Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Εφαρμοσμενων Μαθηματικων και Φυσικων Επιστημων

Διατμηματικο Προγραμμα Μεταπτυχιακων Σπουδων (ΔΠΜΣ)

Μαθηματικη Προτυποποιηση στις Συγχρονες Τεχνολογιες και την Οικονομια

Μέθοδοι bootstrap για τον υπολογισμό διαστημάτων εμπιστοσύνης και ποσοστιαίων σημείων: Συγκριτική μελέτη

Χρυσάνθη Α. Παπαμιχαηλ

Επιτροπή:

Salim Bouzebda	-	Université de Technologie de Compiègne - (Επιβλέπων)
Nikolaos Limnios	-	Université de Technologie de Compiègne - (Επιβλέπων)
Χρυσηΐς Καρωνη	-	Εθνικό Μετσόβιο Πολυτεχνείο

Φεβρουάριος, 2013

National Technical University of Athens
School of Applied Mathematical and Physical Sciences
Interdepartmental Program of Postgraduate Studies (IPPS)
Mathematical Modelling in Modern Technologies and Financial Engineering

Bootstrap methods for confidence intervals and quantiles calculation: A comparison study

Chrysanthi A. Papamichail

Commitee :

Salim Bouzebda	-	Université de Technologie de Compiègne - (Supervisor)
Nikolaos Limnios	-	Université de Technologie de Compiègne - (Supervisor)
Chrysseis Caroni	-	National Technical University of Athens

February, 2013

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors of this project, Prof. N. Limnios and Assoc. Professor S. Bouzebda, who were abundantly helpful and offered invaluable assistance, support and guidance. Besides, I would like to thank Prof. C. Caroni, without whose assistance, this study would not have started. Finally, an honorable mention goes to my coleagues, I. Votsi and S. Georgiadis, for the valuable advice and support.

ΣΥΝΟΨΗ

Σκοπός της παρούσας εργασίας είναι η σύγκριση των ποσοστιαίων σημείων και των διαστημάτων εμπιστοσύνης αφότου εκτιμηθούν με την κλασσική μέθοδο και με μεθόδους bootstrap για την περίπτωση ανεξάρτητων και ισόνομα κατανεμημένων τυχαίων μεταβλητών και για την περίπτωση της αλυσίδας Markov. Μετά από μία σύντομη ιστορική αναφορά στις μεθόδους bootstrap , παρουσιάζεται η βασική θεωρία της μεθόδου bootstrap του Efron και της σταθμισμένης μεθόδου bootstrap. Στην παρούσα μελέτη, εφαρμόστηκαν η μέθοδος bootstrap του Efron και οι σταθμισμένες μέθοδοι με εκθετική κατανομή και με κατανομή Poisson.Περιλαμβάνεται επίσης η βασική θεωρία για την εμπειρική συνάρτηση κατανομής, η οποία χρησιμοποιείται για την εκτίμηση των ποσοστιαίων σημείων. Για τις αλυσίδες Markov , οι οποίες αποτελούν το δεύτερο τύπο δείγματος που χρησιμοποιείται, γίνεται επίσης θεωρητική ανάλυση.

Όσον αφορά στην εφαρμογή, τόσο στην περίπτωση των ανεξάρτητων και ισόνομα κατανεμημένων τυχαίων μεταβλητών όσο και στην περίπτωση της μαρκοβιανής αλυσίδας, εκτιμάται η τυπική απόκλιση με την κλασσική μέθοδο και διάφορες μεθόδους bootstrap, και στη συνέχεια συγκρίνεται με τη θεωρητική τιμή της τυπικής απόκλισης. Επίσης, χρησιμοποιείται η εμπειρική συνάρτηση κατανομής για την εκτίμηση των ποσοστιαίων σημείων, και τέλος, έχοντας τα προηγούμενα αποτελέσματα της τυπικής απόκλισης και των ποσοστιαίων σημείων, λαμβάνουμε μια εκτίμηση των διαστημάτων εμπιστοσύνης.

Λέξεις-κλειδιά: Μέθοδοι bootstrap, σταθμισμένη μέθοδος bootstrap, μέθοδος bootstrap του Efron, εμπειρική συνάρτηση κατανομής, εμπειρική διαδικασία, ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές, μαρκοβιανή αλυσίδα, τυπική απόκλιση, ποσοστιαίο σημείο, διάστημα εμπιστοσύνης

ABSTRACT

The purpose of this project is to compare the quantiles and confidence intervals estimated with classical and bootstrap methods for the case of independent and identically distributed [i.i.d.] variables and the case of a Markov chain. After a short historical overview of bootstrap methods, we expose the basic theory of Efron bootstrap and weighted bootstrap. In the present study, Efron bootstrap and weighted bootstrap with exponential and Poisson distribution were applied. We also include the basic theory for the empirical distribution function, which is used for the estimation of quantiles. Markov chains, which constitute the second type of sample used, are theoretically analyzed, as well.

In regard with the application, for both the case of i.i.d. variables and the case of Markov chain, we estimate the variance with classical method and various bootstrap methods, and then compare it to the theoretical variance. Then we use the empirical distribution function so as to estimate the quantiles, and last, having the previous results of variance and quantiles, we get an estimation of confidence intervals.

Keywords: Bootstrap methods, weighted bootstrap, Efron bootstrap, empirical distribution function, empirical process, independent and identically distributed random variables, Markov chain, variance, quantile, confidence interval

ΕΚΤΕΝΗΣ ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας μεταπτυχιακής εργασίας είναι να συγκρίνουμε τα ποσοστιαία σημεία και τα διαστήματα εμπιστοσύνης αφότου εκτιμηθούν με την κλασσική μέθοδο και με μεθόδους bootstrap όταν έχουμε δείγμα ανεξάρτητων και ισόνομα κατανεμημένων μεταβλητών και όταν έχουμε μαρκοβιανή αλυσίδα.

Αρχικά είναι απαραίτητη μια σύντομη ιστορική αναδρομή, σχετικά με τις μεθόδους bootstrap. Πρόκειται για μια κατηγορία μίας ευρύτερης κλάσης μεθόδων που επαναδειγματοληπτούν από το αρχικό σύνολο δεδομένων και ονομάζονται άρα διαδικασίες επαναδειγματοληψίας. Κάποιες διαδικασίες επαναδειγματοληψίας παρόμοιες με τις μεθόδους bootstrap χρονολογούνται από το πρώτο μισό του 20ού αι. (π.χ. jackknife). Ωστόσο, ήταν ο Efron (1979a) ο οποίος συνδύασε ιδέες και συνέδεσε την απλή μη παραμετρική μέθοδο βοοτστραπ, για ανεξάρτητες και ταυτοτικά κατανεμημένες παρατηρήσεις, η οποία επαναδειγματοληπτεί από τα δεδομένα με αντικατάσταση, με προγενέστερα στατιστικά εργαλεία για την εκτίμηση του τυπικού σφάλματος όπως η μέθοδος jackknife ή η μέθοδος delta. Η ιδέα της δειγματοληψίας με αντικατάσταση από το αρχικό δείγμα εντούτοις προηγείται του Efron.

Οι επαγγελματίες στατιστικοί Bruce και Simon,τη δεκαετία του '90, διέδωσαν μέσω της εταιρίας τους τη χρήση των μεθόδων bootstrap. Επίσης συνεχίζουν να χρησιμοποιούν την προσέγγιση Monte Carlo για τους εκτιμητές που δίνουν οι μέθοδοι βοοτστραπ. Ακόμη, υπήρξαν άλλες διαδικασίες που ονομάζονταν "bootstrap" και διαφέρουν από την ιδέα του Efron, ο οποίος σε κάθε περίπτωση με τη δημοσίευσή του (1979a)έθεσε το θεωρητικό υπόβαθρο των μεθόδων bootstrap, οι οποίες αναπτύχθηκαν στη συνέχεια τόσο από τον ίδιο όσο και από άλλους ερευνητές.

Η γενική ιδέα των μεθόδων bootstrap εξηγείται στη συνέχεια: Αν ένας εκτιμητής της παραμέτρου του δείγματος συμβολίζεται ως $\hat{\theta}$, η κατανομή bootstrap για ($\hat{\theta}$ - θ) είναι η κατανομή που λαμβάνεται δημιουργώντας αρκετά $\hat{\theta}$ ανεξάρτητα, με δειγματοληψία με αντικατάσταση από την εμπειρική κατανομή F_n. Ο εκτιμητής bootstrap του τυπικού σφάλματος της $\hat{\theta}$ είναι τότε η τυπική απόκλιση της κατανομής bootstrap για ($\hat{\theta} - \theta$). Η διαδικασία είναι απλή και περιγράφεται ως εξής:

1. Δημιουργήστε ένα δείγμα με αντικατάσταση από την εμπειρική κατανομή (ένα δείγμα bootstrap),

2. Υπολόγισε την τιμή του $\hat{\theta}$ χρησιμοποιώντας το δείγμα bootstrap στη θέση του αρχικού δείγματος,

3. Επαναλάβετε τα βήματα 1 και 2 k φορές.

Συμπερασματικά, μας ενδιαφέρει η κατανομή της διαφοράς $\hat{\theta} - \theta$. Αυτό που διαθέτουμε είναι μια Monte Carlo προσέγγιση για την κατανομή της διαφοράς $\theta^* - \hat{\theta}$.

Η ιδέα-κλειδί πίσω από το bootstrap είναι ότι για ν αρκετά μεγάλο, περιμένουμε οι δύο κατανομές να είναι σχεδόν ίδιες. Σε μερικές περιπτώσεις, είμαστε σε θέση να υπολογίσουμε τον εκτιμητή bootstrap άμεσα, χωρίς την προσέγγιση Monte Carlo. Η βασική ιδέα άρα πίσω από το bootstrap είναι ότι η μεταβλητότητα του θ^* (βάσει F_n) γύρω από $\hat{\theta}$ θα είναι παρόμοια με τη μεταβλητότητα των $\hat{\theta}$ (βάσει της πραγματικής κατανομής του πληθυσμού F_n) γύρω από την πραγματική τιμή της παραμέτρου, θ. Υπάρχει λόγος να πιστεύουμε ότι αυτό θα ισχύει και για μεγάλα μεγέθη δείγματος, αφού, καθώς το n θα γίνεται όλο και μεγαλύτερο, η F_n έρχεται όλο και πιο κοντά στη F και έτσι δειγματοληψία με αντικατάσταση από F_n είναι σχεδόν σαν τυχαία δειγματοληψία από F.

Ο ισχυρός νόμος των μεγάλων αριθμών για ανεξάρτητες ταυτοτικά κατανεμημένες τυχαίες μεταβλητές συνεπάγεται ότι με πιθανότητα ένα, η F_n συγκλίνει σε F κατά σημείο. Ένα ισχυρότερο αποτέλεσμα, το θεώρημα Glivenko-Cantelli, ισχυρίζεται ότι η εμπειρική κατανομή συγκλίνει ομοιόμορφα με πιθανότητα 1 στην F, όταν οι παρατηρήσεις είναι ανεξάρτητες και ισόνομα κατανεμημένες. Αυτό το θεμελιώδες θεωρητικό αποτέλεσμα προσδίδει αξιοπιστία στην προσέγγιση bootstrap.

Όντας μία από τις πιο σημαντικές ιδέες του τον τελευταίο μισό αιώνα στις στατιστικές εφαρμογές, οι μέθοδοι bootstrap εισήγαγαν πλήθος καινοτόμων προβλημάτων στο χώρο των πιθανοτήτων, τα οποία με τη σειρά τους αποτέλεσαν τη βάση για τη δημιουργία νέων μαθηματικών θεωριών. Οι περισσότερες από αυτές τις θεωρίες έχουν εκπονηθεί για την κυρίαρχη περίπτωση, στη στατιστική πρακτική, όταν το δείγμα αποτελείται από ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές.
Σύμφωνα με τον Efron, υποθέτουμε ότι τα δεδομένα μας αποτελούνται από ένα τυχαίο δείγμα άγνωστης κατανομής πιθανοτήτων F στη γραμμή των πραγματικών αριθμών,

$$X_1, X_2, \ldots, X_n \sim F. \tag{1}$$

Το δείγμα bootstrap δεν έχει απαραίτητα το ίδιο μέγεθος με αυτό του αρχικού δείγματος. Η έννοια του δείγματος bootstrap είναι η ακόλουθη:
Έστω {μ(1), μ(2), ...} μια ακολουθία θετικών ακεραίων και για κάθε $n \in N$, οι τυχαίες μεταβλητές $\{X^*_{n,j}\ ,\ 1 \leqslant \vartheta \leqslant m(n)\}$ Προκύπτουν δειγματολειπτώντας $m(n)$ φορές με αντικατάσταση, από τις n παρατηρήσεις $X_1, \ldots, X_n$ έτσι ώστε για κάθε μία από τις $m(n)$ επιλεγμένες τιμές, κάθε X_k έχει πιθανότητα $1/n$ να έχει επιλεγεί.

Εναλλακτικά, για κάθε $n \in \mathbb{N}$ έχουμε $X_{n,j} = X_{Z(n,j)}$, $1 \leqslant j \leqslant m(n)$, όπου $Z(n,j)$, $1 \leqslant j \leqslant m(n)$ είναι ανεξάρτητες τυχαίες μεταβλητές ομοιόμορφα κατανεμημένες στο $\{1, \ldots, n\}$ και ανεξάρτητες από $\mathbf{X}$· Δεχόμαστε χωρίς βλάβη της γενικότητας ότι ο χώρος πιθανότητας $(\Omega,\ \mathcal{F}, P)$ 'φιλοξενεί' όλες αυτές τις τυχαίες μεταβλητές με από κοινού κατανομές. Τότε, οι $X^*_{n,1}, \ldots, X^*_{n,m(n)}$ είναι υπό προϋποθέσεις ανεξάρτητες και ισόνομα

κατανεμημένες δεδομένου $\mathbf{X_n} = (X_1, \ldots, X_n)$ με $P\{X^*_{n,1} = X_k | \mathbf{X_n}\} = n^{-1}$ σχεδόν βεβαίως , $1 \leqslant k \leqslant n$, $n \in N$. Για κάθε δείγμα μεγέθους $n \in N$, η ακολουθία αναφέρεται ως μη παραμετρικό bootstrap δείγμα του Efron, από $X_1, \ldots, X_n$ με μέγεθος δείγματος bootstrap ίσο με $m(n)$.
Αν γενικά $X_n = x_n$, τότε η μέση τιμή του δείγματος είναι:

$$\bar{x} = \sum_{1}^{n} x_n/n, \tag{2}$$

και η αμερόληπτη εκτιμήτρια της τυπικής απόκλισης $\sigma(F)$:

$$\bar{\sigma} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(n-1)}}, \tag{3}$$

Εφαρμόζοντας τον αλγόριθμο Monte Carlo, όπως περιγράφηκε αδρομερώς παραπάνω, παίρνουμε τον bootstrap -εκτιμητή της τυπικής απόκλισης:

$$\hat{\sigma}_B = (\frac{\sum_{b=1}^{B} \{\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)\}^2}{B-1})^{1/2}, \tag{4}$$

Όπου:

$$\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^{B} \hat{\theta}^*(b)}{B}. \tag{5}$$

η εκτίμηση της μέσης τιμής του δείγματος με τη μέθοδο bootstrap. Έχοντας υπολογίσει τους παραπάνω εκτιμητές, μπορούμε να προχωρήσουμε στον υπολογισμό των διαστημάτων εμπιστοσύνης. Η απλούστερη μέθοδος των bootstrap διαστημάτων εμπιστοσύνης είναι να πάρουμε

$$\theta \in [\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)]$$

ως ένα προσεγγιστικό $1-\alpha$ κεντρικό διάστημα για θ. Πρόκειται για τη μέθοδο εκατοστιαίου σημείου. Έστω $\hat{G}(s)$ η παραμετρική bootstrap συνάρτηση αθροιστικής κατανομής της $\hat{\theta}^*$,

$$\hat{G}(s) = Prob_*\{\hat{\theta}^* < s\}, \tag{6}$$

όπου $Prob_*$ δηλώνει την πιθανότητα που υπολογίστηκε σύμφωνα με την κατανομή bootstrap της $\hat{\theta}^*$. Το διάστημα της μεθόδου εκατοστιαίου σημείου είναι απλά το διάστημα μεταξύ των $(\alpha/2) \cdot 100\%$ και $(1-\alpha/2) \cdot 100\%$ εκατοστιαίων σημείων της κατανομής

bootstrap της $\hat{\theta}^*$. Το διάστημα εκατοστιαίων σημείων έχει άκρα

$$\theta_P[\alpha/2] \equiv \hat{G}^{-1}(\alpha/2). \tag{7}$$

Αυτό συγκρίνεται με το διάστημα,

$$\theta_S[\alpha/2] = \hat{\theta} + \frac{\hat{\sigma}}{\sqrt{n}} \cdot z^{(\alpha/2)}. \tag{8}$$

Αξίζει να σημειωθεί ότι η μέθοδος bootstrap του Efron παρουσιάζει κάποια εγγενή μειονεκτήματα. Για παράδειγμα, κάποιες παρατηρήσεις μπορεί να χρησιμοποιηθούν περισσότερες από μία φορές ενώ άλλες δεν υπόκεινται σε δειγματοληψία. Για την υπέρβαση αυτού του προβλήματος, προτάθηκε μια πιο γενική μορφή bootstrap, η σταθμισμένη bootstrap μέθοδος, η οποία απεδείχθη και πιο αποτελεσματική υπολογιστικά, σε αρκετές εφαρμογές. Για τον ορισμό της μεθόδου αυτής, θεωρούμε $\overline{X}_n$ τη μέση τιμή του δείγματος $X_1, \ldots, X_n$ και $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/n$, την τυπική απόκλιση του δείγματος. Θεωρούμε ένα διάνυσμα τυχαίων βαρών $W_n = (W_{n,1}, W_{n,2}, \ldots, W_{n,n})$, ανεξάρτητων από τα δεδομένα, $X_1, \ldots, X_n$. Υποθέτουμε ότι για κάθε ακέραιο $n \geqslant 1$, τα βάρη W_n είναι εναλλάξιμα. Βάσει του διανύσματος βαρών W_n, η γενικευμένη bootstrap μέση τιμή που αντιστοιχεί στο διάνυσμα αυτό θα είναι:

$$\overline{X}_{\mathcal{W},n} = \frac{1}{n} \sum_{i=1}^{n} W_{n,i} X_i. \tag{9}$$

Τα βάρη W_n θα ικανοποιούν τις εξής συνθήκες:

$$\begin{aligned}
&(\mathcal{W}_I) \qquad W_{n,i} \geqslant 0, i = 1, 2, \ldots, n, n \geqslant 1, \\
&(\mathcal{W}_{II}) \qquad \sum_{i=1}^{n} W_{n,i} = n, \\
&(\mathcal{W}_{III}) \qquad \frac{1}{n} \sum_{i=1}^{n} (W_{n,i} - 1)^2 \to_P c^2 \quad as \quad n \to \infty.
\end{aligned}$$

Στην εφαρμογή που πραγματοποιήθηκε στα πλαίσια της παρούσας εργασίας χρησιμοποιήθηκαν η μέθοδος bootstrap του Efron,η σταθμισμένη μέθοδος με εκθετική κατανομή και η σταθμισμένη μέθοδος με κατανομή Poisson.

Απαραίτητο είναι να αναφερθούμε και στην εμπειρική συνάρτηση κατανομής, η οποία χρησιμοποιήθηκε για την εκτίμηση των ποσοστιαίων σημείων. Ορίζεται λοιπόν ως εξής:

Αν $X_1, \ldots, X_n$ είναι ένα τυχαίο δείγμα από μια συνάρτηση κατανομής F στη γραμμή των πραγματικών αριθμών, η εμπειρική συνάρτηση κατανομής είναι

$$\mathbb{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i \leqslant t\}. \tag{10}$$

Πρόκειται για φυσικό εκτιμητή της υποκείμενης κατανομής F εφόσον είναι εντελώς άγνωστη.

