

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΤΕΧΝΙΚΗΣ
BOOTSTRAP ΣΕ ΔΙΑΚΡΙΤΑ
ΔΕΔΟΜΕΝΑ**

Ανέστης Τουλούμης

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Μάιος 2006

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΤΕΧΝΙΚΗΣ
BOOTSTRAP ΣΕ ΔΙΑΚΡΙΤΑ
ΔΕΔΟΜΕΝΑ**

Ανέστης Τουλούμης

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Μάιος 2006

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε 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 Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**APPLICATION OF BOOTSTRAP
METHOD IN DISCRETE DATA**

By

Anestis Touloumis

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 Science in
Applied Statistics

Piraeus, Greece
May 2006

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

*Στους γονείς μου,
Κωνσταντίνο και Ευφροσύνη,
και στην αδερφή μου,
Όλγα*

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στην κα. Κατέρη Μαρία, επ. Καθηγήτρια του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, για την πολύ σημαντική καθοδήγηση, υποστήριξη, βοήθεια και υπομονή σε όλη την διάρκεια υλοποίησης της παρούσας εργασίας.

Παράλληλα θα ήθελα να ευχαριστήσω τον κ. Μπούτσικα Μιχάλη, Λέκτορα του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, και τον κ. Πολίτη Κωνσταντίνο, επ. Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, για την συμμετοχή τους στην τριμελή επιτροπή.

Τέλος θέλω να ευχαριστήσω όλους τους φίλους μου για την αμέριστη συμπαράσταση τους τελευταίους μήνες.

**ΤΟΥΛΟΥΜΗΣ ΑΝΕΣΤΗΣ
ΑΓΙΟΣ ΝΙΚΟΛΑΟΣ
ΜΑΙΟΣ 2006**

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Περίληψη

Η μέθοδος bootstrap παρουσιάστηκε από τον Efron το 1979 ως μια μη παραμετρική μέθοδο εκτίμησης της τυπικής απόκλισης εκτιμητών. Η απλότητα της ιδέας της συντέλεσε στην ραγδαία ανάπτυξή της και στην εφαρμογή της σε διάφορους τομείς της στατιστικής συμπερασματολογίας, όπως για παράδειγμα στην εκτίμηση παραμέτρων, στην κατασκευή διαστημάτων εμπιστοσύνης, στον έλεγχο υποθέσεων, κ.α. Η μέθοδος bootstrap χρησιμοποιείται ευρέως σε συνεχείς τυχαίες μεταβλητές ενώ η χρήση της είναι πιο περιορισμένη για διακριτές τυχαίες μεταβλητές.

Αντικείμενο της παρούσας εργασίας είναι η εφαρμογή της μεθόδου σε διακριτές τυχαίες μεταβλητές. Αρχικά θα παρουσιάσουμε την μέθοδο bootstrap θεωρητικά. Στην συνέχεια θα ασχοληθούμε με την εφαρμογή της σε διωνυμικές τυχαίες μεταβλητές και πιο συγκεκριμένα θα κατασκευάσουμε bootstrap διαστήματα εμπιστοσύνης ένα διωνυμικό ποσοστό. Επίσης, θα εφαρμόσουμε την μέθοδο bootstrap για την κατασκευή ταυτόχρονων διαστημάτων εμπιστοσύνης για πολυωνυμικά ποσοστά και για αντιθέσεις πολυωνυμικών ποσοστών καθώς και για την κατασκευή διαστημάτων εμπιστοσύνης τη διαφορά δύο ανεξάρτητων διωνυμικών ποσοστών.

Τέλος θα ασχοληθούμε με τους bootstrap ελέγχους υποθέσεων. Θα αναπτύξουμε τον bootstrap έλεγχο για την ισότητα δύο συσχετισμένων ποσοστών και θα παρουσιάσουμε τον δεσμευμένο έλεγχο της ανεξαρτησίας όταν ισχύει το μοντέλο συνάφειας RC.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Abstract

Efron (1979) introduced bootstrap method as a non parametrical method for the estimation of the standard deviation of estimators. The simplicity of the idea behind this method contributed to the huge development of bootstrap and to applications of the method to the statistical areas of estimation, confidence intervals and the hypothesis testing. Even though bootstrap is widely used in continuous random variables, its application in discrete data is limited.

In this thesis, we focus our interest on the application of the method in categorical random variables. After presenting the basic theoretical aspects of bootstrap, we explore the use of bootstrap method in binomial random variables in order to form confidence intervals for a binomial proportion. Moreover, we use bootstrap method to construct simultaneous confidence intervals for multinomial proportions and for contrasts of multinomial proportions as well as for the difference between two independent binomial proportions.

Finally, we refer to bootstrap hypothesis testing. We develop a bootstrap test for the hypothesis of equality of paired proportions and we present the conditional test of independence when the row-column (RC) association model fits well to the data.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Περιεχόμενα

Περίληψη	vii
Abstract	ix
1 Η Μέθοδος Bootstrap	1
1.1 Η Παραδοσιακή προσέγγιση	1
1.2 Μέθοδος Jackknife	4
1.3 Μέθοδος Bootstrap	6
1.3.1 Υπολογισμός της Διασποράς	6
1.3.2 Σχέση μεταξύ Bootstrap και Jackknife	8
1.3.3 Εκτίμηση της Δειγματικής Κατανομής	9
1.3.4 Μεροληψία και Μέσο Τετραγωνικό Σφάλμα	10
1.3.5 Συμπεράσματα	11
1.4 Bootstrap Διαστήματα Εμπιστοσύνης	13
1.4.1 Bootstrap-t	15
1.4.2 Percentile	18
1.4.3 BC (Bias-Corrected Percentile)	20
1.4.4 BC_α (Bias Accelerated Bias-Corrected Percentile)	22
1.4.5 Ακρίβεια των Διαστημάτων Εμπιστοσύνης	26
1.5 Έλεγχος Υποθέσεων	27
1.5.1 Γενική Περιγραφή	28
1.5.2 Υπολογισμός της Ισχύος	30
1.5.3 Υπολογισμός του p-value	31
1.5.4 Συμπεράσματα	33
1.6 Υπολογιστικές Μέθοδοι	33
1.6.1 Πότε Χρειάζονται οι Υπολογιστικές μέθοδοι	34
1.6.2 Μέθοδος Δέλτα	35
1.6.3 Jackknife Προσέγγιση	36
1.6.4 Η Απλή Μέθοδος Monte Carlo	37
1.6.5 Σχόλια	42
1.7 Περιπτώσεις Όπου η Μέθοδος Bootstrap Αποτυγχάνει	43
1.8 Σύντομη Ιστορική Αναδρομή	45

1.9	Εφαρμογή της Μεθόδου Bootstrap σε Κατηγορικά Δεδομένα	46
2	Μέθοδος Bootstrap για Διωνυμικές τ.μ.	47
2.0	Εισαγωγή	47
2.1	Η Διωνυμική Τυχαία Μεταβλητή	47
2.2	Εκτίμηση του Διωνυμικού Ποσοστού	49
2.3	Διαστήματα Εμπιστοσύνης για το Διωνυμικό Ποσοστό	50
2.3.1	Wald Διάστημα Εμπιστοσύνης	50
2.3.2	Clopper-Pearson (Exact) Διάστημα Εμπιστοσύνης	51
2.3.3	Wilson (Score) Διάστημα Εμπιστοσύνης	51
2.3.4	Agresti-Coull Διάστημα Εμπιστοσύνης	52
2.3.5	Άλλες Μέθοδοι Κατασκευής Διαστημάτων Εμπιστοσύνης	52
2.4	Σύγκριση Παραμετρικής και Μη Παραμετρικής Bootstrap	53
2.5	Bootstrap Διαστήματα Εμπιστοσύνης για το Διωνυμικό Ποσοστό	55
2.5.1	Percentile	55
2.5.2	BC	55
2.5.3	BC _α	56
2.5.4	Bootstrap-t	57
2.6	Κριτήρια Σύγκρισης των Διαστημάτων Εμπιστοσύνης	59
2.7	Εκτίμηση των Μεθόδων Κατασκευής Διαστημάτων Εμπιστοσύνης	63
2.8	Συμπεράσματα	84
3	Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά	87
3.0	Εισαγωγή	87
3.1	Πολυωνυμική Κατανομή	87
3.2	Κατανομή Poisson	88
3.3	Δειγματοληπτικά Σχήματα	90
3.3.1	Πολυωνυμικό Δειγματοληπτικό Σχήμα	90
3.3.2	Δειγματοληπτικό Σχήμα Γινομένου Πολυωνυμικών	91
3.3.3	Poisson Δειγματοληπτικό Σχήμα	91
3.4	Η Μέθοδος Bootstrap για την Πολυωνυμική Κατανομή	92
3.5	Ταυτόχρονες Περιοχές Εμπιστοσύνης Για Πολυωνυμικά Ποσοστά	93
3.5.1	Κλασική Προσέγγιση	94
3.5.2	Bootstrap Προσέγγιση	94
3.5.3	Παράδειγμα Ταυτόχρονων Διαστημάτων Εμπιστοσύνης	96
3.6	Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Αντιθέσεις σε Γινόμενο Πολυωνυμικών	97
3.6.1	Παράδειγμα Ταυτόχρονων Διαστημάτων Εμπιστοσύνης Για Αντιθέσεις	99
3.7	Συμπεράσματα	101

4	Διαφορά Ποσοστών	103
4.0	Εισαγωγή	103
4.1	Έλεγχοι Ισότητας Συσχετισμένων Ποσοστών	103
4.1.1	Wald Στατιστικό	104
4.1.2	McNemar και Yates Στατιστικό	106
4.1.3	Ακριβής Διωνυμικός Έλεγχος	107
4.1.4	Προσέγγιση μέσω της Bootstrap Μεθόδου	107
4.2	Κριτήρια Αξιολόγησης των Στατιστικών Ελέγχων	108
4.3	Σύγκριση Στατιστικών Ελέγχου Ισότητας Εξαρτημένων Ποσοστών	109
4.4	Συμπεράσματα για την Σύγκριση Συσχετισμένων Ποσοστών	115
4.5	Ανεξάρτητα Ποσοστά	116
4.6	Παραδοσιακοί Τρόποι Κατασκευής Διαστημάτων Εμπιστοσύνης	116
4.6.1	Μέθοδος Wald	116
4.6.2	Υδριβική Score Μέθοδος του Newcombe	117
4.6.3	Μέθοδος Agresti-Caffo	118
4.7	Μέθοδοι Κατασκευής Διαστημάτων Εμπιστοσύνης μέσω της Bootstrap	118
4.7.1	Bootstrap Percentile Διαστήματα Εμπιστοσύνης	119
4.7.2	Τροποποιημένη Bootstrap Percentile Διαστήματα Εμπιστοσύνης	119
4.7.3	Conlon-Thomas Bootstrap Διαστήματα Εμπιστοσύνης	120
4.8	Μέθοδοι Εκτίμησης των Διαστημάτων Εμπιστοσύνης	120
4.9	Εκτίμηση των Διαστημάτων Εμπιστοσύνης	121
4.10	Συμπεράσματα για τη Διαφορά Δυο Ανεξάρτητων Διωνυμικών Ποσοστών	126
5	Δεσμευμένος Έλεγχος Ανεξαρτησίας (όταν ισχύει το μοντέλο RC)	127
5.0	Εισαγωγή	127
5.1	Το Μοντέλο της Ανεξαρτησίας	128
5.2	Το RC Μοντέλο	129
5.3	Δεσμευμένος Έλεγχος Ανεξαρτησίας	131
5.4	Bootstrap Έλεγχος Ανεξαρτησίας	132
5.5	Αριθμητικό Παράδειγμα	133
	Βιβλιογραφία	137

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

1^ο Κεφάλαιο

Η Μέθοδος Bootstrap

1.1 Η Παραδοσιακή Προσέγγιση

Η αναγνώριση της αβεβαιότητας είναι μια κατηγορηματική παραδοχή της στατιστικής επιστήμης. Ιδέες και αρχές, όπως η εκ των προτέρων πληροφορία, πιθανοθεωρητικά μοντέλα, πιθανοφάνεια, πληρότητα, τυπικά σφάλματα και διαστήματα εμπιστοσύνης έχουν κατασκευαστεί με σκοπό να επισημοποιήσουν την αβεβαιότητα και να την παρουσιάζουν σε στατιστικές αναλύσεις. Ο Rao (1989) συνοψίζοντας τα παραπάνω ανέφερε ότι “ο βασικός στόχος των στατιστικών αναλύσεων είναι να εξάγουμε όσο το δυνατόν περισσότερες πληροφορίες από τα δεδομένα και να συμπεράνουμε τις ιδιότητες του πληθυσμού που παράγονται από τα δεδομένα”.

Πριν από την συλλογή των δεδομένων, μια στατιστική συνάρτηση είναι μια τυχαία ποσότητα που έχει μια κατανομή πιθανότητας, η οποία ονομάζεται δειγματική κατανομή του στατιστικού, και χρησιμοποιείται για την ανάλυση των δεδομένων. Ανάλογα με τον στόχο της ανάλυσης απαιτείται η γνώση της δειγματικής κατανομής ή κάποιου χαρακτηριστικού της. Για παράδειγμα, η κατασκευή διαστημάτων εμπιστοσύνης απαιτεί την γνώση της δειγματικής κατανομής ενώ η εκτίμηση μιας ποσότητας απαιτεί την γνώση συγκεκριμένων χαρακτηριστικών της δειγματικής κατανομής. Επειδή οποιασδήποτε εκτιμητής μπορεί να έχει σφάλμα στην εκτίμηση, μας ενδιαφέρει να έχουμε μια ένδειξη ενός μέτρου ακρίβειας της εκτίμησης. Τα μέτρα ακρίβειας, όπως η διασπορά, η μεροληψία και το μέσο τετραγωνικό σφάλμα, χρησιμοποιούνται συνήθως για να επιλέξουμε τον καλύτερο εκτιμητή από μια κλάση εκτιμητών.

Η δειγματική κατανομή ενός στατιστικού και τα χαρακτηριστικά του συνήθως εξαρτώνται από τον πληθυσμό και είναι άγνωστα. Αυτά πρέπει να εκτιμηθούν ή να προσεγγιστούν από τα παρατηρούμενα δεδομένα στα περισσότερα προβλήματα εκτίμησης ή προβλήματα συμπερασματολογίας. Για να διαλέξουμε έναν εκτιμητή από μια κλάση εκτιμητών, μερικές φορές δεν χρειάζεται να εκτιμήσουμε ή να προσεγγίσουμε τα μέτρα ακρίβειας. Στις περισσότερες περιπτώσεις η σχετική ακρίβεια των εκτιμητών εξαρτάται από τον πληθυσμό

1. Η Μέθοδος Bootstrap

και χρησιμοποιήσουμε τα δεδομένα για να εκτιμήσουμε την σχετική ακρίβεια ώστε να επιλέξουμε τον εκτιμητή. Σύμφωνα με την παραδοσιακή προσέγγιση, ένα μέτρο ακρίβειας εκτιμάται από την εμπειρική αναλογία μιας θεωρητικής εξίσωσης του μέτρου ακρίβειας ή της προσέγγισής του, το οποίο λαμβάνεται από κάποιο μοντέλο.

Ας είναι X_1, \mathbf{K}, X_n ένα δείγμα n ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από μια άγνωστη κατανομή F και $T_n = T(X_1, \mathbf{K}, X_n)$ το στατιστικό που μας ενδιαφέρει. Η διασπορά του T_n είναι

$$\text{Var}[T_n] = \int \left(T_n(\mathbf{x}) - \int T_n(\mathbf{y}) d \prod_{i=1}^n F(y_i) \right)^2 d \prod_{i=1}^n F(x_i) \quad (1.1)$$

όπου $\mathbf{x} = (x_1, \mathbf{K}, x_n)$ και $\mathbf{y} = (y_1, \mathbf{K}, y_n)$.

Όταν το T_n είναι απλό, τότε παίρνουμε μια αναλυτική έκφραση της διασποράς του $\text{Var}[T_n]$ ως συνάρτηση κάποιων άγνωστων ποσοτήτων και στην συνέχεια εκτιμάμε το $\text{Var}[T_n]$ αντικαταστήνοντας τις άγνωστες ποσότητες με τις εκτιμήσεις τους.

Ωστόσο για τα περισσότερα στατιστικά, η έκφραση (1.1) είναι αρκετά πολύπλοκη και είναι πολύ δύσκολο, αν όχι ακατόρθωτο, να βρούμε μια ακριβή και αναλυτική εξίσωση για το $\text{Var}[T_n]$. Κάτω από κάποιες συνθήκες κανονικότητας, μπορούμε συχνά να εξασφαλίσουμε ότι

$$\lim_{n \rightarrow \infty} n \text{Var}[T_n] = S_F^2$$

όπου $S_F^2 = S^2(F)$ είναι μια συνάρτηση της F ή $S_F^2 = S^2(\gamma)$ για ένα δυνάμωμα γ κάτω από τις άγνωστες παραμέτρους. Τότε εκτιμούμε το $\text{Var}[T_n]$ από μια εμπειρική αναλογία της προσέγγισης, για παράδειγμα την $S^2(\hat{F})/n$ ή την $S^2(\hat{\gamma})/n$, όπου \hat{F} και $\hat{\gamma}$ είναι οι εκτιμητές της F και του γ αντίστοιχα.

Οι Shao and Tu (1995) και οι Efron and Tibshirani (1993) υποδεικνύουν μερικές αδυναμίες και μειονεκτήματα της παραδοσιακής προσέγγισης:

- i. Το μεγάλο μέγεθος δείγματος n είναι αναγκαίο ώστε να έχουμε ακριβή διασπορά των εκτιμητών. Όταν χρησιμοποιούμε προσεγγίσεις για την διασπορά και επομένως όταν η τάξη του n δεν είναι η επιθυμητή τότε η ακρίβεια των αποτελεσμάτων είναι αμφισβητήσιμη. Κλασσική είναι η περίπτωση όπου

- χρησιμοποιούμε προσεγγίσεις της κανονικής κατανομής και το n είναι μικρότερο του 30. Τότε συνήθως οδηγούμαστε σε παραπλανητικά αποτελέσματα.
- ii. Η θεωρητική εξίσωση ή η προσέγγισή της βασίζεται σε ένα πιθανοθεωρητικό μοντέλο. Όταν το μοντέλο είναι ελαφρώς λανθασμένο τότε οι εκτιμητές της ακρίβειας μπορεί να μην ισχύουν. Αξίζει να σημειωθεί ότι λανθασμένα αποτελέσματα βασιζόμενα σε ασυμπτωτικές προσεγγίσεις μπορούν να δοθούν και όταν το μέγεθος του δείγματος είναι μεγαλύτερο του 30 αλλά οι προϋποθέσεις της χρήσης των ασυμπτωτικών αποτελεσμάτων δεν ισχύουν.
 - iii. Για να εφαρμόσουμε την παραδοσιακή προσέγγιση σε πολλά διαφορετικά προβλήματα, πρέπει να παραγωγίσουμε μια θεωρητική εξίσωση για κάθε πρόβλημα. Αυτές οι παραγωγίσεις μπορεί να είναι δύσκολες. Επιπλέον, η παραγωγή της θεωρητικής εξίσωσης απαιτεί από τους αναλυτές των δεδομένων να έχουν καλή γνώση της μαθηματικής και στατιστικής επιστήμης.
 - iv. Ίσως όμως το σημαντικότερο πρόβλημα που έχει να αντιμετωπίσει ο στατιστικός είναι να εξετάσει κατά πόσο πληρούνται οι προϋποθέσεις ενός παραμετρικού μοντέλου. Αυτό σημαίνει συνήθως ότι πρέπει να πραγματοποιηθεί ένας μεγάλος αριθμός ελέγχων ή διαγνωστικών ελέγχων προκειμένου να εξασφαλιστούν οι υποθέσεις αυτές. Ακόμα και όταν το αποτέλεσμα των διαγνωστικών τεστ είναι θετικό, υπάρχουν περιπτώσεις όπου δημιουργούνται απορίες για την εγκυρότητα τους. Ένα παράδειγμα μπορεί να προκύψει στην απλή γραμμική παλινδρόμηση, με την εξέταση της κανονικότητας των υπολοίπων. Μπορεί μέσω του στατιστικού Kolmogorov-Smirnov να είναι αποδεκτή αλλά μέσω γραφικών ελέγχων, για παράδειγμα του qq-plot να παρουσιάζονται αποκλίσεις από την διχοτόμο και επομένως εύλογες απορίες για την εγκυρότητα του αποτελέσματος.
 - v. Μερικές φορές η παραγωγή της θεωρητικής εξίσωσης ή η προσέγγιση της είναι πολύ δύσκολη ή και ακατόρθωτη. Αυτό έχει διαπιστωθεί από τότε που παρουσιάστηκαν στατιστικές μέθοδοι που βασίζονται σε εμπειρικές διαδικασίες. Οι οριακές διασπορές των στατιστικών οι οποίες είναι συναρτήσεις των εμπειρικών διαδικασιών είναι συναρτήσεις της κανονικής κατανομής. Εκφράζοντας αυτές ως απλές συναρτήσεις των παραμέτρων του μοντέλου είναι πολύ δύσκολο εγχείρημα.

1. Η Μέθοδος Bootstrap

- vi. Η θεωρητική εξίσωση μπορεί να είναι πολύπλοκη για να είναι χρήσιμη στην εκτίμηση του μέτρου ακρίβειας.

1.2 Μέθοδος Jackknife

Ο στόχος αυτής της παραγράφου είναι να παρουσιάσουμε εν συντομία την μέθοδο jackknife και να επικεντρώσουμε το ενδιαφέρον μας σε αποτελέσματα που θα μας φανούν χρήσιμα παρακάτω. Ο λόγος για τον οποίο παρουσιάζεται η μέθοδος είναι ότι σχετίζεται με την μέθοδο bootstrap, αφού για παράδειγμα, τα BC_α bootstrap διαστήματα εμπιστοσύνης περιέχουν παραμέτρους που μπορούν να υπολογιστούν και με την χρήση της μεθόδου jackknife.

Η μέθοδος jackknife πρωτοπαρουσιάστηκε από τον Quenouille (1949) ως μια τεχνική περιορισμού της μεροληψίας μιας εκτίμησης. Η βασική ιδέα της μεθόδου είναι να αφαιρούμε μια παρατήρηση από το αρχικό δείγμα και υπολογίζοντας την παράμετρο που μας ενδιαφέρει, να συγκεντρώσουμε πληροφορίες σχετικά με την σταθερότητα και την μεταβλητότητα της παραμέτρου.

Έστω X_1, \mathbf{K}, X_n ένα δείγμα n τυχαίων μεταβλητών και $T_n = T(X_1, \mathbf{K}, X_n)$ ένας εκτιμητής μιας άγνωστης παραμέτρου q . Η μεροληψία του T_n ορίζεται ως:

$$bias(T_n) = E[T_n] - q \quad (1.2)$$

Ας είναι $T_{n-1,i} = T_n(X_1, \mathbf{K}, X_{i-1}, X_{i+1}, \mathbf{K}, X_n)$ το δοθέν στατιστικό που βασίζεται στις $n-1$ παρατηρήσεις $X_1, \mathbf{K}, X_{i-1}, X_{i+1}, \mathbf{K}, X_n$, με $i = 1, \mathbf{K}, n$. Ο εκτιμητής jackknife της μεροληψίας του Quenouille είναι ο

$$b_{JACK} = (n-1)(\bar{T}_n - T_n) \quad (1.3)$$

όπου

$$\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T_{n-1,i} \quad (1.4)$$

Αυτό οδηγεί στον εκτιμητή jackknife του q , που έχει μειωμένη μεροληψία

$$T_{JACK} = T_n - b_{JACK} = nT_n - (n-1)\bar{T}_n \quad (1.5)$$

Με απλές αλγεβρικές πράξεις είναι εύκολο να δειχθεί ότι ο εκτιμητής T_{JACK} μπορεί επίσης να γραφτεί ως

$$T_{JACK} = \frac{1}{n} \sum_{i=1}^n (nT_n - (n-1)T_{n-1,i})$$

1^ο Κεφάλαιο

Η χρησιμότητα της τεχνικής jackknife έγινε πιο εμφανής όταν ο Tukey (1958) κατάφερε να την χρησιμοποιήσει για την κατασκευή εκτιμητών διασποράς. Πιο συγκεκριμένα ο Tukey όρισε ως jackknife ψευδοτιμές την ποσότητα $T_{n,i}^{\%}$ που δίνεται από την παρακάτω σχέση

$$T_{n,i}^{\%} = nT_n - (n-1)T_{n-1,i} \quad i = 1, \mathbf{K}, n \quad (1.6)$$

και είκασε ότι ισχύουν δύο υποθέσεις:

- i. Οι ψευδοτιμές $T_{n,i}^{\%}$, με $i = 1, \mathbf{K}, n$, μπορεί να χρησιμοποιηθούν σαν να ήταν ανεξάρτητες και ισόνομες τυχαίες μεταβλητές.
- ii. Η $T_{n,i}^{\%}$ έχει προσεγγιστικά την ίδια διασπορά με την $\sqrt{n}T_n$.

Κάτω από τις συνθήκες (i) και (ii) είναι λογικό να εκτιμήσουμε το $\text{Var}[\sqrt{n}T_n]$ από την δειγματική διασπορά που βασίζεται στα $T_{n,1}^{\%}, \mathbf{K}, T_{n,n}^{\%}$, δηλαδή να εκτιμήσουμε το $\text{Var}[T_n]$ από την σχέση

$$u_{JACK} = \frac{1}{n(n-1)} \sum_{i=1}^n \left(T_{n,i}^{\%} - \frac{1}{n} \sum_{j=1}^n T_{n,j}^{\%} \right)^2 \quad (1.7)$$

Ο εκτιμητής u_{JACK} είναι γνωστός ως jackknife εκτιμητής της διασποράς για το q .

Τέλος, οι Tu and Zhang (1992a) πρότειναν τον διορθωμένο (*adjusted*) jackknife εκτιμητή της λοξότητας

$$sk_{JACK} = \frac{3(n-1)^2}{n^3 (u_{JACK})^{3/2}} \sum_{i \neq j} (T_{n-1,i} - \bar{T}_n)(T_{n-1,j} - \bar{T}_n) \Delta_{ij} - \frac{(n-1)^3}{n^3 (u_{JACK})^{3/2}} \sum_{i \neq j} (T_{n-1,i} - \bar{T}_n)^3 \quad (1.8)$$

όπου $\Delta_{ij} = nT_{n-1,i} - (n-1)(T_{n-1,i} + T_{n-1,j}) + (n-2)T_{n-2,ij}$ και $T_{n-2,ij}$ είναι το στατιστικό T_n όταν από το δείγμα μας έχουμε αφαιρέσει τις τυχαίες μεταβλητές X_i και X_j .

Παράδειγμα 1.1

Έστω ότι ο T_n είναι αμερόληπτος εκτιμητής για την μέση τιμή m του δείγματος X_1, \mathbf{K}, X_n .

Είναι εύκολο να δείξουμε ότι:

$$T_{n-1,i} = \bar{X}_{n-1,i} = \frac{n\bar{X}_n - X_i}{n-1} \quad (1.9)$$

και χρησιμοποιώντας τις σχέσεις (1.4), (1.3) και (1.5) ότι $\bar{T}_n = \bar{X}_n$, $b_{JACK} = 0$ και $T_{JACK} = T_n = \bar{X}_n$, αντίστοιχα. Επίσης από την σχέση (1.6) αντικαθιστώντας το $T_{n-1,i}$ με το δεξιό μέλος της σχέσης (1.9) έχουμε ότι

1. Η Μέθοδος Bootstrap

$$T_{n,i}^* = X_i \text{ και } Var[T_{n,i}^*] = Var[\sqrt{n}\bar{X}_n]$$

Επομένως οι υποθέσεις του Tukey είναι αληθείς σε αυτήν την περίπτωση και το u_{JACK} γίνεται ίσο με $[n(n-1)]^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Συμπεραίνουμε ότι όταν θέλουμε να εκτιμήσουμε την μέση τιμή ενός δείγματος ο jackknife και ο παραδοσιακός εκτιμητής μας δίνουν τους ίδιους εκτιμητές διασποράς.

1.3 Μέθοδος Bootstrap

Από ένα δείγμα μεγέθους n μπορούμε να σχηματίσουμε $2^n - 1$ διαφορετικά, μη κενά υποσύνολα. Η μέθοδος jackknife που περιγράψαμε παραπάνω χρησιμοποιεί μόνο τα n από αυτά. Ωστόσο ο Hartigan (1969) παρουσίασε μια βελτιωμένη παραλλαγή της jackknife χρησιμοποιώντας στατιστικά που βασίζονται σε περισσότερα από n ή και σε όλα τα $2^n - 1$ υποσύνολα. Το μειονέκτημα αυτής της παραλλαγής είναι ότι χρειάζεται περισσότερη υπολογιστική δύναμη από αυτήν που απαιτεί η μέθοδος jackknife.

Η ανάπτυξη των ηλεκτρονικών υπολογιστών βοήθησε την ανάπτυξη στατιστικών μεθόδων που είναι μεν υπολογιστικά επίπονες, όμως είναι πιο αξιόπιστες και έχουν περισσότερες εφαρμογές. Η μέθοδος bootstrap, που παρουσιάστηκε από τον Efron (1979) είναι μία από αυτές τις μεθόδους. Επειδή ο Efron χρησιμοποίησε την μέθοδο για τον υπολογισμό της διασποράς ενός στατιστικού, η παρουσίαση της μεθόδου θα γίνει βάση τον τρόπο που υπολογίζεται η διασπορά μέσω της μεθόδου bootstrap.

1.3.1 Υπολογισμός της Διασποράς

Ας υποθέσουμε ότι οι X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την F και ότι η F εκτιμάται από την \hat{F} . Αντικαθιστώντας με \hat{F} την F στην σχέση (1.1), παίρνουμε ως εκτιμητή μέθοδος της διασποράς τον:

$$\begin{aligned} u_{BOOT} &= \int \left(T_n(x) - \int T_n(y) d \prod_{i=1}^n \hat{F}(y_i) \right)^2 d \prod_{i=1}^n \hat{F}(x_i) \\ &= Var_* [T_n(X_1^*, \mathbf{K}, X_n^*) | X_1, \mathbf{K}, X_n] \end{aligned} \quad (1.10)$$

όπου X_1^*, \mathbf{K}, X_n^* είναι ένα δείγμα από n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την \hat{F} που ονομάζεται bootstrap δείγμα ενώ $Var_*[\cdot | X_1, \mathbf{K}, X_n]$ δηλώνει την δεσμευμένη διασπορά δοθέντων των X_1, \mathbf{K}, X_n . Η σχέση (1.10) είναι η θεωρητική μορφή του εκτιμητή

1^ο Κεφάλαιο

bootstrap της διασποράς του T_n . Αν η u_{BOOT} είναι μια αναλυτική συνάρτηση των X_1, \mathbf{K}, X_n τότε αυτή είναι μια εναλλακτική εκτιμήτρια της $Var[T_n]$.

Παράδειγμα 1.2

Έστω ότι $T_n = \bar{X}_n$ και $\hat{F} = F_n$, όπου F_n η εμπειρική κατανομή ορίζεται ως:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x\} \quad (1.11)$$

όπου $I\{A\}$ είναι η δείκτρια συνάρτηση του συνόλου A . Τότε έχουμε:

$$Var[\bar{X}_n] = \frac{1}{n} Var[X_1] = \int \left(x - \int y dF(y)\right)^2 dF(x)$$

και από την σχέση (1.10) βρίσκουμε ότι:

$$u_{BOOT} = \frac{1}{n^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (1.12)$$

Παρατήρηση 1.1

Επειδή η εμπειρική κατανομή F_n παίζει σημαντικό ρόλο στην μέθοδο bootstrap, θέλουμε στο σημείο αυτό να διευκρινίσουμε και να σχολιάσουμε τον ορισμό της με πιο απλά λόγια. Όταν το δείγμα X_1, \mathbf{K}, X_n αποτελείται από n διακριτές τιμές τότε η F_n ουσιαστικά δίνει βάρος $1/n$ σε κάθε παρατηρηθέν X_i . Στην περίπτωση που έχουμε επαναλαμβανόμενες παρατηρήσεις ή κατηγορικά δεδομένα, το βάρος που δίνει η F_n σε κάθε διακριτή τιμή του δείγματος είναι ίσο με το ποσοστό της συχνότητας της στο δείγμα.

Σε αντίθεση με το παράδειγμα 1.2, η σχέση (1.10) είναι συνήθως περίπλοκη και η u_{BOOT} δεν είναι μια αναλυτική συνάρτηση των X_1, \mathbf{K}, X_n ακόμα και αν η F είναι γνωστή. Σε αυτήν την περίπτωση δεν μπορούμε να εκτιμήσουμε την $Var[T_n]$ ακριβώς. Στην στατιστική, υπάρχει μια τεχνική που ονομάζεται Monte Carlo και χρησιμοποιείται για να προσεγγιστεί η $Var[T_n]$ αριθμητικά όταν η F είναι γνωστή. Η μέθοδος αυτή θα παρουσιαστεί αργότερα, ωστόσο στο σημείο αυτό θα παρουσιάσουμε την βασική της ιδέα. Συλλέγουμε διαρκώς προκαθορισμένο αριθμό καινούριων δειγμάτων από την F και χρησιμοποιώντας την δειγματική διασπορά των τιμών του T_n , που υπολογίστηκαν από τα καινούρια δείγματα, υπολογίζουμε την αριθμητική προσέγγιση της $Var[T_n]$. Η ιδέα αυτή μπορεί να εφαρμοστεί για να προσεγγιστεί η u_{BOOT} όταν η \hat{F} είναι μια γνωστή κατανομή. Παίρνοντας

1. Η Μέθοδος Bootstrap

$\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$, $b=1, \dots, B$, ανεξάρτητα δείγματα από την \hat{F} , δοθέντων των X_1, \mathbf{K}, X_n υπολογίζουμε το $T_n^*(b) = T_n(X_1^*(b), \mathbf{K}, X_n^*(b))$ και προσεγγίζουμε την u_{BOOT} με βάση την μέθοδο Monte Carlo από τον τύπο:

$$u_{BOOT}^{(B)} = \frac{1}{B} \sum_{b=1}^B (T_n^*(b) - \frac{1}{B} \sum_{k=1}^B T_n^*(k))^2$$

Από τον νόμο των μεγάλων αριθμών ισχύει ότι $u_{BOOT} = \lim_{B \rightarrow \infty} u_{BOOT}^{(B)}$ ασυμπτωτικά. Σε αυτήν την ιδιότητα στηρίζεται και δικαιολογείται η χρήση της μεθόδου Monte Carlo. Όπως η u_{BOOT} έτσι και η Monte Carlo προσέγγιση $u_{BOOT}^{(B)}$ καλούνται εκτιμητές bootstrap της διασποράς του T_n . Η $u_{BOOT}^{(B)}$ χρησιμοποιείται κυρίως σε πρακτικές εφαρμογές ενώ σε θεωρητικές μελέτες συνήθως επικεντρωνόμαστε στην u_{BOOT} .

Συνοψίζοντας τα παραπάνω μπορούμε να πούμε ότι η μέθοδος bootstrap είναι ένας συνδυασμός δυο τεχνικών: της αρχής της αντικατάστασης και της αριθμητικής προσέγγισης. Όταν η u_{BOOT} στην σχέση (1.10) είναι μια αναλυτική συνάρτηση των X_1, \mathbf{K}, X_n , η μέθοδος bootstrap συμπίπτει με την παραδοσιακή μέθοδο αντικατάστασης ενώ διαφορετικά η μέθοδος bootstrap προσεγγίζει την $Var[T_n]$ αριθμητικά. Αυτή είναι και η διαφορά της με την παραδοσιακή μέθοδο, που για την προσέγγιση της $Var[T_n]$, εκτιμά πρώτα την $Var[T_n]$ και στην συνέχεια εκτιμά τις άγνωστες ποσότητες στην εξίσωση (1.1).

Η κατανομή \hat{F} που χρησιμοποιείται για την παράγωγη των bootstrap δειγμάτων μπορεί να είναι οποιοσδήποτε εκτιμητήρια της F , παραμετρική ή μη, που βασίζεται στα X_1, \mathbf{K}, X_n . Μια απλή μη παραμετρική εκτιμητήρια της F είναι η εμπειρική κατανομή F_n που ορίζεται από την σχέση (1.11).

1.3.2 Σχέση μεταξύ Bootstrap και Jackknife

Από τον απλό τρόπο που περιγράψαμε τις μεθόδους bootstrap και jackknife υπάρχει μια ένδειξη σύνδεσης μεταξύ τους. Δοθέντος ενός δείγματος μεγέθους n , ο εκτιμητής jackknife του T_n υπολογίζεται με βάση όλα τα δυνατά υποσύνολα μεγέθους $n-1$ του αρχικού δείγματος. Ο υπολογισμός του εκτιμητή bootstrap του T_n βασίζεται στην παραγωγή B bootstrap τυχαίων δειγμάτων $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$, $b=1, \dots, B$, από το αρχικό δείγμα.

Επομένως, η εκτίμηση bootstrap του T_n μπορεί να θεωρηθεί ως μια γενίκευση της εκτίμησης του που προκύπτει με βάση την μέθοδο jackknife. Αυτός είναι και ο λόγος για τον οποίο οι μέθοδοι jackknife και οι bootstrap ονομάζονται μέθοδοι επαναδειγματοληψίας, ενώ οι εκτιμητές του T_n λέγονται εκτιμητές επαναδειγματοληψίας (*resampling estimators*).

Οι Efron and Tibshirani (1993, κεφαλαίο 11) εξετάζοντας την προσέγγιση της jackknife στην bootstrap διαπιστώνουν ότι για γραμμικά στατιστικά η προσέγγιση είναι γραμμική, και μάλιστα διαφέρουν κατά έναν παράγοντα $((n-1)/n)^{1/2}$ που χρησιμοποιεί ο εκτιμητής jackknife, ενώ για την μεροληψία ενός στατιστικού η προσέγγιση της jackknife στην bootstrap είναι δευτεροβάθμια.

1.3.3 Εκτίμηση της Δειγματικής Κατανομής

Η δειγματική κατανομή μιας τυχάιας μεταβλητής $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F)$, δίνεται από τον τύπο:

$$H_F(x) = \text{Prob}\{\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) \leq x\} \quad (1.13)$$

όπου οι X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες τυχάιες μεταβλητές από την F . Η H_F εξαρτάται από το n ωστόσο αυτό παραλείπεται για λόγους συντομίας. Αν θέλουμε να εκτιμήσουμε την δειγματική κατανομή του T_n , θέτουμε απλά $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = T_n$. Δυο συνήθης στατιστικά που χρησιμοποιούμε για την κατασκευή ενός συνόλου εμπιστοσύνης για μια άγνωστη παράμετρο που σχετίζεται με την F , είναι το $T_n - q$ και το $(T_n - q)/S_n$, η τυποποιημένη ποσότητα του $T_n - q$, όπου S_n είναι ένας εκτιμητής της τυπικής απόκλισης της T_n . Σε αυτές τις περιπτώσεις ψάχνουμε να βρούμε την δειγματική κατανομή του $T_n - q$ ή του $(T_n - q)/s_n$ και θέτουμε είτε $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = T_n - q$ ή $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = (T_n - q)/S_n$, αντίστοιχα.

Σύμφωνα με την κλασσική στατιστική, ψάχνουμε μια θεωρητική εξίσωση για την $H_F(x)$, ακριβή ή προσεγγιστική, και αντικαθιστούμε τις άγνωστες ποσότητες με τους εκτιμητές τους. Για παράδειγμα, όταν $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = T_n - q$, η $H_F(x)$ συνήθως προσεγγίζεται από την $\Phi(x\sqrt{n}/\hat{S})$ όπου Φ είναι η τυπική κανονική κατανομή και S είναι μια άγνωστη παράμετρος που εξαρτάται από την F . Αν \hat{S} είναι ένας εκτιμητής του S , τότε η $H_F(x)$

1. Η Μέθοδος Bootstrap

εκτιμάται από την $\Phi(x\sqrt{n}/\hat{S})$. Όταν $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = (T_n - q)/S_n$ η $H_F(x)$ συχνά προσεγγίζεται από την $\Phi(x)$.

Η μέθοδος bootstrap όπως είδαμε λειτουργεί διαφορετικά. Η εκτίμηση bootstrap της $H_F(x)$ προκύπτει αν αντικαταστήσουμε την F με έναν εκτιμητή της \hat{F} . Τότε ο εκτιμητής bootstrap της $H_F(x)$ δίνεται από την σχέση:

$$H_{BOOT} = H_{\hat{F}}(x) = \text{Prob}_* \left\{ \mathfrak{R}_n(X_1^*, \mathbf{K}, X_n^*, \hat{F}) \leq x \mid X_1, \mathbf{K}, X_n \right\}$$

όπου X_1^*, \mathbf{K}, X_n^* είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την \hat{F} και το $\text{Prob}_* \{ \cdot \mid X_1, \mathbf{K}, X_n \}$ δηλώνει την δεσμευμένη πιθανότητα δοθέντων των X_1, \mathbf{K}, X_n . Όπως και στην Παράγραφο 1.3.1 αν η $H_{BOOT}(x)$ είναι μια αναλυτική εξίσωση των X_1, \mathbf{K}, X_n τότε η $H_{BOOT}(x)$ είναι η εκτιμήτρια bootstrap της $H_F(x)$, ενώ σε αντίθετη περίπτωση χρησιμοποιούμε την μέθοδο Monte Carlo για να προσεγγίσουμε την $H_{BOOT}(x)$:

$$H_{BOOT}^{(B)}(x) = \frac{1}{B} \sum_{b=1}^B I \left\{ \mathfrak{R}_n(X_1^*(b), \mathbf{K}, X_n^*(b), \hat{F}) \leq x \right\}$$

όπου $X_1^*(b), \mathbf{K}, X_n^*(b)$, $b = 1, \dots, B$ είναι ανεξάρτητα bootstrap δείγματα από την \hat{F} .

1.3.4 Μεροληψία και Μέσο Τετραγωνικό Σφάλμα

Ένα μέτρο ακρίβειας ενός δοθέντος στατιστικού T_n συχνά είναι ένα χαρακτηριστικό της δειγματικής κατανομής του T_n . Αν ο εκτιμητής της δειγματικής κατανομής του T_n είναι διαθέσιμος τότε είναι φυσικό να εκτιμήσουμε το μέτρο ακρίβειας από το αντίστοιχο χαρακτηριστικό της εκτιμώμενης δειγματικής κατανομής του T_n . Οι εκτιμητές bootstrap του μέτρου ακρίβειας μπορούν να υπολογιστούν χρησιμοποιώντας το χαρακτηριστικό της $H_{BOOT}(x)$ ή της $H_{BOOT}^{(B)}(x)$. Για παράδειγμα, ο εκτιμητής της διασποράς u_{BOOT} ή της $u_{BOOT}^{(B)}$ είναι στην πραγματικότητα η διασπορά της κατανομής $H_{BOOT}(x)$ ή της $H_{BOOT}^{(B)}(x)$ για $\mathfrak{R}_n = T_n$.

Ας θεωρήσουμε την μεροληψία του T_n , που δίνεται από την σχέση:

$$\text{bias}(T_n) = \int x dH_F(x) - q \quad (1.14)$$

1^ο Κεφάλαιο

όπου $H_F(x)$ δίνεται από την σχέση (1.13) για $\mathfrak{X}_n = T_n$. Αντικαθιστώντας την άγνωστη F και το q από τους εκτιμητές τους \hat{F} και T_n στην σχέση (1.14) παίρνουμε τον εκτιμητή bootstrap της μεροληψίας:

$$b_{BOOT} = \int x dH_{\hat{F}}(x) - T_n \quad (1.15)$$

Όταν το ολοκλήρωμα στην σχέση (1.15) δεν έχει αναλυτική μορφή τότε χρησιμοποιούμε την μέθοδο Monte Carlo για την προσέγγιση της b_{BOOT} :

$$\begin{aligned} b_{BOOT}^{(B)} &= \int x dH_{BOOT}^{(B)}(x) - T_n \\ &= \frac{1}{B} \sum_{b=1}^B T_n(X_1^*(b), \mathbf{K}, X_n^*(b)) - T_n \end{aligned}$$

Παράδειγμα 1.3

Έστω ότι $T_n = \bar{X}_n$ και $\hat{F} = F_n$. Τότε

$$\int x dH_{BOOT}(x) = E_*[\bar{X}_n^* | X_1, \mathbf{K}, X_n] = \bar{X}_n$$

και επομένως $b_{BOOT} = 0$. Επομένως ο εκτιμητής bootstrap του μέσου ενός δείγματος είναι αμερόληπτος.

Ένα άλλο μέτρο ακρίβειας είναι το μέσο τετραγωνικό σφάλμα. Το μέσο τετραγωνικό σφάλμα του T_n ως εκτιμητής του q είναι ο

$$E[(T_n - q)^2] = Var[T_n] + (bias(T_n))^2$$

Ο εκτιμητής bootstrap του $E[(T_n - q)^2]$ είναι ο $u_{BOOT} + (b_{BOOT})^2$ και η Monte Carlo προσέγγιση του είναι

$$u_{BOOT}^{(B)} + (b_{BOOT}^{(B)})^2 = \frac{1}{B} \sum_{b=1}^B (T_n(X_1^*(b), \mathbf{K}, X_n^*(b)) - T_n)^2$$

1.3.5 Συμπεράσματα

Έχοντας εξετάσει τα βασικά σημεία της μεθόδου bootstrap μπορούμε να δούμε την φιλοσοφία της. Ας υποθέσουμε ότι έχουμε δεδομένα $\mathbf{Y} = (Y_1, \mathbf{K}, Y_n)$, όχι απαραίτητα ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, και P ένα στατιστικό μοντέλο κάτω από το οποίο παίρνουμε τα δεδομένα. Συνήθως το P περιγράφεται από την κοινή κατανομή του \mathbf{Y} ή από κάποιες ποσότητες που μπορούν να ορίσουν μοναδικά αυτήν την κοινή κατανομή. Έστω μια τυχαία μεταβλητή $\mathfrak{X}_n(\mathbf{Y}, P)$ και ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε το

1. Η Μέθοδος Bootstrap

μοντέλο P από τα δεδομένα \mathbf{Y} . Ας είναι \mathbf{Y}^* ένα bootstrap δείγμα που παράγεται από το εκτιμώμενο μοντέλο \hat{P} . Η δεσμευμένη κατανομή του $\mathfrak{R}_n(\mathbf{Y}^*, \hat{P})$ δοθέντος του \mathbf{Y} είναι ο bootstrap εκτιμητής της $\mathfrak{R}_n(\mathbf{Y}, P)$. Χρησιμοποιώντας ένα διάγραμμα, οι Efron and Tibshirani (1986) παρουσίασαν την μέθοδο ως ακολούθως:

$$\begin{array}{ccccccc} P & \rightarrow & \mathbf{Y} & \hat{P} & \rightarrow & \mathbf{Y}^* & \\ \square & & \square & \Rightarrow & \square & & \square \\ & & \mathfrak{R}_n(\mathbf{Y}, P) & & \mathfrak{R}_n(\mathbf{Y}^*, \hat{P}) & & \end{array}$$

Η ιδέα της μεθόδου bootstrap είναι να χρησιμοποιήσουμε την δειγματική συμπεριφορά της τριπλέτας $(\hat{P}, \mathbf{Y}^*, \mathfrak{R}_n(\mathbf{Y}^*, \hat{P}))$ με σκοπό να μιμηθεί την τριπλέτα $(P, \mathbf{Y}, \mathfrak{R}_n(\mathbf{Y}, P))$, όπου η σχέση ανάμεσα στα \mathbf{Y}^* , \hat{P} και $\mathfrak{R}_n(\mathbf{Y}^*, \hat{P})$ είναι ίδια με την σχέση μεταξύ των \mathbf{Y} , P και $\mathfrak{R}_n(\mathbf{Y}, P)$. Αν $\hat{P} = P$ ακριβώς, τότε η κατανομή του $\mathfrak{R}_n(\mathbf{Y}^*, \hat{P})$ είναι ακριβώς η ίδια με αυτήν της $\mathfrak{R}_n(\mathbf{Y}, P)$. Ακόμα και αν $\hat{P} \neq P$ οι κατανομές των $\mathfrak{R}_n(\mathbf{Y}^*, \hat{P})$ και $\mathfrak{R}_n(\mathbf{Y}, P)$ μερικές φορές μπορεί να είναι οι ίδιες.

Αν και η μέθοδος bootstrap βασίζεται στις αρχές της αντικατάστασης και της μίμησης της δειγματικής συμπεριφοράς, το σύνολο των εφαρμογών της πραγματοποιείται με επαναδειγματοληψία από το δείγμα. Όταν η δεσμευμένη κατανομή του $\mathfrak{R}_n(\mathbf{Y}^*, \hat{P})$ δεν είναι μια αναλυτική έκφραση του \mathbf{Y} , τότε απαιτείται η χρήση της μεθόδου Monte Carlo ή και κάποιας άλλης υπολογιστικής τεχνικής για να υπολογιστούν οι ζητούμενοι εκτιμητές bootstrap. Γι' αυτόν τον λόγο η μέθοδος bootstrap ταξινομείται ως μέθοδος επαναδειγματοληψίας, αν και η επαναδειγματοληψία δεν είναι απόλυτα απαραίτητη για να εφαρμόσουμε την μέθοδο bootstrap. Για παράδειγμα, οι Fisher and Hall (1991) έδειξαν ότι οι ζητούμενοι εκτιμητές bootstrap μπορούν να υπολογιστούν με την κατασκευή όλων των πιθανών bootstrap δειγμάτων όταν το μέγεθος του δείγματος n είναι πολύ μικρό. Περισσότερες λεπτομέρειες για τις υπολογιστικές μεθόδους της μεθόδου bootstrap παρουσιάζονται στην Παράγραφο 1.5.

Η μέθοδος bootstrap εφαρμόζεται σε όλες τις περιπτώσεις όπου το μοντέλο P μπορεί να οριστεί και να εκτιμηθεί από το \hat{P} . Ωστόσο, είναι ιδιαίτερα κρίσιμος ο τρόπος με τον οποίο το μοντέλο P εκτιμάται.

Ας είναι X_1, \mathbf{K}, X_n από μια κατανομή F . Η από κοινού κατανομή των X_1, \mathbf{K}, X_n καθορίζεται από την F . Επομένως, $P = F$. Αν η F ανήκει στην παραμετρική οικογένεια, τότε $P = F_{\theta}$, όπου θ είναι ένα διάνυσμα από άγνωστες παραμέτρους. Στην παραμετρική περίπτωση, το θ πρώτα εκτιμάται από το $\hat{\theta}$ και το P μετά εκτιμάται από το $P = F_{\hat{\theta}}$. Το bootstrap δείγμα X_1^*, \mathbf{K}, X_n^* παράγεται από την $F_{\hat{\theta}}$. Αυτή η μέθοδος bootstrap ονομάζεται παραμετρική μέθοδο bootstrap. Στην μη παραμετρική περίπτωση, το P εκτιμάται από την F_n , την εμπειρική κατανομή. Το bootstrap δείγμα X_1^*, \mathbf{K}, X_n^* τότε παράγεται από την F_n . Αυτή η μέθοδος bootstrap συχνά καλείται μη παραμετρική μέθοδος bootstrap. Αξίζει να σημειώσουμε ότι η μη παραμετρική μέθοδος bootstrap μπορεί να χρησιμοποιηθεί τόσο σε παραμετρικά μοντέλα όσο και σε μη παραμετρικά μοντέλα. Ο Efron (1982) θεωρεί την μη παραμετρική μέθοδο bootstrap ως μια μη παραμετρική μέθοδο μέγιστης πιθανοφάνειας. Επομένως, σύμφωνα με τον Efron η μέθοδος bootstrap μπορεί να θεωρηθεί ως μια γενίκευση της μεθόδου μέγιστης πιθανοφάνειας σε μη παραμετρικό πλαίσιο. Υπό το πρίσμα αυτό, F_n είναι ο παραμετρικός εκτιμητής της F . Αν δεν κάνουμε άλλες υποθέσεις, οι μη παραμετρικοί εκτιμητές bootstrap μπορούν να θεωρηθούν και ως εκτιμητές "μέγιστης πιθανοφάνειας".

Όταν είναι γνωστή η ορθότητα ενός παραμετρικού μοντέλου, η παραμετρική μέθοδος bootstrap είναι ανώτερη από την μη παραμετρική. Γενικά η αποτελεσματικότητα της μεθόδου bootstrap βασίζεται στο πόσο καλά μπορούμε να αναγνωρίσουμε και να εκτιμήσουμε το μοντέλο. Ακόμα και στην μη παραμετρική περίπτωση, μπορούμε να αντικαταστήσουμε την F_n από έναν ομαλό εκτιμητή της F όταν ξέρουμε ότι η F είναι ομαλή.

1.4 Bootstrap Διαστήματα Εμπιστοσύνης

Έστω X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες μεταβλητές από μια άγνωστη κατανομή F και q η παράμετρος που μας ενδιαφέρει. Αν $T_n(X_1, \mathbf{K}, X_n)$ είναι το στατιστικό που συνδέει με την παράμετρο q , $C_n = C_n(X_1, \mathbf{K}, X_n)$ ένα υποσύνολο του \mathbf{R} που εξαρτάται μόνο από τις X_1, \mathbf{K}, X_n και

$$\text{Prob}\{q \in C_n\} \geq 1 - a \quad (1.16)$$

όπου a είναι μια σταθερά που ικανοποιεί την σχέση $0 < a < 1$, τότε το C_n ορίζεται να είναι ένα όριο εμπιστοσύνης για το q σε επίπεδο σημαντικότητας $1 - a$. Η πιθανότητα στην αριστερή πλευρά της παραπάνω ανισότητας λέγεται πιθανότητα κάλυψης (*Coverage Probability*) του

1. Η Μέθοδος Bootstrap

C_n . Αν ισχύει η ισότητα τότε το C λέγεται όριο εμπιστοσύνης με συντελεστή εμπιστοσύνης $1-a$ ή πιο σύντομα ένα $1-a$ όριο εμπιστοσύνης. Ας είναι $\underline{q} = T_{n,1}(X_1, \mathbf{K}, X_n)$ και $\bar{q} = T_{n,2}(X_1, \mathbf{K}, X_n)$ δυο στατιστικά. Τότε τα διαστήματα $(-\infty, \bar{q}]$ και $[\underline{q}, \infty)$ λέγονται μονόπλευρα διαστήματα εμπιστοσύνης και το $[\underline{q}, \bar{q}]$ λέγεται δίπλευρο διάστημα εμπιστοσύνης. Το \underline{q} (\bar{q}) λέγεται επίσης κάτω (άνω) όριο εμπιστοσύνης. Αν \underline{q} και \bar{q} είναι το κάτω και άνω όριο εμπιστοσύνης για το q αντίστοιχα, τότε το $[\underline{q}, \bar{q}]$ είναι ένα $1-2a$ δίπλευρο διάστημα εμπιστοσύνης για το q ίσων ουρών.

Το επιθυμητό επίπεδο ενός διαστήματος εμπιστοσύνης λέγεται ονομαστικό επίπεδο (*nominal level*), το οποίο συνήθως δίνεται. Θα χρησιμοποιούμε $1-a$ και $1-2a$ για να δηλώσουμε το ονομαστικό επίπεδο των μονόπλευρων και δίπλευρων διαστημάτων εμπιστοσύνης, αντίστοιχα. Ένα διάστημα εμπιστοσύνης λέγεται ακριβές αν ο συντελεστής εμπιστοσύνης του είναι ακριβώς ίσος με το ονομαστικό επίπεδο.

Στις περισσότερες περιπτώσεις, τα διαστήματα εμπιστοσύνης κατασκευάζονται θεωρώντας μια pivotal ποσότητα $Z_n = Z_n(X_1, \mathbf{K}, X_n, F)$ της οποίας η κατανομή G_n είναι γνωστή, ανεξάρτητη από την F . Αν μπορούμε να συνάγουμε την σχέση $\underline{q} \leq q \leq \bar{q}$ από την ανισότητα $L \leq Z_n \leq U$ τότε το $[\underline{q}, \bar{q}]$ είναι ένα διάστημα εμπιστοσύνης με ένα συγκεκριμένο επίπεδο σημαντικότητας. Για παράδειγμα, όταν το q είναι μια παράμετρος θέσης τότε το Z_n είναι συνήθως της μορφής $(\hat{q} - q)/\hat{S}$, όπου \hat{q} είναι ο εκτιμητής του q και \hat{S}^2 είναι η εκτίμηση της διασποράς του \hat{q} και το $1-2a$ ακριβές διάστημα εμπιστοσύνης για το q είναι το:

$$\left[\hat{q} - \hat{S} G_n^{-1}(1-a), \hat{q} - \hat{S} G_n^{-1}(a) \right] \quad (1.16)$$

Το να βρούμε μια τέτοια pivotal ποσότητα είναι συνήθως δύσκολο. Αυτό γιατί δεν είναι εύκολο να βρούμε μια Z_n με γνωστή κατανομή G_n . Αν η κατανομή G_n είναι άγνωστη τότε το διάστημα εμπιστοσύνης που δίνεται από την (1.16) δεν μπορεί να χρησιμοποιηθεί και θα πρέπει να προσεγγίσουμε την G_n . Σύμφωνα με την ασυμπτωτική θεωρία θα πρέπει να αντικαταστήσουμε την G_n με το όριο της G . Αν η G_n έχει ένα γνωστό όριο, ανεξάρτητη της F τότε αντικαθιστούμε την G_n με το όριο της στην (1.16). Ωστόσο αν η G_n έχει ως όριο την

1^ο Κεφάλαιο

G_q που εξαρτάται από κάποια άγνωστη q , τότε αντικαθιστούμε στην (1.16) την G_n με $G_{\hat{q}}$, όπου \hat{q} είναι ένας συνεπής εκτιμητής του q . Η τεχνική bootstrap εφαρμόζεται για να πάρουμε ένα διάστημα εμπιστοσύνης απλά αντικαταστήνοντας την G_n με την εκτίμηση bootstrap της G_{BOOT} .

Λόγω της προσέγγισης, η πιθανότητα κάλυψης πολλών ορίων εμπιστοσύνης μπορεί να είναι μεγαλύτερη ή μικρότερη από το ονομαστικό επίπεδο. Για παράδειγμα, το διάστημα εμπιστοσύνης (1.16) είναι ακριβές όταν η G_n είναι γνωστή, μα μπορεί να μην έχει επίπεδο σημαντικότητας $1-2\alpha$ για σταθερό n όταν η G_n είναι άγνωστη και εκτιμάται. Τα κύρια προβλήματα των παραδοσιακών ασυμπτωτικών μεθόδων είναι ότι:

- i. προαπαιτούν να υπάρχει αναλυτική έκφραση του ορίου της G_n
- ii. και ότι δεν επιτυγχάνουν μεγάλη ακρίβεια.

Από την άλλη πλευρά μέσω της μεθόδου bootstrap μπορούν να κατασκευαστούν ευκολότερα διαστήματα εμπιστοσύνης και μεγαλύτερης ακρίβειας.

