ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ & ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΕΚΤΙΜΗΣΗ ΠΟΣΟΣΤΗΜΟΡΙΩΝ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ ΜΕ ΜΕΘΟΔΟΥΣ BOOTSTRAP

Ζυγούρου Αφροδίτη

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής

Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων

για την απόκτηση του Μεταπτυχιακού Διπλώματος στην

Εφαρμοσμένη Στατιστική

Πειραιάς

Ιούλιος 2013

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ & ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΕΚΤΙΜΗΣΗ ΠΟΣΟΣΤΗΜΟΡΙΩΝ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ ΜΕ ΜΕΘΟΔΟΥΣ BOOTSTRAP

Ζυγούρου Αφροδίτη

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής

Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων

για την απόκτηση του Μεταπτυχιακού Διπλώματος στην

Εφαρμοσμένη Στατιστική

Πειραιάς

Ιούλιος 2013

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

Τα μέλη της Επιτροπής ήταν:

-……………………………….. (επιβλέπων)

-………………………………..

-………………………………..

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

UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN APPLIED STATISTICS

ESTIMATION OF THE PERCENTILES OF A DISTRIBUTION WITH BOOTSTRAP METHODS

By

Zygourou Afroditi

MSc Dissertation

submitted to the Department of Statistics and Insurance

Science of the University of Piraeus in partial fulfilment of

the requirements for the degree of Master of Applied Statistics

Piraeus

July 2013

This thesis was approved unanimously by the three-member committee appointed by the Department of Statistics and insurance science, University of Piraeus, in accordance with the rules of the MsC program in Applied Statistics

Committee members were:

-……………………………….. (supervisor)

-………………………………..

-………………………………..

Approval of this Thesis from the Department of Statistics and Insurance Science, university of Piraeus does not imply any endorsement of the opinions of the author.

Ευχαριστίες

Θερμές ευχαριστίες στον επιβλέποντα καθηγητή της διπλωματικής μου, κύριο Πολίτη Κωνσταντίνο για τη βοήθεια που μου παρείχε και τον χρόνο που μου διέθεσε, καθώς και στους κυρίους Χ. Ευαγγελάρα και Μ. Μπούτσικα για τη συμμετοχή τους στην εξεταστική επιτροπή. Ευχαριστώ θερμά και τους γονείς μου, που πάντα με στηρίζουν στις σπουδές μου, την αδερφή μου, τους φίλους μου και συμφοιτητές μου, τόσο για την συμπαράσταση, όσο και για τις χρήσιμες υποδείξεις τους.

Περίληψη

Παρόλο που οι πρώτες τεχνικές επαναδειγματοληψίας είχαν εμφανιστεί πριν από το 1950, η μέθοδος bootstrap εμφανίστηκε λίγο πριν το 1980. Σήμερα, με την ταυτόχρονη ανάπτυξη της τεχνολογίας η μέθοδος bootstrap μπορεί να χρησιμοποιηθεί, έναντι παραδοσιακών τεχνικών, για την εκτίμηση διάφορων παραμέτρων σε πολύπλοκα προβλήματα στατιστικής. Αρχικά παρουσιάζουμε την μέθοδο και κάποιο θεωρητικό υπόβαθρο χρήσιμο για την κατανόησή της. Στη συνέχεια, εφαρμόζουμε τη μέθοδο στην εκτίμηση των ποσοστημορίων, κάνοντας σημειακές εκτιμήσεις και κατασκευάζοντας διαστήματα εμπιστοσύνης. Στόχος είναι να ερευνήσουμε αν τα αποτελέσματα διαφέρουν ανάλογα με την κατανομή από την οποία παίρνουμε το δείγμα μας ή ανάλογα με το ποσοστιαίο σημείο που επιλέγουμε να εκτιμήσουμε.

Abstract

Despite the fact that subsampling techniques had appeared before 1950, bootstrap appeared right before 1980. Today, exploiting the enormous technological progress, bootstrap may be used instead of traditional methods, for the estimation of several parameters, even in complicated statistical problems. Initially, we present the general method and some useful theoretical background. Next, we apply the method in the estimation of percentiles, making point estimations and constructing confidence intervals. Our aim is to review basic results and check if and how they differ, regarding the distribution we choose to take our sample from, or the percentile we choose to estimate.

Περιεχόμενα

Κεφάλαιο 0

Στατιστική είναι η επιστήμη που ασχολείται με τη συλλογή δεδομένων , την επεξεργασία αυτών, και στη συνέχεια με την ανάλυσή τους, με σκοπό την εξαγωγή συμπερασμάτων. Αρχικά η στατιστική είχε μόνο περιγραφική μορφή , αλλά μετά τον 20^{o} αιώνα απέκτησε και μαθηματική- αναλυτική μορφή. Κυριότερο κομμάτι της μαθηματικής στατιστικής είναι η Στατιστική Συμπερασματολογία. Η στατιστική συμπερασματολογία περιλαμβάνει την σημειακή εκτιμητική, τα διαστήματα εμπιστοσύνης και τους ελέγχους υποθέσεων. Στη σημειακή εκτιμητική επιχειρούμε να εκτιμήσουμε μια παράμετρο κάποιας κατανομής, χρησιμοποιώντας στατιστικές συναρτήσεις, των οποίων η τιμή προσεγγίζει την πραγματική τιμή της άγνωστης παραμέτρου. Αντίστοιχα τα διαστήματα εμπιστοσύνης δίνουν μια περιοχή, μέσα στην οποία υπάρχει μεγάλη πιθανότητα να βρίσκεται η άγνωστη τιμή της παραμέτρου και τέλος στους ελέγχους υποθέσεων ελέγχουμε αν κάποιος ισχυρισμός που αφορά μια παράμετρο της κατανομής αληθεύει ή όχι. Ωστόσο ακόμα και αν συλλέξουμε τα δεδομένα μας και προχωρήσουμε στην ανάλυση μας, παραμένει το ερώτημα πόσο ακριβή είναι τα συμπεράσματα μας. Μέτρα ακρίβειας, όπως η μεροληψία, το μέσο τετραγωνικό σφάλμα και η διασπορά δείχνουν εάν οι εκτιμήσεις μας είναι ακριβείς και μπορούμε να προχωρήσουμε στην εξαγωγή συμπερασμάτων.

Η στατιστική συμπερασματολογία χρησιμοποιεί ένα δείγμα από τον πληθυσμό και εξάγει συμπεράσματα για ολόκληρο τον πληθυσμό. Το πιθανοθεωρητικό μοντέλο του πληθυσμού είναι άγνωστο, είτε γιατί αυτό έχει μια γνωστή συναρτησιακή μορφή, αλλά εξαρτάται από άγνωστες παραμέτρους (παραμετρική στατιστική), είτε γιατί η συναρτησιακή του μορφή είναι εντελώς άγνωστη (μη- παραμετρική στατιστική). Στην κλασσική στατιστική εμφανίζονται διάφορα εμπόδια. Συχνά το μέγεθος του δείγματος πρέπει να είναι μεγάλο (μεγαλύτερο από 30 παρατηρήσεις) για να χρησιμοποιήσουμε προσεγγίσεις της κανονικής κατανομής. Επιπλέον συχνά καλούμαστε να εφαρμόσουμε διάφορα διαγνωστικά τεστ για να ελέγξουμε τις υποθέσεις κάποιου παραμετρικού μοντέλου (π.χ. υποθέσεις κανονικότητας) τα οποία καταλήγουν σε διαφορετικά συμπεράσματα, δημιουργώντας ερωτηματικά . Μπορεί για παράδειγμα ο έλεγχος

Kolmogorov-Smirnov να δίνει ότι υπάρχει κανονικότητα, ενώ ο Shapiro- Wilk όχι και αν λάβουμε υπόψιν μας και τα γραφικά τεστ τότε ίσως πάλι δημιουργούνται αντικρουόμενα συμπεράσματα. Επιπλέον σε αρκετές περιπτώσεις απαιτούνται δύσκολοι υπολογισμοί από την πλευρά του αναλυτή.

Το 1979 εισήχθη από τον Bradley Efron η μέθοδος bootstrap, μια μέθοδος βασισμένη στην επαναδειγματοληψία και στην κατασκευή νέων δειγμάτων, από ένα ήδη υπάρχον, αρχικό δείγμα. Παρόμοιες τεχνικές επαναδειγματοληψίας είχαν εμφανιστεί ήδη από το 1949, όπως η μέθοδος jackknife από τον Quenouille, ενώ εκείνη την εποχή άρχισαν να χρησιμοποιούνται και οι υπολογιστές για την προσομοίωση. Από τότε και μετά, με την ταυτόχρονη πρόοδο των υπολογιστών, η bootstrap αποτελεί μια τεχνική που αναπτύσσεται, ενώ σταδιακά προστέθηκαν και άλλες παραλλαγές της μεθόδου, όπως η ABC bootstrap (1992, DiCiccio-Efron) η Bca bootstrap (1996, DiCiccio-Efron). Βασική ιδέα της μεθόδου είναι ότι από ένα αρχικό δείγμα κατασκευάζουμε νέα δείγματα κάνοντας απλή τυχαία δειγματοληψία με επανάθεση. Η μέθοδος πήρε το όνομα της από τον μυθικό βαρόνο Munchausen ο οποίος για να αποφύγει τον πνιγμό σε μία λίμνη, προσπάθησε να τραβήξει τον εαυτό του στην επιφάνεια τραβώντας προς τα πάνω τα λουριά από τις μπότες του.

Στην παρούσα εργασία, το Κεφάλαιο 1^{o} αποτελεί μια εισαγωγή στην μη παραμετρική μέθοδο bootstrap. Εκτός από την μεθοδολογία που ακολουθείται στην bootstrap αναλύονται βασικές έννοιες της στατιστικής που χρησιμοποιεί η μέθοδος, όπως είναι η εμπειρική συνάρτηση κατανομής, οι τρόποι σύγκλισης ακολουθιών τυχαίων μεταβλητών και η μέθοδος Monte Carlo. Γίνεται επίσης μια σύντομη παρουσίαση της μεθόδου jackknife, της παραμετρικής bootstrap και των περιπτώσεων στις οποίες η bootstrap ενδέχεται να αποτύχει.

Στο Κεφάλαιο 2^{o} παρουσιάζονται κάποια γενικά στοιχεία για τα ποσοστιαία σημεία και τα διαστήματα εμπιστοσύνης. Στη συνέχεια παραθέτουμε μερικούς τρόπους κατασκευής διαστημάτων εμπιστοσύνης αρχίζοντας από το βασικό bootstrap διάστημα εμπιστοσύνης και προχωρώντας με το t-bootstrap δ.ε., το percentile bootstrap και τέλος το Bca δ.ε. .

Στο Κεφάλαιο 3^{o} αρχίζει το πρακτικό κομμάτι της εργασίας, όπου δίνουμε σημειακές εκτιμήσεις και διαστήματα εμπιστοσύνης για το 50^{o} ποσοστιαίο σημείο, δηλαδή τη διάμεσο της κατανομής από την οποία προήλθαν τα δεδομένα. Χρησιμοποιούμε τις μεθόδους που αναφέρθηκαν στο Κεφάλαιο 2 και συγκρίνουμε τις σημειακές εκτιμήσεις μεταξύ τους ως προς την τυπική απόκλιση και τα διαστήματα εμπιστοσύνης ως προς την πιθανότητα κάλυψης και το μέσο πλάτος. Τα δεδομένα που χρησιμοποιούμε προέρχονται από την τυπική κανονική κατανομή, την εκθετική και την ομοιόμορφη.

Στο Κεφάλαιο $4^{ο}$ δίνουμε σημειακές εκτιμήσεις και διαστήματα εμπιστοσύνης για το $95^{ο}$ ποσοστημόριο χρησιμοποιώντας δεδομένα από την τυπική κανονική κατανομή, την εκθετική και την Pareto. Στόχος είναι να συγκρίνουμε τα αποτελέσματα, τόσο μεταξύ τους, όσο και με τα αντίστοιχα του $3^{ου}$ κεφαλαίου.

Στο Κεφάλαιο $5^{ο}$ βρίσκεται ο επίλογος της εργασίας και μερικά συμπεράσματα!

Κεφάλαιο 1^o

1.1 Εισαγωγή

Σε ένα τυπικό πρόβλημα στατιστικής μιας ενδιαφέρει να εξάγουμε συμπεράσματα για έναν πληθυσμό με τη βοήθεια ενός δείγματος. Στη διάθεση μας έχουμε συνήθως κάποιες παρατηρήσεις $x_1, x_2, \dots, x_n$ οι οποίες προήλθαν μετά από μια σειρά τυχαίων πειραμάτων. Αυτό που μας ενδιαφέρει κυρίως είναι η κατανομή του πληθυσμού και συγκεκριμένα οι παράμετροι της κατανομής, διότι μέσω της περιγραφικής στατιστικής ενδέχεται να έχουμε μια ιδέα της κατανομής αλλά να ζητάμε μια σαφέστερη εικόνα της. Για να περιγράψουμε τον πληθυσμό και την κατανομή του χρησιμοποιούμε διάφορα αριθμητικά περιγραφικά μέτρα όπως είναι η πληθυσμιακή μέση τιμή, η πληθυσμιακή διακύμανση κλπ, τα οποία ονομάζονται «παράμετροι». Οι παράμετροι είναι συγκεκριμένοι αριθμοί, μοναδικοί για κάθε πληθυσμό και μπορεί να μας είναι είτε γνωστοί είτε άγνωστοι. Τα αντίστοιχα αριθμητικά περιγραφικά μέτρα που χαρακτηρίζουν το δείγμα, όπως η δειγματική μέση τιμή, η δειγματική διασπορά, τα ποσοστημόρια του δείγματος κλπ, τα υπολογίζουμε μέσω του δείγματος μας και είναι μοναδικά για κάθε δείγμα. Σε αυτήν την περίπτωση καλούνται «στατιστικές συναρτήσεις» ή «στατιστικά». Υπολογίζοντας την τιμή μιας στατιστικής συνάρτησης για ένα δείγμα μπορούμε να γενικεύσουμε τα συμπεράσματα μας για ολόκληρο τον πληθυσμό, τα οποία και στη συνέχεια μπορούν να χρησιμοποιηθούν για την λήψη αποφάσεων.

Στην στατιστική συμπερασματολογία όπως και στην λήψη αποφάσεων βλέπουμε το δείγμα δεδομένων ως μια πραγματοποίηση/παρατήρηση ενός τυχαίου χαρακτηριστικού (τυχαία μεταβλητή), το οποίο ορίζεται στον χώρο πιθανότητας (Ω, F, P), όπου το Ω είναι ο δειγματικός χώρος, το F μια $\sigma - άλγεβρα$ του Ω και το P ένα μέτρο πιθανότητας τέτοιο ώστε $P: F \to \mathbb{R}$. Για να είναι το F μια $\sigma - άλγεβρα$ του Ω θα πρέπει:

- $\Omega \in F$
- αν για το ενδεχόμενο A έχουμε $A \in F$ τότε θα πρέπει $A^c \in F$
- αν $A_1, A_{2,\dots} \in F$ τότε και η ένωση $\cup A_i \in F$

Μια στατιστική συνάρτηση $T(X_1 \ , X_2 \ , \ldots\ldots, X_n)$ είναι και αυτή τυχαία μεταβλητή και η κατανομή πιθανοτήτων των τιμών που μπορεί να πάρει, για όλα τα τυχαία δείγματα μεγέθους n από τον ίδιο πληθυσμό, καλείται δειγματική κατανομή. Όπως είναι προφανές η δειγματική κατανομή μιας στατιστικής συνάρτησης είναι πολύ χρήσιμη στην στατιστική συμπερασματολογία. Μπορούμε να χρησιμοποιήσουμε στατιστικά θεωρήματα (π.χ. κεντρικό οριακό θεώρημα), τα οποία για συγκεκριμένες στατιστικές συναρτήσεις δίνουν είτε ακριβώς, είτε κατά προσέγγιση την δειγματική κατανομή. Ένας άλλος τρόπος είναι να πάρουμε πολλά τυχαία δείγματα μεγέθους n και σε καθένα από αυτά να υπολογίσουμε την τιμή της στατιστικής συνάρτησης T, και στη συνέχεια να κατασκευάσουμε το ιστόγραμμα σχετικών συχνοτήτων όλων των τιμών.

Αν πραγματοποιήσουμε την παραπάνω διαδικασία, είτε στον πληθυσμό είτε σε ένα ήδη δοθέν δείγμα δεδομένων, τότε αναφερόμαστε σε τεχνικές επαναδειγματοληψίας. Γιατί να προτιμήσουμε αυτές τις τεχνικές; Γιατί όπως θα δούμε και στις παρακάτω ενότητες, η επαναδειγματοληψία παρακάμπτει προβλήματα που δημιουργούνται με τις παραδοσιακές τεχνικές, όπως είναι το μικρό μέγεθος δείγματος, ή η μη ισχύς παραμετρικών υποθέσεων με τρόπο πολύ αποδοτικότερο από ότι θα επέτρεπαν δύσκολοι μαθηματικοί υπολογισμοί.

1.2 Εισαγωγή στη μέθοδο Bootstrap

Ας υποθέσουμε ότι έχουμε έναν πληθυσμό που ακολουθεί μια κατανομή F. Από τον πληθυσμό αυτό λαμβάνουμε ένα τυχαίο δείγμα $X_1, X_2 \ ,..., X_n$ (όταν αναφερόμαστε σε τυχαίο δείγμα εννοούμε ότι οι τυχαίες μεταβλητές X_i είναι ανεξάρτητες και ακολουθούν την ίδια κατανομή) και ορίζουμε το διάνυσμα $X = (X_1, X_2 \,..., X_n)$ το οποίο είναι μια n-διάστατη μεταβλητή. Μπορούμε επίσης να υποθέσουμε ότι έχουμε συλλέξει τις παρατηρήσεις για κάθε μια από τις τυχαίες μεταβλητές X_i και αυτές είναι οι $x_1, x_2, \ldots, x_n$ και επιθυμούμε να βγάλουμε κάποια συμπεράσματα σχετικά με μια άγνωστη παράμετρο θ της κατανομής F. Η παράμετρος έχει μια σταθερή τιμή, αλλά άγνωστη σε εμάς την οποία και θα επιχειρήσουμε να εκτιμήσουμε. Το ερώτημα που προκύπτει είναι ποια στατιστική συνάρτηση $T = T(X_1, X_2, \ldots, X_n)$ θα χρησιμοποιήσουμε για το σκοπό αυτό και τι χαρακτηριστικά θα διαθέτει αυτή. Σημειώνεται εδώ ότι η παράμετρος θ είναι μια συνάρτηση της κατανομής F, δηλ $\theta = t(F)$, ενώ η στατιστική T είναι συνάρτηση του τυχαίου δείγματος, δηλ $T = T(X_1, X_2, \ldots, X_n)$. Η τιμή της στατιστικής συνάρτησης για μια συλλογή παρατηρήσεων $x_1, x_2, \ldots, x_n$ συμβολίζεται με $T(x_1, x_2, \ldots, x_n)$ και αποτελεί την εκτίμηση $\hat{\theta}$ της παραμέτρου θ. Η εκτίμηση μπορεί να είναι είτε σημειακή είτε σε κάποιο διάστημα. Είναι λοιπόν αναμενόμενο ότι η εκτίμηση $\hat{\theta}$ της παραμέτρου εξαρτάται από την κατανομή F ,για την οποία δεν έχουμε κάνει καμιά υπόθεση, αλλά ακόμα και στη περίπτωση που γνωρίζαμε κάτι για την F, η $T = T(X_1, X_2 ,\ldots, X_n)$ ενδέχεται να είναι πολύ περίπλοκη συνάρτηση και ο υπολογισμός της συνάρτησης κατανομής της σχεδόν αδύνατος.

Για παράδειγμα, εάν γνωρίζουμε ότι οι X_1 , X_2 ,……, X_n προέρχονται από την κανονική συνάρτηση κατανομής και έχουν μέση τιμή μ και διακύμανση σ^2 (συμβολίζουμε X_1, X_2 ,…, $X_n \sim N(\mu, \sigma^2)$ και η παράμετρος που μας ενδιαφέρει είναι η μέση τιμή, δηλ $\theta = \mu$, τότε ο δειγματικός μέσος $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ εκτιμά αμερόληπτα την παράμετρο και η στατιστική συνάρτηση $T = \overline{X}$ ακολουθεί την κατανομή $N(\mu, \frac{\sigma^2}{n})$. Επιπλέον η διασπορά $\frac{\sigma^2}{n}$ εκτιμάται από την δειγματική διασπορά $\frac{S^2}{n}$.Παρόλα αυτά οι υπολογισμοί δεν είναι πάντα το ίδιο εύκολοι και αρκετές φορές δεν είναι εφικτό να προσδιορίσουμε μια στατιστική συνάρτηση T, τόσο σε απαραμετρικά μοντέλα στα οποία η κατανομή F είναι εντελώς άγνωστη, όσο και σε παραμετρικά στα οποία η F θεωρείται γνωστή.

Θα επιχειρήσουμε λοιπόν, δοθέντος ενός δείγματος $(X_1, X_2 ,\ldots, X_n) \sim F$, να εκτιμήσουμε μια παράμετρο $\theta = t(F)$ της κατανομής F, χωρίς να κάνουμε καμιά υπόθεση για την μορφή της συνάρτησης κατανομής. Επιλέγουμε λοιπόν, μιας και δεν έχουμε γνώσεις σχετικά με την F να χρησιμοποιήσουμε μια εκτίμησή της την οποία θα συμβολίσουμε με $\hat{F}$. Τότε η εκτίμηση της παραμέτρου θα συμβολίζεται ως $\hat{\theta} = t(\hat{F})$ και θα είναι συνάρτηση της εκτιμήτριας συνάρτησης $\hat{F}$. Προχωράμε δηλαδή ένα βήμα παρακάτω και αφού εκτιμήσουμε την συνάρτηση κατανομής F,εκτιμούμε και την παράμετρο που μας ενδιαφέρει , η οποία με την σειρά της είναι μια συνάρτηση της $\hat{F}$. Η διαδικασία αυτή αναφέρεται στην ξένη βιβλιογραφία ως «plug-in principle» (Efron-Tibshirani 1993, Trosset 2009) και βασική της ιδέα είναι ότι χρησιμοποιούμε μια ποσότητα του δείγματος για να εκτιμήσουμε την αντίστοιχη ποσότητα του πληθυσμού. Σε αρκετές περιπτώσεις επιλέγεται στη θέση της $\hat{F}$ η εμπειρική συνάρτηση κατανομής $\hat{F}_n$,αυτό ωστόσο δεν είναι δεσμευτικό. Υπάρχει δηλαδή και η περίπτωση, με βάση τα δεδομένα που έχουμε να επιλεγεί κάποια άλλη συνάρτηση. Συγκεκριμένα στην περίπτωση που η παράμετρος που θέλουμε να εκτιμήσουμε δεν είναι ομαλή συνάρτηση (smooth function), όπως για παράδειγμα η διάμεσος, ενδέχεται η χρήση κάποιας άλλης συνάρτησης να είναι καταλληλότερη από την εμπειρική συνάρτηση.

Θα δώσουμε στην συνέχεια τον ορισμό της εμπειρικής συνάρτησης κατανομής ώστε να μπορέσουμε να μελετήσουμε ένα παράδειγμα του «plug-in principle».

Ορισμός1.1 (Εμπειρική συνάρτηση κατανομής) :

Έχοντας ένα τυχαίο δείγμα$(X_1, X_2 ,\ldots, X_n)\sim F$, ορίζουμε ως εμπειρική συνάρτηση κατανομής $\widehat{F}_n$την κατανομή που κατανέμει πιθανότητα $1/n$ σε κάθε ένα από τα X_i . Δηλαδή ,

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\ X_i \leq x\ ) = \#\{X_i \leq x\ \}/n\ ,$$

όπου η $\mathrm{I}(X_i \leq x\ )$ μια δείκτρια συνάρτηση η οποία παίρνει τις τιμές:

$$\mathrm{I}(X_i \leq x\ ) = \begin{cases} 1\ , & \varepsilon\alpha\nu\ \ X_i \leq x \\ 0, & \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\alpha \end{cases}$$

Με άλλα λόγια η $\widehat{F}_n$ δίνει σε ένα σύνολο A του δειγματικού χώρου την εξής πιθανότητα:

$$\widehat{P}\ \{A\} = \#\{X_i \in A\}/n$$

Στην περίπτωση που έχουμε επαναλαμβανόμενες παρατηρήσεις τότε η εμπειρική συνάρτηση κατανομής δίνει πιθανότητα σε καθένα από τα διακριτά X_i ίση με τη συχνότητα εμφάνισής τους στο δείγμα.

Παράδειγμα 1.1

Θα μελετήσουμε το παρακάτω παράδειγμα για να κατανοήσουμε την αρχή plug-in. Ας υποθέσουμε ότι επαναλαμβάνουμε το πείραμα ρίψης ενός αμερόληπτου ζαριού 20 φορές και λαμβάνουμε το εξής δείγμα:

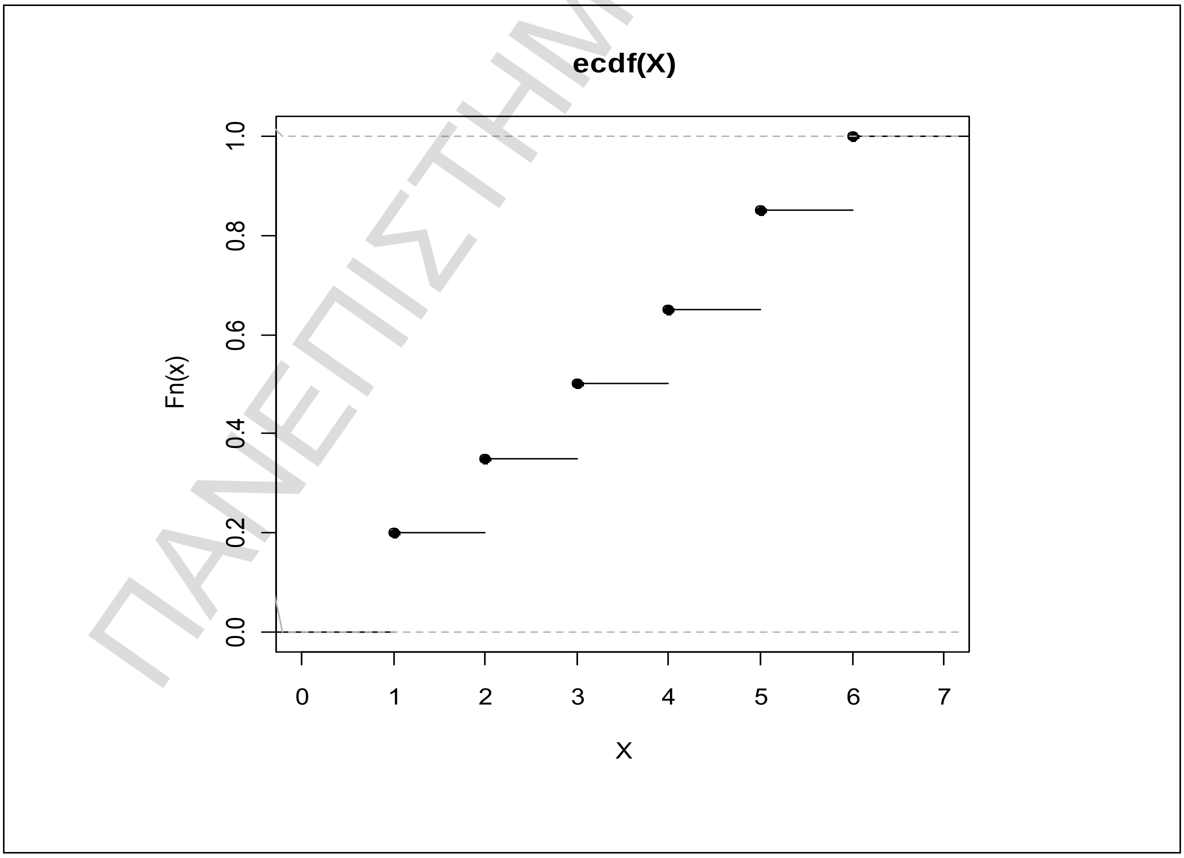
$$\vec{x} = (2,3,6,1,5,2,1,3,5,4,5,4,6,1,6,2,5,3,1,4)$$

Αναφέρουμε ότι τα αποτελέσματα προήλθαν με τυχαίο τρόπο, πραγματοποίησης του πειράματος. Γνωρίζουμε ότι οι πιθανότητες εμφάνισης καθενός από τους αριθμούς 1 έως 6 αντιστοιχεί σε 1/6 και ότι η τυχαία μεταβλητή X, που εκφράζει το αποτέλεσμα ρίψης ενός ζαριού, ακολουθεί την διακριτή ομοιόμορφη στο {1,2,…,6} κατανομή. Η εμπειρική συνάρτηση κατανομής $\hat{F}_{20}$ ωστόσο, δίνει τις παρακάτω πιθανότητες σε κάθε ένα από τα ενδεχόμενα :

x_i	$\#x_i$	$\hat{F}_{20}(x_i)$
1	4	0,2
2	3	0,15
3	3	0,15
4	3	0,15
5	4	0,2
6	3	0,15

Πίνακας 1.1

Ας δούμε και γραφικά την εμπειρική συνάρτηση κατανομής:



Εικόνα 1.1: *Γραφική παράσταση της εμπειρικής συνάρτησης κατανομής για τα δεδομένα στον Πίνακα 1.1*

Σύμφωνα με τον νόμο των μεγάλων αριθμών, όσο μεγαλώνει το μέγεθος του δείγματος η εμπειρική συνάρτηση κατανομής $\widehat{F}_n$ προσεγγίζει την συνάρτηση κατανομής απ' την οποία προήλθαν τα δεδομένα. Θα επιχειρήσουμε να δώσουμε τους «plug-in» εκτιμητές για την μέση τιμή και την διακύμανση. Η αναμενόμενη ή μέση τιμή συμβολίζεται με $\mu = E(X)$ και με βάση το παράδειγμα μας που περιλαμβάνει διακριτά δεδομένα και έξι δυνατά ενδεχόμενα, ισούται με:

$$E(X) = \sum_{i=1}^{6} x_i P(X = x_i) = 3{,}5.$$

Η plug-in εκτίμηση της μέσης τιμής, την οποία θα συμβολίσουμε με $\hat{\mu}_n$ συμπίπτει με τον δειγματικό μέσο και δίνεται από τον τύπο:

$$\hat{\mu}_n = \sum_{i=1}^{6} x_i \, \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{6} x_i = \overline{x}_n$$

Ας κάνουμε τις πράξεις στο παράδειγμα μας:

$$\hat{\mu}_{20} = 2 \times \frac{1}{20} + 3 \times \frac{1}{20} + 6 \times \frac{1}{20} \times \ldots \times 1 \times \frac{1}{20} + 4 \times \frac{1}{20}$$

$$= 1 \times 0{,}2 + 2 \times 0{,}15 + 3 \times 0{,}15 + 4 \times 0{,}15 + 5 \times 0{,}2 + 6 \times 0{,}15 = 3{,}45$$

Αντίστοιχα, αν συμβολίσουμε με $\sigma^2 = VarX_i$ την διακύμανση ενός πληθυσμού, τότε η plug-in εκτίμηση θα συμβολίζεται με $\widehat{\sigma_n^2}$ και θα δίνεται από τον τύπο:

$$\widehat{\sigma_n^2} = \sum_{i=1}^{n} (x_i - \hat{\mu}_n)^2 \cdot \frac{1}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^2$$

Στο παράδειγμα μας η διακύμανση του πληθυσμού δίνεται ως εξής:

$$\sigma^2 = E(X_i^2) - [E(X_i)]^2,$$

και ισούται με :

$$\sigma^2 = \frac{1^2 + \cdots + 6^2}{6} - 3{,}5^2 = 2{,}917$$

Η plug-in εκτίμηση της διακύμανσης δίνεται ως εξής:

$$\widehat{\sigma_{20}^2} = (1^2 \times 0{,}2 + \cdots + 6^2 \times 0{,}15) - 3{,}45^2 = 3{,}048$$

Αξίζει να σημειώσουμε ότι η plug-in εκτίμηση της διασποράς δεν είναι αμερόληπτη εκτίμηση της διασποράς του πληθυσμού και δεν ταυτίζεται με την δειγματική διασπορά. Συγκεκριμένα:

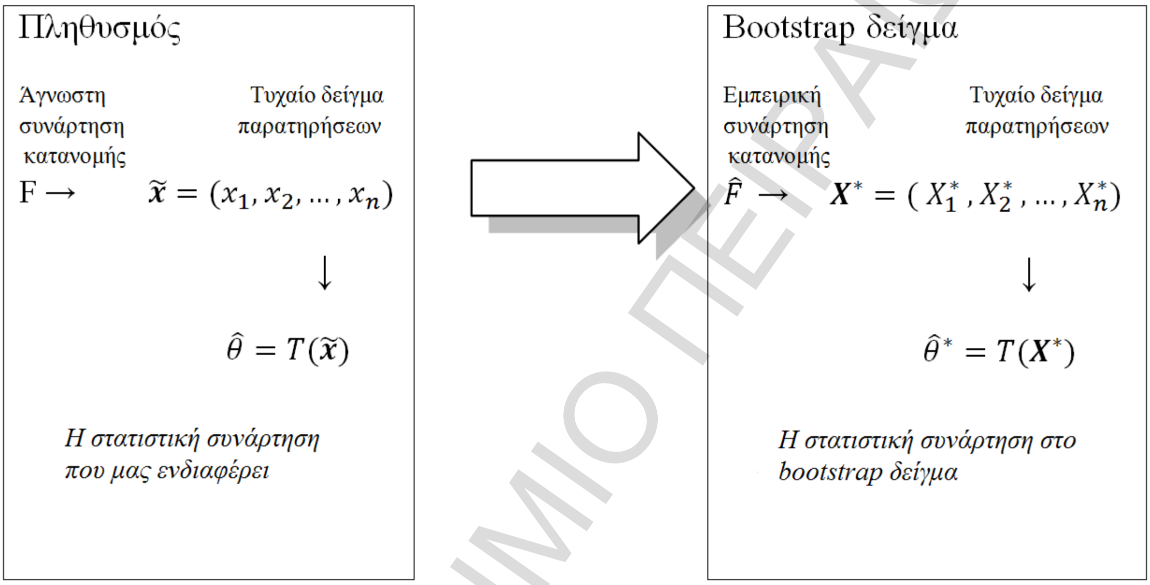
$$E(\widehat{\sigma_n^2}) = E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_n)^2\right] = \frac{n-1}{n} \cdot \sigma^2$$

Ως δειγματική διασπορά εννοούμε την αμερόληπτη εκτίμηση της πληθυσμιακής διασποράς:

$$\widehat{S_n^2} = \frac{n}{n-1}\left[\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_n)^2\right] = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_n)^2.$$

Παραμένει το ζήτημα, ποια στατιστική συνάρτηση $T = T(X_1, X_2, \dots, X_n)$ θα χρησιμοποιήσουμε και τι χαρακτηριστικά θα έχει αυτή. Εάν η κατανομή F των X_1 ,X_2 ,…, X_n ήταν γνωστή θα μπορούσαμε να προσδιορίσουμε τα χαρακτηριστικά και την κατανομή της T, την οποία ας ονομάσουμε F_T. Θα εκτιμήσουμε λοιπόν την κατανομή F_T της $T = T(X_1, X_2, \dots, X_n)$, όπου $X_i \sim F$, από την κατανομή της στατιστικής συνάρτησης T^* = $(T(X_1^*, X_2^*, \dots, X_n^*)$, όπου τα X_i^* ακολουθούν την $\hat{F}$, αντί της άγνωστης F. Δηλαδή, σε κάθε περίπτωση διάφορα χαρακτηριστικά της T, όπως είναι η μεροληψία ,η διασπορά και τα ποσοστιαία σημεία εκτιμώνται με την T^*. Αυτή η τακτική είναι και η βασική ιδέα της μεθόδου bootstrap, η οποία εισήχθη από τον Efron το 1979. Η προσέγγιση αυτή είναι χρήσιμη, ιδιαίτερα όταν η μόνη πληροφορία που έχουμε για την F προέρχεται από το δείγμα των X_i, ενώ είναι λιγότερη καλή όταν υπάρχουν κι άλλες πληροφορίες γύρω από την F, όπως δηλαδή το ότι ανήκει στην παραμετρική οικογένεια κατανομών, ή ότι υπάρχει παλινδρόμηση, δηλαδή το τυχαίο δείγμα των X_i εξαρτάται από μια άλλη μεταβλητή.

