambda(x)$ μπορεί να αυξάνεται ή να μειώνεται με το χρόνο, η αθροιστική συνάρτηση κινδύνου μπορεί μόνο να αυξάνεται ή να παραμένει σταθερή. Τούτο έχει νόημα καθώς ο συνολικός (συσσωρευμένος) κίνδυνος να βιώσει κάποιος το γεγονός που μελετάται θα πρέπει πάντα να αυξάνεται (ή να παραμένει ο ίδιος) με την πάροδο του χρόνου.

Η αθροιστική συνάρτηση κινδύνου συμβολίζεται με $\Lambda(x)$ και δίνεται από την σχέση:

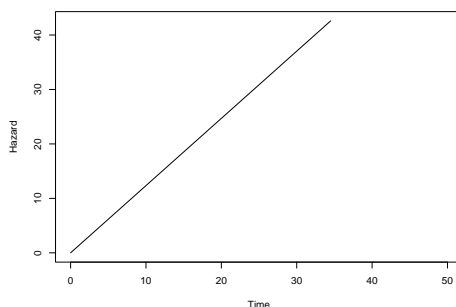
$$\Lambda(x) = \int_0^x \lambda(u) du.$$

Επιπλέον, έχει τις εξής ιδιότητες:

$$\begin{aligned} \Lambda(0) &= 0, \\ \lim_{x \rightarrow \infty} \Lambda(x) &= \infty. \end{aligned}$$



(α') Μοντέλο με σταθερή συνάρτηση κινδύνου (β') Μοντέλο με εκθετική συνάρτηση κινδύνου



(γ') Μοντέλο με γραμμική συνάρτηση κινδύνου ($a = 0, b = 0.9$)

Σχήμα 2.2: Σχηματική απεικόνιση των μοντέλων με σταθερή, εκθετική και γραμμική συνάρτηση κινδύνου

Η σχέση της αθροιστικής συνάρτησης κινδύνου με τη σ.π.π $f(x)$, την α.σ.κ. $F(x)$ και τη σ.ε. $S(x)$ δίνεται μέσω των παρακάτω σχέσεων:

$$f(x) = -\frac{\partial \exp[-\Lambda(x)]}{\partial x}$$

$$F(x) = 1 - \exp[-\Lambda(x)]$$

$$S(x) = \exp[-\Lambda(x)].$$

Από την άλλη, η σ.π.π. $f(x)$, η α.σ.κ. $F(x)$ και η σ.ε. $S(x)$ εκφραζόμενες μέσω της αθροιστικής συνάρτησης κινδύνου δίνονται από τις παρακάτω σχέσεις:

$$\Lambda(x) = -\ln \left\{ \int_x^\infty f(y) dy \right\} = -\ln [1 - F(x)] = -\ln S(x).$$

2.2 Μορφές της συνάρτησης κινδύνου

2.2.1 Κατανομές μονότονων συναρτήσεων κινδύνου

Η σχηματική απεικόνιση κατανομών στις οποίες ο ρυθμός επικινδυνότητας αυξάνεται, όπως επίσης και κατανομών στις οποίες ο ρυθμός επικινδυνότητας μειώνεται, περιέχουν ιδιαίτερο ενδιαφέρον όπως θα δούμε και στη συνέχεια. Η Εκθετική κατανομή (συνεχής κατανομή) και η Γεωμετρική κατανομή (διακριτή κατανομή) που έχουν σταθερές συναρτήσεις κινδύνου, ανήκουν και στις δύο τάξεις που αναφέρθηκαν προηγουμένως. Η υπόθεση ότι μια κατανομή χρόνου ζωής έχει μονότονη συνάρτηση κινδύνου είναι αρκετά ισχυρή, αλλά τέτοιου είδους κατανομές έχουν πολλές σημαντικές και ενδιαφέρουσες ιδιότητες. Η παρούσα υποενοότητα επικεντρώνεται στην περίπτωση των συνεχών κατανομών και το περιεχόμενό της βασίζεται στα συγγράμματα των Barlow and Proschan (1975, 1996) και Rinne (2014). Οι αποδείξεις των θεωρημάτων που ακολουθούν πραγματοποιούνται εκ νέου και παρατίθενται στην αναλυτική τους μορφή.

Στο εξής με τον όρο IHR (Increasing Hazard Rate) θα χαρακτηρίζουμε κατανομές με αύξων ρυθμό επικινδυνότητας και με τον όρο DHR (Decreasing Hazard Rate) κατανομές με φθίνων ρυθμό επικινδυνότητας.

Ορισμός 2.2.1. Έστω $F(\cdot)$ μία συνεχής κατανομή. Η $F(\cdot)$ είναι IHR (DHR), αν και μόνο αν η σχέση,

$$F(y|X > x) = \frac{F(x+y) - F(x)}{1 - F(x)}, \quad (2.8)$$

είναι αύξουσα (φθίνουσα) στο x για $y > 0$, όπου $x \geq 0$ τέτοιο ώστε $F(x) < 1$.

Θεώρημα 2.2.1. (Barlow and Proschan (1996)) Έστω ότι η κατανομή $F(\cdot)$ έχει σ.π.π. $f(\cdot)$. Τότε η $F(\cdot)$ είναι IHR (DHR) αν και μόνο αν η συνάρτηση κινδύνου $\lambda(\cdot)$ είναι αύξουσα (φθίνουσα).

Απόδειξη. Στον Ορισμό 2.2.1 αναφέρθηκε ότι η κατανομή $F(\cdot)$ είναι IHR (DHR) αν η ποσότητα (2.8) είναι αύξουσα (φθίνουσα) στο x .

Παρατηρούμε ότι:

$$\begin{aligned}
\lim_{y \rightarrow 0} \frac{F(y|X > x)}{y} &= \lim_{y \rightarrow 0} \frac{\frac{F(x+y)-F(x)}{1-F(x)}}{y} \\
&= \lim_{y \rightarrow 0} \frac{F(x+y) - F(x)}{y(1-F(x))} \\
&= \frac{1}{1-F(x)} \lim_{y \rightarrow 0} \frac{F(x+y) - F(x)}{y} \\
&= \frac{1}{1-F(x)} \frac{\partial F(x)}{\partial y} \\
&= \frac{f(x)}{1-F(x)} = \frac{f(x)}{S(x)} = \lambda(x).
\end{aligned}$$

Επομένως, αυτό που αρκεί ναδειχθεί είναι ότι, η $\lambda(x)$ είναι αύξουσα (φθίνουσα) στο x συνεπάγεται ότι η (2.8) είναι αύξουσα (φθίνουσα) στο x .

Έστω ότι η $\lambda(x)$ είναι αύξουσα. Επομένως για $x_1 \leq x_2$ έχουμε ότι:

$$\begin{aligned}
x_1 \leq x_2 &\Rightarrow \lambda(x_1) \leq \lambda(x_2) \\
&\Rightarrow \int_0^x \lambda(x_1 + u) du \leq \int_0^x \lambda(x_2 + u) du \\
&\Rightarrow \exp \left\{ - \int_{x_1}^{x_1+x} \lambda(u) du \right\} \geq \exp \left\{ - \int_{x_2}^{x_2+x} \lambda(u) du \right\} \\
&\Rightarrow \frac{F(x_1 + x) - F(x_1)}{1 - F(x_1)} \leq \frac{F(x_2 + x) - F(x_2)}{1 - F(x_2)}.
\end{aligned}$$

Με παρόμοιο τρόπο αποδεικνύεται ότι όταν η $\lambda(x)$ είναι φθίνουσα τότε και η (2.8) είναι φθίνουσα, γεγονός που ολοκληρώνει την απόδειξη. \square

Ορισμός 2.2.2. Η κατανομή $F(\cdot)$ είναι IHR (DHR) αν και μόνο αν η σ.ε. $S(y|x)$ είναι αύξουσα (φθίνουσα) στο x , για κάθε $y > 0, x \geq 0$, τέτοια ώστε $S(x) > 0$, όπου:

$$S(y|x) = \frac{S(x+y)}{S(x)} = \frac{\exp \left\{ - \int_0^{x+y} \lambda(u) du \right\}}{\exp \left\{ - \int_0^x \lambda(u) du \right\}} = \exp \left\{ - \int_x^{x+y} \lambda(u) du \right\}.$$

Το επόμενο θεώρημα δίνεται χωρίς απόδειξη.

Θεώρημα 2.2.2. (Barlow and Proschan (1996)) Η κατανομή $F(\cdot)$ είναι IHR (DHR) αν και μόνο αν ο λογάριθμος της συνάρτησης επιβίωσης είναι κοίλη (κυρτή) συνάρτηση.

Καθώς όμως $\Lambda(x) = -\ln S(x)$, το Θεώρημα 2.2.2 μπορεί να εκφραστεί σύμφωνα με την αθροιστική συνάρτηση κινδύνου μέσω του θεωρήματος που ακολουθεί.

Θεώρημα 2.2.3. (Rinne (2014)) Η κατανομή $F(\cdot)$ είναι IHR (DHR) αν και μόνο αν η $\Lambda(x) = \int_0^x \lambda(u) du$ είναι κυρτή (κοίλη) συνάρτηση.

Απόδειξη. Γνωρίζουμε ότι, $S(x) = 1 - F(x) = \exp[-\Lambda(x)]$.

Άρα,

$$\begin{aligned} \frac{F(x+y) - F(x)}{1 - F(x)} &= \frac{-(1 - F(x+y)) + (1 - F(x))}{1 - F(x)} \\ &= 1 - \frac{1 - F(x+y)}{1 - F(x)} \\ &= 1 - \frac{S(x+y)}{S(x)} \\ &= 1 - \frac{\exp[-\Lambda(x+y)]}{\exp[-\Lambda(x)]} \\ &= 1 - \exp[-\Lambda(x+y) + \Lambda(x)] \\ &= 1 - \exp\{-[\Lambda(x+y) - \Lambda(x)]\}. \end{aligned}$$

Επομένως η $F(\cdot)$ είναι IHR (DHR) αν και μόνο αν το $\Lambda(x+y) - \Lambda(x)$ αυξάνει (φθίνει) στο x , για κάθε $y > 0$. Δηλαδή η $F(\cdot)$ είναι IHR (DHR) αν και μόνο αν η $\Lambda(x)$ είναι κυρτή (κοίλη). \square

Παράδειγμα 2.2.4. (Rinne (2014)) Το παράδειγμα που ακολουθεί αφορά τη μελέτη IHR και DHR για τη Reduced Weibull κατανομή. Η Reduced Weibull κατανομή δεν είναι άλλη παρά η Weibull κατανομή με παράμετρο θέσης ίση με 0, παράμετρο κλίμακας ίση με 1 και παράμετρο μορφής ίση με k . Επομένως η σ.π.π. $f(x)$, η α.σ.κ. $F(x)$, η σ.ε. $S(x)$, η συνάρτηση κινδύνου $\lambda(x)$ και η αθροιστική συνάρτηση κινδύνου $\Lambda(x)$ της δίνονται στις παρακάτω σχέσεις:

$$\begin{aligned} f(x) &= kx^{k-1} \exp(-x^k) \quad \text{για } x \geq 0 \quad \text{και } k > 0, \\ F(x) &= 1 - \exp(-x^k), \\ S(x) &= \exp(-x^k), \\ \lambda(x) &= kx^{k-1}, \\ \Lambda(x) &= x^k. \end{aligned}$$

Ως προς τη συνάρτηση κινδύνου λαμβάνεται ότι:

$$\frac{\partial \lambda(x)}{\partial x} = k(k-1)x^{k-2} \begin{cases} \leq 0, & \text{για } 0 < k \leq 1, \text{ τότε είναι DHR,} \\ \geq 0, & \text{για } k \geq 1, \text{ τότε είναι IHR,} \end{cases}$$

και ως προς τον λογάριθμο της συνάρτησης επιβίωσης λαμβάνεται ότι:

$$\frac{\partial^2 \ln S(x)}{\partial x^2} = -k(k-1)x^{k-2} \begin{cases} \geq 0, & \text{για } 0 < k \leq 1, \text{ τότε είναι DHR,} \\ \leq 0, & \text{για } k \geq 1, \text{ τότε είναι IHR,} \end{cases}$$

όπου $\ln S(x) = -x^k$.

Το επόμενο λήμμα δίνεται χωρίς απόδειξη.

Λήμμα 2.2.5. (Barlow and Proschan (1996)) Αν η κατανομή $F(\cdot)$ είναι IHR (DHR), τότε η ποσότητα $[S(x)]^{\frac{1}{x}}$ είναι αύξουσα (φθίνουσα).

Από το Λήμμα 2.2.5 προκύπτει ότι:

$$S(x) \leq [S(x)]^{\frac{t}{x}}, \quad t > x,$$

και

$$\int_x^\infty u^r S(u) du \leq \int_x^\infty u^r [S(u)]^{\frac{u}{x}} du < \infty,$$

όταν $S(x) < 1$ και $r \geq 0$.

Επομένως οι IHR κατανομές έχουν πεπερασμένες ροπές για όλες τις τάξεις τους, ενώ οι DHR κατανομές, υποχρεωτικά, δεν πρέπει να έχουν πεπερασμένες ροπές.

Στο εξής με τον όρο HRA (Hazard Rate Average) θα αναφερόμαστε στον μέσο ρυθμό επικινδυνότητας μίας κατανομής, ενώ με τους όρους IHRA (Increasing Hazard Rate Average) και DHRA (Decreasing Hazard Rate Average) θα αναφερόμαστε σε κατανομές με αύξων μέσο ρυθμό επικινδυνότητας και κατανομές με φθίνων μέσο ρυθμό επικινδυνότητας αντίστοιχα. Η εισαγωγή του HRA μειώνει την απαίτηση της μονότονης πορείας του ρυθμού επικινδυνότητας όταν η μονοτονία του ζητείται κατά μέσο όρο.

Ορισμός 2.2.3. Μία συνεχής κατανομή είναι IHRA (DHRA) αν η συνάρτηση

$$HRA(x) = -\frac{1}{x} \ln S(x),$$

είναι αύξουσα (φθίνουσα) στο x με $x \geq 0$, ή ισοδύναμα αν

$$[S(x_1)]^{\frac{1}{x_1}} \begin{matrix} \geq \\ (\leq) \end{matrix} [S(x_2)]^{\frac{1}{x_2}}, \quad \text{για } 0 \leq x_1 \leq x_2.$$

Δοθέντος ότι το $-\ln S(x)$ αντιπροσωπεύει το $\Lambda(x) = \int_0^x \lambda(u) du$, όταν υπάρχει η $\lambda(x)$, τότε ισχύει ότι:

$$HRA(x) = \frac{1}{x} \int_0^x \lambda(u) du = \frac{\Lambda(x)}{x}. \quad (2.9)$$

Ορισμός 2.2.4. Μία συνεχής κατανομή είναι IHRA (DHRA) εάν

$$\frac{\Lambda(x_1)}{x_1} \underset{(\leq)}{\geq} \frac{\Lambda(x_2)}{x_2}, \quad \text{για } 0 \leq x_1 \leq x_2.$$

Βάσει της σχέσης (2.9), κατανομή με IHRA (DHRA) σημαίνει ότι:

$$\frac{\partial}{\partial x} \left[\frac{\Lambda(x)}{x} \right] \underset{(\leq)}{\geq} 0, \quad \text{για κάθε } x \geq 0,$$

ή

$$\lambda(x) \underset{(\leq)}{\geq} \frac{\Lambda(x)}{x}, \quad \text{για κάθε } x \geq 0.$$

Δηλαδή η $\lambda(x)$ πρέπει να είναι μεγαλύτερη (μικρότερη) από το $HRA(x)$ για όλα τα x .

Παράδειγμα 2.2.6. (Rinne (2014)) Για την Reduced Weibull κατανομή ισχύουν ότι:

$$\begin{aligned} \Lambda(x) &= x^k, \\ \lambda(x) &= kx^{k-1}, \\ \frac{\Lambda(x)}{x} &= HRA(x) = x^{k-1}. \end{aligned}$$

Επομένως,

$$\lambda(x) = kx^{k-1} \begin{cases} \leq HRA(x) = x^{k-1}, & \text{για } 0 < k \leq 1, \text{ τότε είναι DHRA και DHR,} \\ \geq HRA(x) = x^{k-1}, & \text{για } k \geq 1, \text{ τότε είναι IHRA και IHR.} \end{cases}$$

Είναι προφανές ότι μία IHRA κατανομή χαρακτηρίζεται από τη μείωση του $[S(x)]^{\frac{1}{x}}$ στο $[0, \infty)$, ενώ μία DHRA κατανομή χαρακτηρίζεται από την αύξηση του $[S(x)]^{\frac{1}{x}}$ στο $[0, \infty)$.

Τα επόμενα δύο θεωρήματα δίνονται χωρίς απόδειξη.

Θεώρημα 2.2.7. (Barlow and Proschan (1975)) *Μια κατανομή είναι IHRA (DHRA) αν και μόνο αν ισχύει ότι:*

$$S(ax) \underset{(\leq)}{\geq} [S(x)]^{\frac{1}{a}}, \quad \text{για } 0 < a < 1 \text{ και } x \leq 0.$$

Θεώρημα 2.2.8. (Barlow and Proschan (1975)) *Αν μία κατανομή είναι IHR (DHR) τότε είναι και IHRA (DHRA). Το αντίστροφο δεν ισχύει πάντα.*

Για παράδειγμα η $F(x) = (1 - e^{-x})(1 - e^{-cx})$, $x \geq 0, c > 1$ είναι IHRA αλλά όχι IHR.

2.2.2 Κατανομές μη μονότονων συναρτήσεων κινδύνου

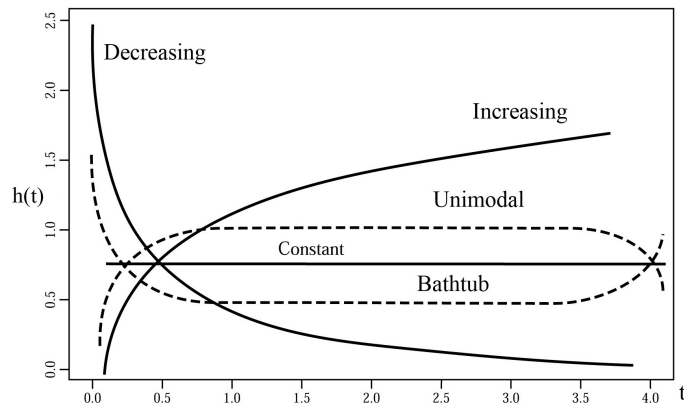
Πολλά φυσικά φαινόμενα παραθέτουν συναρτήσεις κινδύνου οι οποίες δεν είναι μονότονες. Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι συναρτήσεις κινδύνου οι οποίες στην αρχή φθίνουν και στη συνέχεια αυξάνουν (bathtub shape), όπως επίσης και συναρτήσεις κινδύνου που αρχικά αυξάνουν και στη συνέχεια φθίνουν (inverted (upside - down) bathtub shape). Οι συναρτήσεις κινδύνου που διαθέτουν bathtub (inverted bathtub) μορφή, παρουσιάζουν τρεις φάσεις. Στην πρώτη φάση ο ρυθμός επικινδυνότητας μειώνεται (αυξάνεται), στη μεσαία φάση παραμένει σταθερός και στην τρίτη και τελική φάση αυξάνεται (μειώνεται).

Μοντέλα με αυξανόμενο ρυθμό επικινδυνότητας συναντώνται όταν υπάρχει φυσική γήρανση ή φθορά. Συναρτήσεις κινδύνου που φθίνουν είναι λιγότερο συχνές, αλλά βρίσκουν περιστασιακή χρήση όταν υπάρχει πολύ πρόωμη πιθανότητα αποτυχίας, όπως σε ορισμένους τύπους ηλεκτρονικών συσκευών ή σε ασθενείς που υφίστανται ορισμένους τύπους μεταμοσχεύσεων.

Τις περισσότερες φορές συναρτήσεις κινδύνου σε σχήμα bathtub είναι κατάλληλες σε πληθυσμούς που παρακολουθούνται από τη στιγμή της γέννησής τους. Ομοίως, ορισμένοι κατασκευαστικοί εξοπλισμοί μπορούν να παρουσιάσουν πρόωρη αποτυχία (βλάβη) εξαιτίας ελλατωματικών εξαρτημάτων, η οποία ακολουθείται από έναν σταθερό ρυθμό κινδύνου που στα τελευταία στάδια της ζωής του εξοπλισμού αυξάνεται. Τα περισσότερα δεδομένα που αφορούν τη θνησιμότητα σε διάφορους πληθυσμούς, ακολουθούν αυτού του είδους τη συνάρτηση κινδύνου όπου κατά τη διάρκεια μιας πρώιμης περιόδου, οι θάνατοι προκύπτουν, κυρίως, από βρεφικές ασθένειες. Μετά το πέρας των ασθενειών το ποσοστό θνησιμότητας σταθεροποιείται, ακολουθούμενο από ένα αυξανόμενο ποσοστό κινδύνου λόγω της φυσικής διαδικασίας της γήρανσης.

Τέλος, εάν ο ρυθμός επικινδυνότητας αυξάνει νωρίς και τελικά αρχίσει να μειώνεται, η συνάρτηση κινδύνου έχει τη μορφή *inverted bathtub*. Αυτός ο τύπος της συνάρτησης κινδύνου χρησιμοποιείται συχνά στη μοντελοποίηση της επιβίωσης μετά από μία επιτυχή χειρουργική επέμβαση όπου αρχικά υπάρχει αύξηση του κινδύνου λόγω μόλυνσης, αιμορραγίας ή άλλων επιπλοκών αμέσως μετά τη διαδικασία, ακολουθούμενη από μία σταθεροποίηση της κατάστασης του ασθενούς και τέλος της μείωσης του κινδύνου καθώς ο ασθενής αναρρώνει.

Το περιεχόμενο της παρούσας υποενότητας στηρίζεται στα συγγράμματα των Bain (1991), Barlow and Proschan (1975) και Rinne (2014), όπως επίσης και στις εργασίες των Bain (1974), Hjorth (1980), Glaser (1980) και Gore et al. (1986). Στο Σχήμα 2.3, παρατίθενται οι μορφές τις οποίες μπορεί να λάβουν οι συναρτήσεις κινδύνου.



Σχήμα 2.3: Σχηματική απεικόνιση συναρτήσεων κινδύνου (Ramos (2018))

Για συντομία θα αναφερόμαστε στη *bathtub* ιδιότητα με τον όρο *DIHR* (Decreasing - Increasing Hazard Rate) και στην *inverted bathtub* ιδιότητα με τον όρο *IDHR* (Increasing - Decreasing Hazard Rate). Για τον έλεγχο αν μία συνάρτηση κινδύνου είναι *DIHR* ή *IDHR*, εξετάζεται η πρώτη παράγωγος (στην περίπτωση συνεχούς κατανομής).

Ορισμός 2.2.5. Μια κατανομή $F(x)$ με $x \geq 0$ είναι *DIHR* (*IDHR*) αν υπάρχει ένα $x_0 > 0$ τέτοιο ώστε η $\Lambda(x) = -\ln[1 - F(x)]$ είναι κοίλη (κυρτή) συνάρτηση στο $[0, x_0)$ και κυρτή (κοίλη) συνάρτηση στο $[x_0, \infty)$.

Ο Glaser (1980) όρισε ικανές συνθήκες για τον χαρακτηρισμό μιας κατανομής ως *IHR*, *DHR*, *IDHR* και *DIHR*, υποθέτωντας ότι η σ.π.π. $f(x)$ είναι δύο φορές διαφορίσιμη στο $[0, \infty)$. Οι συνθήκες αυτές βασίζονται στο αντίστροφο της

συνάρτησης κινδύνου και επομένως λαμβάνονται τα εξής:

$$g(x) = \lambda^{-1}(x) = \frac{S(x)}{f(x)},$$

με παράγωγο:

$$g'(x) = g(x)p(x) - 1,$$

όπου:

$$p(x) = -\frac{f'(x)}{f(x)},$$

με παράγωγο:

$$p'(x) = [p(x)]^2 - \frac{f''(x)}{f(x)}.$$

Στο ακόλουθο θεώρημα συνοψίζονται οι συνθήκες του Glaser (1980).

Θεώρημα 2.2.9. (Συνθήκες του Glaser (1980))

1. Αν $p'(x) > 0$, για κάθε $x \geq 0$, τότε *IHR* κατανομή.
2. Αν $p'(x) < 0$, για κάθε $x \geq 0$, τότε *DHR* κατανομή.
3. Έστω ότι υπάρχει $x_0 > 0$ τέτοιο ώστε,

$$\begin{aligned} p'(x) < 0, \quad \text{για κάθε } x \in [0, x_0), \quad p'(x_0) = 0 \\ \text{και } p'(x) > 0 \quad \text{για κάθε } x > x_0, \end{aligned} \quad (2.10)$$

τότε:

- (α') Αν υπάρχει $y_0 > 0$ τέτοιο ώστε $g'(y_0) = 0$ τότε *DIHR* κατανομή.
- (β') Αν δεν υπάρχει $y_0 > 0$ τέτοιο ώστε $g'(y_0) = 0$ τότε *IHR* κατανομή.

4. Έστω ότι υπάρχει $x_0 > 0$ τέτοιο ώστε,

$$\begin{aligned} p'(x) > 0, \quad \text{για κάθε } x \in [0, x_0), \quad p'(x_0) = 0 \\ \text{και } p'(x) < 0 \quad \text{για κάθε } x > x_0, \end{aligned} \quad (2.11)$$

τότε:

- (α') Αν υπάρχει $y_0 > 0$ τέτοιο ώστε $g'(y_0) = 0$ τότε *IDHR* κατανομή.
- (β') Αν δεν υπάρχει $y_0 > 0$ τέτοιο ώστε $g'(y_0) = 0$ τότε *DHR* κατανομή.

Για την αποφυγή εύρεσης της ρίζας y_0 της $g'(\cdot)$ ο Glaser (1980) εισήγαγε το παρακάτω λήμμα.

Λήμμα 2.2.10. (Glaser (1980)) Έστω ότι ισχύουν οι σχέσεις (2.10) και (2.11) του θεωρήματος 2.2.9.

1. Έστω ότι το $\epsilon = \lim_{x \rightarrow 0^+} f(x)$ υπάρχει, τότε:

(α') Αν $\epsilon = \infty$ και ισχύει το (2.10) τότε DIHR κατανομή.

(β') Αν $\epsilon = 0$ και ισχύει το (2.10) τότε IHR κατανομή.

(γ') Αν $\epsilon = \infty$ και ισχύει το (2.11) τότε DHR κατανομή.

(δ') Αν $\epsilon = 0$ και ισχύει το (2.11) τότε IDHR κατανομή.

2. Έστω ότι το $\delta = \lim_{x \rightarrow 0^+} g(x)p(x)$ υπάρχει, τότε:

(α') Αν $\delta > 1$ και ισχύει το (2.10) τότε DIHR κατανομή.

(β') Αν $\delta < 1$ και ισχύει το (2.10) τότε IHR κατανομή.

(γ') Αν $\delta > 1$ και ισχύει το (2.11) τότε DHR κατανομή.

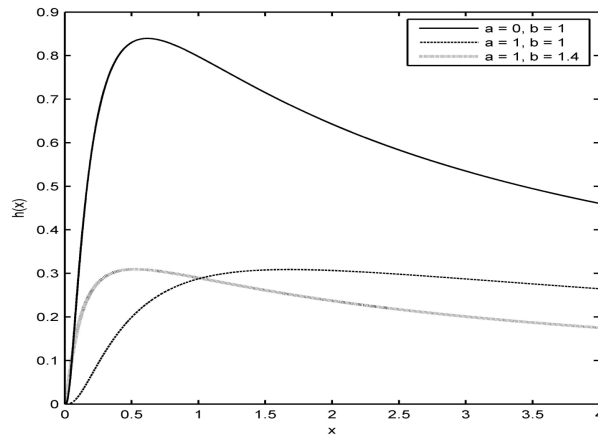
(δ') Αν $\delta < 1$ και ισχύει το (2.11) τότε IDHR κατανομή.

Μεταξύ των κατανομών, σπάνια βρίσκεται κάποια που να είναι DIHR ή IDHR. Εξαιρέση αποτελούν η log - Normal και η inverse Gaussian (Wald) κατανομές οι οποίες είναι IDHR ανεξάρτητα από τις τιμές των παραμέτρων τους, όπως φαίνεται και στα Σχήματα 2.4 και 2.5 παρακάτω.

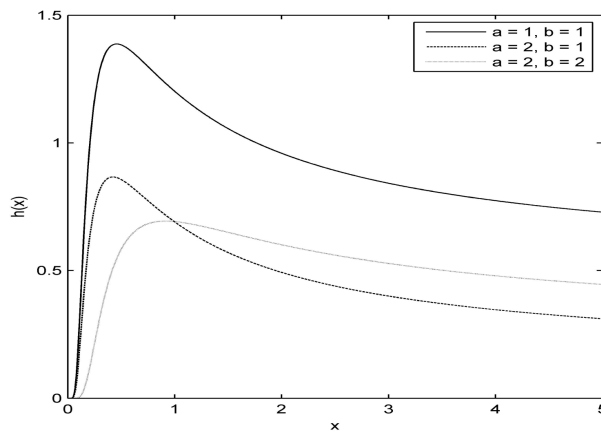
Για την κατασκευή DIHR και IDHR κατανομών υπάρχουν αρκετές μεθοδολογίες. Δύο από αυτές είναι οι εξής:

1. Ορίζεται απευθείας μία συνάρτηση κινδύνου η οποία έχει bathtub ή inverted bathtub σχήμα και στη συνέχεια λαμβάνονται η α.σ.κ. και η σ.π.π. μέσω των σχέσεων (2.6) και (2.7) που ορίστηκαν στην Υποενότητα 2.1.2.
2. Γενικεύεται μία γνωστή κατανομή μέσω της πρόσθεσης επιπλέον παραμέτρων.

Για την κατανόηση των παραπάνω δύο περιπτώσεων δίνονται τα παρακάτω δύο παραδείγματα. Το Παράδειγμα 2.2.11 αφορά την πρώτη περίπτωση, ενώ το Παράδειγμα 2.2.12 αφορά τη δεύτερη περίπτωση.



Σχήμα 2.4: Συναρτήσεις Κινδύνου διάφορων Log – Normal κατανομών (Rinne (2014)).



Σχήμα 2.5: Συναρτήσεις Κινδύνου διάφορων Inverse Gaussian κατανομών (Rinne (2014)).

Παράδειγμα 2.2.11. Οι Bain (1974, 1991) και Gore et al. (1986), χρησιμοποίησαν το δεύτερης τάξης πολυώνυμο το οποίο είναι μια απλή bathtub συνάρτηση κινδύνου που δίνεται από τη σχέση:

$$\lambda(x) = cx^2 + bx + a, \quad x \geq 0, \quad a > 0, \quad b < 0, \quad c > 0.$$

Το ελάχιστο επιτυγχάνεται στο $x = -b/2c$ και η σ.π.π. όπως επίσης και η α.σ.κ.

είναι οι εξής:

$$f(x) = (cx^2 + bx + a) \exp \left\{ -ax - \frac{b}{2}x^2 - \frac{c}{3}x^3 \right\},$$

$$F(x) = 1 - \exp \left\{ -ax - \frac{b}{2}x^2 - \frac{c}{3}x^3 \right\}.$$

Παρατήρηση 2.2.12. Ο Glaser (1980), στην εργασία του, έδειξε ότι δεν είναι δυνατή η κατασκευή κατανομής με *inverted bathtub* συνάρτηση κινδύνου στο $[0, \infty)$ από ένα πολώνυμο.

Παράδειγμα 2.2.13. Στην εργασία του Hjorth (1980), ορίστηκε η παρακάτω συνάρτηση κινδύνου:

$$\lambda(x) = \delta x + \frac{\theta}{1 + \beta x}, \quad x \geq 0, \quad \delta \geq 0, \quad \theta \geq 0, \quad \beta \geq 0, \quad (2.12)$$

όπου δx είναι αύξων όρος και $\theta/(1 + \beta x)$ είναι φθίνων όρος.

Τα αποτελέσματα τα οποία λήφθηκαν είναι τα εξής:

- Για $\beta > 0$,

$$f(x) = \frac{\theta + \delta x (1 + \beta x)}{(1 + \beta x)^{\frac{\theta}{\beta} + 1}} \exp \left\{ -\frac{\delta x^2}{2} \right\}, \quad (2.13)$$

$$F(x) = 1 - \frac{\exp \left\{ -\frac{\delta x^2}{2} \right\}}{(1 + \beta x)^{\frac{\theta}{\beta}}}, \quad (2.14)$$

$$\Lambda(x) = \frac{\delta}{2} \frac{x^2}{2} + \frac{\delta \ln(1 + \beta x)}{\beta}.$$

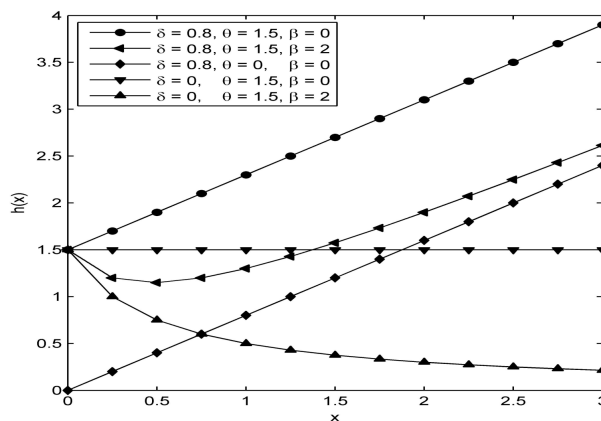
- Για $\beta = 0$, λαμβάνονται τα όρια των (2.13) και (2.14) όταν $\beta \rightarrow 0$.

Ειδικές περιπτώσεις για την συναρτηση κινδύνου (2.12) είναι οι εξής:

1. Αν $\theta = 0$ τότε αύξων ρυθμός επικινδυνότητας (IHR), ίσος με δx , μίας Rayleigh κατανομής.
2. Αν $\delta = 0$ τότε φθίνων ρυθμός επικινδυνότητας (DHR) μίας κατανομής με $F(x) = 1 - (1 + \beta x)^{-\frac{\theta}{\beta}}$ η οποία για $\beta = 1$ είναι η Reduced LOMAX.

3. Αν $\delta = \beta = 0$ τότε σταθερή συνάρτηση κινδύνου μίας εκθετικής κατανομής.
4. Αν $\delta \geq \theta\beta$ τότε IHR κατανομή.
5. Αν $0 < \delta < \theta\beta$ τότε συνάρτηση κινδύνου με bathtub (DIHR) σχήμα και σημείο αλλαγής (ελάχιστο) στο $x = \frac{1}{\beta} \left(\sqrt{\frac{\theta\beta}{\delta}} - 1 \right)$.

Στο Σχήμα 2.6 παρατίθενται συναρτήσεις κινδύνου διάφορων Hjorth κατανομών.



Σχήμα 2.6: Συναρτήσεις κινδύνου διάφορων Hjorth κατανομών (Rinne (2014)).

2.3 Λογοκριμένες και περικομμένες παρατηρήσεις

2.3.1 Λογοκρισία (Censoring)

Η λογοκρισία αποτελεί ζήτημα μείζονος σημασίας στην ανάλυση επιβίωσης, καθώς εμφανίζεται όταν το γεγονός που μας ενδιαφέρει (π.χ. θάνατος, αποτυχία, ανάκαμψη) δεν έχει συμβεί σε όλα τα υποκείμενα της μελέτης μέχρι το τέλος της περιόδου παρακολούθησης. Μπορεί να προκύψει για διάφορους λόγους, όπως λόγω της αποχώρησης των υποκειμένων από τη μελέτη ή λόγω της λήξης της μελέτης πριν από την εμφάνιση του γεγονότος σε όλα τα υποκείμενα που παρακολουθούνται. Κατά την ύπαρξη της λογοκρισίας στα διαθέσιμα δεδομένα, οι παρατηρήσεις δεν εκφράζουν όλες την ολική - πλήρη διάρκεια των γεγονότων

των οποίων έχουν μετρήσει. Υπάρχουν τρία είδη λογοκρισίας, η δεξιά, η αριστερή και η κατά διαστήματα λογοκρισία, παραδείγματα των οποίων δίνονται στις ενότητες που ακολουθούν. Η παρούσα διατριβή θα επικεντρωθεί στην περίπτωση της δεξιάς λογοκρισίας.

Λογοκρισία από δεξιά (Right Censoring)

Κατά τη λογοκρισία από δεξιά, το πραγματικό γεγονός βρίσκεται στα δεξιά του χρόνου λογοκρισίας.

Εστω T_1, T_2, \dots, T_n οι ανεξάρτητοι και ισόνομοι χρόνοι ζωής των ατόμων ή αντικειμένων που υπόκεινται στο υπό μελέτη γεγονός και U_1, U_2, \dots, U_n οι ανεξάρτητοι και ισόνομοι χρόνοι λογοκρισίας. Οι τ.μ. T_i και U_i είναι ανεξάρτητες μεταξύ τους, με τις T_i να έχουν α.σ.κ. $F_T(\cdot)$ και σ.π.π. $f_T(\cdot)$ ενώ οι U_i να έχουν α.σ.κ. $F_U(\cdot)$ και σ.π.π. $f_U(\cdot)$, αντίστοιχα. Ο ακριβής χρόνος ζωής T_i ενός ατόμου που του συνέβη το γεγονός παρατηρείται όταν το T_i είναι μικρότερο ή ίσο του U_i . Αν συμβαίνει το αντίθετο και το U_i είναι μεγαλύτερο του T_i , τότε το άτομο θεωρείται ότι έχει επιζήσει της μελέτης και έχει λογοκριθεί τη χρονική στιγμή U_i .

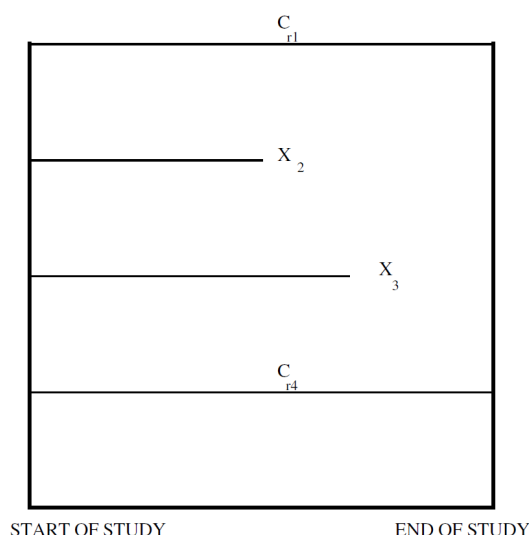
Οι διαθέσιμες παρατηρήσεις παριστάνονται ως ζεύγη τ.μ. (X_i, δ_i) , $i = 1, \dots, n$, όπου X_i είναι ανεξάρτητες και ισόνομες τ.μ. καθώς προκύπτουν από τον συνδυασμό των χρόνων ζωής και των χρόνων λογοκρισίας και δ_i είναι μια δείτρια συνάρτηση που ονομάζεται συντελεστής λογοκρισίας. Μέσω του συντελεστή λογοκρισίας προσδιορίζεται η ύπαρξη ($\delta_i = 0$) ή μη ύπαρξη ($\delta_i = 1$) λογοκριμένων παρατηρήσεων. Κατά συνέπεια, αν $\delta_i = 1$ τότε $X_i = T_i$ και η T_i δεν έχει λογοκριθεί, ενώ αντίστοιχα αν $\delta_i = 0$ τότε $X_i = U_i$ και η T_i έχει λογοκριθεί.

Μέσω μαθηματικών, τα παραπάνω μπορούν να εκφραστούν ως εξής:

$$X_i = \min(T_i, U_i) \quad \text{και} \quad \delta_i = I_{\{T_i \leq U_i\}} = \begin{cases} 1, & \text{αν } T_i \leq U_i, \\ 0, & \text{αν } U_i < T_i. \end{cases}$$

Υπάρχουν τρία είδη δεξιάς λογοκρισίας. Ειδικότερα, η τυχαία λογοκρισία από δεξιά (random right censoring) η οποία παρατηρείται όταν το προκαθορισμένο σημείο είναι τυχαίο και διαφορετικό για κάθε παρατήρηση. Το δεύτερο είδος είναι η δεξιά λογοκρισία τύπου I (type I right censoring) κατά την οποία το προκαθορισμένο σημείο δεν είναι τυχαίο και είναι ίδιο για όλες τις παρατηρήσεις (Σχήμα 2.7), για παράδειγμα $U_1 = U_2 = \dots = U_n = c$. Τέλος, παρατηρείται και η δεξιά λογοκρισία τύπου II (type II right censoring) όπου η μελέτη σταματάει τη στιγμή που παρατηρηθούν n - το πλήθος συμβάντα.

Ως παράδειγμα δεξιάς λογοκρισίας θεωρείται η μελέτη της επίδρασης μίας καρκινογόνου ουσίας σε έναν συγκεκριμένο αριθμό ποντικών. Μέσω αυτού του πειράματος μελετήθηκε η επιβίωση των ποντικών μέχρι μία συγκεκριμένη χρονική στιγμή όπου το πείραμα έλαβε τέλος. Τα ποντίκια τα οποία πέθαιναν πριν το πέρας της έρευνας κάνανε γνωστή με αυτό τον τρόπο στους ερευνητές την πλήρη διάρκεια της ζωής τους από τη στιγμή που λάβανε την καρκινογόνα ουσία μέχρι τη στιγμή που πέθαναν. Όμως τα ποντίκια που επιβίωσαν μετά τη λήξη του πειράματος δεν δώσανε αντίστοιχα αποτελέσματα που να αφορούν τον χρόνο ζωής τους από τη στιγμή που τους δόθηκε η ουσία, με αποτέλεσμα ο χρόνος θανάτου τους να είναι άγνωστος και τα ποντίκια αυτά να θεωρηθούν ότι λογοκρίθηκαν. Το παράδειγμα αυτό αποτελεί παράδειγμα δεξιάς λογοκρισίας τύπου I το οποίο λήφθηκε από το σύγγραμμα των Klein and Moeschberger (2003) και απεικονίζεται στο Σχήμα 2.7.



Σχήμα 2.7: Παράδειγμα δεξιάς λογοκρισίας τύπου I (Klein and Moeschberger (2003)).

Λογοκρισία από αριστερά (Left Censoring)

Κατά τη λογοκρισία από αριστερά, το υπό μελέτη γεγονός έχει ήδη συμβεί σε ένα συγκεκριμένο άτομο ή αντικείμενο πριν αυτό εισέλθει στη μελέτη τη χρονική στιγμή U_i , χωρίς όμως να είναι γνωστή η πραγματική χρονική στιγμή που συνέβη.