Στην ενότητα αυτή θα θεωρήσουμε ότι έχουμε δείγμα $\mathbf{X}=(X_1, \mathbf{K}, X_n)$ n ανεξάρτητων τυχαίων μεταβλητών από μια κατανομή F , η οποία εκτιμάται από την \hat{F} είτε παραμετρικά είτε μη παραμετρικά. Η υπό μελέτη παράμετρος q εκτιμάται από το $\hat{q}=T(X_1, \mathbf{K}, X_n)$ και \hat{S} είναι η εκτίμηση του τυπικού σφάλματος του \hat{q} . Όταν θα αναφέρεται η παράμετρος b , θα θεωρούμε ότι οι εκτιμητές bootstrap προσεγγίζονται μέσω της Monte Carlo προσέγγισης εκτελώντας B bootstrap επαναλήψεις και συλλέγοντας B bootstrap δείγματα $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}, b=1, 2, \dots, B$. Σε κάθε b bootstrap επανάληψη, με $b=1, 2, \dots, B$, ορίζουμε με $\hat{q}^*(b)=T(X_1^*(b), \mathbf{K}, X_n^*(b))$ να είναι η τιμή του \hat{q} για το bootstrap δείγμα $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$ και $\hat{S}^*(b)$ να είναι η εκτιμωμένη τυπική απόκλιση του q για το bootstrap δείγμα $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$.

1.4.1 Bootstrap-t

Η μέθοδος bootstrap-t βασίζεται σε ένα τυποποιημένο pivot στατιστικό $Z=(\hat{q}-q)/\hat{S}$, όπου \hat{q} είναι ένας εκτιμητής του q και \hat{S}^2 είναι ο εκτιμητής της διασποράς του q . Αν η κατανομή

1. Η Μέθοδος Bootstrap

G του Z είναι άγνωστη τότε αυτή εκτιμάται από τον εκτιμητή bootstrap G_{BOOT} που ορίζεται από την σχέση:

$$G_{BOOT}(x) = \text{Prob}_* \{Z^* \leq x\}$$

όπου $Z^* = (\hat{q}^* - \hat{q}) / \hat{S}^*$ και \hat{q}^* και \hat{S}^* είναι οι αναλογίες bootstrap των \hat{q} και \hat{S} αντίστοιχα.

Το κάτω όριο εμπιστοσύνης για το q είναι:

$$q_{BT} = \hat{q} - \hat{S} \times G_{BOOT}^{-1}(1-a)$$

το οποίο θα λέγεται και bootstrap-t κάτω όριο εμπιστοσύνης.

Όμοια ορίζουμε και το μεθόδου bootstrap-t άνω όριο εμπιστοσύνης για το q από την σχέση:

$$\bar{q}_{BT} = \hat{q} - \hat{S} \times G_{BOOT}^{-1}(a)$$

Επομένως ένα $(1-2a)\%$ διάστημα εμπιστοσύνης που προκύπτει με την μέθοδο bootstrap-t είναι το :

$$[q_{BT}, \bar{q}_{BT}] = [\hat{q} - \hat{S} \times G_{BOOT}^{-1}(1-a), \hat{q} - \hat{S} \times G_{BOOT}^{-1}(a)]$$

Ουσιαστικά η όλη ιδέα της μεθόδου bootstrap-t για την κατασκευή διαστημάτων εμπιστοσύνης βασίζεται σε μια τροποποίηση των διαστημάτων εμπιστοσύνης που προκύπτουν από την τυπική κανονική κατανομή $N(0,1)$. Η ποσότητα $Z = (\hat{q} - q) / \hat{S}$ προσεγγίζεται μέσω της μεθόδου bootstrap από την ποσότητα $Z^*(b) = (\hat{q}^*(b) - \hat{q}) / \hat{S}^*(b)$. Δηλαδή σε κάθε bootstrap δείγμα υπολογίζουμε το $\hat{q}^*(b)$ τον εκτιμητή bootstrap του q , το $\hat{S}^*(b)$ τον εκτιμητή bootstrap της τυπικής απόκλισης S και συνεπώς το $Z^*(b)$ την εκτίμηση bootstrap του Z . Στη συνέχεια τις B bootstrap τιμές του Z τις διατάσσουμε σε ένα δυνάμωμα \mathbf{Z}^* και εκτιμούμε το ζητούμενο ποσοστημόριο.

Ας συμβολίσουμε με z_i^* το i -οστό στοιχείο του \mathbf{Z}^* , με $i = 1, 2, \mathbf{K}, B$. Όταν η ποσότητα $B \cdot a$ είναι ακέραιος τότε τα $G_{BOOT}^{-1}(a)$ και $G_{BOOT}^{-1}(1-a)$ είναι τα $z_{B \cdot a}^*$ και $z_{B \cdot (1-a)}^*$ αντίστοιχα.

Πρέπει να σημειώσουμε στο σημείο αυτό ότι αν για να εφαρμοστεί ακριβώς η μέθοδος αυτή θα πρέπει το $a \leq 0.5$ και ο αριθμός $B \cdot a$ να είναι ακέραιος. Αν για κάποιο λόγο η τελευταία συνθήκη δεν είναι εφικτή, τότε μπορούμε να ακολουθήσουμε την μέθοδο που προτείνουν οι

Davison and Hinkley (1997). Πρόκειται για μια απλή μέθοδο που λειτουργεί ικανοποιητικά για τους εκτιμητές στατιστικών που ακολουθούν την κανονική κατανομή και δεν είναι άλλη από αυτή της γραμμικής παρεμβολής στην κλίμακα των ποσοστιαίων σημείων της κανονικής κατανομής. Αν είναι k το ακέραιο μέρος του $B \cdot a$ τότε ορίζουμε ως $G_{BOOT}^{-1}(a)$ την ποσότητα $G_{BOOT(NEW)}^{-1}(a)$ που δίνεται από την σχέση:

$$G_{BOOT(NEW)}^{-1}(a) = z_{B \cdot a}^* + \frac{\Phi^{-1}(a) - \Phi^{-1}\left(\frac{k}{B}\right)}{\Phi^{-1}\left(\frac{k+1}{B}\right) - \Phi^{-1}\left(\frac{k}{B}\right)} (z_{k+1}^* - z_k^*), k = [B \cdot a]$$

Προφανώς το μειονέκτημα της μεθόδου αυτής είναι ότι δεν μπορεί να εφαρμοστεί όταν το k είναι ίσο με $0, B, B-1$.

Οι Efron and Tibshirani (1993) προτείνουν μια διαφορετική μέθοδο όταν το $B \cdot a$ δεν είναι ακέραιος αριθμός. Υποθέτοντας ότι ισχύει $a \leq 0.5$, θέτουν k το ακέραιο μέρος του $B \cdot a$. Επομένως τα $G_{BOOT}^{-1}(a)$ και $G_{BOOT}^{-1}(1-a)$ προσδιορίζονται από τις τιμές z_k^* και z_{B-k}^* αντίστοιχα.

Το τυποποιημένο κανονικό διάστημα εμπιστοσύνης εφαρμόζεται σε όλα τα δείγματα ανεξαρτήτως του μεγέθους του δείγματος, το t διάστημα εμπιστοσύνης εφαρμόζεται σε όλα τα δείγματα σταθερού μεγέθους ενώ το bootstrap- t διάστημα εμπιστοσύνης εφαρμόζεται μόνο δοθέντος του δείγματος. Επίσης τα εκατοστημόρια της κανονικής και της Student κατανομής είναι συμμετρικά γύρω από το μηδέν και αυτό έχει ως αποτέλεσμα τα διαστήματα εμπιστοσύνης είναι συμμετρικά γύρω από το \hat{q} . Σε αντίθεση τα bootstrap- t εκατοστημόρια μπορεί να είναι ασύμμετρα γύρω από το μηδέν, οδηγώντας σε διαστήματα που δεν είναι απαραίτητα ίσων ουρών.

Αν και η μέθοδος bootstrap- t είναι απλή και ευκολονόητη έχει δυο σημαντικά μειονεκτήματα:

- Για να χρησιμοποιήσουμε την μέθοδο bootstrap- t χρειαζόμαστε τον εκτιμητή της διασποράς \hat{S}^2 . Στην περίπτωση που δεν μπορούμε να υπολογίσουμε τον εκτιμητή της διασποράς, τότε μπορούμε να χρησιμοποιήσουμε τον εκτιμητή bootstrap της διασποράς για το \hat{q} . Ωστόσο αν και οι δυο πρέπει να προσεγγιστούν από Monte Carlo τότε πρέπει να χρησιμοποιήσουμε εμφωλευμένη (nested) bootstrap. Για παράδειγμα, θα δημιουργήσουμε B_1 bootstrap δείγματα για να προσεγγίσουμε την G_{BOOT} και για κάθε ένα από τα B_1 bootstrap δείγμα θα δημιουργήσουμε B_2 bootstrap δείγματα για

1. Η Μέθοδος Bootstrap

να προσεγγίσουμε την διασπορά. Επομένως ο συνολικός αριθμός των bootstrap δειγμάτων για να υπολογιστεί το \mathbf{q}_{BT} είναι $B_1 B_2$.

- Η μέθοδος bootstrap-t δεν είναι αμετάβλητη κάτω από επαναπαραμετροποίηση (*reparameterization invariant*). Ωστόσο αυτό δεν είναι σφάλμα της μεθόδου bootstrap αφού οποιοδήποτε διάστημα εμπιστοσύνης βασίζεται πάνω σε ένα τυποποιημένο ρινοί στατιστικό δεν είναι αμετάβλητο.

1.4.2 Percentile

Ας είναι $K_{BOOT}(x)$ η αθροιστική συνάρτηση του \hat{q}^* . Επομένως η $K_{BOOT}(x)$ θα δίνεται από τον τύπο:

$$K_{BOOT}(x) = \text{Prob}_* \{ \hat{q}^* \leq x \} \quad (1.17)$$

Η bootstrap percentile μέθοδος δίνει ως κάτω όριο εμπιστοσύνης για το q :

$$\mathbf{q}_{BP} = K_{BOOT}^{-1}(a) \quad (1.18)$$

Όμοια ορίζουμε και το bootstrap percentile άνω όριο εμπιστοσύνης για το q από την σχέση (1.18):

$$\bar{\mathbf{q}}_{BP} = K_{BOOT}^{-1}(1-a)$$

Το όνομα percentile προέρχεται από το γεγονός ότι το $K_{BOOT}^{-1}(a)$ είναι ένα ποσοστημόριο της κατανομής K_{BOOT} .

Επομένως ένα $(1-2a)\%$ διάστημα εμπιστοσύνης που προκύπτει με την μέθοδο bootstrap percentile είναι το :

$$[\mathbf{q}_{BP}, \bar{\mathbf{q}}_{BP}] = [K_{BOOT}^{-1}(a), K_{BOOT}^{-1}(1-a)] \quad (1.19)$$

Δηλαδή σε κάθε bootstrap δείγμα υπολογίζουμε το $\hat{q}^*(b)$ τον εκτιμητή bootstrap του q . Στην συνέχεια τις B bootstrap τιμές του q τις διατάσσουμε σε ένα δυνάμωμα Θ^* και εκτιμούμε το ζητούμενο ποσοστημόριο. Ας συμβολίσουμε με q_i^* το i -οστό στοιχείο του Θ^* , με $i=1, 2, \dots, B$. Όταν η ποσότητα Ba είναι ακέραιος τότε τα $K_{BOOT}^{-1}(a)$ και $K_{BOOT}^{-1}(1-a)$ είναι τα q_{Ba}^* και $q_{B(1-a)}^*$ αντίστοιχα.

Ένα μεγάλο πλεονέκτημα της bootstrap percentile είναι ότι διατηρεί το εύρος των τιμών, όπως παρατηρούν και οι Efron and Tibshirani (1993). Μερικές φορές υπάρχουν περιορισμοί

στις τιμές που μπορεί να πάρει μια παράμετρος. Ένα διάστημα εμπιστοσύνης που έχει την ιδιότητα να διατηρεί το εύρος των τιμών του στατιστικού λέγεται range-preserving. Η bootstrap percentile μέθοδος έχει αυτήν την ιδιότητα γιατί πρώτον ο εκτιμητής του q , \hat{q} υπακούει στους ίδιους κανόνες που υπακούει και το q και τα $\hat{q}^*(b)$ υπακούουν στους ίδιους περιορισμούς με το q . Επίσης αυτή η μέθοδος είναι αμετάβλητη κάτω από επαναπαραμετροποίηση.

Οι Shao and Tu (1995) εξετάσανε τις υποθέσεις που θα πρέπει να ισχύουν έτσι ώστε να έχουμε καλή επίδοση των ορίων εμπιστοσύνης μέσω της μεθόδου bootstrap percentile. Στη συνέχεια θα παρουσιάσουμε συνοπτικά τα αποτελέσματα τους για το q_{BP} .

Υποθέτουμε ότι υπάρχει ένας αύξοντας μετασχηματισμός $f(x)$ τέτοιος ώστε να ισχύει:

$$\text{Prob}\{f(\hat{q}) - f(q) \leq x\} = \Psi(x) \quad (1.20)$$

για κάθε πιθανή F , περιλαμβάνοντας και την περίπτωση $F = \hat{F}$. Στην παραπάνω σχέση η $\Psi(\cdot)$ είναι μια συνεχής, αύξουσα και συμμετρική ($\Psi(x) = 1 - \Psi(-x)$) κατανομή. Αν $\Psi(x) = \Phi(x)$, την τυπική κανονική κατανομή, τότε η $f(\cdot)$ είναι ο μετασχηματισμός της κανονικοποίησης. Στην συνέχεια αποδείξανε ότι αν οι $f(\cdot)$ και $\Psi(\cdot)$ είναι γνωστές τότε η μέθοδος bootstrap percentile μας δίνει ως κάτω όριο το ακριβές κάτω όριο του διαστήματος εμπιστοσύνης για το q :

$$q_{BP} = f^{-1}\left(f(\hat{q}) + z_a\right) = q_{EXACT}$$

όπου $z_a = \Psi^{-1}(a)$.

Στην ουσία αυτό που υποδεικνύει η παραπάνω σχέση είναι ότι αν ισχύει ακριβώς η (1.20) τότε το q_{BP} ταυτίζεται με το q_{EXACT} ανεξάρτητα από το μέγεθος δείγματος n . Αν η (1.20) ισχύει προσεγγιστικά για μεγάλο n , τότε το κάτω όριο του εμπιστοσύνης της μεθόδου bootstrap percentile είναι ασυμπτωτικά σωστό και η απόδοση του εξαρτάται από το πόσο καλή είναι η προσέγγιση. Για παράδειγμα, όταν η $f(q)$ είναι μη γραμμική και η μεροληψία του $f(\hat{q}) - f(q)$ δεν μηδενίζεται γρήγορα όταν το $n \rightarrow \infty$ τότε η προσέγγιση είναι ικανοποιητική μόνο όταν το n είναι αρκετά μεγάλο.

Οι Efron and Tibshirani (1993) υποστηρίζουν ότι αν η bootstrap κατανομή του q είναι σχεδόν κανονική, τότε τα όρια εμπιστοσύνης που προκύπτουν από την τυπική κανονική

1. Η Μέθοδος Bootstrap

κατανομή και την bootstrap percentile θα είναι παρόμοια. Το κεντρικό οριακό θεώρημα μας λέει ότι όσο το $n \rightarrow \infty$, τόσο το ιστόγραμμα της bootstrap κατανομής θα παίρνει την μορφή της κανονικής κατανομής. Ωστόσο για δείγματα μικρού μεγέθους το ιστόγραμμα της bootstrap κατανομής δεν θα μοιάζει με ιστόγραμμα κανονικής κατανομής.

Επομένως αν και τα διαστήματα εμπιστοσύνης που προκύπτουν μέσω της bootstrap percentile μεθόδου είναι απλά, μπορεί να μην είναι ακριβή αν το n δεν είναι αρκετά μεγάλο. Κάποιες βελτιώσεις μπορούν να γίνουν αν χαλαρώσουμε τις υποθέσεις που κάναμε στην (1.17). Αυτό οδήγησε στην ανάπτυξη των ακόλουθων δυο μεθόδων που επίσης υιοθετούν την χρήση κάποιων ποσοστιαίων σημείων της κατανομής bootstrap ως διαστήματα εμπιστοσύνης. Και οι δυο αυτές μέθοδοι παράγουν αμετάβλητα διαστήματα εμπιστοσύνης.

1.4.3 BC (Bias-Corrected Percentile)

Από το σχόλιο που προηγήθηκε για την μεροληψία είναι λογικό να προσαρμόσουμε και έναν όρο μεροληψίας στην υπόθεση (1.19). Ο Efron (1982) πρότεινε μια πιο γενική υπόθεση:

$$\text{Prob}\{f(\hat{q}) - f(q) + z_0 \leq x\} = \Psi(x) \quad (1.21)$$

με τις $f(\cdot)$ και $\Psi(\cdot)$ να είναι οι ίδιες συναρτήσεις που ορίσαμε στην προηγούμενη παράγραφο και όπου z_0 είναι μια σταθερά που μπορεί να εξαρτάται από την F και το n . Όταν $z_0 = 0$ τότε η υπόθεση (1.21) ταυτίζεται με την (1.19). Αν οι $f(\cdot)$, $\Psi(\cdot)$ και z_0 είναι γνωστές τότε μπορούμε να πάρουμε το ακριβές κάτω όριο εμπιστοσύνης για το q

$$q_{EXACT} = f^{-1}\left(f(\hat{q}) + z_0 + z_a\right) \quad (1.22)$$

Απόδειξη της σχέσης (1.22)

Εφαρμόζοντας την υπόθεση (1.21) για $F = \hat{F}$, παίρνουμε

$$K_{BOOT}(\hat{q}) = \text{Prob}_* \left\{ f(\hat{q}^*) - f(\hat{q}) + z_0 \leq z_0 \right\} = \Psi(z_0)$$

όπου η K_{BOOT} δίνεται από την σχέση (1.17).

Με βάση τα άκρα τις ισότητας βρίσκουμε ότι:

$$z_0 = \Psi^{-1}\left(K_{BOOT}(\hat{q})\right) \quad (1.23)$$

Επίσης από την υπόθεση (1.21) έχουμε ότι:

$$1 - a = \Psi(-z_a)$$

$$\begin{aligned}
 &= \Psi \left(-f(\mathbf{q}_{EXACT}) + f(\hat{\mathbf{q}}) + z_0 \right) \\
 &= \text{Prob} \left\{ f(\hat{\mathbf{q}}^*) - f(\hat{\mathbf{q}}) + z_0 \leq -f(\mathbf{q}_{EXACT}) + f(\hat{\mathbf{q}}) + z_0 \right\} \\
 &= \text{Prob} \left\{ f(\hat{\mathbf{q}}^*) - f(\hat{\mathbf{q}}) \leq -f(\mathbf{q}_{EXACT}) + f(\hat{\mathbf{q}}) \right\} \\
 &= \text{Prob} \left\{ \hat{\mathbf{q}}^* \leq f^{-1} \left(f(\hat{\mathbf{q}}) + z_0 + z_a \right) \right\}
 \end{aligned}$$

από την οποία συνάγεται

$$f^{-1} \left(f(\hat{\mathbf{q}}) + z_0 + z_a \right) = K_{BOOT}^{-1} (1-a) \quad (1.24)$$

Επειδή η (1.24) ισχύει για κάθε a , συνάγεται ότι για ότι για κάθε $0 < x < 1$ ισχύει:

$$K_{BOOT}^{-1} (x) = f^{-1} \left(f(\hat{\mathbf{q}}) + (\Psi^{-1} (x) - z_0) \right) \quad (1.25)$$

Από τον ορισμό του \mathbf{q}_{EXACT} και από την (1.25) καταλήγουμε ότι :

$$\mathbf{q}_{EXACT} = f^{-1} \left(f(\hat{\mathbf{q}}) + z_0 + z_a \right)$$

Τώρα, υποθέτοντας ότι η $\Psi(\cdot)$ είναι γνωστή και χρησιμοποιώντας την μπορούμε να πάρουμε το bootstrap bias-corrected percentile κάτω όριο εμπιστοσύνης για το \mathbf{q} :

$$\mathbf{q}_{BC} = K_{BOOT}^{-1} \left(\Psi(z_a + 2z_0) \right) \quad (1.26)$$

Όμοια ορίζουμε και το bootstrap bias-corrected percentile άνω όριο εμπιστοσύνης για το \mathbf{q} από την σχέση:

$$\bar{\mathbf{q}}_{BC} = K_{BOOT}^{-1} \left(\Psi(z_{1-a} + 2z_0) \right)$$

Επομένως ένα $(1-2a)\%$ διάστημα εμπιστοσύνης που προκύπτει με την μέθοδο bootstrap bias-corrected percentile είναι το :

$$\left[\mathbf{q}_{BC}, \bar{\mathbf{q}}_{BC} \right] = \left[K_{BOOT}^{-1} \left(\Psi(z_a + 2z_0) \right), K_{BOOT}^{-1} \left(\Psi(z_{1-a} + 2z_0) \right) \right]$$

Συνήθως χρησιμοποιούμε $\Psi = \Phi$. Επειδή $\Psi^{-1}(0.5) = 0$ το \mathbf{q}_{BC} συμπίπτει με το \mathbf{q}_{BP} αν $K_{BOOT}^{-1}(\hat{\mathbf{q}}) = 0.5$. Επομένως το \mathbf{q}_{BC} είναι μια παραλλαγή της bootstrap percentile μεθόδου που παίρνει υπόψη της την μεροληψία. Ξανά, το \mathbf{q}_{BC} είναι ακριβές για κάθε n αν ισχύει ακριβώς (1.21) και είναι ασυμπτωτικά σωστό αν η (1.21) ισχύει προσεγγιστικά.

Λαμβάνοντας υπόψη την μεροληψία, η bootstrap BC όντως βελτιώνει την απλή bootstrap percentile μέθοδο. Ωστόσο υπάρχουν πολλές περιπτώσεις όπου η υπόθεση δεν μπορεί να

1. Η Μέθοδος Bootstrap

ικανοποιηθεί και η bootstrap BC δεν δουλεύει πολύ καλά όπως το παράδειγμα του Schenker (1985).

1.4.4 BC_a (Bias Accelerated Bias-Corrected Percentile)

Ο Efron (1987) πρότεινε μια πιο περιεκτική υπόθεση η οποία λαμβάνει υπόψη την λοξότητα της κατανομής. Συγκεκριμένα, όρισε

$$\text{Prob} \left\{ \frac{f(\hat{q}) - f(q)}{1 + af(\hat{q})} + z_0 \leq x \right\} = \Psi(x) \quad (1.27)$$

όπου $f(\cdot)$, $\Psi(\cdot)$ και z_0 είναι οι ίδιες ποσότητες με αυτές τις (1.21) και a είναι μια επιπλέον παράμετρος που εξαρτάται από την F και το n . Σύμφωνα με τον Efron (1987), το a μετράει πόσο γρήγορα αλλάζει η τυπική απόκλιση του $f(\hat{q})$ σε σχέση με την $f(q)$ και για τον λόγο αυτό την ονόμασε σταθερά επιτάχυνσης (*acceleration constant*). Προφανώς η υπόθεση αυτή είναι πιο γενική από τις υποθέσεις (1.19) και (1.21).

Αν οι $f(\cdot)$, $\Psi(\cdot)$, z_0 και a είναι γνωστές, τότε ένα ακριβές κάτω όριο εμπιστοσύνης για το q είναι το:

$$q_{EXACT} = f^{-1} \left(f(\hat{q}) + \frac{(z_0 + z_a)(1 + f(\hat{q}))}{1 - a(z_0 + z_a)} \right)$$

Το q_{EXACT} ισούται με ένα ποσοστημόριο της $K_{BOOT}(\hat{q})$. Πρώτον η υπόθεση (1.23) ισχύει. Επίσης χρησιμοποιώντας την υπόθεση (1.27) και την ίδια απόδειξη όπως στην (1.25), παίρνουμε ότι για κάθε $0 < x < 1$ ισχύει:

$$K_{BOOT}^{-1}(x) = f^{-1} \left(f(\hat{q}) + (\Psi^{-1}(x) - z_0)(1 + af(\hat{q})) \right)$$

Θέτοντας

$$x = \Psi \left(z_0 + \frac{z_0 + z_a}{1 - a(z_0 + z_a)} \right)$$

παίρνουμε ότι:

$$q_{BC_a}(a) = K_{BOOT}^{-1} \left(\Psi \left(z_0 + \frac{z_0 + z_a}{1 - \hat{a}(z_0 + z_a)} \right) \right)$$

1^ο Κεφάλαιο

Αυτό σημαίνει ότι αν τα $\Psi(\cdot)$ και a είναι γνωστά, τότε το $\mathbf{q}_{BC_a}(a)$ είναι ένα ακριβές κάτω όριο εμπιστοσύνης για το q για όλα τα n . Αν μπορούμε να εκτιμήσουμε το a με το \hat{a} και το αντικαταστήσουμε στην (1.26) τότε το βρίσκουμε ως κάτω όριο εμπιστοσύνης το οποίο το είναι ίσο με:

$$\mathbf{q}_{BC_a}(a) = \mathbf{q}_{BC_a}(\hat{a})$$

και ονομάζεται bootstrap accelerated bias-corrected κάτω όριο εμπιστοσύνης που προτάθηκε από τον Efron(1987).

Όμοια ορίζουμε και το bootstrap accelerated bias-corrected άνω όριο εμπιστοσύνης για το q από την σχέση:

$$\bar{\mathbf{q}}_{BC_a}(\hat{a}) = K_{BOOT}^{-1} \left(\Psi \left(z_0 + \frac{z_0 + z_{1-a}}{1 - \hat{a}(z_0 + z_{1-a})} \right) \right)$$

Επομένως ένα $(1-2a)\%$ διάστημα εμπιστοσύνης που προκύπτει με την μέθοδο bootstrap bias-corrected percentile είναι το :

$$\left[\mathbf{q}_{BC_a}(\hat{a}), \bar{\mathbf{q}}_{BC_a}(\hat{a}) \right] = \left[K_{BOOT}^{-1} \left(\Psi \left(z_0 + \frac{z_0 + z_a}{1 - \hat{a}(z_0 + z_a)} \right) \right), K_{BOOT}^{-1} \left(\Psi \left(z_0 + \frac{z_0 + z_{1-a}}{1 - \hat{a}(z_0 + z_{1-a})} \right) \right) \right]$$

Η παράμετρος a ωστόσο δεν είναι εύκολο να προσδιοριστεί ή να εκτιμηθεί και η εκτίμηση της εξαρτάται από το αν η F εκτιμάται παραμετρικά ή εκτιμάται μη παραμετρικά από την εμπειρική κατανομή του δείγματος.

Ας υποθέσουμε ότι έχουμε ένα πολυδιάστατο μοντέλο όπου $F = F_h$,

$\boldsymbol{\eta} = (h_1 \ h_2 \ \mathbf{L} \ h_p) \in \mathbf{R}^p$ και $q = g(\boldsymbol{\eta})$ (DiCiccio and Romano, 1988). Για την περίπτωση

όπου $\hat{q} = g(\hat{\boldsymbol{\eta}})$ και $\hat{\boldsymbol{\eta}}$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας του $\boldsymbol{\eta}$. Αν $l(\boldsymbol{\eta})$ ο

λογάριθμος της συνάρτησης πιθανοφάνειας, $l_i = \partial l(\boldsymbol{\eta}) / \partial h_i$, $m_{ij} = E[l_i l_j]$, $m_{ijk} = E[l_i l_j l_k]$,

m^{ij} το (i, j) στοιχείο του αντίστροφου πίνακα του οποίου το (i, j) στοιχείο είναι το m_{ij} ,

$g_i = \partial g(\boldsymbol{\eta}) / \partial h_i$ και $t_i = \sum_{j=1}^p m^{ij} g_j$, τότε

$$a = a(\boldsymbol{\eta}) = \left(\frac{1}{6} \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p m_{ijk} t_i t_j t_k \right) / \left(\sum_{i=1}^p \sum_{j=1}^p m_{ij} t_i t_j \right)$$

1. Η Μέθοδος Bootstrap

Στο ίδιο αποτέλεσμα καταλήγει και ο Efron (1987) περιορίζοντας την F σε μια λιγότερο ευνοϊκή οικογένεια κατανομών (Stein,1956). Σε αυτήν την περίπτωση, θα πάρουμε ως $\hat{a} = a(\hat{\eta})$.

Ακόμα και στην πιο απλή περίπτωση ο υπολογισμός του a μπορεί να είναι ιδιαίτερα κουραστικός. Μερικά προβλήματα για τον υπολογισμό του υποδείχθηκαν από τους Loh and Wu (1987).

Ας θεωρήσουμε τώρα την μη παραμετρική περίπτωση όπου $q = T(F)$ για κάποια συνάρτηση T , $\hat{q} = T(F_n)$ και F_n είναι η εμπειρική κατανομή (Konishi,1991). Χρησιμοποιώντας Edgeworth expansions βρίσκεται ότι το a δίνεται από την σχέση:

$$a = a(F) = \frac{1}{6\sqrt{n}S^3(F)} \int (f_F(x))^3 dF(x)$$

η οποία είναι παρόμοια με αυτήν που βρίσκει και ο Efron (1987).

Για να υπολογίσουμε το \hat{a} απλώς αντικαθιστούμε το F με F_n στην παραπάνω σχέση. Αυτό απαιτεί την παραγωγισιμότητα της f_F .

Για να αποφευχθεί η παραγωγισιμότητα της f_F , ο Efron (1987) πρότεινε την χρήση της σχέσης:

$$\hat{a} = \left(\frac{1}{6} \sum_{i=1}^n U_i^3 \right) / \left(\sum_{i=1}^n U_i^2 \right)^{3/2} \quad (1.28)$$

όπου

$$U_i = \left[T((1-e)F_n + e d_{x_i}) - T(F_n) \right]$$

για ένα πολύ μικρό e .

Ένας άλλος τρόπος για να εκτιμήσουμε τα a και z_0 είναι μέσω της μεθόδου jackknife. Παρατηρώντας ότι τα a και z_0 είναι συναρτήσεις της μεροληψίας, της διασποράς και της λοξότητας του \hat{q} , οι Tu and Zhang (1992b) πρότειναν την παρακάτω Jackknife προσέγγιση για το a και το z_0 :

$$\hat{a}_{JACK} = \frac{(n-1)}{6nu_{JACK}^{3/2}} \sum_{j=1}^n \left(\hat{q}_{n-1,i} - \frac{1}{n} \sum_{j=1}^n \hat{q}_{n-1,j} \right)^3$$
$$\hat{z}_{JACK} = b_{JACK} / (nu_{JACK}) - sk_{JACK} / 6$$

1^ο Κεφάλαιο

όπου $\hat{q}_{n-1,i} = T(F_{n-1,i})$ και b_{JACK} , u_{JACK} , sk_{JACK} αυτές που έχουν οριστεί από τις σχέσεις (1.3), (1.7) και (1.8) για $q = T(F)$. Σε παρόμοιο αποτέλεσμα έχουν καταλήξει και οι Frangos and Schucany(1990) ενώ οι Efron and Tibshirani (1993) προτείνουν για εκτίμηση του a αυτόν της σχέσης (1.28) και για το z_0 αυτόν της σχέσης (1.23)

Οι Davison and Hinkley (1997) τονίζουν ένα πρόβλημα που μπορεί να προκύψει μέσω τη bootstrap BC_a μεθόδου όταν $\Psi = \Phi$ και η ποσότητα

$$\Phi\left(z_0 + \frac{z_0 + z_{1-a}}{1 - \hat{a}(z_0 + z_{1-a})}\right)$$

είναι πιο κοντά στο μηδέν ή στην μονάδα από ότι το a . Τότε η ποσότητα

$$B \times \Phi\left(z_0 + \frac{z_0 + z_{1-a}}{1 - \hat{a}(z_0 + z_{1-a})}\right)$$

μπορεί να είναι μικρότερη από την μονάδα ή ίση με το B και επομένως ούτε ακόμα και με την βοήθεια της γραμμικής παρεμβολής να μην μπορεί να υπολογιστεί το ποσοστημόριο. Αν αυτό συμβεί και το B δεν μπορεί να αυξηθεί τότε προτείνουν να αναφερθεί η ακραία τιμή του Θ^* και η τιμή του a που συντελεί στην τιμή αυτή. Για παράδειγμα, αν

$$B \times \Phi\left(z_0 + \frac{z_0 + z_{1-a}}{1 - \hat{a}(z_0 + z_{1-a})}\right) = B$$

τότε το άνω όριο θα δίνεται από την σχέση

$$\bar{q}_{BC_a}(\hat{a}) = K_{BOOT}^{-1}\left(\Phi\left(z_0 + \frac{z_0 + z_{1-a_2}}{1 - \hat{a}(z_0 + z_{1-a_2})}\right)\right)$$

όπου

$$a_2 = 1 - \frac{B-1}{B} = \frac{1}{B}$$

Η bootstrap BC_a είναι αμετάβλητη κάτω από επαναπαραμετροποίηση. Περιμένουμε ότι οι κατασκευή ορίων εμπιστοσύνης σύμφωνα με την μέθοδο bootstrap BC_a να είναι πιο ακριβής αφού η κατασκευή των a και z_0 βασίζεται σε υποθέσεις ακρίβειας μεγαλύτερης τάξης.

Ένα μειονέκτημα της bootstrap BC_a είναι ότι ο καθορισμός του a δεν είναι εύκολος, ειδικότερα αν το πρόβλημα που προσπαθούμε να λύσουμε είναι περίπλοκο. Επίσης το αν η μέθοδος bootstrap BC_a παράγει καλύτερα όρια εμπιστοσύνης σε σχέση με την απλή bootstrap percentile μέθοδο εξαρτάται από την ομαλότητα του στατιστικού \hat{q} . Οι Hall and Martin

1. Η Μέθοδος Bootstrap

(1989) έδειξαν ότι τα διαστήματα εμπιστοσύνης για μη ομαλούς εκτιμητές δεν μπορούν να βελτιωθούν.

1.4.5 Ακρίβεια των Διαστημάτων Εμπιστοσύνης

Όταν εκτιμάμε μια παράμετρο είναι ιδιαίτερα σημαντικό να ξέρουμε και την ακρίβεια της εκτίμησης που κάναμε. Το ίδιο μπορούμε να κάνουμε και όταν κατασκευάζουμε διαστήματα εμπιστοσύνης. Αυτό επιτυγχάνεται με την σύγκριση του ρυθμού σύγκλισης της πιθανότητας κάλυψης του διαστήματος εμπιστοσύνης και του ονομαστικού επίπεδου.

Ορισμός 1.1

Ένα σύνολο εμπιστοσύνης C_n του q λέγεται ότι είναι k -τάξης ακριβής αν

$$\text{Prob}\{q \in C_n\} - 1 - a = O(n^{-k/2}).$$

Ο πρώτος που μελέτησε την τάξη ακρίβειας των bootstrap t συνόλων εμπιστοσύνης ήταν ο Hall(1986). Αργότερα επέκτεινε την μελέτη του και στα υπόλοιπα είδη των bootstrap συνόλων εμπιστοσύνης (1988). Θεωρώντας ότι $\hat{q}_n = g(\bar{X}_n)$ όταν η $g(\cdot)$ είναι μια ομαλή συνάρτηση του \bar{X}_n κατέληξε στα παρακάτω αποτελέσματα:

- i. Τα μονόπλευρα bootstrap- t διάστημα εμπιστοσύνης είναι δεύτερης τάξης ακριβή. Το δίπλευρο bootstrap- t διάστημα εμπιστοσύνης ίσων ουρών είναι και αυτό δεύτερης τάξης ακριβές.
- ii. Τα μονόπλευρα bootstrap percentile διάστημα εμπιστοσύνης είναι πρώτης τάξης ακριβή ενώ το δίπλευρο bootstrap percentile διάστημα εμπιστοσύνης ίσων ουρών είναι δεύτερης τάξης ακριβές.
- iii. Τα BC διαστήματα εμπιστοσύνης έχουν την ίδια απόδοση με τα αυτά της bootstrap percentile.
- iv. Τα BC_a διαστήματα εμπιστοσύνης έχουν την ίδια απόδοση με τα αυτά της bootstrap- t .

Στην περίπτωση όπου η $g(\cdot)$ δεν είναι μια ομαλή συνάρτηση του \bar{X}_n , οι Falk and Kaufman (1991) απέδειξαν ότι τα μονόπλευρα και τα δίπλευρα bootstrap percentile διαστήματα εμπιστοσύνης είναι πρώτης τάξης ακριβή ενώ οι Hall and Martin (1991) απέδειξαν ότι και τα BC_a διαστήματα εμπιστοσύνης είναι πρώτης τάξης ακριβή.

- i. Τα μονόπλευρα bootstrap-t και BC_a διαστήματα εμπιστοσύνης είναι δεύτερης τάξης ακριβή ενώ τα αντίστοιχα bootstrap percentile, BC και αυτά που βασίζονται σε προσεγγίσεις της κανονικής κατανομής είναι πρώτης τάξης ακριβή.
- ii. Τα δίπλευρα διαστήματα εμπιστοσύνης που προκύπτουν από τα τέσσερα διαστήματα εμπιστοσύνης καθώς και αυτά που προκύπτουν από προσεγγίσεις της κανονικής κατανομής είναι δεύτερης τάξης ακριβή.

Από εμπειρικές μελέτες οι Shao and Tu (1995) καταλήξανε στα εξής συμπεράσματα:

- i. Τα εμπειρικά αποτελέσματα συμφωνούν με την ασυμπτωτική θεωρία των διαστημάτων εμπιστοσύνης.
- ii. Τα bootstrap-t και BC_a διαστήματα εμπιστοσύνης αποδίδουν καλύτερα από τα αντίστοιχα που βασίζονται σε προσεγγίσεις της κανονικής κατανομής.
- iii. Η απόδοση των BC_a διαστημάτων εμπιστοσύνης εξαρτάται κατά μεγάλο ποσοστό από την εκτίμηση της παραμέτρου επιτάχυνσης a . Στα περισσότερα παραμετρικά προβλήματα το a μπορεί να οριστεί ακριβώς και επομένως τα BC_a διαστήματα εμπιστοσύνης είναι πολύ ακριβή. Στα μη παραμετρικά προβλήματα, όπου ένας ακριβής εκτιμητής του a δεν είναι διαθέσιμος, τα BC_a διαστήματα εμπιστοσύνης δεν είναι ακριβή.
- iv. Τα bootstrap-t διαστήματα εμπιστοσύνης τείνουν να είναι μεγάλα και να έχουν μακριές ουρές.

Οι Efron and Tibshirani (1993) αναφέρουν ότι όταν μπορούν να βρεθούν τα ακριβή όρια των bootstrap διαστημάτων εμπιστοσύνης τότε τα διαστήματα εμπιστοσύνης που βασίζονται σε προσέγγιση της κανονικής κατανομής ή σε Student-t προσέγγιση είναι πρώτης τάξης ακριβή ενώ μερικές από τα bootstrap διαστήματα εμπιστοσύνης είναι δεύτερης τάξης ακριβή.

1.5 Έλεγχος Υποθέσεων

Είναι γνωστό ότι οι έλεγχοι υποθέσεων σχετίζονται με τα σύνολα εμπιστοσύνης. Αν και ο έλεγχος μιας στατιστικής υπόθεσης μπορεί να γίνει με την κατασκευή του κατάλληλου συνόλου εμπιστοσύνης, οι bootstrap έλεγχοι υποθέσεων είναι πολύ σημαντικό ζήτημα για τους παρακάτω λόγους:

- i. Ο έλεγχος μιας στατιστικής υπόθεσης μέσω ελεγχοσυνάρτησης συχνά είναι πολύ πιο εύκολη διαδικασία από το να ελέγξουμε την υπόθεση με την κατασκευή

1. Η Μέθοδος Bootstrap

- συνόλων εμπιστοσύνης, τα οποία σε ορισμένες περιπτώσεις μπορεί να είναι και αδύνατο να κατασκευαστούν
- ii. Οι στατιστικοί έλεγχοι που εφαρμόζονται μέσω ελεγχουσυναρτήσεων μπορεί να είναι καλύτεροι γιατί λαμβάνουν υπόψη τους την ειδική φύση της υπόθεσης.
 - iii. Για τα bootstrap σύνολα εμπιστοσύνης, μπορούμε πάντα να παράγουμε bootstrap δείγματα από την εμπειρική κατανομή F_n , χωρίς κανένα περιορισμό. Για τον έλεγχο υποθέσεων μέσω της bootstrap, έχουμε την επιλογή να παράγουμε bootstrap δείγματα είτε από την F_n είτε από την εκτιμωμένη κατανομή κάτω από τους περιορισμούς της υπόθεσης.

Στην παράγραφο αυτή θα δώσουμε μια σύντομη και γενική περιγραφή των στατιστικών ελέγχων καθώς και τον τρόπο που η μέθοδος bootstrap τους προσεγγίζει. Επίσης θα εξετάσουμε και τον τρόπο υπολογισμού, μέσω της bootstrap, μερικών σημαντικών εννοιών των ελέγχων: της απορριπτικής περιοχής, της κρίσιμης τιμής, της ισχύς και του p-value.

1.5.1 Γενική Περιγραφή

Οι στατιστικοί έλεγχοι υποθέσεων μπορούν να περιγράψουν ως εξής: Ας είναι X_1, \dots, X_n τυχαίες μεταβλητές, όχι απαραίτητα ανεξάρτητες, που έχουν κοινή κατανομή την $F^{(n)}$ και $F^{(n)}$ είναι η συλλογή όλων των πιθανών $F^{(n)}$. Έστω $F_0^{(n)}$ και $F_1^{(n)}$ δυο ξένα υποσύνολα του $F^{(n)}$. Θέλουμε να προσδιορίσουμε, χρησιμοποιώντας τα δεδομένα X_1, \dots, X_n αν η υπόθεση $F^{(n)} \in F_0^{(n)}$ είναι αληθής, δηλαδή αν:

$$H_0 : F^{(n)} \in F_0^{(n)} \text{ έναντι } H_1 : F^{(n)} \in F_1^{(n)} \quad (1.29)$$

Στην σχέση (1.29) η H_0 καλείται μηδενική υπόθεση και η H_1 καλείται εναλλακτική υπόθεση. Στην ειδική περίπτωση όπου οι X_1, \dots, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την F , η $F^{(n)}$ προσδιορίζεται από την F και ο έλεγχος (1.29) απλοποιείται σε:

$$H_0 : F \in F_0 \text{ έναντι } H_1 : F \in F_1 \quad (1.30)$$

όπου F είναι η συλλογή όλων των πιθανών F , F_0 και F_1 είναι δυο ξένα υποσύνολα του F . Ο στατιστικός έλεγχος της υπόθεσης (1.29) ή της υπόθεσης (1.30) ισοδυναμεί με την εύρεση μιας απορριπτικής περιοχής R_n τέτοια ώστε να απορρίπτουμε την μηδενική υπόθεση H_0 αν

1^ο Κεφάλαιο

και μόνο αν $X_1, \mathbf{K}, X_n \in \mathbb{R}_n$. Μια απλή και αποτελεσματική μέθοδος είναι να χρησιμοποιήσουμε ένα στατιστικό $T_n = T_n(X_1, \mathbf{K}, X_n)$ και απορριπτική περιοχή την

$$\mathbb{R}_n = \{x : T_n \geq c_n\} \quad (1.31)$$

όπου η c_n καλείται κρίσιμη τιμή. Το στατιστικό T_n καλείται ελεγχοσυνάρτηση. Αν η \mathbb{R}_n δίνεται από την σχέση (1.31), η απορριπτική περιοχή \mathbb{R}_n ή η κρίσιμη τιμή c_n καθορίζεται από τον έλεγχο της πιθανότητας να απορρίψουμε την H_0 όταν η H_0 είναι σωστή (σφάλμα τύπου I), δηλαδή από την σχέση:

$$\sup_{F^{(n)} \in \mathbb{F}_0} \text{Prob}\{(X_1, \mathbf{K}, X_n) \in \mathbb{R}_n | F^{(n)}\} = a \quad (1.32)$$

όπου a είναι ένας μικρός αριθμός και $\text{Prob}\{\cdot | F^{(n)}\}$ είναι η πιθανότητα που αντιστοιχεί στην $F^{(n)}$. Για δοθέν n , αν η σχέση (1.32) είναι σωστή τότε ο έλεγχος με απορριπτική περιοχή την \mathbb{R}_n λέγεται ένας ακριβής έλεγχος σε επίπεδο σημαντικότητας a . Αν η υπόθεση που θέλουμε να εξετάσουμε σε σημαντικότητα a είναι περίπλοκη, τότε ο στατιστικός έλεγχος είναι δύσκολο ή και ακατόρθωτο να επιτευχθεί. Σε τέτοιες περιπτώσεις μπορούμε να θεωρήσουμε ότι για μεγάλο n , η απορριπτική περιοχή \mathbb{R}_n ή η κρίσιμη τιμή c_n καθορίζεται από την σχέση:

$$\lim_{n \rightarrow \infty} \sup_{F^{(n)} \in \mathbb{F}_0} \text{Prob}\{(X_1, \mathbf{K}, X_n) \in \mathbb{R}_n | F^{(n)}\} = a \quad (1.33)$$

Ωστόσο, όταν η $F_0^{(n)}$ είναι σύνθετη, υπάρχει το ενδεχόμενο να μην μπορεί να πραγματοποιηθεί ο στατιστικός έλεγχος που εκφράζεται από την (1.33). Αυτό μας οδηγεί στον ακόλουθο ορισμό:

ΟΡΙΣΜΟΣ 1.2

Έστω a το δοθέν ονομαστικό επίπεδο. Ένας στατιστικός έλεγχος με απορριπτική περιοχή την \mathbb{R}_n είναι ασυμπτωτικά σωστός αν

$$\lim_{n \rightarrow \infty} \text{Prob}\{(X_1, \mathbf{K}, X_n) \in \mathbb{R}_n | F^{(n)}\} = a$$

για κάποια ακολουθία $F^{(n)} \in F_0^{(n)}$, $n = 1, 2, \dots$. Ο στατιστικός έλεγχος είναι συνεπής αν

$$\lim_{n \rightarrow \infty} \text{Prob}\{(X_1, \mathbf{K}, X_n) \in \mathbb{R}_n | F^{(n)}\} = 1$$

για οποιαδήποτε ακολουθία $F^{(n)} \in F_0^{(n)}$, $n = 1, 2, \dots$

1. Η Μέθοδος Bootstrap

Σύμφωνα με την κλασσική προσέγγιση, βρίσκουμε ένα ασυμπτωτικά σωστό στατιστικό έλεγχο προσεγγίζοντας την κατανομή από την ασυμπτωτική κατανομή της κάτω από τους περιορισμούς της μηδενικής υπόθεσης H_0 . Δεδομένου της υπεροχής της bootstrap μεθόδου έναντι της κλασσικής ασυμπτωτικής προσέγγισης, μπορούμε να προσεγγίσουμε την ζητούμενη κατανομή κάτω από την H_0 . Ο στατιστικός έλεγχος που προκύπτει καλείται bootstrap στατιστικός έλεγχος υποθέσεων.

Ένας τρόπος πραγματοποίησης ενός bootstrap στατιστικού ελέγχου μια υπόθεσης είναι να κατασκευάσουμε bootstrap συνόλων εμπιστοσύνης. Εναλλακτικά, μπορούμε να παράγουμε bootstrap δεδομένα κάτω από τους περιορισμούς της H_0 . Αυτή η προσέγγιση αποκλείει την εμπειρική κατανομή στις περισσότερες περιπτώσεις. Αν η H_0 είναι απλή, τότε μπορούμε εύκολα να αναγνωρίσουμε την κατανομή από την οποία πρέπει να παράγουμε τα bootstrap δεδομένα. Ωστόσο αν η H_0 είναι περίπλοκη τότε μπορεί να είναι δύσκολο να αποφασίσουμε από ποια κατανομή θα πρέπει να παράγουμε τα bootstrap δεδομένα.

1.5.2 Υπολογισμός της Ισχύος

Έστω X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την F_q , όπου $q \in \Theta$. Αν το Θ είναι υποσύνολο του Ευκλείδειου χώρου τότε η F_q ανήκει σε μια παραμετρική οικογένεια. Ωστόσο το πρόβλημα των στατιστικών ελέγχων ανακύπτει όταν το Θ είναι οποιοσδήποτε μετρικός χώρος.

Έστω q η παράμετρος ενδιαφέροντος. Η υπόθεση που θέλουμε να ελέγξουμε εκφράζεται από την σχέση (1.30) με

$$F_0 = \{F_q : q = q_0\}$$

όπου q_0 είναι μια συγκεκριμένη τιμή του Θ και

$$F_1 = \{F_q : q \in \Theta, q \neq q_0\}$$

Έστω T_n ένα δοθέν στατιστικό με κατανομή

$$K_{n,q} = \text{Prob}_\theta \{T_n \leq q\}$$

όπου Prob_θ είναι η πιθανότητα που αντιστοιχεί στην κατανομή F_q .

Αν η K_{n,q_0} είναι συνεχής τότε μπορούμε να πάρουμε τον ακόλουθο εκτιμητή bootstrap της κρίσιμης τιμής c_n :

$$c_{n,BOOT} = K_{n,q_0}^{-1} (1-a)$$

Άρα ο bootstrap έλεγχος έχει απορριπτική περιοχή την:

$$R_n = \{T_n(X_1, \mathbf{K}, X_n) \geq c_{n,BOOT}\}$$

Ένα ερώτημα είναι πως μπορούμε να υπολογίσουμε το $c_{n,BOOT}$. Σπάνια ο έλεγχος αυτός μπορεί να πραγματοποιηθεί αναλυτικά και επομένως θα επικεντρωθούμε στην περίπτωση που υπολογίζεται μέσω της μεθόδου Monte Carlo. Αρχικά παράγουμε B ανεξάρτητα σύνολα bootstrap δειγμάτων $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$, $b=1, \dots, B$, από την F_{q_0} . Σε κάθε b bootstrap επανάληψη, $b=1, \dots, B$, υπολογίζουμε την τιμή του $T_n(X_1^*(b), \mathbf{K}, X_n^*(b))$. Τότε το $c_{n,BOOT}$ υπολογίζεται από την $[B(1-a)+1]$ -οστή τιμή του διατεταγμένου στατιστικού T_n .

Άλλο ένα σημαντικό θέμα είναι ο υπολογισμός της ισχύος του bootstrap ελέγχου. Η συνάρτηση ισχύος του bootstrap ελέγχου είναι η

$$b_n(q) = \text{Prob}_{\theta} \{T_n \geq c_n\} \quad q \in \Theta, q \neq q_0$$

Για δοθέν $q \neq q_0$ η $b_n(q)$ εκτιμάται από τον bootstrap εκτιμητή $b_{n,BOOT}(q)$ χρησιμοποιώντας εμφωλευμένη (*nested*) bootstrap. Για δοθέν $q \neq q_0$ παράγουμε B_1 ανεξάρτητα σύνολα bootstrap δειγμάτων $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$, $b=1, \dots, B_1$, από την F_q . Για κάθε b , παράγουμε B_2 ανεξάρτητα σύνολα από bootstrap δείγματα από την F_{q_0} και χρησιμοποιώντας την μέθοδο που περιγράψαμε παραπάνω βρίσκουμε το $\hat{c}_{n,b}$. Τότε το $b_{n,BOOT}(q)$ υπολογίζεται από την σχέση

$$b_{n,BOOT}(q) = \frac{1}{B_1} \sum_{b=1}^{B_1} I\{T_n(X_1^*(b), \mathbf{K}, X_n^*(b)) \geq \hat{c}_{n,b}\}$$

1.5.3 Υπολογισμός του p-value

Ας είναι $X_1 = x_1, \mathbf{K}, X_n = x_n$ οι παρατηρούμενες τιμές. Για τον έλεγχο της υπόθεσης που μπορεί να περιγράψει από την σχέση (1.29), το p-value ορίζεται να είναι το:

$$p_n = \text{Prob}\{T_n(X_1, \mathbf{K}, X_n) \geq T_n(x_1, \mathbf{K}, x_n) | F^{(n)}\}$$

για οποιοδήποτε $F^{(n)} \in F_0^{(n)}$. Ας υποθέσουμε ότι το p_n ορίζεται μονοσήμαντα. Ένας εκτιμητής bootstrap του p_n είναι:

1. Η Μέθοδος Bootstrap

$$p_{BOOT} = \text{Prob}_* \left\{ T_n(X_1^*, \mathbf{K}, X_n^*) \geq T_n(x_1, \mathbf{K}, x_n) \mid F^{(n)} \right\}$$

όπου X_1^*, \mathbf{K}, X_n^* είναι τα bootstrap δεδομένα, ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την εμπειρική κατανομή που βασίζεται στα x_1, \mathbf{K}, x_n . Μερικά σχετικά παραδείγματα δίνονται από τους Noreen (1989) και Krewski et al. (1991).

Ας υποθέσουμε ότι το $t_0 = E[T_n(X_1, \mathbf{K}, X_n)]$ είναι γνωστό όταν $F \in F_q$. Ας είναι $\hat{t}_0 = E_*[T_n(X_1^*, \mathbf{K}, X_n^*)]$. Ο Noreen (1989) πρότεινε ότι κάτω από την υπόθεση:

$$\text{Prob}_* \left\{ T_n(X_1^*, \mathbf{K}, X_n^*) - \hat{t}_0 \geq x \right\} \approx \text{Prob}_* \left\{ T_n(x_1, \mathbf{K}, x_n) - t_0 \geq x \right\} \quad (1.34)$$

μπορούμε να χρησιμοποιήσουμε ως bootstrap εκτιμητή του p-value το

$$\hat{p}_{BOOT} = \text{Prob}_* \left\{ T_n(X_1^*, \mathbf{K}, X_n^*) \geq T_n(x_1, \mathbf{K}, x_n) - t_0 + \hat{t}_0 \right\}$$

που λαμβάνεται αν μετατοπίσουμε το κέντρο της bootstrap κατανομής. Ο εκτιμητής \hat{p}_{BOOT} ισχύει αν ισχύει η (1.34). Αριθμητικά παραδείγματα δίνονται από τον Noreen (1989). Την bootstrap εκτίμηση του p-value p_{BOOT} ή \hat{p}_{BOOT} την συγκρίνουμε με το ονομαστικό επίπεδο και αναλόγως απορρίπτουμε ή αποδεχόμαστε την μηδενική υπόθεση.

Στην πράξη το p_{BOOT} είναι αυτό που συνήθως εκτιμάται. Αξίζει να δούμε πως υπολογίζεται μέσω της μεθόδου Monte Carlo. Έστω ότι X_1, \mathbf{K}, X_n είναι τα δείγμα μας και $T_n(X_1, \mathbf{K}, X_n)$ είναι το στατιστικό ελέγχου της υπόθεσης (1.30). Παράγουμε B ανεξάρτητα σύνολα bootstrap δειγμάτων $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$, $b=1, \dots, B$, από την F_0 . Σε κάθε b bootstrap επανάληψη, $b=1, \dots, B$, υπολογίζουμε την τιμή του $T_n(X_1^*(b), \mathbf{K}, X_n^*(b))$. Το p_{BOOT} υπολογίζεται από την σχέση:

$$\begin{aligned} p_{BOOT} &= \frac{\sum_{b=1}^B I \left\{ T_n(X_1^*(b), \mathbf{K}, X_n^*(b)) \geq T_n(X_1, \mathbf{K}, X_n) \right\}}{B} \\ &= \frac{\# \left\{ T_n(X_1^*(b), \mathbf{K}, X_n^*(b)) \geq T_n(X_1, \mathbf{K}, X_n) \right\}}{B} \end{aligned} \quad (1.35)$$

Ωστόσο οι Efron and Tibshirani (1993) όπως και οι Davison and Hinkley (1997) δίνουν ως το p_{BOOT} μέσω της σχέσης:

$$P_{BOOT} = \frac{\#\{T_n(X_1^*(b), \mathbf{K}, X_n^*(b)) \geq T_n(X_1, \mathbf{K}, X_n)\} + 1}{B+1} \quad (1.36)$$

Η διαφορά οφείλεται στον τρόπο ορισμού του p-value μέσω της μεθόδου Monte Carlo. Η μέθοδος Monte Carlo προσομοιώνει B τιμές κάτω από την μηδενική υπόθεση και επομένως το σύνολο των τιμών κάτω από την μηδενική υπόθεση είναι $B+1$. Στην σχέση (1.36), η μονάδα στον αριθμητή δηλώνει την τιμή $T_n(X_1, \mathbf{K}, X_n)$. Οι Davison and Hinkley (1997, κεφαλαίο 4) δηλώνουν ότι μπορεί να χρησιμοποιηθεί και η σχέση (1.35), τόσο στην παραμετρική όσο και στη μη παραμετρική bootstrap, χωρίς να υπάρχει διαφορά από θεωρητική σκοπιά.

1.5.4 Συμπεράσματα

Αν και ο έλεγχος στατιστικών υποθέσεων μέσω της μεθόδου bootstrap αυξάνει το οπλοστάσιο μας, το πλήθος των εφαρμογών περιορίζεται. Ένα κομβικό σημείο πριν την εφαρμογή της bootstrap είναι η κατάλληλη επιλογή ελεγχοσυνάρτησης ή/και μεθόδου bootstrap (παραμετρική ή μη) έτσι ώστε να βελτιωθεί η τάξη της ορθότητας του bootstrap ελέγχου. Οι Liu and Shigh (1987) με μέτρο σύγκρισης το μέσο τετραγωνικό σφάλμα έδειξαν ότι οι bootstrap έλεγχοι μπορεί να είναι καλύτεροι από αυτούς που βασίζονται σε προσεγγίσεις των κανονικών κατανομών. Ωστόσο ένας έλεγχος που βασίζεται σε ένα λιγότερο pivotal στατιστικό ελέγχου μπορεί να καταλήξει σε μια μεγάλη διαφορά μεταξύ του πραγματικού επιπέδου σημαντικότητας του ελέγχου και του ονομαστικού επιπέδου. Οι Bunke and Riemer (1983) το απέδειξαν για τον έλεγχο της μέσης τιμής. Οι Ducharme and Jhun (1986) και Quenneville (1986) έδειξαν ότι η αποτελεσματικότητα των bootstrap ελέγχων μπορεί να βελτιωθεί αν χρησιμοποιηθεί ένα τυποποιημένο στατιστικό ελέγχου. Ο Sutton (1983) πρότεινε πως να αποκτάμε ακριβή bootstrap ελέγχους για τον μέσο μιας ασύμμετρης κατανομής χρησιμοποιώντας το τροποποιημένο t-στατιστικό του Johnsons. Περισσότερα πλεονεκτήματα για την χρήση ενός pivot στατιστικού ως στατιστικό ελέγχου δόθηκαν από τους Hall and Wilson (1991). Τέλος, ο Beran (1988a) πρότεινε μια bootstrap prepivoting μέθοδο για να αυξηθεί η ορθότητα των bootstrap ελέγχων.