Η συνάρτηση ποσοστιαίου σημείου που ανήκει στην F και προκύπτει από αντιστροφή της θα είναι

$$\begin{gathered} Q(s) = \inf\{x : F(x) \geqslant s\}, 0 < s < 1, \\ Q(0) = \lim_{t \downarrow 0} Q(t) = Q(0^+), \quad Q(1) = \lim_{t \uparrow 1} Q(t) = Q(1^-) \end{gathered} \tag{11}$$

και, αν σημειώσουμε με $X_{1,n} \leqslant \cdots \leqslant X_{n,n}$ την κατάταξη των $X_1, \ldots, X_n$, η ισοδύναμη 'εμπειρική' εξίσωση θα είναι:

$$Q_n(s) = \begin{cases} X_{k,n}, & \text{ιφ } (k-1)/n < s \leqslant k/n; \quad k = 1, \ldots, n, \\ X_{1,n}, & \text{ιφ } s = 0. \end{cases} \tag{12}$$

Από το νόμο των μεγάλων αριθμών, προκύπτει ότι

$$\mathbb{F}_n(t) \xrightarrow{\text{Σ.β.}} F(t), \quad \text{ερψ} \quad t. \tag{13}$$

Από το κεντρικό οριακό θεώρημα είναι ασυμπτωτικά κανονική,

$$\sqrt{n}(\mathbb{F}_n(t) - F(t)) \xrightarrow{P} N(0, F(t)(1 - F(t))). \tag{14}$$

Η εμπειρική κατανομή ορίζεται ως $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$, όπου δ_X είναι η Dirac κατανομή πιθανότητας που προέρχεται από X.

Αν εφαρμόσουμε bootstrap στο εμπειρικό μέτρο, θα έχει τη μορφή:

$$\hat{\mathbb{P}}_n \equiv \tfrac{1}{n} \sum_{j=1}^n M_{nj} \delta_{X_j^\omega}$$

όπου $\underline{M}_n \sim Mult_n(n, (n^{-1}, \ldots, n^{-1}))$.
Αν $\mathbf{W} = (W_{nj}, j = 1 \ldots, n, n = 1, 2, \ldots)$ ορίζουν τριγωνικό πίνακα μη αρνητικών τυχαίων μεταβλητών με $\sum_{j=1}^n W_{nj} = n\cdot$ τότε η σχέση:

$$\hat{\mathbb{P}}_n \equiv \frac{1}{n} \sum_{j=1}^n W_{nj} \delta_{X_j^\omega} \tag{15}$$

ορίζει ένα εμπειρικό μέτρο σταθμισμένου bootstrap.

Με $\hat{\mathbb{P}}_n$ όπως ορίζεται παραπάνω, $\hat{\mathbb{X}}_n \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \hat{\mathbb{P}}_n^\omega) \Rightarrow G_P$, ς.β. Η αντίστοιχη bootstrap εμπειρική διαδικασία είναι

$$\hat{\mathbb{X}}_n(\omega) = \hat{\mathbb{X}}_n^\omega \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n^\omega) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j^\omega}. \tag{16}$$

.

Συχνά η ποσότητα $\sqrt{n}(\theta_n - \theta)$ όπου θέλουμε να εφαρμόσουμε bootstrap μπορεί να εκφραστεί ως συνάρτηση της εμπειρικής διαδικασίας:

$$\mathbb{X}_n \equiv \sqrt{n}(\mathbb{P}_n - \mathbf{P}).$$

Στα πλαίσια της εφαρμογής σε μαρκοβιανή αλυσίδα χρειάζεται αρχικά να ορίσουμε τι εστί μία αλυσίδα Markov:

ΟΡΙΣΜΟΣ 1. *Η σειρά τυχαίων μεταβλητών $X = (X_n)_{n\in\mathbb{N}}$ ορισμένη σε ένα χώρο πιθανοτήτων $(\Omega, \mathcal{F}, \mathbb{P})$, με τιμές στο πεπερασμένο σύνολο $E = \{1, \ldots, s\}$, είναι μια μαρκοβιανή αλυσίδα αν, για κάθε μη αρνητικό ακέραιο n και κάθε κατάσταση $i, j, i_0, i_1, \ldots, i_{n-1} \in E$, έχουμε:*

$$\begin{aligned} \mathbb{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) \\ = \mathbb{P}(X_{n+1} = j | X_n = 1) = p_{ij;n}. \end{aligned} \tag{17}$$

Η ισότητα ονομάζεται ιδιότητα Markov.

Η ισχυρή ιδιότητα Markov είναι:

ΟΡΙΣΜΟΣ 2. *Η αλυσίδα Markov $(X_n)_{n\in\mathbb{N}}$ θεωρείται ότι έχει την ισχυρή ιδιότητα Markov αν, για κάθε χρόνο στάσης T, για κάθε ακέραιο $m \in \mathbb{N}$ και κατάσταση $j \in E$ έχουμε*

$$\mathbb{P}(X_{m+T} = j | X_k, k \leqslant T) = \mathbb{P}_{X_T}(X_m = j) \text{ Σ.Θ.}$$

Για τις ανάγκες της εφαρμογής αξιοποιήθηκαν θεωρήματα για τις μαρκοβιανές αλυσίδες όπως ο ισχυρός νόμος των μεγάλων αριθμών και το κεντρικό οριακό θεώρημα.

ΠΡΟΤΑΣΗ (Εργοδικό θεώρημα για τις μαρκοβιανές αλυσίδες) 3. *Για μία αμείωτη και απεριοδική μαρκοβιανή αλυσίδα έχουμε*

$$p_{ij}^n \xrightarrow[n\to\infty]{} \nu(j)$$

για κάθε $i, j \in E$.

ΠΡΟΤΑΣΗ 4. *Για μία αμείωτη και απεριοδική μαρκοβιανή αλυσίδα , υπάρχει ν η στάσιμη κατανομή της αλυσίδας τέτοια ώστε p^n να συγκλίνει με εκθετικό ρυθμό στο $\Pi = \mathbf{1}^\top \nu$, όπου $\mathbf{1} = (1, \ldots, 1)$.*

ΠΡΟΤΑΣΗ(Ισχυρός νόμος των μεγάλων αριθμών) 5. *Για μία εργοδική μαρκοβιανή αλυσίδα $(X_n)_{n\in\mathbb{N}}$, με στάσιμη κατανομή ν, έχουμε*

$$\tfrac{1}{n}\textstyle\sum_{k=0}^n g(X_k) \xrightarrow[n\to\infty]{a.s.} \sum_{i\in E} \nu(i) g(i) =: \bar{g}.$$

ΠΡΟΤΑΣΗ (κεντρικό οριακό θεώρημα) 6. *Για μία εργοδική μαρκοβιανή αλυσίδα $(X_n)_{n\in\mathbb{N}}$, με στάσιμη κατανομή ν, έχουμε*

$$\sqrt{n}(\tfrac{1}{n}\textstyle\sum_{k=0}^n g(X_k) - \bar{g}) \xrightarrow[n\to\infty]{D} \mathcal{N}(0, \sigma^2),$$

όπου

$$\sigma^2 := \tilde{g} diag(\nu)[2Z - I]\tilde{g}^\top,$$

όπου $Z = (I - p + \Pi)^{-1}$ είναι ο θεμελιώδης πίνακας του p και $\tilde{g}$είναι το διάνυσμα-γραμμή με στοιχεία $\tilde{g}(i) := g(i) - \bar{g}, \quad i \in E$.

Αφότου λοιπόν αναλύουμε το θεωρητικό υπόβαθρο της εργασίας, προχωρούμε στην υλοποίηση της εφαρμογής. Το πρώτο μέρος της εφαρμογής αφορά σε ανεξάρτητες ταυτοτικά κατανεμημένες τυχαίες μεταβλητές ενώ το δεύτερο σε μαρκοβιανή αλυσίδα. Στο πρώτο μέρος λοιπόν, εκτελούμε ένα πλήθος πειραμάτων, για διαφορετικά μεγέθη δείγματος, για διαφορετικό αριθμό επαναλήψεων της μεθόδου bootstrap, για τη σταθμισμένη μέθοδοbootstrap με βάρη που ακολουθούν εκθετική κατανομή ή κατανομή Poisson, αλλά και για δείγμα τυχαίων μεταβλητών που ακολουθούν κανονική, εκθετική ή ομοιόμορφη κατανομή. Σε κάθε πείραμα τη μέση τιμή και την τυπική απόκλιση του δείγματος με την κλασσική μέθοδο και με τη μέθοδο bootstrap και στη συνέχεια υπολογίζουμε τα σφάλματα, σε σχέση και με τις θεωρητικές τιμές, και τα διαστήματα εμπιστοσύνης.

Για την περίπτωση των ανεξάρτητων και ταυτοτικά κατανεμημένων τυχαίων μεταβλητών θα υπολογίσουμε τα εκατοστιαία σημεία ώστε να συγκρίνουμε με τη θεωρητικά αναμενόμενη τιμή τους. Εδώ χρησιμοποιείται η εμπειρική συνάρτηση κατανομής. Το πείραμα αυτό εκτελείται για διαφορετικά μεγέθη δείγματος και διαφορετικό αριθμό επαναλήψεων bootstrap, για τη μέθοδο Efron bootstrap και σταθμισμένου bootstrap με εκθετική ή Poisson κατανομή βαρών καθώς και για δύο περιπτώσεις τυπικής απόκλισης, μία περίπτωση χρησιμοποιώντας τη θεωρητική τιμή και δεύτερη περίπτωση χρησιμοποιώντας την εκτιμήτρια. Στο πειραματικό μέρος που αφορά στις μαρκοβιανές αλυσίδες υπολογίζουμε τη θεωρητική τυπική απόκλιση, την εκτιμήτριά της αλλά και την εκτιμήτρια με μέθοδο bootstrap, τις οποίες και συγκρίνουμε.

Τέλος, εκτιμούμε τα διαστήματα εμπιστοσύνης για τις δύο εκτιμήτριες της τυπικής απόκλισης και συγκρίνουμε. Τα πειράματα και εδώ εκτελούνται για διαφορετικούς συνδυασμούς παραμέτρων, μέγεθος μαρκοβιανής αλυσίδας, αριθμός επαναλήψεων bootstrap κλπ. (Η μεθοδολογία που ακολουθήθηκε περιγράφεται αναλυτικά στο αντίστοιχο κεφάλαιο.) Για τις εφαρμογές που υλοποιήθηκαν παρατίθενται τα αντίστοιχα αριθμητικά αποτελέσματα και γίνεται μια προσπάθεια γραφικής απεικόνισης μέρους αυτών ώστε να γίνουν πιο εύληπτα. Τέλος, εξάγονται κάποια επιμέρους συμπεράσματα ανά εφαρμογή και κάποια γενικότερα συμπεράσματα ως προς την εργασία συνολικά, τα οποία και σημειώνονται στην τελευταία παράγραφο της διπλωματικής εργασίας.

Contents

1 Introduction

1.1 Historical Overview of Bootstrap Methods

As defined in [1], the bootstrap is a form of a larger class of methods that resample from the original data set and thus are called resampling procedures. Some resampling procedures similar to the bootstrap go back a long way [e.g., the jackknife goes back to Quenouille (1949), and permutation methods go back to the 1930s]. Use of computers to do simulation also goes back to the early days of computing in the late 1940s.

However, it was Efron (1979a) who unified ideas and connected the simple non-parametric bootstrap, for independent and identically distributed (IID) observations, which "resamples the data with replacement", with earlier accepted statistical tools for estimating standard errors such as the jackknife and the delta method. The idea of sampling with replacement from the original data did not begin though with Efron. Also even earlier than the first use of bootstrap sampling, there were a few related techniques that are now often referred to as resampling techniques. These other techniques predate Efron's bootstrap. Among them are the jackknife, cross-validation, random subsampling, and permutation procedures.

The idea of resampling from the empirical distribution to form a Monte Carlo approximation to the bootstrap estimate may have been thought of and used prior to Efron. Simon (1969) has been referenced by some to indicate his use of the idea as a tool in teaching elementary statistics prior to Efron. Bruce and Simon(1991, 1995) popularized the bootstrap approach through their company and their associated software. They also continue to use the Monte Carlo approximation to the bootstrap as a tool for introducing statistical concepts in a first elementary course in statistics [see Simon and Bruce (1991, 1995)]. It is clear, however, that widespread use of the methods (particularly by professional statisticians) along with the many theoretical developments occurred only after Efron's 1979 work. It should be noted here that there have been other procedures called bootstrap as well, but differ from Efron's concept.

1.2 General idea of bootstrap

According to [1], if an estimator of the parameter of the sample is denoted as $\hat{\theta}$, the bootstrap distribution for ($\hat{\theta}$ - θ) is the distribution obtained by generating $\hat{\theta}$'s by sampling independently with replacement from the empirical distribution F_n. The bootstrap estimate of the standard error of $\hat{\theta}$ is then the standard deviation of the bootstrap distribution for ($\hat{\theta}$ - θ) . It should be noted here that almost any parameter of the bootstrap distribution can be used as a "bootstrap" estimate of the corresponding population parameter. We could consider the skewness, the kurtosis, the median, or the 95th percentile of the bootstrap distribution for $\hat{\theta}$. Practical application of the technique usually requires the generation of bootstrap samples or resamples (i.e., samples obtained by independently sampling with replacement from the empirical distribution). From the bootstrap sampling, a Monte Carlo approximation of the bootstrap estimate is obtained. The bootstrap is often referred to as a computer -

intensive method. It gets this label because in most practical problems, where it is useful, the estimation is complex and bootstrap samples are required. The procedure is straightforward and described as follows:

1. Generate a sample with replacement from the empirical distribution (a bootstrap sample),
2. Compute the value of $\hat{\theta}$ obtained by using the bootstrap sample in place of the original sample,
3. Repeat steps 1 and 2 k times.

For standard error estimation, k is recommended to be at least 100. This recommendation can be attributed to the article Efron (1987).

What we would like to know for inference is the distribution of $\hat{\theta}$ - θ . What we have is a Monte Carlo approximation to the distribution of θ*- $\hat{\theta}$, where θ* is the estimation that bootstrap method gives for the parameter θ. The key idea of the bootstrap is that for n sufficiently large, we expect the two distributions to be nearly the same. In a few cases, we are able to compute the bootstrap estimator directly without the Monte Carlo approximation.

The basic idea behind the bootstrap is the variability of θ* (based on F_n) around $\hat{\theta}$ will be similar to (or mimic) the variability of $\hat{\theta}$ (based on the true population distribution F_n) around the true parameter value, θ . There is good reason to believe that this will be true for large sample sizes, since as n gets larger and larger, F_n comes closer and closer to F and so sampling with replacements from F_n is almost like random sampling from F.

The strong law of large numbers for independent identically distributed random variables implies that with probability one, F_n converges to F pointwise [see Chung (1974 pp. 131-132) for details]. Strong laws pertaining to the bootstrap can be found in Athreya (1983). A stronger result, the Glivenko-Cantelli theorem [see Chung (1974 ,p. 133)], asserts that the empirical distribution converges uniformly with probability 1 to the distribution F when the observations are independent and identically distributed. Although not stated explicitly in the early bootstrap literature, this fundamental theoretical result lends credence to the bootstrap approach. The theorem was extended in Tucker (1959) to the case of a random sequence from a strictly stationary stochastic process. In addition to the Glivenko-Cantelli theorem, the validity of the bootstrap requires that the estimator (a functional of the empirical distribution function) converge to the "true parameter value" (i.e., the functional for the "true" population distribution). A functional is simply a mapping that assigns a real value to a function. Most commonly used parameters of distribution functions can be expressed as functionals of the distribution, including the mean, the variance, the skewness, and the kurtosis. Interestingly, sample estimates such as the sample mean can be expressed as the same functional applied to the empirical distribution. The concept of an influence function was first introduced by Hampel (1974) as a method for comparing robust estimators. Influence functions have been used in robust statistical methods and in the detection of outlying observations in data sets. Formal treatment of statistical functionals can be

found in Fernholtz (1983). There are also connections for the influence function with the jackknife and the bootstrap as shown by Efron (1982a).

Convergence of the bootstrap estimate to the appropriate limit (consistency) requires some sort of smoothness condition on the functional corresponding to the estimator. In particular, conditions given in Hall (1992a) employ asymptotic normality for the functional and further allow for the existence of an Edgeworth expansion for its distribution function. So there is more needed. For independent and identically distributed observations we require

(1) the convergence of F_n to F (this is satisfied by virtue of the Glivenko-Cantelli theorem),
(2) an estimate that is the corresponding functional of F_n as the parameter is of F (satisfied for means, standard deviations, variances, medians, and other sample quantiles of the distribution), and
(3) a smoothness condition on the functional.

When the bootstrap fails (i.e., bootstrap estimates are inconsistent), it is often because the smoothness conditions are not satisfied (e.g., extreme order statistics such as the minimum or maximum of the sample). These Edgeworth expansions along with the Cornish - Fisher expansions not only can be used to assure the consistency of the bootstrap, but they also provide asymptotic rates of convergence.

Being one of the most important ideas of the last half century in the practice of statistics, the bootstrap also introduced a wealth of innovative probability problems, which in turn formed the basis for the creation of new mathematical theories. Most of these theories have been worked out for the case, dominant also in statistical practice, when the sample consists of i.i.d. random variables.

2 Basic Theory

2.1 Efron Bootstrap

According to Csorgo and Rosalsky (2003) [3], as applied to a sequence $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of arbitrary random variables defined in $L_2(P)$ on a probability space $(\Omega, \mathcal{F}, P)$, and a bootstrap sample size not necessarily equal to the original sample size, the notion of a bootstrap sample is as follows: Let {m(1), m(2), . . . } be a sequence of positive integers and for each n $\in N$, let the random variables $\{X^*_{n,j}$, 1 $\leqslant$ j $\leqslant$ m(n)$\}$ result from sampling m(n) times with replacement from the n observations $X_1, \ldots, X_n$ such that for each of the m(n) selections, each X_k has probability 1/n of being chosen. Alternatively, for each n$\in \mathbb{N}$ we have $X_{n,j} = X_{Z(n,j)}$, 1 $\leqslant$ j $\leqslant$ m(n), where $Z(n, j)$, 1 $\leqslant$ j $\leqslant$ m(n) are independent random variables uniformly distributed over 1,...,n and independent of $\mathbf{X}$; we may and do assume without loss of generality that the underlying space $(\Omega, \mathcal{F}, P)$ is rich enough to accommodate all these random variables with joint distributions as stated. Then $X^*_{n,1}, \ldots, X^*_{n,m(n)}$ are conditionally independent and identically distributed (i.i.d.) given $\mathbf{X_n}$ =$(X_1, ..., X_n)$ with P$\{X^*_{n,1} = X_k|\mathbf{X_n}\}$ = n^{-1} almost surely, 1$\leqslant$ k$\leqslant$ n, n$\in N$. For any sample size n$\in N$, the sequence is referred to as Efron's nonparametric bootstrap sample from $X_1, \ldots, X_n$ with bootstrap sample size m(n).

As stated in [11], we suppose that our data consists of a random sample from an unknown probability distribution F on the real line,

$$X_1, X_2, \ldots, X_n \sim F. \tag{2.1.1}$$

Having observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, we compute the sample mean

$$\bar{x} = \sum_1^n x_n/n, \tag{2.1.2}$$

and wonder how accurate it is as finite estimate of the true mean $\theta = E_F\{X\}$. The standard error $\sigma(F; n, \bar{x})$, that is the standard deviation of $\bar{x}$ for a sample of size n from distribution F, is

$$\sigma(F) = [E_F X^2 - (E_F X)^2]^{1/2}. \tag{2.1.3}$$

The shortened notation $\sigma(F) \equiv \sigma(F; n, \bar{x})$ is allowable because the sample size n and statistic of interest $\bar{x}$ are known, only F being unknown. The standard error is the traditional measure of $\bar{x}$'s accuracy. Unfortunately, we cannot actually use (2.1.3) to assess the accuracy of $\bar{x}$, but we can use the estimated standard error

$$\bar{\sigma} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{(n-1)}}, \tag{2.1.4}$$

the unbiased estimate of $\sigma(F)$.