Συγκεκριμένα, από ένα αρχικό δείγμα παρατηρήσεων $\tilde{\boldsymbol{x}} = x_1, x_2, \dots, x_n$ μεγέθους n, που προέρχεται από την κατανομή F, κατασκευάζω ένα bootstrap δείγμα του ίδιου μεγέθους πραγματοποιώντας δειγματοληψία με επανάθεση. Αυτό σημαίνει ότι κάποια από τις παρατηρήσεις του αρχικού μας δείγματος μπορεί να εμφανίζεται 0,1,2, ακόμα και n φορές στο bootstrap δείγμα. Αν συμβολίσουμε το bootstrap δείγμα ως $X_1^*, X_2^*, \dots, X_n^*$,τότε αυτό θα προέρχεται από την εμπειρική συνάρτηση κατανομής $\widehat{F}_n$, κατανομή που εκτιμά την κατανομή F. Υπολογίζουμε για το $X_1^*, X_2^*, X_3^* \dots, X_n^*$ την στατιστική συνάρτηση T^* = $T(X_1^*, X_2^*, \dots, X_n^*)$ και επαναλαμβάνοντας την παραπάνω διαδικασία B φορές ,θα έχουμε συνολικά Β bootstrap δείγματα (αναφέρονται στην ξένη βιβλιογραφία ως bootstrap replicates), τα οποία θα συμβολίσουμε ως $X^{*1}, X^{*2}, \dots, X^{*B}$ και αντίστοιχα $T_1^*, T_2^*, \dots, T_B^*$. Σημειώνεται ότι μετά τις Β επαναλήψεις της διαδικασίας αυτό που λαμβάνουμε είναι μια Monte Carlo εκτίμηση της στατιστικής συνάρτησης που μας ενδιαφέρει. Θα παραθέσουμε μία σχηματική αναπαράσταση που περιγράφει την παραπάνω διαδικασία ώστε να γίνει πιο κατανοητή η μέθοδος:



Εικόνα1.2 *Πηγή: Efron&Tibshirani(1993)*

Όπως έγινε κατανοητό τα $X_1^*, X_2^*, \dots, X_n^*$ δεν αποτελούν το πραγματικό δείγμα, αλλά μια εκδοχή του πραγματικού δείγματος στο οποίο έχει γίνει δειγματοληψία με επανάθεση. Για αυτό άλλωστε και η μέθοδος bootstrap είναι μια μέθοδος επαναδειγματοληψίας (resampling), διότι κάνουμε δειγματοληψία μέσα από το δείγμα. Στη συγκεκριμένη περίπτωση μάλιστα, επειδή δεν κάναμε καμία υπόθεση για την κατανομή της F η μέθοδος καλείται απαραμετρική μέθοδος bootstrap. Η εκτιμήτρια $\hat{F}$ που χρησιμοποιείται για την παραγωγή των bootstrap δειγμάτων μπορεί να είναι οποιαδήποτε εκτιμήτρια της F, όχι απαραίτητα η εμπειρική συνάρτηση κατανομής (Efron-Tibshirani 1993). Ωστόσο βέβαια μια μη παραμετρική εκτιμήτρια της F είναι η εμπειρική συνάρτηση κατανομής η οποία κατανέμει πιθανότητα σε καθένα από τα $x_1, x_2, \dots, x_n$ ίση με 1/n και στην περίπτωση που έχουμε επαναλαμβανόμενες τιμές ή κατηγορικά δεδομένα, η εμπειρική συνάρτηση κατανομής δίνει βάρος σε κάθε διακριτή τιμή του δείγματος ίση με το ποσοστό της συχνότητας στο δείγμα. Στην επόμενη ενότητα θα δούμε για ποιόν λόγο η εμπειρική συνάρτηση κατανομής $\hat{F}_n$ αποτελεί μια καλή εκτίμηση της συνάρτησης κατανομής F του αρχικού δείγματος και την χρησιμότητά της στην μέθοδο bootstrap.

1.3 Η Εμπειρική συνάρτηση κατανομής στην μέθοδο Bootstrap

Σε αυτή την ενότητα θα δούμε τους κυριότερους τρόπους σύγκλισης μιας ακολουθίας τυχαίων μεταβλητών καθώς και μερικές ιδιότητες της εμπειρικής συνάρτησης κατανομής. Η εμπειρική συνάρτηση κατανομής παίζει μεγάλο ρόλο στη μέθοδο bootstrap γιατί συνήθως την επιλέγουμε ως εκτιμήτρια της κατανομής του πληθυσμού.

1.3.1 Οι τρόποι σύγκλισης και οι νόμοι των μεγάλων αριθμών

Πριν αναφερθούμε στην εμπειρική συνάρτηση κατανομής κρίνεται χρήσιμο να δώσουμε επιγραμματικά τους τρόπους σύγκλισης μιας ακολουθίας ανεξάρτητων και ισόνομων τυχαίων μεταβλητών καθώς και τους νόμους των μεγάλων αριθμών. Τέσσερεις τρόποι σύγκλισης παρουσιάζονται στη συνέχεια (αναλυτικότερα βλ. Κούτρας 2005):

- Σύγκλιση κατά πιθανότητα: Μία ακολουθία τυχαίων μεταβλητών $\{X_n\}$ ορισμένες στον δειγματικό χώρο Ω, θα λέμε ότι συγκλίνει κατά πιθανότητα σε μια τυχαία μεταβλητή X, επίσης ορισμένη στο δειγματικό χώρο Ω, εάν :

$$\lim_{\nu\to\infty} P(|X_n - X| > \varepsilon) = 0, \text{ για κάθε } \varepsilon > 0$$

- Σύγκλιση σχεδόν βέβαια: Μία ακολουθία τυχαίων μεταβλητών $\{X_n\}$,θα λέμε ότι συγκλίνει σχεδόν βεβαίως ή με πιθανότητα 1, στην τυχαία μεταβλητή X εάν ισχύει:

$$P(\lim_{n\to\infty}|X_n - X| = 0) = 1$$

Η σχεδόν βέβαια σύγκλιση συνεπάγεται ότι όταν το n μεγαλώνει απεριόριστα, σχεδόν για κάθε ω του δειγματικού χώρου Ω η διαφορά $X_n(\omega) - X(\omega)$ γίνεται αυθαίρετα μικρή.

- <u>Σύγκλιση κατά κατανομή:</u> Μία ακολουθία τυχαίων μεταβλητών $\{X_n\}$,θα λέμε ότι συγκλίνει κατά κατανομή στην τυχαία μεταβλητή X, εάν η ακολουθία συναρτήσεων κατανομής πιθανότητας $\{F_n\}$ των X_1 ,..., X_n συγκλίνει στη συνάρτηση κατανομής F της X. Δηλαδή:

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \text{ πού είναι σημείο συνέχειας της } F$$

- <u>Σύγκλιση κατά τετραγωνικό μέσο</u> : Μία ακολουθία τυχαίων μεταβλητών $\{X_n\}$,θα λέμε ότι συγκλίνει κατά τετραγωνικό μέσο στην τυχαία μεταβλητή X, εάν ισχύει:

$$\lim_{n\to\infty} E[|X_n - X|^2] = 0$$

Σημειώνουμε ότι τόσο η σχεδόν βέβαια σύγκλιση, όσο και η σύγκλιση κατά τετραγωνικό μέσο συνεπάγονται την σύγκλιση κατά πιθανότητα, καθώς και ότι η σύγκλιση κατά πιθανότητα συνεπάγεται την σύγκλιση κατά κατανομή.

Επιπλέον, σύμφωνα με τον *ασθενή νόμο των μεγάλων αριθμών*, για κάθε ακολουθία ανεξάρτητων και ισόνομων μεταβλητών $\{X_n\}$ με μέση τιμή $E(X_i) = \mu < \infty$ και $V(X_i) = \sigma^2 < \infty$, ο δειγματικός μέσος $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ συγκλίνει *κατά πιθανότητα* στη μέση τιμή μ. Κατά συνέπεια έχουμε:

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

Σύμφωνα με τον *ισχυρό νόμο των μεγάλων αριθμών* , αν έχουμε μια ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών $\{X_n\}$, με μέση τιμή $E(X_i) = \mu<\infty$ και διακύμανση $V(X_i) = \sigma^2 < \infty$, τότε η ακολουθία των δειγματικών μέσων $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, *συγκλίνει σχεδόν βεβαίως* στη μέση τιμή μ. Οπότε έχουμε:

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| = 0) = 1$$

1.3.2 Η εμπειρική συνάρτηση κατανομής και οι ιδιότητές της

Όπως αναφέρθηκε παραπάνω επιθυμούμε να εκτιμήσουμε την κατανομή από την οποία προήλθε ένα τυχαίο δείγμα. Εάν λοιπόν έχουμε ένα τυχαίο δείγμα $\boldsymbol{X} = (X_1\ , X_2\ ,\ldots, X_n)$ που προέρχεται από την κατανομή F, τότε για δεδομένο $x \in \mathbb{R}$, η εμπειρική συνάρτηση κατανομής

$$\widehat{F}_n(x;\boldsymbol{X}) = \frac{\#\{X_i \leq x\}}{n}, \quad x \in \mathbb{R}$$

είναι αμερόληπτος εκτιμητής της $F(x)$. Δηλαδή $\mathrm{E}[\,\widehat{F}_n(x)] = F(x)$. Ουσιαστικά η εμπειρική συνάρτηση κατανομής είναι μια συνάρτηση του τυχαίου δείγματος $\boldsymbol{X} = (X_1\ , X_2\ ,\ldots, X_n)$ και αποτελεί εκτίμηση της συνάρτησης κατανομής από την οποία προέρχεται το δείγμα.

Για δεδομένες τιμές $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ των $X_1, X_2\ ,\ldots, X_n$ η συνάρτηση $\widehat{F}_n(\cdot;\boldsymbol{x})$ είναι μια συνάρτηση κατανομής αφού έχει τις τρείς χαρακτηριστικές ιδιότητες μιας συνάρτησης κατανομής (π.χ. βλέπε Ηλιόπουλος (2006)):

- Είναι αύξουσα , αφού αν $y_1 < y_2$, τότε $\#\{x_i \leq y_1\} \leq \#\{x_i \leq y_1\}$ και συνεπώς $\widehat{F}_n(y_1;\boldsymbol{x}) \leq \widehat{F}_n(y_2;\boldsymbol{x})$.
- Είναι δεξιά συνεχής αφού καθώς $y \to y_0$, ισχύει ότι $\#\{x_i \leq y\} \to \#\{x_i \leq y_0\}$ και συνεπώς $\lim_{y \to y_0} \widehat{F}_n(y;\boldsymbol{x}) = \widehat{F}_n(y_0;\boldsymbol{x})$.
- Καθώς $y \to -\infty$ ισχύει ότι $\#\{x_i \leq y\} \to 0$ και καθώς $y \to \infty$ ισχύει ότι $\#\{x_i \leq y\} \to n$. Οπότε έχουμε ότι $\lim_{y \to -\infty} \widehat{F}_n(y;\boldsymbol{x}) = 0$ και $\lim_{y \to \infty} \widehat{F}_n(y;\boldsymbol{x}) = 1$.

Επιπλέον η $\widehat{F}_n(\boldsymbol{x})$ συγκλίνει στην $F(x)$ σχεδόν βεβαίως. Ας δούμε λίγο αναλυτικότερα την απόδειξη (π.χ. βλέπε Κούτρας 2005), για να κατανοήσουμε για ποιον λόγο η εμπειρική συνάρτηση κατανομής $\widehat{F}_n(\boldsymbol{x})$ είναι μια καλή επιλογή για την εκτίμηση της $F(x)$, ειδικά όταν έχουμε μεγάλα δείγματα. Υποθέτουμε ότι μελετάμε μια τυχαία μεταβλητή X και θέλουμε να εκτιμήσουμε την αθροιστική συνάρτηση κατανομής $F(x)$ του X, σε ένα συγκεκριμένο σημείο $x \in \mathbb{R}$. Λαμβάνουμε τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών οι οποίες ακολουθούν την κατανομή που περιγράφεται από την F ορίζουμε την δείκτρια συνάρτηση και την εμπειρική συνάρτηση κατανομής $\widehat{F}_n(x)$ ως εξής:

$$\mathrm{I}(X_i \leq x\,) \quad = \begin{cases} 1\,, & \varepsilon\alpha\nu\ X_i \leq x \\ 0, & \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\alpha \end{cases}, \qquad\qquad \widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i \leq x\,)\,.$$

Είναι προφανές ότι η $\hat{F}_n(x)$ εκφράζει το ποσοστό των παρατηρήσεων του δείγματος που είναι μικρότερες ή ίσες του x. Η $\hat{F}_n(x)$ ορίζεται για κάθε x και έτσι σχηματίζεται τελικά μια συνάρτηση. Αφού οι τυχαίες μεταβλητές X_i είναι ανεξάρτητες ,το ίδιο θα ισχύει και για τις δείκτριες συναρτήσεις και επιπλέον έχουμε:

$\mu = E(\mathrm{I}(X_i \leq x\ ) \ = 1\cdot P(\mathrm{I}(X_i \leq x\ ) = 1) + 0 \cdot P(\mathrm{I}(X_i \leq x\ ) \ = 0) = P(\mathrm{I}(X_i \leq x\ ) \ = 1) = P(X_i \leq x) = F(x)$

Και επίσης:

$E(\mathrm{I}(X_i \leq x\ )^{\,2}) = 1^2 \cdot P(\mathrm{I}(X_i \leq x) = 1) + 0^2 \cdot P(\mathrm{I}(X_i \leq x) = 0) = \ P(\mathrm{I}(X_i \leq x) = 1\ \ ) = P(X_i \leq x) = F(x)$

από όπου προκύπτει ότι:

$$\sigma^2 = V(\mathrm{I}(X_i \leq x\ )\ ) = \ E(\mathrm{I}(X_i \leq x\ )^{\,2}) - [E(\mathrm{I}(X_i \leq x)]^2 = \ F(x) - F(x)^2$$
$$= \ F(x)\big(1 - F(x)\big) < \infty$$

Από τον ισχυρό νόμο των μεγάλων αριθμών θα ισχύει ότι:

$\lim_{n\to\infty} \hat{F}_n(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i \leq x\ ) \ = E(\mathrm{I}(X_i \leq x\ ) = \ F(x)$ με πιθανότητα 1.

Επιπλέον, η $\hat{F}_n(x)$ είναι συνεπής εκτιμητής της $F(x)$. Η συνέπεια είναι μια ασυμπτωτική ιδιότητα για πολύ μεγάλα μεγέθη δείγματος (n→∞).

Ορισμός 1.1 (Συνεπής εκτιμητής): Γενικότερα αν η *T=T(*X_1, X_2 *,…,* X_n*)* είναι ένας εκτιμητής της παραμετρικής συνάρτησης $g(\theta)$, τότε η Τ(X_1, X_2 ,…, X_n) καλείται συνεπής εκτιμητής της $g(\theta)$ εάν, καθώς n→∞, η πιθανότητα:

$$P(|\mathrm{T}(X_1\,, X_2\ \,, \dots \dots, X_n\ \ ) - \mathrm{g}(\theta)| > \varepsilon) \to 0, \text{ για κάθε } \varepsilon > 0.$$

Με άλλα λόγια, η πιθανότητα να αποκλίνει ο εκτιμητής από την ποσότητα που εκτιμά περισσότερο από ε, συγκλίνει στο μηδέν για κάθε θετικό και πολύ μικρό ε. Μια ικανή συνθήκη που επαληθεύει την συνέπεια είναι η εξής:

Η $T = T(X_1, X_2, \dots, X_n)$ είναι συνεπής εκτιμητής της $g(\theta)$ εάν ισχύουν τα παρακάτω:

i. $E(T) \rightarrow g(\theta)$, καθώς $n \rightarrow \infty$
ii. $Var(T) \rightarrow 0$, καθώς $n \rightarrow \infty$

Οπότε από τον ορισμό της συνέπειας θα έχουμε:

$$P(\left| \hat{F}_n(x) - F(x) \right| > \varepsilon) \rightarrow 0\, , \text{για κάθε } \varepsilon > 0$$

Σε αυτήν την περίπτωση θα λέμε ότι η $\hat{F}_n(x)$ συγκλίνει στην $F(x)$ κατά πιθανότητα. Όπως έχουμε αναφέρει άλλωστε η σχεδόν βέβαιη σύγκλιση συνεπάγεται την σύγκλιση κατά πιθανότητα, η οποία με τη σειρά της συνεπάγεται την σύγκλιση κατά κατανομή.

Δεδομένου ότι η εμπειρική συνάρτηση κατανομής $\hat{F}_n(x)$ αποτελεί μια εκτίμηση ολόκληρης της συνάρτησης κατανομής $F(x)$, από την οποία προέρχεται το δείγμα, μας ενδιαφέρει να προσδιορίσουμε πόσο απέχει από την $F(x)$. Η «απόσταση» αυτών των δύο συναρτήσεων δίνεται από την *απόσταση Kolmogorov* ως εξής:

$$sup_{x\in\mathbb{R}} | \hat{F}_n(x) - F(x)|$$

Οδηγούμαστε έτσι στο θεώρημα Glivenko-Cantelli σύμφωνα με το οποίο:

<u>Θεώρημα 1.1</u> <u>(Glivenko-Cantelli)</u>: Υποθέτουμε ότι έχουμε την $X_1, X_2, \dots, X_n$ ακολουθία των ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από την συνάρτηση κατανομής F ορισμένη στο $\mathbb{R}$. Τότε η απόσταση Kolmogorov μεταξύ της εμπειρικής συνάρτησης κατανομής $\hat{F}_n(x)$ και της $F(x)$ συγκλίνει ομοιόμορφα στο μηδέν. Δηλαδή:

$$\sup|F_n(x) - F(x)| \rightarrow 0 \;\; \textit{σχεδόν βεβαίως} \;\; \text{καθώς } n \rightarrow \infty.$$

Το θεώρημα Glivenko-Cantelli δίνει μια ισχυρότερη μορφή σύγκλισης, την ομοιόμορφη σύγκλιση. Ουσιαστικά όσο μεγαλώνει το μέγεθος του δείγματος, τότε τόσο πιο κοντά πλησιάζει (και μάλιστα ομοιόμορφα) η εμπειρική συνάρτηση κατανομής την πραγματική κατανομή του δείγματος $\boldsymbol{X} = (X_1 , X_2 ,\ldots, X_n)$.

Αξίζει να σημειώσουμε ότι ο ισχυρός νόμος των μεγάλων αριθμών μας δίνει την κατά σημείο σύγκλιση, διότι κάθε φορά έχουμε ένα δεδομένο $x \in \mathbb{R}$, ενώ το θεώρημα Glivenko-Cantelli μας δίνει μια ισχυρότερη μορφή σύγκλισης , την ομοιόμορφη σύγκλιση.

Γενικότερα όπως γνωρίζουμε από την πραγματική ανάλυση αν έχουμε μια ακολουθία συναρτήσεων $f_n: I \to \mathbb{R}$ και $f: I \to \mathbb{R}$ τότε θα λέμε ότι η f_n *συγκλίνει κατά σημείο* στην f εάν για κάθε $x \in I$ έχουμε ότι:

$$\lim_{n\to\infty} f_n(x) = f\,(x).$$

Ισοδύναμα μπορούμε να πούμε ότι f_n *συγκλίνει κατά σημείο* στην f εάν $\forall x \in I,$ για κάθε $\varepsilon > 0,\ \exists n_0$ τέτοιο ώστε $\forall n \geq n_0$ να ισχύει $|f_n(x) - f(x)| < \varepsilon.$

Θα λέμε ότι η f_n *συγκλίνει ομοιόμορφα* στην f εάν έχουμε ότι:

$$\lim_{n\to\infty} [sup_{x\in I}|f_n(x) - f(x)|] = 0.$$

Ισοδύναμα μπορούμε να πούμε ότι f_n *συγκλίνει ομοιόμορφα* στην f αν για κάθε $\varepsilon > 0,\ \exists n_0$ τέτοιο ώστε $\forall n \geq n_0$ να ισχύει $|f_n(x) - f(x)| < \varepsilon$,για κάθε $x.$

Η διαφορά εντοπίζεται στο ότι στην κατά σημείο σύγκλιση το n_0 εξαρτάται από το ε και από το x, ενώ στην ομοιόμορφη σύγκλιση το n_0 εξαρτάται μόνο από το ε. Έτσι η απαίτηση για ομοιόμορφη σύγκλιση είναι ισχυρότερη.

1.4 Μέθοδος bootstrap και τυπικό σφάλμα

Όπως σε κάθε εκτίμηση μας ενδιαφέρει η ακρίβειά της, έτσι και στις εκτιμήσεις που λαμβάνουμε με την μέθοδο bootstrap μας ενδιαφέρει να υπολογίσουμε το τυπικό σφάλμα. Ας υποθέσουμε λοιπόν ότι έχουμε ένα δείγμα $x_1, x_2, \ldots, x_n$ το οποίο προέρχεται από τη κανονική συνάρτηση κατανομής με μέση τιμή μ και διακύμανση σ^2,δηλαδή $N(\mu, \sigma^2)$. Όπως γνωρίζουμε ο δειγματικός μέσος εκτιμά αμερόληπτα την μέση τιμή του πληθυσμού. Η εκτίμηση του τυπικού σφάλματος για τον δειγματικό μέσο $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ δίνεται από τον τύπο:

$$\widehat{se}(\bar{x}) = \sqrt{\frac{S^2}{n}} \quad , \quad \text{όπου } S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1) \ .$$

Αν τα x_i λοιπόν, προέρχονται από την κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2, τότε ο δειγματικός μέσος $\bar{x}$ ακολουθεί και αυτός την κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2/n , δηληδή $\bar{x} \sim \mathrm{N}(\mu, \sigma^2/n)$.Όπως γίνεται αντιληπτό ο δειγματικός μέσος έχει μικρότερη διακύμανση από τα x_i και μάλιστα όσο μεγαλύτερο είναι το μέγεθος του δείγματος τόσο μικρότερη θα είναι η διακύμανση του δειγματικού μέσου και τόσο καλύτερη θα είναι η εκτίμηση. Στην περίπτωση όμως που το δείγμα μας προέρχεται από κάποια άγνωστη κατανομή F, δεν είναι εφικτό να κάνουμε μια εκτίμηση της διακύμανσης ώστε να εκτιμήσουμε το τυπικό σφάλμα. Δηλαδή αν τα $x_1, x_2, \dots, x_n$ προέρχονται από την F, τότε ο δειγματικός μέσος θα έχει τις παρακάτω παραμέτρους $\bar{x} \sim \left(\mu_F, \frac{\sigma_F^2}{n}\right)$ και το τυπικό σφάλμα θα δίνεται από τον τύπο:

$$se_F(\bar{x}) = [Var(\bar{x})]^{1/2} = \sigma_F/\sqrt{n}.$$

Με βάση το κεντρικό οριακό θεώρημα, για n→ ∞ ο δειγματικός μέσος ακολουθεί ασυμπτωτικά την κανονική κατανομή $\mathrm{N}(\mu_F, \sigma_F^2/n)$ και έτσι μπορούμε να πούμε οτι ο μέσος για τον πληθυσμό αναμένεται με πιθανότητα 68,3% να απέχει από την εκτίμηση του το πολύ ένα τυπικό σφάλμα, ενώ με πιθανότητα 95,4% θα απέχει λιγότερο από δυο τυπικά σφάλματα από την εκτίμηση του. Για τις περισσότερες στατιστικές συναρτήσεις ωστόσο, δεν υπάρχει κάποιος τύπος ώστε να υπολογίζεται άμεσα το τυπικό σφάλμα και εκτός από αυτό δεν είναι πάντα εφικτό να εφαρμόσουμε το κεντρικό οριακό θεώρημα. Με τη βοήθεια της μεθόδου bootstrap απαλλασσόμαστε από τις υποθέσεις του κεντρικού οριακού θεωρήματος και λαμβάνουμε μια εκτίμηση για το τυπικό σφάλμα μιας οποιασδήποτε στατιστικής συνάρτησης $T(X)$ ακόμα και αν δεν γνωρίζουμε την κατανομή F από την οποία προέρχεται το δείγμα.

Ας υποθέσουμε λοιπόν πως από ένα τυχαίο δείγμα $X_1, X_2, \dots, X_n$ που προέρχεται από την κατανομή F, έχουμε τις παρατηρήσεις $x_1, x_2, \dots, x_n$, και μας ενδιαφέρει η παράμετρος $\theta = t(F)$, η οποία εκτιμάται από την $\hat{\theta} = T(x)$. Ουσιαστικά αν το τυπικό σφάλμα της εκτίμησης $\hat{\theta} = T(x)$ είναι το $se_F(\hat{\theta})$, εμείς θα το εκτιμήσουμε με το $se_{\hat{F}}(\hat{\theta}^*)$. Το $se_{\hat{F}}(\hat{\theta}^*)$ είναι μια plug-in εκτίμηση του $se_F(\hat{\theta})$, αφού αντί του αρχικού δείγματος που ακολουθεί την κατανομή F, θα χρησιμοποιήσουμε τα δείγματα μεγέθους n που προέρχονται από την συνάρτηση κατανομής $\hat{F}$, η οποία εκτιμά την F. Ας χρησιμοποιήσουμε την εμπειρική συνάρτηση κατανομής $\hat{F}_n$ στην θέση της $\hat{F}$. Τότε η εκτίμηση $se_{\hat{F}_n}(\hat{\theta}^*)$ καλείται ιδανική εκτίμηση bootstrap για το τυπικό σφάλμα (ideal bootstrap estimate of standard error σύμφωνα με τους Efron-Tibshirani 1993) της $\hat{\theta} = T(x)$ και είναι μη παραμετρική, διότι βασίζεται στην $\hat{F}_n$ χωρίς λάβουμε υπ' όψιν την F. Επειδή όπως αναφέρθηκε παραπάνω δεν υπάρχει πάντα κάποιος τύπος που να μας οδηγεί στην εκτίμηση του τυπικού σφάλματος προτείνεται στη συνέχεια ένας υπολογιστικός τρόπος για την προσέγγιση του $se_{\hat{F}}(\hat{\theta}^*)$, ο οποίος και μας οδηγεί στην bootstrap εκτίμηση του τυπικού σφάλματος την οποία και θα συμβολίσουμε με $\widehat{se}_B$.

Αναλυτικότερα, όπως περιγράφτηκε παραπάνω από τα αρχικά δεδομένα $\tilde{\boldsymbol{x}} = x_1, x_2, \dots, x_n$ λαμβάνουμε ένα δείγμα μεγέθους n με επανάθεση, έστω $X^* = (X_1^*, X_2^*, \dots, X_n^*)$. Το δείγμα των X^* αποτελείται από τα στοιχεία του αρχικού δείγματος $(x_1, x_2, \dots, x_n)$, κάποια από τα οποία εμφανίζονται μηδέν φορές, δύο φορές κοκ, αλλά φυσικά δεν αποτελεί το αυθεντικό σύνολο των δεδομένων. Επαναλαμβάνουμε τη διαδικασία B φορές και έτσι κατασκευάζουμε B ανεξάρτητα bootstrap δείγματα $X^{*1}, X^{*2}, \dots, X^{*B}$. Η μέθοδος bootstrap θα μας βοηθήσει να εκτιμήσουμε το τυπικό σφάλμα της $T(\tilde{\boldsymbol{x}})$, χωρίς περιττούς υπολογισμούς, ανεξάρτητα από το πόσο πολύπλοκος είναι ο τύπος της. Για κάθε ένα από τα bootstrap δείγματα $X^{*1}, X^{*2}, \dots, X^{*B}$ υπολογίζουμε την τιμή $\hat{\theta}^* =$T$(X^{*b})\,, όπου\ b = 1, \dots, B$. Η τυπική απόκλιση των $T(X^{*1}), T(X^{*2}), \dots, T(X^{*B})$ είναι η εκτίμηση του τυπικού σφάλματος της $T(\tilde{\boldsymbol{x}})$. Δηλαδή,

$$\widehat{se}_{boot} = \frac{\sum_{b=1}^{B}[T(X^{*b}) - T(\cdot)]^2}{B-1}$$

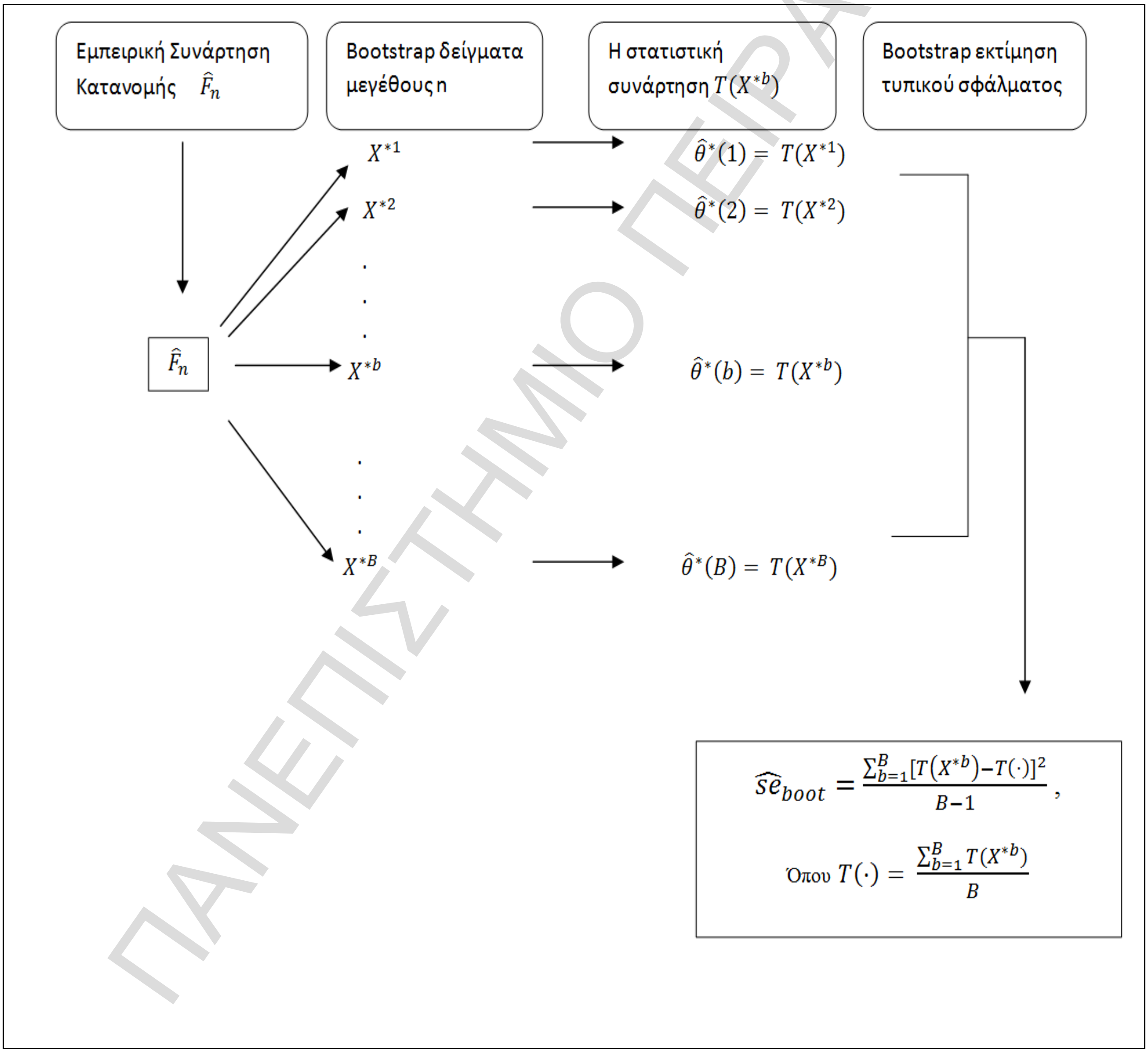
όπου το $T(\cdot)$ δίνεται από την σχέση:

$$\mathrm{T}(\cdot) = \frac{\sum_{b=1}^{B} T(X^{*b})}{B}.$$

Σημειώνουμε ότι η ποσότητα Τ(X^{*b}) προκύπτει επειδή εφαρμόσαμε την στατιστική συνάρτηση $T(X)$ σε κάθε ένα από τα bootstrap δείγματα (b=1,2,....B) . Για παράδειγμα αν η $T(X) = \bar{X}$, τότε για την συνάρτηση $T(X^{*b})$ θα έχω:

$$T(X^{*b}) = \bar{X}^{*} = \sum_{i=1}^{n} \frac{X_i^{*}}{n}$$

Στη συνέχεια παρουσιάζεται μια σχηματική αναπαράσταση της διαδικασίας καθώς και ο αλγόριθμος για την Bootstrap εκτίμηση του τυπικού σφάλματος μιας στατιστικής συνάρτησης $\hat{\theta} = T(x)$. (B.Efron, R.J. Tibshirani 1993)



Εικόνα 1.3 *Πηγή: Efron&Tibshirani(1993)*

<u>**Αλγόριθμος 1.1**: Εκτίμηση τυπικού σφάλματος:</u>

Βήμα 1ο : Διάλεξε B ανεξάρτητα bootstrap δείγματα $X^{*1}, X^{*2}, \dots, X^{*B}$ το κάθενα από τα οποία αποτελείται από n στοιχεία επιλεγμένα τυχαία και με επανάθεση από το αρχικό δείγμα $x_1, x_2, \dots, x_n \sim F$, και προέρχεται από την $\hat{F}$,η οποία ας είναι η εμπειρική συνάρτηση κατανομής.

Βήμα 2ο : Εφάρμοσε την συνάρτηση T σε κάθε ένα από τα bootstrap δείγματα ως εξής :

$$\hat{\theta}^*(b) = T(X^{*b}) \text{ , για } b=1,2,\dots,B$$

Βήμα 3ο : Υπολόγισε το τυπικό σφάλμα $se_F(\hat{\theta})$ από την τυπική απόκλιση των B Bootstrap δειγμάτων:

$$\widehat{se}_B = \{\sum_{b=1}^{B} [\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2 / (B-1)\}^{1/2},$$

όπου :

$$\hat{\theta}^*(.) = \sum_{b=1}^{B} \hat{\theta}^*(b) \,/\, B$$

και η εκτίμηση $\widehat{se}_B$ καλείται μη παραμετρική bootstrap εκτίμηση.

Όσο το πλήθος των επαναλήψεων B τείνει στο άπειρο τότε η εκτίμηση $\widehat{se}_B$ συγκλίνει στην ιδανική bootstrap εκτίμηση $se_{\hat{F}_n}(\hat{\theta}^*)$. Αυτό οφείλεται στο ότι η εμπειρική τυπική απόκλιση, όσο αυξάνονται η επαναλήψεις, προσεγγίζει την τυπική απόκλιση των $\hat{\theta}^*$=T(X^*). Ουσιαστικά η εκτίμηση που προκύπτει από τον παραπάνω αλγόριθμο είναι μια *Monte Carlo* προσέγγιση της "ιδανικής bootstrap εκτίμησης" του τυπικού σφάλματος. Όταν το $B \to \infty$, τότε η bootstrap εκτίμηση για το τυπικό σφάλμα προσεγγίζει την ιδανική εκτίμηση. Δηλαδή:

$$\lim_{B\to\infty} \widehat{se}_B = se_{\hat{F}}(\hat{\theta}^*)$$

Όπως σε κάθε εκτίμηση μας ενδιαφέρει να έχουμε μικρή τυπική απόκλιση και μικρή μεροληψία. Αν είχαμε την δυνατότητα να πραγματοποιήσουμε άπειρες επαναλήψεις τότε η εκτίμηση $\widehat{se}_\infty$ θα είχε την μικρότερη τυπική απόκλιση από οποιαδήποτε άλλη αμερόληπτη εκτίμηση. Ένας εμπειρικός κανόνας που ακολουθείται για να προσδιορίσουμε το πλήθος B των επαναλήψεων της μεθόδου (Efron-Tibshirani,1993) είναι ότι αρκούν B=25 bootstrap δείγματα για να έχουμε μια επαρκή πληροφορία, ενώ για B=50 συνήθως έχουμε μια καλή εκτίμηση του $se_F(\hat{\theta})$, αν όμως θέλουμε να κατασκευάσουμε διαστήματα εμπιστοσύνης ή να πραγματοποιήσουμε ελέγχους υποθέσεων προτείνεται να χρησιμοποιούμε B=200 ή μεγαλύτερο. Ωστόσο δεν είναι απαραίτητο ότι όσο μεγαλώνει το B μειώνεται το σφάλμα του αρχικού bootstrap εκτιμητή. Ο Chernick(1999) προτείνει, μιας και η δύναμη των υπολογιστών όσο περνάνε τα χρόνια ενισχύεται να παράγουμε B=1.000 bootstrap δείγματα αν θέλουμε να εκτιμήσουμε το τυπικό σφάλμα ενός εκτιμητή και B=10.000 για διαστήματα εμπιστοσύνης. Ο Hall (1992a) προτείνει να προχωρήσουμε στις εκτιμήσεις κατασκευάζοντας όλα τα δυνατά bootstrap δείγματα στην περίπτωση που το μέγεθος n του δείγματος είναι μικρό (βλ. Ενότητα 1.8). Το πλήθος των δυνατών δειγμάτων ισούται με:

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

Αξίζει να αναφέρουμε ότι η μεταβλητότητα των bootstrap εκτιμητών οφείλεται αφενός στο ότι έχουμε μόνο ένα δείγμα παρατηρήσεων μεγέθους n αντί του πληθυσμού και αφετέρου στο ότι σταματάμε μετά από την κατασκευή B bootstrap δειγμάτων και όχι μετά από άπειρες επαναλήψεις της μεθόδου. Έχουν προταθεί διάφοροι τρόποι υπολογισμού του B, αλλά προς το παρόν θα χρησιμοποιήσουμε, χωρίς απόδειξη, αυτόν που προτείνουν οι Efron-Tibshirani(1993) με βάση τον συντελεστή μεταβλητότητας (coefficient of variation). Γενικά ο συντελεστής μεταβλητότητας μιας τυχαίας μεταβλητής X με κατανομή Fορίζεται ως ο λόγος της τυπικής απόκλισης προς την απόλυτη τιμή της μέσης τιμής:

$$cv_F(X) = \frac{\sigma_F}{|\, \mu_F \,|}$$

Η αυξημένη μεταβλητότητα, λόγω του ότι σταματήσαμε μετά από B επαναλήψεις της μεθόδου και δεν φτάσαμε στο άπειρο αντικατοπτρίζεται στον παρακάτω τύπο:

$$cv(\widehat{se}_B) = \{cv(\widehat{se}_\infty)^2 + \frac{E(\Delta)+2}{4B}\}^{1/2}$$

Το Δ είναι μια παράμετρος σχετική με την κύρτωση της κατανομής της $\hat{\theta}^*$ και είναι μηδέν για την περίπτωση που η κατανομή του $\hat{\theta}^*$ είναι κανονική ή τείνει στο μηδέν στην περίπτωση που το $n \to \infty$. Ακόμα και η ιδανική bootstrap εκτίμηση $\widehat{se}_\infty$ έχει κάποια μεταβλητότητα και αυτό οφείλεται στο ότι δεν χρησιμοποιήσαμε την συνάρτηση κατανομής του πληθυσμού, αλλά μια εκτίμησή της, την $\widehat{F}_n$. Πρακτικά, με βάση τον παραπάνω τύπο ο συντελεστής $cv(\widehat{se}_B)$ δεν θα είναι πολύ μεγαλύτερος του $cv(\widehat{se}_\infty)$, όταν το B παίρνει τιμές μεγαλύτερες του 200.