1.6 Υπολογιστικές Μέθοδοι

Η μέθοδος bootstrap είναι υπολογιστικά επίπονη (*computer-intensive*) μέθοδος. Η εφαρμογή της μεθόδου bootstrap διευκολύνεται στη σημερινή εποχή καθώς οι σύγχρονοι υπολογιστές

1. Η Μέθοδος Bootstrap

έχουν αποκτήσει μεγάλη υπολογιστική ισχύ. Ωστόσο, η έρευνα για τον υπολογισμό των εκτιμητών bootstrap είναι πολύ σημαντική για τους ακόλουθους λόγους:

- i. Υπάρχουν περιπτώσεις όπου ο απαιτούμενος υπολογισμός είναι ογκώδης ή πρακτικά αδύνατος.
- ii. Πιο αποτελεσματικοί μέθοδοι για τον υπολογισμό είναι ευπρόσδεκτοι, για να σώσουμε χρόνο και υπολογιστικό έργο.
- iii. Ερωτήσεις όπως πόσα bootstrap σύνολα δεδομένων μπορούν να παρθούν μέσω της Monte Carlo διαδικασίας συχνά απασχολούν τους αναλυτές πριν εφαρμόσουν την bootstrap μέθοδο.

Για λόγους απλότητας στην παρουσίαση θα επικεντρωθούμε στην απλή περίπτωση όπου τα δεδομένα X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από μια άγνωστη κατανομή F . Ο Huang (1991) παρουσίασε μια αναλυτική εξίσωση για τους bootstrap εκτιμητές της διασποράς των L-στατιστικών, ωστόσο οι υπολογισμοί που χρειάζονται για να εφαρμόσουμε αυτές τις εξισώσεις απαιτούν μεγάλο αριθμό υπολογιστικών συνδυασμών.

1.6.1 Πότε Χρειάζονται οι Υπολογιστικές μέθοδοι

Όπως αναφέραμε πιο νωρίς δεν είναι ανάγκη να καταφύγουμε πάντα σε υπολογιστικές μεθόδους. Για παράδειγμα, είχαμε αναφέρει ότι η u_{BOOT} που δίνεται από την σχέση (1.10) μπορεί να υπολογιστεί απευθείας αν μπορεί να γραφτεί ως μια αναλυτική συνάρτηση των X_1, \mathbf{K}, X_n . Σύμφωνα με όσα έχουμε αναφέρει οι υπολογισμοί επαναδειγματοληψίας βασίζονται στο παρατηρηθέν δείγμα και κατά κάποιο τρόπο δίνουν την αίσθηση εξαρτώνται από τα δεδομένα. Αυτό είναι πιο φανερό στην μη παραμετρική περίπτωση. Η γνώση που έχουμε για το στατιστικό μοντέλο βασίζεται μόνο στα δεδομένα, αφού δεν κάνουμε καμία υπόθεση από τα δεδομένα. Έτσι εύλογα δημιουργείται το ερώτημα αν και κατά πόσο η συμπερασματολογία που προκύπτει μέσω της μεθόδου bootstrap εξαρτάται από τα αρχικά μας δεδομένα. Μια απάντηση δίνεται από τους Davison and Hinkley (1997): όχι, αν η εξάρτηση ή η δέσμευση δεν μπορεί να εκφραστεί με κάποιον αναλυτικό τρόπο.

Η δεσμευμένη συμπερασματολογία είναι χρήσιμη στην παραμετρική συμπερασματολογία όταν τα επαρκή στατιστικά περιέχουν ένα βοηθητικό στατιστικό του οποίου η κατανομή είναι ελεύθερη παραμέτρων. Τότε μπορούμε να υποστηρίξουμε ότι η συμπερασματολογία των παραμέτρων πρέπει να βασίζεται στην δειγματική κατανομή που δεσμεύεται από το δοθέν δείγμα. Για παράδειγμα, τα διαστήματα εμπιστοσύνης που προκύπτουν από τις bootstrap

μεθόδους θα πρέπει να αντιμετωπίζονται ως διαστήματα εμπιστοσύνης που έχουν κλειστή μορφή (Davison and Hinkley, 1997).

Το πιο κλασσικό παράδειγμα είναι όταν έχουμε ένα δείγμα X_1, \mathbf{K}, X_n όπου το $n = 2m + 1$ με $m \in \mathbb{N}$. Τότε η bootstrap δειγματική κατανομή δίνεται από την σχέση :

$$H_{BOOT}(x_{(k)}) = \sum_{j=0}^m \binom{n}{j} \left(\frac{k-1}{n}\right)^j \left(\frac{n-k+1}{n}\right)^{n-j} - \sum_{j=0}^m \binom{n}{j} \left(\frac{k}{n}\right)^j \left(\frac{n-k}{n}\right)^{n-j}$$

όπου $x_{(k)}$ είναι k -οστή διατεταγμένη τιμή του δείγματος (βλ. Efron and Tibshirani, 1993, Davison and Hinkley, 1997, και Shao and Tu, 1995). Αλλά παραδείγματα δίνονται από τους Davison και Hinkley.

1.6.2 Μέθοδος Δέλτα

Ο Efron (1979) εφάρμοσε την Δέλτα μέθοδο (σειρές Taylor) για να προσεγγίσει τον bootstrap εκτιμητή της μεροληψίας και της διασποράς. Έστω $\{X_1^*, \mathbf{K}, X_n^*\}$ είναι ένα bootstrap δείγμα από την εμπειρική κατανομή F_n που βασίζεται στα $\{X_1, \mathbf{K}, X_n\}$. Ορίζουμε με

$$P_i^* = \frac{\#\{X_j^* = X_i, j=1, \mathbf{K}, n\}}{n}$$

και

$$\mathbf{P}^* = (P_1^*, P_2^*, \mathbf{K}, P_n^*)'$$

Δοθέντων των X_1, \mathbf{K}, X_n το $n\mathbf{P}^*$ ακολουθεί πολυωνυμική κατανομή με παραμέτρους n και $\mathbf{P}^0 = \mathbf{1}/n$ όπου $\mathbf{1} = (1, 1, \mathbf{K}, 1)'$. Τότε

$$E_*[\mathbf{P}^*] = \mathbf{P}^0 \text{ και } Var_*[\mathbf{P}^*] = n^{-2}\mathbf{I} - n^{-3}\mathbf{1}\mathbf{1}'$$

όπου \mathbf{I} είναι ο μοναδιαίος πίνακας και E_* και Var_* είναι η bootstrap αναμενομένη τιμή και διασπορά, δοθέντων των X_1, \mathbf{K}, X_n , αντίστοιχα.

Ο εκτιμητής bootstrap της ροπής μιας τυχαίας μεταβλητής $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F)$ είναι η ίδια η ροπή της bootstrap αναλογίας. Συνήθως το $\mathfrak{R}_n(X_1^*, \mathbf{K}, X_n^*, F_n) = \mathfrak{R}_n(\mathbf{P}^*)$ είναι μια συνάρτηση του \mathbf{P}^* . Για παράδειγμα, αν $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = \bar{X}_n - m$, όπου \bar{X}_n και m είναι ο δειγματικός μέσος και ο πληθυσμιακός μέσος αντίστοιχα, τότε

$$\mathfrak{R}_n(X_1^*, \mathbf{K}, X_n^*, F_n) = \bar{X}_n - \bar{X}_n = (\mathbf{X} - \bar{X}_n \mathbf{1})' \mathbf{P}^*$$

1. Η Μέθοδος Bootstrap

όπου $\mathbf{X} = (X_1, \mathbf{K}, X_n)'$.

Επειδή $E_*[\mathbf{P}^*] = \mathbf{P}^0$, μπορούμε να αναπτύξουμε το $\mathfrak{R}_n(\mathbf{P}^*)$ στο σημείο \mathbf{P}^0 . Παραβλέποντας τους ορούς που περιέχουν παραγώγους τρίτης ή μεγαλύτερης τάξης σε μια πολυδιάστατη Taylor σειρά, παίρνουμε ότι

$$\mathfrak{R}_n(\mathbf{P}^*) \approx \mathfrak{R}_n(\mathbf{P}^0) + \mathbf{U}'(\mathbf{P}^* - \mathbf{P}^0) + \frac{1}{2}(\mathbf{P}^* - \mathbf{P}^0)' \mathbf{V}(\mathbf{P}^* - \mathbf{P}^0) \quad (1.37)$$

όπου $\mathbf{U} = \nabla \mathfrak{R}_n(\mathbf{P}^0)$ και $\mathbf{V} = \nabla^2 \mathfrak{R}_n(\mathbf{P}^0)$. Από την σχέση (1.37) παίρνουμε αμέσως τις προσεγγίσεις των εκτιμητών bootstrap της μεροληψίας και της διασποράς

$$b_{BOOT} = E_*[\mathfrak{R}_n(\mathbf{P}^*)] \approx \mathfrak{R}_n(\mathbf{P}^0) + \frac{1}{2n} \text{tr}(\mathbf{V}) \quad (1.38)$$

και

$$u_{BOOT} = \text{Var}_*[\mathfrak{R}_n(\mathbf{P}^*)] \approx \frac{1}{n^2} \mathbf{U}' \mathbf{U} + \frac{1}{n^2} \sum_{i=1}^n U_i \quad (1.39)$$

όπου U_i είναι το i -οστό στοιχείο του \mathbf{U} .

1.6.3 Jackknife Προσέγγιση

Τα δεξιά μέλη των σχέσεων (1.38) και (1.39) είναι ακριβώς τα ίδια με αυτά του επ' απείρου εκτιμητή jackknife της μεροληψίας και της διασποράς που προτάθηκε από τον Jaeckel (1972). Αν αντικαταστήσουμε την παράγωγο U_i στην σχέση (1.39) από την προσέγγιση των πεπερασμένων διαφορών έχουμε

$$U_i^{\%} = \frac{\mathfrak{R}_n\left(\mathbf{P}^0 - \frac{1}{n-1} \left(\mathbf{e}_i - \frac{i}{n}\right)\right) - \mathfrak{R}_n(\mathbf{P}^0)}{-\frac{1}{n-1}} \quad (1.40)$$

όπου \mathbf{e}_i είναι το μοναδιαίο διάνυσμα στο \mathbf{R}^n τότε

$$U_i^{\%} = (n-1)(T_n - T_{n-1,i})$$

όπου $\mathfrak{R}_n(X_1, \mathbf{K}, X_n, F) = T_n - q$ για δοθέν στατιστικό T_n . Παρατηρούμε ότι η ποσότητα

$\frac{1}{n} \sum_{i=1}^n U_i^{\%}$ είναι σχεδόν ίδια με τον εκτιμητή jackknife της διασποράς u_{JACK} στην σχέση (1.17)

εκτός από ότι το $n \sum_{i=1}^n T_{n-1,i}$ έχει αντικαταστεί από το T_n . Αυτό δείχνει ότι οι εκτιμητές

jackknife μπορούν να θεωρηθούν ως προσεγγίσεις των εκτιμητών bootstrap. Συνεπώς, για την εκτίμηση της μεροληψίας και της διασποράς η μέθοδος jackknife είναι πιο απλή από την μέθοδο bootstrap και είναι πιο καλή όταν το μέγεθος του δείγματος είναι μεγάλο και η $\mathfrak{R}_n(\mathbf{P}^*)$ είναι ικανοποιητικά συνήθη εννοώντας ότι η προσέγγιση που κάνουμε στην σχέση (1.40) ισχύει. Ωστόσο σε μη συνήθη στατιστικά σχέση μεταξύ των μεθόδων bootstrap και jackknife μπορεί να μην υπάρχει.

1.6.4 Η Απλή Μέθοδος Monte Carlo

Ας είναι \hat{P}_n το εκτιμώμενο μοντέλο που χρησιμοποιεί τα δεδομένα X_1, \mathbf{K}, X_n . Οι εκτιμητές bootstrap της μεροληψίας και της κατανομής είναι αντίστοιχα οι

$$b_{BOOT} = E_* \left[\mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_m^*, \hat{P}_n \right) \right] \quad (1.41)$$

$$u_{BOOT} = Var_* \left[\mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_m^*, \hat{P}_n \right) \right] \quad (1.42)$$

και

$$H_{BOOT}(x) = Prob_* \left\{ \mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_m^*, \hat{P}_n \right) \leq x \right\} \quad (1.43)$$

όπου $\{X_1^*, \mathbf{K}, X_m^*\}$ είναι ένα δείγμα από την \hat{P}_n , $\mathfrak{R}_n(\cdot, \cdot)$ είναι μια κατάλληλα ορισμένη συνάρτηση και E_* , Var_* και $Prob_*$ είναι η δεσμευμένη αναμενομένη τιμή, διασπορά και πιθανότητα αντίστοιχα, για δοθέντα X_1, \mathbf{K}, X_n . Γενικά το n δεν είναι απαραίτητα ίσο με το μέγεθος m . Οι Politis and Romano (1993) αποδείξανε ότι αν $m^2/n \rightarrow 0$ όταν $n \rightarrow \infty$, τότε ο εκτιμητής bootstrap της δειγματικής κατανομής είναι συνεπής. Ωστόσο, για λόγους απλότητας θα εξετάσουμε την περίπτωση $n = m$.

Για να εφαρμόσουμε την απλή μέθοδο Monte Carlo για να προσεγγίσουμε τους εκτιμητές bootstrap των σχέσεων (1.41), (1.42) και (1.43), παράγουμε B ανεξάρτητα δείγματα $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$, $b=1, \dots, B$, από ένα εκτιμώμενο μοντέλο \hat{P}_n . Τότε υπολογίζουμε τα $\mathfrak{R}_n^*(b) = \mathfrak{R}_n(X_1^*(b), \mathbf{K}, X_n^*(b))$, για $b=1, \dots, B$, και υπολογίζουμε τα b_{BOOT} , u_{BOOT} και $H_{BOOT}(x)$ από τις σχέσεις

$$b_{BOOT}^{(B)} = \frac{1}{B} \sum_{b=1}^B \mathfrak{R}_n^*(b)$$

1. Η Μέθοδος Bootstrap

$$u_{BOOT}^{(B)} = \frac{1}{B} \sum_{b=1}^B \left(\mathfrak{R}_n^*(b) - \frac{1}{B} \sum_{j=1}^B \mathfrak{R}_n^*(j) \right)^2 \quad (1.44)$$

και

$$H_{BOOT}^{(B)}(x) = \frac{1}{B} \sum_{b=1}^B I \{ \mathfrak{R}_n^*(b) \leq x \} \quad (1.45)$$

Έστω ότι τα X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την κατανομή F και $\hat{P}_n = \hat{F}$ είναι ένας εκτιμητής της F . Αν η \hat{F} είναι η εμπειρική κατανομή, τότε το $\{X_1^*(b), \mathbf{K}, X_n^*(b)\}$ είναι ένα απλό τυχαίο δείγμα από τα $\{X_1, \mathbf{K}, X_n\}$. Η απλή μέθοδος Monte Carlo είναι εύκολη να προγραμματιστεί και να εφαρμοστεί. Το μόνο που θέλουμε είναι να πάρουμε B ανεξάρτητα δείγματα από την \hat{P}_n ή την \hat{F} και να υπολογίσουμε την τυχαία μεταβλητή $\mathfrak{R}_n^*(b)$.

Ένα σημαντικό ζήτημα είναι ο υπολογισμός του μεγέθους της μεθόδου Monte Carlo B . Το πρόβλημα αυτό, συνήθως απασχολεί τους εφαρμοσμένους στατιστικούς, και είναι παρόμοιο με το πρόβλημα του μεγέθους του δείγματος X_1, \mathbf{K}, X_n . Υπάρχουν δυο σημαντικές διαφορές μεταξύ των δυο αυτών προβλημάτων (Efron and Tibshirani, 1993):

- i. Στις περισσότερες περιπτώσεις μπορούμε να έχουμε το B πολύ μεγαλύτερο από το n .
- ii. Μεγαλύτερο μέγεθος του Monte Carlo δείγματος δεν μας εξασφαλίζει μείωση του σφάλματος του αρχικού bootstrap εκτιμητή.

Επομένως, πρέπει να διαλέξουμε ένα B τέτοιο ώστε το σφάλμα της μεθόδου Monte Carlo προσέγγισης να είναι ασήμαντο σε σχέση με το σφάλμα του αρχικού εκτιμητή bootstrap. Στην συνέχεια θα παρουσιάσουμε κάποιους τρόπους για να προσδιορίσουμε το B υποθέτοντας ότι τα bootstrap δεδομένα παράγονται από την εμπειρική κατανομή, η οποία είναι και η πιο δημοφιλής επιλογή όταν τα X_1, \mathbf{K}, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές

Η Monte Carlo προσέγγιση για τον εκτιμητή bootstrap της τυπικής απόκλισης του εκτιμητή \hat{q}_n είναι

$$s_{BOOT}^{(B)} = \sqrt{u_{BOOT}^{(B)}}$$

1^ο Κεφάλαιο

όπου η $u_{BOOT}^{(B)}$ δίνεται από την σχέση (1.44) για $\mathfrak{R}_n = \hat{q}_n$. Ο Serfling (1980) χρησιμοποιώντας τις κανονικές εξισώσεις για τους εκτιμητές των ροπών βρήκε ότι

$$E_* \left[s_{BOOT}^{(B)} \right] = \sqrt{u_{BOOT}}$$

και

$$Var_* \left[s_{BOOT}^{(B)} \right] \approx \frac{r_{BOOT} - u_{BOOT}^2}{4B u_{BOOT}}$$

όπου η u_{BOOT} δίνεται από την σχέση (1.42) και

$$r_{BOOT} = E_* \left[\hat{q}_n^* - E_* \left[\hat{q}^* \right] \right]$$

Επομένως, ο συντελεστής της μεταβλητότητας του $s_{BOOT}^{(B)}$ δοθέντων των X_1, \mathbf{K}, X_n είναι:

$$cv_* \left(s_{BOOT}^{(B)} \right) = \frac{\sqrt{Var_* \left[s_{BOOT}^{(B)} \right]}}{E_* \left[s_{BOOT}^{(B)} \right]} \approx \sqrt{\frac{\hat{d}_n + 2}{4B}} \quad (1.46)$$

όπου $\hat{d}_n = r_{BOOT} / u_{BOOT}^2 - 3$ είναι η δεσμευμένη κύρτωση του \hat{q}_n^* .

Υπάρχουν τουλάχιστον δυο τρόποι καθορισμού του Monte Carlo μεγέθους δείγματος B , βασιζόμενοι στην σχέση (1.46).

Η πρώτη μέθοδος προτάθηκε από τον Efron (1987) με σκοπό να προσδιορίσει το B θέτοντας

$$cv_* \left(s_{BOOT}^{(B)} \right) = e_0 \quad (1.47)$$

όπου e_0 είναι ένα δοθέν επιθυμητό επίπεδο. Στις περισσότερες περιπτώσεις $\hat{d}_n \approx 0$ ή πιο αυστηρά $\hat{d}_n \xrightarrow{P} 0$ όταν $n \rightarrow \infty$. Επομένως η σχέση (1.47) απλοποιείται στην σχέση

$$B = \frac{1}{2} e_0^{-2}$$

Για παράδειγμα αν $e_0 = 0.05$ τότε $B = 200$ ενώ αν $e_0 = 0.1$ τότε $B = 50$.

Δεν αξίζει πάντα να κάνουμε το $cv_* \left(s_{BOOT}^{(B)} \right)$ μικρό σε σχέση με το $cv \left(s_{BOOT} \right)$. Ας είναι $s_{BOOT} = \sqrt{u_{BOOT}}$ ο εκτιμητής bootstrap της τυπικής απόκλισης και $cv \left(s_{BOOT} \right)$ ο συντελεστής της μεταβλητότητας. Μερικές φορές αρκεί να ισχύει ότι $cv_* \left(s_{BOOT}^{(B)} \right) / cv \left(s_{BOOT} \right) \rightarrow 0$. Σε άλλες περιπτώσεις είναι αρκετό να ισχύει $cv_* \left(s_{BOOT}^{(B)} \right) = cv \left(s_{BOOT} \right)$.

1. Η Μέθοδος Bootstrap

Η δεύτερη μέθοδος, που προτάθηκε από τους Shao and Tu (1995), για τον προσδιορισμό του B , είναι να θέσουμε το

$$cv_*(s_{BOOT}^{(B)}) = cv(s_{BOOT})$$

το οποίο μας οδηγεί στην σχέση

$$B = \frac{a_n (\hat{d}_n + 2)}{4(cv(s_{BOOT}))^2} \quad (1.48)$$

όπου $a_n \equiv 1$ ή $\{a_n\}$ είναι μια ακολουθία θετικών ακεραίων αριθμών που συγκλίνουν στο άπειρο, για παράδειγμα $a_n = \log(\log(n))$. Το Monte Carlo μέγεθος δείγματος διαλέγεται σύμφωνα με την (1.48) και εξαρτάται από το n και την ακρίβεια του αρχικού εκτιμητή bootstrap. Ωστόσο απαιτούνται κάποια όρια για το $cv(s_{BOOT})$ και \hat{d}_n που χρειάζονται για να χρησιμοποιηθούν στην σχέση (1.48).

Μια παρόμοια ιδέα μπορεί να χρησιμοποιηθεί για να προσδιοριστεί το Monte Carlo μέγεθος δείγματος προσεγγίζοντας τα bootstrap σύνολα εμπιστοσύνης. Έστω $q_{BOOT}^{(B)}$ είναι η Monte Carlo προσέγγιση του bootstrap BC_a κάτω ορίου με επίπεδο σημαντικότητας $1-a$ υπολογισμένο αντικαταστήοντας την $K_{BOOT}(x)$ από την Monte Carlo προσέγγισή της:

$$K_{BOOT}^{(B)}(x) = \frac{1}{B} \sum_{b=1}^B I\{\hat{q}_n^*(b) \leq x\}$$

Αν υποθέσουμε ότι η συνάρτηση Ψ στην σχέση είναι η Φ και ότι τα z_0 και a είναι γνωστά τότε εφαρμόζοντας την εξίσωση για την συσχέτιση cv ενός δειγματικού ποσοστημορίου, παίρνουμε ότι

$$cv_*(q_{BOOT}^{(B)} - \hat{q}_n^*(b)) \approx \frac{1}{|z_a|f(z_a)} \sqrt{\frac{a(1-a)}{B}} \quad (1.49)$$

όπου $z_a = \Phi^{-1}(a)$. Μπορούμε να υπολογίσουμε το Monte Carlo μέγεθος δείγματος B χρησιμοποιώντας την σχέση (1.47) αν αντικαταστήσουμε την $cv_*(s_{BOOT}^{(B)})$ με $cv_*(q_{BOOT}^{(B)} - \hat{q}_n^*(b))$. Για παράδειγμα, αν $a = 0.025$ και $e_0 = 0.05$ τότε χρησιμοποιώντας την (1.49) παίρνουμε $B = 740$.

Ωστόσο, αν το z_0 το υπολογίζουμε από Monte Carlo προσέγγιση τότε το δεξιό μέλος της (1.49) μπορεί να αντικατασταθεί από την

$$\frac{1}{|z_a|} \sqrt{\frac{1}{B} \left(\frac{1}{(f(0))^2} - \frac{1}{f(0)f(z_a)} + \frac{a(1-a)}{(f(z_a))^2} \right)}$$

Αν πάλι $a = 0.025$ και $e_0 = 0.05$ το $B = 1170$. Είναι καλό να αναζητούμε ακόμα μεγαλύτερη τιμή του B όταν χρησιμοποιείται η μέθοδος Monte Carlo για τον υπολογισμό του a .

Οι Babu and Singh (1983) απέδειξαν ότι αν $\mathfrak{X}_n = (\bar{X}_n - E[X_1]) / \hat{S}_n$ είναι το τυποποιημένο στατιστικό για τον δειγματικό μέσο και B είναι μια συνάρτηση του n τέτοια ώστε $B / (n \log(n)) \rightarrow \infty$, τότε όταν το $n \rightarrow \infty$:

$$\sqrt{n} \sup_x \left| H_{BOOT}^{(B)}(x) - H_{BOOT}(x) \right| \rightarrow 0$$

Αυτό σημαίνει ότι η Monte Carlo προσέγγιση $H_{BOOT}^{(B)}(x)$ έχει ακρίβεια δεύτερης τάξης ως εκτιμητής της δειγματικής κατανομής του \mathfrak{X}_n .

Οι Shi et al. (1983) παρουσίασαν μια μέθοδο για τον υπολογισμό του B βασιζόμενοι στο ρυθμό σύγκλισης των H_{BOOT}^{-1} και $(H_{BOOT}^{(B)})^{-1}$. Ας είναι H_n η κατανομή της τυχαίας μεταβλητής \mathfrak{X}_n , H_{BOOT} ο εκτιμητής bootstrap της H_n που ορίστηκε στην σχέση (1.43) και $H_{BOOT}^{(B)}$ η Monte Carlo προσέγγιση του H_{BOOT} που δίνεται από την σχέση (1.45). Χρησιμοποιώντας την παρουσίαση του Bahadur για τις διαδικασίες των ποσοστιαίων σημείων, οι Shi et al. (1983) έδειξαν ότι

$$\sqrt{n} \sup_x \left| H_{BOOT}^{(B)}(x) - H_{BOOT}(x) \right| = e_n + \sqrt{B \log \log(B)}$$

και

$$\sqrt{n} \sup_{0 < t < 1} \left| (H_{BOOT}^{(B)})^{-1}(t) - H_{BOOT}^{-1}(t) \right| = O\left(e_n + \sqrt{B \log \log(B)}\right)$$

όπου $e_n = \sup_x \left| H_{BOOT}(x) - H_n(x) \right|$. Το σφάλμα της Monte Carlo προσέγγισης είναι αμελητέο σε σχέση με το σφάλμα e_n αν διαλέξουμε το B λύνοντας την εξίσωση:

$$B^{-1} \log \log(B) = o(e_n)$$

Ωστόσο, αν το n είναι πάρα πολύ μεγάλο, τότε μπορούμε να διαλέξουμε το B θέτοντας

$$B^{-1} \log \log(B) = C e_n$$

1. Η Μέθοδος Bootstrap

όπου C είναι μια σταθερά. Αν $e_n = O(n^{-1/2})$ ή $e_n = o_p(n^{-1/2})$, τότε μπορούμε να πάρουμε $B = na_n$ για a_n που ικανοποιεί την $\log(\log(n))/a_n \rightarrow 0$, το οποίο είναι σύμφωνο με το αποτέλεσμα των Babu και Singh. Αν $e_n = O(n^{-1})$ τότε μπορούμε να πάρουμε $B = n^2 \log(\log(n))$. Ένα μειονέκτημα της ασυμπτωτικής αυτής προσέγγισης είναι ότι οι σχέσεις μαρτυρούν απλά την τάξη του B συναρτήσει του n .

Όπως και στον προσδιορισμό του μεγέθους δείγματος n , η επιλογή του B δεν είναι εύκολο πρόβλημα. Παρόλο την πληθώρα μαθηματικών εξισώσεων για τον προσδιορισμό του B συνήθως όλα εξαρτώνται από την εμπειρία. Ως συμβουλή οι Efron and Tibshirani (1986) πρότειναν ότι για τους εκτιμητές bootstrap των ροπών το B θα πρέπει να επιλέγεται μεταξύ 50 και 200 ενώ για την κατανομή bootstrap και την εκτίμηση των ποσοστιαίων σημείων το B θα πρέπει να είναι τουλάχιστον 1000. Ο Chernick (1999), προτείνει ότι για τους εκτιμητές bootstrap των στατιστικών B θα πρέπει να τουλάχιστον 5000, ενώ για τα bootstrap σύνολα εμπιστοσύνης τουλάχιστον 10000.

1.6.5 Σχόλια

Οι υπολογιστικές μέθοδοι χωρίζονται σε δυο κατηγορίες: στις αναλυτικές και στις τεχνικές προσομοιώσεις. Η μέθοδος δέλτα και οι jackknife προσεγγίσεις που είδαμε, είναι δυο από τις αναλυτικές μεθόδους. Άλλη αναλυτική μέθοδος είναι η saddle-point προσέγγιση. Η απλή Monte Carlo προσέγγιση είναι μια τεχνική προσομοίωσης. Άλλες τεχνικές που ανήκουν στην ίδια κατηγορία είναι η Balanced bootstrap resampling, η centering after Monte Carlo, η linear bootstrap, η antithetic bootstrap resampling, η importance bootstrap resampling και η one-step bootstrap. Περισσότερες λεπτομέρειες για τις μεθόδους που δεν αναφέρθηκαν μπορούν να βρεθούν στους Chernick (1999), Davison and Hinkley (1997), Efron and Tibshirani (1993) και Shao and Tu (1995)

Η απλή μέθοδος Monte Carlo είναι η πιο απλή και η πιο δημοφιλής μέθοδος και γενικά μπορεί να αντιμετωπίσει οποιοδήποτε πρόβλημα παρουσιαστεί. Ωστόσο μπορεί να φανεί προβληματική όταν χρειάζεται μεγάλο μέγεθος δείγματος Monte Carlo. Οι άλλες μέθοδοι που αναφέραμε έχουν περιορισμένο πεδίο εφαρμογών. Για παράδειγμα, η χρήση της μεθόδου δέλτα δεν μπορεί να εφαρμοστεί όταν θέλουμε να εκτιμήσουμε την bootstrap δειγματική κατανομή αλλά μπορεί να εκτιμήσει τους bootstrap εκτιμητές των ροπών.

Οι Efron and Tibshirani (1993) παρουσιάζουν την μέθοδο bootstrap μόνο με χρήση των υπολογιστικών τεχνικών και κυρίως με χρήση της απλής Monte Carlo προσέγγισης. Ωστόσο συχνά αναφέρονται στην ιδανική bootstrap εκτίμηση μιας παραμέτρου. Τότε θεωρούν ότι $B = \infty$. Αυτή η υπόθεση είναι σίγουρα καταχρηστική, μα υπονοούν ότι η ιδανική εκτίμηση bootstrap είναι αυτή που προκύπτει όταν είναι δυνατή η εκτίμηση της χωρίς την χρήση υπολογιστικών τεχνικών.

1.7 Περιπτώσεις Όπου η Μέθοδος Bootstrap Αποτυγχάνει.

Στην παράγραφο αυτή θα προσπαθήσουμε να αναλύσουμε μερικές περιπτώσεις όπου η εφαρμογή της μεθόδου bootstrap αποτυγχάνει να δώσει σωστά αποτελέσματα. Αυτό συνήθως συμβαίνει όταν εφαρμόζουμε την μέθοδο bootstrap χωρίς δεύτερη σκέψη. Μερικές φορές πρέπει να λάβουμε υπόψη μας και την φύση του προβλήματος. Πάντως για τα περισσότερα προβλήματα έχουν βρεθεί λύσεις που είτε επιλύουν το πρόβλημα είτε μπορούν να καλυτερέψουν την απόδοση της μεθόδου bootstrap. Περισσότερες λεπτομέρειες μπορούν να βρεθούν στους Karllis (2004), Chernick (1999), και Davison and Hinkley (1997).

Ένα πρόβλημα της τεχνικής bootstrap προκύπτει όταν το μέγεθος του δείγματος n είναι μικρό. Σε αυτήν την περίπτωση η μέθοδος bootstrap θεωρεί ότι η εμπειρική κατανομή του δείγματος είναι μια καλή εκτίμηση της άγνωστης κατανομής του πληθυσμού. Όμως για μικρά δείγματα κάτι τέτοιο δεν είναι αληθές καθώς η προσέγγιση της εμπειρικής κατανομής είναι μάλλον άσχημη. Συνεπώς τα bootstrap δείγματα που παίρνουμε θα έχουν μεγάλο δειγματοληπτικό σφάλμα εξαιτίας του μεγέθους του δείγματος n . Ας υποθέσουμε ότι έχουμε n διακριτές τυχαίες μεταβλητές. Ο Hall (1992a) παραθέτει ότι ο αριθμός των πιθανών δειγμάτων είναι ίσος με

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

Στην συνέχεια αποδεικνύει ότι μέσω της μεθόδου Monte Carlo, για μέγεθος δείγματος $n = 20$ και για $B = 2000$ επαναλήψεις, η πιθανότητα κανένα από τα δυνατά δείγματα να μην εμφανιστεί πάνω από μια φορά είναι μεγαλύτερη από 0.95. Ωστόσο το πρόβλημα αυτό μπορεί να λυθεί όταν το n είναι πάρα πολύ μικρό, χρησιμοποιώντας όλα τα δυνατά bootstrap δείγματα. Οι Fisher and Hall (1990) και οι Diaconis and Holmes (1994) παρουσιάζουν την ιδέα και τους αντίστοιχους αλγόριθμους.

Ένα άλλο πρόβλημα μπορεί να προκύψει όταν θέλουμε να εκτιμήσουμε ποσότητες οι οποίες δεν έχουν καλά ορισμένες ροπές. Για παράδειγμα, αν τα δεδομένα προέρχονται από

1. Η Μέθοδος Bootstrap

την Cauchy κατανομή ξέρουμε πως η αναμενομένη τιμή της κατανομής δεν υπάρχει. Επομένως, η χρήση της μεθόδου bootstrap για την εκτίμηση του τυπικού σφάλματος της δειγματικής μέσης τιμής δεν μπορεί να οδηγήσει σε σωστό αποτέλεσμα.

Η μη-παραμετρική μέθοδος bootstrap ανεξάρτητα του μεγέθους δείγματος εκτιμά σωστά τη μέση τιμή και την τυπική απόκλιση του. Ωστόσο δεν κατορθώνει να εκτιμήσει σωστά τις ακραίες τιμές. Ο Athreya (1987) απέδειξε ότι σε αυτήν την περίπτωση η κατανομή bootstrap είναι μια τυχαία κατανομή πιθανότητας και όχι μια σταθερή κατανομή. Τα πράγματα μπορούν να βελτιωθούν αν χρησιμοποιήσουμε την παραμετρική μέθοδο bootstrap και μεγάλο αριθμό επαναλήψεων.

Η μη παραμετρική μέθοδος bootstrap αποτυγχάνει όταν η άγνωστη κατανομή του πληθυσμού είναι μια συνεχής κατανομή. Ο λόγος είναι ότι προσπαθούμε να προσεγγίσουμε μια συνεχή κατανομή με την εμπειρική κατανομή, δηλαδή μια διακριτή κατανομή. Όταν γνωρίζουμε την κατανομή από την οποία προέχονται τα δεδομένα καλό είναι να χρησιμοποιήσουμε την παραμετρική μέθοδο bootstrap.

Γενικά η μη παραμετρική μέθοδος bootstrap δεν δουλεύει για εξαρτημένα δεδομένα. Ας είναι X_1, \mathbf{K}, X_n το αρχικό μας δείγμα. Η μη παραμετρική μέθοδος bootstrap υποθέτει ότι υπάρχει ανεξαρτησία μεταξύ των $X_i, i=1, \dots, n$ και επομένως η από κοινού πιθανότητα των X_1, \mathbf{K}, X_n θα είναι το γινόμενο των περιθωρίων κατανομών των $X_i, i=1, \dots, n$. Προφανώς κάτι τέτοιο δεν ισχύει αν οι X_i, \mathbf{K}, X_n δεν είναι ανεξάρτητες. Η δυσκολία είναι ότι δεν υπάρχει προφανής τρόπος ώστε να βρούμε την από κοινού κατανομή των X_i, \mathbf{K}, X_n δοθέντος ότι οι X_i, \mathbf{K}, X_n είναι εξαρτημένες μεταβλητές. Ωστόσο υπάρχουν κάποιες περιπτώσεις όπου μπορεί η παραμετρική μέθοδος bootstrap να δώσει σωστά αποτελέσματα.

Άλλη περίπτωση όπου η μέθοδος bootstrap αποτυγχάνει είναι όταν υπάρχουν έκτροπες παρατηρήσεις στα δεδομένα μας. Με την παραδοσιακή προσέγγιση αν ανιχνεύσουμε έκτροπες παρατηρήσεις μπορούμε είτε να τις απομακρύνουμε είτε να τις διορθώσουμε. Όταν όμως πραγματοποιούμε μη παραμετρική μέθοδο bootstrap από την εμπειρική κατανομή του δείγματος τότε τα δεδομένα μας και το μοντέλο ταυτίζονται. Επομένως κανένα από τα εργαλεία που είχαμε στην διάθεση μας για την ανίχνευση μη καλής προσαρμογής των δεδομένων μας δεν μπορεί να μας ωφελήσει. Το μόνο που μπορούμε να κάνουμε είναι να εξετάσουμε από την προσομοίωση αν τα αποτελέσματα εξαρτώνται από τις έκτροπες παρατηρήσεις.

1.8 Σύντομη Ιστορική Αναδρομή

Οι μέθοδοι Monte Carlo είχαν χρησιμοποιηθεί για την εκτίμηση παραμέτρων, αρκετά προτού τις συνδέσει ο Efron (1979) για την χρήση της παραμετρικής συμπερασματολογίας με τον όρο bootstrap. Η ανάπτυξη της μεθόδου bootstrap ήταν ραγδαία και το 1982 η μονογραφία του Efron την σύνδεσε και με την μέθοδο jackknife.

Οι Young και Daniels (1991) ασχοληθήκαν με την μεροληψία της μη παραμετρικής μεθόδου bootstrap όταν χρησιμοποιείται η εμπειρική κατανομή στη θέση της άγνωστης κατανομής.

Ο Hall (1988, 1992b) υποστηρίζει την χρήση pivotal στατιστικών για την κατασκευή των bootstrap διαστημάτων εμπιστοσύνης, για τον έλεγχο υποθέσεων και συνδέει τις Edgeworth expansions για ομαλά στατιστικά.

Ο Efron (1979) προτείνει την χρήση διαστημάτων εμπιστοσύνης που βασίζονται σε ποσοστιαία σημεία της bootstrap κατανομής. Τα BC_a διαστήματα εμπιστοσύνης δίνονται από τον Efron (1987).

Ο Hall (1986) αναλύει την επίδραση της διακριτότητας στα διαστήματα εμπιστοσύνης. Ο Efron (1987) μελετάει τον αριθμό των προσομοιώσεων που απαιτούνται για την εκτίμηση της μεροληψίας και των ποσοστιαίων σημείων της bootstrap κατανομής, ενώ οι Diaconis and Holmes (1994) περιγράφουν πως μπορεί να αποφευχθεί η προσομοίωση και να αντικατασθεί από την απαρίθμηση όλων των δυνατών bootstrap συνόλων.

Οι Bickel and Freedman (1981) και ο Singh (1981) ήταν οι πρώτοι που απέδειξαν την συνέπεια της μεθόδου bootstrap κάτω από κάποιες μαθηματικές συνθήκες. Μάλιστα, οι Bickel and Freedman (1981) παρουσιάζουν και ένα παράδειγμα για την ασυνέπεια των εκτιμητών bootstrap. Η δουλειά τους μελετήθηκε από άλλους στατιστικούς, και αναπτύχθηκε μια θεωρητική βιβλιογραφία για τροποποιήσεις που μας εξασφαλίζουν ότι η μέθοδος bootstrap είναι συνεπής για διάφορες κλάσεις στατιστικών. Η κυρία τροποποίηση είναι η εξομάλυνση των δεδομένων, η οποία μπορεί να βελτιώσει την απόδοση των μη ομαλών στατιστικών, όπως το αντιμετωπίζουν για παράδειγμα οι De Angelis and Young, (1992). Η εργασία του Hall (1992b) είναι το κλειδί στην σύνδεση των Edgeworth Expansions με την μέθοδο bootstrap ενώ ο Mammen (1992) συζητεί πως οι προσομοιώσεις βοηθούν όταν η μέθοδος bootstrap λειτουργεί σωστά και δίνει θεωρητικά αποτελέσματα για διάφορες περιπτώσεις.

1. Η Μέθοδος Bootstrap

Τέλος, οι Shao and Tu (1995) παρουσιάζουν μια αρκετά ενδιαφέρουσα και λεπτομερή θεωρητική προσέγγιση των μεθόδων bootstrap και jackknife.

1.9 Εφαρμογές της Μεθόδου Bootstrap σε Διακριτά Δεδομένα

Οι εφαρμογές της μεθόδου bootstrap σε κατηγορικά δεδομένα είναι αρκετά περιορισμένη και συνδέεται κυρίως με την κατασκευή διαστημάτων εμπιστοσύνης.

Οι πρώτοι που κατασκεύασαν bootstrap διαστήματα εμπιστοσύνης παραμέτρους διακριτών κατανομών είναι οι Conlon and Thomas (1990). Πιο συγκεκριμένα πρότειναν την κατασκευή bootstrap διαστημάτων εμπιστοσύνης για την διαφορά δύο ανεξάρτητων διωνυμικών ποσοστών. Ο Åstrand (2000) πρότεινε την χρήση bootstrap διαστημάτων εμπιστοσύνης για ένα διωνυμικό ποσοστό. Οι Jhun and Jeong (2000) στηριζόμενοι στα αποτελέσματα του Beran (1988a,1988b) κατασκεύασαν ταυτόχρονα bootstrap διαστήματα εμπιστοσύνης για πολυωνυμικά ποσοστά και για τις αντιθέσεις πολυωνυμικών ποσοστών. Οι Morales et al. (2004) μελετήσανε την εργασία των Jhun and Jeong (2000) και πρότειναν την κατασκευή ταυτόχρονων bootstrap διαστημάτων εμπιστοσύνης για πολυωνυμικά ποσοστά με την χρήση της οικογένειας του power-divergence στατιστικού. Την ίδια περίοδο οι Clarke and Smith (2004) παρουσίασαν την κατασκευή bootstrap διαστημάτων εμπιστοσύνης για τις παραμέτρους λογαριθμογραμμικών μοντέλων με δίτιμη μεταβλητή απόκρισης, που είναι non-ignorable non-response. Οι Guerra et al. (1997) παρουσίασαν ομαλά bootstrap διαστήματα εμπιστοσύνης για κατηγορικά δεδομένα. Οι Jeong et al. (2005) μελέτησαν τον bootstrap δεσμευμένο έλεγχο της ανεξαρτησίας όταν ισχύει το μοντέλο συνάφειας RC. Ο Sauermann (1989) χρησιμοποίησε τους bootstrap εκτιμητές προκειμένου να υπολογίσει τους εκτιμητές μέγιστης πιθανοφάνειας σε λογαριθμογραμμικά μοντέλα μεγάλης διάστασης.

Στο δεύτερο κεφάλαιο της παρούσας εργασίας θα ασχοληθούμε με την κατασκευή bootstrap διαστημάτων εμπιστοσύνης για ένα διωνυμικό ποσοστό που προτείνει ο Åstrand (2000) και θα προτείνουμε μια τροποποίηση που βελτιώνει την απόδοσή τους. Στην συνέχεια (3^ο Κεφάλαιο) θα εξετάσουμε τα αποτελέσματα των Jhun and Jeong (2000) για πολυωνυμικά ποσοστά. Στο 4^ο Κεφάλαιο θα αναπτύξουμε τον bootstrap έλεγχο για την διαφορά δυο συσχετισμένων ποσοστών και θα μελετήσουμε την κατασκευή bootstrap διαστημάτων εμπιστοσύνης για τη διαφορά διωνυμικών ποσοστών. Τέλος, στο 5^ο Κεφάλαιο θα δούμε τον δεσμευμένο έλεγχο της ανεξαρτησίας όταν ισχύει το μοντέλο συνάφειας RC.

2^ο Κεφάλαιο

Μέθοδος Bootstrap για Διωνυμικές τ.μ.

2.0 Εισαγωγή

Παραθέτουμε αρχικά τους ορισμούς της διωνυμικής τ.μ. και των παραμέτρων της στην Παράγραφο 2.1. Στην συνέχεια (Παράγραφος 2.2) αναφερόμαστε στην εκτίμηση της διωνυμικής πιθανότητας επιτυχίας. Στην Παράγραφο 2.3 παρουσιάζουμε τις παραδοσιακές μεθόδους κατασκευής διαστημάτων εμπιστοσύνης για την πιθανότητα επιτυχίας. Στην Παράγραφο 2.4 εφαρμόζουμε την παραμετρική και την μη παραμετρική μέθοδο bootstrap σε διωνυμικές τ.μ.. Τα bootstrap διαστήματα εμπιστοσύνης παρουσιάζονται στην Παράγραφο 2.5. Ακολουθεί στην Παράγραφο 2.6 η παρουσίαση των κριτηρίων βάσει των οποίων θα συγκριθούν τα διαστήματα εμπιστοσύνης. Τα αποτελέσματα της σύγκρισης αυτής παρουσιάζονται στην επόμενη παράγραφο. Τέλος, το κεφάλαιο κλείνει με γενικά συμπεράσματα της σύγκρισης αυτής.

2.1 Η Διωνυμική Τυχαία Μεταβλητή

Η διωνυμική διαδικασία (βλ. Simonoff, 2003) χαρακτηρίζεται από τις παρακάτω τέσσερις ιδιότητες:

- i. Υπάρχει ένας σταθερός, πεπερασμένος αριθμός δοκιμών, n .
- ii. Κάθε δοκιμή έχει δύο πιθανά αποτελέσματα, τα οποία ονομάζονται «επιτυχία» και «αποτυχία».
- iii. Η πιθανότητα επιτυχίας, p , είναι ίδια σε όλες τις δοκιμές.
- iv. Οι δοκιμές είναι ανεξάρτητες μεταξύ τους.

Όταν ισχύουν οι παραπάνω ιδιότητες, τότε η τυχαία μεταβλητή X που μετρά το συνολικό αριθμό επιτυχιών στις n δοκιμές καλείται διωνυμική τυχαία μεταβλητή, ή ισοδύναμα γράφουμε $X \sim B(n, p)$.

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

Ένας ισοδύναμος ορισμός που θα μας φανεί χρήσιμος παρακάτω δίνεται από τον Agresti (2002). Έστω x_1, \mathbf{K}, x_n οι αποκρίσεις n ανεξάρτητων και πανομοιότυπων δοκιμών για τις οποίες ισχύει:

$$\text{Prob}\{X_i = 1\} = p \text{ και } \text{Prob}\{X_i = 0\} = 1 - p, \text{ με } i = 1, \dots, n.$$

Με την τιμή 1 συμβολίζουμε την επιτυχία ενώ με την τιμή 0 την αποτυχία. Με τον όρο πανομοιότυπες δοκιμές εννοούμε ότι κάθε δοκιμή έχει σταθερή πιθανότητα επιτυχίας και με τον όρο ανεξάρτητες δοκιμές ότι οι τυχαίες μεταβλητές X_1, \mathbf{K}, X_n είναι ανεξάρτητες. Τέτοιες δοκιμές καλούνται δοκιμές Bernoulli. Ο συνολικός αριθμός επιτυχιών $X = \sum_{i=1}^n X_i$ στις n

ανεξάρτητες δοκιμές Bernoulli ακολουθεί την $B(n, p)$.

Η τυχαία μεταβλητή X έχει συνάρτηση πιθανότητας

$$\text{Prob}\{X = x\} = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, \dots, n$$

όπου η ποσότητα

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

καλείται διωνυμικός συντελεστής. Η αθροιστική συνάρτηση κατανομής της X είναι

$$F(x) = \text{Prob}\{X \leq x\} = \begin{cases} 0 & , x < 0 \\ \sum_{r=0}^k \binom{n}{r} p^r (1-p)^{n-r} & , k \leq x < k+1 \\ & k = 0, \dots, n-1 \\ 1 & , x \geq n \end{cases}$$

Η τυχαία μεταβλητή $X \sim B(n, p)$ έχει αναμενομένη τιμή και διασπορά:

$$E[X] = np \text{ και } \text{Var}[X] = np(1-p)$$

αντίστοιχα.

Ας συμβολίσουμε με $X_{(i)}$ την i -οστή δυνατή διατεταγμένη τιμή της τ.μ. X , δηλαδή $X_{(i)} = i$, και με $p_{(i)}$ την πιθανότητα

$$p_{(i)} = \text{Prob}\{X = X_{(i)} = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n$$

2^ο Κεφάλαιο

Είναι προφανές ότι $\sum_{i=0}^n \text{Prob}\{X = X_{(i)}\} = \sum_{i=0}^n p_{(i)} = 1$.

Με $x_{n,p;a}$ θα συμβολίζουμε το a ποσοστιαίο σημείο της $B(n,p)$ που δίνεται από την σχέση:

$$x_{n,p;a} = \inf_{0 \leq i \leq n} \left\{ i : \sum_{j=0}^i p_{(j)} \geq a \right\} \quad (2.1)$$

Έστω m ανεξάρτητες και ισόνομες διωνυμικές τυχαίες μεταβλητές $X_1, \mathbf{K}, X_m \sim B(n,p)$. Τότε η τυχαία μεταβλητή που μετρά τον συνολικό αριθμό των επιτυχιών στις $m \cdot n$ δοκιμές, είναι επίσης διωνυμική $\sum_{i=1}^m X_i \sim B(mn, p)$.

Στις περισσότερες περιπτώσεις η πιθανότητα επιτυχίας δεν είναι γνωστή και πρέπει να την εκτιμήσουμε. Στην επόμενη παράγραφο θα ασχοληθούμε με τους διάφορους εκτιμητές της πιθανότητας επιτυχίας p .

2.2 Εκτίμηση του Διωνυμικού Ποσοστού

Έστω ότι η $X \sim B(n,p)$ και ότι έχουμε παρατηρήσει την τιμή x μετά την εκτέλεση των n ανεξάρτητων δοκιμών.

Στην παρούσα εργασία θα μας απασχολήσουν οι παρακάτω εκτιμητές:

- i. Ο εκτιμητής μέγιστης πιθανοφάνειας p :

$$p = \frac{x}{n} \quad (2.2)$$

- ii. Ο Laplace (1812) εκτιμητής p_L :

$$p_L = \frac{x+1}{n+2} \quad (2.3)$$

- iii. Ο Wilson (1927) εκτιμητής p_W :

$$p_W = \frac{x+2}{n+4} \quad (2.4)$$

Ο πιο δημοφιλής εκτιμητής του p είναι ο εκτιμητής μέγιστης πιθανοφάνειας. Από τον τύπο (2.2) παρατηρούμε ότι ισούται με το δειγματικό ποσοστό των επιτυχιών στις n ανεξάρτητες δοκιμές. Η ευρεία χρήση του οφείλεται και στο γεγονός ότι ο εκτιμητής p είναι αμερόληπτος εκτιμητής του p σε αντίθεση με p_L και p_W , που είναι μεροληπτικοί. Ο Wilson

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

παρατηρώντας ότι η κατανομή των πιθανοτήτων είναι ασύμμετρη θεώρησε ότι είναι άδικη η εκτίμηση του p από την κεντρική τιμή και γι' αυτό πρότεινε την χρήση του p_w .

Αν παρατηρήσουμε τον τύπο (2.4) συμπεραίνουμε ότι ο εκτιμητής p_w είναι ο εκτιμητής μέγιστης πιθανοφάνειας όταν στις n ανεξάρτητες δοκιμές προσθέσουμε τέσσερις ψευδοδοκιμές, τις οποίες τις μοιράζουμε σε επιτυχίες και αποτυχίες. Όμοιο συμπέρασμα μπορεί να προκύψει και για τον εκτιμητή p_L , με μόνη διαφορά ότι προσθέτουμε δύο ψευδοδοκιμές.

Ένα άλλο ζήτημα που μας απασχολεί για να ολοκληρώσουμε την αναφορά μας στα ποσοστά είναι η κατασκευή διαστημάτων εμπιστοσύνης για την πιθανότητα επιτυχίας p .

2.3 Διαστήματα Εμπιστοσύνης για το Διωνυμικό Ποσοστό

Η κατασκευή διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό επιτυχίας είναι ένα από τα πιο βασικά και ταυτοχρόνως πολυσυζητημένα προβλήματα της στατιστικής επιστήμης. Για να διαπιστώσει ο ενδιαφερόμενος αναγνώστης την πολυπλοκότητα του προβλήματος μπορεί να ανατρέξει στο άρθρο των Brown et al. (2001), όπου στην συζήτηση που ακολουθεί συμμετέχουν οι Agresti, Coull, Casella, Corcoran, Mehta, Ghosh και Santner. Στο άρθρο αυτό παρουσιάζονται και εξετάζονται διάφορες μέθοδοι ωστόσο καμία δεν έχει τύχει καθολικής αποδοχής. Στην παρούσα παράγραφο θα παρουσιάσουμε κάποιες από τις πιο δημοφιλείς μεθόδους.

2.3.1 Wald Διάστημα Εμπιστοσύνης

Το πιο αναγνωρίσιμο $100(1-2\alpha)\%$ διάστημα εμπιστοσύνης για το p , είναι το

$$[\underline{p}_w, \bar{p}_w] = \left[p - z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}, p - z_\alpha \sqrt{\frac{p(1-p)}{n}} \right] \quad (2.5)$$

όπου z_c είναι το c ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Το διάστημα αυτό ονομάζεται Wald διάστημα εμπιστοσύνης για το p γιατί η κατασκευή του οφείλεται στην αντίστροφη του ελέγχου $H_0: p = p_0$ έναντι $H_1: p \neq p_0$, που πραγματοποιείται μέσω του Wald στατιστικού $(p - p_0) / \sqrt{\frac{p(1-p)}{n}}$. Είναι προσεγγιστικό διάστημα εμπιστοσύνης και επομένως η εγκυρότητα του εξαρτάται από το μέγεθος του δείγματος n . Μεταξύ άλλων, οι Agresti and Coull (1998), Newcombe (1998a) και οι Brown

et al. (2001), δείχνουν ότι η απόδοση του Wald διαστήματος εμπιστοσύνης δεν είναι καλή ακόμα και αν το n είναι αρκετά μεγάλο.

2.3.2 Clopper-Pearson (Exact) Διάστημα Εμπιστοσύνης

Το $100(1-2\alpha)\%$ Clopper-Pearson διάστημα εμπιστοσύνης, γνωστό και ως ακριβές διάστημα εμπιστοσύνης (*Exact Confidence Interval*), για το p , είναι οι λύσεις των εξισώσεων

$$\sum_{k=x}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha$$

και

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha$$

όπου x είναι ο αριθμός των επιτυχιών στο δείγμα μας. Όταν $x=0$ τότε το κάτω άκρο είναι ίσο με 0 ενώ όταν $x=n$ το άνω άκρο είναι ίσο με 1. Έχει αποδειχθεί ότι για $x=1, \dots, n-1$ τα άκρα του Clopper-Pearson διαστήματος εμπιστοσύνης δίνονται σε κλειστή μορφή ως ποσοστιαία σημεία της βήτα κατανομής. Πιο συγκεκριμένα για $x=1, \dots, n-1$, το $(1-2\alpha)\%$ Clopper-Pearson διάστημα εμπιστοσύνης είναι το

$$[\underline{p}_{CP}, \bar{p}_{CP}] = [b_a(x, n-x+1), b_{1-a}(x+1, n-x)]$$

όπου $b_c(a, b)$ είναι το c ποσοστιαίο σημείο της βήτα κατανομής με παραμέτρους a και b .

Ο Neyman (1935) είναι από τους πρώτους που έδειξαν ότι η μέθοδος Clopper-Pearson είναι συντηρητική. Αυτό οφείλεται στην διακριτότητα της διωνυμικής κατανομής. Για τον λόγο αυτό έχουν προταθεί διάφορα διαστήματα εμπιστοσύνης που βελτιώνουν την αδυναμία αυτή.

2.3.3 Wilson (Score) Διάστημα Εμπιστοσύνης

Ο Wilson (1927) πρότεινε την κατασκευή διαστημάτων εμπιστοσύνης που βασίζονται στην αντίστροφη του ελέγχου που πραγματοποιείται με την προσέγγιση της κανονικής κατανομής και χρησιμοποιώντας ως τυπική απόκλιση την εκτιμωμένη απόκλιση του p κάτω από την μηδενική υπόθεση. Το $100(1-2\alpha)\%$ Wilson διάστημα εμπιστοσύνης για το p είναι το

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

$$[\underline{p}_{WS}, \bar{p}_{WS}] = \left[\frac{p + \frac{(z_{1-a})^2}{2n} - z_{1-a} \sqrt{\frac{p(1-p) + \frac{(z_{1-a})^2}{4n}}{n}}}{1 + \frac{(z_{1-a})^2}{n}}, \frac{p + \frac{(z_a)^2}{2n} - z_a \sqrt{\frac{p(1-p) + \frac{(z_a)^2}{4n}}{n}}}{1 + \frac{(z_a)^2}{n}} \right]$$

Το διάστημα αυτό είναι γνωστό και ως score διάστημα εμπιστοσύνης γιατί προκύπτει από τον έλεγχο του p μέσω του score στατιστικού.

2.3.4 Agresti-Coull Διάστημα Εμπιστοσύνης

Οι Agresti and Coull (1998) πρότειναν ως 95% διάστημα εμπιστοσύνης για το p , το

$$[\underline{p}_{AC}, \bar{p}_{AC}] = \left[p_W - z_{0.975} \sqrt{\frac{p_W(1-p_W)}{n+4}}, p_W - z_{0.025} \sqrt{\frac{p_W(1-p_W)}{n+4}} \right] \quad (2.6)$$

Αυτό το διάστημα εμπιστοσύνης είναι ουσιαστικά το Wald διάστημα εμπιστοσύνης, αν αντικαταστήσουμε στην σχέση (2.5) το p και το n , με p_W και $n+4$ αντίστοιχα.

Το κίνητρο τους είναι η παρουσίαση ενός προσεγγιστικού διαστήματος εμπιστοσύνης για ποσοστά κυρίως για διδακτικούς σκοπούς. Αν και οι Agresti και Coull προτείνουν την χρήση μόνο σε 95% ονομαστικό επίπεδο, οι Brown et al. (2001) δίνουν μια γενικότερη μορφή του $100(1-2a)\%$ Agresti-Coull διαστήματος εμπιστοσύνης:

$$[\underline{p}_{AC}, \bar{p}_{AC}] = \left[\hat{p}_0 - z_{1-a} \sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n}}, \hat{p}_0 - z_a \sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n}} \right]$$

όπου $\hat{p}_0 = \frac{2x + (z_{1-a})^2}{2(n + (z_{1-a})^2)}$. Ωστόσο επειδή οι Agresti και Coull χρησιμοποιούν αποκλειστικά

την μορφή (2.6) θα αναφερόμαστε στην μέθοδο των Agresti και Coull μόνο για κατασκευή διαστημάτων εμπιστοσύνης ονομαστικού επιπέδου 95%.

2.3.5 Άλλες Μέθοδοι Κατασκευής Διαστημάτων Εμπιστοσύνης

Οι μέθοδοι κατασκευής διαστημάτων εμπιστοσύνης που προαναφέραμε δεν είναι οι μοναδικές που υπάρχουν. Ο Newcombe (1998a) συγκρίνει επτά μεθόδους κατασκευής διαστημάτων εμπιστοσύνης για το p , λαμβάνοντας υπόψη του μεθόδους διόρθωσης συνέχειας και μεθόδους που βασίζονται σε ανισότητες πιθανοφάνειας. Οι Brown et al. (2001, 2002) μελετούν, εκτός από τις μεθόδους που προαναφέραμε, την συμπεριφορά διαστημάτων

εμπιστοσύνης που βασίζονται σε μετασχηματισμούς σταθεροποίησης της διασποράς και του κεντρικού διαστήματος εμπιστοσύνης του Jeffreys που βασίζεται σε μπεύζιανή στατιστική. Ο Pan (2001) πρότεινε ένα διάστημα εμπιστοσύνης για το p , αντικαθιστώντας τα ποσοστιαία σημεία της τυπικής κανονικής της μεθόδου των Agresti και Coull με ποσοστιαία σημεία της Student κατανομής. Το βασικό μειονέκτημα αυτής της μεθόδου είναι ότι τα διαστήματα εμπιστοσύνης είναι πιο πλατιά από τα αντίστοιχα της μεθόδου των Agresti και Coull.