There is a more obvious way to estimate $\sigma(F)$. Let $\hat{F}$ indicate the empirical probability distribution,

$$\hat{F} : probability \quad mass \quad 1/n \quad on \quad x_1, x_2, \ldots, x_n. \tag{2.1.5}$$

Then we can simply replace F by $\hat{F}$ in (2.1.3), obtaining

$$\hat{\sigma} \equiv \sigma(\hat{F}), \tag{2.1.6}$$

as the estimated standard error for $\bar{x}$. This is the bootstrap estimate. Since

$$\hat{\sigma} \equiv \sigma(\hat{F}) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}, \tag{2.1.7}$$

$\hat{\sigma}$ is not quite the same as $\bar{\sigma}$, but the difference is too small to be important in most applications.

It turns out that we can always numerically evaluate the bootstrap estimate $\hat{\sigma} = \sigma(\hat{F})$, without knowing a simple expression for $\sigma(F)$. The evaluation of $\hat{\sigma}$ is a straightforward Monte Carlo exercise. In a good computing environment, the bootstrap effectively gives the statistician a simple formula like (2.1.4) for any statistic, no matter how complicated.

Standard errors are crude but useful measures of statistical accuracy. They are frequently used to give approximate confidence intervals for an unknown parameter θ

$$\theta \in [\hat{\theta} + \frac{\hat{\sigma}}{\sqrt{n}} \cdot z^{(\alpha/2)}, \hat{\theta} + \frac{\hat{\sigma}}{\sqrt{n}} \cdot z^{(1-\alpha/2)}] \tag{2.1.8}$$

where $z^{(\alpha/2)}$ is the $(\alpha/2) \cdot 100\%$ percentile point of a standard normal variate, e.g., $z^{(0.95)} = 1.645$, and $z^{(1-\alpha/2)} = -z^{(\alpha/2)}$. Interval (2.1.8) is sometimes good, and sometimes not so good. The standard interval (2.1.8) is based on taking literally the large sample normal approximation $(\hat{\theta} - \theta)/\hat{\sigma} \sim N(0, 1)$. Applied statisticians use a variety of tricks to improve this approximation.

Here follows a more careful description of the bootstrap estimate of standard error. For now we will assume that the observed data $y = (x_1, x_2, \ldots, x_n)$ consists of independent and identically distributed (iid) observations $X_1, X_2, \ldots, X_n \sim_{iid} F$, as in (2.1.1). Here F represents an unknown probability distribution on $\mathcal{X}$, the common sample space of the observations. We have a statistic of interest, say $\hat{\theta}(y)$, to which we wish to assign an estimated standard error.

Let $\sigma(F)$ indicate the standard error of $\hat{\theta}$, as a function of the unknown sampling distribution F,

$$\sigma(F) = [Var_F\{\hat{\theta}(y)\}]^{1/2}. \tag{2.1.9}$$

Of course $\sigma(F)$ is also a function of the sample size n and the form of the statistic $\hat{\theta}(y)$, but since both of these are known they need to be indicated in the notation. The bootstrap estimate of standard error is

$$\hat{\sigma} = \sigma(\hat{F}), \tag{2.1.10}$$

where $\hat{F}$ is the empirical distribution (2.1.5), putting probability $1/n$ on each observed data point x_i.

In most cases, there is no simple expression for the function $\sigma(F)$ in (2.1.9). Nevertheless, it is easy to numerically evaluate $\hat{\sigma} = \sigma(\hat{F})$ by means of a Monte Carlo algorithm, which depends on the following notation: $y^* = (x_1^*, x_2^*, \ldots, x_n^*)$ indicates n independent draws from $\hat{F}$, called a bootstrap sample. Because $\hat{F}$ is the empirical distribution of the data, a bootstrap sample turns out to be the same as a random sample of size n drawn with replacement from the actual sample $\{x_1, x_2, \ldots, x_n\}$.

The Monte Carlo algorithm proceeds in three steps:
(i) using a random number of bootstrap samples, say $y^*(1), y^*(2), \ldots, y^*(B)$;
(ii) for each bootstrap sample $y^*(b)$, evaluate the statistic of interest, say $\hat{\theta}^*(b) = \hat{\theta}(y^*(b)), b - 1, 2, \ldots, B$; and
(iii) calculate the sample standard deviation of the $\hat{\theta}^*(b)$ values

$$\hat{\sigma}_B = \left(\frac{\sum_{b=1}^{B}\{\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)\}^2}{B-1}\right)^{1/2}, \tag{2.1.11}$$

$$\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^{B}\hat{\theta}^*(b)}{B}. \tag{2.1.12}$$

It is easy to see that as $B \to \infty$, $\hat{\sigma}_B$ will approach $\hat{\sigma} = \sigma(\hat{F})$, the bootstrap estimate of standard error. All we are doing is evaluating a standard deviation by Monte Carlo sampling. For most situations B in the range 50 to 200 is quite adequate. We will usually ignore the difference between $\hat{\sigma}_B$ and $\hat{\sigma}$, calling both simply "$\hat{\sigma}$".

Why is each bootstrap sample taken with the same sample size n as the original data set? Remember that $\sigma(F)$ is actually $\sigma(F, n, \hat{\theta})$, the standard error for the statistic $\hat{\theta}(\cdot)$ based on a random sample of size n from the unknown distribution F. The bootstrap estimate $\hat{\sigma}$ is actually $\sigma(F, n, \hat{\theta})$ evaluated at $F = \hat{F}$.

According to [11], in regard with confidence intervals, obtaining $\hat{\sigma}$, the estimated standard error of an estimator $\hat{\theta}$ is already discussed. In practice, $\hat{\theta}$ and $\hat{\sigma}$ are usually used together to form the approximate confidence interval $\theta \in \hat{\theta} \pm \frac{\hat{\sigma}}{\sqrt{n}} z^{(\alpha/2)}$, (2.1.8), where $z^{(\alpha/2)}$ is the $(\alpha/2) \cdot 100\%$ percentile point of a standard normal distribution. The interval (2.1.8) is claimed to have approximate coverage probability $1 - \alpha$. (2.1.8) is called the standard interval for θ. The simplest method of bootstrap confidence intervals is to take $\theta \in [\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)]$ as an approximate $1-\alpha$ central interval for θ. This is called the percentile method. Define $\hat{G}(s)$ to be the parametric bootstrap

cdf of $\hat{\theta}^*$,

$$\hat{G}(s) = Prob_*\{\hat{\theta}^* < s\}, \tag{2.1.13}$$

where $Prob_*$ indicates probability computed according to the bootstrap distribution of $\hat{\theta}^*$. The percentile method interval is just the interval between the $(\alpha/2) \cdot 100\%$ and $(1-\alpha/2) \cdot 100\%$ percentiles of the bootstrap distribution of $\hat{\theta}^*$. We will use the notation $\theta[\alpha/2]$ for the $\alpha/2$ level endpoint of an approximate confidence interval for $\theta/2$, so $\theta \in [\theta[\alpha/2], \theta[1-\alpha/2]]$ is the central $1-\alpha$ interval. The percentile interval has endpoints

$$\theta_P[\alpha/2] \equiv \hat{G}^{-1}(\alpha/2). \tag{2.1.14}$$

This compares with the standard interval,

$$\theta_S[\alpha/2] = \hat{\theta} + \frac{\hat{\sigma}}{\sqrt{n}} \cdot z^{(\alpha/2)}. \tag{2.1.15}$$

Note that the bootstrap, according to Efron's original formulation [9], presents some drawbacks. Namely, some observations may be used more than once while others are not sampled at all. To overcome that problem, a more general formulation of the bootstrap has been introduced, the weighted (or smooth) bootstrap, which has also been shown to be computationally more efficient in several applications.

2.2 Weighted Bootstrap

According to [14], let $\overline{X}_n$ denote the sample mean of $X_1, \ldots, X_n$ and S_n^2, the sample variance, $\sum_{i=1}^n (X_i - \overline{X}_n)^2/n$.

Consider a vector of random weights $W_n = (W_{n,1}, W_{n,2}, \ldots, W_{n,n})$ independent of the data $X_1, \ldots, X_n$. Assume that for each integer $n \geqslant 1$, the components of W_n are exchangeable. Now form the generalized bootstrapped mean corresponding to the weight vector W_n:

$$\overline{X}_{\mathcal{W},n} = \frac{1}{n} \sum_{i=1}^n W_{n,i} X_i. \tag{2.2.1}$$

Typically, the weights W_n will also satisfy

$$\begin{aligned}
&(\mathcal{W}_I) \qquad W_{n,i} \geqslant 0, i = 1, 2, \ldots, n, n \geqslant 1, \\
&(\mathcal{W}_{II}) \qquad \sum_{i=1}^n W_{n,i} = n, \\
&(\mathcal{W}_{III}) \qquad \frac{1}{n} \sum_{i=1}^n (W_{n,i} - 1)^2 \to_P c^2 \quad as \quad n \to \infty.
\end{aligned}$$

Assuming that $0 < VarX := \sigma^2 < \infty$ uniformly in t,

$$P(\sqrt{n}(\overline{X}_{\mathcal{W},n} - \overline{X}_n)/S_n \leqslant t | X_1, \ldots, X_n) \to \Phi(t) \quad a.s., \tag{2.2.2}$$

as $n \to \infty$, where Φ is the cumulative distribution function of the standard normal random variable. A bootstrap procedure satisfying (2.2.2) is said to be consistent.

An example of weighted bootstrap in [5] (which satisfies the conditions A1-A5 analyzed in paragraph 2.5):

The iid-weighted bootstraps

Let $Y_1, Y_2, \ldots$ be iid, positive random variables where $\| Y_1 \|_{2,1} < \infty$, and define bootstrap weights by $W_{nj} \equiv Y_j/\overline{Y}_n$. By taking, for instance, Y_i iid exponential(1), the weights become $Dirichlet_n(1, \ldots, 1)$, and we have the Bayesian bootstrap of Rubin (1981) and Lo (1987). When the Y_i's are iid $Gamma(4, 1)$ [so that the W_{ni}/n are equivalent to four-spacings from a sample of $4n - 1$ Uniform(0, 1) random variables], this "iid-weighted" bootstrap is second order equivalent to Efron's multinomial bootstrap for bootstrapping the sample mean, as noted by Weng (1989). Intuitively, these bootstraps are "smoother" in some sense than the multinomial bootstrap since they put some (random) weight at all of the X_i^ω's in the sample, whereas the multinomial bootstrap puts positive weight at about $1 - (1 - n^{-1})^n \to 1 - e^{-1} \approx 0.6322$ proportion of the X_i^ω's, on the average.

For the class of weights of this example, Praestgaard (1990) has shown that **implies**, in parallel to the results for Efron's bootstrap in Gine and Zinn (1990) [13] and the almost sure multiplier central limit theorem in Ledoux and Talagrand (1988, 1991). This bootstrap satisfies A1-A5 with $c^2 = VarY_1/(EY_1)^2$.

2.3 Distribution function and quantile function

As follows from [4], the empirical distribution of a random sample is the uniform discrete measures on the observations.

Let $X_1, \ldots, X_n$ be a random sample from a distribution function F on the real line. The empirical distribution function is defined as

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leqslant t\}. \tag{2.3.1}$$

It is the natural estimator for the underlying distribution F if this is completely unknown.

In [9], the quantile function belonging to F is also considered,

$$\begin{gathered} Q(s) = \inf\{x : F(x) \geqslant s\}, 0 < s < 1, \\ Q(0) = \lim_{t\downarrow 0} Q(t) = Q(0^+), \quad Q(1) = \lim_{t\uparrow 1} Q(t) = Q(1^-) \end{gathered} \tag{2.3.2}$$

and, with $X_{1,n} \leqslant \cdots \leqslant X_{n,n}$ denoting the order statistics of $X_1, \ldots, X_n$, its empirical counterpart

$$Q_n(s) = \begin{cases} X_{k,n}, & \text{if } (k-1)/n < s \leqslant k/n; \quad k = 1, \ldots, n, \\ X_{1,n}, & \text{if } s = 0. \end{cases} \tag{2.3.3}$$

According to [4], because $n\mathbb{F}_n(t)$ is binomially distributed with mean $nF(t)$, this estimator is unbiased. By the law of large numbers it is also consistent,

$$\mathbb{F}_n(t) \xrightarrow{\text{a.s.}} F(t), \quad \text{every} \quad t. \tag{2.3.4}$$

By the central limit theorem it is asymptotically normal,

$$\sqrt{n}(\mathbb{F}_n(t) - F(t)) \xrightarrow{P} N(0, F(t)(1-F(t))). \tag{2.3.5}$$

These results get improved by considering $t \mapsto \mathbb{F}_n(t)$ as a random function, rather than as a real-valued estimator for each t separately. This is of interest on its own account but also provides a useful starting tool for the asymptotic analysis of other statistics, such as quantiles, rank statistics, or trimmed means.

The Glivenko-Cantelli theorem extends the law of large numbers and gives uniform convergence. The uniform distance

$$\|\mathbb{F}_n - F\|_\infty = \sup_{t\in\mathbb{R}} |\mathbb{F}_n(t) - F(t)| \tag{2.3.6}$$

is known as the Kolmogorov-Smirnov statistic.

THEOREM (Glivenko-Cantelli) 2.3.1. *If $X_1, X_2, \ldots$ are i.i.d. random variables with distribution function F, then $\|\mathbb{F}_n - F\|_\infty \xrightarrow{a.s.} 0$.*

The extension of the central limit theorem to a "uniform" or "functional" central limit theorem is more involved. A first step is to prove the joint weak convergence of finitely many coordinates. By the multivariate central limit theorem, for every $t_1, \ldots, t_k$,

$$\sqrt{n}(\mathbb{F}_n(t_1) - F(t_1), \ldots, \mathbb{F}_n(t_k) - F(t_k)) \rightsquigarrow (\mathbb{G}_F(t_1), \ldots, \mathbb{G}_F(t_k)), \tag{2.3.7}$$

where the vector on the right has a multivariate-normal distribution, with mean zero and covariances

$$\mathbf{E}\mathbb{G}_F(t_i)\mathbb{G}_F(t_j) = F(t_i \wedge t_j) - F(t_i)F(t_j). \tag{2.3.8}$$

This suggests that the sequence of empirical processes $\sqrt{n}(\mathbb{F}_n - F)$, viewed as random functions, converges in distribution to a Gaussian process $\mathbb{G}_F$ with zero mean and covariance functions as in the preceding display. According to an extension of Donsker's theorem, this is true in the sense of weak convergence of these processes in the Skorohod space $D[-\infty, \infty]$ equipped with the uniform norm. This limit process $\mathbb{G}_F$ is known as an F-Brownian bridge process, and as a standard (or uniform) Brownian bridge if F is the uniform distribution λ on [0, 1]. From the form of the covariance function it is clear that the F-Brownian bridge is obtainable as $\mathbb{G}_\lambda \circ F$ from a standard bridge $\mathbb{G}_\lambda$. The name "bridge" results from the fact that the sample paths of the process are zero ("tied down") at the endpoints 0 and 1. This is a consequence of the fact that the difference of two distribution functions is zero at these points.

THEOREM (Donsker) 2.3.2. *If $X_1, X_2, \ldots$ are i.i.d. random variables with distribution function F, then the sequence of empirical processes $\sqrt{n}(\mathbb{F}_n - F)$ converges in distribution in the space $D[-\infty, \infty]$ to a tight random element $\mathbb{G}_F$, whose marginal distributions are zero-mean normal with covariance function (2.3.8).*

2.4 Bootstrap of empirical measures

We denote the empirical distribution, as denoted in [4], by $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$, where δ_X is the Dirac probability distribution that is degenerate at X. According to [5], the bootstrapped empirical measure can alternatively be expressed as

$$\hat{\mathbb{P}}_n \equiv \tfrac{1}{n}\sum_{j=1}^n M_{nj}\delta_{X_j^\omega}$$

where $\underline{M}_n \sim Mult_n(n, (n^{-1}, \ldots, n^{-1}))$. As observed by Efron [(1982), Section 9.4, pages 71-72], this suggests that there are, in fact, not just one but several ways to bootstrap. If $\mathbf{W} = (W_{nj}, j = 1 \ldots, n, n = 1, 2, \ldots)$ denote a triangular array of nonnegative random variables with $\sum_{j=1}^n W_{nj} = n$; then

$$\hat{\mathbb{P}}_n \equiv \frac{1}{n}\sum_{j=1}^n W_{nj}\delta_{X_j^\omega} \tag{2.4.1}$$

defines a weighted bootstrap empirical measure. We refer to these as bootstraps with exchangeable weights to distinguish them from Efron's (multinomial) bootstrap. Bootstraps with exchangeable weights have not been considered as closely as the Efron bootstrap, and they are not yet widely used in statistical practice.

2.5 Empirical processes indexed by functions-Bootstrap empirical processes

As defined in [4], l et $X_1, \ldots, X_n$ be a random sample from a probability distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$. Given a measurable function $f : \mathcal{X} \mapsto \mathbb{R}$, we write $\mathbb{P}_n f$ for the expectation of f under the empirical measure, and Pf for the expectation under P. Thus

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i), \qquad Pf = \int f dP. \tag{2.5.1}$$

As Efron (1982) stated in [10], in the modern theory of empirical processes it is customary to identify $\mathbf{P}$, $\mathbb{P}_n$ and $\mathbb{X}_n$ with the mappings given by

$$f \to \int f dP = Pf, f \to \int f d\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} f(X_i) = \mathbb{P}_n f$$

and

$$f \to \int f d\mathbb{X}_n = n^{-1/2} \sum_{j=1}^{n} (f(X_j) - Pf) = \mathbb{X}_n(f), \text{ respectively.}$$

Here, $f \in \mathcal{F}$, and $F \subset L_2(\mathrm{P})$ is a collection of functions mapping the sample space $\mathbf{X}$ to $\mathbb{R}$. In this way, $\mathbb{X}_n$ becomes a random element of $l^\infty(\mathcal{F})$, the space of bounded real functions on $\mathcal{F}$. The most straightforward example is to take $\mathbf{X}$ = [0,1] and let $\mathcal{F}$ be the collection of indicator functions of sets of the form [0, c], 0<c⩽1. In this case $\mathbb{P}_n$ becomes the ordinary empirical distribution function. Donsker's theorem states that the empirical process converges in distribution to a Brownian bridge on [0,1]. The same holds for the function-indexed empirical process.

According to [4], by the law of large numbers, the sequence $\mathbb{P}_n f$ converges almost surely to Pf, for every f such that Pf is defined. The abstract Glivenko-Cantelli theorems make this result uniform in f ranging over a class of functions. A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ is called P-Glivenko-Cantelli if

$$\|\mathbb{P}_n f - Pf\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \xrightarrow{\text{a.s.}} 0. \tag{2.5.2}$$

The empirical process evaluated at f is defined as $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$. By the multivariate central limit theorem, given any finite set of measurable functions f_i with $Pf_i^2 < \infty$,

$$(\mathbb{G}_n f_1, \ldots, \mathbb{G}_n f_k) \rightsquigarrow (\mathbb{G}_P f_1, \ldots, \mathbb{G}_P f_k), \tag{2.5.3}$$

where the vector on the right possesses a multivariate-normal distribution with mean zero and covariances

$$E\mathbb{G}_P f \mathbb{G}_P g = Pfg - PfPg. \tag{2.5.4}$$

The abstract Donsker theorems make this result "uniform" in classes of functions. A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ is called P-Donsker if the sequence of processes $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ converges in distribution to a tight limit process in the space $l^\infty(\mathcal{F})$. Then the limit process is a Gaussian process $\mathbb{G}_P$ with zero mean and covariance function as given in the preceding display and is known as a P-Brownian bridge. Of course, the Donsker property includes the requirement that the sample paths $f \mapsto \mathbb{G}f$ are uniformly bounded for every n and every realization of $X_1, \ldots, X_n$. This is the case, for instance, if the class $\mathcal{F}$ has a finite and integrable envelope function F: a function such that $|f(x)| \leqslant F(x) < \infty$, for every x and f. It is not required that the function $x \mapsto F(x)$ be uniformly bounded.