Η χρονική στιγμή T_i κατά την οποία συμβαίνει το υπό μελέτη γεγονός θα είναι γνωστή αν και μόνο αν πραγματοποιηθεί μετά τη χρονική στιγμή U_i , δηλαδή αν και μόνο αν το T_i είναι μεγαλύτερο από το U_i .

Όπως με τη δεξιά λογοκρισία έτσι και εδώ, οι διαθέσιμες παρατηρήσεις παριστάνονται ως ζεύγη τυχαίων μεταβλητών (X_i, δ_i) , $i = 1, \dots, n$, όπου X_i είναι ανεξάρτητες και ισόνομες τ.μ. καθώς προκύπτουν από τον συνδυασμό των χρόνων ζωής και των χρόνων λογοκρισίας (οι οποίοι είναι ανεξάρτητοι μεταξύ τους) και δ_i είναι μια δείκτρια συνάρτηση που δηλώνει τον συντελεστή λογοκρισίας. Αν $X_i = T_i$ τότε η T_i δεν έχει λογοκριθεί και κατά συνέπεια ($\delta_i = 1$), ενώ αν $X_i = U_i$ τότε η T_i έχει λογοκριθεί και κατά συνέπεια ($\delta_i = 0$).

Μαθηματικά τα παραπάνω μπορούν να εκφραστούν ως εξής:

$$X_i = \max(T_i, U_i) \quad \text{και} \quad \delta_i = I_{\{U_i \leq T_i\}} = \begin{cases} 1, & \text{αν } U_i \leq T_i, \\ 0, & \text{αν } T_i < U_i. \end{cases}$$

Παράδειγμα αριστερής λογοκρισίας μπορεί να θεωρηθεί η πρώτη χρονική στιγμή κατά την οποία τα παιδιά αποκτούν οποιαδήποτε ικανότητα. Η μελέτη μπορεί να ξεκινήσει σ' ένα χρονικό στάδιο της ζωής των παιδιών με ζητούμενο τον χρόνο μέχρι να εκδηλωθεί η συγκεκριμένη ικανότητα. Κάποια παιδιά όμως όταν εισέρχονται στη μελέτη μπορεί να έχουν ήδη αναπτύξει την ικανότητα αυτή.

Παρατήρηση 2.3.1. Συχνά, αν σε μια μελέτη εμφανίζεται αριστερή λογοκρισία, μπορεί επίσης να υπάρξει και δεξιά λογοκρισία και οι χρόνοι ζωής να θεωρηθούν διπλά λογοκριμένοι (*doubly censored*) σύμφωνα με τον Turnbull (1974).

Λογοκρισία σε διάστημα (Interval Censoring)

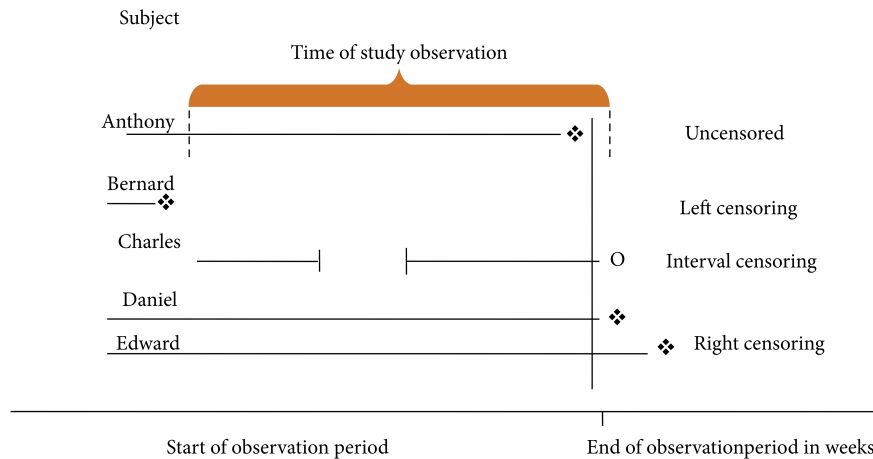
Σε αυτού του είδους τη λογοκρισία είναι γνωστό ότι το υπό μελέτη γεγονός συμβαίνει σε ένα άτομο ή αντικείμενο μέσα σε ένα χρονικό διάστημα, δεν είναι όμως γνωστή η πραγματική του διάρκεια.

Μαθηματικά το παραπάνω μπορεί να εκφραστεί ως ότι οι διαθέσιμες παρατηρήσεις X_i ικανοποιούν τη σχέση:

$$X_i \in (T_i, U_i) \quad i = 1, 2, \dots, n.$$

Παράδειγμα λογοκρισίας σε διάστημα μπορεί να θεωρηθεί όταν ένας ασθενής ελέγχεται σε τακτικά χρονικά διαστήματα και όχι σε συνεχή χρόνο. Αν το γεγονός συμβεί μεταξύ των δύο χρονικών σημείων παρακολούθησης, τότε καθίσταται αδύνατη η γνώση της ακριβούς διάρκειας ζωής του ασθενούς.

Σχηματικά η δεξιά, η αριστερή και η σε διαστήματα λογοκρισία απεικονίζονται στο Σχήμα 2.8.



Σχήμα 2.8: Απεικόνιση δεξιάς, αριστερής και σε διαστήματα λογοκρισίας. (Turkson et al. (2021)).

2.3.2 Περικομμένες παρατηρήσεις (Truncated observations)

Στα περικομμένα δεδομένα, μία παρατήρηση γίνεται αντιληπτή μόνο αν βρίσκεται σε συγκεκριμένο διάστημα, με τη διαφορά ότι δεν υπάρχει λογοκρισία. Αν η παρατήρηση αυτή δεν βρίσκεται στο εν λόγω διάστημα τότε δεν υπάρχει καμία πληροφορία για το υπό μελέτη γεγονός διότι η παρατήρηση δεν είναι διαθέσιμη. Οι περικομμένες παρατηρήσεις μπορούν να προκύψουν από περικομμένες κατανομές.

Η περικομμένη έκδοση μίας κατανομής $F(\cdot)$ στο διάστημα $(a, b]$ δίνεται από τη σχέση:

$$F_Y(y) = \frac{F(y) - F(a)}{F(b) - F(a)}.$$

Για $b = +\infty$ έχουμε περικομμένες παρατηρήσεις από αριστερά, με παράδειγμα την παρατήρηση σωματιδίων που μπορούν να παρατηρηθούν όταν είναι αρκετά μεγάλα. Για $b = 0$ έχουμε περικομμένες παρατηρήσεις από δεξιά με αντίστοιχο παράδειγμα τις μελέτες θνησιμότητας βάσει δεδομένων θανάτων.

Κεφάλαιο 2

2.3. Λογοκριμένες και περιχομμένες παρατηρήσεις

ΚΕΦΑΛΑΙΟ 3

ΕΚΤΙΜΗΣΗ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ ΚΙΝΔΥΝΟΥ

Στο κεφάλαιο αυτό πραγματοποιείται παραμετρική και μη παραμετρική εκτίμηση της συνάρτησης κινδύνου για δεξιά λογοκριμένα δεδομένα. Πιο συγκεκριμένα, η Ενότητα 3.1 εστιάζει στην παραμετρική εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο της μέγιστης πιθανοφάνειας ενώ στην Ενότητα 3.2 γίνεται μη παραμετρική εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο των πυρήνων.

Σε όλη την έκταση του Κεφαλαίου 3 υιοθετούμε ότι οι ανεξάρτητοι και ισόνομοι χρόνοι ζωής $T_i, i = 1, \dots, n$, λογοκρίνονται από δεξιά από τους ανεξάρτητους και ισόνομους χρόνους λογοκρισίας $U_i, i = 1, \dots, n$ (βλέπε Υποενότητα 2.3.1 - Δεξιά Λογοκρισία). Οι τ.μ. T_i και U_i είναι ανεξάρτητες μεταξύ τους, με τις T_i να έχουν α.σ.κ. $F_T(\cdot)$ και σ.π.π. $f_T(\cdot)$ ενώ οι U_i να έχουν α.σ.κ. $F_U(\cdot)$ και σ.π.π. $f_U(\cdot)$. Οι διαθέσιμες παρατηρήσεις παριστάνονται ως ζεύγη τ.μ. $(X_i, \delta_i), i = 1, \dots, n$, όπου X_i είναι ανεξάρτητες και ισόνομες τ.μ. με α.σ.κ. $F(\cdot)$ και σ.π.π. $f(\cdot)$ και δ_i είναι ο συντελεστής λογοκρισίας. Για τις παρατηρήσεις (X_i, δ_i) ισχύει ότι:

$$X_i = \min(T_i, U_i) \quad \text{με} \quad \delta_i = I_{\{T_i \leq U_i\}}.$$

3.1 Παραμετρική εκτίμηση της συνάρτησης κινδύνου

Σκοπός αυτής της ενότητας είναι η εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο της μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation). Αρχικά σχηματίζεται η συνάρτηση πιθανοφάνειας και στη συνέχεια πραγματοποιείται η εκτίμηση της συνάρτησης κινδύνου. Το περιεχόμενο της ενότητας στηρίζεται στα συγγράμματα των Kalbfleisch and Prentice (2002), Klein and Moeschberger (2003) και Lawless (2011).

3.1.1 Συνάρτηση πιθανοφάνειας

Κάθε διαθέσιμη παρατήρηση, λογοκριμένη ή μη, συνεισφέρει πληροφορία στη συνάρτηση πιθανοφάνειας. Η συνεισφορά της (X_i, δ_i) στη συνάρτηση πιθανοφάνειας $L(\theta)$ εξαρτάται από την πιθανότητα που αντιστοιχεί στην περίπτωση της λογοκρίσιμης ή μη στην X_i και κατά συνέπεια δίνεται από τη σχέση:

$$L_i(\theta) = \begin{cases} P(X_i = x, \delta_i = 1; \theta) \\ P(X_i = x, \delta_i = 0; \theta). \end{cases} \quad (3.1)$$

Στην περίπτωση όπου η X_i δεν είναι λογοκριμένη τη χρονική στιγμή t_i ισχύουν ότι:

$$(X_i, \delta_i) = (T_i, 1) = (t_i, 1), \quad (3.2)$$

και

$$T_i \leq U_i. \quad (3.3)$$

Επομένως υπολογίζεται η πιθανότητα:

$$\begin{aligned} P(X_i = x, \delta_i = 1; \theta) &= P(X_i = x, T_i \leq U_i; \theta) \\ &= P(T_i = x | T_i \leq U_i; \theta) P(T_i \leq U_i; \theta) \\ &= \frac{P(T_i \leq U_i | T_i = x; \theta) P(T_i = x; \theta)}{P(T_i \leq U_i; \theta)} \\ &\quad \times P(T_i \leq U_i; \theta) \\ &= P(T_i \leq U_i | T_i = x; \theta) P(T_i = x; \theta) \\ &= P(x \leq U_i; \theta) f_T(x; \theta) \\ &= S_U(x; \theta) f_T(x; \theta) \\ &= (1 - F_U(x; \theta)) f_T(x; \theta), \end{aligned} \quad (3.4)$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκε η (3.3), στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.2) και στην τρίτη γραμμή ο τύπος του Bayes. Επιπλέον όπου $S_U(x; \theta)$ είναι η συνάρτηση επιβίωσης των U_i .

Κάθως όμως τα $F_U(x; \theta)$ και $f_U(x; \theta)$ δεν συνεισφέρουν στην πιθανοφάνεια, η σχέση (3.4) γίνεται:

$$P(X_i = x, \delta_i = 1; \theta) = f_T(x; \theta). \quad (3.5)$$

Στην περίπτωση όπου η X_i λογοκρίνεται τη χρονική στιγμή t_i ισχύουν ότι:

$$(X_i, \delta_i) = (U_i, 0), \quad (3.6)$$

και

$$U_i < T_i = t_i. \quad (3.7)$$

Επομένως υπολογίζεται η πιθανότητα:

$$\begin{aligned} P(X_i = x, \delta_i = 0; \theta) &= P(X_i = x, U_i < T_i; \theta) \\ &= P(U_i = x | U_i < T_i; \theta) P(U_i < T_i; \theta) \\ &= \frac{P(U_i < T_i | U_i = x; \theta) P(U_i = x; \theta)}{P(U_i < T_i; \theta)} P(U_i < T_i; \theta) \\ &= P(U_i < T_i | U_i = x; \theta) P(U_i = x; \theta) \\ &= P(x < T_i; \theta) f_U(x; \theta) \\ &= S_T(x; \theta) f_U(x; \theta) \\ &= (1 - F_T(x; \theta)) f_U(x; \theta), \end{aligned} \quad (3.8)$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκε η (3.7), στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.6) και στην τρίτη γραμμή ο τύπος του Bayes. Επιπλέον όπου $S_T(x; \theta)$ είναι η συνάρτηση επιβίωσης των T_i .

Κάθως όμως τα $F_U(x; \theta)$ και $f_U(x; \theta)$ δεν συνεισφέρουν στην πιθανοφάνεια, η σχέση (3.8) γίνεται:

$$P(X_i = x, \delta_i = 0; \theta) = S_T(x; \theta) = 1 - F_T(x; \theta). \quad (3.9)$$

Επομένως, αντικαθιστώντας τις (3.5) και (3.9), η (3.1) γίνεται:

$$L_i(\theta) = \begin{cases} f_T(X_i; \theta), & \text{αν } \delta_i = 1, \\ 1 - F_T(X_i; \theta), & \text{αν } \delta_i = 0. \end{cases} \quad (3.10)$$

Χρησιμοποιώντας την (3.10), η συνάρτηση πιθανοφάνειας δίνεται από τη σχέση:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n L_i(\theta) = \left(\prod_{\delta_i=1} L_i(\theta) \right) \left(\prod_{\delta_i=0} L_i(\theta) \right) \\ &= \prod_{i=1}^n f_T(X_i; \theta)^{\delta_i} (1 - F_T(X_i, \theta))^{1-\delta_i} \\ &= \prod_{i=1}^n \left(\frac{f_T(X_i; \theta)}{1 - F_T(X_i, \theta)} \right)^{\delta_i} (1 - F_T(X_i, \theta)) \\ &= \prod_{i=1}^n \lambda_T(X_i; \theta)^{\delta_i} S_T(X_i, \theta), \end{aligned} \quad (3.11)$$

η οποία μπορεί να πάρει και την εξής μορφή:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \lambda_T(X_i; \theta)^{\delta_i} S_T(X_i, \theta) \\ &= \prod_{i=1}^n \lambda_T(X_i; \theta)^{\delta_i} \exp \left[- \int_0^{X_i} \lambda_T(u; \theta) du \right], \end{aligned} \quad (3.12)$$

διότι $S(x; \theta) = \exp[-H(x; \theta)] = \exp[-\int h(u; \theta) du]$. Επιπλέον όπου $\lambda_T(x; \theta)$ είναι η συνάρτηση κινδύνου των παρατηρήσεων T_i .

3.1.2 Εκτίμηση μέγιστης πιθανοφάνειας για τη συνάρτηση κινδύνου

Στην προηγούμενη παράγραφο αποδείχθηκε ότι η συνάρτηση πιθανοφάνειας δίνεται από τη σχέση (3.11) ή ισοδύναμα από τη σχέση (3.12). Θέτοντας ως $\hat{\theta}_{ML}$ τον εκτιμητή μέγιστης πιθανοφάνειας, τότε η παραμετρική εκτίμηση της συνάρτησης κινδύνου είναι η $\lambda(x; \hat{\theta}_{ML})$ για $x > 0$.

Για την εύρεση του εκτιμητή μέγιστης πιθανοφάνειας της συνάρτησης κινδύνου εξισώνουμε τη μερική παράγωγο της συνάρτησης πιθανοφάνειας ως προς θ με το μηδέν ως εξής:

$$\frac{\partial L(\theta)}{\partial \theta} = 0,$$

ενώ ισοδύναμα η ίδια διαδικασία μπορεί να πραγματοποιηθεί και για τον λογάριθμο της συνάρτησης πιθανοφάνειας, επομένως:

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0. \quad (3.13)$$

Η λύση της εξίσωσης (3.13) μπορεί να υπολογιστεί με σχετική ευκολία στη μονοδιάστατη περίπτωση της. Όταν όμως πρόκειται για πολυδιάστατες συναρτήσεις πιθανοφάνειας, η χρήση αλγορίθμων όπως ο EM (Expectation - Maximization) οδηγούν στη λύση της εξίσωσης (βλέπε το σύγγραμμα του Pawitan (2001)):

$$\frac{\partial \log L(x; \theta)}{\partial \theta} = 0.$$

Περισσότερες λεπτομέρειες όσον αφορά την παραμετρική εκτίμηση συναρτήσεων κινδύνου μπορούν να βρεθούν στα συγγράμματα των Cox and Oakes (1984) και Kalbfleisch and Prentice (2002).

3.2 Μη παραμετρική εκτίμηση της συνάρτησης κινδύνου

Όταν η α.σ.κ. είναι άγνωστη και κατά συνέπεια δεν υπάρχουν πληροφορίες για την υποκείμενη κατανομή των δεδομένων, τότε μία μη παραμετρική διαδικασία εκτίμησης είναι η μοναδική επιλογή για την εκτίμηση της συνάρτησης κινδύνου που σχετίζεται με τα δεδομένα χρόνων ζωής.

Οι εκτιμητές πυρήνα για τη σ.π.π. και τη συνάρτηση κινδύνου έχουν την ίδια δομή, μοιράζονται το ίδιο σύνολο προβλημάτων και ακολουθούν, σχεδόν, τις ίδιες προσεγγίσεις για την επίλυση των προβλημάτων αυτών. Αρχικά εκτιμητές πυρήνα που αφορούν τη σ.π.π. μελετήθηκαν από τον Rosenblatt (1956) και τον Parzen (1962). Η εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο των πυρήνων ήρθε σε μετέπειτα στάδιο και ξεκίνησε με τις μελέτες των Watson and Leadbetter (1964a,b). Στη συνέχεια, διάφοροι μη παραμετρικοί εκτιμητές πυρήνα για τη συνάρτηση κινδύνου μελετήθηκαν από πολλούς ερευνητές με κάποιους από αυτούς να είναι οι Blum and Susarla (1980), οι Tanner and Wong (1983), ο Ramlau-Hansen (1983), ο Yandell (1983), ο Burke (1983), οι Diehl and Stute (1988), οι Bagkavos and Patil (2008) και ο Bagkavos (2011).

Το περιεχόμενο της ενότητας αυτής στηρίζεται στις εργασίες των Tanner and Wong (1983) και Patil et al. (1994). Στις υποενότητες που ακολουθούν παρουσιάζεται ο εκτιμητής πυρήνα της συνάρτησης κινδύνου που χρησιμοποιείται στην παρούσα μεταπτυχιακή διατριβή καθώς επίσης και η μελέτη της μέσης τιμής, της διακύμανσης αλλά και της κατανομής του.

3.2.1 Εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο των πυρήνων

Η εκτίμηση πυρήνα κλασματικών συναρτήσεων, όπως είναι στην προκειμένη περίπτωση η συνάρτηση κινδύνου, γίνεται συνήθως είτε εκτιμώντας χωριστά τον αριθμητή και τον παρονομαστή είτε θεωρώντας τη συνάρτηση κινδύνου ως συνάρτηση αυτή καθαυτή. Για την εκτίμηση της συνάρτησης κινδύνου με την πρώτη μέθοδο παραπέμπουμε στην εργασία του Bagkavos (2003), ενώ στην παρούσα διατριβή ο εκτιμητής πυρήνα που χρησιμοποιείται προτάθηκε από τους Tanner and Wong (1983).

Όπως έχει αναφερθεί στην εισαγωγή του Κεφαλαίου 3, οι χρόνοι ζωής T_i , οι χρόνοι λογοκρισίας U_i , αλλά και οι διαθέσιμες παρατηρήσεις (X_i, δ_i) με $i = 1, \dots, n$ ακολουθούν τις υποθέσεις της Υποενότητας 2.3.1 για την δεξιά λογοκρισία. Επι-

πλέον, οι παρατηρήσεις X_i έχουν α.σ.κ. $F(x)$ και σ.π.π. $f(x)$ για τις οποίες ισχύει ότι:

$$1 - F(x) = (1 - F_T(x))(1 - F_U(x)), \quad (3.14)$$

και

$$f(x) = f_T(x)(1 - F_U(x)) + f_U(x)(1 - F_T(x)). \quad (3.15)$$

Διατάσσοντας το δείγμα των X_i κατά αύξουσα σειρά μαζί με τον συντελεστή λογοχρισίας $\delta_i = I_{\{T_i \leq U_i\}}$, έτσι ώστε να ισχύει ότι $X_1 < X_2 < \dots < X_n$, οι διαθέσιμες παρατηρήσεις προς ανάλυση δίνονται ως $(X_{(i)}, \delta_{(i)})$.

Ακολουθώντας την εργασία των Patil et al. (1994), για να υπολογιστεί ο εκτιμητής πυρήνα της συνάρτησης κινδύνου, παραθέτουμε τα παρακάτω λήμματα τα οποία και αποδεικνύουμε αναλυτικά.

Λήμμα 3.2.1.

$$F^+(x) = P(X_i \leq x, \delta_i = 0) = \int_0^x (1 - F_T(y)) dF_U(y), \quad (3.16)$$

$$F^-(x) = P(X_i \leq x, \delta_i = 1) = \int_0^x (1 - F_U(y)) dF_T(y). \quad (3.17)$$

Απόδειξη. Αρχικά η απόδειξη της (3.16) πραγματοποιείται ως εξής:

$$\begin{aligned} F^+(x) &= P(X_i \leq x, \delta_i = 0) \\ &= P(X_i \leq x, U_i < T_i) \\ &= P(U_i \leq x, U_i < T_i) \\ &= P(U_i < T_i, U_i \leq x), \end{aligned}$$

όπου στην τρίτη γραμμή χρησιμοποιήθηκε η (3.6).

Χρησιμοποιώντας τη σχέση,

$$P(A, X \in B) = \int_B P(A|X = x) f_x(x) dx, \quad (3.18)$$

έχουμε ότι:

$$\begin{aligned} F^+(x) &= \int_0^x P(U_i < T_i | U_i = y) f_U(y) dy \\ &= \int_0^x P(y < T_i) f_U(y) dy \\ &= \int_0^x (1 - P(T_i < y)) f_U(y) dy, \end{aligned}$$

ή ισοδύναμα ότι:

$$F^+(x) = \int_0^x (1 - F_T(y)) f_U(y) dy = \int_0^x (1 - F_T(y)) dF_U(y).$$

Με παρόμοιο τρόπο αποδεικνύεται η (3.17) ως εξής:

$$\begin{aligned} F^-(x) &= P(X_i \leq x, \delta_i = 1) \\ &= P(X_i \leq x, T_i \leq U_i) \\ &= P(T_i \leq x, T_i \leq U_i) \\ &= P(T_i \leq U_i, T_i \leq x) \\ &= \int_0^x P(T_i \leq U_i | T_i = y) f_T(y) dy \\ &= \int_0^x P(y \leq U_i) f_T(y) dy \\ &= \int_0^x (1 - P(U_i \leq y)) f_T(y) dy \\ &= \int_0^x (1 - F_U(y)) f_T(y) dy \\ &= \int_0^x (1 - F_U(y)) dF_T(y), \end{aligned}$$

όπου στην τρίτη γραμμή χρησιμοποιήθηκε η (3.2), και στην πέμπτη γραμμή η (3.18). \square

Παρατήρηση 3.2.2. Για την α.σ.κ. των X_i , μέσω των (3.16) και (3.17), ισχύει ότι:

$$F(x) = F^+(x) + F^-(x).$$

Ορισμός 3.2.1. Για την εμπειρική α.σ.κ. των X_i η οποία δίνεται από τη σχέση,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}, \quad (3.19)$$

ορίζονται οι εμπειρικές α.σ.κ.

$$F_n^+(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i < x\}} (1 - \delta_i),$$

και

$$F_n^-(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i < x\}} \delta_i, \quad (3.20)$$

έτσι ώστε να ισχύει ότι:

$$F_n(x) = F_n^+(x) + F_n^-(x).$$

Παρατήρηση 3.2.3. Όπως ακριβώς και στην (3.14) έτσι και για την F_n ισχύει ότι:

$$(1 - F_n(x)) = (1 - F_{T,n}(x))(1 - F_{U,n}(x)),$$

όπου $F_{T,n}(x)$ και $F_{U,n}(x)$ οι εμπειρικές εκδοχές των $F_T(x)$ και $F_U(x)$ αντίστοιχα.

Η συνάρτηση κινδύνου των χρόνων επιβίωσης T_i δίνεται από τη σχέση:

$$\lambda_T(x) = \frac{f_T(x)}{S_T(x)} = \frac{f_T(x)}{1 - F_T(x)}. \quad (3.21)$$

Εφαρμόζοντας την (3.14) στην (3.21) λαμβάνουμε ότι:

$$\lambda_T(x) = \frac{(1 - F_U(x)) f_T(x)}{1 - F(x)}. \quad (3.22)$$

Λήμμα 3.2.4. Η αθροιστική συνάρτηση της (3.22) δίνεται από τη σχέση:

$$\Lambda_T(x) = \int_0^x \frac{dF^-(u)}{1 - F(u)}. \quad (3.23)$$

Απόδειξη.

$$\begin{aligned} \Lambda_T(x) &= \int_0^x \lambda_T(u) du \\ &= \int_0^x \frac{(1 - F_U(u)) f_T(u)}{1 - F(u)} du \\ &= \int_0^x \frac{(1 - F_U(u))}{1 - F(u)} dF_T(u) \\ &= \int_0^x \frac{dF^-(u)}{1 - F(u)}, \end{aligned}$$

όπου στην τελευταία γραμμή χρησιμοποιήθηκε η σχέση:

$$dF^-(x) = (1 - F_U(x))dF_T(x),$$

που προκύπτει παραγωγίζοντας την (3.17) ως προς x . □

Η εμπειρική εκδοχή της (3.23) είναι η:

$$\Lambda_{T,n}(x) = \int_0^x \frac{dF_n^-(u)}{1 - F_n(u)}, \quad (3.24)$$

η οποία προκύπτει με όμοιο τρόπο με την (3.23) αν αντικαταστήσουμε τις υπάρχουσες αθροιστικές συναρτήσεις με τις αντίστοιχες εμπειρικές.

Για τους σκοπούς της μη παραμετρικής εκτίμησης, η κλίμακα του πυρήνα δεν είναι μονοσήμαντα ορισμένη. Δηλαδή, για κάθε πυρήνα $K(\cdot)$ θα μπορούσε να είχε οριστεί ο εναλλακτικός πυρήνας:

$$K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right), \quad \text{με } h > 0. \quad (3.25)$$

Αυτοί οι δύο πυρήνες είναι ισοδύναμοι με την έννοια ότι παράγουν τους ίδιους εκτιμητές, εφόσον το εύρος ζώνης (h) αναπροσαρμόζεται. Δηλαδή, για παράδειγμα, αν ο εκτιμητής της συνάρτησης κινδύνου $\lambda(x)$ υπολογίζεται με πυρήνα K και εύρος ζώνης b τότε είναι αριθμητικά πανομοιότυπος με τον υπολογισμό του αντίστοιχου εκτιμητή με πυρήνα K_h και εύρος ζώνης $h^* = b/h$.

Η (3.22) μπορεί να προσεγγιστεί από την:

$$\lambda^*(x) = \int K_h(x - y) \lambda_T(y) dy, \quad (3.26)$$

η οποία αποτελεί τον ορισμό της συνέλιξης της $\lambda_T(x)$ με την $K_h(x)$.

Λήμμα 3.2.5.

$$\lambda_T(x) \approx \int K_h(x - u) d\Lambda_{T,n}(u) + \int K_h(x - u) d[\Lambda_T(u) - \Lambda_{T,n}(u)]. \quad (3.27)$$

Απόδειξη.

$$\begin{aligned} \lambda_T(x) &\approx \lambda^*(x) = \int K_h(x - u) \lambda_T(u) du \\ &= \int K_h(x - u) \frac{(1 - F_U(u)) f_T(u)}{1 - F(u)} du \\ &= \int K_h(x - u) \frac{(1 - F_U(u))}{1 - F(u)} dF_T(u) \\ &= \int K_h(x - u) \frac{dF^-(u)}{1 - F(u)} \\ &= \int K_h(x - u) d\Lambda_T(u), \end{aligned}$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκε η (3.26), στη δεύτερη γραμμή η (3.22), στην τέταρτη γραμμή η παράγωγος της (3.17) και στην πέμπτη γραμμή η παράγωγος της (3.23).

Ισοδύναμα,

$$\begin{aligned}\lambda_T(x) &\approx \int K_h(x-u) d\Lambda_T(u) + \int K_h(x-u) d\Lambda_{T,n}(u) - \int K_h(x-u) d\Lambda_{T,n}(u) \\ &= \int K_h(x-u) d\Lambda_{T,n}(u) + \int K_h(x-u) d[\Lambda_T(u) - \Lambda_{T,n}(u)].\end{aligned}$$

□

Ο ζητούμενος εκτιμητής πυρήνα της συνάρτησης κινδύνου προκύπτει από τον πρώτο όρο του δεξιού μέρους της (3.27) ως εξής:

$$\begin{aligned}\hat{\lambda}(x) &= \int K_h(x-u) d\Lambda_{T,n}(u) \\ &= \int K_h(x-u) \frac{dF_n^-(u)}{1-F_n(u)} \\ &= K_h(x-u) \frac{F_n^-(u)}{1-F_n(u)} \\ &= \frac{K_h(x-u)}{1-F_n(u)} \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}} \delta_i \\ &= \frac{1}{n} \sum_{i=1}^n \frac{K_h(x-X_i) \delta_i}{1-F_n(X_i)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{K_h(x-X_i) \delta_i}{S_n(X_i)},\end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η παράγωγος της (3.24) και στην τέταρτη γραμμή η σχέση (3.19).

Διατάσσοντας τις παρατηρήσεις X_i σε αύξουσα σειρά και αντικαθιστώντας τη σχέση:

$$S_n(X_{(i)}) = 1 - F_n(X_{(i)}) = 1 - \frac{i}{n} = \frac{n-i}{n},$$

λαμβάνουμε ότι:

$$\hat{\lambda}(x) = \frac{1}{n} \sum_{i=1}^n \frac{K_h(x-X_{(i)}) \delta_{(i)}}{\frac{n-i}{n}} = \sum_{i=1}^n \frac{K_h(x-X_{(i)}) \delta_{(i)}}{n-i} = \sum_{i=1}^n \frac{K_h(x-X_{(i)}) \delta_{(i)}}{n-i+1}.$$

Επομένως ο εκτιμητής πυρήνα της συνάρτησης κινδύνου δίνεται από τη σχέση:

$$\hat{\lambda}(x) = \sum_{i=1}^n \frac{K_h(x - X_{(i)})\delta_{(i)}}{n - i + 1}. \quad (3.28)$$

Παρατήρηση 3.2.6. Στον παρονομαστή του εκτιμητή που δίνεται από την (3.28) προστίθεται η μονάδα έτσι ώστε να μην μηδενίζεται όταν $i = n$.

Εκτός από την (3.28), ο εκτιμητής πυρήνα της συνάρτησης κινδύνου μπορεί να βρεθεί και σε άλλες μορφές οι οποίες είναι οι εξής:

$$\hat{\lambda}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F_n(X_i)} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n \frac{K_h(x - X_i)}{1 - F_n(X_i)}, \quad (3.29)$$

και

$$\hat{\lambda}(x) = \sum_{i=1}^n \frac{K_h(x - X_i)\delta_i}{n - R_i + 1}, \quad (3.30)$$

όπου $F_n(x)$ είναι η εμπειρική α.σ.κ. που δίνεται από την (3.19) και R_i είναι η τάξη των X_i που δίνεται από τη σχέση:

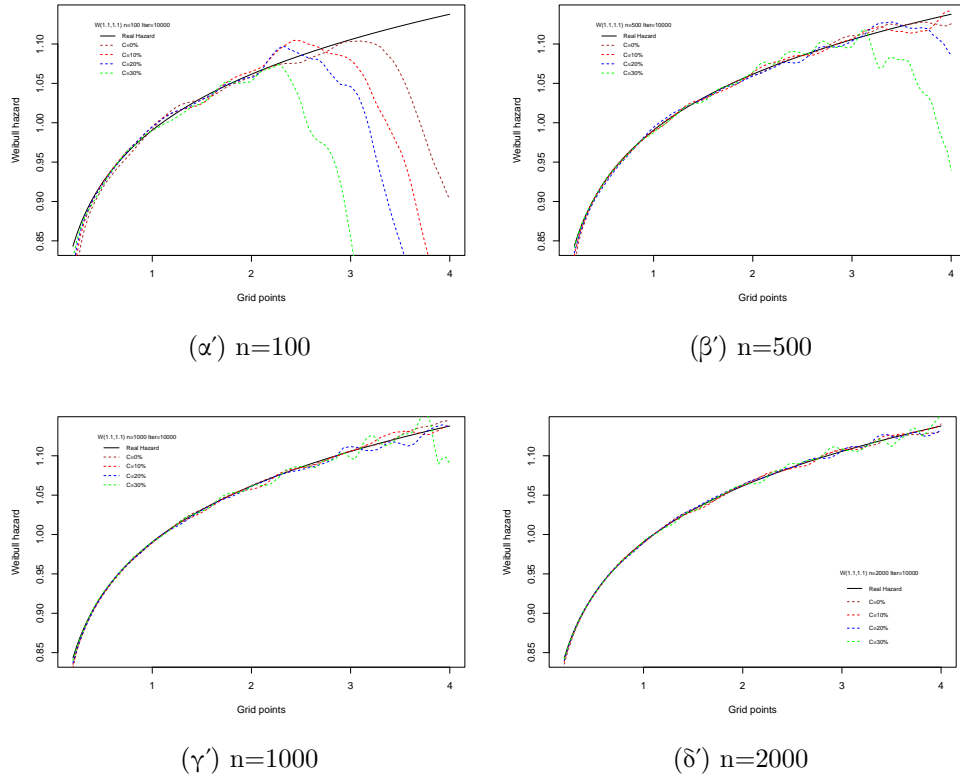
$$R_i = nF_n(X_i) = \sum_{j=1}^n I_{\{X_j \leq X_i\}}. \quad (3.31)$$

Στο Σχήμα 3.1 παρουσιάζεται ο εκτιμητής $\hat{\lambda}(x)$, όπως αυτός προέκυψε μετά από 10.000 επαναλήψεις, χρησιμοποιώντας δεδομένα που προέρχονται από την κατανομή Weibull με παραμέτρους $\kappa = \lambda = 1.1$ (Σύμβ: $W(1.1, 1.1)$). Τα ποσοστά λογοκρισίας (Σύμβ: C) που χρησιμοποιήθηκαν είναι 0% (καθόλου λογοκρισία), 10%, 20% και 30%, με τη μελέτη να πραγματοποιείται για καθένα από αυτά τα ποσοστά για μέγεθος δείγματος $n = 100, 500, 1000, 2000$ αντίστοιχα.

Παρατήρηση 3.2.7. Στο Σχήμα 3.1, παρατηρούμε ότι, ανεξαρτήτου ποσοστού λογοκρισίας και καθώς αυξάνεται το μέγεθος του δείγματος, ο εκτιμητής $\hat{\lambda}(x)$ τείνει όλο και περισσότερο στην πραγματική συνάρτηση κινδύνου, γεγονός που τον καθιστά συνεπή εκτιμητή.

3.2.2 Μέση τιμή και διακύμανση του εκτιμητή $\hat{\lambda}(x)$

Στην υποενότητα αυτή, πραγματοποιείται η μελέτη της μέσης τιμής και της διακύμανσης του εκτιμητή πυρήνα της συνάρτησης κινδύνου που παραθέσαμε στην



Σχήμα 3.1: Προσομοιώσεις για τον εκτιμητή πυρήνα της συνάρτησης κινδύνου για διάφορα μεγέθη δείγματος και ποσοστά λογοχρισίας.

Υποενότητα 3.2.1. Το περιεχόμενο της παρούσας υποενότητας στηρίζεται στην εργασία των Tanner and Wong (1983).

Πριν τον υπολογισμό της μέσης τιμής και της διακύμανσης του εκτιμητή $\hat{\lambda}(x)$, παραθέτουμε και αποδεικνύουμε αναλυτικά το παρακάτω λήμμα. Το λήμμα αυτό αποτελεί το Λήμμα 1 της εργασίας των Tanner and Wong (1983) η απόδειξη του οποίου δεν ήταν διαθέσιμη και για τον λόγο αυτόν την παραθέτουμε αναλυτικά.

Λήμμα 3.2.8. Έστω:

$$m(t) = \frac{f_T(t)(1 - F_U(t))}{f(t)}, \quad \mu \in f(t) > 0, \quad (3.32)$$

τότε:

$$E(\delta_{(i)}|X_{(i)} = t) = m(t), \quad \text{για κάθε } i, \quad (3.33)$$

$$E(\delta_{(i)}\delta_{(j)}|X_{(i)} = t, X_{(j)} = s) = m(t)m(s), \quad \text{για κάθε } i < j, \quad t < s. \quad (3.34)$$

Απόδειξη. Δοθέντος ότι $\delta_{(k)} = \delta_i$ όταν $X_i = X_{(k)}$, η (3.33) προκύπτει ως εξής:

$$\begin{aligned} E(\delta_i|X_i = t) &= 1 \cdot P(\delta_i = 1|X_i = t) + 0 \cdot P(\delta_i = 0|X_i = t) \\ &= P(\delta_i = 1|X_i = t) \\ &= \frac{P(X_i = t|\delta_i = 1)P(\delta_i = 1)}{P(X_i = t)} \\ &= \frac{P(X_i = t|X_i = T_{(i)})P(T_i < U_i)}{f(t)} \\ &= \frac{P(T_i = t)P(U_i > t)}{f(t)} \\ &= \frac{f_T(t)(1 - F_U(t))}{f(t)} \\ &= m(t), \end{aligned}$$

για την οποία στην τρίτη γραμμή χρησιμοποιήθηκε ο τύπος της δεσμευμένης πιθανότητας και στην τέταρτη γραμμή χρησιμοποιήθηκαν οι (3.2) και (3.3).

Η απόδειξη της (3.34) προκύπτει άμεσα από την (3.33) δεδομένου ότι $i \neq j$ και ότι οι X_i είναι ανεξάρτητες και ισόνομες τ.μ. \square

Η μέση τιμή και η διακύμανση του εκτιμητή $\hat{\lambda}(x)$ δίνονται στα Θεωρήματα 3.2.9 και 3.2.10 αντίστοιχα, τα οποία στηρίζονται στο Θεώρημα 1 της εργασίας των Tanner and Wong (1983). Το Θεώρημα 3.2.9 περιέχει την πρώτη σχέση του Θεωρήματος 1, ενώ το Θεώρημα 3.2.10 περιέχει τη δεύτερη σχέση του Θεωρήματος 1. Δοθέντος ότι στην εργασία των Tanner and Wong (1983) η απόδειξη του Θεωρήματος 1 σκιαγραφείται, τα Θεωρήματα 3.2.9 και 3.2.10 αποδεικνύονται αναλυτικά.

Θεώρημα 3.2.9. *Η μέση τιμή του εκτιμητή πυρήνα της συνάρτησης κινδύνου δίνεται από τη σχέση:*

$$E(\hat{\lambda}(x)) = \int_0^\infty (1 - F^n(y)) \lambda_T(y) K_h(x - y) dy. \quad (3.35)$$

Απόδειξη. Χρησιμοποιώντας την (3.28) έχουμε ότι:

$$\begin{aligned} E(\hat{\lambda}(x)) &= E\left(\sum_{i=1}^n \frac{\delta^{(i)}}{n-i+1} K_h(x - X_{(i)})\right) \\ &= \sum_{i=1}^n E\left(\frac{\delta^{(i)}}{n-i+1} K_h(x - X_{(i)})\right), \end{aligned}$$

όπου εφαρμόζοντας τη σχέση,

$$E(h(X, Y)) = E(E(h(X, Y)|X)), \quad (3.36)$$

προκύπτει ότι:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \sum_{i=1}^n E\left[E\left(\frac{\delta^{(i)}}{n-i+1} K_h(x - X_{(i)}) | X_{(i)} = y\right)\right] \\ &= \sum_{i=1}^n \int_0^\infty E\left(\frac{\delta^{(i)}}{n-i+1} K_h(x - X_{(i)}) | X_{(i)} = y\right) f_{X_{(i)}}(y) dy \\ &= \sum_{i=1}^n \int_0^\infty E(\delta^{(i)} | X_{(i)} = y) \frac{1}{n-i+1} K_h(x - y) f_{X_{(i)}}(y) dy \\ &= \sum_{i=1}^n \int_0^\infty \frac{m(y)}{n-i+1} K_h(x - y) f_{X_{(i)}}(y) dy, \end{aligned} \quad (3.37)$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε ο τύπος της μέσης τιμής για συνεχείς κατανομές, στην τρίτη γραμμή χρησιμοποιήθηκε η σχέση:

$$E(f(U)Q|U) = f(U)E(Q|U), \quad (3.38)$$

και στην τέταρτη γραμμή η (3.33).