Η εργασία των Brown et al. (2001) είναι ένα καλό παράδειγμα της διαφορετικής φιλοσοφίας της στατιστικής επιστήμης. Στην συζήτηση που ακολουθεί περιέχονται οι απόψεις διαφόρων στατιστικών και ο τρόπος με τον οποίο ορίζει ο καθένας το «βέλτιστο» διάστημα εμπιστοσύνης.

Μια μαθηματική εξήγηση για την ύπαρξη τόσων διαφορετικών προσεγγίσεων δόθηκε από τον Wang (2005). Περιληπτικά, ο Wang απέδειξε ότι για μονόπλευρα διαστήματα εμπιστοσύνης για ένα διωνυμικό ποσοστό, η μέθοδος των Clopper-Pearson παράγει τα μικρότερου πλάτους διαστήματα εμπιστοσύνης ενώ για δίπλευρα διαστήματα εμπιστοσύνης για ένα διωνυμικό ποσοστό δεν υπάρχει λύση όταν το a ή το n είναι πολύ μικρό. Ο ενδιαφερόμενος αναγνώστης μπορεί να βρει περισσότερες λεπτομέρειες στην εργασία του.

2.4 Σύγκριση Παραμετρικής και Μη Παραμετρικής Bootstrap

Έστω ότι εκτελούμε ένα διωνυμικό πείραμα n Bernoulli δοκιμών με σταθερή πιθανότητα επιτυχίας p και ότι έχουμε μετρήσει x επιτυχίες. Ας είναι $\{x_1, \mathbf{K}, x_n\}$ οι αποκρίσεις των n δοκιμών. Είναι προφανές ότι το σύνολο $\{x_1, \mathbf{K}, x_n\}$ που περιέχει μόνο μονάδες ή μηδενικά

είναι το αρχικό μας δείγμα και ότι ισχύει $x = \sum_{i=1}^n x_i$.

Η μη παραμετρική μέθοδος bootstrap στηρίζεται στην εμπειρική κατανομή F_n . Όπως είδαμε στην παρατήρηση (1.1) στην περίπτωση που έχουμε διακριτά δεδομένα η εμπειρική κατανομή δίνει βάρη σε κάθε διακριτή τιμή του δείγματος ίσα με το ποσοστό της συχνότητας τους στο δείγμα. Επομένως οι μονάδες θα έχουν βάρη x/n και τα μηδενικά $(n-x)/n$. Ας είναι $\{X_1^*, \mathbf{K}, X_n^*\}$ ένα bootstrap σύνολο δεδομένων από την F_n και $\mathfrak{R}(X_1^*, \mathbf{K}, X_n^*, F_n) = \sum_{i=1}^n X_i^* = X^*$ η τυχαία μεταβλητή που μετρά τον αριθμό των επιτυχιών σε

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

ένα bootstrap δείγμα. Είναι εύκολο να δειχτεί ότι η bootstrap δειγματική συνάρτηση κατανομής είναι

$$\begin{aligned} H_{BOOT} &= \text{Prob}_* \left\{ \mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_n^*, F_n \right) \leq k \right\} \\ &= \sum_{r=0}^k \text{Prob}_* \left\{ \mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_n^*, F_n \right) = r \right\} \\ &= \sum_{r=0}^k \binom{n}{r} \left(\frac{x}{n} \right)^r \left(1 - \frac{x}{n} \right)^{n-r} \\ &= \sum_{r=0}^k \binom{n}{r} p^r (1-p)^{n-r} \end{aligned} \quad (2.7)$$

Ας θεωρήσουμε τώρα ότι θέλουμε να εκτελέσουμε την παραμετρική bootstrap. Ας είναι \hat{p} ένας εκτιμητής του p , $\hat{F} \equiv B(n, \hat{p})$ η εκτίμηση της διωνυμικής κατανομής $B(n, p)$, και $\mathfrak{R}(X_1^*, \mathbf{K}, X_n^*, \hat{F}) = \sum_{i=1}^n X_i^* = X^*$. Τότε η εκτίμηση bootstrap της δειγματικής συνάρτησης κατανομής είναι

$$\begin{aligned} H_{BOOT} &= \text{Prob}_* \left\{ \mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_n^*, \hat{F} \right) \leq k \right\} \\ &= \sum_{r=0}^k \text{Prob}_* \left\{ \mathfrak{R}_n \left(X_1^*, \mathbf{K}, X_n^*, \hat{F} \right) = r \right\} \\ &= \sum_{r=0}^k \binom{n}{r} \hat{p}^r (1-\hat{p})^{n-r} \end{aligned} \quad (2.8)$$

Από τις σχέσεις (2.7) και (2.8) μπορούμε να βγάλουμε τα εξής συμπεράσματα

- i. Στην μη παραμετρική bootstrap η τυχαία μεταβλητή $X^* \sim B(n, p)$
- ii. Στην παραμετρική bootstrap η τυχαία μεταβλητή $X^* \sim B(n, \hat{p})$. Αυτό σημαίνει ότι η κατανομή της X^* εξαρτάται αποκλειστικά από τον εκτιμητή της πιθανότητας επιτυχίας p .
- iii. Όταν $\hat{p} = p$, τότε η παραμετρική και η μη παραμετρική bootstrap ταυτίζονται.
- iv. Οι εκτιμητές bootstrap της δειγματικής συνάρτησης κατανομής έχουν αναλυτική μορφή και επομένως δεν χρειάζονται υπολογιστικές τεχνικές για τις εκτιμήσεις τους.

2^ο Κεφάλαιο

Ας εξετάσουμε την περίπτωση $\hat{p} = p$. Όπως είδαμε παραπάνω ανεξάρτητα από την επιλογή της μεθόδου bootstrap, η τ.μ. $X^* \sim B(n, p)$. Επομένως η αναμενομένη τιμή και η διασπορά της X^* είναι

$$E[X^*] = np \text{ και } Var[X^*] = np(1-p)$$

αντίστοιχα.

Έστω p^* η αναλογία bootstrap του p . Τότε $p^* = X^*/n$ και

$$E[p^*] = p \text{ και } Var[p^*] = \frac{p(1-p)}{n}$$

2.5 Bootstrap Διαστήματα Εμπιστοσύνης για το Διωνυμικό Ποσοστό

Έστω $X \sim B(n, p)$ και ότι το p εκτιμάται από το \hat{p} . Σύμφωνα με όσα αναφέραμε στην προηγούμενη παράγραφο θα ισχύει $X^* \sim B(n, p)$. Ο σκοπός μας είναι να περιγράψουμε την κατασκευή $100(1-2a)\%$ bootstrap διαστημάτων εμπιστοσύνης για το p . Η παρουσίαση θα στηριχθεί στα αποτελέσματα του Åstrand (2000).

2.5.1 Percentile

Σύμφωνα με τον Åstrand, ένα $100(1-2a)\%$ bootstrap percentile διάστημα εμπιστοσύνης για το p είναι το:

$$[\underline{p}_{BP}, \bar{p}_{BP}] = \left[\frac{x_{n,p;a}}{n}, \frac{x_{n,p;1-a}}{n} \right]$$

Είναι προφανές ότι το bootstrap percentile διάστημα εμπιστοσύνης ισοδυναμεί με το διάστημα εμπιστοσύνης που προκύπτει παίρνοντας τα ποσοστιαία σημεία $a\%$ και $(1-a)\%$ της διωνυμικής κατανομής $B(n, p)$.

2.5.2 BC

Ένα $100(1-2a)\%$ bootstrap bias-corrected percentile διάστημα εμπιστοσύνης για το p είναι το:

$$[\underline{p}_{BC}, \bar{p}_{BC}] = \left[\frac{x_{n,p; \hat{a}_i}}{n}, \frac{x_{n,p; \hat{a}_i^-}}{n} \right]$$

όπου

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

$$a_1' = \Phi(z_a + 2\hat{z}_0)$$

$$a_1'' = \Phi(z_{1-a} + 2\hat{z}_0)$$

και

$$\hat{z}_0 = \text{Prob}\{X < x\} + 0.5 \text{Prob}\{X = x\} \quad (2.9)$$

Τα ποσοστιαία σημεία a_1' και a_1'' προκύπτουν από την υπόθεση (1.21) για $\Psi = \Phi$. Η εκτίμηση του z_0 δεν προκύπτει από την σχέση (1.23) αλλά από μια διόρθωση της σχέσης αυτής. Η διόρθωση γίνεται για να ληφθεί υπόψη η διακριτικότητα της διωνυμικής κατανομής και είναι παρόμοια με την ιδέα του mid p-value (βλ. Agresti, 2002, σελ. 20).

2.5.3 BC_a

Ένα $100(1-2a)\%$ bootstrap bias-corrected percentile διάστημα εμπιστοσύνης για το p είναι το:

$$\left[\underline{p}_{BC_a}(\hat{a}), \bar{p}_{BC_a}(\hat{a}) \right] = \left[\frac{x_{n,p;a_2'}}{n}, \frac{x_{n,p;a_2''}}{n} \right]$$

όπου

$$a_2' = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_a}{1 - \hat{a}(\hat{z}_0 + z_a)}\right)$$

$$a_2'' = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-a}}{1 - \hat{a}(\hat{z}_0 + z_{1-a})}\right)$$

$$\hat{a} = \frac{1-2p}{6\sqrt{np(1-p)}} \quad (2.10)$$

και το \hat{z}_0 δίνεται από την σχέση (2.9). Η εκτίμηση του a προκύπτει όταν εκτιμούμε το a μέσω της μη παραμετρικής μεθόδου bootstrap.

Ας εξετάσουμε τώρα την εκτίμηση του a μέσω της παραμετρικής μεθόδου bootstrap. Το a εκτιμάται από το

$$\hat{a} = \left(\frac{1}{6} E \left[\frac{\partial^3 l(p)}{\partial p^3} \right] \right) / \left(-E \left[\frac{\partial^2 l(p)}{\partial p^2} \right] \right)^{3/2} \quad (2.11)$$

όπου $l(p)$ είναι ο λογάριθμος της πιθανοφάνειας της $B(n, p)$. Ο λογάριθμος της πιθανοφάνειας είναι ο

$$l(p) = \prod_i \log \binom{n}{x_i} + (n-x) \log(1-p) + x \log(p)$$

και οι πρώτες τρεις μερικές παράγωγοι είναι οι εξής:

$$\frac{\partial l(p)}{\partial p} = -\frac{n}{1-p} + \frac{x}{p(1-p)}$$

$$\frac{\partial^2 l(p)}{\partial p^2} = -\frac{np^2 + x - 2p^3x}{p^2(1-p)^2} \quad (2.12)$$

$$\frac{\partial^3 l(p)}{\partial p^3} = -\frac{2np^3 - 2x + 6px - 6p^2x}{p^3(1-p)^3} \quad (2.13)$$

Από τις σχέσεις (2.12) και (2.13) παίρνουμε ότι

$$E \left[\frac{\partial^3 l(p)}{\partial p^3} \right] = \frac{2n(1-2p)}{(1-p)^2 p^2} \text{ και } \left(-E \left[\frac{\partial^2 l(p)}{\partial p^2} \right] \right)^{3/2} = \left(\frac{n}{(1-p)p} \right)^{3/2} \quad (2.14)$$

Τέλος αντικαθιστώντας τις σχέσεις (2.14) στην σχέση (2.11) παίρνουμε

$$\hat{a} = \frac{1-2p}{3n \sqrt{\frac{p(1-p)}{n}}} \quad (2.15)$$

Επομένως για την εκτίμηση της παραμέτρου a μπορώ να χρησιμοποιήσω τις σχέσεις (2.10) και (2.15).

2.5.4 Bootstrap-t

Η εκτίμηση bootstrap της G , G_{BOOT} , δίνεται από τον τύπο:

$$G_{BOOT}(x) = \text{Prob}_* \{Z^* \leq x\}$$

όπου

$$Z^* = \frac{(p^* - p)}{\sqrt{p^*(1-p^*)/n}} \quad (2.16)$$

Άρα το $100(1-2a)\%$ bootstrap-t διάστημα εμπιστοσύνης για το p είναι το:

$$[p_{BT}, \bar{p}_{BT}] = \left[\max \left\{ 0, p - \sqrt{n} \frac{p}{\sqrt{p(1-p)}} G_{BOOT}^{-1}(1-a) \right\}, \min \left\{ 1, p - \frac{p}{\sqrt{p(1-p)/n}} G_{BOOT}^{-1}(a) \right\} \right]$$

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

Κάνοντας αντικατάσταση του p^* και του p με x^* και με x/n αντίστοιχα στην (2.16) έχουμε:

$$Z^* = \frac{(p^* - p)}{\sqrt{p^*(1-p^*)/n}} = \frac{(x^*/n - x/n)}{\sqrt{x^*/n(1-x^*/n)/n}} = \frac{(x^* - x)}{\sqrt{\frac{x^*(n-x^*)}{n^3}}}$$

και τελικά ότι:

$$Z^* = \sqrt{n} \frac{(x^* - x)}{\sqrt{x^*(n-x^*)}}$$

Προφανώς η Z^* δεν μπορεί να οριστεί όταν $x^*=0$ ή $x^*=n$ και επομένως το $100(1-2a)\%$ bootstrap-t διάστημα εμπιστοσύνης για το p δεν μπορεί να οριστεί.

Έστω ότι $X_{(i)}$ όπως την ορίσαμε στην Παράγραφο 2.1, για $i=1,2,\dots,n-1$. Μπορώ να θέσω

$$Z_{(i)} = \sqrt{n} \frac{(i-x)}{\sqrt{i(n-i)}}$$

για $i=1,2,\dots,n-1$. Η $Z_{(i)}$ είναι αύξουσα συνάρτηση του i γιατί για $i < k$ θα ισχύει $Z_{(i)} < Z_{(k)}$. Προφανώς η G_{BOOT} μπορεί να οριστεί και ουσιαστικά θα έχει την ίδια συμπεριφορά με την K_{BOOT} . Αυτό σημαίνει ότι αν $x_{n,p;a} = X_{(f)}$ και $x_{n,p;1-a} = X_{(g)}$, με $0 < f < g < n$, είναι τα a και $1-a$ ποσοστιαία σημεία της $B(n, p)$, τότε το $100(1-2a)\%$ bootstrap-t διάστημα εμπιστοσύνης για το p είναι το :

$$[\underline{p}_{BT}, \bar{p}_{BT}] = \left[\max \left\{ 0, p - n \frac{p}{\sqrt{p(1-p)}} \times \frac{(g-x)}{\sqrt{g(n-g)}} \right\}, \min \left\{ 1, p - n \frac{p}{\sqrt{p(1-\hat{p})}} \times \frac{(f-x)}{\sqrt{f(n-f)}} \right\} \right]$$

Ωστόσο αυτή είναι μια πολύ ευνοϊκή περίπτωση. Όταν το n είναι πολύ μικρό και το p είναι κοντά στο μηδέν ή κοντά στην μονάδα τότε υπάρχει μεγάλη πιθανότητα τα ποσοστιαία σημεία της $B(n, p)$ να είναι ίσα με μηδέν ή ίσα με n . Συνεπώς τα $Z_{(0)}$ ή/και $Z_{(n)}$ δεν μπορούν να οριστούν.

Όπως αποδείξαμε προηγουμένως η $Z_{(i)}$ είναι αύξουσα και μονότονη ως προς το i . Έτσι αν πάρουμε το όριο των άκρων του $100(1-2a)\%$ bootstrap-t διαστήματος εμπιστοσύνης για το p όταν το i ισούται τείνει στην μονάδα ή στο n μηδέν έχουμε:

$$\lim_{i \rightarrow n} \left[p - n \frac{p}{\sqrt{p(1-p)}} \times \frac{(i-x)}{\sqrt{i(n-i)}} \right] = p - n \frac{p}{\sqrt{p(1-p)}} \lim_{i \rightarrow n} \frac{(i-x)}{\sqrt{i(n-i)}} = -\infty$$

και

$$\lim_{i \rightarrow 0} \left[p - n \frac{p}{\sqrt{p(1-p)}} \times \frac{(i-x)}{\sqrt{i(n-i)}} \right] = p - n \frac{p}{\sqrt{p(1-p)}} \lim_{i \rightarrow 0} \frac{(i-x)}{\sqrt{i(n-i)}} = \infty$$

Αφού έτσι και αλλιώς δεν μας ενδιαφέρουν τιμές που δεν ανήκουν στο $[0,1]$ μπορούμε να θεωρήσουμε ότι στην περίπτωση όπου $x_{n,p;1-a} = n$ ($x_{n,p;a} = 0$) το κάτω (άνω) άκρο είναι ίσο με μηδέν (μονάδα). Η ιδέα για αυτήν την παραδοχή είναι να μπορούμε να χρησιμοποιούμε το $100(1-2a)\%$ bootstrap-t διαστήματος εμπιστοσύνης για το p σε όλες τις περιπτώσεις. Για παράδειγμα, ας πάρουμε $p=0.7$, $n=10$ και θέλουμε να κατασκευάσουμε ένα 95% bootstrap-t διάστημα εμπιστοσύνης για το p . Τα $x_{n,p;a}$ και $x_{n,p;1-a}$ ποσοστιαία σημεία της βρίσκονται από την σχέση :

$$\min \left\{ X_{(i)} = i \mid \sum_{k=0}^i \binom{10}{k} 0.7^k (1-0.7)^{10-k} \geq 0.025 \right\} = 4$$

και την σχέση

$$\min \left\{ X_{(i)} = i \mid \sum_{k=0}^i \binom{10}{k} 0.7^k (1-0.7)^{10-k} \geq 0.975 \right\} = 10$$

Επομένως το 95% bootstrap-t διάστημα εμπιστοσύνης για το p είναι το $[\underline{p}_{BT}, \bar{p}_{BT}] = [0, 0.981]$.

2.6 Κριτήρια Σύγκρισης των Διαστημάτων Εμπιστοσύνης

Η εκτίμηση και ο χαρακτηρισμός ενός διαστήματος εμπιστοσύνης ως «καλό» ή όχι είναι μια πολύ δύσκολη υπόθεση. Όπως προαναφέραμε στην συζήτηση που ακολούθησε την εργασία των Brown et al. (2001) παρουσιάστηκαν διάφορες απόψεις για την καταλληλότητα και την αποτελεσματικότητα ενός διαστήματος εμπιστοσύνης. Ο Newcombe (1998a) θεωρεί ότι τα κριτήρια θα πρέπει να καθοριστούν εκ των προτέρων και κατόπιν να γίνει η σύγκριση των διαστημάτων εμπιστοσύνης. Αυτήν την τακτική θα ακολουθήσουμε και στην παρούσα εργασία. Ως κριτήρια θα χρησιμοποιήσουμε την πιθανότητα κάλυψης (*Coverage Probability*), το αναμενόμενο πλάτος (*Average Width*) και το μέσο τετραγωνικό σφάλμα (*Mean Square Error*) της πιθανότητας κάλυψης των διαστημάτων εμπιστοσύνης.

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

Έστω ότι μια μέθοδος κατασκευής διαστημάτων εμπιστοσύνης παράγει διαστήματα εμπιστοσύνης για την πιθανότητα επιτυχίας p της μορφής

$$[\underline{p}, \bar{p}] = [\underline{p}(x), \bar{p}(x)] \quad \text{για } x = 0, \dots, n$$

Το σημαντικότερο κριτήριο είναι η πιθανότητα κάλυψης του διαστήματος εμπιστοσύνης. Για κάθε διαδικασία κατασκευής διαστημάτων εμπιστοσύνης για το p επιθυμούμε η πιθανότητα κάλυψης να είναι ίση με το προκαθορισμένο ονομαστικό επίπεδο α . Η πιθανότητα κάλυψης του διαστήματος εμπιστοσύνης για σταθερό μέγεθος δείγματος n και για σταθερή πιθανότητα p δίνεται από την ακόλουθη σχέση:

$$CP(n, p) = \sum_{x=0}^n I(x, p) \binom{n}{x} p^x (1-p)^{n-x} \quad (2.17)$$

όπου η συνάρτηση $I(x, p)$ είναι μια δείκτρια συνάρτηση που ισούται με την μονάδα, όταν για την τιμή x το διάστημα εμπιστοσύνης περιέχει το p , και ισούται με το μηδέν διαφορετικά.

Είναι προφανές ότι για τον υπολογισμό της πιθανότητας κάλυψης μέσω της σχέσης (2.17) πρέπει το p να είναι γνωστό. Πρακτικά, αυτό που μας ενδιαφέρει είναι να γνωρίζουμε την πιθανότητα κάλυψης για όλες τις πιθανές τιμές που μπορεί να πάρει το p . Μια εναλλακτική περιγραφή της πιθανότητας κάλυψης παρέχεται παίρνοντας την μέση τιμή όλων των πιθανών τιμών που μπορεί να πάρει το p μέσω του τύπου

$$\overline{CP}_n = \int_0^1 CP(n, p) \cdot g(p) dp \quad (2.18)$$

για κάποια κατανομή $g(p)$. Η κατανομή $g(p)$ είναι η κατανομή από την οποία προέρχεται η πιθανότητα επιτυχίας. Στην παρούσα εργασία θα ασχοληθούμε μόνο με την περίπτωση όπου η $g(p) \equiv U(0,1)$, όπου U είναι η ομοιόμορφη κατανομή. Η επιλογή της ομοιόμορφης κατανομής είναι λογική γιατί με αυτόν τον τρόπο θεωρούμε ότι η πιθανότητα επιτυχίας p μπορεί να είναι οποιοδήποτε σημείο στο διάστημα $(0,1)$, που είναι και το διάστημα ορισμού της διωνυμικής πιθανότητας επιτυχίας. Αντικαθιστώντας την $g(p)$ με την $U(0,1)$ στην σχέση (2.18) έχουμε:

$$\overline{CP}_n = \int_0^1 CP(n, p) \cdot 1 dp = \int_0^1 \sum_{x=0}^n I(x, p) \binom{n}{x} p^x (1-p)^{n-x} dp \quad (2.20)$$

2^ο Κεφάλαιο

Για τον υπολογισμό του ολοκληρώματος στην σχέση (2.20) αρκεί να παρατηρήσουμε ότι για σταθερή τιμή του x η δείκτρια συνάρτηση $I(x, p)$ ισούται με την μονάδα όταν η πιθανότητα επιτυχίας ανήκει στο διάστημα $[\underline{p}(x), \bar{p}(x)]$. Επομένως η αναμενομένη πιθανότητα κάλυψης δίνεται από τον τύπο

$$\overline{CP}_n = \sum_{x=0}^n \binom{n}{x} \int_{\underline{p}(x)}^{\bar{p}(x)} p^x (1-p)^{n-x} dp \quad (2.21)$$

Μια μέθοδος κατασκευής διαστημάτων εμπιστοσύνης είναι αυστηρά συντηρητική αν για όλα τα n η πιθανότητα κάλυψης $CP(n, p) \geq 1-a$. Επειδή όμως για τον υπολογισμό της πιθανότητας κάλυψης εμείς θα χρησιμοποιήσουμε τον τύπο (2.21) θα λέμε ότι μια μέθοδος κατασκευής διαστημάτων εμπιστοσύνης είναι αυστηρά συντηρητική αν για όλα τα n ισχύει $\overline{CP}_n \geq 1-a$. Στην πραγματικότητα θέλουμε η μέθοδος κατασκευής διαστημάτων εμπιστοσύνης να είναι συντηρητική αλλά η τιμή του \overline{CP}_n να απέχει όσο το δυνατόν λιγότερο από το $1-a$.

Ένα άλλο κριτήριο είναι το πλάτος του διαστήματος εμπιστοσύνης. Αυτό που επιθυμούμε είναι το πλάτος του διαστήματος εμπιστοσύνης να είναι όσο το δυνατόν πιο μικρό. Για σταθερό μέγεθος δείγματος n και για σταθερή πιθανότητα p το αναμενόμενο πλάτος ενός διαστήματος εμπιστοσύνης είναι

$$AW(n, p) = \sum_{x=0}^n W(x) \binom{n}{x} p^x (1-p)^{n-x}$$

όπου $W(x)$ είναι το πλάτος του διαστήματος εμπιστοσύνης για την τιμή x .

Όπως και για την πιθανότητα κάλυψης μπορούμε να πάρουμε το μέσο αναμενόμενο πλάτος μέσω του υπολογισμού της μέσης τιμής όλων των πιθανών τιμών που μπορεί να πάρει το p :

$$\overline{AW}_n = \int_0^1 AW(n, p) dp = \int_0^1 \sum_{x=0}^n W(x) \binom{n}{x} p^x (1-p)^{n-x} dp = \sum_{x=0}^n W(x) \binom{n}{x} \int_0^1 p^x (1-p)^{n-x} dp \quad (2.22)$$

Λαμβάνοντας υπόψη ότι

$$W(x) = \bar{p}(x) - \underline{p}(x)$$

και ότι

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

$$\int_0^1 p^x (1-p)^{n-x} dp = \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n-x+2)} = \frac{x!(n-x)!}{(n-x+1)!} = \frac{x!}{n-x+1}$$

το ολοκλήρωμα της σχέσης (2.22), και επομένως και το μέσο αναμενόμενο πλάτος, υπολογίζεται από την σχέση:

$$\overline{AW}_n = \sum_{x=0}^n \binom{n}{x} (\bar{p}(x) - \underline{p}(x)) \frac{x!}{n-x+1} = \frac{\bar{p}(x) - \underline{p}(x)}{n+1} \quad (2.23)$$

Ως μέτρο απόκλισης των πραγματικών πιθανοτήτων κάλυψης από το ονομαστικό επίπεδο χρησιμοποιηθούμε το μέσο τετραγωνικό σφάλμα της πιθανότητας κάλυψης για τα διαστήματα εμπιστοσύνης:

$$\overline{MSRCP}_n = \sqrt{\int_0^1 \left(\sum_{x=0}^n I(x,p) \binom{n}{x} p^x (1-p)^{n-x} - (1-a) \right)^2 dp}$$

Το ολοκλήρωμα αυτό είναι σχετικά δύσκολο να υπολογιστεί. Με βάση την ταυτότητα $(b-c)^2 = b^2 - 2bc + c^2$ διασπάμε το ολοκλήρωμα σε απλούστερα ολοκληρώματα:

$$\overline{MSRCP}_n = \sqrt{\int_0^1 (1-a)^2 dp - 2 \int_0^1 (1-a) \sum_{x=0}^n I(x,p) \binom{n}{x} p^x (1-p)^{n-x} dp + \int_0^1 \left(\sum_{x=0}^n I(x,p) \binom{n}{x} p^x (1-p)^{n-x} \right)^2 dp} \quad (2.24)$$

Ο υπολογισμός των δύο πρώτων όρων είναι άμεσος αφού

$$\int_0^1 (1-a)^2 dp = (1-a)^2 \quad (2.25)$$

και

$$2 \int_0^1 (1-a) \sum_{x=0}^n I(x,p) \binom{n}{x} p^x (1-p)^{n-x} dp = 2(1-a) \overline{CP}_n \quad (2.26)$$

Με βάση την ταυτότητα $\left(\sum_{i=1}^k b_i \right)^2 = \sum_{i=1}^k b_i^2 + 2 \sum_{i=1}^k \sum_{j=i+1}^k b_i b_j$ η ποσότητα μέσα στο ολοκλήρωμα

του τρίτου όρου της σχέσης (2.24) ισούται με

$$\sum_{x=0}^n I(x,p) \left(\binom{n}{x} p^x (1-p)^{n-x} \right)^2 + 2 \sum_{x=0}^n \sum_{y=x+1}^n I(x,p) \binom{n}{x} p^x (1-p)^{n-x} I(y,p) \binom{n}{y} p^y (1-p)^{n-y} \quad (2.27)$$

Ολοκληρώνοντας στο (0,1) τον πρώτο όρο της σχέσης (2.27) παίρνουμε ότι

$$A_1 = \int_0^1 \sum_{x=0}^n I(x,p) \left(\binom{n}{x} p^x (1-p)^{n-x} \right)^2 = \sum_{x=0}^n \binom{n}{x}^2 \int_{\underline{p}(x)}^{\bar{p}(x)} p^{2x} (1-p)^{2(n-x)} dp \quad (2.28)$$

Για τον δεύτερο όρο της σχέσης (2.27) αξίζει να παρατηρήσουμε ότι οι δείκτριες συναρτήσεις $I(x,p)$ και $I(y,p)$ είναι συγχρόνως ίσες με την μονάδα αν και μόνο αν το $p \in D = [\underline{p}(x), \bar{p}(x)] \cap [\underline{p}(y), \bar{p}(y)]$. Επομένως το ολοκλήρωμα του δευτέρου όρου της σχέσης (2.27) στο (0,1) είναι ίσο με

$$A_2 = \sum_{x=0}^n \sum_{y=x+1}^n \int_D \binom{n}{x} \binom{n}{y} p^{x+y} (1-p)^{n-x-y} dp \quad (2.29)$$

Έτσι με αντικατάσταση των (2.25), (2.26), (2.28) και (2.29) στην (2.25) βρίσκουμε ότι

$$\overline{MSRCP}_n = \sqrt{(1-a)^2 - 2(1-a)\overline{CP}_n + A_1 + 2A_2} \quad (2.30)$$

2.7 Εκτίμηση των Μεθόδων Κατασκευής Διαστημάτων Εμπιστοσύνης

Πριν προχωρήσουμε στην ανάλυση των αποτελεσμάτων καλό είναι να διευκρινίσουμε για ποιες τιμές του n ενδιαφερόμαστε. Δείγματα με τάξη μεγέθους μικρότερα από τριάντα θεωρούνται δείγματα μικρού μεγέθους, δείγματα που αποτελούνται από τριάντα έως πενήντα παρατηρήσεις θεωρούνται μεσαίου μεγέθους και δείγματα άνω των πενήντα παρατηρήσεων μεγάλα δείγματα. Θεωρητικά, το πλεονέκτημα της μεθόδου bootstrap είναι η δυνατότητα εφαρμογής της ακόμα και όταν το μέγεθος του δείγματος είναι αρκετά μικρό. Είναι λογικό να εστιάσουμε την προσοχή μας στα δείγματα μικρού και μεσαίου μεγέθους. Θα συγκριθούν bootstrap διαστήματα εμπιστοσύνης με βήμα 5 όταν $5 \leq n \leq 50$, με βήμα 10 όταν $60 \leq n \leq 100$ και με βήμα 100 όταν $100 \leq n \leq 500$. Για λόγους απλότητας τα αποτελέσματα για κάθε κριτήριο θα παρουσιάζονται σε δύο πίνακες, ο πρώτος θα αποτελείται από δείγματα μικρού και μεσαίου μεγέθους ενώ ο δεύτερος θα αποτελείται από δείγματα μεγάλου μεγέθους.

Πίνακας 2.1

Μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών bootstrap διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $n \leq 50$ και $p \sim U(0,1)$.

$1-2a$	n	5	10	15	20	25	30	35	40	45	50
0.90	BT	0.6667	0.7963	0.8465	0.8669	0.8759	0.8702	0.8870	0.8839	0.8858	0.8943
	BP	0.6474	0.7368	0.7869	0.8116	0.8243	0.8243	0.8448	0.8450	0.8476	0.8589
	BC	0.6474	0.7521	0.7994	0.8230	0.8346	0.8405	0.8575	0.8580	0.8590	0.8657
	$BC_{\alpha,j}$	0.6474	0.7828	0.8049	0.8433	0.8441	0.8580	0.8676	0.8671	0.8728	0.8777
	$BC_{\alpha,p}$	0.6474	0.7521	0.7869	0.8116	0.8243	0.8243	0.8515	0.8503	0.8525	0.8633
0.95	BT	0.6667	0.8178	0.8716	0.8906	0.9119	0.9162	0.9268	0.9287	0.9314	0.9354
	BP	0.6474	0.7798	0.8301	0.8452	0.8690	0.8790	0.8919	0.8953	0.8992	0.9058
	BC	0.6474	0.7764	0.8417	0.8595	0.8811	0.8869	0.8984	0.9012	0.9074	0.9128
	$BC_{\alpha,j}$	0.6605	0.7959	0.8513	0.8731	0.8916	0.8985	0.9105	0.9113	0.9191	0.9211
	$BC_{\alpha,p}$	0.6474	0.7764	0.8301	0.8452	0.8720	0.8790	0.8919	0.8953	0.9005	0.9074

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

Ο Πίνακας 2.1 περιέχει την μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών bootstrap διαστημάτων εμπιστοσύνης για το p , σε δείγματα μικρού και μεσαίου μεγέθους.

Τα αποτελέσματα είναι απογοητευτικά. Η μέση πιθανότητα κάλυψης των bootstrap διαστημάτων είναι μικρότερη από το ονομαστικό επίπεδο και μάλιστα αρκετά. Ενώ θα περιμέναμε τα αποτελέσματα αυτά να βελτιωθούν όσο το μέγεθος του δείγματος αυξάνεται, ωστόσο ακόμα και όταν το μέγεθος του δείγματος είναι μεγάλο λαμβάνουμε παρόμοια αποτελέσματα όπως φαίνεται στον Πίνακα 2.2.

Πίνακας 2.2

Μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών bootstrap διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $60 \leq n \leq 100$ και $p \sim U(0,1)$.

$1-2\alpha$	n	60	70	80	90	100
0.90	BT	0.8917	0.8982	0.8947	0.8991	0.8970
	BP	0.8601	0.8724	0.8704	0.8769	0.8764
	BC	0.8660	0.8754	0.8735	0.8807	0.8793
	BC _{α,i}	0.8749	0.8839	0.8804	0.8870	0.8856
	BC _{α,p}	0.8637	0.8724	0.8711	0.8777	0.8769
0.95	BT	0.9380	0.9404	0.9432	0.9433	0.9455
	BP	0.9115	0.9165	0.9215	0.9225	0.9259
	BC	0.9171	0.9226	0.9264	0.9264	0.9324
	BC _{α,i}	0.9272	0.9291	0.9334	0.9333	0.9373
	BC _{α,p}	0.9122	0.9162	0.9212	0.9222	0.9259

Παρατηρούμε ότι ακόμα και όταν $n=100$, η μέση πιθανότητα κάλυψης των bootstrap διαστημάτων εμπιστοσύνης είναι μικρότερη από το ονομαστικό επίπεδο. Το γεγονός αυτό δεν πρέπει να μας εκπλήσσει. Όταν το αρχικό μας δείγμα αποτελείται μόνο από αποτυχίες ή μόνο από επιτυχίες τότε τα bootstrap διαστήματα εμπιστοσύνης παράγουν τα μονοσύνολα $\{0\}$ και $\{1\}$, αντίστοιχα. Από τον τύπο (2.21) βλέπουμε ότι η μέση πιθανότητα κάλυψης όταν η διαδικασία παραγωγής διαστημάτων εμπιστοσύνης παράγει μονοσύνολα στα σημεία $x=0$ και $x=n$, δεν είναι άθροισμα $n+1$ όρων αλλά άθροισμα $n-1$ όρων αφού το ολοκλήρωμα $\int_{p(x)}^{\bar{p}(x)} p^x (1-p)^{n-x} dp$ στις περιπτώσεις αυτές μηδενίζεται. Για να γίνει πιο κατανοητό το συμπέρασμα αυτό στους επόμενους πίνακες παρουσιάσουμε την μέση πιθανότητα κάλυψης για τις παραδοσιακές μεθόδους κατασκευής διαστημάτων εμπιστοσύνης της Ενότητας 2.3.

Πίνακας 2.3

Μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $n \leq 50$ και $p \sim U(0,1)$.

$1-2\alpha$	n	5	10	15	20	25	30	35	40	45	50
0.90	Wald	0.6172	0.7352	0.7806	0.8053	0.8211	0.8320	0.8401	0.8463	0.8513	0.8554
	Exact	0.9765	0.9629	0.9551	0.9498	0.9459	0.9429	0.9404	0.9384	0.9366	0.9351
	Wilson	0.9151	0.9106	0.9083	0.9070	0.9060	0.9053	0.9048	0.9044	0.9040	0.9037
0.95	Wald	0.6406	0.7693	0.8188	0.8458	0.8629	0.8749	0.8838	0.8906	0.8961	0.9006
	Exact	0.9904	0.9838	0.9798	0.9770	0.9750	0.9734	0.9721	0.9710	0.9701	0.9693
	Wilson	0.9552	0.9541	0.9534	0.9530	0.9527	0.9524	0.9522	0.9520	0.9519	0.9518
	AC	0.9649	0.9638	0.9624	0.9613	0.9605	0.9598	0.9592	0.9587	0.9582	0.9578

Πίνακας 2.4

Μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $60 \leq n \leq 500$ και $p \sim U(0,1)$.

$1-2\alpha$	n	60	70	80	90	100	200	300	400	500
0.90	Wald	0.8617	0.8664	0.8699	0.8728	0.8751	0.8862	0.8903	0.8925	0.8938
	Exact	0.9326	0.9306	0.9290	0.9276	0.9264	0.9196	0.9163	0.9143	0.9130
	Wilson	0.9033	0.9029	0.9026	0.9024	0.9022	0.9013	0.9009	0.9007	0.9006
0.95	Wald	0.9075	0.9126	0.9166	0.9197	0.9223	0.9346	0.9391	0.9415	0.9430
	Exact	0.9679	0.9669	0.9660	0.9652	0.9646	0.9609	0.9591	0.9580	0.9572
	Wilson	0.9516	0.9514	0.9513	0.9512	0.9511	0.9507	0.9505	0.9504	0.9503
	AC	0.9571	0.9566	0.9561	0.9557	0.9554	0.9535	0.9526	0.9522	0.9518

Στους πίνακες αυτούς, με έντονα μαύρα γράμματα τονίζεται η τιμή της μέσης τιμής της πιθανότητας κάλυψης που είναι πιο κοντά στο επιθυμητό ονομαστικό επίπεδο για κάθε n . Το πρώτο συμπέρασμα είναι ότι η μέθοδος Wilson μας δίνει τις καλύτερες τιμές. Για κάθε μέγεθος δείγματος που εξετάσαμε η μέση πιθανότητα κάλυψης είναι πολύ κοντά στο ονομαστικό επίπεδο. Η μέθοδος Exact είναι πολύ συντηρητική αφού οι τιμές της μέσης τιμής της πιθανότητας κάλυψης είναι αρκετά μεγαλύτερη από το ονομαστικό επίπεδο. Οι τιμές της μεθόδου των Agresti και Coull (AC) είναι πάντα λίγο μεγαλύτερες από τις αντίστοιχες της μεθόδου Wilson, αν και όσο το μέγεθος του δείγματος μεγαλώνει οι διαφορές αυτές μικραίνουν. Ωστόσο αυτό που έχει ενδιαφέρον είναι ότι τα Wald διαστήματα εμπιστοσύνης έχουν πολύ άσχημη συμπεριφορά, το οποίο έχει ήδη αναφερθεί στην Ενότητα 2.3.1. Η Wald μέθοδος κατασκευής διαστημάτων εμπιστοσύνης έχει το ίδιο μειονέκτημα με τα bootstrap διαστήματα εμπιστοσύνης: όταν ο εκτιμητής μέγιστης πιθανοφάνειας είναι ίσος με την μονάδα ή το μηδέν τότε είναι καταδικασμένη να παράγει τα μονοσύνολα $\{1\}$ ή $\{0\}$ αντίστοιχα, ως διαστήματα εμπιστοσύνης ενώ οι άλλες μέθοδοι κατασκευής διαστημάτων εμπιστοσύνης παράγουν διαστήματα σε αυτές τις περιπτώσεις.

Για να ξεπεραστεί το πρόβλημα αυτό μπορούμε να κάνουμε μια παραδοχή: στο αρχικό μας δείγμα μεγέθους n προσθέτουμε δύο παρατηρήσεις, μια επιτυχία και μια αποτυχία, και στην συνέχεια εφαρμόζουμε την μέθοδο bootstrap στο καινούριο μας δείγμα μεγέθους $n+2$,

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

χρησιμοποιώντας ως εκτιμητή της διωνυμικής πιθανότητας τον εκτιμητή μέγιστης πιθανοφάνειας του καινούριου δείγματος, δηλαδή $p = \frac{x+1}{n+2}$, όπου x είναι ο αριθμός των επιτυχιών στις n δοκιμές. Ο εκτιμητής $p = \frac{x+1}{n+2}$ είναι ουσιαστικά ο εκτιμητής Laplace που δίνεται από την σχέση (2.3) για την πιθανότητα επιτυχίας. Όταν εφαρμόζουμε την μέθοδο bootstrap κάτω από αυτήν την προσθήκη τότε θα την ονομάζουμε τροποποιημένη bootstrap.

Η λογική της τροποποιημένης bootstrap είναι η εξής: Εκ των προτέρων γνωρίζουμε ότι τα δεδομένα μας προέρχονται από κάποια διωνυμική κατανομή η οποία έχει κάποια άγνωστη πιθανότητα επιτυχίας, την οποία και εκτιμούμε. Όταν η πραγματική πιθανότητα επιτυχίας τείνει στην μονάδα (ή στο μηδέν), τότε σε ένα δείγμα μεγέθους n , η πιθανότητα να έχουμε n επιτυχίες (μηδέν επιτυχίες) είναι μεγαλύτερη από την πιθανότητα να έχουμε $n-1$ επιτυχίες (μια επιτυχία). Επομένως είναι πιο πιθανόν να συλλέξουμε ένα δείγμα το οποίο να αποτελείται μόνο από επιτυχίες (αποτυχίες) και συνεπώς η εκτίμηση της διωνυμικής πιθανότητας να είναι η μονάδα (το μηδέν) κάτι το οποίο δεν είναι σύμφωνο με τον ορισμό της διωνυμικής κατανομής αφού $0 < p < 1$. Με την προσθήκη μιας επιτυχίας και μιας αποτυχίας ουσιαστικά αποφεύγουμε την περίπτωση αυτή, και τα bootstrap διαστήματα εμπιστοσύνης παράγουν διαστήματα εμπιστοσύνης μη μηδενικού πλάτους για οποιαδήποτε δείγμα.

Στο σημείο αυτό όμως είναι καλό να διευκρινίσουμε ότι ο σκοπός μας δεν είναι η εκτίμηση του πραγματικού ποσοστού αλλά η κατασκευή διαστημάτων εμπιστοσύνης μέσω της bootstrap. Επιπροσθέτως, θα πρέπει να τονίσουμε ότι η προσθήκη των δύο παρατηρήσεων γίνεται μόνο στο αρχικό μας δείγμα. Αν η προσθήκη γινόταν σε κάθε bootstrap σύνολο δεδομένων τότε ουσιαστικά απλώς θα μετατοπίζαμε την εκτίμηση και θα επιτρέπαμε την ύπαρξη διαστημάτων εμπιστοσύνης μηδενικού πλάτους.

Στον Πίνακα 2.5 παρουσιάζεται η μέση πιθανότητα κάλυψης των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης για δείγματα μικρού μεγέθους.

Πίνακας 2.5

Μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $n \leq 50$ και $p \sim U(0,1)$.

$1-2\alpha$	n	5	10	15	20	25	30	35	40	45	50
0.90	BT	0.9863	0.9737	0.9326	0.9403	0.9243	0.9256	0.9247	0.9167	0.9200	0.9184
	BP	0.9360	0.9359	0.9027	0.9148	0.9038	0.9067	0.9078	0.9016	0.9069	0.9065
	BC	0.9360	0.9203	0.9092	0.9238	0.9066	0.9172	0.9124	0.9055	0.9091	0.9082
	$BC_{\alpha,j}$	0.9456	0.9352	0.9113	0.9177	0.9011	0.9130	0.9083	0.9015	0.9075	0.9074
	$BC_{\alpha,p}$	0.9360	0.9359	0.9027	0.9148	0.9071	0.9124	0.9107	0.9041	0.9067	0.9085
0.95	BT	0.9942	0.9886	0.9885	0.9807	0.9772	0.9744	0.9749	0.9681	0.9700	0.9677
	BP	0.9543	0.9557	0.9607	0.9570	0.9561	0.9552	0.9575	0.9519	0.9552	0.9539
	BC	0.9700	0.9557	0.9681	0.9618	0.9577	0.9571	0.9567	0.9541	0.9581	0.9568
	$BC_{\alpha,j}$	0.9795	0.9710	0.9587	0.9594	0.9587	0.9592	0.9572	0.9516	0.9559	0.9552
	$BC_{\alpha,p}$	0.9700	0.9557	0.9607	0.9565	0.9584	0.9548	0.9563	0.9519	0.9552	0.9548

Η μέση πιθανότητα κάλυψης των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης έχει βελτιωθεί σημαντικά σε σχέση με τον Πίνακα 2.1. Η bootstrap-t (BT) μέθοδος κατασκευής διαστημάτων εμπιστοσύνης είναι η πιο συντηρητική και η μόνη μέθοδος της οποίας η μέση πιθανότητα κάλυψης δεν είναι ποτέ κοντά στο ονομαστικό επίπεδο. Η μέθοδος BC είναι καλή μόνο για $n=5,10$ είναι πάρα πολύ μικρό. Όλες οι άλλες μέθοδοι και κυρίως οι μέθοδοι Bootstrap Percentile (BP) και $BC_{\alpha,p}$ έχουν τις περισσότερες τιμές της μέσης πιθανότητας κάλυψης κοντά στο ονομαστικό επίπεδο. Αξίζει να σημειωθεί ότι τα 90% τροποποιημένα bootstrap διαστήματα εμπιστοσύνης για $n=15,25,40$ έχουν μέση πιθανότητα κάλυψης πιο κοντά στο ονομαστικό επίπεδο από ότι η μέθοδος Wilson όπως και τα 95% τροποποιημένα bootstrap διαστήματα εμπιστοσύνης για $n=5,40$.

Για δείγματα μεγάλου μεγέθους τα αποτελέσματα όσον αφορά τα τροποποιημένα bootstrap διαστήματα διαφέρουν (Πίνακας 2.6). Όταν το ονομαστικό επίπεδο είναι 90% τότε η μέθοδος BP παράγει ελαφρώς καλύτερα αποτελέσματα και ακολουθούν οι μέθοδοι BC, $BC_{\alpha,p}$ και $BC_{\alpha,j}$. Ωστόσο, οι τιμές που είναι καλύτερες από αυτών της μεθόδου Wilson, αφού για $n=60,80,100,300,500$ οι τιμές των διαφόρων τροποποιημένων bootstrap μεθόδων είναι καλύτερες. Όταν το ονομαστικό επίπεδο είναι 95%, τότε η μέθοδος BP είναι καλύτερη για $n \leq 80$ και για $n > 400$ ενώ για $n \geq 90$ η μέθοδος $BC_{\alpha,j}$ είναι καλύτερη. Μάλιστα για $n=90$ η μέση πιθανότητα κάλυψης είναι πιο κοντά από το επιθυμητό επίπεδο. Τέλος η μέθοδος BC είναι συντηρητική αν και τα αποτελέσματα είναι κοντά στο ονομαστικό επίπεδο. Η μέθοδος BT παραμένει αρκετά συντηρητική τόσο σε 90% όσο και σε 95%.

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

Πίνακας 2.6

Μέση πιθανότητα κάλυψης των 90% και 95% ονομαστικών τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $60 \leq n \leq 500$ και $p \sim U(0,1)$.

$1-2\alpha$	n	60	70	80	90	100	200	300	400	500
0.90	BT	0.9098	0.9176	0.9096	0.9135	0.9094	0.9094	0.9074	0.9039	0.9025
	BP	0.8996	0.9069	0.9003	0.9050	0.9016	0.9016	0.9030	0.9009	0.9002
	BC	0.8989	0.9063	0.9003	0.9054	0.9003	0.9027	0.9004	0.8993	0.8997
	$BC_{\alpha,i}$	0.9037	0.9049	0.8995	0.9030	0.8998	0.8998	0.9015	0.8992	0.8998
	$BC_{\alpha,p}$	0.9002	0.9069	0.9003	0.9056	0.9021	0.9021	0.9031	0.9013	0.9002
0.95	BT	0.9666	0.9629	0.9636	0.9601	0.9612	0.9564	0.9547	0.9537	0.9531
	BP	0.9544	0.9519	0.9537	0.9512	0.9531	0.9518	0.9514	0.9511	0.9509
	BC	0.9557	0.9536	0.9554	0.9523	0.9551	0.9523	0.9519	0.9514	0.9511
	$BC_{\alpha,i}$	0.9559	0.9520	0.9538	0.9506	0.9528	0.9510	0.9506	0.9511	0.9506
	$BC_{\alpha,p}$	0.9555	0.9519	0.9537	0.9512	0.9534	0.9518	0.9515	0.9511	0.9509

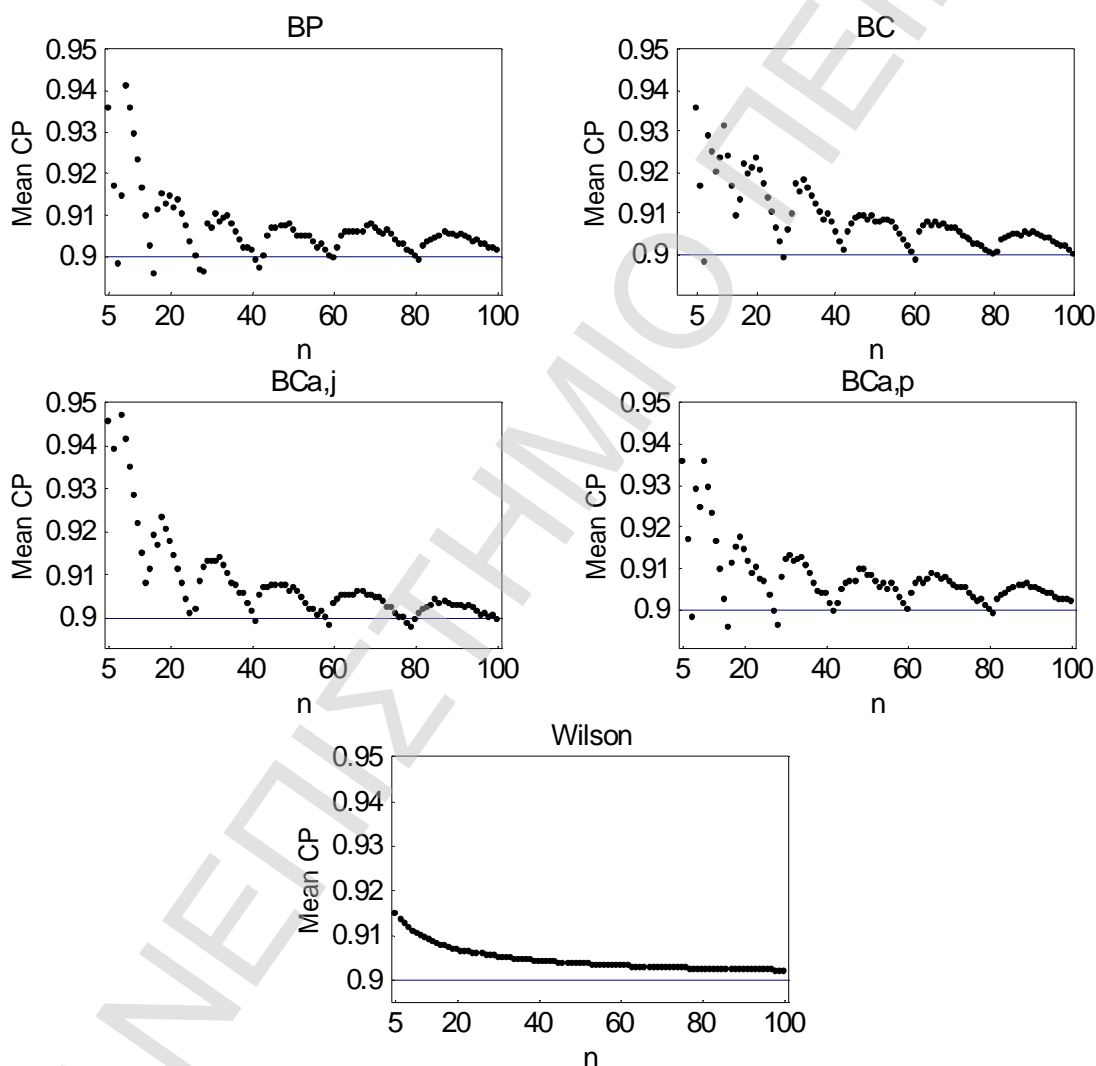
Συνοψίζοντας τους Πίνακες 2.5 και 2.6 μπορούμε να βγάλουμε ορισμένα χρήσιμα συμπεράσματα όσο αφορά τα διαστήματα εμπιστοσύνης που δημιουργούνται από την τροποποιημένη bootstrap.

1. Η μέθοδος BT είναι η πιο συντηρητική και για κανένα από τα n που επιλεχθήκαν δεν δίνει καλύτερα αποτελέσματα. Όλες οι άλλες μέθοδοι είναι ικανοποιητικές, με την έννοια ότι για κάποιες τιμές του n δίνουν μέση πιθανότητα κάλυψης όχι μόνο κοντά στο ονομαστικό επίπεδο αλλά και καλύτερες τιμές από τις παραδοσιακές μεθόδους.
2. Όσο το n μεγαλώνει η μέση πιθανότητα κάλυψης για οποιαδήποτε bootstrap διαδικασία κατασκευής διαστημάτων εμπιστοσύνης δεν μικραίνει πάντα. Για παράδειγμα, το 95% BP διάστημα εμπιστοσύνης για $n=90$ έχει μέση πιθανότητα κάλυψης 0.9512 ενώ για $n=100$ η μέση πιθανότητα κάλυψης είναι 0.9531. Αυτό δεν συμβαίνει με τις παραδοσιακές μεθόδους στις οποίες όσο το n αυξάνεται τόσο η μέση πιθανότητα κάλυψης μικραίνει και τείνει στο επιθυμητό ονομαστικό επίπεδο.
3. Η προηγούμενη παρατήρηση γίνεται πιο σαφής αν για παράδειγμα, προσέξουμε τις τιμές της μέσης πιθανότητας κάλυψης της μεθόδου $BC_{\alpha,j}$ για δείγματα μεγάλου μεγέθους σε 90% ονομαστικό επίπεδο. Παρατηρούμε ότι η μέση πιθανότητα κάλυψης να μην είναι κοντά στο ονομαστικό επίπεδο κατά απόλυτη τιμή αλλά είναι και μικρότερη από αυτή. Ωστόσο οι διαφορές αυτές είναι πάρα πολύ μικρές.

Δημιουργεί έτσι εύλογες απορίες για την καταλληλότητα των διαστημάτων εμπιστοσύνης που προκύπτουν μέσω της τροποποιημένης bootstrap και κυρίως για το αν ο τρόπος με τον οποίο μειώνεται η μέση πιθανότητα κάλυψης μπορεί να προβλεφθεί. Επίσης είναι σημαντικό να εξασφαλίσουμε ότι η πιθανότητα κάλυψης των διαστημάτων εμπιστοσύνης θα είναι τουλάχιστον όσο το επιθυμητό επίπεδο κάλυψης. Τα παραπάνω θα τα ελέγξουμε γραφικά

2^ο Κεφάλαιο

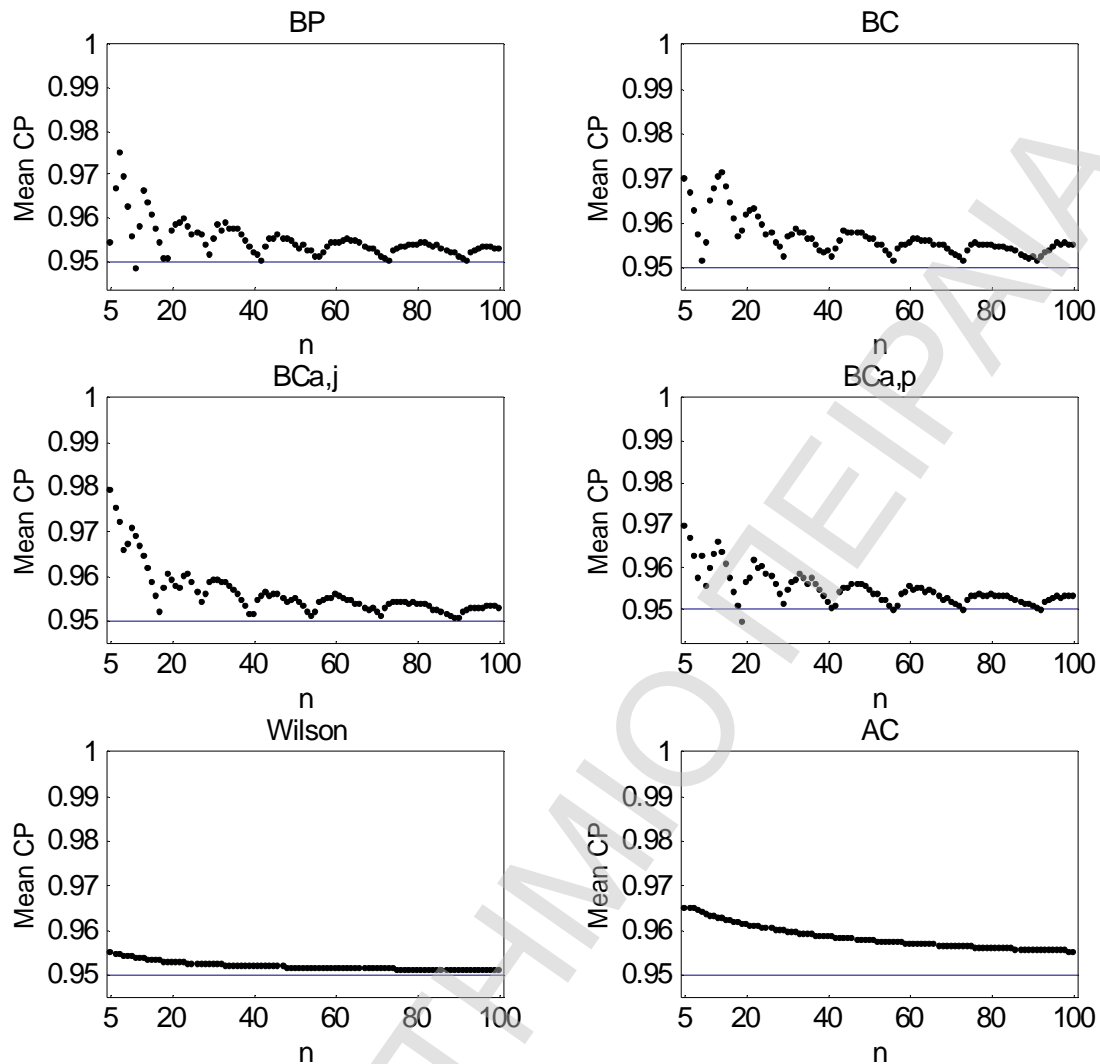
εξετάζοντας την μέση πιθανότητα κάλυψης για $5 \leq n \leq 100$. Για τα 90% ονομαστικού επιπέδου διαστήματα εμπιστοσύνης θα επιλέξουμε την μέθοδο Wilson από τις παραδοσιακές μεθόδους και όλες μεθόδους bootstrap εκτός από την BT. Αποκλείονται οι μέθοδοι που είναι αρκετά συντηρητικές ή που δεν έχουν μέση πιθανότητα κάλυψης κοντά στο επιθυμητό επίπεδο, (Wald). Για τα 95% ονομαστικού επιπέδου διαστήματα εμπιστοσύνης εξετάζουμε και την μέθοδο AC. Στις επόμενες σελίδες δίνονται τα γραφήματα. Η οριζόντια γραμμή είναι το ονομαστικό επίπεδο. Αυτό που επιθυμούμε στα γραφήματα είναι να βρίσκεται η μέση πιθανότητα κάλυψης κοντά στην οριζόντια γραμμή.