For convenience of terminology we define a class $\mathcal{F}$ of vector-valued functions $f : x \mapsto \mathbb{R}^k$ to be Glivenko-Cantelli or Donsker if each of the classes of coordinates $f_i : x \mapsto \mathbb{R}$ with $f = (f_i, \ldots, f_k)$ ranging over $\mathcal{F}(i = 1, 2, \ldots, k)$ is Glivenko-Cantelli or Donsker. It can be shown that this is equivalent to the union of the k coordinate classes being Glivenko-Cantelli or Donsker.

Whether a class of functions is Glivenko-Cantelli or Donsker depends on the "size" of the class. A finite class of integrable functions is always Glivenko-Cantelli, and a finite class of square-integrable functions is always Donsker. On the other hand, the class of the square-integrable functions is Glivenko-Cantelli or Donsker only in trivial cases. A relatively simple way to measure the size of a class $\mathcal{F}$ is in terms of entropy.

Praestgaard and Wellner (1992) [5] establish sufficient conditions on the weights $\mathbf{W}$ for the exchandeably weighted bootstrap to "work" asymptotically, in the sense that $\mathcal{F} \in$ CLT(P) and $PF^2 < \infty$ (and sufficient measurability) implies that, with $\hat{\mathbb{P}}_n$ given above, $\hat{\mathbb{X}}_n \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \hat{\mathbb{P}}_n^\omega) \Rightarrow G_P$, a.s.

The corresponding bootstrap empirical process is

$$\hat{\mathbb{X}}_n(\omega) = \hat{\mathbb{X}}_n^\omega \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n^\omega) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (W_{nj} - 1)\delta_{X_j^\omega}. \tag{2.5.5}$$

For a random variable Y we define $\| Y \|_{2,1} \equiv \int_0^\infty (Pr(| Y | > t))^{1/2} dt$. Often the quantity $\sqrt{n}(\theta_n - \theta)$ to be bootstrapped can be expressed as a function of the empirical process which is defined as

$$\mathbb{X}_n \equiv \sqrt{n}(\mathbb{P}_n - \mathbf{P}).$$

Under the following quite general conditions on $\mathbf{W}$ we shall establish a central limit theorem for the weighted bootstrap:

A1. The vectors $\overline{W}_n$ are exchangeable, $n = 1, 2, \ldots$.
A2. $W_{nj} \geqslant 0$, for all n, j and $\sum_{j=1}^{n} W_{nj} = n$, for all n.
A3. $sup_n \parallel W_{n1} \parallel_{2,1} \equiv M(\mathbf{W}) < \infty$.
A4. $\lim_{\lambda\to\infty} \lim sup_{n\to\infty} sup_{t\geqslant\lambda} t^2 P(W_{n1} \geqslant t) = 0$.
A5. $(1/n)\sum_{j=1}^{n}(W_{nj} - 1)^2 \to c^2 > 0$, in probability.

The main result is the following [5]:

THEOREM 2.5.1. *Let $\mathcal{F} \in M(P)$ be a class of $L_2(P)$ functions, and let $\mathbf{W}$ be a triangular array of bootstrap weights satisfying assumptions A1-A5. Then*

$$\mathcal{F} \in CLT(P) \text{ and } PF^2 < \infty$$

imply that

$$\hat{\mathbb{X}}_n^{\omega} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (W_{nj} - 1)\delta_{X_j^{\omega}} \Rightarrow c \quad G_P \in l^{\infty}(\mathcal{F}) a.s., \tag{2.5.6}$$

where c in (2.5.6) is given by assumption A5.

THEOREM 2.5.2. *Let $\mathcal{F} \in M(P)$ be a class of $L_2(P)$ functions, and let $\mathbf{W}$ be a triangular array of bootstrap weights satisfying assumptions A1-A5. Then*

$$\mathcal{F} \in CLT(P)$$

implies that

$$\hat{\mathbb{X}}_n^{\omega} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (W_{nj} - 1)\delta_{X_j^{\omega}} \Rightarrow c \quad G_P \in l^{\infty}(\mathcal{F}) \quad in \quad probability, \tag{2.5.7}$$

where c in (2.5.7) is given by assumption A5.

The calculations leading to Theorem 2.5.1 also show the following new result about the Efron bootstrap with arbitrary bootstrap sample size. Form an iid sample $\hat{X}_{n1}, \ldots, \hat{X}_{nm}$ from $\mathbb{P}_n^{\omega}$, and let $\hat{\mathbb{P}}_{n,m} \equiv m^{-1}\sum_{j=1}^{m} \delta_{\hat{X}_{nj}}$ denote the bootstrap empirical measure for the bootstrap sample of size m.

COROLLARY 2.5.3. *$\mathcal{F} \in CLT(P)$ and $PF^2 < \infty$ imply that*

$$\sqrt{m}(\hat{\mathbb{P}}_{m,n} - \mathbb{P}_n^{\omega}) = \tfrac{1}{\sqrt{m}} \textstyle\sum_{j=1}^{m} (\delta_{\hat{X}_{nj}} - \mathbb{P}_n^{\omega}) \Rightarrow G_P \text{ in } l^{\infty}(\mathcal{F}) \text{ a.s. as } n \wedge m \to \infty.$$

This result for regular sequences m_n and the corresponding "in probability" result in general were known to Gine and Zinn [Arcones and Gine(1992)].

COROLLARY 2.5.4. *$\mathcal{F} \in CLT(P)$ implies that*

$$\sqrt{m}(\hat{\mathbb{P}}_{m,n} - \mathbb{P}_n^{\omega}) \Rightarrow G_P \text{ in } l^{\infty}(\mathcal{F}) \text{ in probability as } n \wedge m \to \infty.$$

Here [2] follows the development of the complete bootstrapped parallel to the asymptotic theory of weighted empirical and quantile processes. Utilizing this parallel

theory, a general body of techniques, which establish the asymptotic validity of the bootstrap method of constructing confidence bands for statistical functions, is presented. These techniques are demonstrated to be applicable to the construction of asymptotic bootstrap confidence bands for a variety of concrete functions.

As Efron (1979) defined in [9], the bootstrap empirical and quantile processes are, respectively,

$$m^{1/2}\{\tilde{F}_{m,n}(x) - F_n(x)\}, \quad -\infty < x < \infty, \tag{2.5.8}$$

and

$$m^{1/2}\{\tilde{Q}_{m,n}(s) - Q_n(s)\}, \quad 0 \leqslant s \leqslant 1, \tag{2.5.9}$$

where

$$\tilde{F}_{m,n}(x) = m^{-1}\#\{k : 1 \leqslant k \leqslant m, \tilde{X}_k \leqslant x\}$$

$$\tilde{Q}_{m,n}(s) = \tilde{X}_{k,m} \text{ if } (k-1)/m < s \leqslant k/m,\ k = 1, \ldots, m,$$

and

$$\tilde{Q}_{m,n}(0) = \tilde{X}_{1,m}, \text{ with } \tilde{X}_{1,m} \leqslant \cdots \leqslant \tilde{X}_{m,m}$$

standing for the order statistics of the bootstrap sample $\tilde{X}_1, \ldots, \tilde{X}_m$ with resampling size m.

Among many other things, Bickel and Freedman (1981) established the weak convergence of the processes in (2.5.8) and (2.5.9), and they were able to deduce the asymptotic validity of the bootstrap method of forming confidence bands for F and Q.

Here [9] we consider the validity of the bootstrap for general empirical functions on the real line containing as special cases the empirical distribution function and the empirical quantile function. This means that the ultimate aim is to show the asymptotic validity of bootstrap confidence-band estimation of functions on the real line generally different from F and Q. It is noted here that any convergence and other relations will be understood throughout as $n \to \infty$ if not specified otherwise.

Let $\theta_F(\cdot)$ be a statistical function of interest defined in an interval $I \subseteq \mathbb{R}$ and let $\theta_n(\cdot) = \theta_n(\cdot; X_1, \ldots, X_n)$ be an appropriate estimator of $\theta_F(\cdot)$ on I. We can allow I to be the union of a finite number of disjoint (finite or infinite) intervals. Typically, for the process

$$r_n(\cdot) = n^{1/2}(\theta_n(\cdot) - \theta_F(\cdot)) \tag{2.5.10}$$

one can find a sequence of copies $\mathcal{G}_F^{(n)}(\cdot)$ of a separable Gaussian process $\mathcal{G}_F(\cdot)$ on I, i.e. $\mathcal{G}_F^{(n)}(\cdot) =_{\mathcal{D}} \mathcal{G}_F(\cdot)$ for each $n \geqslant 1$, such that on an appropriate probability space $(\Omega, \mathcal{A}, P)$,

$$P\{\sup_{t\in I}|\mathcal{G}_F(t)|<\infty\}=1 \quad \textit{and} \quad \sup_{t\in I}|r_n(t)-\mathcal{G}_F^{(n)}|\xrightarrow{P}0. \tag{2.5.11}$$

Consequently, given $0<\alpha<1$, we have (on any probability space where the X's are defined)

$$P\{\theta_n(t)-cn^{-1/2}\leqslant\theta_F(t)\leqslant\theta_n(t)+cn^{-1/2},t\in I\}\to 1-\alpha, \tag{2.5.12}$$

provided that $G_F(c)=1-\alpha$ and $c=c(\alpha/2,F)$ is a continuity point of the distribution function

$$G_F(x)=P\{\sup_{t\in I}|\mathcal{G}_F(t)|\leqslant x\},\quad x\geqslant 0. \tag{2.5.13}$$

This means that $\{\theta_n(t)\pm cn^{-1/2},t\in I\}$, is an asympotically correct $(1-\alpha)100\%$ confidence band for the statistical function θ_F.

It is rare that this method of forming asymptotically correct confidence bands is feasible, since there are only a few cases when $c=c(\alpha,F)$ is independent of F and its analytical form is known. The most well-known case when this is true is the choice $\theta_F=F$, $\theta_n=F_n$ and F is continuous.

Consider the bootstrapped version of the empirical function $\theta_n(\cdot)$ given by $\tilde{\theta}_{m,n}(\cdot)=\theta_m(\cdot;\tilde{X}_1,\ldots,\tilde{X}_m)$. Suppose we were able to show that on an appropriate extension $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{P})$ of the above probability space $(\Omega,\mathcal{F},P)$ there exist a sequence of versions $\tilde{\mathcal{G}}_F^{(m)}$ of the process $\mathcal{G}_F$, i.e. $\{\tilde{\mathcal{G}}_F^{(m)}(t)\,t\in I\}=_{\mathcal{D}}\{\mathcal{G}_F(t):t\in I\}$ for each n and a sequence of versions $\hat{\theta}_{m,n}$ of the process $\tilde{\theta}_{m,n}$ such that

$$\textit{the sequences}\quad \{X_n\}_{n=1}^{\infty}\quad \textit{and}\quad \{\tilde{\mathcal{G}}_F^{(m)}\}_{n=1}^{\infty}\quad \textit{are independent}, \tag{2.5.14}$$

$$\{(\theta_n(s),\tilde{\theta}_{m,n}(t)):s,t\in I\}\overset{D}{=}\{(\theta_n(s),\hat{\theta}_{m,n}(t)):s,t\in I\} \tag{2.5.15}$$

for each n,

$$\sup_{t\in I}|m^{1/2}(\hat{\theta}_{m,n}(t)-\theta_n(t))-\tilde{\mathcal{G}}_F^{(m)}(t)|\xrightarrow{\tilde{P}}0, \tag{2.5.16}$$

where $m=m(n)\to\infty$ at an appropriate rate. (Here and in what follows $=_{\mathcal{D}}$ stands for the equality of all finite dimensional distributions of the stochastic processes on the two sides.) From (2.5.11), (2.5.14), (2.5.15) and (2.5.16) we can conclude that whenever x is a continuity point of G_F in (2.5.13) then

$$P\{\sup_{t\in I}m^{1/2}|\tilde{\theta}_{m,n}(t)-\theta_n(t)|\leqslant x|X_1,\ldots,X_n\}\xrightarrow{P}G_F(x) \tag{2.5.17}$$

(on any probability space) for the same $m = m(n)$ sequence. Now fix $0 < \alpha < 1$ and suppose we can show that G_F is continuous at both $c = c(\alpha/2, F) = \inf\{x : G_F(x) \geqslant 1 - \alpha\}$ and $d = d(\alpha/2, F) = \sup\{x : G_F(x) \leqslant 1 - \alpha\}$.

Given then the observations $X_1, \ldots, X_n$, for each m let $c_m = c_m(X_1, \ldots, X_n)$ be defined as

$$c_m = \inf\{x : P\{\sup_{t \in I} m^{1/2}|\tilde{\theta}_{m,n}(t) - \theta_n(t)| \leqslant x | X_1, \ldots, X_n\} \geqslant 1 - \alpha\}. \tag{2.5.18}$$

By 2.5.11 and 2.5.17 it is concluded that

$$P\{\sup_{t \in I} n^{1/2}|\theta_n(t) - \theta_F(t)| \leqslant c_m\} \to 1 - \alpha, \tag{2.5.19}$$

provided $m = m(n) \to \infty$ at the rate required by (2.5.14), (2.5.15) and (2.5.16). Hence from (2.5.19) we see that an asymptotically $(1-\alpha)100\%$ confidence band for θ_F is given by

$$\{\theta_n(t) \pm c_m n^{-1/2}, t \in I\}. \tag{2.5.20}$$

However, since, given $X_1, \ldots, X_n$, $\sup\{m^{1/2}|\tilde{\theta}_{m,n}(t) - \theta_n(t)| : t \in I\}$ can take on as many as n^m possible values, which is typically an astronomically large number, c_m must in most practical situations be estimated by Monte Carlo simulation.

Bickel and Freedman (1981) established the validity of the above procedure in two cases. One is when $\theta_F = F$ is an arbitrary distribution function, $I = (-\infty, \infty)$ and $\theta_n = F_n$. The other is when $\theta_F = Q$ as given in (2.3.2) , $I = [a, b] \subset (0, 1), \alpha < b$, and $\theta_n = Q_n$. If $\mathcal{G}_F$ is any separable, mean-zero, almost surely bounded Gaussian process such that $Var\mathcal{G}_F(t) > 0$ for some point t of I, then $\mathcal{G}_F$ is continuous on the whole half-line $(0, \infty)$.

The philosophy of the bootstrap principle includes the appealing heuristic idea that bootstrapped versions $\tilde{r}_{m,n} = m^{1/2}(\tilde{\theta}_{m,n} - \theta_n)$ of processes r_n behave asymptotically the same way as the original processes r_n under the same regularity conditions on the underlying distribution, and, therefore, under the (preferably optimal) regularity conditions of (2.5.11) we also have the final bootstrap confidence-band statement in (2.5.19). One of the most powerful techniques of establishing a result (2.5.11) on the real line (usually under optimal regularity conditions) is the weighted approximation method of Csorgo, Csorgo, Horvath and Mason (1986a). The origin of this method goes back to the works of Chibisov, Pyke and Shorack, Shorack and O' Reilly.

Whatever result (2.5.11) can be proved by the weighted approximation method under some regularity conditions on F the bootstrap is automatically valid under the same conditions, that is, we also have (2.3.2).

We have just defined our understanding of the asymptotic validity of the bootstrap in the present confidence-band context. With the notable exception of Efron (1979) himself, most authors justify the bootstrap by proving that the statement corresponding to (2.5.17) holds in the stronger sense of almost sure convergence rather than convergence in probability. The present approach cannot be adapted to produce this. The in

probability version of the justification of the bootstrap is in fact more universally applicable than the almost sure version. There are cases when the bootstrap construction of confidence intervals and bands cannot be justified by almost sure convergence while it can be in the in probability sense.

2.6 Markov Chains

DEFINITION 2.6.1. *The random variable sequence $X = (X_n)_{n\in\mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in the finite set $E = \{1, \ldots, s\}$, is a Markov chain if, for any nonnegative integer n and any states $i, j, i_0, i_1, \ldots, i_{n-1} \in E$, we have: [6]*

$$\begin{aligned}\mathbb{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) \\ = \mathbb{P}(X_{n+1} = j | X_n = 1) = p_{ij;n}.\end{aligned} \tag{2.6.1}$$

The equality is called Markov property. If $p_{ij;n} = p_{ij}$ does not depend on n, then the Markov chain is called homogeneous (with respect to time). In the sequel, we consider only homogeneous Markov chains. The function $(i, j) \to p_{ij}$, defined on $E \times E$, is called the transition function of the chain. As we are concerned only with finite state space Markov chains, we can represent the transition function by a square matrix $p = (p_{ij})_{i,j\in E} \in \mathcal{M}_E$. The n-step transition function is defined by

$$p_{ij}^{(n)} = \mathbb{P}(X_{n+m} = j | X_m = i), \text{for any } m \in \mathbb{N}.$$

The transition function of a Markov chain satisfies the following properties:
1. $p_{ij} \geqslant 0$,
2. $\sum_{j\in E} p_{ij} = 1$,
3. $\sum_{k\in E} p_{ik}^{(n)} p_{kj}^{(m)} = p_{ij}^{(n+m)}$,
for any $i, j \in E$ and $n, m \geqslant 0$.

PROPOSITION 2.6.2. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain of transition function p and initial distribution α. For any $n \geqslant 1$ and any states $i_0, i_1, \ldots, i_n \in E$, we have:*
1. $\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}$,
2. $\mathbb{P}(X_{n+1} = i_1, \ldots, X_{n+k-1} = i_{k-1}, X_{n+k} = i_k | X_n = i_0) = p_{i_0 i_1} \ldots p_{i_{k-1} i_k}$,
3. $\mathbb{P}(X_{n+m} = j | X_m = i) = \mathbb{P}(X_n = j | X_0 = i) = p_{ij}^n$.

DEFINITION (stopping time or Markov time) 2.6.3. *A random variable T, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with values in $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, is called a stopping time with respect to the sequence $(X_n)_{n\in\mathbb{N}}$ if the occurrence of the event $\{T = n\}$ is determined by the past of the chain up to time n, $(X_k; k \leqslant n)$. More precisely, let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n), n \geqslant 0$, be the σ-algebra generated by $X_0, \ldots, X_n$, i.e., the information known at time n. The random variable T is called a stopping time if, for every $n \in \mathbb{N}, \{T = n\} \in \mathcal{F}_n$.*

DEFINITION (strong Markov property) 2.6.4. *The Markov chain $(X_n)_{n\in\mathbb{N}}$ is said to have the strong Markov property if, for any stopping time T, for any integer $m \in \mathbb{N}$ and state $j \in E$ we have* $\mathbb{P}(X_{m+T} = j | X_k, k \leqslant T) = \mathbb{P}_{X_T}(X_m = j)$ *a.s.*

PROPOSITION 2.6.5. *Any Markov chain has the strong Markov property.*

Let us define:

- $\eta_i = min\{n|n \in \mathbb{N}^*, X_n = i\}$ (with min $\varnothing = \infty$), the first passage time of the chain in state i. If $X_0(\omega) = i$ then η_i is the recurrence time of state i. Note that $\eta_i > 0$.

- $N_i(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=i\}}$, the time spent by the chain in state i, during the time interval $[0, n-1]$. If $n = \infty$, then we note $N_i = N_i(\infty)$, with N_i taking values in $\bar{\mathbb{N}}$.

- $N_{ij}(n) = \sum_{k=1}^{n} \mathbf{1}_{\{X_{k-1}=i, X_k=j\}}$, the number of direct transition from i to j, up to time n. If $n = \infty$, then we note $N_{ij} = N_{ij}(\infty)$, with N_{ij} taking values in $\bar{\mathbb{N}}$.

DEFINITION (recurrent and transient Markov chain) 2.6.6. *A state $i \in E$ is called recurrent if $\mathbb{P}_i(\eta_i < \infty) = 1$; in the opposite case, when $\mathbb{P}_i(\eta_i < \infty) < 1$, the state i is called transient. A recurrent state i is called positive recurrent if $\mu_{ii}^* = \mathbb{E}_i[\eta_i] < \infty$ and null recurrent if μ_{ii}^*.*

The Markov chain is said to be (positive/null) recurrent (resp. transient), if all the states are (positive/null) recurrent (resp. transient).