Αντικαθιστώντας στην (3.37) τη σ.π.π. του διατεταγμένου στατιστικού, η οποία δίνεται από τη σχέση:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} F^{i-1}(x) (1 - F(x))^{n-i} f(x), \quad (3.39)$$

έχουμε ότι:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \sum_{i=1}^n \int_0^\infty \frac{m(y)}{n-i+1} K_h(x - y) \frac{n!}{(i-1)!(n-i)!} \\ &\quad \times F^{i-1}(y) (1 - F(y))^{n-i} f(y) dy. \end{aligned} \quad (3.40)$$

Παρατηρούμε ότι:

$$\frac{1}{n-i+1} \frac{n!}{(i-1)!(n-i)!} = \frac{n!}{(i-1)!(n-i+1)!} = \binom{n}{i-1},$$

το οποίο αν αντικατασταθεί στην (3.40) λαμβάνουμε το εξής:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \sum_{i=1}^n \int_0^{\infty} \binom{n}{i-1} F^{i-1}(y) (1-F(y))^{n-i} \\ &\quad \times f(y)m(y)K_h(x-y) dy \\ &= \int_0^{\infty} \left[\sum_{i=1}^n \binom{n}{i-1} F(y)^{i-1} (1-F(y))^{n-i} \right] \\ &\quad \times f(y)m(y)K_h(x-y) dy. \end{aligned} \quad (3.41)$$

Επιπλέον, πραγματοποιώντας την αλλαγή μεταβλητής $k = i - 1$ προκύπτει ότι:

$$\sum_{i=1}^n \binom{n}{i-1} F^{i-1}(y) (1-F(y))^{n-i} = \sum_{k=0}^{n-1} \binom{n}{k} F^k(y) (1-F(y))^{n-k-1}, \quad (3.42)$$

και επομένως η (3.41) γίνεται:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \int_0^{\infty} \left[\sum_{k=0}^{n-1} \binom{n}{k} F^k(y) (1-F(y))^{n-k-1} \right] \\ &\quad \times f(y)m(y)K_h(x-y) dy. \end{aligned} \quad (3.43)$$

Στη συνέχεια, μέσω του διωνυμικού θεωρήματος αποδεικνύεται το εξής:

$$\begin{aligned} \sum_{k=0}^n \binom{n}{k} F^k(y) (1-F(y))^{n-k} &= (F(y) + 1 - F(y))^n \Rightarrow \\ \Rightarrow \sum_{k=0}^n \binom{n}{k} F^k(y) (1-F(y))^{n-k} &= 1 \\ \Rightarrow \sum_{k=0}^{n-1} \binom{n}{k} F^k(y) (1-F(y))^{n-k} &+ \binom{n}{n} F^n(y) (1-F(y))^{n-n} = 1 \\ \Rightarrow \sum_{k=0}^{n-1} \binom{n}{k} F^k(y) (1-F(y))^{n-k} &= 1 - F^n(y) \\ \Rightarrow \sum_{k=0}^{n-1} \binom{n}{k} F^k(y) (1-F(y))^{n-k-1} &= \frac{1 - F^n(y)}{1 - F(y)}. \end{aligned} \quad (3.44)$$

Επομένως λόγω της (3.44), η (3.43) γίνεται:

$$\begin{aligned}
E\left(\hat{\lambda}(x)\right) &= \int_0^\infty \frac{1 - F^n(y)}{1 - F(y)} f(y) m(y) K_h(x - y) dy \\
&= \int_0^\infty \frac{1 - F^n(y)}{1 - F(y)} f(y) \frac{f_T(y)(1 - F_U(y))}{f(y)} K_h(x - y) dy \\
&= \int_0^\infty \frac{1 - F^n(y)}{(1 - F_T(y))(1 - F_U(y))} f_T(y)(1 - F_U(y)) K_h(x - y) dy \\
&= \int_0^\infty (1 - F^n(y)) \frac{f_T(y)}{1 - F_T(y)} K_h(x - y) dy \\
&= \int_0^\infty (1 - F^n(y)) \lambda_T(y) K_h(x - y) dy,
\end{aligned}$$

γεγονός που αποδεικνύει την (3.35).

Παραπάνω, στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.32), στην τρίτη γραμμή η (3.14) και στην τελευταία γραμμή η (3.21). \square

Θεώρημα 3.2.10. Η διακύμανση του εκτιμητή πυρήνα της συνάρτησης κινδύνου δίνεται από τη σχέση:

$$\begin{aligned}
Var\left(\hat{\lambda}(x)\right) &= \int I_n(F(y)) \lambda_T(y) K_h^2(x - y) dy \\
&+ 2 \int \int_{y \leq z} \left\{ F^n(z) - F^n(y) F^n(z) - \frac{1 - F(y)}{F(z) - F(y)} (F^n(z) - F^n(y)) \right\} \\
&\quad \lambda_T(y) \lambda_T(z) K_h(x - y) K_h(x - z) dy dz, \quad (3.45)
\end{aligned}$$

όπου:

$$I_n(F(y)) = \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n}{k} F^k(y) (1 - F(y))^{n-k}. \quad (3.46)$$

Απόδειξη. Από τον τύπο της διακύμανσης ισχύει ότι:

$$Var\left(\hat{\lambda}(x)\right) = E\left(\hat{\lambda}^2(x)\right) - \left[E\left(\hat{\lambda}(x)\right)\right]^2. \quad (3.47)$$

Αρχικά για τον πρώτο όρο του δεξιού μέρους της (3.47) έχουμε ότι:

$$E\left(\hat{\lambda}^2(x)\right) = E\left[\left(\sum_{i=1}^n \frac{\delta_{(i)}}{n-i+1} K_h(x - X_{(i)})\right)^2\right],$$

ή ισοδύναμα ότι:

$$E\left(\hat{\lambda}^2(x)\right) = E\left[\sum_{i=1}^n \frac{\delta_{(i)}^2}{(n-i+1)^2} K_h^2(x - X_{(i)}) + 2 \sum_{i < j} \frac{\delta_{(i)} \delta_{(j)}}{(n-i+1)(n-j+1)} K_h(x - X_{(i)}) K_h(x - X_{(j)})\right],$$

στο οποίο χρησιμοποιήθηκε η σχέση:

$$\left(\sum_{i=1}^n x_i\right)^2 = \sum_{i=1}^n x_i^2 + 2 \sum_{i < j} x_i x_j.$$

Επομένως προκύπτει ότι:

$$\begin{aligned} E\left(\hat{\lambda}^2(x)\right) &= E\left[\sum_{i=1}^n \frac{\delta_{(i)}^2}{(n-i+1)^2} K_h^2(x - X_{(i)})\right] \\ &\quad + 2E\left[\sum_{i < j} \frac{\delta_{(i)} \delta_{(j)}}{(n-i+1)(n-j+1)} \right. \\ &\quad \left. \times K_h(x - X_{(i)}) K_h(x - X_{(j)})\right] \\ &\equiv E\left(\hat{\lambda}_1^2(x)\right) + 2E\left(\hat{\lambda}_2^2(x)\right). \end{aligned} \quad (3.48)$$

Χειριζόμαστε ξεχωριστά τους δύο όρους του δεξιού μέρους της (3.48) και ξεκινώντας από τον πρώτο όρο έχουμε ότι:

$$\begin{aligned} E\left(\hat{\lambda}_1^2(x)\right) &= E\left[\sum_{i=1}^n \frac{\delta_{(i)}^2}{(n-i+1)^2} K_h^2(x - X_{(i)})\right] \\ &= \sum_{i=1}^n E\left[\frac{\delta_{(i)}^2}{(n-i+1)^2} K_h^2(x - X_{(i)})\right]. \end{aligned}$$

Χρησιμοποιώντας το γεγονός ότι η $\delta_{(i)}$ είναι δείκτρια συνάρτηση και κατά συνέπεια λαμβάνει τιμές 0 ή 1, θα ισχύει ότι $\delta_{(i)}^2 = \delta_{(i)}$. Συνεπώς:

$$\begin{aligned} E\left(\hat{\lambda}_1^2(x)\right) &= \sum_{i=1}^n E\left[\frac{\delta_{(i)}}{(n-i+1)^2} K_h^2(x - X_{(i)})\right] \\ &= \sum_{i=1}^n E\left[E\left[\frac{\delta_{(i)}}{(n-i+1)^2} K_h^2(x - X_{(i)}) \mid X_{(i)} = y\right]\right], \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.36).

Ισοδύναμα έχουμε ότι:

$$\begin{aligned} E\left(\hat{\lambda}_1^2(x)\right) &= \int_0^\infty \sum_{i=1}^n \frac{K_h^2(x-y)}{(n-i+1)^2} E\left(\delta_{(i)} | X_{(i)} = y\right) f_{X_{(i)}}(x) dy \\ &= \int_0^\infty \sum_{i=1}^n \frac{K_h^2(x-y)}{(n-i+1)^2} m(y) f_{X_{(i)}}(x) dy, \end{aligned}$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκαν η (3.38) και ο τύπος της μέσης τιμής για συνεχείς κατανομές και στην δεύτερη γραμμή χρησιμοποιήθηκε η (3.33).

Αντικαθιστώντας την (3.39) λαμβάνουμε ότι:

$$\begin{aligned} E\left(\hat{\lambda}_1^2(x)\right) &= \int_0^\infty \sum_{i=1}^n \frac{n!}{(n-i+1)^2(i-1)!(n-i)!} F^{i-1}(y) \\ &\quad \times (1-F(y))^{n-i} m(y) K_h^2(x-y) f(y) dy. \end{aligned} \quad (3.49)$$

Επιπλέον, καθώς,

$$\frac{n!}{(n-i+1)^2(i-1)!(n-i)!} = \frac{1}{n-i+1} \binom{n}{i-1},$$

και λαμβάνοντας υπόψη την (3.42) στην (3.49) προκύπτει ότι:

$$\begin{aligned} E\left(\hat{\lambda}_1^2(x)\right) &= \int_0^\infty \sum_{i=1}^n \frac{1}{n-i+1} \binom{n}{i-1} F^{i-1}(y) (1-F(y))^{n-i} \\ &\quad \times m(y) K_h^2(x-y) f(y) dy \\ &= \int_0^\infty \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n}{k} F^k(y) (1-F(y))^{n-k-1} \\ &\quad \times m(y) K_h^2(x-y) f(y) dy, \quad \text{για } k = i-1 \\ &= \int_0^\infty \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n}{k} F^k(y) (1-F(y))^{n-k-1} \\ &\quad \times \frac{f_T(y)(1-F_U(y))}{f(y)} K_h^2(x-y) f(y) dy \\ &= \int_0^\infty \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n}{k} F^k(y) (1-F(y))^{n-k-1} \frac{f_T(y)}{1-F_T(y)} \\ &\quad \times (1-F(y)) K_h^2(x-y) dy, \end{aligned}$$

ή ισοδύναμα:

$$\begin{aligned} E\left(\hat{\lambda}_1^2(x)\right) &= \int_0^\infty \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n}{k} F^k(y) (1-F(y))^{n-k} \\ &\quad \times \lambda_T(y) K_h^2(x-y) dy \\ &= \int_0^\infty I_n(F(y)) \lambda_T(y) K_h^2(x-y) dy, \end{aligned} \quad (3.50)$$

όπου στο πρώτο βήμα χρησιμοποιήθηκε η σχέση της συνάρτησης κινδύνου για τους χρόνους ζωής T_i και στην τελευταία γραμμή έγινε αντικατάσταση της (3.46).

Στη συνέχεια ασχολούμαστε με το δεύτερο μέρος της (3.48). Επομένως έχουμε ότι:

$$\begin{aligned} E\left(\hat{\lambda}_2^2(x)\right) &= E\left[\sum_{i<j} \sum \frac{\delta_{(i)}\delta_{(j)}}{(n-i+1)(n-j+1)} K_h(x-X_{(i)}) K_h(x-X_{(j)})\right] \\ &= \sum_{i<j} \sum E\left[\frac{\delta_{(i)}\delta_{(j)}}{(n-i+1)(n-j+1)} K_h(x-X_{(i)}) K_h(x-X_{(j)})\right]. \end{aligned}$$

Χρησιμοποιώντας την (3.36) λαμβάνουμε ότι:

$$\begin{aligned} E\left(\hat{\lambda}_2^2(x)\right) &= \sum_{i<j} \sum E\left\{E\left[\frac{\delta_{(i)}\delta_{(j)}}{(n-i+1)(n-j+1)} K_h(x-X_{(i)})\right.\right. \\ &\quad \left.\left.\times K_h(x-X_{(j)}) \mid X_{(i)}=y, X_{(j)}=z\right]\right\} \\ &= \sum_{i<j} \sum \int_0^\infty \int_0^\infty \frac{E(\delta_{(i)}\delta_{(j)} \mid X_{(i)}=y, X_{(j)}=z)}{(n-i+1)(n-j+1)} K_h(x-y) \\ &\quad \times K_h(x-z) f_{X_{(i)}, X_{(j)}}(y, z) dy dz, \end{aligned}$$

ή ισοδύναμα, μέσω του τύπου της μέσης τιμής για συνεχείς κατανομές, έχουμε ότι:

$$\begin{aligned} E\left(\hat{\lambda}_2^2(x)\right) &= \sum_{i<j} \sum \int_0^\infty \int_0^\infty \frac{K_h(x-y) K_h(x-z)}{(n-i+1)(n-j+1)} m(y)m(z) \\ &\quad \times f_{X_{(i)}, X_{(j)}}(y, z) dy dz. \end{aligned}$$

Αντικαθιστώντας την από κοινού σ.π.π. των $X_{(i)}, X_{(j)}$ που δίνεται από τη σχέση:

$$\begin{aligned} f_{X_{(i)}, X_{(j)}}(y, z) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(y) (F(z)-F(y))^{j-i-1} \\ &\quad \times (1-F(z))^{n-j} f(y)f(z)I_{(x,\infty)}, \end{aligned}$$

λαμβάνουμε ότι:

$$\begin{aligned}
E\left(\hat{\lambda}_2^2(x)\right) &= \sum_{i < j} \sum \int_0^\infty \int_0^\infty \frac{n!}{(n-i+1)(n-j+1)(i-1)!} \\
&\quad \times \frac{1}{(j-i-1)!(n-j)!} F^{i-1}(y) (F(z) - F(y))^{j-i-1} \\
&\quad \times (1 - F(z))^{n-j} m(y)m(z)K_h(x-y)K_h(x-z) \\
&\quad \times f(y)f(z) dy dz \\
&= \sum_{i < j} \sum \int_0^\infty \int_0^\infty \frac{n!}{(n-i+1)(n-j+1)(i-1)!} \\
&\quad \times \frac{1}{(j-i-1)!(n-j)!} F^{i-1}(y) (F(z) - F(y))^{j-i-1} \\
&\quad \times (1 - F(z))^{n-j} \frac{f_T(y)(1 - F_U(y))}{f(y)} \frac{f_T(z)(1 - F_U(z))}{f(z)} \\
&\quad \times K_h(x-y)K_h(x-z) f(y)f(z) dy dz \\
&= \sum_{i < j} \sum \int_0^\infty \int_0^\infty \frac{n!}{(n-i+1)(n-j+1)(i-1)!} \\
&\quad \times \frac{1}{(j-i-1)!(n-j)!} F^{i-1}(y) (F(z) - F(y))^{j-i-1} \\
&\quad \times (1 - F(z))^{n-j} \frac{f_T(y)}{1 - F_T(y)} \frac{f_T(z)}{1 - F_T(z)} (1 - F(y)) \\
&\quad \times (1 - F(z)) K_h(x-y)K_h(x-z) dy dz \\
&= \sum_{i < j} \sum \int_0^\infty \int_0^\infty \frac{n!}{(n-i+1)(n-j+1)(i-1)!} \\
&\quad \times \frac{1}{(j-i-1)!(n-j)!} F^{i-1}(y) (F(z) - F(y))^{j-i-1} \\
&\quad \times (1 - F(z))^{n-j+1} (1 - F(y)) \lambda_T(y)\lambda_T(z) \\
&\quad \times K_h(x-y)K_h(x-z) dy dz \\
&= \int_0^\infty \int_0^\infty \sum_{i < j} \sum \frac{n!}{(n-i+1)(n-j+1)(i-1)!} \\
&\quad \times \frac{1}{(j-i-1)!(n-j)!} F^{i-1}(y) (F(z) - F(y))^{j-i-1} \\
&\quad \times (1 - F(z))^{n-j+1} (1 - F(y)) \lambda_T(y)\lambda_T(z) \\
&\quad \times K_h(x-y)K_h(x-z) dy dz,
\end{aligned}$$

ή ισοδύναμα:

$$\begin{aligned}
E\left(\hat{\lambda}_2^2(x)\right) &= \int_0^\infty \int_0^\infty \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{n!}{(n-i+1)(n-j+1)(i-1)!} \\
&\quad \times \frac{1}{(j-i-1)!(n-j)!} F^{i-1}(y) (F(z) - F(y))^{j-i-1} \\
&\quad \times (1 - F(z))^{n-j+1} (1 - F(y)) \lambda_T(y) \lambda_T(z) \\
&\quad \times K_h(x-y) K_h(x-z) dy dz.
\end{aligned} \tag{3.51}$$

Έπειτα, υπολογίζουμε τη σχέση:

$$\begin{aligned}
Q(F(y), F(z)) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{n!}{(n-i+1)(n-j+1)(i-1)!(j-i-1)!(n-j)!} \\
&\quad \times F^{i-1}(y) (F(z) - F(y))^{j-i-1} (1 - F(z))^{n-j+1} (1 - F(y)) \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{n!}{(n-i+1)(n-j+1)(i-1)!(j-i-1)!(n-j)!} \\
&\quad \times F^{i-1}(y) (F(z) - F(y))^{j-i-1} (1 - F(z))^{n-j+1} (1 - F(y)),
\end{aligned}$$

όπου στη δεύτερη γραμμή αντικαθιστώντας τη σχέση:

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n X_i X_j = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_j,$$

και ακολουθώντας την ίδια διαδικασία με την απόδειξη της (3.35), προκύπτει το εξής αποτέλεσμα:

$$\begin{aligned}
Q(F(y), F(z)) &= \frac{1}{2} \left[1 - F^n(y) + F^n(z) - F^n(y)F^n(z) \right. \\
&\quad \left. - 2(1 - F(y)) \frac{F^n(z) - F^n(y)}{F(z) - F(y)} \right].
\end{aligned} \tag{3.52}$$

Λόγω της (3.52), η (3.51) γίνεται:

$$\begin{aligned}
E\left(\hat{\lambda}_2^2(x)\right) &= \int_0^\infty \int_0^\infty Q(F(y), F(z)) \lambda_T(y) \lambda_T(z) \\
&\quad \times K_h(x-y) K_h(x-z) dy dz,
\end{aligned}$$

ή ισοδύναμα:

$$\begin{aligned}
E\left(\hat{\lambda}_2^2(x)\right) &= \frac{1}{2} \int_0^\infty \int_0^\infty \left[1 - F^n(y) + F^n(z) - F^n(y)F^n(z) \right. \\
&\quad \left. - 2(1 - F(y)) \frac{F^n(z) - F^n(y)}{F(z) - F(y)} \right] \\
&\quad \times \lambda_T(y)\lambda_T(z)K_h(x-y)K_h(x-z) dy dz. \tag{3.53}
\end{aligned}$$

Αντικαθιστώντας τις (3.50) και (3.53) στην (3.48) λαμβάνουμε:

$$\begin{aligned}
E\left(\hat{\lambda}^2(x)\right) &= E\left(\hat{\lambda}_1^2(x)\right) + 2E\left(\hat{\lambda}_2^2(x)\right) \\
&= \int_0^\infty I_n(F(y)) \lambda_T(y)K_h^2(x-y) dy \\
&\quad + \int_0^\infty \int_0^\infty \left[1 - F^n(y) + F^n(z) - F^n(y)F^n(z) \right. \\
&\quad \left. - 2(1 - F(y)) \frac{F^n(z) - F^n(y)}{F(z) - F(y)} \right] \\
&\quad \times \lambda_T(y)\lambda_T(z)K_h(x-y)K_h(x-z) dy dz. \tag{3.54}
\end{aligned}$$

Συνεχίζοντας με τον δεύτερο όρο του δεξιού μέρους της (3.47) και χρησιμοποιώντας την (3.35) έχουμε

$$\begin{aligned}
\left[E\left(\hat{\lambda}(x)\right)\right]^2 &= \left(\int_0^\infty (1 - F^n(y)) \lambda_T(y)K_h(x-y) dy\right)^2 \\
&= \int_0^\infty \int_0^\infty (1 - F^n(y))(1 - F^n(z)) \lambda_T(y)\lambda_T(z) \\
&\quad \times K_h(x-y)K_h(x-z) dy dz \\
&= \int_0^\infty \int_0^\infty (1 - F^n(z) - F^n(y) + F^n(y)F^n(z)) \\
&\quad \times \lambda_T(y)\lambda_T(z) \times K_h(x-y)K_h(x-z) dy dz. \tag{3.55}
\end{aligned}$$

Η αντικατάσταση των (3.54) και (3.55) στην (3.47) ολοκληρώνει την απόδειξη. \square

Το παρακάτω λήμμα αποτελεί το Λήμμα 6 από την εργασία των Watson and Leadbetter (1964b) και δίνεται χωρίς απόδειξη.

Λήμμα 3.2.11. Για κάθε καθορισμένο b , με $0 \leq b < 1$ ισχύει ότι:

$$nI_n(F(y)) \rightarrow (1 - F(y))^{-1},$$

ομοιόμορφα για $0 \leq F(y) \leq b$.

Το θεώρημα που ακολουθεί είναι αντίστοιχο του Θεωρήματος 2 της εργασίας των Tanner and Wong (1983). Καθώς η απόδειξη του Θεωρήματος 2 σκιαγραφείται, το Θεώρημα 3.2.12 αποδεικνύεται αναλυτικά.

Θεώρημα 3.2.12. Για τη μεροληψία και τη διακύμανση του εκτιμητή πυρήνα της συνάρτησης κινδύνου λαμβάνουμε ότι:

$$E(\hat{\lambda}(x)) - \lambda_T(x) = \frac{h^2}{2} \lambda_T''(x) \mu_2(K) + o(h^2), \quad (3.56)$$

και

$$\text{Var}(\hat{\lambda}(x)) = \frac{1}{nh} \frac{\lambda(x)}{1 - F(x)} R(K) + o((nh)^{-1}), \quad (3.57)$$

όπου $\mu_2(K)$ είναι η δεύτερη ροπή του πυρήνα K και $R(K)$ είναι το ολοκλήρωμα της πραγματικής συνάρτησης K^2 σε όλο το πεδίο ορισμού της.

Απόδειξη. Αρχικά από την (3.35) έχουμε ότι:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \int_0^\infty (1 - F^n(y)) \lambda_T(y) K_h(x - y) dy \\ &= \int_0^\infty \lambda_T(y) K_h(x - y) dy - \int_0^\infty F^n(y) \lambda_T(y) K_h(x - y) dy. \end{aligned}$$

Όταν $n \rightarrow \infty$ και δοθέντος ότι $0 < F(y) < 1$, ο δεύτερος όρος της παραπάνω σχέσης είναι ασυμπτωτικά αμελητέος και για τον λόγο αυτόν δεν λαμβάνεται υπόψη.

Επομένως, για την παραπάνω σχέση έχουμε ότι:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \int_0^\infty \lambda_T(y) K_h(x - y) dy \\ &= \frac{1}{h} \int_0^\infty \lambda_T(y) K\left(\frac{x - y}{h}\right) dy \\ &= \int_0^\infty \lambda_T(x - ht) K(t) dt. \end{aligned}$$

Αναπτύσσοντας σε σειρά Taylor γύρω από το x τον όρο $\lambda_T(x - ht)$, λαμβάνουμε ότι:

$$\lambda_T(x - ht) = \lambda_T(x) - (ht)\lambda'_T(x) + \frac{(ht)^2}{2}\lambda''_T(x) + o(h^2).$$

Κατά συνέπεια, προκύπτει το εξής:

$$\begin{aligned} E(\hat{\lambda}(x)) &= \lambda_T(x) \int_0^\infty K(t) dt - h\lambda'_T(x) \int_0^\infty tK(t) dt + \frac{h}{2}\lambda''_T(x) \int_0^\infty t^2K(t) dt \\ &\quad + o(h^2) \\ &= \lambda_T(x) \int_0^\infty K(t) dt - h\lambda'_T(x)\mu_1(K) + \frac{h}{2}\lambda''_T(x)\mu_2(K) + o(h^2), \end{aligned}$$

για το οποίο, κάνοντας χρήση των ιδιοτήτων ότι ο πυρήνας ολοκληρώνει στη μονάδα και ότι η πρώτη ροπή του είναι ίση με μηδέν (Βλέπε Ορισμός 1.0.1) παίρνουμε ότι:

$$E(\hat{\lambda}(x)) = \lambda_T(x) + \frac{h}{2}\lambda''_T(x)\mu_2(K) + o(h^2),$$

πράγμα το οποίο ολοκληρώνει την απόδειξη της σχέσης (3.56).

Συνεχίζοντας με την απόδειξη της (3.57), παρατηρούμε αρχικά ότι ο δεύτερος όρος του δεξιού μέρους της διακύμανσης του εκτιμητή πυρήνα της συνάρτησης κινδύνου που δίνεται από την (3.45) μπορεί να παραληφθεί, καθώς όταν $n \rightarrow \infty$ αυτός θα τείνει στο μηδέν. Επομένως κρατώντας τον πρώτο όρο έχουμε ότι:

$$Var(\hat{\lambda}(x)) = \int I_n(F(y)) \lambda_T(y) K_h^2(x - y) dy. \quad (3.58)$$

Κάνοντας χρήση του Λήμματος 3.2.11 όπως επίσης και της (3.26), ελέγχουμε τη σύγκλιση της σχέσης που ακολουθεί:

$$\begin{aligned} nh \left(\int K^2(t) dt \right)^{-1} Var(\hat{\lambda}(x)) &= nh \left(\int K^2(t) dt \right)^{-1} \int I_n(F(y)) \lambda_T(y) \\ &\quad \times K_h^2(x - y) dy \\ &= h \left(\int K^2(t) dt \right)^{-1} \int nI_n(F(y)) \lambda_T(y) \\ &\quad \times K_h^2(x - y) dy, \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.58).

Για την παραπάνω σχέση, από το Λήμμα 3.2.11, προκύπτει ότι:

$$h \left(\int K^2(t) dt \right)^{-1} \int nI_n(F(y)) \lambda_T(y) K_h^2(x - y) dy \rightarrow (1 - F(y))^{-1} \lambda_T(x).$$

Κατα συνέπεια,

$$\text{Var}(\hat{\lambda}(x)) \rightarrow (nh)^{-1} \left(\int K^2(t) dt \right) \lambda_T(x) (1 - F(y))^{-1},$$

από το οποίο όταν $n \rightarrow \infty$, $h \rightarrow 0$ και $nh \rightarrow \infty$ λαμβάνουμε ότι:

$$\text{Var}(\hat{\lambda}(x)) = (nh)^{-1} \left(\int K^2(t) dt \right) \lambda_T(x) (1 - F(y))^{-1} + o((nh)^{-1}),$$

γεγονός το οποίο ολοκληρώνει την απόδειξη. \square

Παρατήρηση 3.2.13. Στο Θεώρημα 3.2.12, παρατηρούμε ότι η μεροληψία του εκτιμητή πυρήνα της συνάρτησης κινδύνου είναι τάξης h^2 ενώ η διακύμανσή του είναι τάξης $(nh)^{-1}$. Επομένως για μεγάλες τιμές δείγματος ($n \rightarrow \infty$), $h \rightarrow 0$ και $nh \rightarrow \infty$, ο εκτιμητής είναι ασυμπτωτικά αμερόληπτος και η διακύμανσή του συγκλίνει στο 0. Επιπλέον, αξιοσημείωτος είναι ο τρόπος με τον οποίο η μεροληψία εξαρτάται από την πραγματική συνάρτηση κινδύνου $\lambda(x)$. Η μεροληψία είναι μεγάλη όταν η απόλυτη τιμή της δεύτερης παραγώγου της συνάρτησης κινδύνου ($\lambda''(x)$) είναι μεγάλη. Τέλος, η διακύμανση του εκτιμητή $\hat{\lambda}(x)$ εξαρτάται από τη λογοκρισία σύμφωνα τη σχέση (3.14), η οποία υπάρχει στον παρονομαστή του κλάσματος $\lambda(x)/(1 - F(x))$.

Επίσης παρατηρούμε ότι, αυξάνοντας το εύρος ζώνης αυξάνεται και η μεροληψία ενώ αντίθετα η διακύμανση ελαττώνεται. Αυτό έχει ως συμπέρασμα ότι η βέλτιστη επιλογή του εύρους ζώνης για τον εκτιμητή πυρήνα της συνάρτησης κινδύνου βασίζεται στην εναλλαγή μεταξύ μεροληψίας και διακύμανσης (*bias-variance tradeoff*). Ένα πολύ μικρό εύρος ζώνης μπορεί να οδηγήσει σε υπερ-εξομάλυνση και υψηλή μεροληψία, ενώ ένα πολύ μεγάλο εύρος ζώνης μπορεί να οδηγήσει σε υπο-εξομάλυνση και υψηλή διακύμανση. Αξίζει να σημειωθεί ότι η εναλλαγή μεταξύ της μεροληψίας και της διακύμανσης δεν είναι η μοναδική παράμετρος που πρέπει να ληφθεί υπόψη κατά την επιλογή ενός εύρους ζώνης για έναν εκτιμητή ρυθμού επικινδυνότητας. Άλλοι παράγοντες, όπως η μορφή της υποκείμενης συνάρτησης κινδύνου και τα διαθέσιμα δεδομένα, μπορούν επίσης να επηρεάσουν την επιλογή του εύρους ζώνης.

3.2.3 Ασυμπτωτική κατανομή του εκτιμητή $\hat{\lambda}(x)$

Για την μελέτη της ασυμπτωτικής κατανομής του εκτιμητή $\hat{\lambda}(x)$, το κλασικό Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.) δεν μπορεί να εφαρμοστεί. Αυτό συμβαίνει

διότι, δοθείσης της ύπαρξης διατεταγμένων στατιστικών στον εκτιμητή, κάθε τυχαία μεταβλητή θα εξαρτάται από τις επόμενες ως εξής:

$$\frac{K_h(x - X_{(1)})\delta_{(1)}}{n - 1 + 1}, \frac{K_h(x - X_{(2)})\delta_{(2)}}{n - 2 + 1}, \dots, \frac{K_h(x - X_{(n)})\delta_{(n)}}{n - n + 1}.$$

Για να ξεπεραστεί το παραπάνω πρόβλημα, η ασυμπτωτική κατανομή του $\hat{\lambda}(x)$ προσδιορίζεται μέσω της μεθόδου της προβολής του Hajek και κάνοντας χρήση του Κ.Ο.Θ. του Lyapunov (βλέπε Billingsley (1995), σελ. 362). Η κεντρική ιδέα της μεθόδου αυτής έγκειται στο ότι η εφαρμοσιμότητα του Κ.Ο.Θ. μπορεί να επεκταθεί σε αθροίσματα τυχαίων μεταβλητών που είναι ασυμπτωτικά ισοδύναμα με αθροίσματα τυχαίων μεταβλητών που είναι ανεξάρτητα και ισόνομα.

Το περιεχόμενο αυτής της υποενότητας στηρίζεται στις εργασίες των Hájek (1968) και Tanner and Wong (1983).

Αρχικά, ο τύπος της προβολής που εισήγαγε ο Hájek (1968) στην αντίστοιχη εργασία του, όπως επίσης η μέση τιμή και η διακύμανση του αποτυπώνονται στο παρακάτω θεώρημα το οποίο και αποδεικνύουμε αναλυτικά.

Θεώρημα 3.2.14. Έστω Y_1, Y_2, \dots, Y_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές και W μία στατιστική συνάρτηση για την οποία ισχύει ότι $E(W^2) < \infty$. Τότε για τον τύπο προβολής,

$$\hat{W} = \sum_{i=1}^n E(W|Y_i) - (n-1)E(W), \quad (3.59)$$

ισχύει ότι:

$$E(\hat{W}) = E(W), \quad (3.60)$$

$$Var(\hat{W}) = Var(W) - E(W - \hat{W})^2. \quad (3.61)$$

Απόδειξη. Αρχικά θα αποδείξουμε ότι η (3.59) αποτελεί όντως τύπο προβολής, βασιζόμενοι στο κεφάλαιο 11 του συγγράμματος του Van der Vaart (2000).

Έστω Q το σύνολο που περιέχει όλα τα αθροίσματα της μορφής $\sum_{i=1}^n g_i(Y_i)$, $i = 1, \dots, n$, όπου g_i είναι μετρήσιμες συναρτήσεις με $E(g_i^2) < \infty$.

Σύμφωνα με το Θεώρημα 11.1 του συγγράμματος του Van der Vaart (2000), το \hat{W} είναι προβολή του W στο Q αν και μόνο αν $\hat{W} \in Q$ και το $W - \hat{W}$ είναι ορθογώνιο στο Q , το οποίο ισχύει όταν:

$$E\left[(W - \hat{W}) \sum_{i=1}^n g_i(Y_i)\right] = 0. \quad (3.62)$$

Δοθείσης της φύσης της (3.59), είναι προφανές ότι $\hat{W} \in \mathcal{Q}$. Επομένως, αρκεί να δειχθεί η (3.62).

$$\begin{aligned}
E \left[\left(W - \hat{W} \right) \sum_{i=1}^n g_i(Y_i) \right] &= \sum_{i=1}^n E \left[\left(W - \hat{W} \right) g_i(Y_i) \right] \\
&= \sum_{i=1}^n E \left\{ E \left[\left(W - \hat{W} \right) g_i(Y_i) | Y_i \right] \right\} \\
&= \sum_{i=1}^n E \left\{ g_i(Y_i) E \left[\left(W - \hat{W} \right) | Y_i \right] \right\} \\
&= \sum_{i=1}^n E \left\{ g_i(Y_i) \left[E(W|Y_i) - E(\hat{W}|Y_i) \right] \right\}, \quad (3.63)
\end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.36) και στην τρίτη γραμμή η (3.38).

Επιπλέον, ισχύει ότι:

$$\begin{aligned}
E(\hat{W}|Y_i) &= E \left\{ \left[\sum_{j=1}^n E(W|Y_j) - (n-1)E(W) \right] | Y_i \right\} \\
&= E \left\{ \left[\sum_{j=1}^n E(W|Y_j) \right] | Y_i \right\} - E \{ [(n-1)E(W)] | Y_i \} \\
&= \sum_{j=1}^{n-1} E [E(W|Y_j) | Y_i] + E [E(W|Y_i) | Y_i] - (n-1)E(W) \\
&= \sum_{j=1}^{n-1} E(W) + E(W|Y_i) - (n-1)E(W) \\
&= (n-1)E(W) + E(W|Y_i) - (n-1)E(W) \\
&= E(W|Y_i). \quad (3.64)
\end{aligned}$$

Αντικαθιστώντας την (3.64) στην (3.63), αποδεικνύεται η (3.62).

Συνεχίζοντας με την απόδειξη της (3.60) έχουμε ότι:

$$E(\hat{W}) = E \left[\sum_{i=1}^n E(W|Y_i) - (n-1)E(W) \right],$$

η οποία ισοδύναμα γίνεται:

$$\begin{aligned}
 E(\hat{W}) &= E\left[\sum_{i=1}^n E(W|Y_i)\right] - (n-1)E[E(W)] \\
 &= E[nE(W|Y_1)] - (n-1)E(W) \\
 &= nE[E(W|Y_1)] - (n-1)E(W) \\
 &= nE(W) - nE(W) + E(W) = E(W),
 \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε το γεγονός ότι εφόσον οι Y_i είναι ανεξάρτητες και ισόνομες τ.μ. τότε:

$$E(W|Y_1) = E(W|Y_2) = \dots = E(W|Y_n).$$

Τέλος, για $G = \sum_{i=1}^n g_i(Y_i)$ έχουμε ότι:

$$\begin{aligned}
 E(W - G)^2 &= E\left((W - \hat{W} + \hat{W} - G)^2\right) \\
 &= E\left[\left(W - \hat{W}\right)^2 + 2\left(W - \hat{W}\right)\left(\hat{W} - G\right) + \left(\hat{W} - G\right)^2\right] \\
 &= E\left(W - \hat{W}\right)^2 + 2E\left[\left(W - \hat{W}\right)\left(\hat{W} - G\right)\right] \\
 &\quad + E\left(\hat{W} - G\right)^2. \tag{3.65}
 \end{aligned}$$

Για τον μεσαίο όρο του δεξιού μέρους της (3.65) ισχύει ότι:

$$\begin{aligned}
 E\left[\left(W - \hat{W}\right)\left(\hat{W} - G\right)\right] &= E\left\{E\left[\left(W - \hat{W}\right)\left(\hat{W} - G\right)|Y_i]\right\} \\
 &= E\left\{E\left[\left(W - \hat{W}\right)|Y_i]\right. \right. \\
 &\quad \left. \left. \times E\left[\left(\hat{W} - G\right)|Y_i]\right\}\right\}. \tag{3.66}
 \end{aligned}$$

Η παραπάνω σχέση προέκυψε αντικαθιστώντας την (3.36) στην πρώτη γραμμή και λαμβάνοντας υπόψη στη δεύτερη γραμμή ότι $(W - \hat{W})$ και $(\hat{W} - G)$ είναι μεταξύ τους ανεξάρτητες ως συναρτήσεις ανεξάρτητων τ.μ.

Στη συνέχεια παρατηρούμε ότι λόγω της (3.64) ισχύει ότι:

$$E \left[(W - \hat{W}) | Y_i \right] = E(W | Y_i) - E(\hat{W} | Y_i) = 0,$$

το οποίο αν αντικατασταθεί στην (3.66) λαμβάνουμε ότι:

$$E \left[(W - \hat{W}) (\hat{W} - G) \right] = 0,$$

και κατά συνέπεια η (3.65) γίνεται:

$$E(W - G)^2 = E(W - \hat{W})^2 + E(\hat{W} - G)^2. \quad (3.67)$$

Επιλέγοντας, χωρίς βλάβη της γενικότητας, ότι:

$$G = EW = E\hat{W} = 0,$$

και χρησιμοποιώντας τον τύπο που συνδέει τη διακύμανση με τη μέση τιμή, μέσω της (3.67) αποδεικνύεται η (3.61). \square

Προσαρμόζοντας τα παραπάνω στην περίπτωση του εκτιμητή $\hat{\lambda}(x)$, θέτουμε

$$Y_i = (X_i, \delta_i), \quad i = 1, \dots, n,$$

οι οποίες είναι ανεξάρτητες και ισόνομες τ.μ. και

$$W = \hat{\lambda}(x) = \sum_{j=1}^n W_j = \sum_{j=1}^n \frac{K_h(x - X_j) \delta_j}{n - R_j + 1}, \quad (3.68)$$

τον εκτιμητή της συνάρτησης κινδύνου που δίνεται από την (3.30).

Για τον πυρήνα $K(\cdot)$ της (3.68) ισχύουν οι συνθήκες που δόθηκαν στον Ορισμό 1.0.1, ενώ επιπλέον υιοθετούμε και ότι υπάρχει αρκετά μικρό εύρος ζώνης h τέτοιο ώστε το $h^{-1}K(h^{-1}(y-x))/(1-F(y))$, με $0 < F(y) < 1$, να είναι ομοιόμορφα φραγμένο για $|y-x| > M$, για κάθε $M > 0$. Στην παρούσα ενότητα το φράγμα αυτό θα συμβολίζεται με G_M .

Σκοπός είναι να δείξουμε ότι η τυποποιημένη μορφή της (3.68) ακολουθεί ασυμπτωτικά την τυπική κανονική κατανομή, όταν $n \rightarrow \infty, h \rightarrow 0, nh \rightarrow \infty$. Για να αποδειχθεί αυτό μέσω της μεθόδου της προβολής του Hajek, θα πρέπει αρχικά να δείξουμε ότι η τυποποιημένη μορφή της προβολής \hat{W} είναι ασυμπτωτικά ισόδυναμη με την τυποποιημένη μορφή του W και στη συνέχεια ότι ακολουθεί την

τυπική κανονική κατανομή. Κατά αυτόν τον τρόπο θα έχει αποδειχθεί ότι και η W θα ακολουθεί την τυπική κανονική κατανομή.

Το επόμενο λήμμα, το οποίο και αποδεικνύουμε αναλυτικά, περιέχει δύο σχέσεις οι οποίες, σύμφωνα με την εργασία των Tanner and Wong (1983), απαιτούνται για την απόδειξη της ασυμπτωτικής κατανομής του $\hat{\lambda}(x)$.

Λήμμα 3.2.15. Δοθέντος των (X_i, δ_i) ισχύει ότι:

$$R_i \sim 1 + \text{Bin}(n-1, F(X_i)), \quad (3.69)$$

και δοθέντος των (X_i, δ_i) και (X_j, δ_j) ισχύει ότι:

$$R_j = \begin{cases} 1 + \text{Bin}(n-2, F(X_j)), & \text{όταν } X_j \leq X_i \\ 2 + \text{Bin}(n-2, F(X_j)), & \text{όταν } X_j > X_i, \end{cases} \quad (3.70)$$

όπου Bin είναι η Διωνυμική κατανομή.

Απόδειξη. Από την (3.31) γνωρίζουμε ότι:

$$R_i = \sum_{j=1}^n I_{\{X_j \leq X_i\}} \quad \text{και} \quad R_j = \sum_{i=1}^n I_{\{X_i \leq X_j\}},$$

με τις $I_{\{X_j \leq X_i\}}$ και $I_{\{X_i \leq X_j\}}$ να είναι δείκτριες συναρτήσεις και κατά συνέπεια Bernoulli τυχαίες μεταβλητές με πιθανότητα επιτυχίας $p_i = F(X_i)$ και $p_j = F(X_j)$ αντίστοιχα. Επομένως τα επιμέρους αθροίσματα των συναρτήσεων αυτών θα ακολουθούν τη διωνυμική κατανομή με n και πιθανότητες επιτυχίας $p_i = F(X_i)$ και $p_j = F(X_j)$.

Ξεκινώντας από την (3.69) έχουμε το εξής:

$$R_i = \sum_{j=1}^n I_{\{X_j \leq X_i\}} = 1 + \sum_{j \neq i}^{n-1} I_{\{X_j \leq X_i\}},$$

γεγονός που αποδεικνύει το ζητούμενο.

Συνεχίζοντας με την (3.70) στην περίπτωση που $X_j \leq X_i$,

$$\begin{aligned} R_j &= \sum_{i=1}^n I_{\{X_i \leq X_j\}} = 1 + \sum_{i \neq j}^{n-1} I_{\{X_i \leq X_j\}} = 1 + 0 + \sum_{i \neq j}^{n-2} I_{\{X_i \leq X_j\}} \\ &= 1 + \sum_{i \neq j}^{n-2} I_{\{X_i \leq X_j\}}, \end{aligned}$$

αποδεικνύοντας το ζητούμενο.

Τέλος για την περίπτωση που $X_i < X_j$, ισχύει ότι:

$$R_j = \sum_{i=1}^n I_{\{X_i \leq X_j\}} = 1 + 1 + \sum_{i \neq j}^{n-2} I_{\{X_i < X_j\}} = 2 + \sum_{i \neq j}^{n-2} I_{\{X_i < X_j\}},$$

το οποίο ολοκληρώνει την απόδειξη. \square

Το λήμμα που ακολουθεί είναι αντίστοιχο του Λήμματος 2 της εργασίας των Tanner and Wong (1983). Καθώς στην εργασία των Tanner and Wong (1983), η απόδειξη του Λήμματος 2 σκιαγραφείται, το Λήμμα 3.2.16 αποδεικνύεται αναλυτικά.

