UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET

Jelena S. Stanojević

Statistički problemi ocenjivanja količnika disperzija i visokih kvantila raspodela

Doktorska disertacija

Beograd, maj 2015.

UNIVERSITY OF BELGRADE
FACULTY OF MATHEMATICS

Jelena S. Stanojević

Statistical problems of estimation of the ratio of variances and the large quantiles of the distributions

Doctoral Dissertation

Belgrade, May 2015.

Mentor: Profesor dr Pavle Mladenović,
Matematički fakultet, Univerzitet u Beogradu.

Članovi komisije:

1. Profesor dr Slobodanka Janković, Matematički fakultet, Univerzitet u Beogradu.
2. Profesor dr Ljiljana Petrović, Ekonomski fakultet, Univerzitet u Beogradu.

Zahvalnost

Zahvaljujem se svom mentoru Prof. dr Pavlu Mladenovicu na korisnim primedbama i komentarima, što je doprinelo boljem kvalitetu disertacije. Takođe se zahvaljujem Prof. dr Ljiljani Petrović i Prof. dr Slobodanki Janković na ukazanim propustima u tekstu.

Posebnu zahvalnost dugujem kolegenici i drugarici dr. Tijani Levajković na stalnoj podršci, sugestijama i tehničkoj pomoći u vezi disertacije, kao i kolegenici dr. Vesni Rajić na korisnim komentarima i Prof. dr Branislavu Boričiću na podršci koju mi je pružio.

I na kraju, veliko hvala mojoj porodici, sestri Mariji i mojim roditeljima, ocu Sretenu i majci Svetlani, na ogromnom razumevanju, podršci i ljubavi koju su mi pružili sve ovo vreme. Njima posvećujem ovaj rad i žalim što tata nije doživeo da ga pročita, a sigurno bi mu se radovao.

Jelena Stanojević

Beograd, maj 2015.

Naslov teze: STATISTIČKI PROBLEMI OCENJIVANJA KOLIČNIKA DISPERZIJA I VISOKIH KVANTILA RASPODELA

Apstrakt:
Glavni sadržaj doktorske teze odnosi se na predlog novih, transformisanih intervala poverenja za količnik disperzija dva uzorka. Naime, u literaturi su do sada predložene metode bazirane na F statistici. Međutim, nedostatak razmatranih intervala jeste velika osetljivost u odnosu na pretpostavke o parametrima raspodele. Predložena statistika u literaturi može biti modifikovana. U tezi je nađen Edgeworthov razvoj t-statistike i na bazi toga su upoređivani intervali. Takođe je ukazano na osnovu simulacija, da interval baziran na Johnsonovoj transformaciji daje bolji rezultat u smislu verovatnoće pokrivanja u odnosu na F interval i interval baziran na Hallovoj transformaciji. U radu je posvećena pažnja i već poznatim intervalima poverenja za matematičko očekivanje i disperziju za problem jednog uzorka, kao i za problem dva uzorka. Posebno je razmatran problem razlike proporcija dva uzorka, sa numeričkim rezultatima i podacima iz oblasti osiguranja.
Pored pomenutog, u tezi su prikazane postojeće metode za ocenu indeksa ekstremne vrednosti i visokih kvantila. Posebno je razmatrana direktno simulirana ocena kvantila i verovatnoća pokrivanja njenog odstupanja od tačne vrednosti, za slučaj Pareto i Gama raspodele, kao i uopšteni slučaj uopštene Pareto raspodele. Rezultati su dobijeni na osnovu teorijskih rezultata teorije velikog odstupanja i dato je njeno uopštenje na topološkim prostorima.
Uz teoriju verovatnoće i elementarnih principa klasične analize, u istraživanju su korišćene i metode teorijske statistike i statističkog zaključivanja.

Ključne reči: Ocena količnika disperzija, ocena indeksa ekstremne vrednosti i visokih kvantila, direktno simulirana ocena, teorija velikog odstupanja, Tihonovljev tpološki prostor, idempotentna mera, idempotentna integracija

Naučna oblast: Matematika

Uža naučna oblast: Verovatnoća i statistika

UDK broj: [519.222+519.234]:519.243(043.3)

AMS klasifikacija: 62G32, 60G70, 62G15.

Dissertation title:
STATISTICAL PROBLEMS OF ESTIMATION OF THE RATIO OF VARIANCES AND THE LARGE QUANTILES OF THE DISTRIBUTIONS

Abstract:
The main goal of the thesis is the development of a new suggested transform confidence intervals for the ratio of the variances of the two samples. Since now, the methods based on the F statistic have been suggested in the literature. However, the defect of that intervals is the huge sensitivity in relation with assumption of parameters distribution. Suggested statistic could be modified. Edgeworth expansion of the t-statistic has found the place in the thesis and based on that intervals have been compared. Also, on the base of the simulation it was point out that Johnsons transformation give better result in the sense of probability covering in regard to F interval and interval based on Halls transformation. Moreover, the confidence intervals for the mean and variances for the one and two sample problems have been considered in the dissertation. Especially, the problem of the difference of the proportions for the two samples, with the numerical results and data from the insurance. In addition, the existing methods for the estimation of the extreme value index and the high quantiles have been reviewed. Particularly, the direct simulation estimation of the quantile and probability covering of its deviation from the rights value, for Pareto and Gamma distributions, and also for general Pareto distribution have been discussed. The results were obtained by large deviation theory and their generalization on the topological spaces is stated. In this research, beside the probability theory and elementary principles of the classical analysis, methods of the statistical theory and statistical conclusions have been applied.

Keywords: Estimation of the ratio of two variances, estimation of the extreme value index and high quantiles, direct simulation estimation, large deviation theory, Tihonov topological space, idempotent measure, idempotent integration.

Scientific field: Mathematics

Specialized scientific field: Probability and Statistics

UDK number: [519.222+519.234]:519.243(043.3)

AMS classfication: 62G32, 60G70, 62G15.

Uvod

Teorija ekstremnih vrednosti se bavi problemima raspodela i njihovih parametara, maksimuma i minimuma slučajnih vrednosti. Samo predvidjanje i modeliranje ekstremnih dogadjaja od velike je važnosti u prektičnim problemima u hidrologiji, ekonomiji, finansijskoj matematici, osiguranju i drugim oblastima. To je razlog njenog razvoja, posebno poslednjih decenija.
Predmet ove disertacije jeste analiza i svojstva ocena indeksa ekstremne vrednosti i visokih kvantila. Takodje od interesa je posebna ocena kvantila i njena brzina konvergencije u slučajevima Pareto, uopštene Pareto i Gama raspodele.

U prvom poglavlju ove disertacije biće predstavljeni, već poznati u literaturi, intervali poverenja za sredinu, disperziju i binomnu proporciju (problem jednog uzorka), kao i za razliku sredina, disperzija i proporcija (problem dva uzorka). Pri tome će biti korišćena literatura: Zhou i Dinh, 2005, Johnson, 1978, Hall, 1992a,b, Chen, 1995, Zhou i Gao, 2000, Zhou et al., 1997. Analiza različitih intervala poverenja za razliku proporcija i njihovo poređenje jeste predmet publikovanog rada Rajić i Stanojević, 2011. Posebna pažnaja biće posvećena analizi pokrivenosti intervala poverenja za količnik disperzija dva uzorka. Naime, za odgovarajuću t-statistiku biće nadjen Edgeworthov razvoj te statistike i analiziraće se pokrivenost intervala dobijenih odgovarajućim transformacijama i postojećeg F-intervala, što je jedan od originalnih rezultata ove disertacije, publikovan u radu Ćojbašić i Stanojević, 2013. Simulacijom različitih tipova asimetričnih raspodela i različitih parova veličina uzoraka, biće analiziran problem pokrivenosti i preporučen najbolji interval, takodje rezultat publikovan u radu Ćojbašić i Stanojević, 2013. Na kraju prvog poglavlja biće predstavljen još jedan od originalnih publikovanih rezultata disertacije, prezentovanje predloženih metoda na podacima iz osiguranja. Biće korišćena sledeća literatura: Gnedenko, 1943, de

Haan, 1970, Ferreira et al., 2003, Matthys i Beirlant, 2003, Danielsson et al., 2001, Dekkers i de Haan, 1993, de Haan i Resnick, 1980, Pickands, 1975, Dekkers et al., 1975 i drugi.

U drugom i trećem poglavlju ove disertacije biće predstavljene postojeće ocene i intervali poverenja za indeks ekstremne vrednosti i visoke kvantile. Naime, u drugom poglavlju biće analizirane metode za dobijanje ocena pomenutih parametara: metoda ekstremne vrednosti, POT metod i Q(vantil) bazirani metod, koji uključuje Hillovu ocenu, Pickandsovu ocenu, ocenu momenta i ocenu maksimalne verodostojnosti. Za te ocene biće analizirani uslovi za slabu i jaku konzistentnost, kao i asimtotska svojstva ocena indeksa ekstremne vrednosti i biće date odgovarajuće teoreme i za konvergenciju ocena visokih kvantila. Na kraju drugog poglavlja biće data i MVRB ocena indeksa ekstremne vrednosti i visokih kvantila, koja uključuje i analizu još nekih parametara drugog reda i biće dato poređenje postojećih ocena. Koristiće se sledeća literatura: Gnedenko, 1943, de Haan, 1970, Ferreira et al., 2003, Matthys i Beirlant, 2003, Danielsson et al., 2001, Dekkers i de Haan, 1993, de Haan i Resnick, 1980, Pickands, 1975, Dekkers et al., 1975 i drugi.

U četvrtom poglavlju ove disertacije, u okviru prvog odeljka, biće dati osnovni principi velikog odstupanja, osnovna teorema ove oblasti, teorema Kramera, kao i uopštenje na prostoru R^k u vidu Gärtner-Ellisove teoreme i biće dat princip kontrakcije, čije je uopštenje na topološkom prostoru dato u okviru poslednjeg, trećeg odeljka ovog poglavlja. Koristiće se sledeća literatura: Mörters, 2008, Jin u C.Fu, 2002, Buckwel, 1990, Dembo i Zeitoni, 1998, Deuschel i Stroock, 1989 i Puhalskii, 2000. Primena izložene teorije velikog odstupanja biće data u okviru drugog odeljka ovog poglavlja. Naime, biće data primena na visokim kvantilima, u okviru teoreme, za direktno-simuliranu ocenu kvantila, u slučaju strogo rastuće funkcije raspodele za negativno zavisne slučajne promenljive. Ista teorema biće primenjena za dobijanje brzine konvergencije spomenute ocene, u slučaju Pareto raspodele, kao i za njeno uopštenje, na uopštenoj Pareto raspodeli i takođe na Gama raspodeli. Dobijeni rezultati su publikovani u radu Stanojević, 2014 i predstavljaju originalne rezultate ove disertacije. I na kraju, u trećem odeljku ovog četvrtog poglavlja, biće razmatrana konvergencija u velikom odstupanju u topološkom Tihonovljevom prostoru. Radi razumevanja, prvo će biti dati pojmovi kao što su idempotentna mera i idempotentna integracija, nezavisnost idempotentnih veličina, mera Luzina, E_0-gornje i E_0-donje semineprekidne funkcije, kao i još neki pojmovi ključni za razumevanje ovog odeljka. Osnovni razlog za izdvajanje definicija i teorema u ovom odeljku, koje su tehničke prirode, leži upravo u Teoremi 4.3.5, a može se iskoristiti za definiciju pojma konvergencije u velikom odstupanju, koja je identična definiciji principa velikog odstupanja. Naime, konvergencija u velikom odstupanju na Tihonovljevom prostoru jeste ekvivalentna principu velikog odstupanja (koji je razmatran u prvom odeljku ovog pogavlja). I na kraju treba reći da će u pojedinim delovima ovog poglavlja, uz odgovarajuće teoreme, biti data sugestija sa mogućim uopštenjima, koja mogu predstavljati jedan od mogućih pravaca daljeg istraživanja.

Sadržaj

Glava 1

Intervalno ocenjivanje sredine, disperzije i binomne proporcije

1.1 Intervali poverenja sredine za probleme jednog i dva uzorka

U ovom odeljku biće reči o nekim postojećim tehnikama za testiranje sredine (problem jednog uzorka) i razlike sredina (problem dva uzorka), koje se baziraju na t-statistici i *bootstrap*-t statistici, kao i o novim metodama (koje su predložili Zhou i Dinh, 2005). Kako su dosadašnja israživanja pokazala da novi intervali poverenja pokazuju bolju moć pokrivanja i manju širinu intervala, oni se predlažu za upotrebu. Takođe su u ovom delu rada analizirani intervali poverenja kada su u pitanju podaci iz asimetrične i pozitivno asimetrične raspodele, koji su već poznati u literaturi. Pogledati na primer, Johnson, 1978, Hall, 1992b, Efron i Tibshirani, 1993, Zhou i Gao, 2000.

U ovoj disertaciji se koriste sledeće oznake: slučajne promenljive X_1, X_2, ..., X_n koje su nezavisne i jednako raspodeljene, skraćeno u oznaci i.i.d. slučajne promenljive, uzoračka sredina $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, uzoračka disperzija $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, korigovana uzoračka disperzija $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ (koja je nepristrasna ocena disperzije raspodele iz koje je uzet uzorak). Takođe, gustina slučajne promenljive $Z \sim N(0,1)$, sa standardnom normalnom raspodelom, je $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$, dok je njena funkcija raspodele $\Phi(x) = P(Z < x) = \int_{-\infty}^{x} \phi(t)\, dt$, $x \in \mathbb{R}$. Njeno matematičko očekivanje i disperzija su redom, 0 i 1. Standardna normalna raspodela je primer simetrične raspodele, tj. $f(x) = f(-x)$, gde je f gustina raspodele. Raspodela slučajne promenljive X za koju važi $EX^3 > 0$ ($EX^3 < 0$) jeste pozitivno (odnosno negativno) asimetrična raspodela. Odgovarajući kvantili se nazivaju percentili raspodele. Tako je odgovarajući percentil z_α, kvantil normalne raspodele za koji važi: $\Phi(z_\alpha) = P(Z < z_\alpha) = 1 - \alpha$, tj. $P(Z > z_\alpha) = \alpha$, dok je $t_{n,\alpha}$ kvantil Studentove raspodele, sa n stepeni slobode, za koji važi: $P(T > t_{n,\alpha}) = \alpha$.

1.1.1 Postojeće metode za problem jednog uzorka

Neka su $X_1, X_2, ...X_n$ i.i.d. iz raspodele sa matematičkim očekivanjem μ i disperzijom σ^2. Interval poverenja za μ, koji se najčešće koristi u literaturi i praksi, kada je disperzija nepoznata, čak i kada je veličina uzorka mala ($n < 30$), bazira se na t-statistici, predložio je Student, 1908. Statistika je oblika:

$$T = \frac{\sqrt{n}(\overline{X} - \mu)}{S}. \tag{1.1}$$

Odgovarajući $(1-\alpha)100\%$ dvostrani interval poverenja za μ, baziran na t-statistici (1.1) je oblika:

$$\Big(\overline{X} - n^{-1/2} t_{n-1,\alpha/2} S, \overline{X} + n^{-1/2} t_{n-1,\alpha/2} S\Big),$$

gde je $t_{n-1,\alpha/2}$ odgovarajući percentil Studentove raspodele sa $(n-1)$ stepeni slobode. Odgovarajući jednostrani intervali poverenja su oblika:

$$\Big(-\infty, \overline{X} + n^{-1/2} t_{n-1,\alpha} S\Big),$$

$$\Big(\overline{X} - n^{-1/2} t_{n-1,\alpha} S, +\infty\Big).$$

U slučaju kada je disperzija poznata, koristi se normalna raspodela, i odgovarajući dvostrani interval poverenja jeste:

$$\Big(\overline{X} - n^{-1/2} z_{\alpha/2} \sigma, \overline{X} + n^{-1/2} z_{\alpha/2} \sigma\Big),$$

dok su jednostrani intervali:

$$\Big(-\infty, \overline{X} + n^{-1/2} z_\alpha \sigma\Big),$$

$$\left(\overline{X} - n^{-1/2} z_\alpha \sigma, +\infty\right).$$

Za velike uzorke, odgovarajući interval poverenja je:

$$\left(\overline{X} - n^{-1/2} z_{\alpha/2} S, \overline{X} + n^{-1/2} z_{\alpha/2} S\right). \tag{1.2}$$

U literaturi je već poznato da je verovatnoća pokrivanja poslednjeg intervala tačno $1-\alpha$. U slučaju kada je reč o podacima koji potiču iz raspodele koja nije normalna, verovatnoća pokrivanja je približno $1-\alpha$. Mnogi autori, kao na primer Gayen, 1949, Barrett i Goldsmith, 1976, Johnson, 1978, Chen, 1995 i Boos i Hughes-Oliver, 2000, ispitivali su moć pokrivanja za interval (1.2) u zavisnosti od asimetrije raspodele (engleski: skewness) i veličine uzorka. Pronašli su da je za t-interval ona slaba, u slučaju asimetričnih raspodela, a da se popravlja sa rastom veličine uzorka.

Za ocenjivanje sredine uzorka u slučaju asimetričnih raspodela predložena su tri pristupa. Prvi pristup jeste interval poverenja (1.2), koji se dobija koristeći centralnu graničnu teoremu. Drugi pristip koji se koristi bazira se na transformaciji postojećih podataka. Od svih transformacija najčešće se koristi logaritamska funkcija. Treći pristup podrazumeva korišćenje neparametarskih metoda.

Svaki od ova tri pristupa ima svoje slabosti. Pri značajnom odstupanju od normalnosti, pristup baziran na t-statistici je mnogo robusan (videti Boos i Hughes-Oliver, 2000). Za aproksimaciju u (1.2) centralna granična teorema ne daje odgovor kolika veličina uzorka n treba da bude, već to zavisi od asimetrije i nešto manje od spljoštenosti (engleski: kurtosis) raspodele (videti Barret i Goldsmith, 1976, Boss i Hughes-Oliver, 2000). Drugi pristup, ocenjivanje sredine za problem jednog uzorka na transformisanoj skali, nije uvek isto sa ocenjivanjem na originalnoj skali, zato pristup transformacije uzorka može biti neodgovarajući (pogledati Zhou et al., 1997). Nedostatak trećeg pristupa je što se kod neparametarskih metoda analizira medijana a ne sredina.

Za problem jednog uzorka Johnson, 1978, Hall, 1992b i Chen, 1995 su najviše istraživali modifikaciju t-statistike, kako bi se uklonio efekat asimetrije. Pristup je baziran na Edgeworthovom razvoju (videti Hall, 1992a). Oni su pokazali, da kada je veličina uzorka mala i kada je raspodela asimetrična,

t-statistika treba da bude zamenjena sledećom statistikom:

$$t_1 = \Big\{(\overline{X} - \mu) + \frac{\widehat{\mu}_3}{6nS^2} + \frac{\widehat{\mu}_3}{3S^4}(\overline{X} - \mu)^2\Big\}\Big(S^2/n\Big)^{-1/2},$$

gde je $\widehat{\mu}_3$ ocena trećeg centralnog momenta.

Ako je $U = (\overline{X} - \mu)/S$, Edgeworthov razvoj raspodele U statistike (videti Hall, 1992b) je:

$$P(n^{1/2}U \leq x) = \Phi(x) + n^{-1/2}\gamma(ax^2 + b)\phi(x) + O(n^{-1}),$$

gde su $a = 1/3$ i $b = 1/6$, γ je koeficijent asimetrije koji treba da se oceni. Hall, 1992b je predložio dve transformacije:

$$T_1 = T_1(U) = U + a\widehat{\gamma}U^2 + \frac{1}{3}a^2\widehat{\gamma}^2U^3 + n^{-1}b\widehat{\gamma}, \tag{1.3}$$

$$T_2 = T_2(U) = (2an^{-1/2}\widehat{\gamma})^{-1}\{e^{2an^{-1/2}\widehat{\gamma}U} - 1\} + n^{-1}b\widehat{\gamma}. \tag{1.4}$$

Ove transformacije zadovoljavaju da je $T(U) \approx U$, za U u okolini nule i $T(0) = 0$ (osim za $n^{-1}b\widehat{\gamma}$), pogledati Hoaglin, 1985.

Zhou i Dinh, 2005 su predložili novu, jednostavniju transformaciju:

$$T_3 = T_3(U) = U + U^2 + \frac{1}{3}U^3 + n^{-1}b\widehat{\gamma}. \tag{1.5}$$

$(1 - \alpha)100\%$ interval poverenja za sredinu μ je oblika:

$$\overline{X} - T_i^{-1}(n^{-1/2}z_{\alpha/2})S \leq \mu \leq \overline{X} - T_i^{-1}(n^{-1/2}z_{1-\alpha/2})S,$$

gde su $T_i^{-1}(\cdot)$, $i = 1, 2, 3$, inverzne funkcije funkcija $T_i(\cdot)$, koje se mogu analitički rešiti i imaju sledeći oblik:

$$\begin{aligned}
T_1^{-1}(t) &= \Big(\frac{1}{3}\widehat{\gamma}\Big)\Big\{1 + 3\frac{1}{3}\widehat{\gamma}\Big(t - n^{-1}\frac{1}{6}\widehat{\gamma}\Big)\Big\}^{1/3} - \Big(\frac{1}{3}\widehat{\gamma}\Big)^{-1}, \\
T_2^{-1}(t) &= \Big(2\frac{1}{3}n^{-1/2}\widehat{\gamma}\Big)\ln\Big\{2\frac{1}{3}n^{-1/2}\widehat{\gamma}\Big(t - n^{-1}\frac{1}{6}\widehat{\gamma}\Big) + 1\Big\}, \\
T_3^{-1}(t) &= \Big\{1 + 3\Big(t - n^{-1}\frac{1}{6}\widehat{\gamma}\Big)\Big\}^{1/3} - 1.
\end{aligned}$$

Hall, 1992b, Zhou i Gao, 2000 i drugi, proučavali su valjanost metode transformacija. Obimne simulacije dali su Zhou i Dinh, 2005, kako bi uporedili ove metode sa postojećim metodama. Pokazali su da *bootstrap-t* intervali daju konzistentan i dobar rezultat, u smislu verovatnoće pokrivanja. U sledećem odeljku dato je detaljnije objašnjenje *bootstrap* principa i računa na osnovu njega. Ovde je dovoljno reći da je taj princip analogan principu 'Ruske lutke babuške', gde se na primer karakteristika broja pega na početnoj lutki računa kao funkcija broja pega na dve sledeće lutke, koje se izvlače iz početne. Naime, ako se sa χ_0 označi populacija, izvlači se sločajan uzorak χ iz χ_0 i uzorak χ^* iz uzorka χ i dalje se vrši izračunavanje. Sa *bootstrap-t* intervalom uporedive su metoda koja koristi T_3 transformaciju ili T_1 transformaciju, i pokazano je da su te metode ponekad bolje. Za uzorke obima većeg od 100, interval baziran na T_3 transformaciji daje užu pokrivenost u smislu srednje dužine intervala poverenja, u poređenju sa *bootstrap-t* intervalom i intervalom baziranim na T_1 transformaciji. Utvrđeno je da je uobičajeni-t interval neadekvatan, u slučaju kada je koeficijent $\widehat{\gamma}/\sqrt{n}$ veći od 0,3. Stoga se za podatke iz znatno asimetričnih raspodela i za relativno male veličine uzorka ($\widehat{\gamma}/\sqrt{n} \geq 0,3$), intervali bazirani na T_1 ili T_3 transformaciji ili *bootstrap-t* intervali, preporučuju u odnosu na standardni t-interval.

1.1.2 Jednostrani intervali poverenja za sredinu pozitivno asimetrične raspodele. Testiranje sredine asimetrične raspodele

Jednostrani intervali poverenja za sredinu pozitivno asimetrične raspodele

U ovom odeljku biće razmatrani jednostrani intervali poverenja za sredinu i njihova verovatnoća pokrivanja, u slučaju asimetričnih raspodela. Intervali su konstruisani koristeći Hallovu i Johnsonovu transformaciju i *bootstrap* metod. Dalje su dati već poznati rezultati, da donje krajnji interval poverenja dobijen na osnovu Hallove transformacije pokazuje bolju verovatnoću pokrivanja od onog baziranog na Johnsonovoj transformaciji, dok Hallov gornje krajnji interval poverenja ima slabu verovatnoću pokrivanja. Takođe je pokazano kroz simulacije da je i za gornji i donji krajnji interval poverenja najveća verovatnoća pokrivanja kada se koristi *bootstrap* metod i Hallova transformacija.

Već je rečeno da je standardni t-interval:

$$\Lambda_0 = \left(\overline{X} - n^{-1/2} t_{n-1,\alpha/2} S, \overline{X} + n^{-1/2} t_{n-1,\alpha/2} S\right).$$

U slučaju kada podaci potiču iz pozitivno asimetrične raspodele problem sa standardnim t intervalom je to što se bazira na asimptotski normalnoj aproksimaciji, koja može biti potpuno neadekvatna. Bez pretpostavke o obliku raspodele, Johnson, 1978 je predložio transformaciju za korigovanje pristrasnosti i asimetrije raspodele. Hall, 1992b je istakao da Johnsonova transformacija nije monotona, niti invertibilna, i predložio je monotonu transformaciju za eliminisanje asimetričnosti raspodele. Njegova metoda koriguje prvi član u Edgeworthovom razvoju raspodele t-statistike. Ideja je da se transformiše t-statistika u drugu, čija je raspodela aproksimativno normalna. Hall, 1992b je pokazao da donje krajnji interval poverenja baziran na njegovoj transformaciji može dati visoku verovatnoću pokrivanja za sredinu mnogih asimetričnih raspodela. Iako Johnsonova transformacija nije invertibilna, Abramovitch i Singh, 1985 su pokazali da je ona invertibilna do na $n^{-1/2}$ i iskoristili su je za konstrukciju intervala poverenja sredine.

Dalje će biti razmatrana Johnsonova i Hallova transformacija i *bootstrap* metod, za konstrukciju jednostranog, $(1-\alpha)100\%$ intervala poverenja za sredinu pozitivno asimetrične raspodele.

Neka je $X_1, ..., X_n$ slučajan uzorak iz pozitivno asimetrične raspodele sa sredinom μ i disperzijom σ^2. Ako se ignoriše asimetričnost raspodele, može se konstruisati interval poverenja baziran na t-statistici:

$$T = \frac{\sqrt{n}(\overline{X} - \mu)}{S^*}.$$

Ako je raspodela slučajne promenljive X apsolutno neprekidna i ako je $E(X^4) < +\infty$, Edgeworthov razvoj raspodele statistike T (videti Barndorff-Nielsen i Cox, 1989, Bhattacharya i Ghosh, 1978) je:

$$P(T \leq x) = \Phi(x) + n^{-1/2}\gamma(ax^2 + b)\phi(x) + O(n^{-1}), \tag{1.6}$$

gde su $a = 1/3$, $b = 1/6$ i $\gamma = E((X-\mu)/\sigma)^3$ je koeficijent asimetrije. Iz jednakosti (1.6) se može videti da tačnost normalne aproksimacije raspodele statistike T zavisi od koeficijenta asimetrije γ.
$(1-\alpha)100\%$ interval poverenja baziran na statistici T jeste Λ_0, za koji važi:

$$P(\mu \in \Lambda_0) = 1 - \alpha + O(n^{-1/2}).$$

U cilju eliminisanja efekta asimetrije γ, Hall, 1992b je predložio transformaciju statistike T, koja je oblika:

$$g(T) = T + n^{-1/2}\widehat{\gamma}(aT^2 + b) + n^{-1}(1/3)(a\widehat{\gamma})^2T^3,$$

gde je:

$$\widehat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^3/S^{*3}.$$

Ova transformacija je monotona i njena inverzna funkcija je:

$$T = g^{-1}(x) = n^{1/2}(a\widehat{\gamma})^{-1}[(1 + 3a\widehat{\gamma}(n^{-1/2}x - n^{-1}b\widehat{\gamma}))^{1/3} - 1],$$

za koju je pokazano da važi:

$$P(g(T) \leq x) = \Phi(x) + O(n^{-1}). \tag{1.7}$$

Gornje krajnji $(1-\alpha)100\%$ interval poverenja za μ, baziran na inverznoj transformaciji, je oblika:

$$\Lambda_1+ = (-\infty, \overline{X} - n^{-1/2}S^*g^{-1}(z_\alpha)),$$

i donje krajnji $(1-\alpha)100\%$ interval poverenja za μ, baziran na inverznoj transformaciji, je oblika:

$$\Lambda_1- = (\overline{X} - n^{-1/2}S^*g^{-1}(z_{1-\alpha}), +\infty).$$

Koristeći jednakost (1.7), može se pokazati da je verovatnoća pokrivanja intervala poverenja:

$$\begin{aligned} P(\mu \in \Lambda_1+) &= 1 - \alpha + O(n^{-1}), \\ P(\mu \in \Lambda_1-) &= 1 - \alpha + O(n^{-1}). \end{aligned}$$

Umesto korišćenja normalne raspodele, Hall, 1992b, Efron i Tibshirani, 1993 su predložili da se koristi *bootstrap* pristup za određivanje empirijskog 100α-tog percentila raspodele statistike $g(T)$, bez pretpostavki o obliku raspodele. Neka je $\chi^* = (X_1^*, ..., X_n^*)$ slučajni uzorak veličine n, izvučen sa ponavljanjem iz originalnog uzorka $\chi = (X_1, ..., X_n)$. Neka je $g^*(T^*)$ *bootstrap* verzija statistike $g(T)$ izračunata na osnovu uzorka χ^*. Raspodela statistike $g^*(T^*)$ se naziva *bootstrap* raspodela statistike $g(T)$. Neka je $\widehat{t^{(\alpha)}}$ 100α-ti percentil *bootstrap* raspodele statistike $g(T)$. On se izračunava na sledeći način. Generiše se B *bootstrap* uzoraka, za b-ti uzorak $(X_1^{*b}, ..., X_n^{*b})$ se računa *bootstrap* verzija statistike $g(T)$, u oznaci:

$$g^*(T^*)_b = T_b^* + n^{-1/2}\widehat{\gamma}_b^*(a(T^*)^2 + b) + n^{-1}(1/3)(a\widehat{\gamma}_b^*)^2(T_b^*)^3,$$

gde su $\widehat{\mu}_b^* = (1/n)\sum_{i=1}^n X_i^{*b}$, $(\widehat{\sigma}_b^*)^2 = (1/n)\sum_{i=1}^n(X_i^{*b} - \widehat{\mu}_b^*)^2$, $\widehat{\gamma}_b^* = (1/n)\sum_{i=1}^n(X_i^{*b} - \widehat{\mu}_b^*)^3/(\widehat{\sigma}_b^*)^3$ i $T_b^* = n^{1/2}(\widehat{\mu}_b^* - \widehat{\mu})/\widehat{\sigma}_b^*$. U slučaju kada je $B\alpha$ ceo broj, $\widehat{t^{(\alpha)}}$ je $B\alpha$-ta najveća vrednost od $g^*(T^*)_b$. U slučaju kada $B\alpha$ nije ceo broj, $\widehat{t^{(\alpha)}}$ je k-ta najveća vrednost od $g^*(T^*)_b$, gde je k najveći ceo broj manji ili jednak od $(B+1)\alpha$. Gornje krajnji interval poverenja, baziran na *bootstrap* metodu, je:

$$\Lambda_2+ = \Big(-\infty, \overline{X} - n^{-1/2}S^*g^{-1}\Big(\widehat{t^{(\alpha)}}\Big)\Big),$$

dok je donje krajnji interval poverenja:

$$\Lambda_2- = \Big(\overline{X} - n^{-1/2}S^*g^{-1}\Big(\widehat{t^{(1-\alpha)}}\Big), +\infty\Big).$$

Oba intervala Λ_2- i Λ_2+ imaju grešku verovatnoće pokrivanja reda $n^{-3/2}$, manju od one date korišćenjem transformacija. Važi:

$$\begin{aligned} P(\mu \in \Lambda_2+) &= 1 - \alpha + O(n^{-3/2}), \\ P(\mu \in \Lambda_2-) &= 1 - \alpha + O(n^{-3/2}). \end{aligned}$$

Dokaz se može pogledati u Zhou i Gao, 2000.

U radu Zhou i Gao, 2000 je takođe sugerisana alternativna transformacija koja uklanjanja član $n^{-1}(a\widehat{\gamma})^2T^3/3$ iz Hallove transformacije, i ona je oblika:

$$g_1(T) = T + n^{-1/2}\widehat{\gamma}(aT^2 + b).$$

Ovu transformaciju je zapravo predložio Johnson, 1978 i dalje su je proučavali Abramovich i Singh, 1985. Oni su pokazali da je, do na $n^{-1/2}$, rešenje jednačine $g_1(T) = x$ oblika:

$$g_1^{-1}(x) = x - n^{-1/2}\widehat{\gamma}(ax^2 + b).$$

Gornje krajnji i donje krajnji $(1-\alpha)100\%$ intervali poverenja za μ, bazirani na Johnsonovoj transformaciji, su:

$$\begin{aligned} \Lambda_3+ &= (-\infty, \overline{X} - n^{-1/2}S^*g_1^{-1}(z_\alpha)), \\ \Lambda_3- &= (\overline{X} - n^{-1/2}S^*g_1^{-1}(z_{1-\alpha}), +\infty). \end{aligned}$$

Gornje krajnji i donje krajnji $(1-\alpha)100\%$ intervali poverenja, kada se koristi *bootstrap* metod i Johnsonova transformacija, su:

$$\begin{aligned}\Lambda_4+ &= (-\infty, \overline{X} - n^{-1/2}S^* g_1^{-1}(\widehat{t}_1^{(\alpha)})),\\ \Lambda_4- &= (\overline{X} - n^{-1/2}S^* g_1^{-1}(\widehat{t}_1^{(1-\alpha)}), +\infty),\end{aligned}$$

gde je $\widehat{t}_1^{(\alpha)}$ 100α-ti percentil *bootstrap* raspodele statistike $g_1(T)$. Treba naglasiti da postoji potencijalni problem sa ovom *bootstrap* metodom, iz razloga što $g_1(T)$ nije monotona transformacija. U radu Zhou i Gao, 2000 je pokazano da ako je $\widehat{t}_1^{(\alpha)} + \widehat{t}_1^{(1-\alpha)} > \frac{\sqrt{n}}{a\widehat{\gamma}}$, što se dešava u slučajevima kada je mala veličina uzoraka i velika asimetrija, tada *bootstrap* u kombinaciji sa Johnsonovom transformacijom može dati pogrešan dvostrani interval poverenja, čija je donja krajnja tačka veća od gornje krajnje tačke. Još je primećeno, iz razloga što je $z_\alpha + z_{1-\alpha} = 0$, da Johnsonova transformacija, koja se bazira na normalnim percentilima, nema navedeni problem.

Simulacije u radu Zhou i Gao, 2000 pokazuju da za male uzorke Hallova transformacija daje dobru verovatnoću pokrivanja za donje krajnje intervale poverenja, dok gornje krajnji intervali poverenja imaju slabu verovatnoću pokrivanja. Hallova transformacija u kombinaciji sa *bootstrap* metodom poboljšava vrednost verovatnoće pokrivanja u odnosu na Hallovu originalnu metodu za gornje i donje krajnje intervale poverenja i ima najbolju verovatnoću pokrivanja i za gornje i donje krajnje intervale poverenja, među svim metodama razmatranim u radu. Kao što je i očekivano, verovatnoća pokrivanja i transformisanih intervala i njihovih *bootstrap* verzija, opada sa porastom asimetrije.

Testiranje sredine asimetrične raspodele

Standardna statistika za testiranje sredine za raspodelu koja nije normalna, u slučaju kada je σ poznato i veličina uzorka velika, jeste $\sqrt{n}(\overline{X} - \mu)/\sigma$, dok je u slučaju sa nepoznatim σ test statistika dobro poznata Studentova t statistika, $\sqrt{n}(\overline{X} - \mu)/S$. Kao što je rečeno, kada je veličina uzorka mala i polazna raspodela asimetrična, Johnson, 1978 je predložio da se statistika $\sqrt{n}(\overline{X} - \mu)/S$ zameni statistikom:

$$t_1 = \left\{(\overline{X} - \mu) + \frac{\mu_3}{6\sigma^2 n} + \frac{\mu_3}{3\sigma^4}(\overline{X} - \mu)^2\right\}\left(S^2/n\right)^{-1/2}, \tag{1.8}$$

gde je μ_3 treći centralni momenat.

Parametri μ_3 i σ^2 iz (1.8) se ocenjuju sa uzoračkim trećim centralnim momentom $\widehat{\mu}_3$ i korigovanom uzoračkom disperzijom S^2. Za testiranje hipoteze $H_0 : \mu = \mu_0$, Johnsonova t-statistika je oblika:

$$t_2 = \left\{(\overline{X} - \mu_0) + \frac{\widehat{\mu}_3}{6S^2 n} + \frac{\widehat{\mu}_3}{3S^4}(\overline{X} - \mu_0)^2\right\}\left(S^2/n\right)^{-1/2}.$$

Sutton, 1993 je, koristeći Monte Karlo analizu za asimetrične raspodele (Eksponencijalnu, Gama, χ^2, Weibullovu i Lognormalnu) proučavao Studentov t-test i Johnsonov test. Zaključio je da je u mnogim slučajevima povećanje verovatnoće pokrivanja Jonsonovog testa u odnosu na t-test veliko. Takođe je potvrdio da se Johnsonova procedura može koristiti umesto t-testa kod jednostranih testova, kada je polazna raspodela asimetrična, dok je procedura neadekvatna kada je asimetrija velika i veličina uzorka mala. Zbog slabe verovatnoće pokrivanja testa predložen je kompozitni test, čija se kritična oblast sastoji od unije kritične oblasti Johnsonovog testa i dva *bootstrap* testa zasnovana na Johnsonovoj modifikovanoj t-statistici. Pokazano je da se kompozitni test bolje ponaša od Jonsonovog testa u mnogim slučajevima. Monte Karlo analiza je pokazala da je oblast odbacivanja najbolja za $\alpha = 0,01$ i $\alpha = 0,05$ (pogledati Noreen, 1989).

Sutton je istraživao mnoge metode kako bi unapredio Jonsonovu proceduru. Predložio je odbacivanje hipoteze $H_0 : \mu = \mu_0$ naspram $H_1 : \mu > \mu_0$ sa nivoom značajnosti α, ako je i $t_2 \geq t_{n-1,\alpha}$ i $t_2/S_{t_2,b_s} \geq z_\alpha$, ili $\{(n-1)/(n-3)\}^{1/2} t_2/S_{t_2,b_s} \geq t_{n-1,\alpha}$, gde je S_{t_2,b_s} *bootstrap* standardna devijacija statistike t_2.

Prilikom testiranja $H_0 : \mu = \mu_0$ naspram $H_1 : \mu > \mu_0$, može se koristiti i sledeći Edgeworthov razvoj statistike $\sqrt{n}(\overline{X} - \mu)/S$:

$$P\left\{\frac{\sqrt{n}(\overline{X} - \mu)}{S} \leq x - \frac{1}{6\sqrt{n}}\widehat{\beta}_1(1 + 2x^2)\right\} = \Phi(x) + o(n^{-1/2}),$$

gde su:

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}), \\ \widehat{\beta}_1 &= \sum_{i=1}^{n}(X_i - \overline{X})^3/(n/S^3). \end{aligned}$$

Oblast odbacivanja H_0 je:

$$\sqrt{n}(\overline{X} - \mu_0)/S > z_\alpha - \frac{1}{6\sqrt{n}}\widehat{\beta}_1(2z_\alpha^2 + 1),$$

gde z_α zadovoljava: $\Phi(z_\alpha) = 1 - \alpha$.

Kako kritična tačka prethodne oblasti odlučivanja zavisi od koeficijenta asimetrije, predložena je sledeća test statistika:

$$t_3 = t + a(1 + 2t^2) + 4a^2(t + 2t^3),$$

za $a = \widehat{\beta}_1/(6\sqrt{n})$ i $t = \sqrt{n}(\overline{X} - \mu_0)/S$, koja se može koristiti i u slučaju kada je veličina uzorka mala.

Tako je nova oblast odbacivanja: $t_3 > z_\alpha$, za testiranje $H_0 : \mu = \mu_0$ naspram $H_1 : \mu > \mu_0$, za sredinu pozitivno asimetrične raspodele, sa blago visokim ili visokim koeficijentom asimetrije i kada je veličina uzorka mala. Monte Karlo simulacije su pokazale da, kada je veličina uzorka mala, t_3 test je mnogo precizniji i moćniji od Johnsonovog testa, čak i u slučajevima kada je raspodela manje asimetrična od eksponencijalne (pogledati Chen, 1995). Treba primetiti da je statistika t_3 modifikacija standardne t statistike i Johnsonove t_2 statistike. Simulacije su takođe pokazale da kada polazna raspodela ima veoma visok pozitivan koeficijent asimetrije, Jonsonov test i Suttonova modifikacija mogu imati veliku grešku prve vrste, dok se t_3 test pokazao dobrim. On je i jednostavniji za primenu od Suttonove procedure.

1.1.3 Problem dva uzorka. Edgeworthov razvoj *t*-statistike

Ako su disperzije datih uzoraka iz dve populacije sa normalnim raspodelama poznate, odgovarajući dvostrani interval poverenja za razliku sredina $\mu_1 - \mu_2$ je oblika:

$$\Big(\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\Big),$$

gde su n_1 i n_2 veličine uzoraka, dok se jednostrani intervali poverenja dobijaju iz jednakosti:

$$P\Big(\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_{1-\alpha}\Big) = 1 - \alpha,$$

$$P\Big(\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_\alpha\Big) = 1 - \alpha,$$

gde je $\overline{X}_1$ sredina prvog uzorka $\{X_{11}, ..., X_{1n_1}\}$ i $\overline{X}_2$ sredina dugog uzorka $\{X_{21}, ..., X_{2n_2}\}$. U slučaju kada su disperzije nepoznate a uzorci velike veličine ($n_1 \geq 30, n_2 \geq 30$), odgovarajući interval poverenja baziran na centralnoj graničnoj teoremi, za razliku sredina $\mu_1 - \mu_2$ je oblika:

$$\Big(\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\Big), \tag{1.9}$$

gde su S_1^2 i S_2^2 odgovarajuće korigovane uzoračke disperzije, dok ako je veličina jednog ili oba uzorka mala, disperzije nisu poznate, tada se pretpostavlja da su jednake, i koristi se interval:

$$\begin{aligned}\Big(\overline{X}_1 &- \overline{X}_2 - S_{n_1,n_2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{n_1+n_2-2,\alpha/2}, \\ \overline{X}_1 &- \overline{X}_2 + S_{n_1,n_2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{n_1+n_2-2,\alpha/2}\Big),\end{aligned}$$

gde je:

$$S_{n_1,n_2}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2},$$

i $t_{n_1+n_2-2,\alpha/2}$ je odgovarajući percentil Studentove raspodele, tj.

$$P\Big\{ - t_{n_1+n_2-2,\alpha/2} < t_{n_1+n_2-2} < t_{n_1+n_2-2,\alpha/2}\Big\} = 1 - \alpha,$$

gde je:

$$t_{n_1+n_2-2} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_{n_1,n_2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Jednostrani $(1-\alpha)100\%$ intervali nalaze se iz jednakosti:

$$\begin{aligned}P(t_{n_1+n_2-2} > -t_{n_1+n_2-2,\alpha}) &= 1 - \alpha, \\ P(t_{n_1+n_2-2} < t_{n_1+n_2-2,\alpha}) &= 1 - \alpha.\end{aligned}$$

Analogno problemu jednog uzorka, sada za problem dva uzorka standardna t-statistika je:

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \tag{1.10}$$

gde su n_1 i n_2 veličine uzoraka. Odgovarajući t-interval poverenja za razliku sredina $\mu_1 - \mu_2$, već poznat u literaturi, je:

$$\left(\overline{X}_1 - \overline{X}_2 - t_{\nu,\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \overline{X}_1 - \overline{X}_2 + t_{\nu,\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right), \qquad (1.11)$$

za $\nu = n_1 + n_2 - 2$, a odgovarajući interval poverenja za velike uzorke već spomenut, je (1.9).