DEFINITION (irreducible Markov chain) 2.6.7. *If for any states i, j there is a positive integer n such that $p_{ij}^{(n)} > 0$, then the Markov chain is said to be irreducible.*

DEFINITION 2.6.8. *A probability distribution ν on E is said to be stationary or invariant for the Markov chain $(X_n)_{n\in\mathbb{N}}$ if, for any $j \in E$,*

$$\sum_{j\in E} \nu(i) p_{ij} = \nu(j),$$

or, in matrix form,

$$\nu p = \nu,$$

where $\nu = (\nu(1), \ldots, \nu(s))$ is a row vector.

PROPOSITION 2.6.9. *For a recurrent state i, we have: $\nu(i) = 1/\mu_{ii}^*$.*

DEFINITION 2.6.10. *A state $i \in E$ is said to be periodic of period $d > 1$, or d-periodic, if g.c.d.$\{n|n > 1, p_{ii}^n > 0\} = d$. If $d = 1$, then the state i is said to be aperiodic.*

DEFINITION 2.6.11. *An aperiodic recurrent state is called ergodic. An irreducible Markov chain with one state ergodic (and then all states ergodic) is called ergodic.*

PROPOSITION (ergodic theorem for Markov chains) 2.6.12. *For an irreducible and aperiodic Markov chain we have*

$$p_{ij}^n \underset{n\to\infty}{\longrightarrow} \nu(j)$$

for any $i, j \in E$.

PROPOSITION 2.6.13. *For an irreducible and aperiodic Markov chain, there exists ν the stationary distribution of the chain such that p^n converges at an exponential rate to $\Pi = \mathbf{1}^\top \nu$, where $\mathbf{1} = (1, \ldots, 1)$.*

PROPOSITION (strong law of large numbers) 2.6.14. *For an ergodic Markov chain* $(X_n)_{n\in\mathbb{N}}$*, with stationary distribution* ν*, we have*

$$\tfrac{1}{n}\sum_{k=0}^{n} g(X_k) \xrightarrow[n\to\infty]{a.s.} \sum_{i\in E} \nu(i)g(i) =: \bar{g}.$$

PROPOSITION [16] (central limit theorem) 2.6.15. *For an ergodic Markov chain* $(X_n)_{n\in\mathbb{N}}$*, with stationary distribution* ν*, we have*

$$\sqrt{n}(\tfrac{1}{n}\sum_{k=0}^{n} g(X_k) - \bar{g}) \xrightarrow[n\to\infty]{D} \mathcal{N}(0,\sigma^2),$$

where

$$\sigma^2 := \tilde{g}diag(\nu)[2Z - I]\tilde{g}^{\top},$$

where $Z = (I - p + \Pi)^{-1}$ *is the fundamental matrix of* p *and* $\tilde{g}$ *is the row vector with components* $\tilde{g}(i) := g(i) - \bar{g}, \quad i \in E$.

3 Application - i.i.d. case

This is the first part of arithmetical - experimental approach to the calculation of confidence intervals. Here, we used independent and identically distributed (iid) variables $\{X_i\}_{i=1,\ldots,n}$, firstly in classical method of calculation and secondly within weighted bootstrap method. In both cases, classical and bootstrap one, it was necessary to calculate the mean value of the sample and the standard deviation.
For the classical method, the sample mean is defined as

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n},$$

and the standard deviation as

$$S_n = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}.$$

For the method of weighted bootstrap, bootstrap mean value is defined, according to (2.2.1), as

$$\overline{X}_{\mathcal{W},n} = \sum_{i=1}^{n} W_{n,i} X_i,$$

for the weight vector $W_n = (W_{n,1}, W_{n,2}, \ldots, W_{n,n})$, where the weights W satisfy conditions $\mathcal{W}_I - \mathcal{W}_{III}$.
The bootstrap standard deviation is calculated as

$$S_B = \sqrt{\sum_{i=1}^{n} W_{n,i} \cdot (X_i - \overline{X}_B)^2}.$$

Next, the errors of mean and variance were calculated, for classic and bootstrap method:
$y_1 = \overline{X}_n - \mu$, where μ is the theoretical mean value,
$y_3 = \overline{X}_{\mathcal{W},n} - \overline{X}_n$,
$y_2 = S_n - \sigma$, where σ is the theoretical variance,
$y_4 = S_B - S_n$.

Then, the confidence intervals could be calculated, both for classic and bootstrap method:

$$confInterval = \overline{X}_n \pm \frac{z^{(\alpha/2)} \cdot S_n}{\sqrt{n}},$$

$$confIntervalBootstrap = \overline{X}_{\mathcal{W},n} \pm \frac{z^{(\alpha/2)} \cdot S_B}{\sqrt{n}},$$

where $z^{(\alpha/2)}$ is the $(\alpha/2) \cdot 100\%$ percentile point of a standard normal distribution. According to standard method for confidence intervals (paragraph 2.1), these interval ends give an approximate $1-\alpha$ central interval for $\overline{X}$. All the above calculations were conducted:

- for two different percentile points, $\alpha = 0.1$ and $\alpha = 0.05$ (thus for central intervals of 90% and 95%), which give $z(0.95) = 1.645$ and $z(0.975) = 1.96$,

- for four different sizes of the sample, which are $n = 1000, \quad n = 2000, \quad n = 5000$, and,

- for four different number of times that we get Bootstrap weights, which are $k = 1000, \quad k = 2000, \quad k = 5000$,

- for random variables following the law of normal distribution $(X_i \sim \mathcal{N}(0,1))$, exponential distribution $(X_i \sim Exp(1))$ or uniform distribution $(X_i \sim U(0,1))$ and,

- for bootstrap weights following the law of exponential distribution $(W_{n,i} \sim Exp(1))$ or the law of Poisson distribution $(W_{n,i} \sim P(1))$.

For the results we got, we also calculated the ratio of the estimated confidence interval over the confidence interval estimated with bootstrap method, which is (right margin-left margin)/(right margin B.-left margin B.). This ratio (λ) as a function of the sample size (n) is presented graphically. The results and graphs are presented below:

X_i	$\sim$	Exp(1)		W_i	$\sim$	Exp(1)		conf. int	90%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	0,0264	-0,0088	0,1703	-0,0365	0,8903	1,1625	0,8857	1,1495	1,0322
500	500	-0,0316	0,0001	-0,0707	0,0020	0,9000	1,0368	0,9000	1,0370	0,9979
1000	1000	0,0026	0,0006	-0,0665	0,0014	0,9540	1,0512	0,9546	1,0518	0,9985
1000	2000	0,0150	-0,0002	0,0180	-0,0018	0,9620	1,0680	0,9619	1,0677	1,0018
1000	5000	-0,0115	-0,0004	0,0010	-0,0007	0,9364	1,0406	0,9361	1,0401	1,0007
2000	1000	-0,0005	-0,0002	0,0262	-0,0010	0,9618	1,0372	0,9616	1,0370	1,0010
2000	2000	0,0025	-0,0006	-0,0334	-0,0006	0,9669	1,0381	0,9664	1,0374	1,0006
2000	5000	-0,0628	0,0003	-0,0286	-0,0002	0,9015	0,9729	0,9018	0,9732	1,0002
5000	1000	0,0150	-0,0007	0,0434	-0,0014	0,9907	1,0393	0,9901	1,0385	1,0013
5000	2000	0,0032	0,0000	-0,0176	-0,0003	0,9803	1,0261	0,9804	1,0260	1,0003
5000	5000	0,0132	-0,0001	-0,0004	-0,0005	0,9899	1,0365	0,9899	1,0363	1,0005

X_i	$\sim$	Exp(1)		W_i	$\sim$	Exp(1)		conf. int	95%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,0077	0,0000	-0,0775	0,0011	0,8644	1,1202	0,8643	1,1203	0,9988
500	500	0,0522	-0,0004	-0,0024	-0,0019	0,9648	1,1396	0,9645	1,1391	1,0019
1000	1000	-0,0104	0,0004	-0,0562	-0,0005	0,9311	1,0481	0,9315	1,0485	1,0005
1000	2000	-0,0544	0,0000	-0,0916	-0,0010	0,8893	1,0019	0,8894	1,0018	1,0011
1000	5000	0,0731	0,0003	0,0881	-0,0007	1,0057	1,1405	1,0060	1,1408	1,0006
2000	1000	-0,0148	0,0001	-0,0137	0,0004	0,9420	1,0284	0,9421	1,0285	0,9996
2000	2000	0,0028	-0,0003	0,0081	-0,0007	0,9586	1,0470	0,9583	1,0467	1,0007
2000	5000	-0,0411	-0,0004	-0,0099	-0,0005	0,9155	1,0023	0,9151	1,0019	1,0005
5000	1000	-0,0072	0,0003	-0,0070	0,0007	0,9653	1,0203	0,9656	1,0206	0,9993
5000	2000	-0,0279	0,0003	-0,0279	-0,0005	0,9452	0,9990	0,9455	0,9993	1,0005
5000	5000	0,0023	0,0001	0,0138	-0,0002	0,9742	1,0304	0,9743	1,0305	1,0002

X_i	$\sim$	Exp(1)		W_i	$\sim$	P(1)		conf. int	90%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	0,0296	-0,0031	-0,0071	-0,0032	0,9141	1,1451	0,9114	1,1416	1,0032
500	500	-0,0105	-0,0013	-0,0155	-0,0042	0,9171	1,0619	0,9161	1,0603	1,0043
1000	1000	-0,0003	0,0019	0,0230	0,0009	0,9465	1,0529	0,9483	1,0549	0,9991
1000	2000	0,0607	0,0014	0,1265	-0,0004	1,0021	1,1193	1,0035	1,1207	1,0004
1000	5000	-0,0708	0,0006	-0,0460	-0,0004	0,8796	0,9788	0,8802	0,9794	1,0004
2000	1000	-0,0158	-0,0002	-0,0756	-0,0011	0,9502	1,0182	0,9500	1,0180	1,0012
2000	2000	-0,0118	-0,0004	-0,0048	-0,0004	0,9516	1,0248	0,9512	1,0244	1,0004
2000	5000	0,0393	-0,0001	0,0030	-0,0003	1,0024	1,0762	1,0023	1,0761	1,0003
5000	1000	0,0058	0,0002	-0,0219	0,0000	0,9830	1,0286	0,9832	1,0288	1,0000
5000	2000	-0,0134	0,0000	-0,0086	-0,0001	0,9635	1,0097	0,9635	1,0097	1,0001
5000	5000	0,0085	-0,0003	0,0367	-0,0004	0,9844	1,0326	0,9841	1,0323	1,0004

X_i	$\sim$	Exp(1)		W_i	$\sim$	P(1)		conf. int	95%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,0700	0,0032	-0,0681	-0,0005	0,8008	1,0592	0,8041	1,0623	1,0005
500	500	-0,0117	0,0001	-0,0155	-0,0026	0,9020	1,0746	0,9023	1,0745	1,0026
1000	1000	0,0494	0,0020	0,0659	0,0041	0,9833	1,1155	0,9851	1,1177	0,9962
1000	2000	0,0307	0,0005	0,0369	-0,0006	0,9664	1,0950	0,9670	1,0954	1,0006
1000	5000	-0,0068	0,0003	0,0010	-0,0003	0,9312	1,0552	0,9315	1,0555	1,0003
2000	1000	0,0133	-0,0001	0,0423	-0,0008	0,9676	1,0590	0,9676	1,0588	1,0008
2000	2000	0,0204	-0,0004	0,0093	-0,0008	0,9762	1,0646	0,9758	1,0642	1,0008
2000	5000	-0,0320	-0,0002	-0,0466	-0,0003	0,9262	1,0098	0,9260	1,0096	1,0003
5000	1000	0,0099	0,0003	0,0224	-0,0002	0,9816	1,0382	0,9819	1,0385	1,0002
5000	2000	0,0122	0,0003	-0,0150	0,0002	0,9849	1,0395	0,9852	1,0398	0,9998
5000	5000	0,0040	-0,0002	-0,0276	-0,0003	0,9770	1,0310	0,9769	1,0307	1,0003

X_i	$\sim$	N(1,1)		W_i	$\sim$	Exp(1)		conf. int	90%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,1069	0,0038	-0,0468	0,0003	0,7822	1,0040	0,7860	1,0078	0,9997
500	500	-0,0803	-0,0011	0,0194	0,0000	0,8447	0,9947	0,8436	0,9936	1,0000
1000	1000	-0,0481	-0,0003	-0,0214	0,0004	0,9010	1,0028	0,9007	1,0025	0,9996
1000	2000	-0,0149	0,0006	0,0337	-0,0007	0,9313	1,0389	0,9320	1,0394	1,0007
1000	5000	0,0007	-0,0003	0,0584	-0,0009	0,9456	1,0558	0,9454	1,0554	1,0009
2000	1000	0,0123	-0,0005	0,0177	-0,0004	0,9749	1,0497	0,9744	1,0492	1,0004
2000	2000	-0,0234	0,0006	0,0197	-0,0004	0,9391	1,0141	0,9397	1,0147	1,0004
2000	5000	0,0326	0,0000	0,0384	-0,0001	0,9944	1,0708	0,9944	1,0708	1,0001
5000	1000	0,0168	-0,0004	0,0345	-0,0003	0,9927	1,0409	0,9923	1,0405	1,0003
5000	2000	0,0192	-0,0006	0,0400	-0,0001	0,9950	1,0434	0,9944	1,0428	1,0001
5000	5000	0,0091	-0,0001	0,0394	-0,0002	0,9849	1,0333	0,9848	1,0332	1,0002

X_i	$\sim$	N(1,1)		W_i	$\sim$	Exp(1)		conf. int	95%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	0,0503	-0,0058	0,0130	-0,0087	0,9099	1,1907	0,9053	1,1837	1,0087
500	500	0,0040	-0,0015	-0,0725	-0,0013	0,9227	1,0853	0,9213	1,0837	1,0014
1000	1000	0,0695	-0,0003	0,0088	-0,0003	1,0070	1,1320	1,0067	1,1317	1,0003
1000	2000	0,0199	0,0005	0,0429	-0,0004	0,9553	1,0845	0,9558	1,0850	1,0004
1000	5000	-0,0376	-0,0004	0,0162	-0,0004	0,8994	1,0254	0,8990	1,0250	1,0004
2000	1000	0,0250	0,0015	0,0088	-0,0004	0,9808	1,0692	0,9823	1,0707	1,0004
2000	2000	-0,0250	0,0004	0,0086	0,0002	0,9308	1,0192	0,9312	1,0196	0,9998
2000	5000	0,0031	-0,0001	0,0487	-0,0005	0,9571	1,0491	0,9571	1,0489	1,0005
5000	1000	-0,0017	0,0009	0,0516	0,0000	0,9692	1,0274	0,9701	1,0283	1,0000
5000	2000	-0,0204	0,0002	0,0488	-0,0003	0,9505	1,0087	0,9507	1,0089	1,0003
5000	5000	0,0056	0,0002	0,0523	-0,0001	0,9764	1,0348	0,9766	1,0350	1,0001

X_i	$\sim$	N(1,1)		W_i	$\sim$	P(1)		conf. int	90%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,0503	0,0043	-0,0706	-0,0031	0,8416	1,0578	0,8463	1,0617	1,0033
500	500	-0,0403	-0,0007	-0,0027	-0,0002	0,8863	1,0331	0,8856	1,0324	1,0002
1000	1000	0,0495	-0,0004	0,0148	-0,0004	0,9967	1,1023	0,9963	1,1019	1,0004
1000	2000	-0,0149	-0,0007	0,0295	-0,0002	0,9315	1,0387	0,9309	1,0379	1,0002
1000	5000	0,0308	0,0005	-0,0524	-0,0006	0,9815	1,0801	0,9820	1,0806	1,0006
2000	1000	-0,0065	0,0006	-0,0530	-0,0002	0,9587	1,0283	0,9593	1,0289	1,0002
2000	2000	0,0131	-0,0005	-0,0612	0,0000	0,9786	1,0476	0,9781	1,0471	1,0000
2000	5000	0,0031	-0,0001	-0,0835	0,0000	0,9694	1,0368	0,9693	1,0367	1,0000
5000	1000	-0,0162	0,0001	-0,0739	-0,0001	0,9623	1,0053	0,9624	1,0054	1,0001
5000	2000	-0,0202	0,0001	-0,0646	0,0000	0,9580	1,0016	0,9581	1,0017	1,0000
5000	5000	0,0106	0,0000	-0,0826	0,0001	0,9893	1,0319	0,9893	1,0319	0,9999

X_i	$\sim$	N(1,1)		W_i	$\sim$	P(1)		conf. int	95%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,0502	0,0039	0,0231	-0,0027	0,8080	1,0916	0,8123	1,0951	1,0026
500	500	0,0357	0,0012	-0,0234	-0,0007	0,9501	1,1213	0,9514	1,1224	1,0007
1000	1000	0,0196	-0,0005	0,0317	-0,0007	0,9557	1,0835	0,9552	1,0830	1,0007
1000	2000	-0,0243	0,0005	-0,0230	0,0000	0,9151	1,0363	0,9156	1,0368	1,0000
1000	5000	-0,0037	0,0005	-0,0534	-0,0008	0,9376	1,0550	0,9382	1,0554	1,0008
2000	1000	-0,0115	-0,0015	-0,0477	0,0002	0,9468	1,0302	0,9453	1,0287	0,9998
2000	2000	-0,0059	0,0001	-0,0764	0,0001	0,9536	1,0346	0,9537	1,0347	0,9999
2000	5000	0,0185	0,0000	-0,0716	-0,0003	0,9778	1,0592	0,9778	1,0592	1,0003
5000	1000	0,0074	0,0002	-0,0727	-0,0004	0,9817	1,0331	0,9819	1,0333	1,0004
5000	2000	0,0040	-0,0002	-0,0695	0,0000	0,9782	1,0298	0,9780	1,0296	1,0000
5000	5000	0,0035	-0,0004	-0,0891	-0,0001	0,9783	1,0287	0,9779	1,0283	1,0001

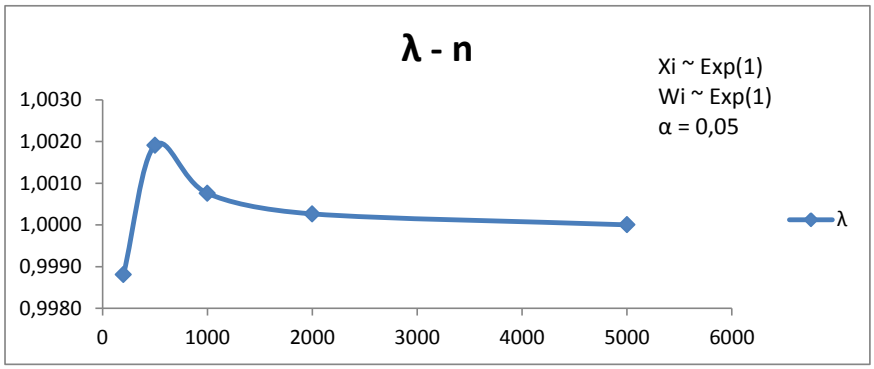
X_i	$\sim$	U(0,1)		W_i	$\sim$	Exp(1)		conf. int	90%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,0327	-0,0013	-0,0032	-0,0012	0,4341	0,5005	0,4329	0,4991	1,0042
500	500	0,0179	0,0007	-0,0012	0,0003	0,4967	0,5391	0,4974	0,5398	0,9990
1000	1000	-0,0064	-0,0002	0,0072	-0,0001	0,4782	0,5090	0,4780	0,5088	1,0003
1000	2000	-0,0043	-0,0003	0,0063	-0,0001	0,4804	0,5110	0,4801	0,5107	1,0003
1000	5000	-0,0029	-0,0002	0,0006	-0,0001	0,4821	0,5121	0,4819	0,5119	1,0003
2000	1000	-0,0087	0,0000	-0,0002	0,0000	0,4807	0,5019	0,4807	0,5019	1,0000
2000	2000	0,0001	0,0003	0,0003	-0,0001	0,4895	0,5107	0,4898	0,5110	1,0003
2000	5000	-0,0015	0,0002	0,0033	0,0000	0,4878	0,5092	0,4880	0,5094	1,0000
5000	1000	0,0029	0,0000	0,0010	-0,0001	0,4962	0,5096	0,4962	0,5096	1,0003
5000	2000	0,0042	-0,0001	0,0020	0,0001	0,4974	0,5110	0,4973	0,5109	0,9997
5000	5000	0,0004	0,0001	0,0020	0,0000	0,4936	0,5072	0,4937	0,5073	1,0000