Λήμμα 3.2.16.

$$E(W_i | Y_i) = n^{-1} V_n(Y_i), \quad (3.71)$$

όπου:

$$V_n(Y_i) = (1 - F(X_i))^{-1} (1 - F^n(X_i)) \delta_i K_h(x - X_i),$$

και για $j \neq i$ ισχύει ότι:

$$\begin{aligned} E(W_j | Y_i) &= (n-1)^{-1} \int (1 - F(y))^{-1} (1 - F^{n-1}(y)) m(y) \\ &\quad \times K_h(x - y) f(y) dy + (n(n-1))^{-1} U_n(Y_i), \end{aligned} \quad (3.72)$$

όπου:

$$\begin{aligned} U_n(Y_i) &= - \int (1 - F(y))^{-2} [1 - F^n(y) - nF^{n-1}(y)(1 - F(y))] \\ &\quad \times I_{\{y \leq X_i\}} m(y) f(y) K_h(x - y) dy. \end{aligned}$$

Απόδειξη. Από την (3.68) έχουμε ότι:

$$W_i = \frac{K_h(x - X_i) \delta_i}{n - R_i + 1}.$$

Επομένως,

$$\begin{aligned} E(W_i | Y_i) &= E \left[\frac{K_h(x - X_i) \delta_i}{n - R_i + 1} | Y_i \right] \\ &= K_h(x - X_i) \delta_i E \left[\frac{1}{n - (R_i - 1)} | Y_i \right]. \end{aligned} \quad (3.73)$$

Ασχολούμαστε με την αναμενόμενη τιμή του δεξιού μέρους της (3.73), λαμβάνοντας υπόψη από την (3.69) ότι:

$$R_i - 1 \sim \text{Bin}(n - 1, F(X_i)).$$

Άρα έχουμε ότι:

$$\begin{aligned} E \left[\frac{1}{n - (R_i - 1)} | Y_i \right] &= \sum_{k=1}^n \frac{1}{n - k + 1} f_{R_i - 1}(X_i) \\ &= \sum_{k=1}^n \frac{1}{n - k + 1} \binom{n-1}{k-1} F^{k-1}(X_i) \\ &\quad \times (1 - F(X_i))^{n-k} \\ &= \sum_{k=1}^n \frac{1}{n} \binom{n}{k-1} F^{k-1}(X_i) (1 - F(X_i))^{n-k} \\ &= \frac{1}{n} \sum_{k=1}^n \binom{n}{k-1} F^{k-1}(X_i) (1 - F(X_i))^{n-k} \\ &= \frac{1}{n} \left(\frac{1 - F^n(X_i)}{1 - F(X_i)} \right), \end{aligned} \quad (3.74)$$

όπου στην πρώτη γραμμή εφαρμόστηκε ο τύπος της μέσης τιμής για διακριτές κατανομές, στη δεύτερη γραμμή αντικαταστάθηκε η συνάρτηση πιθανότητας (σ.π.) της $R_i - 1$ που δίνεται από τη σχέση:

$$f_{R_i - 1}(X_i) = \binom{n-1}{k-1} F^{k-1}(X_i) (1 - F(X_i))^{n-k},$$

στην τρίτη γραμμή εφαρμόστηκε η σχέση:

$$\frac{1}{n - k + 1} \binom{n-1}{k-1} = \frac{1}{n} \binom{n}{k-1},$$

και τέλος στην τελευταία γραμμή αντικαταστάθηκε η (3.44).

Εφαρμόζοντας την (3.74) στην (3.73) λαμβάνουμε ότι:

$$E(W_i | Y_i) = \frac{1}{n} \left(\frac{1 - F^n(X_i)}{1 - F(X_i)} \right) K_h(x - X_i) \delta_i = n^{-1} V_n(Y_i).$$

Συνεχίζουμε με την απόδειξη της (3.72) λαμβάνοντας υπόψη από την (3.68) ότι:

$$W_j = \frac{K_h(x - X_j) \delta_j}{n - R_j + 1},$$

και από την (3.70) ότι:

$$R_i - 1 \sim \text{Bin}(n - 2, F(X_j)) \quad \text{όταν} \quad X_j \leq X_i,$$

και

$$R_j - 2 \sim \text{Bin}(n - 2, F(X_j)) \quad \text{όταν} \quad X_j > X_i.$$

Έστω $Y_{(i)}$ και $Y_{(j)}$ για τα οποία ισχύει ότι:

$$Y_{(i)} = (X_{(i)}, \delta_{(i)}) \quad \text{και} \quad Y_{(j)} = (X_{(j)}, \delta_{(j)}) \quad \text{με} \quad i, j = 1, \dots, n \quad \text{και} \quad i \neq j,$$

όπου $X_{(i)}$ και $X_{(j)}$ η i -οστή και j -οστή διατεταγμένη παρατήρηση αντίστοιχα.

$$\begin{aligned} E(W_j | Y_i) &= E[E(W_j | Y_i) | Y_j = Y_{(j)}, Y_i = Y_{(i)}] \\ &= E\left[E\left(\frac{K_h(x - X_{(j)}) \delta_{(j)}}{n - R_j + 1} | Y_{(i)}\right) | X_j \leq X_i\right] \\ &\quad + E\left[E\left(\frac{K_h(x - X_{(j)}) \delta_{(j)}}{n - R_j + 1} | Y_{(i)}\right) | X_j > X_i\right] \\ &\equiv S_1 + S_2. \end{aligned} \tag{3.75}$$

Ασχολούμαστε αρχικά με τον πρώτο όρο του δεξιού μέρους της (3.75).

Για $X_j \leq X_i$ ισχύει ότι:

$$R_j - 1 \sim \text{Bin}(n - 2, F(X_j)),$$

με σ.π.

$$f_{R_j-1}(X_j) = \binom{n-2}{k-1} F^{k-1}(X_j) (1 - F(X_j))^{n-k-1}. \tag{3.76}$$

Επομένως έχουμε ότι:

$$\begin{aligned} S_1 &= \sum_{k=1}^n \int_0^\infty \frac{K_h(x-y) E(\delta_{(j)} | X_{(j)} = y)}{n-k+1} \binom{n-2}{k-1} F^{k-1}(y) \\ &\quad \times (1 - F(y))^{n-k-1} f(y) dy, \end{aligned} \tag{3.77}$$

στην οποία αντικαταστήσαμε την (3.76) και τον τύπο της μέσης τιμής.

Αντικαθιστώντας επίσης την (3.33), η (3.77) γίνεται:

$$\begin{aligned}
S_1 &= \sum_{k=1}^n \int_0^\infty \frac{K_h(x-y) m(y)}{n-k+1} \binom{n-2}{k-1} F^{k-1}(y) \\
&\quad \times (1-F(y))^{n-k-1} f(y) dy \\
&= \int_0^\infty \sum_{k=1}^n \frac{1}{n-k+1} \binom{n-2}{k-1} F^{k-1}(y) (1-F(y))^{n-k-1} \\
&\quad \times K_h(x-y) m(y) f(y) dy \\
&= \int_0^\infty \sum_{i=0}^{n-1} \frac{1}{n-i} \binom{n-2}{i} F^i(y) (1-F(y))^{n-i-2} \\
&\quad \times K_h(x-y) m(y) f(y) dy \\
&= \int_0^\infty \sum_{i=0}^{n-1} \left(\frac{1}{n} \binom{n}{i} - \frac{i}{n(n-1)} \binom{n}{i} \right) F^i(y) (1-F(y))^{n-i-2} \\
&\quad \times K_h(x-y) m(y) f(y) dy \\
&= \int_0^\infty \left[\sum_{i=0}^{n-1} \frac{1}{n} \binom{n}{i} F^i(y) (1-F(y))^{n-i-2} \right. \\
&\quad \left. - \sum_{i=0}^{n-1} \frac{i}{n(n-1)} \binom{n}{i} F^i(y) (1-F(y))^{n-i-2} \right] \\
&\quad \times K_h(x-y) m(y) f(y) dy \\
&= (1-F(y))^{-2} \int_0^\infty \left[\sum_{i=0}^{n-1} \frac{1}{n} \binom{n}{i} F^i(y) (1-F(y))^{n-i} \right. \\
&\quad \left. - \sum_{i=0}^{n-1} \frac{i}{n(n-1)} \binom{n}{i} F^i(y) (1-F(y))^{n-i} \right] \\
&\quad \times K_h(x-y) m(y) f(y) dy, \tag{3.78}
\end{aligned}$$

όπου στην τρίτη γραμμή έγινε αλλαγή μεταβλητής $i = k - 1$ και στην τέταρτη γραμμή χρησιμοποιήθηκε η σχέση:

$$\frac{1}{n-i} \binom{n-2}{i} = \frac{1}{n} \binom{n}{i} - \frac{i}{n(n-1)} \binom{n}{i}.$$

Κάνοντας χρήση του διωνυμικού θεωρήματος, υπολογίζουμε αρχικά ότι:

$$\begin{aligned} \sum_{i=0}^{n-1} \frac{1}{n} \binom{n}{i} F^i(y) (1 - F(y))^{n-i} &= \frac{1}{n} \sum_{i=0}^{n-1} \binom{n}{i} F^i(y) (1 - F(y))^{n-i} \\ &= \frac{1}{n} (1 - F^n(y)), \end{aligned} \quad (3.79)$$

για το οποίο στη δεύτερη γραμμή χρησιμοποιήθηκε το αποτέλεσμα που βρέθηκε στην τέταρτη γραμμή της (3.44).

Θέτοντας:

$$N_1 = \sum_{i=0}^{n-1} \frac{i}{n(n-1)} \binom{n}{i} F^i(y) (1 - F(y))^{n-i},$$

υπολογίζουμε ότι:

$$\begin{aligned} N_1 &= \frac{1}{n(n-1)} \sum_{i=0}^{n-1} i \binom{n}{i} F^i(y) (1 - F(y))^{n-i} \\ &= \frac{1}{n(n-1)} \left[0 + \sum_{i=1}^{n-1} i \binom{n}{i} F^i(y) (1 - F(y))^{n-i} \right] \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n-1} i \binom{n}{i} F^i(y) (1 - F(y))^{n-i} \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n-1} n \binom{n-1}{i-1} F^i(y) (1 - F(y))^{n-i} \\ &= \frac{1}{n-1} \sum_{z=0}^{n-2} \binom{n-1}{z} F^{z+1}(y) (1 - F(y))^{n-z-1} \\ &= \frac{F(y)}{m} \sum_{z=0}^{m-1} \binom{m}{z} F^z(y) (1 - F(y))^{m-z} \\ &= \frac{1}{m} F(y) (1 - F^m(y)) \\ &= \frac{1}{n-1} F(y) (1 - F^{n-1}(y)), \end{aligned} \quad (3.80)$$

στην πέμπτη γραμμή έγινε η αλλαγή μεταβλητής $z = i - 1$, στην έκτη γραμμή έγινε η αλλαγή μεταβλητής $m = n - 1$ και στη τελευταία γραμμή χρησιμοποιήθηκε το αποτέλεσμα που προέκυψε στην τέταρτη γραμμή της (3.44).

Συνεπώς, εφαρμόζοντας τις (3.79) και (3.80) στην (3.78) προκύπτει ότι:

$$\begin{aligned}
S_1 &= (1 - F(y))^{-2} \int_0^\infty \left[\frac{1}{n} (1 - F^n(y)) \right. \\
&\quad \left. - \frac{1}{n-1} F(y) (1 - F^{n-1}(y)) \right] \\
&\quad \times K_h(x-y) m(y) f(y) dy \\
&= \frac{1}{n(n-1)} \int_0^\infty (1 - F(y))^{-2} [-1 + F^n(y) \\
&\quad + nF^{n-1}(y) (1 - F(y))] \\
&\quad \times I_{(y \leq X_i)} K_h(x-y) m(y) f(y) dy \\
&= -n(n-1)^{-1} \int_0^\infty (1 - F(y))^{-2} [1 - F^n(y) \\
&\quad - nF^{n-1}(y) (1 - F(y))] \\
&\quad \times I_{(y \leq X_i)} K_h(x-y) m(y) f(y) dy \\
&= n(n-1)^{-1} U_n(Y_i). \tag{3.81}
\end{aligned}$$

Με παρόμοια διαδικασία με προηγούμενους, ασχολούμαστε με τον δεύτερο όρο του δεξιού μέρους της (3.75). Αρχικά για $X_j > X_i$ ισχύει ότι:

$$R_j - 2 \sim \text{Bin}(n-2, F(X_j)),$$

με σ.π.

$$f_{R_j-2}(X_j) = \binom{n-2}{k-2} F^{k-2}(X_j) (1 - F(X_j))^{n-k}. \tag{3.82}$$

Επομένως,

$$\begin{aligned}
S_2 &= E \left[E \left(\frac{K_h(x - X_{(j)}) \delta_{(j)}}{n - R_j + 1} \middle| Y_{(i)} \right) \middle| X_j > X_i \right] \\
&= \sum_{k=2}^n \int_0^\infty \frac{K_h(x-y) E(\delta_{(j)} | X_{(j)} = y)}{n-k+1} \binom{n-2}{k-2} F^{k-2}(y) \\
&\quad \times (1 - F(y))^{n-k} f(y) dy,
\end{aligned}$$

όπου στο δεύτερο βήμα χρησιμοποιήθηκε η (3.82).

Ισοδύναμα,

$$S_2 = \int_0^\infty \sum_{k=2}^n \frac{1}{n-k+1} \binom{n-2}{k-2} F^{k-2}(y) (1-F(y))^{n-k} \\ \times K_h(x-y) m(y) f(y) dy, \quad (3.83)$$

η οποία προκύπτει αντικαθιστώντας στην προηγούμενη σχέση την (3.82).

Θέτοντας:

$$N_2 = \sum_{k=2}^n \frac{1}{n-k+1} \binom{n-2}{k-2} F^{k-2}(y) (1-F(y))^{n-k},$$

υπολογίζουμε ότι:

$$\begin{aligned} N_2 &= \sum_{i=0}^{n-2} \frac{1}{n-i-1} \binom{n-2}{i} F^i(y) (1-F(y))^{n-i-2} \\ &= (1-F(y))^{-2} \sum_{i=0}^{n-2} \frac{1}{n-i-1} \binom{n-2}{i} F^i(y) (1-F(y))^{n-i} \\ &= (1-F(y))^{-2} \sum_{i=0}^{n-2} \frac{1}{n-1} \binom{n-1}{i} F^i(y) (1-F(y))^{n-i} \\ &= \frac{(1-F(y))^{-2}}{n-1} \sum_{i=0}^{n-2} \binom{n-1}{i} F^i(y) (1-F(y))^{n-i} \\ &= \frac{(1-F(y))^{-2}}{m} \sum_{i=0}^{m-1} \binom{m}{i} F^i(y) (1-F(y))^{m-i+1} \\ &= \frac{(1-F(y))^{-1}}{m} \sum_{i=0}^{m-1} \binom{m}{i} F^i(y) (1-F(y))^{m-i} \\ &= \frac{(1-F(y))^{-1}}{m} (1-F^m(y)) \\ &= \frac{(1-F(y))^{-1}}{n-1} (1-F^{n-1}(y)), \end{aligned} \quad (3.84)$$

όπου στην πρώτη γραμμή έγινε η αλλαγή μεταβλητής $i = k-2$, στην τρίτη γραμμή χρησιμοποιήθηκε ότι:

$$\frac{1}{n-i-1} \binom{n-2}{i} = \frac{1}{n-1} \binom{n-1}{i},$$

στην πέμπτη γραμμή έγινε η αλλαγή μεταβλητής $m = n - 1$ και στην έβδομη γραμμή χρησιμοποιήθηκε το αποτέλεσμα που βρέθηκε στην τέταρτη γραμμή της (3.44).

Αντικαθιστώντας την (3.84) στην (3.83) λαμβάνουμε το εξής:

$$S_2 = (n - 1)^{-1} \int_0^\infty (1 - F(y))^{-1} (1 - F^{n-1}(y)) \times K_h(x - y) m(y) f(y) dy. \quad (3.85)$$

Τέλος, αντικαθιστώντας τις (3.81) και (3.85) στην (3.75) λαμβάνουμε ότι:

$$E(W_j|Y_i) = (n - 1)^{-1} \int_0^\infty (1 - F(y))^{-1} (1 - F^{n-1}(y)) K_h(x - y) \times m(y) f(y) dy + (n(n - 1))^{-1} U_n(Y_i),$$

το οποίο και ολοκληρώνει την απόδειξη. \square

Μέσω των Λημμάτων 3.2.17 και 3.2.18 που ακολουθούν, παραθέτουμε δύο σχέσεις που υπάρχουν στην εργασία των Tanner and Wong (1983). Οι σχέσεις αυτές είναι χρήσιμες για την απόδειξη της ασυμπτωτικής κατανομής του εκτιμητή $\hat{\lambda}(x)$ και δυνάμει ότι η απόδειξή τους δεν είναι διαθέσιμη, τις αποδεικνύουμε αναλυτικά.

Λήμμα 3.2.17.

$$E(W|Y_i) = E(W_i|Y_i) + (n - 1)E(W_j|Y_i).$$

Απόδειξη. Από την (3.68) έχουμε ότι:

$$\begin{aligned} E(W|Y_i) &= E\left(\sum_{j=1}^n W_j|Y_i\right) = E\left[\left(W_i + \sum_{j \neq i}^{n-1} W_j\right) | Y_i\right] \\ &= E(W_i|Y_i) + E\left(\sum_{j \neq i}^{n-1} W_j|Y_i\right) = E(W_i|Y_i) + \sum_{j \neq i}^{n-1} E(W_j|Y_i) \\ &= E(W_i|Y_i) + (n - 1)E(W_j|Y_i), \end{aligned}$$

\square

Χρησιμοποιώντας τον τύπο της προβολής του Hajek που δίνεται από την (3.59), το Λήμμα 3.2.16, το Λήμμα 3.2.17 και το Θεώρημα 3.2.9 αποδεικνύεται το παρακάτω λήμμα.

Λήμμα 3.2.18.

$$\hat{W} - E\hat{W} = \sum_{i=1}^n [n^{-1}V_n(Y_i) + n^{-1}U_n(Y_i) + \Delta_n], \quad (3.86)$$

όπου:

$$\Delta_n = - \int_0^{\infty} F^{n-1}(y)K_h(x-y)m(y)f(y) dy. \quad (3.87)$$

Απόδειξη.

$$\begin{aligned} \hat{W} - E\hat{W} &= \sum_{i=1}^n E(W|Y_i) - (n-1)E(W) - E(\hat{W}) \\ &= \sum_{i=1}^n E(W|Y_i) - (n-1)E(W) - E(W) \\ &= \sum_{i=1}^n E(W|Y_i) - nE(W) = \sum_{i=1}^n [E(W|Y_i) - E(W)] \\ &= \sum_{i=1}^n [E(W_i|Y_i) + (n-1)E(W_j|Y_i) - E(W)] \\ &= \sum_{i=1}^n \left[n^{-1}V_n(Y_i) + n^{-1}U_n(Y_i) \right. \\ &\quad \left. + \int_0^{\infty} (1-F(y))^{-1} (1-F^{n-1}(y)) \right. \\ &\quad \left. \times K_h(x-y)m(y)f(y) dy - E(W) \right], \end{aligned} \quad (3.88)$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκε η (3.59), στη δεύτερη γραμμή η (3.60) και στην τρίτη γραμμή οι (3.71) και (3.72).

Ασχολούμαστε με τη διαφορά του τρίτου και του τέταρτου όρου του αθροίσματος του δεξιού μέρους της (3.88). Θέτουμε:

$$G = \int_0^{\infty} (1-F(y))^{-1} (1-F^{n-1}(y)) K_h(x-y)m(y)f(y) dy - E(W).$$

Επομένως, λαμβάνουμε ότι:

$$\begin{aligned} G &= \int_0^{\infty} (1-F(y))^{-1} (1-F^{n-1}(y)) K_h(x-y)m(y)f(y) dy \\ &\quad - \int_0^{\infty} (1-F(y))^{-1} (1-F^n(y)) K_h(x-y)m(y)f(y) dy, \end{aligned}$$

ή ισοδύναμα ότι:

$$\begin{aligned}
G &= - \int_0^\infty \left(\frac{F^{n-1}(y) + F^n(y)}{1 - F(y)} \right) K_h(x - y) m(y) f(y) dy \\
&= - \int_0^\infty \frac{F^{n-1}(y) (1 - F(y))}{1 - F(y)} K_h(x - y) m(y) f(y) dy \\
&= - \int_0^\infty F^{n-1}(y) K_h(x - y) m(y) f(y) dy \\
&= \Delta_n.
\end{aligned}$$

Αντικαθιστώντας την παραπάνω σχέση στην (3.88) λαμβάνουμε την (3.86). \square

Τα Λήμματα 3.2.19 έως 3.2.21 στηρίζονται στο Θεώρημα 3 της εργασίας των Tanner and Wong (1983). Καθώς η απόδειξη του Θεωρήματος 3 σκιαγραφείται, τα Λήμματα 3.2.19 έως 3.2.21 αποδεικνύονται αναλυτικά.

Λήμμα 3.2.19. Για $n \rightarrow \infty, h \rightarrow 0$ και $nh \rightarrow \infty$ ισχύουν τα παρακάτω:

$$|U_n(Y_i)| = O(\log n), \quad (3.89)$$

$$|\Delta_n| = O((n(n+1))^{-1}). \quad (3.90)$$

Απόδειξη. Έστω M τέτοιο ώστε $F(x + M) < 1$.

Από το Λήμμα 3.2.16 έχουμε ότι:

$$\begin{aligned}
-U_n(Y_i) &= \int_0^\infty (1 - F(y))^{-2} [1 - F^n(y) - nF^{n-1}(y)(1 - F(y))] \\
&\quad \times I_{(y \leq X_i)} K_h(x - y) m(y) f(y) dy \\
&= \int_0^\infty \left[\frac{1 - F^n(y)}{1 - F(y)} - nF^{n-1}(y) \right] (1 - F(y))^{-1} \\
&\quad \times I_{(y \leq X_i)} K_h(x - y) m(y) dF(y),
\end{aligned}$$

για το οποίο λαμβάνοντας υπόψιν το φράγμα G_M που δόθηκε στη σελίδα 63 παίρνουμε ότι:

$$\begin{aligned}
-U_n(Y_i) &\leq \int_{|y-x| \leq M} \left[\frac{1 - F^n(y)}{1 - F(y)} - nF^{n-1}(y) \right] dF(y) \\
&\quad + G_M \int_{|y-x| > M} \left[\frac{1 - F^n(y)}{1 - F(y)} - nF^{n-1}(y) \right] dF(y). \quad (3.91)
\end{aligned}$$

Θέτοντας:

$$d_n = \int_{|y-x| \leq M} \left[\frac{1 - F^n(y)}{1 - F(y)} - nF^{n-1}(y) \right] dF(y),$$

λαμβάνουμε ότι:

$$d_n = O(1),$$

καθώς,

$$\begin{aligned} d_n &= \int_{|y-x| \leq M} \left[\frac{1 - F^n(y)}{1 - F(y)} - nF^{n-1}(y) \right] dF(y) \\ &= \int_{|y-x| \leq M} \frac{1}{1 - F(y)} dF(y) - \int_{|y-x| \leq M} \frac{F^n(y)}{1 - F(y)} dF(y) \\ &\quad - \int_{|y-x| \leq M} nF^{n-1}(y) dF(y) \\ &= -\log(1 - F(y)) - \int_{|y-x| \leq M} \frac{F^n(y)}{1 - F(y)} dF(y) - F^n(y). \end{aligned}$$

Από το παραπάνω και δοθέντος ότι $0 < F(y) < 1$, προκύπτει αμέσως το ζητούμενο.

Αντικαθιστώντας το παραπάνω αποτέλεσμα στην (3.91) λαμβάνουμε ότι:

$$0 \leq -U_n(Y_i) \leq d_n + G_M \int_{|y-x| > M} \left[\frac{1 - F^n(y)}{1 - F(y)} - nF^{n-1}(y) \right] dF(y),$$

από το οποίο προκύπτει το εξής:

$$|U_n(Y_i)| = O(\log n).$$

Συνεχίζουμε με την απόδειξη της (3.90), όπου από την (3.87) έχουμε ότι:

$$\begin{aligned} -\Delta_n &= \int_0^\infty F^{n-1}(y) K_h(x - y) m(y) f(y) dy \\ &= \int_0^\infty F^{n-1}(y) (1 - F(y)) \frac{K_h(x - y)}{1 - F(y)} m(y) dF(y), \end{aligned}$$

για το οποίο, με παρόμοιο τρόπο με την (3.91), λαμβάνουμε ότι:

$$\begin{aligned} -\Delta_n &\leq \int_{|y-x| \leq M} F^{n-1}(y) (1 - F(y)) dF(y) \\ &\quad + G_M \int_{|y-x| > M} F^{n-1}(y) (1 - F(y)) dF(y). \end{aligned} \quad (3.92)$$

Θέτοντας:

$$p_n = \int_{|y-x| \leq M} F^{n-1}(y)(1-F(y)) dF(y) = \frac{(n+1-nF(y))F^n(y)}{n(n+1)},$$

λαμβάνουμε ότι:

$$\frac{p_n}{F^{n-1}(y)} = \frac{(n+1-nF(y))F^n(y)}{n(n+1)F^{n-1}(y)} = \frac{\left(\frac{1}{n} + \frac{1}{n^2} - \frac{F(y)}{n}\right) F(y)}{1 + \frac{1}{n}} \xrightarrow{n \rightarrow \infty} 0.$$

Επομένως,

$$p_n = o(F^{n-1}(y)),$$

και αφού

$$x - M \leq y \leq x + M,$$

προκύπτει το εξής:

$$p_n = O(F^{n-1}(x+M)).$$

Με βάση το παραπάνω αποτέλεσμα η (3.92) γίνεται:

$$0 \leq -\Delta_n \leq O(F^{n-1}(x+M)) + G_M \int_{|y-x| > M} F^{n-1}(y)(1-F(y)) dF(y),$$

από το οποίο προκύπτει ότι:

$$|\Delta_n| = O((n(n+1))^{-1}).$$

□

Θεώρημα 3.2.20. Έστω N ο αριθμός των μη λογοκριμένων παρατηρήσεων του δείγματος $Y_i = (X_i, \delta_i), i = 1, \dots, n$. Τότε:

$$N \sim Bin(n, p),$$

με πιθανότητα επιτυχίας:

$$p(t) = \int_0^t f_T(x)(1-F_U(x)) dx.$$

Για δοθέν $N = \nu$, οι $(X_i : \delta_i = 1)$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με πυκνότητα

$$\tilde{f}(x) = \frac{f_T(x)(1-F_U(x))}{\int_0^t f_T(x)(1-F_U(x)) dx} = \frac{f_T(x)(1-F_U(x))}{p(t)}, \quad (3.93)$$

για $\nu = 1, 2, \dots, n$.

Απόδειξη. Ο αριθμός των μη λογοκριμένων παρατηρήσεων δίνεται από τη σχέση:

$$N(t) = \sum_{i=1}^n I_{\{X_i \leq t, \delta_i = 1\}},$$

επομένως $N \sim Bin(n, p(t))$ καθώς πρόκειται για άθροισμα n-Bernoulli δοκιμών με πιθανότητα επιτυχίας $p(t)$, η οποία είναι η:

$$\begin{aligned} p(t) &= P(X_i \leq t, \delta_i = 1) \\ &= P(X_i \leq t, T_i \leq U_i) \\ &= P(T_i \leq U_i, T_i \leq t) \\ &= \int_0^t P(T_i \leq U_i, T_i = x) f_T(x) dx \\ &= \int_0^t P(U_i \geq x) f_T(x) dx \\ &= \int_0^t [1 - P(U_i \leq x, T_i = x)] f_T(x) dx \\ &= \int_0^t f_T(x) (1 - F_U(x)) dx, \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.3), στην τρίτη γραμμή η (3.2) και στην τέταρτη γραμμή η (3.18).

Έπειτα, θεωρώντας ότι x είναι ένας σταθερός αριθμός, για $\xi \in (x, x + h)$, έχουμε ότι:

$$\begin{aligned} P(x \leq X_i \leq x + h, \delta_i = 1) &= P(x \leq T_i \leq x + h, U_i \geq T_i) \\ &\simeq P(x \leq T_i \leq x + h, U_i \geq x) \\ &= P(x \leq T_i \leq x + h) P(U_i \geq x) \\ &= f_T(\xi) h (1 - F_U(x)), \end{aligned}$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκαν οι (3.2) και (3.3), και στην τρίτη γραμμή χρησιμοποιήθηκε η ανεξαρτησία των τ.μ. T_i και U_i .

Επομένως, σε αυτήν την περίπτωση, η από κοινού κατανομή των X_i και δ_i είναι η εξής:

$$f_{X_i, \delta_i=1}(x) = \lim_{h \rightarrow 0} \frac{f_T(\xi) h (1 - F_U(x))}{h} = f_T(x) (1 - F_U(x)). \quad (3.94)$$

Με τον ίδιο ακριβώς τρόπο μπορεί να αποδειχθεί ότι:

$$f_{X_i, \delta_i=0}(x) = f_U(x) (1 - F_T(x)).$$

Τέλος, η απόδειξη της (3.93) στηρίζεται στην περικοπή των δεδομένων για τα οποία ισχύει ότι $\delta_i = 1, i = 1, \dots, n$. Χρησιμοποιώντας την (3.94), η σ.π.π. τους δίνεται από τη σχέση:

$$f_{X_i|\delta=1}(x) = \frac{f_{X_i,\delta=1}(x)}{f_{\delta=1}(x)} = \frac{f_{X_i,\delta=1}(x)}{p(t)} = \frac{f_T(x)(1 - F_U(x))}{\int_0^t f_T(x)(1 - F_U(x)) dx} = \tilde{f}(x).$$

□

Λήμμα 3.2.21. Για $n \rightarrow \infty, h \rightarrow 0$ και $nh \rightarrow \infty$ ισχύει ότι:

$$E|V_n(Y_i)|^r = a_{r,h}(1 - F(x))^r m(x)f(x) + o(a_{r,h}), \quad (3.95)$$

όπου:

$$a_{r,h} = \int_0^\infty K_h^r(y) dy = h^{-(r-1)} \int_0^\infty K^r(y) dy, \quad r = 1, 2, \dots$$

Απόδειξη. Η απόδειξη βασίζεται στον υπολογισμό της ζητούμενης αναμενόμενης τιμής χρησιμοποιώντας ως συνθήκη τον αριθμό των μη λογοκριμένων παρατηρήσεων του δείγματος $Y_i = (X_i, \delta_i)$. Λαμβάνοντας υπόψη το Λήμμα 3.2.16 και κάνοντας χρήση του Θεωρήματος 3.2.20 έχουμε ότι:

$$\begin{aligned} E|V_n|^r &= E \left| (1 - F(X_i))^{-1} (1 - F^n(X_i)) \delta_i K_h(x - X_i) \right|^r \\ &= E \left| (1 - F(X_i))^{-r} (1 - F^n(X_i))^r \delta_i^r K_h^r(x - X_i) \right| \\ &= E \left\{ E \left[(1 - F(X_i))^{-r} (1 - F^n(X_i))^r \delta_i^r K_h^r(x - X_i) \mid (X_i, \delta_i) = (y, 1) \right] \right\} \\ &= E \left[\int_0^\infty (1 - F(y))^{-r} (1 - F^n(y))^r K_h^r(x - y) \tilde{f}(y) dy \right] \\ &= E \left[\frac{1}{p(t)} \int_0^\infty (1 - F(y))^{-r} (1 - F^n(y))^r K_h^r(x - y) f_T(y) (1 - F_U(y)) dy \right] \\ &= \frac{p(t)}{p(t)} \int_0^\infty (1 - F(y))^{-r} (1 - F^n(y))^r K_h^r(x - y) f_T(y) (1 - F_U(y)) dy \\ &= \int_0^\infty (1 - F(y))^{-r} (1 - F^n(y))^r K_h^r(x - y) f_T(y) (1 - F_U(y)) dy \\ &= \int_0^\infty (1 - F(y))^{-r} (1 - F^n(y))^r K_h^r(x - y) m(y)f(y) dy, \end{aligned}$$

όπου στην όγδοη γραμμή χρησιμοποιήθηκε η (3.33).

Επομένως, προκύπτει το εξής:

$$\begin{aligned} E|V_n|^r &\leq \int_{|y-x|\leq M} (1-F(y))^{-r} K_h^r(x-y) m(y) f(y) dy \\ &\quad + G_M^r \int_{|y-x|>M} (1-F^n(y))^r m(y) f(y) dy. \end{aligned} \quad (3.96)$$

Αρχικά, το ολοκλήρωμα του δεύτερου όρου του δεξιού μέρους της (3.96) θεωρείται αμελητέο και δεν λαμβάνεται υπόψη καθώς για $n \rightarrow \infty$ θα τείνει σε κάτι μικρότερο της μονάδας.

Ασχολούμαστε με τον πρώτο όρο του δεξιού μέρους της (3.96). Θέτουμε:

$$B = \int_{|y-x|\leq M} (1-F(y))^{-r} K_h^r(x-y) m(y) f(y) dy,$$

και λαμβάνουμε το εξής:

$$\begin{aligned} B &= \frac{1}{h^r} \int_{|y-x|\leq M} (1-F(y))^{-r} K^r\left(\frac{x-y}{h}\right) m(y) f(y) dy \\ &= h^{-(r-1)} \int_{|ht|\leq M} (1-F(x-ht))^{-r} K^r(t) m(x-ht) f(x-ht) dt, \end{aligned} \quad (3.97)$$

στο οποίο θέσαμε $x-y=ht$.

Αναπτύσσοντας σε σειρά Taylor γύρω από το x τον όρο:

$$(1-F(x-ht))^{-r} m(x-ht) f(x-ht),$$

η (3.97) γίνεται:

$$\begin{aligned} B &= h^{-(r-1)} (1-F(x))^r m(x) f(x) \int_{|ht|\leq M} K^r(t) dt + o(a_{r,h}) \\ &= a_{r,h} (1-F(x))^r m(x) f(x) + o(a_{r,h}). \end{aligned}$$

Αντικαθιστώντας την παραπάνω στην (3.96), λαμβάνοντας υπόψιν το φράγμα G_M που δόθηκε στη σελίδα 63 και επιλέγοντας M τέτοιο ώστε $F(x+M) < 1$, προκύπτει αμέσως το ζητούμενο αποτέλεσμα και ολοκληρώνεται η απόδειξη. \square

Η ασυμπτωτική κατανομή του εκτιμητή πυρήνα της συνάρτησης κινδύνου δίνεται μέσω του θεωρήματος που ακολουθεί. Το θεώρημα αυτό, το οποίο αποδεικνύουμε αναλυτικά, αποτελεί τμήμα του Θεωρήματος 3 της εργασίας των Tanner and Wong (1983).

Θεώρημα 3.2.22. *Ο εκτιμητής πυρήνα της συνάρτησης κινδύνου $\hat{\lambda}(x)$, ακολουθεί ασυμπτωτικά κανονική κατανομή όταν $n \rightarrow \infty$, $h \rightarrow 0$ και $nh \rightarrow \infty$. Συγκεκριμένα ισχύει ότι:*

$$\hat{\lambda}(x) \sim N \left(E \left(\hat{\lambda}(x) \right), \text{Var} \left(\hat{\lambda}(x) \right) \right),$$

όπου $E \left(\hat{\lambda}(x) \right)$ η μέση τιμή του εκτιμητή $\hat{\lambda}(x)$ που δίνεται από την (3.35) και $\text{Var} \left(\hat{\lambda}(x) \right)$ η διακύμανσή του που δίνεται από την (3.45).

Απόδειξη. Ξεκινάμε την απόδειξη έχοντας ως πρωταρχικό στόχο να δείξουμε ότι οι τυποποιημένες μορφές των \hat{W} και W είναι ασυμπτωτικά ισοδύναμες. Για να το αποδείξουμε αυτό, υπολογίζουμε αρχικά την διακύμανση του \hat{W} χρησιμοποιώντας τα αποτελέσματα των Λημμάτων 3.2.18, 3.2.19 και 3.2.21.

$$\begin{aligned} \text{Var} \left(\hat{W} \right) &= \text{Var} \left(\hat{W} - E \left(\hat{W} \right) + E \left(\hat{W} \right) \right) = \text{Var} \left(\hat{W} - E \left(\hat{W} \right) \right) + 0 \\ &= \text{Var} \left[\sum_{i=1}^n \left(n^{-1} V_n(Y_i) + n^{-1} U_n(Y_i) + \Delta_n \right) \right] \\ &= \sum_{i=1}^n \text{Var} \left(n^{-1} V_n(Y_i) + n^{-1} U_n(Y_i) + \Delta_n \right) \\ &= n \text{Var} \left(n^{-1} V_n(Y_i) + n^{-1} U_n(Y_i) + \Delta_n \right) \\ &= n \text{Var} \left(n^{-1} V_n(Y_i) \right) + 0 + 0 = n^{-1} \text{Var} \left(V_n(Y_i) \right) \\ &= n^{-1} \left[E \left| V_n(Y_i) \right|^2 - \left(E \left| V_n(Y_i) \right| \right)^2 \right], \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.86) και στην πέμπτη γραμμή οι (3.89) και (3.90).

Ισοδύναμα έχουμε ότι:

$$\begin{aligned} \text{Var} \left(\hat{W} \right) &= n^{-1} a_{2,h} \left(1 - F(x) \right)^{-2} m(x) f(x) + o \left((nh)^{-1} \right) \\ &= n^{-1} a_{2,h} \left(1 - F(x) \right)^{-2} f_T(x) \left(1 - F_U(x) \right) + o \left((nh)^{-1} \right) \\ &= n^{-1} a_{2,h} \left(1 - F(x) \right)^{-1} \frac{f_T(x)}{1 - F_T(x)} + o \left((nh)^{-1} \right) \\ &= n^{-1} a_{2,h} \lambda_T(x) \left(1 - F(x) \right)^{-1} + o \left((nh)^{-1} \right), \end{aligned}$$

όπου στην πρώτη γραμμή η (3.95), στη δεύτερη γραμμή η (3.32), στην τρίτη γραμμή η (3.14) και στην τέταρτη γραμμή η (3.21).

Στη συνέχεια, χρησιμοποιώντας το γεγονός ότι:

$$a_{2,h} = h^{-1} \int K^2(y) dy,$$

προκύπτει το εξής:

$$Var(\hat{W}) = (nh)^{-1} \left(\int K^2(y) dy \right) \lambda_T(x) (1 - F(x))^{-1} + o((nh)^{-1}). \quad (3.98)$$

Συγκρίνοντας την (3.98) με την (3.57) και λαμβάνοντας υπόψη ότι $\hat{\lambda}(x) = W$ συμπεραίνουμε το εξής:

$$\frac{Var(\hat{W})}{Var(W)} \rightarrow 1, \quad (3.99)$$

και κατά συνέπεια:

$$Var(W) - Var(\hat{W}) \rightarrow 0. \quad (3.100)$$

Επομένως, λόγω της (3.99), προκύπτει ότι:

$$E \left(\frac{W - E(W)}{\sqrt{Var(W)}} - \frac{\hat{W} - E(\hat{W})}{\sqrt{Var(\hat{W})}} \right)^2 \rightarrow E \left(\frac{W - E(W) - \hat{W} + E(\hat{W})}{\sqrt{Var(W)}} \right)^2.$$

Όμως,

$$\begin{aligned} E \left(\frac{W - E(W) - \hat{W} + E(\hat{W})}{\sqrt{Var(W)}} \right)^2 &= E \left(\frac{W - \hat{W}}{\sqrt{Var(W)}} \right)^2 = E \left[\frac{(W - \hat{W})^2}{Var(W)} \right] \\ &= \frac{E(W - \hat{W})^2}{Var(W)} = \frac{Var(W) - Var(\hat{W})}{Var(W)}, \end{aligned}$$

το οποίο προκύπτει χρησιμοποιώντας την (3.60).

Κάνοντας χρήση των σχέσεων (3.61) και (3.100), προκύπτει ότι:

$$\frac{Var(W) - Var(\hat{W})}{Var(W)} \rightarrow 0.$$

Από το παραπάνω αποτέλεσμα λαμβάνουμε ότι:

$$\frac{W - E(W)}{\sqrt{Var(W)}} - \frac{\hat{W} - E(\hat{W})}{\sqrt{Var(\hat{W})}} \xrightarrow{p} 0, \quad (3.101)$$

το οποίο αποδεικνύει ότι η τυποποιημένη μορφή του \hat{W} είναι ασυμπτωτικά ισοδύναμη με την τυποποιημένη μορφή του $W = \hat{\lambda}(x)$. Κατά συνέπεια δείχνοντας ότι η τυποποιημένη μορφή του \hat{W} ακολουθεί την τυπική κανονική κατανομή, το ίδιο θα ισχύει και για την τυποποιημένη μορφή της W .

Παρατήρηση 3.2.23. Μέσω του αποτελέσματος που δίνεται από την (3.101) και κάνοντας χρήση των (3.60) και (3.99) εύκολα αποδεικνύεται η ασυμπτωτική ισοδυναμία του W με το \hat{W} .