Za rešavenje problema dva uzorka takođe se mogu koristiti prethodno opisana tri pristupa. Korišćenje centralne granične teoreme za dobijanje intervala poverenja (1.9) jeste prvi pristup. Transformacije posmatranih podataka i korišćenje sredina transformisanih podataka, kako bi se umanjio efekat asimetrije, jeste drugi mogući pristup. Korišćenje standardnih neparametarskih metoda, kao što je Wilcoxon test, jeste treći pristup, slično problemu jednog uzorka.

Nedostaci svakog pristupa ponaosob su sledeći. Simulacije pokazuju da pokrivenost intervala poverenja datog u (1.11) zavisi od relativnog koeficijenta asimetrije dva uzorka (pogledati Zhou i Dinh, 2005). Drugi pristup, testiranje razlike sredina na transformisanoj skali, nije uvek isto sa testiranjem na originalnoj skali, zato pristup transformacije uzoraka može biti neodgovarajući (pogledati Zhou et al., 1997). Treći pristup, korišćenje Wilcoxon testa, ima nedostatak što je to test za jednakost raspodela, nije test za jednakost sredina, osim u slučaju kada posmatrane raspodele imaju isti oblik.

U ovom odeljku biće analizirane transformacije T_1, T_2 i T_3, predstavljene za problem jednog uzorka, sada za problem dva uzorka. U cilju dobijanja bolje pokrivenosti, u slučaju asimetričnih raspodela, interval poverenja baziran na t-statistici može se modifikovati. Neka su X_{11}, X_{12} , ... , X_{1n_1} i X_{21}, X_{22}, ... , X_{2n_2} i.i.d. slučajne veličine redom iz raspodele F, sa sredinom μ_1, disperzijom σ_1^2, koeficijentom asimetrije γ_1 i iz raspodele G, sa sredinom μ_2, disperzijom σ_2^2 i koeficijentom asimetrije γ_2. Neka su $\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ i $S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2$, za $i = 1, 2$.

Teorema 1.1.1. (Zhou i Dinh, 2005) *Neka je* $\lambda_N = n_1/(n_1 + n_2) = n_1/N$. *Pretpostavlja se da je* $\lambda_N = \lambda + O(N^{-r})$ *za neko* $r \geq 0$. *Za regularne uslove* *(videti Hall,1992a), raspodela* t-*statistike date u (1.10) ima sledeći razvoj:*

$$
\begin{aligned}
P(T \leq x) &= P(N^{1/2}U \leq x) \\
&= \Phi(x) + \frac{1}{\sqrt{N}}\frac{A}{6}(2x^2+1)\phi(x) + O(N^{-min(1,r+1/2)}),
\end{aligned}
\tag{1.12}
$$

gde je:

$$A = \Big(\frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1-\lambda}\Big)^{-3/2}\Big(\frac{\sigma_1^3\gamma_1}{\lambda^2} - \frac{\sigma_2^3\gamma_2}{(1-\lambda)^2}\Big).$$

Slično problemu jednog uzorka za $a = 1/3$, $b = 1/6$ i $\gamma = A$, mogu se definisati tri transformacije T_i, $i = 1, 2, 3$, date sa (1.3), (1.4) i (1.5), redom. Ovde se mogu dati tri intervala poverenja za razliku sredina $\mu_1 - \mu_2$, bazirana na transformacijama, na sledeći način. Neka je $\widehat{\sigma} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$. Tada je $(1-\alpha)100\%$ interval poverenja za razliku sredina $\mu_1 - \mu_2$ dat sa:

$$\overline{X}_1 - \overline{X}_2 - N^{1/2}T_i^{-1}(N^{-1/2}z_{\alpha/2})\widehat{\sigma} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 - N^{1/2}T_i^{-1}(N^{-1/2}z_{1-\alpha/2})\widehat{\sigma},$$

gde su $T_i^{-1}(\cdot)$ inverzne funkcije funkcija $T_i(\cdot)$, koje se mogu analitički rešiti:

$$
\begin{aligned}
T_1^{-1}(t) &= \Big(\frac{1}{3}\widehat{A}\Big)^{-1}\Big\{1 + 3\frac{1}{3}\widehat{A}\Big(t - N^{-1}\frac{1}{6}\widehat{A}\Big)\Big\}^{1/3} - \Big(\frac{1}{3}\widehat{A}\Big)^{-1}, \\
T_2^{-1}(t) &= \Big(2\frac{1}{3}N^{-1/2}\widehat{A}\Big)^{-1}\ln\Big\{2\frac{1}{3}N^{-1/2}\widehat{A}\Big(t - N^{-1}\frac{1}{6}\widehat{A}\Big) + 1\Big\}, \\
T_3^{-1}(t) &= \Big\{1 + 3\Big(t - N^{-1}\frac{1}{6}\widehat{A}\Big)\Big\}^{1/3} - 1.
\end{aligned}
$$

Ovde je $\widehat{A}$ ocena koeficijenta A, za koji je:

$$\widehat{A} = \frac{(N/n_1)^2S_1^3\widehat{\gamma}_1 - (N/n_2)^2S_2^3\widehat{\gamma}_2}{\{(N/n_1)S_1^2 + (N/n_2)S_2^2\}^{3/2}},$$

gde je za $i = 1, 2$:

$$\widehat{\gamma}_i = \frac{n_i}{(n_i-1)(n_i-2)}\sum_{j=1}^{n_i}\Big(\frac{X_{ij} - \overline{X}_i}{S_i}\Big)^3,$$

Za konstrukciju intervala poverenja pomoću normalne aproksimacije, koeficijent $A/\sqrt{N}$ (kao i $\gamma/\sqrt{n}$, za problem jednog uzorka) ima važnu ulogu, što je pokazano u radu Zhou i Dinh, 2005. Uobičajeni t-interval pokazuje dobre rezultate u situaciji kada je $\widehat{A}/\sqrt{N}$ ($\widehat{\gamma}/\sqrt{n}$) manje od $0,3$. Ako to nije slučaj, t-interval se može unaprediti koristeći *bootstrap-t* metod ili T_1 ili T_3 transformacije. U Zhou i Dinh, 2005 je pokazano da je u slučaju asimetričnih raspodela najbolja pokrivenost dobijena pomoću *bootstrap-t* intervala. U smislu dužine intervala najbolje rezultate daje T_3 transformacija i preporučuje se za asimetrične raspodele. Simulacije su takođe pokazale da su transformisani intervali bolji kada je koeficijent A pozitivan.

1.1.4 Testiranje razlike sredina dve raspodele, problem dva uzorka

Problem dva uzorka je dosta prisutan u statističkoj praksi. Najčešća pretpostavka o dva uzorka je da oni dolaze iz nezavisnih populacija, sa normalnom raspodelom, sa različitim sredinama. U većini slučajeva su disperzije nepoznate i moraju se oceniti. Kada su ispunjene pretpostavke o jednakim disperzijama koristi se *pooled* ocena zajedničke disperzije i test statistika ima t raspodelu. Kada su različite disperzije, dve disperzije se ocenjuju odvojeno i test statistika ne mora imati tačno t raspodelu.

Neka su $X_{11}, ..., X_{1n_1} \sim N(\mu_1, \sigma_1^2)$ i $X_{21}, ..., X_{2n_2} \sim N(\mu_2, \sigma_2^2)$. Dalje će biti testirana razlika sredina dve nezavisne, normalne raspodele, tj. testira se hipoteza $H_0 : \mu_1 - \mu_2 = \triangle_0 = 0$.

Slučaji kada su obe disperzije poznate ili obe nepoznate

Jasno je da je najjednostavniji slučaj kada su obe disperzije poznate. Tada se koristi z-statistika koja je oblika:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - \triangle_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Pod pretpostavkom hipoteze H_0, Z ima $N(0,1)$ raspodelu.

Naravno, komplikovaniji slučaj nastaje kada su disperzije nepoznate. Tada z-test nije adekvatan za upotrebu, iz razloga što disperzije moraju biti ocenjene. Ako se pretpostavi da su disperzije jednake, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, za ocenu

σ^2 se kombinuju dva uzorka. Test statistika je:

$$Z_1 = \frac{\overline{X}_1 - \overline{X}_2 - \triangle_0}{\sqrt{S^2_{n_1,n_2}(\frac{1}{n_1} + \frac{1}{n_2})}},$$

koja ima tačno t raspodelu sa $\nu_1 = n_1 + n_2 - 2$ stepeni slobode, pod pretpostavkom hipoteze H_0, gde je $S^2_{n_1,n_2}$ jedinstvena ocena disperzije σ^2, data formulom:

$$S^2_{n_1,n_2} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2},$$

gde su S_1^2 i S_2^2 korigovane uzoračke disperzije oba uzorka redom.

Dosta realniji slučaj je kada su dve disperzije različite, tj. $\sigma_1^2 \neq \sigma_2^2$. Ovo je u literaturi poznato kao Behrens-Fisherov problem. U ovom slučaju test statistika je:

$$Z_2 = \frac{\overline{X}_1 - \overline{X}_2 - \triangle_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

koja nema tačno t raspodelu. Kako $(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})^{-1}(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})$ nema odgovarajuću χ^2 raspodelu, to je problem teži za rešavanje. Satterthwaite, 1941, 1946 je predložio metod koji aproksimira tačnu raspodelu sa odgovarajućom χ^2 raspodelom i koristeći tu aproksimaciju dobija se:

$$\nu_2\Big(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\Big)^{-1}\Big(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\Big) \sim \chi^2_{\nu_2}.$$

Aproksimativna raspodela statistike Z_2 je t raspodela sa $\widehat{\nu}_2$ stepeni slobode, gde je:

$$\widehat{\nu}_2 = \Big(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\Big)^2\Big(\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}\Big)^{-1}.$$

Kako je $\nu_2 = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2\left(\frac{(\sigma_1^2/n_1)^2}{n_1-1} + \frac{(\sigma_2^2/n_2)^2}{n_2-1}\right)^{-1}$, ν_2 se ocenjuje sa $\widehat{\nu}_2$.

Slučaj kada je količnik disperzija poznat

U ovom slučaju pretpostavlja se da je količnik disperzija poznat, tj. $\sigma_2^2/\sigma_1^2 = c$, za $c \neq 1$. Kako statistika Z_2 uopšte ne koristi informaciju da je količnik disperzija poznat, treba iskoristiti drugu statistiku. Može se primetiti da sledeća statistika ima standardnu normalnu raspodelu:

$$Z_3 = \frac{\overline{X}_1 - \overline{X}_2 - \triangle_0}{\sigma_1\sqrt{\frac{1}{n_1} + \frac{c}{n_2}}}.$$

Jedini parametar koji treba oceniti u statistici Z_3 je σ_1. Ako se iskoristi transformacija $\widetilde{X}_{2i} := X_{2i}/c^{1/2}$, za $i = 1, 2, ..., n_2$, tada važi da $\widetilde{X}_{21}, ..., \widetilde{X}_{2n_2} \sim N(\mu_2/c^{1/2}, \sigma_1^2)$. Odgovarajuća test statistika je oblika:

$$Z_4 = \frac{\overline{X}_1 - \overline{X}_2 - \triangle_0}{\widetilde{S}_{n_1,n_2}\sqrt{\frac{1}{n_1} + \frac{c}{n_2}}}, \tag{1.13}$$

gde je $\widetilde{S}^2_{n_1,n_2}$ uzoračka jedinstvena ocena disperzije za X_{1i}, $i = 1, ..., n_1$ i $\widetilde{X}_{2j}$, $j = 1, ..., n_2$. Može se pokazati da ta statistika ima tačno t raspodelu sa $(n_1 + n_2 - 2)$ stepeni slobode. Analiza je sledeća. Statistika Z_3 ima standardnu normalnu raspodelu, $(n_1 - 1)S_1^2/\sigma_1^2$ ima χ^2 raspodelu sa $(n_1 - 1)$ stepeni slobode i $(n_2 - 1)\widetilde{S}_2^2/\sigma_1^2$ ima χ^2 raspodelu sa $(n_2 - 1)$ stepeni slobode, gde je $\widetilde{S}_2^2$ korigovana uzoračka disperzija za $\widetilde{X}_{2j}$, $j = 1, ..., n_2$. Kako su dve χ^2 slučajne veličine nezavisne, važi:

$$\begin{aligned} \chi^2 &= \frac{(n_1 - 1)S_1^2}{\sigma_1^2} + \frac{(n_2 - 1)\widetilde{S}_2^2}{\sigma_1^2} \\ &= \frac{(n_1 - 1)S_1^2}{\sigma_1^2} + \frac{(n_2 - 1)S_2^2}{c\sigma_1^2} \sim \chi^2_{n_1+n_2-2}. \end{aligned} \tag{1.14}$$

Iz (1.13) i (1.14) dobija se test statistika (pogledati Sprott, 1993):

$$Z_4 = \frac{Z_3}{[\chi^2/(n_1 + n_2 - 2)]^{1/2}} \sim t_{n_1+n_2-2}.$$

Za test statistiku Z_4 odgovarajuća pravila odlučivanja su sledeća:

- Ako je alternativna hipoteza $H_1 : \mu_1 - \mu_2 \neq \triangle_0$, onda je oblast odbacivanja $Z_4 > t_{n_1+n_2-2,\alpha/2}$ ili $Z_4 < -t_{n_1+n_2-2,\alpha/2}$,
- Ako je alternativna hipoteza $H_1 : \mu_1 - \mu_2 > \triangle_0$, onda je oblast odbacivanja $Z_4 > t_{n_1+n_2-2,\alpha}$,
- Ako je alternativna hipoteza $H_1 : \mu_1 - \mu_2 < \triangle_0$, onda je oblast odbacivanja $Z_4 < -t_{n_1+n_2-2,\alpha}$,

gde je odgovarajuća vrednost $t_{n_1+n_2-2,\alpha}$ uzeta iz tablice Studentove raspodele.

1.2 Intervali poverenja za disperziju

Za intervalno ocenjivanje disperzije, u slučaju jednog uzorka, postojeće metode su zasnovane na χ^2-statistici (videti Pearson, 1900), a u slučaju dva uzorka, kada se ocenjuje količnik disperzija, postojeće metode se baziraju na F-statistici (videti Snedecor, 1934).

U ovom odeljku biće dati intervali poverenja za nepoznatu disperziju, razliku disperzija i količnik disperzija, baziranih na uobičajenoj t-statistici i metodima ponovljenih uzoraka.

Moguće je konstruisati *bootstrap-t* interval bazirano na t-statistici, koji ima kraću dužinu intervala (pogledati Beran, 1987, Bickel, 1992, Efron i Tibshirani, 1993, Hall, 1988, Liu i Singh, 1987). Posle korekcije t-statistike moguće je ukloniti efekat asimetrije i dobiti najbolju verovatnoću pokrivanja. Takođe će biti sugerisane neke metode transformacija koje su dobijene Edgewortovim razvojem t-statistike.

1.2.1 Intervali poverenja za disperziju, problem jednog uzorka

Neka su $X_1, ..., X_n$ i.i.d. iz raspodele F, sa sredinom μ i disperzijom σ^2, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ korigovana uzoračka disperzija i $\overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. U slučaju kada je μ poznato, kako statistika $\frac{n\overline{S}^2}{\sigma^2}$ ima χ^2 raspodelu sa n stepeni slobode, dvostrani $(1-\alpha)100\%$ interval poverenja za disperziju je:

$$\Big(\frac{n\overline{S}^2}{\chi^2_{n,\alpha/2}}, \frac{n\overline{S}^2}{\chi^2_{n,1-\alpha/2}}\Big),$$

dok je važniji jednostrani interval poverenja oblika:

$$\Big[0, \frac{n\overline{S}^2}{\chi^2_{n,1-\alpha}}\Big),$$

gde je $\chi^2_{n,\alpha}$ odgovarajući kvantil χ^2 raspodele sa n stepeni slobode, za koji važi $P(\chi^2_n \geq \chi^2_{n,\alpha}) = \alpha$. U slučaju kada je μ nepoznato, kako statistika $\frac{(n-1)S^2}{\sigma^2}$ ima χ^2 raspodelu sa $(n-1)$ stepeni slobode, dvostrani $(1-\alpha)100\%$

interval poverenja za disperziju je:

$$\Big(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2,}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2,}}\Big),$$

dok je jednostrani interval poverenja oblika:

$$\Big[0, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha,}}\Big).$$

Intervali poverenja za disperziju su takođe analizirani u radu Ćojbašić i Tomović, 2007. Ovde se navode osnovne ideje iz tog rada.

Neka su $X_1, ..., X_n$ i.i.d. iz raspodele F, sa sredinom μ i disperzijom σ^2. Ako je $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ korigovana uzoračka disperzija, poznato je da statistika $(n-1)S^2/\sigma^2$ ima χ^2 raspodelu sa $(n-1)$ stepeni slobode. Moguće je konstruisati interval poverenja za disperziju koristeći prethodnu statistiku. Osnovna ideja je da se χ^2 raspodela sa $(n-1)$ stepeni slobode može aproksimirati normalnom raspodelom sa parametrima $n-1$ i $2(n-1)$. Sledi, da za dovoljno veliko n, raspodela standardizovane promenljive:

$$Z = \frac{(n-1)S^2/\sigma^2 - (n-1)}{\sqrt{2(n-1)}} = \frac{S^2 - \sigma^2}{\sqrt{var(S^2)}}$$

konvergira ka standardnoj normalnoj raspodeli. Za konstrukciju jednostranog *bootstrap-t* intervala poverenja za disperziju σ^2 razmatrana je statistika:

$$T = \frac{S^2 - \sigma^2}{\sqrt{\widehat{var}(S^2)}}, \tag{1.15}$$

gde je $\widehat{var}(S^2)$ konzistentna ocena disperzije S^2.

$(1-\alpha)100\%$ jednostrani *bootstrap-t* interval za nepoznati parametar σ^2 je oblika:

$$I_{boot} = (0, S^2 - \widehat{t}^{(\alpha)}\sqrt{\widehat{var}(S^2)}),$$

gde je $\widehat{t}^{(\alpha)}$ α percentil statistike T^*, date sa:

$$T^* = \frac{S^{2^*} - \sigma^2}{\sqrt{\widehat{var}(S^{2^*})}},$$

gde je S^{2^*} *bootstrap* replika statistike S^2.

t-statistika je od izuzetne važnosti, jer ju je moguće modifikovati da bi se uklonio efekat asimetrije i dobila najbolja verovatnoća pokrivanja. Transformacija se bazira na Edgeworthovom razvoju raspodele t-statistike. Dalje je dat Edgeworthov razvoj statistike (1.15). Koristiće se uslov Kramera, koji znači da je $\limsup_{|t|\to\infty} |\psi(t)| < 1$, gde je $\psi(t)$ odgovarajuća karakteristična funkcija. Ovaj uslov važi ukoliko je raspodela odgovarajuće slučajne veličine nesingularna, što znači da slučajna veličina ima odgovarajuću funkciju gustine.

Teorema 1.2.1. (Ćojbašić i Tomović, 2007) *Ako je zadovoljen uslov Kramera (pogledati Hall, 1992b) i ako je $E(X_i^6) < +\infty$, tada raspodela* t-*statistike date u jednakosti (1.15) ima sledeći razvoj:*

$$P(T \leq x) = \Phi(x) + \frac{1}{\sqrt{n}} q(x)\phi(x) + O(n^{-1}),$$

gde je $q(x) = \frac{M_3'}{6}(2x^2+1)$ i $M_3' = E((1/n)\sum_{i=1}^n X_i'^3)$, gde su:

$$X_i' = \frac{(X_i - \overline{X})^2 - ((n-1)/n)\sigma^2}{\sqrt{V}}, \tag{1.16}$$

$i = 1, 2, ..., n$, za $V = E((X_i - \overline{X})^2 - ((n-1)/n)\sigma^2)^2$.

Dalje su istaknuti samo najvažniji koraci dokaza, dok se u navedenom radu može pogledati ceo dokaz.

Dokaz: Neka su slučajne promenljive X_i' date izrazom (1.16). Tada se test statistika (1.15) može zapisati u obliku $T = \sqrt{n-1}(Y_1/\sqrt{Y_2 - Y_1^2})$, gde su $Y_1 = (1/n)\sum_{i=1}^n X_i'$ i $Y_2 = (1/n)\sum_{i=1}^n X_i'^2$. Definisana je statistika W_n:

$$\begin{aligned} W_n &= \sqrt{n-1}\Big(\frac{\partial g}{\partial Y_1}(U)(Y_1 - U_1) + \frac{\partial g}{\partial Y_2}(U)(Y_2 - U_2) \\ &+ \frac{1}{2}\Big[\frac{\partial^2 g}{\partial Y_1^2}(U)(Y_1 - U_1)^2 + 2\frac{\partial^2 g}{\partial Y_1 Y_2}(U)(Y_1 - U_1)(Y_2 - U_2) \\ &+ \frac{\partial^2 g}{\partial Y_2^2}(U)(Y_2 - U_2)^2\Big]\Big), \end{aligned}$$

gde je $g(Y) = Y_1/\sqrt{Y_2 - Y_1^2}$ i $E(Y_1, Y_2) \equiv U \equiv (U_1, U_2) = (0, 1)$. Dalje, za statistiku

$$W_n = \sqrt{n-1}(\frac{3}{2}Y_1 - \frac{1}{2}Y_1 Y_2)$$

je pokazano da važi:

$$
\begin{aligned}
EW_n &= -\frac{1}{2}\frac{1}{\sqrt{n}}M_3' + O(n^{-1}),\\
EW_n^2 &= 1 + O(n^{-1}),\\
EW_n^3 &= -\frac{7}{2}\frac{1}{\sqrt{n}}M_3' + O(n^{-1}),
\end{aligned}
$$

i da je njena karakteristična funkcija:

$$\psi_n(t) = \exp\Big(-\frac{t^2}{2}\Big)\Big\{1 + \frac{1}{\sqrt{n}}r_1(it) + O(n^{-1})\Big\},$$

gde je:

$$r_1(it) = -\frac{1}{2}M_3'(it) - \frac{2}{6}M_3'(it)^3.$$

Kako je $\psi_n(t) = \int_{-\infty}^{+\infty} e^{itx}dP(W_n \le x)$ i kako za standardnu normalnu raspodelu važi: $e^{-t^2/2} = \int_{-\infty}^{+\infty} e^{itx}d\Phi(x)$, sledi da je $T = W_n + O(n^{-1})$, odakle rezultat sledi. ∎

U cilju eliminisanja asimetrije raspodele Ćojbašić i Tomović, 2007 su iskoristili Hallovu ideju (Hall, 1992b) i analizirali su transformaciju t-statistike koja ima oblik:

$$T_1(U) = U + \frac{1}{3}\widehat{M_3'}U^2 + \frac{1}{27}\widehat{M_3'}^2U^3 + \frac{1}{6}\frac{1}{n}\widehat{M_3'},$$

gde je $\widehat{M_3'}$ ocena momenta M_3'.

Takođe su predložili novu transformaciju koja ima oblik:

$$T_2(U) = U + \frac{1}{3}\widehat{M_3'}U^2 + U^3 + \frac{1}{6}\frac{1}{n}\widehat{M_3'}.$$

Transformisani $(1-\alpha)100\%$ intervali poverenja za disperziju σ^2, bazirani na transformacijama T_1 i T_2, su oblika:

$$I_{T_i} = \Big(0, S^2 - \sqrt{n}T_i^{-1}\Big(\frac{\widehat{t^{(\alpha)}}}{\sqrt{n}}\Big)\sqrt{\widehat{var}(S^2)}\Big),$$

gde su $T_i^{-1}(\cdot)$, $i = 1,2$, inverzne funkcije funkcija $T_i(\cdot)$, oblika:

$$T_1^{-1}(t) = \Big(\frac{1}{3}\widehat{M_3'}\Big)^{-1}\Big\{1 + 3\frac{1}{3}\widehat{M_3'}\Big(t - \frac{1}{6}\frac{1}{n}\widehat{M_3'}\Big)\Big\}^{1/3} - \Big(\frac{1}{3}\widehat{M_3'}\Big)^{-1},$$

$T_2^{-1}(t)$ je jednako rešenju jednačine $T_2(U) = t$, ako ona ima jedno rešenje, a uzima se najmanje rešenje jednačine, ako ona ima tri realna rešenja.

1.2.2 Intervali poverenja za razliku disperzija, problem dva uzorka

Neka su $X_{11}, X_{12}, ..., X_{1n_1}$ i $X_{21}, X_{22}, ..., X_{2n_2}$ i.i.d., redom iz raspodele F, sa sredinom μ_1 i disperzijom σ_1^2, i raspodele G, sa sredinom μ_2 i disperzijom σ_2^2. Ako su korigovane uzoračke disperzije, $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2$, $S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2$ i uzoračke sredine, $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$, $\overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$, iz prethodnog odeljka sledi da slučajne veličine:

$$\begin{aligned} Z_1 &= \frac{S_1^2 - \sigma_1^2}{\sqrt{var(S_1^2)}}, \\ Z_2 &= \frac{S_2^2 - \sigma_2^2}{\sqrt{var(S_2^2)}}, \end{aligned}$$

konvergiraju ka standardnoj normalnoj raspodeli. Tada, razlika korigovanih uzoračkih disperzija ima aproksimativno normalnu raspodelu: $S_1^2 - S_2^2 \sim N(\sigma_1^2 - \sigma_2^2, var(S_1^2) + var(S_2^2))$. Ćojbašić i Tomović, 2007 su razmatrali *bootstrap-t* i transformisane *bootstrap-t* intervale poverenja za razliku disperzija, bazirane na statistici:

$$T = \frac{S_1^2 - S_2^2 - (\sigma_1^2 - \sigma_2^2)}{\sqrt{\widehat{var}(S_1^2) + \widehat{var}(S_2^2)}}. \tag{1.17}$$

Takođe su dali Edgeworthov razvoj t-statistike (1.17), o čemu govori sledeća teorema.

Teorema 1.2.2. (Ćojbašić i Tomović, 2007) *Neka je $\lambda_N = n_1/N = n_1/(n_1 + n_2)$ i pretpostavlja se da je $\lambda_N = \lambda + O(N^{-r})$ za neko $r \geq 0$. Ako je ispunjen uslov Kramera (pogledati Hall, 1992a) i ako je $E(X_{ij}^6) < +\infty$, raspodela* t*-statistike date u jednakosti (1.17) ima sledeći razvoj:*

$$P(T \leq x) = \Phi(x) + \frac{1}{\sqrt{N}} q(x)\phi(x) + O\Big(\frac{1}{N^{-min(1,r+1/2)}}\Big),$$

gde je funkcija $q(x)$:

$$q(x) = \frac{A}{6}(2x^2 + 1),$$

za:

$$A = \frac{V_1^{3/2} M_{31}'/\lambda^2 - V_2^{3/2} M_{32}'/(1-\lambda)^2}{(V_1/\lambda + V_2/(1-\lambda))^{3/2}},$$

gde su $M'_{31} = E((1/n_1)\sum_{j=1}^{n_1} X'^3_{1j})$, $M'_{32} = E((1/n_2)\sum_{j=1}^{n_2} X'^3_{2j})$, $X'_{ij} = \frac{(X_{ij}-\overline{X}_i)^2-((n_i-1)/n_i)\sigma_i^2}{\sqrt{V_i}}$, $V_i = E((X_{ij}-\overline{X}_i)^2 - ((n_i-1)/n_i)\sigma_i^2)^2$, $i = 1,2, j = 1,2,...,n_i$.

Treba istaći da su Ćojbašić i Tomović, 2007 uočili da se statistika (1.17) može napisati kao:

$$T = \sqrt{N}\frac{Y_1\sqrt{V_1} - Y_3\sqrt{V_2}}{\sqrt{(V_1/\lambda_N)(Y_2 - Y_1^2) + (V_2/(1-\lambda_N))(Y_4 - Y_3^2)}},$$

gde je:

$$Y_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X'_{1j},\ Y_2 = \frac{1}{n_1}\sum_{j=1}^{n_1} X'^2_{1j},$$
$$Y_3 = \frac{1}{n_2}\sum_{j=1}^{n_2} X'_{2j},\ Y_4 = \frac{1}{n_2}\sum_{j=1}^{n_2} X'^2_{2j},$$

i $\lambda_N = n_1/N = n_1/(n_1+n_2)$. Posle navedenog definisanja promenljivih dokaz je analogan dokazu prethodne teoreme (pogledati takođe Zhou i Dinh, 2005). Kompletan dokaz može se naći u Appendix A, rada Ćojbašić i Tomović, 2007.

Dalje su dati dvostrani intervali poverenja za razliku disperzija, korišćenjem transformacija T_1 i T_2, iz prethodnog odeljka. Posle generisanja B *bootsrap* uzoraka, za svaki se može izračunati vrednost statistike:

$$T^* = \frac{S_1^{2^*} - S_2^{2^*} - (S_1^2 - S_2^2)}{\sqrt{\widehat{var}(S_1^{2^*}) + \widehat{var}(S_2^{2^*})}},$$

gde je $S_i^{2^*}$ (i=1,2) *bootstrap* replika statistike S_i^2 i $\widehat{var}(S_i^{2^*})$ (i=1,2) je ocena disperzije statistike $S_i^{2^*}$ (i=1,2) u *bootstrap* uzorku.

$(1-2\alpha)100\%$ dvostrani *bootstrap-t* interval poverenja za razliku disperzija $\sigma_1^2 - \sigma_2^2$ je:

$$I_{boot} = \Big(S_1^2 - S_2^2 - \widehat{t}^{(1-\alpha)}\sqrt{\widehat{var}(S_1^2) + \widehat{var}(S_2^2)}, S_1^2 - S_2^2 - \widehat{t}^{(\alpha)}\sqrt{\widehat{var}(S_1^2) + \widehat{var}(S_2^2)}\Big),$$

gde su $\widehat{t}^{(1-\alpha)}$ i $\widehat{t}^{(\alpha)}$ redom, $1-\alpha$ i α percentili statistike T^*.

$(1-2\alpha)100\%$ T_1 interval poverenja, dobijen korišćenjem Hallove transformacije, je:

$$I_{T_1} = \Big(S_1^2 - S_2^2 - \sqrt{N}T_1^{-1}\Big(\frac{\widehat{t}^{(1-\alpha)}}{\sqrt{N}}\Big)\widehat{se}, S_1^2 - S_2^2 - \sqrt{N}T_1^{-1}\Big(\frac{\widehat{t}^{(\alpha)}}{\sqrt{N}}\Big)\widehat{se}\Big),$$

gde je:

$$T_1^{-1}(t) = \Big(\frac{1}{3}\widehat{A}\Big)^{-1}\Big(1 + 3\frac{1}{3}\widehat{A}\Big(t - \frac{1}{n}\frac{1}{6}\widehat{A}\Big)\Big)^{1/3} - \Big(\frac{1}{3}\widehat{A}\Big)^{-1},$$

$\widehat{A}$ je ocena za A, $\widehat{se} = \sqrt{\widehat{var}(S_1^2) + \widehat{var}(S_2^2)}$ i $N = n_1 + n_2$.

$(1-2\alpha)100\%$ T_2 interval poverenja je oblika:

$$I_{T_2} = \Big(S_1^2 - S_2^2 - \sqrt{N}T_2^{-1}\Big(\frac{\widehat{t}^{(1-\alpha)}}{\sqrt{N}}\Big)\widehat{se}, S_1^2 - S_2^2 - \sqrt{N}T_2^{-1}\Big(\frac{\widehat{t}^{(\alpha)}}{\sqrt{N}}\Big)\widehat{se}\Big).$$

$T_2^{-1}(t)$ je jednako rešenju jednačine $T_2(U) = t$, ako ona ima jedno realno rešenje, a ukoliko jednačina ima tri realna rešenja razmatra se najmanje od svih u slučaju gornjeg ograničenja, i najveće od svih u slučaju donjeg ograničenja.

1.2.3 Pokrivenost intervala poverenja za količnik disperzija

U ovom odeljku biće analizirana pokrivenost intervala poverenja za količnik disperzija i biće istaknute sve prednosti dobijenih intervala na osnovu transformacija, nad postojećim intervalom najčešće korišćenim u literaturi. Analiza pokrivenosti navedenih intervala predstavlja jedan od originalnih rezultata ove disertacije. Postojeće metode za intervalno ocenjivanje količnika disperzija baziraju se na Fišerovoj F-statistici. Ti intervali su jako osetljivi u slučaju kada parametarske pretpostavke nisu ispunjene. Ovde su razmatrani intervali poverenja bazirani na t-statistici, kada se može modifikovati, kako bi se uklonili efekti asimetrije. Zanimljive rezultate o intervalnim ocenama disperzije u slučaju jednog i dva uzorka, koji mogu biti od koristi, dali su Barham i Jeyaratnam, 1999, Abu-Shawiesh et al., 2011, Kittani i Zghoul, 2010, Ćojbašić i Tomović, 2007 i drugi.

Neka su $X_{11}, X_{12}, ..., X_{1n_1}$ i $X_{21}, X_{22}, ..., X_{2n_2}$ i.i.d. iz dve normalne raspodele sa sredinama μ_1 i μ_2 i disperzijama σ_1^2 i σ_2^2, redom. Postojeći metod za intervalno ocenjivanje količnika disperzija bazira se na F-statistici i $(1-2\alpha)100\%$ dvostrani interval poverenja je:

$$(S_1^2/S_2^2 F_{n_1-1,n_2-1;\alpha}, S_1^2/S_2^2 F_{n_1-1,n_2-1;1-\alpha}),$$

gde je $F_{n_1-1,n_2-1;\alpha}$ određeno iz $P(F_{n_1-1,n_2-1} > F_{n_1-1,n_2-1;\alpha}) = \alpha$ i F_{n_1-1,n_2-1} je slučajna veličina koja ima F raspodelu sa (n_1-1) i (n_2-1) stepeni slobode. S_1^2 i S_2^2 su odgovarajuće korigovane uzoračke disperzije.

Raspodela F-statistike može se aproksimirati normalnom raspodelom, što su analizirali mnogi autori. Patel et al., 1976 su analizirali normalnu aproksimaciju za velike n_1 i n_2, Johnson et al., 1995 su razmatrali dve normalne aproksimacije i χ^2 aproksimaciju za veliko n_1 i fiksirano n_2, Ferreira, 2011 je iskoristio činjenicu da se χ^2 može aproksimirati normalnom raspodelom i onda iskoristio χ^2 aproksimaciju F-statistike.

Ako se razmotri statistika:

$$Z = \frac{S_1^2/S_2^2 - E(S_1^2/S_2^2)}{\sqrt{var(S_1^2/S_2^2)}},$$

ona se može napisati u obliku:

$$Z = \frac{S_1^2/S_2^2 - (\sigma_1^2/\sigma_2^2)(n_2+1)/(n_2-1)}{\sqrt{(2\sigma_1^4/\sigma_2^4)((n_1+n_2-2)/((n_1-1)(n_2-1)))}}.$$

Ovaj oblik proizilazi iz formule za aproksimaciju očekivane vrednosti količnika slučajnih veličina, koju je dao Pearson, 1897, a može se naći u radu Grossman i Norton, 1980:

$$E\Big(\frac{X}{Y}\Big) \sim \frac{E(X)}{E(Y)}\Big[1 + \frac{var(Y)}{E^2(Y)} - \frac{cov(X,Y)}{E(X)E(Y)}\Big], Y > 0,$$

i formule za aproksimaciju disperzije količnika dve slučajne veličine, koju su dali Staurt i Ord, 1998:

$$var\Big(\frac{X}{Y}\Big) \sim \frac{E^2(X)}{E^2(Y)}\Big[\frac{var(X)}{E^2(X)} - 2\frac{cov(X,Y)}{E(X)E(Y)} + \frac{var(Y)}{E^2(Y)}\Big].$$

U ovom slučju t-statistika se definiše na sledeći način:

$$T = \frac{S_1^2/S_2^2 - (\sigma_1^2/\sigma_2^2)(n_2+1)/(n_2-1)}{\sqrt{\widehat{var}(S_1^2/S_2^2)}}. \tag{1.18}$$

U sledećoj teoremi dat je Edgeworthov razvoj raspodele T statistike, što je jedan od originalnih rezultata ove disertacije.

Teorema 1.2.3. (Ćojbašić Rajić i Stanojević, 2013) *Ako je zadovoljen uslov Kramera (videti Hall, 1992a) i ako je $E(X_{ij}^6) < +\infty$, raspodela* t*-statistike date u (1.18) ima sledeći razvoj:*

$$P(T \leq x) = \Phi(x) + \frac{1}{\sqrt{n_1}} q(x)\phi(x) + O(n_1^{-1}),$$

gde je:

$$q(x) = \frac{M_3^{'}}{6}(2x^2 + 1),$$

$M_3^{'} = E((1/n_1)\sum_{i=1}^{n_1} X_i^{'3})$, *za slučajne veličine* $X_i^{'}$, $i = 1, 2, ..., n_1$, *definisane na sledeći način:*

$$X_i^{'} = \frac{(X_{1i} - \overline{X}_1)^2 / \sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2 - (n_1 - 1)(n_2 + 1)\sigma_1^2/(n_1(n_2 - 1)^2\sigma_2^2)}{\sqrt{V}},$$

$i = 1, 2, ..., n_1$, *i*

$$V = E\Big(\frac{(X_{1i} - \overline{X}_1)^2}{\sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2} - \frac{(n_1 - 1)(n_2 + 1)\sigma_1^2}{n_1(n_2 - 1)^2\sigma_2^2}\Big)^2$$

je konstanta.

Dokaz: Statistika (1.18) se može napisati:

$$T = \sqrt{n_1 - 1}\frac{Y_1}{\sqrt{Y_2 - Y_1^2}},$$

gde su $Y_1 = (1/n_1)\sum_{i=1}^{n_1} X_i^{'}$ i $Y_2 = (1/n_1)\sum_{i=1}^{n1} X_i^{'2}$. Ako se uvedu oznake: $g(Y) = Y_1/\sqrt{Y_2 - Y_1^2}$ i $E(Y_1, Y_2) \equiv U \equiv (U_1, U_2) = (0, 1)$ i definiše statistika W:

$$\begin{aligned} W &= \sqrt{n_1 - 1}\Big(\frac{\partial g}{\partial Y_1}(U)(Y_1 - U_1) + \frac{\partial g}{\partial Y_2}(U)(Y_2 - U_2) \\ &+ \frac{1}{2}\Big[\frac{\partial^2 g}{\partial Y_1^2}(U)(Y_1 - U_1)^2 + 2\frac{\partial^2 g}{\partial Y_1 Y_2}(U)(Y_1 - U_1)(Y_2 - U_2) \\ &+ \frac{\partial^2 g}{\partial Y_2^2}(U)(Y_2 - U_2)^2\Big]\Big), \end{aligned}$$

onda je:

$$W = \sqrt{n_1 - 1}\Big(\frac{3}{2}Y_1 - \frac{1}{2}Y_1Y_2\Big).$$

Prva tri momenta statistike W su:

$$\begin{aligned}
EW &= -\frac{1}{2}\frac{1}{\sqrt{n_1}}M_3' + O(n_1^{-1}),\\
EW^2 &= 1 + O(n_1^{-1}),\\
EW^3 &= -\frac{7}{2}\frac{1}{\sqrt{n_1}}M_3' + O(n_1^{-1}),
\end{aligned}$$

a njena karakteristična funkcija je:

$$\psi(t) = \exp\Big(-\frac{t^2}{2}\Big)\Big\{1 + \frac{1}{\sqrt{n_1}}r_1(it) + O(n_1^{-1})\Big\},$$

gde je:

$$r_1(it) = -\frac{1}{2}M_3'(it) - \frac{2}{6}M_3'(it)^3.$$

Kako je $\psi(t) = \int_{-\infty}^{+\infty} e^{itx}dP(W \le x)$ i kako za standardnu normalnu raspodelu važi: $e^{-t^2/2} = \int_{-\infty}^{+\infty} e^{itx}d\Phi(x)$, to se statistika T može napisati u obliku: $T = W + O(n_1^{-1})$, odakle sledi tvrđenje teoreme. ∎

Johnson, 1978 i Hall, 1992b su predložili transformacije t-statistike kako bi eliminisali asimetriju raspodele. Hallova transformacija u ovom slučaju ima oblik:

$$g_1(T) = T + \frac{1}{3}\frac{1}{\sqrt{n_1}}\widehat{M_3'}T^2 + \frac{1}{27}\frac{1}{n_1}\widehat{M_3'}^2T^3 + \frac{1}{6}\frac{1}{\sqrt{n_1}}\widehat{M_3'},$$

gde je $\widehat{M_3'}$ ocena za M_3'. Ova transformacija je monotona i ima inverznu funkciju:

$$T = g_1^{-1}(x) = \sqrt{n_1}\Big(\frac{\widehat{M_3'}}{3}\Big)^{-1}\Big[\Big(1 + \widehat{M_3'}\Big(\frac{x}{\sqrt{n_1}} - \frac{\widehat{M_3'}}{6n_1}\Big)\Big)^{1/3} - 1\Big].$$

Johnsonova transformacija u ovom slučaju ima oblik:

$$g_2(T) = T + \frac{1}{3}\frac{1}{\sqrt{n_1}}\widehat{M_3'}T^2 + \frac{1}{6}\frac{1}{\sqrt{n_1}}\widehat{M_3'}.$$

Inverzna funkcija funkcije $g_2(T) = x$ je oblika:

$$g_2^{-1}(x) = x - \frac{1}{3\sqrt{n_1}}\widehat{M_3'}x^2 - \frac{1}{6\sqrt{n_1}}\widehat{M_3'}.$$

Simulacijom različitih tipova asimetričnih raspodela se pokazuje da Johnsonova transformacija daje najbolju verovatnoću pokrivanja među razmatranim transformacijama, osim za normalnu raspodelu, kada F interval daje najbolju verovatnoću pokrivanja, pogledati Ćojbašić Rajić i Stanojević, 2013, što je jedan od originalnih rezultata ove disertacije. U slučaju visoke asimetrije raspodele verovatnoća pokrivanja je daleko od nominalne vrednosti. I pored toga, može se zaključiti da predložene metode pokazuju poboljšanje u odnosu na postojeću metodu u literaturi. Rezultati 95% intervala poverenja za različite raspodele i različite parove veličina uzoraka dati su u Tabeli 1.2.1.