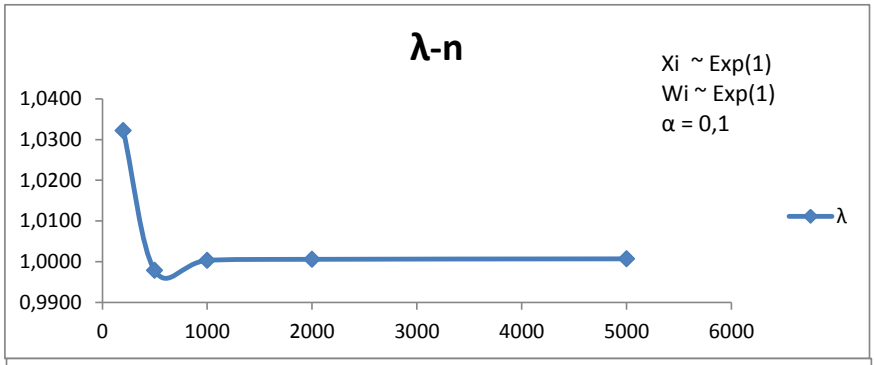
X_i	$\sim$	U(0,1)		W_i	$\sim$	Exp(1)		conf. int	95%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	-0,0333	0,0009	-0,0108	-0,0008	0,4282	0,5052	0,4292	0,5060	1,0029
500	500	-0,0131	-0,0004	-0,0030	-0,0001	0,4619	0,5119	0,4615	0,5115	1,0004
1000	1000	-0,0199	0,0000	0,0044	0,0000	0,4619	0,4983	0,4619	0,4983	1,0000
1000	2000	-0,0020	-0,0001	0,0046	-0,0002	0,4798	0,5162	0,4797	0,5161	1,0007
1000	5000	0,0002	0,0001	0,0081	-0,0002	0,4818	0,5186	0,4819	0,5187	1,0007
2000	1000	-0,0037	0,0004	-0,0027	-0,0001	0,4838	0,5088	0,4842	0,5092	1,0003
2000	2000	-0,0001	0,0001	-0,0017	-0,0002	0,4873	0,5125	0,4874	0,5126	1,0007
2000	5000	0,0038	0,0000	0,0004	-0,0001	0,4911	0,5165	0,4911	0,5165	1,0003
5000	1000	0,0062	0,0001	-0,0002	-0,0001	0,4982	0,5142	0,4983	0,5143	1,0003
5000	2000	-0,0008	0,0002	-0,0003	-0,0001	0,4912	0,5072	0,4914	0,5074	1,0003
5000	5000	0,0013	0,0001	0,0006	-0,0001	0,4933	0,5093	0,4934	0,5094	1,0003

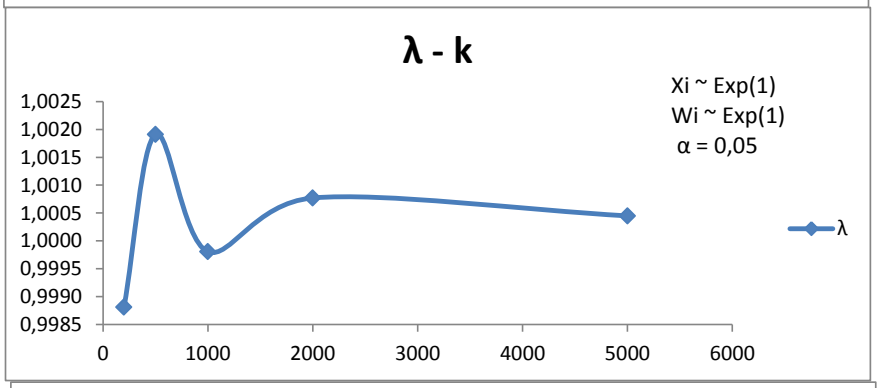
X_i	$\sim$	U(0,1)		W_i	$\sim$	P(1)		conf. int	90%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	0,0003	0,0000	0,0003	-0,0006	0,4667	0,5339	0,4668	0,5338	1,0021
500	500	-0,0061	0,0004	0,0119	-0,0005	0,4718	0,5160	0,4722	0,5164	1,0017
1000	1000	0,0012	0,0002	-0,0004	-0,0003	0,4862	0,5162	0,4864	0,5164	1,0010
1000	2000	-0,0018	0,0002	0,0000	-0,0002	0,4832	0,5132	0,4834	0,5134	1,0007
1000	5000	0,0065	0,0001	-0,0009	-0,0001	0,4915	0,5215	0,4916	0,5216	1,0003
2000	1000	0,0014	0,0002	-0,0001	0,0001	0,4908	0,5120	0,4910	0,5122	0,9997
2000	2000	-0,0057	0,0001	0,0014	-0,0001	0,4836	0,5050	0,4837	0,5051	1,0003
2000	5000	0,0017	0,0000	-0,0005	0,0000	0,4911	0,5123	0,4911	0,5123	1,0000
5000	1000	-0,0029	0,0000	0,0035	-0,0001	0,4903	0,5039	0,4903	0,5039	1,0003
5000	2000	0,0004	-0,0001	-0,0014	0,0000	0,4937	0,5071	0,4936	0,5070	1,0000
5000	5000	0,0011	0,0000	0,0004	0,0000	0,4944	0,5078	0,4944	0,5078	1,0000

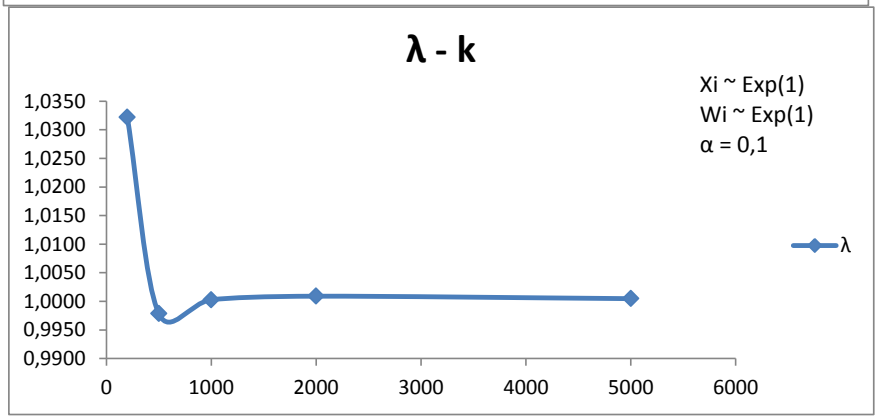
X_i	$\sim$	U(0,1)		W_i	$\sim$	P(1)		conf. int	95%	
n	k	y_1^*	y_2^*	y_3^*	y_4^*	CIL^*	CIR^*	CIL_B^*	CIR_B^*	λ
200	200	0,0139	-0,0016	0,0064	0,0003	0,4730	0,5548	0,4714	0,5532	0,9990
500	500	-0,0137	-0,0004	0,0006	-0,0004	0,4609	0,5117	0,4606	0,5112	1,0014
1000	1000	0,0065	0,0003	-0,0008	-0,0002	0,4887	0,5243	0,4890	0,5246	1,0007
1000	2000	0,0083	0,0001	-0,0030	-0,0003	0,4906	0,5260	0,4907	0,5261	1,0011
1000	5000	0,0064	0,0003	0,0057	-0,0003	0,4882	0,5246	0,4885	0,5249	1,0010
2000	1000	-0,0149	-0,0003	0,0004	-0,0003	0,4724	0,4978	0,4721	0,4975	1,0010
2000	2000	-0,0050	-0,0001	0,0021	0,0000	0,4823	0,5077	0,4822	0,5076	1,0000
2000	5000	-0,0035	0,0001	-0,0018	0,0000	0,4839	0,5091	0,4840	0,5092	1,0000
5000	1000	-0,0069	-0,0002	-0,0013	-0,0001	0,4851	0,5011	0,4849	0,5009	1,0003
5000	2000	-0,0007	0,0000	-0,0006	0,0000	0,4913	0,5073	0,4913	0,5073	1,0000
5000	5000	-0,0040	0,0001	0,0007	0,0000	0,4880	0,5040	0,4881	0,5041	1,0000

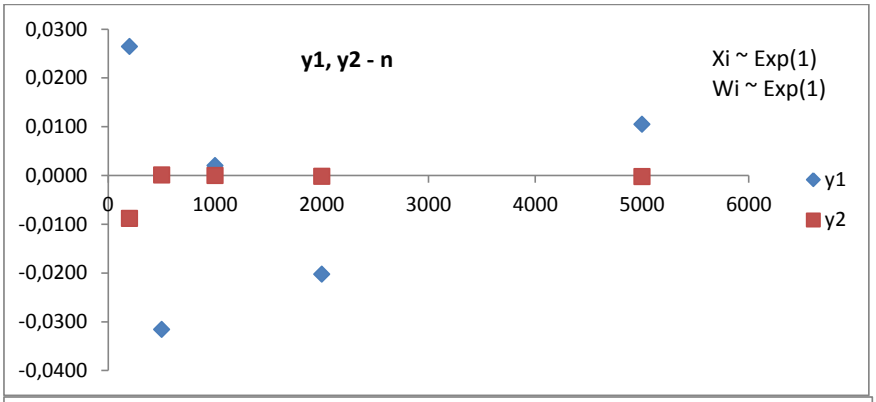
*CIL : left margin for confidence interval - classical
*CIR : right margin for confidence interval - classical
*CIL_B : left margin for confidence interval - bootstrap
*CIR_B : right margin for confidence interval - bootstrap
*y_1 : error for mean value-classical
*y_2 : error for variance-classical
*y_3 : error for mean value-bootstrap
*y_4 : error for variance-bootstrap

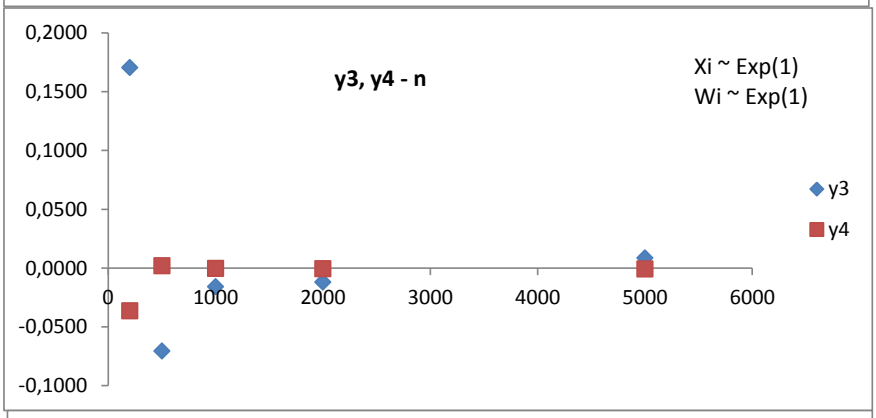

λ - n
Xi ~ Exp(1)
Wi ~ Exp(1)
α = 0,05
1,0030
1,0020
1,0010
1,0000
0,9990
0,9980
0
1000
2000
3000
4000
5000
6000
λ

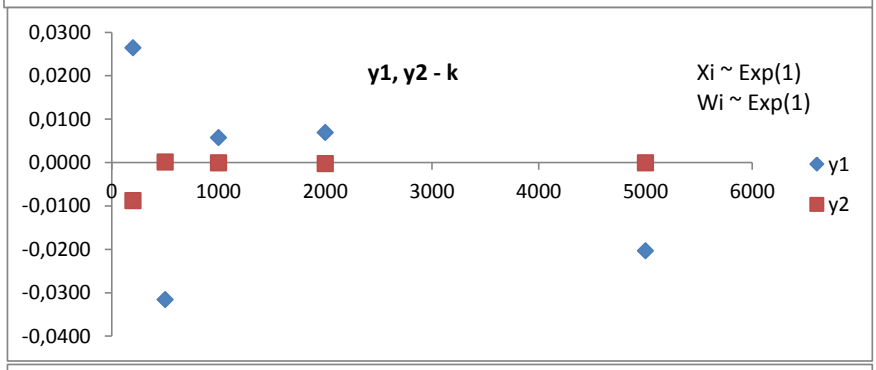

λ-n
Xi ~ Exp(1)
Wi ~ Exp(1)
α = 0,1
1,0400
1,0300
1,0200
1,0100
1,0000
0,9900
0
1000
2000
3000
4000
5000
6000
λ

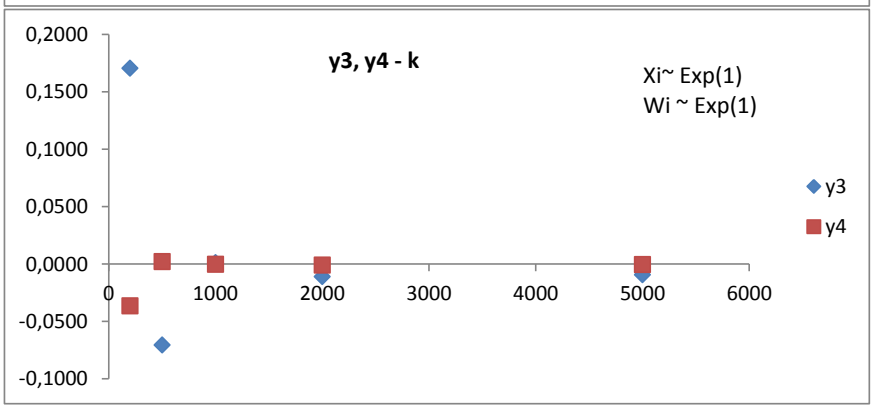

λ - k
Xi ~ Exp(1)
Wi ~ Exp(1)
α = 0,05
1,0025
1,0020
1,0015
1,0010
1,0005
1,0000
0,9995
0,9990
0,9985
0
1000
2000
3000
4000
5000
6000
λ


λ - k
Xi ~ Exp(1)
Wi ~ Exp(1)
α = 0,1
1,0350
1,0300
1,0250
1,0200
1,0150
1,0100
1,0050
1,0000
0,9950
0,9900
0
1000
2000
3000
4000
5000
6000
λ


y1, y2 - n
Xi ~ Exp(1)
Wi ~ Exp(1)
y1
y2
0,0300
0,0200
0,0100
0,0000
-0,0100
-0,0200
-0,0300
-0,0400
0
1000
2000
3000
4000
5000
6000


y3, y4 - n
Xi ~ Exp(1)
Wi ~ Exp(1)
y3
y4
0,2000
0,1500
0,1000
0,0500
0,0000
-0,0500
-0,1000
0
1000
2000
3000
4000
5000
6000


y1, y2 - k
Xi ~ Exp(1)
Wi ~ Exp(1)
y1
y2
0,0300
0,0200
0,0100
0,0000
-0,0100
-0,0200
-0,0300
-0,0400
0
1000
2000
3000
4000
5000
6000


y3, y4 - k
Xi~ Exp(1)
Wi ~ Exp(1)
y3
y4
0,2000
0,1500
0,1000
0,0500
0,0000
-0,0500
-0,1000
0
1000
2000
3000
4000
5000
6000

From the preceding tables of results, for all distributions of random variables or bootstrap weights considered here and also for both percentile points used, follows that, as we increase the sample size ($n \uparrow$) and the number of iterations ($k \uparrow$), errors tend to be smaller and margins of confidence intervals get generally smaller, mainly for great increase of sample size. Remarking the results of each line, we note that in most cases bootstrap method gives better results, which is, in our experiments, smaller errors and narrower confidence intervals.

Calculation of quantiles

The purpose of this part is to calculate the quantiles with the use of the empirical function. Based on the formula (2.2.2) we have to form the quantity

$$Q_i = \frac{\sqrt{n}\cdot(\overline{X}^i_{\mathcal{W},n} - \overline{X}_n)}{S_n}$$

for every set of bootstrap weights. The values of the sample mean and sample variance are calculated as described in chapter 3. When the value of Q_i is calculated for all bootstrap iterations, the empirical function (2.3.1) is used so as to get an estimation of the quantile.
In our experiment, the sample follows normal random distribution: $\mathcal{X}_i \sim \mathcal{N}(1,1)$. Since $\alpha = 0.1$ (central interval of 90%) empirical function is set equal to 0.95, so theoretical $z_{1-\alpha/2} = 1.645$.

The experiment is conducted

- for Efron bootstrap and weighted bootstrap, with the cases of weights following the exponential law ($\mathcal{W}_{n,i} \sim Exp(1)$) or the Poisson law ($\mathcal{W}_{n,i} \sim P(1)$),
- for sample size $n = 1000$, $n = 2000$ and $n = 5000$
- for a number of bootstrap iterations: k =1000, 2000, 5000

  and

- for two values of variance, either the estimated variance (S_n) or the theoretical variance (σ) of the sample.