Παρατηρούμε ότι το Δ_n , το οποίο δίνεται από τη σχέση (3.87), θεωρείται αμελητέο καθώς για $n \rightarrow \infty$, αυτό θα τείνει στο μηδέν. Επομένως μία ικανή συνθήκη για την εφαρμογή του Θεωρήματος του Lyapunov θα είναι ότι:

$$\left(Var(\hat{W})\right)^{-3/2} nE |n^{-1}V_n(Y_i) + n^{-1}U_n(Y_i)|^3 \rightarrow 0. \quad (3.102)$$

Χρησιμοποιώντας τα φράγματα τα οποία δείξαμε στα Λήμματα 3.2.19 και 3.2.21, η ποσότητα που δίνεται από την (3.102) είναι $O\left((nh)^{-1/2}\right)$, επομένως συγκλίνει στο μηδέν όταν $n \rightarrow \infty$, $h \rightarrow 0$ και $nh \rightarrow \infty$ που είναι και το ζητούμενο. Οπότε από το Θεώρημα του Lyapunov, το οποίο ικανοποιείται, θα έχουμε ότι:

$$\frac{\hat{W} - E(\hat{W})}{\sqrt{Var(\hat{W})}} \xrightarrow{d} N(0, 1),$$

και κατά συνέπεια:

$$\frac{W - E(W)}{\sqrt{Var(W)}} \xrightarrow{d} N(0, 1).$$

Αντικαθιστώντας όπου $W = \hat{\lambda}(x)$, λαμβάνουμε ότι η κατανομή του εκτιμητή πυρήνα της συνάρτησης κινδύνου δίνεται από τις σχέσεις:

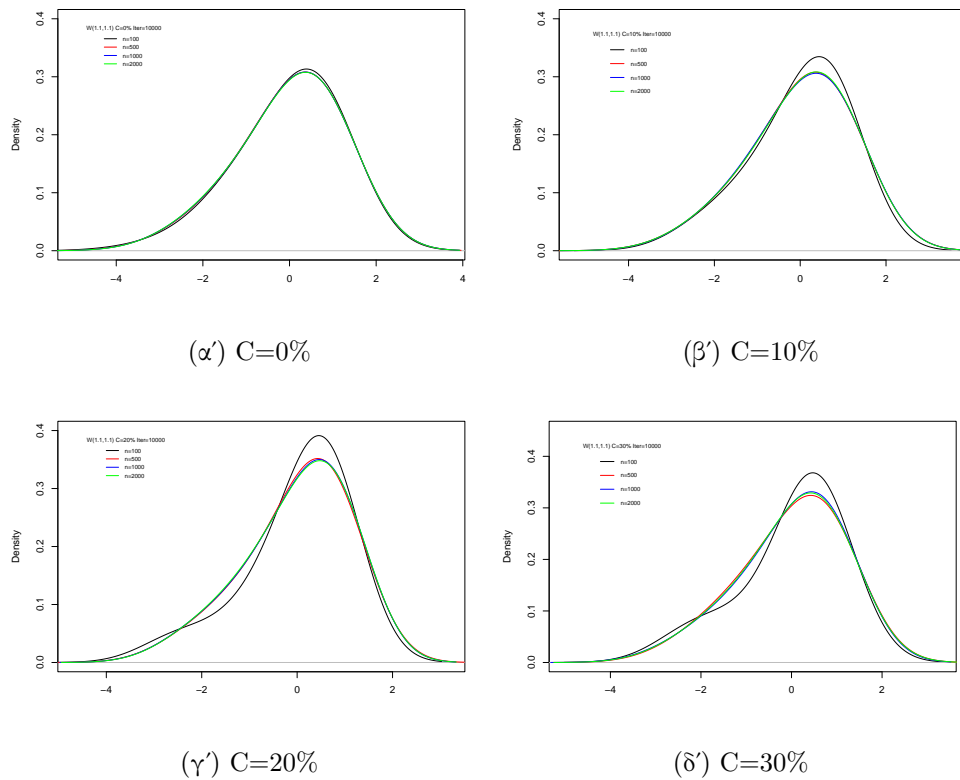
$$\frac{\hat{\lambda}(x) - E(\hat{\lambda}(x))}{\sqrt{Var(\hat{\lambda}(x))}} \xrightarrow{d} N(0, 1),$$

ή ισοδύναμα

$$\hat{\lambda}(x) \stackrel{d}{\sim} N \left(E \left(\hat{\lambda}(x) \right), Var \left(\hat{\lambda}(x) \right) \right).$$

□

Το Θεώρημα 3.2.24 επιβεβαιώνεται στο Σχήμα 3.2 στο οποίο παρουσιάζεται η κατανομή του τυποποιημένου εκτιμητή πυρήνα της συνάρτησης κινδύνου, όπως αυτή προέκυψε μετά από 10.000 επαναλήψεις, χρησιμοποιώντας δεδομένα που προέρχονται από την κατανομή Weibull με παραμέτρους $\kappa = \lambda = 1.1$ (Σύμβ: $W(1.1, 1.1)$). Τα ποσοστά λογοκρισίας (Σύμβ: C) που χρησιμοποιήθηκαν είναι 0% (καθόλου λογοκρισία), 10%, 20% και 30%, με τη μελέτη να πραγματοποιείται για καθένα από αυτά τα ποσοστά για μέγεθος δείγματος $n = 100, 500, 1000, 2000$ αντίστοιχα.



Σχήμα 3.2: Κατανομή του τυποποιημένου εκτιμητή πυρήνα της συνάρτησης κινδύνου για διάφορα μεγέθη δείγματος και ποσοστά λογοκρισίας.

ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ

Στη βιβλιογραφία έχουν προταθεί αρκετοί έλεγχοι καλής προσαρμογής για τον έλεγχο της καταλληλότητας συναρτήσεων ρυθμού επικινδυνότητας για λογοκριμένα ή μη λογοκριμένα δεδομένα.

Χρονολογικά, ο πρώτος που πρότεινε μία τέτοιου είδους διαδικασία, βασιζόμενος στην εργασία του Akritas (1988), ήταν ο Hjort (1990) ο οποίος ανέπτυξε έλεγχο καλής προσαρμογής με σκοπό να προσδιορίσει την καταλληλότητα του μοντέλου των αναλογικών κινδύνων για δεδομένα χρόνων ζωής. Έπειτα οι Bagdonavičius et al. (2010), Bagdonavicius and Nikulin (2011a), Bagdonavicius and Nikulin (2011b) και Bagdonavicius, Kruopis and Nikulin (2013), χρησιμοποιώντας τον χ^2 έλεγχο καλής προσαρμογής, πραγματοποίησαν ελέγχους σε παραμετρικά μοντέλα ρυθμού επικινδυνότητας, όπως είναι τα μοντέλα των αναλογικών κινδύνων, τα Accelerated Failure Time (AFT) μοντέλα, τα γενικευμένα μοντέλα αναλογικών κινδύνων, τα μοντέλα με διασταυρωμένες επιδράσεις των συναρτήσεων επιβίωσης και τα μοντέλα μετασχηματισμών. Τέλος, οι Bagdonavicius, Levulienė and Nikulin (2013) ανέπτυξαν μία παραλλαγή του τύπου-Pearson ελέγχου καλής προσαρμογής που εισήγαγαν, για να πραγματοποιήσουν ελέγχους σε AFT μοντέλα.

Όπως αρκετοί έλεγχοι καλής προσαρμογής που υπάρχουν στη βιβλιογραφία, έτσι και ο έλεγχος για τη συνάρτηση κινδύνου που προτείνεται στην παρούσα διατριβή, βασίζεται στο ολοκληρωτικό τετραγωνικό σφάλμα Integrated Square Error (ISE). Το ISE ως μετρική συνάρτηση χρησιμοποιήθηκε αρχικά από τους Bickel and Rosenblatt (1973), οι οποίοι κατασκεύασαν τεστ καλής προσαρμογής για τη μονοδιάστατη συνάρτηση πυκνότητας πιθανότητας. Μελετώντας την ισχύ του ελέγχου έναντι εναλλακτικών Pitman κατέληξαν στο συμπέρασμα ότι ο έλεγχός τους είναι λιγότερο ισχυρός από ελέγχους που βασίζονται στην εμπειρική συνάρτηση κατανομής, όπως είναι αυτοί των Kolmogorov-Smirnov και Cramer-von Mises, είναι όμως ισχυρότερος από τον χ^2 έλεγχο. Στη συνέχεια, ο Rosenblatt (1975) πραγματοποίησε τον παραπάνω έλεγχο για σ.π.π. δύο διαστάσεων και ει-

σήγαγε μία νέα τάξη τοπικών εναλλακτικών υπό τις οποίες έδειξε ότι ο έλεγχός του είναι ισχυρότερος έναντι ελέγχων που βασίζονται σε εμπειρικές συναρτήσεις κατανομής. Έπειτα η Fan (1994) ασχολήθηκε με ελέγχους καλής προσαρμογής για την παραμετρική σ.π.π. με τη μέθοδο των πυρήνων χρησιμοποιώντας ως εναλλακτικές τόσο αυτές που προτάθηκαν στην εργασία του Rosenblatt (1975) όσο και τις εναλλακτικές Pitman. Οι έλεγχοί της, πλην ενός, αποδείχθηκαν ότι είναι ισχυρότεροι από τους αντίστοιχους των Kolmogorov-Smirnov υπό τις εναλλακτικές του Rosenblatt, ενώ υπό τις εναλλακτικές Pitman οι έλεγχοί της είναι λιγότερο ισχυροί από αυτούς των Kolmogorov-Smirnov. Σε συνέχεια της προηγούμενης εργασίας της, η Fan (1998) πρότεινε ελέγχους καλής προσαρμογής βασισμένους στην εκτιμώμενη με τη μέθοδο των πυρήνων σ.π.π., σταθεροποιώντας τις παραμέτρους εξομάλυνσης. Επιπλέον οι Bagkavos et al. (2013) πρότειναν έναν μη παραμετρικό έλεγχο καλής προσαρμογής για τη συνάρτηση επιβίωσης για δεξιά λογοκριμένα δεδομένα, ο οποίος στηρίζεται στο ISE, έναντι εναλλακτικών οι παράμετροι των οποίων επιλέγονται με τέτοιον τρόπο ώστε η απόκλιση Kullback-Leibler μεταξύ της μηδενικής και των εναλλακτικών συναρτήσεων επιβίωσης να αποκλίνει όλο και περισσότερο. Τέλος οι Vital and Patil (2020) ασχολήθηκαν με τον έλεγχο για την συνάρτηση κινδύνου βασιζόμενοι στο ISE, όπου η άγνωστη συνάρτηση κινδύνου εκτιμάται από το ιστόγραμμα και επέλεξαν εναλλακτικές τύπου Pitman.

Στο κεφάλαιο αυτό εισάγουμε έναν νέο έλεγχο καλής προσαρμογής για τη συνάρτηση κινδύνου ο οποίος βασίζεται στο ISE. Στις ενότητες που ακολουθούν διατυπώνονται οι υποθέσεις του ελέγχου και μελετάται η κατανομή της στατιστικής συνάρτησης μέσω Θεωρημάτων και Λημμάτων τα οποία διατυπώνουμε και αποδεικνύουμε αναλυτικά.

4.1 Διατύπωση των υποθέσεων και της στατιστικής συνάρτησης

4.1.1 Διατύπωση των υποθέσεων

Οι χρόνοι ζωής T_i , οι χρόνοι λογοκρισίας U_i , αλλά και οι διαθέσιμες παρατηρήσεις (X_i, δ_i) με $i = 1, \dots, n$, ακολουθούν τις υποθέσεις της Ενότητας 2.3.1 για τη δεξιά λογοκρισία. Επιπλέον οι τ.μ. X_i έχουν α.σ.κ. $F(x)$ και σ.π.π. $f(x)$ για τις οποίες ισχύουν οι (3.14) και (3.15). Διατάσσοντας το δείγμα των X_i σε αύξουσα σειρά μαζί με τον συντελεστή λογοκρισίας $\delta_i = I_{\{T_i \leq U_i\}}$ έτσι ώστε να ισχύει ότι $X_1 < X_2 < \dots < X_n$, οι διαθέσιμες παρατηρήσεις προς ανάλυση δίνονται ως $(X_{(i)}, \delta_{(i)})$. Επιπλέον, έστω $\lambda(x)$ η συνάρτηση κινδύνου που δίνεται από

την (2.5).

Σκοπός μας είναι να πραγματοποιήσουμε έλεγχο της μηδενικής υπόθεσης ότι η πραγματική συνάρτηση κινδύνου προέρχεται από μία κλάση πλήρως προσδιορισμένων συναρτήσεων κινδύνου της μορφής $\lambda_0(x)$, έναντι της εναλλακτικής ότι δεν προέρχεται. Μοντελοποιώντας το παραπάνω λαμβάνουμε ότι η μηδενική και η εναλλακτική υπόθεση δίνονται ως εξής:

$$H_0 : \lambda(x) = \lambda_0(x) \quad \text{έναντι} \quad H_1 : \lambda(x) \neq \lambda_0(x).$$

4.1.2 Διατύπωση της στατιστικής συνάρτησης

Ο έλεγχος καλής προσαρμογής που προτείνεται στην παρούσα διατριβή χρησιμοποιεί ως μετρική συνάρτηση των δύο καμπυλών υπό την μηδενική υπόθεση το ολοκληρωμένο τετραγωνικό σφάλμα ISE, διότι ποσοτικοποιεί την απόδοση της εκτίμησης για τα διαθέσιμα δεδομένα.

Επομένως, βασιζόμενοι στο ISE, επιλέγουμε ως στατιστική συνάρτηση την

$$T_n = \int (\lambda(x) - \lambda_0(x))^2 dx. \quad (4.1)$$

Καθώς όμως η πραγματική συνάρτηση κινδύνου $\lambda(x)$ είναι άγνωστη, θα εκτιμηθεί με τη μέθοδο των πυρήνων από τον εκτιμητή $\hat{\lambda}(x)$, για τον οποίο έχει γίνει εκτενή μελέτη στην Ενότητα 3.2 και δίνεται από τις (3.28), (3.29) και (3.30). Για το λόγο αυτό λαμβάνουμε ως στατιστική συνάρτηση μία εκδοχή της (4.1), την οποία ονομάζουμε \hat{T}_n και δίνεται από τη σχέση:

$$\hat{T}_n = \int (\hat{\lambda}(x) - \lambda_0(x))^2 dx. \quad (4.2)$$

Προτού προχωρήσουμε στη μελέτη της κατανομής της στατιστικής συνάρτησης που δίνεται από την (4.2), υιοθετούμε τις παρακάτω συνθήκες για όλο το τέταρτο κεφάλαιο. Αρχικά συμβολίζουμε με $\mu_i(K)$ την i -οστή ροπή, $i = 0, 1, 2$, του πυρήνα K και με $R(K)$ το ολοκλήρωμα της πραγματικής συνάρτησης K^2 σε όλο το πεδίο ορισμού της. Οι συνθήκες είναι οι εξής:

1. Η $\lambda(x)$ είναι δύο φορές διαφορίσιμη και η $\lambda''(x)$ είναι φραγμένη και ομοιόμορφα συνεχής.
2. Για $l = 0, 1, 2$, η l -οστή παράγωγος του πυρήνα K , $K^{(l)}$, είναι φραγμένη και απολύτως ολοκληρώσιμη με πεπερασμένες δευτερες ροπές.

3. $R(K) < +\infty$ και $\mu_0(K) = 1, \mu_1(K) = 0, \mu_2(K) < +\infty$, δηλαδή ο πυρήνας K είναι δεύτερης τάξης.
4. Υπάρχει αρκετά μικρό εύρος ζώνης h τέτοιο ώστε το

$$h^{-1}K(h^{-1}(y-x))/(1-F(y)),$$

με $0 < F(y) < 1$ να είναι ομοιόμορφα φραγμένο για $|y-x| > M$, για κάθε $M > 0$.

Οι συνθήκες 1-3 ικανοποιούνται από όλους σχεδόν τους πυρήνες που χρησιμοποιούνται στην πράξη, όπως χαρακτηριστικά αναφέρεται στην εργασία των Kim et al. (2005). Η συνθήκη 4 σημαίνει ουσιαστικά ότι θα πρέπει να υπάρχουν αρκετά λογοκριμένα δεδομένα στο δεξί άκρο της περιοχής εκτίμησης προκειμένου να εφαρμοστούν οι ασυμπτωτικές μέθοδοι. Η συνθήκη αυτή ικανοποιείται αυτομάτως όταν ο πυρήνας K είναι φραγμένος.

4.2 Μελέτη της κατανομής της στατιστικής συνάρτησης

Στην περίπτωση του εκτιμητή $\hat{\lambda}(x)$, το ISE ορίζεται ως εξής:

$$I_n = \int (\hat{\lambda}(x) - \lambda(x))^2 dx.$$

Θέτουμε:

$$k = \frac{\mu_2}{2},$$

$$B(u) = \int K(t)K(u+t) dt,$$

και

$$d(n) = \begin{cases} nh^{1/2}, & \text{αν } nh^5 \rightarrow 0, \\ n^{1/2}h^{-2}, & \text{αν } nh^5 \rightarrow +\infty, \\ n^{9/10}, & \text{αν } nh^5 \rightarrow \lambda, \quad 0 < \lambda < +\infty. \end{cases}$$

Επιπλέον, έστω Z μία ασυμπτωτικά τυπική κανονική μεταβλητή. Τότε η ασυμπτωτική κατανομή του I_n δίνεται στο παρακάτω θεώρημα.

Θεώρημα 4.2.1. Έστω $n \rightarrow \infty$, $h \rightarrow 0$ και $nh \rightarrow \infty$, τότε υπό τις συνθήκες 1-4 (σελ. 87-88) ισχύει ότι:

$$d(n)(I_n - EI_n) \rightarrow \begin{cases} 2^{1/2}\sigma_1 Z, & \text{αν } nh^5 \rightarrow 0, \\ 2k\sigma_4 Z, & \text{αν } nh^5 \rightarrow +\infty, \\ (2\sigma_1^2\lambda^{-1/5} + 4k^2\sigma_4^2\lambda^{4/5})^{1/2} Z, & \text{αν } nh^5 \rightarrow \lambda, \end{cases}$$

με $0 < \lambda < +\infty$ και

$$\begin{aligned} \sigma_1^2 &= R\left(\lambda_T(x)(1-F(x))^{-1}\right)R(B(u)), \\ \sigma_4^2 &= \left(\int \left(\lambda''(x)\right)^2 \frac{\lambda_T(x)}{1-F(x)} dx\right)R(B(u)) - \left(\int \lambda''(x)\lambda_T(x) dx\right)^2. \end{aligned}$$

Απόδειξη του Θεωρήματος 4.2.1

Η απόδειξη του Θεωρήματος 4.2.1 ξεκινάει ορίζοντας έναν όρο ο οποίος προσεγγίζει ασυμπτωτικά το μέσο ολοκληρωμένο σφάλμα I_n . Σε αυτή την ασυμπτωτική προσέγγιση ο εκτιμητής $\hat{\lambda}(x)$ που υπάρχει στο I_n , αντικαθίσταται μέσω του Λήμματος 4.2.3, από τον εκτιμητή $\tilde{\lambda}(x)$, ο οποίος είναι ασυμπτωτικά ισοδύναμος του. Η διαδικασία αυτή πραγματοποιείται με σκοπό τη διόρθωση της παραβίασης της ανεξαρτησίας καθώς, δοθείσης της ύπαρξης της εμπειρικής συνάρτησης κατανομής στον εκτιμητή $\hat{\lambda}(x)$, κάθε τυχαία μεταβλητή θα εξαρτάται από τις επόμενες.

Παρατήρηση 4.2.2. Σύμφωνα με τις εργασίες των *Watson and Leadbetter (1964a,b)*, το βέλτιστο εύρος ζώνης για την εκτίμηση της συνάρτησης κινδύνου με τη μέθοδο των πυρήνων προκύπτει ελαχιστοποιώντας το μέσο τετραγωνικό σφάλμα και είναι της τάξης του $n^{-1/5}$.

Λήμμα 4.2.3. Έστω $n \rightarrow \infty$, $h \rightarrow 0$ και $nh \rightarrow \infty$. Τότε για βέλτιστο εύρος ζώνης $h \sim n^{-\frac{1}{5}}$, ο εκτιμητής πυρήνα $\hat{\lambda}(x)$ της συνάρτησης κινδύνου προσεγγίζεται ασυμπτωτικά από τον εκτιμητή:

$$\tilde{\lambda}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1-F(X_i)} K\left(\frac{x-X_i}{h}\right). \quad (4.3)$$

Απόδειξη. Για τον εκτιμητή της συνάρτησης κινδύνου που δίνεται από την (3.29) ισχύει ότι:

$$\hat{\lambda}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1-F_n(X_i)} K\left(\frac{x-X_i}{h}\right),$$

όπου $F_n(X_i)$ η εμπειρική α.σ.κ. που δίνεται από την (3.19).

Από τον Serfling (2009) γνωρίζουμε ότι:

$$\sup_{-\infty < x < +\infty} |F_n(x) - F(x)| = O_p\left(n^{-1/2}\right). \quad (4.4)$$

Επιπλέον, για $0 < F(X_i) < 1$ λαμβάνουμε το εξής:

$$\begin{aligned} 1 - F_n(X_i) &= 1 - F(X_i) + F(X_i) - F_n(X_i) \\ &= (1 - F(X_i)) \left(1 + \frac{F(X_i) - F_n(X_i)}{1 - F(X_i)}\right), \end{aligned}$$

Κάνοντας χρήση της (3.25) και λαμβάνοντας υπόψη το παραπάνω αποτέλεσμα έχουμε ότι:

$$\frac{K_h(x - X_i) \delta_i}{1 - F_n(X_i)} = \frac{K_h(x - X_i) \delta_i}{1 - F(X_i)} \left(1 + \frac{F(X_i) - F_n(X_i)}{1 - F(X_i)}\right)^{-1}. \quad (4.5)$$

Επίσης,

$$\begin{aligned} \left(1 + \frac{F(X_i) - F_n(X_i)}{1 - F(X_i)}\right)^{-1} &= \sum_{j=0}^{\infty} (-1)^{-j} \left(\frac{F(X_i) - F_n(X_i)}{1 - F(X_i)}\right)^j \\ &= 1 + \sum_{j=1}^{\infty} (-1)^{-j} \left(\frac{F(X_i) - F_n(X_i)}{1 - F(X_i)}\right)^j \\ &= 1 + O_p\left(n^{-1/2}\right), \end{aligned} \quad (4.6)$$

όπου στην πρώτη γραμμή έγινε χρήση της σειράς Maclaurin που δίνεται ως εξής:

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i, \quad \text{για } -1 < x < 1,$$

και στην τρίτη γραμμή χρησιμοποιήθηκε η (4.4).

Αντικαθιστώντας την (4.6) στην (4.5) λαμβάνουμε ότι

$$\frac{K_h(x - X_i) \delta_i}{1 - F_n(X_i)} = \frac{K_h(x - X_i) \delta_i}{1 - F(X_i)} \left(1 + O_p\left(n^{-1/2}\right)\right),$$

το οποίο αν αντικατασταθεί στην (3.29) προκύπτει ότι:

$$\hat{\lambda}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right) \left(1 + O_p\left(n^{-1/2}\right)\right). \quad (4.7)$$

Επομένως, για $n \rightarrow \infty$, $h \rightarrow 0$, $nh \rightarrow \infty$, για βέλτιστο εύρος ζώνης $h \sim n^{-\frac{1}{5}}$, η (4.7) προσεγγίζεται ασυμπτωτικά από την:

$$\tilde{\lambda}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right),$$

γεγονός που ολοκληρώνει την απόδειξη. \square

Υπό το Λήμμα 4.2.3, για $n \rightarrow \infty$, $h \rightarrow 0$, $nh \rightarrow \infty$ και για βέλτιστο εύρος ζώνης $h \sim n^{-\frac{1}{5}}$, το I_n προσεγγίζεται ασυμπτωτικά από το:

$$I_s = \int \left(\tilde{\lambda}(x) - \lambda(x) \right)^2 dx,$$

το οποίο, μετά από πράξεις, διασπάται στο παρακάτω:

$$\begin{aligned} I_s &= \int \left(\tilde{\lambda}(x) - E\tilde{\lambda}(x) \right)^2 dx + \int \left(E\tilde{\lambda}(x) - \lambda(x) \right)^2 dx \\ &\quad + 2 \int \left(\tilde{\lambda}(x) - E\tilde{\lambda}(x) \right) \left(E\tilde{\lambda}(x) - \lambda(x) \right) dx. \end{aligned}$$

Στο εξής, για λόγους ευκολίας θέτουμε:

$$w_i(x) = \frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right). \quad (4.8)$$

Για $Y_i = (X_i, \delta_i)$, $i = 1, \dots, n$, και θέτοντας,

$$\begin{aligned} R_i(x, Y_i) &= \frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right) - E\left[\frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right) \right] \\ &= w_i(x) - E(w_i(x)), \end{aligned} \quad (4.9)$$

όπως επίσης και ότι:

$$H_n(Y_i, Y_j) = \int R_i(x, Y_i) R_j(x, Y_j) dx, \quad (4.10)$$

Στη συνέχεια λαμβάνουμε το εξής:

$$\begin{aligned}
\int \left(\tilde{\lambda}(x) - E\tilde{\lambda}(x) \right)^2 dx &= \int \left\{ \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right. \\
&\quad \left. - E \left[\frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right] \right\}^2 dx \\
&= \frac{1}{n^2 h^2} \int \left\{ \sum_{i=1}^n \left\{ \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right. \right. \\
&\quad \left. \left. - E \left[\frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right] \right\} \right\}^2 dx \\
&= \frac{1}{n^2 h^2} \int \left(\sum_{i=1}^n R_i(x, Y_i) \right)^2 dx \\
&= \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{j=1}^n \int R_i(x, Y_i) R_j(x, Y_j) dx \\
&= \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{j=1}^n H_n(Y_i, Y_j) \\
&= \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{i \neq j}^n H_n(Y_i, Y_j) + \frac{1}{n^2 h^2} \sum_{i=j}^n H_n(Y_i, Y_j) \\
&= \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{i \neq j}^n H_n(Y_i, Y_j) + \frac{1}{n^2 h^2} \sum_{i=1}^n H_n(Y_i, Y_i),
\end{aligned}$$

για το οποίο στο πρώτο βήμα έγινε αντικατάσταση της (4.10), στο τρίτο βήμα της (4.9), στο τέταρτο βήμα χρησιμοποιήθηκε η σχέση:

$$\left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n X_i \sum_{j=1}^n X_j = \sum_{i=1}^n \sum_{j=1}^n X_i X_j,$$

και στο πέμπτο βήμα χρησιμοποιήθηκε η (4.10).

Επομένως το I_s γίνεται:

$$\begin{aligned}
I_s &= \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{i \neq j}^n H_n(Y_i, Y_j) + \frac{1}{n^2 h^2} \sum_{i=1}^n H_n(Y_i, Y_i) \\
&\quad + \int \left(E\tilde{\lambda}(x) - \lambda(x) \right)^2 dx + 2 \int \left(\tilde{\lambda}(x) - E\tilde{\lambda}(x) \right) \left(E\tilde{\lambda}(x) - \lambda(x) \right) dx.
\end{aligned}$$

ή ισοδύναμα:

$$I_s \equiv \hat{I}_1 + \hat{I}_2 + \hat{I}_3 + 2\hat{I}_4. \quad (4.11)$$

Παρατηρούμε ότι το I_s διασπάται σε τέσσερις όρους:

- α) Ένα συμμετρικό U-στατιστικό. (Σύμβ: \hat{I}_1).
- β) Τη διαγώνιο του συμμετρικού U-στατιστικού. (Σύμβ: \hat{I}_2)
- γ) Το ολοκλήρωμα του τετραγώνου της μεροληψίας του $\hat{\lambda}(x)$. (Σύμβ: \hat{I}_3)
- δ) Το ολοκλήρωμα του γινομένου του στοχαστικού και του συστηματικού μέρους της απόκλισης της $\hat{\lambda}(x)$ από την $\lambda(x)$. (Σύμβ: \hat{I}_4)

Αυτοί οι όροι μελετώνται στα Λήμματα 4.2.6 έως 4.2.9 και τα αντίστοιχα αποτελέσματα τους αντικαθίστανται στο I_s προκειμένου να αποφανθούμε για την ασυμπτωτική κατανομή του και κατά συνέπεια για την ασυμπτωτική κατανομή του I_n .

Προτού εξετάσουμε κάθε όρο του I_s ξεχωριστά, αποδεικνύονται τα παρακάτω λήμματα.

Λήμμα 4.2.4. Για i, j, k, l, r καθορισμένα και διαφορετικά μεταξύ τους, όπως επίσης και για $E(H_n^2(Y_i, Y_j)) < \infty$ για κάθε n , ισχύουν ότι:

$$E(H_n(Y_i, Y_j)) = 0, \quad (4.12)$$

$$E(H_n(Y_i, Y_j) H_n(Y_i, Y_k)) = 0, \quad (4.13)$$

$$E(H_n(Y_i, Y_j) H_n(Y_i, Y_k) H_n(Y_i, Y_l) H_n(Y_i, Y_r)) = 0, \quad (4.14)$$

$$E(H_n^3(Y_i, Y_j) H_n(Y_i, Y_k)) = 0, \quad (4.15)$$

$$E(H_n^2(Y_i, Y_j) H_n(Y_i, Y_k) H_n(Y_i, Y_l)) = 0. \quad (4.16)$$

Απόδειξη. Καθώς $i \neq j \neq k \neq l \neq r$ και καθώς τα Y_i , $i = 1, \dots, n$ είναι ανεξάρτητες και ισόνομες τ.μ., τότε τα $w_i(x)$, $w_j(x)$, $w_k(x)$, $w_l(x)$ και $w_r(x)$ είναι ανεξάρτητες τ.μ. ως συναρτήσεις ανεξάρτητων τυχαίων μεταβλητών. Παραθέτουμε την απόδειξη της (4.12), με τις αποδείξεις των υπολοίπων να είναι παρόμοιες

(αν και ιδιαίτερα πιο εκτενείς).

$$\begin{aligned}
 E(H_n(Y_i, Y_j)) &= E\left(\int R_i(x, Y_i) R_j(x, Y_j) dx\right) \\
 &= E\left(\int (w_i(x) - E(w_i(x)))(w_j(x) - E(w_j(x))) dx\right) \\
 &= \int E(w_i(x)w_j(x) - w_i(x)E(w_j(x)) - w_j(x)E(w_i(x)) \\
 &\quad + E(w_i(x))E(w_j(x))) dx,
 \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (4.10) και στην τρίτη γραμμή η (4.9).

Ισοδύναμα, έχουμε ότι:

$$\begin{aligned}
 E(H_n(Y_i, Y_j)) &= \int (E(w_i(x))E(w_j(x)) - E(w_i(x))E(w_j(x)) \\
 &\quad - E(w_i(x))E(w_j(x)) + E(w_i(x))E(w_j(x))) dx \\
 &= 0,
 \end{aligned}$$

το οποίο ολοκληρώνει την απόδειξη της (4.12). \square

Λήμμα 4.2.5. Για i, j καθορισμένα με $i \neq j$, ισχύουν τα εξής:

$$E(H_n^2(Y_i, Y_j)) = h^3 R\left(\lambda_T(x)(1 - F(x))^{-1}\right) R(B(u)) + O(h^5), \quad (4.17)$$

$$E(H_n^4(Y_i, Y_j)) = O(h^5), \quad (4.18)$$

όπου:

$$R(f(x)) = \int_{x \in A} f^2(x) dx \quad \mu \in f : A \rightarrow \mathbb{R}, \quad (4.19)$$

και

$$B(u) = \int K(t)K(u+t) dt. \quad (4.20)$$

Απόδειξη. Η απόδειξη βασίζεται στον υπολογισμό των ζητούμενων αναμενόμενων τιμών χρησιμοποιώντας ως συνθήκη τον αριθμό των μη λογοκριμένων παρατηρήσεων του δείγματος $Y_i = (X_i, \delta_i)$. Κάνοντας χρήση του Θεωρήματος 3.2.19, ξεκινάμε με την απόδειξη της (4.17).

Αρχικά έχουμε ότι:

$$E(H_n^2(Y_i, Y_j)) = \int \int [E(R_i(x, Y_i) R_j(y, Y_j))]^2 dx dy. \quad (4.21)$$

Για καθορισμένο i , ισχύει ότι:

$$\begin{aligned}
E(R_i(x, Y_i) R_j(y, Y_i)) &= E\{[w_i(x) - E(w_i(x))][w_i(y) - E(w_i(y))]\} \\
&= E[w_i(x)w_i(y) - w_i(x)E(w_i(y)) \\
&\quad - w_i(y)E(w_i(x)) + E(w_i(x))E(w_i(y))] \\
&= E[w_i(x)w_i(y)] - E[w_i(x)E(w_i(y))] \\
&\quad - E[w_i(y)E(w_i(x))] \\
&\quad + E[E(w_i(x))E(w_i(y))] \\
&= E[w_i(x)w_i(y)] - E(w_i(x))E(w_i(y)) \\
&\quad - E(w_i(y))E(w_i(x)) + E(w_i(x))E(w_i(y)) \\
&= E[w_i(x)w_i(y)] - E(w_i(x))E(w_i(y)). \tag{4.22}
\end{aligned}$$

Εφόσον το i είναι καθορισμένο, η $(X_i, \delta_i = 1)$ είναι Bernoulli τ.μ. με μέση τιμή $p(t)$, όπως αυτή προσδιορίστηκε στο Θεώρημα 3.2.20. Επομένως:

$$\begin{aligned}
E(w_i(x)w_i(y)) &= E\left[\frac{\delta_i}{1-F(X_i)}K\left(\frac{x-X_i}{h}\right)\frac{\delta_i}{1-F(X_i)}K\left(\frac{y-X_i}{h}\right)\right] \\
&= E\left\{E\left[\frac{\delta_i^2}{(1-F(X_i))^2}K\left(\frac{x-X_i}{h}\right)\right.\right. \\
&\quad \left.\left.\times K\left(\frac{y-X_i}{h}\right)\middle|(X_i, \delta_i) = (z, 1)\right]\right\} \\
&= E\left[\int\frac{1}{(1-F(z))^2}K\left(\frac{x-z}{h}\right)K\left(\frac{y-z}{h}\right)\tilde{f}(z)dz\right] \\
&= E\left[\int\frac{1}{(1-F(z))^2}K\left(\frac{x-z}{h}\right)K\left(\frac{y-z}{h}\right)\right. \\
&\quad \left.\times\frac{f_T(z)(1-F_U(z))}{p(t)}dz\right] \\
&= p(t)\frac{1}{p(t)}\int\frac{1}{1-F(z)}\frac{1-F_U(z)}{1-F(z)}K\left(\frac{x-z}{h}\right) \\
&\quad \times K\left(\frac{y-z}{h}\right)f_T(z)dz \\
&= \int\frac{1}{1-F(z)}\frac{f_T(z)}{1-F_T(z)}K\left(\frac{x-z}{h}\right)K\left(\frac{y-z}{h}\right)dz \\
&= \int\frac{\lambda_T(z)}{1-F(z)}K\left(\frac{x-z}{h}\right)K\left(\frac{y-z}{h}\right)dz,
\end{aligned}$$

όπου στο πρώτο βήμα χρησιμοποιήθηκε η (4.8), στο δεύτερο βήμα η (3.36), στο τέταρτο βήμα η (3.95) και στο έκτο βήμα η (3.21).

Πραγματοποιώντας την αλλαγή μεταβλητής $x - z = ht$, η παραπάνω γίνεται:

$$E(w_i(x)w_i(y)) = h \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t) K\left(\frac{y - x + ht}{h}\right) dt. \quad (4.23)$$

Ομοίως,

$$\begin{aligned} E(w_i(x)) &= E\left[\frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right)\right] \\ &= E\left\{E\left[\frac{\delta_i}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right) \mid (X_i, \delta_i) = (z, 1)\right]\right\} \\ &= E\left[\int \frac{1}{1 - F(z)} K\left(\frac{x - z}{h}\right) \tilde{f}(z) dz\right] \\ &= E\left[\int \frac{1}{1 - F(z)} K\left(\frac{x - z}{h}\right) \frac{f_T(z)(1 - F_U(z))}{p(t)} dz\right] \\ &= p(t) \frac{1}{p(t)} \int \frac{f_T(z)}{1 - F_T(z)} K\left(\frac{x - z}{h}\right) dz \\ &= \int \lambda_T(z) K\left(\frac{x - z}{h}\right) dz \\ &= h \int \lambda_T(x - ht) K(t) dt, \end{aligned} \quad (4.24)$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκε η (4.8), στη δεύτερη γραμμή χρησιμοποιήθηκε η (3.36), στην τέταρτη γραμμή η (3.95), στην πέμπτη γραμμή η (3.14), στην έκτη γραμμή η (3.21) και στην τελευταία γραμμή έγινε η αλλαγή μεταβλητής $x - z = ht$.

Από την (4.24), εύκολα προκύπτει ότι:

$$\begin{aligned} E(w_i(x)) E(w_i(y)) &= \left(h \int \lambda_T(x - ht) K(t) dt\right) \\ &\quad \times \left(h \int \lambda_T(y - ht) K(t) dt\right) \\ &= h^2 \left(\int \lambda_T(x - ht) K(t) dt\right) \\ &\quad \times \left(\int \lambda_T(y - ht) K(t) dt\right). \end{aligned} \quad (4.25)$$

Αντικαθιστούμε τις (4.23) και (4.25) στην (4.22) για να λάβουμε ότι:

$$\begin{aligned} E(R_i(x, Y_i) R_j(y, Y_j)) &= h \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t) K\left(\frac{y - x + ht}{h}\right) dt \\ &\quad - h^2 \left(\int \lambda_T(x - ht) K(t) dt \right) \\ &\quad \times \left(\int \lambda_T(y - ht) K(t) dt \right). \end{aligned} \quad (4.26)$$

Στη συνέχεια, αντικαθιστώντας την (4.26) στην (4.21) παίρνουμε το εξής:

$$\begin{aligned} E(H_n^2(Y_i, Y_j)) &= \int \int \left[h \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t) K\left(\frac{y - x + ht}{h}\right) dt \right. \\ &\quad \left. - h^2 \left(\int \lambda_T(x - ht) K(t) dt \right) \left(\int \lambda_T(y - ht) K(t) dt \right) \right]^2 dx dy. \end{aligned}$$

Θέτοντας:

$$G = E(H_n^2(Y_i, Y_j)),$$

και πραγματοποιώντας την αλλαγή μεταβλητής $y - x = uh$ προκύπτει ότι:

$$\begin{aligned} G &= h \int \int \left[h \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t) K(u + t) dt \right. \\ &\quad \left. - h^2 \left(\int \lambda_T(x - ht) K(t) dt \right) \right. \\ &\quad \left. \times \left(\int \lambda_T(x + uh - ht) K(t) dt \right) \right]^2 dx du \\ &= h^3 \int \int \left[\int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t) K(u + t) dt \right. \\ &\quad \left. - h \left(\int \lambda_T(x - ht) K(t) dt \right) \right. \\ &\quad \left. \times \left(\int \lambda_T(x + uh - ht) K(t) dt \right) \right]^2 dx du, \end{aligned}$$

ή ισοδύναμα ότι:

$$\begin{aligned}
G = & h^3 \int \int \left[\left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right)^2 \right. \\
& - 2h \left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right) \\
& \times \left(\int \lambda_T(x-ht) K(t) dt \right) \left(\int \lambda_T(x+uh-ht) K(t) dt \right) \\
& + h^2 \left(\int \lambda_T(x-ht) K(t) dt \right. \\
& \left. \left. \times \int \lambda_T(x+uh-ht) K(t) dt \right)^2 \right] dxdu. \tag{4.27}
\end{aligned}$$

Παρατηρούμε ότι οι όροι:

$$h^2 \left(\int \lambda_T(x-ht) K(t) dt \int \lambda_T(x+uh-ht) K(t) dt \right)^2,$$

και

$$\begin{aligned}
h \left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right) & \left(\int \lambda_T(x-ht) K(t) dt \right) \\
& \times \left(\int \lambda_T(x+uh-ht) K(t) dt \right),
\end{aligned}$$

είναι ασυμπτωτικά αμελητέοι, για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$, σε σχέση με το:

$$\left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right)^2.$$

Λαμβάνοντας υπόψη το παραπάνω, η (4.27) γίνεται:

$$\begin{aligned}
G = & \int \int \left[h^3 \left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right)^2 \right. \\
& - 2h^4 \left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right) \\
& \times \left(\int \lambda_T(x-ht) K(t) dt \right) \left(\int \lambda_T(x+uh-ht) K(t) dt \right) \\
& \left. + h^5 \left(\int \lambda_T(x-ht) K(t) dt \int \lambda_T(x+uh-ht) K(t) dt \right)^2 \right] dxdu,
\end{aligned}$$

ή ισοδύναμα:

$$\begin{aligned} G &= h^3 \iint \left(\int \frac{\lambda_T(x-ht)}{1-F(x-ht)} K(t)K(u+t) dt \right)^2 dxdu + O(h^5) \\ &= h^3 \left[\int \left(\frac{\lambda_T(x)}{1-F(x)} \right)^2 dx \right] \left[\int \left(\int K(t)K(u+t) dt \right)^2 du \right] + O(h^5), \end{aligned}$$

για το οποίο στο δεύτερο βήμα, αναπτύξαμε σε σειρά Taylor γύρω από το x τον όρο:

$$\frac{\lambda_T(x-ht)}{1-F(x-ht)}.$$

Εφαρμόζοντας τις (4.19) και (4.20) λαμβάνουμε ότι:

$$E(H_n^2(Y_i, Y_j)) = h^3 R \left(\lambda_T(x) (1-F(x))^{-1} \right) R(B(u)) + O(h^5),$$

το οποίο και ολοκληρώνει την απόδειξη της (4.17).

Για την απόδειξη της (4.18), αρχικά έχουμε ότι:

$$E(H_n^4(Y_i, Y_j)) = \int \int \int \int [E(R_i(x, Y_i) R_i(y, Y_i) R_i(z, Y_i) R_i(u, Y_i))]^2 dx dy dz du.$$

Το αποτέλεσμα της σχέσης:

$$[E(R_i(x, Y_i) R_i(y, Y_i) R_i(z, Y_i) R_i(u, Y_i))]^2,$$

αποτελείται από διάφορους όρους, καθένας από τους οποίους είναι τάξης h^5 . Σκιαγραφώντας την απόδειξη για τον πρώτο όρο έχουμε ότι:

$$S_t = \int \int \int \int [E(w_i(x)w_i(y)w_i(z)w_i(u))]^2 dx dy dz du. \quad (4.28)$$

Θέτοντας:

$$T_1 = E(w_i(x)w_i(y)w_i(z)w_i(u)),$$

και χρησιμοποιώντας την (4.9), υπολογίζουμε το εξής:

$$\begin{aligned} T_1 &= E \left[\frac{\delta_i^4}{(1-F(X_i))^4} K \left(\frac{x-X_i}{h} \right) K \left(\frac{y-X_i}{h} \right) K \left(\frac{z-X_i}{h} \right) K \left(\frac{u-X_i}{h} \right) \right] \\ &= E \left\{ E \left[\frac{\delta_i^4}{(1-F(X_i))^4} K \left(\frac{x-X_i}{h} \right) K \left(\frac{y-X_i}{h} \right) K \left(\frac{z-X_i}{h} \right) \right. \right. \\ &\quad \left. \left. \times K \left(\frac{u-X_i}{h} \right) \mid (X_i, \delta_i) = (v, 1) \right] \right\}, \end{aligned}$$

όπου στο δεύτερο βήμα χρησιμοποιήθηκε η (3.36).