Tabela 1.2.1: Pokrivenost 95% intervala poverenja

Raspodela	Koeficijent asimetrije	n_1, n_2	Hallova tr.	Johnsonova tr.	F int.
1. Eksponencijalna		10,10	0,744	0,826	0,790
sa sredinom 1	2	20,20	0,758	0,842	0,819
		50,50	0,773	0,860	0,828
2. Weibullova sa	0,62	100,100	0,808	0,864	0,832
parametrom oblika 2		10,100	0,749	0,829	0,803
1. Eksponencijalna		10,10	0,694	0,732	0,720
sa sredinom 1	2	20,20	0,696	0,767	0,737
		50,50	0,732	0,780	0,740
2. Eksponencijalna	2	100,100	0,738	0,806	0,774
sa sredinom 2		10,100	0,703	0,758	0,725
1. Normalna sa $\sigma^2 = 1$	0	10,10	0,910	0,967	0,953
		20,20	0,917	0,961	0,951
2. Normalna sa $\sigma^2 = 2$	0	50,50	0,918	0,962	0,953
		100,100	0,948	0,963	0,956
		10,100	0,909	0,962	0,953
1. Lognormalna		10,10	0,417	0,593	0,521
sa $\sigma^2 = 1$	6,18	20,20	0,425	0,623	0,538
		50,50	0,487	0,637	0,601
2. Weibullova sa	0,62	100,100	0,523	0,650	0,627
parametrom oblika 2		10,100	0,432	0,603	0,523

U radu Ćojbašić Rajić i Stanojević, 2013 prezentovana je metoda na podacima iz osiguranja. U pomenutom radu analizirane su različite intervalne ocene na podacima iz osiguranja, a deskriptivne statistike analiziranih skupova podataka date su u Tabeli 1.2.2.

Tabela 1.2.2: Deskriptivne statistike za dva skupa podataka

Skup	N	Sredina	Varijansa	Koef. asimetrije
Iznos šteta 2009	423	20.237,63	267.474,736	1,67
Iznos šteta 2007	223	17.792,56	196.853,551	1,62

Pokrivenost intervala poverenja za količnik disperzija data je u Tabeli 1.2.3, na osnovu koje se uočava da Johnsonova transformacija daje najbolje rezultate.

Tabela 1.2.3: Pokrivenost 95% intervala poverenja za količnik disperzija dva skupa podataka

Veličina uzorka n_1, n_2	Metoda	Verovatnoća pokrivanja
$n_1 = 10, n_2 = 10$	Hallova transformacija	0,707
	Johnsonova transformacija	0,779
	F interval	0,748
$n_1 = 20, n_2 = 20$	Hallova transformacija	0,721
	Johnsonova transformacija	0,809
	F interval	0,762
$n_1 = 50, n_2 = 50$	Hallova transformacija	0,735
	Johnsonova transformacija	0,828
	F interval	0,798
$n_1 = 100, n_2 = 100$	Hallova transformacija	0,768
	Johnsonova transformacija	0,855
	F interval	0,836
$n_1 = 10, n_2 = 100$	Hallova transformacija	0,709
	Johnsonova transformacija	0,808
	F interval	0,747

1.3 Intervali poverenja za razliku proporcija

U ovom odeljku biće analizirani intervali poverenja za razliku proporcija dva skupa zasnovani na t-statistici. Za intervalnu ocenu razlike proporcija postojeći metod je zasnovan na normalnoj raspodeli. Prikazan je Edgeworthov razvoj raspodele t-statistike i koristeći taj razvoj predloženi su novi intervali poverenja. Takođe je analizirano poređenje pokrivenosti predloženih intervala u oblasti osiguranja, što je jedan od originalnih rezultata ove disertacije.

Poznato je da ako se iz populacije sa velikim brojem elemenata bira uzorak sa vraćanjem od n elemenata, slučajna promenljiva X koja predstavlja broj elemenata sa određenom karakteristikom ima binomnu raspodelu $B(n, p)$. Parametar p $(0 < p < 1)$ je obično nepoznat i treba ga oceniti. On predstavlja proporciju elemenata u populaciji sa određenom karakteristikom. Kako je $E(X) = np$, sledi da je p sredina od X/n. Kada se parametar p zameni sa ocenom iz uzorka $\widehat{p} = \frac{x}{n}$ dobija se dvostrani $(1-\alpha)100\%$ intreval poverenja za proporciju:

$$\Big(\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\Big).$$

Čest problem u praksi jeste da se uporede proporcije dva skupa. U tu svrhu konstruiše se interval poverenja za razliku proporcija ili se testira hipoteza o razlici proporcija dva skupa. U statističkoj literaturi je najčešće korišćen Waldov interval poverenja, čijom primenom se ostvaruje pokrivenost značajno manja od nominalne verovatnoće pokrivenosti (za više detalja može se pogledati Agresti i Caffo, 2000 i Brown et al., 2001). Neka su p_1 i p_2 proporcije dva osnovna skupa i neka su:

$$\widehat{p}_1 = \frac{\widehat{X}_1}{n_1},$$

$$\widehat{p}_2 = \frac{\widehat{X}_2}{n_2},$$

proporcije dva velika, nezavisna uzorka uzeta iz dva odgovarajuća skupa, gde su $\widehat{X}_1$ i $\widehat{X}_2$ brojevi elemenata u uzorcima sa posmatranom karakteristikom, n_1 i n_2 su veličine uzoraka. $(1-\alpha)100\%$ Waldov interval poverenja za razliku proporcija $p_1 - p_2$ je oblika:

$$\left[\widehat{p}_1 - \widehat{p}_2 - z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}, \widehat{p}_1 - \widehat{p}_2 + z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}\right], \tag{1.19}$$

gde je $z_{\alpha/2}$ odgovarajući percentil standardne normalne raspodele.

Rajić i Stanojević, 2011 su razmatrale intervale poverenja za razliku proporcija zasnovane na t-statistici i razmatrale su pokrivenost predloženih intervala, što je jedan od originalnih rezultata ove disertacije. Kako je tu statistiku moguće modifikovati kako bi se otklonio efekat asimetrije i dobila preciznija ocena, dat je Edgeworthov razvoj raspodele pomenute t-statistike i na osnovu tog razvoja predložene su u literaturi nove transformisane statistike.

Posmatra se sledeća t raspodela, na čijoj se normalnoj aproksimaciji zasniva interval poverenja za razliku proporcija dva skupa:

$$T = \frac{\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}}}, \tag{1.20}$$

gde je $q_i = 1 - p_i$, i=1,2.

U slučaju kada veličina uzorka nije dovoljno velika, a posebno kada se radi o znatno asimetričnim raspodelama, aproksimacija gore navedene statistike normalnom raspodelom je suviše gruba. Da bi se eliminisao efekat asimetrije nalazi se Edgeworthov razvoj raspodele statistike (1.20). Neka je $R_n(p_1, p_2, t)$ periodična funkcija koja uzima vrednosti iz intervala $[-0,5; 0,5]$, i neka su:

$$\begin{aligned}
\delta &= \left(\frac{n}{n_2}\right)^2 p_2q_2(1-2p_2) - \left(\frac{n}{n_1}\right)^2 p_1q_1(1-2p_1),\\
\sigma &= \left(\frac{n}{n_2}p_2q_2 + \frac{n}{n_1}p_1q_1\right)^2,\\
a &= \frac{\delta}{6\sigma^2},\\
b &= \frac{n(1-2p_2)}{2n_2} - \frac{\delta}{6\sigma^2},
\end{aligned}$$

gde je $n = n_1 + n_2$. Neka je funkcija $Q(t) = \sigma^{-1}(a + bt^2)$.

Teorema 1.3.1. (Zhou et al., 2004) *Neka su* p_1 *i* p_2 *racionalni brojevi i neka* $min(n_1, n_2) \to +\infty$ *i* $n_2 = O(n_1)$. *Tada:*

$$P(T \leq t) = \Phi(t) + n^{-1/2}Q(t)\phi(t) + (n\sigma^2)^{-1/2}R_n(p_1, p_2, t)\phi(t) + O(n^{-1}\ln\ln n).$$

Dalje je dat $(1-\alpha)100\%$ interval poverenja za razliku proporcija $p_1 - p_2$ (Zhou et al., 2004), koji je dobijen korišćenjem Hallove ideje (Hall, 1982) i eliminisanjem greške $R_n(p_1, p_2, t)$ u Edgewothovom razvoju:

$$\Big[\widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2}(z_{1-\alpha/2} - n^{-1/2}\widehat{Q}(z_{1-\alpha/2})),$$
$$\widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2}(z_{\alpha/2} - n^{-1/2}\widehat{Q}(z_{\alpha/2}))\Big], \qquad (1.21)$$

gde je $\widehat{Q}(t) = \widehat{\sigma}^{-1}(\widehat{a} + \widehat{b}t^2)$ i $\widehat{a}$, $\widehat{b}$, $\widehat{\sigma}$, $\widehat{\delta}$ su ocene parametara a, b, σ i δ.

Hall je definisao monotonu transformaciju statistike T (Hall, 1992b i Hall, 1992a), koja u ovom slučaju ima sledeći oblik:

$$g_1(T) = n^{-1/2}\widehat{a}\widehat{\sigma} + T + n^{-1/2}(\widehat{b}\widehat{\sigma})T^2 + n^{-1}\frac{1}{3}(\widehat{b}\widehat{\sigma})^2T^3,$$

gde je $\widehat{\sigma} = ((n/n_2)\widehat{p}_2\widehat{q}_2 + (n/n_1)\widehat{p}_1\widehat{q}_1)^{1/2}$.

Dvostrani $(1-\alpha)100\%$ interval poverenja za razliku proporcija $p_1 - p_2$, dobijen pomoću prethodne transformacije (Zhou et al., 2004), je oblika:

$$\Big[\widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2}g_1^{-1}(z_{1-\alpha/2}), \widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2}g_1^{-1}(z_{\alpha/2})\Big], \quad (1.22)$$

gde je $g_1^{-1}(t)$ inverzna funkcija funkcije $g_1(T)$.

Johnson je takođe predložio transformaciju t-statistike u cilju eliminisanja efekata asimetrije (Johnson, 1978). U slučaju razlike proporcija ona je oblika:

$$g_2(T) = n^{-1/2}\widehat{a}\widehat{\sigma} + T + n^{-1/2}(\widehat{b}\widehat{\sigma})T^2,$$

tako da je dvostrani $(1-\alpha)100\%$ interval poverenja za razliku proporcija $p_1 - p_2$:

$$\Big[\widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2} g_2^{-1}(z_{1-\alpha/2}), \widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2} g_2^{-1}(z_{\alpha/2})\Big], \quad (1.23)$$

gde je $g_2^{-1}(t)$ inverzna funkcija funkcije $g_2(T)$.

Ćojbašić i Tomović, 2007 su predložili transformaciju t-statistike koja je u slučaju konstrukcije intervala poverenja za razliku proporcija oblika:

$$g_3(T) = n^{-1/2}\widehat{a}\widehat{\sigma} + T + n^{-1/2}(\widehat{b}\widehat{\sigma})T^2 + n^{-1}T^3.$$

Dvostrani $(1-\alpha)100\%$ interval poverenja za razliku proporcija $p_1 - p_2$ je oblika:

$$\Big[\widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2} g_3^{-1}(z_{1-\alpha/2}), \widehat{p} - \Big(\frac{\widehat{p}_2\widehat{q}_2}{n_2} + \frac{\widehat{p}_1\widehat{q}_1}{n_1}\Big)^{1/2} g_3^{-1}(z_{\alpha/2})\Big], \quad (1.24)$$

gde je $g_3^{-1}(t)$ inverzna funkcija funkcije $g_3(T)$.

Primer 1.3.1. Rajić i Stanojević, 2011 su analizirale upravljanje rizikom šteta u tarifi požar-civil. U tu svrhu ispitivale su pokrivenost intervala poverenja iz ovog odeljka. Posmatranje je vršeno na dva skupa podataka iz kojih su birani nezavisni uzorci (veličine 10, 20, 50 i 100). Prvi skup podataka predstavljao je podatke o visini šteta u tarifi požar-civil za prvi kvartal 2009. godine, dok je drugi skup podataka predstavljao podatke o visini šteta u tarifi požar-civil za prvi kvartal 2007. godine. U oba skupa podataka je razmatran procenat šteta koje su veće od 30000 dinara.

Koeficijent asimetrije prvog skupa je 3,79, što ukazuje na znatnu asimetriju. Skup ima 480 podataka, a procenat šteta koje su veće od 30000 dinara iznosi 0,28 (to je zapravo procenat šteta sa određenom karakteristikom). Koeficijent asimetrije drugog skupa je 4,54, što takođe ukazuje na znatnu asimetriju. Skup ima 240 podataka, a procenat šteta koje su veće od 30000 dinara iznosi 0,21.

Podaci o nastalim štetama su dobijeni iz jedne osiguravajuće kompanije. U Tabeli 1.3.1 i Tabeli 1.3.2 data je pokrivenost intervala poverenja za razliku proporcija. Može se uočiti da predloženi intervali ostvaruju bolju pokrivenost od Waldovog intervala poverenja.

Tabela 1.3.1: Pokrivenost 95% intervala poverenja za razliku proporcija dva skupa podataka

	Waldov interval (1.19)	Hallov interval (1.21)	Hallov interval (1.22)
$n_1 = 10$, $n_2 = 10$	0,680	0,810	0,891
$n_1 = 10$, $n_2 = 20$	0,698	0,825	0,896
$n_1 = 10$, $n_2 = 100$	0,699	0,836	0,900
$n_1 = 20$, $n_2 = 20$	0,710	0,840	0,902
$n_1 = 50$, $n_2 = 50$	0,721	0,856	0,956
$n_1 = 100$, $n_2 = 100$	0,789	0,901	0,947

Tabela 1.3.2: Pokrivenost 95% intervala poverenja za razliku proporcija dva skupa podataka

	Johnsonov interval (1.23)	Ćojbašić i Tomović (1.24)
$n_1 = 10$, $n_2 = 10$	0,822	0,872
$n_1 = 10$, $n_2 = 20$	0,839	0,880
$n_1 = 10$, $n_2 = 100$	0,841	0,891
$n_1 = 20$, $n_2 = 20$	0,861	0,899
$n_1 = 50$, $n_2 = 50$	0,893	0,914
$n_1 = 100$, $n_2 = 100$	0,914	0,933

Iz Tabele 1.3.1 i Tabele 1.3.2, u svim slučajevima koji su razmatrani, može se videti da Waldov interval poverenja (1.19) ostvaruje pokrivenost ispod 80%, zbog čega se i ne preporučuje njegovo korišćenje. Među svim predloženim metodama Hallov transformisani interval poverenja (1.22) ostvaruje najbolju pokrivenost.

Glava 2

Ocena indeksa ekstremne vrednosti i visokih kvatnila

Činjenica je da je mnoge prirodne i slučajne događaje teško predvideti. Tako na primer, u okvirima ekonomske prakse, sama ekonomska aktivnost je dostigla svoj nagli razvoj u drugoj polovini 20-tog veka. Ona je dovela do toga da je ukupna dobit svetske populacije porasla, a sa njom je porastao i rizik upravljanja. Kako bi se adekvatno reagovalo na te promene poslednjih godina su razvijeni mnogi instrumenti osiguranja i analizu svih aspekata ekonomskih procesa što je preuzeo risk menadžment. Posebni slučajevi, kada uobičajena iskustva i sigurnosne mere ne mogu da se iskoriste, nastaju kao takozvani 'najlošiji događaji'. Njihova analiza, modeliranje i predviđanje su upravo predmet upravljanja rizikom. Te nepopularne i retke scenarije je važno analizirati pre nego li se realizuju, iz razloga što mogu dovesti do potpunog kraha sistema. Iz razloga što se koriste ograničeni podaci, razvijene su tehnike bazirane na stohastičkoj teoriji ekstremnih vrednosti (EVT), kako bi se opisali i predvideli ekstremni događaji sa manjom ili većom preciznošću.

Statistički rečeno, osnovni problem jeste kako modelirati repove i oceniti ekstremne kvantile raspodele procesa rizika. Neka su $X_1, X_2, ..., X_n$ i.i.d. sa raspodelom F, koji, primera radi, mogu predstavljati dnevne gubitke na tržištu. Za analizu najlošijeg scenarija treba razmatrati nivo x_p koji se prelazi sa verovatnoćom $p \in (0,1)$ blisko 0, tj. $F(x_p) = 1 - p$, blisko 1. Ako se definiše uopštena inverz funkcija $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, može se videti da nivo x_p odgovara kvantilu $F^{\leftarrow}(1-p)$ sa malom verovatnoćom prelaska p. Ocena ovog visokog kvantila je analogna problemu modeliranja repa raspodele $\overline{F}(x) := 1 - F(x) = P(X_i > x)$ za veliki prag x.

Parametar koji se zove indeks ekstremne vrednosti (EVI), određuje debljinu i ponašanje repa raspodele, što je već poznato u literaturi. Postoji dosta radova vezanih za ocenu ovog parametra, ali se još uvek uči i dalje istražuje o tome. Uslovi koji omogućuju konvergenciju raspodele uzoračkog maksimuma $X_{(n,n)} = \max\{X_1, ..., X_n\}$ ka graničnoj raspodeli ekstremnog tipa, pokazuju se važnim za mnoge ocene. Pored indeksa ekstremne vrednosti, koji određuje asimtotsko ponašanje kvantila i repova raspodele i dodatni parametri (odnosa i lokacije) značajni su za ocenu kvantila. Pokazuje se da kvalitet ocene zavisi od modela iz koga se ocenjuje indeks i odgovarajućih ocena drugih parametara.

U ovom poglavlju biće dat pregled klasičnih metoda za ocenu indeksa ekstremne vrednosti kao i ekstremnih kvantila. Takođe će biti predstavljena dva nelinearna eksponencijalna regresiona modela, što su dali Feuerverger i Hall, 1999 i Beirlant et al., 1999, pomoću kojih je moguća konstrukcija ocene maksimalne verodostojnosti za realnu vrednost indeksa ekstemne vrednosti i ekstremnih kvantila.

2.1 Metode ekstrmnih vrednosti za ocenu visokih kvantila

U ovom odeljku biće dat pregled klasičnih metoda za konstrukciju repa $\overline{F}(x)$ neprekidne raspodele F i visokih kvantila $F^{\leftarrow}(1-p)$, za $p \in (0, 1)$ blisko 0. Pretpostavlja se da su date $X_1, ..., X_n$ i.i.d. iz raspodele F, sa odgovarajućim centriranim normiranim uzoračkim maksimumom $X_{(n,n)} = \max\{X_1, ..., X_n\}$, koji konvergira u raspodeli ka nedegenerisanoj graničnoj raspodeli. Gnedenko, 1943 je pokazao da ta nedegenerisana granična raspodela jeste tipa ekstremne vrednosti. To znači da za neko $\gamma \in \mathbb{R}$ postoji niz konstanti $a_n > 0$, $b_n \in \mathbb{R}$, tako da je:

$$\lim_{n\to\infty} P\Big(\frac{X_{(n,n)} - b_n}{a_n} \leq x\Big) \longrightarrow H_\gamma(x), \tag{2.1}$$

za sve tačke neprekidnosti raspodele ekstremne vrednosti $H_\gamma(x)$, koja je definisana:

$$H_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & \gamma \neq 0, 1+\gamma x > 0; \\ \exp(-e^{-x}), & \gamma = 0. \end{cases}$$

Funkcija raspodele F pripada oblasti privlačenja raspodele ekstremnih vrednosti H_γ i označava se $F \in MDA(H_\gamma)$. Parametar γ se naziva indeks ekstremne vrednosti. U ovom radu se razmatraju modeli debelih repova, tj. modeli raspodele F sa parametrom $\gamma > 0$. Karakteristika tih raspodela je da rep $\overline{F}$ opada sporo. Primeri takvih raspodela su Frešeova klasa, Pareto, Burr, Studentova t, α-stabilna ($\alpha < 2$) i loggama raspodela. Drugi tip raspodele ekstremne vrednosti jeste Gumbelova klasa raspodela, sa $\gamma = 0$, što su eksponencijalna, normalna, lognormalna, gama raspodela, koje karakteriše eksponencijalna brzina opadanja. I poslednji tip raspodela ekstremnih vrednosti jeste Weibullova klasa, koju karakteriše $\gamma < 0$, sa konačnom desnom krajnjom tačkom $x_+ := F^{\leftarrow}(1)$. Primeri raspodela ovog tipa su uniformna, beta i druge.

Modeli debelih repova su korisni u mnogim poljima istraživanja, kao što su osiguranje, telekomunikacije, internet saobraćaj, između ostalih, i skoro uvek je potrebno oceniti visoke kvantile. Dakle, statistika ekstrema kao primarni problem ima ocenu indeksa ekstremne vrednosti. Klasične ocene ekstremnih parametara pokazuju bias (pristrasnost, nekonzistentnost) i mogu biti veoma osetljive na izbor broja statistika poretka koje se koriste u ocenjivanju, broja k. Za malu vrednost k ocene imaju veliku disperziju, a za veliku vrednost k ocene imaju veliki bias. Razvijene ocene oblika i razmere drugog reda dozvoljavaju razvoj ocena smanjenog biasa drugog reda, koje su manje osetljive na izbor vrednosti k. Kasnije će u disertaciji biti više reči o optimalnom izboru vrednosti k.

Uvede se sledeće oznake: definiše se $U := (\frac{1}{1-F})^{\leftarrow}$, gde je sa $F^{\leftarrow}(x)$ označena uopštena inverz funkcija funkcije F. Dalje važi:

$$F \in MDA(H_\gamma), \gamma > 0 \Longleftrightarrow 1 - F \in RV_{-1/\gamma} \Longleftrightarrow U \in RV_\gamma, \tag{2.2}$$

što je rezultat dat u Gnedenko, 1943 i de Haan, 1970, gde RV_γ označava klasu pravilno promenljivih funkcija u beskonačnosti, sa indeksom pravilne promenljivosti γ, za svako realno γ, tj. kaže se da je funkcija g pravilno promenljiva sa indeksom promenljivosti γ ako je pozitivno merljiva i takva da je granična vrednost $\lim\limits_{t\to+\infty} g(tx)/g(t) = x^\gamma$, za $x > 0$.

Kao što je rečeno, razmatra se ocena visokih kvantila x_p, standardni parametar u mnogim oblastima primene. Važi $F(x_p) = 1-p$, gde je x_p kvantil koji se prelazi sa malom verovatnoćom p. Tačnije, vrši se ekstrapolacija unutar uzorka, i ocenjuje se:

$$x_p = U\Big(\frac{1}{p}\Big),\ p = p_n \to 0,\ np_n \to K,\ n \to \infty,\ K \in [0,1]. \tag{2.3}$$

Ako je $X_{(1,1)} < ... < X_{(n,n)}$ niz statistika poretka, Weissman(1978) je predložio sledeću semi-parametarsku ocenu za x_p:

$$Q_{\widehat{\gamma}}^{(p)}(k) := X_{(n-k,n)}\, c_n^{\widehat{\gamma}},\ \ c_n := \frac{k}{np} \to \infty,\ \ n \to \infty, \tag{2.4}$$

gde je $\widehat{\gamma}$ konzistentna ocena indeksa ekstremne vrednosti γ. de Hann i Rootzén, 1983, Ferreira et al., 2003 i Matthys i Beirlant, 2003 su dali semi-parametarske ocene visokih kvantila za $\gamma \in R$. Kao što je uobičajeno za semi-parametarske modele i ovde se pretpostavlja da je $k = k(n)$ intermidijatni niz, tj. niz celih brojeva iz skupa $[1, n]$ za koji važi:

$$k(n) \to \infty,\ \frac{k(n)}{n} \to 0,\ n \to \infty.$$

Kako je:

$$P\Big(\frac{X_{(n,n)} - b_n}{a_n} \le x\Big) = (1 - \overline{F}(a_n x + b_n))^n,$$

to je (2.1) ekvivalentno sa:

$$\lim_{n\to\infty} n\overline{F}(a_n x + b_n) \to -\ln H_\gamma(x) = (1 + \gamma x)^{-1/\gamma},$$

za sve tačke neprekidnosti funkcije H_γ, gde se za $\gamma = 0$ uzima e^{-x}. Tako za veliki prelaz $y = a_n x + b_n$ važi aproksimacija funkcije $\overline{F}$:

$$\overline{F}(y) \approx \frac{1}{n}\Big(1 + \gamma\frac{y - b_n}{a_n}\Big)^{-1/\gamma}, \tag{2.5}$$

što daje sledeću aproksimaciju visokog kvantila, sa verovatnoćom prelaska p, koja je bliska 0:

$$x_p = F^{\leftarrow}(1-p) \approx a_n \frac{(pn)^{-\gamma} - 1}{\gamma} + b_n.$$

Sa odgovarajućim ocenama za indeks ekstremne vrednosti γ i normirajućih konstanti a_n i b_n, dobija se odgovarajuća ocena za x_p.

Weissman, 1978 je razmatrao ocenu ekstremnih kvantila za tri odvojene klase ekstremnih vrednosti, pretpostavljajući apriorno poznavanje indeksa γ. U praksi ta apriorna informacija nije poznata, pa nije odmah očigledno kojoj klasi posmatrana raspodela pripada. Tako su metode koje tretiraju sve indekse ekstremne vrednosti od velikog značaja. Obimna je literatura o oceni indeksa ekstremne vrednosti, dok je literatura o modeliranju repa i visokih kvantila prilično ograničena. Dalje će biti prikazane klasične metode, počevši sa generalisanim Pareto modelom za prelazak preko visokog praga. Uopšteno, postoje tri klase metoda za ocenu ekstremnih kvantila: metod blokova (koji ovde nije razmatran), POT metod (pikovi iznad praga) i Q(vantil)-bazirani metod, pogledati na primer Pickands, 1975, Matthys i Beirlant, 2003, Hill, 1975, Beirlant et al., 1996, de Haan i Resnick, 1980, Dekkers et al., 1975.

2.2 POT metod

POT metod koristi aproksimaciju (2.5). Ako se sa

$$F_u(x) = P(X - u \leq x \mid X > u)$$

označi funkcija raspodele prelaska X preko praga u, i sa $G_{\gamma,\mu,\sigma}(x)$ označi generalisana Pareto raspodela (GPD), koja je definisana:

$$\overline{G}_{\gamma,\mu,\sigma}(x) = \begin{cases} \left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-1/\gamma}, & \gamma \neq 0, \\ \exp\left(-\frac{x-\mu}{\sigma}\right), & \gamma = 0, \end{cases} \tag{2.6}$$

rezultat granične raspodele prelaska preko visokog praga, koji kaže da (2.1) važi akko

$$\lim_{u \to x_+} \sup_{0<x<x_+-u} \mid F_u(x) - G_{\gamma,0,\sigma(u)}(x) \mid = 0, \tag{2.7}$$

za pozitivnu funkciju razmere $\sigma(u)$, dao je Pickands, 1975.

GPD sa parametrima γ, $\mu = 0$ i $\sigma = \sigma(u)$ je dobra aproksimacija raspodele F_u za N_u ekscesa $Y_j = X_{i_j} - u$, ako se fiksira prag u i izdvoje iz uzorka $X_1, ..., X_n$ one opservacije $X_{i_1}, ..., X_{i_{N_u}}$ koje prelaze prag u. Takođe, GPD može dobro fitovati ekscese $Y_1, ..., Y_{N_u}$ sa metodom maksimalne verodostojnosti, što je pokazao Smith, 1987. On je takođe pokazao da su ocene za γ i σ asimtotski normalne ako je $\gamma > -1/2$. Tako je ocena POT maksimalne verodostojnosti $\widehat{\gamma}^{MLP}$, za EVI, upravo $(1+\gamma)^2/N_u$. Za $\gamma \in (-1,-1/2)$ ocena maksimalne verodostojnosti konvergira graničnoj raspodeli koja nije normalna, brzinom $n^{-\gamma}$. Metodu momenta za ocene γ i σ, koja funkcioniše za slučaj kada je $\gamma < 1/2$, dali su Hosking i Wallis, 1987. Oni su takođe pokazali metodom verovatnosnog težinskog momenta (PWM), da je odgovarajući indeks ekstremne vrednosti dobra alternativa za ocenu maksimalne verodostojnosti u slučaju $\gamma < 1$. Alternativni metod percentila (EPM) koji ne zahteva nikakva ograničenja za γ, predložili su Castillo i Hadi, 1997, dok su Coles i Powell, 1996 predložili Bajesovu metodu.

Jednom od gore spomenutih metoda mogu se oceniti γ i σ sa $\widehat{\gamma}_u$ i $\widehat{\sigma}_u$ redom, te uslovni rep $\overline{F}_u$ funkcije F može biti ocenjen:

$$\widehat{\overline{F}}(x) = \left(1+\widehat{\gamma}_u \frac{x}{\widehat{\sigma}_u}\right)^{-1/\widehat{\gamma}_u},\ 0 < x < x_+ - u,$$

i bezuslovni rep $\overline{F}(x) = \overline{F}(u)\overline{F}_u(x-u)$ može biti ocenjen:

$$\widehat{\overline{F}}(x) = \frac{N_u}{n}\left(1+\widehat{\gamma}_u \frac{x-u}{\widehat{\sigma}_u}\right)^{-1/\widehat{\gamma}_u},\ u < x < x_+, \tag{2.8}$$

dok se $\overline{F}(u)$ ocenjuje empirijskom verovatnoćom prelaska N_u/n. Dalje, koristeći jednakost (2.8) dobija se spomenuta POT ocena za visoke kvantile preko praga u:

$$\widehat{F}_u^{\leftarrow}(1-p) = u + \widehat{\sigma}_u \frac{\left(\frac{N_u}{np}\right)^{\widehat{\gamma}_u} - 1}{\widehat{\gamma}_u},\ p < \frac{N_u}{n}.$$

Treba istaći da se za u bira jedna od statistika poretka $X_{(1,n)}, ..., X_{(n,n)}$. Uzimajući da je u $(k+1)$-va najveća opservacija, tj. $u = X_{(n-k,n)}$ i da je $N_u = k$, definisana je ocena kvantila:

$$\widehat{x}_{p,k+1}^{POT} := X_{(n-k,n)} + \widehat{\sigma}_{k+1} \frac{\left(\frac{k}{np}\right)^{\widehat{\gamma}_{k+1}} - 1}{\widehat{\gamma}_{k+1}},\ p < \frac{k}{n},$$

koristeći indeks $k+1$ za u.

Posle brojnih simulacija u radu Matthys i Beirlant, 2003, između tri POT metode ocenjivanja (EPM, ML i PWM), pokazano je da ML metod daje najbolju ocenu kada je $\gamma > 0$, dok se EPM pokazuje boljom metodom kada je $\gamma < 0$. I treba reći da do sada nije istaknuta adekvatna procedura za izbor optimalnog praga $u = X_{(n-k,n)}$, za bilo koju POT ocenu.

2.3 Q(vantil)-bazirani metod

Kao što je već rečeno, poslednjih decenija sve više pažnje je posvećeno proučavanju raspodela sa debelim repovima, a samim tim i problemu ocene indeksa ekstremne vrednosti. Danas postoji više poznatih ocena, koje su izražene preko statistika poretka posmatranog uzorka. Neka su $X_1, ..., X_n$ i.i.d. sa raspodelom F, koja zadovoljava uslov $1 - F(x) \in RV_{-1/\gamma}$, za veliko x, gde je $\gamma > 0$. Ako je $X_{(1,n)}, ..., X_{(n,n)}$ niz statistika poretka niza $X_1, ..., X_n$, sledeću ocenu parametra γ predložio je Hill, 1975:

$$\gamma_{n,k}^{(1)} = \frac{1}{k} \sum_{i=1}^{k} \ln X_{(n-i+1,n)} - \ln X_{(n-k,n)},$$

gde je k broj koji zadovoljava $1 \leq k \leq n$. Problem izbora k je prilično komplikovan, pogledati na primer: Danielsson et al., 2001, Dekkers i de Haan, 1993, Draisma et al., 1999, Dres et al., 1998, Hall, 1990 i Geluk, 2000. Kasnije će biti više reči o izboru k. Nakon predložene Hillove ocene, u radovima Beirlant et al., 1996, Csörgo et al., 1985, de Haan i Resnick, 1980, Dekkers et al., 1975, Pickands, 1975, kao i u njihovim referencama, predložene su modifikacije te ocene.
Dalje slede druge ocene parametra γ, koje se baziraju na statistikama poretka:

$$\begin{aligned}
\gamma_{n,k}^{(2)} &= (\ln 2)^{-1} \ln \frac{X_{(n-[k/4],n)} - X_{(n-[k/2],n)}}{X_{(n-[k/2],n)} - X_{(n-k,n)}}, \\
\gamma_{n,k}^{(3)} &= \gamma_{n,k}^{(1)} + 1 - \frac{1}{2}\Big(1 - (\gamma_{n,k}^{(1)})^2/M_n\Big)^{-1}, \\
\gamma_{n,k}^{(4)} &= \frac{M_n}{2\gamma_{n,k}^{(1)}},
\end{aligned}$$

gde je:

$$M_n = \frac{1}{k} \sum_{i=1}^{k} (\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)})^2.$$

Ocenu $\gamma_{n,k}^{(2)}$ dao je Pickands, 1975, $\gamma_{n,k}^{(3)}$ je data u Dekkers et al., 1975 i ocenu $\gamma_{n,k}^{(4)}$ dao je de Vries (ona ovde nije razmatrana). Sve ove ocene su upoređivane i pokazano je da se nijedna od ovih ocena ne pokazuje najboljom. Za različite vrednosti parametra γ različite ocene imaju najmanju asimtotsku srednje kvadratnu grešku. Izučavanje asimtotskog ponašanja ovih ocena nije nimalo lak zadatak. U radovima Davydov et al., 2000 i Davydov i Paulauskas, 1999 predložene su nove ocene i dati su rezultati asimtotske normalnosti tih ocena, o čemu će biti reči kasnije.

Q(vantil)-bazirani metodi za ocenu parametra γ su Hillova ocena, Pickandsova ocena, ocena momenta i ocena maksimalne verodostojnosti za eksponencijalno regresioni model, o kojima će dalje biti reči.

2.3.1 Hillova ocena

Za debele repove klasična ocena indeksa ekstremne vrednosti koja se koristi za semi-parametarsku ocenu kvantila, jeste Hillova ocena $\gamma_{n,k}^{(1)} = \widehat{\gamma}(k) := H(k)$,

$$H(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \frac{1}{k}\sum_{i=1}^{k} U_i, \tag{2.9}$$

što je sredina log-ekscesa $V_{ik} := \ln X_{(n-i+1,n)} - \ln X_{(n-k,n)}$, $1 \le i \le k < n$, kao i sredina skaliranog log-prostora:

$$U_i := i(\ln X_{(n-i+1,n)} - \ln X_{(n-i,n)}),\ 1 \le i \le k < n. \tag{2.10}$$

Na osnovu Hillove ocene $H(k)$ dobija se klasična ocena kvantila $Q_H^{(p)}(k)$. Ako važi uslov prvog reda (2.2) i za intermidijatno k, poznato je da su $H(k)$ i $Q_H^{(p)}(k)$ konzistentne ocene za γ i x_p. Glavni problem sa ovim semi-parametarskim ocenama jeste to što pokazuju veliku disperziju za malo k i veliki bias za veliko k.

Dalje se pretpostavlja uslov drugog reda, kako bi se dobile informacije o ponašanju raspodele ocena:

$$\begin{aligned}
&\lim_{t\to+\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \Longleftrightarrow \\
&\lim_{t\to+\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},
\end{aligned} \tag{2.11}$$

za $x > 0$, parametar oblika drugog reda $\rho \le 0$ i funkciju $|A|$ koja je pravilno promenljiva sa indeksom promenljivosti ρ (Geluk i de Haan, 1980). Korisna je pretpostavka da se koristi Hallova klasa modela sa teškim repom, koju su dali Hall, 1982, Hall i Welsh, 1985, gde je za $\gamma > 0$, $\rho < 0$, $C > 0$ i $D_1 \neq 0$:

$$U(t) = Ct^{\gamma}\Big(1 + D_1 t^{\rho} + o(t^{\rho})\Big),\ t \to +\infty. \tag{2.12}$$

U tom slučaju važi uslov drugog reda (2.11), za $A(t) = \rho D_1 t^{\rho} := \gamma\beta t^{\rho}$.

Teorema 2.3.1. (de Haan i Peng, 1998) *Pod uslovom drugog reda (2.11) i za intermidijantno k, važi asimtotska normalnost $H(k)$ u (2.9), može se napisati:*

$$H(k) =_d \gamma + \frac{\gamma}{\sqrt{k}} Z_k + \frac{A(n/k)}{1-\rho}(1 + o_p(1)), \tag{2.13}$$

gde je $Z_k = \sqrt{k}(\sum_{i=1}^{k} E_i/k - 1)$ i $\{E_i\}$, $i = 1, 2, ...k$, su i.i.d. eksponencijalne slučajne veličine. Ako je k takvo da je $\sqrt{k}A(n/k) \to \lambda \neq 0$ konačno, za $n \to \infty$, $\sqrt{k}(H(k) - \gamma)$ je asimtotski normalno sa disperzijom γ^2 i nenula sredinom $\lambda/(1-\rho)$.

Rezultat (2.13) je naveo da se razmotri mogućnost korišćenja asimptotskog biasa, gradeći ocene drugog reda sa smanjenim biasom. Pogledati na primer Peng, 1998, Beirlant et al., 1999, Gomes et al., 2000. U tim radovima autori su uspeli da uklone dominantnu komponentu asimtotskog biasa, ali sa rastućom asimtotskom disperzijom. Gomes et al., 2004b, Caeiro et al., 2005 i Gomes et al., 2007a, su predložili ocenu minimalne-varijanse smanjenog-biasa (MVRB), bazirane na oceni parametara drugog reda, koja je sposobna da smanji bias bez povećanja asimtotske disperzije. Ona drži disperziju na vrednosti γ^2, što je upravo disperzija Hillove ocene. O ovoj (MVRB) oceni biće više reči u četvrtom odeljku ovog poglavlja.

Na osnovu radova koji se bave semi-parametarskim ocenama visokih kvantila za teške repove, na primer Gomes i Figueiredo, 2006 i Caerio i Gomes, 2007, može se dati sledeći rezultat.

Teorema 2.3.2. (Caeiro i Gomes, 2008) *Pod uslovima prethodne teoreme, za uslov (2.3), poznati indeks ekstremne vrednosti γ i c_n iz jednakosti (2.4), važi:*

$$Q_{\gamma}^{(p)}(k) =_d x_p\Big(1 + \frac{\gamma}{\sqrt{k}} B_k + \frac{1 - c_n^{\rho}}{\rho} A(n/k)(1 + o_p(1))\Big), \tag{2.14}$$

gde su B_k asimtotski standardne normalne slučajne veličine. Ako je $\sqrt{k}A(n/k) \to \lambda$ konačno, onda je $\sqrt{k}(Q_{\gamma}^{(p)}(k)/x_p - 1)$ je asimptotski *normalno sa disperzijom γ^2 i sredinom λ/ρ. Ako je γ nepoznato i ocenjeno sa konzistentnom ocenom $\hat{\gamma}$, važi:*

$$Q_{\hat{\gamma}}^{(p)}(k) =_d x_p\Big(1 + (\hat{\gamma} - \gamma)\ln c_n + \frac{\gamma}{\sqrt{k}}B_k + \frac{1 - c_n^{\rho}}{\rho}A(n/k)(1 + o_p(1))\Big). \quad (2.15)$$

Takođe, ako je $\sqrt{k}A(n/k) \to \lambda$ konačno i $\ln c_n/\sqrt{k} \to 0$ za $n \to \infty$, tada $\frac{\sqrt{k}}{\ln c_n}(Q_{\gamma}^{(p)}(k)/x_p - 1)$ ima asimtotski istu raspodelu kao $\sqrt{k}(\hat{\gamma} - \gamma)$.

Iz jednakosti (2.15) sledi da ponašanje $\hat{\gamma}$ određuje ponašanje $Q_{\hat{\gamma}}^{(p)}$. Sumiranje $(1 - c_n^{\rho})A(n/k)/\rho$ je asimtotski ekvivalentno sa $A(n/k)/\rho$ i dominantna komponenta biasa $Q_{\gamma}^{(p)}$ u jednakosti (2.14) ne utiče na graničnu raspodelu $Q_{\hat{\gamma}}^{(p)}$. U radu Matthys et al., 2004 je uočeno da uklanjanje tog člana za konačni uzorak, vodi u poboljšanje stabilnosti ocena kvantila kao funkcije od k.
Ovde treba reći da je slučaj nedostajućih podataka razmatran u radu Mladenović i Piterbarg, 2008 i slučaj zavisnih slučajnih veličina je razmatran u radu Hsing, 1991.

2.3.2 Pickandsova ocena

Atraktivnu ocenu indeksa ekstremne vrednosti predložio je Pickands, 1975. Neka je $k = k(n)$ intermidijatni niz, tj. niz celih brojeva tako da $k(n) \to \infty$ i $k(n)/n \to 0$, $n \to \infty$. Ocena Pickandsa je data:

$$\widehat{\gamma}_n := (\ln 2)^{-1} \ln \frac{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}{X_{(n-2k+1,n)} - X_{(n-4k+1,n)}}, \quad (2.16)$$

gde je $X_{(1,n)} \leq ... \leq X_{(n,n)}$ niz statistika poretka niza $X_1, ..., X_n$. Treba reći da je ranije dat nešto drugačiji zapis ove ocene. Pickands je pokazao da je ova ocena slabo konzistentna. U slučaju kada niz $k(n)$ raste dovoljno brzo, važiće jaka konzistentnost. U radu Pickands, 1975 su dati i uslovi pod kojima važi asimtotska normalnost.