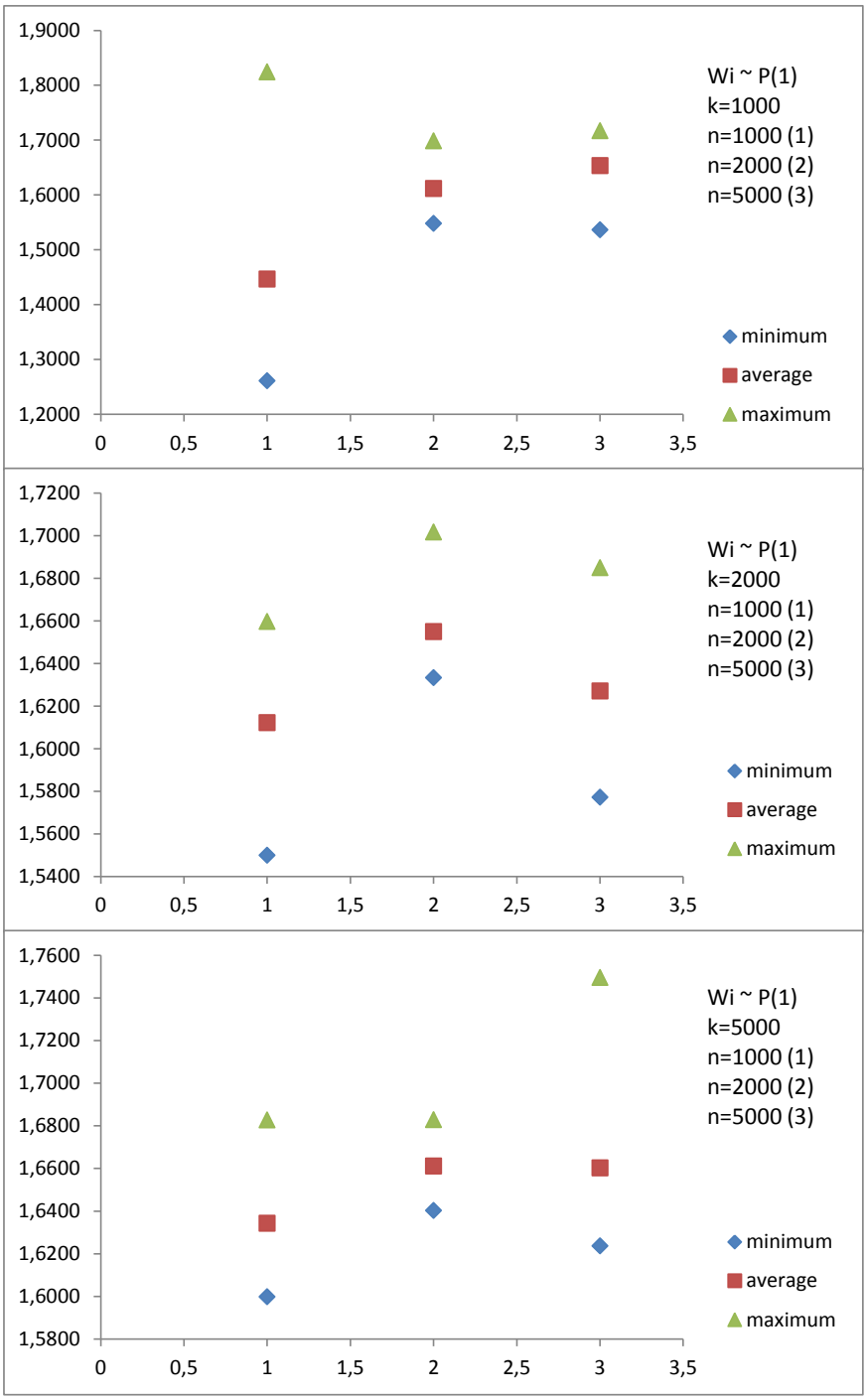
Each case of the experiment is conducted 5 times and for each case we calculated the average, the minimum and the maximum value. Based on these values, some graphs were made. The results and graphs are:

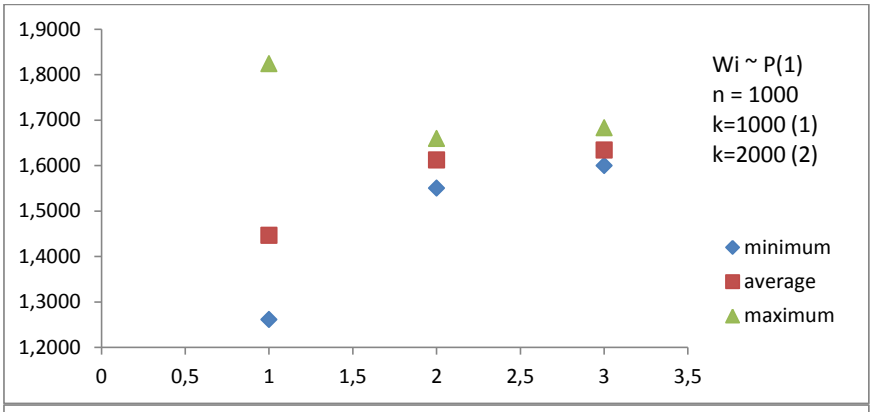
Weighted	Bootstrap	$W_i \sim P(1)$			with	estim.	variance		
	n=1000	n=1000	n=1000	n=2000	n=2000	n=2000	n=5000	n=5000	n=5000
	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000
	1,4041	1,5895	1,6827	1,5783	1,6384	1,6829	1,6468	1,6372	1,7495
	1,8242	1,5500	1,6290	1,6984	1,6333	1,6483	1,7103	1,5908	1,6237
	1,4534	1,6315	1,6350	1,6127	1,6374	1,6799	1,6560	1,5772	1,6320
	1,2611	1,6596	1,6247	1,6210	1,6637	1,6403	1,5363	1,6452	1,6396
	1,2889	1,6300	1,5998	1,5478	1,7016	1,6545	1,7169	1,6849	1,6565
minimum	1,2611	1,5500	1,5998	1,5478	1,6333	1,6403	1,5363	1,5772	1,6237
average	1,4463	1,6121	1,6342	1,6116	1,6549	1,6612	1,6533	1,6271	1,6603
maximum	1,8242	1,6596	1,6827	1,6984	1,7016	1,6829	1,7169	1,6849	1,7495

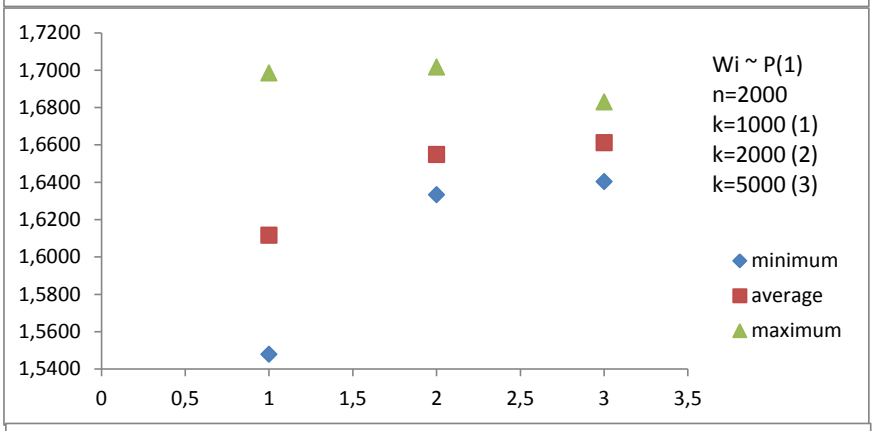
Weighted	Bootstrap	$W_i \sim P(1)$			with	theor.	variance		
	n=1000	n=1000	n=1000	n=2000	n=2000	n=2000	n=5000	n=5000	n=5000
	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000
	1,6704	1,6485	1,6219	1,5028	1,6132	1,6027	1,5925	1,6779	1,6196
	1,3025	1,5501	1,6484	1,6054	1,5869	1,6364	1,4909	1,5756	1,7054
	1,4286	1,5798	1,6402	1,6472	1,5637	1,5927	1,6970	1,5936	1,6232
	1,2926	1,6635	1,6613	1,5711	1,6922	1,5845	1,6384	1,7124	1,5994
	1,2882	1,6084	1,6155	1,5708	1,7221	1,6157	1,5246	1,5866	1,5913
minimum	1,2882	1,5501	1,6155	1,5028	1,5637	1,5845	1,4909	1,5756	1,5913
average	1,3965	1,6101	1,6375	1,5795	1,6356	1,6064	1,5887	1,6292	1,6278
maximum	1,6704	1,6635	1,6613	1,6472	1,7221	1,6364	1,6970	1,7124	1,7054

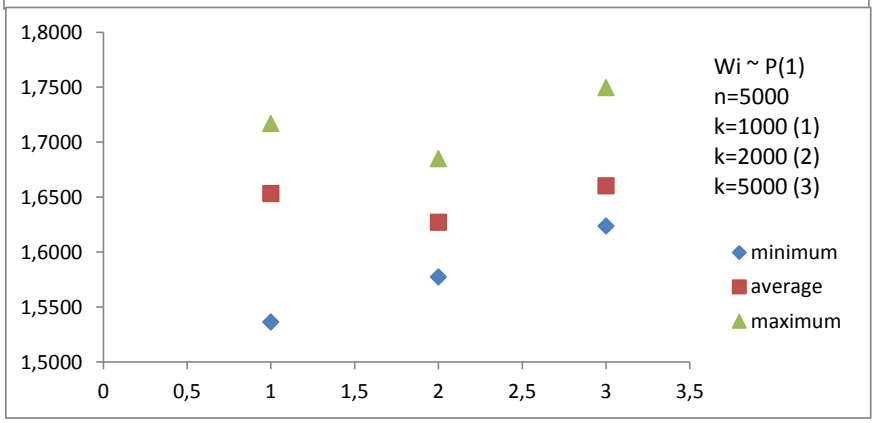
Weighted	Bootstrap	$W_i \sim Exp(1)$			with	estim.	variance		
	n=1000	n=1000	n=1000	n=2000	n=2000	n=2000	n=5000	n=5000	n=5000
	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000
	1,4849	1,6688	1,6365	1,6050	1,6913	1,6399	1,5751	1,6671	1,6677
	1,4979	1,7453	1,6338	1,6500	1,6825	1,6328	1,5726	1,6720	1,6764
	1,2580	1,6172	1,6991	1,5693	1,6204	1,6280	1,7647	1,5946	1,6165
	1,7499	1,6465	1,6555	1,7064	1,6719	1,6304	1,6938	1,6031	1,6338
	1,4357	1,6116	1,6380	1,6944	1,6442	1,5846	1,7072	1,5620	1,6518
minimum	1,2580	1,6116	1,6338	1,5693	1,6204	1,5846	1,5726	1,5620	1,6165
average	1,4853	1,6579	1,6526	1,6450	1,6621	1,6231	1,6627	1,6198	1,6492
maximum	1,7499	1,7453	1,6991	1,7064	1,6913	1,6399	1,7647	1,6720	1,6764

Weighted	Bootstrap	$W_i \sim Exp(1)$			with	theor.	variance		
	n=1000	n=1000	n=1000	n=2000	n=2000	n=2000	n=5000	n=5000	n=5000
	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000
	1,2639	1,6109	1,6066	1,6969	1,5535	1,6197	1,6155	1,6672	1,6349
	1,3980	1,6749	1,6253	1,6126	1,6093	1,6246	1,5663	1,4992	1,6232
	1,3562	1,5574	1,6303	1,5834	1,6104	1,6232	1,7032	1,6406	1,5901
	1,4021	1,6172	1,6409	1,7451	1,6144	1,6577	1,6308	1,6556	1,5675
	1,2327	1,6671	1,6232	1,6532	1,6269	1,5625	1,6049	1,6770	1,6292
minimum	1,2327	1,5574	1,6066	1,5834	1,5535	1,5625	1,5663	1,4992	1,5675
average	1,3306	1,6255	1,6253	1,6582	1,6029	1,6175	1,6241	1,6279	1,6090
maximum	1,4021	1,6749	1,6409	1,7451	1,6269	1,6577	1,7032	1,6770	1,6349


1,9000
1,8000
1,7000
1,6000
1,5000
1,4000
1,3000
1,2000
0
0,5
1
1,5
2
2,5
3
3,5
Wi ~ P(1)
k=1000
n=1000 (1)
n=2000 (2)
n=5000 (3)
minimum
average
maximum
1,7200
1,7000
1,6800
1,6600
1,6400
1,6200
1,6000
1,5800
1,5600
1,5400
0
0,5
1
1,5
2
2,5
3
3,5
Wi ~ P(1)
k=2000
n=1000 (1)
n=2000 (2)
n=5000 (3)
minimum
average
maximum
1,7600
1,7400
1,7200
1,7000
1,6800
1,6600
1,6400
1,6200
1,6000
1,5800
0
0,5
1
1,5
2
2,5
3
3,5
Wi ~ P(1)
k=5000
n=1000 (1)
n=2000 (2)
n=5000 (3)
minimum
average
maximum


1,9000
1,8000
1,7000
1,6000
1,5000
1,4000
1,3000
1,2000
0
0,5
1
1,5
2
2,5
3
3,5
Wi ~ P(1)
n = 1000
k=1000 (1)
k=2000 (2)
minimum
average
maximum


1,7200
1,7000
1,6800
1,6600
1,6400
1,6200
1,6000
1,5800
1,5600
1,5400
0
0,5
1
1,5
2
2,5
3
3,5
Wi ~ P(1)
n=2000
k=1000 (1)
k=2000 (2)
k=5000 (3)
minimum
average
maximum


1,8000
1,7500
1,7000
1,6500
1,6000
1,5500
1,5000
0
0,5
1
1,5
2
2,5
3
3,5
Wi ~ P(1)
n=5000
k=1000 (1)
k=2000 (2)
k=5000 (3)
minimum
average
maximum

Efron	Bootstrap				with	estim.	variance		
	n=1000	n=1000	n=1000	n=2000	n=2000	n=2000	n=5000	n=5000	n=5000
	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000
	1,3674	1,6655	1,6209	1,6929	1,6728	1,6377	1,7112	1,6350	1,6555
	1,8528	1,7080	1,5866	1,7105	1,5715	1,6509	1,5952	1,5681	1,6321
	1,7157	1,6364	1,6374	1,6311	1,6550	1,5960	1,6472	1,6154	1,6022
	1,5680	1,5788	1,6473	1,6712	1,6624	1,5934	1,6030	1,6010	1,6250
	1,3322	1,6546	1,6410	1,6369	1,6926	1,6547	1,5925	1,6171	1,6138
minimum	1,3322	1,5788	1,5866	1,6311	1,5715	1,5934	1,5925	1,5681	1,6022
average	1,5672	1,6487	1,6266	1,6685	1,6509	1,6265	1,6298	1,6073	1,6257
maximum	1,8528	1,7080	1,6473	1,7105	1,6926	1,6547	1,7112	1,6350	1,6555

Efron	Bootstrap				with	theor	variance		
	n=1000	n=1000	n=1000	n=2000	n=2000	n=2000	n=5000	n=5000	n=5000
	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000	k=1000	k=2000	k=5000
	1,2101	1,5988	1,6328	1,5740	1,6565	1,5798	1,5193	1,6886	1,5891
	1,5269	1,7058	1,6916	1,5671	1,5977	1,6436	1,6094	1,6457	1,6497
	1,2956	1,6090	1,5540	1,6288	1,5905	1,6003	1,6035	1,5441	1,6407
	1,2935	1,6571	1,5889	1,6757	1,5981	1,6051	1,6860	1,5953	1,6117
	2,0408	1,6387	1,6368	1,5774	1,6271	1,6621	1,7112	1,6181	1,5600
minimum	1,2101	1,5988	1,5540	1,5671	1,5905	1,5798	1,5193	1,5441	1,5600
average	1,4734	1,6419	1,6208	1,6046	1,6140	1,6182	1,6259	1,6184	1,6102
maximum	2,0408	1,7058	1,6916	1,6757	1,6565	1,6621	1,7112	1,6886	1,6497

Here we have the results of estimated quantiles for approximate central interval of 90%. The experiment was conducted for different bootstrap methods and combinations of sample size-bootstrap iterations. The results are close to the value $z = 1,645$, which corresponds to central interval of 90%. Also, they tend to be closer to this theoretical value as sample size and number of bootstrap iterations grow. Since we repeat each experiment for 5 times, this tendency becomes more clear.

4 Application - Markov Chains

The purpose here is to estimate the variance, the quantiles and the confidence intervals in the Markov chain setting (and not i.i.d. variables). The steps followed are:

1. First, it is necessary to produce the trajectory of a Markov Chain.
2. As second follows the calculation of the theoretical variance.
3. Then, the variance is estimated.
4. Using the method of weighted bootstrap, with weights following the law of Poisson distribution and Exponential distribution, we get another estimator of the variance.
5. After the estimation of variance, we estimate the quantiles.
6. The calculation of confidence intervals constitutes the last part.

To be more specific:

1. So as to produce the trajectory of a Markov chain, we use the transition probability matrix $\mathbf{P} = (P_{ij}), i, j \in E$ where $E = \{1, 2, 3, 4\}$:

$$\mathbf{P} = \begin{pmatrix} 0,5 & 0,20 & 0,20 & 0,10 \\ 0,25 & 0,25 & 0,25 & 0,25 \\ 0,3 & 0,25 & 0,20 & 0,25 \\ 0,15 & 0,35 & 0,25 & 0,25 \end{pmatrix}$$

First, we choose an initial value, $X_0 = 1$. Then, for as many times as until we get the total number of random variables that the MC should have, we generate a random independent uniform variable U, such that $U < \sum_{k=1}^{j} P_{i,k}$, where j will be the next state of the trajectory. In this way the random variables of MC are produced.

2. The theoretical variance is calculated according to Proposition 2.6.15, where the transition probability matrix $\mathbf{P}$ is used in any calculations needed.

3. For the estimation of variance, first we have to estimate the transition probability matrix (here we don't use the one given as data, we estimate it though):

$$\hat{P}_{ij}(n) = \frac{N_{ij}(n)}{N_i(n)}, \quad for \quad i \neq j$$

and

$$\hat{P}_{ij}(n) = 1 - \sum_{k=1}^{4} \hat{P}_{ik}(n), with \quad k \neq j, \quad for \quad i = j$$

where

$$N_{ij}(n) = \sum_{k=1}^{n} \mathbf{1}_{\{X_{k-1}=i, X_k=j\}} \quad \text{and} \quad N_i(n) = \sum_{k=1}^{n} \mathbf{1}_{\{X_{k-1}=i\}},$$

according to proposition (2.6.5).

The row vector ν consists of the elements $\nu(i) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k=i\}}$ and the stationary matrix is given by $\mathbf{\Pi} = \mathbf{1}^{\top} \nu$, where $\mathbf{1} = (1, \ldots, 1)$. According to propositions

(2.6.14) and (2.6.15), always for a fixed j, an estimator of the mean value ($\hat{\bar{g}}$) and an estimator of variance are calculated.

4. Here we estimate again the variance, but at this time, we use the method of weighted bootstrap. Therefore, for a number of iterations (N), we generate a sample of weights every time, $(W_1^{(l)}, \ldots, W_n^{(l)})$, for $l = 1, \ldots, N$ and calculate the following quantity, for each iteration:

$$\alpha_n^{(l)}(j) = \sqrt{n}(\tfrac{1}{n}\sum_{i=1}^n W_i^{(l)} \mathbf{1}_{\{X_i=j\}} - \tfrac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i=j\}})$$

Since it is known that it follows the normal distribution, we can calculate its variance, which is the estimator of the variance we are interested in.

5. In order to estimate the quantiles [15], we use the smallest z, such that

$$\tfrac{1}{N}\sum_{l=1}^N \mathbf{1}\{\psi_n^{(l)}(j) \leqslant z\} \geqslant 1-\alpha.$$

where

$$\psi_n^{(l)}(j) = |c^{-1}\alpha_n^{(l)}(j)|, \quad for \quad l = 1, \ldots, N.$$

Here z is a $(1-\alpha/2)\cdot 100\%$ percentile point of normal distribution for an approximate central interval of $1-\alpha$. In our case $c = 1$.

6. The basic formula for the calculation of confidence intervals is

$$ConfInterval = \bar{X} \pm \tfrac{\hat{\sigma}\cdot z^{(\alpha/2)}}{\sqrt{n}}.$$

The experiment described above was conducted:

- for central confidence interval of 95% and 97,5% and
- for the case of weighted bootstrap with exponential distribution $W_{i,n} \sim exp(1)$ and weighted bootstrap with Poisson distribution $W_{i,n} \sim P(1)$,
- for Markov chain of size n =1000, 2000, 5000 and
- for number of bootstrap iterations d =1000, 2000, 5000.