Ισοδύναμα έχουμε ότι:

$$\begin{aligned}
T_1 &= E \left[\int \frac{1}{(1-F(v))^4} K\left(\frac{x-v}{h}\right) K\left(\frac{y-v}{h}\right) K\left(\frac{z-v}{h}\right) K\left(\frac{u-v}{h}\right) \tilde{f}(v) dv \right] \\
&= E \left[\int \frac{1}{(1-F(v))^4} K\left(\frac{x-v}{h}\right) K\left(\frac{y-v}{h}\right) K\left(\frac{z-v}{h}\right) \right. \\
&\quad \left. \times K\left(\frac{u-v}{h}\right) \frac{f_T(v)(1-F_U(v))}{p(t)} dv \right] \\
&= p(t) \frac{1}{p(t)} \int \frac{1}{(1-F(v))^3} \frac{1-F_U(v)}{1-F(v)} f_T(v) K\left(\frac{x-v}{h}\right) K\left(\frac{y-v}{h}\right) \\
&\quad \times K\left(\frac{z-v}{h}\right) K\left(\frac{u-v}{h}\right) dv \\
&= \int \frac{1}{(1-F(v))^3} \frac{f_T(v)}{1-F_T(v)} K\left(\frac{x-v}{h}\right) K\left(\frac{y-v}{h}\right) K\left(\frac{z-v}{h}\right) K\left(\frac{u-v}{h}\right) dv \\
&= \int \frac{\lambda_T(v)}{(1-F(v))^3} K\left(\frac{x-v}{h}\right) K\left(\frac{y-v}{h}\right) K\left(\frac{z-v}{h}\right) K\left(\frac{u-v}{h}\right) dv,
\end{aligned}$$

όπου στο δεύτερο βήμα χρησιμοποιήθηκε η (3.95), στο τέταρτο βήμα η (3.14) και στο πέμπτο βήμα η (3.21).

Το παραπάνω, πραγματοποιώντας την αλλαγή μεταβλητής $x-v=ht$ στο τελευταίο βήμα, γίνεται:

$$\begin{aligned}
T_1 &= h \int \frac{\lambda_T(x-ht)}{(1-F(x-ht))^3} K(t) K\left(\frac{y-x+ht}{h}\right) K\left(\frac{z-x+ht}{h}\right) \\
&\quad \times K\left(\frac{u-x+ht}{h}\right) dt. \tag{4.29}
\end{aligned}$$

Αντικαθιστώντας την (4.29) στην (4.28), προκύπτει ότι:

$$\begin{aligned}
S_t &= h^2 \int \int \int \int \left[\int \frac{\lambda_T(x-ht)}{(1-F(x-ht))^3} K(t) K\left(\frac{y-x+ht}{h}\right) \right. \\
&\quad \left. \times K\left(\frac{z-x+ht}{h}\right) K\left(\frac{u-x+ht}{h}\right) dt \right]^2 dx dy dz du,
\end{aligned}$$

ή ισοδύναμα:

$$\begin{aligned}
S_t &= h^5 \int \int \int \int \left[\int \frac{\lambda_T(x-ht)}{(1-F(x-ht))^3} K(t)K(w+t) \right. \\
&\quad \left. \times K(s+t)K(r+t) dt \right]^2 dx dw ds dr \\
&\leq h^5 \left[\left(\frac{\lambda_T(x)}{(1-F(x))^3} \right)^2 dx \right] \\
&\quad \times \int \int \int \int [K(t)K(w+t)K(s+t)K(r+t)]^2 dt dw ds dr \\
&= O(h^5),
\end{aligned}$$

όπου στο πρώτο βήμα πραγματοποιήθηκε αλλαγή μεταβλητών $y-x=wh$, $z-x=sh$, $u-x=rh$ και στο δεύτερο βήμα αναπτύχθηκε σε σειρά Taylor γύρω από το x ο όρος:

$$\frac{\lambda_T(x-ht)}{(1-F(x-ht))^3}.$$

Με τον ίδιο ακριβώς τρόπο αποδεικνύεται ότι και οι υπόλοιποι όροι που προκύπτουν από την ανάπτυξη του ολοκληρώματος της $E(H_n^4(Y_i, Y_j))$ είναι τάξης h^5 , το οποίο και ολοκληρώνει την απόδειξη. \square

Λήμμα 4.2.6. Όταν $n \rightarrow +\infty$ τότε

$$\hat{I}_1 \sim N(0, 2n^{-2}h^{-1}\sigma_1^2).$$

Απόδειξη. Η απόδειξη της κατανομής του \hat{I}_1 είναι ανάλογη της απόδειξης του Θεωρήματος 1 της εργασίας του Hall (1984). Για το U-στατιστικό:

$$U_n = \sum_{i < j} \sum H_n(X_i, X_j),$$

με i, j καθορισμένα, αν:

$$[E(G_n^2(X_i, X_j)) + n^{-1}E(H_n^4(X_i, X_j))] / [E(H_n^2(X_i, X_j))]^2 \rightarrow 0, \quad (4.30)$$

όπου,

$$G_n(x, y) = E(H_n^*(x, X_i) H_n^*(y, X_i)), \quad (4.31)$$

και με δεδομένο ότι το H_n είναι συμμετρική συνάρτηση, $E[H_n(X_i, X_j) | X_i] = 0$ και $E(H_n^2(X_i, X_j)) < \infty$ για κάθε n , τότε το U_n ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή μηδέν και διακύμανση ίση με:

$$\frac{1}{2}n^2 E H_n^2(X_i, X_j).$$

Αρχικά,

$$\hat{I}_1 = \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{i \neq j}^n H_n(Y_i, Y_j) = \frac{2}{n^2 h^2} \sum_{1 \leq i < j \leq n} H_n(Y_i, Y_j) \equiv \frac{2}{n^2 h^2} U_n.$$

Το U_n είναι συμμετρικό U-στατιστικό καθώς:

$$\begin{aligned} H_n(Y_i, Y_j) &= \int R_i(x, Y_i) R_j(x, Y_j) dx \\ &= \int R_j(x, Y_j) R_i(x, Y_i) dx \\ &= H_n(Y_j, Y_i). \end{aligned}$$

Επομένως, έχουμε ότι:

$$\begin{aligned} \hat{I}_1 &= \frac{1}{n^2 h^2} U_n = \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{i \neq j}^n H_n(Y_i, Y_j) \\ &= \frac{2}{n^2 h^2} \sum_{i=2}^n \sum_{j=1}^{i-1} H_n(Y_i, Y_j) \\ &= \frac{2}{n^2 h^2} \sum_{i=2}^n P_i, \end{aligned}$$

όπου:

$$P_i = \sum_{j=1}^{i-1} H_n(Y_i, Y_j). \quad (4.32)$$

Στη συνέχεια, από την (4.12) και τις υποθέσεις του Θεωρήματος 1 της εργασίας του Hall (1984), εύκολα διαπιστώνεται ότι η ποσότητα $\sum_{i=2}^k P_i, k = 2, \dots, n$, είναι ένα τετραγωνικά ολοκληρώσιμο Martingale με διαφορές P_i . Επιπλέον, στην περίπτωση μας, η συνθήκη Lindeberg (βλέπε σχέση (3.7), Hall and Heyde (1980), σελ. 54) που χρησιμοποιείται στο Πόρισμα 3.1 του συγγράμματος των Hall and Heyde (1980) αντικαθίσταται από τη συνθήκη Lindeberg (βλέπε σχέση (3.6), Hall and Heyde (1980), σελ. 53) η οποία είναι ισοδύναμη της. Η μελέτη της

ασυμπτωτικής κατανομής του \hat{I}_1 ξεκινάει με τη μελέτη της ασυμπτωτικής κατανομής του U_n . Επομένως, χρησιμοποιώντας το Πόρισμα 3.1 του συγγράμματος των Hall and Heyde (1980), πρέπει να δείξουμε ότι:

α) Για κάθε $\epsilon > 0$, όταν $n \rightarrow +\infty$,

$$s_n^{-2} \sum_{i=1}^n E [P_i^2 I(|P_i| > \epsilon s_n)] \xrightarrow{p} 0,$$

όπου:

$$s_n^2 = E(U_n^2),$$

και

β)

$$s_n^{-2} V_n^2 = \sum_{i=1}^n E(P_i^2 | Y_1, Y_2, \dots, Y_{i-1}) \xrightarrow{p} 1.$$

Τότε, από τα (α) και (β) προκύπτει ότι $U_n/s_n \sim N(0, 1)$.

Ξεκινώντας έχουμε ότι:

$$\begin{aligned} E(P_i^2) &= E\left(\sum_{j=1}^{i-1} H_n(Y_i, Y_j)\right)^2 \\ &= E\left(\sum_{j=1}^{i-1} H_n^2(Y_i, Y_j) + 2 \sum_{j < k} H_n(Y_i, Y_j) H_n(Y_i, Y_k)\right) \\ &= \sum_{j=1}^{i-1} E(H_n^2(Y_i, Y_j)) + 2 \sum_{j < k} E(H_n(Y_i, Y_j) H_n(Y_i, Y_k)) \\ &= \sum_{j=1}^{i-1} E(H_n^2(Y_i, Y_j)) \\ &= (i-1)E(H_n^2(Y_i, Y_j)), \end{aligned} \tag{4.33}$$

διότι τα $Y_i, i = 1, \dots, n$ είναι ανεξάρτητες και ισόνομες τ.μ. Επιπλέον στην πρώτη γραμμή χρησιμοποιήθηκε η (4.32), στη δεύτερη γραμμή η (3.48) και στην τέταρτη γραμμή η (4.13).

Επομένως, χρησιμοποιώντας την (4.33), λαμβάνουμε ότι:

$$\begin{aligned}
s_n^2 &= E(U_n^2) \\
&= E\left(\sum_{i=2}^n P_i\right)^2 \\
&= E\left(\sum_{i=2}^n P_i^2 + 2\sum_{i<k} P_i P_k\right) \\
&= \sum_{i=2}^n E P_i^2 \\
&= \sum_{i=2}^n (i-1)E(H_n^2(Y_i, Y_j)) \\
&= \frac{1}{2}n(n-1)E(H_n^2(Y_i, Y_j)), \tag{4.34}
\end{aligned}$$

όπου στην τρίτη γραμμή χρησιμοποιήθηκε η (3.48) και στην τέταρτη γραμμή η (4.13).

Προχωρώντας με την αποδειξη του (α), από την ανισότητα του Chebyshev (βλέπε Petron (1995)) λαμβάνουμε ότι:

$$s_n^{-2} \sum_{i=1}^n E[P_i^2 I(|P_i| > \epsilon s_n)] \leq \epsilon^{-2} s_n^{-4} \sum_{i=1}^n E(|P_i|^4). \tag{4.35}$$

Στη συνέχεια έχουμε ότι:

$$\sum_{i=1}^n E(|P_i|^4) = \sum_{i=1}^n E\left|\sum_{j=1}^{i-1} H_n(Y_i, Y_j)\right|^4. \tag{4.36}$$

Ασχολούμαστε με τον δεξιό όρο της (4.35). Αρχικά λαμβάνουμε ότι:

$$\begin{aligned}
E\left|\sum_{j=1}^{i-1} H_n(Y_i, Y_j)\right|^4 &= \sum_{j=1}^{i-1} E|H_n^4(Y_i, Y_j)| \\
&\quad + 3 \sum_{\substack{1 \leq j, k \leq i-1 \\ j \neq k}} E|H_n^2(Y_i, Y_j) H_n^2(Y_i, Y_k)|. \tag{4.37}
\end{aligned}$$

Το παραπάνω προκύπτει αντικαθιστώντας τις (4.14), (4.15) και (4.16) στο

$$\begin{aligned}
E \left| \sum_{j=1}^{i-1} H_n(Y_i, Y_j) \right|^4 &= \sum_{j=1}^{i-1} E |H_n^4(Y_i, Y_j)| + 4 \sum_{\substack{1 \leq j, k \leq i-1 \\ j \neq k}} E |H_n^3(Y_i, Y_j) H_n(Y_i, Y_k)| \\
&+ 3 \sum_{\substack{1 \leq j, k \leq i-1 \\ j \neq k}} E |H_n^2(Y_i, Y_j) H_n^2(Y_i, Y_k)| \\
&+ 6 \sum_{\substack{1 \leq j, k, l \leq i-1 \\ j \neq k \neq l}} E |H_n^2(Y_i, Y_j) H_n(Y_i, Y_k) H_n(Y_i, Y_l)| \\
&+ 24 \sum_{\substack{1 \leq j, k, l, r \leq i-1 \\ j \neq k \neq l \neq r}} E |H_n(Y_i, Y_j) H_n(Y_i, Y_k) \\
&\times H_n(Y_i, Y_l) H_n(Y_i, Y_r)|.
\end{aligned}$$

Συνδυάζοντας τις (4.35) και (4.37), με j και k καθορισμένα, παίρνουμε ότι:

$$\begin{aligned}
\sum_{i=1}^n E \left(|P_i|^4 \right) &= \sum_{i=1}^n E \sum_{j=1}^{i-1} E |H_n^4(Y_i, Y_j)| \\
&+ 3 \sum_{i=1}^n E \sum_{\substack{1 \leq j, k \leq i-1 \\ j \neq k}} E |H_n^2(Y_i, Y_j) H_n^2(Y_i, Y_k)| \\
&= \sum_{i=1}^n (i-1) E |H_n^4(Y_i, Y_j)| \\
&+ 3 \sum_{i=1}^n (i-1)(i-2) E |H_n^2(Y_i, Y_j) H_n^2(Y_i, Y_k)| \\
&\leq n^2 CE [H_n^4(Y_i, Y_j)] + n^3 CE [H_n^2(Y_i, Y_j) H_n^2(Y_i, Y_k)],
\end{aligned}$$

όπου C είναι μία θετική σταθερά. Από τις (4.17) και (4.18) του Λήμματος 4.2.5 και για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$ έχουμε ότι, $n^3 CE [H_n^2(Y_i, Y_j) H_n^2(Y_i, Y_k)] = O(n^{9/5})$ το οποίο είναι μικρότερο από $n^3 CE [H_n^4(Y_i, Y_j)] = O(n^2)$. Επομένως, καταλήγουμε στο ότι:

$$\sum_{i=1}^n E \left(|P_i|^4 \right) \leq n^3 CE [H_n^4(Y_i, Y_j)]. \quad (4.38)$$

Αντικαθιστώντας την (4.38) στην (4.35) λαμβάνουμε το εξής:

$$s_n^{-2} \sum_{i=1}^n E [P_i^2 I(|P_i| > \epsilon s_n)] \leq \epsilon^2 s_n^{-4} n^3 CE [H_n^4(Y_i, Y_j)].$$

Όμως από τις (4.30) και (4.34) εύκολα προκύπτει ότι:

$$n^3 s_n^{-4} E [H_n^4(Y_i, Y_j)] \rightarrow 0,$$

από το οποίο αποδεικνύεται η (α).

Συνεχίζοντας με την απόδειξη του (β), αρχικά έχουμε ότι:

$$V_n^2 = \sum_{i=2}^n \left[E \left(\sum_{j=1}^{i-1} H_n(Y_i, Y_j) \right)^2 \mid Y_1, \dots, Y_{i-1} \right],$$

ή ισοδύναμα ότι:

$$\begin{aligned} V_n^2 &= \sum_{i=2}^n E \left(\sum_{j=1}^{i-1} H_n^2(Y_i, Y_j) \mid Y_1, \dots, Y_{i-1} \right) \\ &\quad + 2 \sum_{i=2}^n E \left(\sum_{j=1}^{i-1} \sum_{k=j+1}^{i-1} H_n(Y_i, Y_j) H_n(Y_i, Y_k) \mid Y_1, \dots, Y_{i-1} \right) \\ &= \sum_{i=2}^n E \left(\sum_{j=1}^{i-1} H_n^2(Y_i, Y_j) \mid Y_1, \dots, Y_{i-1} \right) \equiv \sum_{i=2}^n u_i, \end{aligned}$$

όπου:

$$u_i = E \left(\sum_{j=1}^{i-1} H_n^2(Y_i, Y_j) \mid Y_1, \dots, Y_{i-1} \right) = E (P_i^2 \mid Y_1, \dots, Y_{i-1}). \quad (4.39)$$

Το (β) θα αποδειχθεί αν δείξουμε ότι:

$$s_n^{-4} E (V_n^2 - s_n^2)^2 \rightarrow 0. \quad (4.40)$$

Αρχικά,

$$\begin{aligned} E (V_n^4) &= E \left(\sum_{i=2}^n u_i \right)^2 = E \left(\sum_{i=2}^n u_i^2 + 2 \sum_{2 \leq i < j \leq n} u_i u_j \right) \\ &= \sum_{i=2}^n E (u_i^2) + 2 \sum_{2 \leq i < j \leq n} E (u_i u_j), \end{aligned} \quad (4.41)$$

όπου στην πρώτη γραμμή χρησιμοποιήθηκε η (3.48).

Για την (4.39), έχουμε ότι:

$$\begin{aligned} u_i &= E(P_i^2 | Y_1, \dots, Y_{i-1}) = \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} G_n(X_j, X_k) \\ &= 2 \sum_{1 \leq j < k \leq i-1} G_n(X_j, X_k) + \sum_{j=1}^{i-1} G_n(X_j, X_j), \end{aligned} \quad (4.42)$$

όπου η $G_n(x, y)$ δίνεται από την (4.31) με

$$H_n^*(x, X_i) = \int A(x, X_i) A(y, X_i),$$

και

$$A(x, X_i) = \frac{1}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right) - E\left[\frac{1}{1 - F(X_i)} K\left(\frac{x - X_i}{h}\right)\right].$$

Για το $G_n(X_j, X_k)$ (βλέπε Hall (1984), σελ. 5), ισχύει ότι αν $j_1 \leq k_1$ και $j_2 \leq k_2$ τότε:

$$\begin{aligned} E[G_n(X_{j_1}, X_{k_1}) G_n(X_{j_2}, X_{k_2})] &= E(G_n^2(X_1, X_1)), \quad \text{αν } j_1 = k_1 = j_2 = k_2, \\ &= [E(G_n(X_1, X_1))]^2, \quad \text{αν } j_1 = k_1 \neq j_2 = k_2, \\ &= E(G_n^2(X_1, X_2)), \quad \text{αν } j_1 = j_2, k_1 = k_2, j_1 < k_1, \\ &= 0, \quad \text{αλλιώς.} \end{aligned}$$

Επομένως, στην περίπτωση μας, αν $i \leq r$ έχουμε ότι:

$$\begin{aligned} E(u_i, u_r) &= E\left[\left(2 \sum_{1 \leq j < k \leq i-1} G_n(X_j, X_k) + \sum_{j=1}^{i-1} G_n(X_j, X_j)\right)\right. \\ &\quad \times \left.\left(2 \sum_{1 \leq j < k \leq r-1} G_n(X_j, X_k) + \sum_{j=1}^{r-1} G_n(X_j, X_j)\right)\right] \\ &= 4 \sum_{1 \leq j < k \leq i-1} E(G_n^2(X_j, X_k)) + \sum_{j=1}^{i-1} \sum_{k=1}^{r-1} [E(G_n(X_j, X_j))]^2 \\ &\quad + \sum_{j=1}^{i-1} \left\{ E(G_n^2(X_j, X_j)) - [E(G_n(X_j, X_j))]^2 \right\}. \end{aligned}$$

όπου στο πρώτο βήμα χρησιμοποιήθηκε η (4.42).

Λαμβάνοντας υπόψη τη σχέση που συνδέει τη μέση τιμή με τη διακύμανση, η παραπάνω σχέση γίνεται:

$$\begin{aligned}
E(u_i, u_r) &= 2 \sum_{j=1}^{i-2} \sum_{k=j+1}^{i-1} E(G_n^2(X_j, X_k)) + \sum_{j=1}^{i-1} \sum_{k=1}^{r-1} [E(G_n(X_j, X_j))]^2 \\
&\quad + \sum_{j=1}^{i-1} \text{Var}[G_n(X_j, X_j)] \\
&= 2(i-1)(i-2)E(G_n^2(X_j, X_k)) \\
&\quad + (i-1)(r-1)[E(G_n(X_j, X_j))]^2 \\
&\quad + (i-1)\text{Var}[G_n(X_j, X_j)], \tag{4.43}
\end{aligned}$$

Αντικαθιστώντας την (4.43) στην (4.41) προκύπτει το εξής:

$$\begin{aligned}
E(V_n^4) &= E\left(\sum_{i=2}^n u_i\right)^2 \\
&= \left[\frac{1}{2}n(n-1)E(G_n(X_j, X_j))\right]^2 \\
&\quad + 2 \sum_{i=2}^n (i-1)(i-2)(2n-2i+1)E(G_n^2(X_j, X_k)) \\
&\quad + \sum_{i=2}^n (i-1)(2n-2i+1)\text{Var}[G_n(X_j, X_j)],
\end{aligned}$$

απ' όπου:

$$\begin{aligned}
E(V_n^2 - s_n^2)^2 &\leq C [n^4 E(G_n^2(X_j, X_k)) + n^3 E(G_n^2(X_j, X_j))] \\
&\leq C [n^4 E(G_n^2(X_j, X_k)) + n^3 E(H_n^2(X_j, X_k))], \tag{4.44}
\end{aligned}$$

με C θετική σταθερά.

Εργαζόμενοι με τον ίδιο τρόπο όπως και στην απόδειξη της (4.25), αλλά και σύμφωνα με την απόδειξη της σχέσης (4.6) της εργασίας του Hall (1984), ισχύει ότι:

$$E(G_n^2(X_j, X_k)) = O(h^7). \tag{4.45}$$

Κατά συνέπεια, από τις (4.30), (4.34) και (4.45) προκύπτει η (4.40), κάτι το οποίο ολοκληρώνει την απόδειξη του (β).

Επομένως, από το Θεώρημα 1 της εργασίας του Hall (1984), λαμβάνουμε ότι:

$$U_n \sim N \left(0, \frac{1}{2} n^2 E (H_n^2 (X_i, X_j)) \right),$$

και κατά συνέπεια ότι:

$$\frac{2}{n^2 h^2} U_n \equiv \hat{I}_1 \sim N (0, 2n^{-2} h^{-4} E (H_n^2 (X_i, X_j))).$$

Αντικαθιστώντας την (4.19) προκύπτει, τελικά, ότι:

$$\hat{I}_1 \sim N (0, 2n^{-2} h^{-1} \sigma_1^2).$$

□

Λήμμα 4.2.7. Όταν $n \rightarrow +\infty$ τότε

$$\hat{I}_2 = \sigma_2^2 + O_p (n^{-3/2} h^{-1}).$$

Απόδειξη. Αρχικά έχουμε ότι:

$$\begin{aligned} \hat{I}_2 &= \frac{1}{n^2 h^2} \sum_{i=1}^n H_n (Y_i, Y_i) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \int R_i (x, Y_i) R_j (x, Y_i) dx \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \int R_i^2 (x, Y_i) dx \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \int (w_i(x) - E(w_i(x)))^2 dx \equiv \frac{1}{n^2 h^2} \sum_{i=1}^n L_i, \end{aligned}$$

όπου $L_i = \int (w_i(x) - E(w_i(x)))^2 dx$ και επιπλέον στη δεύτερη γραμμή χρησιμοποιήθηκε η (4.10) και στην τέταρτη γραμμή η (4.9).

Στη συνέχεια,

$$\begin{aligned} E(L_i) &= E \int (w_i(x) - E(w_i(x)))^2 dx \\ &= E \int (w_i^2(x) - 2w_i(x)E(w_i(x)) + (E(w_i(x)))^2) dx \\ &= \int E(w_i^2(x)) dx - 2 \int E(w_i(x)) dx + \int (E(w_i(x)))^2 dx, \end{aligned}$$

ή ισοδύναμα:

$$E(L_i) = \int E(w_i^2(x)) dx - \int (E(w_i(x)))^2 dx. \quad (4.46)$$

Για $x = y$ η (4.23) γίνεται:

$$E(w_i^2(x)) = h \int K^2(t) \frac{\lambda_T(x - ht)}{1 - F(x - ht)} dt. \quad (4.47)$$

Αντικαθιστώντας τις (4.24) και (4.47) στην (4.46) παίρνουμε το εξής:

$$\begin{aligned} E(L_i) &= h \int \int K^2(t) \frac{\lambda_T(x - ht)}{1 - F(x - ht)} dt dx \\ &\quad - h^2 \int \left(\int \lambda_T(x - ht) dt \right)^2 dx. \end{aligned} \quad (4.48)$$

Για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$, αναπτύσσοντας σε σειρά Taylor γύρω από το x τον όρο:

$$\frac{\lambda_T(x - ht)}{1 - F(x - ht)},$$

λαμβάνουμε ότι:

$$\frac{\lambda_T(x - ht)}{1 - F(x - ht)} = \frac{\lambda_T(x)}{1 - F(x)} + O(h).$$

Επομένως, θέτοντας ως H_1 τον πρώτο όρο της (4.48) έχουμε ότι:

$$\begin{aligned} H_1 &= h \int \int K^2(t) \frac{\lambda_T(x - ht)}{1 - F(x - ht)} dt dx \\ &= h \int \int K^2(t) \left(\frac{\lambda_T(x)}{1 - F(x)} + O(h) \right) dt dx \\ &= h \int \int K^2(t) \frac{\lambda_T(x)}{1 - F(x)} dt dx + O(h^2) \\ &= h \int K^2(t) dt \int \frac{\lambda_T(x)}{1 - F(x)} dx + O(h^2) \\ &= hR(K) \int \frac{\lambda_T(x)}{1 - F(x)} dx + O(h^2). \end{aligned} \quad (4.49)$$

Επιπλέον για τον δεύτερο όρο της (4.48), τον οποίο και συμβολίζουμε με H_2 ,

έχουμε ότι:

$$\begin{aligned} H_2 &= h^2 \int \left(\int \lambda_T(x - ht) dt \right)^2 dx \\ &= h^2 \int \left(\int \int \lambda_T(x - ht) \lambda_T(x - hu) K(t)K(u) dt du \right) dx, \end{aligned}$$

για τον οποίο χρησιμοποιώντας το Λήμμα 2 της εργασίας του Hall (1984) λαμβάνουμε ότι:

$$H_2 = h^2 \left(\int \int K(t)K(t+u) dt du \right) \int \lambda_T(x)\lambda_T(x+hu) dx. \quad (4.50)$$

Αντικαθιστώντας τις (4.49) και (4.50) στην (4.48) προκύπτει το εξής:

$$\begin{aligned} E(L_i) &= hR(K) \int \frac{\lambda_T(x)}{1-F(x)} dx \\ &\quad - h^2 \left(\int \int K(t)K(t+u) dt du \right) \int \lambda_T(x)\lambda_T(x+hu) dx \\ &\quad + O(h^2). \end{aligned} \quad (4.51)$$

Με παρόμοιο τρόπο με την απόδειξη της σχέσης (5.2) της εργασίας του Hall (1984) προκύπτει ότι:

$$\begin{aligned} E(L_i^2) &= \int \int E \left[(w_i(x) - E(w_i(x)))^2 (w_i(y) - E(w_i(y)))^2 \right] dx dy \\ &\leq C_1 \int \int \left[E(w_i^2(x)w_i^2(y)) + (E(w_i(x)))^2 E(w_i^2(y)) \right. \\ &\quad \left. + E(w_i(x)) E(w_i(x)w_i^2(y)) + (E(w_i(x)))^2 (E(w_i(y)))^2 \right] dx dy \\ &\leq C_2 \int \int (E(w_i^2(x)w_i^2(y)) + E(w_i(x)w_i^2(y))) dx dy \\ &\quad + C_2 \left(\int E(w_i(x)) dx \right)^2, \end{aligned}$$

ή ισοδύναμα ότι:

$$\begin{aligned} E(L_i^2) &\leq C_2 \int \int E(w_i^2(x)w_i^2(y)) dx dy \\ &\quad + C_2 \int \int E(w_i(x)w_i^2(y)) dx dy + C_2 \left(\int E(w_i(x)) dx \right)^2, \end{aligned} \quad (4.52)$$

όπου C_1 και C_2 είναι θετικές σταθερές.

Εξετάζουμε κάθε όρο της (4.52) ξεχωριστά. Αρχικά, θέτοντας:

$$Q_1 = \left(\int E(w_i(x)) dx \right)^2,$$

έχουμε ότι:

$$\begin{aligned} Q_1 &= \int \int E(w_i(x)) E(w_i(y)) \\ &= h^2 \int \int \left(\int \lambda_T(x - ht) K(t) dt \right) \left(\int \lambda_T(y - ht) K(t) dt \right) dx dy, \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε η (4.24).

Το παραπάνω ισοδυναμεί με το:

$$Q_1 = O(h^2). \quad (4.53)$$

Επιπλέον, με παρόμοιο τρόπο όπως στην απόδειξη του Λήμματος 4.2.5, και θέτοντας:

$$Q_2 = \int \int E(w_i(x)w_i^2(y)) dx dy,$$

έχουμε ότι:

$$\begin{aligned} Q_2 &= \int \int E \left[\frac{\delta_i^3}{(1 - F(X_i))^3} K \left(\frac{x - X_i}{h} \right) K^2 \left(\frac{y - X_i}{h} \right) \right] dx dy \\ &= \int \int E \left\{ E \left[\frac{\delta_i^3}{(1 - F(X_i))^3} K \left(\frac{x - X_i}{h} \right) \right. \right. \\ &\quad \left. \left. \times K^2 \left(\frac{y - X_i}{h} \right) \middle| (X_i, \delta_i) = (z, 1) \right] \right\} dx dy \\ &= \int \int E \left[\int \frac{1}{(1 - F(z))^3} K \left(\frac{x - z}{h} \right) K^2 \left(\frac{y - z}{h} \right) \tilde{f}(z) dz \right] dx dy \\ &= \int \int \int \frac{\lambda_T(z)}{(1 - F(z))^2} K \left(\frac{x - z}{h} \right) K^2 \left(\frac{y - z}{h} \right) dz dx dy, \end{aligned}$$

για το οποίο εφαρμόζοντας την αλλαγή μεταβλητής $x - z = ht$ λαμβάνουμε ότι:

$$Q_2 = h \int \int \int \frac{\lambda_T(x - ht)}{(1 - F(x - ht))^2} K(t) K^2 \left(\frac{y - x + ht}{h} \right) dt dx dy.$$

Εφαρμόζοντας επιπλέον την αλλαγή μεταβλητής $y - x = uh$ προκύπτει το εξής:

$$Q_2 = h^2 \int \int \int \frac{\lambda_T(x - ht)}{(1 - F(x - ht))^2} K(t) K^2(u + t) dt dx du.$$

Για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$, αναπτύσσοντας σε σειρά Taylor γύρω από το x τον όρο:

$$\frac{\lambda_T(x - ht)}{(1 - F(x - ht))^2},$$

λαμβάνουμε ότι:

$$\frac{\lambda_T(x - ht)}{(1 - F(x - ht))^2} = \frac{\lambda_T(x)}{(1 - F(x))^2} + O(h).$$

Επομένως,

$$\begin{aligned} Q_2 &= h^2 \int \int \int \left(\frac{\lambda_T(x)}{(1 - F(x))^2} + O(h) \right) K(t) K^2(u + t) dt dx du \\ &= h^2 \int \int \int \frac{\lambda_T(x)}{(1 - F(x))^2} K(t) K^2(u + t) dt dx du + O(h^3) \\ &= O(h^2) + O(h^3) = O(h^2). \end{aligned} \quad (4.54)$$

Με τον ίδιο ακριβώς τρόπο, και θέτοντας:

$$Q_3 = \int \int E(w_i^2(x) w_i^2(y)) dx dy,$$

προκύπτει ότι:

$$\begin{aligned} Q_3 &= h^2 \int \int \int \frac{\lambda_T(x - ht)}{(1 - F(x - ht))^3} K^2(t) K^2(u + t) dt dx du \\ &= h^2 \int \int \int \left(\frac{\lambda_T(x)}{(1 - F(x))^3} + O(h) \right) K^2(t) K^2(u + t) dt dx du \\ &= h^2 \int \int \int \frac{\lambda_T(x)}{(1 - F(x))^3} K^2(t) K^2(u + t) dt dx du + O(h^3) \\ &= O(h^2) + O(h^3) = O(h^2). \end{aligned} \quad (4.55)$$

Αντικαθιστώντας τις (4.53), (4.54) και (4.55) στην (4.52), παίρνουμε το εξής:

$$E(L_i^2) = O(h^2). \quad (4.56)$$

Επιπλέον, χρησιμοποιώντας την (4.51) λαμβάνουμε ότι:

$$\begin{aligned}
 (EL_i)^2 &= \left[hR(K) \int \frac{\lambda_T(x)}{1-F(x)} dx - h^2 \left(\int \int K(t)K(t+u) dtdu \right) \right. \\
 &\quad \left. \times \int \lambda_T(x)\lambda_T(x+hu) dx + O(h^2) \right]^2 \\
 &= [O(h) + O(h^2) + O(h^2)]^2 \\
 &= [O(h)]^2 = O(h^2). \tag{4.57}
 \end{aligned}$$

Επομένως,

$$\begin{aligned}
 Var \{ \hat{I}_2 \} &= Var \left\{ \frac{1}{n^2 h^2} \sum_{i=1}^n L_i \right\} \\
 &= n^{-4} h^{-4} \sum_{i=1}^n Var \{ L_i \} \\
 &= n^{-3} h^{-4} Var \{ L_i \} \\
 &= n^{-3} h^{-4} [E(L_i^2) - (E(L_i))^2] \\
 &= n^{-3} h^{-4} O(h^2) = O(n^{-3} h^{-2}),
 \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε το γεγονός ότι εφόσον οι Y_i είναι ανεξάρτητες και ισόνομες τ.μ. τότε το ίδιο θα ισχύει και για τις τ.μ. L_i , στην τέταρτη γραμμή χρησιμοποιήθηκε η σχέση που συνδέει τη μέση τιμή με τη διακύμανση και στην πέμπτη γραμμή χρησιμοποιήθηκαν οι σχέσεις (4.56) και (4.57).

Θέτοντας:

$$\begin{aligned}
 \sigma_2^2 &= (nh^2)^{-1} E(L_i) \\
 &= (nh^2)^{-1} \left[hR(K) \int \frac{\lambda_T(x)}{1-F(x)} dx - h^2 \left(\int \int K(t)K(t+u) dtdu \right) \right. \\
 &\quad \left. \times \int \lambda_T(x)\lambda_T(x+hu) dx + O(h^2) \right],
 \end{aligned}$$

και κάνοντας χρήση του Λήμματος 2 της εργασίας του Hall (1984) προκύπτει ότι:

$$\hat{I}_2 = \sigma_2^2 + O_p(n^{-3/2} h^{-1}).$$

□

Λήμμα 4.2.8.

$$\begin{aligned} E \int (\tilde{\lambda}(x) - \lambda(x))^2 dx &= \int (E(\tilde{\lambda}(x)) - \lambda(x))^2 + \sigma_2^2 + O(h^2) \\ &\equiv \hat{I}_3 + \sigma_2^2 + O_p(n^{-3/2}h^{-1}). \end{aligned}$$

Απόδειξη. Αρχικά, θέτουμε:

$$J = E \int (\tilde{\lambda}(x) - \lambda(x))^2 dx.$$

Επομένως, έχουμε ότι:

$$\begin{aligned} J &= E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x)) + E(\tilde{\lambda}(x)) - \lambda(x))^2 dx \\ &= E \int \left[(\tilde{\lambda}(x) - E(\tilde{\lambda}(x)))^2 + (E(\tilde{\lambda}(x)) - \lambda(x))^2 \right. \\ &\quad \left. + 2(\tilde{\lambda}(x) - E(\tilde{\lambda}(x)))(E(\tilde{\lambda}(x)) - \lambda(x)) \right] dx \\ &= E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x)))^2 dx + E \int (E(\tilde{\lambda}(x)) - \lambda(x))^2 dx \\ &\quad + 2E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x)))(E(\tilde{\lambda}(x)) - \lambda(x)) dx. \end{aligned} \quad (4.58)$$

Κάνοντας χρήση της (3.56), και θέτοντας:

$$J_1 = E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x)))(E(\tilde{\lambda}(x)) - \lambda(x)) dx,$$

λαμβάνουμε ότι:

$$\begin{aligned} J_1 &= E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x))) \left(\frac{h^2}{2} \lambda''(x) \mu_2(K) \right) dx \\ &= \frac{h^2}{2} \mu_2(K) E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x))) \lambda''(x) dx \\ &= 0, \end{aligned}$$

καθώς,

$$E \int (\tilde{\lambda}(x) - E(\tilde{\lambda}(x))) \lambda''(x) dx = 0.$$

Επομένως η (4.58) γίνεται:

$$J = E \int \left(\tilde{\lambda}(x) - E \left(\tilde{\lambda}(x) \right) \right)^2 dx + E \int \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right)^2 dx.$$

Καθώς,

$$E \int \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right)^2 dx = \hat{I}_2,$$

και δοθέντος του Λήμματος 4.2.7 έχουμε ότι:

$$\begin{aligned} J &= E \int \left(\tilde{\lambda}(x) - E \left(\tilde{\lambda}(x) \right) \right)^2 dx + \hat{I}_2 \\ &= E \int \left(\tilde{\lambda}(x) - E \left(\tilde{\lambda}(x) \right) \right)^2 dx + \sigma_2^2 + O_p \left(n^{-3/2} h^{-1} \right) \\ &= \hat{I}_3 + \sigma_2^2 + O_p \left(n^{-3/2} h^{-1} \right), \end{aligned}$$

το οποίο και ολοκληρώνει την απόδειξη. \square

Λήμμα 4.2.9. Υπό τις συνθήκες 1,2,3 των σελίδων 87-88 και υπό την H_0 : $\lambda(x) = \lambda_0(x)$ ισχύει ότι:

$$\hat{I}_4 \sim N \left(0, n^{-1} h^4 k^2 \sigma_4^2 \right).$$

Απόδειξη. Αρχικά έχουμε ότι:

$$\begin{aligned} \hat{I}_4 &= \int \left(\tilde{\lambda}(x) - E \left(\tilde{\lambda}(x) \right) \right) \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dx \\ &= \int \left[\frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right. \\ &\quad \left. - E \left(\frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right) \right] \\ &\quad \times \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dx \\ &= \frac{1}{nh} \int \left(\sum_{i=1}^n w_i(x) - \sum_{i=1}^n E(w_i(x)) \right) \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dx \\ &= (nh)^{-1} \sum_{i=1}^n \int \left(w_i(x) - E(w_i(x)) \right) \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dx, \end{aligned}$$

όπου στο δεύτερο βήμα χρησιμοποιήθηκε η (4.10) και στο τρίτο βήμα η (4.8).

Θέτοντας:

$$Z_i = \int (w_i(x) - E(w_i(x))) \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) dx,$$

λαμβάνουμε ότι:

$$\hat{I}_4 = (nh)^{-1} \sum_{i=1}^n Z_i.$$

Επιπλέον για την τ.μ. Z_i έχουμε ότι:

$$\begin{aligned} Z_i &= \int (w_i(x) - E(w_i(x))) \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) dx \\ &= \int w_i(x) \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) dx - \int E(w_i(x)) \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) dx. \end{aligned}$$

Θέτοντας επιπλέον ότι:

$$D_i = \int w_i(x) \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) dx,$$

προκύπτουν το εξής:

$$Z_i = D_i - E(D_i), \quad (4.59)$$

με

$$E(Z_i) = 0. \quad (4.60)$$

Χρησιμοποιώντας την (4.59) και θέτοντας επιπλέον ότι:

$$D_{i,1} = E(D_i) \quad \text{και} \quad D_{i,2} = E(D_i^2),$$

λαμβάνουμε το εξής:

$$\begin{aligned} E(Z_i^2) &= E(D_i - E(D_i))^2 \\ &= E\left(D_i^2 - 2D_i E(D_i) + (E(D_i))^2\right) \\ &= E(D_i^2) - 2(E(D_i))^2 + (E(D_i))^2 \\ &= E(D_i^2) - (E D_i)^2 \\ &= D_{i,2} - D_{i,1}^2. \end{aligned} \quad (4.61)$$

Με παρόμοιο τρόπο όπως και στην απόδειξη του Λήμματος 4.2.5 και χρησιμοποιώντας ως συνθήκη τον αριθμό των μη λογοκριμένων παρατηρήσεων του δείγματος $Y_i = (X_i, \delta_i)$, υπολογίζουμε τα $D_{i,1}$ και $D_{i,2}$.

Ξεκινάμε με τον υπολογισμό του $D_{i,1}$.

$$\begin{aligned} D_{i,1} &= E \left[\int \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dx \right] \\ &= \int E \left\{ E \left[\frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \right. \right. \\ &\quad \left. \left. \times \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) \mid (X_i, \delta_i) = (z, 1) \right] \right\} dx \\ &= \int \int \lambda_T(z) K \left(\frac{x - z}{h} \right) \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dz dx \\ &= h \int \int \lambda_T(x - ht) K(t) \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) dt dx, \end{aligned}$$

όπου στην τελευταία γραμμή εφαρμόστηκε η αλλαγή μεταβλητής $x - z = ht$.

Για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$, αναπτύσσοντας σε σειρά Taylor γύρω από το x τον όρο:

$$\lambda_T(x - ht),$$

λαμβάνουμε ότι:

$$\lambda_T(x - ht) = \lambda_T(x) + O(h).$$

Επομένως, κάνοντας χρήση και της (3.56), έχουμε ότι:

$$\begin{aligned} D_{i,1} &= h \int \int (\lambda_T(x) + O(h)) K(t) \left(\frac{h^2}{2} \mu_2(K) \lambda''(x) + o(h^2) \right) dt dx \\ &= h \int \int \left(\frac{h^2}{2} \lambda_T(x) \mu_2(K) \lambda''(x) + o(h^2) + O(h^3) + O(h^3) \right) \\ &\quad \times K(t) dt dx \\ &= h \int \int \left(\frac{h^2}{2} \lambda_T(x) \mu_2(K) \lambda''(x) + o(h^2) \right) K(t) dt dx \\ &= \frac{h^3}{2} \mu_2(K) \int \int \lambda''(x) \lambda_T(x) K(t) dt dx + o(h^3) \\ &= \frac{h^3}{2} \mu_2(K) \int \lambda''(x) \lambda_T(x) dx \int K(t) dt + o(h^3), \end{aligned}$$

ή ισοδύναμα ότι:

$$D_{i,1} = \frac{h^3}{2} \mu_2(K) \int \lambda''(x) \lambda_T(x) dx + o(h^3). \quad (4.62)$$

Συνεχίζοντας με τον υπολογισμό του $D_{i,2}$ έχουμε ότι:

$$\begin{aligned} D_{i,2} &= E \left[\int \frac{\delta_i}{1 - F(X_i)} K \left(\frac{x - X_i}{h} \right) \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) dx \right]^2 \\ &= E \left[\int \int \frac{\delta_i^2}{(1 - F(X_i))^2} K \left(\frac{x - X_i}{h} \right) K \left(\frac{y - X_i}{h} \right) \right. \\ &\quad \left. \times \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) \left(E(\tilde{\lambda}(y)) - \lambda(y) \right) dx dy \right] \\ &= \int \int E \left[\frac{\delta_i^2}{(1 - F(X_i))^2} K \left(\frac{x - X_i}{h} \right) K \left(\frac{y - X_i}{h} \right) \right. \\ &\quad \left. \times \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) \left(E(\tilde{\lambda}(y)) - \lambda(y) \right) \right] dx dy \\ &= \int \int E \left\{ E \left[\frac{\delta_i^2}{(1 - F(X_i))^2} K \left(\frac{x - X_i}{h} \right) K \left(\frac{y - X_i}{h} \right) \right. \right. \\ &\quad \left. \left. \times \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) \left(E(\tilde{\lambda}(y)) - \lambda(y) \right) \mid (X_i, \delta_i) = (z, 1) \right] \right\} dx dy \\ &= \int \int E \left[\int \frac{1}{(1 - F(z))^2} K \left(\frac{x - z}{h} \right) K \left(\frac{y - z}{h} \right) \right. \\ &\quad \left. \times \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) \left(E(\tilde{\lambda}(y)) - \lambda(y) \right) \tilde{f}(z) dz \right] dx dy \\ &= \int \int \int \frac{\lambda_T(z)}{1 - F(z)} K \left(\frac{x - X_i}{h} \right) K \left(\frac{y - X_i}{h} \right) \\ &\quad \times \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) \left(E(\tilde{\lambda}(y)) - \lambda(y) \right) dz dx dy \\ &= h \int \int \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t) K \left(\frac{y - x + ht}{h} \right) \\ &\quad \times \left(E(\tilde{\lambda}(x)) - \lambda(x) \right) \left(E(\tilde{\lambda}(y)) - \lambda(y) \right) dt dx dy, \end{aligned}$$

όπου στο τέταρτο βήμα χρησιμοποιήθηκε η (3.36) και στο έβδομο βήμα έγινε η αλλαγή μεταβλητής $x - z = ht$.