Razmatra se idealan model, gde je n opservacija $X_1, ..., X_n$ i.i.d. iz raspodele F. Uzeto je k nezavisnih jednako raspodeljenih slučajnih veličina iz F: $Y_1, ..., Y_k$. Cilj je da se nađe nivo x_{k,p_0}, gde je p_0 dosta manje od 1, tako da je $P(\max(Y_1, ..., Y_k) \leq x_{k,p_0}) = 1 - p_0$, tj.

$$F^k(x_{k,p_0}) = 1 - p_0.$$

Dalje se U definiše promoću F na sledeći način:

$$U(x) := \Big(\frac{1}{1-F}\Big)^{\leftarrow}(x).$$

Ako se definišu:

$$\begin{aligned} p: &= 1-(1-p_0)^{1/k}, \\ x_p: &= x_{k,p_0}, \end{aligned}$$

tada je:

$$x_p = U\Big(\frac{1}{p}\Big).$$

Cilj je dati ocenu za x_p u funkciji statistika poretka. Neka je F_n empirijska funkcija raspodele i

$$U_n := \Big(\frac{1}{1-F_n}\Big)^{\leftarrow}.$$

Tada je $F_n(X_{(k,n)}) = k/n$, $k = 1, 2, ..., n$ i

$$X_{(n-k+1,n)} = U_n\Big(\frac{n}{k}\Big), \quad k = 1, 2, ..., n.$$

Potrebna su dodatna ograničenja za funkciju F, kada je $p < 1/n$. F se uzima iz oblasti privlačenja neke raspodele ekstremnih vrednosti.

Neka $n \to \infty$ i $p = p_n \to 0$. Za nepoznatu funkciju U važi:

$$x_p = U\Big(\frac{1}{p}\Big) = \frac{U(1/p) - U(n/k)}{U(n/k) - U(n/(2k))}\Big(U\Big(\frac{n}{k}\Big) - U\Big(\frac{n}{2k}\Big)\Big) + U\Big(\frac{n}{k}\Big),\ k = 1, 2, ...$$

Ako se U zameni svojom empirijskom funkcijom U_n i iskoristi Lema 2.3.2 (koja je data kasnije), dobija se sledeća ocena za x_p:

$$\widehat{x}_{p,n} := \frac{(k/(np_n))^{\widehat{\gamma}_n} - 1}{1 - 2^{-\widehat{\gamma}_n}}\Big(X_{(n-k+1,n)} - X_{(n-2k+1,n)}\Big) + X_{(n-k+1,n)}, \qquad (2.17)$$

gde je $\widehat{\gamma}_n$ data formulom (2.16). Treba primetiti da se najveća vrednost opservacije ne koristi eksplicitno. Jedan mogući argument za to jeste što najveća opservacija može dati dosta veliku disperziju.

Konzistentnost i asimptotska normalnost

Za dokaz slabe i jake konzistentnosti i asimtotske normalnosti potrebni su sledeći rezultati. Dokazi se mogu pogledati u Dekkers i de Haan, 1989.

Lema 2.3.1. *Ako je $F(x) = 1-e^{-x}$ (standardna eksponencijalna raspodela), $k \leq n$ i $k = k(n) \to \infty$, $n \to \infty$, tada:*

$$\sqrt{2k}\,\left(X_{(n-k+1,n)} - X_{(n-2k+1,n)} - \ln 2\right)$$

ima asimtotski standardnu normalnu raspodelu.

Posledica 2.3.1. $X_{(n-k+1,n)} - X_{(n-2k+1,n)} \to \ln 2$, *u verovatnoći, kada* $n \to \infty$.

Lema 2.3.2. *Ako je $U := (1/(1-F))^{\leftarrow}$ uopštena inverz funkcija, relacija (2.1) važi akko za $x, y > 0$, $y \neq 1$,*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\gamma - 1}{y^\gamma - 1},$$

gde se za $\gamma = 0$ uzima granična vrednost $\frac{\ln x}{\ln y}$.

Teorema 2.3.3. (Dekkers i de Haan, 1989) *(Slaba konzistentnost) Ako važi (2.1), $k(n) \to \infty$ i $k(n)/n \to 0$, $n \to \infty$, tada $\widehat{\gamma}_n \to \gamma$ u verovatnoći, za $n \to \infty$.*

Dokaz: Neka su $A_1, A_2, ...$ i.i.d. eksponencijalne slučajne veličine i neka je $(A_{(k,n)})$ odgovarajući niz statistika poretka. Tada:

$$(X_{(n-k+1,n)})_{k=1}^n =_d (U(e^{A_{(n-k+1,n)}}))_{k=1}^n,$$

gde $=_d$ označava jednakost u raspodeli. Treba primetiti da $k(n)/n \to 0$, povlači $e^{A_{(n-k+1,n)}} \to \infty$ skoro sigurno za $n \to \infty$. Važi:

$$\frac{U\left(e^{A_{(n-k+1,n)}}\right) - U\left(e^{A_{(n-2k+1,n)}}\right)}{U\left(e^{A_{(n-2k+1,n)}}\right) - U\left(e^{A_{(n-4k+1,n)}}\right)} =$$

$$\frac{U\left(e^{A_{(n-2k+1,n)}} \cdot e^{A_{(n-k+1,n)} - A_{(n-2k+1,n)}}\right) - U\left(e^{A_{(n-2k+1,n)}}\right)}{U\left(e^{A_{(n-2k+1,n)}}\right) - U\left(e^{A_{(n-2k+1,n)}} \cdot e^{A_{(n-4k+1,n)} - A_{(n-2k+1,n)}}\right)} \to \frac{2^\gamma - 1}{1 - 2^{-\gamma}} = 2^\gamma,$$

u verovatnoći, iz Posledice 2.3.1 i Leme 2.3.2. Rezultat sledi. ■

Teorema 2.3.4. (Dekkers i de Haan, 1989) *(Jaka konzistentnost) Ako važi (2.1), $k(n)/n \to 0$ i $k(n)/\ln\ln n \to \infty$, $n \to \infty$, tada $\widehat{\gamma}_n \to \gamma$ skoro sigurno, za $n \to \infty$.*

Teorema 2.3.5. (Asimptotska normalnost) *Ako se pretpostavi da U ima pozitivan izvod i da postoji pozitivna funkcija a, tako da je za $x > 0$ i $\gamma \in \mathbb{R}$:*

$$\lim_{t\to\infty} \frac{(tx)^{1-\gamma}U'(tx) - t^{1-\gamma}U'(t)}{a(t)} = \pm \ln x, \tag{2.18}$$

tada:

$$\sqrt{k}(\widehat{\gamma}_n - \gamma)$$

ima asimtotski normalnu raspodelu sa sredinom 0 i disperzijom $\gamma^2(2^{2\gamma+1} + 1)/(2(2^\gamma - 1)\ln 2)^2$ za niz $k = k(n) \to \infty$ koji zadovoljava $k(n) = o(n/g^{\leftarrow}(n))$, gde je $g(t) := t^{3-2\gamma}(U'(t)/a(t))^2$, $n \to \infty$.

Dokaz prethodne teoreme pogledati u Dekkers i de Haan, 1989.

U narednoj teoremi su formulisani uslovi za funkciju U preko funkcije raspodele F i njene gustine. Takođe je potrebno poznavanje teorije pravilno promenljivih funkcija i funkcija klase Π i Γ (pogledati Geluk i de Haan, 1987 i Mladenović, 2002). Ovde je data definicija.

Definicija 2.3.1. *Rastuća funkcija $H : (c, +\infty) \to \mathbb{R}$ je Π-promenljiva ako postoji nenegativna funkcija $a : (c, +\infty) \to \mathbb{R}_+$ i funkcija $b : (c, +\infty) \to \mathbb{R}$, takve da za svako $x > 0$ važi jednakost:*

$$\lim_{t\to+\infty} \frac{H(tx) - b(t)}{a(t)} = \ln x.$$

Koristi se oznaka $H \in \Pi$ ili $H \in \Pi(a)$ i funkcija a se naziva pomoćna funkcija za Π-promenljivu funkciju H.

Definicija 2.3.2. *Rastuća funkcija $H : (c, x_0) \to \mathbb{R}$ je Γ-promenljiva, ako su ispunjena sledeća dva uslova:*

a) $\lim_{x\to x_0} H(x) = +\infty$,

b) *postoji funkcija $g : (c, x_0) \to \mathbb{R}_+$, takva da za svaki realan broj x važi jednakost:*

$$\lim_{t\to x_0} \frac{H(t + xg(t))}{H(t)} = e^x.$$

Koristi se oznaka $H \in \Gamma$ ili $H \in \Gamma(g)$ i funkcija g se zove pomoćna funkcija za Γ-promenljivu funkciju H.

Teorema 2.3.6. (Dekkers i de Haan, 1989) *Pretpostavlja se da U ima pozitivan izvod $U^{'}$. Ekvivalentna su sledeća tvrđenja (sa oba izbora znaka):*

(a) $\pm t^{1-\gamma} U^{'}(t) \in \Pi(a)$.

(b)
- *Za $\gamma > 0$:* $\pm t^{1+1/\gamma} F^{'}(t) \in \Pi(b)$.
- *Za $\gamma < 0$:* $U(+\infty) := \lim\limits_{t \to +\infty} U(t) < +\infty$ *i* $\mp t^{-1-1/\gamma} F^{'}(U(+\infty) - t^{-1}) \in \Pi(b)$.
- *Za $\gamma = 0$: Neka su $f_0 = (1-F)/F^{'}$ i $x^* := \sup\{x | F(x) < 1\}$. Tada postoji pozitivna funkcija α sa $\alpha(t) \to 0$, $t \to x^*$, takva da za $x > 0$ uniformno lokalno važi:*

$$\lim_{t \to x^*} \Big[\Big(\frac{1 - F(t + x f_0(t))}{1 - F(t)} - e^{-x}\Big)/\alpha(t)\Big] = \pm \frac{x^2}{2} e^{-x}.$$

Posledica 2.3.2. *Ako F zadovoljava prethodnu teoremu pod (b), tada U zadovoljava jednakost (2.18) sa:*

$$a(t) = \begin{cases} \gamma^3 t^{1-\gamma} (U(t))^{1-1/\gamma} b(U(t)), & \gamma > 0, \\ \alpha(U(t)) f_0(U(t)) = tU^{'}(t)\alpha(U(t)), & \gamma = 0, \\ -\gamma^3 t^{1-\gamma} (U(\infty) - U(t))^{1-1/\gamma} b(1/(U(\infty) - U(t))), & \gamma < 0. \end{cases}$$

Normalna raspodela zadovoljava uslove Teoreme 2.3.6 i tada važi asimtotska normalnost ocene $\widehat{\gamma}_n$, za niz $k(n) \to \infty$ koji zadovoljava $k(n) = o(\ln^2 n)$.

Za raspodelu kao što je Košijeva važi sledeća teorema.

Teorema 2.3.7. (Dekkers i de Haan, 1989) *Pretpostavlja se da važi jedan od sledećih uslova:*

(a) *Za neko $\gamma > 0$, $\rho > 0$ i $c > 0$ funkcija $t^{1+1/\gamma} F^{'}(t) - c$ je konstantnog znaka i:*

$$\lim_{t \to +\infty} \frac{(tx)^{1+1/\gamma} F^{'}(tx) - c}{t^{1+1/\gamma} F^{'}(t) - c} = x^{-\rho},$$

(pravilno promenljiva sa eksponentom $-\rho$, oznaka $\pm(t^{1+1/\gamma} F^{'}(t) - c) \in RV_{-\rho}$).

(b) *Za neko* $\gamma < 0$, $\rho > 0$ *i* $c > 0$ *funkcija:*

$$\pm(t^{-1-1/\gamma}F'(U(+\infty) - t^{-1}) - c) \in RV_{-\rho}.$$

Tada:

$$\sqrt{k}(\widehat{\gamma}_n - \gamma)$$

ima asimtotski normalnu raspodelu sa sredinom 0 *i disperzijom* $\gamma^2(2^{2\gamma+1} + 1)/(2(2^\gamma - 1)\ln 2)^2$, *za niz* $k = k(n) \to \infty$ *koji zadovoljava* $k(n) = o(n/g^{\leftarrow}(n))$, $n \to \infty$, *gde je* $g^{\leftarrow}$ *uopštena inverz funkcija funkcije*

$$g(t) := t^{3-2\gamma}(U'(t)/(t^{1-\gamma}U'(t) - c^\gamma|\gamma|^{1+\gamma}))^2.$$

Primer 2.3.1. *(Košijeva raspodela) Uslovi Teoreme 2.3.7 su zadovoljeni za* $\rho = 2$ *i* $c = \pi^{-1}$. *Tada je* $g(t) \sim ct^5$, *tako da teorema važi za niz* $k = k(n) \to \infty$ *koji zadovoljava* $k(n) = o(n^{4/5})$.

Primer 2.3.2. *Ako su* Z_1, Z_2 *i.i.d. eksponencijalne slučajne veličine, tada raspodela* $\exp(Z_1 + Z_2)$ *zadovoljava Teoremu 2.3.5 za* $\gamma = 1$ *i* $a(t) = 1$.

Primer 2.3.3. *Za eksponencijalnu i uniformnu raspodelu važi* $t^{1-\gamma}U'(t) \equiv 1$. *Sledi da zaključci Teoreme 2.3.5 važe za sve nizove* $k = k(n) \to \infty$, $k(n)/n \to 0$, $n \to \infty$. *Isto važi za generalisanu Pareto raspodelu*

$$F_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma},\ \gamma \in \mathbb{R},\ 1 + \gamma x \geq 0.$$

Ocena visokih kvantila i krajnje tačke

Sledeća teorema daje uslove asimtotske normalnosti ocene kvantila i krajnje tačke i tako, teorijski može pomoći u konstrukciji intervala poverenja kvantila x_p, kada $p = p_n \to 0$, $np_n \to \infty$, $n \to \infty$.

Teorema 2.3.8. (Dekkers i de Haan, 1989) *Pretpostavlja se da* F *ima pozitivan izvod* F', *tako da* U' *postoji. Ako je* $U' \in RV_{\gamma-1}$, *(tj.* $F' \in RV_{-1-1/\gamma}$ *za* $\gamma > 0$, $1/F' \in \Gamma$ *za* $\gamma = 0$ *i* $F'(x^* - 1/x) \in RV_{1+1/\gamma}$ *za* $\gamma < 0$*), tada je:*

$$\sqrt{2k}\frac{X_{(n-k+1,n)} - U(1/p_n)}{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}$$

asimtotski normalno sa sredinom 0 *i disperzijom* $2^{2\gamma+1}\gamma^2/(2^\gamma - 1)^2$, *uz uslov* $p_n \to 0$, $np_n \to \infty$, $n \to \infty$ *i* $k = k(n) := [np_n]$.

Teorema 2.3.9. (Dekkers i de Haan, 1989) *Pretpostavlja se da važe uslovi Teoreme 2.3.7 ili Teoreme 2.3.5 sa* $\gamma < 0$. *Tada je* $x^* < +\infty$, *gde je* $x^* = x^*(F) := \sup\{x|F(x) < 1\}$. *Definiše se:*

$$\widehat{x}_n^* := \frac{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}{2^{-\widehat{\gamma}_n} - 1} + X_{(n-k+1,n)}. \tag{2.19}$$

Tada:

$$\sqrt{2k}\frac{\widehat{x}_n^* - x^*}{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}$$

ima asimtotski normalnu raspodelu sa sredinom 0 *i disperzijom* $\frac{3\gamma^2 2^{2\gamma-1}}{(2^\gamma-1)^6}$.

Slede dve teoreme o oceni (2.17) za fiksirano k.

Teorema 2.3.10. (Dekkers i de Haan, 1989) *Pretpostavlja se da važi uslov ekstremne vrednosti (2.1),* $p = p_n$ *i* $\lim_{n\to\infty} np_n = c \in (0, +\infty)$. *Definiše se* $\widehat{x}_{p_n,n}$ *kao u (2.17), tj.:*

$$\widehat{x}_{p_n,n} := \frac{(k/(np_n))^{\widehat{\gamma}_n} - 1}{1 - 2^{-\widehat{\gamma}_n}}(X_{(n-k+1,n)} - X_{(n-2k+1,n)}) + X_{(n-k+1,n)}.$$

Tada, za fiksirano $k > c$,

$$\frac{\widehat{x}_{p_n,n} - x_{p_n}}{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}$$

konvergira u raspodeli ka raspodeli slučajne veličine:

$$\left[\left(\frac{k}{c}\right)^\gamma - 2^{-\gamma}\right]/(1 - 2^{-\gamma}) + \left(1 - \left(\frac{Q_k}{c}\right)^\gamma\right)/(e^{\gamma H_k} - 1),$$

gde su H_k *i* Q_k *nezavisne,* Q_k *ima standardnu gama* $(2k + 1)$ *raspodelu i* H_k *raspodelu sume* $\sum_{j=k+1}^{2k} Z_j/j$, *gde su* $Z_1, Z_2, ...$ *i.i.d. standardne eksponencijalne slučajne veličine.*

Teorema 2.3.11. (Dekkers i de Haan, 1989) *Pretpostavlja se da važi uslov ekstremne vrednosti (2.1) sa* $\gamma < 0$. *Tada je* $x^* < +\infty$, *gde je* $x^* = x^*(F) := \sup\{x|F(x) < 1\}$. *Definiše se* $\widehat{x}_n^*$ *kao u jednakosti (2.19). Tada:*

$$\frac{\widehat{x}_n^* - x^*}{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}$$

konvergira u raspodeli ka raspodeli slučajne veličine $(1-2^\gamma)^{-1} + (e^{\gamma H_k} - 1)^{-1}$, *za* H_k *iz Teoreme 2.3.10.*

2.3.3 Ocena momenta

Pretpostavlja se da je $x^* = x^*(F) > 0$, za $x^*(F) := \sup\{x|F(x) < 1\}$ (što se može postići translacijom) i definiše se:

$$M_n^{(1)} : \; = \; \frac{1}{k}\sum_{i=1}^{k} \ln X_{(n-i+1,n)} - \ln X_{(n-k,n)},\ k < n,$$
$$M_n^{(2)} : \; = \; \frac{1}{k}\sum_{i=1}^{k} (\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)})^2.$$

Data je sledeća ocena momenta, u radu Dekkers et al., 1975:

$$\widehat{\gamma}_n := M_n^{(1)} + 1 - \frac{1}{2}\Big(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\Big)^{-1}.$$

Pod određenim uslovima moguće je pokazati slabu i jaku konzistentnost ove ocene, što je prikazano u narednom delu disertacije, a takođe su dati uslovi pod kojima je ocena asimtotski normalna. Može se koristiti ocena momenta za dobijanje asimtotskog intervala poverenja za visoke kvantile funkcije F i za $x^*(F)$, u slučaju kada je $\gamma < 0$.

Slaba i jaka konzistentnost

Teorema 2.3.12. (Dekkers et al., 1989) *Ako važi (2.1), $x^*(F) > 0$, $k(n)/n \to 0$ i $k(n) \to \infty$, $n \to \infty$, tada:*

$$\lim_{n\to\infty} \widehat{\gamma}_n = \gamma,\ p.$$

Ako važi (2.1), $x^*(F) > 0$, $k(n)/n \to 0$ i $k(n)/(\ln n)^\delta \to \infty$, $n \to \infty$, za neko $\delta > 0$, tada:

$$\lim_{n\to\infty} \widehat{\gamma}_n = \gamma,\ s.s.$$

Za dokaz prethodne teoreme potrebne su sledeće leme, pogledati Dekkers et al., 1989.

Lema 2.3.3. *Pretpostavlja se da su $U_1, U_2, ...$ i.i.d. sa uniformnom $[0,1]$ raspodelom. Neka je $\Gamma_n(t)$ empirijska funkcija raspodele bazirana na $U_1, ..., U_n$*

(n=1,2,...). Tada za $0 < k(n) \leq n$, $k(n)/(\ln n)^{\delta} \to \infty$ *za neko* $\delta > 0$ *i* $a < \delta/(2(1+\delta))$*:*

$$\lim_{n\to\infty} \left(\frac{n}{k(n)}\right)^{1-a} \int_0^{k(n)/n} t^{-a-1}(\Gamma_n(t) - t)dt = 0, \ s.s.$$

Lema 2.3.4. *Neka je* $0 < k(n) \leq n$ *i* $k(n)/(\ln n)^{\delta} \to \infty$, $n \to \infty$ *za neko* $\delta > 0$.

(a) *Ako je funkcija raspodele oblika* $F(x) = x^{\alpha}$ $(0 < x < 1)$ *za neko* $\alpha > 0$, *onda važi:*

$$\lim_{n\to\infty} \frac{1}{k(n)} \sum_{i=1}^{k(n)} \frac{X_{(i,n)}}{X_{(k(n)+1,n)}} = \frac{\alpha}{\alpha+1}, \ \ s.s.$$

(b) *Ako je funkcija raspodele oblika* $F(x) = 1 - x^{-\alpha}$ $(x > 1)$ *za neko* $\alpha > 2(1+\delta)/\delta$, *onda važi:*

$$\lim_{n\to\infty} \frac{1}{k(n)} \sum_{i=0}^{k(n)-1} \frac{X_{(n-i,n)}}{X_{(n-k(n),n)}} = \frac{\alpha}{\alpha-1}, \ \ s.s.$$

Lema 2.3.5. *Neka je* $0 < k(n) \leq n$ *i* $k(n) \to \infty$, $n \to \infty$.

(a) *Ako je funkcija raspodele oblika* $F(x) = x^{\alpha}$ $(0 < x < 1)$ *za neko* $\alpha > 0$, *onda važi:*

$$\lim_{n\to\infty} \frac{1}{k(n)} \sum_{i=1}^{k(n)} \frac{X_{(i,n)}}{X_{(k(n)+1,n)}} = \frac{\alpha}{\alpha+1}, \ \ p.$$

(b) *Ako je funkcija raspodele oblika* $F(x) = 1 - x^{-\alpha}$ $(x > 1)$ *za neko* $\alpha > 1$, *onda važi:*

$$\lim_{n\to\infty} \frac{1}{k(n)} \sum_{i=0}^{k(n)-1} \frac{X_{(n-i,n)}}{X_{(n-k(n),n)}} = \frac{\alpha}{\alpha-1}, \ \ p.$$

Lema 2.3.6. *Pretpostavlja se da važi (2.1) i* $x^*(F) > 0$. *Neka je* $U = (1/(1-F))^{\leftarrow}$. *Tada, za neku pozitivnu funkciju a:*

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t)}{a(t)/U(t)} = \begin{cases} \ln x, & \gamma \geq 0, \\ \frac{x^{\gamma}-1}{\gamma}, & \gamma < 0, \end{cases}$$

za sve $x > 0$. *Šta više, za svako* $\epsilon > 0$ *postoji* t_0 *tako da je, za* $t \geq t_0$ *i* $x \geq 1$,

(a)

$$(1-\epsilon)\frac{1-x^{-\epsilon}}{\epsilon}-\epsilon<\frac{\ln U(tx)-\ln U(t)}{a(t)/U(t)}<(1+\epsilon)\frac{x^{\epsilon}-1}{\epsilon}+\epsilon,$$

uz uslov $\gamma \geq 0$, *i*

(b)

$$1-(1+\epsilon)x^{\gamma+\epsilon}<\frac{\ln U(tx)-\ln U(t)}{\ln U(+\infty)-\ln U(t)}<1-(1-\epsilon)x^{\gamma-\epsilon},$$

uz uslov $\gamma < 0$.

Asimtotska normalnost

Teorema 2.3.13. (Dekkers et al., 1989) *Pretpostavlja se da važi (2.1) i* $U := (1/(1-F))^{\leftarrow}$.

(a) *Za* $\gamma > 0$:

$$\pm t^{-\gamma}U(t) \in \Pi(b_1),$$

za neku pozitivnu funkciju b_1.

(b) *Za* $\gamma = 0$: *Postoje pozitivne funkcije* b_2 *i* b_3 *tako da je:*

$$\lim_{t\to+\infty}\frac{\ln U(tx)-\ln U(t)-b_2(t)\ln x}{b_3(t)}=\pm\frac{(\ln x)^2}{2}.$$

(c) *Za* $\gamma < 0$:

$$\mp t^{-\gamma}(U(+\infty)-U(t)) \in \Pi(b_4),$$

za neku pozitivnu funkciju b_4.

Pretpostavlja se takođe da je $\lim\limits_{n\to\infty} k(n) = +\infty$ *i:*

(d) *Za* $\gamma > 0$:

$$k(n) = o(n/g^{\leftarrow}(n)),$$

gde je $g(t) := t^{1-2\gamma}(U(t)/b_1(t))^2$.

(e) *Za* $\gamma = 0$:

$$k(n) = o(n/g^{\leftarrow}(n)),$$

gde je $g(t) := tb_2^2(t)/b_3^2(t)$.

(f) *Za* $\gamma < 0$*:*

$$k(n) = o(n/g^{\leftarrow}(n)),$$

gde je $g(t) := t^{1-2\gamma}[(\ln U(+\infty) - \ln U(t))/b_4(t)]^2$.

Tada:

$$\sqrt{k(n)}\Big(\frac{M_n^{(1)}}{f(\ln X_{(n-k(n),n)})} - \rho_1(\gamma), \frac{M_n^{(2)}}{(f(\ln X_{(n-k(n),n)}))^2} - \rho_2(\gamma)\Big),$$

za $f(t) := a(1/(1-F(e^t)))/U(1/(1-F(e^t)))$ *ima asimtotski normalnu raspodelu, za* $n \to \infty$*, sa sredinama* 0 *i kovarijacionom matricom* (s_{ij}) *sa, za* $\gamma \leq 0$*,*

$$\begin{aligned} s_{11} &= (1-\gamma)^{-2}(1-2\gamma)^{-1}, \\ s_{12} &= 4(1-\gamma)^{-2}(1-2\gamma)^{-1}(1-3\gamma)^{-1}, \\ s_{22} &= 4(5-11\gamma)(1-\gamma)^{-2}(1-2\gamma)^{-2}(1-3\gamma)(1-4\gamma), \end{aligned}$$

i za $\gamma \geq 0$*,*

$$s_{11} = 1,\ s_{12} = 4,\ s_{22} = 20.$$

Funkcije ρ_1 *i* ρ_2 *su definisane:*

$$\begin{aligned} \rho_1(\gamma) &= \begin{cases} 1, & \gamma \geq 0, \\ 1/(1-\gamma), & \gamma < 0, \end{cases} \\ \rho_2(\gamma) &= \begin{cases} 2, & \gamma \geq 0, \\ 2/((1-\gamma)(1-2\gamma)), & \gamma < 0. \end{cases} \end{aligned}$$

Posledica 2.3.3. *Ako su ispunjeni uslovi Teoreme 2.3.13 i ako je, za slučaj* $\gamma = 0$*,*

$$k(n) = o(n/g_1^{\leftarrow}(n)),$$

gde je $g_1(t) = t(U(t)/a(t))^2$*, tada:*

$$\sqrt{k(n)}(\widehat{\gamma}_n - \gamma)$$

ima asimtotski normalnu raspodelu sa sredinom 0 *i disperzijom*

$$\begin{cases} 1+\gamma^2, & \gamma \geq 0, \\ (1-\gamma)^2(1-2\gamma)\Big(4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\Big), & \gamma < 0. \end{cases}$$

Teorema 2.3.14. (Dekkers et al., 1989) *Uslovi (a), (b), (c) Teoreme 2.3.13 povlače (2.1) za isto γ. Uslovi (a), (b), (c) Teoreme 2.3.13 su ekvivalentni, redom:*

(a) *Za $\gamma > 0$:*
$$\mp t^{1/\gamma}(1 - F(t)) \in \Pi.$$

(b) *Za $\gamma = 0$: Postoje pozitivne funkcije f i α takve da je $\lim_{t\to x^*} \alpha(t) = 0$ i da je za $x > 0$:*
$$\lim_{t\to x^*} \frac{\frac{1-F(e^{t+xf(t)})}{1-F(e^t)} - e^{-x}}{\alpha(t)} = \pm\frac{x^2}{2}e^{-x}.$$

(c) *Za $\gamma < 0$:*
$$\pm t^{-1/\gamma}(1 - F(x^* - t^{-1})) \in \Pi.$$

Za dokaz Teoreme 2.3.13 potrebne su sledeće leme. Dokaz se može pogledati u Dekkers et al., 1989.

Lema 2.3.7. *Neka su $Y_{(1,n)} \le ... \le Y_{(n,n)}$ statistike poretka raspodele $1 - x^{-1}$ $(x > 1)$. Neka je $0 < k(n) \le n$ i $k(n) \to \infty$, $n \to \infty$.*

(a)
$$\sqrt{k(n)}\Big(\frac{1}{k(n)}\sum_{i=0}^{k(n)-1} \ln Y_{(n-i,n)} - \ln Y_{(n-k(n),n)} - 1,$$
$$(20)^{-1/2}\Big(\frac{1}{k(n)}\sum_{i=0}^{k(n)-1} (\ln Y_{(n-i,n)} - \ln Y_{(n-k(n),n)})^2 - 2\Big)\Big)$$

ima asimtotski normalnu rasodelu, za $n \to \infty$, sa sredinama 0, disperzijom 1 i kovarijansom $2\sqrt{5}$.

(b) *Za $\gamma < 0$:*
$$\sqrt{k(n)}\Big(\frac{1}{k(n)}\sum_{i=0}^{k(n)-1} 1 - \Big(\frac{Y_{(n-i,n)}}{Y_{(n-k(n),n)}}\Big)^{\gamma} + \frac{\gamma}{1-\gamma},$$
$$\frac{1}{k(n)}\sum_{i=0}^{k(n)-1} \Big(1 - \Big(\frac{Y_{(n-i,n)}}{Y_{(n-k(n),n)}}\Big)^{\gamma}\Big)^2 - \frac{2\gamma^2}{(1-\gamma)(1-2\gamma)}\Big)$$

ima asimtotski normalnu raspodelu, za $n \to \infty$, sa sredinama 0, disperzijama γ^2 i γ^4 redom, i kovarijansom

$$\frac{2\gamma^3(5-30\gamma+40\gamma^2)^{1/2}}{\sqrt{5}(5-26\gamma+33\gamma^2)^{1/2}}.$$

Lema 2.3.8. *Pretpostavlja se da važi uslov (a), (b) ili (c) Teoreme 2.3.13 sa gornjim znakom (tj. + za $\gamma \geq 0$ i - za $\gamma < 0$). Za bilo koje $\epsilon > 0$ postoji t_0 tako da je za svako $t \geq t_0$ i $x \geq 1$:*

(a) *Za slučaj $\gamma > 0$:*

$$\begin{aligned}(1-\epsilon)\frac{1-x^{-\epsilon}}{\epsilon}-\epsilon &< \frac{\ln U(tx)-\ln U(t)-\gamma\ln x}{t^\gamma b_1(t)/U(t)} \\ &< (1+\epsilon)\frac{x^\epsilon-1}{\epsilon}+\epsilon.\end{aligned}$$

(b) *Za slučaj $\gamma = 0$:*

$$\begin{aligned}\frac{(1-\epsilon^2)(\ln x)^2}{2}-2\epsilon\ln x-\epsilon &< \frac{\ln U(tx)-\ln U(t)-b_2(t)\ln x}{b_3(t)} \\ &< \frac{(1+\epsilon)^2 x^\epsilon(\ln x)^2}{2}+2\epsilon\ln x+\epsilon.\end{aligned}$$

(c) *Za slučaj $\gamma < 0$:*

$$\begin{aligned}(1-\epsilon)x^\gamma\frac{1-x^{-\epsilon}}{\epsilon}-\epsilon x^\gamma &< \frac{\ln U(tx)-\ln U(t)-(1-x^\gamma)(\ln U(+\infty)-\ln U(t))}{t^\gamma b_4(t)/U(+\infty)} \\ &< (1+\epsilon)x^\gamma\frac{x^\epsilon-1}{\epsilon}+\epsilon x^\gamma.\end{aligned}$$

Ocena kvantila i krajnje tačke: konačan slučaj.

Za ocenu indeksa ekstremne vrednosti γ i visokih kvantila, u radu Dekkers i de Haan, 1989 koristi se razlika visokih statistika poretka. Slične ocene za visoke kvantile koje koriste sumu visokih statistika poretka, predložene su u radu Dekkers et al., 1989.

Kao i ranije, pretpostavlja se da su $X_1, ..., X_n$ i.i.d. sa funkcijom raspodele F, koja zadovoljava (2.1). Cilj je da se nađe nivo x_p, za p dosta manje od 1, tako da je:

$$F(x_p) = 1 - p.$$

Sa ranije definisanom funkcijom U, važi:

$$x_p = U\Big(\frac{1}{p}\Big).$$

Dekkers i de Haan, 1989 su predložili ocenu x_p:

$$\widehat{x}_{p,n} := \frac{a_n^{\widehat{\gamma}_n} - 1}{\widehat{\gamma}_n} \cdot \frac{X_{(n-k,n)} M_n^{(1)}}{\rho_1(\widehat{\gamma}_n)} + X_{(n-k,n)},$$

gde je $\widehat{\gamma}_n$ neka konzistentna ocena γ, $M_n^{(1)}$ i ρ_1 su ranije definisani i:

$$a_n := \frac{k}{np}.$$

Asimtotski interval poverenja za x_p može se konstruisati koristeći sledeći rezultat, što može biti pravac budućeg istraživanja.

Teorema 2.3.15. (Dekkers et al., 1989) *Pretpostavlja se da $p = p_n \to 0$, $np_n \to c \in (0, +\infty)$, $n \to \infty$. Neka je k fiksirano, $k > c$. Tada, uz uslov (2.1), važi:*

$$\frac{\widehat{x}_{p,n} - x_p}{X_{(n-k,n)} M_n^{(1)}} \to_d$$
$$\to_d \begin{cases} \dfrac{\left(\frac{k}{c}\right)^{\gamma} - 1}{\gamma \rho_1(\gamma)} + \dfrac{1 - \left(\frac{1}{c} \cdot Q_k\right)^{\gamma}}{\gamma} \Big/ \Big(\frac{1}{k} \sum_{i=0}^{k-1} Z_i\Big), & \gamma \geq 0 \\ \dfrac{\left(\frac{k}{c}\right)^{\gamma} - 1}{\gamma \rho_1(\gamma)} + \dfrac{1 - \left(\frac{1}{c} \cdot Q_k\right)^{\gamma}}{\gamma} \Big/ \Big(\frac{1}{k} \sum_{i=0}^{k-1} \frac{\exp(\gamma \sum_{j=i}^{k-1} Z_j / j) - 1}{\gamma}\Big), & \gamma < 0 \end{cases},$$

kada $n \to \infty$, za Q_k, $Z_0, Z_1, ...Z_{k-1}$ nezavisne, Q_k gama sa k stepeni slobode i Z_i, $i = 0, 1, ..., k-1$ su i.i.d. eksponencijalne slučajne veličine.

Dokaz prethodne teoreme bazira se na sledećoj lemi.

Lema 2.3.9. (Beirlant i Teugels, 1986) *Pod uslovima i oznakama Teoreme 2.3.15 važi:*

$$\frac{M_n^{(1)}}{a(n)/U(n)} \to_d \begin{cases} \frac{1}{k}\sum_{i=0}^{k-1} Z_i, & \gamma \geq 0 \\ Q_k^{-\gamma}\frac{1}{k}\sum_{i=0}^{k-1}\frac{\exp(\gamma\sum_{j=i}^{k-1} Z_j/j)-1}{\gamma}, & \gamma < 0 \end{cases}.$$

U slučaju kada je $\gamma < 0$ prethodni rezultat se može prilagoditi za graničnu situaciju kada je $p = 0$, da bi se dobio interval poverenja za gornju krajnju tačku raspodele, $x^* = x^*(F) = U(+\infty)$.

Teorema 2.3.16. (Dekkers et al., 1989) *Pretpostavlja se da važi (2.1) sa $\gamma < 0$. Tada je $x^* = x^*(F) := \sup\{x|F(x) < 1\}$ konačno. Neka je:*

$$\widehat{x}_n^* := X_{(n-k,n)} M_n^{(1)}\Big(1 - \frac{1}{\widehat{\gamma}_n}\Big) + X_{(n-k,n)}. \tag{2.20}$$

Pod uslovima Teoreme 2.3.15, važi:

$$\frac{\widehat{x}_n^* - x^*}{X_{(n-k,n)} M_n^{(1)}} \to_d \Big(1 - \frac{1}{\gamma}\Big)\Big(\frac{1}{k}\sum_{i=0}^{k-1}\exp\Big(\gamma\sum_{j=i}^{k-1}\frac{Z_j}{j}\Big) - 1\Big)^{-1}.$$

Ocena kvantila i krajnje tačke: beskonačan slučaj.

Ovde se može razmotriti ocena x_p za $n \to \infty$, kada je dozvoljeno da broj statistika poretka k i $M_n^{(1)}$ rastu bez ograničenja. Sledeće teoreme mogu se iskoristiti za konstrukciju intervala poverenja kvantila x_p, kada $p_n \to 0$, $np_n \to \infty$, $n \to \infty$, što je predmet budućeg istraživnja.

Teorema 2.3.17. (Dekkers et al., 1989) *Neka F ima pozitivan izvod F', tako da U' postoji. Ako je $U' \in RV_{\gamma-1}$ (tj. $F' \in RV_{-1/\gamma-1}$ za $\gamma > 0$, $1/F' \in \Gamma$ za $\gamma = 0$ i $F'(x^* - 1/x) \in RV_{1/\gamma+1}$ za $\gamma < 0$), tada:*

$$\sqrt{k(n)}\frac{X_{(n-k(n),n)} - U(1/p_n)}{X_{(n-k(n),n)} M_n^{(1)}}$$

ima asimtotski normalnu raspodelu sa sredinom 0 *i disperzijom* $(1-\min(0,\gamma))^2$, *uz uslov $p_n \to 0$, $np_n \to \infty$, $n \to \infty$ i $k(n) := [np_n]$.*

Sledi teorema o oceni krajnje tačke raspodele.

Teorema 2.3.18. (Dekkers et al., 1989) *Neka* $k = k(n) \to \infty$ *i* $k(n)/n \to 0$, $n \to \infty$. *Pretpostavlja se da važe uslovi Teoreme 2.3.13 sa* $\gamma < 0$. *Pretpostavlja se još da* U *ima pravilno promenljivi izvod* U'. *Tada, za* $\widehat{x}_n^*$ *definisano u (2.20):*

$$\sqrt{k(n)}\frac{\widehat{x}_n^* - x^*}{X_{(n-k(n),n)}M_n^{(1)}(1-\widehat{\gamma}_n)}$$

ima asimtotski normalnu raspodelu sa sredinom 0 *i disperzijom*

$$\frac{1}{\gamma^2}\Big[\frac{1}{1-2\gamma} + \frac{1-2\gamma}{\gamma^2}\Big(4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\Big) - \frac{4}{1-3\gamma}\Big].$$

Za dokaz prethodne teoreme potrebna je sledeća lema.

Lema 2.3.10. (Dekkers et al., 1989) *Pretpostavlja se da važe uslovi Teoreme 2.3.18. Koristeći funkciju* U *iz Leme 2.3.6, slučajni vektor:*

$$\sqrt{k}\Big(\frac{X_{(n-k,n)}M_n^{(1)}}{-\gamma(U(+\infty)-U(n/k))} - (1-\gamma)^{-1}, \widehat{\gamma}_n - \gamma, \frac{X_{(n-k,n)} - U(n/k)}{-\gamma(U(+\infty)-U(n/k))}\Big)$$

ima asimtotski normalnu raspodelu sa sredinama 0 *i kovarijacionom matricom* (s_{ij}) *sa:*

$$\begin{aligned}
s_{11} &= \frac{1+\gamma^2(1-2\gamma)}{(1-\gamma)^2(1-2\gamma)}, \;\; s_{12} = -2 + \frac{2(1-2\gamma)}{(1-3\gamma)}, \\
s_{13} &= \frac{\gamma}{(1-\gamma)}, \, s_{22} = (1-\gamma)^2(1-2\gamma)\Big[4 - \frac{8(1-2\gamma)}{(1-3\gamma)} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\Big], \\
s_{23} &= 0, \;\; s_{33} = 1.
\end{aligned}$$

2.3.4 Ocena maksimalne verodostojnosti za eksponencijalno regresioni model

U ovom odeljku predstavljene su dve nelinearne eksponencijalne regresione metode koje daju ocene maksimalne verodostojnosti indeksa ekstremne vrednosti i visokih kvantila. Prva je log-odnos prostora statistika poretka, dok je druga prostor statistika poretka.

Dalje se koriste sledeće oznake: konvergencija u raspodeli u oznaci $=_d$, $U_{(j,n)}$ $(1 \leq j \leq n)$ su statistike poretka i.i.d. iz uniformne (0,1) raspodele, $V_{(j,k)}$ $(1 \leq j \leq k)$ su takođe statistike poretka i.i.d. iz uniformne (0,1) raspodele i $E_{(j,k)}$ $(1 \leq j \leq k)$ su statistike poretka i.i.d. sa standarnom eksponencijalnom raspodelom, $U(r) := F^{\leftarrow}(1 - 1/r) = x_{1/r}$,

$$\begin{aligned} X_{(n-j+1,n)} &=_d U(U^{-1}_{(j,n)}),\ j \leq n, \\ \frac{U_{(j,n)}}{U_{(k+1,n)}} &=_d V_{(j,k)},\ j \leq k < n, \\ -\ln(V_{(j,k)}) &=_d E_{(k-j+1,k)},\ j \leq k. \end{aligned}$$

Slučajne veličine $V_{(j,k)}$ su nezavisne sa $U_{(k+1,n)}$. Ključni rezultat, koji se kasnije koristiti, jeste Rényi reprezentacija standardnih eksponencijalnih statistika poretka:

$$E_{(k-j+1,k)} =_d \sum_{i=j}^{k} \frac{f_{k-i+1}}{i},\ j \leq k, \tag{2.21}$$

gde su f_i i.i.d. sa standardnom eksponencijalnom raspodelom.

Pored rezultata (2.1) i (2.7) može se formulisati ekvivalentan uslov pomoću funkcije kvantila repa U. Sledeći rezultat dao je de Haan, 1970, koji kaže da $F \in MDA(H_\gamma)$ akko postoji pozitivna merljiva funkcija a_U, takva da je za svako $t > 0$:

$$\lim_{r \to +\infty} \frac{U(tr) - U(r)}{a_U(r)} = \begin{cases} \frac{t^\gamma - 1}{\gamma}, & \gamma \neq 0, \\ \ln t, & \gamma = 0. \end{cases} \tag{2.22}$$

Normirajuća funkcija $a_U(r)$ ekvivalentna je funkciji $\sigma(U(r))$, za $r \to +\infty$, gde je funkcija $\sigma(\cdot)$ kao u (2.7). Može se primetiti da za GPD (2.6) važi $\sigma(u) = \sigma + \gamma(u - \mu)$, $U(r) = \mu + \sigma(r^\gamma - 1)/\gamma$ i jednakost (2.22) važi za sve $t > 0$ i sve $r > 0$ sa funkcijom $a_U(r) = \sigma r^\gamma = \sigma(U(r))$.