Γράφημα 2.1

Μέση πιθανότητα κάλυψης των 90% ονομαστικών διαστημάτων εμπιστοσύνης για $5 \leq n \leq 100$.

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.



Γράφημα 2.2

Μέση πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $5 \leq n \leq 100$.

Από τα Γραφήματα 2.1 και 2.2 παρατηρούμε η μέση πιθανότητα κάλυψης για τα bootstrap διαστήματα εμπιστοσύνης ελαττώνεται σε κυματοειδή μορφή ενώ η μέθοδος του Wilson και των Agresti και Coull ελαττώνονται σε γραμμική μορφή. Αυτό το γεγονός δικαιολογεί σε μεγάλο βαθμό γιατί συναντήσαμε τιμές της μέσης πιθανότητας κάλυψης των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης πιο κοντά στο ονομαστικό επίπεδο από ότι η αντίστοιχη τιμή της μεθόδου του Wilson. Επίσης, η μέθοδος BP έχει παρόμοιο Γράφημα με την μέθοδο $BC_{a,p}$, ενώ η μέθοδος BC με την μέθοδο $BC_{a,j}$. Τα 90% ονομαστικού επιπέδου τροποποιημένα bootstrap διαστήματα εμπιστοσύνης έχουν αρκετές τιμές που είναι μικρότερες από 0.90. Αυτό είναι κάτι το οποίο μπορεί να επιφέρει προβληματισμό για την χρησιμότητα τους. Εν αντίθεση τα 95% ονομαστικού επιπέδου τροποποιημένα bootstrap διαστήματα εμπιστοσύνης έχουν μόνο δύο τιμές μικρότερες από το ονομαστικό επίπεδο και μάλιστα πάρα

2^ο Κεφάλαιο

πολύ μικρές κατά απόλυτη τιμή. Τέλος, είναι ιδιαίτερα χρήσιμο να παρατηρήσουμε ότι η μέθοδος BP έχει τιμές μέσης πιθανότητας κάλυψης αρκετά μικρότερες από ότι η μέθοδος των Agresti και Coull όταν $n \geq 15$.

Τις παραπάνω μεθόδους θα τις συγκρίνουμε και ως προς το αναμενόμενο πλάτος τους σύμφωνα με τον τύπο (2.23).

Πίνακας 2.7

Μέσο αναμενόμενο πλάτος των 90% και 95% ονομαστικών διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $n \leq 50$ και $p \sim U(0,1)$.

$1-2a$	n	5	10	15	20	25	30	35	40	45	50
0.90	Wilson	0.4849	0.3718	0.3128	0.2751	0.2484	0.2282	0.2123	0.1993	0.1884	0.1791
	BP	0.5238	0.4091	0.3088	0.2857	0.2479	0.2319	0.2162	0.1986	0.1915	0.1817
	BC	0.5238	0.3939	0.3162	0.2944	0.2507	0.2399	0.2192	0.2009	0.1933	0.1833
	BC _{a,j}	0.5714	0.4242	0.3309	0.2987	0.2536	0.2440	0.2222	0.2033	0.1970	0.1870
	BC _{a,p}	0.5238	0.4091	0.3088	0.2857	0.2507	0.2359	0.2177	0.1998	0.1915	0.1825
0.95	Wilson	0.5581	0.4354	0.3687	0.3254	0.2943	0.2708	0.2520	0.2367	0.2239	0.2129
	AC	0.5864	0.4565	0.3841	0.3374	0.3041	0.2789	0.2590	0.2428	0.2292	0.2177
	BP	0.5714	0.4394	0.3824	0.3333	0.2991	0.2742	0.2583	0.2358	0.2276	0.2149
	BC	0.6190	0.4394	0.3971	0.3420	0.3020	0.2782	0.2583	0.2381	0.2303	0.2172
	BC _{a,j}	0.6667	0.5152	0.3971	0.3550	0.3134	0.2903	0.2673	0.2427	0.2359	0.2217
	BC _{a,p}	0.6190	0.4394	0.3824	0.3333	0.3020	0.2742	0.2568	0.2358	0.2276	0.2157

Από τον Πίνακα 2.7 παρατηρούμε ότι η μέθοδος BP δίνει το μικρότερο μέσο αναμενόμενο πλάτος από τις υπόλοιπες bootstrap μεθόδους τις περισσότερες φορές τόσο σε 90% ονομαστικό επίπεδο όσο και σε 95%. Επιπλέον, έχει πάντα μικρότερο μέσο αναμενόμενο μήκος από την μέθοδο των Agresti και Coull ενώ σε κάποιες περιπτώσεις είναι και καλύτερη από την μέθοδο του Wilson. Η μόνη μέθοδος bootstrap που την συναγωνίζεται είναι η BC_{a,p} σε 95% ονομαστικό επίπεδο.

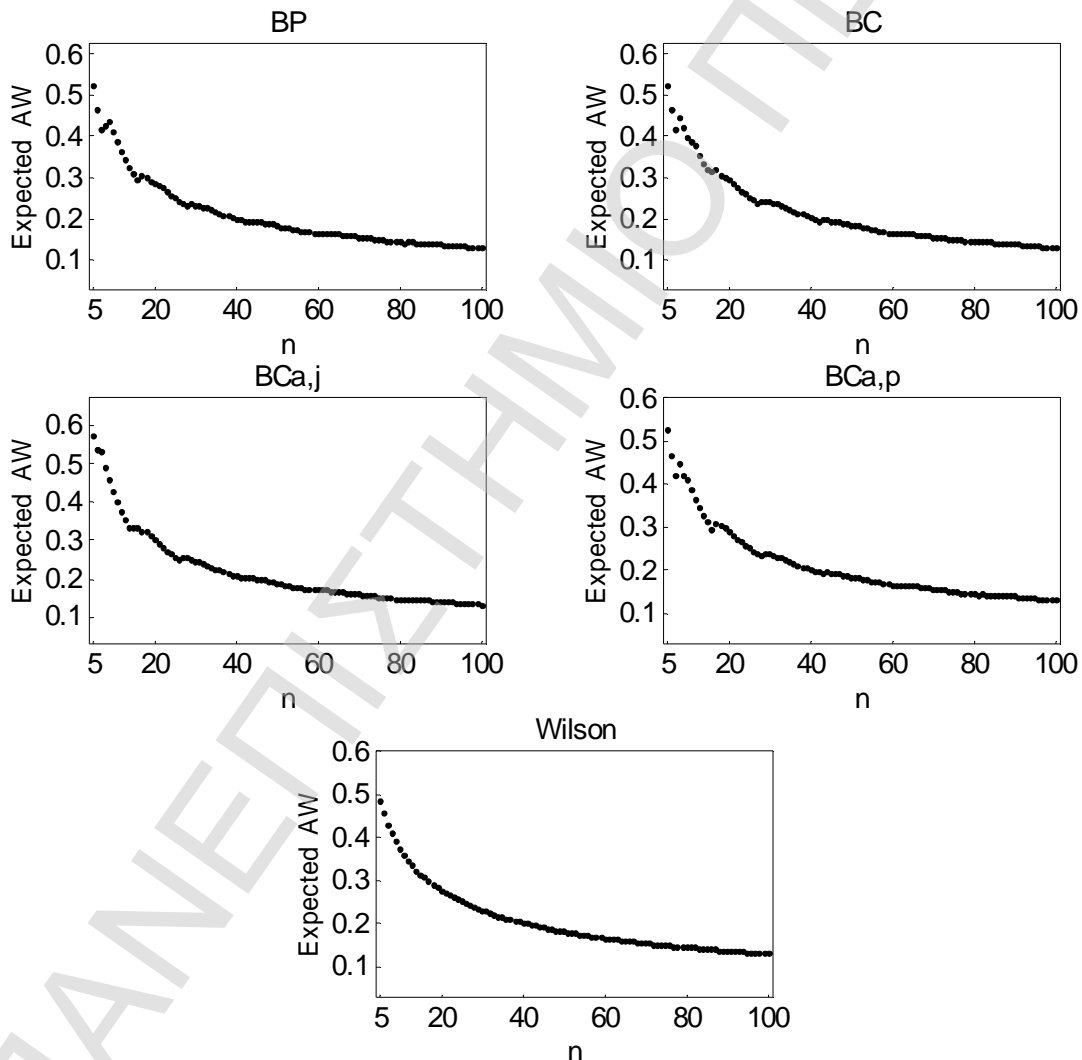
Όταν το δείγμα μας είναι μεγάλο τα προηγούμενα αποτελέσματα εδραιώνονται ακόμα περισσότερο (βλ. Πίνακας 2.8). Η μέθοδος BP υπερέρχει εμφανώς των υπολοίπων bootstrap μεθόδων και της μεθόδου των Agresti-Coull και όσο το n αυξάνει τόσο πλησιάζει τις τιμές της μεθόδου Wilson, για παράδειγμα για $n = 500$ οι δύο μέθοδοι διαφέρουν μόνο κατά 0.0001. Στην επόμενη σελίδα δίνονται και τα γραφήματα του μέσου αναμενόμενου πλάτους για $5 \leq n \leq 100$.

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

Πίνακας 2.8

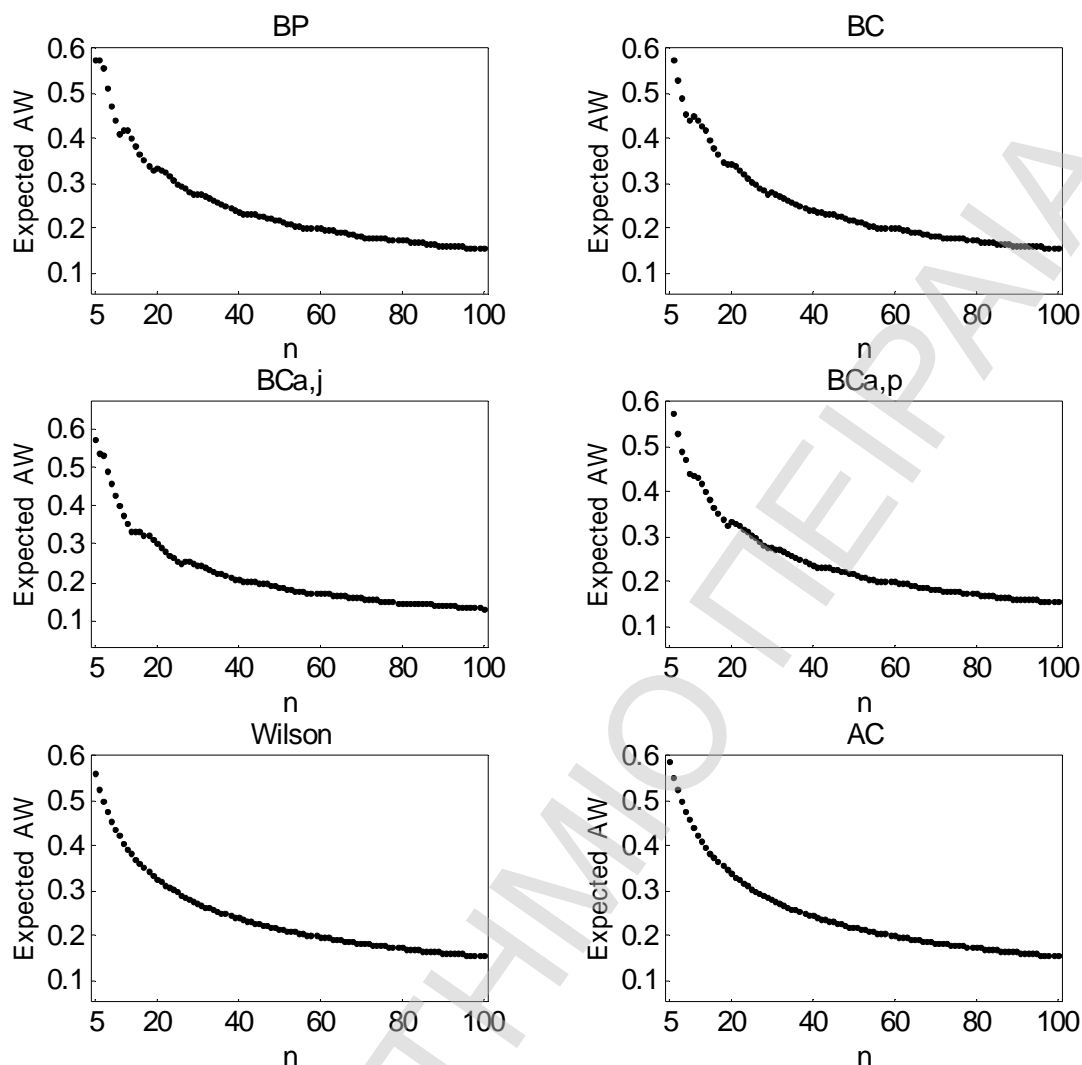
Μέσο αναμενόμενο πλάτος των 90% και 95% ονομαστικών διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $60 \leq n \leq 500$ και $p \sim U(0,1)$.

$1-2\alpha$	n	60	70	80	90	100	200	300	400	500
0.90	Wilson	0.1640	0.1522	0.1426	0.1347	0.1279	0.0909	0.0743	0.0644	0.0577
	BP	0.1623	0.1545	0.1415	0.1362	0.1277	0.0915	0.0743	0.0642	0.0575
	BC	0.1623	0.1549	0.1421	0.1369	0.1277	0.0916	0.0743	0.0642	0.0575
	$BC_{\alpha,j}$	0.1687	0.1569	0.1442	0.1378	0.1291	0.0921	0.0745	0.0646	0.0576
	$BC_{\alpha,p}$	0.1623	0.1545	0.1415	0.1364	0.1279	0.0915	0.0743	0.0642	0.0575
0.95	Wilson	0.1951	0.1811	0.1698	0.1603	0.1523	0.1083	0.0886	0.0768	0.0687
	AC	0.1989	0.1843	0.1725	0.1627	0.1543	0.1091	0.0890	0.0771	0.0689
	BP	0.1978	0.1811	0.1716	0.1598	0.1536	0.1088	0.0889	0.0769	0.0688
	BC	0.1988	0.1823	0.1728	0.1605	0.1547	0.1090	0.0890	0.0771	0.0689
	$BC_{\alpha,j}$	0.2036	0.1851	0.1752	0.1624	0.1561	0.1096	0.0892	0.0774	0.0691
	$BC_{\alpha,p}$	0.1988	0.1811	0.1716	0.1598	0.1538	0.1088	0.0889	0.0770	0.0689



Γράφημα 2.3

Μέσο αναμενόμενο πλάτος των 90% ονομαστικών διαστημάτων εμπιστοσύνης για $5 \leq n \leq 100$ και $p \sim U(0,1)$.



Γράφημα 2.4

Μέσο αναμενόμενο πλάτος των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $5 \leq n \leq 100$ και $p \sim U(0,1)$.

Τα γραφήματα επιβεβαιώνουν τα όσα αναφέραμε νωρίτερα, αν και οι διαφορές τους δεν είναι ευδιάκριτες.

Το τελευταίο κριτήριο που θα εξετάσουμε είναι η ρίζα του μέσου τετραγωνικού σφάλματος η οποία δίνεται από τον τύπο 2.23. Η ποσότητα αυτή μετράει κατά πόσο τα σφάλματα κατανέμονται ομοιόμορφα γύρω από το ονομαστικό επίπεδο (βλ. Agresti and Coull, 1998). Βάσει όσων σχολιάσαμε νωρίτερα για την μέση πιθανότητα κάλυψης των διαστημάτων εμπιστοσύνης παρατηρήσαμε στα Γραφήματα 2.1 και 2.2 τότε αναμένουμε ότι οι μέθοδοι bootstrap θα έχουν μεγαλύτερη τιμή στην ρίζα του μέσου τετραγωνικού σφάλματος. Η κυματοειδής μορφή των σχημάτων είναι ο βασικός λόγος για τον οποίο

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.

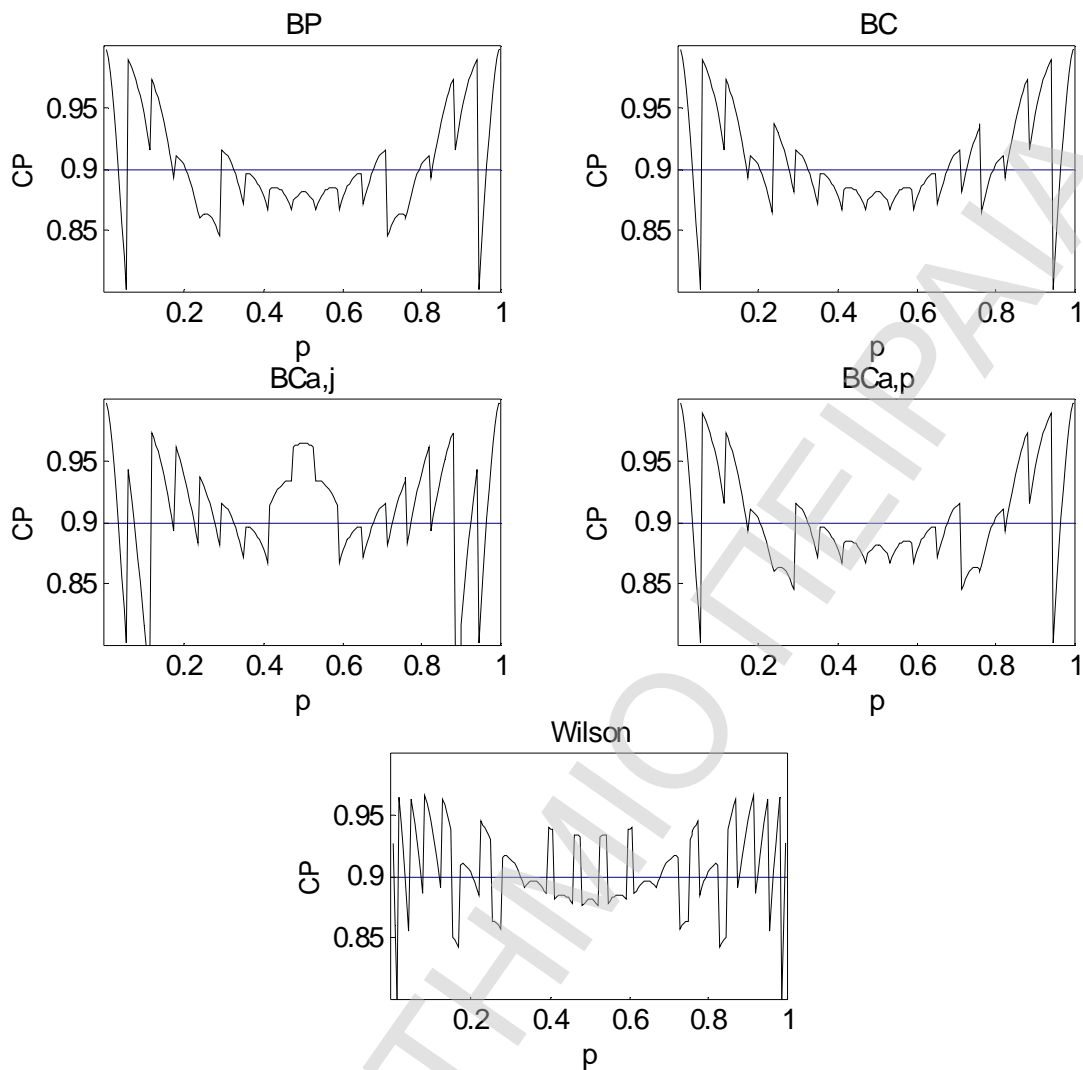
δημιουργείται αυτό το φαινόμενο. Στον Πίνακα 2.10 παραθέτουμε τις τιμές της ρίζας του μέσου τετραγωνικού σφάλματος για δείγματα μικρού και μεσαίου μεγέθους.

Πίνακας 2.10

Ρίζα του μέσου τετραγωνικού σφάλματος των 90% και 95% ονομαστικών διαστημάτων εμπιστοσύνης για το διωνυμικό ποσοστό p , όταν $5 \leq n \leq 50$ και $p \sim U(0,1)$

1-2α	n	5	10	15	20	25	30	35	40	45	50
0.90	Wilson	0.0529	0.0361	0.0331	0.0295	0.0257	0.0234	0.0231	0.0219	0.0203	0.0204
	BP	0.0500	0.0517	0.0407	0.0357	0.0312	0.0294	0.0273	0.0254	0.0279	0.0264
	BC	0.0500	0.0488	0.0388	0.0401	0.0328	0.0320	0.0284	0.0272	0.0269	0.0256
	$BC_{\alpha,j}$	0.0614	0.0573	0.0456	0.0458	0.0382	0.0408	0.0361	0.0339	0.0345	0.0326
	$BC_{\alpha,p}$	0.0500	0.0517	0.0407	0.0357	0.0302	0.0316	0.0281	0.0257	0.0252	0.0263
0.95	Wilson	0.0295	0.0218	0.0186	0.0171	0.0156	0.0142	0.0132	0.0123	0.0117	0.0116
	AC	0.0289	0.0217	0.0193	0.0176	0.0169	0.0162	0.0155	0.0150	0.0143	0.0141
	BP	0.0365	0.0246	0.0217	0.0225	0.0197	0.0193	0.0156	0.0147	0.0146	0.0135
	BC	0.0407	0.0246	0.0241	0.0216	0.0176	0.0177	0.0158	0.0148	0.0159	0.0148
	$BC_{\alpha,i}$	0.0444	0.0421	0.0337	0.0339	0.0290	0.0283	0.0261	0.0240	0.0242	0.0224
	$BC_{\alpha,p}$	0.0407	0.0246	0.0217	0.0220	0.0198	0.0200	0.0154	0.0147	0.0146	0.0132

Πριν κλείσουμε την σύγκριση των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης με τις παραδοσιακές μεθόδους αξίζει να δούμε την πιθανότητα κάλυψης κάθε μεθόδου για σταθερό μέγεθος δείγματος και σταθερή πιθανότητα επιτυχίας. Για να έχουμε μια πιο ολοκληρωμένη άποψη θα βρούμε την πιθανότητα κάλυψης από τον τύπο (2.17) για $p \in [0.005, 0.995]$. Επειδή οι τιμές του p είναι άπειρες θα πάρουμε τις 199 τιμές του p που είναι της μορφής $0.005 \cdot i, i=1, \dots, 199$. Στην αρχή εξετάζουμε τα 90% διαστήματα εμπιστοσύνης για $n=15, 30, 50$ και 100

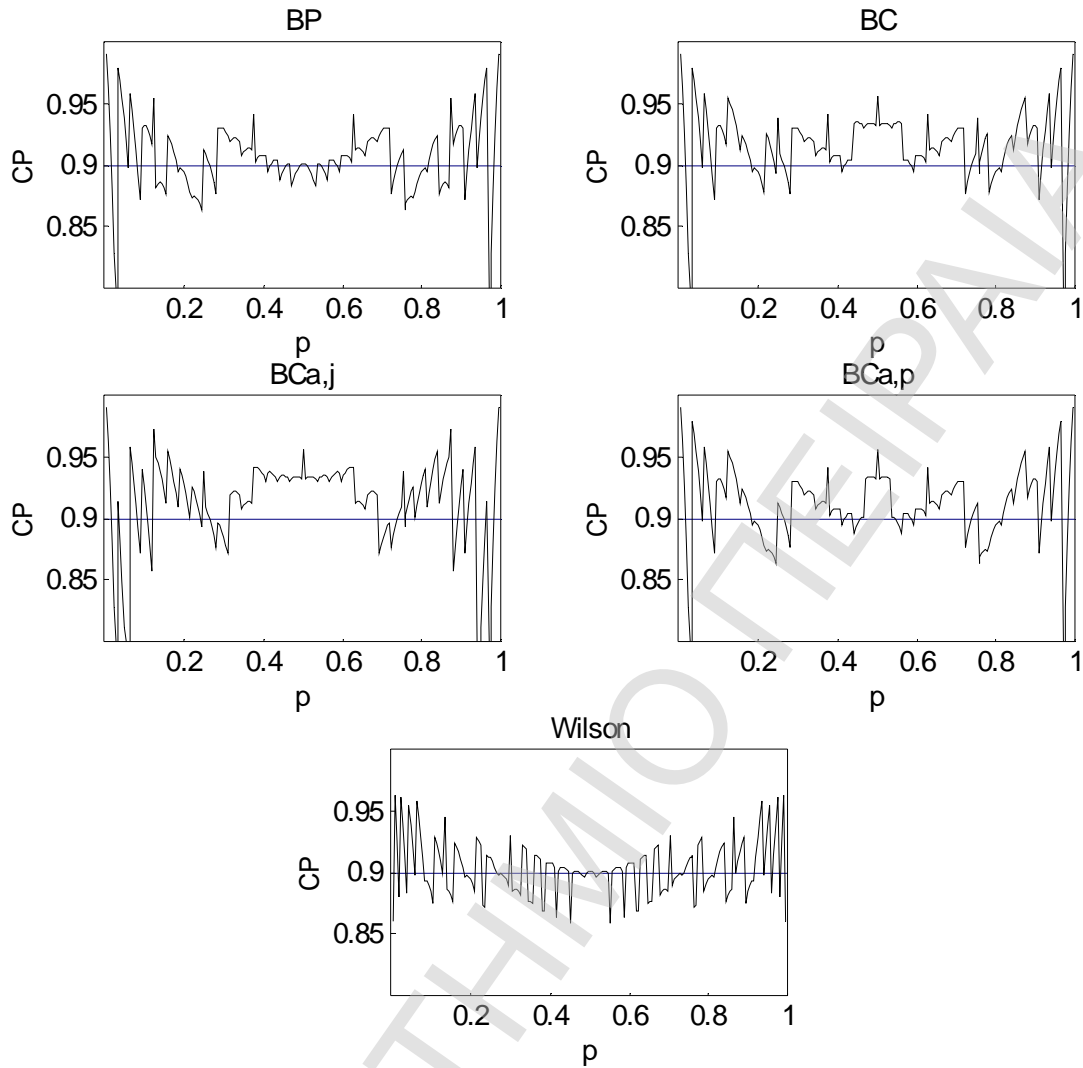


Γράφημα 2.5

Πιθανότητα κάλυψης των 90% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 15$.

Παρατηρούμε ότι η μέθοδος bootstrap δεν συμπεριφέρεται καλά αφού υπάρχουν τιμές της πιθανότητας κάλυψης που είναι αρκετά χαμηλότερες από το επιθυμητό επίπεδο. Επιπλέον υπάρχουν και αρκετές τιμές που είναι πολύ κοντά στην μονάδα, κυρίως όταν το ποσοστό είναι κοντά στο μηδέν ή στην μονάδα. Τα ίδια πράγματα βλέπουμε να ισχύουν και για $n = 30$.

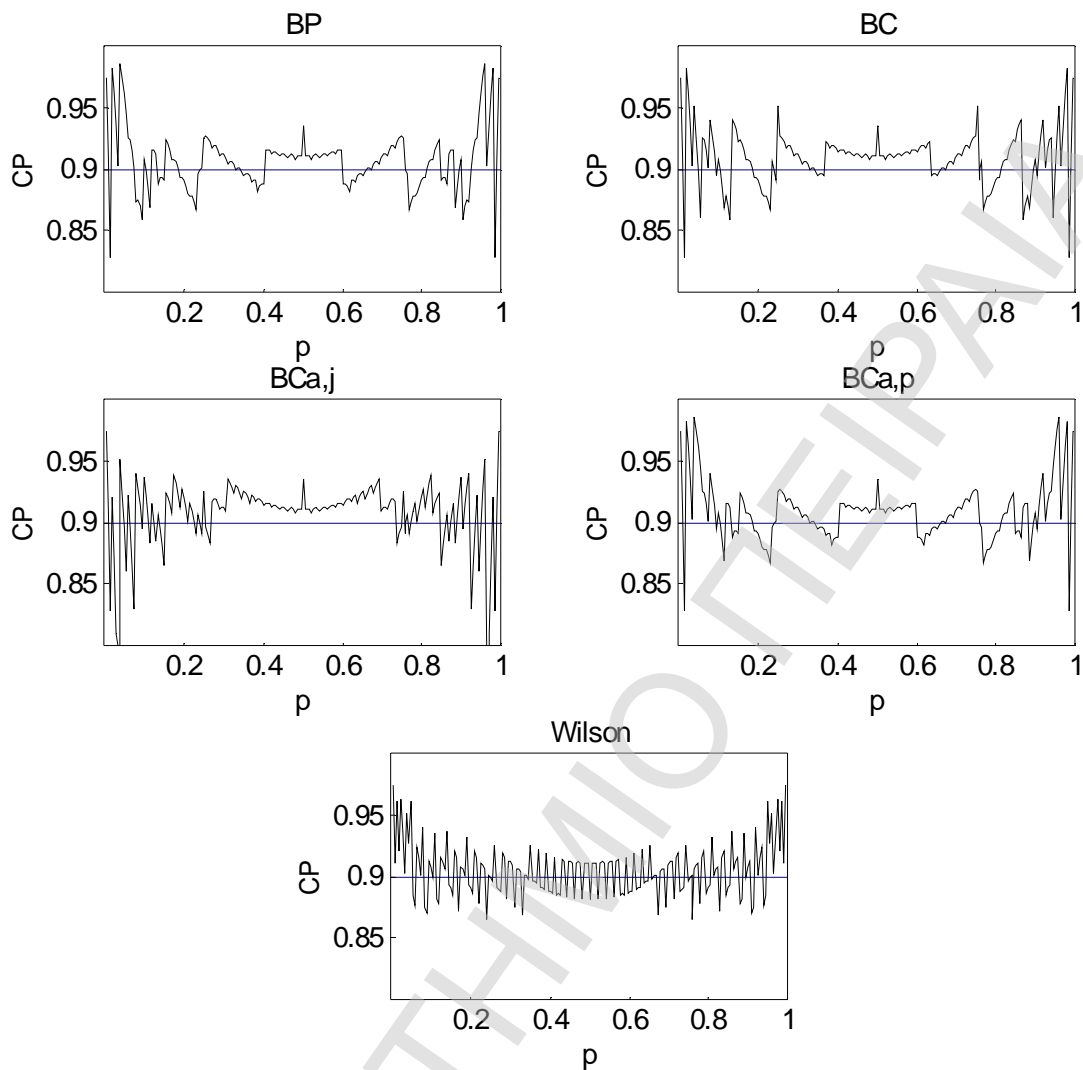
2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.



Γράφημα 2.6

Πιθανότητα κάλυψης των 90% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 30$

Αν και το μέγεθος του δείγματος αυξήθηκε τα συμπεράσματα δεν έχουν αλλάξει ιδιαίτερα. Τα 90% ονομαστικά τροποποιημένα bootstrap διαστήματα εμπιστοσύνης παρουσιάζουν την ίδια ανωμαλία σε ποσοστά που είναι κοντά στην μονάδα και στα άκρα και συνεπώς παραμένουν ακατάλληλα. Στο επόμενο Γράφημα έχουμε αυξήσει το μέγεθος του δείγματος σε $n = 50$. Παρόλο που με την αύξηση θα περιμέναμε να λειτουργούν καλύτερα η εικόνα είναι πολύ άσχημη ενώ η μέθοδος Wilson σταδιακά βελτιώνεται ολοένα και περισσότερο.

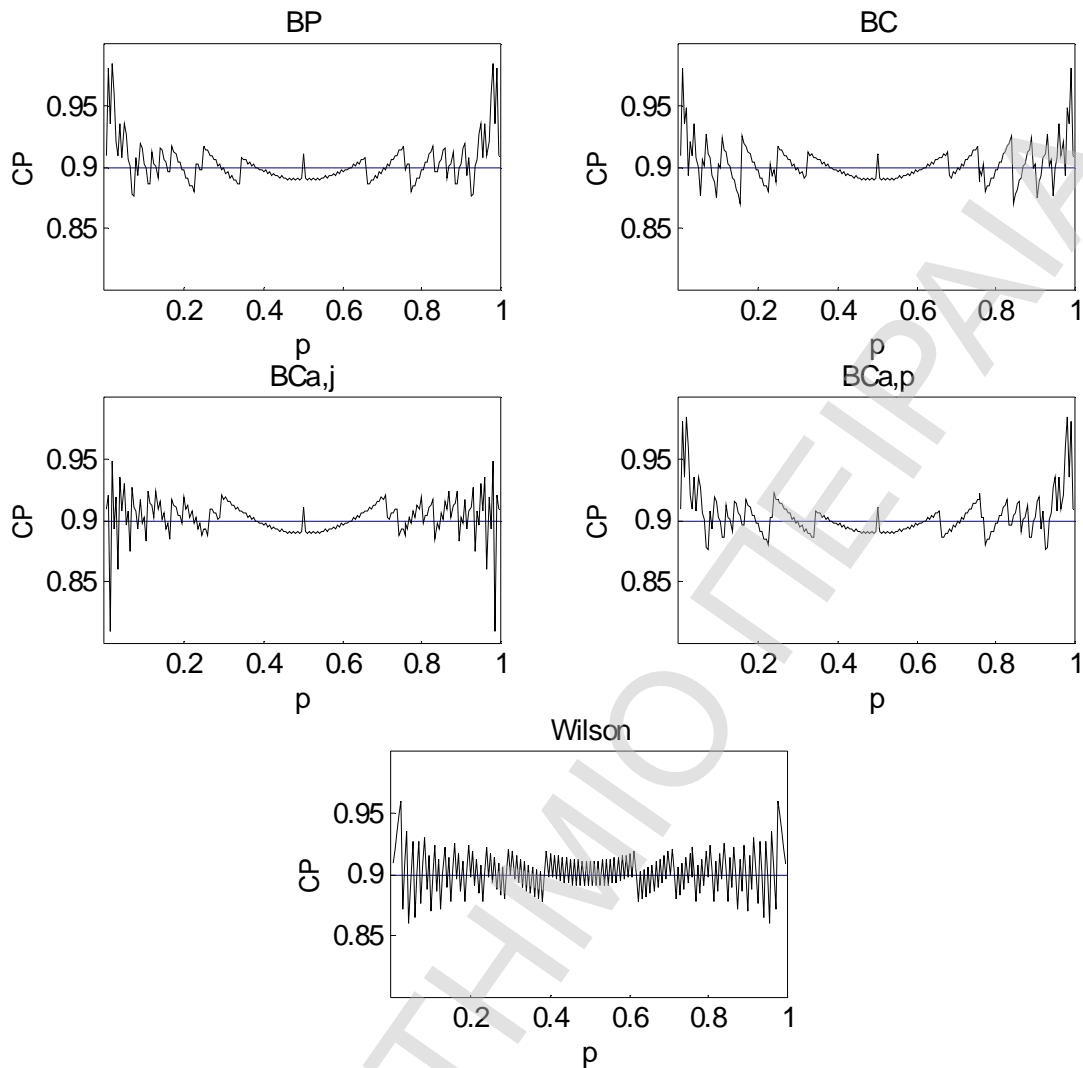


Γράφημα 2.7

Πιθανότητα κάλυψης των 90% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 50$

Η ύστατη προσπάθεια είναι να αυξήσουμε ακόμα περισσότερο τα δείγμα μας, για την ακρίβεια να το διπλασιάσουμε. Στα παρακάτω γραφήματα φαίνεται το πραγματικό πρόβλημα των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης. Η πιθανότητα κάλυψης όταν το ποσοστό είναι κοντά στο μηδέν ή στην μονάδα είναι είτε αρκετά κάτω ή άνω από το ονομαστικό επίπεδο. Για τους παραπάνω λόγους η μέση πιθανότητα κάλυψης τείνει κοντά στο ονομαστικό επίπεδο, χωρίς όμως αυτό να είναι ουσιαστικά σωστό.

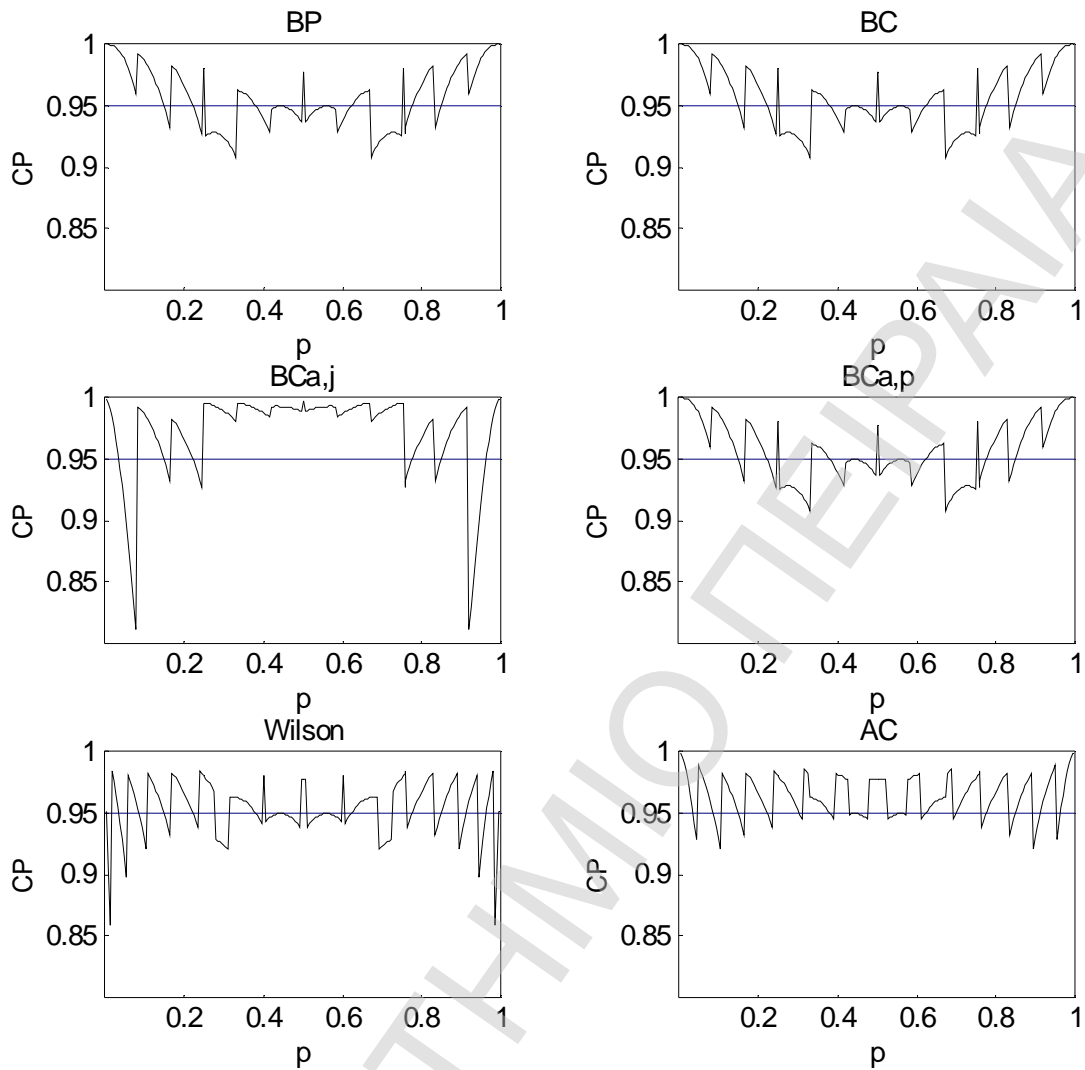
2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.



Γράφημα 2.8

Πιθανότητα κάλυψης των 90% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 100$

Στην συνέχεια θα εξετάσουμε την πιθανότητα κάλυψης των 95% ονομαστικού επιπέδου διαστημάτων εμπιστοσύνης. Τα τροποποιημένα bootstrap διαστήματα εμπιστοσύνης περιμένουμε να έχουν καλύτερη συμπεριφορά από τα αντίστοιχα 90% γιατί η μέση πιθανότητα κάλυψης ήταν σχεδόν πάντα μεγαλύτερη από το ονομαστικό επίπεδο. Θα δοκιμάσουμε πρώτα να πάρουμε μικρού μεγέθους δείγμα, και πιο συγκεκριμένα την τιμή $n = 10$. Στο επόμενο γράφημα βλέπουμε ότι η πιθανότητα κάλυψης των τροποποιημένων bootstrap, εκτός από την μέθοδο $BC_{a,j}$ είναι σχετικά καλή. Πάλι για τιμές κοντά στο μηδέν και στην μονάδα η πιθανότητα κάλυψης είναι κοντά στην μονάδα αλλά οι υπόλοιπες τιμές κατανέμονται γύρω από το ονομαστικό επίπεδο. Η χαμηλότερη τιμή της μέσης πιθανότητας κάλυψης των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης είναι αρκετά μεγαλύτερη από την χαμηλότερη τιμή της μεθόδου Wilson.

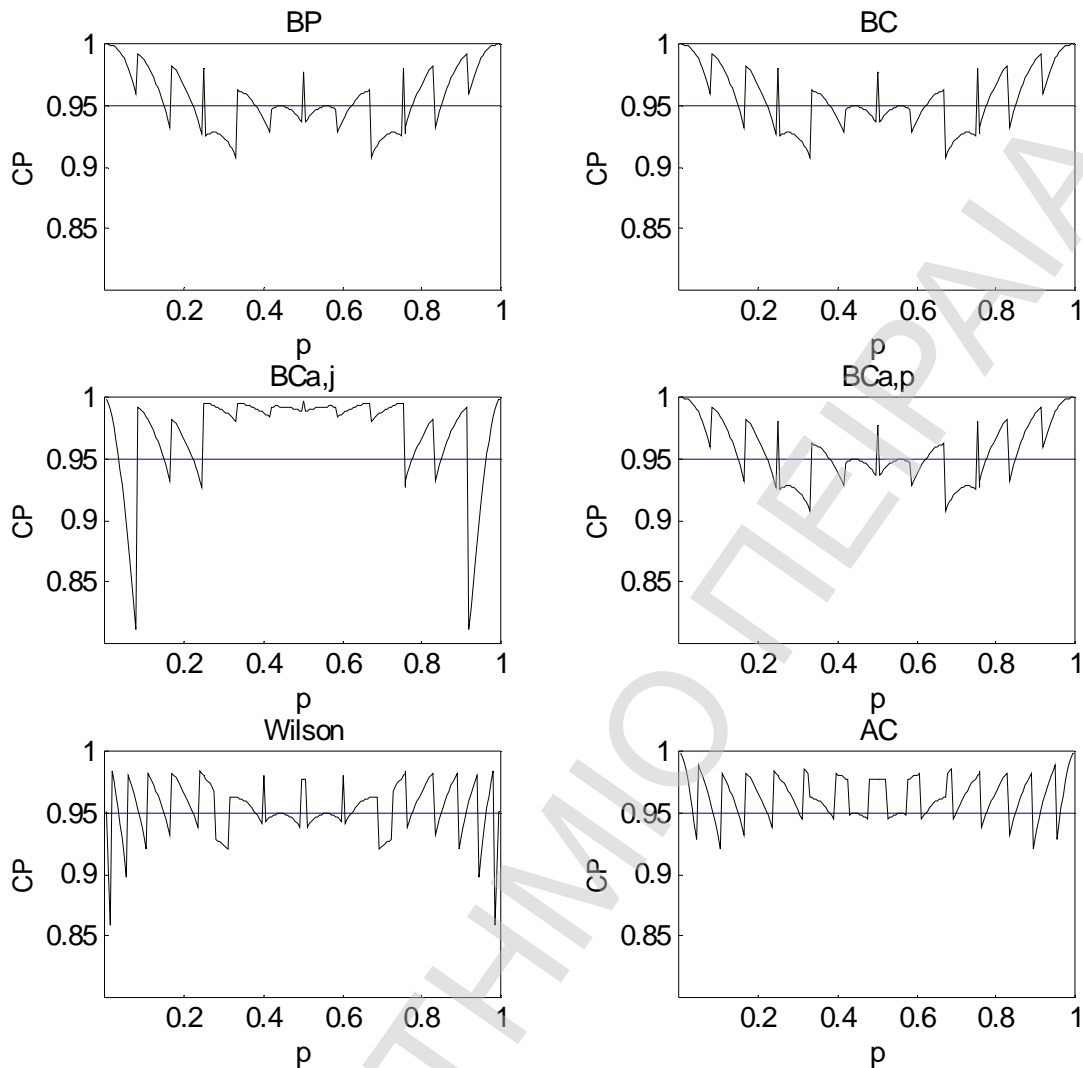


Γράφημα 2.9

Πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 10$

Όταν $n = 20$, τότε φαίνεται ότι η μέθοδος BP συμπεριφέρεται αρκετά καλύτερα από ότι η μέθοδος BC, BC_{a,p} και η μέθοδος Wilson με την έννοια ότι η χαμηλότερη πιθανότητα κάλυψης είναι μεγαλύτερη από την αντίστοιχη πιθανότητα αυτών των μεθόδων. Επίσης, η μέθοδος των Agresti-Coull φαίνεται να είναι πιο συντηρητική από ότι η BP.

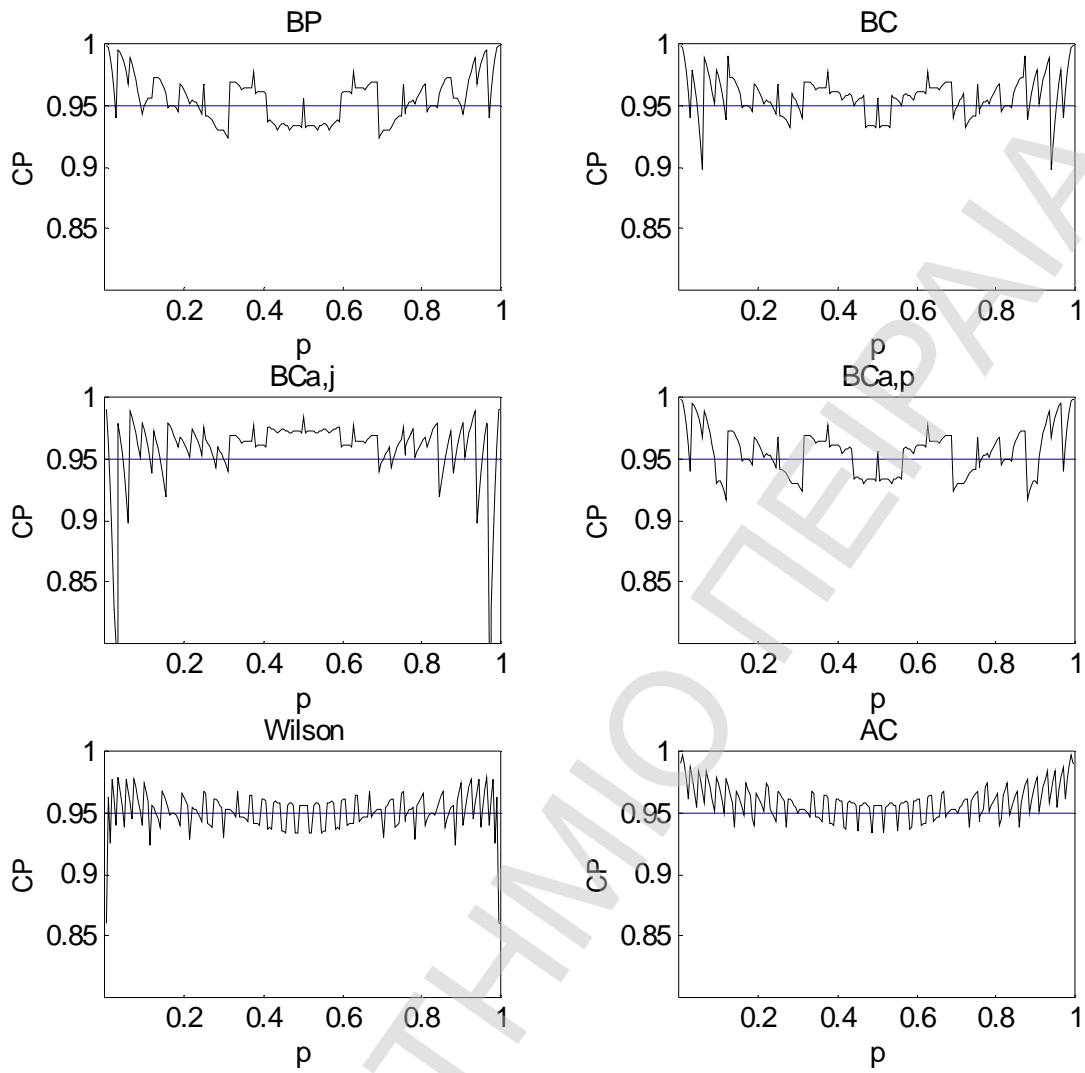
2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.



Γράφημα 2.10

Πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 20$

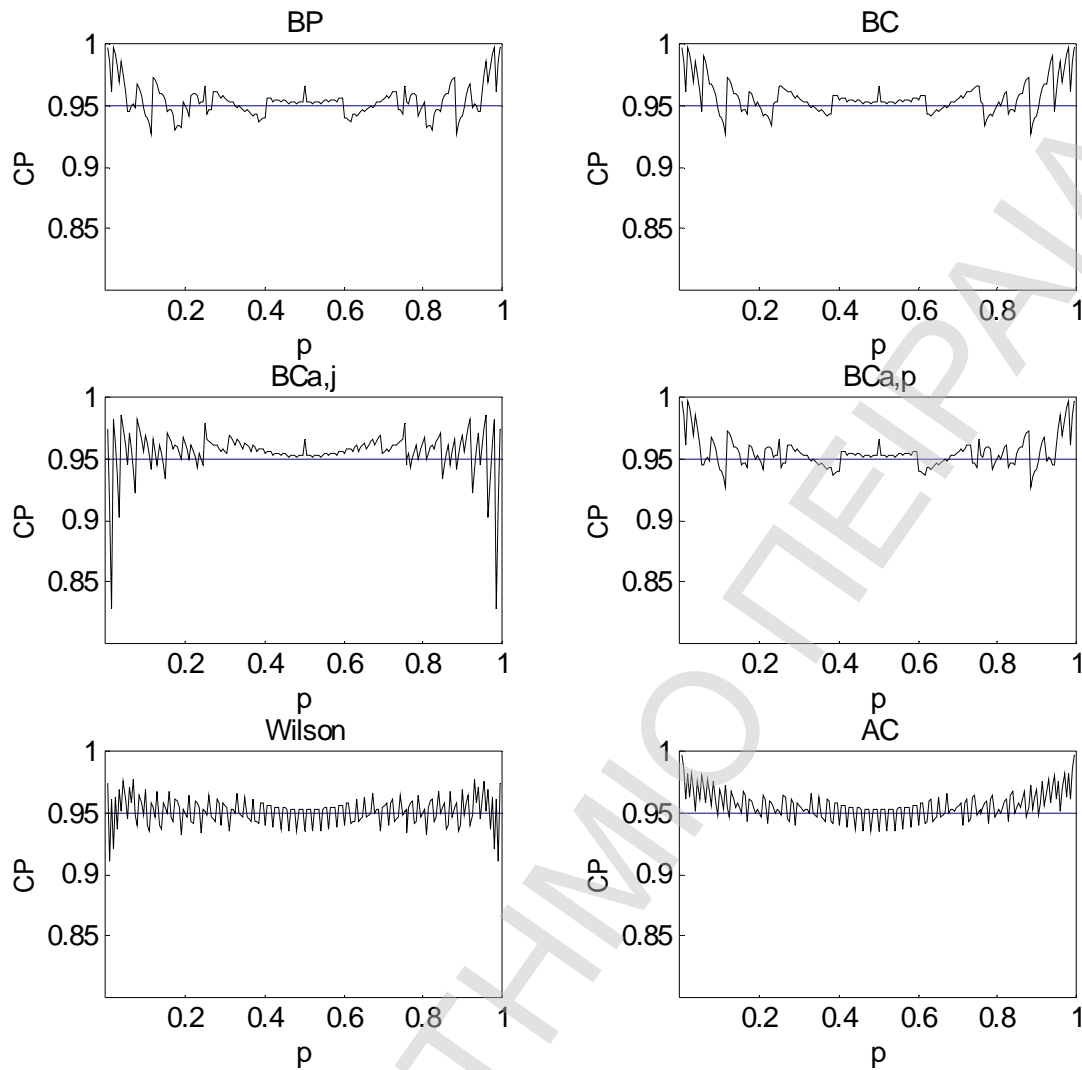
Τα παραπάνω συμπεράσματα επιβεβαιώνονται και για $n = 30$ (βλ. Γράφημα 2.10). Αυτό μας προϊδεάζει ότι η δομή των τροποποιημένων bootstrap διαστημάτων εμπιστοσύνης είναι σταθερή. Αυξάνοντας το μέγεθος του δείγματος και παίρνοντας $n = 50$ παρατηρούμε ότι οι μέθοδοι bootstrap είναι αρκετά παρόμοιοι με εξαίρεση την $BC_{a,j}$, το οποίο φαίνεται στο Γράφημα 2.11. Επίσης η σχέση μεταξύ της μεθόδου BP και των παραδοσιακών μεθόδων παραμένει η ίδια.



Γράφημα 2.11

Πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 30$

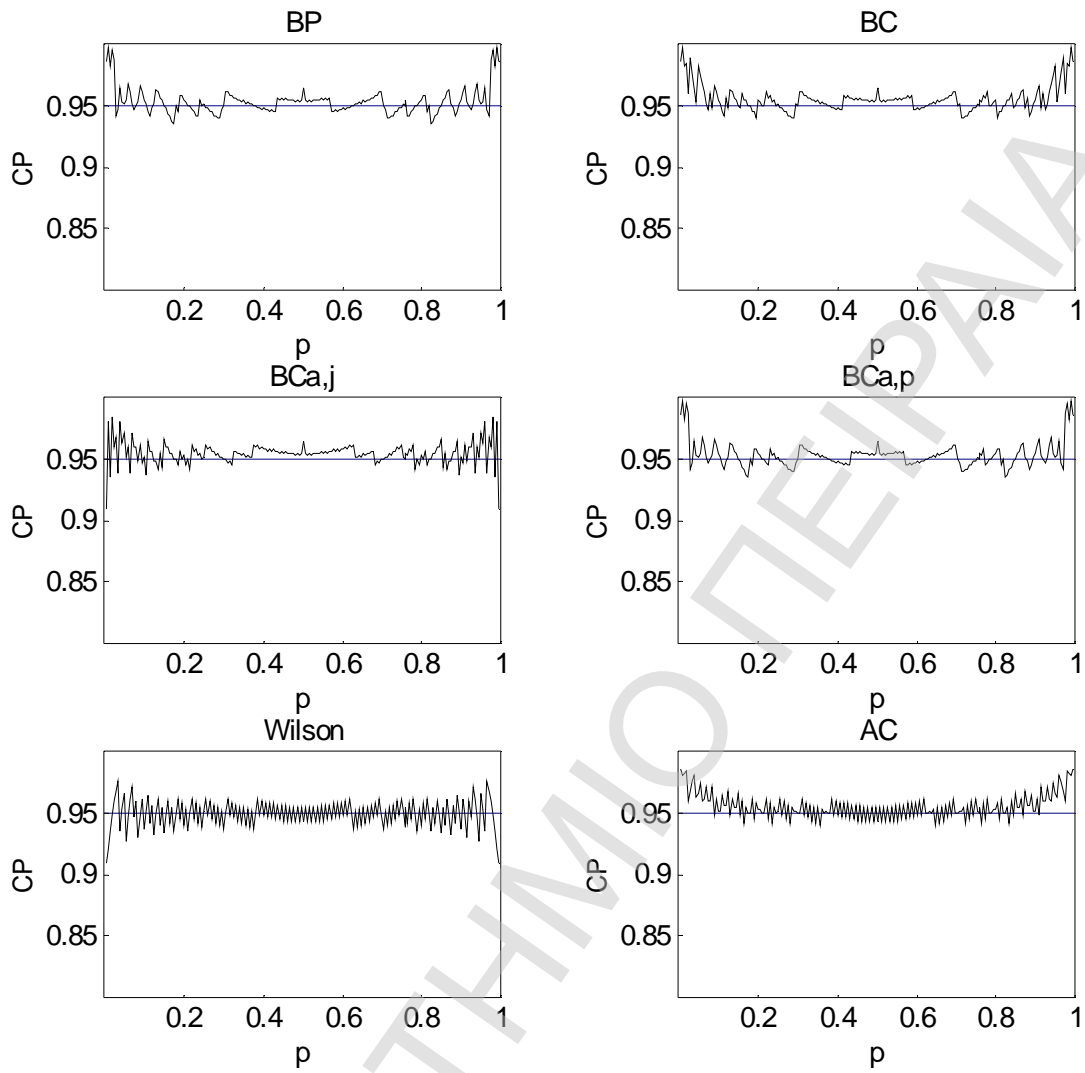
2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.