Ας υποθέσουμε ότι σκοπός μας είναι να προσεγγίσουμε τον αριθμό B θέτοντας ένα επιθυμητό επίπεδο για τον συντελεστή μεταβλητότητας του $\widehat{se}_B$. Έστω ότι θέλουμε το $cv(\widehat{se}_B) = 0.01$. Θα χρησιμοποιήσουμε την περίπτωση που Δ=0. Τότε:

$$cv(\widehat{se}_B) = 0.01$$

$$\sqrt{\frac{2}{4B}} = 0.01,$$

και συνεπώς προκύπτει:

$$B = 5000$$

Τέλος, μιας και χρησιμοποιήσαμε την Monte Carlo εκτίμηση θα ήταν χρήσιμο να αναφερθούμε επιγραμματικά στη μέθοδο. Ας υποθέσουμε ότι έχουμε μια τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας $f_X(x) > 0$, τότε η αναμενόμενη τιμή μιας συνάρτησης g του X θα είναι:

$$E\big(g(X)\big) = \sum_{x \in X} g(x)\, f_X(x),$$

εάν η X είναι διακριτή τυχαία μεταβλητή, ή:

$$E\big(g(X)\big)=\int_{x \in X} g(x)\, f_X(x) dx,$$

εάν η X είναι συνεχής τυχαία μεταβλητή.

Αν λάβουμε ένα δείγμα μεγέθους n από την τυχαία μεταβλητή X, έστω $\big(x_1, x_{2,}, \dots, x_n\big)$ τότε θα έχουμε την Monte Carlo εκτίμηση της $E\big(g(X)\big)$:

$$\hat{g}_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

Αν η τιμή $E\big(g(X)\big)$ υπάρχει, τότε από τον ασθενή νόμο των μεγάλων αριθμών θα έχουμε ότι:

$$\lim_{n\to\infty} P(|\hat{g}_n(x) - E\big(g(X)\big)\, |\geq \varepsilon\,)=0\ ,$$

για οποιοδήποτε αυθαίρετα μικρό ε. Συνεπώς όσο μεγαλώνει το n η πιθανότητα να απέχει η εκτίμηση $\hat{g}_n(x)$ από την $E\big(g(X)\big)$ τείνει στο μηδέν. Το συμπέρασμα μας ενισχύεται και από τον ισχυρό νόμο των μεγάλων αριθμών. Αξίζει να αναφέρουμε ότι οι Monte Carlo εκτιμητές είναι αμερόληπτοι:

$$E\big(\hat{g}_n(x)\big) = E[\tfrac{1}{n}\sum_{i=1}^{n} g(x_i)] = \tfrac{1}{n}\sum_{i=1}^{n} E[g(x_i)] = E(g(X))$$

1.5 Μέθοδος bootstrap και μεροληψία

Έστω ότι έχουμε το τυχαίο δείγμα $\widetilde{\boldsymbol{x}} = (x_1, x_2, \dots, x_n)$ από μία κατανομή F . Εκτιμούμε την πραγματική τιμή της παραμέτρου $\theta = t(\mathrm{F})$ με την στατιστική συνάρτηση $\hat{\theta}=\mathrm{T}(\widetilde{\boldsymbol{x}})$. Θα ορίσουμε ως μεροληψία ή μέσο σφάλμα του εκτιμητή την διαφορά:

$$bias_F(\hat{\theta}\,,\theta) = E(\hat{\theta} - \theta) = E_F[T(\widetilde{\boldsymbol{x}})] - t(F).$$

Στην περίπτωση που η διαφορά αυτή προκύπτει ίση με το μηδέν ο εκτιμητής καλείται αμερόληπτος. Ο υπολογισμός του μέσου σφάλματος είναι χρήσιμος αφενός γιατί μας δίνει μια εικόνα της ακρίβειας του εκτιμητή, αφετέρου διότι με τον υπολογισμό του μπορούμε να βελτιώσουμε την εκτίμηση(bias correction). Με την μέθοδο bootstrap δεν προκύπτουν εν γένει αμερόληπτοι εκτιμητές, ωστόσο στην περίπτωση που αντί της F, χρησιμοποιήσω μια εκτίμησή της $\hat{F}$, τότε το μέσο σφάλμα τείνει να είναι μικρό. Σε αυτή την περίπτωση καλώ την ποσότητα:

$$bias_{\hat{F}} = E_{\hat{F}}[T(X^*)] - t(\hat{F})$$

ιδανική bootstrap εκτίμηση (ideal bootstrap estimate) της μεροληψίας $bias_F$. Σημειώνεται ότι η $t(\hat{F})$ είναι η plug-in εκτίμηση της παραμέτρου θ, η οποία προκύπτει αντικαθιστώντας την F με την $\hat{F}$ και ενδέχεται να διαφέρει από την $\hat{\theta} = T(\widetilde{\boldsymbol{x}})$. Στη θέση της $\hat{F}$ μπορούμε να θεωρήσουμε ότι έχουμε την εμπειρική συνάρτηση κατανομής $\hat{F}_n$. Επιπλέον, όταν η παράμετρος θ είναι η μέση τιμή και η $t(F)$ ο δειγματικός μέσος, τότε η $bias_{\hat{F}}$ =0 ,διότι ο δειγματικός μέσος είναι αμερόληπτος εκτιμητής της μέσης τιμής και συνεπώς $bias_F = 0$. Όπως και στην ενότητα που μελετήσαμε το τυπικό σφάλμα, έτσι κι εδώ θα προσεγγίσουμε την ιδανική bootstrap εκτίμηση της μεροληψίας $bias_{\hat{F}}$ με μια Monte Carlo εκτίμηση. Θα παράγουμε λοιπόν B ανεξάρτητα bootstrap δείγματα μεγέθους n ,τα οποία θα είναι τα $X^{*1}, X^{*2}, \dots, X^{*B} \sim \widehat{F_n}$ και σε κάθε ένα από αυτά θα εφαρμόσουμε τη στατιστική συνάρτηση $T(X)$, λαμβάνοντας τα $\hat{\theta}^*(1) = T(X^{*1}), \hat{\theta}^*(2) = T(X^{*2})$,…, $\hat{\theta}^*(B) = T(X^{*B})$. Στη συνέχεια θα εκτιμήσουμε την αναμενόμενη τιμή $E_{\hat{F}_n}[T(X^{*b})]$ με την ποσότητα:

$$\hat{\theta}^*(.) = \sum_{b=1}^{B} \hat{\theta}^*(b) \,/\, \mathrm{B} = \sum_{b=1}^{B} T(X^{*b})/B\ ,$$

και:

$$\widehat{bias}_B = \hat{\theta}^*(.) - t(\hat{F})$$

Σε αντιστοιχία με τον αλγόριθμο για την εκτίμηση του τυπικού σφάλματος που δόθηκε στην Ενότητα 1.3 παραθέτουμε τον αλγόριθμο για την εκτίμηση της bootstrap εκτίμησης της μεροληψίας :

<u>**Αλγόριθμος 1.2** Εκτίμηση της μεροληψίας</u>

Βήμα 1$^{\text{ο}}$: Διάλεξε B ανεξάρτητα bootstrap δείγματα $X^{*1}, X^{*2}, \dots, X^{*B}$ το καθένα από τα οποία αποτελείται από n στοιχεία, επιλεγμένα τυχαία και με επανάθεση από το αρχικό δείγμα $x_1, x_2, \dots, x_n \sim$ F, και προέρχονται από την $\hat{F}$, η οποία ας είναι η εμπειρική συνάρτηση κατανομής.

Βήμα 2$^{\text{ο}}$: Εφάρμοσε την συνάρτηση T σε κάθε ένα από τα bootstrap δείγματα ως εξής :

$$\hat{\theta}^*(b) = \mathrm{T}(X^{*b})\ ,\ \text{για } b=1,2,\dots,B$$

Βήμα 3$^{\text{ο}}$: Υπολόγισε την αναμενόμενη τιμή $E_{\hat{F}}[T(X^*)]$ από τον μέσο όρο

$$\hat{\theta}^*(.) = \sum_{b=1}^{B} \hat{\theta}^*(b) \,/\, \mathrm{B} = \sum_{b=1}^{B} T(X^{*b})/B\ .$$

Βασιζόμενοι σε B bootstrap δείγματα παίρνουμε μια Monte Carlo προσέγγιση του $bias_{\hat{F}}$, την οποία καλούμε $\widehat{bias}_B$ και ισούται με $\widehat{bias}_B = \hat{\theta}^*(.) - t(\hat{F})$

Αν και η μέθοδος bootstrap χρησιμοποιείται κυρίως σε περιπτώσεις που η ακριβής κατανομή μιας στατιστικής συνάρτησης είναι ιδιαίτερα πολύπλοκη, μια άλλη χρήση της μεθόδου είναι για την περίπτωση που το πλήθος των διαθέσιμων δεδομένων είναι πολύ μικρό. Σε περιπτώσεις που έχουμε μικρά δείγματα και η ακριβής συμπερασματολογία δεν είναι εφικτή η bootstrap επιτρέπει την εξαγωγή συμπερασμάτων ακόμα και όταν το δείγμα δεν είναι αντιπροσωπευτικό του πληθυσμού απ' τον οποίο προήλθε. Ακολουθεί σχετικό παράδειγμα.

Παράδειγμα 1.2

Στη συνέχεια θα παρουσιάσουμε ένα απλό παράδειγμα, στο οποίο θα εφαρμόσουμε την μέθοδο bootstrap με τη βοήθεια του στατιστικού πακέτου R, και θα επιχειρήσουμε να εκτιμήσουμε το τυπικό σφάλμα και τη μεροληψία για τον δειγματικό μέσο. Έστω λοιπόν ότι έχουμε ένα δείγμα κάποιων απόφοιτων μαθηματικών τμημάτων, για τους οποίους γνωρίζουμε τον βαθμό πτυχίου και θέλουμε να εξάγουμε κάποια συμπεράσματα για τον σύνολο των απόφοιτων μαθηματικών. Το δείγμα μας, χάριν ευκολίας θα είναι μικρού μεγέθους και θα εκφράζει τον βαθμό πτυχίου. Ας είναι λοιπόν το μέγεθος δείγματος 5 και συγκεκριμένα ας έχουμε τις παρατηρήσεις:

$$x_1 = 5.5\,, x_2 = 6.3\,, x_3 = 6.9\,, x_4 = 7.8\,, x_5 = 7.1.$$

Επιλέγω B= 10 τυχαία δείγματα μεγέθους 5,από τα $x_1, \dots, x_5$ με επανάθεση. Για το δείγμα $X_1^*, X_2^*, \dots, X_5^*$ καθένα από τα X_i^* παίρνει τις τιμές x_i με πιθανότητα 1/5. Έστω ότι μας ενδιαφέρει να μελετήσουμε την T=$\bar{X}$. Για τα δείγματα που προκύπτουν υπολογίζω την $T^* = \overline{X^*}$ η οποία εκτιμά την $\bar{X}$. Τα αποτελέσματα δίνονται στον πίνακα:

Επανάληψη	Δείγμα	$X_1^*, X_2^*, \dots, X_5^*$	$T^* = \bar{X}^*$
1	X^{*1}	5.5 , 5.5, 7.8, 6.3, 6.3	6.28
2	X^{*2}	7.8, 7.1, 7.1, 5.5, 6.3	6.76
3	X^{*3}	6.9, 7.8, 7.1, 6.3, 6.3	6.88
4	X^{*4}	5.5, 7.8, 5.5, 7.8, 6.9	6.7
5	X^{*5}	6.9, 6.9, 5.5, 7.8, 5.5	6.52
6	X^{*6}	7.1, 5.5, 7.8, 7.8, 6.9	7.02
7	X^{*7}	7.8, 5.5, 5.5, 7.8, 6.3	6.58
8	X^{*8}	7.1, 6.3, 6.9,5.5, 7.8	7.04
9	X^{*9}	5.5, 6.3, 5.5, 6.3, 7.1	6.14
10	X^{*10}	7.1, 5.5 , 5.5, 5.5, 6.3	5.98

Πίνακας 1.2: *Κατασκευή bootstrap δειγμάτων*

Το δείγμα 6.28, 6.76, .6.88, … ,5.98 προέρχεται από την συνάρτηση κατανομής της $\bar{X}^*$και κάθε ένα από τα X_i^* λαμβάνουν κάποια απ’ τις τιμές 5.5 , 6.3 , 6.9 ,7.8 ,7.1 με πιθανότητα $1/_5$.

Μπορούμε να επαναλάβουμε την διαδικασία αρκετές φορές και έτσι μέσω της προσομοίωσης παίρνουμε μία Monte Carlo εκτίμηση της $T^* = \bar{X}^*$ η οποία με τη σειρά της αποτελεί μια bootstrap εκτίμηση της $T = \bar{X}$. Όσο περισσότερες επαναλήψεις έχουμε τόσο θα βελτιώνεται η Monte Carlo εκτίμηση, αυτό ωστόσο δεν ισχύει και για την bootstrap εκτίμηση, διότι αυτή εξαρτάται αποκλειστικά από το αρχικό μέγεθος n του δείγματος (π.χ βλέπε Μπούτσικας 2005).

Στην περίπτωση που θέλουμε να χρησιμοποιήσουμε αρκετά bootstrap δείγματα θα χρησιμοποιήσουμε το στατιστικό πακέτο R. Για χάριν συντομίας θα πληκτρολογήσουμε τον παρακάτω κώδικα, ο οποίος θα μας δώσει μερικά bootstrap δείγματα, ζητώντας μόνο 5 επαναλήψεις.

```
B<-5
x<-c(5.5,6.3,6.9,7.8,7.1)
>output <- matrix(nrow = B, ncol = length(x))
for(i in 1:10){
boot.deigma<-sample(x,size=length(x),replace=T)
output[i,]<-boot.deigma
print(output[i,])}
```

```
[1] 6.9 7.1 6.3 7.1 7.1
[1] 6.9 7.8 5.5 7.1 7.1
[1] 6.3 5.5 7.8 7.1 7.8
[1] 7.1 7.8 6.9 6.9 5.5
[1] 6.9 6.9 5.5 7.1 7.1
```

Εικόνα 1.4 *Κατασκευή 5 δειγμάτων bootstrap*

Στη συνέχεια θα τρέξουμε τον παρακάτω κώδικα για να ζητήσουμε από το πρόγραμμα να μας εκτιμήσει την μέση τιμή με τη μέθοδο bootstrap , μέσα από 1000 δείγματα bootstrap. Η εκτίμηση προκύπτει 6,74.

```
x<-c(5.5,6.3,6.9,7.8,7.1)
B<-numeric(100)
for (i in 1:100) {
B[i]<-mean(sample(x,5,replace=T))
}
B[i]
[1] 6.74
```

Εικόνα 1.5: *Εκτίμηση μέσης τιμής*

Επαναλαμβάνουμε τη διαδικασία κατασκευής bootstrap δειγμάτων για B=200 και B=1000 επαναλήψεις ζητώντας εκτιμήσεις για το τυπικό σφάλμα και τη μεροληψία εκτίμησης της μέσης τιμής με σκοπό να συγκρίνουμε τα αποτελέσματα: Δίνουμε ενδεικτικά τον κώδικα για B=200, και συνοπτικά τα αποτελέσματα και για τις δύο περιπτώσεις:

```
x<-c(5.5,6.3,6.9,7.8,7.1)
B<-200
boot.mean<-function(x,i){mean(x[i])}
b1<-boot(x,statistic=boot.mean,B)
b1
```

- B=200

```
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot (data = x, statistic = boot.mean, R = B)

Bootstrap Statistics :
    original  bias     std. error
t1*     6.72 -0.0141   0.3617423
```

- B=1000

```
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = x, statistic = boot.mean, R = B)

Bootstrap Statistics :
    original  bias    std. error
t1*     6.72 0.01032   0.3465718
```

Εικόνα 1.6

Παρατηρούμε ότι η εκτίμηση με τη μέθοδο bootstrap υποεκτιμά τον πληθυσμιακό μέσο κατά 0.0141 όταν B=200 και τον υπερεκτιμά κατά 0.01032 όταν το B=1000. Η μεροληψία που φαίνεται παραπάνω οφείλεται στο σφάλμα της Monte Carlo εκτίμησης και όχι της μεθόδου bootstrap, διότι η ο δειγματικός μέσος εκτιμά αμερόληπτα τη μέση τιμή.

1.6 Παραμετρική μέθοδος Bootstrap

Στις προηγούμενες ενότητες γνωρίσαμε την μη παραμετρική μέθοδο bootstrap η οποία μας επιτρέπει , μέσω ενός δείγματος παρατηρήσεων $\tilde{\boldsymbol{x}} = (\, x_1, x_2, \dots, x_n)$ να εκτιμήσουμε μια παράμετρο $\theta = t(F)$ του πληθυσμού, χωρίς να κάνουμε καμιά υπόθεση για την κατανομή του και κυρίως χωρίς να είμαστε εξαρτημένοι από περίπλοκους τύπους και δυσεπίλυτες εξισώσεις όπως στην περίπτωση του τυπικού σφάλματος και της μεροληψίας (βλ. ενότητες 1.4 και 1.5). Αν υποθέσουμε λοιπόν ότι το δείγμα των παρατηρήσεων $\tilde{\boldsymbol{x}} = (\, x_1, x_2, \dots, x_n)$ προέρχεται από μια κατανομή F, τότε θα χρησιμοποιήσουμε μια παραμετρική εκτίμηση της F, την οποία θα ονομάσουμε $\hat{F}_{par}$. Στη συνέχεια αντί να πραγματοποιήσουμε δειγματοληψία με επανάθεση, όπως στην περίπτωση της μη παραμετρικής μεθόδου, διαλέγουμε B δείγματα μεγέθους n από την $\hat{F}_{par}$ και σε αυτά υπολογίζουμε την στατιστική συνάρτηση που μας ενδιαφέρει. Το κάθε bootstrap δείγμα θα συμβολίζεται ως εξής:

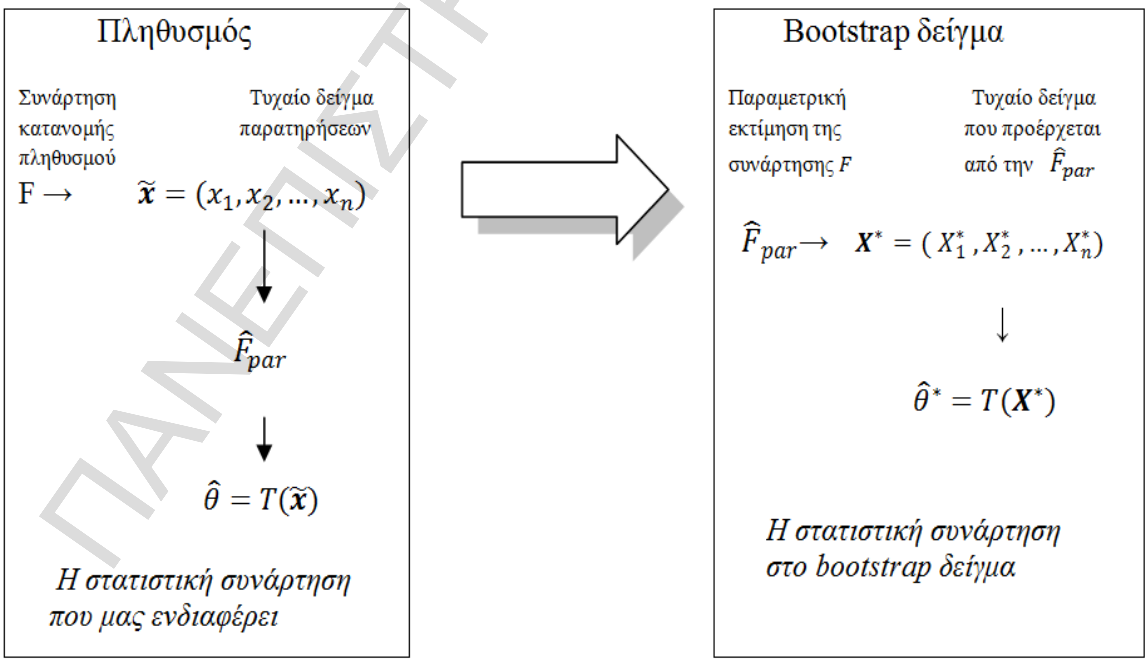
$$\hat{F}_{par} \rightarrow (X_1^*, X_2^*, \dots, X_n^*),$$

ενώ αφού παράγουμε B τέτοια δείγματα θα έχουμε το σύνολο $\boldsymbol{X}^* = (\, X_1^*\,, X_2^*\,, \dots, X_n^*)$.

Αναλυτικότερα, στην παραμετρική μέθοδο θεωρούμε ότι η συνάρτηση κατανομής F ανήκει σε μια παραμετρική οικογένεια κατανομών και λαμβάνουμε την $\hat{F}_{par}$ εκτιμώντας τις παραμέτρούς της από το αρχικό δείγμα $\tilde{\boldsymbol{x}} = (\, x_1, x_2, \dots, x_n)$. Θα συμβολίσουμε την εκτίμηση της παραμέτρου θ με $\hat{\theta}$, η οποία συνήθως, αλλά όχι απαραιτήτως, θα προκύπτει με την μέθοδο της μέγιστης πιθανοφάνειας. Αντικαθιστώντας την θ με την $\hat{\theta}$ λαμβάνουμε το προσαρμοσμένο μοντέλο (fitted model) το οποίο μπορούμε να χρησιμοποιήσουμε για να μελετήσουμε τα χαρακτηριστικά της παραμέτρου θ. Χρησιμοποιούμε το προσαρμοσμένο μοντέλο για να παράγουμε τα δείγματα bootstrap και από εκεί και μετά επαναλαμβάνουμε τα βήματα που συζητήθηκαν και στην μη παραμετρική μέθοδο. Μπορούμε επίσης να χρησιμοποιήσουμε τα bootstrap δείγματα που παρήγαμε για να εκτιμήσουμε τα χαρακτηριστικά της στατιστικής συνάρτησης T που χρησιμοποιήσαμε για την εκτίμηση της παραμέτρου. Δηλαδή αν επιθυμούμε να εκτιμήσουμε το τυπικό σφάλμα ή την μεροληψία εφαρμόζουμε τους αλγόριθμους 1.1 και 1.2 αντίστοιχα (από το Βήμα 2 και μετά). Σημειώνουμε ότι τα δείγματα bootstrap που παράγαμε μας δίνουν την εκτίμηση $\hat{\theta}^*$ της παραμέτρου, η οποία δεν πρέπει να συγχέεται με την $\hat{\theta}$.

Όπως αναφέρει και ο Chernick(2008), αν θεωρήσουμε ότι η F είναι συνεχής και συγκεκριμένα είναι η κανονική κατανομή, μπορούμε να χρησιμοποιήσουμε την $\hat{F}_{par}$, η οποία και αυτή θα είναι κανονική κατανομή και οι παράμετροί της θα είναι οι εκτιμητές μέγιστης πιθανοφάνειας των μ και σ^2 . Έπειτα η δειγματοληψία με επανάθεση από την $\hat{F}_{par}$, οδηγεί σε εκτιμητές που είναι εκτιμητές μέγιστης πιθανοφάνειας και τέλος η Monte Carlo είναι μια προσέγγιση των εκτιμητών αυτών.

Παρουσιάζουμε παρακάτω τη σχηματική αναπαράσταση της μεθόδου:



Εικόνα 1.7: *Παραμετρική μέθοδος bootstrap*

Ίσως φανεί περίεργη στον αναγνώστη η χρήση της μεθόδου όταν μπορούμε να χρησιμοποιήσουμε ένα παραμετρικό μοντέλο και την μέγιστη πιθανοφάνεια για να εκτιμήσουμε μια παράμετρο. Στους Davison και Hinkley (1997) βλέπουμε ότι παραμετρική bootstrap χρησιμοποιείται ως ένας τρόπος για να ελέγξουμε αν το παραμετρικό μοντέλο είναι έγκυρο. Όταν η bootstrap χρησιμοποιείται με μη παραμετρικό τρόπο μας γλυτώνει από τις παραμετρικές υποθέσεις που πρέπει να γίνουν για τον πληθυσμό και όλα τα διαγνωστικά τεστ (π.χ. υποθέσεις κανονικότητας), αλλά όταν χρησιμοποιείται με παραμετρικό τρόπο μας δίνει πιο ακριβή αποτελέσματα συγκριτικά με την χρήση των μαθηματικών τύπων ή ακόμα παρακάμπτει την χρήση τους όταν είναι πολύ δύσκολο να προκύψουν με μαθηματικές πράξεις.

1.7 Μέθοδος jackknife

Έστω ένα τυχαίο δείγμα $\widetilde{\boldsymbol{x}} = x_1, x_2, \ldots, x_n$ και $\hat{\theta} = T(\widetilde{\boldsymbol{x}}) = T\big(x_1, x_{2,\ldots,} x_n\big)$ ο εκτιμητής μιας παραμέτρου θ, για τον οποίο επιθυμούμε να εκτιμήσουμε το τυπικό σφάλμα και την μεροληψία. Η μέθοδος που προτάθηκε από τον Quenouille(1949) βασίζεται στην κατασκευή δειγμάτων που παραλείπουν κάθε φορά από μια παρατήρηση. Το δείγμα x_i ορίζεται ως εξής: $x_{(i)} = (x_1, x_2, \ldots, x_{i-1,} x_{i+1}, \ldots, x_n)$ και ας είναι $\hat{\theta}_{(i)} = T(x_{(i)})$, $i = 1, \ldots, n$, βασισμένη στην i επανάληψη της μεθόδου και σε $n-1$ παρατηρήσεις. Η μέθοδος δηλαδή κατασκευάζει n δυνατά δείγματα $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ τα οποία προέρχονται από το αρχικό δείγμα $x_1, x_{2,} \ldots, x_n$, αφήνοντας μια παρατήρηση κάθε φορά και υπολογίζοντας σε κάθε δείγμα την στατιστική συνάρτηση που μας ενδιαφέρει.

Όπως και στη μέθοδο bootstrap, μπορούμε μιας και δεν γνωρίζουμε από ποια κατανομή προέρχεται το δείγμα, να χρησιμοποιήσουμε την εμπειρική συνάρτηση κατανομής η οποία θα ορίζεται ως $\hat{F}_{(i)}$ και θα είναι η εμπειρική συνάρτηση για τα $n-1$ στοιχεία του δείγματος $x_{(i)}$. Οι παρακάτω τύποι μας δίνουν την δυνατότητα να εκτιμήσουμε, μέσω της μεθόδου jackknife το τυπικό σφάλμα και την μεροληψία που προέκυψε λόγω της χρήσης της εμπειρικής συνάρτησης κατανομής.

Η jackknife εκτίμηση της μεροληψίας δίνεται από τον τύπο:

$$\widehat{bias}_{jack} = (n-1)\hat{\theta}_{(.)} - \hat{\theta} \,,$$

όπου $\hat{\theta}_{(.)}$έχουμε:

$$\hat{\theta}_{(.)} = \sum_{i=1}^{n} \hat{\theta}_{(i)} / n = \bar{T}_n$$

Η εκτίμηση της μεροληψίας μας δίνει την δυνατότητα να διορθώσουμε την εκτίμηση μας, οπότε:

$$T_{jack} = T_n - \widehat{bias}_{jack} = nT_n - (n-1)\bar{T}_n \,,$$

ή αλλιώς :

$$\hat{\theta}_{jack} = n\hat{\theta} - (n-1)\hat{\theta}_{(.)}$$

Το τυπικό σφάλμα δίνεται από τον παρακάτω τύπο:

$$\widehat{se}_{jack} = [\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2]^{1/2}$$

Μπορούμε επίσης να μελετήσουμε την μέθοδο jackknife με τη βοήθεια των ψευδομεταβλητών οι οποίες ορίζονται από τον Tukey (1958) ως εξής:

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\,\hat{\theta}_{(i)}$$

και έχουν τις παρακάτω ιδιότητες:

- Οι ψευδομεταβλητές $\tilde{\theta}_i$,$i = 1, \dots, n$ μπορούν να χρησιμοποιηθούν σαν να είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές.
- Η $\tilde{\theta}_i$ έχει προσεγγιστικά την ίδια διακύμανση με την $\sqrt{n}\hat{\theta}$

Από τα παραπάνω μπορούμε να εκτιμήσουμε την διακύμανση της $\sqrt{n}\hat{\theta}$ από την δειγματική διακύμανση που βασίζεται στα $\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_n$ και κατ' επέκταση και την διακύμανση της $\hat{\theta}$, η οποία και προκύπτει:

$$Var_{jack} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\hat{\theta}_{(j)} - \frac{1}{n}\sum_{j=1}^{n}\hat{\theta}_{(j)})^2.$$

Αυτή η εκτίμηση καλείται jackknife εκτίμησης της διακύμανσης της $\hat{\theta}$.

Στην συνέχεια θα δώσουμε ένα απλό παράδειγμα κατασκευής jackknife δειγμάτων με την βοήθεια της R.

Παράδειγμα 1.3

Θα επανέλθουμε στα δεδομένα του παραδείγματος 1.2 και από τις αρχικές παρατηρήσεις $x_1 = 5.5$, $x_2 = 6.3$, $x_3 = 6.9$, $x_4 = 7.8$, $x_5 = 7.1$ θα κατασκευάσουμε όλα τα jackknife δείγματα, εφαρμόζοντας τη μέθοδο που παρουσιάζεται παραπάνω και επιδιώκοντας να μελετήσουμε την διάμεσο.

```
x<-c(5.5,6.3,6.9,7.8,7.1)
output<-matrix(0,nrow=5, ncol=4)
jack <- numeric(4)
for (i in 1:length(x))
{ for (j in 1:length(x))
{if(j < i) jack[j] <- x[j] else if(j > i) jack[j-1] <- x[j]}
output[i,]<-jack
print(output[i,])}
```

$x_{(1)}$	6.3	6.9	7.8	7.1
$x_{(2)}$	5.5	6.9	7.8	7.1
$x_{(3)}$	5.5	6.3	7.8	7.1
$x_{(4)}$	5.5	6.3	6.9	7.1
$x_{(5)}$	5.5	6.3	6.9	7.8

Εικόνα1.8 *Παραπάνω φαίνονται όλα τα jackknife δείγματα* $\boldsymbol{x}_{(i)}$

Σε αυτό το παράδειγμα η παράμετρος θ που μας ενδιαφέρει είναι η διάμεσος. Είναι προφανές ότι η δειγματική διάμεσος αντιστοιχεί στην τιμή 6.9. Θα παρουσιάσουμε την εκτίμηση $\hat{\theta}_{jack}$ της παραμέτρου.

```
x<-c(5.5, 6.3, 6.9,7.8,7.1)
diamesos<- function(x) median(x)
jack <- numeric(length(x)-1)
pseudo<-numeric(length(x))
for (i in 1:length(x))
{ for (j in 1:length(x))
{if(j < i) jack[j] <- x[j] else if(j > i) jack[j-1] <- x[j]}
pseudo[i]<-length(x)*diamesos(x) -(length(x)-1)*diamesos(jack)
print(pseudo[i])}
mean(pseudo)
```

$\tilde{\theta}_1$	6.5
$\tilde{\theta}_2$	6.5
$\tilde{\theta}_3$	7.7
$\tilde{\theta}_4$	8.1
$\tilde{\theta}_5$	8.1

$\hat{\theta}_{jack} = 7.38$

Εικόνα 1.9 **Εκτίμηση της διαμέσου με τη μέθοδο jackknife**

Δημιουργείται το ερώτημα ποια από τις μεθόδους bootstrap ή jackknife είναι καταλληλότερη για την εκτίμηση μιας παραμέτρου θ . Είναι σαφές ότι είναι ευκολότερο να χρησιμοποιήσουμε την μέθοδο jackknife γιατί αρκούν n δυνατά δείγματα, ενώ με την μέθοδο bootstrap θα χρειαστεί να κατασκευάσουμε εκατοντάδες δείγματα. Ουσιαστικά όμως η μέθοδος jackknife είναι μια προσέγγιση της μεθόδου bootstrap. Όπως αναφέρεται στην βιβλιογραφία(Efron & Tibshirani 1993) εάν η εκτίμηση $\hat{\theta} = T(\tilde{\boldsymbol{x}})$ είναι μια γραμμική συνάρτηση (μπορεί δηλαδή να γραφεί στην μορφή $\hat{\theta} = T(\tilde{\boldsymbol{x}}) = \mu + \frac{1}{2}\sum_1^n \alpha(x_i)$), τότε αποδεικνύεται ότι η jackknife εκτίμηση για το τυπικό σφάλμα συμφωνεί με την εκτίμηση που προκύπτει από την μέθοδο bootstrap, με εξαίρεση τον συντελεστή $\{\frac{(n-1)}{n}\}^{1/2}$ που βρίσκουμε στην περίπτωση της jackknife. Για γραμμικές λοιπόν συναρτήσεις, με την μέθοδο jackknife δεν υπάρχει απώλεια πληροφορίας, παρόλο που τα δείγματα που χρησιμοποιούμε είναι σε πλήθος λιγότερα, κάτι το οποίο δεν ισχύει για τις μη γραμμικές συναρτήσεις.

Για να συνοψίσουμε, η μέθοδος Jackknife συχνά αποτελεί μια καλή προσέγγιση της bootstrap εκτίμησης για το τυπικό σφάλμα και τη μεροληψία αρκεί η $\hat{\theta} = T(\tilde{\boldsymbol{x}})$ να είναι ομαλή (smooth) συνάρτηση. Για παράδειγμα, η μέθοδος jackknife δεν θα έδινε ικανοποιητικά αποτελέσματα στην περίπτωση που η παράμετρος που μας ενδιέφερε ήταν η διάμεσος και αυτό γιατί η διάμεσος δεν είναι ομαλή συνάρτηση. Αρκεί να λάβουμε υπ’ όψιν ότι τα δείγματα που προκύπτουν με την εν λόγω μέθοδο έχουν μεγάλο ποσοστό ομοιότητας με τα αρχικά δεδομένα˙ σε κάθε jackknife δείγμα λείπει μονάχα μια παρατήρηση των αρχικών δεδομένων. Ως εκ τούτου, μικρές αλλαγές στα αρχικά δεδομένα μπορεί να οδηγήσουν σε τελείως διαφορετικά jackknife δείγματα, κάτι το οποίο συνεπάγεται μεγάλες αλλαγές στον υπολογισμό της $\hat{\theta}_{(i)} = T(x_{(i)})$. Αντίθετα, τα δείγματα που κατασκευάζουμε ακολουθώντας την μέθοδο bootstrap δεν είναι τόσο όμοια με το αρχικό σετ δεδομένων και έτσι η bootstrap κρίνεται προτιμότερη από την jackknife σε αυτήν την περίπτωση.

1.8 Πότε αποτυγχάνει η μέθοδος bootstrap;

Η μέθοδος δεν εφαρμόζεται το ίδιο αποτελεσματικά σε όλες τις περιπτώσεις. Θα δούμε στη συνέχεια τέτοια παραδείγματα, ωστόσο καλό είναι να λάβουμε υπ' όψιν ότι ακόμα και για αυτές τις περιπτώσεις έχουν προταθεί αλγόριθμοι ή παραλλαγές της κλασικής μεθόδου που βελτιώνουν τα αποτελέσματα (Chernick 1999 & Davison and Hinkley 1997). Η εφαρμογή της μεθόδου αποτυγχάνει όταν έχουμε εξαρτημένα δεδομένα και αυτό διότι η ανεξαρτησία προϋποθέτει ότι η από κοινού συνάρτηση κατανομής των x_i μπορεί να γραφεί ως γινόμενο των περιθωρίων κατανομών τους, κάτι που προφανώς δεν ισχύει αν έχουμε εξαρτημένα δεδομένα. Στην περίπτωση που υπάρχει ασθενής εξάρτηση η μη παραμετρική μέθοδος δίνει ικανοποιητικά αποτελέσματα. Γενικότερα αν έχουμε κάποια πληροφορία για την κατανομή F του αρχικού δείγματος ,καλό είναι να ακολουθείται η παραμετρική μέθοδος και όχι η μη παραμετρική. Αν γνωρίζουμε ότι η τα δεδομένα προέρχονται από μια συνεχή κατανομή δεν είναι ωφέλιμο να προσπαθούμε να την προσεγγίσουμε με την εμπειρική συνάρτηση κατανομής η οποία είναι διακριτή.