Uslov (2.22) izražen pomoću funkcije kvantila repa, za fiksirano $k < n$ i $1 \leq j \leq k$, daje sledeću aproksimaciju:

$$\begin{aligned} X_{(n-j+1,n)} - X_{(n-k,n)} &=_d U(U^{-1}_{(j,n)}) - U(U^{-1}_{(k+1,n)}) \qquad (2.23) \\ &=_d U(V^{-1}_{(j,k)} U^{-1}_{(k+1,n)}) - U(U^{-1}_{(k+1,n)}) \\ &\approx_d a_{n,k+1} \frac{V^{-\gamma}_{(j,k)} - 1}{\gamma}, \end{aligned}$$

gde je $a_{n,k+1}$ umesto $a_U(U^{-1}_{(k+1,n)})$.

Kako je $a_{n,k+1} = a_U(U^{-1}_{(k+1,n)}) \sim \sigma(X_{(n-k,n)})$, (2.23) je kopija GPD aproksimacije repa (2.7), za $u = X_{(n-k,n)}$, izraženo pomoću statistika poretka.

Za prvu metodu log-odnosa prostora statistika poretka, može se dobiti:

$$\ln \frac{X_{(n-j+1,n)} - X_{(n-k,n)}}{X_{(n-j,n)} - X_{(n-k,n)}} \approx_d \ln \frac{V^{-\gamma}_{(j,k)} - 1}{V^{-\gamma}_{(j+1,k)} - 1}, \ 1 \le j < k.$$

Primenjujući Teoremu o srednjoj vrednosti funkcije na desnu stranu prethodnog izraza, koristeći $E^*_{(j,k)} \in (E_{(k-j,k)}, E_{(k-j+1,k)})$ i $V^*_{(j,k)} = \exp(-E^*_{(j,k)})$, dobija se:

$$\begin{aligned} \ln \frac{V^{-\gamma}_{(j,k)} - 1}{V^{-\gamma}_{(j+1,k)} - 1} &=_d \ln(e^{\gamma E_{(k-j+1,k)}} - 1) - \ln(e^{\gamma E_{(k-j,k)}} - 1) \\ &=_d (E_{(k-j+1,k)} - E_{(k-j,k)}) \cdot \frac{\gamma e^{\gamma E^*_{(j,k)}}}{e^{\gamma E^*_{(j,k)}} - 1} \\ &=_d \frac{f_{k-j+1}}{j} \cdot \frac{\gamma}{1 - (V^*_{(j,k)})^{\gamma}}. \end{aligned}$$

Jasno je da poslednja jednakost sledi iz Rényi reprezentacije standardnih eksponencijalnih statistika poretka, (2.21). Posle ocene $V^*_{(j,k)}$ sa $j/(k+1)$, sledi sledeći nelinearni eksponencijalni regresioni model za log-odnos prostora:

$$j \ln \frac{X_{(n-j+1,n)} - X_{(n-k,n)}}{X_{(n-j,n)} - X_{(n-k,n)}} \approx_d \frac{\gamma}{1 - (\frac{j}{k+1})^{\gamma}} f_{k-j+1}, 1 \le j < k, \tag{2.24}$$

gde je f_j $(1 \le j \le k)$ niz definisan ranije.

Ocena maksimalne verodostojnosti za indeks ekstremne vrednosti $\widehat{\gamma}^A_{k+1}$, za fiksirano $k < n$, može se izračunati na bazi modela (2.24), računajući log-odnos prostora $Y_j := j \ln \frac{X_{(n-j+1,n)} - X_{(n-k,n)}}{X_{(n-j,n)} - X_{(n-k,n)}}$ za $1 \le j < k$ i maksimiziranjem:

$$L_\gamma(Y) := \sum_{j=1}^{k-1} \Big(\ln \Big(\frac{1 - (\frac{j}{k+1})^{\gamma}}{\gamma} \Big) - \frac{1 - (\frac{j}{k+1})^{\gamma}}{\gamma} Y_j \Big).$$

Dobijena ocena $\widehat{\gamma}^A_{k+1}$ je invarijantna u odnosu na translacije i reskaliranja podataka. Ocena je takođe konzistentna i ima asimtotski normalnu raspodelu, pod nekim tehničkim uslovima na repu raspodele, što se može pogledati u Matthys i Beirlant, 2003. Za slučaj kada je $\gamma = 0$ asimtotska disperzija za $\sqrt{k}(\widehat{\gamma}^A_{k+1} - \gamma)$ jeste 1, kao i u slučaju ocene momenta i POT ocene maksimalne verodostojnosti. Za druge vrednosti γ numerički se može izračunati. U radu Matthys i Beirlant, 2003 upoređivane su vrednosti asimtotske disperzije za $\widehat{\gamma}^A_{k+1}$ sa disperzijom ocene momenta i POT ocene maksimalne verodostojnosti za $-3 \leq \gamma \leq 3$ $(-1/2 < \gamma \leq 3$ za POT ocenu maksimalne verodostojnosti). Vidi se da je za $\gamma > 0$ asimtotska disperzija ocene $\widehat{\gamma}^A_{k+1}$ skoro jednaka disperziji POT ocene maksimalne verodostojnosti. Za $\gamma < 0$ asimtotska disperzija ocene $\widehat{\gamma}^A_{k+1}$ je niža od asimtotske disperzije ocene momenta.

Takođe u radu Matthys i Beirlant, 2003 data je vrednost asimtotskog biasa ocene $\widehat{\gamma}^A_{k+1}$, koja se može numerički izračunati. Pokazano je da je ta vrednost jednaka vrednosti asimtotskog biasa POT ocene maksimalne verodostojnosti, za slučaj kada je $\gamma > 0$. U istom radu mogu se videti detalji o bias korekciji i izboru praga u, što u ovoj disertaciji nije razmatrano.

Skoro kao pravilo treba posmatrati, za svaku ocenu indeksa ekstremne vrednosti, odnos disperzije i biasa. Disperzija se smanjuje kada se za ocenu uzima više statistika poretka, dok bias raste kada se prag izabere na visokom nivou. Vrednost gde je srednjekvadratna greška ocene minimalna može se uzeti kao optimalni izbor za prag.

Dalje je predstavljen drugi eksponencijalni regresioni model, koji daje ocenu faktora $a_{n,k+1} = a_U(U^{-1}_{k+1,n})$, što je potrebno za ocenu visokih kvantila. Umesto do sada razmatranog log-odnosa, ovde se koristi (2.23) kako bi se aproksimirao prostor statistika poretka direktno:

$$X_{(n-j+1,n)} - X_{(n-j,n)} \approx_d a_{n,k+1} \frac{V^{-\gamma}_{(j,k)} - V^{-\gamma}_{(j+1,k)}}{\gamma}, \; 1 \leq j < k.$$

Posle primene Teoreme o srednjoj vrednosti na desnu stranu prethodnog izraza, koristeći raniju notaciju za $E^*_{(j,k)}$ i $V^*_{(j,k)}$, dobija se:

$$\begin{aligned}
\frac{V_{(j,k)}^{-\gamma} - V_{(j+1,k)}^{-\gamma}}{\gamma} &=_d \frac{e^{\gamma E_{(k-j+1,k)}} - e^{\gamma E_{(k-j,k)}}}{\gamma} \\
&=_d (E_{(k-j+1,k)} - E_{(k-j,k)})e^{\gamma E^*_{(j,k)}} \\
&=_d \frac{f_{k-j+1}}{j} \cdot (V^*_{(j,k)})^{-\gamma}.
\end{aligned}$$

Posle ocene $V^*_{(j,k)}$ sa $j/(k+1)$, dobija se nelinearni regresioni model:

$$j(X_{(n-j+1,n)} - X_{(n-j,n)}) \approx_d a_{n,k+1}\Big(\frac{j}{k+1}\Big)^{-\gamma} f_{k-j+1},\ 1 \le j \le k, \tag{2.25}$$

gde je f_j $(1 \le j \le k)$ niz definisan ranije.

Posle maksimiziranja odgovarjuće funkcije prostora $Z_j := j(X_{(n-j+1,n)} - X_{(n-j,n)})$, za $1 \le j \le k$, na osnovu modela (2.25), može se dobiti ocena za $a_{n,k+1}$:

$$\widehat{a}_{n,k+1} = \frac{1}{k}\sum_{j=1}^{k} Z_j \Big(\frac{j}{k+1}\Big)^{\gamma}. \tag{2.26}$$

Iz jednakosti (2.22) dobija se aproksimacija za visoki kvantil:

$$\begin{aligned}
x_p - X_{(n-k,n)} &=_d U(p^{-1}) - U(U^{-1}_{(k+1,n)}) \\
&=_d U((U_{(k+1,n)}/p)U^{-1}_{(k+1,n)}) - U(U^{-1}_{(k+1,n)}) \\
&\approx_d \widehat{a}_{n,k+1}\frac{(U_{(k+1,n)}/p)^{\gamma} - 1}{\gamma}.
\end{aligned}$$

Sada se $U_{(k+1,n)}$ ocenjuje očekivanom vrednošću $(k+1)/(n+1)$, γ sa $\widehat{\gamma}^A_{k+1}$ i $a_{n,k+1}$ se ocenjuje zamenom $\widehat{\gamma}^A_{k+1}$ u formulu (2.26):

$$\widehat{a}^A_{n,k+1} := \frac{1}{k}\sum_{j=1}^{k} j(X_{(n-j+1,n)} - X_{(n-j,n)})\Big(\frac{j}{k+1}\Big)^{\widehat{\gamma}^A_{k+1}}.$$

Tako se dobija ocena kvantila:

$$\widehat{x}^A_{p,k+1} := X_{(n-k,n)} + \widehat{a}^A_{n,k+1}\frac{\Big(\frac{k+1}{p(n+1)}\Big)^{\widehat{\gamma}^A_{k+1}} - 1}{\widehat{\gamma}^A_{k+1}},\ k < n.$$

2.4 Još neke, već predložene ocene indeksa ekstremne vrednosti

Ocena indeksa ekstremne vrednosti pomoću metode MVRB, minimalne-varijanse redukovanog-biasa u okviru trećeg reda, od skora je proučavana u literaturi. Ona je dalje u disertaciji predstavljena. Pokazano je da ocena indeksa ekstremne vrednosti zavisi od adekvatnih ocena parametara drugog reda. Proučavana je u radovima Gomes et al., 2004a, Caeiro et al., 2005 i Gomes et al., 2007a. Asimtotska disperzija ove ocene jednaka je asimtotskoj disperziji klasične Hillove ocene, s tim što se parametri drugog reda ocenjuju na višem nivou nego što se koristi za ocenu parametara prvog reda. Mogućnost redukavanja biasa bez povećanja asimptotske disperzije, jeste poboljšanje ove ocene. Takođe ove MVRB ocene omogućavaju dobijanje nove klase ocena visokih kvantila, sa redukovanim biasom. Korišćenjem Monte Karlo simulacija u radu Caeiro i Gomes, 2008 upoređivane su ove nove klase među sobom i sa već predloženim ocenama.

Ako se pogleda (2.13), može se videti da je dominantna komponenta biasa Hillove ocene upravo $A(n/k)/(1-\rho) = \gamma\beta(n/k)^{\rho}/(1-\rho)$, za model (2.12). Ta komponenta se može oceniti i ukloniti iz Hillove ocene, tako da se dobiju asimptotski ekvivalentne ocene, pogledati Caeiro et al., 2005:

$$\begin{aligned}\overline{H}_{\widehat{\beta},\widehat{\rho}}(k) : &= H(k)\Big(1 - \frac{\widehat{\beta}}{1-\widehat{\rho}}\Big(\frac{n}{k}\Big)^{\widehat{\rho}}\Big),\\ \underline{H}_{\widehat{\beta},\widehat{\rho}}(k) : &= H(k)\exp\Big(-\frac{\widehat{\beta}}{1-\widehat{\rho}}\Big(\frac{n}{k}\Big)^{\widehat{\rho}}\Big),\end{aligned}$$

gde su $\widehat{\rho}$ i $\widehat{\beta}$ konzistentne ocene parametara drugog reda ρ i β, ako treba zadržati asimptotsku disperziju γ^2. To nameće korišćenje većeg broja statistika poretka od broja k i ocenu $\widehat{\rho}$ za ρ, takvu da je $\widehat{\rho} - \rho = o_p(1/\ln n)$. Takođe, za ocenu visokih kvantila, kako je

$$x_p/X_{(n-k,n)} \sim_p c_n^{\gamma}(1 + (c_n^{\rho} - 1)A(n/k)/\rho),$$

predložene su nove ocene:

$$\begin{aligned}\overline{Q}_{\widehat{\gamma}}^{(p)}(k;\widehat{\beta},\widehat{\rho}) : &= X_{(n-k,n)} c_n^{\widehat{\gamma}}\Big(1 + \widehat{\gamma}\widehat{\beta}\Big(\frac{n}{k}\Big)^{\widehat{\rho}}\frac{c_n^{\widehat{\rho}} - 1}{\widehat{\rho}}\Big),\\ \underline{Q}_{\widehat{\gamma}}^{(p)}(k;\widehat{\beta},\widehat{\rho}) : &= X_{(n-k,n)} c_n^{\widehat{\gamma}}\exp\Big(\widehat{\gamma}\widehat{\beta}\Big(\frac{n}{k}\Big)^{\widehat{\rho}}\frac{c_n^{\widehat{\rho}} - 1}{\widehat{\rho}}\Big).\end{aligned}$$

Prva ocena je asimptotski ekvivalentna, do drugog reda, ranije predloženim ocenama u Matthys et al., 2004 i Gomes i Pestana, 2007b. Naravno, može se zameniti $\widehat{\gamma}$ nekom od MVRB ocena $\overline{H}_{\widehat{\beta},\widehat{\rho}}(k)$ i $\underline{H}_{\widehat{\beta},\widehat{\rho}}(k)$, što se označava sa $\widetilde{H}(k)$. U disertaciji će biti razmatran indeks $\widetilde{H}(k)$, kao i $\overline{Q}^{(p)}_{\widetilde{H}}$, i pretpostavlja se okvir trećeg reda, kako bi se dobile pune informacije o vodećem članu asimtotskog biasa.

2.4.1 Okvir trećeg reda

Ovde će biti korišćena sub-klasa Hallove klase, za koju važi:

$$U(t) = Ct^{\gamma}(1 + D_1 t^{\rho} + D_2 t^{\rho+\rho^*} + o(t^{\rho+\rho^*})), \quad t \to +\infty, \tag{2.27}$$

$C > 0$, $D_1 \neq 0$, $\rho < 0$, $\rho^* < 0$, sa ciljem da se dobije asimtotski bias ocena MVRB. Treba primetiti da se za $h_\theta(x) := (x^\theta - 1)/\theta$, $\theta < 0$, $A(t) = \rho D_1 t^\rho = \gamma\beta t^\rho$, $\rho^{'} = \max(\rho, \rho^*) \geq \rho$ i:

$$B(t) = \beta^{'} t^{\rho^{'}} = \begin{cases} ((1 + \rho^*/\rho) D_2/D_1) t^{\rho^*}, & \rho < \rho^*, \\ (2D_2/D_1 - D_1) t^{\rho}, & \rho = \rho^*, \\ -D_1 t^{\rho}, & \rho > \rho^* \vee D_2 = 0, \end{cases}$$

za svako $x > 0$, može pisati:

$$\ln \frac{U(tx)}{U(t)} - \gamma \ln x = A(t) h_\rho(x) + A(t) B(t) h_{\rho+\rho'}(x)(1 + o(1)), \tag{2.28}$$

što je uslov trećeg reda korišćen u radu Gomes et al., 2004a, ekvivalentan onome u radu Gomes et al., 2002 i Fraga Alves et al., 2003. Ovde se razmatra validnost (2.27), što je ekvivalentno razmatranju (2.28), za $\rho \leq \rho^{'}$ i $A(t) = \alpha t^\rho$, za realno α.

Može se primetiti da klasa (2.27) obuhvata mnoge modele teških repova koji se javljaju u praksi. Tako na primer Frešeovu, sa $U(t) = (\ln(t/(t-1)))^{-\gamma}$, Burovu, sa $U(t) = (t^{-\rho} - 1)^{-\gamma/\rho}, t > 1$, generalisanu Pareto, sa $U(t) = (t^\gamma - 1)/\gamma, t > 1$ i Studentovu t_ν, sa funkcijom raspodele:

$$F(x) = F(x|\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \int_{-\infty}^{x} (1 + z^2/\nu)^{-(\nu+1)/2} dz,\ x \in \mathbb{R},\ \nu > 0.$$

Za sve ove klasične modele važi da je $\rho = \rho^{'} = \rho^*$. Može se naći primer gde je $\rho^{'} \neq \rho$. U radu Gomes i Oliveira, 2003 su uočili da translacija podataka može promeniti asimtotsko ponašanje repa i vrednosti parametara drugog reda, tj. ako su X početni podaci i $Y = X + a$, tada je $U_Y(t) = U_X(t) + a$ i važi:

$$U_Y(t) = Ct^\gamma(1 + D_1 t^\gamma + at^{-\gamma}/C + D_2 t^{\rho+\rho^*} + o(t^{\rho+\rho^*})),\ t \to +\infty.$$

2.4.2 Ocene parametara drugog reda

Da bi se dale ocene redukovanog biasa indeksa ekstremne vrednosti i kvantila, potrebno je prvo dati ocene parametara drugog reda ρ i β.

Ocena parametra ρ drugog reda

U radu Fraga Alves et al., 2003 razmatrane su ocene parametra drugog reda ρ, ali parametrizovanih parametrom τ, videti Caeiro i Gomes, 2006. Ako se označi sa $M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}^j$, j-ti moment log-ekscesa, $j = 1, 2, 3$, tada ocene za ρ zavise od statistika:

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^\tau - \left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}}, & \tau \neq 0, \\ \dfrac{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)}, & \tau = 0, \end{cases}$$

što konvergira ka $3(1-\rho)/(3-\rho)$, za svako realno τ, u slučaju kada važi uslov drugog reda (2.11), k je intermidijatno i $\sqrt{k}A(n/k) \to \infty$, za $n \to \infty$. Tada je ocena za ρ:

$$\widehat{\rho}_\tau(k) = \widehat{\rho}(k;\tau) := -\min\left(0, 3(T_n^{(\tau)}(k) - 1)/(T_n^{(\tau)}(k) - 3)\right). \qquad (2.29)$$

Teorema 2.4.1. (Fraga Alves et al., 2003) *Ako važi uslov drugog reda (2.11), za $\rho < 0$, intermidijatno k i $\sqrt{k}A(n/k) \to \infty$, tada $\widehat{\rho}(k;\tau)$ u (2.29) konvergira u verovatnoći ka ρ, za $n \to \infty$. Pod okvirima trećeg reda u (2.28), važi:*

$$\widehat{\rho}(k;\tau) =_d \rho + \left(\frac{\gamma\sigma_\rho W_k^\rho}{\sqrt{k}A(n/k)} + v_1 A(n/k) + v_2 B(n/k)\right)(1 + o_p(1)),$$

gde je W_k^ρ slučajna veličina sa asimtotski normalnom raspodelom,

$$\sigma_\rho = \frac{(1-\rho)^3}{\rho}\sqrt{(2\rho^2-2\rho+1)},$$

$$v_1 \equiv v_1(\gamma,\rho,\tau) = \frac{\rho\Big[\tau(1-2\rho)^2(3-\rho)(3-2\rho)+6\rho(4(2-\rho)(1-\rho)^2-1)\Big]}{12\gamma(1-\rho)^2(1-2\rho)^2},$$

$$v_2 = \frac{\rho^{'}(\rho+\rho^{'})(1-\rho)^3}{\rho(1-\rho-\rho^{'})^3}.$$

Još, ako je $\sqrt{k}A^2(n/k) \to \lambda_A$ *i* $\sqrt{k}A(n/k)B(n/k) \to \lambda_B$ *konačno, tada* $\sqrt{k}A(n/k)(\widehat{\rho}(k;\tau)-\rho) \to_d N(\lambda_A v_1 + \lambda_B v_2, \gamma^2\sigma_p^2)$.

Ocena parametra β drugog reda

Uvodi se oznaka $N_n^{(\alpha)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}U_i$, gde je U_i ranije definisano u (2.10). U radu Gomes i Martins, 2002 razmatrana je ocena za β:

$$\widehat{\beta}_{\widehat{\rho}}(k) = \widehat{\beta}(k;\widehat{\rho}) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}(\frac{i}{k})^{-\widehat{\rho}}\right)N_n^{(1)}(k) - N_n^{(1-\widehat{\rho})}(k)}{\left(\frac{1}{k}\sum_{i=1}^{k}(\frac{i}{k})^{-\widehat{\rho}}\right)N_n^{(1-\widehat{\rho})}(k) - N_n^{(1-2\widehat{\rho})}(k)}. \tag{2.30}$$

Teorema 2.4.2. (Gomes et al., 2004b) *Ako važi uslov drugog reda (2.11), za* $A(t) = \gamma\beta t^\rho$, $\rho < 0$, *k intermidijatno i ako* $\sqrt{k}A(n/k) \to \infty$, *tada za* $\widehat{\rho}_n(k;\tau)$ *i* $\widehat{\beta}_{\widehat{\rho}}(k)$ *date u jednakostima (2.29) i (2.30) redom, i* $\widehat{\rho} = \widehat{\rho}_n(k;\tau)$ *tako da važi* $\widehat{\rho} - \rho = o_p(1/\ln n)$, *za* $n \to \infty$, *tada je* $\widehat{\beta}_{\widehat{\rho}}(k)$ *konzistentna ocena za* β *i važi:*

$$\widehat{\beta}_{\widehat{\rho}}(k) - \beta \sim_p -\beta\ln(n/k)(\widehat{\rho}-\rho) = o_p(1).$$

2.4.3 Asimptotska svojstva ocene indeksa ekstremne vrednosti u okvirima trećeg reda

U ovom odeljku razmatra se asimtotsko ponašanje, u okvirima trećeg reda, ocena MVRB, označenih sa $\widetilde{H}$. Pretpostavka je da su poznate vrednsti parametara drugog reda β i ρ.

Teorema 2.4.3. (Caeiro i Gomes, 2008)

(a) *Pod okvirima drugog reda u (2.12) i za intermidijatno k, važi:*

$$\widetilde{H}_{\beta,\rho}(k) =_d \gamma + \frac{\gamma}{\sqrt{k}} Z_k + o_p(A(n/k)),$$

gde je Z_k *slučajna veličina sa asimtotski standardnom normalnom raspodelom, iz (2.13). Takođe, ako se izabere* k *tako da je* $\sqrt{k}A(n/k) \to \lambda$ *konačno, za* $n \to \infty$, $\sqrt{k}(\widetilde{H}_{\beta,\rho}(k) - \gamma)$ *ima asimptotski normalnu raspodelu sa sredinom 0 i disperzijom* γ^2 *i u slučaju kada je* $\lambda \neq 0$.

(b) *Pod pretpostavkom (2.27), dobija se asimptotska raspodela:*

$$\overline{H}_{\beta,\rho}(k) =_d \gamma + \frac{\gamma}{\sqrt{k}} Z_k^* + \frac{A(n/k)B(n/k)}{1-\rho-\rho'} \Big(1 - \frac{(1-\rho-\rho')A(n/k)}{\gamma(1-\rho)^2 B(n/k)}\Big)(1 + o_p(1)),$$

$$\underline{H}_{\beta,\rho}(k) =_d \gamma + \frac{\gamma}{\sqrt{k}} Z_k^* + \frac{A(n/k)B(n/k)}{1-\rho-\rho'} \Big(1 - \frac{(1-\rho-\rho')A(n/k)}{2\gamma(1-\rho)^2 B(n/k)}\Big)(1 + o_p(1)),$$

gde Z_k^* *ima asimptotski standardnu normalnu raspodelu. Ako je* $\sqrt{k}A(n/k)B(n/k) \to \lambda_B$ *konačno (i tada* $\sqrt{k}A^2(n/k) \to \lambda_A$ *takođe konačno),* $\sqrt{k}(\overline{H}_{\beta,\rho}(k) - \gamma)$ *i* $\sqrt{k}(\underline{H}_{\beta,\gamma}(k) - \gamma)$ *su asimptotski normalne sa istim disperzijama* γ^2 *i asimptotskim biasom* $b_{\overline{H}} = \lambda_B/(1-\rho-\rho') - \lambda_A/(\gamma(1-\rho)^2)$ *i* $b_{\underline{H}} = \lambda_B/(1-\rho-\rho') - \lambda_A/(2\gamma(1-\rho)^2)$ *redom.*

Može se primetiti da $\overline{H}$ i $\underline{H}$ imaju iste asimptotske disperzije i da za bias važi $b_{\underline{H}} = b_{\overline{H}} + \lambda_A/(2\gamma(1-\rho)^2)$, za $\lambda_A \geq 0$. Tako da ako su oba biasa pozitivna, $\overline{H}$ ima asimptotski bolje osobine od $\underline{H}$.

Teorema 2.4.4. (Caeiro i Gomes, 2008)

(a) *Pod početnim uslovima Teoreme 2.4.3, razmatraju se ocene indeksa ekstremne vrednosti* $\widetilde{H}_{\widehat{\beta},\widehat{\rho}}$. *Sa* $\widehat{\beta}$ *i* $\widehat{\rho}$ *se označavaju konzistentne ocene za* β *i* ρ *redom, obe izračunate na nivou* k_1 *višeg reda od* k*, na kome se računa indeks ekstremne vrednosti i važi* $\widehat{\rho}-\rho = o_p(1/\ln n)$*, za* $n \to \infty$*. Tada,* $\sqrt{k}(\widetilde{H}_{\widehat{\beta},\widehat{\rho}}(k)-\gamma)$ *je asimptotski normalno, sa disperzijom* γ^2 *i srednjom vrednošću* 0*, čak i kada je* $\sqrt{k}A(n/k) \to \lambda \neq 0$*, za* $n \to \infty$*.*

(b) *Ako se radi u okviru trećeg reda (2.27), razmatra se* $\widehat{\beta}_{\widehat{\rho}}(k)$ *iz (2.30),* $\widehat{\beta} = \widehat{\beta}_{\widehat{\rho}}(k_1)$ *i izabere se* k *tako da* $\sqrt{k}A(n/k) \to \infty$*, ali* $\sqrt{k}A(n/k)B(n/k) \to \lambda_B$ *konačno, tada* $\sqrt{k}(\overline{H}_{\widehat{\beta},\widehat{\rho}}(k) - \gamma)$ *i* $\sqrt{k}(\underline{H}_{\widehat{\beta},\widehat{\rho}}(k) - \gamma)$ *su asimptotski normalne sa disperzijom* γ^2 *i asimptotskim biasom* $b_{\overline{H}}$ *i* $b_{\underline{H}}$ *redom, dato u Teoremi 2.4.3, s tim što je* $(\widehat{\rho}-\rho)\ln n = o_p(1/\sqrt{k}A(n/k))$*.*

2.4.4 Asimptotska svojstva ocene visokih kvantila u okvirima trećeg reda

U ovom odeljku su date dve teoreme o ponašanju raspodele ocene visokih kvantila, u okviru modela (2.27).

Teorema 2.4.5. (Caeiro i Gomes, 2008) *Pod okvirom trećeg reda (2.27), za intermidijatno* k *i kada god je* $\ln(np) = o(\sqrt{k})$*, može se napisati:*

$$
\begin{aligned}
Q^{(p)}_{H(k)}(k)/x_p =_d 1 \ &+ \ (H(k)-\gamma)\ln c_n + \frac{\gamma}{\sqrt{k}}B_k - h_\rho(c_n)A(n/k) + O_p\Big(\frac{A(n/k)}{\sqrt{k}}\Big) \\
&- \ \Big(h_{\rho+\rho'}(c_n)A(n/k)B(n/k) + \frac{1}{2}h^2_\rho(c_n)A^2(n/k)\Big)(1+o_p(1)),
\end{aligned}
$$

gde je B_k *asimptotski standardna normalna slučajna veličina,* $h_\theta(x) = (x^\theta - 1)/\theta$, $\theta < 0$*. Takođe, ako je* $\sqrt{k}A(n/k) \to \lambda$ *konačno i* $\ln c_n/\sqrt{k} \to 0$*, za* $n \to \infty$*, tada* $\frac{\sqrt{k}}{\ln c_n}(Q^{(p)}_{H(k)}(k)/x_p - 1)$ *ima asimptotski istu raspodelu kao* $\sqrt{k}(H(k)-\gamma)$*, tj. asimptotski je normalna sa disperzijom* γ^2 *i srednjom vrednošću* $\lambda/(1-\rho)$*.*

Teorema 2.4.6. (Caeiro i Gomes, 2008)

(a) *Pod uslovima Teoreme 2.4.5, razmatra se ocena indeksa ekstremne vrednosti* $\widetilde{H} = \widetilde{H}_{\widehat{\beta},\widehat{\rho}}$, *gde je* $(\widehat{\beta},\widehat{\rho})$ *konzistentna ocena za* (β,ρ), *obe izračunate na nivou* k_1, *gde je* $k = o(k_1)$ *i tako da je* $(\widehat{\rho}-\rho)\ln n = o_p(1)$. *Tada, ako* $\sqrt{k}A(n/k) \to \lambda$, $\frac{\sqrt{k}}{\ln c_n}(\overline{Q}^{(p)}_{\widetilde{H}(k)}(k)/x_p - 1)$ *ima asimptotski istu raspodelu kao* $\sqrt{k}(\widetilde{H}(k)-\gamma)$, *tj. obe su asimptotski normalne sa disperzijom* γ^2 *i srednjom vrednošću 0 (čak i kada je* $\lambda \neq 0$*).*

(b) *Ako se* k *izabere tako da je* $\sqrt{k}A(n/k) \to \infty$ *ali* $\sqrt{k}A(n/k)B(n/k) \to \lambda_B$ *konačno,* $\frac{\sqrt{k}}{\ln c_n}(\overline{Q}^{(p)}_{\widetilde{H}(k)}(k)/x_p - 1)$ *i* $\sqrt{k}(\widetilde{H}(k)-\gamma)$ *takođe imaju asimptotski istu raspodelu, tj. asimptotski su normalne sa disperzijom* γ^2 *i asimptotskim biasom datim u Teoremi 2.4.3, stim što je* $(\widehat{\rho}-\rho)\ln n = o_p(1/\sqrt{k}A(n/k))$.

2.5 Poređenje ocena

Kao što je rečeno, za raspodele sa pravilno promenljivim repom predloženo je više ocena indeksa ekstremne vrednosti. Najpoznatija ocena jeste Hillova ocena, funkcija k najvećih statistika poretka. Asimptotski optimalna vrednost za k se može odrediti koristeći parametre drugog reda (u smislu minimalne srednjenjekvadratne greške). Tako se mogu porediti dve ocene, poređenjem srednjekvadratnih grešaka, na dva načina. Prvi je poređenje za fiksirano $k = k(n) \to \infty$, za koje su obe ocene asimptotski normalne, dok je drugi način za poređenje upoređivanje za nizove $k(n)$, koji su različiti za dve ocene i koji su izabrani tako da je za svaku asimptotska srednjekvadratna greška najmanja. Treba reći da ovaj drugi pristup poređenja ocena, iako prirodan, nije mnogo razvijan.

Pretpostavlja se da je $X_1, X_2, ...$ i.i.d. sa funkcijom raspodele F, za koju je $1-F$ pravilno promenljiva funkcija u beskonačnosti, tj.

$$\lim_{t\to+\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}, \tag{2.31}$$

za $x > 0$, gde je $\gamma > 0$ (tj. $1-F \in RV_{-1/\gamma}$). Do sada su razmatrane sledeće ocene za parametar γ:

$$\begin{aligned}
\widehat{\gamma}_n^{(1)} &= \frac{1}{k}\sum_{i=1}^{k}\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)},\\
\widehat{\gamma}_n^{(2)} &= (\ln 2)^{-1}\ln\frac{X_{(n-[k/4],n)} - X_{(n-[k/2],n}}{X_{(n-[k/2],n)} - X_{(n-k,n)}},\\
\widehat{\gamma}_n^{(3)} &= \widehat{\gamma}_n^{(1)} + 1 - \frac{1}{2}(1-(\widehat{\gamma}_n^{(1)})^2/M_n) - 1,
\end{aligned}$$

gde je:

$$M_n = \frac{1}{k}\sum_{i=1}^{k}(\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)})^2.$$

Određivanjem brzine konvergencije u (2.31), takođe se pretpostavlja uslov drugog reda, tj. da postoji funkcija a konstantnog znaka, takva da je:

$$\lim_{t\to+\infty}\frac{\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma}}{a(t)} = x^{-1/\gamma}\frac{x^{\rho}-1}{\rho}, \tag{2.32}$$

za $x > 0$, gde je $\rho \le 0$ parametar drugog reda.

Neka je kao i ranije $U = (1/(1-F))^{\leftarrow}$ i $A(t) = \gamma^2 a(U(t))$. Jednakost (2.32) je ekvivalentna sledećoj jednakosti:

$$\lim_{t\to+\infty}\frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = x^{\gamma}\frac{x^{\rho\gamma}-1}{\rho\gamma}, \tag{2.33}$$

lokalno uniformno za $x > 0$.

U radu de Haan i Peng, 1998 zaključeno je da ne postoji jedna ocena koja dominira nad drugim, već da to zavisi od situacije. Tek kada se napravi napredak u određivanju optimalne vrednosti za $k(n)$ i parametar drugog reda ρ učini adaptivnim, može se razmišljati o izboru najbolje ocene u svim situacijama.

Teorema 2.5.1. (de Haan i Peng, 1998) *Pretpostavlja se da važi jednakost (2.33). Neka je $k = k(n)$ niz celih brojeva za koji važi $k(n) \to \infty$ i $k(n)/n \to 0$ i:*

$$\lim_{n\to\infty}\sqrt{k}A\left(\frac{n}{k}\right) = \lambda \in (-\infty, +\infty).$$

Definiše se $C_i = C_i(\gamma,\rho)$ na sledeći način:

$$\begin{aligned} C_1(\gamma,\rho) &= \frac{1}{1-\gamma\rho}, \\ C_2(\gamma,\rho) &= \frac{1}{(2^\gamma-1)\ln 2}\frac{2^{\rho\gamma}-1}{\rho\gamma}(2^{\gamma+\rho\gamma}-1), \\ C_3(\gamma,\rho) &= \frac{1}{1-\rho\gamma}+\frac{\rho}{(1-\rho\gamma)^2}. \end{aligned}$$

Tada:

$$\sqrt{k}(\widehat{\gamma}_n^{(i)}-\gamma) \to_d N(b_i,\sigma_i^2),$$

za $i = 1,2,3$, gde su:

$$b_1 = \lambda C_1,\ b_2 = \lambda C_2,\ b_3 = \lambda C_3,$$

$$\sigma_1^2 = \gamma^2,\ \sigma_2^2 = \frac{\gamma^2(1+2^{2\gamma+1})}{(2^\gamma-1)^2\ln^2 2},\ \sigma_3^2 = 1+\gamma^2.$$

Može se izračunati da je asimptotska srednjekvadratna greška za $\widehat{\gamma}_n^{(i)}$, za svaku ocenu, $k^{-1}(\sigma_i^2+b_i^2)$. Takođe, za $\rho = 0$, važi $b_1 = b_2 = b_3$ i σ_1^2 je najmanja vrednost od svih.

Posledica 2.5.1. *Neka je $\rho < 0$ i važi jednakost (2.33). Ako je RMMSE(i,j) granična vrednost za količnik minimalnih srednjekvadratnih grešaka i-te i j-te ocene, tada važi:*

$$RMMSE(1,2) = \frac{(2^\gamma-1)^2\ln^2 2}{1+2^{2\gamma+1}}\cdot\left[\frac{(1-\rho\gamma)^2(2^{\rho\gamma}-1)^2(2^{\gamma+\rho\gamma}-1)^2}{(\rho\gamma)^2(1+2^{2\gamma+1})}\right]^{1/(2\rho\gamma-1)},$$

za $1+\rho \neq 0$ i

$$RMMSE(1,3) = \frac{\gamma^2}{1+\gamma^2}\left[\frac{(\gamma+\rho\gamma-\rho\gamma^2)^2}{(1+\gamma^2)(1-\rho\gamma)^2}\right]^{1/(2\rho\gamma-1)},$$

za $1+\rho-\gamma\rho \neq 0$.

Na kraju ovog poglavlja treba istaći da razmatrane ocene indeksa ekstremne vrednosti i visokih kvantila, kao i njihove osobine, mogu poslužiti za analizu i dobijanje intervala poverenja pomenutih parametara. Tako se na primer, Teorema 2.3.1 i Teorema 2.3.2, koje govore o asimptotskoj konvergenciji ocena, mogu iskoristiti u tu svrhu, što je jedan od mogućih pravaca daljeg istraživanja. U narednom poglavlju dat je jedan pristup, koji je već poznat u literaturi, za određivanje intervala poverenja razmatranih parametara.

Glava 3

Intervali poverenja za indeks ekstremne vrednosti i visoke kvantile

U ovom poglavlju biće dati intervali poverenja za indeks ekstremne vrednosti i visoke kvantile, kao i optimalna brzina konvergencije ocena. Pristup za dobijanje intervala poverenja, koji je analiziran u literaturi, koristi aproksimaciju normalnom raspodelom i ima neoptimalnu brzinu. U radu Ferreira i Vries, 2003, predložen je pristup sa optimalnom brzinom, ali u tom slučaju mora biti ocenjen bias nepoznatog znaka.

3.1 Optimalni intervali poverenja za indeks ekstremne vrednosti

Neka je kao i ranije $X_1, X_2, ..., X_n$ i.i.d. niz sa raspodelom F, koja pripada oblasti privlačenja raspodele ekstremnih vrednosti, sa pozitivnim indeksom promenljivosti. Ako se taj uslov preformuliše pomoću termina pravilne promenljivosti, to znači da je rep raspodele F pravilno promenljiva funkcija sa indeksom promenljivosti $-1/\gamma$, tj. za neko $\gamma > 0$:

$$\lim_{t \to +\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad x > 0. \tag{3.1}$$

Definiše se $U = (1/(1-F))^{\leftarrow}$, kao i ranije, odgovarajuća inverz funkcija. Dalje se pretpostavlja da postoji funkcija A, konstantnog znaka u beskonačnosti, $\lim_{t\to+\infty} A(t) = 0$ i da važi:

$$\lim_{t\to+\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \quad x > 0,\ \rho < 0. \tag{3.2}$$

Već je ranije analizirana Hillova ocena:

$$\widehat{\gamma}(k) = \frac{1}{k} \sum_{i=1}^{k} \ln X_{(n-i+1,n)} - \ln X_{(n-k,n)}. \tag{3.3}$$

Neka je $k = k(n) = k_n$ intermidijatni niz, za $n \to \infty$ i:

$$H_{n,k_n} = \sqrt{k_n}\Big(\frac{\widehat{\gamma}(k_n)}{\gamma} - 1\Big). \tag{3.4}$$

Ranije je rečeno da pod uslovima (3.2) i $A(n/k_n)\sqrt{k_n} \to \lambda \in (-\infty, +\infty)$, H_{n,k_n} konvergira u raspodeli ka normalnoj slučajnoj veličini sa sredinom $\lambda/(\gamma(1-\rho))$ i disperzijom 1. Najbolja brzina konvergencije se postiže kada $A(n/k_n)\sqrt{k_n} \to \lambda \neq 0$ i u tom slučaju granična raspodela ima nenula sredinu.

Cheng i Peng, 2001 i Caers et al., 1998, u cilju konstruisanja intervala poverenja za γ, zbog jednostavnosti, koriste (3.4) sa $\lambda = 0$, kada se upotrebljava Hillova ocena (3.3). Treba istaći da je za konstrukciju intervala poverenja sa optimalnom brzinom potrebno oceniti i druge parametre. Znači, da bi se konstruisao interval poverenja, treba rešiti nekoliko problema: na adekvatan način se treba odrediti optimalni niz k_n, treba oceniti dva dodatna parametra konzistentnim ocenama, parametar drugog reda ρ i znak asimtotskog biasa. Za izbor niza k_n pogledati Danielsson et al., 2001. U njemu je predložena *bootstrap* metodologija za određivanje optimalnog niza k_n. Ta metodologija je sasvim drugačija od one predložene u radu Ferreira i Vries, 2003, koja je razmatrana u ovoj disertaciji. Za ocenu parametra ρ pogledati Danielsson et al., 2001 i Fraga Alves et al., 2003. Za znak asimptotskog biasa data je ocena i pokazana je konzistentnost u radu Ferreira i de Vries, 2003. Relevantni radovi o intervalu poverenja ocene su Caers et al., 1998 i Cheng i Peng, 2001. Cheng i Peng, 2001 takođe ističu važnost znaka asimptotskog biasa, ali ne razmatraju eksplicitno ocenu. Ovde će biti razmatran slučaj kada je $\rho < 0$, jer je za taj slučaj, otimalni rezultat za izbor k_n dobro poznat.

3.1.1 Dvostrani interval poverenja za indeks ekstremne vrednosti

Ovde se dalje razmatra interval poverenja za indeks ekstremne vrednosti. Neka $A(n/k_n)\sqrt{k_n} \to \lambda \in R$. Kao i do sada, sa Φ se označava standardna normalna funkcija raspodele i sa $z_\alpha = \Phi^{\leftarrow}(1-\alpha)$ odgovarajući kvantil. Bazirajući se na (3.4) i njegovoj graničnoj raspodeli, rešavajući $-z_\alpha < H_{n,k_n} - \lambda/(\gamma(1-\rho)) < z_\alpha$ po γ, dobija se prvi pristup za konstruisanje intervala poverenja za γ, sa nivoom zanačajnosti α, pogledati Ferreira i de Vries, 2003:

$$\frac{\widehat{\gamma}(k_n)\sqrt{k_n} - \frac{\lambda}{1-\rho}}{z_{\alpha/2} + \sqrt{k_n}} < \gamma < \frac{\widehat{\gamma}(k_n)\sqrt{k_n} - \frac{\lambda}{1-\rho}}{-z_{\alpha/2} + \sqrt{k_n}}, \tag{3.5}$$

uz uslov $\sqrt{k_n} - z_{\alpha/2} > 0$.