Γράφημα 2.12

Πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 50$

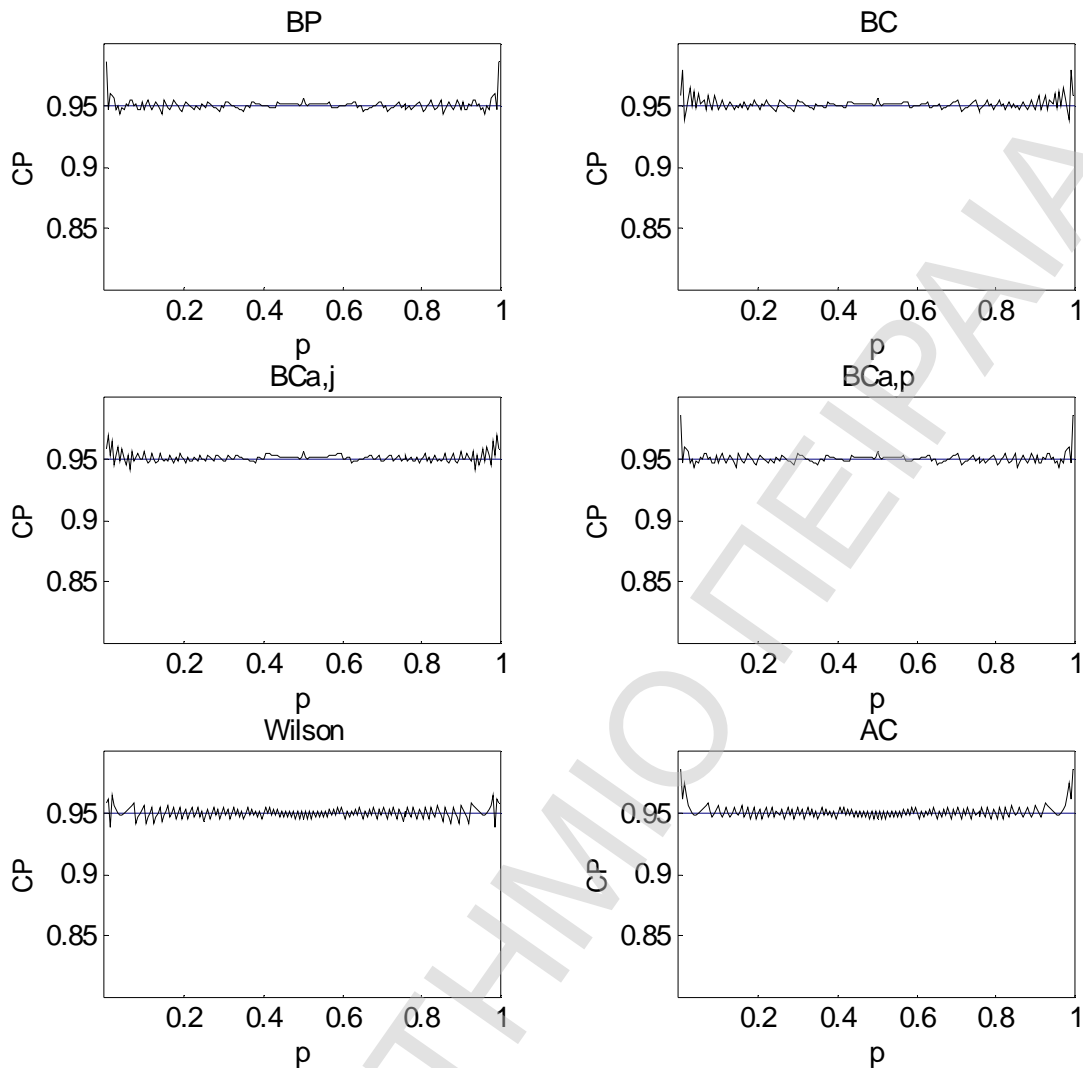
Στην συνέχεια εξετάζουμε και δύο ακόμα περιπτώσεις όταν το μέγεθος του δείγματος είναι αρκετά μεγάλο. Το Γράφημα 2.13 αντιστοιχεί σε $n = 100$ και το Γράφημα 2.14 σε $n = 500$. Στην πρώτη περίπτωση παρατηρούμε μια μεγάλη άνοδο της μεθόδου $BC_{a,j}$ ενώ στην δεύτερη περίπτωση παρατηρούμε σχεδόν ταύτιση της πιθανότητας κάλυψης μεταξύ όλων των μεθόδων εκτός από την μέθοδο του Wilson.



Γράφημα 2.13

Πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 100$

2. Μέθοδος Bootstrap για Διωνυμικές τ.μ.



Γράφημα 2.14

Πιθανότητα κάλυψης των 95% ονομαστικών διαστημάτων εμπιστοσύνης για $n = 500$

2.8 Συμπεράσματα

Το γεγονός ότι οι bootstrap μέθοδοι κατασκευής διαστημάτων εμπιστοσύνης δεν απέδωσαν ήταν λίγο πολύ αναμενόμενο. Για αρχή, οι μέθοδοι αυτοί στηρίζονται στα ποσοστιαία σημεία της διωνυμικής κατανομής για την κατασκευή των διαστημάτων εμπιστοσύνης και επομένως δεν περιμέναμε να έχουν καλή απόδοση. Δεύτερον, όταν το αρχικό μας δείγμα αποτελείται μόνο από επιτυχίες ή μόνο από αποτυχίες τότε παράγουν διαστήματα εμπιστοσύνης μηδενικού πλάτους και επομένως η πιθανότητα κάλυψης τους είναι μικρή.

Για τους παραπάνω λόγους προτάθηκε η τροποποιημένη bootstrap η οποία και επιλύει το πρόβλημα των διαστημάτων εμπιστοσύνης με μηδενικό πλάτος. Η πιθανότητα κάλυψης των bootstrap διαστημάτων εμπιστοσύνης βελτιώνεται σημαντικά, ωστόσο τα προβλήματα συνεχίζουν να υπάρχουν. Το πιο σημαντικό ποιος είναι ο ρυθμός σύγκλισης των διαστημάτων

2^ο Κεφάλαιο

αυτών στο ονομαστικό επίπεδο. Η κυματοειδής σύγκλιση που φαίνεται να υπάρχει δημιουργεί κάποια ερωτηματικά ως προς τις ιδιότητες τους και οδηγεί σε μεγάλη τιμή της ρίζας του μέσου τετραγωνικού σφάλματος σε σύγκριση με τις παραδοσιακές μεθόδους.

Τα 90% ονομαστικού επιπέδου τροποποιημένα bootstrap διαστήματα εμπιστοσύνης παρουσιάζουν μια περίεργη συμπεριφορά. Η μέση πιθανότητα κάλυψης είναι πολλές φορές μικρότερη από το ονομαστικό επίπεδο και όπως φάνηκε από τα γραφήματα της πιθανότητας κάλυψης υπάρχουν σχεδόν πάντα τιμές της πιθανότητας επιτυχίας τέτοιες ώστε η πιθανότητα κάλυψης να είναι αρκετά μικρή.

Τα 95% ονομαστικού επιπέδου τροποποιημένα bootstrap διαστήματα εμπιστοσύνης έχουν σαφώς καλύτερη συμπεριφορά από ότι τα αντίστοιχα 90%. Η μέση πιθανότητα κάλυψης είναι αρκετά κοντά στο ονομαστικό επίπεδο. Ειδικότερα η μέθοδος BP είναι η πιο σταθερή από τις υπόλοιπες καθώς έχει το μικρότερο μέσο αναμενόμενο πλάτος και πλησιάζει στο ονομαστικό επίπεδο καλύτερα από ότι οι υπόλοιπες μέθοδοι bootstrap. Άλλο ένα πλεονέκτημα της είναι ότι είναι συγκρίσιμη με τις παραδοσιακές μεθόδους. Η μέση πιθανότητα κάλυψης είναι μικρότερη από αυτή της μεθόδου των Agresti και Coull ενώ για ορισμένες τιμές του n είναι μικρότερη και από την μέθοδο του Wilson, ακόμα και αν το δείγμα μας είναι μικρού μεγέθους. Επίσης, από τα γραφήματα της πιθανότητας κάλυψης φαίνεται ότι η μικρότερη τιμή της είναι μεγαλύτερη από την μικρότερη πιθανότητα κάλυψης της μεθόδου Wilson. Το μειονέκτημα της είναι φυσικά η μεγαλύτερη τιμή της ρίζας του μέσου τετραγωνικού σφάλματος κάτι το οποίο εξηγείται από την κυματοειδή συμπεριφορά της μέσης πιθανότητας κάλυψης.

Συνοψίζοντας τα παραπάνω θεωρούμε ότι η τροποποιημένη bootstrap BP είναι μια μέθοδος που αξίζει να εξεταστεί καλύτερα αφού έχει ορισμένα πλεονεκτήματα έναντι των παραδοσιακών μεθόδων.

3^ο Κεφάλαιο

Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

3.0 Εισαγωγή

Στο προηγούμενο κεφαλαίο εξετάσαμε την εφαρμογή της μεθόδου bootstrap σε διωνυμικές τυχαίες μεταβλητές και την κατασκευή απλών διαστημάτων εμπιστοσύνης για διωνυμικά ποσοστά. Η διωνυμική κατανομή δεν είναι η μοναδική κατανομή από την οποία λαμβάνουμε διακριτά δεδομένα. Στις Παραγράφους 3.1 και 3.2 θα εξετάσουμε δυο ακόμα κατανομές, την πολυωνυμική κατανομή και την κατανομή Poisson, αντίστοιχα. Στην Παράγραφο 3.3 θα παρουσιάσουμε τα κύρια δειγματοληπτικά σχήματα που χρησιμοποιούνται στην ανάλυση κατηγορικών δεδομένων. Η εφαρμογή της μεθόδου bootstrap σε πολυωνυμικές τ.μ. παρουσιάζεται στην Παράγραφο 3.4. Στις Ενότητες 3.5 και 3.6 παρουσιάζονται τα ταυτόχρονα διαστήματα εμπιστοσύνης για πολυωνυμικά ποσοστά και για αντιθέσεις πολυωνυμικών ποσοστών, αντίστοιχα. Τα συμπεράσματα για τα εν λόγω bootstrap διαστήματα εμπιστοσύνης συνοψίζονται στην Παράγραφο 3.7.

3.1 Πολυωνυμική Κατανομή

Η πολυωνυμική κατανομή είναι μια γενίκευση της διωνυμικής κατανομής. Στην διωνυμική κατανομή θεωρούμε ότι εκτελούμε ένα πείραμα ανεξάρτητα και πανομοιότυπα n φορές με σταθερή πιθανότητα επιτυχίας και ότι σε κάθε δοκιμή υπάρχουν δυο δυνατά αποτελέσματα. Στην πολυωνυμική κατανομή θεωρούμε πάλι ότι εκτελούμε n ανεξάρτητες και πανομοιότυπες δοκιμές, όμως σε κάθε δοκιμή έχουμε k δυνατά αποτελέσματα, με $k > 2$, κάθε ένα από τα οποία έχει σταθερή πιθανότητα επιτυχίας.

Πιο αυστηρά, το τυχαίο διάνυσμα $\mathbf{Y}_{\%} = (Y_1, \dots, Y_k)$ θα λέμε ότι ακολουθεί την πολυωνυμική κατανομή με παραμέτρους n και $\boldsymbol{\pi}_{\%} = (p_1, \mathbf{K}, p_k)$ και θα συμβολίζουμε $\mathbf{Y}_{\%} \sim Multi(n, \boldsymbol{\pi}_{\%})$, αν η συνάρτηση πυκνότητας πιθανότητάς του είναι η :

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

$$\text{Prob}\{(Y_1, \dots, Y_{k-1}) = (y_1, \dots, y_{k-1})\} = \frac{n!}{\prod_{i=1}^k y_i!} \prod_{i=1}^k p_i^{y_i}$$

όπου οι μεταβλητές y_1, \dots, y_k παίρνουν τις τιμές $0, \dots, n$ υπό την δέσμευση ότι ισχύει

$$\sum_{i=1}^k y_i = n \text{ και ότι οι παράμετροι } p_1, \dots, p_k \text{ είναι θετικοί αριθμοί με } \sum_{i=1}^k p_i = 1.$$

Είναι προφανές ότι η περιθώρια κατανομή της Y_i είναι η διωνυμική $B(n, p_i)$, για κάθε $i \leq n$. Επομένως η αναμενομένη τιμή του αριθμού των παρατηρήσεων της i -κατηγορίας είναι $E[Y_i] = np_i$ και η διασπορά της είναι $\text{Var}[Y_i] = np_i(1-p_i)$. Η συνδιασπορά του αριθμού των παρατηρήσεων της i -οστής και της j -οστής κατηγορίας είναι $\text{cov}[Y_i, Y_j] = -np_i p_j$ για $i, j = 1, \dots, k$ και $i \neq j$. Το γεγονός ότι η συνδιασπορά είναι αρνητική είναι λογικό αν σκεφτούμε ότι οι παρατηρήσεις μας αθροίζουν στο n . Επομένως όσο περισσότερες παρατηρήσεις έχουμε στην i -οστή κατηγορία τόσο λιγότερες θα έχουμε στην j -οστή κατηγορία.

Όπως και στην περίπτωση της διωνυμικής κατανομής, οι πιθανότητες επιτυχίας κάθε κατηγορίας δεν είναι γνωστές εκ των προτέρων και συνεπώς πρέπει να εκτιμηθούν. Ο εκτιμητής μέγιστης πιθανοφάνειας για την πιθανότητα επιτυχίας p_i της i -οστής κατηγορίας είναι ο

$$\hat{p}_i = p_i = \frac{y_i}{n}, i = 1, \dots, k$$

δηλαδή το δειγματικό ποσοστό της i -οστής κατηγορίας.

3.2 Κατανομή Poisson

Τόσο στην διωνυμική όσο και στην πολυωνυμική κατανομή το πλήθος των πιθανών αποτελεσμάτων ή κατηγοριών είναι γνωστό εκ των προτέρων. Ωστόσο, αυτό δεν συμβαίνει πάντα στην ανάλυση των κατηγορικών δεδομένων καθώς είναι πιθανόν να μην είμαστε σε θέση να διακρίνουμε όλες τις δυνατές περιπτώσεις. Για παράδειγμα μπορεί να ενδιαφερόμαστε για τον καθημερινό αριθμό αυτοκινητιστικών δυστυχημάτων σε έναν συγκεκριμένο αυτοκινητόδρομο. Προφανώς γνωρίζουμε ότι ο αριθμός αυτός θα είναι μη αρνητικός ακέραιος αλλά δεν μπορούμε να γνωρίζουμε την μέγιστη δυνατή του τιμή.

3^ο Κεφάλαιο

Όταν η τυχαία μεταβλητή Y μετράει τον αριθμό των εμφανίσεων κάποιου γεγονότος θα λέμε ότι ακολουθεί την κατανομή Poisson με αναμενομένη τιμή I και θα συμβολίζουμε $Y \sim P(I)$, αν έχει συνάρτηση πυκνότητας πιθανότητας:

$$\text{Prob}\{Y = y\} = \frac{e^{-I} I^y}{y!}, \quad y=0,1,2,\dots$$

όπου I παράμετρος τέτοια ώστε $I > 0$.

Η αναμενομένη τιμή της τυχαίας μεταβλητής Y που ακολουθεί την κατανομή Poisson $P(I)$ είναι $E[Y] = I$ και η διασπορά της είναι ίση με $\text{Var}[Y] = I$. Ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου I είναι η δειγματική μέση τιμή, δηλαδή $\hat{I} = \bar{Y}$.

Η κατανομή Poisson μπορεί να θεωρηθεί και ως μια προσέγγιση της διωνυμικής κατανομής. Αν ο αριθμός των επιτυχιών στις n δοκιμές μπορεί να περιγράψει από μια διωνυμική τυχαία μεταβλητή, με αριθμό δοκιμών $n \rightarrow \infty$ και πιθανότητα επιτυχίας $p \rightarrow 0$ τέτοιοι ώστε $np \rightarrow I$, τότε η κατανομή των επιτυχιών είναι Poisson με αναμενομένη τιμή I . Αυτή η προσέγγιση της Poisson κατανομής εξηγεί τον λόγο για τον οποίο η Poisson κατανομή είναι μια υποψήφια κατανομή της τυχαίας μεταβλητής που μετράει τον αριθμό εμφανίσεων σπάνιων γεγονότων.

Επίσης η Poisson κατανομή μπορεί να συσχετιστεί με την πολυωνυμική κατανομή. Ας υποθέσουμε ότι οι έχουμε K ανεξάρτητες τυχαίες μεταβλητές Y_1, \mathbf{K}, Y_K οι οποίες ακολουθούν την Poisson κατανομή $P(I_i)$ για $i=1, \dots, K$. Η τυχαία μεταβλητή $n = \sum_{i=1}^K Y_i$ ακολουθεί και αυτή Poisson κατανομή με παράμετρο $\sum_{i=1}^K I_i$. Αν θεωρήσουμε ότι το άθροισμα των Poisson τυχαίων μεταβλητών δεν είναι τυχαίο αλλά σταθερό και ίσο με n , τότε οι τυχαίες μεταβλητές Y_1, \mathbf{K}, Y_K δεν θα ακολουθούν Poisson κατανομές αφού κάθε ένα από τα Y_1, \mathbf{K}, Y_K δεν μπορεί να είναι μεγαλύτερο από το n . Επομένως δοθέντος του n οι Y_1, \mathbf{K}, Y_K δεν είναι ανεξάρτητες. Η δεσμευμένη κατανομή των Y_1, \mathbf{K}, Y_K δοθέντος ότι $n = \sum_{i=1}^K Y_i$ βρίσκεται από το θεώρημα του Bayes ως εξής:

$$\text{Prob}\left\{(Y_1 = y_1, \dots, Y_K = y_K) \mid \sum_{i=1}^K Y_i = n\right\} = \frac{\text{Prob}\{(Y_1 = y_1, \dots, Y_K = y_K)\}}{\text{Prob}\left\{\sum_{i=1}^K Y_i = n\right\}}$$

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

$$= \frac{\prod_{i=1}^K \frac{e^{-l_i} l_i^{y_i}}{y_i!}}{e^{\sum_{i=1}^K l_i} \left(\sum_{i=1}^K l_i \right)^n} = \frac{n!}{\prod_{i=1}^K y_i!} \prod_{i=1}^K \left(\frac{l_i}{\sum_{i=1}^K l_i} \right)^{y_i}$$

Θέτοντας

$$p_j = \frac{l_j}{\sum_{i=1}^K l_i}, j = 1, \dots, K$$

παρατηρούμε η δεσμευμένη κατανομή των Y_1, \mathbf{K}, Y_K δοθέντος του $n = \sum_{i=1}^K Y_i$ κατανομή είναι η πολυωνυμική κατανομή $Multi(n, \boldsymbol{\pi})$, με $\boldsymbol{\pi} = (p_1, \dots, p_K)$.

3.3 Δειγματοληπτικά Σχήματα

Ένα πολύ σημαντικό μέρος της ανάλυσης κατηγορικών δεδομένων βασίζεται στην ανάλυση πινάκων συνάφειας που αποτελεί και τον συνηθέστερο τρόπο παρουσίασης των κατηγορικών δεδομένων. Για λόγους απλότητας στην παρουσίαση, θα περιοριστούμε στους διδιάστατους πίνακες συνάφειας $I \times J$. Η γενική μορφή ενός $I \times J$ πίνακα είναι η παρακάτω:

$$\begin{pmatrix} n_{11} & n_{12} & \mathbf{L} & n_{1J} \\ n_{21} & n_{22} & \mathbf{L} & n_{2J} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ n_{I1} & n_{I2} & \mathbf{L} & n_{IJ} \end{pmatrix}$$

όπου n_{ij} είναι η συχνότητα του (i, j) κελιού. Τα πιθανά δειγματοληπτικά σχήματα ενός πίνακα συνάφειας είναι το πολυωνυμικό δειγματοληπτικό σχήμα, το δειγματοληπτικό σχήμα γινόμενου ανεξάρτητων πολυωνυμικών και το Poisson δειγματοληπτικό σχήμα.

3.3.1 Πολυωνυμικό Δειγματοληπτικό Σχήμα

Στο πολυωνυμικό δειγματοληπτικό σχήμα θεωρούμε ότι το συνολικό μέγεθος του δείγματος n είναι σταθερό ενώ τα περιθώρια αθροίσματα γραμμών και στηλών δεν είναι σταθερά. Με άλλα λόγια, θεωρούμε ότι το διάνυσμα των συχνοτήτων των κελιών του πίνακα $(n_{11}, \mathbf{K}, n_{IJ})$ ακολουθεί την $Multi(n, (p_{11}, \mathbf{K}, p_{IJ}))$. Επομένως η από κοινού συνάρτηση πυκνότητας των συχνοτήτων $(n_{11}, \mathbf{K}, n_{IJ})$ δοθέντος του n είναι:

$$\text{Prob}\{n_{11}, \mathbf{K}, n_{IJ} | n\} = \frac{n!}{\prod_{i=1}^I \prod_{j=1}^J n_{ij}!} \prod_{i=1}^I \prod_{j=1}^J p_{ij}^{n_{ij}}$$

Ο εκτιμητής μέγιστης πιθανοφάνειας των p_{ij} είναι ο $\hat{p}_{ij} = p_{ij} = \frac{n_{ij}}{n}, \forall i=1, \dots, I, j=1, \dots, J$.

3.3.2 Δειγματοληπτικό Σχήμα Γινομένου Πολυωνυμικών

Στο δειγματοληπτικό σχήμα γινομένου πολυωνυμικών θεωρούμε ότι τα περιθώρια αθροίσματα των γραμμών είναι σταθερά. Με άλλα λόγια θεωρούμε ότι οι γραμμές του πίνακα ακολουθούν ανεξάρτητες πολυωνυμικές κατανομές. Έτσι το διάνυσμα των συχνοτήτων των κελιών κάθε γραμμής του πίνακα $(n_{i1}, \mathbf{K}, n_{iJ})$ ακολουθεί την πολυωνυμική κατανομή $Multi(n_i, (p_{i1}, \mathbf{K}, p_{iJ}))$, για κάθε $i=1, \dots, I$, με $\sum_{j=1}^J p_{ji} = 1$ και $n_i = \sum_{j=1}^J n_{ij}$.

Επομένως η από κοινού συνάρτηση πυκνότητας των συχνοτήτων της i -γραμμής είναι :

$$\text{Prob}\{n_{i1}, \mathbf{K}, n_{iJ}\} = \frac{n_i!}{\prod_{j=1}^J n_{ij}!} \prod_{j=1}^J p_{ji}^{n_{ij}}$$

Ο εκτιμητής μέγιστης πιθανοφάνειας των p_{ji} είναι ο $\hat{p}_{ji} = p_{ji} = \frac{n_{ij}}{n_i}, \forall i=1, \dots, I, j=1, \dots, J$

Αυτό το δειγματοληπτικό σχήμα είναι γνωστό και ως δειγματοληπτικό σχήμα γινομένου ανεξάρτητων πολυωνυμικών.

3.3.3 Poisson Δειγματοληπτικό Σχήμα

Στο Poisson δειγματοληπτικό σχήμα θεωρούμε ότι οι συχνότητες n_{ij} , ακολουθούν ανεξάρτητες Poisson κατανομές με παραμέτρους I_{ij} , για $i=1, \dots, I$ και $j=1, \dots, J$ και

$\sum_{i=1}^I \sum_{j=1}^J I_{ij} = n$. Στην Παράγραφο 3.2 είδαμε ότι, δοθέντος του n , οι τυχαίες μεταβλητές

$n_{ij} \sim P(I_{ij})$, για $i=1, \dots, I$ και $j=1, \dots, J$, ακολουθούν πολυωνυμική κατανομή. Επομένως η

από κοινού συνάρτηση πυκνότητας των $(n_{11}, \mathbf{K}, n_{IJ})$ δοθέντος του n είναι η:

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

$$\text{Pr ob}\{n_{11}, \mathbf{K}, n_{JJ}\} = \frac{n!}{\prod_{i=1}^I \prod_{j=1}^J n_{ij}!} \prod_{i=1}^I \prod_{j=1}^J \left(\frac{I_{ij}}{\sum_{i=1}^I \sum_{j=1}^J I_{ij}} \right)^{n_{ij}}$$

Θέτοντας

$$p_{ij} = \frac{I_{ij}}{\sum_{i=1}^I \sum_{j=1}^J I_{ij}} = \frac{I_{ij}}{n}$$

βρίσκουμε ότι ο εκτιμητής μέγιστης πιθανοφάνειας του p_{ij} είναι ο $\hat{p}_{ij} = \frac{n_{ij}}{n}$.

Και τα τρία δειγματοληπτικά σχήματα βασίζονται στην πολυωνυμική κατανομή. Επομένως είναι λογικό να εξετάσουμε την εφαρμογή της μεθόδου bootstrap σε πολυωνυμικές κατανομές.

3.4 Η Μέθοδος Bootstrap για την Πολυωνυμική Κατανομή

Έστω ότι εκτελούμε ένα n δοκιμές και ότι τα δυνατά αποτελέσματα σε κάθε δοκιμή είναι r στο πλήθος. Τότε το αρχικό μας διάνυσμα θα είναι ένα n -διάστατο διάνυσμα της μορφής:

$$\left(\underbrace{1, \dots, 1}_{n_1 \text{-for } \mathcal{E}}, \dots, \underbrace{r, \dots, r}_{n_r \text{-for } \mathcal{E}} \right)$$

όπου τα n_1, \dots, n_r είναι οι συχνότητες των κατηγοριών $1, \dots, r$, αντίστοιχα.

Η εμπειρική κατανομή F_n θα δίνει σε κάθε παρατήρηση βάρη ίσα με το δειγματικό ποσοστό της κάθε κατηγορίας που ανήκει η παρατήρηση. Επομένως αν $\{x_1^*, \mathbf{K}, x_n^*\}$ είναι ένα bootstrap σύνολο δεδομένων από την F_n και $\{n_1^*, \mathbf{K}, n_r^*\}$ είναι η αναλογία bootstrap των $\{n_1, \dots, n_r\}$, τότε η bootstrap δειγματική κατανομή πυκνότητας πιθανότητας των $\{n_1^*, \mathbf{K}, n_r^*\}$ είναι η:

$$h_{BOOT} = \text{Prob}_* \left\{ (n_1^* = y_1, \mathbf{K}, n_r^* = y_r, \hat{F}) \right\} = \frac{n!}{\prod_{i=1}^r y_i!} \prod_{i=1}^r \left(\frac{n_i}{n} \right)^{y_i} \quad (3.1)$$

Δηλαδή, το (n_1^*, \dots, n_r^*) ακολουθεί την $Multi(n, (p_1, \mathbf{L}, p_r))$ όπου $\hat{p}_i = p_i = \frac{n_i}{n}$, για κάθε $i = 1, \dots, r$, είναι το δειγματικό ποσοστό της i -οστής κατηγορίας. Επομένως, η εφαρμογή της μη παραμετρικής μεθόδου bootstrap οδηγεί σε ακριβή εκτίμηση της δειγματικής κατανομής

3^ο Κεφάλαιο

του αρχικού δείγματος χρησιμοποιώντας ως εκτιμητές των πιθανοτήτων κάθε κατηγορίας τον εκτιμητή μέγιστης πιθανοφάνειας.

Είναι εύκολο να δείξουμε ότι αν χρησιμοποιήσουμε την παραμετρική μέθοδο bootstrap και ως εκτιμητή των πιθανοτήτων των κατηγοριών τον εκτιμητή μέγιστης πιθανοφάνειας θα καταλήξουμε και πάλι στην σχέση (3.1).

Όπως και στην περίπτωση της εφαρμογής της μεθόδου bootstrap σε διωνυμικές τυχαίες μεταβλητές έτσι και σε πολυωνυμικά διανύσματα, ο εκτιμητής bootstrap της δειγματικής κατανομής ταυτίζεται με τον εκτιμητή της δειγματικής κατανομής που προκύπτει από τις κλασικές μεθόδους, ανεξάρτητα από την χρήση παραμετρικής ή μη παραμετρικής μεθόδου bootstrap, όταν στην μη παραμετρική μέθοδο bootstrap χρησιμοποιούμε ως εκτιμητή των πιθανοτήτων κάθε κατηγορίας τον εκτιμητή μέγιστης πιθανοφάνειας.

Στο προηγούμενο κεφάλαιο εξετάσαμε την κατασκευή των απλών διαστημάτων εμπιστοσύνης μέσω της bootstrap κατανομής. Από την Παράγραφο 3.3.1 ξέρουμε ότι οι περιθώριες κατανομές των συχνοτήτων ακολουθούν διωνυμική κατανομή. Επομένως η συμπερασματολογία για μια από τις κατηγορίες μέσω της μεθόδου bootstrap έχει καλυφθεί στο προηγούμενο κεφάλαιο. Αυτό που έχει ενδιαφέρον είναι να εξετάσουμε την κατασκευή χωρίων εμπιστοσύνης για όλες τις πιθανότητες των r κατηγοριών και ταυτόχρονων διαστημάτων εμπιστοσύνης για αντιθέσεις ποσοστών μέσω της μεθόδου bootstrap.

3.5 Ταυτόχρονες Περιοχές Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

Ας υποθέσουμε ότι έχουμε έναν πολυωνυμικό πληθυσμό μεγέθους n . Έστω p_i , $i=1, \dots, r$, είναι η πιθανότητα της i -οστής κατηγορίας του πληθυσμού, με $p_i > 0$ και $\sum_{i=1}^r p_i = 1$. Ας είναι n_i ο αριθμός των ατόμων που emπίπτουν στην i -οστή κατηγορία και p_i ο εκτιμητής μέγιστης πιθανοφάνειας της πιθανότητας p_i , για $i=1, \dots, r$.

Ενδιαφερόμαστε για την κατασκευή μιας $100(1-a)\%$ ταυτόχρονης περιοχής εμπιστοσύνης για το διάνυσμα (p_1, \mathbf{K}, p_r) . Δηλαδή, θέλουμε να κατασκευάσουμε μια περιοχή που να περιέχει όλα τα p_i , $i=1, \dots, r$, με πιθανότητα $1-a$, ταυτόχρονα. Είναι γνωστό, ότι αν κατασκευάσουμε $100(1-a)\%$ απλά διαστήματα εμπιστοσύνης για κάθε p_i , $i=1, \dots, r$, και παρουσιάσουμε ως $100(1-a)\%$ ταυτόχρονη περιοχή εμπιστοσύνης την

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

περιοχή που δημιουργείται από τα $100(1-a)\%$ απλά διαστήματα εμπιστοσύνης, τότε η περιοχή αυτή θα έχει συντελεστή εμπιστοσύνης πολύ μικρότερο από $1-a$ (βλ. Κουνιάς 2002, Κεφαλαίο 2).

3.5.1 Κλασσική Προσέγγιση

Η ανισότητα του Bonferroni παίζει σημαντικό ρόλο στην κατασκευή $100(1-a)\%$ ταυτόχρονων διαστημάτων εμπιστοσύνης για παραμέτρους. Για παράδειγμα είναι αρκετά γνωστή η χρήση της για την κατασκευή ταυτόχρονων διαστημάτων εμπιστοσύνης για τις παραμέτρους της γραμμικής παλινδρόμησης.

Την μέθοδο του Bonferroni μπορούμε να την χρησιμοποιήσουμε και για την κατασκευή $100(1-a)\%$ ταυτόχρονων διαστημάτων εμπιστοσύνης για τα πολυωνυμικά ποσοστά $(\mathbf{p}_1, \mathbf{K}, \mathbf{p}_r)$. Έστω ότι το r δηλώνει τον αριθμό των κελιών, τότε μια μορφή των $100(1-a)\%$ Bonferroni ταυτόχρονων διαστημάτων εμπιστοσύνης είναι

$$\left[p_i \in p_i \pm t_{n-1; a/2r} \sqrt{\frac{p_i(1-p_i)}{n} + \frac{1}{2n}}, \forall i=1, \dots, r \right] \quad (3.2)$$

όπου $t_{n-1; a/2r}$ είναι το $100(a/2r)\%$ ποσοστιαίο σημείο της t κατανομής με $n-1$ βαθμούς ελευθερίας. Υπάρχουν αρκετές παραλλαγές των Bonferroni διαστημάτων. Ωστόσο εδώ θα χρησιμοποιήσουμε την κατανομή t με διόρθωση συνέχειας $1-a$ (βλ. Κουνιάς 2002, Κεφαλαίο 3).

Το πρόβλημα των Bonferroni ταυτόχρονων διαστημάτων εμπιστοσύνης είναι ότι η περιοχή εμπιστοσύνης που δίνεται είναι συντηρητική. Αυτό σημαίνει ότι η περιοχή εμπιστοσύνης δεν περιέχει τα ποσοστά $(\mathbf{p}_1, \mathbf{K}, \mathbf{p}_r)$ με πιθανότητα ακριβώς ίση με $1-a$ αλλά μεγαλύτερη (βλ. Κουνιάς 2002, Κεφαλαίο 2).

3.5.2 Bootstrap Προσέγγιση

Οι Jeong and Jhun (2000) πρότειναν την κατασκευή $100(1-a)\%$ bootstrap ταυτόχρονων περιοχών εμπιστοσύνης για το $(\mathbf{p}_1, \mathbf{K}, \mathbf{p}_r)$ μέσω του στατιστικού καλής προσαρμογής του

Pearson, $X^2 = n \sum_{i=1}^r \frac{(p_i - p_i)^2}{p_i}$ και της εκτίμησης bootstrap του. Πιο συγκεκριμένα πρότειναν

την αντικατάσταση της $100(1-a)\%$ περιοχής εμπιστοσύνης

$$\left[(p_1, \mathbf{K}, p_r) : n \sum_{i=1}^r \frac{(p_i - p_i)^2}{p_i} \leq c_{r-1; 1-a}^2 \right] \quad (3.3)$$

από την $100(1-a)\%$ bootstrap ταυτόχρονη περιοχή εμπιστοσύνης

$$\left[(p_1, \mathbf{K}, p_r) : n \sum_{i=1}^r \frac{(p_i^* - p_i)^2}{p_i^*} \leq c_{1-a}^{2*} \right] \quad (3.4)$$

όπου p_i^* , ($i=1, \dots, r$), δηλώνει τον εκτιμητή μέγιστης πιθανοφάνειας για την παραμετρική μέθοδο bootstrap και c_{1-a}^{2*} είναι το $100(1-a)$ ποσοστιαίο σημείο της bootstrap κατανομής

$$X^{2*} = n \sum_{i=1}^r \frac{(p_i^* - p_i)^2}{p_i^*}, \text{ δοθέντων των } (p_1, \mathbf{K}, p_r).$$

Αργότερα οι Morales et al. (2004) χρησιμοποίησαν την οικογένεια του power-divergence στατιστικού για να προσδιορίσουν την περιοχή εμπιστοσύνης (3.4). Η οικογένεια του power-divergence στατιστικού δίνεται από την σχέση:

$$CR(I) = \frac{2}{I(I+1)} \sum_{i=1}^r n_i \left[\left(\frac{n_i}{m_i} \right)^I - 1 \right] \quad (3.5)$$

όπου $-\infty < I < \infty$ και m_i είναι η αναμενομένη συχνότητα κάτω από κάποιο μοντέλο. Οι Cressie and Read (1984) απέδειξαν ότι το $CR(I) \sim X_{r-1}^2$ ασυμπτωτικά. Η οικογένεια του power-divergence στατιστικού μας δίνει για διάφορες τιμές του I γνωστά στατιστικά. Για παράδειγμα, για $I = -2$ η σχέση (3.5) μας δίνει το τροποποιημένο X^2 του Newman, για $I = 0$ δίνει το γνωστό G^2 ενώ για $I = 1$ το X^2 στατιστικό. Έχει αποδειχθεί (Cressie and Read, 1984) ότι για τιμές του I κοντά στο $2/3$ η σχέση (3.5) έχει την καλύτερη σύγκλιση. Το στατιστικό που προκύπτει για $I = 2/3$ λέγεται στατιστικό των Cressie και Read.

Αν συμβολίσουμε με $CR^*(I)$ την αναλογία bootstrap του $CR(I)$ για δοσμένη τιμή του I και με $T_{1-a}^*(I)$ το $1-a$ ποσοστιαίο σημείο της bootstrap κατανομής $CR^*(I)$ τότε οι οικογένεια των ταυτόχρονων διαστημάτων εμπιστοσύνης δίνεται από τον τύπο

$$\left[(p_1, \mathbf{K}, p_r) : CR(I) \leq T_{1-a}^*(I) \right]$$

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

Οι Morales, Pardo και Santamaría απέδειξαν ότι τα ταυτόχρονα διαστήματα εμπιστοσύνης δίνουν καλύτερες τιμές όταν το $I \in [0, 2]$.

Τέλος οι Jeong and Jhun (2000) πρότειναν την κατασκευή ενός ταυτόχρονου συνόλου εμπιστοσύνης για το (p_1, \mathbf{K}, p_r) μέσω της ποσότητας

$$T_{(r)} = \max_{1 \leq i \leq r} \left[|T_i| = \frac{\sqrt{n} |p_i - p_i|}{\sqrt{p_i(1-p_i)}} \right]$$

Ωστόσο τα T_i είναι εξαρτημένα και επομένως μια κλειστή μορφή της δειγματικής κατανομής του $T_{(r)}$ είναι δύσκολο να βρεθεί. Μπορούμε όμως να πάρουμε έναν εκτιμητή bootstrap της δειγματικής κατανομής του από την δεσμευμένη ποσότητα

$$T_{(r)}^* = \max_{1 \leq i \leq r} \left[|T_i^*| = \frac{\sqrt{n} |p_i^* - p_i|}{\sqrt{p_i^*(1-p_i^*)}} \right]$$

Χρησιμοποιώντας τον εκτιμητή bootstrap της δειγματικής κατανομής του $T_{(r)}$, ως είναι $T_{i;1-a}^*$ το $100(1-a)\%$ ποσοστιαίο σημείο της bootstrap κατανομής του $T_{(r)}^*$. Τότε ένα $100(1-a)\%$ ταυτόχρονο σύνολο εμπιστοσύνης για το (p_1, \mathbf{K}, p_r) δίνεται από την περιοχή:

$$\left[p_i \in p_i \pm T_{i;1-a}^* \sqrt{\frac{p_i(1-p_i)}{n}}, \forall i = 1, \dots, r \right] \quad (3.6)$$

Οι σχέσεις (3.3) και (3.4) δικαιολογούνται ασυμπτωτικά από τον Singh (1981).

3.5.3 Παράδειγμα Ταυτόχρονων Διαστημάτων Εμπιστοσύνης

Οι Jeong and Jhun (2000), απέδειξαν ότι η πιθανότητα κάλυψης των bootstrap ταυτόχρονων διαστημάτων εμπιστοσύνης της μορφής (3.6) είναι ικανοποιητική και ότι κατορθώνει να είναι κοντά στο ονομαστικό επίπεδο. Ωστόσο δεν αναφέρουν καθόλου την περίπτωση όπου ένα ή περισσότερα πολυωνυμικά ποσοστά έχουν την αναλογία bootstrap τους ίση με το μηδέν σε κάποια bootstrap επανάληψη. Στην περίπτωση αυτή, η εκτίμηση bootstrap της τυπικής απόκλισης του ποσοστού μηδενίζεται και επομένως η ποσότητα $T_{(r)}^*$ και κατά συνέπεια το bootstrap ταυτόχρονο διάστημα εμπιστοσύνης δεν μπορεί να οριστεί. Η πιθανότητα να είναι η αναλογία bootstrap ενός τουλάχιστον ποσοστού ίση με το μηδέν είναι μεγάλη, όταν τουλάχιστον ένα από τα ποσοστά είναι κοντά στο μηδέν ή το μέγεθος δείγματος είναι αρκετά

μικρό. Για να λειτουργήσει η μέθοδος που προτείνουν πρέπει το δείγμα να είναι αρκετά μεγάλο, για παράδειγμα μεγαλύτερο από 50, ή/και να μην υπάρχουν μικρές εκτιμήσεις των πολυωνυμικών ποσοστών ή όταν το μέγεθος δείγματος είναι μικρό, να μην υπάρχουν πολλές κατηγορίες.

Με βάση τα παραπάνω αποφασίστηκε να συγκριθούν η μέθοδος bootstrap (3.6) με την μέθοδο του Bonferroni (3.2). Για να πραγματοποιηθεί η σύγκριση πήραμε 4 πολυωνυμικά ποσοστά μέσω Simplex από την ομοιόμορφη $U(0,1)$ και στην συνέχεια πήραμε ένα δείγμα μεγέθους 40. Υπολογίζοντας από το αρχικό μας δείγμα τους εκτιμητές μέγιστης πιθανοφάνειας των ποσοστών βρήκαμε ότι $\hat{\pi}_{\%} = (0.15, 0.20, 0.30, 0.35)$. Από το διάνυσμα $\hat{\pi}_{\%}$, προσομοιώσαμε 1000 bootstrap σύνολα δεδομένων παραμετρικά από την $Multi(40, \hat{\pi}_{\%})$ προκειμένου να βρούμε τα bootstrap ταυτόχρονα διαστήματα εμπιστοσύνης (3.5) για τα 4 πολυωνυμικά ποσοστά. Στον Πίνακα 3.1 παρουσιάζονται τα 90% και 95% ονομαστικού επιπέδου ταυτόχρονα διαστήματα εμπιστοσύνης για τις δυο μεθόδους, αντίστοιχα. Παρατηρούμε ότι τα ταυτόχρονα διαστήματα εμπιστοσύνης της μεθόδου Bonferroni είναι πιο πλατιά από τα αντίστοιχα της μεθόδου bootstrap. Το συμπέρασμα είναι λογικό αν σκεφτούμε ότι τα bootstrap ταυτόχρονα διαστήματα εμπιστοσύνης κατασκευάζονται έτσι ώστε να έχουν ονομαστικό επίπεδο ακριβώς ίσο με το επιθυμητό ενώ τα Bonferroni έτσι ώστε να έχουν ονομαστικό επίπεδο τουλάχιστον όσο το επιθυμητό.

Πίνακας 3.1

Ταυτόχρονα διαστήματα εμπιστοσύνης των μεθόδων bootstrap και Bonferroni 90% και 95% ονομαστικού επιπέδου. Στις τελευταίες στήλες δίνεται το πλάτος των διαστημάτων εμπιστοσύνης.

$1-a$	i	\hat{p}_i	\hat{p}_i^*	Bootstrap		Bonferroni		Πλάτος Bootstrap	Πλάτος Bonferroni
0.90	1	0.15	0.1494	0.0286	0.2899	0.0224	0.2761	0.2537	0.2613
	2	0.20	0.2008	0.0615	0.3535	0.0558	0.3392	0.2834	0.2920
	3	0.30	0.3108	0.1511	0.4906	0.1461	0.4757	0.3296	0.3395
	4	0.35	0.3389	0.1784	0.5264	0.1735	0.5114	0.3379	0.3480
0.95	1	0.15	0.1494	0.0057	0.2927	0.0152	0.3032	0.2870	0.2880
	2	0.20	0.2008	0.0372	0.3578	0.0466	0.3684	0.3206	0.3218
	3	0.30	0.3108	0.1245	0.4973	0.1338	0.5079	0.3728	0.3741
	4	0.35	0.3389	0.1513	0.5335	0.1607	0.5442	0.3822	0.3835

3.6 Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Αντιθέσεις σε Γινόμενο Πολυωνυμικών

Έστω ότι το δειγματοληπτικό μας πλαίσιο είναι το γινόμενο πολυωνυμικών. Ας είναι p_{ij} η πιθανότητα της j -οστής κατηγορίας του i -πληθυσμού. Για ένα δείγμα μεγέθους n_{\square} από τον

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνυμικά Ποσοστά

i -πληθυσμός είναι n_{ij} ο αριθμός των αποκρίσεων της j -κατηγορίας του i -πληθυσμού. Ο εκτιμητής μέγιστης πιθανοφάνειας των πιθανοτήτων p_{ij} είναι $p_{ij} = n_{ij} / n_{i\cdot}$.

Ας υποθέσουμε ότι ενδιαφερόμαστε για την κατασκευή ταυτόχρονων διαστημάτων εμπιστοσύνης για όλες τις γραμμικές συναρτήσεις $q = \sum_{i=1}^r \sum_{j=1}^c b_{ij} p_{ij}$, με $\sum_{i=1}^r b_{ij} = 0$ για $j=1, \dots, c$. Το q εκτιμάται από το $\hat{q} = \sum_{i=1}^r \sum_{j=1}^c b_{ij} p_{ij}$.

Ο Goodman (1964) βασιζόμενος στο F του Scheffe πρότεινε ως $100(1-a)\%$ ταυτόχρονα διαστήματα εμπιστοσύνης για τα q το

$$\hat{q} - \sqrt{c_{(r-1)(j-1);1-a}^2} s(\hat{q}) \leq q \leq \hat{q} + \sqrt{c_{(r-1)(j-1);1-a}^2} s(\hat{q}) \quad (3.7)$$

$$\text{όπου } s^2(\hat{q}) = \sum_{i=1}^r \frac{1}{n_i} \left(\sum_{j=1}^c (b_{ij})^2 p_{ij} - \left(\sum_{j=1}^c b_{ij} p_{ij} \right)^2 \right).$$

Το μειονέκτημα της μεθόδου αυτής είναι ότι η πιθανότητα είναι $1-a$ ότι όλες οι αντιθέσεις θα περιέχονται στα αντίστοιχα διαστήματα. Στην πράξη εμείς δεν ενδιαφερόμαστε για όλες τις αντιθέσεις παρά μόνο για ένα σύνολο αυτών.

Εστω ότι ενδιαφερόμαστε για την κατασκευή $100(1-a)\%$ ταυτόχρονων διαστημάτων εμπιστοσύνης για τις K αντιθέσεις q_1, \mathbf{K}, q_K . Τότε σύμφωνα με την μέθοδο Bonferroni, ένα $100(1-a)\%$ Bonferroni ταυτόχρονο διάστημα εμπιστοσύνης για τα q_1, \mathbf{K}, q_K δίνεται από την σχέση:

$$\hat{q}_k - z_k s(\hat{q}_k) \leq q_k \leq \hat{q}_k + z_k s(\hat{q}_k) \quad (3.8)$$

όπου z_k είναι το $100(1-b_k)\%$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής και $\sum_{k=1}^K b_k = a/2$. Το μειονέκτημα των Bonferroni διαστημάτων είναι ότι είναι συντηρητικά και όταν το K είναι πολύ μεγάλο είναι πάρα πολύ πλατιά. Τα διαστήματα εμπιστοσύνης του Goodman είναι πιο ισχυρά από ότι τα διαστήματα εμπιστοσύνης του Bonferroni αν ο αριθμός των αντιθέσεων είναι αρκετά μεγάλος.

Τα προβλήματα αυτά μπορούμε να τα λύσουμε μέσω της μεθόδου bootstrap. Οι Jeong and Jhun (2000) πρότειναν την χρήση της μεθόδου bootstrap για την κατασκευή ταυτόχρονων

διαστημάτων εμπιστοσύνης για ένα πεπερασμένο αριθμό των γραμμικών αντιθέσεων q_1, \mathbf{K}, q_K . Θεωρώντας την ποσότητα

$$B_{(K)} = \max_{1 \leq k \leq K} \left[\frac{|\hat{q}_1 - q_1|}{s(\hat{q}_1)}, \dots, \frac{|\hat{q}_K - q_K|}{s(\hat{q}_K)} \right]$$

Μπορούμε να χρησιμοποιήσουμε ως εκτιμητή της bootstrap δειγματικής κατανομής $B_{(K)}$ την ποσότητα

$$B_{(K)}^* = \max_{1 \leq k \leq K} \left[\frac{|\hat{q}_1^* - \hat{q}_1|}{s(\hat{q}_1^*)}, \dots, \frac{|\hat{q}_K^* - \hat{q}_K|}{s(\hat{q}_K^*)} \right]$$

όπου $\hat{q}_k^* = \sum_{i=1}^r \sum_{j=1}^c b_{ij}^k p_{ij}^*$ και $s^2(\hat{q}_k^*) = \sum_{i=1}^r \frac{1}{n_i} \left(\sum_{j=1}^c (b_{ij}^k)^2 p_{ij}^* - \left(\sum_{j=1}^c b_{ij}^k p_{ij}^* \right)^2 \right)$ με p_{ij}^* να είναι ο εκτιμητής μέγιστης πιθανοφάνειας για το bootstrap δείγμα.

Ο Beran (1988a, 1988b) είχε δείξει ότι η bootstrap προσέγγιση μπορεί να χρησιμοποιηθεί για την κατασκευή ταυτόχρονων διαστημάτων εμπιστοσύνης πεπερασμένου πλήθους γραμμικών συνδυασμών. Οι Jeong and Jhun (2000) εφαρμόζοντας στην ουσία τα αποτελέσματα του Beran πρότειναν την κατασκευή του $100(1-a)\%$ bootstrap ταυτόχρονου διαστήματος εμπιστοσύνης για τα q_1, \mathbf{K}, q_K

$$\hat{q}_k - s(\hat{q}_k)Q(1-a) \leq q_k \leq q_k + s(\hat{q}_k)Q(1-a), k = 1, \dots, K \quad (3.9)$$

όπου $Q(1-a)$ είναι το $100(1-a)\%$ ποσοστιαίο σημείο της bootstrap κατανομής $B_{(K)}^*$.

Συμπεραίνουμε λοιπόν ότι το μεγάλο πλεονέκτημα των ταυτόχρονων διαστημάτων εμπιστοσύνης μέσω της μεθόδου bootstrap είναι η κατασκευή διαστημάτων εμπιστοσύνης για τον αριθμό των αντιθέσεων που μας ενδιαφέρει. Επομένως δεν περιμένουμε να υπάρχει μεγάλη απώλεια στην πιθανότητα κάλυψης των διαστημάτων εμπιστοσύνης.

3.6.1 Παράδειγμα Ταυτόχρονων Διαστημάτων Εμπιστοσύνης Για Αντιθέσεις

Οι Jeong and Jhun (2000), απέδειξαν ότι η πιθανότητα κάλυψης των bootstrap ταυτόχρονων διαστημάτων εμπιστοσύνης της μορφής (3.9) είναι ικανοποιητική και ότι κατορθώνει να είναι κοντά στο ονομαστικό επίπεδο. Ωστόσο δεν αναφέρουν καθόλου την περίπτωση όπου τουλάχιστον μια αντίθεση μηδενίζεται. Στην περίπτωση αυτή, η εκτίμηση bootstrap της τυπικής απόκλισης της αντίθεσης μηδενίζεται και επομένως η ποσότητα $B_{(K)}^*$ και κατά

3. Ταυτόχρονα Διαστήματα Εμπιστοσύνης για Πολυωνμικά Ποσοστά

συνέπεια το bootstrap ταυτόχρονο διάστημα εμπιστοσύνης δεν μπορεί να οριστεί. Η πιθανότητα να είναι η αναλογία bootstrap μιας αντίθεσης ίση με το μηδέν είναι μικρή και εξαλείφεται όταν όλοι οι συντελεστές των αντιθέσεων είναι διάφοροι του μηδενός. Επιπλέον οι Jeong and Jhun (2000), λαμβάνουν μόνο συγκεκριμένες τιμές των συντελεστών των αντιθέσεων και θεωρούν ότι τα δείγματα των ανεξάρτητων πολυωνμικών είναι ίσου μεγέθους.

Με βάση τα παραπάνω αποφασίστηκε να συγκριθούν η μέθοδος bootstrap (3.9) με την μέθοδο του Bonferroni (3.7) και του Goodman (3.8). Θεωρήσαμε έναν πίνακα διάστασης 3x3 και για κάθε γραμμή πήραμε πολυωνμικά ποσοστά μέσω Simplex από την ομοιόμορφη $U(0,1)$. Επίσης, θεωρήσαμε ότι $(n_1, n_2, n_3) = (20, 25, 30)$. Οι πραγματικές και οι εκτιμώμενες πιθανότητές τους από το τυχαίο δείγμα φαίνονται στον Πίνακα 3.2.

Πίνακας 3.2

Οι πραγματικές (p_{ij}) και οι εκτιμώμενες (\hat{p}_{ij}) πιθανότητες του 3x3 πίνακα.

$p_{ij} =$	0.4693	0.1665	0.3642	$\hat{p}_{ij} =$	0.4500	0.2500	0.3000
	0.2910	0.3650	0.3440		0.4800	0.0800	0.4400
	0.3243	0.2452	0.4305		0.2333	0.2000	0.5667

Στην συνέχεια παίρνουμε 1000 τυχαία δείγματα από τις εκτιμώμενες πιθανότητες, και εφαρμόζουμε την μέθοδο bootstrap ώστε να εκτιμήσουμε τα διαστήματα εμπιστοσύνης για τις αντιθέσεις. Στον Πίνακα 3.3 δίνονται οι τιμές των συντελεστών των αντιθέσεων.

Πίνακας 3.3

Οι συντελεστές b_{ij}^K των τριών αντιθέσεων

K	1			2			3		
b_{ij}^K	1	2	3	1	2	3	1	2	3
1	1	2	3	1	2	3	0	0	0
2	-1	-2	-3	0	0	0	1	2	3
3	0	0	0	-1	-2	-3	-1	-2	-3

Πίνακας 3.4

Ταυτόχρονα διαστήματα εμπιστοσύνης για τις αντιθέσεις ονομαστικού επιπέδου 90% και 95%

$1-\alpha$		Goodman		Bonferroni		Bootstrap	
0.90	$q_1 \in$	-0.3139	0.0939	-0.2850	0.0650	-0.8243	0.6043
	$q_2 \in$	-0.6488	-0.3178	-0.6245	-0.5163	-1.0630	0.0964
	$q_3 \in$	-0.5398	-0.2068	-0.5163	0.2304	-0.9566	0.2099
0.95	$q_1 \in$	-0.3352	0.1152	-0.3029	0.0829	-0.9261	0.7061
	$q_2 \in$	-0.6661	-0.3006	-0.6399	-0.3268	-1.1457	0.1791
	$q_3 \in$	-0.5572	-0.1894	-0.5308	-0.2158	-1.0398	0.2931

Τα 90% και 95% ταυτόχρονα διαστήματα εμπιστοσύνης για τις αντιθέσεις παρουσιάζονται στον Πίνακα 3.4. Παρατηρούμε ότι τα ταυτόχρονα διαστήματα εμπιστοσύνης της μεθόδου bootstrap είναι πιο πλατιά από τα αντίστοιχα των μεθόδων Goodman και Bonferroni. Αυτό το έχουν παρατηρήσει και οι Jeong και Jhun, δίνοντας την εξήγηση ότι με την μέθοδο bootstrap κατασκευάζουμε ταυτόχρονα διαστήματα εμπιστοσύνης για το πλήθος των αντιθέσεων που μας ενδιαφέρει και ονομαστικού επιπέδου ίσου με το επιθυμητό. Αυτό έχει ως συνέπεια να παράγονται διαστήματα εμπιστοσύνης τα οποία είναι πιο πλατιά.

3.7 Συμπεράσματα

Η μέθοδος που προτείνουν οι Jeong και Jhun για την κατασκευή bootstrap ταυτόχρονων διαστημάτων εμπιστοσύνης για τα πολυωνυμικά ποσοστά είναι αμφιλεγόμενη. Όταν το δείγμα μας είναι μικρό ή κάποιο από τα πολυωνυμικά ποσοστά είναι κοντά στο μηδέν, τότε η πιθανότητα να εμφανιστεί μηδενική εκτίμηση bootstrap για ένα από τα ποσοστά είναι μεγάλη. Συνεπώς, το φάσμα εφαρμογής της μεθόδου είναι πολύ μικρότερο από ότι παρουσιάζεται στην εργασία τους.

Τα bootstrap ταυτόχρονα διαστήματα εμπιστοσύνης για τις αντιθέσεις λειτουργούν καλύτερα. Αν και υπάρχει πιθανότητα να εμφανιστεί μηδενική εκτίμηση bootstrap για κάποια αντίθεση αυτή η πιθανότητα είναι μικρότερη και μπορεί να αποφευχθεί. Το μεγάλο όμως μειονέκτημα είναι τα πλάτος των διαστημάτων εμπιστοσύνης. Είναι πολύ πιο πλατιά από τις υπόλοιπες μεθόδους. Η «θυσία» αυτή, σύμφωνα με τους Jeong και Jhun, μας επιτρέπει να πλησιάσουμε περισσότερο το επιθυμητό ονομαστικό επίπεδο.

Η χρήση τους πρέπει να είναι περιορισμένη και να χρησιμοποιείται μόνο στην περίπτωση σχεδόν μηδενικής πιθανότητας εμφάνισης μηδενικής εκτίμησης bootstrap.

4^ο Κεφάλαιο

Διαφορά Ποσοστών

4.0 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με την διαφορά δύο ποσοστών. Στις Παραγράφους 4.1 έως 4.5 θα εξετάσουμε την περίπτωση που τα ποσοστά είναι συσχετισμένα μεταξύ τους και θα μελετήσουμε τον bootstrap έλεγχο για την ισότητα τους. Στις Παραγράφους 4.5 έως 4.10 θα εξετάσουμε την περίπτωση που τα ποσοστά προέρχονται από ανεξάρτητες διωνυμικές κατανομές, θα μελετήσουμε την κατασκευή διαστημάτων εμπιστοσύνης μέσω της μεθόδου bootstrap και θα τις συγκρίνουμε με τις παραδοσιακές μεθόδους.

4.1 Έλεγχοι Ισότητας Συσχετισμένων Ποσοστών

Στη παράγραφο αυτή μας ενδιαφέρει η σύγκριση δύο κατηγορικών αποκρίσεων για δύο δείγματα όταν η κάθε παρατήρηση στο ένα δείγμα ταιριάζει με μια παρατήρηση από το άλλο δείγμα. Δηλαδή τα δύο δείγματα που έχουμε είναι συσχετισμένα μεταξύ τους και οι παρατηρήσεις δίνονται σε ζεύγη.

Ας υποθέσουμε ότι εκτελούμε ένα πείραμα σε ένα δείγμα και μετά από την μεσολάβηση κάποιου γεγονότος επαναλαμβάνουμε το πείραμα στο ίδιο δείγμα. Ας θεωρήσουμε επίσης ότι η μεταβλητή απόκρισης μπορεί να ταξινομηθεί σε δύο κατηγορίες. Τότε ουσιαστικά έχουμε δύο συσχετισμένα δείγματα. Για παράδειγμα, επιλέγεται ένα τυχαίο δείγμα ψηφοφόρων οι οποίοι απαντούν για το αν θα ψηφίσουν τον ίδιο ψηφοφόρο πριν και μετά από μια τηλεοπτική συνέντευξη του με δυνατές απαντήσεις «Ναι» και «Όχι». Αυτό που θα μας ενδιαφέρει στην συγκεκριμένη περίπτωση είναι αν υπάρχει μεταστροφή στην πρόθεση ψήφου των ψηφοφόρων. Ας δούμε το θεωρητικό πλαίσιο ενός τέτοιου πειράματος.

Έστω ότι έχουμε έναν 2×2 πίνακα συνάφειας και με n_{ij} συμβολίζουμε την συχνότητα του (i, j) κελιού για $i=1,2$ και $j=1,2$. Με n δηλώνουμε το συνολικό πλήθος παρατηρήσεων

4. Διαφορά Ποσοστών

για το οποίο ισχύει $n = \sum_{i=1}^2 \sum_{j=1}^2 n_{ij}$. Επίσης με p_{ij} συμβολίζουμε την πιθανότητα του (i, j)

κελιού και ισχύει ότι $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$. Δηλαδή θεωρούμε ότι το δειγματοληπτικό σχήμα είναι

πολυωνυμικό με $(n_{11}, n_{12}, n_{21}, n_{22}) \sim Multi(n, (p_{11}, p_{12}, p_{21}, p_{22}))$. Οι εκτιμητές μέγιστης

πιθανοφάνειας των πιθανοτήτων p_{ij} είναι τα δειγματικά ποσοστά $p_{ij} = \frac{n_{ij}}{n}$, $i, j = 1, 2$. Τέλος,

ας θεωρήσουμε ότι $p_{i\cdot} = \sum_{j=1}^2 p_{ij}$ είναι το περιθώριο άθροισμα των πιθανοτήτων της i -γραμμής

και $p_{\cdot j} = \sum_{i=1}^2 p_{ij}$ είναι το περιθώριο άθροισμα της j -στήλης.