Πραγματοποιώντας στην παραπάνω σχέση την αλλαγή μεταβλητής $y - x = hu$, λαμβάνουμε τελικά ότι:

$$D_{i,2} = h^2 \int \int \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t)K(u + t) \\ \times \left(E \left(\tilde{\lambda}(x) \right) - \lambda(x) \right) \left(E \tilde{\lambda}(x + hu) - \lambda(x + hu) \right) dt d u d x,$$

Στη συνέχεια, κάνοντας χρήση της (3.56) προκύπτει ότι:

$$D_{i,2} = h^2 \int \int \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t)K(u + t) \\ \times \left(\frac{h^2}{2} \lambda''(x) \mu_2(K) + o(h^2) \right) \left(\frac{h^2}{2} \lambda''(x + hu) \mu_2(K) + o(h^2) \right) dt d u d x \\ = h^2 \int \int \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} K(t)K(u + t) \\ \times \left(\frac{h^4}{4} \mu_2^2(K) \lambda''(x) \lambda''(x + hu) + o(h^4) \right) dt d u d x \\ = \frac{h^6}{4} \mu_2^2(K) \int \int \int \frac{\lambda_T(x - ht)}{1 - F(x - ht)} \lambda''(x) \lambda''(x + hu) \\ \times K(t)K(u + t) dt d u d x + o(h^6).$$

Για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$, αναπτύσσοντας σε σειρά Taylor γύρω από το x τους όρους:

$$\frac{\lambda_T(x - ht)}{1 - F(x - ht)} \quad \text{και} \quad \lambda''(x + hu),$$

λαμβάνουμε αντίστοιχα ότι:

$$\frac{\lambda_T(x - ht)}{1 - F(x - ht)} = \frac{\lambda_T(x)}{1 - F(x)} + O(h) \quad \text{και} \quad \lambda''(x + hu) = \lambda''(x) + O(h).$$

Επομένως, παίρνουμε το εξής:

$$D_{i,2} = \frac{h^6}{4} \mu_2^2(K) \int \int \int \left(\frac{\lambda_T(x)}{1 - F(x)} + O(h) \right) \left(\lambda''(x) + O(h) \right) \lambda''(x) \\ \times K(t)K(u + t) dt d u d x + o(h^6) \\ = \frac{h^6}{4} \mu_2^2(K) \int \int \int \left(\frac{\lambda_T(x)}{1 - F(x)} \lambda''(x) + O(h) \right) \lambda''(x) \\ \times K(t)K(u + t) dt d u d x + o(h^6),$$

το οποίο ισοδύναμα γίνεται:

$$\begin{aligned}
D_{i,2} &= \frac{h^6}{4} \mu_2^2(K) \int \int \int (\lambda''(x))^2 \frac{\lambda_T(x)}{1-F(x)} K(t)K(u+t) dt du dx \\
&\quad + o(h^6) \\
&= \frac{h^6}{4} \mu_2^2(K) \int (\lambda''(x))^2 \frac{\lambda_T(x)}{1-F(x)} dx \int \int K(t)K(u+t) du dt \\
&\quad + o(h^6) \\
&= \frac{h^6}{4} \mu_2^2(K) \left[\int (\lambda''(x))^2 \frac{\lambda_T(x)}{1-F(x)} dx \right] R(B(u)) + o(h^6). \quad (4.63)
\end{aligned}$$

Από τις (4.62), (4.63) και αντικαθιστώντας την σ_4^2 , η (4.61) γίνεται:

$$\begin{aligned}
EZ_i^2 &= \frac{h^6}{4} \mu_2^2(K) \left(\int (\lambda''(x))^2 \frac{\lambda_T(x)}{1-F(x)} dx \right) R(B(u)) + o(h^6) \\
&\quad - \left[\frac{h^3}{2} \mu_2(K) \int \lambda''(x) \lambda_T(x) dx + o(h^3) \right]^2 \\
&= \frac{h^6}{4} \mu_2^2(K) \left(\int (\lambda''(x))^2 \frac{\lambda_T(x)}{1-F(x)} dx \right) R(B(u)) + o(h^6) \\
&\quad - \left[\frac{h^6}{4} \mu_2^2(K) \left(\int \lambda''(x) \lambda_T(x) dx \right)^2 + o(h^6) \right] \\
&= \frac{h^6}{4} \mu_2^2(K) \left[\left(\int (\lambda''(x))^2 \frac{\lambda_T(x)}{1-F(x)} dx \right) R(B(u)) \right. \\
&\quad \left. - \left(\int \lambda''(x) \lambda_T(x) dx \right)^2 \right] + o(h^6) \\
&= \frac{h^6}{4} \mu_2^2(K) \sigma_4^2 + o(h^6). \quad (4.64)
\end{aligned}$$

Εργαζόμενοι κατά τον ίδιο τρόπο, εύκολα μπορεί να αποδειχθεί ότι:

$$E(Z_i^4) = O(h^{12}).$$

Θέτοντας:

$$s_n^2 = \sum_{i=1}^n E(Z_i^2),$$

και αντικαθιστώντας την (4.64) στην (4.35), λαμβάνουμε ότι:

$$s_n^{-2} \sum_{i=1}^n E [Z_i^2 I(|Z_i| > \epsilon s_n)] \leq \epsilon^{-2} s_n^{-4} \sum_{i=1}^n E(Z_i^4) \rightarrow 0,$$

όταν $n \rightarrow +\infty$.

Επομένως, το \hat{I}_4 ακολουθεί την κανονική κατανομή με μέση τιμή μηδέν και διακύμανση που δίνεται από τη σχέση:

$$\begin{aligned} \text{Var}\{\hat{I}_4\} &= \text{Var} \left\{ (nh)^{-1} \sum_{i=1}^n Z_i \right\} \\ &= (nh)^{-2} \sum_{i=1}^n \text{Var}\{Z_i\} \\ &= n^{-1} h^{-2} \text{Var}\{Z_i\} \\ &= n^{-1} h^{-2} [E Z_i^2 - (E Z_i)^2] \\ &= n^{-1} h^{-2} E Z_i^2 \\ &= n^{-1} h^4 \left(\frac{\mu_2(K)}{2} \right)^2 \sigma_4^2 \\ &= n^{-1} h^4 k^2 \sigma_4^2, \end{aligned}$$

όπου στη δεύτερη γραμμή χρησιμοποιήθηκε το γεγονός ότι εφόσον οι Y_i είναι ανεξάρτητες και ισόνομες τ.μ. τότε το ίδιο θα ισχύει και για τις Z_i , στην τέταρτη γραμμή χρησιμοποιήθηκε η σχέση που συνδέει τη μέση τιμή με τη διακύμανση, στην πέμπτη γραμμή χρησιμοποιήθηκε η (4.60) και στην έκτη γραμμή χρησιμοποιήθηκε η (4.64).

Επομένως,

$$\hat{I}_4 \sim N(0, n^{-1} h^4 k^2 \sigma_4^2).$$

□

Εφαρμόζοντας τα Λήμματα 4.2.6, 4.2.7, 4.2.8 και 4.2.9 και λαμβάνοντας υπόψη ότι η τ.μ. Z ακολουθεί ασυμπτωτικά τυπική κανονική κατανομή, εύκολα προκύπτει ότι:

$$\hat{I}_1 = 2^{1/2} n^{-1} h^{-1/2} \sigma_1 Z,$$

και

$$\hat{I}_4 = n^{-1/2} h^2 k \sigma_4 Z.$$

Επομένως, χρησιμοποιώντας την (4.11), έχουμε ότι:

$$\begin{aligned} I_s - E(I_s) &= \hat{I}_1 + \hat{I}_2 + \hat{I}_3 + 2\hat{I}_4 - E \int (\tilde{\lambda}(x) - \lambda(x))^2 dx \\ &= 2^{1/2}n^{-1}h^{-1/2}\sigma_1 Z + \sigma_2^2 + O_p(n^{-3/2}h^{-1}) + \hat{I}_3 + 2n^{-1/2}h^2k\sigma_4 Z \\ &\quad - \hat{I}_3 - \sigma_2^2 + O_p(n^{-3/2}h^{-1}) \\ &= 2^{1/2}n^{-1}h^{-1/2}\sigma_1 Z + 2n^{-1/2}h^2k\sigma_4 Z + O_p(n^{-3/2}h^{-1}), \end{aligned}$$

απ' όπου παρατηρούμε ότι για βέλτιστο εύρος ζώνης $h \sim n^{-1/5}$, ισχύει ότι $O_p(n^{-3/2}h^{-1}) = O_p(n^{-13/10})$ το οποίο είναι ασυμπτωτικά αμελητέο. Δοθέντος ότι το I_n είναι ασυμπτωτικά ισοδύναμο με το I_s , ολοκληρώνεται η απόδειξη του Θεωρήματος 4.2.1.

Παρατήρηση 4.2.10. Είναι προφανές από το Θεώρημα 4.2.1. ότι η ασυμπτωτική κατανομή του I_n εξαρτάται από την ποσότητα της εξομάλυνσης που εφαρμόζεται στα δεδομένα.

Αν $nh^5 \rightarrow 0$, τότε τα δεδομένα υπόκεινται σε υποεξομάλυνση και το στοχαστικό μέρος της απόκλισης κυριαρχεί του συστηματικού μέρους. Στην περίπτωση αυτή, η στοχαστική συμπεριφορά του $d(n) (I_n - (I_n))$ καθορίζεται από τον όρο:

$$\int (\hat{\lambda}(x) - E(\hat{\lambda}(x)))^2 dx.$$

Στην περίπτωση της υπερεξομάλυνσης ($nh^5 \rightarrow +\infty$), το στοχαστικό μέρος του $d(n) (I_n - E(I_n))$ είναι ασυμπτωτικά αμελητέο σε σύγκριση με το συστηματικό μέρος. Τότε η οριακή κατανομή του $d(n) (I_n - (I_n))$ καθορίζεται από τον όρο:

$$\int (\hat{\lambda}(x) - E(\hat{\lambda}(x))) (E(\hat{\lambda}(x)) - \lambda(x)) dx.$$

Για βέλτιστη εξομάλυνση, δηλαδή όταν $nh^5 \rightarrow \lambda$, κανένας από τους δύο όρους δεν κυριαρχεί καθώς, ασυμπτωτικά, έχουν την ίδια συνεισφορά. Στην περίπτωση αυτή η ασυμπτωτική κατανομή του $d(n) (I_n - (I_n))$ προσδιορίζεται ως το άθροισμα των κατανομών των δύο παραπάνω όρων.

Εφαρμογή του Θεωρήματος 4.2.1

Στη συνέχεια, το Θεώρημα 4.2.1 εφαρμόζεται για τη δημιουργία ενός ελέγχου καλής προσαρμογής. Υπό την μηδενική υπόθεση H_0 , το Θεώρημα 4.2.1 και το

Λήμμα 4.2.8, η στατιστική συνάρτηση \hat{T}_n που δίνεται από την (4.2) γράφεται ως:

$$\hat{T}_{n,*} = d(n) \left(\int (\hat{\lambda}(x) - \lambda_0(x))^2 dx - c(n) \right),$$

όπου:

$$c(n) = \int \left(E(\hat{\lambda}(x)) - \lambda_0(x) \right)^2 dx + \sigma_2^2,$$

και

$$\sigma_2^2 = (nh^2)^{-1} \left[hR(K) \int \frac{\lambda_T(x)}{1-F(x)} dx - h^2 \left(\int \int K(t)K(t+u) dt du \right) \int \lambda_T(x)\lambda_T(x+hu) dx \right].$$

Επομένως, για τον έλεγχο της μηδενικής υπόθεσης $H_0 : \lambda(x) = \lambda_0(x)$ έναντι της εναλλακτικής $H_1 : \lambda(x) \neq \lambda_0(x)$ με επίπεδο σημαντικότητας a έχουμε ότι:

$$\frac{\hat{T}_{n,*}}{\sqrt{\text{Var}\{\hat{T}_{n,*}\}}} \xrightarrow{d} N(0,1),$$

όπου:

$$\text{Var}\{\hat{T}_{n,*}\} = \begin{cases} 2\sigma_1^2, & \text{αν } nh^5 \rightarrow 0, \\ 4k^2\sigma_4^2, & \text{αν } nh^5 \rightarrow +\infty, \\ 2\sigma_1^2\lambda^{-1/5} + 4k^2\sigma_4^2\lambda^{4/5}, & \text{αν } nh^5 \rightarrow \lambda. \end{cases}$$

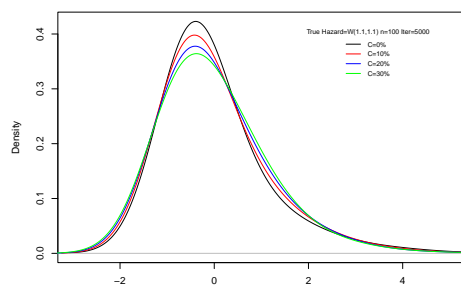
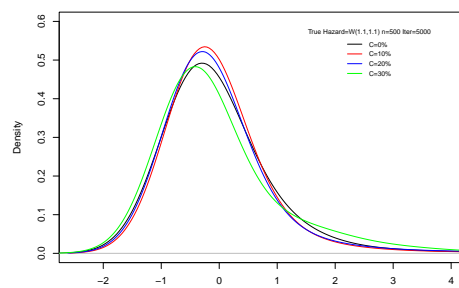
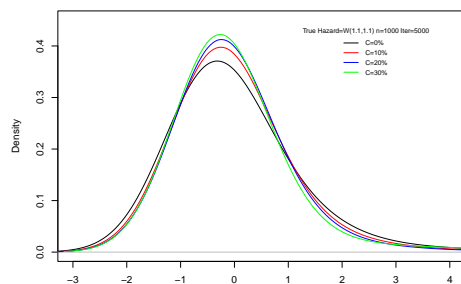
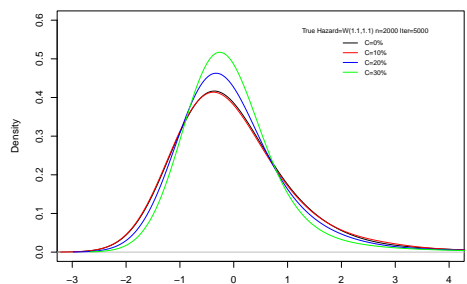
Κατά συνέπεια ο έλεγχος υποδηλώνει απόρριψη της H_0 όταν:

$$\frac{\hat{T}_{n,*}}{\sqrt{\text{Var}\{\hat{T}_{n,*}\}}} > z_a,$$

όπου z_a είναι ποσοστιαίο σημείο της τυπικής κανονικής κατανομής για επίπεδο σημαντικότητας a για το οποίο ισχύει ότι $P(Z > z_a) = a$.

Στο Σχήμα 4.1 παρουσιάζεται η κατανομή της τυποποιημένης στατιστικής συνάρτησης του προτεινόμενου ελέγχου καλής προσαρμογής, όπως αυτή προέκυψε μετά από 5.000 επαναλήψεις. Για την υλοποίηση τόσο του εκτιμητή πυρήνα, όσο και της συνάρτησης κινδύνου υπό τη μηδενική υπόθεση, χρησιμοποιήθηκαν δεδομένα που προέρχονται από την κατανομή Weibull με παραμέτρους

$\kappa = \lambda = 1.1$ (Σύμβ: $W(1.1, 1.1)$). Τα ποσοστά λογοκρισίας (Σύμβ: C) που χρησιμοποιήθηκαν είναι 0% (καθόλου λογοκρισία), 10%, 20% και 30%, με τη μελέτη να πραγματοποιείται για καθένα από αυτά τα ποσοστά για μέγεθος δείγματος $n = 100, 500, 1000, 2000$ αντίστοιχα.

(α') $C=0\%$ (β') $C=10\%$ (γ') $C=20\%$ (δ') $C=30\%$

Σχήμα 4.1: Κατανομή της τυποποιημένης στατιστικής συνάρτησης του προτεινόμενου έλεγχου καλής προσαρμογής για διάφορα μεγέθη δείγματος και ποσοστά λογοκρισίας.

ΚΕΦΑΛΑΙΟ 5

ΠΡΟΣΟΜΟΙΩΣΕΙΣ

Ο κύριος στόχος του κεφαλαίου αυτού είναι η μελέτη της συνάρτησης ισχύος του προτεινόμενου ελέγχου καλής προσαρμογής. Στη συνέχεια πραγματοποιούνται προσομοιώσεις για τη συνάρτηση ισχύος για διάφορα μεγέθη δείγματος, ποσοστά λογοκρισίας και διάφορες κατανομές χρόνων ζωής με σκοπό να δοθούν αριθμητικά αποτελέσματα για την απόδοση του ελέγχου.

Προκειμένου να προκύψει ένας υπολογιστικά βολικότερος τύπος για την πραγματοποίηση του ελέγχου, χρησιμοποιείται η σ.π.π. του δείγματος ως συνάρτηση βάρους για τα ολοκληρώματα που περιέχονται στη στατιστική συνάρτηση \hat{T}_n , όπως αυτή έχει οριστεί στη σχέση (4.2). Το γεγονός της στάθμησης του ολοκληρώματος μίας συνάρτησης οδηγεί στην αναμενόμενη τιμή του, η οποία στη συνέχεια μπορεί να εκτιμηθεί από τη δειγματική μέση τιμή της. Επομένως λαμβάνουμε τη δειγματική εκδοχή του \hat{T}_n , η οποία δίνεται από τη σχέση:

$$\tilde{T}_n = \frac{1}{n} \sum_{i=1}^n \left(\hat{\lambda}(X_i) - \lambda(X_i) \right)^2.$$

5.1 Υλοποίηση του ελέγχου

Ξεκινώντας την υλοποίηση του προτεινόμενου ελέγχου καλής προσαρμογής, ο πυρήνας Epanechnikov που δίνεται από τη σχέση:

$$K(x) = \begin{cases} \frac{3}{4} (1 - x^2), & \text{για } |x| \leq 1, \\ 0, & \text{αλλιώς,} \end{cases}$$

χρησιμοποιείται στην παρούσα ενότητα. Για αυτόν τον πυρήνα προκύπτει, μετά από υπολογισμούς ότι:

$$R(K) = \frac{3}{5}. \quad (5.1)$$

Η υλοποίηση του εκτιμητή πυρήνα της συνάρτησης κινδύνου, που δίνεται από την σχέση (3.28), πραγματοποιείται χρησιμοποιώντας το ασυμπτωτικά βέλτιστο εύρος ζώνης του μέσου ολοκληρωμένου τετραγωνικού σφάλματος (η σχέση (9) της εργασίας των Hua et al. (2018)), το οποίο λαμβάνεται με τη μέθοδο plug in, και δίνεται από τη σχέση:

$$h_{opt} = \left[\frac{R(K)}{nR(\lambda_T'')\mu_2^2(K)} \int \frac{\lambda_T(x)}{1-F(x)} dx \right]^{1/5}. \quad (5.2)$$

Συνεχίζουμε με την υλοποίηση της (5.2). Αρχικά, μετά από υπολογισμούς έχουμε ότι:

$$\mu_2(K) = \frac{1}{5}.$$

Έπειτα, η εκτίμηση του $R(\lambda_T'')$ δίνεται από τη σχέση:

$$\hat{R}(\lambda_T'') = \int_0^\xi \left(\hat{\lambda}_T''(x|\hat{b}_n^*) \right)^2 dx, \quad (5.3)$$

και το ολοκλήρωμα:

$$\int_0^T \frac{\lambda_T(x)}{1-F(x)} dx,$$

εκτιμάται εφαρμόζοντας τον ολοκληρωτικό κανόνα του Simpson (η σχέση (9) της εργασίας του Weisstein (2006)) στο:

$$\frac{\hat{\lambda}_T(x|\hat{b}_n^*)}{1-\hat{F}(x)}. \quad (5.4)$$

Στις σχέσεις (5.3) και (5.4), το εύρος ζώνης b_n^* εκτιμάται από μία εκδοχή της διαδικασίας rule-of-thumb για την επιλογή εύρους ζώνης για την Γκαουσιανή εκτίμηση πυρήνα που προτάθηκε από τον Silverman (1986). Η εκδοχή που αναφέρθηκε, δόθηκε από τον Scott (1992) και υπάρχει στο πακέτο **stats** της R μέσω της συνάρτησης **bw.nrd**.

Για τον παρονομαστή της (5.4), σύμφωνα με τη σχέση (3.14), ισχύει ότι:

$$1 - \hat{F}(x) = \left(1 - \hat{F}_T(x)\right) \left(1 - \hat{F}_U(x)\right).$$

Στην παραπάνω σχέση, η ποσότητα $1 - \hat{F}_T(x)$, εκτιμάται από τον εκτιμητή Kaplan-Meier (βλέπε Kaplan and Meier (1958) και Marron and Padgett (1987)) ο

οποίος δίνεται από τη σχέση:

$$1 - \hat{F}_T(x) = \begin{cases} 1, & 0 \leq x \leq X_{(1)}, \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}}, & X_{(k-1)} < x \leq X_{(k)}, k = 2, \dots, n, \\ 0, & X_{(n)} \leq x. \end{cases}$$

Αντίστοιχα, η ποσότητα $1 - \hat{F}_U(x)$, εκτιμάται από τον (ελαφρώς τροποποιημένο) εκτιμητή Kaplan-Meier (βλέπε Blum and Susarla (1980) και Marron and Padgett (1987)), που δίνεται από τη σχέση:

$$1 - \hat{F}_U(x) = \begin{cases} 1, & 0 \leq x \leq X_{(1)}, \\ \prod_{i=1}^{k-1} \left(\frac{n-i+1}{n-i+2} \right)^{1-\delta_{(i)}}, & X_{(k-1)} < x \leq X_{(k)}, k = 2, \dots, n, \\ \prod_{i=1}^n \left(\frac{n-i+1}{n-i+2} \right)^{1-\delta_{(i)}}, & X_{(n)} \leq x. \end{cases}$$

Οι σχέσεις (5.1), (5.3) και (5.4) ολοκληρώνουν την υλοποίηση της σχέσης (5.2).

Στη συνέχεια μελετάται η επίδοση της ελεγχουσυνάρτησης του προτεινόμενου ελέγχου καλής προσαρμογής τόσο όσον αφορά το επίπεδο σημαντικότητας όσο και την ισχύ. Για τον λόγο αυτόν χρησιμοποιούνται τρεις οικογένειες κατανομών χρόνων ζωής, η Εκθετική με παράμετρο λ (Σύμβ: $Exp(\lambda)$) και οι Weibull και Γάμμα με παραμέτρους k και λ (Σύμβ: $W(k, \lambda)$ και $G(k, \lambda)$ αντίστοιχα). Η μηδενική υπόθεση διατυπώνεται ως $H_0 : \lambda(x) = \lambda_0(x)$ όπου $\lambda_0(x) = 1$, η οποία αντιστοιχεί στη συνάρτηση κινδύνου οποιασδήποτε από τις $Exp(1)$, $W(1, 1)$ και $G(1, 1)$ καθώς είναι ισοδύναμες. Προκειμένου να εκτιμηθεί η ισχύς του ελέγχου, γίνεται προσέγγιση της πιθανότητας απόρριψης της μηδενικής υπόθεσης δοθέντος ότι η εναλλακτική υπόθεση είναι αληθής, για κάθε μία από τις 12 εναλλακτικές που υιοθετούνται. Κάθε εναλλακτική συνάρτηση κινδύνου, έστω $\lambda_1(x)$, αποτελεί τμήμα μίας σειράς εναλλάκτικων υποθέσεων οι οποίες όλες ανήκουν στην ίδια οικογένεια κατανομών με την $\lambda_0(x)$ και οι παράμετροι τους επιλέγονται με τέτοιο τρόπο ώστε η απόκλιση Kullback-Leibler ανάμεσα στις $\lambda_0(x)$ και $\lambda_1(x)$ να αυξάνει όλο και περισσότερο (βλέπε Παράρτημα Α').

Καθώς ο έλεγχος αφορά συναρτήσεις κινδύνου για δεξιά λογοκριμένα δεδομένα, ο τύπος που χρησιμοποιήθηκε για τη μέτρηση της απόκλισης Kullback-Leibler μεταξύ της μηδενικής και των εναλλακτικών συναρτήσεων κινδύνου, προτάθηκε από τους Park and Shin (2014) και δίνεται από τη σχέση:

$$KL_{(0,C)}(\lambda_0(x) : \lambda_1(x)) = \int_0^C f_0(x) \left(\frac{\lambda_1(x)}{\lambda_0(x)} - \log \frac{\lambda_1(x)}{\lambda_0(x)} - 1 \right) dx, \quad (5.5)$$

όπου $f_0(x)$ είναι η σ.π.π. της κατανομής υπό τη μηδενική υπόθεση και C είναι ο χρόνος λογοκρισίας ο οποίος θεωρείται σταθερός.

Επομένως, για κάθε κατανομή, μέγεθος δείγματος και ποσοστό λογοκρισίας, η υπό συνθήκη πιθανότητα απόρριψης της $H_0 : \lambda(x) = \lambda_0(x)$ δοθέντος ότι η $H_1 : \lambda(x) = \lambda_1(x)$ είναι αληθής, προσεγγίζεται από $m = 10.000$ ανεξάρτητα τυχαία δείγματα από τη σχέση:

$$P(\tilde{T}_n > \text{cut-off}) = \frac{\#\{\tilde{T}_n > \text{cut-off}\}}{m}. \quad (5.6)$$

Όπου cut-off είναι τα σημεία αποκοπής της κατανομής του \tilde{T}_n για επίπεδο σημαντικότητας a , τα οποία δίνονται από τη σχέση:

$$\text{cut-off} = z_a \sqrt{\text{Var}(\tilde{T}_n)} + E(\tilde{T}_n),$$

με z_a είναι ποσοστιαίο σημείο της τυπικής κανονικής κατανομής για το οποίο ισχύει ότι $P(Z > z_a) = a$.

Υιοθετώντας επίπεδο σημαντικότητας $a = 5\%$ (πιθανότητα σφάλματος τύπου I), τα σημεία αποκοπής (cut-off points) στην (5.6) εκτιμώνται από το 95% ποσοστημόριο της κατανομής του $\tilde{T}(h)$. Η κατανομή της στατιστικής συνάρτησης προσεγγίζεται παράγοντας 100.000 τιμές της για κάθε μέγεθος δείγματος, χωρίς λογοκρισία και υπό την μηδενική υπόθεση. Για τον προσδιορισμό των σημείων αποκοπής χρησιμοποιείται η συνάρτηση **quantile** της R με επιλογή type=7 (βλέπε Ορισμό 7, Hyndman and Fan (1996), σελ. 363).

Παρατήρηση 5.1.1. Τα σημεία αποκοπής που λαμβάνονται διαφέρουν ανάλογα με το μέγεθος του δείγματος και όχι ανάλογα με το επίπεδο λογοκρισίας. Η επιλογή σημείων αποκοπής χωρίς την ύπαρξη λογοκρισίας γίνεται με σκοπό να ληφθεί μία ένδειξη για το πως η λογοκρισία επηρεάζει την παραγόμενη ισχύ.

5.2 Αποτελέσματα των προσομοιώσεων

Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα των προσομοιώσεων του προτεινόμενου έλεγχου καλής προσαρμογής για τη συνάρτηση του ρυθμού επικινδυνότητας σε επίπεδο σημαντικότητας $a = 5\%$. Τα ευρήματα προέκυψαν μετά από 100.000 επαναλήψεις για τον προσδιορισμό των σημείων αποκοπής και 10.000 επαναλήψεων για τον προσδιορισμό των τιμών της εμπειρικής συνάρτησης ισχύος. Τα μεγέθη δείγματος τα οποία λαμβάνονται είναι $n = 50, 80, 100, 150, 200, 300$

και τα ποσοστά λογοκρισίας είναι $C = 0\%, 10\%, 20\%, 30\%$. Στο Παράρτημα Α' παρατίθενται οι πίνακες των αποτελεσμάτων.

Η μηδενική υπόθεση του ελέγχου δίνεται ως:

$$H_0 : \lambda(x) = \lambda_0(x),$$

όπου $\lambda_0(x) = 1$ η οποία αντιστοιχεί στη συνάρτηση κινδύνου κάθε μίας από τις κατανομές $Exp(1)$, $W(1, 1)$ και $G(1, 1)$.

Οι εναλλακτικές υποθέσεις δίνονται ως:

$$H_1 : \lambda(x) = \lambda_1(x),$$

όπου οι $\lambda_1(x)$ αποτελούν συναρτήσεις κινδύνου των $Exp(\lambda)$, $W(\kappa, \lambda)$ και $G(\kappa, \lambda)$ αντίστοιχα. Οι παράμετροι τις εκάστοτε εναλλακτικής συνάρτησης κινδύνου επιλέγονται με τέτοιο τρόπο ώστε η απόκλιση Kullback-Leibler ανάμεσα στη μηδενική και την εναλλακτική υπόθεση να αυξάνει όλο και περισσότερο (βλέπε Παράρτημα Α'). Συγκεκριμένα ισχύει ότι:

- Αν η $\lambda_1(x)$ αποτελεί συνάρτηση κινδύνου της Εκθετικής κατανομής τότε :

$$\lambda_1(x) = \lambda, \quad \text{για } \lambda = 1, 1.1, \dots, 2.2$$

- Αν η $\lambda_1(x)$ αποτελεί συνάρτηση κινδύνου της Weibull κατανομής τότε:

$$\lambda_1(x) = \left(\frac{\kappa}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{\kappa-1}, \quad \text{για } \kappa = \lambda = 1, 1.1, \dots, 2.2$$

- Αν η $\lambda_1(x)$ αποτελεί συνάρτηση κινδύνου της Γάμμα κατανομής τότε

$$\lambda_1(x) = \frac{x^{\kappa-1} e^{-\frac{x}{\lambda}}}{\lambda^{\kappa} (\Gamma(\kappa) - \gamma(\kappa, \frac{x}{\lambda}))}, \quad \text{για } \kappa = \lambda = 1, 1.1, \dots, 2.2$$

όπου Γ είναι η συνάρτηση γάμμα και γ είναι η κατώτερη ελλιπής συνάρτηση γάμμα (lower incomplete gamma function).

Τα αποτελέσματα που προέκυψαν σε κάθε περίπτωση παρουσιάζονται στις επόμενες υποενότητες.

Παρατήρηση 5.2.1. Και στις τρεις περιπτώσεις εναλλακτικών υποθέσεων που αναφέρονται παραπάνω, επιλέγονται ως παράμετροι και αυτοί που αντιστοιχούν στη συνάρτηση κινδύνου υπό τη μηδενική υπόθεση. Η επιλογή αυτή πραγματοποιείται με σκοπό τον έλεγχο της συνέπειας του τεστ ως προς το θεωρητικό επίπεδο σημαντικότητας (nominal level).

5.2.1 Αποτελέσματα του ελέγχου για εναλλακτικές συναρτήσεις κινδύνου από την Εκθετική κατανομή

Για την πραγματοποίηση του ελέγχου της υπόθεσης:

$$H_0 : \lambda(x) = 1 \quad \text{έναντι} \quad H_1 : \lambda(x) = \lambda, \quad \text{για } \lambda = 1, 1.1, \dots, 2.2,$$

τα αποτελέσματα τα οποία προέκυψαν δίνονται στο Σχήμα 5.1, το οποίο περιέχει την ισχύ του ελέγχου για τα διάφορα μεγέθη δείγματος και ποσοστά λογοκρισίας που αναφέρθηκαν στην αρχή της Ενότητας 5.2.

Παρατηρούμε, όπως αναμενόταν, ότι η ισχύς αυξάνεται καθώς αυξάνεται η απόκλιση από την H_0 . Παράλληλα η ισχύς αυξάνεται όταν αυξάνεται και το μέγεθος του δείγματος. Επιπλέον παρατηρούμε ότι ο έλεγχος είναι συνεπής ως προς το θεωρητικό επίπεδο σημαντικότητας (nominal level) για ποσοστά λογοκρισίας 0%, 10% και 20%, ενώ για ποσοστό λογοκρισίας 30% το θεωρητικό επίπεδο διπλασιάζεται όταν το μέγεθος του δείγματος είναι ίσο με 300. Στη συνέχεια, παρατηρούμε ότι, όσο περισσότερο αυξάνεται η λογοκρισία τόσο περισσότερο αυξάνεται η ισχύς. Ωστόσο όμως με τη χρήση σημείων αποκοπής που να βασίζονται στην κατανομή του στατιστικού υπό την ύπαρξη λογοκρισίας, να μπορέσουμε να λάβουμε θεωρητικό επίπεδο σημαντικότητας και τιμές ισχύος πολύ κοντά σε αυτές που λήφθηκαν χωρίς την ύπαρξη λογοκρισίας. Τέλος, παρά τη λογοκρισία, το Σχήμα 5.1 υποδηλώνει ότι όσο μεγαλύτερη είναι η απόκλιση από τη μηδενική υπόθεση, τόσο πιο “ευαίσθητος” γίνεται ο έλεγχος στην ανίχνευση των εναλλακτικών συναρτήσεων κινδύνου.

5.2.2 Αποτελέσματα του ελέγχου για εναλλακτικές συναρτήσεις κινδύνου από την Weibull κατανομή

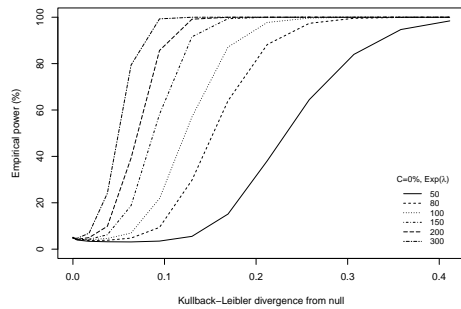
Για την πραγματοποίηση του ελέγχου της υπόθεσης:

$$H_0 : \lambda(x) = 1 \quad \text{έναντι} \quad H_1 : \lambda(x) = \left(\frac{\kappa}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{\kappa-1},$$

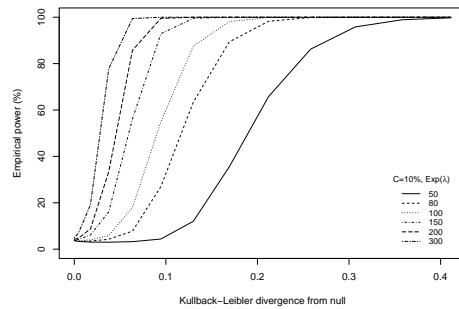
για $\kappa = \lambda = 1, 1.1, \dots, 2.2,$

τα αποτελέσματα τα οποία προέκυψαν δίνονται στο Σχήμα 5.2, το οποίο περιέχει την ισχύ του ελέγχου για τα διάφορα μεγέθη δείγματος και ποσοστών λογοκρισίας που αναφέρθηκαν στην αρχή της Ενότητας 5.2.

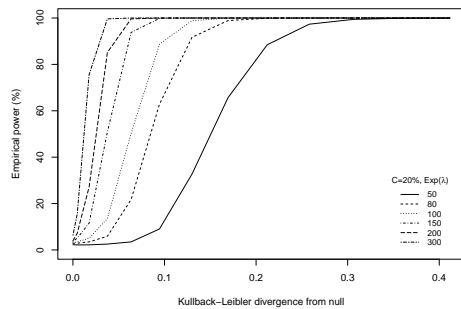
Αρχικά, παρατηρούμε ότι η ισχύς αυξάνει όσο αυξάνεται η απόκλιση από την H_0 , όπως επίσης και ότι η ισχύς αυξάνεται όσο αυξάνεται και το μέγεθος του δείγματος. Επιπλέον, παρατηρούμε ότι, η ισχύς αυξάνεται με μικρότερο ρυθμό



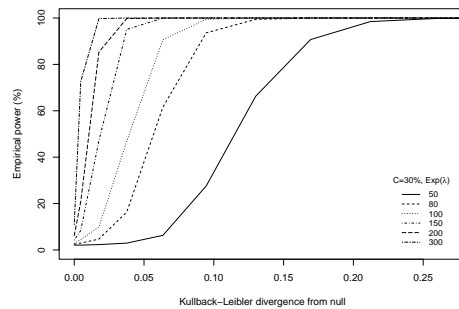
(α') C=0%



(β') C=10%



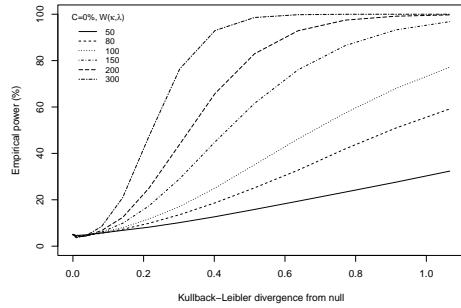
(γ') C=20%



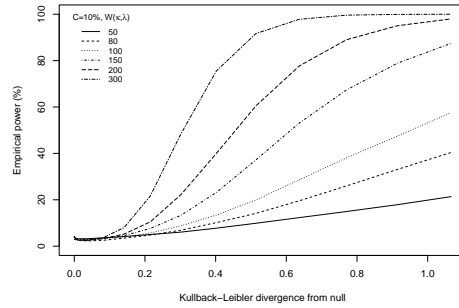
(δ') C=30%

Σχήμα 5.1: Ποσοστά απόρριψης του προτεινόμενου ελέγχου καλής προσαρμογής, επιπέδου σημαντικότητας $\alpha=5\%$, για έλεγχο της H_0 ($Exp(1)$) έναντι της H_1 ($Exp(\lambda)$) για διάφορα μεγέθη δείγματος και ποσοστά λογοκρισίας.

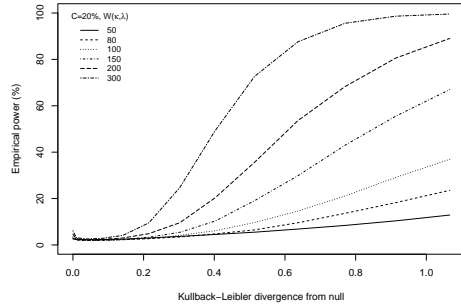
καθώς το ποσοστό λογοκρισίας αυξάνεται. Έπειτα ο έλεγχος είναι συνεπής ως προς το θεωρητικό επίπεδο σημαντικότητας (nominal level) για ποσοστά λογοκρισίας 0%, 10% και 20%, ενώ για ποσοστό λογοκρισίας 30% το θεωρητικό επίπεδο μπορεί να φτάσει ακόμη και τον διπλασιασμό για μέγεθος δείγματος ίσο με 300. Ωστόσο όμως με τη χρήση σημείων αποκοπής που να βασίζονται στην κατανομή του στατιστικού υπό την ύπαρξη λογοκρισίας, να μπορούσαμε να λάβουμε θεωρητικό επίπεδο σημαντικότητας και τιμές ισχύος πολύ κοντά σε αυτές που λήφθηκαν χωρίς την ύπαρξη λογοκρισίας. Τέλος, παρά τη λογοκρισία, το Σχήμα 5.2 υποδηλώνει ότι όσο μεγαλύτερη είναι η απόκλιση από τη μηδενική υπόθεση, τόσο πιο “ευαίσθητος” γίνεται ο έλεγχος στην ανίχνευση των εναλλακτικών συναρτήσεων κινδύνου.



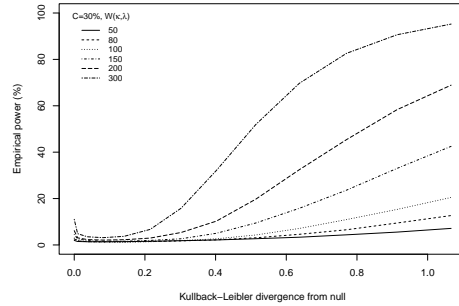
(α') C=0%



(β') C=10%



(γ') C=20%



(δ') C=30%

Σχήμα 5.2: Ποσοστά απόρριψης του προτεινόμενου ελέγχου καλής προσαρμογής, επιπέδου σημαντικότητας $\alpha=5\%$, για έλεγχο της $H_0 (W(1, 1))$ έναντι της $H_1 (W(\kappa, \lambda))$ για διάφορα μεγέθη δείγματος και ποσοστά λογοχρισίας.