As done in the case of i.i.d. random variables, the ratio λ was calculated here as well. The results are presented in the following tables. Also, some graphs depicting the errors estimated as a function of n or k and λ as a function of n (averaged over d) or d (averaged over n) are presented here:

W_i	$\sim$	Exp(1)		$f_n = 0,95$		j=2			
n^*	1000	1000	1000	2000	2000	2000	5000	5000	5000
$d*$	1000	2000	5000	1000	2000	5000	1000	2000	5000
g_{th}	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213
$\bar{g}$	0,2600	0,2460	0,2500	0,2405	0,2550	0,2600	0,2494	0,2516	0,2436
σ_{th}	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447
$\hat{\sigma}$	0,4551	0,4429	0,4714	0,4392	0,4429	0,4419	0,4425	0,4341	0,4353
$\hat{\sigma}_W$	0,4281	0,4239	0,4367	0,4308	0,4271	0,4371	0,4366	0,4269	0,4296
z	0,8373	0,8095	0,8454	0,8599	0,8167	0,8534	0,8307	0,8299	0,8504
CL^*	0,2479	0,2347	0,2374	0,2321	0,2469	0,2516	0,2442	0,2465	0,2384
CR^*	0,2721	0,2573	0,2626	0,2489	0,2631	0,2684	0,2546	0,2567	0,2488
CL_B^*	0,2487	0,2351	0,2383	0,2322	0,2472	0,2517	0,2443	0,2466	0,2384
CR_B^*	0,2713	0,2569	0,2617	0,2488	0,2628	0,2683	0,2545	0,2566	0,2488
λ	1,0631	1,0448	1,0795	1,0195	1,0370	1,0110	1,0135	1,0169	1,0133

W_i	$\sim$	Exp(1)		$f_n = 0,975$		j=2			
n^*	1000	1000	1000	2000	2000	2000	5000	5000	5000
$d*$	1000	2000	5000	1000	2000	5000	1000	2000	5000
g_{th}	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213
$\bar{g}$	0,2430	0,2580	0,2550	0,2495	0,2355	0,2400	0,2612	0,2502	0,2514
σ_{th}	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447
$\hat{\sigma}$	0,4422	0,4455	0,4336	0,4251	0,4211	0,4311	0,4512	0,4434	0,4471
$\hat{\sigma}_W$	0,4344	0,4342	0,4379	0,4512	0,4158	0,4288	0,4517	0,4331	0,4364
z	0,9525	0,9550	0,9915	0,9838	0,9406	0,9529	0,9915	0,9617	0,9706
CL^*	0,2297	0,2445	0,2414	0,2401	0,2266	0,2308	0,2549	0,2442	0,2453
CR^*	0,2563	0,2715	0,2686	0,2589	0,2444	0,2492	0,2675	0,2562	0,2575
CL_B^*	0,2299	0,2449	0,2413	0,2396	0,2268	0,2309	0,2549	0,2443	0,2454
CR_B^*	0,2561	0,2711	0,2687	0,2594	0,2442	0,2491	0,2675	0,2561	0,2574
λ	1,0180	1,0260	0,9902	0,9422	1,0127	1,0054	0,9989	1,0238	1,0245

$*n$: the size of the Markov Chain

$*d$: the number of iterations of the bootstrap method

$*CL$: the left margin of the confidence interval, estimated as $CL = \bar{g} - \frac{\hat{\sigma} \cdot z}{\sqrt{n}}$

$*CR$: the right margin of the confidence interval, estimated as $CR = \bar{g} + \frac{\hat{\sigma} \cdot z}{\sqrt{n}}$

$*CL_B$: the left margin of the confidence interval, estimated as $CL_B = \bar{g} - \frac{\hat{\sigma}_W \cdot z}{\sqrt{n}}$

$*CR_B$: the right margin of the confidence interval, estimated as $CR_B = \bar{g} + \frac{\hat{\sigma}_W \cdot z}{\sqrt{n}}$

W_i	$\sim$	P(1)		$f_n = 0,95$		j=2			
n^*	1000	1000	1000	2000	2000	2000	5000	5000	5000
$d*$	1000	2000	5000	1000	2000	5000	1000	2000	5000
g_{th}	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213
$\bar{g}$	0,2840	0,2330	0,2660	0,2505	0,2600	0,2620	0,2618	0,2400	0,2478
σ_{th}	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447
$\hat{\sigma}$	0,4398	0,4194	0,4556	0,4494	0,4273	0,4277	0,4425	0,4207	0,4347
$\hat{\sigma}_W$	0,4526	0,4177	0,4491	0,4353	0,4389	0,4390	0,4355	0,4192	0,4328
z	0,8905	0,8086	0,8873	0,8908	0,8391	0,8619	0,8453	0,8133	0,8959
CL^*	0,2716	0,2223	0,2532	0,2415	0,2520	0,2538	0,2565	0,2352	0,2423
CR^*	0,2964	0,2437	0,2788	0,2595	0,2680	0,2702	0,2671	0,2448	0,2533
CL^*_B	0,2713	0,2223	0,2534	0,2418	0,2518	0,2535	0,2566	0,2352	0,2423
CR^*_B	0,2967	0,2437	0,2786	0,2592	0,2682	0,2705	0,2670	0,2448	0,2533
λ	0,9717	1,0041	1,0145	1,0324	0,9736	0,9743	1,0161	1,0036	1,0044

W_i	$\sim$	P(1)		$f_n = 0,975$		j=2			
n^*	1000	1000	1000	2000	2000	2000	5000	5000	5000
$d*$	1000	2000	5000	1000	2000	5000	1000	2000	5000
g_{th}	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213	0,3213
$\bar{g}$	0,2510	0,2610	0,2640	0,2465	0,2755	0,2835	0,2578	0,2478	0,2584
σ_{th}	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447	0,4447
$\hat{\sigma}$	0,4412	0,4343	0,4515	0,4339	0,4424	0,4592	0,4562	0,4440	0,4475
$\hat{\sigma}_W$	0,4320	0,4424	0,4417	0,4325	0,4498	0,4500	0,4402	0,4386	0,4365
z	0,9836	0,9860	0,9833	0,9733	1,0265	1,0002	0,9478	0,9801	0,9836
CL^*	0,2373	0,2475	0,2500	0,2371	0,2653	0,2732	0,2517	0,2416	0,2522
CR^*	0,2647	0,2745	0,2780	0,2559	0,2857	0,2938	0,2639	0,2540	0,2646
CL^*_B	0,2376	0,2472	0,2503	0,2371	0,2652	0,2734	0,2519	0,2417	0,2523
CR^*_B	0,2644	0,2748	0,2777	0,2559	0,2858	0,2936	0,2637	0,2539	0,2645
λ	1,0213	0,9817	1,0222	1,0032	0,9835	1,0204	1,0363	1,0123	1,0252

$*n$: the size of the Markov Chain

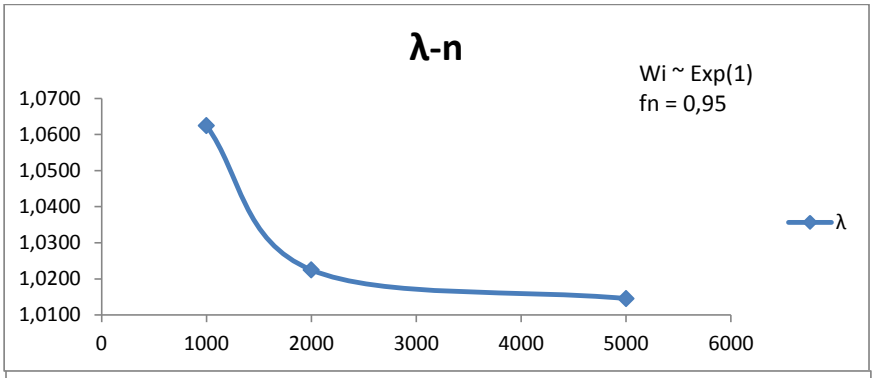
$*d$: the number of iterations of the bootstrap method

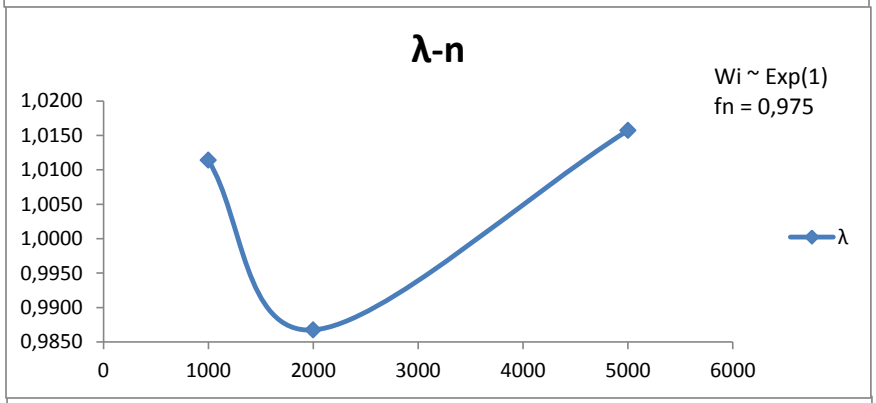
$*CL$: the left margin of the confidence interval, estimated as $CL = \bar{g} - \frac{\hat{\sigma} \cdot z}{\sqrt{n}}$

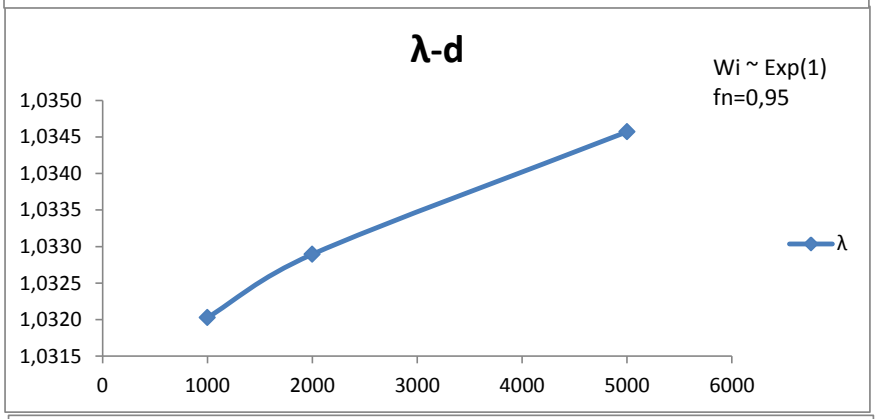
$*CR$: the right margin of the confidence interval, estimated as $CR = \bar{g} + \frac{\hat{\sigma} \cdot z}{\sqrt{n}}$

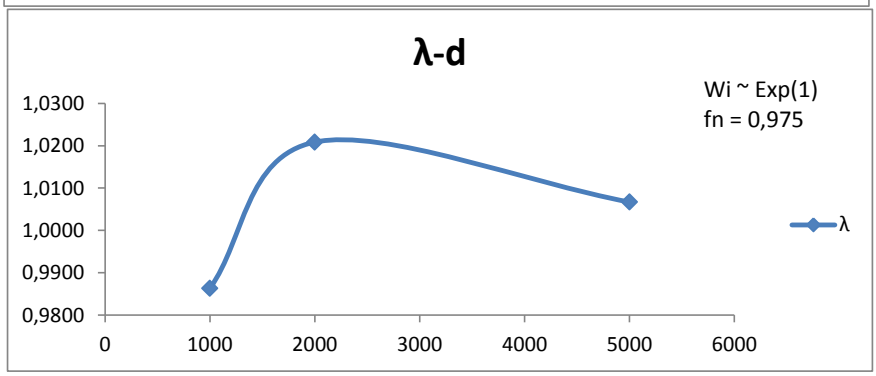
$*CL_B$: the left margin of the confidence interval, estimated as $CL_B = \bar{g} - \frac{\hat{\sigma}_W \cdot z}{\sqrt{n}}$

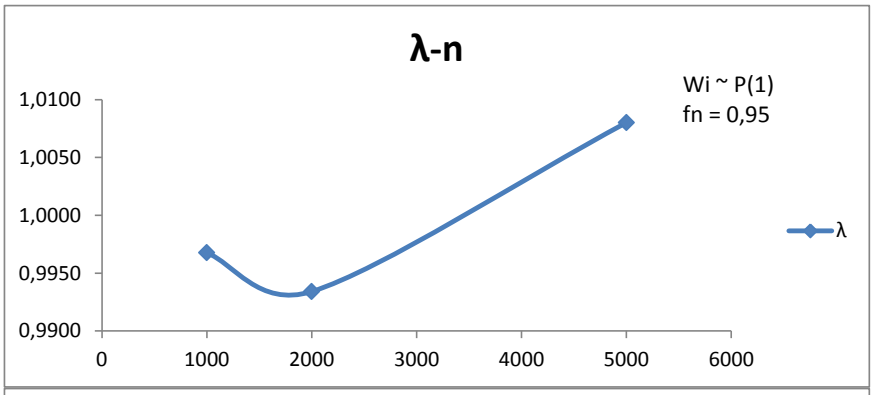
$*CR_B$: the right margin of the confidence interval, estimated as $CR_B = \bar{g} + \frac{\hat{\sigma}_W \cdot z}{\sqrt{n}}$

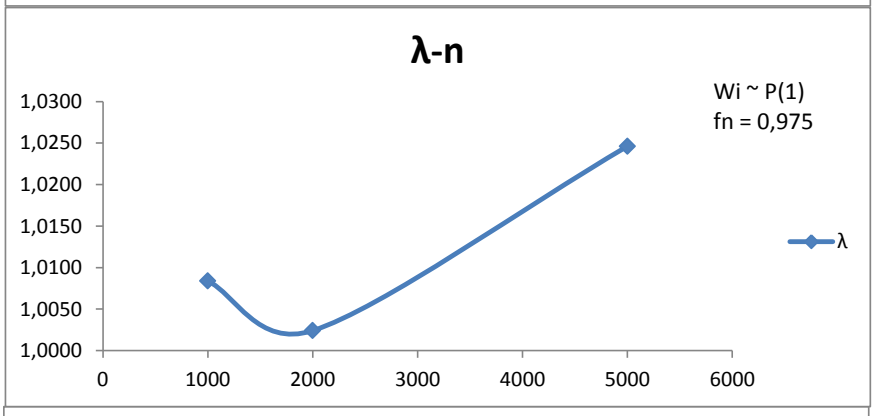

λ-n
Wi ~ Exp(1)
fn = 0,95
1,0700
1,0600
1,0500
1,0400
1,0300
1,0200
1,0100
0
1000
2000
3000
4000
5000
6000
λ

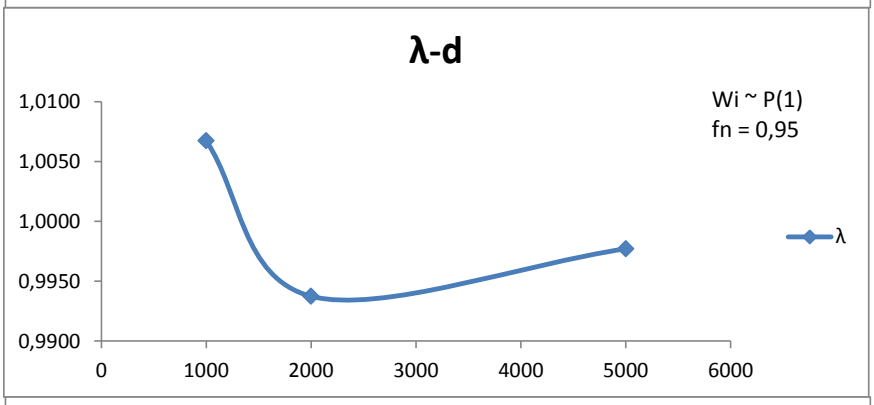

λ-n
Wi ~ Exp(1)
fn = 0,975
1,0200
1,0150
1,0100
1,0050
1,0000
0,9950
0,9900
0,9850
0
1000
2000
3000
4000
5000
6000
λ

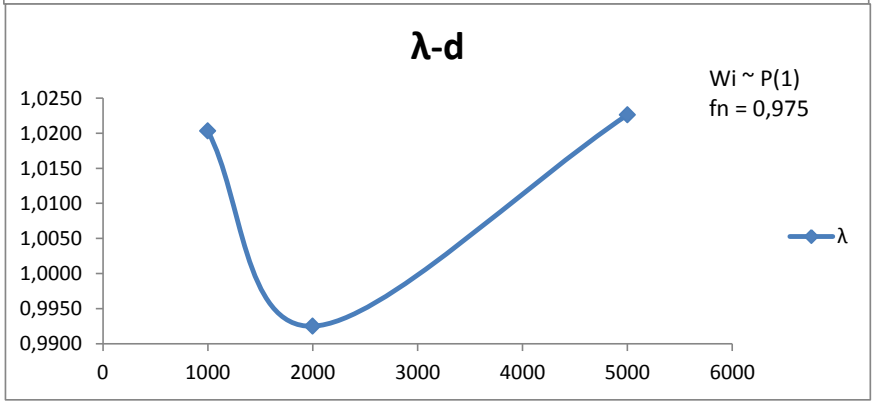

λ-d
Wi ~ Exp(1)
fn=0,95
1,0350
1,0345
1,0340
1,0335
1,0330
1,0325
1,0320
1,0315
0
1000
2000
3000
4000
5000
6000
λ


λ-d
Wi ~ Exp(1)
fn = 0,975
1,0300
1,0200
1,0100
1,0000
0,9900
0,9800
0
1000
2000
3000
4000
5000
6000
λ


λ-n
Wi ~ P(1)
fn = 0,95
1,0100
1,0050
1,0000
0,9950
0,9900
0
1000
2000
3000
4000
5000
6000
λ


λ-n
Wi ~ P(1)
fn = 0,975
1,0300
1,0250
1,0200
1,0150
1,0100
1,0050
1,0000
0
1000
2000
3000
4000
5000
6000
λ


λ-d
Wi ~ P(1)
fn = 0,95
1,0100
1,0050
1,0000
0,9950
0,9900
0
1000
2000
3000
4000
5000
6000
λ


λ-d
Wi ~ P(1)
fn = 0,975
1,0250
1,0200
1,0150
1,0100
1,0050
1,0000
0,9950
0,9900
0
1000
2000
3000
4000
5000
6000
λ

It is concluded here for another time that bootstrap methods give better estimations compared to classical estimation method. In most calculations done in this experiment, bootstrap variance is smaller than the other estimated variance, thus gives also narrower confidence intervals, which is graphically depicted with $\lambda > 1$. Also, we see that for random variables belonging to a Markov chain, the conclusions are similar as in the case of i.i.d. variables.

5 Conclusions

Within this project, bootstrap methods were studied, starting from the basic theory and then through applying their basic principles.

This made it possible to study the applications of bootstrap methods on the calculation of statistical measures like variance, quantiles and confidence intervals.

It was observed that generally we got better results when using bootstrap methods.

Bibliography

[1] Michael R. Chernick, (2007). *Bootstrap Methods. A Guide for Practitioners and Researchers*. Wiley Series in Probability and Statistics.

[2] Sandor Csorgo and David Mason, (1989). *Bootstrapping Empirical Functions*. University of Sweged and University of Delaware.

[3] Sandor Csorgo and Andrew Rosalsky, (2003). *A survey of Limit Laws for Bootstrapped Sums*.

[4] A. W. van der Vaart, (1997). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Leiden.

[5] Jens Praestgaard and Jon A. Wellner, (1992). *Exchangeably Weighted Bootstraps of the General Empirical Process*. University of Iowa and University of Washington.

[6] Vlad Stefan Barbu and Nikolaos Limnios, (2008). *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications*.

[7] Massart P., (1990). *The tight constantin the Dvoretsky-Kiefer-Wolfowitz inequality. Annals of Probability*. 1269-1283

[8] Breiman L., (1965). *On some limit theorems similar to the arc-sin law. Theory Probab. Appl. 10* 323-331.

[9] Efron B., (1979). *Bootstrap Methods: Another Look at the jackknife*.

[10] Efron B., (1982a). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia

[11] Efron B. and Tibshirani R., (1986). *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy*.

[12] Hoffmann-Jorgensen, J., (1984). *Stochastic Process on Polish Spaces. Unpublished manuscript*

[13] Gine, E. and Zinn, J. (1990). *Bootstrapping general empirical functions*.

[14] Mason David and Newton Michael, (1992). *A Rank Statistics Approach to the Consistency of General Bootstrap*.

[15] Salim Bouzebda and Nikolaos Limnios, (2013). *On general bootstrap of empirical estimator of a semi-Markov kernel with applications*. Universite de technologie de Compiegne

[16] S. Trevezas and N. Limnios, (2008) *Variance estimation in the central limit theorem for Markov chains,*

[17] Quenouille , M. H. (1949). *Approximate tests of correlation in time series*. J. Roy. Statist. Soc. B 11 , 18-84 .

[18] Simon . J. L. (1969). Basic Research Methods in Social Science . Random House , New York .

[19] Simon . J. L. , and Bruce , P. (1991). *Resampling: A tool for everyday statistical work*. Chance 4 , 22-32.

[20] Simon . J. L. , and Bruce , P. (1995). *The new biostatistics of resampling*. M. D. Comput. 12 , 115-121.

[21] Efron , B. (1987). *Better bootstrap confidence intervals (with discussion)*. J. Am. Statist. Assoc . 82 , 171-200 .

[22] Chung , K. L. (1974). A Course in Probability Theory , 2nd ed . Academic Press , New York .

[23] Hampel , F. R. (1974). *The influence curve and its role in robust estimation*. J. Am. Statist. Assoc . 69 , 383-393 .

[24] Fernholtz , L. T. (1983). von Mises Calculus for Statistical Functionals . Lecture Notes in Statistics, Vol. 19. Springer-Verlag , New York .

[25] Hall , P. (1992a). The Bootstrap and Edgeworth Expansion . Springer-Verlag , New York .

[26] Athreya , K. B. (1983). *Strong law for the bootstrap*. Statist. Probab. Lett. 1 , 147-150.

[27] Tucker , H. G. (1959). *A generalization of the Glivenkob•"Cantelli theorem*. Ann. Math. Statist. 30 , 828-830.