Hall i Welsh, 1985 i Dekkers i de Haan, 1993, su odredili k_n^0, optimalni niz u smislu minimalne srednjekvadratne greške. Ako se pretpostavi da se pravilno promenljiva funkcija A ponaša asimtotski kao stepena funkcija:

$$A(t) \sim ct^\rho,\ c \neq 0,\ \rho < 0, \tag{3.6}$$

za $t \to +\infty$, niz k_n^0 se računa pod ovde uvedenim uslovom. Dalje se pretpostavlja jači uslov, Hallov model, tj.:

$$1 - F(x) = Cx^{-1/\gamma}(1 + Dx^{\rho/\gamma} + o(x^{\rho/\gamma})),\ C > 0, D \neq 0, x \to +\infty.$$

Treba primetiti da iz (3.2) i (3.6) sledi da je $D = c\gamma^{-1}\rho^{-1}C^\rho$. Tako se dobija (pogledati Hall i Welsh, 1985):

$$k_n^0 \sim \Big(\frac{\gamma^2(1-\rho)^2}{-2\rho c^2}\Big)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}. \tag{3.7}$$

Vrednost λ za niz k_n^0 je asimptotski $sgn(c)\gamma(1-\rho)/\sqrt{-2\rho}$. U ovom slučaju, umesto intervala (3.5) dobija se interval:

$$\frac{\widehat{\gamma}(k_n^0)\sqrt{k_n^0}}{z_{\alpha/2} + sgn(c)/\sqrt{-2\rho} + \sqrt{k_n^0}} < \gamma < \frac{\widehat{\gamma}(k_n^0)\sqrt{k_n^0}}{-z_{\alpha/2} + sgn(c)/\sqrt{-2\rho} + \sqrt{k_n^0}},$$

pogledati Ferreira i de Vries, 2003. Sada je potrebno aproksimirati k_n^0 i oceniti ρ i $sgn(c)$, kako bi se dobio interval poverenja iz prethodne nejednakosti. U radu Hall i Welsh, 1985 je određen adaptivni izbor za $\widehat{k}_n^0$, za koji je:

$$\frac{\widehat{k}_n^0}{k_n^0} \to_p 1, \tag{3.8}$$

gde $\to_p$ označava konvergenciju u verovatnoći. Za ρ je potrebna konzistentna ocena i za $sgn(c)$ ocena $\widehat{sgn}$ zadovoljava:

$$P(\widehat{sgn} = sgn(c)) \to 1,\ n \to \infty. \tag{3.9}$$

Tako važi sledeća teorema.

Teorema 3.1.1. (Ferreira i de Vries, 2003) *Pretpostavlja se da važe (3.2) i (3.6). Neka $\widehat{k}_n^0$ zadovoljava (3.8), $\widehat{sgn}$ zadovoljava (3.9) i neka je $\widehat{\rho}$ konzistentna ocena za ρ. Tada, za $n \to \infty$:*

$$\sqrt{\widehat{k}_n^0}\Big(\frac{\widehat{\gamma}(\widehat{k}_n^0)}{\gamma} - 1\Big) - \frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}}$$

konvergira u raspodeli ka standardnoj normalnoj slučajnoj veličini. Važi:

$$P\Bigg(\frac{\widehat{\gamma}(\widehat{k}_n^0)\sqrt{\widehat{k}_n^0}}{z_{\alpha/2} + \frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}} + \sqrt{\widehat{k}_n^0}} < \gamma < \frac{\widehat{\gamma}(\widehat{k}_n^0)\sqrt{\widehat{k}_n^0}}{-z_{\alpha/2} + \frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}} + \sqrt{\widehat{k}_n^0}}\Bigg) \to 1 - \alpha, \tag{3.10}$$

za $n \to \infty$, što daje asimptotski interval poverenja za γ, sa koeficijentom pouzdanosti $1 - \alpha$.

Treba primetiti da za slučaj kada je γ blizu nule, može se posmatrati alternativni jednostrani interval poverenja.

Pokrivenost dvostranog intervala poverenja

Ako se interval baziran na (3.5) označi sa $(\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})$, gde je k_n takvo da je $\lambda = 0$ i ρ zamenjeno sa konzistentnom ocenom. Dalje se može pokazati da je interval (3.10) moćniji od intervala $(\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})$.

Važi:

$$\lim_{n\to\infty} P(\gamma \in (\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})) = 1 - \alpha,$$

za svako $\gamma > 0$. Tada, verovatnoća pokrivanja pogrešne vrednosti γ' treba da bude što je moguće manja za svako $\gamma' > 0$, tj.

$$P(\gamma' \in [\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]) \tag{3.11}$$

što manje. Ova verovatnoća konvergira ka 0 u slučaju kada je niz k_n takav da je $A(n/k_n)\sqrt{k_n} \to \lambda$ i za $\gamma' \neq \gamma$, kako gornja i donja granična vrednost intervala poverenja konvergiraju ka γ, u verovatnoći. Dalje se porede verovatnoće pogrešnog pokrivanja, kako je $\gamma_n'/\gamma \to 1$ za $n \to \infty$.

Verovatnoća pogrešnog pokrivanja za interval poverenja (3.10) je:

$$P\Big(-z_{\alpha/2}\frac{\gamma_n'}{\gamma} \leq \sqrt{\widehat{k}_n^0}\Big(\frac{\widehat{\gamma}(\widehat{k}_n^0)}{\gamma}-1\Big)-\frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}}+\Big(1-\frac{\gamma_n'}{\gamma}\Big)\Big(\sqrt{\widehat{k}_n^0}-\frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}}\Big) \leq z_{\alpha/2}\frac{\gamma_n'}{\gamma}\Big).$$

Za niz γ_n' za koji $\sqrt{k_n^0}(1-\gamma_n'/\gamma) \to \nu \neq 0, \pm\infty$, ova verovatnoća konvergira ka $\Phi(z_{\alpha/2}-\nu)-\Phi(-z_{\alpha/2}-\nu)$. Ako se izabere k_n tako da je $A(n/k_n)\sqrt{k_n} \to 0$, tada je verovatnoća (3.11) ekvivalentna verovatnoći:

$$P\Big(-z_{\alpha/2}\frac{\gamma_n^*}{\gamma}+\sqrt{k_n}\Big(\frac{\gamma_n^*}{\gamma}-1\Big) \leq \sqrt{k_n}\Big(\frac{\widehat{\gamma}(k_n)}{\gamma}-1\Big) \leq z_{\alpha/2}\frac{\gamma_n^*}{\gamma}+\sqrt{k_n}\Big(\frac{\gamma_n^*}{\gamma}-1\Big)\Big),$$

što konvergira ka $\Phi(z_{\alpha/2}-\nu^*)-\Phi(-z_{\alpha/2}-\nu^*)$, za niz γ_n^* za koji je $\sqrt{k_n}(1-\gamma_n^*/\gamma) \to \nu^* \neq 0, \pm\infty$. Uzima se zajednički niz γ_n^*, da bi se poredile dve verovatnoće. Tada verovatnoća pokrivanja pogrešne vrednosti konvergira ka 0 u prvom slučaju, dok je u drugom slučaju jednaka $\Phi(z_{\alpha/2}-\nu^*)-\Phi(-z_{\alpha/2}-\nu^*) > 0$.

Ocena znaka biasa za Hillovu ocenu

U ovom delu koristi se Hillova parametrizacija. Sledeći rezultat može se pogledati u Drees et al., 2000.

Lema 3.1.1. *Neka je k_n intermidijatni niz. Pod uslovom (3.2), postoji prostor verovatnoća za $X_1, X_2, ...$ i niz slučajnih veličina sa Braunovim kretanjem W_n (proces za koj je $W_n(0) = 0$, sa trajektorijom koja je skoro svuda neprekidna, sa nezavisnim priraštajima, koji imaju odgovarajuću normalnu raspodelu), tako da je:*

$$\sup_{t_n \leq t \leq 1} t^{1/2}\Big|\widehat{\gamma}([k_n t])-\Big(\gamma+\frac{\gamma}{\sqrt{k_n}}\frac{W_n(t)}{t}+A\Big(\frac{n}{k_n}\Big)\frac{t^{-\rho}}{1-\rho}\Big)\Big| = o_p\Big(k_n^{-1/2}+A\Big(\frac{n}{k_n}\Big)\Big),$$

za sve $t_n \to 0$, koji zadovoljava $k_n t_n \to +\infty$.

Član $t^{-\rho}/(1-\rho)A(n/k_n)\sqrt{k_n}$ je upravo bias Hillove ocene. Treba primetiti da ako $A(n/k_n)\sqrt{k_n} \to \lambda \neq 0$, njegov asimptotki znak je određen znakom funkcije $A(n/k_n)$, što je jednako $sgn(c)$ uz uslov (3.6). Tako na primer, za $t=1$, znak očekivane vrednosti granične promenljive $\sqrt{k_n}(\widehat{\gamma}-\gamma)$ jednak je $sgn(\lambda/(1-\rho)) = sgn(c)$.

Neka su a_n, b_n i c_n intermidijatni nizovi takvi da je ispunjeno

$$a_n < b_n \leq c_n,\ a_n/b_n \to \nu \in [0,1), \tag{3.12}$$

za svako n. U radu Ferreira i de Vries, 2003 predložena je sledeća ocena znaka biasa:

$$\widehat{sgn} = sgn\Big(\widehat{\gamma}(c_n) - \frac{1}{b_n - a_n + 1}\sum_{i=a_n}^{b_n}\widehat{\gamma}(i)\Big).$$

Teorema 3.1.2. (Ferreira i de Vries, 2003) *Pretpostavlja se da važe (3.2) i (3.12) i još da b_n zadovoljava $A(n/b_n)\sqrt{b_n} \to \infty$. Tada:*

$$P\Big(\widehat{sgn} = sgn(c)\Big) \to 1,\ n \to \infty.$$

Može se primetiti još da je isključen slučaj $\rho = 0$ u uslovu (3.2). Rezultat ovog dela važi za $\rho = 0$ uz uslov $b_n < c_n$ u (3.12).

3.1.2 Jednostrani interval poverenja za indeks ekstremne vrednosti

U ovom odeljku biće analiziran jednostrani interval poverenja baziran na asimptotski normalnoj aproksimaciji Hillove ocene. Biće dat rezultat iz Cheng i Peng, 2001, o verovatnoći pokrivanja jednostranog intervala poverenja. Otkriveno je i da red optimalne verovatnoće pokrivanja za jednostrani interval poverenja zavisi od znaka drugog reda pravilne promenljivosti.

Dalje se takođe pretpostavlja da je $X_1, ..., X_n$ i.i.d. niz sa funkcijom raspodele F, za koju je $1 - F \in RV_{-1/\gamma}$. Postavlja se jači uslov za funkciju F od prethodnog. Pretpostavlja se da za $x \to +\infty$:

$$1 - F(x) = Cx^{-1/\gamma}\Big(1 + Dx^{-\beta} + o(x^{-\beta})\Big),\ C > 0, \beta > 0, D \neq 0, \tag{3.13}$$

gde je $\beta = -\rho/\gamma$ iz Hallovog modela. Može se primetiti da je prethodni uslov poseban slučaj uopštene pravilne promenljivosti drugog reda (pogledati de Haan i Stadtm*ü*ller, 1996). Već je poznat sledeći rezultat, da za intermidijatni niz k, slučajna veličina

$$\sqrt{k}(\widehat{\gamma}_n - \gamma) \rightarrow_d N(0, \gamma^2), \tag{3.14}$$

akko $k = o(n^{2\beta\gamma/(1+2\beta\gamma)})$, gde je $\widehat{\gamma}_n$ odgovarajuća Hillova ocena.

Tada je nominalni α jednostrani interval poverenja za γ, baziran na normalnoj aproksimaciji (3.14), upravo:

$$I(\alpha) = \Big(0, \widehat{\gamma}_n + \frac{z_\alpha \widehat{\gamma}_n}{\sqrt{k}}\Big), \tag{3.15}$$

gde z_α zadovoljava $\Phi(z_\alpha) = \alpha$. Dalje je data verovatnoća pokrivanja datog jednostranog intervala poverenja i dat je teorijski optimalni izbor za veličinu uzorka k, u smislu minimalne apsolutne greške pokrivanja.

Verovatnoća pokrivanja jednostranog intervala poverenja

Edgeworthov razvoj (pogledati Hall, 1992a) je jedan od osnovnih načina za analizu verovatnoće pokrivanja intervala poverenja. U radu Cheng i Pan, 1998 dat je Edgeworthov razvoj za Hillovu ocenu. Sledi taj rezultat.

Teorema 3.1.3. (Cheng i Pan, 1998) *Pretpostavlja se da važi (3.13) i $k \to \infty$, $k/n \to 0$. Tada:*

$$\begin{aligned} P\Big(\frac{\sqrt{k}(\widehat{\gamma}_n - \gamma)}{\gamma} \le x\Big) &= \Phi(x) + \phi(x)\Big(\frac{1-x^2}{3\sqrt{k}} + \frac{D\beta\gamma}{(1+\beta\gamma)C^{\beta\gamma}}\sqrt{k}\Big(\frac{n}{k}\Big)^{-\beta\gamma}\Big) \\ &+ o\Big(\frac{1}{\sqrt{k}} + \sqrt{k}\Big(\frac{n}{k}\Big)^{-\beta\gamma}\Big), \end{aligned} \tag{3.16}$$

uniformno na R, gde su $\Phi(x)$ i $\phi(x)$ redom, funkcija raspodele i gustina standardne normalne raspodele.

Treba napomenuti da je u radu Cheng i Pan, 1998 uzet u obzir član $\phi(x)\frac{1-x^2}{3\sqrt{k}}$ a ignorisan član $\phi(x)\frac{D\beta\gamma}{(1+\beta\gamma)C^{\beta\gamma}}\sqrt{k}\left(\frac{n}{k}\right)^{-\beta\gamma}$ iz Edgeworthovog razvoja, prethodne teoreme, da bi se konstruisao što ispravniji interval poverenja. Dok se u radu Cheng i Peng, 2001 koriste oba člana da bi se izabrala optimalna veličina uzorka kako bi se minimizovala apsolutna greška pokrivanja, što će ovde biti iskorišćeno.

Koristeći Edgeworhov razvoj (3.16) biće predstavljen rezultat koji daje verovatnoću pokrivanja za interval $I(\alpha)$, (3.15).

Teorema 3.1.4. (Cheng i Peng, 2001) *Pretpostavlja se da važi (3.13) i $k \to \infty$, $k/n \to 0$. Tada:*

$$\begin{aligned} P(\gamma \in I(\alpha)) = \alpha \quad &- \quad \phi(z_\alpha)\Big(\frac{1+2z_\alpha^2}{3\sqrt{k}} + \frac{D\beta\gamma}{(1+\beta\gamma)C^{\beta\gamma}}\sqrt{k}\Big(\frac{n}{k}\Big)^{-\beta\gamma}\Big) \\ &+ \quad o\Big(\frac{1}{\sqrt{k}} + \sqrt{k}\Big(\frac{n}{k}\Big)^{-\beta\gamma}\Big). \end{aligned}$$

Optimalna vrednost k koja minimizira apsolutnu vrednost greške pokrivanja u intervalu iz prethodne teoreme jeste:

$$k^* = \begin{cases} \left(\frac{(1+2z_\alpha^2)(1+\beta\gamma)C^{\beta\gamma}}{3D\beta\gamma(1+2\beta\gamma)}\right)^{1/(1+\beta\gamma)} n^{\beta\gamma/(1+\beta\gamma)}, & D > 0, \\ \left(\frac{(1+2z_\alpha^2)(1+\beta\gamma)C^{\beta\gamma}}{-3D\beta\gamma}\right)^{1/(1+\beta\gamma)} n^{\beta\gamma/(1+\beta\gamma)}, & D < 0, \end{cases} \tag{3.17}$$

koja zadovoljava uslov $k = o(n^{2\beta\gamma/(1+2\beta\gamma)})$. I još, optimalna verovatnoća pokrivanja intervala $I(\alpha)$ je:

$$\begin{aligned} &P(\gamma \in I(\alpha)) = \\ &= \begin{cases} \alpha - 2\left(\frac{(1+\beta\gamma)(1+2z_\alpha^2)}{3(1+2\beta\gamma)}\right)^{(1+2\beta\gamma)/(2(1+\beta\gamma))}\left(\frac{D\beta\gamma}{C^{\beta\gamma}}\right)^{1/(2(1+\beta\gamma))} & \\ \qquad \cdot\phi(z_\alpha)n^{-\beta\gamma/(2(1+\beta\gamma))}(1+o(1)), & D > 0, \\ \qquad \alpha + o(n^{-\beta\gamma/(2(1+\beta\gamma))}), & D < 0. \end{cases} \end{aligned}$$

Iz poslednja dva izraza može se zaključiti da optimalna verovatnoća pokrivanja intervala $I(\alpha)$ zavisi od znaka pravilne promenljivosti drugog reda.

Izbor optimalne vrednosti k

U terminu verovatnoće pokrivanja, optimalna veličina uzorka zavisi od nekih nepoznatih veličina. Ovde će biti data ocena za optimalnu veličinu za jednostrani interval poverenja. Neka:

$$M_n^{(j)}(k) = \frac{1}{k}\sum_{i=1}^{k}(\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)})^j,\ j=1,2,$$

$$\widehat{\rho}_n = -(\ln 2)^{-1}\ln\left|\frac{M_n^{(2)}(n/(2\sqrt{\ln n})) - 2[M_n^{(1)}(n/(2\sqrt{\ln n}))]^2}{M_n^{(2)}(n/\sqrt{\ln n}) - 2[M_n^{(1)}(n/\sqrt{\ln n})]^2}\right|,$$

$$\widehat{\delta}_n = \Big(M_n^{(2)}\frac{n}{\sqrt{\ln n}} - 2\Big[M_n^{(1)}\frac{n}{\sqrt{\ln n}}\Big]^2\Big)(1+\widehat{\rho}_n)(\sqrt{\ln n})^{\widehat{\rho}_n}$$
$$\Big(2\Big[M_n^{(1)}\Big(\frac{n}{\sqrt{\ln n}}\Big)\Big]^2\widehat{\rho}_n\Big)^{-1}.$$

Na osnovu rezultata datih u Peng, 1998 sledi:

$$\begin{aligned}\widehat{\rho}_n &\to_p \beta\gamma,\\ \widehat{\delta}_n &\to_p D\beta\gamma(1+\beta\gamma)^{-1}C^{-\beta\gamma},\end{aligned}$$

kada $n \to \infty$. Ocena za optimalnu uzoračku veličinu k^*, data u (3.17), je:

$$\widehat{k} = \begin{cases}\left(\frac{1+2z_\alpha^2}{3\widehat{\delta}_n(1+2\widehat{\rho}_n)}\right)^{1/(1+\widehat{\rho}_n)} n^{\widehat{\rho}_n/(1+\widehat{\rho}_n)}, & \widehat{\delta}_n > 0,\\ \left(\frac{1+2z_\alpha^2}{-3\widehat{\delta}_n}\right)^{1/(1+\widehat{\rho}_n)} n^{\widehat{\rho}_n/(1+\widehat{\rho}_n)}, & \widehat{\delta}_n < 0,\end{cases}$$

gde z_α zadovoljava $\Phi(z_\alpha) = \alpha$.

3.2 Optimalni intervali poverenja za visoke kvantile

Ocenjivanje visokih kvantila ima važnu ulogu u primenama, kao na primer u kontekstu risk menadžmenta. To zahteva ekstrapolaciju nepoznate funkcije raspodele. U ovom delu disertacije biće predstavljene dve metode za konstrukciju intervala poverenja visokih kvantila: metod normalne aproksimacije i metod količnika verodostojnosti, kao i konstrukcija jednostranog intervala poverenja za visoke kvantile raspodela sa debelim repom.

Ovde se pretpostavlja da je $X_1, ..., X_n$ i.i.d. niz sa funkcijom raspodele F, koja zadovoljava $1 - F(x) \in RV_{-\alpha}$, tj.:

$$1 - F(x) = e(x)x^{-\alpha},\ x > 0, \tag{3.18}$$

gde je $\alpha > 0$ i $e(x)$ je sporo promenljiva funkcija, tj. $\lim_{t\to+\infty} e(tx)/e(t) = 1$, $x > 0$. Dalje se pretpostavlja da $p_n \in (0,1)$ i $p_n \to 0$, za $n \to \infty$. Odgovarajući $(1-p_n)100\%$ kvantil raspodele F je definisan $x_p = (1-F)^{\leftarrow}(p_n)$. Cilj je da se odredi interval poverenja za x_p.

Dalje se pretpostavlja da F zadovoljava specijalan slučaj (3.18), za $e(x) = c$, gde je c konstanta,

$$1 - F(x) = cx^{-\alpha},\ x > T. \tag{3.19}$$

Neka je δ_i odgovarajući indikator, tj. $\delta_i = I(X_i > T)$. Tada je funkcija maksimalne verodostojnosti veličine $\{(\delta_i, \max(X_i, T))\}_{i=1}^n$:

$$L(\alpha, c) = \prod_{i=1}^{n} (c\alpha X_i^{-\alpha-1})^{\delta_i} (1 - cT^{-\alpha})^{1-\delta_i}.$$

Dalje se uzima da je $T = X_{(n-k,n)}$, gde je $k = k_n$ intermidijatni niz, tj. $k \to \infty$, $k/n \to 0$. Gornja funkcija maksimalne verodostojnosti sada postaje:

$$L(\alpha, c) = \prod_{i=1}^{n} (c\alpha X_i^{-\alpha-1})^{\delta_i} (1 - cX_{(n-k,n)}^{-\alpha})^{1-\delta_i}.$$

U naredna dva dela ovog poglavlja biće predstavljene dve metode za konstrukciju intervala poverenja za x_p i u trećem delu biće predstavljen metod za konstrukciju jednostranog intervala poverenja.

3.2.1 Metod normalne aproksimacije

Neka je $(\widehat{\alpha}_n, \widehat{c}_n)$ ocena maksimalne verodostojnosti za (α, c). Tako je $L(\widehat{\alpha}_n, \widehat{c}_n) = \max_{\alpha>0, c>0} L(\alpha, c)$. Neposredno se može proveriti da je:

$$\begin{aligned} \widehat{\alpha}_n &= \Big(\frac{1}{k}\sum_{i=1}^{k} \ln X_{(n-i+1,n)} - \ln X_{(n-k,n)}\Big)^{-1}, \\ \widehat{c}_n &= \frac{k}{n} X_{(n-k,n)}^{\widehat{\alpha}_n}. \end{aligned}$$

Iz (3.19) sledi da je prirodna ocena za x_p upravo:

$$\widehat{x}_p = (p_n/\widehat{c}_n)^{-1/\widehat{\alpha}_n}.$$

U cilju dobijanja asimptotske normalnosti za $\widehat{x}_p$ potrebni su jači uslovi od (3.18), već korišćeni ranije, tj. pretpostavlja se da je:

$$\lim_{t\to+\infty}\frac{U(tx)/U(t)-x^{1/\alpha}}{A(t)}=x^{1/\alpha}\frac{x^{\rho}-1}{\rho}, \tag{3.20}$$

gde je $A(t)\to 0$, $t\to+\infty$, za neko $\rho<0$, pravilno promenljiva funkcija sa indeksom ρ, gde je $U(x)=(1/(1-F))^{\leftarrow}(x)$.

Teorema 3.2.1. (Ferreira et al., 2003) *Pretpostavlja se da važi (3.20) i da je niz k intermidijatni. Ako je $\sqrt{k}A(n/k)\to 0$, $np_n=O(k)$ i $\ln(\frac{k}{np_n})/\sqrt{k}\to 0$, za $n\to\infty$, tada:*

$$\frac{\widehat{\alpha}_n\sqrt{k}}{\ln(k/(np_n))}\ln\frac{\widehat{x}_p}{x_p}\to_d N(0,1).$$

Tako je, bazirajući se na prethodnoj teoremi, interval poverenja za x_p sa nivoom značajnosti α':

$$I^n_{\alpha'}=\Big(\widehat{x}_p\exp\Big(-z_{\alpha'}\ln\Big(\frac{k}{np_n}\Big)/(\widehat{\alpha}_n\sqrt{k})\Big),\widehat{x}_p\exp\Big(z_{\alpha'}\ln\Big(\frac{k}{np_n}\Big)/(\widehat{\alpha}_n\sqrt{k})\Big)\Big),$$

gde $z_{\alpha'}$ zadovoljava $\Phi(z_{\alpha'})-\Phi(-z_{\alpha'})=\alpha'$. Za ovaj interval poverenja važi da je $P(x_p\in I^n_{\alpha'})\to\alpha'$, kada $n\to\infty$, tj. ima aimptotsku verovatnoću pokrivanja α'. Sledeća teorema daje pokrivenost intervala $I^n_{\alpha'}$.

Teorema 3.2.2. (Peng i Qi, 2006) *Pod uslovima prethodne teoreme važi:*

$$\begin{aligned}
P\ &\Big(\ \frac{\widehat{\alpha}_n\sqrt{k}}{\ln(k/(np_n))}\ln\frac{\widehat{x}_p}{x_p}\le x\Big)-\Phi(x)\\
=\ &\frac{1}{3\sqrt{k}}\phi(x)(1+2x^2)-\phi(x)\frac{\alpha}{1-\rho}\sqrt{k}A(n/k)-\frac{1}{2}x\phi(x)\Big(\ln\frac{k}{np_n}\Big)^{-2}\\
+\ &o\Big(\Big(\ln\frac{k}{np_n}\Big)^{-2}+\frac{1}{\sqrt{k}}+\sqrt{k}|A(n/k)|\Big),
\end{aligned}$$

uniformno za $-\infty<x<+\infty$. Takođe važi:

$$P(x_p\in I^n_{\alpha'})=\alpha'-z_{\alpha'}\phi(z_{\alpha'})\Big(\ln\frac{k}{np_n}\Big)^{-2}+o\Big(\Big(\ln\frac{k}{np_n}\Big)^{-2}+\frac{1}{\sqrt{k}}+\sqrt{k}|A(n/k)|\Big).$$

3.2.2 Metod količnika verodostojnosti

Neka su $\widehat{\alpha}_n$ i $\widehat{c}_n$ definisane u prethodnoj metodi. Ako se označi:

$$l_1 = \max_{\alpha>0,c>0} \ln L(\alpha, c) = \ln L(\widehat{\alpha}_n, \widehat{c}_n),$$

dalje treba maksimizirati $\ln L(\alpha, c)$, uz uslove:

$$\alpha > 0,\ c > 0,\ \alpha \ln x_p + \ln\left(\frac{p_n}{c}\right) = 0,$$

gde prethodna jednakost sledi iz $p_n = 1 - F(x_p) = c x_p^{\alpha}$. Sa $l_2(x_p)$ je označena funkcija maksimalne verodostojnosti. Može se pokazati da je:

$$l_2(x_p) = \ln L(\overline{\alpha}(\lambda), \overline{c}(\lambda)),$$

gde su:

$$\begin{aligned}\overline{\alpha}(\lambda) &= \frac{k}{\sum\limits_{i=1}^{k}(\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)}) + \lambda \ln X_{(n-k,n)} - \lambda \ln x_p},\\ \overline{c}(\lambda) &= X_{(n-k,n)}^{\overline{\alpha}(\lambda)} \frac{k-\lambda}{n-\lambda},\end{aligned}$$

i λ zadovoljava sledeće uslove:

$$\overline{\alpha}(\lambda) \ln x_p + \ln \frac{p_n}{\overline{c}(\lambda)} = 0, \tag{3.21}$$

$$\overline{\alpha}(\lambda) > 0,\ \lambda < k. \tag{3.22}$$

Dalje je označeno:

$$l(x_p) = -2(l_2(x_p) - l_1).$$

Teorema 3.2.3. (Peng i Qi, 2006) *Pretpostavlja se da važe uslovi Teoreme 3.2.1. Tada postoji jedinstveno rešenje za (3.21) i (3.22), u oznaci* $\widehat{\lambda}(x_p)$ *i važi:*

$$l(x_{p,0}) \to_d \chi_1^2,$$

za $\lambda = \widehat{\lambda}(x_{p,0})$ *u* $l_2(x_{p,0})$, *gde je* $x_{p,0}$ *tačna vrednost za* x_p.

Dalje je dat interval poverenja za x_p sa nivoom značajnosti α', baziran na prethodnoj graničnoj vrednosti:

$$I_{\alpha'}^{l} = \{x_p : l(x_p) \le u_{\alpha'}\},$$

gde je $u_{\alpha'}$ kritična tačka za χ_1^2 sa nivoom α'. Ovaj interval poverenja ima asimptotsku verovatnoću pokrivanja α', to je $P(x_p \in I_{\alpha'}^{l}) \to \alpha'$, kada $n \to \infty$.

3.2.3 Jednostrani interval poverenja za x_p

Kao i u prethodnim pododeljcima pretpostavlja se da je k_n intermidijatni niz, da je $1 - F(x) \in RV_{-1/\gamma}$ i da važi jednakost (3.2), za $U = (1/(1-F))^{\leftarrow}$. Da bi se ocenio visoki kvantil $x_p = F^{\leftarrow}(1 - p_n)$ u radovima Dekkers i de Haan, 1989 i Ferreira et al., 2003, predlažena je ocena:

$$\widehat{x}(k_n) = X_{(n-k_n,n)}\Big(\frac{k_n}{np_n}\Big)^{\widehat{\gamma}(k_n)},$$

gde je $\widehat{\gamma}(k_n)$ odgovarajuća Hillova ocena (2.9) odnosno (3.3).

Sledeći rezultat može se pogledati u Ferreira et al., 2003.

Lema 3.2.1. *Pretpostavlja se da važi (3.2), (3.6) i* $np_n \to const(\geq 0)$. *Neka je* k_n *intermidijatni niz takav da važe sledeća dva uslova:*

$$A(n/k_n)\sqrt{k_n} \to \lambda \in (-\infty, +\infty),$$

$$\ln(k_n/np_n)/\sqrt{k_n} \to 0.$$

Tada:

$$\frac{\sqrt{k_n}}{\gamma \ln(\frac{k_n}{np_n})}\Big(\frac{\widehat{x}(k_n)}{x_p} - 1\Big) \tag{3.23}$$

konvergira u raspodeli ka normalnoj slučajnoj veličini sa sredinom $\lambda/(\gamma(1-\rho))$ *i disperzijom* 1.

Sa ${}^{q}k_n^0$ se označava niz k_n koji minimizira srednjekvadratnu grešku granične raspodele u (3.23). Tada se u radu Ferreira et al., 2003 nalazi da je za slučaj $\gamma > 0$, ${}^{q}k_n^0 \sim k_n^0$, gde je k_n^0 iz jednakosti (3.7). Konzistentna ocena za ${}^{q}k_n^0$ je $\widehat{k_n^0}$ iz Odeljka 3.1.1., takva da je:

$$\frac{\widehat{k_n^0}}{{}^{q}k_n^0} \to_p 1. \tag{3.24}$$

Kada se zameni k_n sa $\widehat{k_n^0}$ sledi da važi rezultat prethodne leme, detalji dokaza se mogu pogledati u radu Ferreira i de Vries, 2004. Takođe, analogno rezultatima iz Odeljku 3.1.1. sledi sledeća teorema, sa jednostranim intervalom poverenja, čije je izdvajanje motivisano primenom VaR ocene (što je odgovarajući kvantil date raspodele). Kompletan dokaz može se pogledati u radu Ferreira et al., 2003.

Teorema 3.2.4. *Pretpostavlja se da važe (3.2), (3.6),* $np_n \to const(\geq 0)$ *i* $\ln(p_n) = o(n^{-\rho/(1-2\rho)})$*,* $n \to \infty$*. Neka* $\widehat{k}_n^0$ *zadovoljava uslov (3.24),* $\widehat{sgn}$ *zadovoljava uslov (3.9) i neka je* $\widehat{\rho}$ *konzistentna ocena parametra* ρ*. Tada, za* $n \to \infty$*:*

$$\frac{\sqrt{\widehat{k}_n^0}}{\widehat{\gamma}(\widehat{k}_n^0)\ln(\frac{\widehat{k}_n^0}{np_n})}\Big(\frac{\widehat{x}(\widehat{k}_n^0)}{x_p} - 1\Big) - \frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}}$$

konvergira u raspodeli ka standardnoj normalnoj slučajnoj veličini. Takođe, za $n \to \infty$*:*

$$P\Big(x_p < \widehat{x}(\widehat{k}_n^0)\Big(1 + \frac{\widehat{\gamma}(\widehat{k}_n^0)\ln(\frac{\widehat{k}_n^0}{np_n})}{\sqrt{\widehat{k}_n^0}}\Big(-z_\alpha + \frac{\widehat{sgn}}{\sqrt{-2\widehat{\rho}}}\Big)\Big)^{-1}\Big) \to 1 - \alpha,$$

što daje jednostrani interval poverenja za x_p*, sa verovatnoćom pokrivanja* $1 - \alpha$*.*

Treba primetiti da je u ovom poglavlju predstavljen poznati pristup za određivanje intervala poverenja indeksa ekstremne vrednosti i visokih kvantila. Teorema 3.1.3 i Teorema 3.2.2 su iskorišćene za izračunavanje verovatnoće pokrivanja odgovarajućih intervala poverenja spomenutih parametara. Takođe, kako daju Edgeworthov razvoj odgovarajućih statistika, mogle bi biti iskorišćene za dobijanje boljih intervala poverenja, poput korišćenja Teoreme 1.2.3. iz prvog poglavlja. To je jedan od mogućih pristupa za dalje istraživanje.

Na kraju sledi poslednje poglavlje, koje sadrži dosta teorijskih i tehničkih rezultata, motivisanih samom primenom teorije u praksi. Takođe, ono sadrži i originalne rezultate ove disertacije, koje je moguće uopštiti u duhu iznete teorije, što je ujedno jedan od mogućih pristupa za dalje istraživanje.

Glava 4

Teorija velikog odstupanja

Teorija velikog odstupanja je jedna od osnovnih tehnika moderne verovatnoće, koja se bavi problemom brzine opadanja verovatnoće retkih događaja. Jedan od prvih istraživača ove teme jeste dobitnik Abelove nagrade, S.R.S. Varadhan. Primene ove teorije su brojne, u samoj teoriji ali i u praktičnim problemima. Na primer, istaknute su primene u statističkoj mehanici, teoriji informacija, kao i u osiguranju.

Ovo poglavlje sadrži tri odeljka. Prvi odeljak sadrži osnovne principe i teoreme teorije velikog odstupanja. U drugom odeljku biće analizirana primena te teorije na kvantile određenih raspodela, gde su dobijeni originalni rezultati ove disertacije. Dok se u trećem odeljku razmatra konvergencija u velikom odstupanju i osnovni principi i teoreme, od kojih je većina tehničke prirode i predstavljaju uopštenje teorije za topološke prostere, gde se spominju prostor Hausdorffa i Tihonovljev prostor.

4.1 Osnovne teoreme teorije velikog odstupanja

Neka su X i $X_1, X_2, ...$ i.i.d. sa sredinom μ i disperzijom $\sigma^2 < +\infty$ i neka je S_n n-ta parcijalna suma, $S_n = \sum\limits_{i=1}^{n} X_i$. Na osnovu slabog zakona velikih brojeva, važi da empirijska sredina $\frac{1}{n}S_n$ konvergira u verovatnoći ka μ, tj. za svako $\epsilon > 0$, $\lim\limits_{n\to\infty} P\{|\frac{1}{n}S_n - \mu| > \epsilon\} = 0$. Dokaz se može izvesti primenom nejednakosti Čebišova.

Naime, važi:

$$\begin{aligned} P\Big\{\frac{1}{n}S_n > \mu+\epsilon\Big\} &= P\Big\{\Big(\sum_{i=1}^{n} X_i\Big) - n\mu > n\epsilon\Big\} \\ &\leq \frac{E[(\sum_{i=1}^{n} X_i - n\mu)^2]}{n^2\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0, \end{aligned}$$

i analogno sledi dokaz za suprotni događaj, $\frac{1}{n}S_n < \mu - \epsilon$.

Poznat je i jači rezultat, strogi zakon velikih brojeva, koji kaže:

$$P\{\lim_{n\to\infty} \frac{1}{n}S_n = \mu\} = 1.$$

Da bi se ovo dokazalo, koristi se Borel-Kantelijeva lema. Naime, iz $\sum_{n=1}^{\infty} P\{\frac{1}{n}S_n > \mu+\epsilon\} < +\infty$, sledi da za samo konačno mnogo n, skoro sigurno važi $\frac{1}{n}S_n > \mu+\epsilon$, odakle sledi: $\limsup_{n\to\infty} \frac{1}{n}S_n \leq \mu$, i potpuno analogno prethodnom, važi strogi zakon za suprotan događaj, $\frac{1}{n}S_n < \mu - \epsilon$.

Teoriju velikog odstupanja naime zanima da pronađe kojom brzinom tačno opada verovatnoća velikog odstupanja:

$$P\{\frac{1}{n}S_n > \mu+\epsilon\}.$$

Može se pokazati da to zavisi od finijih osobina slučajne veličine X nego što je konačnost disperzije. Tako je dalje posvećena pažnja slučajnim veličinama koje zadovoljavaju uslov:

$$\varphi(\lambda) := \ln E(e^{\lambda X}) < +\infty,\ \lambda \in \mathbb{R}. \tag{4.1}$$

U tom slučaju verovatnoća velikog odstupanja opada eksponencijalnom brzinom i upravo Kramerova teorema govori koliko brzo.

Teorema 4.1.1. (Kramerova teorema) *Neka su $X_1, X_2, ...$ i.i.d. sa sredinom μ, koje zadovoljavaju uslov (4.1) i S_n je njihova parcijalna suma. Tada, za svako $x > \mu$ važi:*

$$\lim_{n\to\infty} \frac{1}{n} \ln P\{\frac{1}{n}S_n \geq x\} = -\varphi^*(x),$$

gde je φ^* Legendreova transformacija od φ, tj.

$$\varphi^*(x) := \sup_{\lambda\in\mathbb{R}}\{\lambda x - \varphi(\lambda)\}.$$

Dokaz prethodne teoreme može se pogledati u radu Mörters, 2008.

Treba reći da je kasnije u Odeljku 4.3.2., gde se razmatra konvergencija u velikom odstupanju u Tihonovljevom prostoru, data teorama koja se može posmatrati kao analogon Kramerove teoreme za prostor $\mathbb{R}^k$, pogledati Teoremu 4.3.11.

Primer 4.1.1. *Neka je dat jedan nefer novčić, koji ima realizaciju* X*, nulu sa verovatnoćom* $1-p$ *i 1 sa verovatnoćom* p*. Može se izračunati da je* $\varphi(\lambda) = \ln(pe^\lambda + (1-p))$ *i da je za* $0 < x < 1$, $\varphi^*(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}$.

Znači, može se zaključiti da verovatnoća događaja velikog odstupanja, tipa $\{S_n - \mu n \geq nx\}, x > 0$, tj. da parcijalna suma prelazi srednju vrednost za više od nx, opada eksponencijalno. Dok centralna granična teorema govori da je red kojim parcijalna suma normalno prelazi srednju vrednost, upravo $\sqrt{n}$. Preciznije,

$$P\{S_n - \mu n \geq \sqrt{n}x\} \to 1 - \Phi(x/\sigma) > 0,$$

gde je Φ funkcija raspodele standardne normalne veličine. Odakle sledi da je za svaki niz a_n, za koji važi $\sqrt{n} \ll a_n \ll n$, takođe $P\{S_n - \mu n \geq a_n\} \to 0$, i može se primetiti da ni centralna granična teorema ni Kramerova teorema, ne govore koja je brzina ove konvergencije. Ovo pitanje je u fokusu umerenog principa velikog odstupanja.

Teorema 4.1.2. (Umereni princip velikok odstupanja) *Pod pretpostavkama Kramerove teoreme, ako je* $\sqrt{n} \ll a_n \ll n$*, važi za svako* $x > 0$,

$$\lim_{n\to\infty} \frac{n}{a_n^2} \ln P\{S_n - \mu n \geq x a_n\} = -\frac{x^2}{2\sigma^2}.$$

Dokaz prethodne teoreme može se pogledati u radu Mörters, 2008.

Treba istaći da se umereni princip velikog odstupanja bavi objašnjenjem opadanja odstupanja od sredine, korišćenjem teorema centralnog tipa i utvrđuje kada se može koristiti normalna aproksimacija kako bi se izračunala verovatnoća repa, i takođe se bavi univerzalnošću brzine opadanja.

Sledi jednostavan rezultat formulisan u narednoj lemi, koji je bitan za teoriju velikog odstupanja.

Lema 4.1.1. (Laplasov princip) *Za niz* $a_n \to \infty$ *i konačan broj* N *nenegativnih nizova* $(b_n^{(1)}), ..., (b_n^{(N)})$, *važi:*

$$\lim_{n\to\infty} \sup \frac{1}{a_n} \ln \sum_{i=1}^{N} b_n^{(i)} = \max_{1\le i\le N} \lim_{n\to\infty} \sup \frac{1}{a_n} \ln b_n^{(i)}.$$

Dokaz: Za svako fiksirano n važi:

$$0 \le \ln \sum_{i=1}^{N} b_n^{(i)} - \max_{1\le i\le N} \ln b_n^{(i)} \le \ln N,$$

i ukoliko se podele obe strane sa a_n, uzimajući limsup, može se pokazati:

$$\lim_{n\to\infty} \sup \frac{1}{a_n} \ln \sum_{i=1}^{N} b_n^{(i)} = \lim_{n\to\infty} \sup \frac{1}{a_n} \max_{1\le i\le N} \ln b_n^{(i)} = \max_{1\le i\le N} \lim_{n\to\infty} \sup \frac{1}{a_n} \ln b_n^{(i)},$$

što je i trebalo dokazati. ∎

Dalje, na osnovu Kramerove teoreme i prethodne leme, sledi sledeći rezultat.

Lema 4.1.2. *Za svaki Borelov skup* $A \subset R$, *važi:*

$$\begin{aligned} \lim_{n\to\infty} \sup \frac{1}{n} \ln P\Big\{\frac{1}{n} S_n \in A\Big\} &\le -\inf_{x \in clA} \varphi^*(x), \\ \lim_{n\to\infty} \inf \frac{1}{n} \ln P\Big\{\frac{1}{n} S_n \in A\Big\} &\ge -\inf_{x \in intA} \varphi^*(x), \end{aligned}$$

gde $intA$ *i* clA *označavaju redom unutrašnjost i topološko zatvaranje skupa* A.

Dokaz prethodne leme može se pogledati u radu Mörters, 2008.

Dalje, za uopštenje teorije velikog odstupanja može se razmatrati niz slučajnih veličina $X_1, X_2, ...$ u metričkom prostoru M i može se razmatrati događaj tipa $\{X_n \in A\}$, gde je $A \subset M$ Borelov skup. Treba naći funkciju koja zamenjuje φ^*. Kasnije će biti više reči o uopštenju na metričkom prostoru i samoj konvergenciji u velikom odstupanju.

Definicija 4.1.1. *Neka je* M *fiksirani metrički prostor. Funkcija* $I : M \to [0, +\infty]$ *se naziva*

- funkcija brzine *ako je donje semineprekidna, što znači da je skup* $\{x \in M : I(x) \leq a\}$ *zatvoren za svako* $a \geq 0$,
- dobra funkcija brzine *ako je skup* $\{x \in M : I(x) \leq a\}$ *kompaktan za svako* $a \geq 0$.