5.2.3 Αποτελέσματα του ελέγχου για εναλλακτικές συναρτήσεις κινδύνου από την Γάμμα κατανομή

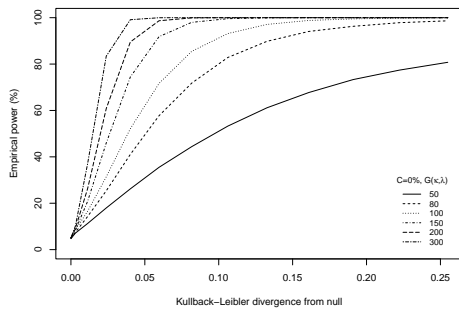
Για την πραγματοποίηση του ελέγχου της υπόθεσης:

$$H_0 : \lambda(x) = 1 \quad \text{έναντι} \quad H_1 : \lambda(x) = \frac{x^{k-1} e^{-\frac{x}{\lambda}}}{\lambda^k (\Gamma(\kappa) - \gamma(\kappa, \frac{x}{\lambda}))},$$

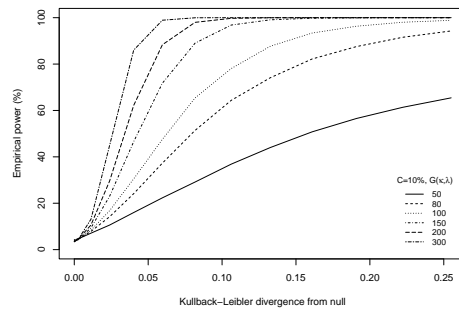
για $\kappa = \lambda = 1, 1.1, \dots, 2.2,$

τα αποτελέσματα τα οποία προέκυψαν δίνονται στο Σχήμα 5.3, το οποίο περιέχει την ισχύ του ελέγχου για τα διάφορα μεγέθη δείγματος και ποσοστών λογοχρισίας που αναφέρθηκαν στην αρχή της Ενότητας 5.2.

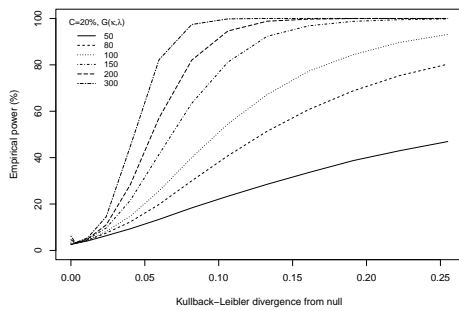
Ξεκινώντας την ερμηνεία των αποτελεσμάτων παρατηρούμε ότι, η ισχύς του ελέγχου αυξάνεται όσο αυξάνεται η απόκλιση από την H_0 και επιπλέον η ισχύς του ελέγχου αυξάνεται όσο αυξάνεται και το μέγεθος του δείγματος. Επιπλέον παρατηρούμε ότι ο έλεγχος είναι συνεπής ως προς το θεωρητικό επίπεδο σημαντικότητας (nominal level) για ποσοστά λογοκρισίας 0%, 10% και 20% , ενώ για ποσοστό λογοκρισίας 30% το θεωρητικό επίπεδο διπλασιάζεται όταν το μέγεθος του δείγματος είναι ίσο με 300. Συνεχίζοντας, βλέπουμε ότι, όσο αυξάνεται το ποσοστό λογοκρισίας, τόσο μικρότερος γίνεται ο ρυθμός με τον οποίο αυξάνεται η ισχύς. Ωστόσο όμως με τη χρήση σημείων αποκοπής που να βασίζονται στην κατανομή του στατιστικού υπό την ύπαρξη λογοκρισίας, να μπορέσουμε να λάβουμε θεωρητικό επίπεδο σημαντικότητας και τιμές ισχύος πολύ κοντά σε αυτές που λήφθηκαν χωρίς την ύπαρξη λογοκρισίας. Τέλος, παρά τη λογοκρισία, το Σχήμα 5.3 υποδηλώνει ότι όσο μεγαλύτερη είναι η απόκλιση από τη μηδενική υπόθεση, τόσο πιο “ευαίσθητος” γίνεται ο έλεγχος στην ανίχνευση των εναλλακτικών συναρτήσεων κινδύνου.



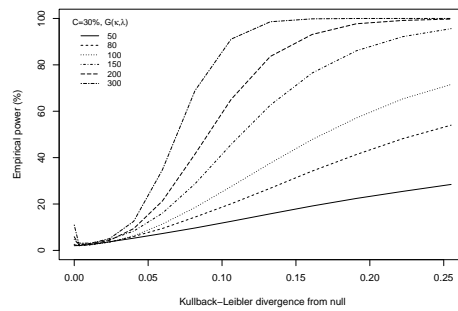
(α') C=0%



(β') C=10%



(γ') C=20%



(δ') C=30%

Σχήμα 5.3: Ποσοστά απόρριψης του προτεινόμενου ελέγχου καλής προσαρμογής, επιπέδου σημαντικότητας $\alpha=5\%$, για έλεγχο της $H_0 (G(1, 1))$ έναντι της $H_1 (G(\kappa, \lambda))$ για διάφορα μεγέθη δείγματος και ποσοστά λογοχρισίας.

ΕΠΙΛΟΓΟΣ - ΘΕΜΑ ΠΡΟΣ ΠΕΡΕΤΑΙΡΩ ΕΡΕΥΝΑ

Στον κεφάλαιο αυτό, θα παρουσιάσουμε ένα θέμα προς περεταίρω έρευνα. Ο έλεγχος καλής προσαρμογής για τη συνάρτηση του ρυθμού επικινδυνότητας που προτείνουμε στην παρούσα διατριβή βασίζεται στον έλεγχο της υπόθεσης:

$$H_0 : \lambda(x) = \lambda_0(x) \quad \text{έναντι} \quad H_1 : \lambda(x) \neq \lambda_0(x),$$

όπου $\lambda_0(x)$ μία πλήρως προσδιορισμένη συνάρτηση κινδύνου.

Η έρευνα αυτή μπορεί να επεκταθεί και στην κατασκευή ελέγχου καλής προσαρμογής, για τον έλεγχο της μηδενικής υπόθεσης ότι η πραγματική συνάρτηση κινδύνου προέρχεται από μία κλάση παραμετρικών συναρτήσεων κινδύνου της μορφής $\lambda(x; \theta)$ με $\theta \in \Theta$, έναντι της εναλλακτικής ότι δεν προέρχεται. Μοντελοποιώντας το παραπάνω λαμβάνουμε ότι η μηδενική και η εναλλακτική υπόθεση εκφράζονται ως εξής:

$$H_0 : \lambda(x) = \lambda(x; \theta) \quad \text{έναντι} \quad H_1 : \lambda(x) \neq \lambda(x; \theta)$$

Καθώς η παράμετρος θ είναι άγνωστη, εκτιμάται με τη μέθοδο της μέγιστης πιθανοφάνειας. Για τον λόγο αυτόν θέτουμε ως $\hat{\theta}$ τον εκτιμητή μέγιστης πιθανοφάνειας (ε.μ.π.) της θ .

Ο συγκεκριμένος έλεγχος μπορεί επίσης να βασιστεί στο ISE και κατά συνέπεια να ληφθεί η εξής στατιστική συνάρτηση ελέγχου:

$$T(h) = \int \left(\lambda(x) - \lambda(x; \hat{\theta}) \right) dx.$$

Καθώς όμως η πραγματική συνάρτηση κινδύνου $\lambda(x)$ είναι άγνωστη, εκτιμάται με τη μέθοδο των πυρήνων από τον εκτιμητή που δίνεται από την (3.28). Επομένως,

Κεφάλαιο 6

η προτεινόμενη ελεγχοσυνάρτηση λαμβάνει την μορφή:

$$\hat{T}(h) = \int \left(\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right) dx.$$

Για τη μελέτη της κατανομής της ελεγχοσυνάρτησης αποδεικνύεται αρχικά το παρακάτω λήμμα, υπό τις Υποθέσεις 1-4 που δόθηκαν στην Υποενότητα 4.1.2 (βλέπε σελ. 87-88).

Λήμμα 6.0.1. Έστω $n \rightarrow \infty$, $h \rightarrow 0$ και $nh \rightarrow \infty$, με βέλτιστο εύρος ζώνης $h \sim n^{-\frac{1}{5}}$. Τότε υπό τις Υποθέσεις 1-4 ισχύει ότι:

$$I_p = \int \left(\hat{\lambda}(x) - \lambda(x) \right)^2 dx + O_p(1), \quad (6.1)$$

όπου $\lambda(x)$ μία πλήρως προσδιορισμένη συνάρτηση κινδύνου.

Απόδειξη. Αρχικά ξεκινάμε με τη διάσπαση του

$$I_p = \int \left(\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right) dx,$$

η οποία πραγματοποιείται ως εξής:

$$\begin{aligned} I_p &= \int \left(\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right)^2 dx \\ &= \int \left(\hat{\lambda}(x) - E\hat{\lambda}(x) + E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right)^2 dx \\ &= \int \left(\hat{\lambda}(x) - E\hat{\lambda}(x) \right)^2 dx + \int \left(E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right)^2 dx \\ &\quad + 2 \int \left(\hat{\lambda}(x) - E\hat{\lambda}(x) \right) \left(E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right) dx. \end{aligned} \quad (6.2)$$

Παίρνοντας τη διαφορά των (6.2) και (4.3) λαμβάνουμε ότι:

$$\begin{aligned} I_p - I_n &= \int \left(E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right)^2 dx - \int \left(E\hat{\lambda}(x) - \lambda(x) \right)^2 dx \\ &\quad + 2 \int \left(\hat{\lambda}(x) - E\hat{\lambda}(x) \right) \left(E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) \right) dx \\ &\quad - 2 \int \left(\hat{\lambda}(x) - E\hat{\lambda}(x) \right) \left(E\hat{\lambda}(x) - \lambda(x) \right) dx. \end{aligned} \quad (6.3)$$

Κεφάλαιο 6

Χρησιμοποιώντας την (3.56) παίρνουμε ότι:

$$E\hat{\lambda}(x) - \lambda(x) = O(h^2). \quad (6.4)$$

Επιπλέον,

$$\begin{aligned} \hat{\lambda}(x) - E\hat{\lambda}(x) &= \sum_{i=1}^n \frac{K_h(x - X_{(i)}) \delta_{(i)}}{n - i + 1} - E \left(\sum_{i=1}^n \frac{K_h(x - X_{(i)}) \delta_{(i)}}{n - i + 1} \right) \\ &= \sum_{i=1}^n \left[\frac{K_h(x - X_{(i)}) \delta_{(i)}}{n - i + 1} - E \left(\frac{K_h(x - X_{(i)}) \delta_{(i)}}{n - i + 1} \right) \right] \\ &= O(1). \end{aligned} \quad (6.5)$$

Τέλος έχουμε ότι:

$$E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) = E\hat{\lambda}(x) - E\lambda(x; \hat{\theta}) + E\lambda(x; \hat{\theta}) - \lambda(x; \hat{\theta}),$$

από όπου παίρνοντας ότι:

$$E\lambda(x; \hat{\theta}) - \lambda(x; \hat{\theta}) = o_p((\sqrt{n})^{-1}),$$

λαμβάνουμε:

$$\begin{aligned} E\hat{\lambda}(x) - \lambda(x; \hat{\theta}) &= E\hat{\lambda}(x) - E\lambda(x; \hat{\theta}) + o_p((\sqrt{n})^{-1}) \\ &= E \left(\sum_{i=1}^n \frac{K_h(x - X_{(i)}) \delta_{(i)}}{n - i + 1} \right) - E \left(\frac{f(x; \hat{\theta})}{1 - F(x; \hat{\theta})} \right) \\ &\quad + o_p((\sqrt{n})^{-1}) \\ &= \sum_{i=1}^n E \left(\frac{K_h(x - X_{(i)}) \delta_{(i)}}{n - i + 1} \right) - E \left(\frac{f(x; \hat{\theta})}{1 - F(x; \hat{\theta})} \right) \\ &\quad + o_p((\sqrt{n})^{-1}) \\ &= O(1). \end{aligned}$$

Αντικαθιστώντας τις (6.4), (6.5) και (6.4) στην (6.3), λαμβάνουμε την (6.1) και ολοκληρώνεται η απόδειξη. \square

Κεφάλαιο 6

Μέσω του Λήμματος 6.0.1 αυτό που επιτυγχάνουμε είναι ότι, για να προσδιορίσουμε την κατανομή του I_p , και κατά συνέπεια του $\hat{T}(h)$, αρκεί να προσδιορίσουμε την κατανομή του

$$I_n = \int (\hat{\lambda}(x) - \lambda(x))^2 dx,$$

του οποίου ο χειρισμός είναι ευκολότερος. Η κατανομή του αποδείχθηκε αναλυτικά στο κεφάλαιο 4 της παρούσας διατριβής, επομένως μέσω προσομοιώσεων και της κατάλληλης μελέτης είναι εφικτή η επιβεβαίωση των αποτελεσμάτων που προκύπτουν από τα παραπάνω και για το λόγο αυτό προτείνεται ως θέμα για περαιτέρω έρευνα.

ΠΑΡΑΡΤΗΜΑ Α'

ΠΙΝΑΚΕΣ

Στο παράρτημα αυτό παραθέτουμε, με τη μορφή πινάκων, τις τιμές των αποκλίσεων Kullback-Leibler, τα σημεία αποκοπής και τις τιμές της εμπειρικής ισχύος της ελεγχουσυνάρτησης του προτεινόμενου έλεγχου καλής προσαρμογής για τη συνάρτηση κινδύνου. Πιο συγκεκριμένα, στην Ενότητα Α'.1 παρουσιάζονται οι αποκλίσεις Kullback-Leibler ανάμεσα στη συνάρτηση κινδύνου υπό τη μηδενική υπόθεση ($\lambda_0(x)$) και τις αντίστοιχες εναλλακτικές ($\lambda_1(x)$), στην Ενότητα Α'.2 παρουσιάζονται τα σημεία αποκοπής της ελεγχουσυνάρτησης υπό την μηδενική υπόθεση H_0 και στις Ενότητες Α'.3, Α'.4 και Α'.5 παρουσιάζονται οι πίνακες με την εμπειρική ισχύ (%) για καθεμία από τις διαφορετικές περιπτώσεις εναλλακτικών υποθέσεων H_1 (Εκθετική, Weibull και Γάμμα αντίστοιχα). Τα αποτελέσματα προέκυψαν με τη χρήση του λογισμικού της R και μέσω προσομοιώσεων.

Α'.1 Αποκλίσεις Kullback-Leibler

Τα αποτελέσματα που παρατίθενται στους πίνακες της παρούσας ενότητας προέκυψαν από τη σχέση (5.5) για $C = 10$ και για όλες τις περιπτώσεις εναλλακτικών υποθέσεων που αναφέρονται στην εισαγωγή του Παραρτήματος Α'.

$\lambda_0(x)$	$\lambda_1(x)$			
Exp(1)	Exp(1.1)	Exp(1.2)	Exp(1.3)	Exp(1.4)
	0.004689607	0.017677641	0.037634027	0.063524879
Exp(1)	Exp(1.5)	Exp(1.6)	Exp(1.7)	Exp(1.8)
	0.094530600	0.129990469	0.169364059	0.212203701
Exp(1)	Exp(1.9)	Exp(2)	Exp(2.1)	Exp(2.2)
	0.258134394	0.306838888	0.358046399	0.411523956

Πίνακας Α'.1: Αποκλίσεις Kullback-Leibler ανάμεσα στη $\lambda_0(x)$ ($Exp(1)$) και τις $\lambda_1(x)$ ($Exp(\lambda)$, $\lambda = 1.1, \dots, 2.2$).

$\lambda_0(x)$	$\lambda_1(x)$			
W(1,1)	W(1.1,1.1)	W(1.2,1.2)	W(1.3,1.3)	W(1.4,1.4)
	0.009577776	0.037193649	0.081402172	0.140989706
W(1,1)	W(1.5,1.5)	W(1.6,1.6)	W(1.7,1.7)	W(1.8,1.8)
	0.214909317	0.302238763	0.402152732	0.513903305
W(1,1)	W(1.9,1.9)	W(2,2)	W(2.1,2.1)	W(2.2,2.2)
	0.636807237	0.770235771	0.913607616	1.066383429

Πίνακας Α'.2: Αποκλίσεις Kullback-Leibler ανάμεσα στη $\lambda_0(x)$ ($W(1,1)$) και τις $\lambda_1(x)$ ($W(\kappa, \lambda)$, $\kappa = \lambda = 1.1, \dots, 2.2$).

$\lambda_0(x)$	$\lambda_1(x)$			
G(1,1)	G(1.1,1.1)	G(1.2,1.2)	G(1.3,1.3)	G(1.4,1.4)
	0.00300776	0.01128252	0.02391492	0.04020978
G(1,1)	G(1.5,1.5)	G(1.6,1.6)	G(1.7,1.7)	G(1.8,1.8)
	0.05962427	0.08172678	0.10616865	0.13266438
G(1,1)	G(1.9,1.9)	G(2,2)	G(2.1,2.1)	G(2.2,2.2)
	0.16097737	0.19090931	0.22229254	0.25498395

Πίνακας Α'.3: Αποκλίσεις Kullback-Leibler ανάμεσα στη $\lambda_0(x)$ ($G(1,1)$) και τις $\lambda_1(x)$ ($G(\kappa, \lambda)$, $\kappa = \lambda = 1.1, \dots, 2.2$).

Α'.2 Σημεία αποκοπής

Στην ενότητα αυτή παρατίθενται τα σημεία αποκοπής της ελεγχουσυνάρτησης υπό την μηδενική υπόθεση (H_0), όπως αυτά προέκυψαν για καθένα από τους ελέγχους που πραγματοποιήθηκαν.

n	50	80	100
cut-off	0.7991653	0.5347137	0.4427233
n	150	200	300
cut-off	0.3274572	0.2682844	0.2017026

Πίνακας Α'.4: Σημεία αποκοπής της ελεγχουσυνάρτησης υπό την $H_0 : \lambda(x) = 1$ ($Exp(1)$).

n	50	80	100
cut-off	0.8039226	0.5357003	0.4425589
n	150	200	300
cut-off	0.3278811	0.2662460	0.2017746

Πίνακας Α'.5: Σημεία αποκοπής της ελεγχοσυνάρτησης υπό την $H_0 : \lambda(x) = 1$ ($W(1, 1)$).

n	50	80	100
cut-off	0.7981444	0.5326710	0.4430020
n	150	200	300
cut-off	0.3271491	0.2670478	0.2038451

Πίνακας Α'.6: Σημεία αποκοπής της ελεγχοσυνάρτησης υπό την $H_0 : \lambda(x) = 1$ ($G(1, 1)$).

Παρατήρηση Α'.2.1. Παρά το γεγονός ότι οι κατανομές $Exp(1)$, $W(1, 1)$ και $G(1, 1)$ είναι ισοδύναμες, παρατηρούμε ότι σε κάθε περίπτωση έχουμε διαφορετικά σημεία αποκοπής. Αυτό οφείλεται στο ότι για κάθε έλεγχο που πραγματοποιήθηκε, λήφθηκαν εκ νέου σημεία αποκοπής, στον κώδικα των οποίων γίνεται χρήση και των εντολών **rexp**, **rweibull** και **rgamma**. Όπως είναι γνωστό οι εντολές αυτές αποτελούν γεννήτριες τυχαίων αριθμών από την εκάστοτε κατανομή και για τον λόγο αυτόν, όπως επίσης και για τον λόγο ότι σε κάθε έλεγχο χρησιμοποιήθηκε διαφορετική εντολή **set.seed** (ορίζει την αρχική τιμή της γεννήτριας τυχαίων αριθμών της R σε συγκεκριμένη τιμή), είναι λογικό να λαμβάνονται ελαφρώς διαφορετικά σημεία αποκοπής σε κάθε περίπτωση.

Α'.3 Ισχύς για εναλλακτικές από την Εκθετική κατανομή

n	Exp(1)	Exp(1.1)	Exp(1.2)	Exp(1.3)	Exp(1.4)
50	4.92	3.94	3.38	3.19	3.12
80	5.15	4.20	3.73	3.84	4.77
100	5.03	4.26	4.10	4.61	6.93
150	4.77	4.05	4.28	6.21	18.81
200	4.83	4.26	4.99	9.87	39.30
300	4.82	4.68	6.90	24.13	79.39
n	Exp(1.5)	Exp(1.6)	Exp(1.7)	Exp(1.8)	Exp(1.9)
50	3.48	5.52	15.14	38.10	64.47
80	9.38	29.75	63.99	88.22	97.37
100	21.83	57.03	87.12	97.75	99.64
150	58.14	91.57	99.37	99.99	100.00
200	85.65	99.15	99.98	100.00	100.00
300	99.29	100.00	100.00	100.00	100.00
n	Exp(2)	Exp(2.1)	Exp(2.2)		
50	83.98	94.69	98.42		
80	99.54	99.92	100.00		
100	99.98	100.00	100.00		
150	100.00	100.00	100.00		
200	100.00	100.00	100.00		
300	100.00	100.00	100.00		

Πίνακας Α'.7: Εμπειρική ισχύς (%) της ελεγχουσυνάρτησης για $a = 5\%$ (Exp(λ)) με 0% λογοκρισία.

n	Exp(1)	Exp(1.1)	Exp(1.2)	Exp(1.3)	Exp(1.4)
50	3.84	3.32	3.05	3.01	3.24
80	3.66	3.36	3.49	4.26	7.78
100	3.85	3.61	4.01	5.88	17.90
150	4.13	4.25	6.12	16.08	56.16
200	4.23	4.80	8.74	33.38	85.74
300	4.86	6.99	19.18	77.74	99.44
n	Exp(1.5)	Exp(1.6)	Exp(1.7)	Exp(1.8)	Exp(1.9)
50	4.37	12.02	35.58	65.80	86.24
80	26.95	63.39	89.41	98.26	99.88
100	54.82	87.54	98.11	99.91	99.98
150	92.86	99.56	99.99	100.00	100.00
200	99.49	100.00	100.00	100.00	100.00
300	100.00	100.00	100.00	100.00	100.00
n	Exp(2)	Exp(2.1)	Exp(2.2)		
50	95.81	98.91	99.75		
80	99.99	99.99	100.00		
100	100.00	100.00	100.00		
150	100.00	100.00	100.00		
200	100.00	100.00	100.00		
300	100.00	100.00	100.00		

Πίνακας Α'.8: Εμπειρική ισχύς (%) της ελεγχοσυνάρτησης για $\alpha = 5\%$ (Exp(λ)) με 10% λογοκρισία.

n	Exp(1)	Exp(1.1)	Exp(1.2)	Exp(1.3)	Exp(1.4)
50	2.25	2.17	2.20	2.53	3.46
80	2.86	2.92	3.47	5.90	21.71
100	3.10	3.52	5.05	13.53	50.25
150	3.74	5.07	11.56	50.84	93.77
200	3.94	7.21	27.17	85.16	99.62
300	6.36	15.41	75.61	99.62	100.00
n	Exp(1.5)	Exp(1.6)	Exp(1.7)	Exp(1.8)	Exp(1.9)
50	9.05	32.68	65.66	88.47	97.34
80	62.85	91.68	98.90	99.92	99.99
100	88.70	98.86	99.94	100.00	100.00
150	99.84	100.00	100.00	100.00	100.00
200	100.00	100.00	100.00	100.00	100.00
300	100.00	100.00	100.00	100.00	100.00
n	Exp(2)	Exp(2.1)	Exp(2.2)		
50	99.43	99.89	99.97		
80	100.00	100.00	100.00		
100	100.00	100.00	100.00		
150	100.00	100.00	100.00		
200	100.00	100.00	100.00		
300	100.00	100.00	100.00		

Πίνακας Α'.9: Εμπειρική ισχύς (%) της ελεγχουσυνάρτησης για $\alpha = 5\%$ (Exp(λ)) με 20% λογοκρισία.

n	Exp(1)	Exp(1.1)	Exp(1.2)	Exp(1.3)	Exp(1.4)
50	2.10	2.08	2.31	2.94	6.29
80	2.50	2.96	4.67	16.37	61.58
100	2.89	4.42	9.90	46.89	90.66
150	4.28	8.42	47.07	95.08	99.93
200	6.18	20.91	85.33	99.76	100.00
300	10.95	72.67	99.74	100.00	100.00
n	Exp(1.5)	Exp(1.6)	Exp(1.7)	Exp(1.8)	Exp(1.9)
50	27.57	66.26	90.72	98.46	99.77
80	93.59	99.44	99.98	100.00	100.00
100	99.43	99.99	100.00	100.00	100.00
150	100.00	100.00	100.00	100.00	100.00
200	100.00	100.00	100.00	100.00	100.00
300	100.00	100.00	100.00	100.00	100.00
n	Exp(2)	Exp(2.1)	Exp(2.2)		
50	99.98	100.00	100.00		
80	100.00	100.00	100.00		
100	100.00	100.00	100.00		
150	100.00	100.00	100.00		
200	100.00	100.00	100.00		
300	100.00	100.00	100.00		

Πίνακας Α'.10: Εμπειρική ισχύς (%) της ελεγχοσυνάρτησης για $a = 5\%$ ($\text{Exp}(\lambda)$) με 30% λογοκρισία.

Α'.4 Ισχύς για εναλλακτικές από την Weibull κατανομή

n	W(1,1)	W(1.1,1.1)	W(1.2,1.2)	W(1.3,1.3)
50	4.99	4.62	4.87	5.67
80	5.09	4.45	4.61	5.59
100	5.10	4.45	4.76	5.97
150	5.07	4.10	4.32	6.08
200	5.08	3.97	4.43	6.74
300	4.81	3.64	4.53	8.50
n	W(1.4,1.4)	W(1.5,1.5)	W(1.6,1.6)	W(1.7,1.7)
50	6.82	8.18	10.16	12.70
80	7.28	9.86	13.58	18.67
100	8.02	11.72	17.09	25.06
150	9.88	17.31	29.11	44.97
200	12.42	24.95	43.97	65.82
300	20.97	47.30	76.43	92.97
n	W(1.8,1.8)	W(1.9,1.9)	W(2,2)	W(2.1,2.1)
50	15.83	19.39	23.31	27.58
80	25.16	32.75	41.92	50.92
100	35.11	46.22	57.40	67.98
150	61.57	76.01	86.45	93.24
200	82.95	92.86	97.45	99.15
300	98.59	99.80	99.99	100.00
n	W(2.2,2.2)			
50	32.36			
80	59.18			
100	77.17			
150	96.80			
200	99.72			
300	100.00			

Πίνακας Α'.11: Εμπειρική ισχύς (%) της ελεγχοσυνάρτησης για $a = 5\%$ ($W(k,\lambda)$), με $k = \lambda$ με 0% λογοκρισία.

n	W(1,1)	W(1.1,1.1)	W(1.2,1.2)	W(1.3,1.3)
50	3.66	3.17	3.21	3.56
80	3.00	2.46	2.24	2.52
100	3.88	2.95	2.88	3.19
150	3.95	2.96	2.89	3.37
200	3.94	2.65	2.47	3.22
300	4.22	2.53	2.58	3.63
n	W(1.4,1.4)	W(1.5,1.5)	W(1.6,1.6)	W(1.7,1.7)
50	4.14	5.01	6.07	7.76
80	3.46	4.79	6.90	10.16
100	4.01	5.67	8.80	13.43
150	4.57	7.63	13.33	23.18
200	5.21	10.45	22.28	40.00
300	8.01	21.58	48.52	75.56
n	W(1.8,1.8)	W(1.9,1.9)	W(2,2)	W(2.1,2.1)
50	9.87	12.31	14.94	17.87
80	14.12	19.57	25.97	33.05
100	19.93	28.64	38.07	47.45
150	37.18	53.10	67.35	78.83
200	60.47	77.70	89.00	95.02
300	91.67	97.82	99.60	99.93
n	W(2.2,2.2)			
50	21.36			
80	40.40			
100	57.55			
150	87.46			
200	97.98			
300	100.00			

Πίνακας Α'.12: Εμπειρική ισχύς (%) της ελεγχοσυνάρτησης για $a = 5\%$ ($W(k,\lambda)$), με $k = \lambda$ με 10% λογοχρυσία.

n	W(1,1)	W(1.1,1.1)	W(1.2,1.2)	W(1.3,1.3)
50	2.62	2.32	2.22	2.31
80	2.85	2.07	1.90	1.99
100	3.13	2.26	1.89	1.90
150	3.70	2.47	2.04	2.00
200	4.66	2.70	2.32	2.36
300	6.15	3.15	2.58	2.82
n	W(1.4,1.4)	W(1.5,1.5)	W(1.6,1.6)	W(1.7,1.7)
50	2.49	3.00	3.68	4.52
80	2.28	2.70	3.44	4.78
100	2.18	2.73	4.12	6.15
150	2.38	3.33	5.45	10.27
200	2.99	4.87	9.57	20.29
300	4.10	9.52	24.69	49.20
n	W(1.8,1.8)	W(1.9,1.9)	W(2,2)	W(2.1,2.1)
50	5.50	6.85	8.42	10.39
80	6.49	9.58	13.63	18.25
100	9.66	14.61	21.14	29.14
150	19.08	29.87	43.01	55.53
200	35.63	53.62	68.31	80.59
300	72.66	87.62	95.58	98.64
n	W(2.2,2.2)			
50	12.89			
80	23.55			
100	36.93			
150	67.00			
200	89.04			
300	99.57			

Πίνακας Α'.13: Εμπειρική ισχύς (%) της ελεγχοσυνάρτησης για $a = 5\%$ ($W(k,\lambda)$), με $k = \lambda$ με 20% λογοχρυσία.

n	W(1,1)	W(1.1,1.1)	W(1.2,1.2)	W(1.3,1.3)
50	1.95	1.49	1.38	1.38
80	2.59	1.73	1.41	1.32
100	2.99	1.87	1.37	1.24
150	4.50	2.61	1.88	1.61
200	6.16	3.06	2.31	2.09
300	11.04	4.86	3.52	3.14
n	W(1.4,1.4)	W(1.5,1.5)	W(1.6,1.6)	W(1.7,1.7)
50	1.42	1.58	1.83	2.15
80	1.37	1.49	1.75	2.20
100	1.17	1.37	1.89	2.65
150	1.61	1.95	2.66	5.03
200	2.25	2.96	5.36	10.26
300	3.71	6.71	15.82	32.11
n	W(1.8,1.8)	W(1.9,1.9)	W(2,2)	W(2.1,2.1)
50	2.69	3.33	4.36	5.54
80	3.08	4.59	6.54	9.46
100	4.30	7.11	10.93	15.32
150	9.47	15.79	23.61	33.02
200	19.75	32.43	45.39	58.34
300	51.95	69.67	82.60	90.63
n	W(2.2,2.2)			
50	7.14			
80	12.67			
100	20.48			
150	42.51			
200	68.89			
300	95.23			

Πίνακας Α'.14: Εμπειρική ισχύς (%) της ελεγχοσυνάρτησης για $a = 5\%$ ($W(k,\lambda)$), με $k = \lambda$ με 30% λογοχρισία.

Α'.5 Ισχύς για εναλλακτικές από την Γάμμα κατανομή

n	G(1,1)	G(1.1,1.1)	G(1.2,1.2)	G(1.3,1.3)
50	4.72	7.11	11.26	17.97
80	4.90	7.43	14.27	25.37
100	5.13	8.52	16.51	31.40
150	4.98	8.34	20.80	46.00
200	5.11	9.09	26.27	60.83
300	5.02	10.14	37.01	83.43
n	G(1.4,1.4)	G(1.5,1.5)	G(1.6,1.6)	G(1.7,1.7)
50	26.21	35.49	44.40	53.26
80	41.07	57.74	71.79	82.84
100	52.17	71.67	85.50	93.16
150	74.45	91.83	97.99	99.58
200	89.54	98.68	99.91	99.99
300	99.13	99.98	100.00	100.00
n	G(1.8,1.8)	G(1.9,1.9)	G(2,2)	G(2.1,2.1)
50	61.16	67.74	73.31	77.40
80	89.82	94.07	96.28	97.85
100	97.11	98.81	99.48	99.83
150	99.93	99.98	100.00	100.00
200	100.00	100.00	100.00	100.00
300	100.00	100.00	100.00	100.00
n	G(2.2,2.2)			
50	80.76			
80	98.65			
100	99.91			
150	100.00			
200	100.00			
300	100.00			

Πίνακας Α'.15: Εμπειρική ισχύς (%) της ελεγχουσυνάρτησης για $\alpha = 5\%$ ($G(k,\lambda)$), με $k = \lambda$ με 0% λογοκρισία.

n	G(1,1)	G(1.1,1.1)	G(1.2,1.2)	G(1.3,1.3)
50	3.46	4.73	7.10	10.59
80	3.38	4.53	7.84	14.18
100	3.43	4.59	8.38	16.79
150	3.58	4.35	9.38	23.01
200	3.93	4.76	10.60	29.60
300	4.23	4.75	13.09	44.95
n	G(1.4,1.4)	G(1.5,1.5)	G(1.6,1.6)	G(1.7,1.7)
50	15.99	22.33	29.14	36.85
80	24.13	37.30	51.40	64.35
100	30.52	47.56	65.37	78.06
150	46.57	71.76	89.02	96.82
200	62.19	88.34	97.99	99.72
300	86.02	98.91	99.96	100.00
n	G(1.8,1.8)	G(1.9,1.9)	G(2,2)	G(2.1,2.1)
50	44.02	50.79	56.54	61.37
80	74.12	82.21	87.54	91.56
100	87.70	93.39	96.32	98.04
150	99.08	99.78	99.92	99.98
200	99.96	100.00	100.00	100.00
300	100.00	100.00	100.00	100.00
n	G(2.2,2.2)			
50	65.44			
80	94.29			
100	98.86			
150	100.00			
200	100.00			
300	100.00			

Πίνακας Α'.16: Εμπειρική ισχύς (%) της ελεγχουσυνάρτησης για $a = 5\%$ ($G(k,\lambda)$), με $k = \lambda$ με 10% λογοχρυσία.

n	G(1,1)	G(1.1,1.1)	G(1.2,1.2)	G(1.3,1.3)
50	2.48	3.01	4.10	6.34
80	2.66	2.97	4.37	7.48
100	3.02	3.08	4.46	8.18
150	3.57	3.06	4.77	9.68
200	4.61	3.31	4.93	11.03
300	6.18	3.36	5.38	14.58
n	G(1.4,1.4)	G(1.5,1.5)	G(1.6,1.6)	G(1.7,1.7)
50	9.32	13.37	18.33	23.35
80	12.31	19.81	30.00	40.77
100	14.78	25.60	40.02	54.49
150	21.59	41.41	63.40	81.17
200	28.38	57.12	82.03	94.58
300	44.92	82.23	97.41	99.81
n	G(1.8,1.8)	G(1.9,1.9)	G(2,2)	G(2.1,2.1)
50	28.49	33.59	38.71	42.97
80	51.34	60.77	68.77	75.35
100	67.14	77.32	84.36	89.66
150	92.28	96.85	98.76	99.47
200	98.81	99.71	99.97	99.99
300	100.00	100.00	100.00	100.00
n	G(2.2,2.2)			
50	46.93			
80	80.30			
100	93.11			
150	99.81			
200	100.00			
300	100.00			

Πίνακας Α'.17: Εμπειρική ισχύς (%) της ελεγχουσυνάρτησης για $a = 5\%$ ($G(k,\lambda)$), με $k = \lambda$ με 20% λογοχρισία.

n	G(1,1)	G(1.1,1.1)	G(1.2,1.2)	G(1.3,1.3)
50	2.04	2.10	2.59	3.71
80	2.44	2.05	2.44	3.62
100	2.64	1.97	2.38	3.55
150	4.73	2.61	3.02	4.67
200	5.89	2.49	2.72	4.32
300	10.95	3.14	3.11	5.08
n	G(1.4,1.4)	G(1.5,1.5)	G(1.6,1.6)	G(1.7,1.7)
50	5.27	7.27	9.66	12.54
80	6.00	9.44	14.36	20.10
100	6.28	11.35	18.49	27.54
150	8.31	16.01	28.71	45.69
200	9.38	21.27	41.55	65.12
300	12.51	34.71	68.91	91.10
n	G(1.8,1.8)	G(1.9,1.9)	G(2,2)	G(2.1,2.1)
50	15.77	19.15	22.41	25.45
80	26.80	34.20	41.34	48.18
100	37.61	47.82	57.10	65.33
150	62.64	76.48	86.08	92.16
200	83.63	93.11	97.72	99.15
300	98.58	99.84	99.99	100.00
n	G(2.2,2.2)			
50	28.44			
80	54.00			
100	71.53			
150	95.61			
200	99.75			
300	100.00			

Πίνακας Α'.18: Εμπειρική ισχύς (%) της ελεγχουσυνάρτησης για $\alpha = 5\%$ ($G(k,\lambda)$), με $k = \lambda$ με 30% λογοχρυσία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Akritas, M. G. (1988). Pearson-type goodness-of-fit tests: the univariate case, *Journal of the American Statistical Association* **83**(401): 222–230.
- Bagdonavičius, V. B., Levulienė, R. J. and Nikulin, M. S. (2013). Chi-squared goodness-of-fit tests for parametric accelerated failure time models, *Communications in Statistics-Theory and Methods* **42**(15): 2768–2785.
- Bagdonavicius, V., Kruopis, J. and Nikulin, M. (2013). *Nonparametric tests for Censored Data*, Wiley-ISTE.
- Bagdonavičius, V., Levulienė, R. and Nikulin, M. (2010). Goodness-of-fit criteria for the cox model from left truncated and right censored data., *Journal of Mathematical Sciences* **167**(4): 436–443.
- Bagdonavicius, V. and Nikulin, M. (2011a). Chi-squared goodness-of-fit test for right censored data, *The International Journal of Applied Mathematics and Statistics* **24**(1): 1–11.
- Bagdonavičius, V. and Nikulin, M. (2011b). Chi-squared tests for general composite hypotheses from censored samples, *Comptes Rendus Mathematique* **349**(3-4): 219–223.
- Bagkavos, D. (2003). *Bias reduction in nonparametric hazard rate estimation*, PhD thesis, University of Birmingham.
- Bagkavos, D. (2011). Local linear hazard rate estimation and bandwidth selection, *Annals of the Institute of Statistical Mathematics* **63**(5): 1019–1046.

- Bagkavos, D., Ioannides, D. and Kalamatianou, A. (2013). A goodness of fit test for the survival function under random right censoring, *Electronic Journal of Statistics* **7**: 2550–2576.
- Bagkavos, D. and Patil, P. N. (2008). Local polynomial fitting in failure rate estimation, *IEEE Transactions on Reliability* **57**(1): 41–52.
- Bain, L. (1991). *Statistical analysis of reliability and life-testing models: theory and methods, Second Edition*, Routledge.
- Bain, L. J. (1974). Analysis for the linear failure-rate life-testing distribution, *Technometrics* **16**(4): 551–559.
- Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability*, Society for Industrial and Applied Mathematics.
- Barlow, R. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*, Holt, Reinhart and Winston, New York.
- Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *The Annals of Statistics* **1**(6): 1071–1095.
- Billingsley, P. (1995). *Measure and Probability*, John Wiley & Sons: New York.
- Blum, J. and Susarla, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data, *Multivariate Analysis* **5**: 213–222.
- Burke, M. D. (1983). Approximations of some hazard rate estimators in a competing risks model, *Stochastic Processes and Their Applications* **14**(2): 157–174.
- Cox, D. and Oakes, D. (1984). *Analysis of survival data*, Chapman & Hall.
- Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring, *Journal of Multivariate Analysis* **25**(2): 299–310.
- Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method, *Econometric Theory* **10**(2): 316–356.
- Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters, *Econometric Theory* **14**(5): 604–621.

- Glaser, R. E. (1980). Bathtub and related failure rate characterizations, *Journal of the American Statistical Association* **75**(371): 667–672.
- Gore, A., Paranjape, S., Rajarshi, M. and Gadgil, M. (1986). Some methods for summarizing survivorship data in nonstandard situations, *Biometrical journal* **28**(5): 577–586.
- Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives, *The Annals of Mathematical Statistics* **39**(2): 325–346.
- Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis* **14**(1): 1–16.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*, Academic Press.
- Hjort, N. L. (1990). Nonparametric bayes estimators based on beta processes in models for life history data, *The Annals of Statistics* **18**(3): 1259–1294.
- Hjorth, U. (1980). A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates, *Technometrics* **22**(1): 99–107.
- Hua, H., Bagkavos, D. and Patil, P. N. (2018). An L1 analysis of a kernel-based hazard rate estimator, *Australian & New Zealand Journal of Statistics* **60**(1): 43–64.
- Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages, *The American Statistician* **50**(4): 361–365.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**(282): 457–481.
- Kim, C., Bae, W., Choi, H. and Park, B. U. (2005). Non-parametric hazard function estimation using the Kaplan-Meier estimator, *Journal of Nonparametric Statistics* **17**(8): 937–948.
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Vol. 1230, Springer.

- Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons.
- Marron, J. and Padgett, W. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples, *The Annals of Statistics* **15**(4): 1520–1535.
- Park, S. and Shin, M. (2014). Kullback-leibler information of a censored variable and its applications, *Statistics* **48**(4): 756–765.
- Parzen, E. (1962). On estimation of a probability density function and mode, *The Annals of Mathematical Statistics* **33**(3): 1065–1076.
- Patil, P., Wells, M. and Marron, J. (1994). Some heuristics of kernel based estimators of ratio functions, *Journal of Nonparametric Statistics* **4**(2): 203–209.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press.
- Petrov, V. V. (1995). *Limit Theorems of Probability Theory*, Oxford Science Publications.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions, *The Annals of Statistics* **11**(2): 453–466.
- Ramos, P. L. (2018). *Bayesian and Classical Inference for the Generalized Gamma Distribution and Related Models*, PhD thesis, Universidade Federal de São Carlos.
- Rinne, H. (2014). *The Hazard Rate : Theory and Inference (With supplementary MATLAB-Programs)*, Justus-Liebig-University. <http://geb.uni-giessen.de/geb/volltexte/2014/10793>.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics* **27**(3): 832–837.
- Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence, *The Annals of Statistics* **3**(1): 1–14.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons.

- Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Vol. 26, Chapman & Hall.
- Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method, *The Annals of Statistics* **11**(3): 989–993.
- Turkson, A. J., Ayiah-Mensah, F. and Nimoh, V. (2021). Handling censoring and censored data in survival analysis: A standalone systematic literature review, *International Journal of Mathematics and Mathematical Sciences* **2021**: 1–16.
- Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data, *Journal of the American Statistical Association* **69**(345): 169–173.
- Van der Vaart, A. W. (2000). *Asymptotic Statistics*, Vol. 3, Cambridge University Press.
- Vital, R.-A. and Patil, P. (2020). Goodness-of-fit test for hazard rate, *Journal of Nonparametric Statistics* **32**(2): 403–427.
- Watson, G. and Leadbetter, M. (1964a). Hazard analysis I, *Biometrika* **51**(1/2): 175–184.
- Watson, G. S. and Leadbetter, M. (1964b). Hazard analysis II, *Sankhyā: The Indian Journal of Statistics, Series A* **26**(1): 101–116.
- Weisstein, E. W. (2006). Simpson’s Rule. From MathWorld—A Wolfram Web Resource, <https://mathworld.wolfram.com/SimpsonsRule.html>.
- Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data, *The Annals of Statistics* **11**(4): 1119–1135.