Δυσκολία στην εφαρμογή της μεθόδου εντοπίζεται και στην περίπτωση που έχουμε ελλιπή δεδομένα η έκτροπες παρατηρήσεις. Σε κάθε περίπτωση υποθέτουμε έχουμε ένα αρχικό δείγμα παρατηρήσεων $x_1, x_2, \dots, x_n$ και πάνω σε αυτό βασίζουμε την μέθοδο και τα συμπεράσματα μας. Υπάρχει ωστόσο το ενδεχόμενο το αρχικό μας δείγμα να μην είναι ολοκληρωμένο, όπως για παράδειγμα αν έχουμε επαναλαμβανόμενες μετρήσεις και κάποια μέτρηση απουσιάζει. Από την άλλη όταν έχουμε έκτροπες παρατηρήσεις και εφαρμόζουμε την bootstrap είναι χρήσιμο να ελέγξουμε το αποτέλεσμα της προσομοίωσης και να δούμε αν τα αποτελέσματα εξαρτώνται από συγκεκριμένες παρατηρήσεις.

Κλασική περίπτωση στην οποία η bootstrap δεν ανταποκρίνεται ικανοποιητικά είναι στην περίπτωση των ακραίων τιμών (Athreya 1987). Για παράδειγμα εάν θέλουμε να εκτιμήσουμε την παράμετρο θ από ένα δείγμα που ακολουθεί την ομοιόμορφη κατανομή $U(0, \theta)$, τότε η μέθοδος αποτυγχάνει διότι η κατανομή bootstrap δεν προσεγγίζει ικανοποιητικά την δειγματική κατανομή. Επιπλέον δεδομένα που προέρχονται από κατανομές με βαριές ουρές (Κεφάλαιο 4^{ο}), όπου η πιθανότητα ακραίων παρατηρήσεων είναι μεγάλη, επηρεάζουν την ποιότητα της μη παραμετρικής bootstrap, ειδικά όταν η παράμετρος που μας ενδιαφέρει είναι η μέση τιμή.

Τέλος όπως έχει αναφερθεί ήδη, έχει προταθεί μια διαφορετική προσέγγιση όταν το μέγεθος του δείγματος είναι πολύ μικρό από τον Hall (1992a). Το πρόβλημα έγκειται στο ότι με μικρό μέγεθος δείγματος, λόγω της επανάθεσης πολλές παρατηρήσεις θα επαναλαμβάνονται και πολλά bootstrap δείγματα θα ταυτίζονται, ακόμα και αν επιδιώξουμε να κάνουμε πολλές επαναλήψεις για να λάβουμε την Monte Carlo εκτίμηση. Επιπλέον στη περίπτωση του μικρού δείγματος η εμπειρική συνάρτηση συνήθως δεν είναι μια καλή εκτίμηση της κατανομής του πληθυσμού. Έτσι επιλέγουμε να χρησιμοποιήσουμε όλα τα δυνατά δείγματα μεγέθους n που λαμβάνονται με επανάθεση από το αρχικό δείγμα και είναι $\binom{2n-1}{n}$ στο πλήθος. Ακόμα και για $n = 10$ ο παραπάνω αριθμός είναι αρκετά μεγάλος (92.378) . Για $n = 20$ και B=2000 επαναλήψεις την μεθόδου Bootstrap, ο Hall(1992a) αποδεικνύει ότι η πιθανότητα να μην επανεμφανιστεί κάποιο bootstrap δείγμα είναι μεγαλύτερη από 0,95.

Κεφάλαιο 2ο

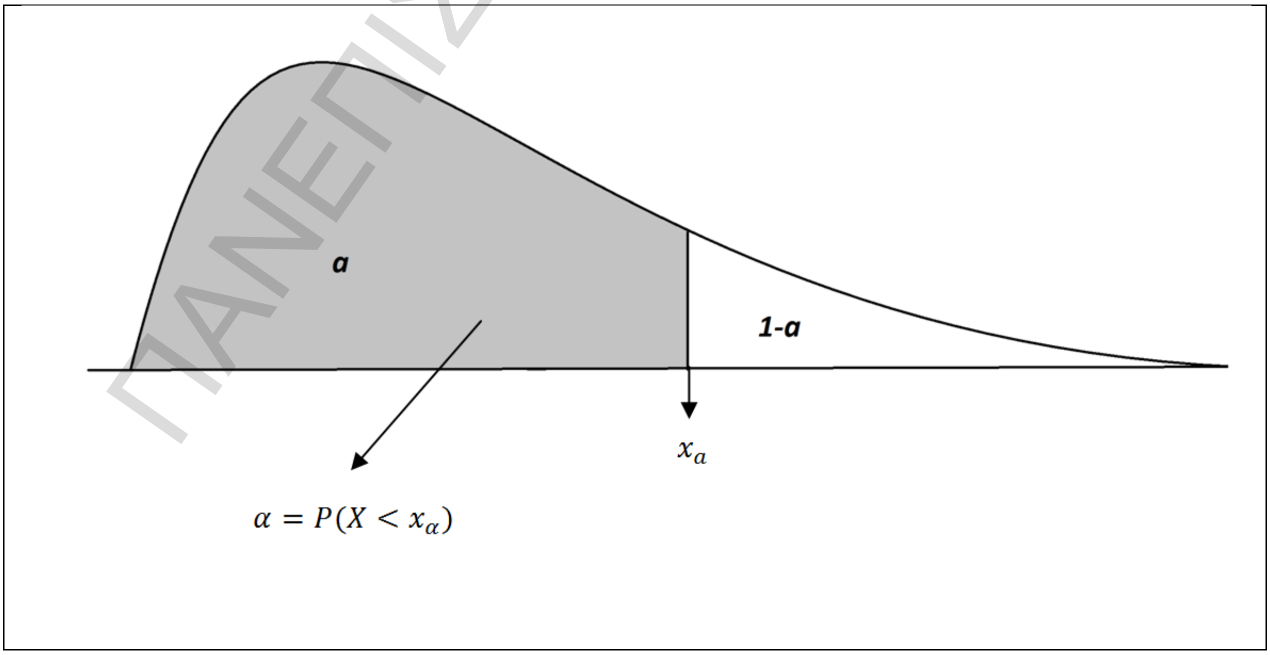
2.1 Ποσοστιαία Σημεία

Σκοπός μας σε αυτό το κεφάλαιο είναι να χρησιμοποιήσουμε την μέθοδο bootstrap ώστε να εκτιμήσουμε τα ποσοστιαία σημεία της κατανομής μιας τυχαίας μεταβλητής X. Δίνουμε τον ορισμό του ποσοστιαίου σημείου:

Ορισμός 2.1 (Ποσοστιαίο σημείο): Έστω $\alpha\epsilon(0,1)$. Το σημείο x_a είναι ένα α-ποσοστιαίο σημείο της κατανομής της τυχαίας μεταβλητής X εάν:

$$P(X < x_a) = a$$

Στην παρακάτω εικόνα βλέπουμε την συνάρτηση πυκνότητας πιθανότητας μια συνεχούς τυχαίας μεταβλητής η οποία έχει α-ποσοστιαίο σημείο x_a



Εικόνα 2.1

Όπως φαίνεται και στην Εικόνα 2.1, το εμβαδόν κάτω από την καμπύλη πυκνότητας και αριστερά του x_a είναι α, ενώ δεξιά του x_a ισούται με 1- α. Εάν η F είναι συνεχής και γνησίως αύξουσα τότε ισχύει ότι το α-ποσοστιαίο σημείο είναι η λύση της $F(x_a) = a$, δηλαδή $x_a = F^{-1}(a)$. Είναι επίσης γνωστό ότι το 25^{o} ποσοστημόριο συμβολίζεται με $x_{0,25}$ και είναι το πρώτο τεταρτημόριο. Επιπλέον, το 50^{o} ποσοστιαίο σημείο εκφράζει τη διάμεσο ή αλλιώς το 2^{o} τεταρτημόριο, ενώ το 75^{o} είναι το τρίτο τεταρτημόριο. Για τις διακριτές κατανομές το α-ποσοστιαίο σημείο δεν υπάρχει για κάθε $a\epsilon(0,1)$, αλλά ακόμα και όταν υπάρχει δεν ορίζεται μονοσήμαντα από τον Ορισμό 2.1. Για τις συνηθισμένες κατανομές τα ποσοστιαία σημεία έχουν συγκεκριμένο συμβολισμό, όπως για παράδειγμα το α-ποσοστιαίο σημείο της τυπικής κανονικής κατανομής $N(0,1)$ συμβολίζεται με z_{1-a}.

2.2 Εισαγωγή στα διαστήματα εμπιστοσύνης

Σε αρκετές περιπτώσεις δεν αρκεί για την εκτίμηση μιας παραμέτρου θ να βρούμε μια κατάλληλη στατιστική συνάρτηση $T(X)$, όπου $X = (X_1, X_{2,\ldots,}X_{n)}$, και για δεδομένες παρατηρήσεις $\widetilde{\boldsymbol{x}} = (x_1, x_2, \ldots, x_n)$ να δώσουμε μια σημειακή εκτίμηση. Πρέπει να δοθεί ταυτόχρονα η ακρίβεια της εκτίμησης και το πιθανό σφάλμα της. Έτσι λοιπόν χρησιμοποιούμε τα διαστήματα εμπιστοσύνης, διαστήματα δηλαδή που περιέχουν την εκτιμώμενη παράμετρο με δεδομένη πιθανότητα. Θα λέμε ότι το 100(1-α)% διάστημα εμπιστοσύνης για μια παράμετρο ενός πληθυσμού, είναι ένα διάστημα που υπολογίζεται από ένα τυχαίο δείγμα από τον πληθυσμό και έχει πιθανότητα 1-α να περιέχει την πραγματική τιμή της παραμέτρου. Η πιθανότητα 1-α θα λέγεται συντελεστής εμπιστοσύνης του διαστήματος. Τα πιο συνηθισμένα επίπεδα εμπιστοσύνης είναι το 95% και το 99%.

Ορισμός 2.2 (Διάστημα εμπιστοσύνης): Εάν $T_1(X)$ και $T_2(X)$ είναι στατιστικές συναρτήσεις με $T_1(X) < T_2(X)$, τότε το τυχαίο διάστημα $[T_1(X), T_2(X)]$ λέγεται διάστημα εμπιστοσύνης για την παράμετρο θ με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$ εάν:

$$P_\theta(T_1(X) \leq \theta \leq T_2(X)) = 1 - \alpha$$

Σε ορισμένες περιπτώσεις , ιδιαίτερα σε διακριτές κατανομές δεν είναι πάντα εφικτό να βρούμε ένα διάστημα $[T_1(X), T_2(X)]$ για το οποίο η παραπάνω πιθανότητα να είναι ακριβώς $1-\alpha$. Αυτό που επιδιώκουμε τότε είναι η παραπάνω πιθανότητα να είναι τουλάχιστον $1-\alpha$. Με άλλα λόγια ζητάμε :

$$P_\theta(T_1(X) \leq \theta \leq T_2(X)) \geq 1-\alpha$$

Επιπλέον για την περίπτωση που η παραπάνω πιθανότητα είναι :

$$P_\theta(T_1(X) \leq \theta \leq T_2(X)) \approx 1-\alpha, \quad ή \ \lim_{n\to\infty} P_\theta(T_1(X) \leq \theta \leq T_2(X))=1-\alpha,$$

το διάστημα $[T_1(X), T_2(X)]$ θα λέγεται αντίστοιχα 100(1-α)% προσεγγιστικό ή ασυμπτωτικό διάστημα εμπιστοσύνης για την παράμετρο θ. Ας αναφέρουμε επίσης ότι δεν μας ενδιαφέρει πάντα το διάστημα εμπιστοσύνης, αλλά ένα κάτω και ένα άνω φράγμα για την παράμετρο θ. Ένα κάτω φράγμα λοιπόν με συντελεστή εμπιστοσύνης 100(1-α)% ορίζεται από την σχέση:

$$P_\theta(T_1(X) \leq \theta)=1-\alpha,$$

ενώ αντίστοιχα η σχέση:

$$P_\theta(\theta \leq T_2(X))=1-\alpha,$$

θα ορίζει το άνω φράγμα με συντελεστή εμπιστοσύνης 100(1-α)% .

Η διαδικασία για να υπολογίσουμε ένα διάστημα εμπιστοσύνης περιγράφεται από τα παρακάτω βήματα (π.χ. βλέπε Κουρούκλης 2006):

1. Προσπαθούμε να βρούμε μια τυχαία μεταβλητή $T(X,\theta)$ της οποίας η τιμή ενδέχεται να εξαρτάται από θ, αλλά η κατανομή της δεν εξαρτάται από το θ. Αυτή η τυχαία μεταβλητή θα λέγεται ποσότητα οδηγός (pivotal quantity).
2. Προσδιορίζουμε τις σταθερές $c_1 < c_2$, οι οποίες εξαρτώνται από την κατανομή της T και την επιλογή του α, έτσι ώστε:

   $$P_\theta(c_1 \leq T(X,\theta) \leq c_2) = 1-\alpha$$
3. Στη συνέχεια λύνουμε την διπλή ανισότητα και εφόσον είναι εφικτό καταλήγουμε στην εξής σχέση:

   $$P_\theta(T_1(X) \leq \theta \leq T_2(X)) \geq 1-\alpha$$

   Τότε το διάστημα $[T_1(X), T_2(X)]$ είναι διάστημα εμπιστοσύνης για την θ με συντελεστή εμπιστοσύνης 100(1-α)%.

Ας υποθέσουμε τώρα ότι η εκτιμήτρια $\hat{\theta}$ κατανέμεται κανονικά με παραμέτρους θ και se^2, δηλαδή $\hat{\theta} \sim N(\theta, se^2)$, όπου η θ είναι άγνωστη, ενώ το τυπικό σφάλμα είναι γνωστή ποσότητα. Η τυχαία ποσότητα Z που ορίζεται παρακάτω θα ακολουθεί την τυποποιημένη κανονική κατανομή:

$$Z = \frac{\hat{\theta} - \theta}{se} \sim \mathrm{N}(0,1)$$

Αν συμβολίσουμε με z_α το $100 \cdot \alpha$ ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής, τότε θα ισχύει ότι:

$$P_\theta(z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{se} \leq z_{1-a/2}) = 1 - a$$

$$P_\theta\{z_{\alpha/2} \cdot se - \hat{\theta} \leq -\theta \leq z_{1-a/2} \cdot se - \hat{\theta}\} = 1 - a$$

$$P_\theta\{-z_{\alpha/2} \cdot se + \hat{\theta} \geq \theta \geq -z_{1-a/2} \cdot se + \hat{\theta}\} = 1 - a$$

$$P_\theta\{\theta \in [\hat{\theta} - z_{1-a/2} \cdot se, \hat{\theta} - z_{\alpha/2} \cdot se]\} = 1\text{-}\alpha$$

Θα συμβολίσουμε στη συνέχεια ως $\hat{\theta}_{low} = \hat{\theta} - z_{1-a/2} \cdot se$ το κάτω φράγμα εμπιστοσύνης με συντελεστή εμπιστοσύνης $1 - \alpha/2$ και ως $\hat{\theta}_{up} = \hat{\theta} - z_{a/2} \cdot se$ το άνω φράγμα εμπιστοσύνης με συντελεστή εμπιστοσύνης $1 - \alpha/2$, οπότε μπορούμε να εκφράσουμε το παραπάνω διάστημα εμπιστοσύνης όπως παρακάτω:

$$[\hat{\theta} - z_{1-a/2} \cdot se, \hat{\theta} - z_{\alpha/2} \cdot se] = [\hat{\theta}_{low}, \hat{\theta}_{up}]$$

Το παραπάνω διάστημα εμπιστοσύνης καλείται διάστημα εμπιστοσύνης ίσων ουρών (equal tailed) διότι όπως παρατηρούμε το άνω και κάτω φράγμα έχουν τον ίδιο συντελεστή εμπιστοσύνης.

2.3 Το bootstrap διάστημα εμπιστοσύνης

Θα υποθέσουμε και πάλι ότι έχουμε τις παρατηρήσεις $\tilde{\boldsymbol{x}} = (x_1, x_2, \ldots, x_n)$ οι οποίες προήλθαν κατόπιν τυχαίας δειγματοληψίας από την άγνωστη κατανομή F. Ας είναι $\theta = t(F)$ η παράμετρος που μας ενδιαφέρει, $\hat{\theta}$ η plug-in εκτίμησή της, $T = T(X_1, X_2, \dots, X_n)$ στατιστική συνάρτηση που χρησιμοποιήσαμε για την εκτίμηση της παραμέτρου και $\widehat{se}$ μια εκτίμηση για το τυπικό σφάλμα της $\hat{\theta}$, η οποία προήλθε με τη μέθοδο bootstrap. Σε αρκετές περιπτώσεις και ενώ το μέγεθος n του δείγματος μεγαλώνει τότε η κατανομή της $\hat{\theta}$ προσεγγίζει όλο και πιο πολύ την κανονική κατανομή με μέση τιμή κοντά στο θ και διακύμανση κοντά στο $\widehat{se}^2$. Δηλαδή:

$$\hat{\theta} \dot{\sim} N(\theta, \widehat{se}^2), \text{ ή ισοδύναμα } \frac{\hat{\theta}-\theta}{\widehat{se}} \dot{\sim} N(0,1)$$

Η ποσότητα $\frac{\hat{\theta}-\theta}{\widehat{se}}$ είναι η ποσότητα οδηγός και το διάστημα εμπιστοσύνης με 1-α πιθανότητα κάλυψης (coverage probability) ή 100(1-α)% συντελεστή εμπιστοσύνης προκύπτει:

$$[\hat{\theta} - z_{1-a/2} \cdot \widehat{se}\, , \hat{\theta} - z_{\alpha/2} \cdot \widehat{se}]\, ,$$

ή επειδή $z_a = -z_{1-\alpha}$ το ίδιο διάστημα εμπιστοσύνης μπορεί να συμβολιστεί και ως:

$$\hat{\theta} \pm z_{1-a/2} \cdot \widehat{se}$$

Ουσιαστικά το παραπάνω διάστημα, το οποίο χρησιμοποιεί την προσέγγιση της τυποποιημένης κανονικής κατανομής καλείται κανονικοποιημένο (standardized) και είναι ένα προσεγγιστικό διάστημα διότι η πιθανότητα κάλυψης στις περισσότερες περιπτώσεις δεν είναι ακριβώς 1-α και όμοια θα είναι προσεγγιστικά και τα διαστήματα για τα οποία θα μιλήσουμε στη συνέχεια. Ωστόσο η παραπάνω διαδικασία που περιγράψαμε δεν είναι πάντα εφικτή διότι συχνά δεν είναι εύκολο να βρούμε μια ποσότητα οδηγό, της οποίας η κατανομή να μην εξαρτάται από την παράμετρο αλλά να είναι γνωστή και αρκετές φορές δεν μπορούμε να βασιστούμε σε υποθέσεις κανονικότητας. Θα χρησιμοποιήσουμε την ιδέα της μεθόδου bootstrap για να κατασκευάσουμε τα διαστήματα εμπιστοσύνης. Αναλυτικότερα, θα πρέπει να χρησιμοποιήσουμε τα ποσοστιαία σημεία της κατανομής της $\hat{\theta} - \theta$, όπου $\hat{\theta} = T(X_1, \dots, X_n)$. Παραθέτουμε την διαδικασία όπως αναφέρεται στην βιβλιογραφία (π.χ. βλέπε Μπούτσικας 2008):

Αρκεί να εκτιμήσουμε τα σημεία $c_{a/2}$ και $c_{1-a/2}$, έτσι ώστε να ισχύει:

$$P(\hat{\theta} - \theta < c_{a/2}) = \alpha/2 \text{ και } P(\hat{\theta} - \theta > c_{1-a/2}) = 1 - \alpha/2.$$

Τότε θα έχουμε ότι $P(c_{1-a/2} < \hat{\theta} - \theta < c_{a/2}) = 1 - \alpha$ και το διάστημα $(\hat{\theta} - c_{a/2},\ \hat{\theta} - c_{1-a/2})$ θα είναι ένα διάστημα εμπιστοσύνης για την παράμετρο *θ,* με συντελεστή εμπιστοσύνης 1-α. Θεωρούμε ότι η παράμετρος θ είναι συνάρτηση μιας κατανομής F και ότι η εκτίμηση $\hat{\theta} = T$, ακολουθεί μια κατανομή, την F_T. Οπότε έχουμε:

$$P\left(\hat{\theta} - \theta_F > c_{\frac{a}{2}}\right) = \frac{a}{2} \Leftrightarrow F_T\left(c_{\frac{a}{2}} + \theta_F\right) = 1 - \frac{a}{2} \Leftrightarrow c_{\frac{a}{2}} = F_T^{-1}\left(1 - \frac{a}{2}\right) - \theta_F.$$

Αντικαθιστούμε την F με την $\hat{F}$ και επομένως την T με την T^*, οπότε θα είναι:

$$\hat{c}_{a/2} = F_{T^*}^{-1}\left(1 - \frac{a}{2}\right) - \theta_{\hat{F}},$$

όπου $T^* = T(X_1^*, \dots, X_n^*)$. Επειδή είναι δύσκολος ο υπολογισμός της $F_{T^*}^{-1}\left(1 - \frac{a}{2}\right)$, χρησιμοποιούμε και πάλι προσομοίωση. Παράγουμε $T_1^*, T_2^*, \dots, T_B^*$, τυχαίους αριθμούς και τότε η F_{T^*} προσεγγίζεται από την εμπειρική συνάρτηση κατανομής:

$$\hat{F}_{T^*} = \frac{1}{B}\sum_{i=1}^{B} I(T_i^* < x).$$

Διατάσσοντας τα T_i^*, από το μικρότερο στο μεγαλύτερο, έχουμε το σύνολο των $T_{(i)}^*$, έχουμε ότι:

$$\hat{c}_{a/2} = T^*_{\left(B\left(1-\frac{\alpha}{2}\right)\right)} - \theta_{\hat{F}} \qquad \text{και} \qquad \hat{c}_{1-a/2} = T^*_{(B \cdot \alpha/2)} - \theta_{\hat{F}}.$$

Το αντίστοιχο δ.ε. θα είναι:

$$[\hat{\theta} - (T^*_{\left(B\left(1-\frac{\alpha}{2}\right)\right)} - \theta_{\hat{F}}),\ \hat{\theta} - (T^*_{\left(B \cdot \frac{\alpha}{2}\right)} - \theta_{\hat{F}})].$$

Αν μάλιστα έχουμε θέσει ότι $\theta_{\hat{F}} = \hat{\theta}$, τότε το διάστημα εμπιστοσύνης θα είναι ίσο με:

$$[2\hat{\theta} - T^*_{\left(B\left(1-\frac{\alpha}{2}\right)\right)}, 2\hat{\theta} - T^*_{\left(B \cdot \frac{\alpha}{2}\right)}].$$

Το παραπάνω δ.ε. καλείται βασικό bootstrap διάστημα εμπιστοσύνης.

2.4 Bootstrap-t

Η μέθοδος bootstrap –t βασίζεται στη ποσότητα οδηγό $Z = \frac{\hat{\theta}-\theta}{\widehat{se}}$, όπου $\hat{\theta}$ είναι μια εκτίμηση για την παράμετρο θ και $\widehat{se}$ μια εκτίμηση για το τυπικό σφάλμα του $\hat{\theta}$. Η Z για $n \to \infty$ ακολουθεί την τυποποιημένη κανονική ενώ αν το $\hat{\theta} = \bar{x}$, η Z ακολουθεί την κατανομή t του Student με n-1 βαθμούς ελευθερίας (εισήχθη από τον Gosset 1908). Σε αυτή τη περίπτωση το αντίστοιχο 1-2α διάστημα εμπιστοσύνης, για την παράμετρο θ θα είναι το εξής:

$$[\hat{\theta} - t_{(1-a,n-1)} \cdot \widehat{se} , \hat{\theta} - t_{(a,n-1)} \cdot \widehat{se}] ,$$

Ωστόσο με τη μέθοδο bootstrap μπορούμε να εκτιμήσουμε την κατανομή της Z απευθείας από τα δεδομένα. Υποθέτουμε ότι η εκτίμηση της θ είναι η $\hat{\theta} = T(\tilde{\boldsymbol{x}})$. Από ένα λοιπόν αρχικό σύνολο δεδομένων $\tilde{\boldsymbol{x}} = (x_1, x_2, \dots, x_n)$ το οποίο προέρχεται από μια άγνωστη κατανομή F, κατασκευάζουμε B bootstrap δείγματα μεγέθους n πραγματοποιώντας δειγματοληψία με επανάθεση. Σε κάθε ένα από τα δείγματα $X^{1*}, X^{2*}, \dots, X^{*B}$ υπολογίζουμε την στατιστική συνάρτηση:

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{\widehat{se}^*(b)},$$

όπου $\hat{\theta}^*(b) = T(X^{*b})$ είναι η τιμή της στατιστικής συνάρτησης T για το b bootstrap δείγμα και όπου $\widehat{se}^*(b)$ είναι η εκτίμηση του τυπικού σφάλματος του $\hat{\theta}^*(b)$ για το b bootstrap δείγμα.

Όπως αναφέρεται στους Efron-Tibshirani(1993) το α-ποσοστιαίο σημείο της κατανομής της $Z^*(b)$ υπολογίζεται για την τιμή $\hat{t}_a$ για την οποία ισχύει:

$$\#\{Z^*(b) \leq \hat{t}_a\}/B = \alpha$$

Για παράδειγμα, εάν το B=1000 τότε το 5ο ποσοστημόριο θα είναι η $50^{η}$ μεγαλύτερη παρατήρηση των $Z^*(b)$ και ομοίως το 95^{o} ποσοστημόριο θα είναι η $950^{η}$ μεγαλύτερη τιμή των τιμών της $Z^*(b)$. Επομένως το 100(1-α)% διάστημα εμπιστοσύνης θα είναι το :

$$[\hat{\theta} - \hat{t}_{1-a/2} \cdot \widehat{se} , \hat{\theta} - \hat{t}_{a/2} \cdot \widehat{se}$$

Δημιουργείται πρόβλημα στην περίπτωση που το $B \cdot \alpha/2$ δεν είναι ακέραιος. Σε αυτήν την περίπτωση υποθέτοντας ότι το $\alpha \leq 0.05$ προτείνεται να χρησιμοποιήσουμε το ακέραιο μέρος του γινομένου $B \cdot \alpha/2$ και να υπολογίσουμε τα αντίστοιχα ποσοστιαία σημεία.

Η bootstrap-t θεωρείται εύκολη μέθοδος, ωστόσο είναι πολύ ευαίσθητη στην περίπτωση που έχουμε έκτροπες παρατηρήσεις. Για μεγάλα δείγματα έχει αποδειχτεί ότι η bootstrap –t τείνει να είναι πιο κοντά στον επιθυμητό συντελεστή εμπιστοσύνης σε σχέση με το βασικό bootstrap διάστημα εμπιστοσύνης ή με το διάστημα που προκύπτει χρησιμοποιώντας τα ποσοστιαία σημεία της κατανομής $Student - t$. Ωστόσο όπως αναφέρεται και στην βιβλιογραφία (Efron-Tibshirani,1993), μπορεί να κερδίζουμε σε ακρίβεια με την μέθοδο bootstrap-t χάνουμε όμως όσον αφορά την γενίκευση. Ενώ ο πίνακας με τα ποσοστιαία σημεία της τυποποιημένης κανονικής ισχύει για όλα τα δείγματα, ανεξαρτήτου μεγέθους, ο πίνακας της κατανομής t ισχύει για όλα τα δείγματα σταθερού μεγέθους n, ο bootstrap-t πίνακας με τα ποσοστιαία σημεία εφαρμόζεται μόνο για το δοθέν πρόβλημα. Η bootstrap-t μέθοδος βρίσκει εφαρμογή ιδιαίτερα αν η παράμετρος που μας ενδιαφέρει είναι παράμετρος θέσης, όπως η διάμεσος, ή μέση τιμή ή τα ποσοστιαία σημεία αν και καμιά φορά οδηγούμαστε σε πολλές επαναλήψεις τις μεθόδου bootstrap. Είναι προφανές ότι στο στατιστικό $Z^*(b) = (\hat{\theta}^*(b) - \hat{\theta})/\widehat{se}^*(b)$ απαιτείται η εκτίμηση του τυπικού σφάλματος της $\hat{\theta}^*$. Έχουμε αναφέρει και στο πρώτο κεφάλαιο ότι αν η παράμετρος δεν είναι η μέση τιμή κάτι τέτοιο δεν είναι πάντα εφικτό ή απαιτεί δύσκολους υπολογισμούς. Αναγκαστικά θα πρέπει να καταφύγουμε ξανά στην μέθοδο bootstrap και να εκτιμήσουμε έτσι το τυπικό σφάλμα σε κάθε ένα από τα bootstrap δείγματα. Και ενώ για το τυπικό σφάλμα μπορεί να αρκούν B_2=20 επαναλήψεις, για τα διαστήματα εμπιστοσύνης και τα ποσοστιαία σημεία χρειαζόμαστε τουλάχιστον B_1=1.000 επαναλήψεις, οπότε συνολικά ξεπερνάμε τις $B_1 \cdot B_2$ =20.000 επαναλήψεις. Επιπλέον, αν μετασχηματίσουμε την παράμετρο οδηγούμαστε σε διαφορετικά διαστήματα εμπιστοσύνης, άλλες φορές καλύτερα, άλλες όχι, χωρίς να είναι προφανές ποιον μετασχηματισμό πρέπει να ακολουθήσουμε για την κατασκευή των bootstrap-t διαστημάτων εμπιστοσύνης. Γενικότερα η μέθοδος δεν είναι ανεκτική στον μετασχηματισμό της παραμέτρου (transformation respecting).

2.5 Bootstrap Percentile

Σε αυτήν την ενότητα θα δούμε μια άλλη μέθοδο κατασκευής διαστημάτων εμπιστοσύνης χρησιμοποιώντας τα ποσοστιαία σημεία της κατανομής bootstrap μιας στατιστικής συνάρτησης. Ας είναι λοιπόν το θ η παράμετρος που μας ενδιαφέρει, η $\hat{\theta}$ μια εκτιμήτρια συνάρτηση και $\hat{\theta}^*(b)$ η τιμή της στατιστικής συνάρτησης για το b bootstrap δείγμα που κατασκευάσαμε. Να διαχωρίσουμε εδώ ότι ενώ η δειγματική κατανομή μιας στατιστικής συνάρτησης συλλέγει τιμές του στατιστικού από τα δείγματα που έχουμε, η κατανομή bootstrap μιας στατιστικής συνάρτησης συλλέγει τις τιμές του στατιστικού από τα bootstrap δείγματα που έχουμε κατασκευάσει μέσω της επαναδειγματοληψίας. Όπως είναι λοιπόν σαφές η τυχαία μεταβλητή $\hat{\theta}^*$ ακολουθεί την bootstrap κατανομή. Ας ονομάσουμε $\widehat{G}$ την αθροιστική συνάρτηση κατανομής της $\hat{\theta}^*$. Τότε από τον ορισμό του ποσοστιαίου σημείου γνωρίζουμε ότι το 100· α ποσοστιαίο σημείο θα είναι το:

$$\hat{G}^{-1}(a) = \hat{\theta}^*_{\alpha},$$

Και συνεπώς το 1-α percentile διάστημα εμπιστοσύνης θα είναι το παρακάτω:

$$[\hat{\theta}_{\%,low}, \hat{\theta}_{\%,up}] = \left[\hat{G}^{-1}\left(\frac{a}{2}\right), \hat{G}^{-1}\left(1-\frac{a}{2}\right)\right] = [\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}]$$

Η παραπάνω έκφραση (Efron&Tibshirani 1993) αναφέρεται στο «ιδανικό» διάστημα εμπιστοσύνης (ideal percentile confidence interval), δηλαδή στην περίπτωση που πραγματοποιήσουμε άπειρες επαναλήψεις της μεθόδου bootstrap. Εμείς όμως σταματάμε μετά από Β επαναλήψεις , αφού δηλαδή κατασκευάσουμε τα εξής B bootstrap δείγματα $\boldsymbol{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$ και εφαρμόσουμε σε αυτά την στατιστική συνάρτηση $T(\boldsymbol{X}) = \hat{\theta}$, δηλαδή $\hat{\theta}^*(b) = T(X^{*b})$. Στη συνέχεια, διατάσσουμε τις τιμές των $\hat{\theta}^*(b)$ σε αύξουσα σειρά.Ας είναι για παράδειγμα οι διατεταγμένες τιμές των $\hat{\theta}^*(b)$ οι εξής: $\hat{\theta}^*_{(1)}, \hat{\theta}^*_{(2)}, \ldots, \hat{\theta}^*_{(B)}$. Έτσι μετά από B επαναλήψεις θα έχουμε ότι το αντίστοιχο 100(1-α)% διάστημα εμπιστοσύνης θα είναι:

$$[\hat{\theta}^*_{(\frac{\alpha}{2}*B)}, \hat{\theta}^*_{((1-\frac{\alpha}{2})*B)}]$$

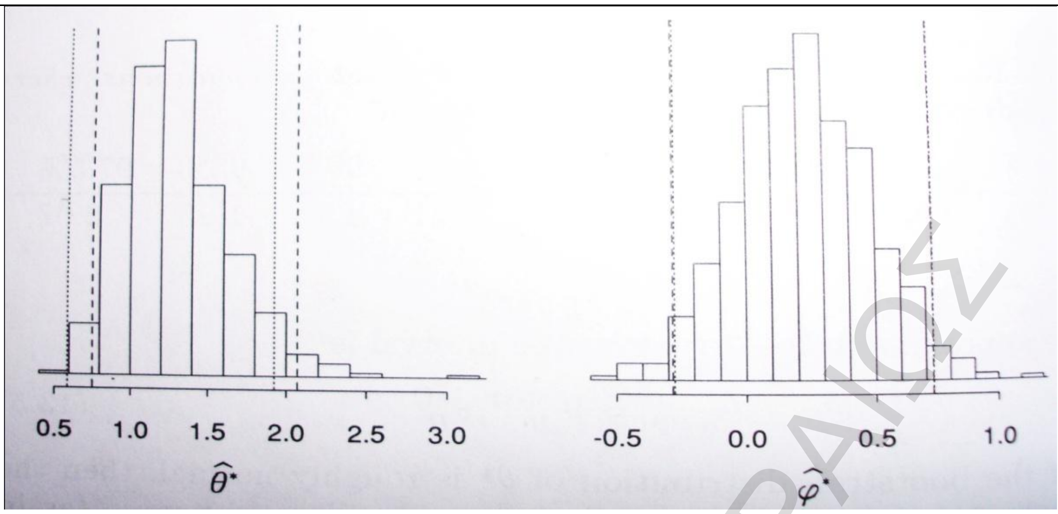
Για παράδειγμα εάν το B=2000 και ζητάμε το 95% percentile διάστημα εμπιστοσύνης για την παράμετρο θ, τότε το διάστημα αυτό θα είναι το :

$$[\hat{\theta}^*_{(50)}, \hat{\theta}^*_{(1950)}]$$

Εάν η κατανομή bootstrap της $\hat{\theta}^*$ προσεγγίζει την κανονική κατανομή τότε το διάστημα εμπιστοσύνης που δίνει η μέθοδος percentile είναι πολύ κοντά με το τυποποιημένο κανονικό διάστημα εμπιστοσύνης(standard normal confidence interval), προσεγγίζοντας δηλαδή το τυπικό σφάλμα με τη μέθοδο bootstrap αλλά χρησιμοποιώντας τα ποσοστιαία σημεία της τυποποιημένης κανονικής. Σύμφωνα με το κεντρικό οριακό θεώρημα όταν το μέγεθος του δείγματος n τείνει στο άπειρο τότε το ιστόγραμμα των bootstrap τιμών (replications) φαίνεται να παίρνει το σχήμα της κανονικής κατανομής και για αυτόν τον λόγο τα παραπάνω διαστήματα εμπιστοσύνης συμφωνούν. Στην περίπτωση που το μέγεθος n του αρχικού δείγματος είναι μικρό τα παραπάνω διαστήματα εμπιστοσύνης διαφέρουν. Ποια διαστήματα εμπιστοσύνης θα πρέπει να χρησιμοποιήσουμε; Θα παραθέσουμε το παρακάτω παράδειγμα των Efron- Tibshirani(Κεφ.13,1993) για να απαντηθεί το ερώτημα.

Παράδειγμα 2.1 (Efron- Tibshirani 1993)

Θα παράγουμε ένα δείγμα $X_1, X_2, \ldots, X_{10}$ από την τυποποιημένη κανονική κατανομή και θα υποθέσουμε ότι η παράμετρος που μας ενδιαφέρει είναι ή $\theta = e^{\mu}$, όπου μ είναι η μέση τιμή του πληθυσμού. Η πραγματική τιμή της παραμέτρου είναι $e^0 = 1$ ενώ η εκτίμηση της παραμέτρου στο δείγμα $\hat{\theta} = e^{\bar{x}}$ προκύπτει ίση με 1,25. Μετά από 1000 επαναλήψεις το 95% percentile διάστημα εμπιστοσύνης προκύπτει $[\hat{\theta}_{\%,low}, \hat{\theta}_{\%,up}]$=[0.75, 2.07], ενώ το αντίστοιχο τυποποιημένο, για $\widehat{se}_{1000} = 0.34$ είναι [0.59, 1.92]. Σε αυτή την περίπτωση είναι προτιμότερο να κρατήσουμε το percentile διάστημα εμπιστοσύνης, επειδή όπως φαίνεται και στο ιστόγραμμα των $\hat{\theta}^*$ η κανονική κατανομή δεν είναι μια ικανοποιητική προσέγγιση του δείγματος μας. Αντίθετα αν μετασχηματίσουμε την παράμετρο ως εξής: $\varphi = \log(\theta)$ και κατασκευάσουμε το ιστόγραμμα των $\hat{\varphi}^* = \log(\hat{\theta}^*)$, τότε αυτό φαίνεται να έχει το σχήμα της κανονικής κατανομής.



*Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.*Efron –Tibshirani 1993 *(Οι τελείες δείχνουν το κανονικοποιημένο διάστημα εμπιστοσύνης και οι διακεκομμένες το percentile διάστημα εμπιστοσύνης, σε επίπεδο 95% και μετά B=1000 επαναλήψεις)*

Το τυποποιημένο κανονικό διάστημα εμπιστοσύνης για την μετασχηματισμένη παράμετρο προκύπτει [-0.28, 0.73] ενώ το percentile διάστημα εμπιστοσύνης είναι το [-0.29,0.73] και όπως παρατηρούμε σχεδόν ταυτίζονται. Γενικότερα ένας μετασχηματισμός που θα «κανονικοποιεί» την κατανομή της παραμέτρου θα δίνει παρόμοια διαστήματα εμπιστοσύνης και με τις δύο μεθόδους. Στη συνέχεια μπορούμε να χρησιμοποιήσουμε το τυποποιημένο διάστημα εμπιστοσύνης για την φ, έτσι ώστε να υπολογίσουμε το διάστημα εμπιστοσύνης για την παράμετρο θ, με απλούς υπολογισμούς χρησιμοποιώντας την εκθετική συνάρτηση(η διαδικασία αυτή αναφέρεται ως mapping). Συγκεκριμένα:

$$[e^{-0.28}, e^{0.73}] = [0.76\,, 2.08]$$

Παρατηρούμε ότι το παραπάνω διάστημα εμπιστοσύνης σχεδόν ταυτίζεται με το διάστημα εμπιστοσύνης που προέκυψε με τη μέθοδο percentile για την παράμετρο θ. Μέσα από τον μετασχηματισμό που κάναμε καταφέραμε να βελτιώσουμε το αρχικό τυποποιημένο διάστημα εμπιστοσύνης, αλλά όπως είναι προφανές η δυσκολία έγκειται στην επιλογή του εκάστοτε μετασχηματισμού. Μπορούμε να θεωρήσουμε τη μέθοδο percentile είτε σαν μια μέθοδο που συμφωνεί με τα τυποποιημένα διαστήματα εμπιστοσύνης , όταν αυτά είναι ικανοποιητικά, είτε σαν έναν αλγόριθμο ο οποίος βελτιώνει τα τυποποιημένα διαστήματα εμπιστοσύνης, όταν αυτό χρειάζεται κάνοντας αυτόματα τον μετασχηματισμό που θα κάναμε σε αυτή την περίπτωση.

Δεν χρειάζεται να ξέρουμε ποιος θα είναι αυτός ο μετασχηματισμός, αρκεί απλά να υποθέσουμε ότι αυτός υπάρχει.

Μερικές φορές υπάρχουν περιορισμοί στις τιμές που μπορεί να πάρει μια παράμετρος, όπως για παράδειγμα ο συντελεστής συσχέτισης που παίρνει τιμές από -1 έως 1 και όπως είναι λογικό επιθυμούμε τα διαστήματα εμπιστοσύνης που κατασκευάζουμε να υπακούουν σε αυτούς τους περιορισμούς. Η μέθοδος percentile πλεονεκτεί έναντι των διαστημάτων εμπιστοσύνης που προκύπτουν με την τυποποιημένη κανονική διότι διατηρεί το εύρος των τιμών. Η ιδιότητα αυτή καλείται range-preserving. Αυτό οφείλεται τόσο στο ότι η plug-in εκτίμηση $\hat{\theta}$ όσο και οι τιμές $\hat{\theta}^*(b)$ υπακούουν στους ίδιους περιορισμούς που υπακούει και η θ. Τα range-preserving διαστήματα εμπιστοσύνης θεωρούνται πιο αξιόπιστα και ακριβή. Φυσικά η bootstrap percentile δεν είναι πάντα η προτιμότερη επιλογή, υπάρχουν και περιπτώσεις στις οποίες αποτυγχάνει γι’ αυτό κι έχουν προταθεί και άλλες μέθοδοι.

2.6 BCa bootstrap (Bias Corrected and Accelerated)

Σε αυτή την ενότητα θα ασχοληθούμε με τα BCa (bias corrected and accelerated) διαστήματα εμπιστοσύνης, τα οποία αποτελούν μία βελτιωμένη εκδοχή των percentile διαστημάτων εμπιστοσύνης. Ας υποθέσουμε ότι το $\hat{\theta}^*_{(\alpha)}$ είναι το $100\cdot\alpha$ ποσοστιαίο σημείο των B παρατηρήσεων $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$. Η μέθοδος BCa χρησιμοποιεί και αυτή τα ποσοστιαία σημεία της κατανομής bootstrap των $\hat{\theta}^*$, αλλά αυτά δεν είναι απαραίτητα αυτά που χρησιμοποιεί η μέθοδος percentile (βλ. ενότητα 2.3)

Ορισμός 2.3 (BCa διαστήμα εμπιστοσύνης). Το BCa διαστήμα εμπιστοσύνης με πιθανότητα κάλυψης 1-α δίνεται από:

$$\left[\, \hat{\theta}_{low}, \hat{\theta}_{up}\right] = [\hat{\theta}^*_{(a_1)}, \hat{\theta}^*_{(a_2)}],$$

όπου τα α_1 και α_2 δίνονται από:

$$\alpha_1 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z_{a/2}}{1 - \hat{\alpha}(\hat{z}_0 + z_{a/2})})$$

$$\alpha_2 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-a/2}}{1-\hat{\alpha}(\hat{z}_0 + z_{1-a/2})}).$$

Η Φ(·) είναι η τυποποιημένη κανονική αθροιστική συνάρτηση κατανομής και ως z_a συμβολίζουμε το 100· α ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής. Γνωρίζουμε ότι αν $z_{0,95} = 1{,}65$ τότε $\Phi(1{,}65) = 0{,}95$ κοκ. Επιπλέον, το $\hat{\alpha}$ είναι καλείται επιτάχυνση (acceleration) και το $\hat{z}_0$ διόρθωση μεροληψίας (bias- correction). Ας δούμε πώς υπολογίζονται οι αριθμοί $\hat{\alpha}$ και $\hat{z}_0$. Η τιμή του $\hat{z}_0$ προκύπτει από την παρακάτω αναλογία :

$$\hat{z}_0 = \Phi^{-1}(\frac{\#\{\hat{\theta}^*(b) < \hat{\theta}\}}{B}),$$

όπου $\hat{\theta}^*(b)$ είναι η εκτίμηση της παραμέτρου με την μέθοδο bootstrap, $\hat{\theta}$ είναι η εκτίμηση της παραμέτρου που προκύπτει από τα αρχικά δεδομένα και Φ^{-1} είναι η αντίστροφη συνάρτηση της τυποποιημένης κανονικής αθροιστικής συνάρτησης κατανομής, δηλαδή Φ^{-1}(0,95)= 1,65. Ο αριθμός $\hat{z}_0$ μετράει την μεροληψία της διαμέσου και αν προκύψει μηδέν τότε αυτό σημαίνει ότι οι μισές παρατηρήσεις των $\hat{\theta}^*$ είναι ίσες ή μικρότερες της εκτίμησης $\hat{\theta}$.

Ο αριθμός $\hat{\alpha}$ (acceleration) μετράει τον ρυθμό μεταβολής του τυπικού σφάλματος του $\hat{\theta}$ σε σχέση με την πραγματική τιμή της θ, μετρημένο στην κανονική κλίμακα (normalized scale). Η υπόθεση ότι $\hat{\theta} \sim N(\theta, se^2)$, συνεπάγεται ότι το τυπικό σφάλμα του $\hat{\theta}$ είναι ίδιο για κάθε θ, κάτι το οποίο διορθώνει η τιμή του α , μιας και η υπόθεση αυτή δεν είναι πάντα αληθής. Ο τρόπος υπολογισμού που προτείνεται από τους Efron- Tibshirani(1993) χρησιμοποιεί τις jackknife τιμές για την $\hat{\theta} = T(\tilde{x})$. Αν λοιπόν θέσουμε ότι το δείγμα $x_{(i)}$ είναι το αρχικό δείγμα παρατηρήσεων χωρίς την παρατήρηση i, τότε θα συμβολίσουμε με $\hat{\theta}_{(i)}$ τιμή της $T(\tilde{x})$ για το $x_{(i)}$ δείγμα, δηλαδή $\hat{\theta}_{(i)} = T(\, x_{(i)})$ και με $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n$. Το $\hat{\alpha}$ θα ορίζεται ως εξής:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6\{\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2\}^{3/2}}$$

Σημειώνουμε ότι όπως και στα percentile bootstrap διαστήματα εμπιστοσύνης, εάν η κατανομή του bootstrap δείγματος είναι G_{boot}, τότε το $1 - \alpha \cdot 100\%$ BCa διάστημα εμπιστοσύνης μπορεί να γραφτεί και ως εξής :

$$[\,\hat{\theta}_{low}, \hat{\theta}_{up}] = [\hat{\theta}^*_{(a_1)}, \hat{\theta}^*_{(a_2)}]\text{=}[G^{-1}_{boot}(a_1)\,, G^{-1}_{boot}(a_2)],$$

όπου τα a_1, a_2 ορίζονται όπως και παραπάνω. Ας παρατηρήσουμε ότι αν τα $\hat{\alpha}$ και $\hat{z}_0$ προκύψουν ίσα με το μηδέν τότε:

$$\alpha_1 = \Phi(z_{a/2}) = \alpha/2 \text{ και } \alpha_2 = \Phi\left(z_{1-\frac{a}{2}}\right) = 1-\alpha/2$$

και τότε το διάστημα εμπιστοσύνης ταυτίζεται με το percentile bootstrap διάστημα εμπιστοσύνης.

Στους DiCiccio-Efron (1996) αναφέρεται πως τα BCa διαστήματα εμπιστοσύνης είναι σταθερά στον μετασχηματισμό της παραμέτρου (transformation respecting). Αν μετασχηματίσουμε την παράμετρο θ στην $\varphi = m(\theta)$, όπου η $m(\theta)$ είναι μια μονότονη συνάρτηση, τότε κατά συνέπεια θα αλλάξουν και τα $\hat{\theta}$, $\hat{\theta}^*$ σε $\hat{\varphi} = m(\hat{\theta})$ και $\hat{\varphi}^* = m(\hat{\varphi})$ αντίστοιχα. Με το ίδιο τρόπο θα αλλάξουν και οι κρίσιμες τιμές (endpoints) του διαστήματος εμπιστοσύνης και θα είναι:

$$\hat{\varphi}_{BC_a}[a] = m(\hat{\theta}_{BC_a}[a])$$

Για παράδειγμα αν έχουμε υπολογίσει ένα διάστημα εμπιστοσύνης για την παράμετρο θ και αποφασίσουμε να αλλάξουμε την παράμετρο σε $\sqrt{\theta}$, τότε για να υπολογίσουμε εκ νέου τις κριτικές τιμές αρκεί να πάρουμε την τετραγωνική ρίζα των κριτικών τιμών της θ. Επιπλέον όπως αναφέρθηκε και στο *Παράδειγμα 2.1* αρκετές φορές για να λάβουμε πιο ακριβή διαστήματα εμπιστοσύνης πρέπει να μετασχηματίσουμε την παράμετρο μας, έτσι ώστε να έρθει πιο κοντά στην κανονική κατανομή και με βάση τα νέα διαστήματα εμπιστοσύνης για την μετασχηματισμένη παράμετρο να υπολογίσουμε τα αντίστοιχα της παραμέτρου που μας ενδιαφέρει. Αντίστοιχα, ένας λάθος μετασχηματισμός μπορεί να δώσει λάθος διαστήματα. Όπως και στη percentile λοιπόν, δεν είναι απαραίτητο να κάνουμε τον μετασχηματισμό διότι η μέθοδος κατασκευάζει αυτόματα τα βελτιωμένα διαστήματα εμπιστοσύνης και δεν επηρεάζεται από λάθος μετασχηματισμούς. Γενικότερα, αν και τα BCa διαστήματα εμπιστοσύνης φαίνονται πιο δυσνόητα, εντούτοις δεν είναι πολύ δύσκολα στον υπολογισμό από τα percentile και τείνουν να δίνουν πιο ακριβή αποτελέσματα, γι' αυτό και στην πράξη προτιμώνται.

2.7 Ακρίβεια και σύγκριση διαστημάτων εμπιστοσύνης

Όπως σε κάθε σημειακή εκτίμηση έτσι και στα διαστήματα εμπιστοσύνης μας ενδιαφέρει η ακρίβεια. Όπως έχει αναφερθεί και στις προηγούμενες ενότητες τα διαστήματα εμπιστοσύνης που προκύπτουν με τις διάφορες παραλλαγές της μεθόδου bootstrap είναι προσεγγιστικά, το οποίο σημαίνει ότι η πιθανότητα κάλυψης τους δεν είναι ακριβώς $(1-\alpha)$. Για ένα $(1-\alpha)\cdot 100\%$ διάστημα εμπιστοσύνης θα πρέπει η πραγματική τιμή της παραμέτρου να βρίσκεται μέσα στο διάστημα με πιθανότητα $(1-\alpha)$. Οι μέθοδοι επαναδειγματοληψίας θεωρητικά κατασκευάζουν ακριβή διαστήματα εμπιστοσύνης στην περίπτωση που έχουμε την δυνατότητα να κάνουμε άπειρες επαναλήψεις και η κατανομή του πληθυσμού είναι συμμετρική, πρακτικά όμως, επειδή εμείς σταματάμε μετά από B επαναλήψεις τα διαστήματα είναι προσεγγιστικά, και το πόσο κοντά είναι η πραγματική πιθανότητα κάλυψης στο ονομαστικό επίπεδο αποτελεί ένα κριτήριο για το αν τελικά η μέθοδος που χρησιμοποιήσαμε δούλεψε ικανοποιητικά. Συγκεκριμένα μπορούμε να συγκρίνουμε τις διάφορες παραλλαγές της μεθόδου bootstrap που περιγράψαμε στις προηγούμενες ενότητας χρησιμοποιώντας τον ρυθμό σύγκλισης της πιθανότητας κάλυψης ενός διαστήματος εμπιστοσύνης προς το ονομαστικό του επίπεδο.

Για μεγαλύτερη ευκολία, για να μην έχουμε δύο ακραία σημεία (endpoints) για το διάστημα εμπιστοσύνης, θα χρησιμοποιήσουμε το μονόπλευρο διάστημα εμπιστοσύνης με επιθυμητή πιθανότητα κάλυψης α και το ακραίο σημείο $\hat{\theta}[\alpha]$.Δηλαδή:

$$P\big(\theta \le \hat{\theta}[\alpha]\big) \approx \alpha, \forall \alpha$$

Ορισμός 2.4 : Το παραπάνω προσεγγιστικό διάστημα εμπιστοσύνης θα λέμε ότι είναι πρώτης τάξης ακριβές εάν :

$$P\big(\theta \le \hat{\theta}[\alpha]\big) = a + O(n^{-\frac{1}{2}})$$

και δεύτερης τάξης ακριβές εάν:

$$P\big(\theta \le \hat{\theta}[\alpha]\big) = a + O(n^{-1}),$$

όπου οι παραπάνω πιθανότητες αναφέρονται στον πληθυσμό ή την κατανομή του πληθυσμού. Γενικότερα αναφερόμαστε σε ακρίβεια $\kappa - \tau\acute{\alpha}\xi\eta\varsigma$ εάν:

$$P\big(\theta \leq \hat{\theta}[\alpha]\big) = a + O(n^{-\frac{\kappa}{2}})$$

Σημειώνουμε, όταν αναφέρουμε ότι μια ποσότητα A_n είναι $O(n^d)$ τότε θα ισχύει ότι:

$$\lim_{n\to\infty} n^{-d} A_n = a,$$

για κάποιο a πεπερασμένο.

Η διαφορά ανάμεσα στην πρώτη και δεύτερη τάξη ακρίβειας είναι ότι η τελευταία μας δίνει καλύτερες προσεγγίσεις στα όρια ενός διαστήματος (endpoints) όταν αυτά υπάρχουν. Σύμφωνα με τους Efron - Tibshirani(1998) και DiCiccio-Efron(1998) τα τυποποιημένα (standardized) διαστήματα εμπιστοσύνης είναι πρώτης τάξης ακριβή, αλλά όχι δεύτερης τάξης εκτός κι αν η πραγματική κατανομή του πληθυσμού είναι η κανονική. Ωστόσο υπάρχουν μέθοδοι bootstrap που δίνουν και ακρίβεια δεύτερης τάξης, ανεξάρτητα απ’ την κατανομή του πληθυσμού. Η τυποποιημένη μέθοδος (standardized) συγκεκριμένα, μας δίνει διαστήματα τα οποία δεν είναι δεύτερης τάξης ακριβή, αλλά επίσης δεν είναι ανθεκτικά στους μετασχηματισμούς (transformation respecting), ενώ αντίθετα η μέθοδος BCa μας παρέχει και τα δύο. Η percentile bootstrap είναι transformation respecting, αλλά όχι δεύτερης τάξης ακριβής. Η μέθοδος bootstrap-t μας δίνει διαστήματα τα οποία είναι δεύτερης τάξης ακριβή (Hall 1988), αλλά δεν είναι transformation respecting, ενώ ο αλγόριθμος ενδέχεται να είναι ασταθής και να δώσει πολύ μακριά διαστήματα, ιδιαίτερα στην μη παραμετρική περίπτωση, όπου και προτείνεται η BCa μέθοδος ως πιο συντηρητική.

Στους Shao-Tu (1995) αναφέρεται ότι τα BCa και τα bootstrap-t διαστήματα εμπιστοσύνης είναι καλύτερα σε σχέση με εκείνα που βασίζονται σε προσεγγίσεις της κανονικής κατανομής. Επιπλέον, τα BCa διαστήματα είναι πολύ ακριβή στην μη παραμετρική περίπτωση, όπου η εκτίμηση του συντελεστή επιτάχυνσης $\hat{\alpha}$ ορίζεται εύκολα, αλλά όχι τόσο ακριβή στην παραμετρική περίπτωση, όπου δεν είναι διαθέσιμος ένας ακριβής εκτιμητής $\hat{\alpha}$. Και εδώ αναφέρεται ότι η μέθοδος t-bootstrap μπορεί να δώσει μακριά διαστήματα εμπιστοσύνης.

Ένα άλλο βασικό κριτήριο σύγκρισης διαστημάτων εμπιστοσύνης, εκτός από την πιθανότητα κάλυψης, είναι το αναμενόμενο πλάτος των διαστημάτων εμπιστοσύνης. Προφανώς ένα διάστημα εμπιστοσύνης με μεγάλο πλάτος δεν αποτελεί μια καλή εκτίμηση. Στη συνέχεια μπορούμε να μελετήσουμε και την τυπική απόκλιση του αναμενόμενου πλάτους του διαστήματος εμπιστοσύνης ώστε να έχουμε μια πληρέστερη εικόνα για την αξιολόγηση του. Στο επόμενο κεφάλαιο θα επιχειρήσουμε να λάβουμε ένα δείγμα από μια γνωστή κατανομή, με γνωστές παραμέτρους και να ελέγξουμε πόσο ικανοποιητικά αποτελέσματα δίνουν οι παραπάνω μέθοδοι που αναλύσαμε.

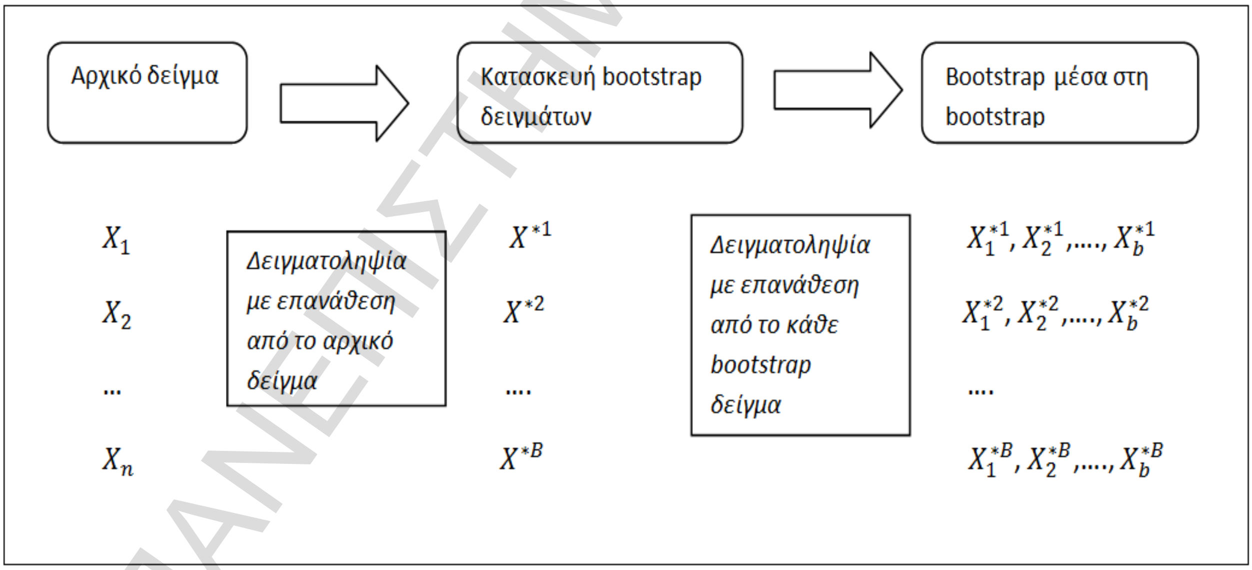
Κεφάλαιο 3^{o}

Σε αυτό το κεφάλαιο θα πάρουμε δείγματα από μερικές γνωστές κατανομές και θα κατασκευάσουμε διαστήματα εμπιστοσύνης για την διάμεσο με όλες τις μεθόδους που αναπτύχθηκαν στο προηγούμενο κεφάλαιο. Σκοπός μας είναι να αξιολογήσουμε τις μεθόδους και τα διαστήματα εμπιστοσύνης που προκύπτουν ελέγχοντας την πραγματική πιθανότητα κάλυψης καθώς και το πλάτος των διαστημάτων που προκύπτουν.

3.1 Εφαρμογή της bootstrap για την διάμεσο της κανονικής κατανομής

Η διάμεσος, ή αλλιώς το 50^{o} ποσοστιαίο σημείο είναι η αριθμητική τιμή που χωρίζει στα δύο την κατανομή μιας μεταβλητής, όταν οι τιμές της τοποθετηθούν σε αύξουσα σειρά. Η διάμεσος είναι ένα πολύ χρήσιμο περιγραφικό μέτρο θέσης, ιδιαίτερα για τις ασύμμετρες κατανομές, όπου περιγράφει καλύτερα τον πληθυσμό συγκριτικά με την μέση τιμή και επιπλέον δεν επηρεάζεται από τις έκτροπες παρατηρήσεις. Με την μέθοδο bootstrap θα κατασκευάσουμε διαστήματα εμπιστοσύνης για την διάμεσο χρησιμοποιώντας διάφορες παραλλαγές της μεθόδου, πραγματοποιώντας δειγματοληψία με επανάθεση από ένα τυχαίο δείγμα που προέρχεται από μια γνωστή κατανομή, συγκρίνοντας τα αποτελέσματα μας ανάμεσα στις μεθόδους, αλλά και πως συμπεριφέρεται η bootstrap αν μεταβάλλουμε το αρχικό μέγεθος δείγματος n και το πλήθος των επαναλήψεων B.

Θα χρησιμοποιήσουμε στην R το πακέτο boot και συγκεκριμένα την εντολή boot.ci η οποία κατασκευάζει μη παραμετρικά διαστήματα εμπιστοσύνης ίσων ουρών, με πέντε διαφορετικές μεθόδους. Τα διαστήματα που προκύπτουν είναι το κανονικοποιημένο διάστημα (στην εντολή ορίζεται ως norm), το βασικό bootstrap διάστημα εμπιστοσύνης (basic), το bootstrap-t διάστημα ή αλλιώς studentized, το percentile bootstrap και τέλος το BCa διάστημα εμπιστοσύνης. Ωστόσο επειδή η παράμετρος που μελετάμε είναι η διάμεσος, η διαδικασία που θα ακολουθήσουμε παρουσιάζει μια ιδιαιτερότητα. Συγκεκριμένα, για τον υπολογισμό των κανονικοποιημένων διαστημάτων εμπιστοσύνης απαιτείται να υποθέσουμε ότι η δειγματική κατανομή προσεγγίζει την τυποποιημένη κανονική, κάτι το οποίο ίσως είναι εύκολο να γίνει μέσω του κεντρικού οριακού θεωρήματος, εάν μιλάμε για την μέση τιμή, αλλά όχι αυτονόητο όταν μελετάμε την διάμεσο. Επιπλέον, για τον παραπάνω υπολογισμό απαιτείται το τυπικό σφάλμα της διαμέσου, για το οποίο δεν μπορούμε να χρησιμοποιήσουμε κάποιον έτοιμο τύπο, ειδικά όταν το δείγμα δεν προέρχεται από κανονική κατανομή. Το ίδιο κώλυμα εμφανίζεται και στην κατασκευή των bootstrap-t διαστημάτων. Αναλυτικότερα, όπως είδαμε και στο Κεφάλαιο 2 για την κατασκευή των studentized διαστημάτων καλούμαστε να υπολογίσουμε σε κάθε bootstrap δείγμα την τιμή του στατιστικού $Z^* = \frac{\hat{\theta}^* - \hat{\theta}}{\widehat{se}^*}$. Για να εκτιμήσουμε την τυπική απόκλιση της διαμέσου σε κάθε bootstrap δείγμα, πρέπει να θεωρήσουμε το κάθε bootstrap δείγμα που έχουμε σαν αρχικό και να πραγματοποιήσουμε εκ νέου δειγματοληψία με επανάθεση με σκοπό την εκτίμηση του τυπικού σφάλματος, δηλαδή να κάνουμε bootstrap μέσα στη bootstrap. Σχηματικά:



Εικόνα 3.1: *Bootstrap μέσα στην bootstrap*

Αρχικά θα χρησιμοποιήσουμε δεδομένα από την τυπική κανονική κατανομή για να δούμε πως ανταποκρίνεται η μέθοδος bootstrap. Η διαδικασία που θα ακολουθήσουμε θα είναι η εξής: Από ένα τυχαίο δείγμα που προέρχεται από την τυπική κανονική κατανομή ,θα κατασκευάζουμε ταυτόχρονα 5 διαστήματα εμπιστοσύνης με ονομαστικό επίπεδο κάλυψης 95% ,με τις πέντε μεθόδους ταυτόχρονα που αναφέρονται στο Κεφάλαιο 2. Θα επαναλάβουμε τη διαδικασία αρχικά για 300 και στη συνέχεια 500 δοκιμές (trials), ώστε να έχουμε την δυνατότητα να υπολογίσουμε την πιθανότητα κάλυψης και το μέσο πλάτος των διαστημάτων εμπιστοσύνης και επιπλέον σε κάθε δοκιμή θα αποθηκεύουμε την bootstrap εκτίμηση για την διάμεσο $\hat{\theta}^*$. Να αναφέρουμε ότι η $\hat{\theta}^*$ είναι η σημειακή εκτίμηση που μας έδωσε η bootstrap σε κάθε μία από τις δοκιμές που πραγματοποιήσαμε και όχι η τιμή του στατιστικού σε κάθε ένα από τα bootstrap δείγματα (replications) που χρησιμοποιήσαμε για την κατασκευή ενός διαστήματος εμπιστοσύνης. Για να διαπιστώσουμε πώς το μέγεθος του αρχικού μας δείγματος επηρεάζει τα αποτελέσματά μας, θα πάρουμε διαδοχικά $n = 50, n = 200, n = 500$ και ομοίως θα επαναλάβουμε την διαδικασία πραγματοποιώντας $B = 1.000, B = 5.000, B = 10.000$ επαναλήψεις της μεθόδου bootstrap. Επιλέγουμε μεγάλα δείγματα ($n > 30$), διότι υπάρχουν ενδείξεις ότι για μικρά δείγματα η μέθοδος δεν θα δώσει ικανοποιητικά αποτελέσματα (Efron 1980). Η δυσκολία έγκειται στο ότι η διάμεσος από την μία δεν είναι ομαλή στατιστική συνάρτηση, από την άλλη αν το μέγεθος δείγματος είναι μικρό, τότε σε κάθε επανάληψη της bootstrap υπάρχει μεγάλη πιθανότητα να επαναλαμβάνονται οι ίδιες τιμές. Να σημειώσουμε εδώ, ότι δεν πρέπει να συγχέουμε τον αριθμό των δοκιμών (trials) με τον αριθμό των επαναλήψεων της μεθόδου (replications). Κάθε δοκιμή θα περιλαμβάνει ένα καινούριο δείγμα από την τυπική κανονική, αλλά για κάθε δείγμα θα πραγματοποιούμε συγκεκριμένο αριθμό επαναλήψεων έτσι ώστε να λάβουμε ένα διάστημα εμπιστοσύνης και μια σημειακή εκτίμηση για την διάμεσο. Επιπλέον, είναι προφανές ότι η τιμή της παραμέτρου που μελετάμε δεν είναι μια άγνωστη τιμή για εμάς, μιας και τα δεδομένα προέρχονται από μια γνωστή κατανομή και ως εκ τούτου πρωταρχικός μας στόχος δεν είναι να κάνουμε μια εκτίμηση, αλλά να δούμε πόσο καλά αποδίδει η bootstrap και αν υστερεί ή υπερτερεί σε σχέση με τις παραδοσιακές μεθόδους.

Ο κώδικας που χρησιμοποιήθηκε παρακάτω παρατίθεται στο παράρτημα (Π1).

- **300 Δοκιμές**

Πραγματοποιήσαμε την παραπάνω διαδικασία, πραγματοποιώντας 300 δοκιμές και λάβαμε τα παρακάτω αποτελέσματα. Ας δούμε αρχικά μερικά περιγραφικά μέτρα για τις σημειακές εκτιμήσεις:

- *n=50*

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	0.000512	0.197766	-0.1321	0.02025	0.1296	-0.5671	0.575
B=5000	-0.01380	0.209027	-0.1379	0.00025	0.12685	-0.6615	0.5642
B=10.000	-0.01109	0.189404	-0.1371	-0.01415	0.10437	-0.61095	0.6218

Πίνακας 3.1

- *n=200*

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	0.003201	0.095324	-0.0542	0.00415	0.06615	-0.3137	0.2554
B=5000	-0.00698	0.093598	-0.0644	-0.0044	0.05937	-0.2747	0.2348
B=10.000	0.000173	0.10317	-0.0737	0.00017	0.07510	-0.2687	0.3005

Πίνακας 3.2

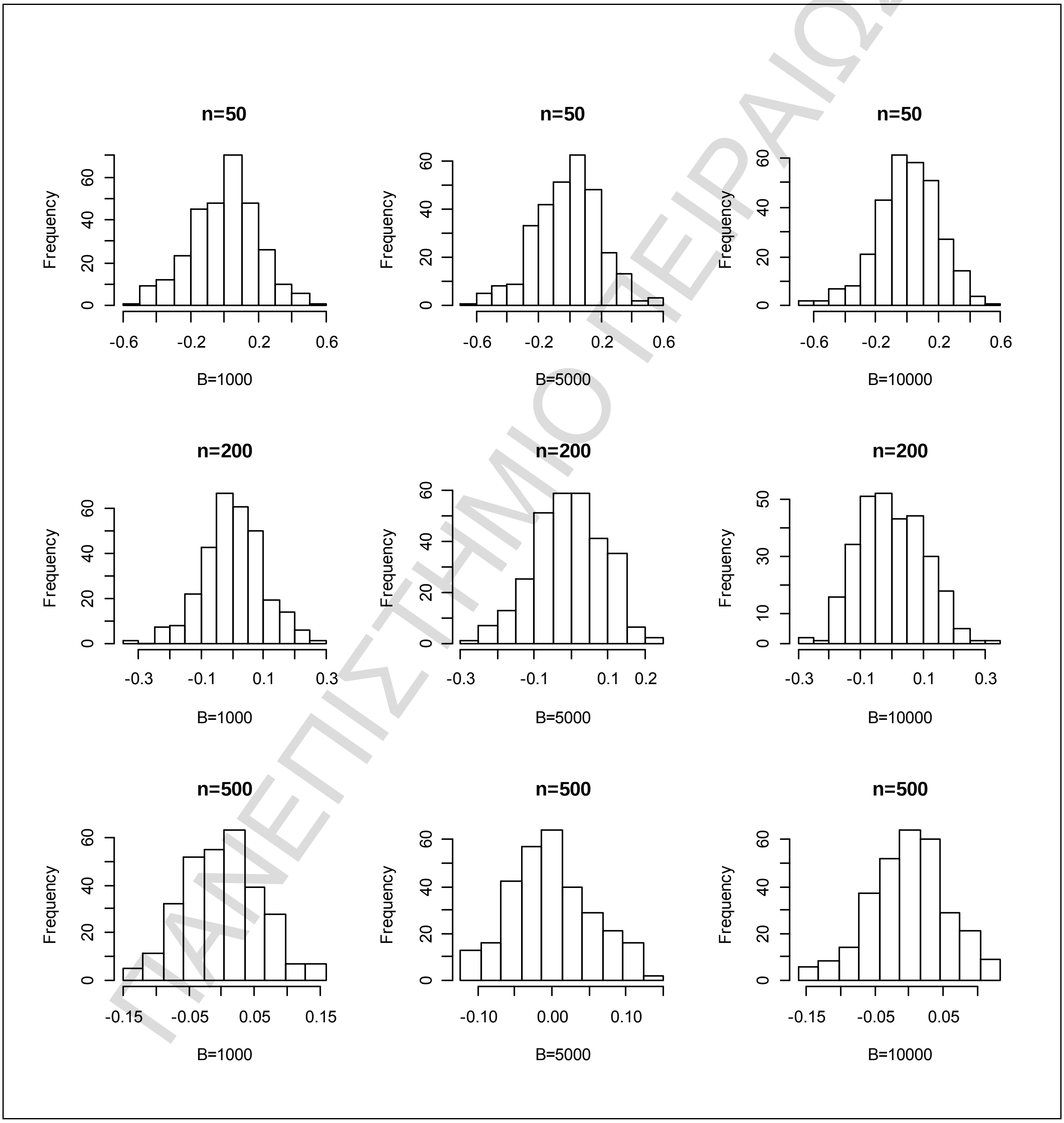
- *n=500*

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	0.000810	0.05829	-0.0412	0.0014	0.03963	-0.1500	0.1590
B=5000	-0.000607	0.05627	-0.0389	-0.0048	0.03656	-0.1234	0.1507
B=10.000	-0.00065	0.05642	-0.0373	0.0017	0.03298	-0.1615	0.1334

Πίνακας 3.3

Στους παραπάνω πίνακες έχουμε συνοπτικά μερικά περιγραφικά μέτρα που αφορούν 300 σημειακές εκτιμήσεις της διαμέσου που λάβαμε με την μέθοδο bootstrap. Παρατηρούμε ότι όσο μεγαλύτερο είναι το μέγεθος του αρχικού μας δείγματος , τότε σταθερά η τυπική απόκλιση της εκτίμησης μειώνεται. Η μεγαλύτερη τυπική απόκλιση που παρατηρείται είναι για δείγματα μεγέθους $n = 50$, ενώ η μικρότερη είναι για δείγματα μεγέθους $n = 500$. Η μικρότερη τυπική απόκλιση εντοπίζεται για *n=500* και *B=5.000*. Ωστόσο, παρατηρούμε ότι η αύξηση των επαναλήψεων B από 1.000 σε 5.000 και στη συνέχεια σε 10.000 δεν μειώνει απαραίτητα την τυπική απόκλιση. Κάτι τέτοιο δεν είναι παράδοξο, διότι η τυπική απόκλιση εξαρτάται από το μέγεθος του αρχικού δείγματος.