Dalje se uvodi značenje principa velikog odstupanja.

DEFINICIJA 4.1.2. *Za niz slučajnih veličina* $X_1, X_2, ...$ *sa vrednostima u metričkom prostoru* M, *se kaže da zadovoljava princip velikog odstupanja, sa brzinom* $a_n \to \infty$ *i funkcijom brzine* I, *ako za sve Borelove skupove* $A \subset M$, *važi:*

$$\begin{aligned}\lim_{n\to\infty} \sup \frac{1}{a_n} \ln P\{X_n \in A\} &\leq -\inf_{x \in clA} I(x),\\ \lim_{n\to\infty} \inf \frac{1}{a_n} \ln P\{X_n \in A\} &\geq -\inf_{x \in intA} I(x).\end{aligned}$$

Može se primetiti da je funkcija brzine principa velikog odstupanja jedistveno određena, kroz donju semineprekidnost. Takođe treba primetiti da parcijalna suma $S_n = \sum_{i=1}^{n} X_i$ zadovoljava princip velikog odstupanja sa brzinom n i dobrom funkcijom brzine φ^*, iz Kramerove teoreme i da za svaki niz $\sqrt{n} \ll a_n \ll n$ slučajna veličina $\frac{S_n - \mu n}{a_n}$ zadovoljava princip velikog odstupanja sa brzinom $\frac{a_n^2}{n}$ i dobrom funkcijom brzine $I(x) = \frac{x^2}{2\sigma^2}$, iz umerenog principa velikog odstupanja.

Treba reći da se kasnije definiše pojam konvergencije u velikom odstupanju, koji je na Tihonovljevom prostoru ekvivalentan principu velikog odstupanja, pogledati Teoremu 4.3.5.

U nastavku pažnja će biti posvećena dvema bitnim teorijskim temama koje su vezane za uopštenje. Treba odgovoriti na sledeća dva pitanja: da li se mogu dokazati nejednakosti iz poslednje definicije ne za Borelov skup A već za užu familiju, i pitanje ako važi princip velikog odstupanja za niz X_n i ako je $f : M \to M'$ neprekidna funkcija, da li važi princip vlikog odstupanja za $f(X_n)$?

U vezi prvog pitanja, jasno je da se mogu posmatrati otvoreni skupovi za donju granicu i zatvoreni skupovi za gornju granicu. Ali treba proveriti donju granicu za otvorene lopte i gornju granicu za kompaktne skupove. Sledi lema koja pomaže u rešavanju problema donje granice.

Lema 4.1.3. *Ako je I funkcija brzine i A Borelov skup, takav da je za svako $x \in A$ i $\epsilon > 0$, $B(x,\epsilon) \subset A$,*

$$\lim_{n\to\infty} \inf \frac{1}{a_n} \ln P\{X_n \in B(x,\epsilon)\} \geq -I(x),$$

tada:

$$\lim_{n\to\infty} \inf \frac{1}{a_n} \ln P\{X_n \in A\} \geq -\inf_{x \in intA} I(x).$$

Dokaz prethodne leme može se pogledati u radu Mörters, 2008.

Zamena zatvorenih skupova sa kompaktnim skupovima za gornju granicu, zahteva dodatne uslove, koje je uvek teže proveriti.

Definicija 4.1.3. *Niz slučajnih veličina X_n se naziva eksponencijalno čvrst ako za svako $N < +\infty$ postoji kompaktan skup $K \subset M$ tako da je:*

$$\lim_{n\to\infty} \sup \frac{1}{a_n} \ln P\{X_n \notin K\} < -N.$$

Lema 4.1.4. *Ako je familija X_n eksponencijalno čvrsta i zadovoljava veliko odstupanje za gornju granicu za svaki kompaktan skup, tada zadovoljava veliko odstupanje i za svaki zatvoreni skup.*

Dokaz prethodne leme može se pogledati u radu Mörters, 2008.

Treba primetiti, da ukoliko niz X_n zadovoljava princip velikog odstupanja sa brzinom a_n i funkcijom brzine I, ako je niz eksponencijalno čvrst, tada je I dobra funkcija brzine. Obrnuto, ako je M separabilan metrički prostor i I dobra funkcija brzine, tada je niz eksponencijalno čvrst.

Za drugo pitanje, ako važi princip velikog odstupanja za niz X_n i ako je $f : M \to M'$ neprekidna funkcija, da li važi princip vlikog odstupanja za $f(X_n)$, prilično je jasan odgovor.

Lema 4.1.5. (Princip kontrakcije) *Ako $X_1, X_2, ...$ zadovoljavaju princip velikog odstupanja sa brzinom a_n i dobrom funkcijom brzine I i ako je $f : M \to M'$ neprekidno preslikavanje, tada niz $f(X_1), f(X_2), ...$ zadovoljava princip velikog odstupanja sa brzinom a_n i dobrom funkcijom brzine J, koja je data:*

$$J(y) = \inf_{x \in f^{-1}(y)} I(x).$$

Dokaz prethodne leme može se videti u radu Mörters, 2008.

Treba reći da je analogan rezultat principu kontrakcije, u Tihonovljevom prostoru, dat u Teoremi 4.3.6, u Odeljku 4.3.2.

U radu Varadhan, 1984, dat je precizan rezultat o eksponencijalnom rastu funkcije oblika:

$$E[e^{nf(X_n)}],\ n \to \infty.$$

Jedna od karakterističnih primena ovog rezultata jeste u modelima statističke mehanike, za izračunavanje slobodne energije.

Teorema 4.1.3. (Varadhanova lema) *Ako je X_n familija slučajnih veličina koje uzimaju vrednosti iz metričkog prostora M, koja zadovoljava princip velikog odstupanja sa brzinom n i funkcijom brzine I i $f : M \to \mathbb{R}$ je neprekidna funkcija, koja je ograničena odozgo, tada:*

$$\lim_{n\to\infty} \frac{1}{n} \ln E[e^{nf(X_n)}] = \sup_{x\in M}\{f(x) - I(x)\}. \tag{4.2}$$

Dokaz prethodne teoreme može se pogledati u radu Mörters, 2008.

Treba primetiti da prethodni rezultat (4.2) proširuje Laplasov princip na prirodan način. Može se pisati:

$$E[e^{nf(X_n)}] \approx \sum_x e^{nf(x)} P\{X_n \approx x\} \approx \sum_x e^{n(f(x)-I(x))},$$

i sada najveći eksponent u sumi određuje brzinu sume.

Takođe, sledi rezultat koji je inverzan Varadhanovoj lemi, Brycova lema.

Teorema 4.1.4. (Brycova lema) *Ako je X_n eksponencijalno čvrsta familija slučajnih veličina, koje uzimaju vrednosti u metričkom prostoru M tako da:*

$$\Lambda(f) := \lim_{n\to\infty} \frac{1}{n} \ln E[e^{nf(X_n)}]$$

postoji za sve neprekidne i ograničene funkcije $f : M \to \mathbb{R}$, tada slučajne veličine X_n zadovoljavaju princip velikog odstupanja sa brzinom n i dobrom funkcijom brzine:

$$J(x) = \sup\{f(x) - \Lambda(f)\}.$$

Pored toga, važi i rezultat:

$$\Lambda(f) = \sup_{x\in M}\{f(x) - J(x)\}.$$

Značaj Brycove leme leži u činjenici da egzistencija $\Lambda(f)$ treba da se proveri samo za ograničene, neprekidne funkcije. Kako je ova klasa suviše široka za praktičnu primenu, preferira se rad sa mnogo restriktivnijim uslovima, u kojima egzistencija $\Lambda(f)$ treba da se proveri za užu klasu funkcija. Jedan od uspešnijih rezultata u ovom smislu jeste Gärtner-Ellis teorema, koja pod odgovarajućim pretpostavkama, dozvoljava testiranje samo linearnih funkcija f.

Teorema 4.1.5. (Gärtner-Ellis teorema) *Pretpostavlja se da je X_n familija slučajnih veličina koje uzimaju vrednosti iz $\mathbb{R}^k$ tako da:*

$$\Lambda(x) := \lim_{n\to\infty} \frac{1}{n} \ln E[e^{n\langle x, X_n\rangle}]$$

postoji i konačno je za svako $x \in \mathbb{R}^k$. Ako je Λ još i diferencijabilna, tada slučajne veličine X_n zadovoljavaju princip velikog odstupanja sa brzinom n i dobrom funkcijom brzine:

$$J(x) = \sup_{y\in\mathbb{R}^k} \{\langle x, y\rangle - \Lambda(y)\}.$$

Dokaz prethodne teoreme može se pogledati u radu Mörters, 2008.

Može se primetiti da je Gärtner-Ellis teorema za slučaj kada je $k = 1$, prirodno uopštenje Kramerove teoreme za zavisne slučajne veličine. U tom slučaju J je dato pomoću Legendreove transformacije.

4.2 Primena teorije velikog odstupanja na kvantile raspodele

Neka je F nepoznata funkcija raspodele i odgovarajući p-kvantil je $x_p = F^{\leftarrow}(p)$, gde je:

$$F^{\leftarrow}(p) = \inf\{t : F(t) \geq p\}, \quad p \in (1, 0).$$

Standardna ocena kvantila x_p je direktno-simulirana ocena, koja je data formulom:

$$\widehat{x}_n = \inf\{t : F_n(t) \geq p\},$$

gde je $F_n(t)$ empirijska funkcija raspodele uzorka $\{Y_i, i = 1, 2, ..., n\}$.

U ovom odeljku biće korišćen pojam negativne zavisnosti, što je uopštenje pojma negativne kvadratne zavisnosti, koju je definisao Lehmann, 1966, pogledati takođe Nelsen, 1999.

Definicija 4.2.1. *Za par* (X, Y) *se kaže da je negativno kvadratno zavisan ako je:*

$$P(X \leq x, Y \leq y) \leq P(X \leq x)P(Y \leq y),$$

za sve x, y.

Definicija 4.2.2. (Negativna zavisnost) *Slučajne veličine* X_i, $i = 1, 2, ..., n$ *su negativno zavisne ako za sve* $x_1, x_2, ..., x_n$ *važe sledeće dve nejednakosti:*

$$P(X_1 \leq x_1, ..., X_n \leq x_n) \leq P(X_1 \leq x_1) \cdot ... \cdot P(X_n \leq x_n),$$

$$P(X_1 \geq x_1, ..., X_n \geq x_n) \leq P(X_1 \geq x_1) \cdot ... \cdot P(X_n \geq x_n).$$

Treba primetiti da je svaki par iz skupa negativno zavisnih slučajnih veličina negativno kvadratno zavisan. U specijalnom slučaju

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1) \cdot ... \cdot P(X_n \leq x_n)\,,$$

dobija se nezavisnost.

Lema 4.2.1. *Ako su* X_i, $i = 1, ..., n$ *negativno zavisne i* $E|X_i| < +\infty$, $i = 1, ..., n$, *tada je:*

$$E(X_iX_j) \leq E(X_i)E(X_j), i \neq j, i, j = 1, ..., n,$$

gde jednakost važi ako su $X_i, i = 1, ..., n$ *nezavisne. Takođe važi, ako su* $X_i, i = 1, ..., n$ *nenegativne slučajne veličine i* $E(X_1 \cdot ... \cdot X_n) < +\infty$, *tada:*

$$E(X_1 \cdot ... \cdot X_n) \leq E(X_1) \cdot ... \cdot E(X_n).$$

Lema 4.2.2. *Neka su slučajni vektori* $X^{(j)} = (X_1^{(j)}, ..., X_n^{(j)})$, $j = 1, ..., d$, *uzajamno nezavisni i neka su slučajne veličine* $X_1^{(j)}, ..., X_n^{(j)}$ *negativno zavisne za svako* $j = 1, ..., d$. *Neka su* $f_1(\cdot), ..., f_n(\cdot)$ *realne funkcije od* d *argumenata saglasne svakoj koordinati, tada su* $f_i(X_i^{(1)}, ..., X_i^{(d)})$, $i = 1, ..., n$ *negativno zavisne.*

Dokaz prethodne dve leme može se naći u radu Jin and C.Fu, 2002.

Dalje je dat osnovni rezultat, tj. da je konvergencija direktno-simulirane ocene kvantila ka tačnoj vrednosti kvantila upravo eksponencijalna, gde je primenjen princip velikog odstupanja. Osnov te teorije koja se koristi može se naći u radovima Bucklew, 1990, Dembo i Zeitouni, 1998, Deuschel i Stroock, 1989.

Posmatra se slučajna veličina Y sa srednjom vrednošću $\mu = EY$ i funkcijom generatrise momenata $M(\lambda) = E[e^{\lambda Y}]$. Važi da je $M(\lambda) > 0$ za sve $\lambda \in \mathbb{R}$ i domen $\{\lambda : M(\lambda) < +\infty\}$ funkcije generatrise momenata je interval koji sadrži 0. Konjugat funkcija je

$$I(x) = \varphi^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \varphi(\lambda)\},$$

gde je $\varphi(\lambda) = \ln M(\lambda)$.

Neka su $Y_1, Y_2, ..., Y_n$ i.i.d. i $\frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ odgovarajuća uzoračka sredina. Ako $M(\lambda)$ postoji u ϵ okolini tačke $\lambda = 0$ za neko $\epsilon > 0$ tada je:

$$\begin{aligned} -\inf_{x \in int\Gamma} I(x) &\leq \lim_{n\to\infty} \inf \frac{\ln(P[\frac{1}{n}S_n \in \Gamma])}{n} \\ &\leq \lim_{n\to\infty} \sup \frac{\ln(P[\frac{1}{n}S_n \in \Gamma])}{n} \\ &\leq -\inf_{x \in cl\Gamma} I(x), \end{aligned}$$

i $I(x)$ je odgovarajuća funkcija data:

$$\begin{aligned} I(x) &= \sup_{\lambda \in (-\epsilon,\epsilon)} \left\{ \lambda x - \lim_{n\to\infty} \frac{\ln(E[e^{\lambda S_n}])}{n} \right\} \\ &= \sup_{\lambda \in (-\epsilon,\epsilon)} \{\lambda x - \varphi(\lambda)\}, \end{aligned}$$

gde druga jednakost sledi iz $E[e^{\lambda S_n}] = (E[e^{\lambda Y}])^n$, jer su $Y_1, ..., Y_n$ nezavisne jednako raspodeljene slučajne veličine.
Posebno za $\Gamma = \{x : x \in [\mu - \epsilon, \mu + \epsilon]\}$ važi:

$$\begin{aligned} -\inf_{x \in int\Gamma} I(x) &= -\inf_{x \in cl\Gamma} I(x) \\ \lim_{n\to\infty} \frac{\ln(P[\frac{1}{n}S_n \in \Gamma])}{n} &= -\inf_{x \in cl\Gamma} I(x). \end{aligned}$$

Lema 4.2.3. *Neka su $\{Y_n, n \geq 1\}$ negativno zavisne i i.i.d. sa funkcijom generatrise momenata $M(\lambda) = E[e^{\lambda Y_1}]$. Neka je $S_n = \sum_{i=1}^{n} Y_i$. Ako $M(\lambda)$ postoji u ϵ okolini tačke $\lambda = 0$ za neko $\epsilon > 0$, tada:*

$$P[\frac{1}{n}S_n \geq x] \leq e^{-n\Delta_+(x,n)}, \forall x > E(Y_1),$$

$$P[\frac{1}{n}S_n \leq x] \leq e^{-n\Delta_-(x,n)}, \forall x < E(Y_1),$$

gde su:

$$\begin{aligned}
\Delta_+(x,n) &= \sup_{0\leq\lambda\leq\epsilon}\Big(\lambda x - \frac{\ln E\{e^{\lambda S_n}\}}{n}\Big) \\
&\geq \sup_{0\leq\lambda\leq\epsilon}(\lambda x - \ln E\{e^{\lambda Y_1}\}) > 0, \\
\Delta_-(x,n) &= \sup_{-\epsilon\leq\lambda\leq 0}\Big(\lambda x - \frac{\ln E\{e^{\lambda S_n}\}}{n}\Big) \\
&\geq \sup_{-\epsilon\leq\lambda\leq 0}(\lambda x - \ln E\{e^{\lambda Y_1}\}) > 0.
\end{aligned}$$

Obrnuto, ako je $E[|Y_1|] < +\infty$ i za svako $x > E(Y_1)$, postoji $\alpha(x) > 0$ tako da je ispunjen uslov:

$$P[\frac{1}{n}S_n \geq x] \leq e^{-n\alpha(x)},$$

i ako za svako $x < E(Y_1)$, postoji $\beta(x) > 0$ tako da je:

$$P[\frac{1}{n}S_n \leq x] \leq e^{-n\beta(x)},$$

tada funkcija generatrise momenata $M(\lambda)$ postoji u ϵ okolini tačke $\lambda = 0$ za neko $\epsilon > 0$.

Dokaz prethodne leme može se pogledati u Jin i Michael C.Fu, 2002.

Dalje sledi teorema koja pokazuje da standardna direktno-simulirana ocena kvantila $\widehat{x}(n)$ konvergira ka x_p eksponencijalnom brzinom u verovatnoći, kada $n \to \infty$, za slučaj neprekidne funkcije raspodele F.

Teorema 4.2.1. (Jin i Michael C.Fu, 2002) *Ako je funkcija raspodele F strogo rastuća i $\{Y_n, n \geq 1\}$ su negativno zavisne, tada je:*

$$P\Big[\big|\widehat{x}(n) - x_p\big| \geq \epsilon\Big] \leq e^{-n\Delta_+(\epsilon,n)} + e^{-n\Delta_-(\epsilon,n)}, \forall \epsilon > 0, \tag{4.3}$$

gde su:

$$\begin{aligned}\Delta_+(\epsilon, n) &= \sup_{-\infty<\lambda\leq 0}\Big(\lambda p - \frac{\ln E\{\exp[\lambda\sum_{i=1}^n I(Y_i \leq x_p + \epsilon)]\}}{n}\Big),\\ \Delta_-(\epsilon, n) &= \sup_{0\leq\lambda<\infty}\Big(\lambda p - \frac{\ln E\{\exp[\lambda\sum_{i=1}^n I(Y_i \leq x_p - \epsilon)]\}}{n}\Big).\end{aligned}$$

Takođe, brzina je povećana negativnom zavisnošću u smislu da je ispunjeno:

$$\begin{aligned}\Delta_+(\epsilon, n) &\geq \sup_{-\infty<\lambda\leq 0}(\lambda p - \ln E\{\exp[\lambda I(Y \leq x_p + \epsilon)]\}) > 0,\\ \Delta_-(\epsilon, n) &\geq \sup_{0\leq\lambda<\infty}(\lambda p - \ln E\{\exp[\lambda I(Y \leq x_p - \epsilon)]\}) > 0,\end{aligned}$$

gde se desna strana odnosi na brzinu u slučaju linearno nezavisnog jednako raspodeljenog uzorka.

Dokaz: Iz definicije direktno-simulirane ocene $\widehat{x}(n)$ sledi:

$$\begin{aligned}P\big[\widehat{x}(n) - x_p \leq -\epsilon\big] &= P\big[F_n^{-1}(p) \leq x_p - \epsilon\big]\\ &= P\big[F_n(x_p - \epsilon) \geq p\big]\\ &= P\Big[\frac{\sum_{i=1}^n I(Y_i \leq x_p - \epsilon)}{n} \geq p\Big]\\ &\leq \exp\Big[-n\sup_{0\leq\lambda<+\infty}\Big(\lambda p - \frac{\ln E\{\exp[\lambda\sum_{i=1}^n I(Y_i \leq x_p - \epsilon)]\}}{n}\Big)\Big],\end{aligned}$$

gde poslednja nejednakost sledi iz Leme 4.2.3.
Na potpuno analogan način može se pokazati:

$$P\big[\widehat{x}(n) - x_p \geq \epsilon\big] \leq \exp\Big[-n\sup_{-\infty<\lambda\leq 0}\Big(\lambda p - \frac{\ln E\{\exp[\lambda\sum_{i=1}^n I(Y_i \leq x_p + \epsilon)]\}}{n}\Big)\Big].$$

Dalje je:

$$\begin{aligned}P\Big[\big|\widehat{x}(n) - x_p\big| \geq \epsilon\Big] &= P\Big[\widehat{x}(n) - x_p \leq -\epsilon\Big] + P\Big[\widehat{x}(n) - x_p \geq \epsilon\Big]\\ &\leq \exp\Big[-n\sup_{0\leq\lambda<+\infty}\Big(\lambda p - \frac{\ln E\{\exp[\lambda\sum_{i=1}^n I(Y_i \leq x_p - \epsilon)]\}}{n}\Big)\Big]\\ &+ \exp\Big[-n\sup_{-\infty<\lambda\leq 0}\Big(\lambda p - \frac{\ln E\{\exp[\lambda\sum_{i=1}^n I(Y_i \leq x_p + \epsilon)]\}}{n}\Big)\Big],\end{aligned}$$

čime je dokazana nejednakost (4.3).

Dalje se može primetiti da su indikatori $I(Y_i \leq x_p - \epsilon)$ i $I(Y_i \leq x_p + \epsilon)$ opadajuće funkcije po Y_i za svako i, i na osnovu Leme 4.2.1 i Leme 4.2.2 važi:

$$\sup_{-\infty<\lambda\leq 0}\Big(\lambda p - \frac{\ln E\{\exp[\lambda \sum_{i=1}^{n} I(Y_i \leq x_p+\epsilon)]\}}{n}\Big)$$
$$\geq \sup_{-\infty<\lambda\leq 0}(\lambda p - \ln E\{\exp[\lambda I(Y \leq x_p+\epsilon)]\}) > 0,$$

$$\sup_{0\leq\lambda<+\infty}\Big(\lambda p - \frac{\ln E\{\exp[\lambda \sum_{i=1}^{n} I(Y_i \leq x_p-\epsilon)]\}}{n}\Big)$$
$$\geq \sup_{0\leq\lambda<+\infty}(\lambda p - \ln E\{\exp[\lambda I(Y \leq x_p-\epsilon)]\}) > 0,$$

gde pozitivnost sledi iz činjenice da je funkcija raspodele F strogo rastuća, te su

$$E[I(Y \leq x_p - \epsilon)] = F(x_p - \epsilon) < F(x_p) = p \quad \text{i}$$
$$E[I(Y \leq x_p + \epsilon)] = F(x_p + \epsilon) > F(x_p) = p.$$

Ovim je dokaz završen. ■

Dalje su date tri teoreme koje govore o brzini konvergencije standardne ocene kvantila za slučaj Pareto, uopštene Pareto i Gama raspodele, što je originalni rezultat ove disertacije.

Teorema 4.2.2. (Stanojević, 2014) *Neka su $\{Y_n, n \geq 1\}$ negativno zavisne slučajne veličine sa zajedničkom Pareto funkcijom raspodele:*

$$\overline{F}(x) \;=\; Kx^{-\alpha},\ x \geq \sqrt[\alpha]{K},\ K, \alpha > 0.$$

Važi sledeće ograničenje brzine konvergencije standardne ocene kvantila x_p:

$$P\Big[\big|\widehat{x}(n) - x_p\big| \geq \epsilon\Big] \;\leq\; \Big(\frac{pK(x_p+\epsilon)^{-\alpha}}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)}\Big)^{-pn}\cdot\Big(\frac{K(x_p+\epsilon)^{-\alpha}}{1-p}\Big)^{n}$$
$$+\;\Big(\frac{pK(x_p-\epsilon)^{-\alpha}}{(1-K(x_p-\epsilon)^{-\alpha})(1-p)}\Big)^{-pn}\cdot\Big(\frac{K(x_p-\epsilon)^{-\alpha}}{1-p}\Big)^{n}.$$

Dokaz: Na osnovu Teoreme(4.2.1) važi:

$$P\Big[\big|\widehat{x}(n)-x_p\big|\geq\epsilon\Big]\leq e^{-n\Delta_+(\epsilon,n)}+e^{-n\Delta_-(\epsilon,n)},\forall\epsilon>0,$$

gde je:

$$\Delta_+(\epsilon,n)\geq\sup_{-\infty<\lambda\leq 0}(\lambda p-\ln E\{\exp[\lambda I(Y\leq x_p+\epsilon)]\})=\Delta+,$$

$$\Delta_-(\epsilon,n)\geq\sup_{0\leq\lambda<+\infty}(\lambda p-\ln E\{\exp[\lambda I(Y\leq x_p-\epsilon)]\})=\Delta-.$$

Dalje se računaju vrednosti $\Delta+$ i $\Delta-$, što daje grublju ocenu tražene verovatnoće, ali i dalje zadovoljavajuću.

Ako se označi $p=P[Y\leq x_p+\epsilon]=F(x_p+\epsilon)=1-K(x_p+\epsilon)^{-\alpha}$ i razmatra raspodela indikatora:

$$I(Y\leq x_p+\epsilon):\begin{pmatrix}0 & 1\\ \text{1-p} & \text{p}\end{pmatrix},$$

gde je $p=1-K(x_p+\epsilon)^{-\alpha}$ i $1-p=K(x_p+\epsilon)^{-\alpha}$, dobija se:

$$\Delta+=\sup_{-\infty<\lambda\leq 0}(\lambda p-\ln(e^{\lambda}(1-K(x_p+\epsilon)^{-\alpha})+K(x_p+\epsilon)^{-\alpha})).$$

Posle diferenciranja i sređivanja izraza, važi:

$$\frac{pK(x_p+\epsilon)^{-\alpha}-e^{\lambda}(1-K(x_p+\epsilon)^{-\alpha})(1-p)}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)}=0,$$

$$\begin{aligned}e^{\lambda}&=\frac{pK(x_p+\epsilon)^{-\alpha}}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)},\\ \lambda&=\ln\frac{pK(x_p+\epsilon)^{-\alpha}}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)}.\end{aligned}$$

Kako je $1-Kx_p^{-\alpha}=p$ i $K(x_p+\epsilon)^{-\alpha}<Kx_p^{-\alpha}=1-p$, važi nejednakost:

$$(1-K(x_p+\epsilon)^{-\alpha})(1-p)>0,$$

što ne utiče na znak prvog izvoda. Takođe je:

$$\frac{pK(x_p+\epsilon)^{-\alpha}}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)} < 1,$$

odakle sledi da se maksimum postiže u izračunatoj vrednosti λ, koja je negativna. To je ujedno i vrednost u kojoj $\Delta+$ postiže sup na traženom intervalu. Sledi da je:

$$\Delta+ = p\ln\frac{pK(x_p+\epsilon)^{-\alpha}}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)} - \ln\frac{K(x_p+\epsilon)^{-\alpha}}{(1-p)}.$$

Posle primene analognog postupka, prilikom izračunavanje $\Delta+$, sada na izračunavanje vrednosti $\Delta-$ i analize u kojoj je $K(x_p-\epsilon)^{-\alpha} > Kx_p^{-\alpha} = 1-p$ i $(1-K(x_p-\epsilon)^{-\alpha})(1-p) > 0$, važi:

$$\begin{aligned} e^{\lambda} &= \frac{pK(x_p-\epsilon)^{-\alpha}}{(1-K(x_p-\epsilon)^{-\alpha})(1-p)}, \\ \lambda &= \ln\frac{pK(x_p-\epsilon)^{-\alpha}}{(1-K(x_p-\epsilon)^{-\alpha})(1-p)} > 0. \end{aligned}$$

Za ovu vrednost λ, $\Delta-$ postiže sup na traženom intervalu. Sledi da je:

$$\Delta- = p\ln\frac{pK(x_p-\epsilon)^{-\alpha}}{(1-K(x_p-\epsilon)^{-\alpha})(1-p)} - \ln\frac{K(x_p-\epsilon)^{-\alpha}}{(1-p)}.$$

Na osnovu dobijenih rezultata, važi:

$$\begin{aligned} P\Big[\big|\widehat{x}(n)-x_p\big| \geq \epsilon\Big] &\leq e^{-n\Delta+} + e^{-n\Delta-} \\ &= \Big(\frac{pK(x_p+\epsilon)^{-\alpha}}{(1-K(x_p+\epsilon)^{-\alpha})(1-p)}\Big)^{-pn} \cdot \Big(\frac{K(x_p+\epsilon)^{-\alpha}}{1-p}\Big)^{n} \\ &+ \Big(\frac{pK(x_p-\epsilon)^{-\alpha}}{(1-K(x_p-\epsilon)^{-\alpha})(1-p)}\Big)^{-pn} \cdot \Big(\frac{K(x_p-\epsilon)^{-\alpha}}{1-p}\Big)^{n}, \end{aligned}$$

što je i trebalo pokazati. ■

Na osnovu prethodne teoreme, za određenu Pareto funkciju raspodele sa poznatim vrednostima parametara, može se izračunati verovatnoća pokrivenosti za standardnu ocenu kvantila x_p. U Tabeli 4.1.1, Tabeli 4.1.2, Tabeli

4.1.3 i Tabeli 4.1.4 dati su rezultati za dva izbora parametara K, α i p, i za svaku kombinaciju parametara su istaknute dve vrednosti ϵ. Na osnovu rezultata dobijenih u tabelama, može se videti da pokrivenost raste sa porastom veličine uzorka i da je za finiju ocenu pokrivenosti potreban dosta veći uzorak, što bi i trebalo očekivati.

Tabela 4.1.1.

K=3,$\alpha = 2$	p=0,5	$x_p = 2,44949$
	n	$e^{-n\triangle+} + e^{-n\triangle-}$
$\epsilon = 0,1$:	250	0,863757
	400	0,524585
	500	0,376996
	700	0,19566
	1000	0,074017
$\epsilon = 0,05$:	1000	0,867838
	2000	0,377556
	3000	0,164684
	4000	0,072016
	7000	0,006112

Tabela 4.1.2.

K=3,$\alpha = 2$	p=0,75	$x_p = 3,464102$
	n	$e^{-n\triangle+} + e^{-n\triangle-}$
$\epsilon = 0,1$:	1500	0,868021
	2000	0,657577
	3000	0,377705
	5000	0,125039
	7000	0,04158
$\epsilon = 0,05$:	7000	0,756232
	9000	0,572858
	12000	0,377739
	17000	0,188766
	22000	0,094373

Tabela 4.1.3.

K=5,$\alpha = 6$	p=0,5	$x_p = 1,467799$
	n	$e^{-n\triangle +} + e^{-n\triangle -}$
$\epsilon = 0,1:$	10	0,765314
	15	0,516138
	20	0,362218
	50	0,059878
	100	0,00355
$\epsilon = 0,05:$	40	0,842994
	50	0,683893
	100	0,249537
	200	0,038469
	500	0,00022

Tabela 4.1.4.

K=5,$\alpha = 6$	p=0,75	$x_p = 1,647549$
	n	$e^{-n\triangle +} + e^{-n\triangle -}$
$\epsilon = 0,1:$	40	0,802931
	50	0,645005
	100	0,226595
	200	0,033402
	500	0,000167
$\epsilon = 0,05:$	200	0,657895
	300	0,380365
	500	0,129106
	800	0,026438
	1000	0,009371

Dalje je dat analogan rezultat za uopšteniji slučaj, uopštenu Pareto raspodelu.

Teorema 4.2.3. (Stanojević, 2014) *Neka su* $\{Y_n, n \geq 1\}$ *negativno zavisne slučajne veličine sa zajedničkom uopštenom Pareto funkcijom raspodele:*

$$F(x) = L(x)x^{-\alpha}, \ \alpha > 0,$$

gde je $L(x)$ sporo promenljiva funkcija. Važi sledeće ograničenje brzine konvergencije standardne ocene kvantila x_p:

$$\begin{aligned}
&P\{\left|\widehat{x}_p(n)-x_p\right|\geq\epsilon\}\leq e^{-n\Delta+}+e^{-n\Delta-}\\
&=\Big(\frac{p(1-L(x_p+\epsilon)(x_p+\epsilon)^{-\alpha})}{(1-p)L(x_p+\epsilon)(x_p+\epsilon)^{-\alpha}}\Big)^{-pn}\cdot\Big(\frac{1-L(x_p+\epsilon)(x_p+\epsilon)^{-\alpha}}{1-p}\Big)^{n}\\
&+\Big(\frac{p(1-L(x_p-\epsilon)(x_p-\epsilon)^{-\alpha})}{(1-p)L(x_p-\epsilon)(x_p-\epsilon)^{-\alpha}}\Big)^{-pn}\cdot\Big(\frac{1-L(x_p-\epsilon)(x_p-\epsilon)^{-\alpha}}{1-p}\Big)^{n}.
\end{aligned}$$

Dokaz ovog slučaja potpuno je analogan dokazu za Pareto funkciju raspodele, pogledati Stanojević, 2014.

Teorema 4.2.4. (Stanojević, 2014) *Neka su $\{Y_n,n\geq1\}$ negativno zavisne slučajne veličine sa zajedničkom Gama raspodelom:*

$$f(x,\alpha,\beta)=x^{\alpha-1}\frac{\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)},\quad x>0,\alpha>0,\beta>0.$$

Važi sledeće ograničenje brzine konvergencije standardne ocene kvantila x_p:

$$\begin{aligned}
P\Big[\left|\widehat{x}(n)-x_p\right|\geq\epsilon\Big]\ &\leq\ \Big(\frac{pA}{\Gamma(\alpha)(1-\frac{A}{\Gamma(\alpha)})(1-p)}\Big)^{-pn}\cdot\Big(\frac{A}{\Gamma(\alpha)(1-p)}\Big)^{n}\\
&+\ \Big(\frac{pB}{\Gamma(\alpha)(1-\frac{B}{\Gamma(\alpha)})(1-p)}\Big)^{-pn}\cdot\Big(\frac{B}{\Gamma(\alpha)(1-p)}\Big)^{n},
\end{aligned}$$

gde su:

$$A=e^{-\beta(x_p+\epsilon)}\Big([\beta(x_p+\epsilon)]^{\alpha-1}+(\alpha-1)[\beta(x_p+\epsilon)]^{\alpha-2}+o([\beta(x_p+\epsilon)]^{\alpha-2})\Big),$$

$$B=e^{-\beta(x_p-\epsilon)}\Big([\beta(x_p-\epsilon)]^{\alpha-1}+(\alpha-1)[\beta(x_p-\epsilon)]^{\alpha-2}+o([\beta(x_p-\epsilon)]^{\alpha-2})\Big).$$

Dokaz: Potrebno je izračunati:

$$\begin{aligned}
\Delta+ &= \sup_{-\infty<\lambda\leq0}(\lambda p-\ln E\{\exp[\lambda I(Y\leq x_p+\epsilon)]\}),\\
\Delta- &= \sup_{0\leq\lambda<+\infty}(\lambda p-\ln E\{\exp[\lambda I(Y\leq x_p-\epsilon)]\}).
\end{aligned}$$

Ako se označi $p = P[Y \leq x_p + \epsilon]$, dobija se:

$$\begin{aligned} p &= \int_0^{x_p+\epsilon} x^{\alpha-1}\frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}dx \\ &=_{s.y=\beta x} \frac{1}{\Gamma(\alpha)}\int_0^{\beta(x_p+\epsilon)} y^{\alpha-1}e^{-y}dy \\ &= \frac{1}{\Gamma(\alpha)}\Big(\int_0^{+\infty} y^{\alpha-1}e^{-y}dy - \int_{\beta(x_p+\epsilon)}^{+\infty} y^{\alpha-1}e^{-y}dy\Big) \\ &= \frac{1}{\Gamma(\alpha)}\Big(\Gamma(\alpha) - A\Big) = 1 - \frac{A}{\Gamma(\alpha)}, \end{aligned}$$

gde je:

$$\begin{aligned} A = e^{-\beta(x_p+\epsilon)}\Big([\beta(x_p+\epsilon)]^{\alpha-1} &+ (\alpha-1)[\beta(x_p+\epsilon)]^{\alpha-2} \\ &+ o([\beta(x_p+\epsilon)]^{\alpha-2})\Big). \end{aligned}$$

Naime, poslednja jednakost sledi iz rezultata

$$\int_t^{+\infty} s^{r-1}e^{-s}ds = e^{-t}\Big(t^{r-1} + (r-1)t^{r-2} + o(t^{r-2})\Big),$$

koji se može pogledati u Dekkers i de Haan, 1989.
Raspodela indikatora je:

$$I(Y \leq x_p + \epsilon) : \begin{pmatrix} 0 & 1 \\ \text{1-p} & \text{p} \end{pmatrix},$$

gde su $p = 1 - \frac{A}{\Gamma(\alpha)}$ i $1 - p = \frac{A}{\Gamma(\alpha)}$.
Dalje je:

$$\Delta + = \sup_{-\infty<\lambda\leq 0}\Big(\lambda p - \ln\Big(e^\lambda\Big(1 - \frac{A}{\Gamma(\alpha)}\Big) + \frac{A}{\Gamma(\alpha)}\Big)\Big).$$

Posle diferenciranja i primene postupka analize kao u Teoremi 4.2.2, dobija se:

$$\begin{aligned} e^\lambda &= \frac{pA}{\Gamma(\alpha)(1 - \frac{A}{\Gamma(\alpha)})(1-p)}, \\ \lambda &= \ln\frac{pA}{\Gamma(\alpha)(1 - \frac{A}{\Gamma(\alpha)})(1-p)}, \end{aligned}$$

i dalje je:

$$\Delta+ = p\ln\frac{pA}{\Gamma(\alpha)(1-\frac{A}{\Gamma(\alpha)})(1-p)} - \ln\frac{A}{\Gamma(\alpha)(1-p)}.$$

Ostaje da se izračuna $\Delta-$. Posle primene analognog postupka i analize primenjene prilikom izračunavanja $\Delta+$, dobija se:

$$\begin{aligned} e^{\lambda} &= \frac{pB}{\Gamma(\alpha)(1-\frac{B}{\Gamma(\alpha)})(1-p)}, \\ \lambda &= \ln\frac{pB}{\Gamma(\alpha)(1-\frac{B}{\Gamma(\alpha)})(1-p)}, \end{aligned}$$

i važi:

$$\Delta- = p\ln\frac{pB}{\Gamma(\alpha)(1-\frac{B}{\Gamma(\alpha)})(1-p)} - \ln\frac{B}{\Gamma(\alpha)(1-p)},$$

gde je:

$$\begin{aligned} B = e^{-\beta(x_p-\epsilon)}\Big([\beta(x_p-\epsilon)]^{\alpha-1} &+ (\alpha-1)[\beta(x_p-\epsilon)]^{\alpha-2} \\ &+ o([\beta(x_p-\epsilon)]^{\alpha-2})\Big). \end{aligned}$$

Na osnovu dobijenih rezultata, verovatnoća pokrivanja u ovom slučaju je:

$$\begin{aligned} P\Big[\big|\widehat{x}(n) - x_p\big| \geq \epsilon\Big] &\leq e^{-n\Delta+} + e^{-n\Delta-} \qquad (4.4) \\ &= \Big(\frac{pA}{\Gamma(\alpha)(1-\frac{A}{\Gamma(\alpha)})(1-p)}\Big)^{-pn} \cdot \Big(\frac{A}{\Gamma(\alpha)(1-p)}\Big)^{n} \\ &+ \Big(\frac{pB}{\Gamma(\alpha)(1-\frac{B}{\Gamma(\alpha)})(1-p)}\Big)^{-pn} \cdot \Big(\frac{B}{\Gamma(\alpha)(1-p)}\Big)^{n}, \end{aligned}$$

što je i trebalo pokazati. ■

Na osnovu prethodne teoreme, za određenu Gama funkciju raspodele sa poznatim vrednostima parametara, može se izračunati verovatnoća pokrivenosti za standardnu ocenu kvantila x_p. U Tabeli 4.2.1, Tabeli 4.2.2, Tabeli 4.2.3 i Tabeli 4.2.4 su dati rezultati za dva izbora parametara α, β i p, i za svaku kombinaciju parametara su istaknute dve vrednosti ϵ. Takođe, na osnovu dobijenih rezultata, može se videti da pokrivenost raste sa porastom veličine uzorka i da je za finiju ocenu pokrivenosti potreban dosta veći uzorak, što bi i trebalo očekivati.

Tabela 4.2.1.

$\alpha = 1, \beta = 2$	p=0,45	$x_p = 0,298919$
	n	$e^{-n\triangle+} + e^{-n\triangle-}$
$\epsilon = 0,1:$	40	0,731441
	60	0,45301
	80	0,284644
	130	0,093933
	200	0,021738
$\epsilon = 0,05$	150	0,79643
	300	0,323328
	400	0,179091
	500	0,099959
	800	0,018092

Tabela 4.2.2.

$\alpha = 1, \beta = 2$	p=0,95	$x_p = 1,497866$
	n	$e^{-n\triangle+} + e^{-n\triangle-}$
$\epsilon = 0,1:$	1000	0,700845
	1300	0,514664
	1800	0,309401
	2500	0,153501
	3400	0,048615
$\epsilon = 0,05:$	4000	0,706745
	5000	0,549344
	6000	0,428318
	6500	0,257413
	6520	0,25634

Tabela 4.2.3.

$\alpha = 1, \beta = 3$	p=0,45	$x_p = 0,199279$
	n	$e^{-n\triangle+} + e^{-n\triangle-}$
$\epsilon = 0,1:$	20	0,621191
	30	0,370728
	50	0,145009
	80	0,040106
	100	0,017578
$\epsilon = 0,05:$	80	0,657281
	100	0,501443
	140	0,29437
	200	0,135063
	250	0,071707

Tabela 4.2.4.

$\alpha = 1, \beta = 3$	p=0,95	$x_p = 0,998577$
	n	$e^{-n\triangle+} + e^{-n\triangle-}$
$\epsilon = 0,1:$	400	0,774824
	600	0,48703
	900	0,245569
	1300	0,100568
	1700	0,042035
$\epsilon = 0,05:$	1500	0,825545
	2000	0,618025
	3000	0,349103
	4000	0,199183
	4400	0,159559

Takođe, u radu Stanojevicć, 2014 data je analiza realnih podataka i demonstrirano je kako se prethodni rezultati o brzini konvergencije ocene kvantila za Pareto i Gama raspodelu mogu primeniti u praksi.

Naime, analizirani su realni podaci, skup podataka X koji predstavljaju vreme kvara vazdušnog uslovnog sistema jednog aviona, koji u časovima iznose: 33, 47, 55, 56, 104, 176, 182, 220, 239, 246, 320, a dati su u radu Bain i Engelhardt, 1991. X se može modelirati Gama$(1; 152,5)$ raspodelom. U radu Jovanović i Rajić, 2013, je utvrđena valjanost Gama raspodele za date podatke i izračunata je Kolmogorov-Smirnov (KS) vrednost za empirijsku funkciju raspodele i ocenjivanu funkciju raspodele, koja iznosi $0,23$ sa p-vrednošću većom od $0,05$. Jasno je da je Gama funkcija dobra ocena ovog skupa podateka.
U radu Stanojević, 2014, je izračunata vrednost za $\widehat{x}_p(n)$, kada su $p = 0,90$ i $n = 11$ i dobijeno je $\widehat{x}_{0,90}(11) = 246$. Takođe je izračunat kvantil x_p za Gama(1;152,5) raspodelu i verovatnoću $p = 0,90$ i dobijeno je $x_{0.90} = 351,1442$. Koristeći rezultat (4.4), može se zaključiti da postoji $94,95\%$ šanse da kvantil $x_{0.90}$ odstupa od direktno-simulirane ocene $\widehat{x}_{0.90}(11)$ za više od 160.