Αυτό που μας ενδιαφέρει είναι η σύγκριση των περιθωρίων ποσοστών $p_{1\cdot}$ και $p_{\cdot 1}$, δηλαδή η διαφορά :

$$d = p_{1\cdot} - p_{\cdot 1} = (p_{11} + p_{12}) - (p_{11} + p_{21}) = p_{12} - p_{21} \quad (4.1)$$

Είναι προφανές ότι όταν ισχύει $p_{1\cdot} = p_{\cdot 1}$ τότε θα ισχύει και $p_{2\cdot} = p_{\cdot 2}$, δηλαδή θα υπάρχει περιθώρια ομοιογένεια (*marginal homogeneity*) και επειδή ο πίνακας μας είναι 2×2 συνάγεται ότι θα υπάρχει και συμμετρία (*symmetry*).

Η υπόθεση που θέλουμε να ελέγξουμε είναι:

$$H_0 : d = 0 \text{ έναντι } H_1 : d \neq 0$$

η οποία μέσω της σχέσης (4.1) μπορεί να γραφεί εναλλακτικά:

$$H_0 : p_{12} = p_{21} \text{ έναντι } H_1 : p_{12} \neq p_{21} \quad (4.2)$$

Τα ποσοστά p_{12} και p_{21} είναι συσχετισμένα μεταξύ τους και επομένως οι μέθοδοι σύγκρισης δύο ανεξάρτητων ποσοστών δεν μπορούν να εφαρμοστούν. Η υπόθεση (4.2) ελέγχεται με το κριτήριο McNemar (McNemar, 1947). Οι συνήθεις ελεγχοσυναρτήσεις βασίζονται είτε στο X^2 κριτήριο καλής προσαρμογής είτε σε στατιστικά που βασίζονται σε προσεγγίσεις της κανονικής κατανομής.

4.1.1 Wald Στατιστικό

Ας είναι \hat{d} ο εκτιμητής μέγιστης πιθανοφάνειας του d . Τότε ισχύει :

$$\hat{d} = \frac{n_{12} - n_{21}}{n} \quad (4.3)$$

4^ο Κεφάλαιο

Η διασπορά του \hat{d} είναι

$$Var[\hat{d}] = Var(p_{12} - p_{21}) = Var[p_{12}] + Var[p_{21}] - 2cov(p_{12}, p_{21})$$

Αντικαταστάοντας τους εκτιμητές μέγιστης πιθανοφάνειας $p_{ij} = \frac{n_{ij}}{n}$ και χρησιμοποιώντας τις ιδιότητες της πολυωνυμικής κατανομής βρίσκουμε ότι :

$$\begin{aligned} Var[\hat{d}] &= Var\left[\frac{n_{12}}{n}\right] + Var\left[\frac{n_{21}}{n}\right] - 2cov\left(\frac{n_{12}}{n}, \frac{n_{21}}{n}\right) = \frac{1}{n^2} [Var[n_{12}] + Var[n_{21}] - 2cov(n_{12}, n_{21})] \\ &= \frac{1}{n^2} [np_{12}(1-p_{12}) + np_{21}(1-p_{21}) - 2np_{12}p_{21}] = \frac{1}{n} [p_{12} - p_{12}^2 + p_{21} - p_{21}^2 - p_{12}p_{21}] \end{aligned}$$

και με πράξεις ότι

$$Var[\hat{d}] = \frac{1}{n} [(p_{12} + p_{21}) - (p_{12} - p_{21})^2] \quad (4.4)$$

Κάτω από την μηδενική υπόθεση του ελέγχου (4.2) η διασπορά του \hat{d} στην σχέση (4.4) απλοποιείται σε

$$Var[\hat{d}] = \frac{1}{n} (p_{12} + p_{21})$$

Αν αντικαταστήσουμε τα p_{12} και p_{21} με τους εκτιμητές μέγιστης πιθανοφάνειας τότε βρίσκουμε την εκτίμηση του $Var[\hat{d}]$ κάτω από την μηδενική υπόθεση

$$\hat{Var}[\hat{d}] = \frac{1}{n^2} (n_{12} + n_{21}) \quad (4.5)$$

Για μεγάλα δείγματα n το \hat{d} ακολουθεί κανονική κατανομή (βλ. Agresti, 2002, σελ.410-411). Επομένως το τυποποιημένο στατιστικό

$$Z = \frac{\hat{d}}{\sqrt{\hat{Var}[\hat{d}]}} = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$$

ακολουθεί την τυπική κανονική κατανομή $N(0,1)$ και ονομάζεται στατιστικό του Wald. Η κρίσιμη περιοχή της υπόθεσης (4.2) είναι η $\mathfrak{R} = \{|Z| > z_{\alpha/2}\}$.

Αυτό που είναι αρκετά ενδιαφέρον είναι ότι το στατιστικό Z βασίζεται μόνο στις συχνότητες των μη διαγώνιων στοιχείων του πίνακα, δηλαδή στις κατηγορίες όπου η μεταβλητή απόκρισης είναι διαφορετική. Ακόμα και αν οι συχνότητες n_{11} , n_{22} είναι μικρές ή μεγάλες η τιμή του Z δεν επηρεάζεται. Από τις σχέσεις (4.3) και (4.5) παρατηρούμε ότι οι

4. Διαφορά Ποσοστών

συχνότητες των n_{11}, n_{22} συνεισφέρουν στην εκτίμηση του d και στην εκτίμηση της διασποράς του.

4.1.2 McNemar και Yates Στατιστικό

Μια πολύ απλή εξήγηση είναι να πάρουμε το τετράγωνο του στατιστικού Z , το οποίο ακολουθεί χι-τετράγωνο κατανομή με έναν βαθμό ελευθερίας (βλ. Agresti, 2002, σελ.410-411), δηλαδή

$$Z^2 = \frac{(n_{12} - n_{21})^2}{(n_{12} + n_{21})} \sim X_1^2$$

Αυτό το στατιστικό είναι γνωστό ως McNemar στατιστικό.

Ωστόσο αξίζει να δούμε τον McNemar έλεγχο ως έναν έλεγχο καλής προσαρμογής. Επειδή τα διαγώνια στοιχεία του πίνακα δεν επηρεάζουν την μηδενική υπόθεση μπορούμε να υποστηρίξουμε ότι το δειγματοληπτικό σχέδιο είναι

$$(n_{11} + n_{22}, n_{12}, n_{21}) \sim Multi(n, (p_{11} + p_{22}, p_{12}, p_{21}))$$

Κάτω από την μηδενική υπόθεση του ελέγχου (4.2) το δειγματοληπτικό σχήμα είναι

$$(n_{11} + n_{22}, n_{12}, n_{21}) \sim Multi(n, (1 - 2p, p, p))$$

όπου $p = p_{12} = p_{21}$. Το p εκτιμάται από το $\hat{p} = \frac{n_{12} + n_{21}}{2n}$.

Επομένως το X^2 στατιστικό είναι ίσο με:

$$\begin{aligned} X^2 &= \frac{(n_{11} + n_{22} - n(1 - 2\hat{p}))^2}{n(1 - 2\hat{p})} + \frac{(n_{12} - n\hat{p})^2}{n\hat{p}} + \frac{(n_{21} - n\hat{p})^2}{n\hat{p}} \\ &= \frac{\left(n_{11} + n_{22} - n\left(1 - 2\frac{n_{12} + n_{21}}{2n}\right)\right)^2}{n\left(1 - 2\frac{n_{12} + n_{21}}{2n}\right)} + \frac{\left(n_{12} - n\frac{n_{12} + n_{21}}{2n}\right)^2}{n\frac{n_{12} + n_{21}}{2n}} + \frac{\left(n_{21} - n\frac{n_{12} + n_{21}}{2n}\right)^2}{n\frac{n_{12} + n_{21}}{2n}} \\ &= 0 + \frac{\left(\frac{n_{12} - n_{21}}{2}\right)^2}{\frac{n_{12} + n_{21}}{2}} + \frac{\left(\frac{n_{21} - n_{12}}{2}\right)^2}{\frac{n_{12} + n_{21}}{2}} = 2\frac{\left(\frac{n_{12} - n_{21}}{2}\right)^2}{\frac{n_{12} + n_{21}}{2}} \end{aligned}$$

το οποίο απλοποιείται σε

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \sim X_{k-s-1}^2 \equiv X_1^2$$

4^ο Κεφάλαιο

αφού $k = 3$ γιατί έχουμε την τριωνυμική κατανομή και $s = 1$ γιατί η μόνη παράμετρος προς εκτίμηση κάτω από την μηδενική υπόθεση είναι το p .

Όταν οι αναμενόμενες συχνότητες είναι πολύ μικρές τότε κάνουμε την διόρθωση της συνέχειας του Yates (βλ. Κόλυβα-Μαχαίρα και Μπόρα-Σέντα, 1998, σελ.416) και υπολογίζουμε την ποσότητα X^2 που ακολουθεί επίσης την X_1^2 και είναι

$$X^2 = \frac{(|n_{12} - n_{21}| - 1)^2}{n_{12} + n_{21}}$$

4.1.3 Ακριβής Διωνυμικός Έλεγχος

Δοθέντος του αθροίσματος $n_{12} + n_{21}$, κάτω από την μηδενική υπόθεση ισχύει $n_{12} \square B(n_{12} + n_{21}, 0.5)$, (βλ. Agresti, 2002, σελ. 413). Επομένως, εναλλακτικά ο έλεγχος ισότητας των ποσοστών μπορεί να πραγματοποιηθεί με χρήση του ακριβούς διωνυμικού ελέγχου.

4.1.4 Προσέγγιση μέσω της Bootstrap Μεθόδου

Όπως είδαμε στην Παράγραφο 3.3 όταν έχουμε κατηγορικά δεδομένα που προέρχονται από πολυωνυμική κατανομή τότε η εκτίμηση bootstrap της δειγματικής κατανομής είναι πολυωνυμική με παραμέτρους n και \underline{p} , όπου \underline{p} είναι το διάνυσμα των εκτιμητών μέγιστης πιθανοφάνειας των πιθανοτήτων. Από την άλλη το p-value του bootstrap ελέγχου είναι η πιθανότητα να παρατηρήσουμε κάτι το οποίο είναι πιο ακραίο από αυτό που παρατηρήσαμε δοθέντος του αρχικού μας δείγματος. Αυτή η πιθανότητα μπορεί να υπολογιστεί ακριβώς με βάση τον ακόλουθο αλγόριθμο:

1ο Βήμα: Δοθέντων των αρχικών δεδομένων $n_{11}, n_{12}, n_{21}, n_{22}$ υπολογίζουμε το

$$\hat{d} = (n_{12} - n_{21}) / n.$$

2ο Βήμα: Υπολογίζουμε το p μέσω του εκτιμητή:

$$\hat{p} = (n_{12} + n_{21}) / 2n$$

3ο Βήμα: Παίρνουμε όλα τα δυνατά δείγματα μεγέθους n , τα οποία έστω ότι είναι B στο πλήθος.

4ο Βήμα: Για κάθε $\{n_{11}^{*b} + n_{22}^{*b}, n_{12}^{*b}, n_{21}^{*b}\}$ bootstrap σύνολο δεδομένων, $b = 1, \dots, B$,

υπολογίζουμε την ποσότητα $\hat{d}^{*b} = (n_{12}^{*b} - n_{21}^{*b}) / n$ και την πιθανότητα

4. Διαφορά Ποσοστών

$$\text{Prob}_* \{n_{11}^* + n_{22}^* = n_{11}^{*b} + n_{22}^{*b}, n_{12}^* = n_{12}^{*b}, n_{21}^* = n_{21}^{*b}\} \quad \text{από} \quad \text{την} \quad \text{κατανομή}$$
$$\text{Multi}(n, (1-2\hat{p}, \hat{p}, \hat{p})).$$

5ο Βήμα: Αθροίζουμε τις πιθανότητες όλων των bootstrap δειγμάτων για τα οποία ισχύει

$\hat{d}^{*b} \geq |\hat{d}|$. Το p-value του bootstrap ελέγχου βρίσκεται από τον τύπο:

$$P_{BOOT} = \sum_{\{(n_{11}^{*b} + n_{22}^{*b}, n_{12}^{*b}, n_{21}^{*b}) : \hat{d}^{*b} \geq |\hat{d}|\}} \text{Prob}_* \{n_{11}^* + n_{22}^* = n_{11}^{*b} + n_{22}^{*b}, n_{12}^* = n_{12}^{*b}, n_{21}^* = n_{21}^{*b}\}$$

4.2 Κριτήρια Αξιολόγησης των Στατιστικών Ελέγχων

Σε κάθε έλεγχο μας ενδιαφέρουν τα σφάλματα τύπου I και II, και η ισχύς του ελέγχου. Ας συμβολίσουμε με a την πιθανότητα σφάλματος τύπου I, με $b(q)$ την πιθανότητα σφάλματος τύπου II όταν η τιμή της πραγματική τιμή της παραμέτρου είναι q και με $P(q)$ την ισχύ του ελέγχου στην τιμή q , δηλαδή $P(q) = 1 - b(q)$. Η ιδανική περίπτωση για ένα στατιστικό έλεγχο είναι να έχει μέγεθος (*size*) μικρότερο ή ίσο με το ονομαστικό επίπεδο του ελέγχου και μεγάλη ισχύ. Πριν εκτελέσουμε τον έλεγχο ορίζουμε το a , ενώ η πιθανότητα σφάλματος τύπου II και η ισχύς του ελέγχου είναι άγνωστες ποσότητες, αφού είναι άγνωστη και η πραγματική τιμή της παραμέτρου q . Αν η πραγματική τιμή της παραμέτρου q είναι γνωστή τότε μπορούμε να υπολογίσουμε την ισχύ του ελέγχου.

Στην παρούσα εργασία θα εκτιμήσουμε την ισχύ του ελέγχου μέσω της εμπειρικής ισχύς. Όταν ο έλεγχος είναι της μορφής

$$H_0 : q = q_0 \quad \text{έναντι} \quad H_1 : q \neq q_0$$

τότε η εμπειρική ισχύς στην τιμή q_1 υπολογίζεται υποθέτοντας ότι η πραγματική τιμή της q ισούται με $q_1 \neq q_0$ και υπολογίζοντας στην συνέχεια την πιθανότητα απόρριψης της μηδενικής υπόθεσης. Με ανάλογο τρόπο μπορούμε να υπολογίσουμε εμπειρικό μέγεθος του ελέγχου στο q_0 από την πιθανότητα απόρριψης της μηδενικής υπόθεσης.

Είναι επιθυμητό το στατιστικό έλεγχο να έχει πάντα μέγεθος μικρότερο από το a . Ωστόσο, επειδή εμείς ασχολούμαστε με το εμπειρικό μέγεθος μας αρκεί να είναι κοντά στο ονομαστικό επίπεδο του ελέγχου. Συνεπώς, θα λέμε ότι ένα στατιστικό έλεγχο είναι ισχυρό αν το εμπειρικό του μέγεθος είναι κοντά στο ονομαστικό επίπεδο και η ισχύς του είναι μεγαλύτερη από όλα τα στατιστικά έλεγχο που έχουν εμπειρικό μέγεθος κοντά στο ονομαστικό επίπεδο του ελέγχου για όλες τις τιμές της παραμέτρου q_1 .

4.3 Σύγκριση Στατιστικών Ελέγχου Ισότητας Εξαρτημένων Ποσοστών.

Θα πραγματοποιηθεί η σύγκριση των στατιστικών του McNemar, του χι-τετράγωνο, του χι-τετράγωνο με την διόρθωση του Yates, του ακριβές διωνυμικού ελέγχου και του bootstrap ελέγχου για μικρού και μεσαίου μεγέθους δείγματα μέσω του εμπειρικού μεγέθους και της εμπειρικής ισχύς τους. Η σύγκριση θα γίνει σε 5% και 10% ονομαστικό επίπεδο.

Για κάθε μέγεθος δείγματος n , θα παραχθούν όλα τα δυνατά δείγματα πλην του δείγματος που αποτελείται μόνο από τις σύμφωνες κατηγορίες. Ο λόγος είναι ότι τα παραπάνω στατιστικά δεν μπορούν να υπολογιστούν όταν οι μη σύμφωνες κατηγορίες έχουν μηδενικές συχνότητες. Η εμπειρική ισχύς θα υπολογιστεί από το άθροισμα της πιθανότητας εμφάνισης των δειγμάτων που απορρίπτουν την μηδενική υπόθεση. Η πιθανότητα εμφάνισης είναι η πιθανότητα να εμφανιστεί το συγκεκριμένο δείγμα όταν ισχύει η εναλλακτική υπόθεση. Ομοίως το εμπειρικό μέγεθος θα υπολογιστεί από το άθροισμα της πιθανότητας εμφάνισης των δειγμάτων που απορρίπτουν την μηδενική υπόθεση όταν ισχύει η μηδενική υπόθεση. Επίσης ορίζουμε με $y = p_{12} + p_{21}$, το άθροισμα των πιθανοτήτων των μη σύμφωνων κατηγοριών έτσι ώστε να μπορούμε να μελετήσουμε την συμπεριφορά των στατιστικών ελέγχου για διάφορες τιμές των p_{12} και p_{21} .

Πίνακας 4.1

Αριθμός δειγμάτων που απορρίπτουν την μηδενική υπόθεση για σταθερό n

a	n	Bootstrap	Wald	McNemar	Yates	Exact
0.05	5	0	4	4	0	0
	10	18	20	20	14	14
	15	26	54	54	44	44
	20	102	108	108	92	94
	25	172	182	182	160	162
	30	262	276	276	248	250
	40	504	524	524	486	488
	50	830	858	858	808	812
0.1	5	4	6	6	2	2
	10	22	28	28	18	18
	15	60	68	68	54	54
	20	118	128	128	110	110
	25	196	208	208	186	186
	30	296	310	310	282	284
	40	558	576	576	538	542
	50	906	930	930	882	886

Ο Πίνακας 4.1 μας δείχνει τον αριθμό δειγμάτων που απορρίπτουν την υπόθεση της ισότητας των ποσοστών. Παρατηρούμε ότι το στατιστικό του Yates και ο ακριβής διωνυμικός έλεγχος απορρίπτουν την μηδενική υπόθεση λιγότερες φορές από ότι οι υπόλοιπες μέθοδοι.

4. Διαφορά Ποσοστών

Πίνακας 4.2
Εμπειρική Ισχύς όταν $5 \leq n \leq 25$ και $\alpha = 0.05$

<i>n</i>	<i>d</i>	<i>Y</i>	Bootstrap	Wald	McNemar	Yates	Exact
5	0.1	0.2	0.0000	0.0021	0.0021	0.0000	0.0000
		0.3	0.0000	0.0063	0.0063	0.0000	0.0000
		0.4	0.0000	0.0143	0.0143	0.0000	0.0000
	0.2	0.5	0.0000	0.0270	0.0270	0.0000	0.0000
		0.3	0.0000	0.0147	0.0147	0.0000	0.0000
		0.4	0.0000	0.0270	0.0270	0.0000	0.0000
	0.3	0.5	0.0000	0.0441	0.0441	0.0000	0.0000
		0.4	0.0000	0.0503	0.0503	0.0000	0.0000
		0.5	0.0000	0.0745	0.0745	0.0000	0.0000
10	0.1	0.2	0.0074	0.0356	0.0356	0.0011	0.0011
		0.3	0.0181	0.0601	0.0601	0.0039	0.0039
		0.4	0.0309	0.0741	0.0741	0.0090	0.0090
	0.2	0.5	0.0414	0.0732	0.0732	0.0154	0.0154
		0.3	0.0571	0.1538	0.1538	0.0152	0.0152
		0.4	0.0778	0.1581	0.1581	0.0271	0.0271
	0.3	0.5	0.0893	0.1402	0.1402	0.0386	0.0386
		0.4	0.1786	0.3257	0.3257	0.0715	0.0715
		0.5	0.1858	0.2701	0.2701	0.0904	0.0904
15	0.1	0.2	0.0356	0.0957	0.0957	0.0101	0.0101
		0.3	0.0549	0.1008	0.1008	0.0213	0.0213
		0.4	0.0642	0.0861	0.0861	0.0304	0.0304
	0.2	0.5	0.0666	0.0731	0.0731	0.0355	0.0355
		0.3	0.1841	0.2897	0.2897	0.0881	0.0881
		0.4	0.1766	0.2173	0.2173	0.0989	0.0989
	0.3	0.5	0.1614	0.1717	0.1717	0.0983	0.0983
		0.4	0.4102	0.4845	0.4845	0.2665	0.2665
		0.5	0.3466	0.3638	0.3638	0.2383	0.2383
20	0.1	0.2	0.0735	0.1493	0.1493	0.0318	0.0318
		0.3	0.0719	0.1174	0.1174	0.0457	0.0457
		0.4	0.0665	0.0936	0.0936	0.0501	0.0501
	0.2	0.5	0.0691	0.0821	0.0821	0.0495	0.0496
		0.3	0.2711	0.3767	0.3767	0.1956	0.1956
		0.4	0.2110	0.2703	0.2703	0.1737	0.1737
	0.3	0.5	0.1924	0.2184	0.2184	0.1504	0.1507
		0.4	0.5123	0.6034	0.6034	0.4549	0.4549
		0.5	0.4285	0.4716	0.4716	0.3669	0.3671
25	0.1	0.2	0.1110	0.1875	0.1875	0.0632	0.0632
		0.3	0.0930	0.1313	0.1313	0.0698	0.0698
		0.4	0.0899	0.1063	0.1063	0.0658	0.0661
	0.2	0.5	0.0859	0.0961	0.0961	0.0590	0.0616
		0.3	0.3628	0.4520	0.4520	0.3046	0.3046
		0.4	0.2949	0.3313	0.3313	0.2413	0.2419
	0.3	0.5	0.2534	0.2734	0.2734	0.1947	0.2007
		0.4	0.6572	0.7092	0.7092	0.6017	0.6020
		0.5	0.5520	0.5760	0.5760	0.4722	0.4790

Ο Πίνακας 4.2 περιέχει την εμπειρική ισχύ των μεθόδων όταν $5 \leq n \leq 25$ και $\alpha = 0.05$. Το στατιστικό του Yates και ο ακριβής διωνυμικός έλεγχος έχουν την μικρότερη εμπειρική ισχύ ενώ το χι-τετράγωνο του McNemar και το στατιστικό του Wald έχουν ακριβώς την ίδια ισχύ. Επίσης, η ισχύς του ελέγχου bootstrap είναι μεγαλύτερη από τον ακριβή διωνυμικό έλεγχο αλλά μικρότερη από τον έλεγχο McNemar.

Πίνακας 4.3
Εμπειρική Ισχύς όταν $30 \leq n \leq 50$ και $\alpha = 0.05$

<i>n</i>	<i>d</i>	<i>y</i>	Bootstrap	Wald	McNemar	Yates	Exact
30	0.1	0.2	0.1446	0.2166	0.2166	0.0992	0.0993
		0.3	0.1195	0.1484	0.1484	0.0917	0.0919
		0.4	0.1100	0.1232	0.1232	0.0783	0.0812
		0.5	0.0937	0.1149	0.1149	0.0675	0.0766
	0.2	0.3	0.4597	0.5264	0.5264	0.4029	0.4032
		0.4	0.3706	0.3967	0.3967	0.3012	0.3071
		0.5	0.2954	0.3366	0.3366	0.2371	0.2579
	0.3	0.4	0.7694	0.7946	0.7946	0.7104	0.7131
		0.5	0.6356	0.6745	0.6745	0.5625	0.5859
40	0.1	0.2	0.2145	0.2718	0.2718	0.1723	0.1724
		0.3	0.1719	0.1919	0.1919	0.1289	0.1336
		0.4	0.1336	0.1657	0.1657	0.1027	0.1154
		0.5	0.1110	0.1493	0.1493	0.0923	0.0989
	0.2	0.3	0.6325	0.6635	0.6635	0.5620	0.5674
		0.4	0.4782	0.5313	0.5313	0.4131	0.4397
		0.5	0.3794	0.4504	0.4504	0.3394	0.3544
	0.3	0.4	0.8904	0.9088	0.9088	0.8507	0.8630
		0.5	0.7630	0.8159	0.8159	0.7254	0.7423
50	0.1	0.2	0.2930	0.3325	0.3325	0.2394	0.2412
		0.3	0.2091	0.2444	0.2444	0.1639	0.1787
		0.4	0.1588	0.2057	0.2057	0.1344	0.1429
		0.5	0.1363	0.1764	0.1764	0.1181	0.1188
	0.2	0.3	0.7454	0.7761	0.7761	0.6824	0.6993
		0.4	0.5736	0.6424	0.6424	0.5288	0.5467
		0.5	0.4735	0.5413	0.5413	0.4385	0.4401
	0.3	0.4	0.9472	0.9630	0.9630	0.9304	0.9387
		0.5	0.8604	0.8962	0.8962	0.8390	0.8408

Τα συμπεράσματα που αναφέραμε ισχύουν και όταν το δείγμα κυμαίνεται από 30 έως 40 όπως φαίνεται και στον Πίνακα 4.3. Για σταθερές τιμές των *d* και *y*, η εμπειρική ισχύς όλων των μεθόδων αυξάνει όσο αυξάνει το *n*. Αυτό είναι λογικό αν σκεφτούμε ότι όσο το μέγεθος του δείγματος μεγαλώνει τόσο οι έλεγχοι μπορούν να ανιχνεύσουν διαφορά μεταξύ των ποσοστών.

Στον Πίνακα 4.4 παρουσιάζονται οι τιμές του εμπειρικού μεγέθους των στατιστικών ελέγχου. Το πρώτο που πρέπει να παρατηρήσουμε είναι ότι το εμπειρικό μέγεθος των στατιστικών του Wald και του McNemar είναι ακριβώς το ίδιο και σε γενικές γραμμές κοντά στο ονομαστικό επίπεδο. Το εμπειρικό μέγεθος του ακριβές διωνυμικού ελέγχου και του στατιστικού του Yates είναι παρόμοια και μικρότερο από το ονομαστικό επίπεδο αλλά όπως είδαμε στους προηγούμενους πίνακες η εμπειρική ισχύς τους είναι μικρή. Ο bootstrap έλεγχος έχει πάντα εμπειρικό μέγεθος μικρότερο από το ονομαστικό επίπεδο και μεγαλύτερο από τον ακριβή διωνυμικό έλεγχο και το στατιστικό του Yates. Επίσης, όταν το άθροισμα των ποσοστών των σύμφωνων κατηγοριών είναι κοντά στο 0.5, δίνει μεγάλα μεγέθη ελέγχου.

4. Διαφορά Ποσοστών

Πίνακας 4.4
Εμπειρικό μέγεθος ελέγχου όταν $\alpha = 0.05$

n	$1-\gamma$	Bootstrap	Wald	McNemar	Yates	Exact
5	0.9	0.0000	0.0000	0.0000	0.0000	0.0000
	0.8	0.0000	0.0008	0.0008	0.0000	0.0000
	0.7	0.0000	0.0037	0.0037	0.0000	0.0000
	0.6	0.0000	0.0102	0.0102	0.0000	0.0000
	0.5	0.0000	0.0215	0.0215	0.0000	0.0000
10	0.9	0.0000	0.0015	0.0015	0.0000	0.0000
	0.8	0.0018	0.0129	0.0129	0.0002	0.0002
	0.7	0.0078	0.0328	0.0328	0.0013	0.0013
	0.6	0.0175	0.0489	0.0489	0.0043	0.0043
	0.5	0.0271	0.0527	0.0527	0.0090	0.0090
15	0.9	0.0007	0.0061	0.0061	0.0000	0.0000
	0.8	0.0083	0.0317	0.0317	0.0016	0.0016
	0.7	0.0218	0.0491	0.0491	0.0067	0.0067
	0.6	0.0326	0.0485	0.0485	0.0132	0.0132
	0.5	0.0387	0.0439	0.0439	0.0184	0.0184
20	0.9	0.0023	0.0136	0.0136	0.0003	0.0003
	0.8	0.0157	0.0444	0.0444	0.0048	0.0048
	0.7	0.0248	0.0483	0.0483	0.0130	0.0130
	0.6	0.0280	0.0436	0.0436	0.0193	0.0193
	0.5	0.0338	0.0421	0.0421	0.0226	0.0226
25	0.9	0.0049	0.0223	0.0223	0.0009	0.0009
	0.8	0.0213	0.0485	0.0485	0.0088	0.0088
	0.7	0.0273	0.0448	0.0448	0.0178	0.0178
	0.6	0.0335	0.0424	0.0424	0.0223	0.0224
	0.5	0.0379	0.0436	0.0436	0.0236	0.0248
30	0.9	0.0083	0.0308	0.0308	0.0019	0.0019
	0.8	0.0244	0.0479	0.0479	0.0128	0.0128
	0.7	0.0307	0.0430	0.0430	0.0208	0.0209
	0.6	0.0370	0.0433	0.0433	0.0234	0.0244
	0.5	0.0367	0.0475	0.0475	0.0237	0.0278
40	0.9	0.0153	0.0427	0.0427	0.0050	0.0050
	0.8	0.0285	0.0443	0.0443	0.0187	0.0187
	0.7	0.0365	0.0436	0.0436	0.0233	0.0244
	0.6	0.0359	0.0489	0.0489	0.0242	0.0285
	0.5	0.0347	0.0528	0.0528	0.0267	0.0296
50	0.9	0.0207	0.0473	0.0473	0.0088	0.0088
	0.8	0.0330	0.0429	0.0429	0.0218	0.0221
	0.7	0.0365	0.0474	0.0474	0.0239	0.0275
	0.6	0.0348	0.0526	0.0526	0.0266	0.0296
	0.5	0.0362	0.0536	0.0536	0.0293	0.0296

Στους Πίνακες 4.5 και 4.6 παρουσιάζεται η εμπειρική ισχύς όταν $\alpha = 0.1$. Όσον αφορά την εμπειρική ισχύ ισχύουν όσα και προηγουμένως.

Πίνακας 4.5
Εμπειρική Ισχύς όταν $5 \leq n \leq 25$ και $\alpha = 0.1$

<i>n</i>	<i>d</i>	<i>Y</i>	Bootstrap	Wald	McNemar	Yates	Exact
5	0.1	0.2	0.0021	0.0245	0.0245	0.0000	0.0000
		0.3	0.0063	0.0504	0.0504	0.0003	0.0003
		0.4	0.0143	0.0827	0.0827	0.0011	0.0011
	0.2	0.5	0.0270	0.1145	0.1145	0.0027	0.0027
		0.3	0.0147	0.0918	0.0918	0.0010	0.0010
		0.4	0.0270	0.1278	0.1278	0.0024	0.0024
	0.3	0.5	0.0441	0.1597	0.1597	0.0053	0.0053
		0.4	0.0503	0.2051	0.2051	0.0053	0.0053
		0.5	0.0745	0.2370	0.2370	0.0103	0.0103
10	0.1	0.2	0.0359	0.1240	0.1240	0.0074	0.0074
		0.3	0.0620	0.1510	0.1510	0.0181	0.0181
		0.4	0.0812	0.1454	0.1454	0.0309	0.0309
	0.2	0.5	0.0905	0.1252	0.1252	0.0414	0.0414
		0.3	0.1573	0.3130	0.3130	0.0571	0.0571
		0.4	0.1714	0.2659	0.2659	0.0778	0.0778
	0.3	0.5	0.1695	0.2158	0.2158	0.0893	0.0893
		0.4	0.3424	0.4872	0.4872	0.1786	0.1786
		0.5	0.3132	0.3774	0.3774	0.1858	0.1858
15	0.1	0.2	0.1000	0.2097	0.2097	0.0356	0.0356
		0.3	0.1185	0.1780	0.1780	0.0549	0.0549
		0.4	0.1197	0.1513	0.1513	0.0642	0.0642
	0.2	0.5	0.1145	0.1484	0.1484	0.0670	0.0670
		0.3	0.3222	0.4247	0.4247	0.1841	0.1841
		0.4	0.2796	0.3262	0.3262	0.1767	0.1767
	0.3	0.5	0.2432	0.2930	0.2930	0.1621	0.1621
		0.4	0.5600	0.6163	0.6163	0.4102	0.4102
		0.5	0.4644	0.5211	0.5211	0.3474	0.3474
20	0.1	0.2	0.1670	0.2591	0.2097	0.0356	0.0794
		0.3	0.1584	0.1989	0.1780	0.0549	0.0902
		0.4	0.1419	0.1843	0.1513	0.0642	0.0896
	0.2	0.5	0.1268	0.1751	0.1484	0.0670	0.0900
		0.3	0.4501	0.5109	0.4247	0.1841	0.3139
		0.4	0.3592	0.4219	0.3262	0.1767	0.2628
	0.3	0.5	0.2978	0.3708	0.2930	0.1621	0.2335
		0.4	0.6991	0.7473	0.6163	0.4102	0.5851
		0.5	0.5730	0.6505	0.5211	0.3474	0.4879
25	0.1	0.2	0.2255	0.2941	0.2941	0.1277	0.1277
		0.3	0.1880	0.2341	0.2341	0.1214	0.1214
		0.4	0.1589	0.2108	0.2108	0.1160	0.1160
	0.2	0.5	0.1442	0.1839	0.1839	0.1159	0.1159
		0.3	0.5481	0.6049	0.6049	0.4268	0.4268
		0.4	0.4259	0.5024	0.5024	0.3484	0.3484
	0.3	0.5	0.3584	0.4190	0.4190	0.3101	0.3101
		0.4	0.7924	0.8419	0.8419	0.7163	0.7163
		0.5	0.6694	0.7289	0.7289	0.6161	0.6161

4. Διαφορά Ποσοστών

Πίνακας 4.6
Εμπειρική Ισχύς όταν $30 \leq n \leq 50$ και $\alpha = 0.1$

<i>n</i>	<i>d</i>	<i>Y</i>	Bootstrap	Wald	McNemar	Yates	Exact
30	0.1	0.2	0.2753	0.3317	0.3317	0.1747	0.1747
		0.3	0.2126	0.2694	0.2694	0.1524	0.1524
		0.4	0.1791	0.2265	0.2265	0.1441	0.1441
	0.2	0.5	0.1648	0.1966	0.1966	0.1378	0.1378
		0.3	0.6262	0.6892	0.6892	0.5259	0.5259
		0.4	0.4923	0.5605	0.5605	0.4332	0.4332
	0.3	0.5	0.4217	0.4689	0.4689	0.3776	0.3776
		0.4	0.8594	0.8978	0.8978	0.8145	0.8145
			0.5	0.7535	0.7922	0.7922	0.7142
40	0.1	0.2	0.3566	0.4189	0.4189	0.2627	0.2627
		0.3	0.2633	0.3197	0.3197	0.2180	0.2180
		0.4	0.2249	0.2604	0.2604	0.1935	0.1935
	0.2	0.5	0.1986	0.2270	0.2270	0.1728	0.1729
		0.3	0.7486	0.8024	0.8024	0.6913	0.6913
		0.4	0.6174	0.6610	0.6610	0.5741	0.5741
	0.3	0.5	0.5272	0.5656	0.5656	0.4888	0.4889
		0.4	0.9418	0.9562	0.9562	0.9258	0.9258
		0.5	0.8647	0.8854	0.8854	0.8416	0.8416
50	0.1	0.2	0.4257	0.4970	0.4970	0.3481	0.3481
		0.3	0.3188	0.3639	0.3639	0.2791	0.2791
		0.4	0.2644	0.2973	0.2973	0.2340	0.2343
	0.2	0.5	0.2259	0.2587	0.2587	0.1995	0.2045
		0.3	0.8397	0.8717	0.8717	0.8067	0.8067
		0.4	0.7127	0.7455	0.7455	0.6783	0.6785
	0.3	0.5	0.6104	0.6493	0.6493	0.5753	0.5813
		0.4	0.9773	0.9826	0.9826	0.9709	0.9709
		0.5	0.9253	0.9386	0.9386	0.9115	0.9130

Στον Πίνακα 4.7 παρουσιάζονται οι τιμές του εμπειρικού μεγέθους των στατιστικών ελέγχου. Παρατηρούμε ότι το εμπειρικό μέγεθος των στατιστικών του Wald και του McNemar είναι ακριβώς το ίδιο και σε γενικές γραμμές κοντά στο ονομαστικό επίπεδο. Το εμπειρικό μέγεθος του ακριβές διωνυμικού ελέγχου και του στατιστικού του Yates είναι παρόμοια και μικρότερο από το ονομαστικό επίπεδο αλλά όπως είδαμε στους προηγούμενους πίνακες η εμπειρική ισχύς τους είναι μικρή. Ο bootstrap έλεγχος έχει πάντα μεγαλύτερο μέγεθος από αυτές τις μεθόδους και μάλιστα μικρότερο και από το ονομαστικό επίπεδο.

Πίνακας 4.7
Εμπειρικό μέγεθος ελέγχου όταν $\alpha = 0.1$

<i>n</i>	$1-y$	Bootstrap	Wald	McNemar	Yates	Exact
5	0.9	0.0021	0.0021	0.0021	0.0000	0.0000
	0.8	0.0136	0.0136	0.0136	0.0000	0.0000
	0.7	0.0368	0.0368	0.0368	0.0002	0.0002
	0.6	0.0678	0.0678	0.0678	0.0006	0.0006
	0.5	0.0996	0.0996	0.0996	0.0020	0.0020
10	0.9	0.0159	0.0158	0.0158	0.0000	0.0000
	0.8	0.0643	0.0633	0.0633	0.0018	0.0018
	0.7	0.1074	0.1006	0.1006	0.0078	0.0078
	0.6	0.1284	0.1075	0.1075	0.0175	0.0175
	0.5	0.1348	0.0963	0.0963	0.0271	0.0271
15	0.9	0.0386	0.0382	0.0382	0.0007	0.0007
	0.8	0.1039	0.0959	0.0959	0.0083	0.0083
	0.7	0.1301	0.1024	0.1024	0.0218	0.0218
	0.6	0.1346	0.0946	0.0946	0.0326	0.0326
	0.5	0.1355	0.1001	0.1001	0.0389	0.0389
20	0.9	0.0630	0.0613	0.0613	0.0023	0.0023
	0.8	0.1234	0.1029	0.1029	0.0171	0.0171
	0.7	0.1344	0.0965	0.0965	0.0320	0.0320
	0.6	0.1379	0.1025	0.1025	0.0402	0.0402
	0.5	0.1458	0.1079	0.1079	0.0467	0.0467
25	0.9	0.0843	0.0798	0.0798	0.0050	0.0050
	0.8	0.1312	0.0999	0.0999	0.0252	0.0252
	0.7	0.1370	0.1002	0.1002	0.0384	0.0384
	0.6	0.1452	0.1067	0.1067	0.0468	0.0468
	0.5	0.1514	0.1020	0.1020	0.0553	0.0553
30	0.9	0.1009	0.0920	0.0920	0.0086	0.0086
	0.8	0.1342	0.0978	0.0978	0.0315	0.0315
	0.7	0.1416	0.1049	0.1049	0.0435	0.0435
	0.6	0.1505	0.1034	0.1034	0.0535	0.0535
	0.5	0.1603	0.0990	0.0990	0.0607	0.0607
40	0.9	0.1211	0.1009	0.1009	0.0170	0.0170
	0.8	0.1387	0.1018	0.1018	0.0400	0.0400
	0.7	0.1506	0.1034	0.1034	0.0533	0.0533
	0.6	0.1634	0.0989	0.0989	0.0619	0.0619
	0.5	0.1698	0.0975	0.0975	0.0652	0.0653
50	0.9	0.1300	0.1001	0.1001	0.0248	0.0248
	0.8	0.1448	0.1049	0.1049	0.0468	0.0468
	0.7	0.1600	0.0996	0.0996	0.0602	0.0602
	0.6	0.1696	0.0976	0.0976	0.0650	0.0652
	0.5	0.1714	0.0974	0.0974	0.0649	0.0672

4.4 Συμπεράσματα για την Σύγκριση Συσχετισμένων Ποσοστών

Ο bootstrap έλεγχος για την ισότητα των ποσοστών είναι ανώτερος από τις υπόλοιπες μεθόδους όταν το μέγεθος του δείγματος είναι μικρό. Έχει μεγαλύτερη ισχύ από ότι η ακριβής μέθοδος και το στατιστικό του Yates και το εμπειρικό μέγεθος του ελέγχου είναι μικρότερο από το ονομαστικό επίπεδο.

Σε σχέση με το McNemar στατιστικό και το Wald στατιστικό, που είναι ισοδύναμα, ο bootstrap έλεγχος έχει μικρότερη εμπειρική ισχύ. Έχει το πλεονέκτημα ότι η χρήση του δεν

4. Διαφορά Ποσοστών

απαιτεί προϋποθέσεις κανονικότητας και δεν συσχετίζεται με το μέγεθος δείγματος του. Το εμπειρικό μέγεθος του McNemar και του Wald στατιστικού είναι πολύ κοντά στο ονομαστικό επίπεδο αλλά η ορθότητα της χρήσης του δεν μπορεί να πιστοποιηθεί μόνο από αυτό το γεγονός.

Τέλος, το πιο σημαντικό πλεονέκτημα του bootstrap ελέγχου είναι ότι λαμβάνει υπόψη του τις σύμφωνες κατηγορίες κάτι που οι άλλες μέθοδοι δεν λαμβάνουν. Μπορεί οι διάφορες στην τιμή του p-value να είναι μικρές αλλά είναι σημαντικό να συμμετέχουν και αυτές οι κατηγορίες στην διαμόρφωση του ελέγχου.

4.5 Ανεξάρτητα Ποσοστά

Ας είναι X_1 και X_2 δύο ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν τις κατανομές $B(n_1, p_1)$ και $B(n_2, p_2)$, αντίστοιχα. Με d συμβολίζουμε την διαφορά των πιθανοτήτων επιτυχίας των τυχαίων μεταβλητών X_1 και X_2 , δηλαδή $d = p_1 - p_2$. Έστω x_i είναι η παρατηρούμενη τιμή της τυχαίας μεταβλητής X_i , για $i=1,2$. Τότε ο εκτιμητής μέγιστης πιθανοφάνειας του d , $\hat{d} = d$ είναι ίσος με την διαφορά $p_1 - p_2$, όπου $p_i = \frac{x_i}{n_i}$, είναι ο εκτιμητής μέγιστης πιθανοφάνειας της πιθανότητας επιτυχίας p_i , για $i=1,2$.

Θα εξετάσουμε τρόπους κατασκευής $(1-2a)\%$ διαστημάτων εμπιστοσύνης για την διαφορά p των πιθανοτήτων επιτυχίας των δύο διωνυμικών τυχαίων μεταβλητών X_1 και X_2 , με την χρήση παραδοσιακών μεθόδων και με την χρήση της μεθόδου bootstrap.

4.6 Παραδοσιακοί Τρόποι Κατασκευής Διαστημάτων Εμπιστοσύνης

Ο στόχος της ενότητας αυτής είναι να παρουσιάσουμε τις παραδοσιακές μεθόδους κατασκευής διαστημάτων εμπιστοσύνης για την διαφορά δύο ανεξάρτητων διωνυμικών ποσοστών. Επιλεχθήκαν οι πιο δημοφιλείς μέθοδοι (βλ. Newcombe, 1998b).

4.6.1 Μέθοδος Wald

Το σύνηθες $100(1-2a)\%$ διάστημα εμπιστοσύνης για το d , είναι το

$$\left[(p_1 - p_2) - z_{1-a} \sqrt{\sum_{i=1}^2 \text{Var}[p_i]}, (p_1 - p_2) - z_a \sqrt{\sum_{i=1}^2 \text{Var}[p_i]} \right]$$

όπου $Var[p_i] = \frac{p_i(1-p_i)}{n}$ είναι η εκτιμωμένη διασπορά του p_i , $i=1,2$ και z_c είναι το c ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Στην βιβλιογραφία (βλ. Ghosh, 1979, Vollset, 1993, και Newcombe, 1998a και 1998b) αναφέρεται ότι η αποτελεσματικότητα της μεθόδου Wald δεν είναι καλή.

4.6.2 Υδριβική Score Μέθοδος του Newcombe

Ο Newcombe (1998b) μελέτησε τις μεθόδους κατασκευής $100(1-2a)\%$ διαστημάτων εμπιστοσύνης για την διαφορά δύο ανεξάρτητων διωνυμικών ποσοστών. Η μελέτη του περικλείει την μέθοδο του Wald με και χωρίς διόρθωση συνέχειας, την μέθοδο των Clopper-Pearson, τέσσερις μεθόδους που βασίζονται στην πιθανοφάνεια της κοινής κατανομής των X_1 και X_2 και τρεις παραλλαγές της μεθόδου του Wilson. Στην συνέχεια, πρότεινε μια μέθοδο κατασκευής $100(1-2a)\%$ διαστήματος εμπιστοσύνης για το d που στηρίζεται στην μέθοδο του Wilson για ένα διωνυμικό ποσοστό και η οποία ονομάζεται Υδριβική Score μέθοδος του Newcombe (NH) (*Newcombe's Hybrid Score Method*). Το $100(1-2a)\%$ NH διάστημα εμπιστοσύνης για το d είναι το

$$\left[(p_1 - p_2) - z_{1-a} \sqrt{\frac{l_1(1-l_1)}{n_1} + \frac{u_2(1-u_2)}{n_2}}, (p_1 - p_2) - z_a \sqrt{\frac{u_1(1-u_1)}{n_1} + \frac{l_2(1-l_2)}{n_2}} \right]$$

όπου l_i, u_i είναι το κάτω και το άνω άκρο του $100(1-2a)\%$ Wilson διαστήματος εμπιστοσύνης για το p_i , $i=1,2$.

Η μέθοδος NH είναι συντηρητική όταν το p_1 και το p_2 είναι κοντά στην μονάδα (βλ. Agresti and Caffo, 2000). Επιπλέον, δεν είναι γνωστές οι θεωρητικές ιδιότητες της μεθόδου.

4.6.3 Μέθοδος Agresti-Caffo

Οι Agresti and Caffo (2000) πρότειναν ως 95% διάστημα εμπιστοσύνης για την διαφορά, το

$$\left[(p_{L,1} - p_{L,2}) - z_{0.975} \sqrt{\sum_{i=1}^2 Var[p_{L,i}]}, (p_{L,1} - p_{L,2}) - z_{0.025} \sqrt{\sum_{i=1}^2 Var[p_{L,i}]} \right]$$

4. Διαφορά Ποσοστών

όπου $p_{L,i} = \frac{x_i + 1}{n_i + 2}$ είναι ο Laplace εκτιμητής της πιθανότητας επιτυχίας p_i , $i=1,2$ και

$Var[p_{L,i}] = \frac{p_{L,i}(1-p_{L,i})}{n_i + 2}$ είναι η εκτιμωμένη διασπορά του.

Στην ουσία η μέθοδος αυτή δεν είναι τίποτα άλλο παρά η επέκταση της μεθόδου των Agresti και Coull για δύο ανεξάρτητα διωνυμικά ποσοστά. Η διαφορά είναι ότι σε κάθε διωνυμικό δείγμα προσθέτουμε δύο ψευδο-παρατηρήσεις, μια επιτυχία και μια αποτυχία, αντί για τέσσερις.

Το μειονέκτημα της μεθόδου των Agresti και Caffo είναι ότι η χρήση της περιορίζεται μόνο στην κατασκευή 95% ονομαστικού επίπεδου διαστήματα εμπιστοσύνης καθώς οι θεωρητικές ιδιότητες της δεν έχουν ερευνηθεί.

4.7 Μέθοδοι Κατασκευής Διαστημάτων Εμπιστοσύνης μέσω της Bootstrap

Μια βασική ιδιότητα δύο ανεξάρτητων τυχαίων μεταβλητών X και Y είναι ότι η από κοινή τους πιθανότητα ισούται με το γινόμενο των πιθανοτήτων τους, $\text{Prob}\{X = x\}\text{Prob}\{Y = y\}$. Αυτήν την ιδιότητα εκμεταλλευόμαστε προκειμένου να κατασκευάσουμε διαστήματα εμπιστοσύνης μέσω της μεθόδου bootstrap για την διαφορά δύο ανεξάρτητων διωνυμικών ποσοστών.

Τα ποσοστά επιτυχίας p_1 και p_2 είναι άγνωστα και έστω \hat{p}_1 και \hat{p}_2 , δύο εκτιμήσεις τους. Η από κοινού κατανομή των τυχαίων μεταβλητών X_1 και X_2 , δηλαδή το γινόμενο των διωνυμικών κατανομών $B(n_1, \hat{p}_1)$ και $B(n_2, \hat{p}_2)$, είναι η :

$$\text{Prob}\{X_1 = x_1, X_2 = x_2\} = \binom{n_1}{x_1} \hat{p}_1^{x_1} (1 - \hat{p}_1)^{n_1 - x_1} \binom{n_2}{x_2} \hat{p}_2^{x_2} (1 - \hat{p}_2)^{n_2 - x_2} \quad (4.6)$$

Δοθέντων των n_1 και n_2 , κάθε πιθανό αποτέλεσμα (x_1, x_2) ανήκει στο σύνολο $A = \{0, 1, \dots, n_1\} \times \{0, 1, \dots, n_2\}$. Το σύνολο A είναι διακριτό και επομένως είναι εύκολο να βρούμε την κατανομή του εκτιμητή d της διαφοράς των ποσοστών (Conlon-Thomas, 1990) από την κατανομή:

$$\text{Prob}\{d = d\} = \sum_{\{(x_1, x_2) \in A: \frac{x_1}{n_1} - \frac{x_2}{n_2} = d\}} \text{Prob}\{X_1 = x_1, X_2 = x_2\} \quad (4.7)$$

4.7.1 Bootstrap Percentile Διαστήματα Εμπιστοσύνης

Όταν η διαφορά των ποσοστών εκτιμάται από το $d = p_1 - p_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$ τότε η κατανομή (4.6)

γράφεται :

$$\text{Prob}\{X_1 = x_1, X_2 = x_2\} = \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} (1-p_2)^{n_2-x_2}$$

Η εκτίμηση bootstrap της κατανομής αυτής είναι η

$$\text{Prob}_* \{X_1^* = x_1^*, X_2^* = x_2^*\} = \binom{n_1}{x_1^*} (p_1)^{x_1^*} (1-p_1)^{n_1-x_1^*} \binom{n_2}{x_2^*} (p_2)^{x_2^*} (1-p_2)^{n_2-x_2^*}$$

και επομένως η bootstrap κατανομή της (4.7) είναι η

$$\text{Prob}_* \{d^* = d^*\} = \sum_{\left\{ \binom{x_1^*}{x_1^*}, \binom{x_2^*}{x_2^*} : \frac{x_1^*}{n_1} - \frac{x_2^*}{n_2} = d^* \right\}} \text{Prob}_* \{X_1^* = x_1^*, X_2^* = x_2^*\}$$

Το $(1-2a)\%$ bootstrap percentile διάστημα εμπιστοσύνης για το p είναι το:

$$\left[\inf \left\{ d^* : \sum_{d^*} \text{Prob}_* \{d^* = d^*\} \geq a \right\}, \inf \left\{ d^* : \sum_{d^*} \text{Prob}_* \{d^* = d^*\} \geq 1-a \right\} \right] \quad (4.8)$$

Προφανώς όταν το $p_i = 0$ ή 1 , $i = 1, 2$ τότε το διάστημα εμπιστοσύνης βρίσκεται με την χρήση των ποσοστιαίων σημείων της διωνυμικής μεταβλητής που το ποσοστό επιτυχίας της εκτιμάται ότι είναι διάφορο του μηδέν ή του ένα.

4.7.2 Τροποποιημένη Bootstrap Percentile Διαστήματα Εμπιστοσύνης

Στο δεύτερο κεφαλαίο είδαμε ότι η τροποποιημένη bootstrap είναι ανώτερη της κλασσικής μεθόδου bootstrap και επομένως θα ήταν καλό να την εξετάσουμε και για την διαφορά των ποσοστών. Προσθέτουμε λοιπόν δύο παρατηρήσεις σε κάθε διωνυμική μεταβλητή, μια επιτυχία και μια αποτυχία και συνεχίζουμε όπως στην προηγούμενη παράγραφο.

Το $(1-2a)\%$ bootstrap percentile διάστημα εμπιστοσύνης για το p στην τροποποιημένη bootstrap είναι πάλι το διάστημα εμπιστοσύνης (4.8) με την διαφορά ότι όπου $p_i = p_{L,i}$ είναι

ο εκτιμητής Laplace, που δίνεται από την σχέση (2.3) και $d^* = \frac{x_1^* + 1}{n_1 + 2} - \frac{x_2^* + 1}{n_2 + 2}$, $x_i^* = x_i^* + 1$ και

$$n_i = n_i + 2.$$

4. Διαφορά Ποσοστών

4.7.3 Conlon-Thomas Bootstrap Διαστήματα Εμπιστοσύνης

Οι Conlon and Thomas (1990) ήταν οι πρώτοι που ασχοληθήκαν με την κατασκευή bootstrap percentile διαστημάτων εμπιστοσύνης. Το $(1-2a)\%$ bootstrap percentile διάστημα εμπιστοσύνης σύμφωνα με την μέθοδο των Conlon-Thomas είναι της μορφής (4.8) με την διαφορά ότι τα δειγματικά ποσοστά εκτιμώνται από την σχέση:

$$p'_i = \begin{cases} n_i (1-(2a)^{1/n_i}) & \text{αν } x_i = 0 \\ \frac{x_i}{n_i} & \text{αν } x_i \neq 0, n_i, i = 1, 2 \\ (2a)^{1/n_i} & \text{αν } x_i = n_i \end{cases}$$

Όταν $x_i = 0$ ή $x_i = n_i$, η διόρθωση που προτείνουν, έχει παρουσιαστεί από τον Louis (1981) και εφαρμόζεται για να παράγονται διαστήματα εμπιστοσύνης όταν έχουμε μηδενικές ή μοναδιαίες εκτιμήσεις των δειγματικών ποσοστών

4.8 Μέθοδοι Εκτίμησης των Διαστημάτων Εμπιστοσύνης

Για την εκτίμηση των διαστημάτων εμπιστοσύνης που περιγράψαμε προηγουμένως θα ασχοληθούμε με τα κριτήρια που χρησιμοποιήσουμε και στο 2^ο κεφάλαιο: την πιθανότητα κάλυψης (*Coverage Probability*) και το αναμενόμενο πλάτος (*Average Width*). Σύμφωνα με το δεύτερο κεφαλαίο τα bootstrap διάστημα εμπιστοσύνης συμπεριφέρονται καλύτερα όταν το ονομαστικό επίπεδο είναι 95%.

Για κάθε διαδικασία κατασκευής διαστημάτων εμπιστοσύνης για την διαφορά των ποσοστών επιθυμούμε η πιθανότητα κάλυψης να είναι ίση με το προκαθορισμένο ονομαστικό επίπεδο a . Η πιθανότητα κάλυψης του διαστήματος εμπιστοσύνης για p , $CP(n_1, p_1, n_2, p_2)$ για σταθερά μεγέθη δείγματος (n_1, n_2) και για σταθερές πιθανότητες (p_1, p_2) δίνεται από την ακόλουθη σχέση:

$$CP(n_1, p_1, n_2, p_2) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} I(x_1, p_1, x_2, p_2) \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} (1-p_2)^{n_2-x_2}$$

όπου η συνάρτηση $I(x_1, p_1, x_2, p_2)$ είναι μια δείκτρια συνάρτηση που ισούται με την μονάδα, όταν για τις τιμές (x_1, x_2) το διάστημα εμπιστοσύνης περιέχει το $p_1 - p_2$, και ισούται με το μηδέν διαφορετικά (p_1, p_2) .

4^ο Κεφάλαιο

Για τον υπολογισμό της πιθανότητας κάλυψης $CP(n_1, p_1, n_2, p_2)$ θα παραχθούν 10000 τυχαία (p_1, p_2) από το $[0,1] \times [0,1]$. Σε κάθε δείγμα θα εκτιμήσουμε το $CP(n_1, p_1, n_2, p_2)$ από το μέσο όρο \overline{CP}_{n_1, n_2} , για σταθερά (n_1, n_2) . Επειδή η τιμή \overline{CP}_{n_1, n_2} είναι απλά ο μέσος όρος, θα βρούμε επιπλέον το ποσοστό του $CP(n_1, p_1, n_2, p_2)$ που είναι μικρότερες από 0.93 και μεγαλύτερες από 0.97. Τα ποσοστά αυτά θέλουμε να είναι μικρά καθώς το αντίθετο σημαίνει ότι υπάρχει μεγάλη διασπορά στις τιμές των $CP(n_1, p_1, n_2, p_2)$. Τέλος, θα βρούμε και την μικρότερη και την μεγαλύτερη πιθανότητα κάλυψης από τις προσομοιωμένες τιμές. Αυτό θα μας βοηθήσει να καταλάβουμε πόσο διασπαρμένες είναι οι τιμές της πιθανότητας κάλυψης.

Το άλλο κριτήριο είναι το πλάτος του διαστήματος εμπιστοσύνης. Επιθυμούμε το πλάτος του διαστήματος εμπιστοσύνης να είναι όσο το δυνατόν πιο μικρό. Για σταθερά για σταθερά μεγέθη δείγματος (n_1, n_2) και για σταθερές πιθανότητες (p_1, p_2) , το αναμενόμενο πλάτος ενός διαστήματος εμπιστοσύνης είναι

$$AW(n_1, p_1, n_2, p_2) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} W(x_1, n_1, x_2, n_2) \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} (1-p_2)^{n_2-x_2}$$

όπου $W(x_1, n_1, x_2, n_2)$ είναι το πλάτος του διαστήματος εμπιστοσύνης για τις τιμές (x_1, x_2) .

Όπως και για την πιθανότητα κάλυψης το αναμενόμενο πλάτος θα υπολογιστεί μέσω των προσομοιωμένων τιμών λαμβάνοντας τον μέσο όρο των τιμών του αναμενόμενου πλάτος για κάθε προσομοιωμένη τιμή.

4.9 Εκτίμηση των Διαστημάτων Εμπιστοσύνης

Στον Πίνακα 4.8 παρουσιάζονται οι τιμές των κριτηρίων για τις συγκρινόμενες μεθόδους όταν $5 \leq n_1, n_2 \leq 20$. Παρατηρούμε μια συμπεριφορά ανάλογη με αυτήν που παρατηρήσαμε στο δεύτερο κεφάλαιο. Το διάστημα εμπιστοσύνης του Wald και το bootstrap percentile διάστημα εμπιστοσύνης (BP) έχουν μικρή πιθανότητα κάλυψης. Η μέθοδος των Conlon-Thomas (CT) έχει πιθανότητα κάλυψης κάτω από το ονομαστικό επίπεδο κάλυψης αλλά όσο τα (n_1, n_2) η πιθανότητα κάλυψης αυξάνει. Απεναντίας η μέθοδος των Agresti-Caffo (AC), η υδριβική μέθοδος του Newcombe (HN) και η τροποποιημένη bootstrap (MBP) παράγουν διαστήματα εμπιστοσύνης που έχουν πιθανότητα κάλυψης κοντά στο ονομαστικό επίπεδο. Από τις μεθόδους αυτές η HN καλύτερα και να έπεται η MBP όσον αφορά το μέσο αναμενόμενο πλάτος. Η μέθοδος των Agresti-Caffo έχει το μικρότερο ποσοστό πιθανότητας

4. Διαφορά Ποσοστών

κάλυψης κάτω από 0.93 και την μεγαλύτερη τιμή $\min CP_{n_1, n_2}$. Από την άλλη το ποσοστό των πιθανοτήτων κάλυψης που είναι άνω του 0.97 είναι αρκετά μεγάλο ακόμα και όταν τα (n_1, n_2) αυξάνονται. Απεναντίας τα τροποποιημένα bootstrap διαστήματα εμπιστοσύνης όσο τα (n_1, n_2) τόσο μειώνεται το ποσοστό των πιθανοτήτων κάλυψης που είναι μεγαλύτερα από 0.97. Για παράδειγμα, όταν $n_1 = n_2 = 20$ τότε το ποσοστό αυτό είναι το μικρότερο από τις τρεις μεθόδους.