Ας δούμε και τα αντίστοιχα ιστογράμματα:



Πίνακας 3.4: *Ιστογράμματα συχνοτήτων των* $\widehat{\boldsymbol{\theta}}^*$

Παρατηρούμε ότι τα περισσότερα ιστογράμματα τα οποία φαίνονται συμμετρικά. Για παράδειγμα στην περίπτωση *n=200* και *B=1.000* ή *5.000* η κατανομή των $\hat{\theta}^*$ φαίνεται πιο συμμετρική από ότι στην περίπτωση n=50, B=1.000. Υπογραμμίζουμε σε αυτό το σημείο, ότι η κατανομή των bootstrap δειγμάτων δεν είναι απαραίτητο να ταυτίζεται με την κατανομή του πληθυσμού και ως εκ τούτου δεν περιμένουμε ότι όλα τα ιστογράμματα θα μοιάζουν στην κανονική κατανομή. Στον επόμενο πίνακα δίνεται η πιθανότητα κάλυψης των διαστημάτων εμπιστοσύνης για κάθε μία από τις μεθόδους και το μέσο πλάτος των διαστημάτων εμπιστοσύνης, μετά από 300 δοκιμές:

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=50	B=1000	0,9233	0,8568	0,92	0,96	**0,9567**
	B=5000	0,920266	0,86711	0,900332	0,94684	0,93356
	B=10.000	0,909699	0,87291	0,90301	**0,94983**	0,93646
n=200	B=1000	0,9333	0,8860	0,9231	0,9565	**0,9498**
	B=5000	0,9233	0,8867	0,91	0,9633	95,67
	B=10.000	0,9060	0,8792	0,9195	0,9430	94,63
n=500	B=1000	0,9541	0,9410	0,9541	0,9607	0,9541
	B=5000	0,9534	0,93	0,9633	**0,951**	**0,951**
	B=10.000	0,9406	0,9076	0,9439	0,9670	0,9604

Πίνακας 3.5: *Πιθανότητες κάλυψης για την διάμεσο της τυπικής κανονικής, πλήθος δοκιμών 300.Σημειώνονται με έντονα γράμματα πιθανότητες κάλυψης που είναι πολύ κοντά στο 0,95*

Συμπερασματικά μπορούμε να πούμε ότι σε ένα δεδομένο μέγεθος δείγματος η αύξηση του αριθμού B των επαναλήψεων δεν βελτιώνει απαραίτητα την πιθανότητα κάλυψης. Για παράδειγμα η πιθανότητα κάλυψης στην περίπτωση που *n=500* και *B=5.000* (norm) είναι 0,9534, ενώ όταν μεγαλώνουμε τον αριθμό των επαναλήψεων σε *B=10.000* η πιθανότητα κάλυψης μειώνεται σε 0,9406. Σε γενικές γραμμές όσο αυξάνουμε το μέγεθος του αρχικού μας δείγματος τα αποτελέσματα τείνουν να βελτιώνονται. Όταν το αρχικό μας δείγμα αποτελείται από 50 παρατηρήσεις τα αποτελέσματα είναι απογοητευτικά στις περισσότερες μεθόδους, με εξαίρεση τη μέθοδο percentile και την bca. Η percentile δίνει ικανοποιητικά αποτελέσματα για *B= 5.000* και *B=10.000*, ενώ η bca έρχεται πολύ κοντά στο ονομαστικό επίπεδο της πιθανότητας κάλυψης με *B=1.000* επαναλήψεις. Η μέθοδος basic δίνει σταθερά τα χειρότερα διαστήματα εμπιστοσύνης συγκριτικά με τις άλλες μεθόδους, όσον αφορά την πιθανότητα κάλυψης και μόνο όταν αυξάνεται τόσο το μέγεθος δείγματος, όσο και το πλήθος των επαναλήψεων πλησιάζει το 95%. Επιπλέον οι μέθοδοι percentile και bca φαίνεται να ξεπερνάνε σε αρκετές περιπτώσεις το ονομαστικό επίπεδο κάλυψης των διαστημάτων εμπιστοσύνης. Σε αυτήν την περίπτωση η μέθοδος λέγεται συντηρητική (conservative).

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=50	B=1000	0,7163	0,70158	**0,8107**	0,7017	0,70158
	B=5000	0,7291	0,7146	0,77515	0,7146	0,70307
	B=10.000	0,7220	0,7024	0,79491	0,7051	0,69713
n=200	B=1000	0,3537	0,3478	0,3765	0,3478	0,34769
	B=5000	0,3508	0,3443	0,3627	0,3443	0,3440
	B=10.000	0,3583	0,3542	0,3890	0,3542	0,3528
n=500	B=1000	0,2220	0,2205	0,2417	0,2205	0,2202
	B=5000	0,2204	**0,2174**	0,2346	**0,2174**	**0,2174**
	B=10.000	0,2224	0,2199	0,228	0,21993	0,2198

Πίνακας 3.6:*Μέσο πλάτος διαστημάτων εμπιστοσύνης για την διάμεσο της κανονικής,300 δοκιμές. Σημειώνονται με έντονα γράμματα το μικρότερο και το μεγαλύτερο μέσο πλάτος.*

Στον παραπάνω πίνακα, παρατηρώντας το μέσο πλάτος των διαστημάτων εμπιστοσύνης, βλέπουμε ότι σταθερά όταν αυξάνουμε το αρχικό μέγεθος δείγματος, το πλάτος μειώνεται. Ωστόσο και εδώ, η αύξηση του B δεν συνεπάγεται την μείωση του μέσου πλάτους. Επιπλέον η μέθοδος t-bootstrap (studentized) έχει την τάση να δίνει τα πιο πλατιά διαστήματα εμπιστοσύνης, κάτι το οποίο έχουμε αναφέρει ότι είναι πιθανό να συμβεί με αυτήν την μέθοδο (Ενότητα 2.5) .

- **500 Δοκιμές**

Παραθέτουμε τα αντίστοιχα αποτελέσματα, για 500 δοκιμές. Σκοπός είναι να δούμε αν αυτά συμφωνούν με τα προηγούμενα. Ας δούμε πρώτα τα περιγραφικά μέτρα:

- *n=50*

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	0.001925	0.1924958	-0.127387	0.007750	0.132837	-0.675150	0.589350
B=5000	0.003181	0.1883313	-0.11572	-0.00695	0.12490	-0.56315	0.65440
B=10.000	-0.01109	0.1894037	-0.13705	-0.01415	0.10437	-0.61095	0.62180

Πίνακας 3.7

- *n=200*

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	0.001950	0.0970245	-0.064763	0.000625	0.066337	-0.4077	0.3266
B=5000	-0.000996	0.0927108	-0.06606	-0.001475	0.055413	-0.2863	0.2563
B=10.000	0.001052	0.0921743	-0.06163	0.000325	0.062588	-0.3148	0.2574

Πίνακας 3.8

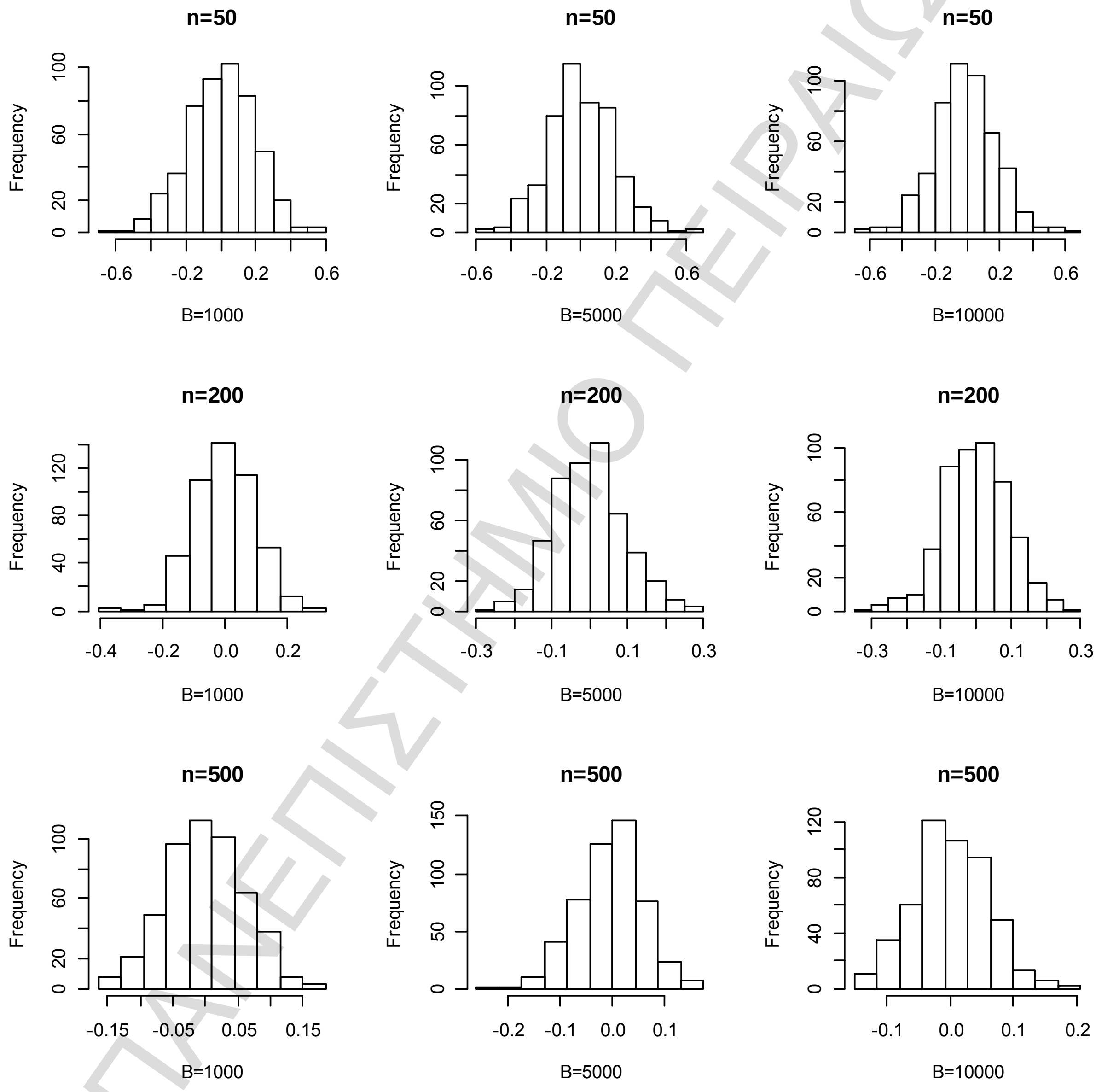
- *n=500*

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	0.001409	0.059912	-0.037275	0.001825	0.0411	-0.1639	0.1853
B=5000	-0.003779	0.061844	-0.04511	-0.001025	0.03679	-0.2615	0.17495
B=10.000	0.0001093	0.059777	-0.03693	-0.000725	0.04151	-0.1531	0.20705

Πίνακας 3.9

Τα συμπεράσματα και εδώ είναι αντίστοιχα με τα προηγούμενα και συμφωνούν με αυτά των 300 δοκιμών.

Παραθέτουμε και τα ιστογράμματα:



Πίνακας 3.10*: Ιστογράμματα συχνοτήτων των* $\widehat{\boldsymbol{\theta}}^*$*, διάμεσος τυπικής κανονικής,500 δοκιμές*

Τα ιστογράμματα (*n=50* και *B=1.000 και B=10.000*) φαίνονται ικανοποιητικά, ενώ συμμετρικό είναι και το ιστόγραμμα για *n=500* και *B=1.000*. Υπάρχουν και περιπτώσεις όπου εμφανίζονται λιγότερο συμμετρικά όπως για *n=500* και *B=5.000/B=10.000*. Τα ιστογράμματα έχουν παρόμοια μορφή αν τα συγκρίνουμε με αυτά των 300 δοκιμών.

Ας δούμε τον αντίστοιχο πίνακα για την πιθανότητα κάλυψης:

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=50	B=1000	0,938	0,878	0,932	0,966	0,962
	B=5000	0,925852	0,881764	0,921844	**0,947896**	0,94389
	B=10.000	0,923848	0,869739	0,913828	0,953908	0,951904
n=200	B=1000	0,934426	0,887295	0,928279	0,942623	**0,95082**
	B=5000	0,92	0,898	0,922	0,938	0,936
	B=10.000	0,936	0,906	0,952	0,956	0,956
n=500	B=1000	0,928	0,892	0,926	**0,95**	0,956
	B=5000	0,923228	0,905512	0,913386	0,935039	0,9331
	B=10.000	0,924	**0,92**	0,928	**0,948**	0,942

Πίνακας 3.11- *500 δοκιμές για την διάμεσο της τυπικής κανονικής.Σημειώνονται με έντονα γράμματα πιθανότητες κάλυψης που είναι πολύ κοντά στο 0,95*

Αντίστοιχα είναι τα αποτελέσματα και για τις 500 δοκιμές. Και εδώ παρατηρούμε ότι η αύξηση του Β δεν βελτιώνει πάντοτε την πιθανότητα κάλυψης. Για παράδειγμα για *n =500* και *B=1.000* η πιθανότητα κάλυψης (perc) είναι 0,95 ενώ για *n=500* και *B=5.000* η πιθανότητα κάλυψης είναι 0,935. Επιπλέον ούτε και η αύξηση του δείγματος μας δίνει απαραίτητα καλύτερα επίπεδα κάλυψης. Για παράδειγμα αν πραγματοποιήσουμε 1.000 επαναλήψεις της μεθόδου bootstrap με ένα αρχικό δείγμα μεγέθους *n=50* (norm) η πιθανότητα κάλυψης είναι 0,938, ενώ με την ίδια μέθοδο, με 1.000 επαναλήψεις και αρχικό δείγμα *n=500* η πιθανότητα κάλυψης είναι 0,924. Για τα μικρότερα δείγματα, μεγέθους 50 καλύτερα φαίνεται ότι αποδίδει η μέθοδος percentile για B=5.000 επαναλήψεις , όπου η πιθανότητα κάλυψης είναι 0,947896.

Και εδώ το βασικό bootstrap διάστημα εμπιστοσύνης είναι αρκετά μακριά από το επιθυμητό επίπεδο κάλυψης, ενώ οι μέθοδοι bca και percentile δίνουν τα καλύτερα αποτελέσματα ακόμα και για μικρότερα μεγέθη δείγματος. Γενικά η μέθοδος basic δίνει πιθανότητες κάλυψης κάτω από το ονομαστικό επίπεδο, κάτι το οποίο ισχύει και με την norm, ενώ οι άλλες μέθοδοι σε μερικές περιπτώσεις το ξεπερνάνε.

Παραθέτουμε τον πίνακα με το μέσο πλάτος:

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=50	B=1000	0,7147598	0,698859	0,816952	0,69886	0,697033
	B=5000	0,710741	0,699611	0,808184	0,69961	**0,68939**
	B=10.000	0,716065	0,702972	0,806698	0,702975	**0,693422**
n=200	B=1000	0,350501	0,346355	0,378249	0,346351	**0,345482**
	B=5000	0,349512	0,345777	0,377093	0,345771	**0,344892**
	B=10.000	0,356646	0,350348	0,39002	0,350348	**0,348371**
n=500	B=1000	0,220529	**0,218112**	0,236509	0,218113	0,218216
	B=5000	0,224663	**0,222741**	0,239102	0,222744	0,340399
	B=10.000	0,222658	0,21998	0,237482	0,219983	**0,219857**

Πίνακας 3.12*-500 δοκιμές*

Ανάλογα συμπεράσματα, έχουμε και εδώ με την περίπτωση που οι δοκιμές ήταν 300. Τα στενότερα διαστήματα εμπιστοσύνης προκύπτουν με την μέθοδο Bca, ενώ σταθερά τα πιο πλατιά διαστήματα προκύπτουν με την μέθοδο t-bootstrap. Επίσης, η βασική μέθοδος, από τη μια δίνει αρκετά στενά διαστήματα εμπιστοσύνης, από την άλλη αποτυγχάνει να προσεγγίσει το ονομαστικό επίπεδο κάλυψης και είναι εμφανές ότι για να επιτύχουμε ικανοποιητικό επίπεδο κάλυψης θα πρέπει να πάρουμε μεγάλο δείγμα και να κάνουμε πολλές επαναλήψεις.

Θα προσπαθήσουμε στη συνέχεια να συγκρίνουμε τα αποτελέσματα που λάβαμε από τη μέθοδο bootstrap με αντίστοιχα αποτελέσματα που προκύπτουν από απλή προσομοίωση καθώς και με παραδοσιακές μεθόδους. Στόχος μας είναι να συγκρίνουμε τα διαστήματα εμπιστοσύνης ως προς το πλάτος. Αρχικά θα παραθέσουμε μερικά από τα αποτελέσματα που προέκυψαν από απλή προσομοίωση. Για την κατασκευή τους, παρήγαμε 300 δείγματα μεγέθους *n=50* και σε κάθε δείγμα βρήκαμε την διάμεσο. Κρατήσαμε αυτές τις τιμές και τις χρησιμοποιήσαμε για τον υπολογισμό του 95% διαστήματος εμπιστοσύνης. Επαναλάβαμε την διαδικασία για *n=200* και n=*500*. Λαμβάνουμε τα παρακάτω διαστήματα εμπιστοσύνης (Παράρτημα Π2):

		Σημειακή Εκτίμηση	(2.5% , 97.5%)	Πλάτος δ.ε.
300 Δοκιμές	n=50	-0.00989437	(-0.3139120 , 0.3038648)	0.6177768
	n=200	-0.004564359	(-0.1713420 , 0.1728675)	0.3442095
	n=500	0.005301973	(-0.1034876 , 0.1257844)	0.229272
500 Δοκιμές	n=50	-0.01181141	(-0.3387118 0.3054474)	0.6441592
	n=200	-0.00166967	(-0.1702327, 0.1581996)	0.3284323
	n=500	-0,0004.4775	(-0.1130758 , 0.1064830)	0.2195588

Πίνακας 3.13 *Διαστήματα εμπιστοσύνης για την διάμεσο της τυπικής κανονικής.*

Συγκρίνοντας τα αποτελέσματα με τους Πίνακες 3.6 και 3.12 παρατηρούμε ότι η παραπάνω διαδικασία δίνει ικανοποιητικά αποτελέσματα και παρόμοια με την μέθοδο bootstrap ιδιαίτερα για την περίπτωση που η bootstrap χρησιμοποιεί δείγματα μεγέθους n=200 και n=500 (και για τις μεθόδους percentile και bca). Για τα μικρότερα δείγματα η παραπάνω διαδικασία φαίνεται να αποδίδει καλύτερα από την bootstrap, αλλά παράλληλα το πλάτος ενός διαστήματος δεν είναι το μοναδικό κριτήριο αξιολόγησής του, και έτσι δεν μπορούμε με σιγουριά να αποφανθούμε, αν θα ήταν καλύτερα να αποφύγουμε το υπολογιστικό κόστος της bootstrap και να προτιμήσουμε απλή προσομοίωση.

Στη συνέχεια θα συγκρίνουμε τα παραπάνω αποτελέσματα και τα αποτελέσματα της bootstrap με κάποια πιο παραδοσιακή προσέγγιση κατασκευής διαστημάτων εμπιστοσύνης (π.χ. βλ. Wade-Koutoumanou 2010). Αν υποθέσουμε ότι έχουμε ένα πληθυσμό, η κατανομή του οποίου είναι άγνωστη, τότε κάθε τιμή του πληθυσμού μπορεί να συνδεθεί με κάποιο ποσοστιαίο σημείο. Για παράδειγμα, η τιμή x θα είναι το α-ποσοστιαίο σημείο (συμβολίζουμε x_α), όταν 100·α% των δεδομένων είναι μικρότερα του *x*. Συνεπώς, η πιθανότητα οι τιμές να είναι μεγαλύτερες του *x* είναι *1-α*. Για ένα σύνολο παρατηρήσεων $x_1, x_2, \ldots, x_n$ θα θεωρήσουμε τις τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ κάθε μία από τις οποίες παίρνει την i διατεταγμένη τιμή από τις αρχικές παρατηρήσεις. Η πιθανότητα το x_α ποσοστιαίο σημείο να βρίσκεται ανάμεσα στις X_j *και* X_κ , όπου *j<κ*, γράφεται ως $P(X_j \leq x_\alpha \leq X_\kappa)$ και επιπλέον παρατηρούμε ότι κάθε δειγματική τιμή θα είναι είτε κάτω από το x_α ποσοστιαίο σημείο, είτε θα το ξεπερνάει. Παρατηρούμε ότι η πιθανότητα η δειγματική τιμή να βρίσκεται κάτω απ' το ποσοστιαίο σημείο είναι σταθερή και ίση με *α,* το μέγεθος του δείγματος είναι σταθερό και ίσο με n και επιπλέον αν το δείγμα είναι τυχαία επιλεγμένο, τότε οι δειγματικές τιμές είναι ανεξάρτητες μεταξύ τους και έτσι μπορούμε να χρησιμοποιήσουμε την διωνυμική κατανομή με πιθανότητα επιτυχίας *p=α*. Συνεπώς η πιθανότητα το x_α να βρίσκεται ανάμεσα στις X_j *και* X_κ είναι:

$$P\left(X_j \leq x_\alpha \leq X_{\kappa-1}\right) = \sum_{i=j}^{\kappa-1} \binom{n}{i} p^i (1-p)^{n-i}, \text{ όπου } p=\alpha.$$

Υπό προϋποθέσεις και για μεγάλο μέγεθος δείγματος γνωρίζουμε, από το θεώρημα de Moivre-Laplace ότι διωνυμική κατανομή $B(n,p)$ προσεγγίζεται από την κανονική κατανομή με μέση τιμή $\mu = np$ και διασπορά $\sigma^2 = np(1-p)$. Για να είναι ικανοποιητική η προσέγγιση εφαρμόζουμε τον κανόνα $n \cdot p > 5$ και $n \cdot (1-p) > 5$. Έτσι μπορούμε να κατασκευάσουμε διαστήματα εμπιστοσύνης για τα ποσοστιαία σημεία χρησιμοποιώντας την κανονική κατανομή. Συγκεκριμένα το 95% διάστημα εμπιστοσύνης δίνεται από τον παρακάτω τύπο:

- Κάτω όριο : $\mu - 1{,}96 \cdot \sigma$
- Άνω όριο : $1 + \mu + 1{,}96 \cdot \sigma$

Σημειώνουμε ότι οι παραπάνω τύποι δεν μας δίνουν τα άκρα του διαστήματος εμπιστοσύνης, αλλά τις παρατηρήσεις που πρέπει να διαλέξουμε από τα δεδομένα μας, όταν αυτά βρίσκονται σε αύξουσα σειρά. Επιπλέον τα διαστήματα εμπιστοσύνης που προκύπτουν είναι προσεγγιστικά και όχι ακριβή. Στην περίπτωση μας το ποσοστιαίο σημείο που μας ενδιαφέρει είναι το 50$^{\text{ο}}$, οπότε για *p=0,5* έχουμε ότι $\mu = n \cdot 0{,}5$ και $\sigma = 0{,}5\sqrt{n}$. Δηλαδή το 95% διάστημα εμπιστοσύνης θα δίνεται από τις παρατηρήσεις:

- Κάτω όριο: $0{,}5 \cdot n - 1.96 \cdot 0{,}5\sqrt{n}$

- Άνω όριο: $1 + 0{,}5 \cdot n + 1.96 \cdot 0{,}5\sqrt{n}$

Στις δοκιμές μας θα χρησιμοποιήσουμε $n = 50, n = 200\ , n = 500\ και\ n = 1000$ και θα προσπαθήσουμε να συγκρίνουμε τα αποτελέσματα μας με τις παραπάνω μεθόδους .

Μέγεθος δείγματος	Παρατηρήσεις	Δ.ε.	*Πλάτος*
n=50	($18^η$,$33^η$)	(-0.67999570, 0.41307588)	1.093072
n=200	($86^η$,$115^η$)	(-0.37534493, 0.03297182)	0.4083168
n=500	($228^η$, $273^η$)	(-0.165695, 0.079129621)	0.2448246
n=1000	($469^η$,$532^η$)	(-0.079648083, 0.101385924)	0.181034

Πίνακας 3.14: *Διαστήματα εμπιστοσύνης για την διάμεσο, χρησιμοποιώντας την κανονική κατανομή.*

Από τον παραπάνω πίνακα παρατηρούμε ότι τα διαστήματα εμπιστοσύνης που προκύπτουν είναι ικανοποιητικά, ειδικά όταν αυξάνουμε το μέγεθος το δείγματος αρκετά, γεγονός που μας οδηγεί στο συμπέρασμα ότι δεν ήταν απαραίτητο να καταφύγουμε στην μέθοδο bootstrap για την κατασκευή των διαστημάτων εμπιστοσύνης και θα μπορούσαμε να είχαμε ακολουθήσει την παραδοσιακή οδό. Ωστόσο για μικρά δείγματα (n=50) η μέθοδος bootstrap (ειδικότερα η percentile και η Bca) υπερτερεί της παραπάνω διαδικασίας, δίνοντας στενότερα διαστήματα και μάλιστα με πιθανότητα κάλυψης πολύ κοντά στο ονομαστικό επίπεδο. Επιπλέον, ανάλογα με το μέγεθος δείγματος και το ποσοστιαίο σημείο που θέλουμε να εκτιμήσουμε δεν είμαστε πάντα σε θέση να εφαρμόζουμε τα παραπάνω διότι υπάρχουν κάποιοι περιορισμοί στην εφαρμογή του θεωρήματος de Moivre- Laplace, τους οποίους έχουμε ήδη αναφέρει.

3.2 Εφαρμογή της bootstrap για την διάμεσο της εκθετικής κατανομής

Σε αυτή την ενότητα θα επιχειρήσουμε και πάλι να δούμε πώς ανταποκρίνεται η μέθοδος bootstrap στην εκτίμηση της διαμέσου, μόνο που αυτή τη φορά τα δεδομένα μας θα προέρχονται απ' την εκθετική κατανομή με παράμετρο λ=1. Στη συνέχεια θα συγκρίνουμε τα διαστήματα εμπιστοσύνης, ως προς την πιθανότητα κάλυψης και το μέσο πλάτος και θα δούμε πως αυτά επηρεάζονται από την αλλαγή του αρχικού μεγέθους δείγματος *n* και του αριθμού *B* των επαναλήψεων. Η διαδικασία που θα ακολουθήσουμε είναι όμοια με αυτήν που περιγράφεται στην Ενότητα 3.1.

Ας αναφέρουμε ότι μια συνεχής τυχαία μεταβλητή που ακολουθεί την εκθετική κατανομή έχει συνάρτηση πυκνότητας:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Επιπλέον, λύνοντας την εξίσωση *F(x)=α,* το *α*- ποσοστιαίο σημείο της κατανομής δίνεται από την συνάρτηση:

$$F^{-1}(a) = \frac{-\ln(1-a)}{\lambda}, \text{ για } 0 < \alpha < 1$$

και επομένως η διάμεσος, που αποτελεί την παράμετρο που μας ενδιαφέρει υπολογίζεται από το πηλίκο $ln2/\lambda$. Στην περίπτωση μας , όπου χρησιμοποιήσαμε δεδομένα από την εκθετική με παράμετρο λ=1, η πραγματική τιμή της διαμέσου είναι $ln2 = 0{,}6931.$ Τα αρχικά αποτελέσματα μας θα βασιστούν σε 300 δοκιμές. Ας δούμε μερικά περιγραφικά μέτρα για τις σημειακές εκτιμήσεις :

- **300 Δοκιμές**

- *n=50*

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	0.6941	0.1558505	0.5897	0.6904	0.7822	0.3397	1.1220
B=5000	0.7005	0.1593942	0.5907	0.6872	0.8005	0.2807	1.2805
B=10.000	0.7160	0.1616184	0.5951	0.7055	0.8165	0.3441	1.2055

Πίνακας 3.15

- *n=200*

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	0.6951	0.0764983	0.6498	0.6883	0.7319	0.4822	0.9477
B=5000	0.6968	0.07258668	0.6457	0.6908	0.7480	0.5282	0.9444
B=10.000	0.6879	0.07807842	0.6345	0.6766	0.7368	0.5103	0.9209

Πίνακας 3.16

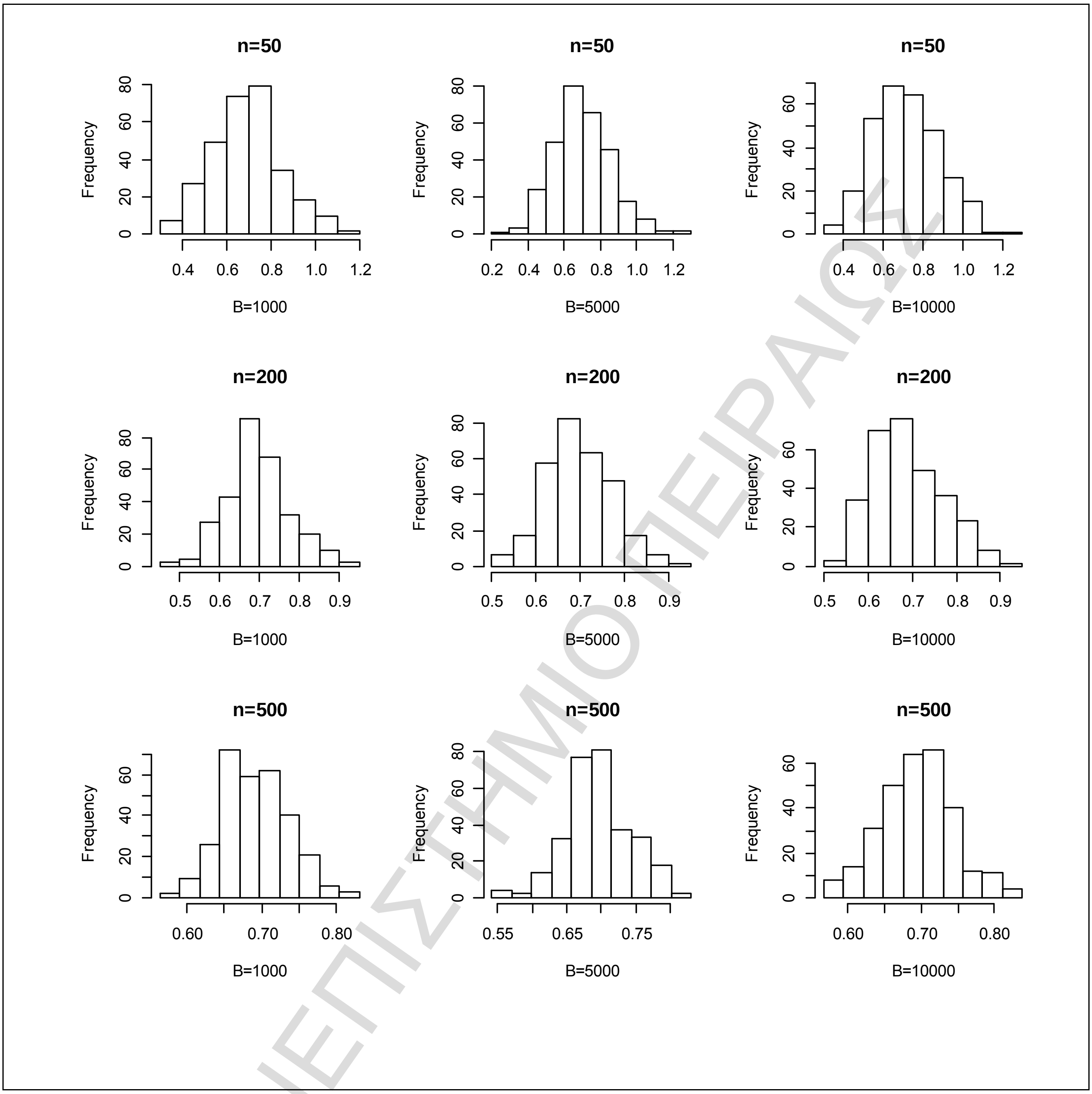
- *n=500*

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	0.6914	0.04478473	0.6588	0.6876	0.7225	0.5632	0.8318
B=5000	0.6959	0.0475141	0.6669	0.6946	0.7248	0.5412	0.8300
B=10.000	0.6947	0.05070491	0.6579	0.6932	0.7251	0.5679	0.8380

Πίνακας 3.17

Παρατηρούμε και εδώ ότι η αύξηση των bootstrap επαναλήψεων δεν μας οδηγεί σε μείωση της τυπικής απόκλισης, κάτι το οποίο είναι αναμενόμενο. Ωστόσο, η αύξηση του μεγέθους δείγματος μειώνει την τυπική απόκλιση της σημειακής εκτίμησης.

Στον παρακάτω πίνακα περιλαμβάνονται τα ιστογράμματα για τις 300 σημειακές εκτιμήσεις της διαμέσου. Επισημαίνουμε ότι κάθε σημειακή εκτίμηση είναι βασισμένη σε ένα δείγμα μεγέθους *n* και προέκυψε μετά από την κατασκευή *B* bootstrap δειγμάτων.



*Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.**Πίνακας 3.18:** Ιστογράμματα συχνοτήτων των $\widehat{\boldsymbol{\theta}}^*$για τη διάμεσο της εκθετικής, 300 δοκιμές*

Παρατηρούμε ότι στις σε αρκετές περιπτώσεις οι παρατηρήσεις μας δεν είναι συμμετρικά κατανεμημένες. Συγκεκριμένα στις περιπτώσεις n=500 και B=1000, n=200 και B=10.000, έχουμε έντονη θετική ασυμμετρία, ενώ ασύμμετρη είναι και η περίπτωση n=500 και B=1.000.

Στον επόμενο πίνακα θα παρουσιάσουμε την πιθανότητα κάλυψης για κάθε μια από τις μεθόδους. Αυτό που αναμένουμε είναι ότι οι μέθοδοι percentile και bca θα αποδίδουν καλύτερα απ’ τις υπόλοιπες διότι χρησιμοποιούν ποσοστιαία σημεία της κατανομής bootstrap.

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=50	B=1000	0,883333	0,843333	0,893333	0,93	0,936667
	B=5000	0,896667	0,84	0,926667	0,933333	0,943333
	B=10.000	0,92	0,836667	0,92	0,946667	**0,95**
n=200	B=1000	0,913333	0,893333	0,923333	0,936667	0,936667
	B=5000	0,92	0,903333	0,93	**0,95**	0,946667
	B=10.000	0,913333	0,863333	0,913333	0,936667	0,936667
n=500	B=1000	**0,95**	0,92	**0,943333**	0,963333	0,963333
	B=5000	0,916667	0,90	0,92	0,923333	0,923333
	B=10.000	0,916667	0,90	0,93	0,936667	0,943333

Πίνακας 3.19:*Πιθανότητες κάλυψης, για τη διάμεσο της εκθετικής, 300 δοκιμές. Με έντονα γράμματα σημειώνονται οι πιθανότητες κάλυψης που είναι κοντά στο 0,95.*

Το ονομαστικό επίπεδο των διαστημάτων εμπιστοσύνης που έχουμε ζητήσει είναι 0,95. Παρατηρούμε ότι για *n=50* , *B=10.000* και *n=200* , *B=5.000* το προσεγγιστικό επίπεδο εμπιστοσύνης, με βάση τις 300 δοκιμές είναι ακριβώς 95%. Επιπλέον πολύ καλά αποτελέσματα έχουμε για *n=200* & *n=500*, και *B=10.000*, χωρίς ωστόσο αυτό να σημαίνει ότι όσο αυξάνουμε το μέγεθος δείγματος και τις επαναλήψεις θα παίρνουμε καλύτερα αποτελέσματα. Παρόλα αυτά για πολύ μικρά μεγέθη δείγματος (n=50) τα αποτελέσματα είναι απογοητευτικά ειδικά στις μεθόδους, norm, basic και t-bootstrap και βελτιώνονται μόνο όταν αυξήσουμε το μέγεθος σε *n=200* ή *n=500*. Η μέθοδος basic και εδώ δίνει την χειρότερη πιθανότητα κάλυψης. Η μέθοδος t-bootstrap δίνει τα καλύτερα αποτελέσματα για *n=500* και *B=1.000.*

Στον επόμενο πίνακα έχουμε μερικά στοιχεία για το μέσο πλάτος τω διαστημάτων εμπιστοσύνης που προκύπτουν με την κάθε μέθοδο. Τα συμπεράσματα είναι και εδώ ανάλογα με την ενότητα 3.1. Η αύξηση του B δεν μας δίνει απαραίτητα στενότερα διαστήματα εμπιστοσύνης, αλλά η αύξηση του μεγέθους δείγματος ναι.

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=50	B=1000	0,55861495	0,545764	0,633521	0,545755	0,533856
	B=5000	0,560819	0,551546	0,643589	0,551557	0,533792
	B=10.000	0,579054	0,567318	0,666221	0,568854	0,551418
n=200	B=1000	0,278225	0,273833	0,305003	0,273828	0,27332
	B=5000	0,281271	0,276831	0,309529	0,276829	0,274917
	B=10.000	0,28172	0,274003	0,274005	0,274005	0,272845
n=500	B=1000	0,177065	**0,175769**	0,190681	**0,175767**	**0,175422**
	B=5000	0,1771023	0,176345	0,186985	0,17635	0,176086
	B=10.000	0,178908	0,176557	0,193778	0,176552	0,176099

Πίνακας 3.20:*Μέσο πλάτος διαστημάτων εμπιστοσύνης για την διάμεσο της εκθετικής, 300 δοκιμές, με έντονα γράμματα έχουμε τα δ.ε με το μικρότερο πλάτος.*

Η μέθοδος BCa δίνει, ανεξάρτητα απ' το μέγεθος δείγματος και τον αριθμό B των επαναλήψεων, τα λιγότερο πλατιά διαστήματα εμπιστοσύνης. Το μικρότερο μέσο πλάτος παρατηρείται με την μέθοδο BCa n=500, B=1.000, ενώ για τις ίδιες παραμέτρους πολύ λεπτά διαστήματα δίνουν και οι μέθοδοι basic και percentile. Το μεγαλύτερο μέσο πλάτος έχει η μέθοδος t-bootstrap, κάτι το οποίο παρατηρήσαμε και στην ενότητα 3.1 που τα δεδομένα μας προέρχονταν από την τυπική κανονική. Επίσης η μέθοδος basic παρόλο που απέτυχε να προσεγγίσει σωστά το ονομαστικό επίπεδο κάλυψης φαίνεται να δίνει πολύ μικρά διαστήματα εμπιστοσύνης.

Ας δούμε τα αντίστοιχα αποτελέσματα για 500 δοκιμές. Αρχίζουμε με μερικά περιγραφικά μέτρα:

- **500 Δοκιμές**
- *n=50*

	Μέση τιμή	Τυπική	$1^{ο}$ τεταρτ.	Διάμεσος	$3^{ο}$ τεταρτ.	min	max

		απόκλιση					
B=1000	0.6932	0.1536102	0.5832	0.6888	0.7836	0.3376	1.1475
B=5000	0.7054	0.1643851	0.5893	0.6938	0.8046	0.2807	1.3435
B=10.000	0.7002	0.1643688	0.5872	0.6858	0.8030	0.2192	1.2545

Πίνακας 3.21

- *n=200*

	Μέση τιμή	Τυπική απόκλιση	1^{o} τεταρτ.	Διάμεσος	3^{o} τεταρτ.	min	max
B=1000	0.6963	0.07445486	0.6498	0.6940	0.7403	0.4822	0.9477
B=5000	0.6951	0.07518754	0.6415	0.6911	0.7469	0.4677	0.9711
B=10.000	0.6877	0.07962712	0.6365	0.6827	0.7370	0.4834	1.0760

Πίνακας 3.22

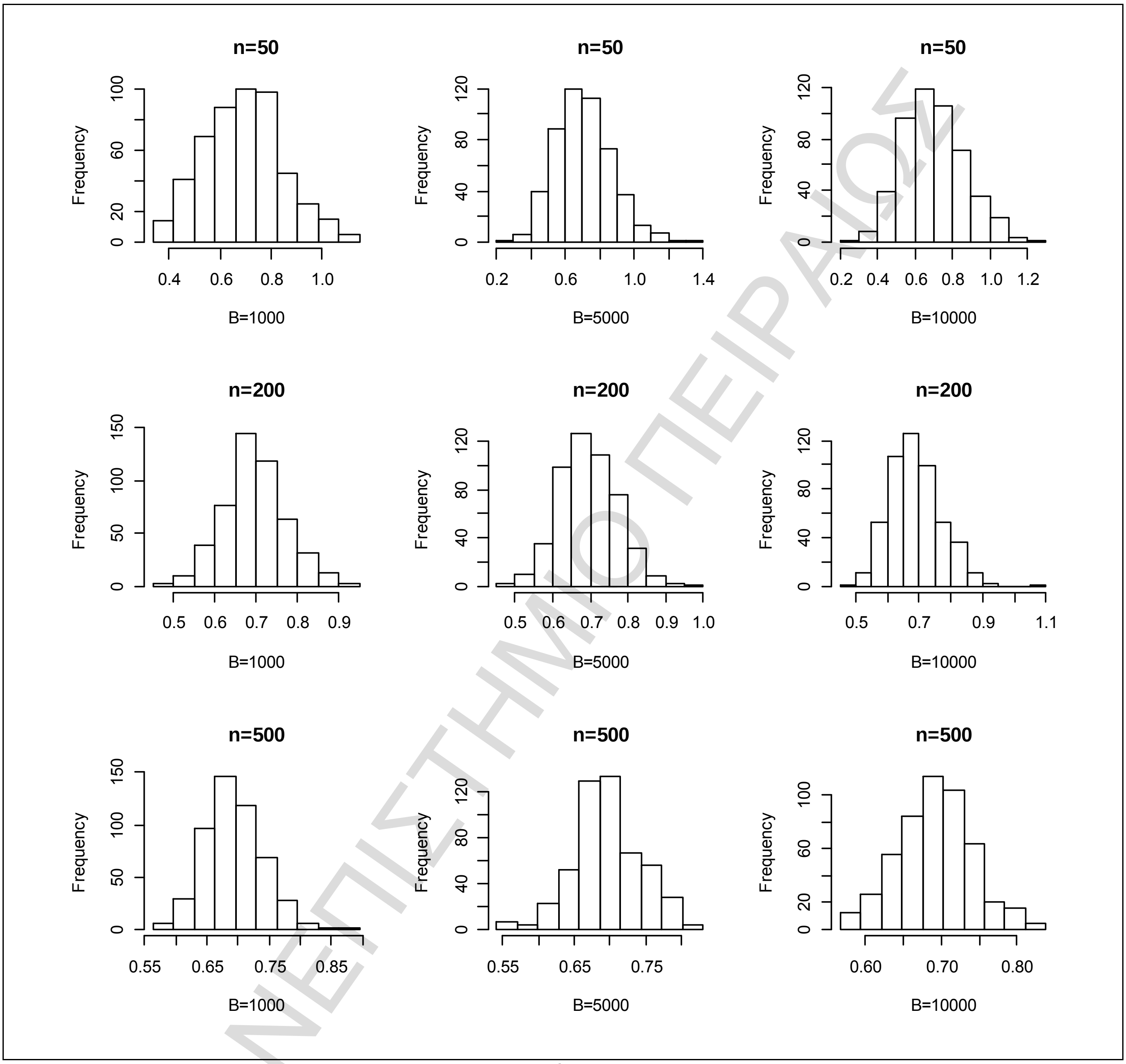
- *n=500*

	Μέση τιμή	Τυπική απόκλιση	1^{o} τεταρτ.	Διάμεσος	3^{o} τεταρτ.	min	max
B=1000	0.6929	0.0458358	0.6606	0.6892	0.7225	0.5632	0.8958
B=5000	0.6967	0.04738707	0.6684	0.6948	0.7257	0.5412	0.8300
B=10.000	0.6924	0.0483569	0.6571	0.6907	0.7221	0.5679	0.8380

Πίνακας 3.23

Τα συμπεράσματα είναι ανάλογα με αυτά που βγάλαμε στις 300 δοκιμές. Από ότι φαίνεται η μέση τιμή των σημειακών εκτιμήσεων είναι πολύ κοντά στην πραγματική τιμή της παραμέτρου (0,6931) για n=50 και B=1.000, ωστόσο η τυπική απόκλιση σε αυτήν την περίπτωση είναι μεγαλύτερη συγκριτικά με τις άλλες. Ικανοποιητικά είναι τα αποτελέσματα για n=200, αλλά ακόμα καλύτερα είναι τα αποτελέσματα για *n=500* , τόσο για B=1.000 όσο και B=10.000. Παρατηρούμε ότι η μέση τιμή των σημειακών εκτιμήσεων, αλλά και η διάμεσος βρίσκονται πολύ κοντά στην τιμή που θέλουμε και επιπλέον η τυπική απόκλιση είναι σαφέστατα μικρότερη συγκριτικά με τους Πίνακες 3.21,3.22. Μικρότερη τυπική απόκλιση έχουμε για *n=50* και *B=1.000* .

Στην συνέχεια παραθέτουμε τα αντίστοιχα ιστογράμματα των σημειακών εκτιμήσεων:



Πίνακας 3.24 *Ιστογράμματα συχνοτήτων των* $\widehat{\boldsymbol{\theta}}^{*}$ *για τη διάμεσο της εκθετικής, 500 δοκιμές*

Τα ιστογράμματα είναι αντίστοιχα με αυτά των 300 δοκιμών, ωστόσο εξακολουθούν να υπάρχουν ασυμμετρίες όπως για n=200 και B=1.000 και n=500, B=1.000. Ακολουθεί ο πίνακας με την πιθανότητα κάλυψης:

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=50	B=1000	0,898	0,838	0,91	0,934	0,938
	B=5000	0,888	0,838	0,926	0,938	**0,942**
	B=10.000	0,908	0,838	0,918	0,938	0,94
n=200	B=1000	0,922	0,898	0,93	**0,946**	0,94
	B=5000	0,922	0,906	0,93	**0,95**	**0,948**
	B=10.000	0,908	0,864	0,904	0,932	0,93
n=500	B=1000	**0,944**	0,914	0,936	0,958	**0,956**
	B=5000	0,916	0,90	0,916	0,924	0,924
	B=10.000	0,924	0,908	0,932	**0,944**	0,972

Πίνακας 3.25: *Πιθανότητα κάλυψης δ.ε. για τη διάμεσο της εκθετικής-500 δοκιμές*

Όπως βλέπουμε στον παραπάνω πίνακα η μέθοδος basic αποτυγχάνει να προσεγγίσει ικανοποιητικά το ονομαστικό επίπεδο κάλυψης. Συγκεκριμένα, φτάνει μόλις στο 0,914 για n=500 και B=1.000. Η μέθοδος t-bootstrap δίνει τα καλύτερα αποτελέσματα για n=500και B=10.000, ενώ με την μέθοδο percentile για n=200 και B=5.000 (για αυτές τις 500 δοκιμές) έχουμε πιθανότητα κάλυψης στο 0.95, συμπέρασμα που συμφωνεί και με τον πίνακα των 300 δοκιμών. Για το ίδιο μέγεθος δείγματος και αριθμό επαναλήψεων η μέθοδος BCa δίνει επίσης πολύ καλά αποτελέσματα (0.948). Παρατηρούμε επίσης ότι το φαινόμενο η προσεγγιστική πιθανότητα κάλυψης να ξεπεράσει το ονομαστικό επίπεδο είναι περιορισμένο και συμβαίνει σε ελάχιστες περιπτώσεις.

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=50	B=1000	0,5595776	0,547234	0,638247	0,547186	0,53665
	B=5000	0,567127	0,555988	0,651554	0,555989	0,53844
	B=10.000	0,57415	0,560335	0,658058	0,561258	0,544346
n=200	B=1000	0,278226	0,274311	0,303363	0,274309	0,273456
	B=5000	0,283227	0,27836	0,314246	0,27836	0,276582
	B=10.000	0,283193	0,277073	0,277073	0,277072	0,27557
n=500	B=1000	0,177812	0,176675	0,191207	0,176675	0,176519
	B=5000	0,1768328	0,176226	0,186635	0,176231	0,175871
	B=10.000	0,177874	0,175682	0,192296	0,175678	**0,175263**

Πίνακας 3.26:*Μέσο πλάτος-500 δοκιμές*

Από τον *Πίνακα 3.2*6 μπορούμε για άλλη μια φορά να παρατηρήσουμε πως τα διαστήματα εμπιστοσύνης γίνονται πιο στενά όσο αυξάνεται το μέγεθος του αρχικού δείγματος. Τα συμπεράσματα μας είναι ανάλογα με αυτά του Πίνακα 3.20. Το μικρότερο πλάτος, στο σύνολο των δοκιμών παρατηρείται στον παραπάνω πίνακα και είναι 0,175263 για n=500 και B=10.000. Γενικότερα μπορούμε να πούμε, ότι όσον αφορά το μέσο πλάτος των διαστημάτων η μέθοδος bootstrap δούλεψε ικανοποιητικά για την εκθετική κατανομή, ωστόσο όσον αφορά την πιθανότητα κάλυψης, τα αποτελέσματα δεν ήταν ικανοποιητικά, ιδιαίτερα για μικρά δείγματα και ειδικότερα όταν επιχειρήσαμε να κατασκευάσουμε το κανονικοποιημένο διάστημα, το βασικό και το t-bootstrap.

3.3 Εφαρμογή της bootstrap για την διάμεσο της ομοιόμορφης κατανομής

Σε αυτή την ενότητα θα ακολουθήσουμε παρόμοια διαδικασία και χρησιμοποιώντας τη μέθοδο bootstrap θα κάνουμε σημειακές εκτιμήσεις και θα κατασκευάσουμε διαστήματα εμπιστοσύνης για την διάμεσο της ομοιόμορφης κατανομής. Θα πραγματοποιήσουμε 500 δοκιμές, λαμβάνοντας σε κάθε δοκιμή ένα δείγμα μεγέθους *n* από την ομοιόμορφη κατανομή *U(0,1)*. Σε κάθε μία από τις δοκιμές, θα λαμβάνουμε ένα δείγμα, από το οποίο θα κατασκευάζουμε *B* νέα δείγματα, μεγέθους n πραγματοποιώντας δειγματοληψία με επανάθεση. Από κάθε μία δοκιμή παίρνουμε μια σημειακή εκτίμηση για την διάμεσο και πέντε διαφορετικά διαστήματα εμπιστοσύνης με ονομαστικό πιθανότητα κάλυψης 95%. Σημειώνουμε ότι η ομοιόμορφη κατανομή *U(α,β)* έχει συνάρτηση πυκνότητας:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}\,, & \alpha \le x \le \beta \\ 0, & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

και η διάμεσος της δίνεται από το τύπο: *(α+β)/2*. Η διάμεσος της *U(0,1)* είναι το 0,5 και ταυτίζεται με την μέση τιμή. Ας δούμε μερικά περιγραφικά μέτρα για τις σημειακές εκτιμήσεις της διαμέσου:

- n=50

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	0.5000	0.0794277	0.4490	0.4945	0.5501	0.2905	0.7547
B=5000	0.5007	0.0754179	0.4479	0.5047	0.5547	0.2801	0.7178
B=10.000	0.4981	0.0765245	0.4449	0.4998	0.5454	0.2772	0.7518

Πίνακας 3.27

- n=200

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	0.4996	0.03697857	0.4746	0.4987	0.5262	0.3623	0.6120
B=5000	0.4981	0.03710071	0.3660	0.4969	0.5218	0.3660	0.6096
B=10.000	0.5016	0.03756796	0.4743	0.4998	0.5284	0.4042	0.6606

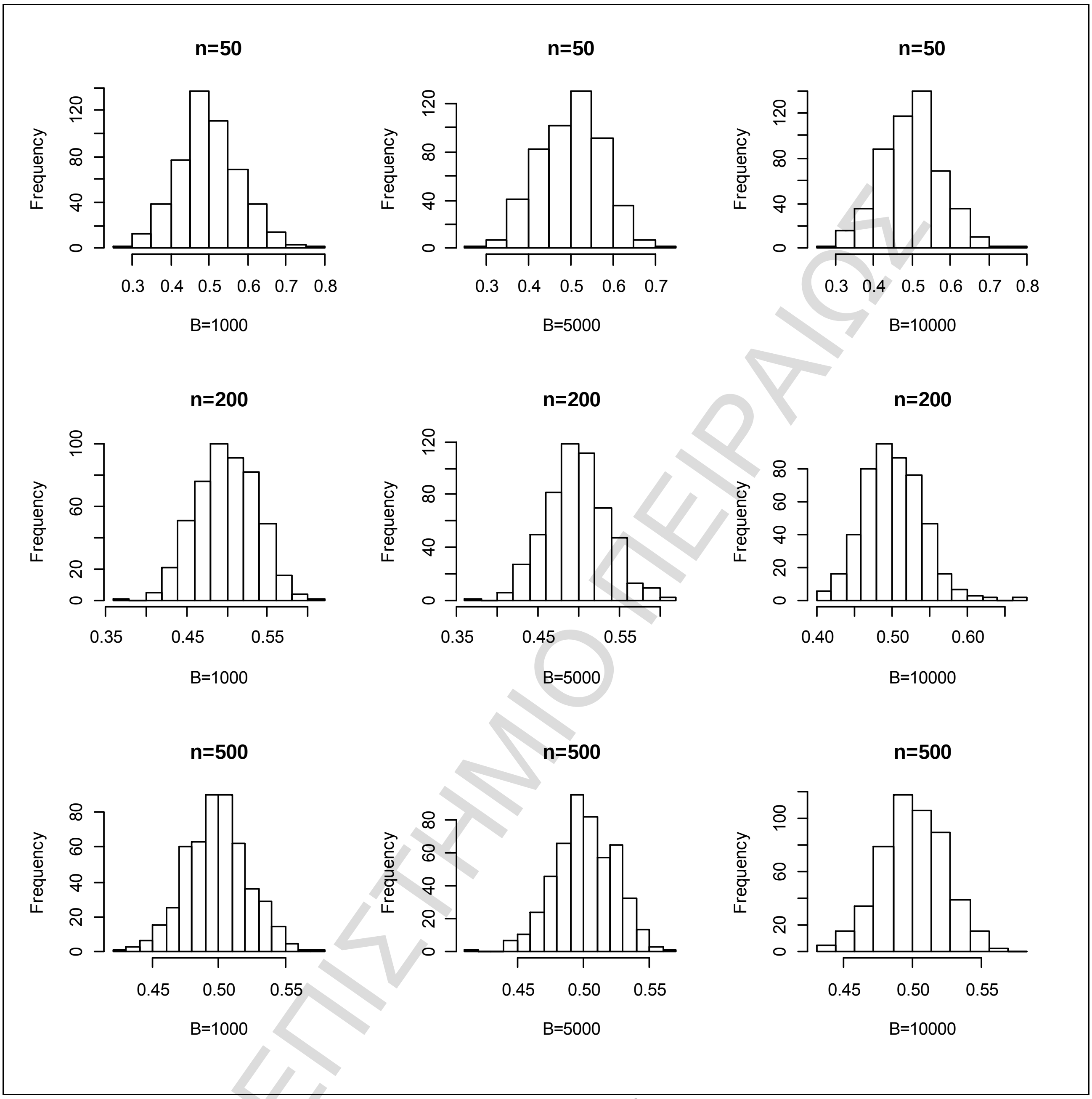
Πίνακας 3.28

- n=500

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	0.4984	0.02355074	0.4832	0.4981	0.5133	0.4255	0.5759
B=5000	0.5007	0.02244084	0.4865	0.5002	0.5168	0.4158	0.5658
B=10.000	0.5004	0.02292556	0.4851	0.5000	0.5160	0.4303	0.5586

Πίνακας 3.29

Παρατηρούμε ότι οι σημειακές εκτιμήσεις είναι πολύ κοντά στην πραγματική τιμή της παραμέτρου (0,5) ακόμα και όταν έχουμε μικρά δείγματα και πραγματοποιούμε λίγες επαναλήψεις. Επιπλέον η τυπική απόκλιση είναι αρκετά μικρή και μικραίνει σταθερά όταν αυξάνεται το μέγεθος δείγματος (για δεδομένο *B*). Συγκριτικά με τις προηγούμενες κατανομές που ερευνήσαμε, παρατηρούμε ότι η τυπική απόκλιση των σημειακών εκτιμήσεων είναι μικρότερη όταν τα δεδομένα μας προέρχονται από την ομοιόμορφη *U(0,1)*.



Πίνακας.3.30 *Ιστογράμματα συχνοτήτων των* $\widehat{\boldsymbol{\theta}}^*$, ***U(0,1).*** *500 δοκιμές*

Πιο συμμετρικό από τα παραπάνω ιστογράμματα φαίνεται να είναι αυτό για *n=500* και *B=1000,* ενώ τα ιστογράμματα για μικρά δείγματα (*n=50*) φαίνονται να είναι λιγότερο συμμετρικά. Τα ιστογράμματα για *n=200* είναι ικανοποιητικά. Στον επόμενο πίνακα δίνεται η πιθανότητα κάλυψης για κάθε μια από μεθόδους:

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=50	B=1000	0,9	0,818	0,916	0,946	0,948
	B=5000	0,924	0,868	0,936	0,952	0,948
	B=10.000	0,900391	0,837891	0,916016	0,951172	0,947266
n=200	B=1000	0,939638	0,905433	0,937626	0,94165	0,95171
	B=5000	0,912477	0,877095	0,925512	0,927374	0,932961
	B=10.000	0,921277	0,895745	0,921277	0,940426	0,942553
n=500	B=1000	0,936	0,918	0,938	**0,95**	0,946
	B=5000	**0,944223**	0,920319	0,950199	0,956175	0,956175
	B=10.000	0,938	0,912	0,942	**0,95**	0,946

Πίνακας 3.31: *Πιθανότητα Κάλυψης δ.ε. για τη διάμεσο της U(0,1). 500 δοκιμές*

Για μικρό μέγεθος δείγματος η πιθανότητα κάλυψης απέχει αρκετά από το ονομαστικό επίπεδο (0,95) και μόνο οι μέθοδοι percentile και bca το προσεγγίζουν αρκετά ακόμα και αν πραγματοποιήσουμε λίγες επαναλήψεις (replications). Επιπλέον παρατηρούμε ότι αυτές οι δύο μέθοδοι σε αρκετές περιπτώσεις ξεπερνάνε το ονομαστικό επίπεδο κάλυψης κάτι το οποίο δεν συμβαίνει για την μέθοδο norm και basic. Ειδικότερα το basic διάστημα εμπιστοσύνης φαίνεται να έχει τις χειρότερες επιδόσεις, ωστόσο είναι εμφανές ότι όσο μεγαλώνουμε το μέγεθος του δείγματος η μέθοδος δίνει καλύτερα αποτελέσματα. Αν συγκρίνουμε μάλιστα τα αποτελέσματα της basic με τα αντίστοιχα της κανονικής κατανομής (Πίνακας 3.11) και της εκθετικής κατανομής (Πίνακας 3.25) βλέπουμε ότι για την ομοιόμορφη τα αποτελέσματα μας είναι λίγο καλύτερα απ’ την εκθετική, αλλά όχι τόσο ικανοποιητικά όσο αυτά της τυποποιημένης κανονικής.

Στη συνέχεια παραθέτουμε τον πίνακα με το μέσο πλάτος των διαστημάτων εμπιστοσύνης:

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=50	B=1000	0,274212	0,264829	0,324182	0,264825	0,264139
	B=5000	0,273182	0,268049	0,317998	0,268044	0,264735
	B=10.000	0,269396	0,264549	0,3117	0,264551	0,260041
n=200	B=1000	0,1389618	0,138131	0,149458	0,138101	0,141829
	B=5000	0,13874	0,137079	0,151485	0,137076	0,136782
	B=10.000	0,139536	0,137001	0,153949	0,136996	0,136754
n=500	B=1000	0,088916	0,088251	0,095926	0,088251	0,088312
	B=5000	0,086765	0,086364	0,092294	**0,086362**	0,086445
	B=10.000	0,088907	0,08796	0,095382	0,087959	0,087865

Πίνακας 3.32 *Μέσο πλάτος δ.ε. για τη διάμεσο της U(0,1). 500 δοκιμές*

Παρατηρούμε ότι η ομοιόμορφη κατανομή μας δίνει τα διαστήματα εμπιστοσύνης με το μικρότερο πλάτος για τη διάμεσο. Όπως είναι αναμενόμενο η αύξηση του μεγέθους δείγματος μας οδηγεί σε διαστήματα εμπιστοσύνης με μικρότερο πλάτος , κάτι το οποίο δεν συμβαίνει απαραιτήτως και με την αύξηση του B. Το μικρότερο πλάτος παρατηρείται για την μέθοδο percentile *(n=500, B=5.000)* και είναι 0,086362. Διαστήματα με πολύ μικρό πλάτος παράγει και η μέθοδος basic, αν και απ' ότι είδαμε στον Πίνακα 3.31 υστερεί στην πιθανότητα κάλυψης. Η μέθοδος t-bootstrap δίνει πλατύτερα διαστήματα εμπιστοσύνης σε σχέση με τις άλλες μεθόδους.

Κεφάλαιο 4ο

Στο 4ο κεφάλαιο θα εφαρμόσουμε τη μέθοδο bootstrap, επαναλαμβάνοντας τις διαδικασίες που ακολουθήσαμε στο Κεφάλαιο 3ο , μόνο που αυτή τη φορά σκοπός μας θα είναι η σημειακή εκτίμηση του 95ου ποσοστιαίου σημείου καθώς και η κατασκευή διαστημάτων εμπιστοσύνης. Σε ένα δεύτερο στάδιο μας ενδιαφέρει να ελέγξουμε εάν η μέθοδος αποδίδει διαφορετικά για το 95ο ποσοστιαίο σημείο, σε σύγκριση με το 50ο .Θα χρησιμοποιήσουμε δεδομένα από την τυπική κανονική, την εκθετική κατανομή και την κατανομή Pareto.

4.1 Εφαρμογή της bootstrap για την εκτίμηση του 95ου ποσοστημορίου της τυπικής κανονικής κατανομής

Για τα παρακάτω αποτελέσματα πραγματοποιήσαμε 300 δοκιμές, πήραμε δηλαδή 300 διαφορετικά δείγματα από την κανονική κατανομή και σε κάθε δοκιμή ζητούσαμε αντίστοιχα 1.000, 5.000 και 10.000 επαναλήψεις της μεθόδου bootstrap. Κάθε δοκιμή μας δίνει μια σημειακή εκτίμηση για το 95ο ποσοστιαίο σημείο και ένα διάστημα εμπιστοσύνης με ονομαστική πιθανότητα κάλυψης 95%. Το αρχικό δείγμα σε κάθε δοκιμή είναι *n=200* ή *n=500* και όχι μικρότερο. Η πραγματική τιμή για το 95ο ποσοστημόριο της τυποποιημένης κανονικής είναι 1.6449. Ας δούμε μερικά περιγραφικά μέτρα για τις σημειακές εκτιμήσεις:

- **n=200**

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	1.628	0.1696111	1.507	1.625	1.742	1.134	2.104
B=5000	1.613	0.1662583	1.499	1.613	1.720	1.203	2.119
B=10.000	1.630	0.1766094	1.507	1.614	1.751	1.109	2.190

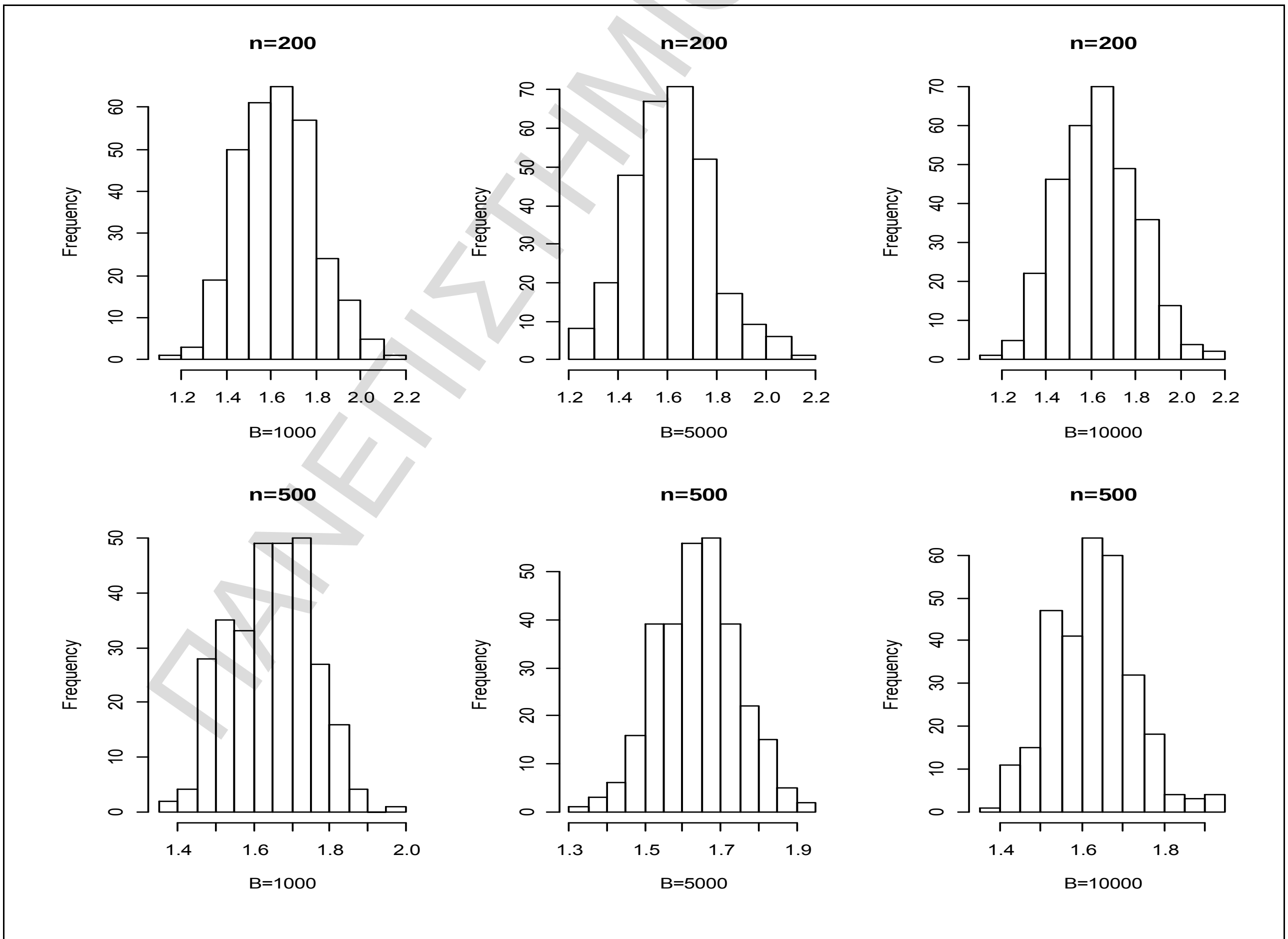
Πίνακας 4.1.

- **n=500**

	Μέση τιμή	Τυπική απόκλιση	1ο τεταρτ.	Διάμεσος	3ο τεταρτ.	min	max
B=1000	1.643	0.1073602	1.565	1.650	1.724	1.388	1.962
B=5000	1.641	0.1050934	1.571	1.645	1.714	1.330	1.938
B=10.000	1.626	0.09895976	1.551	1.629	1.679	1.382	1.933

Πίνακας 4.2

Παρατηρούμε ότι η μέση τιμή των σημειακών εκτιμήσεων είναι αρκετά κοντά στην τιμή 1.6449, ειδικά όταν το αρχικό μας δείγμα είναι μεγάλο (n=500). Για n=200 οι σημειακές εκτιμήσεις είναι λίγο πιο μακριά από την πραγματική τιμή της παραμέτρου και η τυπική απόκλιση είναι μεγαλύτερη. Η καλύτερη εκτίμηση είναι για n=200 και B=5.000, όπου παρατηρούμε ότι τόσο η μέση τιμή όσο και η διάμεσος των σημειακών εκτιμήσεων είναι πολύ κοντά στο 1,6449. Στη συνέχεια παραθέτουμε και τα αντίστοιχα ιστογράμματα:



Πίνακας 4.3 *Ιστογράμματα συχνοτήτων των $\widehat{\boldsymbol{\theta}}^{*}$,300 δοκιμές από την τυπική κανονική*

Το ιστόγραμμα για n=200 και B=10.000 φαίνεται ότι είναι καλύτερο συγκριτικά με τα αντίστοιχα ιστογράμματα που χρησιμοποιούν το ίδιο μέγεθος δείγματος, αλλά έχουν μικρότερο αριθμό επαναλήψεων Β. Ωστόσο, συνολικά τα ιστογράμματα για n=500 είναι πιο συμμετρικά σε σχέση με αυτά όπου το n=200. Ίσως είναι λίγο καλύτερο το ιστόγραμμα για n=500 και B=5.000

Ας δούμε στον επόμενο πίνακα τα αποτελέσματα για την πραγματική πιθανότητα κάλυψης. Ικανοποιητικά θεωρούνται τα αποτελέσματα που βρίσκονται κοντά στο 0,95. Έχουμε παραλείψει τα αποτελέσματα για δείγματα μεγέθους n=50, διότι ένα τέτοιο δείγμα θα έδινε απογοητευτικά αποτελέσματα.

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=200	B=1000	0,900	0,817	0,850	**0,943**	**0,947**
	B=5000	0,900	0,843	0,870	0,926	0,926
	B=10.000	0,880	0,832	0,841	0,909	0,900
n=500	B=1000	0,893	0,852	0,866	0,933	0,933
	B=5000	0,903	0,853	0,870	0,927	**0,937**
	B=10.000	0,930	0,867	0,883	**0,960**	0,963

Πίνακας 4.5 : *Πιθανότητα κάλυψης δ.ε για το $95^{ο}$ ποσοστημόριο της τυπικής κανονικής*

Από τον παραπάνω πίνακα παρατηρούμε ότι οι μέθοδοι basic και t-bootstrap αποδίδουν καλύτερα όσο μεγαλώνουμε το μέγεθος δείγματος, αλλά και πάλι ακόμα και για πολλές επαναλήψεις δεν προσεγγίζουν ικανοποιητικά το 0,95. Η norm δίνει καλύτερα αποτελέσματα από τις basic και t- bootstrap(stud) , αλλά και πάλι τα αποτελέσματα είναι αρκετά μακριά από το ονομαστικό επίπεδο κάλυψης και μόνο για n=500 και B=10.000 το κανονικοποιημένο διάστημα φτάνει στο 0,93. Οι μέθοδοι bca και percentile εμφανώς υπερτερούν από τις άλλες τρεις χωρίς όμως να δίνουν τόσο καλά αποτελέσματα όπως στην περίπτωση εκτίμησης του $50^{ου}$ ποσοστιαίου σημείου (Πίνακας 3.5 και 3.11). Παρατηρούμε ότι ακόμα και για δείγματα n=200 και n=500 εμφανίζονται πιθανότητες κάλυψης κάτω από το 0.93 κάτι το οποίο δεν υπήρχε ως φαινόμενο όταν μελετούσαμε τη διάμεσο. Και σε αυτήν την περίπτωση όμως οι μέθοδοι percentile και bca εμφανίζουν πιθανότητα κάλυψης μεγαλύτερη από το ονομαστικό επίπεδο.

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=200	B=1000	0,579567	0,563457	0,595233	0,563437	0,559293
	B=5000	0,591094	0,56299	0,600498	0,562896	0,563161
	B=10.000	0,601806	0,572935	0,612806	0,573777	0,564803
n=500	B=1000	0,380503	0,368265	0,387836	0,368279	0,369953
	B=5000	0,378713	0,36402	0,385247	0,36399	0,37234
	B=10.000	0,38105	**0,36401**	0,385397	0,36402	0,37276

Πίνακας 4.6: *Μέσο πλάτος των δ.ε για το 95^{ο} ποσοστημόριο της τυπικής κανονικής*

Το μέσο πλάτος των διαστημάτων εμπιστοσύνης μειώνεται όσο μεγαλώνουμε το μέγεθος του αρχικού δείγματος, αλλά δεν μειώνεται όσο αυξάνουμε τον αριθμό B των επαναλήψεων. Το μικρότερο μέσο πλάτος έχει το βασικό bootstrap διάστημα εμπιστοσύνης, ωστόσο η ίδια μέθοδος αποτυγχάνει στο να προσεγγίσει το ονομαστικό επίπεδο κάλυψης. Ικανοποιητικά διαστήματα, με κριτήριο το μέσο πλάτος δίνει και η μέθοδος percentile. Συγκριτικά με τα αντίστοιχα αποτελέσματα για τη διάμεσο (Πίνακας 3.6 και 3.12) τα παραπάνω διαστήματα έχουν ελαφρώς μεγαλύτερο πλάτος.

Ας δούμε τώρα μερικά αποτελέσματα με μια παραδοσιακή προσέγγιση, χρησιμοποιώντας τους τύπους που εμφανίζονται στην Ενότητα 3.1. Για να κατασκευάσουμε το 95% διάστημα εμπιστοσύνης για το 95^{ο} ποσοστιαίο σημείο θα θεωρήσουμε ότι η κατανομή του δείγματος προσεγγίζεται από την διωνυμική με πιθανότητα επιτυχίας *p=0.95*. Το μέγεθος δείγματος που θα πάρουμε θα είναι n=200, n=500 και n=1.000 . Επειδή οι υποθέσεις $n \cdot p > 5$ και $n \cdot (1 - p) > 5$ ισχύουν μπορούμε να προσεγγίσουμε την διωνυμική κατανομή με την τυποποιημένη κανονική και να χρησιμοποιήσουμε τους παρακάτω τύπους:

- Κάτω όριο : $\mu - 1{,}96 \cdot \sigma$
- Άνω όριο : $1 + \mu + 1{,}96 \cdot \sigma$

Τα διαστήματα εμπιστοσύνης που προκύπτουν φαίνονται στον παρακάτω πίνακα:

Μέγεθος δείγματος	Παρατηρήσεις	Δ.ε.	*Πλάτος*
n=200	($184^η$,$197^η$)	(1.461934933, 2.179482550)	0.7175476
n=500	($465^η$, $486^η$)	(1.571519377, 2.009101373)	0.437582
n=1000	($935^η$, $965^η$)	(1.476609292, 1.805040354)	0.3284311

Πίνακας 4.7

Στον παραπάνω πίνακα παρατηρούμε ότι τα διαστήματα εμπιστοσύνης έχουν μεγαλύτερο μέσο πλάτος σε σχέση με αυτά που προκύπτουν με τη μέθοδο bootstrap. Όπως έχουμε υπογραμμίσει όμως και στο προηγούμενο κεφάλαιο το μέσο πλάτος δεν είναι το μοναδικό κριτήριο αξιολόγησης ενός διαστήματος εμπιστοσύνης, μας ενδιαφέρει σίγουρα και το πραγματικό επίπεδο κάλυψης. Εάν επαναλάβουμε τη διαδικασία αρκετές φορές θα διαπιστώσουμε ότι, τουλάχιστον για τα μεγέθη δείγματος που επιλέξαμε τα αποτελέσματα ποικίλουν και μερικές φορές δεν περιέχουν καν την πραγματική τιμή της παραμέτρου. Αυτό είναι αναμενόμενο διότι τα αποτελέσματα εξαρτώνται σε μεγάλο βαθμό από το αρχικό δείγμα, κάτι το οποίο δεν είναι ισχύει όταν εφαρμόζουμε την μέθοδο bootstrap λόγω του ότι κατασκευάζουμε νέα δείγματα.