Πίνακας 4.8

95% ονομαστικού επιπέδου διαστήματα εμπιστοσύνης για την διαφορά ποσοστών

(n_1, n_2)	Μέθοδος	\overline{CP}_{n_1, n_2}	\overline{AW}_{n_1, n_2}	$CP_{n_1, n_2} \leq 0.93$	$CP_{n_1, n_2} \geq 0.97$	$\min CP_{n_1, n_2}$	$\max CP_{n_1, n_2}$
(5,5)	Wald	0.8144	0.8129	96.09%	0.00%	0.0262	0.9417
	AC	0.9654	0.9057	2.60%	37.18%	0.8884	1.0000
	HN	0.9535	0.8601	18.55%	25.32%	0.8356	1.0000
	BP	0.7175	0.8111	99.74%	0.00%	0.0000	0.9363
	CT	0.9070	1.1444	26.07%	64.36%	0.0005	1.0000
	MBP	0.9594	0.8889	12.21%	43.88%	0.8415	1.0000
(10,5)	Wald	0.8410	0.7383	97.91%	0.00%	0.0496	0.9510
	AC	0.9638	0.7980	1.30%	28.19%	0.8956	1.0000
	HN	0.9534	0.7616	5.05%	18.18%	0.8612	1.0000
	BP	0.7845	0.7273	96.82%	0.00%	0.0000	0.9493
	CT	0.9136	0.9515	27.48%	37.55%	0.0010	1.0000
	MBP	0.9633	0.7951	5.84%	32.77%	0.7823	1.0000
(10,10)	Wald	0.8896	0.6488	87.79%	0.00%	0.0314	0.9493
	AC	0.9600	0.6732	0.95%	16.83%	0.9081	1.0000
	HN	0.9535	0.6540	7.80%	16.46%	0.8443	1.0000
	BP	0.8252	0.6380	81.64%	0.00%	0.0002	0.9508
	CT	0.9179	0.7289	32.89%	20.72%	0.0000	1.0000
	MBP	0.9577	0.6722	6.79%	18.17%	0.7988	1.0000
(20,5)	Wald	0.8274	0.6697	95.52%	0.00%	0.0560	0.9679
	AC	0.9646	0.7271	0.80%	33.72%	0.8984	1.0000
	HN	0.9535	0.6879	2.67%	10.53%	0.8650	1.0000
	BP	0.7951	0.6595	97.95%	0.00%	0.0005	0.9392
	CT	0.9215	0.8722	22.70%	59.23%	0.0000	1.0000
	MBP	0.9535	0.7223	2.67%	10.53%	0.8650	1.0000
(20,10)	Wald	0.8998	0.5703	91.96%	0.00%	0.0900	0.9640
	AC	0.9588	0.5876	0.44%	14.75%	0.9111	1.0000
	HN	0.9533	0.5720	1.10%	10.33%	0.8770	1.0000
	BP	0.8644	0.5636	88.63%	0.00%	0.0012	0.9512
	CT	0.9306	0.6299	24.79%	18.31%	0.0007	1.0000
	MBP	0.9585	0.5860	2.32%	11.93%	0.8227	1.0000
(20,20)	Wald	0.9223	0.4807	41.30%	0.00%	0.0618	0.9502
	AC	0.9559	0.4878	0.24%	7.38%	0.9212	1.0000
	HN	0.9529	0.4807	2.33%	9.01%	0.8473	1.0000
	BP	0.8834	0.4760	46.72%	0.00%	0.0005	0.9513
	CT	0.9340	0.5009	16.92%	6.61%	0.0002	1.0000
	MBP	0.9528	0.4820	2.94%	7.04%	0.7992	1.0000

Στον Πίνακα 4.9 φαίνονται τα αποτελέσματα για $n_1 = 30$. Παρατηρούμε ότι σε γενικές γραμμές δεν έχουν αλλάξει πολλά πράγματα. Οι μέθοδοι που είχαν καλύτερα αποτελέσματα

4^ο Κεφάλαιο

νωρίτερα έχουν και τώρα. Το μόνο αλλάζει είναι ότι η τροποποιημένη bootstrap παράγει σταθερά διαστήματα εμπιστοσύνης με το μικρότερο ποσοστό πιθανότητας κάλυψης μεγαλύτερο από 0.97.

Πίνακας 4.9

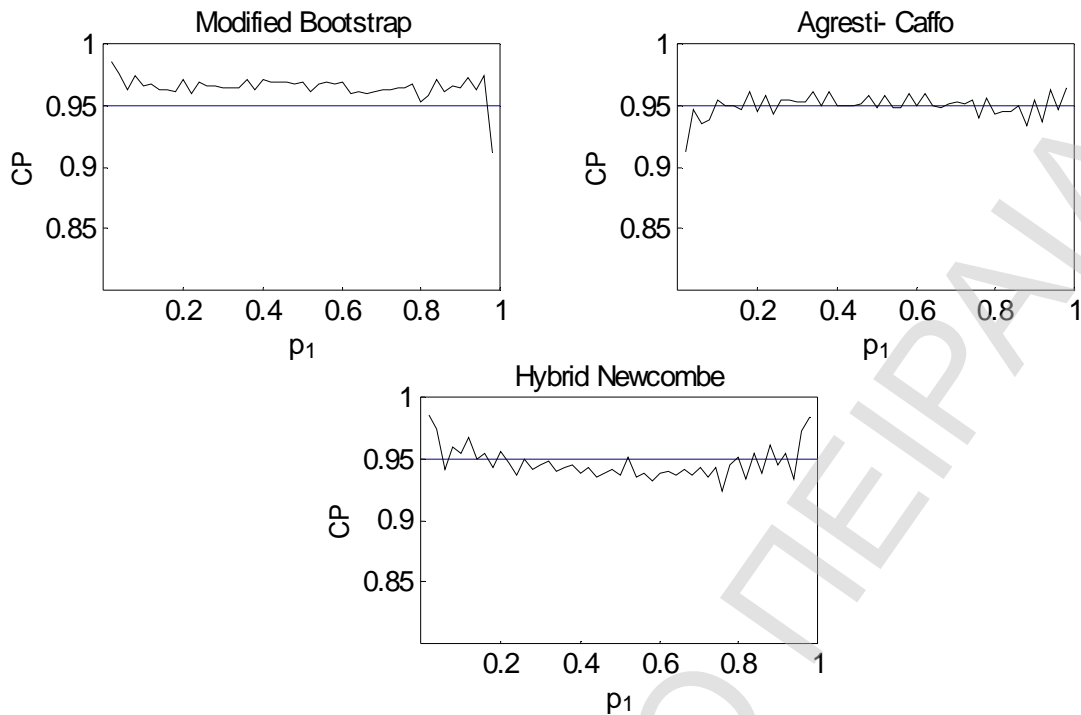
95% ονομαστικού επιπέδου διαστήματα εμπιστοσύνης για την διαφορά ποσοστών $n_1 = 30$

(n_1, n_2)	Μέθοδος	\overline{CP}_{n_1, n_2}	\overline{AW}_{n_1, n_2}	$CP_{n_1, n_2} \leq 0.93$	$CP_{n_1, n_2} \geq 0.97$	min CP_{n_1, n_2}	max CP_{n_1, n_2}
(30,5)	Wald	0.8952	0.6381	95.45%	0.00%	0.0821	0.9617
	AC	0.9591	0.6998	0.62%	43.03%	0.9038	1.0000
	HN	0.9530	0.6564	2.12%	6.12%	0.8685	0.9994
	BP	0.8740	0.6258	95.92%	0.00%	0.0017	0.9544
	CT	0.9325	0.8434	18.68%	64.57%	0.0005	1.0000
	MBP	0.9593	0.6923	2.15%	37.83%	0.8045	1.0000
(30,10)	Wald	0.9265	0.5358	93.19%	0.00%	0.1182	0.9659
	AC	0.9553	0.5535	0.45%	19.23%	0.8950	1.0000
	HN	0.9526	0.5369	1.46%	6.77%	0.8915	0.9994
	BP	0.9023	0.5316	89.94%	0.00%	0.0021	0.9543
	CT	0.9399	0.5969	25.50%	25.71%	0.0011	1.0000
	MBP	0.9550	0.5516	1.63%	11.49%	0.8002	1.0000
(30,20)	Wald	0.9234	0.4408	34.03%	0.00%	0.1693	0.9570
	AC	0.9553	0.4465	0.22%	6.66%	0.9191	1.0000
	HN	0.9526	0.4406	0.55%	6.34%	0.8856	1.0000
	BP	0.9023	0.4390	30.50%	0.00%	0.0021	0.9542
	CT	0.9399	0.4584	9.90%	6.20%	0.0022	1.0000
	MBP	0.9550	0.4456	0.84%	4.56%	0.7915	1.0000
(30,30)	Wald	0.9304	0.3977	10.30%	0.00%	0.0702	0.9612
	AC	0.9543	0.4013	0.08%	4.81%	0.9246	1.0000
	HN	0.9524	0.3975	0.81%	4.95%	0.8562	1.0000
	BP	0.9023	0.3942	15.02%	0.00%	0.0028	0.9542
	CT	0.9402	0.4046	5.10%	6.00%	0.0029	1.0000
	MBP	0.9528	0.4002	1.65%	3.56%	0.7965	1.0000

Το μεγάλο μειονέκτημα της τροποποιημένης bootstrap είναι ότι η μικρότερη τιμή πιθανότητας κάλυψης είναι κοντά στο 0.8 σταθερά. Ωστόσο τα να απορρίψουμε την μέθοδο για τον λόγο αυτό δεν φαίνεται καλή ιδέα. Στα Γραφήματα που ακολουθούν φαίνεται η πιθανότητα κάλυψης για $p_2 = 0.4$ και για συγκεκριμένες τιμές των (n_1, n_2) .

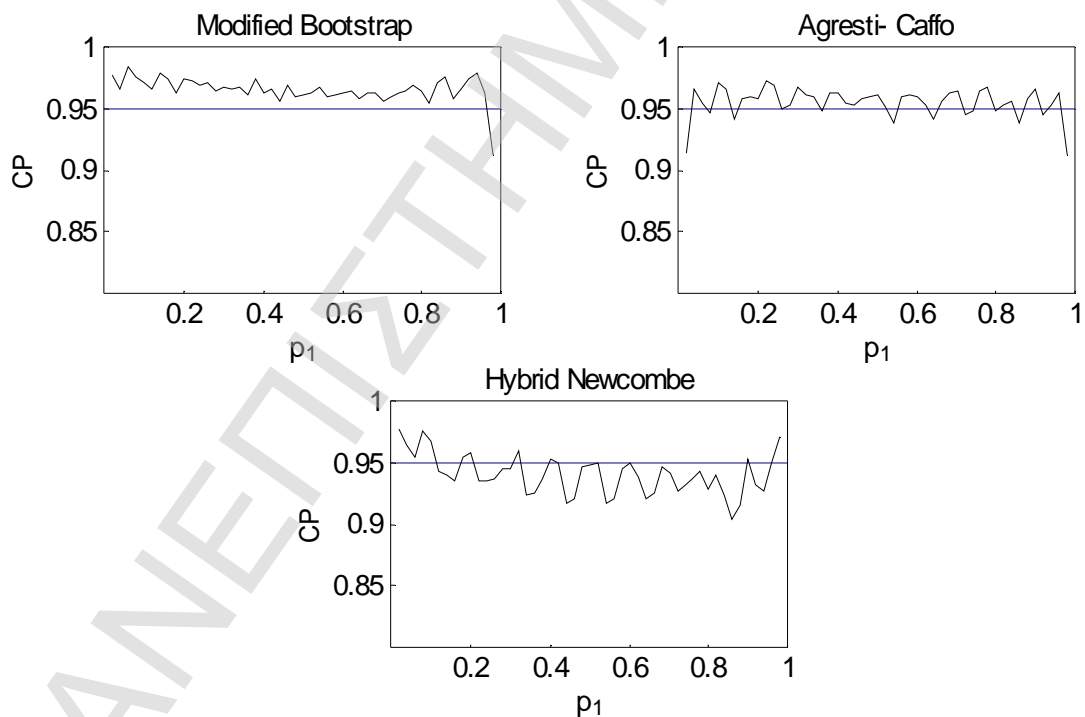
Τα Γραφήματα φαίνονται παράδοξα με την έννοια ότι φαίνεται η MBP να είναι κοντά και πάνω από το επιθυμητό ονομαστικό επίπεδο (οριζόντια γραμμή), ενώ οι άλλες μέθοδοι φαίνονται ικανοποιητικές για μερικές τιμές των (n_1, n_2) . Πάντως πρέπει να τονίσουμε ότι δεν αποκλείεται η περίπτωση να ισχύουν και τα αντίστροφα για τιμές των (n_1, n_2) και p_2 .

4. Διαφορά Ποσοτών



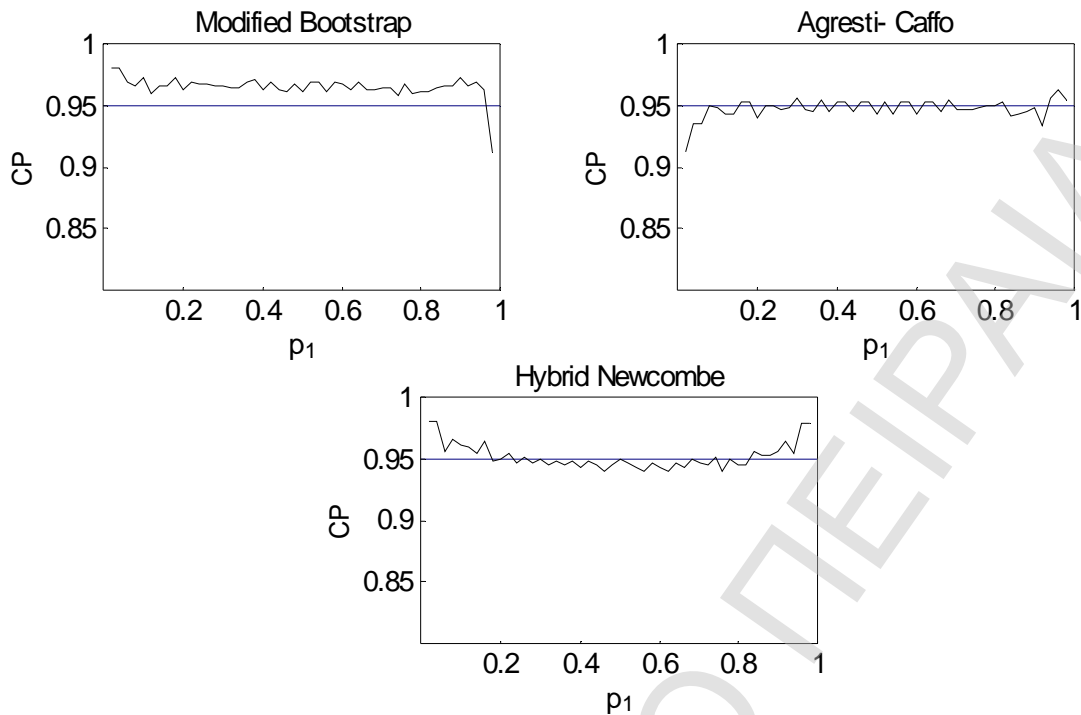
Γράφημα 4.1

$CP(n_1, p_1, n_2, p_2)$ για $p_2 = 0.4$ και $(n_1, n_2) = (10, 5)$



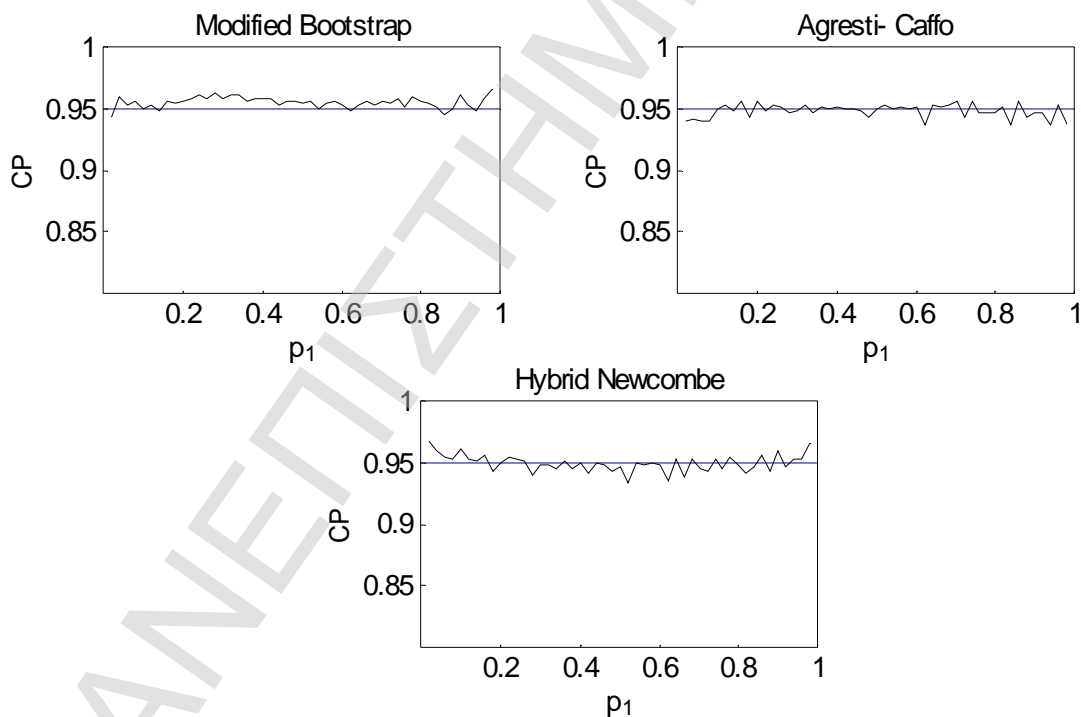
Γράφημα 4.2

$CP(n_1, p_1, n_2, p_2)$ για $p_2 = 0.4$ και $(n_1, n_2) = (20, 5)$



Γράφημα 4.3

$CP(n_1, \rho_1, n_2, p_2)$ για $p_2 = 0.4$ και $(n_1, n_2) = (30, 5)$



Γράφημα 4.4

$CP(n_1, \rho_1, n_2, p_2)$ για $p_2 = 0.4$ και $(n_1, n_2) = (20, 10)$

4. Διαφορά Ποσοστών

4.10 Συμπεράσματα για τη Διαφορά Δύο Ανεξάρτητων Διωνυμικών Ποσοστών

Τα βασικά κριτήρια που πρέπει να ισχύουν για να έχει καλή απόδοση ένα διάστημα εμπιστοσύνης ισχύει στην περίπτωση της τροποποιημένης bootstrap. Δεν μπορεί να παραγνωρισθεί το γεγονός ότι έχει ανεπιθύμητες ιδιότητες, όπως για παράδειγμα η τιμή του $\min CP_{n_1, n_2}$ που είναι κοντά στο 0.8. Από τα Γραφήματα όμως είδαμε ότι μπορεί να υπάρχουν περιπτώσεις όπου το 95% διάστημα εμπιστοσύνης της τροποποιημένης bootstrap φαίνεται να είναι καλύτερο από τις μεθόδους των Newcombe και Agresti-Caffo.

5^ο Κεφάλαιο

Δεσμευμένος Έλεγχος Ανεξαρτησίας (όταν ισχύει το μοντέλο RC)

5.0 Εισαγωγή

Για την ανάλυση ενός $I \times J$ πίνακα συνάφειας που έχει διατάξιμες μεταβλητές ταξινόμησης, η χρήση μοντέλων που χρησιμοποιούν την πληροφορία της διάταξης αντί των απλών λογαριθμικών μοντέλων που χρησιμοποιούνται για ονομαστικές μεταβλητές, μας παρέχει πολλά πλεονεκτήματα. Για διατάξιμα κατηγορικά δεδομένα, υπάρχει μια πληθώρα μοντέλων τα οποία είναι πιο οικονομικά και έχουν απλούστερη ανάλυση από ότι αυτά των ονομαστικών μεθόδων (Davis, 1988).

Ο Goodman (1979) παρουσίασε δύο μοντέλα συνάφειας γνωστά και ως “Linear by Linear Association”(LL) και “Row and Column Effects” (RC) μοντέλα. Αυτά τα μοντέλα μας επιτρέπουν να εκτιμήσουμε τις τιμές των σκορ των κατηγοριών γραμμών και στηλών, και μας δίνουν την δυνατότητα να χρησιμοποιήσουμε το χι-τετράγωνο στατιστικό για τον δεσμευμένο έλεγχο της υπόθεσης της ανεξαρτησίας μεταξύ των μεταβλητών των γραμμών και των στηλών. Το LL μοντέλο είναι ένα λογαριθμογραμμικό μοντέλο που έχει μόλις έναν βαθμό ελευθερίας λιγότερο από το μοντέλο της ανεξαρτησίας και προσφέρεται για δεσμευμένο έλεγχο της ανεξαρτησίας που ασυμπτωτικά θα ακολουθεί την χι τετράγωνο κατανομή με έναν βαθμό ελευθερίας .

Ο Agresti (1990) έδειξε ότι όταν το LL μοντέλο ισχύει τότε ο διατάξιμος έλεγχος που βασίζεται στο μοντέλο είναι ασυμπτωτικά πιο ισχυρός. Για το μοντέλο RC ο Haberman (1981) συμπέρανε ότι η ασυμπτωτική κατανομή για την ελεγχοσυνάρτηση της υπόθεσης της ανεξαρτησίας είναι πιο περίπλοκη. Ο Haberman (1981) έδειξε ότι ο έλεγχος της ανεξαρτησίας που βασίζεται στο canonical correlation μοντέλο είναι ασυμπτωτικά ισοδύναμο με τον έλεγχο της ανεξαρτησίας όταν το μοντέλο RC ισχύει. Οι προσεγγιστικές κριτικές τιμές γι’ αυτούς τους ελέγχους μπορούν να παρθούν από την κατανομή της μέγιστης ιδιοτιμής ενός κεντρικού Wishart πίνακα. Ωστόσο, η δυσκολία του υπολογισμού πολλαπλών

5. Δεσμευμένος Έλεγχος Ανεξαρτησίας (όταν ισχύει το Μοντέλο RC)

ολοκληρωμάτων πρέπει να αναληφθεί προκειμένου να υπολογιστούν για την κατανομή του μέγιστης ιδιοτιμής του πίνακα. Οι Hanumara and Tompson (1968) πρότειναν μια προσεγγιστική εξίσωση για να υπολογιστεί η αθροιστική συνάρτηση κατανομής της μέγιστης ιδιοτιμής. Αυτή η προσεγγιστική μέθοδος έχει κάποιους περιορισμούς και επομένως σε πολλές περιπτώσεις δεν είναι χρήσιμη. Με βάση το σκεπτικό αυτό, οι Jeong et al. (2005) πρότειναν να πραγματοποιηθεί ο έλεγχος αυτός μέσω της μεθόδου bootstrap.

5.1 Το Μοντέλο της Ανεξαρτησίας

Ας θεωρήσουμε έναν $I \times J$ πίνακα συνάφειας όπου με X θα συμβολίζουμε την μεταβλητή γραμμής και με Y την μεταβλητή των στηλών. Με n_{ij} δηλώνουμε την παρατηρούμενη συχνότητα του κελιού της i -γραμμής και της j -στήλης του πίνακα, με m_{ij} τις αντίστοιχες αναμενόμενες συχνότητες κάτω από κάποιο μοντέλο και με $n_{i.} = \sum_{j=1}^J n_{ij}$ ($n_{.j} = \sum_{i=1}^I n_{ij}$) τα περιθώρια αθροίσματα των γραμμών (στηλών) για $1 \leq i \leq I$ ($1 \leq j \leq J$). Υποθέτουμε ότι τα n_{ij} ακολουθούν την πολυωνυμική κατανομή με συνολικό μέγεθος δειγματος $n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ και με πιθανότητες $p_{ij} \geq 0$, $1 \leq i \leq I, 1 \leq j \leq J$. Επίσης με $p_{ij} = \frac{n_{ij}}{n}$, $p_{i.} = \frac{n_{i.}}{n}$ και $p_{.j} = \frac{n_{.j}}{n}$ συμβολίζουμε τις σχετικές συχνότητες που αντιστοιχούν στα n_{ij} , $n_{i.}$ και $n_{.j}$.

Η μαθηματική έκφραση του ελέγχου της ανεξαρτησίας είναι:

$$H_0 : p_{ij} = p_{i.}p_{.j} \text{ έναντι } H_1 : p_{ij} \neq p_{i.}p_{.j}$$

για $1 \leq i \leq I, 1 \leq j \leq J$, όπου $p_{i.}$ είναι οι περιθώριες πιθανότητες των γραμμών και $p_{.j}$ είναι οι περιθώριες πιθανότητες των στηλών.

Οι εκτιμητές μέγιστης πιθανοφάνειας των p_{ij} όταν ισχύει η ανεξαρτησία είναι οι $\hat{g}_{ij} = p_{i.}p_{.j}$ για $1 \leq i \leq I, 1 \leq j \leq J$ και επομένως τα m_{ij} δίνονται από την σχέση

$$m_{ij} = np_{i.}p_{.j} \quad 1 \leq i \leq I, 1 \leq j \leq J$$

Η λογαριθμογραμμική έκφραση του μοντέλου της ανεξαρτησίας είναι

$$\log m_{ij} = I + I_i^X + I_j^Y$$

όπου τα I_i^X και I_j^Y είναι οι κύριες επιδράσεις των γραμμών και των στηλών. Από τις παραμέτρους I_i^X , $i=1, \dots, I$ ($I_j^Y, j=1, \dots, J$) μόνο οι $I-1$ ($J-1$) είναι ανεξάρτητες. Για να

5^ο Κεφάλαιο

είναι προσδιορίσιμες οι παραπάνω παράμετροι πρέπει να θέσουμε κάποιους περιορισμούς, όπως για παράδειγμα

$$\sum_{i=1}^I I_i^X = \sum_{j=1}^J I_j^Y = 0$$

ή

$$I_i^X = I_j^Y = 0$$

Για να ελέγξουμε την καλή προσαρμογή του μοντέλου της ανεξαρτησίας στα δεδομένα μας χρησιμοποιούμε συνήθως είτε το χι-τετράγωνο στατιστικό του Pearson

$$X^2(I) = n \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - \hat{g}_{ij})^2}{\hat{g}_{ij}}$$

είτε το στατιστικό του λόγου πιθανοφάνειας

$$G^2(I) = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \left(\frac{p_{ij}}{\hat{g}_{ij}} \right)$$

Τα στατιστικά $X^2(I)$ και $G^2(I)$ παράγουν συνεπείς και ασυμπτωτικά αμερόληπτους ελέγχους για την ανεξαρτησία. Όταν το μέγεθος του δείγματος είναι αρκετά μεγάλο τότε τα δύο στατιστικά είναι ισοδύναμα. Ωστόσο αν χρησιμοποιήσουμε την πληροφορία της διάταξης για τις μεταβλητές μπορούμε να έχουμε μεγαλύτερη ισχύ για να ανιχνεύσουμε συγκεκριμένους τύπους υποθέσεων εναλλακτικών της ανεξαρτησίας.

5.2 Το Μοντέλο RC

Ο Goodman (1979) πρότεινε το μοντέλο RC για διατάξιμες μεταβλητές ταξινόμησης, του οποίου η λογαριθμική έκφραση δίνεται από την σχέση:

$$\log m_{ij} = I + I_i^X + I_j^Y + f m_i n_j$$

όπου οι f , m_i και n_j είναι παράμετροι.

Για να είναι προσδιορίσιμοι οι παράμετροι του μοντέλου χρειάζεται εκτός από τους περιορισμούς για τα $\{I_i^X\}$ και τα $\{I_j^Y\}$, να θέσουμε περιορισμούς θέσης και κλίμακας στις παραμέτρους $\{m_i\}$ και $\{n_j\}$:

$$\sum_{i=1}^I w_{1i} m_i = \sum_{j=1}^J w_{2j} n_j = 0 \text{ και } \sum_{i=1}^I w_{1i} m_i^2 = \sum_{j=1}^J w_{2j} n_j^2 = 1$$

5. Δεσμευμένος Έλεγχος Ανεξαρτησίας (όταν ισχύει το Μοντέλο RC)

όπου $\{w_{1i}, 1 \leq i \leq I\}$ και $\{w_{2j}, 1 \leq j \leq J\}$ είναι συστήματα βαρών. Συνήθως επιλέγουμε $w_{1i} = w_{2j} = 1$ ή $w_{1i} = p_{.i}$ και $w_{2j} = p_{.j}$, $1 \leq i \leq I, 1 \leq j \leq J$. Πάντως η επιλογή του συστήματος στάθμισης δεν επηρεάζει την προσαρμογή του μοντέλου ούτε την αναμενομένη πιθανότητα των κελιών αλλά μόνο τις τιμές των σκορ και την ερμηνεία τους.

Οι βαθμοί ελευθερίας του μοντέλου είναι $(I-2)(J-2)$. Το μοντέλο αυτό δεν είναι λογαριθμικό γιατί τα m_{ij} είναι πολλαπλασιαστικές συναρτήσεις των παραμέτρων $\{m_i\}$ και $\{n_j\}$. Το μοντέλο RC αντιμετωπίζει τις μεταβλητές ταξινομήσεις ως ονομαστικές μεταβλητές, δηλαδή είναι ανεπηρέαστο σε αναδιατάξεις των γραμμών ή/και των στηλών του πίνακα συνάφειας.

Η παράμετρος f καλείται παράμετρος εσωτερικής συνάφειας (*intrinsic association*) και είναι ένα σταθμικό μέτρο συσχέτισης μεταξύ των γραμμών και των στηλών του πίνακα και έχει άμεση σχέση με την συσχέτιση σε μια διδιάστατη κανονική κατανομή. Οι τιμές των παραμέτρων m_i και n_j προσδιορίζουν την φύση της συνάφειας, που περιγράφεται από τον λογάριθμο των τοπικών odds ratio $f(m_{i+1} - m_i)(n_{j+1} - n_j)$.

Οι εκτιμήσεις των παραμέτρων δεν δίνονται σε κλειστή μορφή και επομένως απαιτείται η χρήση επαναληπτικών μεθόδων. Οι Goodman (1979) και Becker (1990) χρησιμοποίησαν την μονοδιάστατη μέθοδο του Newton ενώ ο Haberman (1979) τον αλγόριθμο των Newton-Raphson.

Το μοντέλο RC θέτει συγκεκριμένη δομή στον όρο της αλληλεπίδρασης του κορεσμένου μοντέλου μπορεί να θεωρηθεί ως μια ειδική περίπτωση του κορεσμένου μοντέλου. Η λογαριθμογραμμική έκφραση του κορεσμένου μοντέλου ενός $I \times J$ πίνακα συνάφειας είναι:

$$\log m_{ij} = I + I_i^X + I_j^Y + I_{ij}^{XY} = \text{independence} + I_{ij}^{XY}$$

Επομένως αν θέσουμε $I_{ij}^{XY} = f m_i n_j$, αρκούν $(I-2) + (J-2) + 1$ παράμετροι για τις $(I-1)(J-1)$ παραμέτρους I_{ij}^{XY} . Επομένως ο έλεγχος καλής προσαρμογής του μοντέλου της ανεξαρτησίας σε έναν διδιάστατο πίνακα συνάφειας είναι ισοδύναμος με τον έλεγχο της υπόθεσης

$$H_0 : f = 0 \text{ έναντι } H_1 : f \neq 0 \quad (5.1)$$

στο μοντέλο RC όπου η παράμετρος f είναι η παράμετρος εσωτερικής συνάφειας.

5.3 Δεσμευμένος Έλεγχος Ανεξαρτησίας

Όταν το μοντέλο RC έχει καλή προσαρμογή στα δεδομένα μας τότε μπορούμε να το χρησιμοποιήσουμε για να ελέγξουμε αν και το μοντέλο της ανεξαρτησίας έχει καλή προσαρμογή στα δεδομένα μας. Όπως είδαμε στην παράγραφο 5.1 ο δεσμευμένος έλεγχος της ανεξαρτησίας έχει μεγαλύτερη ισχύ από τον απλό χι-τετράγωνο έλεγχο καλής προσαρμογής.

Ας υποθέσουμε ότι το μοντέλο RC προσαρμόζεται καλά στα δεδομένα μας και ας συμβολίσουμε με \hat{f}_{ij} τον εκτιμητή μέγιστης πιθανοφάνειας της πιθανότητας p_{ij} κάτω από το μοντέλο RC. Ο έλεγχος της καλής προσαρμογής του μοντέλου της ανεξαρτησίας όταν ισχύει το μοντέλο RC είναι ισοδύναμος με τον έλεγχο (5.1) και μπορεί να πραγματοποιηθεί είτε μέσω του στατιστικού του Pearson

$$X^2(I|RC) = X^2(I) - X^2(RC) = n \sum_{i=1}^I \sum_{j=1}^J \frac{(\hat{f}_{ij} - \hat{g}_{ij})^2}{\hat{g}_{ij}}$$

είτε μέσω του λόγου πιθανοφάνειας

$$G^2(I|RC) = G^2(I) - G^2(RC) = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \left(\frac{\hat{f}_{ij}}{\hat{g}_{ij}} \right)$$

όπου \hat{f}_{ij} και \hat{g}_{ij} είναι οι εκτιμητές μέγιστης πιθανοφάνειας των πιθανοτήτων p_{ij} κάτω από το μοντέλο RC και το μοντέλο της ανεξαρτησίας αντίστοιχα.

Ο Haberman (1981) απέδειξε ότι η ασυμπτωτική κατανομή των ελεγχουσυναρτήσεων $X^2(I|RC)$ και $G^2(I|RC)$ δεν είναι εν γένει χι-τετράγωνο κατανομή. Αυτό συμβαίνει μόνο αν $J = 2$. Πιο συγκεκριμένα απέδειξε το ακόλουθο θεώρημα για την ασυμπτωτική κατανομή των στατιστικών $X^2(I|RC)$ και κάτω $G^2(I|RC)$ από το μοντέλο RC.

Θεώρημα 5.1:

Ας είναι $F(I-1, J-1)$ η μέγιστη ιδιοτιμή ενός $(I-1) \times (I-1)$ Wishart πίνακα με $(J-1)$ βαθμούς ελευθερίας και $I(I-1, J-1, a)$ το ποσοστιαίο σημείο της κατανομής του $F(I-1, J-1)$. Τότε

$$\text{Prob}\{X^2(I|RC) > I(I-1, J-1, a)\} - \text{Prob}\{F(I-1, J-1) > I(I-1, J-1, a)\} \rightarrow 0$$

και

5. Δεσμευμένος Έλεγχος Ανεξαρτησίας (όταν ισχύει το Μοντέλο RC)

$$\text{Prob}\{G^2(I|RC) > I(I-1, J-1, a)\} - \text{Prob}\{F(I-1, J-1) > I(I-1, J-1, a)\} \rightarrow 0. \quad \square$$

Σύμφωνα με το θεώρημα, το $X^2(I|RC)$ και το $G^2(I|RC)$ έχουν την ίδια ασυμπτωτική κατανομή με αυτήν της μέγιστης ιδιοτιμής του $(I-1) \times (I-1)$ Wishart πίνακα με $(J-1)$ βαθμούς ελευθερίας. Ωστόσο η κατανομή της μέγιστης ιδιοτιμής του Wishart πίνακα είναι αρκετά περιπλοκή. Για να υπολογιστούν τα ποσοστιαία σημεία μπορούν να χρησιμοποιηθούν οι προσεγγιστικές μέθοδοι των Hanumara and Thompson (1968), αν και όπως είπαμε στην παράγραφο 5.1 δεν καλύπτουν όλες τις περιπτώσεις.

5.4 Bootstrap Έλεγχος Ανεξαρτησίας

Οι Jeong et al. (2005) προτείνουν να ελεγχθεί η υπόθεση της προσαρμογής του μοντέλου της ανεξαρτησίας δεδομένου ότι ισχύει το μοντέλο RC μέσω ενός bootstrap ελέγχου. Η υλοποίηση του αλγορίθμου γίνεται μέσω των ακόλουθων βημάτων:

Αλγόριθμος 5.1

Αλγόριθμος για τον bootstrap έλεγχο της ανεξαρτησίας όταν το μοντέλο RC ισχύει.

- 1ο Βήμα:** Δοθέντων των συχνοτήτων του $I \times J$ πίνακα $\{n_{ij} : 1 \leq i \leq I, 1 \leq j \leq J\}$, προσαρμόζουμε το μοντέλο RC και το μοντέλο της ανεξαρτησίας και υπολογίζουμε τα αντίστοιχα αναμενόμενα τυχαία διάνυσμα των πιθανοτήτων $\{\hat{f}_{ij} : 1 \leq i \leq I, 1 \leq j \leq J\}$ και $\{\hat{g}_{ij} = p_i \cdot p_{\cdot j} : 1 \leq i \leq I, 1 \leq j \leq J\}$. Στην συνέχεια, υπολογίζουμε το στατιστικού του λόγου πιθανοφάνειας $G^2(I|RC)$.
- 2ο Βήμα:** Από το τυχαίο διάνυσμα $\{\hat{g}_{ij} = p_i \cdot p_{\cdot j} : 1 \leq i \leq I, 1 \leq j \leq J\}$, παίρνουμε ένα bootstrap σύνολο δεδομένων.
- 3ο Βήμα:** Για το bootstrap σύνολο δεδομένων υπολογίζουμε το $G^{2*}(I|RC)$.
- 4ο Βήμα:** Επαναλαμβάνουμε B φορές το 2ο και 3ο βήμα.
- 5ο Βήμα:** Υπολογίζουμε το bootstrap p-value με βάση τον τύπο
- $$P_{\text{BOOT}} = \frac{\#\{G^{2*}(I|RC) \geq G^2(I|RC)\}}{B}$$

Στην παράγραφο 5.1 είδαμε ότι ο έλεγχος υποθέσεων μέσω της bootstrap μεθόδου μπορεί να γίνει είτε μέσω της εμπειρικής κατανομής είτε μέσω των περιορισμών της μηδενικής υπόθεσης. Οι Jeong, Jhun και Kim, χρησιμοποιούν την δεύτερη προσέγγιση. Τα bootstrap

σύνολα δεδομένων παράγονται κάτω από το μοντέλο της ανεξαρτησίας, δηλαδή του μοντέλου που ελέγχουμε με την μηδενική υπόθεση.

Ο έλεγχος που διενεργούμε είναι δεσμευμένος. Υποθέτουμε ότι το μοντέλο RC προσαρμόζεται καλά στα δεδομένα μας και στην συνέχεια εξετάζουμε την καλή προσαρμογή του μοντέλου της ανεξαρτησίας. Όπως αναφέραμε σε προηγούμενη παράγραφο, το μοντέλο της ανεξαρτησίας είναι εμφωλεμένο του μοντέλου RC. Επομένως όταν το μοντέλο της ανεξαρτησίας ισχύει τότε το μοντέλο RC θα έχει καλή προσαρμογή στα δεδομένα μας. Αυτό το γεγονός χρησιμοποιούν οι Jeong, Jhun και Kim και προτείνουν την παραγωγή των bootstrap συνόλων δεδομένων μέσω του μοντέλου της ανεξαρτησίας.

Ένα κρίσιμο ερώτημα είναι κατά πόσο ο αλγόριθμος μας εγγυάται την καλή προσαρμογή του μοντέλου RC για κάθε bootstrap σύνολο δεδομένων. Είναι προφανές ότι η καλή προσαρμογή του μοντέλου RC για κάθε bootstrap σύνολο δεδομένων δεν μπορεί να εξασφαλιστεί καθώς πάντα υπάρχει η πιθανότητα να παραχθεί ένα bootstrap σύνολο δεδομένων τέτοιο ώστε να μην ισχύει το μοντέλο RC. Το ζήτημα που τίθεται σε αυτήν την περίπτωση είναι κατά πόσον είναι πρέπον να εξετάζουμε έναν δεσμευμένο έλεγχο χωρίς να ισχύει η δέσμευση του ελέγχου;

Μια προσέγγιση είναι να ανεχθούμε να μην ισχύει η δέσμευση μας σε ποσοστό περίπου ίσο με το σφάλμα τύπου I. Όταν το ποσοστό αυτό είναι πολύ μικρότερο από το επίπεδο σημαντικότητας τότε τίθενται πολλές απορίες για την καταλληλότητα του bootstrap ελέγχου.

Μια δεύτερη προσέγγιση είναι να αλλάξουμε το δεύτερο βήμα του αλγορίθμου με το να παράγουμε τα bootstrap σύνολα δεδομένων μέσω των $\{\hat{f}_{ij} : 1 \leq i \leq I, 1 \leq j \leq J\}$ δηλαδή μέσω των πιθανοτήτων του μοντέλου RC. Το μειονέκτημα της προσέγγισης αυτής είναι ότι δεν παράγουμε τα bootstrap σύνολα δεδομένων κάτω από την μηδενική υπόθεση το οποίο πιθανώς να έχει επίπτωση στην ισχύ του ελέγχου.

Μια τρίτη προσέγγιση είναι να παράγουμε τα bootstrap σύνολα δεδομένων μέσω της εμπειρικής κατανομής. Είναι ο πιο απλοϊκός bootstrap έλεγχος που μπορούμε να εφαρμόσουμε. Ωστόσο το αποτέλεσμα θα εξαρτάται από το κατά πόσο τα μοντέλα της ανεξαρτησίας και το μοντέλο RC προσαρμόζονται καλά στα δεδομένα μας.

5.5 Αριθμητικό Παράδειγμα

Για να εξετάσουμε τις τρεις διαφορετικές προσεγγίσεις που περιγράψαμε στην προηγούμενη παράγραφο, δημιουργήσαμε έναν 3x3 πίνακα συνάφειας όπου το μοντέλο της ανεξαρτησίας

5. Δεσμευμένος Έλεγχος Ανεξαρτησίας (όταν ισχύει το Μοντέλο RC)

θα είναι αποδεκτό σε επίπεδο σημαντικότητας 5% και μη αποδεκτό σε επίπεδο σημαντικότητας 10%. Τα δεδομένα φαίνονται στον Πίνακα 5.1.

Πίνακας 5.1

X \ Y	Y			Σύνολο
	1	2	3	
1	19	17	16	52
2	12	10	15	37
3	11	30	21	62
Σύνολο	42	57	52	151

Βρίσκουμε ότι $G^2(I) = 8.082$ με 4 βαθμούς ελευθερίας και $p\text{-value} = 0.089$. Επομένως το μοντέλο της ανεξαρτησίας είναι αποδεκτό σε επίπεδο σημαντικότητας 5% και μη αποδεκτό σε επίπεδο σημαντικότητας 10%. Το μοντέλο RC έχει καλή προσαρμογή στα δεδομένα μας αφού $G^2(RC) = 0.922$ με 1 β.ε. και $p\text{-value} = 0.337$. $G^2(RC) = 0.922$. Επίσης $G^2(I|RC) = 7.160$. Στον Πίνακα 5.2 παρουσιάζονται τα αποτελέσματα για τους bootstrap ελέγχους που περιγράψαμε στην προηγούμενη παράγραφο. Έγιναν 1000 bootstrap επαναλήψεις και καταγράψαμε τον αριθμό των επαναλήψεων όπου το μοντέλο της ανεξαρτησίας και το μοντέλο RC είχαν καλή προσαρμογή στα δεδομένα.

Πίνακας 5.2

Ονομαστικό Επίπεδο	P_{BOOT}	Μοντέλο Επαναδειγματοληψίας	Ποσοστό που ισχύει το μοντέλο της Ανεξαρτησίας	Ποσοστό που ισχύει το μοντέλο RC
5%	0.107	I	97%	99%
10%			90%	98%
5%	0.674	RC	41%	98%
10%			31%	91%
5%	0.694	F_n	39%	87%
10%			10%	78%

Παρατηρούμε ότι όταν εκτελούμε τον bootstrap έλεγχο με μοντέλο επαναδειγματοληψίας το μοντέλο της ανεξαρτησίας τότε το μοντέλο RC έχει καλή προσαρμογή σχεδόν σε όλα τα bootstrap σύνολα δεδομένων, ενώ με μοντέλο επαναδειγματοληψίας το μοντέλο RC και την εμπειρική κατανομή F_n , η αποδοτικότητα του μοντέλου RC χειροτερεύει σταδιακά. Όταν η επαναδειγματοληψία πραγματοποιείται με βάση την εμπειρική κατανομή, το μοντέλο RC έχει

πολύ άσχημη προσαρμογή, γεγονός που μας οδηγεί στην απόρριψη της εμπειρικής κατανομής ως δειγματοληπτικό μοντέλο.

Στην πρώτη περίπτωση το μοντέλο της ανεξαρτησίας ισχύει σε μεγάλο ποσοστό και η υπόθεση της ανεξαρτησίας δεν μπορεί να απορριφθεί. Στην δεύτερη περίπτωση η υπόθεση της ανεξαρτησίας γίνεται αποδεκτή αν και το ποσοστό που ισχύει το μοντέλο της ανεξαρτησίας στα bootstrap δείγματα είναι αρκετά μικρό. Αυτό μπορεί να εξηγηθεί διαισθητικά ως εξής: όταν το μοντέλο της ανεξαρτησίας δεν ισχύει τότε το στατιστικό $G^2(I|RC)$ έχει μεγάλη τιμή, αφού το $G^2(I)$ είναι μεγάλο και το $G^2(RC)$ είναι μικρό, καθώς ισχύει το μοντέλο RC. Όταν στα bootstrap δείγματα το μοντέλο της ανεξαρτησίας γίνεται αποδεκτό τότε τόσο το $G^{2*}(I)$ όσο και το $G^{2*}(RC)$ είναι μικρά. Αυτό σημαίνει ότι και το $G^{2*}(I|RC)$ είναι μικρό και μάλιστα μικρότερο από το $G^2(I|RC)$. Το αντίστροφο συμβαίνει όταν τα bootstrap σύνολα δεδομένων λαμβάνονται από το μοντέλο RC.

Με βάση τον παραπάνω ισχυρισμό δημιουργείται η υποψία ότι ο bootstrap έλεγχος που προτείνουν οι Jeong et al. (2005) εξαρτάται από τις τιμές των $G^2(I)$ και $G^2(RC)$. Επομένως, αξίζει να μελετηθεί ο bootstrap έλεγχος όταν το μοντέλο της ανεξαρτησίας απορρίπτεται ή γίνεται αποδεκτό οριακά.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Καρλής, Δ., (2004). *Υπολογιστικά Μαθηματικά*, (Σημειώσεις) Οικονομικό Πανεπιστήμιο Αθηνών, Αθήνα.
- Κόλυβα-Μαχαίρα, Φ., και Μπόρα-Σέντα, Ε., (1998). *Στατιστική*, Εκδόσεις Ζήτη, Θεσσαλονίκη.
- Κουνιάς, Ε., (2002). *Εφαρμοσμένη Ανάλυση Παλινδρόμησης*, Πανεπιστήμιο Αθηνών, Αθήνα.

Ξένα

- Agresti, A., (1990). *Categorical Data Analysis*, Wiley Interscience, New York.
- Agresti, A., (2002). *Categorical Data Analysis*, Wiley Interscience, New York.
- Agresti, A., and Caffo, B.A., (2000). Simple and Effective Confidence Intervals for Proportions and Differences of Proportions from Adding two Success and two Failures, *The American Statistician*, **54**, 280-288.
- Agresti, A., and Coull, B.A., (1998). Approximation is better than “exact” for interval estimation of binomial proportions, *The American Statistician*, **52**, 119-126.
- Athreya, K.B., (1987). Bootstrap estimation of the mean in the infinite variance case, *The Annals of Statistics*, **15**, 724-731.
- Åstrand M., (2000). *On Bootstrapping Survival Data*, Ph.D. Thesis, Chalmers University Of Technology and Gotenburg University, Sweden.
- Babu, G.J., and Singh, K., (1983). Inference on means using the bootstrap, *The Annals of Statistics*, **11**, 999-1003.
- Beran, R. (1988a). Prepivoting test statistics: A bootstrap view of asymptotic refinements, *Journal of the American Statistical Association*, **83**, 687-697.
- Beran, R., (1988b). Balanced simultaneous confidence sets, *Journal of the American Statistical Association*, **83**, 679-686.
- Becker, M., (1990). Maximum likelihood estimation of the RC M association model. *Applied Statistics*, **39**: 152-167.
- Bickel, P.J., and Freedman, D.A., (1981). Some asymptotic theory for the bootstrap, *The Annals of Statistics*, **9**, 1196-1217.
- Bunke, O., and Riemer, S., (1983). A note on bootstrap and other empirical procedures for testing linear hypotheses without normality, *Statistics*, **14**, 517-526.
- Brown, L.D., Cai, T.T., and DasGupta, A., (2001). Interval estimation for a binomial proportion, *Statistical Science*, **16**, 101-133.
- Brown, L.D., Cai, T.T., and DasGupta, A., (2002). Confidence Intervals for a binomial proportion and asymptotic expansions, *The Annals of Statistics*, **30**, 160-201.

- Chernick, M.R., (1999). *Bootstrap methods. A practitioner's guide*, John Wiley & Sons, New York.
- Clarke, P.S., and Smith P.W., (2004). Interval Estimation for Log-linear models with one variable subject to non-ignorable non-response, *Journal of the Royal Statistical Society J*, **66**, 357-368.
- Cressie, R.C., and Read, N.A.C., (1984). *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer Verlag, New York.
- Conlon, M., and Thomas R.G., (1990). A New Confidence Interval For the difference of two Binomial Proportions. *Computational Statistics & Data Analysis*, **9**, 237-241.
- Davis, C., (1988). *Estimation of row and column scores in the linear by linear association model for two-way ordinal contingency tables*, In: Proceedings of the 13th Annual SAS Users Group International Conference, SAS Institute Inc, 946-951.
- Davison, A.C., and Hinkley, D.V., (1997). *Bootstrap Methods and Their Application*, U.K. Cambridge University Press, Cambridge.
- De Angelis, D., and Young, G. A., (1992). Smoothing the bootstrap, *International Statistical Review*, **60**, 45-56.
- Diaconis, P., and Holmes, S., (1994). Gray Codes for randomization procedures, *Statistics and Computing*, **4**, 287-302.
- DiCiccio, T.J., and Romano, J.P., (1988). A review of bootstrap confidence intervals (with Discussion), *Journal of the Royal Statistical Society, Series B*, **50**, 338-354.
- Ducharme, G.R., and Jhun, M., (1986). A note on the bootstrap procedure in testing linear hypotheses, *Statistics*, **17**, 527-531.
- Efron, B., (1979). Bootstrap methods: Another look at the jackknife, *The Annals of Statistics*, **7**, 1-26.
- Efron, B., (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*, SI AM, Philadelphia.
- Efron, B., (1987). Better bootstrap confidence intervals (with Discussion), *Journal of the American Statistical Association*, **82**, 171-200.
- Efron, B., and Tibshirani, R.J., (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, **1**, 54-77.
- Efron, B., and Tibshirani, R.J., (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Falk, M., and Kaufman, E., (1991). Coverage probabilities of bootstrap confidence intervals for quantiles, *The Annals of Statistics*, **19**, 485-495.
- Fisher, N.I., and Hall, P., (1991). On bootstrap hypothesis testing, *Australian Journal of Statistical Planning & Inference*, **32**, 177-190.
- Fisher, N.I., and Hall, P., (1991). Bootstrap algorithms for small samples, *Journal of Statistical Planning & Inference*, **27**, 157-169.
- Frangos, C.C., and Schucany, W.R., (1990). Jackknife estimation of the bootstrap acceleration constant, *Computational Statistics & Data Analysis*, **9**, 271-282.

- Ghosh, B.K., (1979). A comparison of some approximate confidence intervals for the binomial parameter, *Journal of the American Statistical Association*, **74**, 849-900.
- Goodman, L.A., (1964). Simultaneous confidence intervals for contrasts among multinomial populations, *Annals of Mathematical Statistics*, **35**, 716-725.
- Goodman, L.A., (1979). Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association*, **74**, 537-552.
- Guerra, R., Polansky, A.M., and Schucany, W.R., (1997). Smoothed bootstrap Confidence Intervals with discrete data, *Computational Statistics & Data Analysis*, **26**, 163-176
- Haberman, S.J., (1981). Tests for independence in two way contingency tables based on canonical correlations and on linear interaction, *The Annals of Statistics*, **9**, 1178-1186.
- Hall, P., (1986). On the bootstrap and confidence intervals (with Discussion), *The Annals of Statistics*, **14**, 1431-1452.
- Hall, P., (1988). Theoretical Comparison of bootstrap confidence intervals (with Discussion), *The Annals of Statistics*, **16**, 927-985.
- Hall, P., (1992a). *Efficient bootstrap simulations, Exploring the Limits of Bootstrap*, R. LePage and L. Billard eds., 127-143, Wiley, New York.
- Hall, P., (1992b). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.
- Hall, P., and Martin, M.A., (1989). A note on the accuracy of bootstrap percentile method confidence intervals for a quantile, *Statist. Prob. Letters*, **8**, 197-200.
- Hall, P., and Wilson, S.R., (1991). Two guide lines for bootstrap hypothesis testing, *Biometrics*, **47**, 757-762.
- Hanumara, R.C., and Thompson, W.A., (1968). Percentage points of the extreme roots of a Wishart matrix, *Biometrika*, **55**, 505-512.
- Hartigan, J. A., (1969). Using subsample values as typical value, *Journal of the American Statistical Association*, **64**, 1303-1317.
- Huang, J. S., (1991). Efficiency computation of the performance of bootstrap and jackknife estimators of the variance of L-statistics, *J. Statist. Compu. Simul.*, **38**, 45-56.
- Jaekel, L., (1972). *The infinitesimal jackknife, Memorandum*, MM 72-1215-11, Bell Lab., Murray Hill, NJ.
- Jeong, H.C, Jhun, M., and Kim, D., (2005). Bootstrap tests for independence in two-way ordinal contingency tables, *Computational Statistics & Data Analysis*, **48**, 623-631.
- Jhun, M., and Jeong, H.C., (2000). Applications of bootstrap methods for categorical data analysis, *Computational Statistics & Data Analysis*, **35**, 83-91.
- Konishi, S., (1991). Normalizing transformations and bootstrap confidence intervals, *The Annals of Statistics*, **19**, 2209-2225.
- Laplace, P.S.,(1812), *Théorie Analytique des Probabilités*, Paris, Courcier.
- Liu, R.Y., and Singh, K., (1987). On a partial correction by the bootstrap, *The Annals of Statistics*, **15**, 1713-1718.
- Loh, W.Y., and Wu, C.F.J., (1987). Discussion of "Better bootstrap confidence intervals"

- by B. Efron, *Journal of the American Statistical Association*, **82**, 188-190.
- Louis, T.A., (1981). Confidence Interval for a binomial parameter after observing no successes, *The American Statistician*, **35**, 181.
- McNemar, Q., (1947). Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika*, **12**, 153-157.
- Mammen, E. (1992). *When does Bootstrap Work? Asymptotic Results and Simulation*, Springer-Verlag, Heidelberg.
- Morales, D., Pardo, L., and Santamaría, L., (2004). Bootstrap confidence regions in multinomial sampling, *Applied Mathematics and Computation*, **155**, 295-315.
- Newcombe R.G., (1998a). Two sided confidence intervals for single proportions; comparison of seven methods, *Statistics in Medicine*, **17**, 857-872.
- Newcombe R.G., (1998b). Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods, *Statistics in Medicine*, **52**, 873-890.
- Neyman, J., (1935). On the Problem of Confidence Limits, *Annals of Mathematical Statistics*, **6**, 111-116.
- Noreen, E.W., (1989). *Computer Intensive Methods for Testing Hypotheses: An Introduction*, Wiley, New York.
- Pan, W., (2002). Approximate confidence intervals for one proportions and difference of two proportions, *Computational Statistics & Data Analysis*, **40**, 143-257.
- Politis, D.N., and Romano, J.P., (1993). Estimating the distribution of a studentized statistic by subsampling, *Bull. Intern. Statist. Inst.*, 49th Session, 2, 315-316.
- Quenneville, B., (1986). Bootstrap procedures for testing linear hypothesis without normality, *Statistics*, **17**, 533-538.
- Quenouille, M., (1949). Approximation tests of correlation in time series, *Journal of the Royal Statistical Society*, **11**, 18-84.
- Rao, C.R., (1989). *Statistics and Truth. Putting Chance to Work*, International Co-operative Publishing House, Burtonsville, Md.
- Sauermann, W., (1989). Bootstrapping The Maximum Likelihood Estimator in High-dimensional Log-Linear Models, *The Annals of Statistics*, **17**, 1198-1216.
- Schenker, N., (1985). Qualms about bootstrap confidence intervals, *Journal of the American Statistical Association*, **80**, 360-361.
- Serfling, R.J., (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- Shao, J. and Tu, D. (1995). *The Jackknife and the Bootstrap*, Springer, New York.
- Shi, X., Chen, J., and Wu, C.F.J., (1990). Weak and strong representations for quantile processes from finite populations with applications to simulation size in resampling methods, *Canadian Journal of Statistics*, **18**, 141-148.
- Simonoff, J.S., (2003). *Analyzing Categorical Data*, Springer, New York.
- Singh, K., (1981). On the asymptotic accuracy of Efron's bootstrap, *The Annals of*

- Statistics*, **9**, 1187-1195.
- Stein, C., (1956). Efficient nonparametric testing and estimation, *Proceedings of the Third Berkeley Symposium*, 187-196, Univ. of California Press, Berkeley.
- Sutton, C.D., (1993). Computer-intensive methods for tests about the mean of an asymmetric distribution, *Journal of the American Statistical Association*, **88**, 802-810.
- Tu, D. and Zhang, L. (1992a). On the estimation of skewness of a statistic using the jackknife and the bootstrap, *Statistical Papers*, **33**, 39-56.
- Tu, D., and Zhang, L., (1992b). Jackknife approximations for some nonparametric confidence intervals of functional parameters based on normalizing transformations, *Computational Statistics*, **7**, 3-15.
- Tukey, J., (1958). Bias and confidence in not quite large samples, *Annals of the Mathematical Statistics*, **29**, 614.
- Vollset, S.E., (1993), Confidence Intervals for a binomial Proportion, *Statistics in Medicine*, **12**, 809-824.
- Wang, W., (2005). Smallest confidence intervals for one binomial proportion, *Journal of Statistical Planning & Inference*, to appear.
- Wilson, E.B., (1927). Probable Inference, the Law of Succession, and Statistical Inference, *Journal of the American Statistical Association*, **22**, 209-212.