4.2 Εφαρμογή της bootstrap για την εκτίμηση του $95^{ου}$ ποσοστημορίου της εκθετικής κατανομής

Σε αυτή την ενότητα θα χρησιμοποιήσουμε δεδομένα από την εκθετική κατανομή και η παράμετρος που μας ενδιαφέρει είναι το $95^ο$ ποσοστιαίο σημείο. Πραγματοποιούμε 300 δοκιμές, και σε κάθε δοκιμή λαμβάνουμε ένα τυχαίο δείγμα από την εκθετική κατανομή (λ=1), μεγέθους n, με τη βοήθεια του οποίου κατασκευάζουμε *B* νέα δείγματα πραγματοποιώντας δειγματοληψία με επανάθεση. Σε κάθε ένα από τα *B* δείγμα βρίσκουμε το 95^{ο} ποσοστιαίο σημείο και από εκεί οδηγούμαστε σε σημειακές εκτιμήσεις και διαστήματα εμπιστοσύνης. Αναμένουμε ότι τα αποτελέσματα μας δεν θα είναι τόσο ικανοποιητικά όσο τα αντίστοιχα για την διάμεσο διότι όπως αναφέρεται στην βιβλιογραφία (Chernick 1999 και Davison & Hinkley 1997) η μέθοδος bootstrap δεν αποδίδει το ίδιο καλά όταν μελετάμε ακραίες τιμές όπως είναι για παράδειγμα 95^{ο} ή ακόμα και το 99^{ο} ποσοστημόριο. Ακόμα και με μεγάλα δείγματα και αρκετές επαναλήψεις μπορούμε οδηγηθούμε σε ανακριβείς εκτιμήσεις και αυτό διότι αυτές οι στατιστικές συναρτήσεις επηρεάζονται πιο εύκολα από ακραίες παρατηρήσεις.

Στην περίπτωση μας, για την πραγματική τιμή της παραμέτρου που θέλουμε να εκτιμήσουμε, χρησιμοποιούμε και πάλι τον τύπο:

$$F^{-1}(a) = \frac{-\ln(1-a)}{\lambda}$$

και θέτοντας $\lambda=1$ και $\alpha=0,95$, το 95^{ο} ποσοστιαίο σημείο προκύπτει ότι είναι ίσο με 2,9957.

Ας δούμε πρώτα μερικά περιγραφικά μέτρα για τις σημειακές εκτιμήσεις:

- n=200

	Μέση τιμή	Τυπική απόκλιση	1^{ο} τεταρτ.	Διάμεσος	3^{ο} τεταρτ.	min	max
B=1000	2.991	0.3638736	2.740	2.970	3.231	1.966	4.609
B=5000	2.948	0.3607221	2.716	2.930	2.716	1.956	3.931
B=10.000	2.943	0.3444781	2.686	2.902	3.132	2.135	4.095

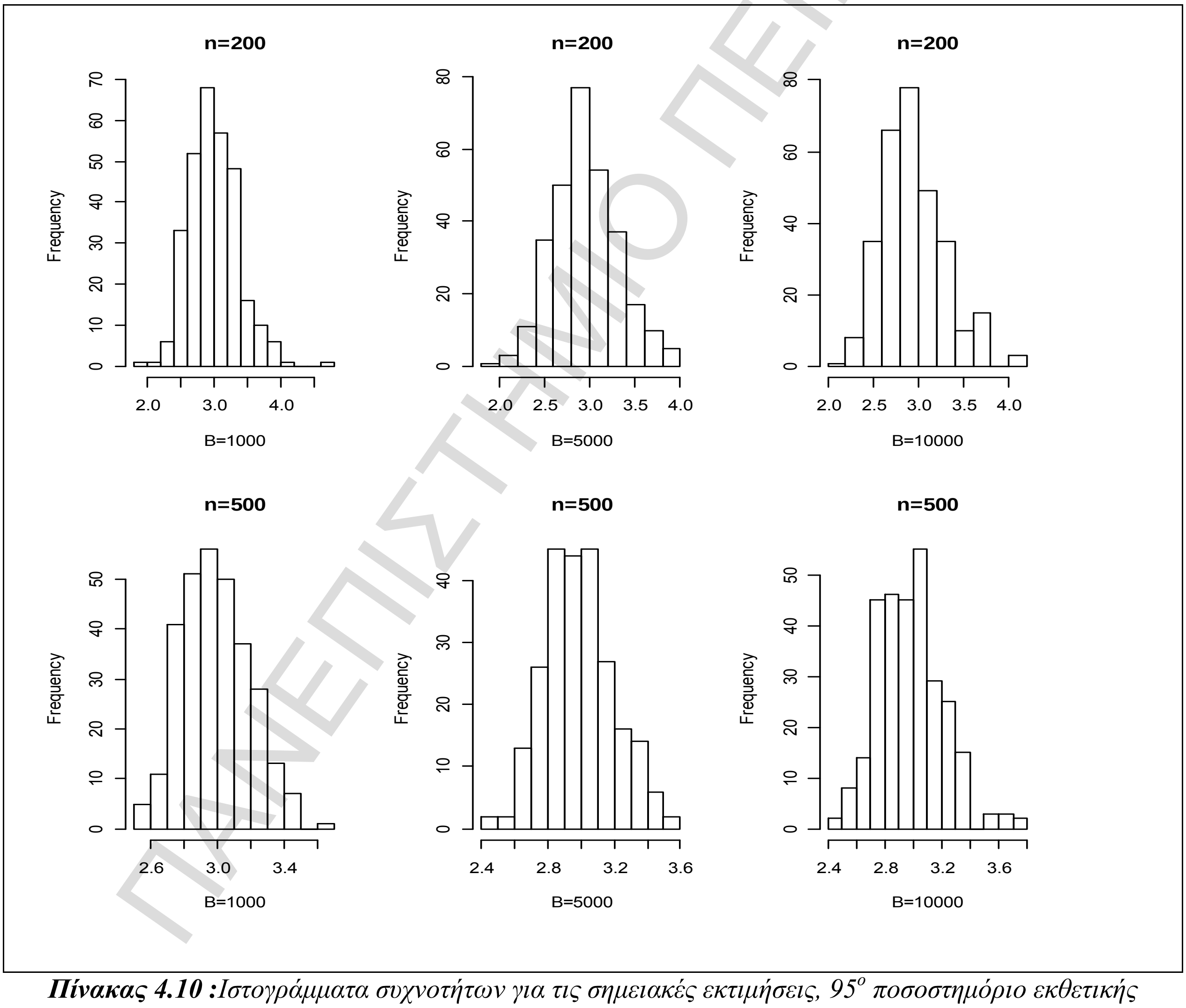
Πίνακας 4.8

- n=500

	Μέση τιμή	Τυπική απόκλιση	1^{ο} τεταρτ.	Διάμεσος	3^{ο} τεταρτ.	min	max
B=1000	2.990	0.1984952	2.846	2.981	3.137	2.512	3.608
B=5000	2.986	0.2067732	2.841	2.969	3.104	2.414	3.597
B=10.000	2.979	0.2261378	2.804	2.961	3.115	2.438	3.787

Πίνακας 4.9

Οι μέση τιμή των σημειακών εκτιμήσεων είναι πολύ κοντά στην πραγματική τιμή της παραμέτρου ειδικά όταν το αρχικό δείγμα είναι μεγέθους 200 ή 500, ακόμα και όταν το B=1.000. Ωστόσο παρατηρούμε ότι οι πλέον οι σημειακές εκτιμήσεις έχουν μεγαλύτερη τυπική απόκλιση, από τα αντίστοιχα αποτελέσματα για την διάμεσο (Πίνακες 3.12-3.14), η οποία μειώνεται αρκετά όταν το μέγεθος του αρχικού δείγματος αυξάνεται. Και σε αυτόν τον πίνακα τα αποτελέσματα επιβεβαιώνουν ότι η αύξηση του B δεν συνεπάγεται την μείωση της τυπικής απόκλισης.



Πίνακας 4.10 :*Ιστογράμματα συχνοτήτων για τις σημειακές εκτιμήσεις, 95^{o} ποσοστημόριο εκθετικής κατανομής*

Στον παραπάνω πίνακα έχουμε μερικά ιστογράμματα για τις σημειακές εκτιμήσεις. Τα ιστογράμματα παρουσιάζουν ασυμμετρίες ειδικά αυτά που έχουν *n=200* και ιδιαίτερα αν τα συγκρίνουμε αυτά της Ενότητας 3.2. Λίγο καλύτερα φαίνονται να είναι τα διάγραμματα για *n=200* και *B=5.000*, αλλά σε καμία περίπτωση δεν είναι συμμετρικό.

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=200	B=1000	0,916667	0,84	0,886667	**0,943333**	0,94
	B=5000	0,875	0,799	0,845	0,924	0,924
	B=10.000	0,905229	0,820261	0,879085	0,915033	0,908497
n=500	B=1000	0,940	0,890	0,913	0,973	0,970
	B=5000	0,921	0,860	0,901	0,930	**0,942**
	B=10.000	0,925	0,870	0,887	0,921	0,932

Πίνακας 4.11: *Πιθανότητες κάλυψης δ.ε για το 95^{o} ποσοστημόριο της εκθετικής κατανομής. Με έντονη γραμματοσειρά είναι οι τιμές που βρίσκονται κοντά στο 0,95.*

Στον παραπάνω πίνακα έχουμε συνοπτικά τις πιθανότητες κάλυψης για κάθε μια από τις πέντε μεθόδους και τη δυνατότητα να ελέγξουμε πως επηρεάζονται τα αποτελέσματα μας ανάλογα με τις αλλαγές στο μέγεθος του δείγματος και στον αριθμό των επαναλήψεων. Και εδώ δεν συμπεριλάβαμε τις δοκιμές για τα μικρά δείγματα (n=50) γιατί ήταν απογοητευτικά, κάτι το οποίο ήταν αναμενόμενο, αλλά συνολικά τα αποτελέσματα μας είναι εμφανώς χειρότερα από τα αντίστοιχα που λάβαμε για την εκτίμηση της διαμέσου της εκθετικής (Πίνακας 3.19). Τα καλύτερα αποτελέσματα λαμβάνουμε για *n=200* και *B=1000* καθώς και *n=500* B=*5.000*. Παρατηρούμε ότι ακόμα και οι μέθοδοι percentile και bca που στις προηγούμενες ενότητες έδιναν καλά αποτελέσματα ακόμα και μικρά δείγματα, πλέον απαιτούν μεγαλύτερο μέγεθος δείγματος. Επιπλέον βλέπουμε ότι η πιθανότητα κάλυψης σπάνια πλέον ξεπερνάει το ονομαστικό επίπεδο, κάτι που ήταν πιο σύνηθες στο προηγούμενο κεφάλαιο.

Στη συνέχεια έχουμε μερικά αποτελέσματα για το μέσο πλάτος των διαστημάτων εμπιστοσύνης:

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=200	B=1000	1,27106	1,22391	1,42043	1,223887	1,227473
	B=5000	1,234	1,180	1,348	1,180	1,181
	B=10.000	1,225908	1,186925	1,326974	1,186915	1,174399
n=500	B=1000	0,777	0,760	0,844	0,760	0,768
	B=5000	0,783215	0,755975	0,839368	0,756012	0,77293
	B=10.000	0,774	**0,741**	0,847	**0,741**	0,768

Πίνακας 4.12 *Μέσο πλάτος δ.ε για το 95^o ποσοστημόριο της εκθετικής.*

Παρατηρούμε ότι τα πλάτη των διαστημάτων εμπιστοσύνης είναι μεγάλα για *n=200* και μικραίνουν αρκετά όταν αυξάνουμε το μέγεθος δείγματος σε *n=500.* Το μικρότερο πλάτος παρατηρείται για *n=500* και *B=10.000* με τις μεθόδους basic και percentile bootstrap, ενώ η μέθοδος t-bootstrap (stud) δίνει τα πιο πλατιά διαστήματα. Ωστόσο συγκριτικά με τους πίνακες 3.17 και 3.25 όπου έχουμε τα αντίστοιχα αποτελέσματα για την διάμεσο, τα διαστήματα εμπιστοσύνης του Πίνακα 4.12 είναι πιο πλατιά, γεγονός που επιβεβαιώνει ότι η μέθοδος bootstrap δεν αποδίδει το ίδιο καλά για το 95^o ποσοστιαίο σημείο.

4.3 Εφαρμογή της bootstrap για την εκτίμηση του 95^{ου} ποσοστημορίου της κατανομής Pareto

Στην τελευταία ενότητα θα μελετήσουμε δεδομένα από την κατανομή Pareto. Γνωρίζουμε ήδη ότι η μέθοδος bootstrap και η τεχνική της επαναδειγματοληψίας δεν αποδίδει ικανοποιητικά όταν μελετάμε ακραίες τιμές. Επιπλέον όμως, αναφέρεται στην βιβλιογραφία (Hall–Yi Jing 1998) όταν η δειγματική κατανομή είναι μια κατανομή με βαριά ουρά (heavy-tailed), όπως η κατανομή Pareto,η μέθοδος bootstrap αποτυγχάνει να προσεγγίσει την κατανομή του δειγματικού μέσου και αυτό γιατί υπάρχει μεγάλη πιθανότητα μέσα στο δείγμα να βρίσκεται κάποια ακραία τιμή. Μια κατανομή λέγεται ότι έχει βαριά ουρά όταν συγκλίνει πολύ αργά στο μηδέν για $x \to \infty$. Σκοπός μας είναι, αφενός να δούμε την απόδοση της bootstrap στο 95^o

ποσοστιαίο σημείο και αφετέρου να ελέγξουμε εάν η επιλογή μιας κατανομής με βαριά ουρά επιδείνωσε τα αποτελέσματα μας.

Η συνάρτηση πυκνότητας μιας τυχαίας μεταβλητής που ακολουθεί την κατανομή Pareto με παραμέτρους α και β, δίνεται από τον παρακάτω τύπο:

$$f(x) = \frac{\alpha\beta^{\alpha}}{(\beta+x)^{\alpha+1}} \text{ , για } x \geq 0.$$

Η αθροιστική συνάρτηση της κατανομής θα είναι :

$$F(x) = 1 - \frac{\beta^{a}}{(\beta+x)^{a}} \text{ , για } x \geq 0.$$

Η μέση τιμή μιας τυχαίας μεταβλητής που ακολουθεί την κατανομή Pareto με παραμέτρους α και β είναι:

$$E(X) = \frac{\beta}{\alpha-1},$$

ενώ η διακύμανση δίνεται από τον τύπο:

$$V(X) = \frac{\alpha\beta^{2}}{(\alpha-1)^{2}(\alpha-2)}.$$

Ικανή συνθήκη για την ύπαρξη της μέσης τιμής είναι η τιμή της παραμέτρου α να είναι μεγαλύτερη της μονάδας ($\alpha>1$) και για να έχουμε πεπερασμένη διακύμανση απαιτείται το α να είναι μεγαλύτερο του 2 ($\alpha>2$) .

Επιλέξαμε να πάρουμε δεδομένα από την κατανομή Pa(3,3) για την οποία η πραγματική τιμή του 95^{ου} ποσοστιαίου σημείου είναι 5,1433. Χρησιμοποιήσαμε το πακέτο actuar στην R, στο οποίο βρίσκεται ορισμένη η Pareto. Πραγματοποιήσαμε 250 δοκιμές, και σε κάθε δοκιμή πήραμε ένα τυχαίο δείγμα μεγέθους *n* από την Pa(3,3). Σε κάθε δοκιμή πραγματοποιούμε *B* επαναλήψεις της μεθόδου bootstrap και παίρνουμε μια σημειακή εκτίμηση για το 95^{ο} ποσοστιαίο σημείο και ένα διάστημα εμπιστοσύνης με ονομαστικό επίπεδο κάλυψης 95%. Στους παρακάτω πίνακες έχουμε μερικά περιγραφικά μέτρα για τις σημειακές εκτιμήσεις που λάβαμε, μετά από 250 δοκιμές (Παράρτημα Π3):

- **n=200**

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	5.089	0.9880915	4.347	5.006	5.689	4.347	9.026
B=5000	5.021	0.8236704	4.497	4.984	5.542	3.121	7.723
B=10.000	5.065	0.8831327	4.440	4.984	5.682	3.221	7.949

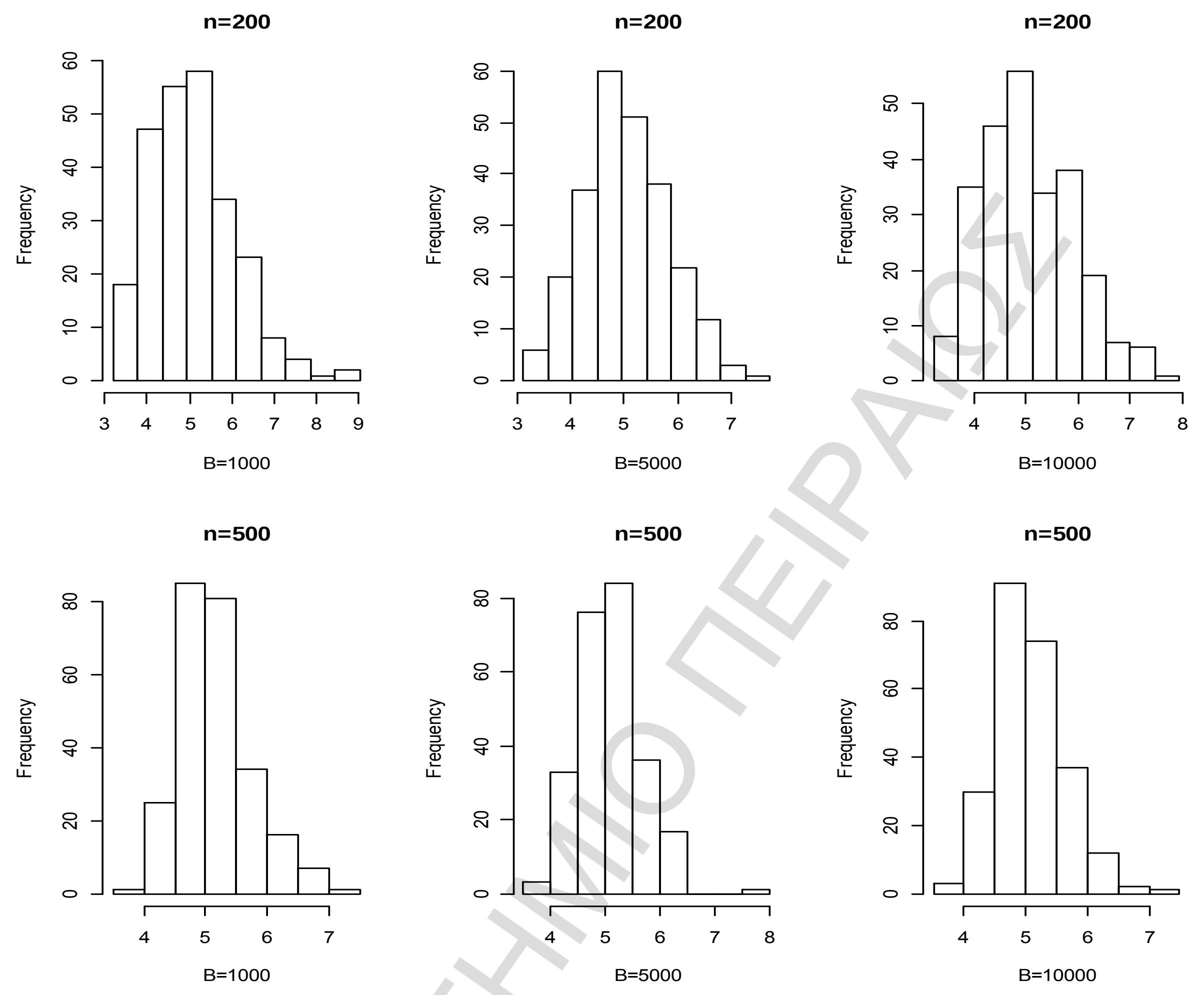
Πίνακας 4.13

- **n=500**

	Μέση τιμή	Τυπική απόκλιση	1^o τεταρτ.	Διάμεσος	3^o τεταρτ.	min	max
B=1000	5.137	0.5822395	4.713	5.056	5.455	3.813	7.454
B=5000	5.103	0.5661506	4.713	5.078	5.440	3.896	7.651
B=10.000	5.073	0.5401398	4.670	5.016	5.413	3.856	7.041

Πίνακας 4.14

Παρατηρούμε ότι τα αποτελέσματα που προέκυψαν από το δείγματα μεγέθους *n=200* δεν είναι τόσο κοντά στην τιμή 5,1433 όσο και αυτά που έχουν προκύψει από δείγματα μεγέθους *n=500* και επιπλέον έχουν μεγαλύτερη τυπική απόκλιση. Παραδόξως, η εκτίμηση που προκύπτει για *n=500* και *B=1.000* είναι πιο κοντά στην πραγματική τιμή της παραμέτρου, και όχι οι τιμές εκείνες που προκύπτουν μετά από 5.000 και 10.00 επαναλήψεις της μεθόδου bootstrap. Στον επόμενο πίνακα έχουμε τα αντίστοιχα ιστογράμματα:



Πίνακας 4.15 : *Ιστογράμματα συχνοτήτων για τις σημειακές εκτιμήσεις του 95^{ου} ποσοστημορίου, κατανομή Pareto(3,3)*

Τα παραπάνω ιστογράμματα, όπως είναι αναμενόμενο φαίνονται ασύμμετρα. Ικανοποιητικό είναι αυτό για n=500 και B=1.000 και στη συνέχεια εκείνο για n=500 και B=5.000. Δεν είναι τυχαίο άλλωστε που αυτές οι δύο περιπτώσεις φαίνεται να δίνουν τις καλύτερες σημειακές εκτιμήσεις (Πίνακας 4.14). Μετά τις σημειακές εκτιμήσεις θα μελετήσουμε τα αντίστοιχα διαστήματα εμπιστοσύνης:

Πιθανότητα κάλυψης						
		norm	basic	stud	perc	bca
n=200	B=1000	0,884	0,82	0,912	0,92	0,916
	B=5000	0,924	0,852	0,908	0,932	**0,948**
	B=10.000	0,936	0,820	0,932	**0,952**	0,948
n=500	B=1000	0,920	0,868	0,924	0,956	**0,952**
	B=5000	0,928	0,868	0,912	0,944	0,956
	B=10.000	0,936	0,872	0,932	**0,952**	0,964

Πίνακας 4.16: *Πιθανότητα κάλυψης δ.ε. για το 95^o ποσοστημόριο της Pa(3,3)*

Στον παραπάνω πίνακα είναι εμφανές ότι η απόδοση της μεθόδου basic είναι φτωχή. Η πραγματική πιθανότητα κάλυψης απέχει πολύ από το ονομαστικό επίπεδο, παρόλο που είναι η εμφανής η τάση που έχει η μέθοδος να βελτιώνεται όσο αυξάνουμε το μέγεθος δείγματος. Το κανονικοποιημένο διάστημα εμπιστοσύνης απέχει πολύ απ0ό το 0.95, ειδικά όταν το δείγμα έχει μέγεθος 200 και σταματάμε την επαναδειγματοληψία στις 1.000 επαναλήψεις. Ωστόσο φαίνεται ότι είτε το μέγεθος δείγματος είναι στις 200 παρατηρήσεις, είτε στις 500 η αύξηση των επαναλήψεων βελτιώνει αρκετά την πιθανότητα κάλυψης. Η μέθοδος t-bootstrap είναι καλύτερη από τις norm και basic, άλλα όχι τόσο καλή όσο η percentile και η Bca. Οι δύο τελευταίες σταθερά επιτυγχάνουν καλή πιθανότητα κάλυψης. Στον Πίνακα 4.16 βλέπουμε ότι σε αρκετές περιπτώσεις είμαστε κοντά στο 0.95 . Αυτό που παρατηρείται σε αυτή τη δοκιμή είναι ότι μερικές φορές ξεπερνάμε ελαφρά το ονομαστικό επίπεδο κάλυψης, κάτι το οποίο δεν συναντήσαμε ιδιαίτερα συχνά στους Πίνακες 4.5 και 4.11, όπου τα δεδομένα προέρχονται από την κανονική και την εκθετική κατανομή αντίστοιχα.

Στη συνέχεια έχουμε τον πίνακα με το μέσο πλάτος των διαστημάτων εμπιστοσύνης:

Μέσο πλάτος						
		norm	basic	stud	perc	bca
n=200	B=1000	3,597268	3,4677	5,067272	3,46776	3,574004
	B=5000	3,482824	3,320744	4,980396	3,320748	3,37664
	B=10.000	3,591892	3,413344	5,259816	3,413368	3,437128
n=500	B=1000	2,163684	2,09642	3,33136	2,0964	2,133732
	B=5000	2,119792	**2,030224**	3,076332	**2,030268**	2,117688
	B=10.000	2,143448	2,031876	3,240588	2,031864	2,119352

Πίνακας 4.16: *Μέσο πλάτος δ.ε για το 95^{ο} ποσοστημόριο της Pa(3,3)*

Από τα στοιχεία του παραπάνω πίνακα μπορούμε να δούμε ότι η κατανομή Pareto μας έδωσε τα πιο πλατιά διαστήματα εμπιστοσύνης. Το πλάτος των διαστημάτων εμπιστοσύνης μειώνεται όταν αυξάνουμε το μέγεθος του δείγματος. Το μικρότερο πλάτος παρατηρείται για την μέθοδο basic για μέγεθος δείγματος n=500 και B=5.000. Πολύ μικρό πλάτος δίνει και η μέθοδος percentile για n=500 και B=5.000. Η μέθοδος t-bootstrap δίνει, με διάφορα από τις άλλες- τα πιο πλατιά διαστήματα εμπιστοσύνης, κάτι το οποίο έχουμε δει να συμβαίνει σε αρκετές περιπτώσεις, απλά εδώ το φαινόμενο είναι πιο έντονο. Η μέθοδος Bca σε όλες τις περιπτώσεις δίνει πλατύτερα διαστήματα εμπιστοσύνης από την percentile.

Είναι εμφανές ότι τα αποτελέσματα για την κατανομή Pareto δεν ήταν τόσο ικανοποιητικά όσο για τις προηγούμενες κατανομές, αλλά και συνολικά τα αποτελέσματα για το 95^{ο} ποσοστιαίο σημείο δεν ήταν τόσο καλά όσο για το 50^{ο} . Αν διαλέγαμε να εκτιμήσουμε το 99^{ο} ποσοστιαίο σημείο τα αποτελέσματα που θα περιμέναμε θα ήταν ακόμα χειρότερα. Ωστόσο, χρησιμοποιήσαμε μια μέθοδο που παρότι ήταν χρονοβόρα, μας απάλλαξε από δύσκολους υπολογισμούς.

Κεφάλαιο 5ο

Είναι σαφές ότι η μέθοδος bootstrap από το 1979, οπότε και εισήχθη με την μορφή που της έδωσε ο Efron, δεν έχει σταματήσει να αναπτύσσεται, ταυτόχρονα με την ανάπτυξη της τεχνολογίας που πλέον μας επιτρέπει να κάνουμε προσομοίωση και να κατασκευάζουμε νέα δείγματα μέσω της επαναδειγματοληψίας πολύ ταχύτερα απ' ότι στο παρελθόν. Πλέον, μπορούμε να επεξεργαζόμαστε ζητήματα, χωρίς να εξαρτόμαστε από πολύπλοκους μαθηματικούς τύπους. Όπως αναφέρουν και οι Efron –Tibshirani(1993), η μέθοδος bootstrap καλείται να υπολογίσει γνώριμα στοιχεία της στατιστικής, όπως είναι το τυπικό σφάλμα, η μεροληψία, τα διαστήματα εμπιστοσύνης κ.α. με έναν τρόπο που δεν είναι και τόσο γνώριμος! Ωστόσο δεν πρέπει σε καμία περίπτωση δεν πρέπει να δημιουργείται η εντύπωση ότι η μέθοδος μπορεί να εφαρμοστεί παντού με απόλυτη επιτυχία.

Η εφαρμογή της μεθόδου εξαρτάται από την φύση του προβλήματος. Θα πρέπει σε κάθε περίπτωση η φύση του προβλήματος να είναι τέτοια που να αξίζει να προτιμήσουμε τη bootstrap από κάποια άλλη παραδοσιακή προσέγγιση. Για παράδειγμα η μη παραμετρική bootstrap μπορεί να εφαρμοστεί όταν έχουμε αμφιβολίες για την ισχύ παραμετρικών υποθέσεων, μιας και αν εφαρμόσουμε κάποιο λανθασμένο πιθανοθεωρητικό μοντέλο για τον πληθυσμό, τα αποτελέσματα που θα προκύψουν θα είναι παραπλανητικά. Ένα ερώτημα που δημιουργείται είναι αν θα πρέπει να χρησιμοποιηθεί η μέθοδος bootstrap ή κάποια άλλη αντίστοιχη τεχνική όπως η jackknife. Επιπλέον ακόμα κι αν επιλέξουμε να εφαρμόσουμε την bootstrap, παραμένει το ερώτημα αν αυτό θα γίνει παραμετρικά, μη παραμετρικά η χρησιμοποιώντας κάποια από τις παραλλαγές της μεθόδου που παρουσιάστηκαν αργότερα. Η αλήθεια είναι ότι δεν υπάρχει εύκολη απάντηση σε αυτό το ερώτημα. Αυτό που είναι αποδεκτό είναι ότι αν δεν συντρέχει κάποια από τις περιπτώσεις που αναφέρονται στην Ενότητα 1.9, η μη παραμετρική μέθοδος δίνει ικανοποιητικά αποτελέσματα όταν το μέγεθος του δείγματος είναι μεγάλο και στην πράξη, η παραμετρική μέθοδος αποφεύγεται όταν μπορούμε να χρησιμοποιήσουμε την μέγιστη πιθανοφάνεια για να εκτιμήσουμε μια παράμετρο, παρόλο που όπως έχουμε αναφέρει (Ενότητα 1.6) η παραμετρική bootstrap θα έδινε καλύτερα αποτελέσματα. Στους Davison & Hinkley (1997) υπάρχει ένα παράδειγμα στο οποίο χρησιμοποιείται η παραμετρική bootstrap για να ελέγξει την εγκυρότητα του παραμετρικού μοντέλου. Όσον αφορά τη χρήση της bootstrap έναντι της jackknife στον Chernick(2008) , αναφέρεται ότι είναι προτιμότερη η χρήση της bootstrap όταν θέλουμε να εκτιμήσουμε το τυπικό σφάλμα.

Τέλος να αναφέρουμε ότι υπάρχουν αρκετές παραλλαγές της μεθόδου, που δεν περιέχονται στην παρούσα εργασία, αλλά ενδέχεται, ανάλογα με την περίπτωση να είναι καταλληλότερες από την κλασική bootstrap. Επιγραμματικά να αναφέρουμε την Μπεϋζιανή bootstrap (Rubin 1981), την ομαλή bootstrap (smoothed bootstrap, Efron 1982) και η m-από-n bootstrap (Athreya 1987).

Παράρτημα

Π1:

Ενδεικτικά, δίνουμε τον παρακάτω κώδικα που χρησιμοποιήθηκε για την κατασκευή 300 διαστημάτων εμπιστοσύνης , με δείγματα μεγέθους n=200, από την κανονική κατανομή και B=5000 επαναλήψεις. Για την κατασκευή των studentized διαστημάτων εμπιστοσύνης απαιτείται να κάνουμε bootstrap μέσα στην bootstrap. Εδώ διαλέξαμε να κάνουμε 50 επαναλήψεις.

```
library(boot)
N.CI <-300
list.CI <- list(N.CI)
for (i in 1:N.CI ) {
x<-rnorm (200,0,1)
myfunction<- function(x, idx)
    {boot.sample<-x[idx]
   m <- median(boot.sample)
    N<-length(boot.sample)
    R<-50
    reboot.medians<-rep(0,R)
  for (k in 1:R){
      reboot.sample<-sample(boot.sample,replace=T)
      reboot.medians[k]<-median(reboot.sample)
  }
reboot.samples<-matrix(data=sample(boot.sample, size=R*N,replace=T),nrow=N)
reboot.medians<-apply(reboot.samples,MARGIN=2,median)
m.var<-var(reboot.medians)
results<-c(m,m.var)
}

b<-boot(x,statistic=myfunction,R=5000)
list.CI[[i]] <-boot.ci(b,conf=0.95,type=c("norm","basic","stud", "perc","bca"))
}
list.CI
```

Π2: Ενδεικτικά με τον παρακάτω κώδικα κατασκευάστηκαν R σετ με τυχαία δείγματα μεγέθους B. Σε κάθε σετ κρατήσαμε την διάμεσο και στη συνέχεια από όλα τα σετ πήραμε μία μέση τιμή για την διάμεσο και το 95% δ.ε.

```
B<-200   #μέγεθος δείγματος#
R<-500   #αριθμός δοκιμών#
A<-matrix(nrow=B, ncol=R)
for (i in 1:R ) {
x<-rnorm(B,0,1)
A[,i]<-x
}
mymed<-matrix(1,R)
for (i in 1:R ){
mymed[i]<-median(A[,i])
}
mymed
N<-quantile(mymed, c(0.025, 0.975))
N
mean(mymed)
```

Π3: Ενδεικτικά δίνουμε τον κώδικα που χρησιμοποιήθηκε για να κατασκευάσουμε 200 διαστήματα εμπιστοσύνης, από την κατανομή Pareto(3,3) με τη βοήθεια του πακέτου actuar. Στο παράδειγμα το μέγεθος δείγματος είναι n=200 και B=5.000

```
library(actuar)
library(boot)

N.CI <-250
list.CI <- list(N.CI)
for (i in 1:N.CI ) {
x<-rpareto(200,3,3)
myfunction<- function(x, idx)
   {boot.sample<-x[idx]
  m <- quantile((boot.sample),0.95 )
   N<-length(boot.sample)
   R<-50
   reboot.quantiles<-rep(0,R)
  for (k in 1:R){
```

```
    reboot.sample<-sample(boot.sample,replace=T)
    reboot.quantiles[k]<-quantile((reboot.sample),0.95)
  }
reboot.samples<-matrix(data=sample(boot.sample, size=R*N,replace=T),nrow=N)
reboot.quantiles<-apply(reboot.samples,MARGIN=2,quantile)
m.var<-var(reboot.quantiles)
results<-c(m,m.var)
}
b<-boot(x,statistic=myfunction,R=5000)
list.CI[[i]] <-boot.ci(b,conf=0.95,type=c("norm","basic","stud", "perc","bca"))
}
list.CI
```

Βιβλιογραφία

Ελληνική Βιβλιογραφία

Ηλιόπουλος Γ. (2006) Βασικές μέθοδοι εκτίμησης παραμετρων, Εκδόσεις Σταμούλη

Μπούτσικας Μ. (2004), Μέθοδοι Προσομοίωσης και υπολογιστικές στατιστικές τεχνικές, Πανεπιστημιακές Σημειώσεις για το ΠΜΣ Εφαρμοσμένης Στατιστικής του Πανεπιστημίου Πειραιά.

Κουρούκλης Σ. (2006) , Στατιστική Ι, Πανεπιστημιακές Παραδόσεις, Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών

Κούτρας Μ. (2004), Εισαγωγή στις Πιθανότητες, Μέρος Ι, Εκδόσεις Σταμούλη

Κούτρας Μ. (2005), Εισαγωγή στις Πιθανότητες, Μέρος 2, Εκδόσεις Σταμούλη

Τουλούμης Α. (2006), Εφαρμογή της τεχνικής bootstrap σε διακριτά δεδομένα, Διπλωματική εργασία για το ΠΜΣ Εφαρμοσμένης Στατιστικής του Πανεπιστημίου Πειραιώς

Ξενόγλωση βιβλιογραφία

Athreya K.B (1987), Bootstrap estimation of the mean in the infinite variance case, Annals of statistics, 15,724-731

Bennett P.J. (2009), Introduction to the bootstrap and robust statistics, Lecture notes,McMaster University, Canada

Chernick M. (1999), Bootstrap methods, a practitioner’s guide, John Wiley and Sons

Chernick M. (2008), Bootstrap methods, a guide for practitioners and researchers, John Wiley and Sons

Davison A.-Hinkley D. (1997), Bootstrap methods and their application, Cambridge University Press

Efron B. (1980), The jackknife, the bootstrap and other resampling plans, Technical report No.63

Efron B.- DiCiccio (1992) More accurate confidence intervals in exponential families. Biometrika 79 231-245

Efron B.- DiCiccio (1996) ,Bootstrap confidence intervals, Statistical Science,11, 189-228

Efron B.- Tibshirani R, (1993),An Introduction to the bootstrap , Chapman and Hall

Falk M.-Kaufman E. (1991), Coverage probability of bootstrap confidence intervals for quantiles, The annals of statistics, 19,485-495

Hall P. (1992a) , Efficient bootstrap simulations, Exploring the limits of bootstrap, LePage and L.Billard eds, Wiley

Rubin D. (1981), The Bayesian bootstrap, Annals of statistics, 9, 130-134

Shao J.- Tu D. (1995) ,The jackknife and bootstrap, Springer

Trosset M. (2009), An introduction to statistical inference and its applications R, Chapman and Hall

Tukey J. (1958), Bias and confidence intervals in not quite large samples, Annals of statistics, 29, 614

Ιστοσελίδες

https://epilab.ich.ucl.ac.uk/coursematerial/statistics/non_parametric/confidence_interval.html, Wade-Koutoumanou (2010), Lecture notes, UCL, τελευταία ενημέρωση 9/2010