4.3 Konvergencija u velikom odstupanju

U ovom odeljku biće dati osnovni pojmovi, definicije i teoreme, koji će biti od izuzetne važnosti za razumevanje glavne teme ovog odeljka, konvergencije u velikom odstupanju u Tihonovljevom prostoru. Naime, prvo će biti opisan pojam prostora Hausdorffa. Tačke x i y u topološkom prostoru X se mogu odvojiti okolinama ako postoji okolina U od x i okolina V od y tako da su U i V bez preseka. Prostor X je Hausdorffov prostor ako se bilo koje dve različite tačke iz X mogu razdvojiti okolinama. Dalje, X je kompletno regularan prostor ako za svaki zatvoren skup F i svaku tačku x koja ne pripada F, postoji neprekidna funkcija f iz X na realnu pravu tako da je $f(x) = 0$, i za svako y iz F, $f(y) = 1$. Drugim rečima, može se reći da se x i F mogu razdvojiti neprekidnom funkcijom. Prostor X je Tihonovljev prostor, ako je kompletno regularan i Hausdorffov prostor. Takođe, u ovom odeljku biće dati pojmovi kao što su idempotentna mera i idempotentna integracija, nezavisnost idempotetnih veličina, mera Luzina, E_0-gornje i E_0-donje semineprekidne funkcije, kao i ostali ključni pojmovi za dalje razumevanje.

4.3.1 Idempotentna mera i idempotentna integracija

U ovom delu biće dati pojmovi idempotentne mere, anlogno idempotentnog merljivog prostora i idempotentnog prostora verovatnoće, kao i osnovne teoreme.

Uvodi se notacija koja će važiti do kraja ovog poglavlja. Neka je Ω skup i neka je Ξ kolekcija podskupova skupa Ω koji sadrži $\emptyset$. Neka je $P(\Omega)$ partitivni skup skupa Ω, neka Φ i Ψ označavaju neprazne skupove i J arbitrarni skup indeksa. Sledi osnovna definicija idempotentne mere.

DEFINICIJA 4.3.1. *Funkcija $\mu : P(\Omega) \to \overline{\mathbb{R}}_+$ (gde je $\overline{\mathbb{R}}_+ = [0, +\infty]$), je idempotentna mera na skupu Ω ako važe sledeći uslovi:*

(μ_0) $\mu(\emptyset) = 0$,

(μ_1) $\mu(A \cup B) = \mu(A) \vee \mu(B)$, *gde $\vee$ označava maksimum,*

(μ_2) $\mu(\cup_\phi A_\phi) = \sup_\phi \mu(A_\phi)$ *za rastući skup $\{A_\phi, \phi \in \Phi\}$ podskupova skupa Ω.*
Ako je još i:

(Π) $\mu(\Omega) = 1$,
idempotentna mera se naziva idemotentna verovatnoća, u oznaci Π.
Ako uz prva tri uslova važi i:

(μ_3) $\mu(\cap_\phi F_\phi) = \inf_\phi \mu(F_\phi)$
za svaki opadajući skup $\{F_\phi, \phi \in \Phi\}$ elemenata iz Ξ, tada se kaže da je idempotentna mera τ-glatka u odnosu na Ξ, ili kraće, da je jedna Ξ-idempotentna mera.

Može se primetiti da osobina (μ_1) pokazuje da je μ rastuća i subaditivna funkcija skupa, u smislu da je $\mu(A) \leq \mu(B)$ ako je $A \subset B$ i da je $\mu(A \cup B) \leq \mu(A) + \mu(B)$.

Sledi lema koja je direktna posledica prethodne definicije.

Lema 4.3.1. *Uslovi (μ_1) i (μ_2) su ekvivalentni uslovu*

$$\mu(\bigcup_{j \in J} A_j) = \sup_j \mu(A_j), \tag{4.5}$$

za svaku kolekciju $\{A_j, j \in J\}$ *podskupova skupa* Ω, *što je ekvivalentno reprezentaciji:*

$$\mu(A) = \sup_{\omega \in A} \mu(\{\omega\}),\ A \subset \Omega.$$

Funkcija $\mu(\{\omega\})$ se naziva gustinom mere μ. Uslov (4.5) se naziva τ-maksitivna, i uslov (μ_2) se naziva τ-glatka osobina za rastuće skupove, što ne treba mešati sa pojmom τ-glatkom osobinom, koja se odnosi na opadajuće skupove elemenata Ξ. Sledi definicija koja se odnosi na funkcije skupa koje su definisane na podskupovima skupa Ω.

DEFINICIJA 4.3.2. *Funkcija skupa* $\mu : \Xi \to \overline{\mathbb{R}}_+$ *je maksitivna (respektivno,* τ*-maksitivna) na* Ξ *ako je* $\mu(A \cup B) = \mu(A) \vee \mu(B)$ *za svako* $A \in \Xi$ *i* $B \in \Xi$ *tako da* $A \cup B \in \Xi$ *(respektivno,* $\mu(\cup_{j \in J} A_j) = \sup_{j \in J} \mu(A_j)$ *za svaku kolekciju skupova* $A_j \in \Xi, j \in J$, *tako da* $\cup_{j \in J} A_j \in \Xi$*).*

Za datu kolekciju Ξ, sa Ξ_u (respektivno, Ξ_i) se označava kolekcija neograničenih unija (respektivno, preseka) elemenata Ξ. Takođe se označava $\Xi_{iu} = (\Xi_i)_u$ i $\Xi_{ui} = (\Xi_u)_i$. Kolekcija Ξ_{iu} je zatvorena za neograničene unije i preseke.

DEFINICIJA 4.3.3. *Kolekcija* Ξ *podskupova od* Ω *se naziva* paving *na* Ω *ako sadrži* $\emptyset$ *i zatvorena je za formiranje konačnih unija i preseka. Kolekcija* Ξ *podskupova* Ω *se naziva* π-sistemom *ako je zatvorena za formiranje konačnih preseka.*

Dalje će biti data definicija specijalnije kolekcije podkupova skupa Ω, drugačiji pojam od pavinga i π-sistema, tačna definicija τ-algebre i A-merljivih podskupova, što će biti od važnosti za dalje izlaganje.

DEFINICIJA 4.3.4. *Kolekcija* A *podskupova skupa* Ω *je* τ*-algebra ako sadrži* $\emptyset$ *i ako je zatvorena u odnosu na formiranje komplemenata i konačnih unija. Elementi kolekcije* A *se nazivaju* A*-merljivi podskupovi od* Ω.

Primer τ-algebre jeste partitivni skup $P(\Omega)$ i on se naziva diskretna τ-algebra.

Sledi definicija atoma.

DEFINICIJA 4.3.5. *Kolekcija* Ξ *podskupova od* Ω *se naziva atomskom ako postoji subkolekcija* $\Xi' = \{A_\alpha\}$, *koja sadrži neprazne podskupove od* Ω, *tako da je ispunjeno* $A_\alpha \cap A_{\alpha'} = \emptyset$ *ili* $A_\alpha = A_{\alpha'}$ *za svako* α *i* α', *i* $F \in \Xi$ *akko* $F = \cup A_\alpha$, *gde se unija uzima preko* $A_\alpha \subset F$, $A_\alpha \in \Xi'$. *Elementi kolekcije* Ξ' *se nazivaju atomima od* Ξ.

Iz prethodne definicije sledi naredna teorema koja utvrđuje uslov za τ-algebru.

Teorema 4.3.1. (Puhalskii, 2000) *Kolekcija A, koja sadrži $\emptyset$ i Ω, jeste τ-algebra akko je atomska.*

U daljem izlaganju sledi definicija i osnovne teoreme vezane za pojam merljivih preslikavanja prostora sa idempotentnim merama. Neka su Ω i Ω' skupovi, i neka su Ξ i Ξ' kolekcije podskupova od Ω i Ω' redom, koje sadrže $\emptyset$.

Definicija 4.3.6. *Za funkciju $f : \Omega \to \Omega'$ se definiše kolekcija podskupova od Ω generisana funkcijom f, kao kolekcija $f^{-1}(\Xi') = \{f^{-1}(B), B \subset \Xi'\}$.*

Treba reći da se funkcije definisane na idempotentnim prostorima verovatnoće odnose na idempotetne promenljive.

Definicija 4.3.7. *Funkcija $f : \Omega \to \Omega'$ je Ξ/Ξ'-merljiva ako je $f^{-1}(\Xi') \subset \Xi$.*

Lema 4.3.2. *Funkcija $f : \Omega \to \Omega'$ je Ξ_{iu}/Ξ'_{iu}-merljiva akko je Ξ_{iu}/Ξ'-merljiva.*

Posledica prethodne leme i definicija jeste zaključak, funkcija $f : \Omega \to \Omega'$ je A/A'-merljiva, gde su A i A' τ-algebre na skupovima Ω i Ω' redom, akko inverzne slike atoma A' pripadaju A.

Da bi slika τ-glatke idempotentne mere bila takođe τ-glatka idempotentna mera, potrebno je postaviti neke uslove za preslikavanje. U definiciji koncepta merljivosti koristi se teorema Luzina. To je teorema teorije mere, koja kaže da je realna funkcija Borel merljiva akko je neprekidna na zatvorenim i kompaktnim skupovima. Prvi korak je upoznavanje koncepta čvrste mere.

Definicija 4.3.8. *Neka je μ jedna Ξ-idempotentna mera. Kolekcija Θ podskupova od Ω je čvrsta za μ ako je $T \cap F \in \Xi$ za $T \in \Theta$ i $F \in \Xi$, i za ograničeno $\epsilon > 0$ postoji $T \in \Theta$ tako da je $\mu(T^c) \leq \epsilon$. Takođe se kaže da je μ čvrsta u odnosu na Θ, ili kraće Θ-čvrsta.*

Dalje sledi definicija merljivosti Luzina.

Definicija 4.3.9. *Neka je Θ čvrsta kolekcija za Ξ-idempotentnu meru μ. Funkcija $f : \Omega \to \Omega'$ se naziva Luzin $(\Xi, \Theta)/\Xi'$-merljiva ako je restrikcija funkcije f na ograničenom $T \in \Theta$ upravo Ξ_T/Ξ'-meljiva, gde je $\Xi_T = \{T \cap F, F \in \Xi\}$.*

U sledećoj teoremi se vidi svrha koncepta merljivosti Luzina.

Teorema 4.3.2. (Puhalskii, 2000) *Neka je μ jedna Ξ-idempotentna mera i Θ čvrsta kolekcija za μ. Ako je funkcija $f : \Omega \to \Omega'$ Luzin $(\Xi, \Theta)/\Xi'$-merljiva, tada je slika $\mu' = \mu \circ f^{-1}$, τ-glatka u odnosu na Ξ' idempotentnu meru na Ω'. Ako $f(T) \cap F' \in \Xi'$ kad god je $T \in \Theta$ i $F' \in \Xi'$, tada je μ' $f(\Theta)$-čvrsta.*

Uvodi se sledeća terminologija.

DEFINICIJA 4.3.10. *Neka je (Ω, Π) idempotentni prostor verovatnoće i neka je $f : \Omega \to \Omega'$ jedna Ω' idempotentna promenljiva. Idempotentna verovatnoća $\Pi \circ f^{-1}$ se naziva idempotentna raspodela (ili idempotentni zakon) f nad Π. Ako je Ω' metrički prostor i $\lim\limits_{r \to +\infty} \Pi(f \notin B_r(z)) = 0$, gde je z fiksirani element iz E i $B_r(z)$ označava zatvorenu r-loptu za z, tada se f naziva odgovarajućom idempotentnom promenljivom.*

Dalje slede osnovne teoreme i definicije u vezi idempotentne integracije. Neka je (Ω, μ) idempotentan merljivi prostor tako da je $\mu(\Omega) < +\infty$. Prihvata se konvencija da je $\infty \cdot 0 = 0$.

DEFINICIJA 4.3.11. *Za funkciju f na skupu Ω sa vrednostima u $\overline{\mathbb{R}}_+$ definiše se idempotentna integracija u odnosu na μ na sledeći način:*

$$\bigvee_{\Omega} f d\mu = \sup_{a \in \overline{\mathbb{R}}_+} a\mu(f \geq a).$$

Za $A \subset \Omega$, označava se $\bigvee_A f d\mu = \bigvee_\Omega f \mathbf{1}(A) d\mu$.

Ako je μ idempotentna verovatnoća onda se idempotentni integral naziva idempotentno očekivanje. Takođe se idempotentni integral označava sa $\bigvee_\Omega f(\omega) d\mu(\omega)$ i ako je μ idempotentna verovatnoća, sa Sf, $Sf(\omega)$ ili $S_\Pi f$ se označava idempotentna verovatnoća Π za koju se izračunava idempotentni integral. Iz definicije sledi sledeća lema.

Lema 4.3.3. *Neka je $f : \Omega \to \overline{\mathbb{R}}_+$. Važi sledeća ekvivalentna reprezentacija:*

$$\bigvee_{\Omega} f d\mu = \sup_{a \in \overline{\mathbb{R}}_+} a\mu(f = a) = \sup_{\omega \in \Omega} f(\omega)\mu(\{\omega\}) = \sup_{\omega \in \Omega} f(\omega)\mu([\omega]_{f^{-1}(P(\overline{\mathbb{R}}_+))}).$$

Sledi teorema čiji dokaz direktno sledi iz prethodne leme. Razmatraju se samo integrali funkcije koja uzima vrednosti u $\mathbb{R}_+$ (gde je $\mathbb{R}_+ = [0, +\infty)$), za vrednosti iz skupa $\overline{\mathbb{R}}_+$ rezultat se može analogno formulisati.

Teorema 4.3.3. (Puhalskii, 2000) *Neka su f, g odgovarajuće $\mathbb{R}_+$ funkcije na Ω. Važe sledeće osobine.*

1. $\bigvee_\Omega 0 d\mu = 0$,
2. $\bigvee_\Omega f d\mu \leq \bigvee_\Omega g d\mu$ *ako* $f \leq g$,
3. $\bigvee_\Omega (cf) d\mu = c \bigvee_\Omega f d\mu$, $c \in \mathbb{R}_+$,
4. $\bigvee_\Omega (f \vee g) d\mu = \bigvee_\Omega f d\mu \vee \bigvee_\Omega g d\mu$,
5. $\bigvee_\Omega (f + g) d\mu \leq \bigvee_\Omega f d\mu + \bigvee_\Omega g d\mu$,
6. $|\bigvee_\Omega f d\mu - \bigvee_\Omega g d\mu| \leq \bigvee_\Omega |f - g| d\mu$, *uz uslov da je leva strana nejednakosti dobro definisana,*
7. $\bigvee_\Omega \sup_{j \in J} f_j d\mu = \sup_{j \in J} \bigvee_\Omega f_j d\mu$, *gde* $f_j : \Omega \to \mathbb{R}_+, j \in J$.

Lema 4.3.4. *Ako* $f : \Omega \to \mathbb{R}_+$, *tada:*

$$\mu(f \geq a) \leq \frac{1}{a} \bigvee_\Omega f \mathbf{1}(f \geq a) d\mu, a > 0.$$

Prethodna lema je nejednakost tipa Čebišova.

Teorema 4.3.4. (Puhalskii, 2000) *Neka je μ' idempotentna mera na Ω' tako da je $\mu' = \mu \circ f^{-1}$ za neku funkciju $f : \Omega \to \Omega'$. Tada za funkciju $g : \Omega' \to \mathbb{R}_+$, važi:*

$$\bigvee_{\Omega'} g d\mu' = \bigvee_\Omega g \circ f d\mu.$$

Sledi rezultat analogan Hölder-ovoj nejednakosti, koji je koristan.
Za funkciju $f : \Omega \to \mathbb{R}_+$ i $p > 0$, definiše se $\|f\|_p = (\bigvee_\Omega f^p d\mu)^{1/p}$ i $\|f\|_\infty = \sup_{\omega : \mu(\{\omega\}) > 0} f(\omega)$.

Lema 4.3.5. *Neka su $f, g : \Omega \to \mathbb{R}_+$. Tada:*

1. *Neka* $p \in [1, +\infty]$ *i* $q \in [1, +\infty]$ *tako da je* $1/p + 1/q = 1$, *tada:*

$$\bigvee_\Omega f g d\mu \leq \|f\|_p \|g\|_q.$$

2. *Ako* $\mu(\Omega) = 1$, *tada, za* $0 < p < q$, *važi:* $\|f\|_p \leq \|f\|_q$.

Dalje se uvodi pojam nezavisnosti.

Neka je (Ω, Π) idempotentni prostor verovatnoće.

Definicija 4.3.12. *Konačna kolekcija* $\{A_i, i = 1, ..., k\}$ *podskupova na* Ω *je nezavisna ako:*

$$\Pi(\bigcap_{i=1}^{k} A_i) = \prod_{i=1}^{k} \Pi(A_i).$$

Kolekcija $\{A_\alpha\}$ *podskupova na* Ω *je nezavisna ako je svaka konačna subkolekcija nezavisna.*
Kolekcija $\{\Xi_\alpha\}$ *familije podskupova na* Ω *je nezavisna ako je svaka kolekcija skupova* $\{A_\alpha\}$*, gde je* $A_\alpha \in \Xi_\alpha$*, nezavisna.*

Definicija 4.3.13. *Neka je skup* $\Omega^{'}$ *opremljen* τ*-algebrom* $A^{'}$*. Kolekcija* $\{f_\alpha\}$ *idmpotentnih veličina na* Ω *sa vrednostima u* $\Omega^{'}$ *je* $A^{'}$*-nezavisna (ili nazavisna ako* $A^{'} = P(\Omega^{'})$*) ako je kolekcija* τ*-algebri* $f_\alpha^{-1}(A^{'})$ *nezavisna. Idempotentna promenljiva* $f : \Omega \to \Omega^{'}$ *i kolekcija* Ξ *podskupova na* Ω *su* $A^{'}$*-nezavisne (ili nezavisne ako* $A^{'} = P(\Omega^{'})$*) ako* τ*-algebra* $f^{-1}(A^{'})$ *i* Ξ *obrazuju nezavisnu kolekciju.*

Pojmu čvrsta F-idempotentna verovatnoća u teoriji velikog odstupanja se daje posebno ime, što je od značaja za dalje izlaganje.

Definicija 4.3.14. *Čvrsta* F*-idempotentna verovatnoća se naziva* deviability *(devijantna ili nepravilna). Ako je idempotentna raspodela neke idempotentne veličine deviability, onda je ona deviability raspodela.*

Dalje je dat dovoljan uslov za neprekidnost u odnosu na Π i drugi pojmovi vezani za to. Neka je $E_0 \subset E$.

Definicija 4.3.15. *Kaže se da je skup* $H \subset E$*,* E_0*-zatvoren ako sadrži sve svoje tačke nagomilavanja koje su u* E_0*, tj.* $clH \cap E_0 \subset H$*. Kaže se da je skup* $H \subset E$*,* E_0*-otvoren ako je svaka tačka iz* $H \cap E_0$ *jedna unutrašnja tačka iz* H*, tj.* $H \cap E_0 \subset intH$*.*

Posledica prethodne defenicije kaže da je H E_0-otvoren akko je njen komplement u E E_0-zatvoren.

Definicija 4.3.16. *Funkcija* $h : E \to \mathbb{R}_+$ *se naziva* E_0*-gornje (*E_0*-donje) semineprekida ako je skup* $\{z \in E : f(z) \geq a\}$ *(*$\{z \in E : f(z) \leq a\}$*),* $a \in \mathbb{R}_+$ *upravo* E_0*-zatvoren. Za funkciju* $h : E \to \mathbb{R}_+$ *se kaže da je* E_0*-neprekidna ako je* $h^{-1}(G)$ *upravo* E_0*-otvoren za svaki otvoren skup* $G \subset \mathbb{R}_+$*.*

Kao posledica prethodnih definicija jeste tvrđenje da je indikator funkcija $\mathbf{1}(A)$, $A \subset E$, upravo E_0-gornje (E_0-donje) semineprekidna akko A je E_0-zatvoren (E_0-otvoren).

Ovde se završava izlaganje ključnih pojmova, definicija, teorema i njihovih posledica, koje su od fundamentalnog značaja za razumevanje narednog odeljka.

4.3.2 Konvergencija u velikom odstupanju u Tihonovljevom prostoru

U ovom odeljku biće iznete osnovne definicije i teoreme konvergencije u velikom odstupanju u Tihonovljevom prostoru. Uvodi se sledeća notacija. Neka je E topološki prostor sa Borelovom σ-algebrom $B(\mathrm{E})$. Neka je Φ direktan skup, $\{P_\phi, \phi \in \Phi\}$ skup mera verovatnoća na $(E, B(E))$ i $\{r_\phi, \phi \in \Phi\}$ skup realnih brojeva većih od 1 koji konvergiraju ka $+\infty$ kada $\phi \in \Phi$. Važi da $\mathbb{C}_b^+(E)$, $\overline{\mathbb{C}}_b^+(E)$ i $\underline{\mathbb{C}}_b^+(E)$ označavaju redom skupove $\mathbb{R}_+$-neprekidno ograničene funkcije na E, $\mathbb{R}_+$-ograničene gornje semineprekidne funkcije na E i $\mathbb{R}_+$-ograničene donje semineprekidne funkcije na E. Neka je još Π jedna F-idempotentna verovatnoća na E, gde F označava kolekciju zatvorenih skupova E.

Definicija 4.3.17. *Kaže se da skup $\{P_\phi, \phi \in \Phi\}$ konvergira u velikom odstupanju (LD) brzinom r_ϕ ka Π ako za svako $h \in \mathbb{C}_b^+(E)$:*

$$\lim_{\phi \in \Phi} \Big(\int_E h(z)^{r_\phi} dP_\phi(z) \Big)^{1/r_\phi} = \bigvee_E h(z) d\Pi(z). \tag{4.6}$$

Treba istaći da ako je E Tihonovljev topološki prostor tada je F-idempotentna verovatnoća Π jedinstveno određena desnom stranom jednakosti (4.6). Konvergencija u velikom odstupanju označavaće se sa $P_\phi \to_{r_\phi}^{ld} \Pi$. Kako je skup r_ϕ fiksiran prihvata se jednostavniji zapis $P_\phi \to^{ld} \Pi$. U daljem radu koriste se sledeće oznake: $P_\phi^{1/r_\phi}(A) = (P_\phi(A))^{1/r_\phi}$ i $\|f\|_\phi = \Big(\int_E f(z)^{r_\phi} dP_\phi(z) \Big)^{1/r_\phi}$, gde $f : E \to \mathbb{R}_+$. Sledi Portmanteau teorema za konvergenciju u velikom odstupanju.

Teorema 4.3.5. (Puhalskii, 2000) *Neka je E Tihonovljev topološki prostor. Sledeća tvrđenja su ekvivalentna:*

1. $P_\phi \to^{ld} \Pi$,

2. *(i)* $\liminf_\phi \|g\|_\phi \geq \bigvee_E g d\Pi$, *za sve* $g \in \underline{\mathbb{C}}_b^+(E)$,
   (ii) $\limsup_\phi \|f\|_\phi \leq \bigvee_E f d\Pi$, *za sve* $f \in \overline{\mathbb{C}}_b^+(E)$.

$2'$. *Nejednakost u delu 2 važi za sve donje semineprekidne na* Π *ograničene Borel-merljive funkcije* $g : E \to \mathbb{R}_+$ *i sve gornje semineprekidne na* Π *ograničene Borel-merljive funkcije* $g : E \to \mathbb{R}_+$*, redom.*

3. *(i)* $\liminf_\phi P_\phi^{1/r_\phi}(G) \geq \Pi(G)$, *za sve otvorene skupove* $G \subset E$,
   (ii) $\limsup_\phi P_\phi^{1/r_\phi}(F) \leq \Pi(F)$, *za sve zatvorene skupove* $F \subset E$.

$3'$. *Nejednakost u delu 3 važi za sve otvorene na* Π *Borel-merljive skupove* G *i zatvorene na* Π *Borel-merljive skupove* F*, redom.*

4. $\lim_\phi P_\phi^{1/r_\phi}(H) = \Pi(H)$, *za sve neprekidne na* Π *Borel-merljive skupove* $H \subset E$.

5. $\lim_\phi \|h\|_\phi = \bigvee_E h d\Pi$, *za sve neprekidne na* Π *ograničene Borel-merljive funkcije* $h : E \to \mathbb{R}_+$.

6. $\lim_\phi \|h\|_\phi = \bigvee_E h d\Pi$, *za sve ograničene Borel-merljive funkcije* $h : E \to \mathbb{R}_+$ *koje su uniformno neprekidne na* E.

Može se primetiti da se deo 3 prethodne teoreme može iskoristiti za definiciju pojma ograničene konvergencije u velikom odstupanju, koji je identičan definiciji principa velikog odstupanja, pogledati Varadhan, 1984. Tako da posebno na Tihonovljevom prostoru, konvergencija u velikom odstupanju je ekvivalentna principu velikog odstupanja.

Lema 4.3.6. *Neka je* E *Tihonovljev topološki prostor. Ako* $P_\phi \to^{ld} \Pi$, *tada:*

$$\Pi(z) = \lim_{U \in U_z'} \liminf_{\phi \in \Phi} P_\phi^{1/r_\phi}(U) = \lim_{U \in U_z'} \limsup_{\phi \in \Phi} P_\phi^{1/r_\phi}(clU),$$

gde je U_z' *kolekcija otvorenih okolina* z. *Posebno, ako je* E *metrički prostor, tada:*

$$\Pi(z) = \lim_{r \to 0} \limsup_{\phi \in \Phi} P_\phi^{1/r_\phi}(B_r(z)).$$

Lema 4.3.7. *Neka je E Tihonovljev topološki prostor. Neka je Borelov podskup E_0 iz E opremljen relativnom topologijom. Neka je još $P_\phi(E \setminus E_0) = \Pi(E \setminus E_0) = 0$ i restrikcija Π na E_0, koja je označena sa $\widetilde{\Pi}$, τ-glatka na kolekciji zatvorenih podskupova od E_0. Tada $P_\phi \to^{ld} \Pi$ akko $\widetilde{P}_\phi \to^{ld} \widetilde{\Pi}$, gde $\widetilde{P}_\phi$ označava restrikciju P_ϕ na E_0.*

Dalje će biti potrebna konvergenicija integrala, ne neophodno ograničenih funkcija. To je koncept analogan uniformnoj integrabilnosti.

Definicija 4.3.18. *Za Borel-merljivu funkciju $f : E \to \mathbb{R}_+$ se kaže da je uniformno eksponencijalno integrabilna (reda r_ϕ) u odnosu na skup $\{P_\phi\}$ ako:*

$$\lim_{a\to\infty} \limsup_{\phi\in\Phi} \Big(\int_E f(z)^{r_\phi} \mathbf{1}(f(z) > a) dP_\phi(z) \Big)^{1/r_\phi} = 0.$$

Uniformna eksponencijalna integrabilnost važi ako za neko $\epsilon > 0$:

$$\limsup_{\phi} \Big(\int_E f(z)^{r_\phi(1+\epsilon)} dP_\phi(z) \Big)^{1/r_\phi} < +\infty.$$

Lema 4.3.8. *Neka je E Tihonovljev topološki prostor i neka $P_\phi \to^{ld} \Pi$, kada $\phi \in \Phi$ i Π je na $E_0 \subset E$. Tada važe sledeća tvrđenja:*

1. *Za sve E_0-neprekidne i uniformno eksponencijalno integrabilne u odnosu na $\{P_\phi\}$ Borel-merljive funkcije $h : E \to \mathbb{R}_+$:*

$$\lim_{\phi\in\Phi} \Big(\int_E h(z)^{r_\phi} dP_\phi(z) \Big)^{1/r_\phi} = \bigvee_E h(z) d\Pi(z).$$

2. *Za sve E_0-donje semineprekidne Borel-merljive funkcije $g : E \to \mathbb{R}_+$:*

$$\liminf_{\phi\in\Phi} \Big(\int_E g(z)^{r_\phi} dP_\phi(z) \Big)^{1/r_\phi} \geq \bigvee_E g(z) d\Pi(z).$$

Dokaz prethodne leme, kao i sledeće, može se pogledati u Puhalskii, 2000.

Lema 4.3.9. *Neka je E Tihonovljev topološki prostor i neka $P_\phi \to^{ld} \Pi$. Neka su $h_\phi : E \to \mathbb{R}_+$ uniformno ograničene i Borel-merljive funkcije tako da za funkciju $h : E \to \mathbb{R}_+$:*

$$\lim_{\phi\in\Phi} h_\phi(z_\phi) = h(z)$$

za Π-skoro svako $z \in E$ i svaki skup $z_\phi \to z$ za $\phi \in \Phi$. Tada:

$$\lim_{\phi \in \Phi} \Big(\int_E h_\phi(z)^{r_\phi} dP_\phi(z) \Big)^{1/r_\phi} = \bigvee_E h(z) d\Pi(z).$$

Kao posledica prethodne leme može se dobiti sledeća verzija principa kontrakcije čuvanja LD konvergencije nad preslikavanjem.

Teorema 4.3.6. (Puhalskii, 2000) *Neka je E Hausdorffov topološki prostor i neka je E' Tihonovljev topološki prostor. Neka je Π deviability na E. Neka su Borel-merljive funkcije $f_\phi : E \to E'$, $\phi \in \Phi$ i Π-Luzin-merljiva funkcija $f : E \to E'$ je takva da za Π-skoro svako $z \in E$ i svaki skup $z_\phi \to z$ važi da $f_\phi(z_\phi) \to f(z)$. Ako $P_\phi \to^{ld} \Pi$, tada $P_\phi \circ f_\phi^{-1} \to^{ld} \Pi \circ f^{-1}$.*

Dalje sledi često korišćena lema.

Lema 4.3.10. *Neka je E Hausdorffov topološki prostor, E' je Tihonovljev topološki prostor i Π je deviability na E. Ako $P_\phi \to^{ld} \Pi$ za $\phi \in \Phi$ i $f : E \to E'$ je Borel-merljiva i neprekidna Π-skoro svuda, tada $P_\phi \circ f^{-1} \to^{ld} \Pi \circ f^{-1}$.*

Definicija 4.3.19. F-*idempotentna verovatnoća Π na E se naziva akomulativna tačka velikog odstupanja na $\{P_\phi, \phi \in \Phi\}$ (za brzinu r_ϕ) ako postoji subpodskup $\{P_{\phi'}, \phi' \in \Phi'\}$ na $\{P_\phi, \phi \in \Phi\}$ koji LD konvergira (brzinom $r_{\phi'}$) ka Π.*

Dalje, neka $\mathbb{K}$ označava kolekciju kompaktnih podskupova na E.

Definicija 4.3.20. *Skup $\{P_\phi, \phi \in \Phi\}$ se naziva eksponencijalno čvrstim (reda r_ϕ) ako $\inf_{K \in \mathbb{K}} \lim_{\phi \in \Phi} \sup P_\phi^{1/r_\phi}(K^c) = 0$.*

Teorema 4.3.7. (Puhalskii, 2000) *Neka je E Tihonovljev topološki prostor.*

1. *Ako je skup $\{P_\phi, \phi \in \Phi\}$ eksponencijalno čvrst, to je LD relativno kompaktna akomulativna tačka deviability.*

2. *Neka je E još i lokalno kompaktan Hausdorffov topološki prostor. Ako je skup $\{P_\phi, \phi \in \Phi\}$ LD relativno kompaktan, tada je eksponencijalno čvrst.*

Prethodna teorema dozvoljava uvođenje sledećeg korisnog koncepta.

DEFINICIJA 4.3.21. *Neka je E Tihonovljev topološki prostor i $E_0 \subset E$. Kaže se da je skup $\{P_\phi, \phi \in \Phi\}$ E_0-eksponencijalno čvrst ako je eksponencijalno čvrst i ako je svaka LD akomulativna tačka Π sadržana u E_0.*

Sledi lema koja predstavlja verziju principa kontrakcije.

Lema 4.3.11. *Neka je E Tihonovljev topološki prostor, $E_0 \subset E$ i E' je lokalno kompaktan Hausdorffov topološki prostor. Ako je skup $\{P_\phi, \phi \in \Phi\}$ E_0-eksponencijalno čvrst i funkcija $f : E \to E'$ je Borel-merljiva i E_0-neprekidna, tada je skup $\{P_\phi \circ f^{-1}, \phi \in \Phi\}$ eksponencijalno čvrst.*

Dokaz prethodne leme može se pogledati u Puhalskii, 2000.

Dalje se pretpostavlja da je E metrički prostor i uvodi se 'metrika' za konvergenciju u velikom odstupanju. Prvo se definiše Prohorova metrika, a zatim metrika analogna njoj. Za $A \subset E$ se uvodi oznaka $A^\epsilon = \{z \in E : \rho(z, A) \leq \epsilon\}$. Lévy-Prohorova metrika definiše rastojanje između dve verovatnosne mere μ i ν:

$$\begin{aligned} p(\mu, \nu) = \inf\{\epsilon > 0 : \mu(A) &\leq \nu(A^\epsilon) + \epsilon, \\ \nu(A) &\leq \mu(A^\epsilon) + \epsilon, \text{ za sve } A \subset E\}. \end{aligned}$$

Dalje se pretpostavlja da je dat skup $\{P_\phi, \phi \in \Phi\}$ Borel mera na E, skup realnih brojeva $\{r_\phi, \phi \in \Phi\}$ većih od 1 koji konvergiraju ka $+\infty$ i F-idempotentna verovatnoća Π na E. Tada se definiše metrika analogna Prohorovoj metrici:

$$\begin{aligned} p^{ld}_\phi(P_\phi, \Pi) = \inf\{\epsilon > 0 : P_\phi^{1/r_\phi}(F) &\leq \Pi(F^\epsilon) + \epsilon, \\ \Pi(F) &\leq P_\phi^{1/r_\phi}(F^\epsilon) + \epsilon, \text{ za sve zatvorene } F \subset E\}. \end{aligned}$$

Sledeća lema sledi iz regularnosti Borelovih mera i τ-maksitivnosti idempotentnih mera.

Lema 4.3.12. *Može se ekvivalentno napisati:*

$$\begin{aligned} p^{ld}_\phi(P_\phi, \Pi) = \inf\{\epsilon > 0 : P_\phi^{1/r_\phi}(A) &\leq \Pi(A^\epsilon) + \epsilon, \\ \Pi(A) &\leq P_\phi^{1/r_\phi}(A^\epsilon) + \epsilon, \textit{ za sve } A \in B(E)\}. \end{aligned}$$

Teorema 4.3.8. (Puhalskii, 2000) *Skup $\{P_\phi\}$ LD konvergira ka Π akko $p^{ld}_\phi(P_\phi, \Pi) \to 0$, za $\phi \in \Phi$.*

Dalje se definiše metrika analogna Kantorovich-Wasserstein metrici, na sledeći način:

$$\rho^{ld}_{\phi_{BL}}(P_\phi, \Pi) = \sup_{f \in \mathbb{C}_b^+(E): \|f\|_{BL} \leq 1} \Big| \Big(\int_E f(z)^{r_\phi} dP_\phi(z) \Big)^{1/r_\phi} - \bigvee_E f(z) d\Pi(z) \Big|,$$

gde je:

$$\|f\|_{BL} = (\sup_{z \in E} f(z)) \vee \Big(\sup_{z \neq z'} \frac{|f(z) - f(z')|}{\rho(z, z')} \Big).$$

Teorema 4.3.9. (Puhalskii, 2000) *Skup $\{P_\phi\}$ LD konvergira ka Π akko $\rho^{ld}_{\phi_{BL}}(P_\phi, \Pi) \to 0$, za $\phi \in \Phi$.*

Dalje se razmatra sekvencijalna kompaktnost za metričke prostore. Pretpostavlja se da je $\Phi = N$ i skup $\{r_\phi\}$ se zamenjuje nizom $\{r_n\}$. Mere verovatnoća se označavaju sa P_n.

Definicija 4.3.22. *Niz $\{P_n, n \in N\}$ je LD relativno sekvencijalno kompaktan (sa brzinom r_n) ako svaki subniz $\{P_{n'}\}$ niza $\{P_n\}$ sadrži dalji subniz $\{P_{n''}\}$ koji LD konvergira (brzinom $r_{n''}$) ka F-idempotentnoj verovatnoći na E.*

Teorema 4.3.10. (Puhalskii, 2000)

1. *Neka je E metrički prostor. Ako je niz verovatnoća $\{P_n, n \in N\}$ na $(E, \mathrm{B}(E))$ eksponencijalno čvrst, tada je LD relativno sekvencijalno kompaktan, LD akomulativn tačke su deviability.*

2. *Neka je E homeomorfno na kompletnom separabilnom metričkom prostoru. Ako je niz $\{P_n, n \in N\}$ LD relativno sekvencijalno kompaktan, onda je eksponencijalno čvrst.*

Dalje, neka je $L(X)$ raspodela slučajne veličine X i neka E_ϕ označava očekivanje u odnosu na meru verovatnoća P_ϕ.

Teorema 4.3.11. (Puhalskii, 2000) *Neka je $\{X^\phi, \phi \in \Phi\}$ skup $\mathbb{R}^k$ slučajnih veličina definisanih redom na prostorima verovatnoće $(\Omega_\phi, \mathrm{F}_\phi, P_\phi)$ tako da je za svako $\lambda \in \mathbb{R}^k$:*

$$\lim_{\phi \in \Phi} \frac{1}{r_\phi} \ln E_\phi e^{r_\phi \lambda \cdot X^\phi} = G(\lambda),$$

gde je $G(\lambda)$ *odgovarajuća* $\overline{\mathbb{R}}$ *niže semineprekidna i esencijalno glatka konveksna funkcija tako da je* $0 \in int(domG)$. *Tada* $\mathrm{L}(X^{\phi}) \to^{ld} \Pi$ *bzrinom* r_{ϕ}, *gde je* Π *deviability tako da je* $\Pi(x) = e^{-\sup_{\lambda \in \mathbb{R}^k}(\lambda \cdot x - G(\lambda))}$.

Teorema 4.3.12. (Puhalskii, 2000) *Neka je* E *Tihonovljev prostor. Neka je* Υ *podskup od* $\mathbb{C}_b^+(E)$ *koji sadrži uniformno ograničene i ekvineprekidne funkcije , tj.*
$\sup_{f \in \Upsilon} \sup_{z \in E} f(z) < +\infty$ *i za svako* $\epsilon > 0$ *i* $z \in E$ *postoji otvorena okolina* U_z *od* z *tako da je* $\sup_{f \in \Upsilon} \sup_{z' \in U_z} |f(z) - f(z')| \leq \epsilon$. *Ako* $P_{\phi} \to^{ld} \Pi$, *tada:*

$$\lim_{\phi} \sup_{f \in \Upsilon} \Big| \Big(\int_E f^{r_{\phi}} dP_{\phi} \Big)^{1/r_{\phi}} - \bigvee_E f d\Pi \Big| = 0.$$

Definicija 4.3.23. *Kaže se da skup* $\{P_{\phi}, \phi \in \Phi\}$ *verovatnoća na* $(E, \mathrm{B}(E))$ *neodređeno LD konvergira brzinom* r_{ϕ} *ka* K-*idempotentnoj verovatnoći* Π *na* E *za svako* $f \in \mathbb{C}_{\mathrm{K}}^+(E)$, *ako:*

$$\lim_{\phi \in \Phi} \Big(\int_E f(z)^{r_{\phi}} dP_{\phi}(z) \Big)^{1/r_{\phi}} = \bigvee_E f(z) d\Pi(z).$$

Ako je E lokalno kompaktan i Hausdorffov prostor, tada su osobine neodređene LD konvergencije slične osobinama slabe LD konvergencije.

Teorema 4.3.13. (Puhalskii, 2000) *Neka je* E *lokalno kompaktan Hausdorffov topološki prostor. Tada je skup verovatnoća* $\{P_{\phi}, \phi \in \Phi\}$ *na* $(E, \mathrm{B}(E))$ *neodređeno LD relativno kompaktan.*

Intuitivno se može formulisati LD konvergencija mera verovatnoća kao LD konvergencija u raspodeli sa odgovarajućim slučajnim veličinama.

Definicija 4.3.24. *Neka je* $\{X_{\phi}, \phi \in \Phi\}$ *skup slučajnih veličina definisanih na odgovarajućim prostorima verovatnoća* $(\Omega_{\phi}, \mathrm{F}_{\phi}, P_{\phi})$ *i vrednostima na topološkom prostoru* E *i* X *je idempotentna veličina definisana na idempotentnom prostoru verovatnoće* (Ω, Π) *sa vrednostima na* E, *čija je idempotentna raspodela* τ*-glatka sa vrednostima na kolekciji zatvorenih podskupova od* E. *Kaže se da skup* $\{X_{\phi}, \phi \in \Phi\}$ *LD konvergira u raspodeli ka* X *ako* $P_{\phi} \circ X_{\phi}^{-1} \to^{ld} \Pi \circ X^{-1}$.

Do sada je $\to^{ld}$ označavalo LD konvergenciju u raspodeli. Iz konteksta bi trebalo da bude jasno da li se radi o LD konvergenciji mera verovatnoća ili LD konvergenciji u raspodeli slučajnih veličina. Ovde će dalje biti data verzija Leme 4.3.8.

Kaže se da je skup $\{\xi_\phi, \phi \in \Phi\}$ $\mathbb{R}_+$ slučajnih veličina redom na prostorima verovatnoća $(\Omega_\phi, F_\phi, P_\phi)$ uniformno eksponencijalno integrabilan u odnosu na $\{P_\phi, \phi \in \Phi\}$ (brzinom r_ϕ) ako:

$$\lim_{A\to\infty} \lim_{\phi\in\Phi} \sup(E_\phi \xi_\phi^{r_\phi} \mathbf{1}(\xi_\phi > A))^{1/r_\phi} = 0.$$

Lema 4.3.13. *Neka $\xi_\phi \to^{ld} \xi$, gde je ξ jedna $\mathbb{R}_+$ idempotentna veličina na idempotentnom prostoru verovatnoće (Ω, Π). Ako je skup $\{\xi_\phi, \phi \in \Phi\}$ uniformno eksponencijalno integrabilan u odnosu na $\{P_\phi, \phi \in \Phi\}$, tada*

$$\lim_{\phi\in\Phi} (E_\phi \xi_\phi^{r_\phi})^{1/r_\phi} = S\xi.$$

Dalje slede tri leme, tehničke prirode. Neka su X i Y dve slučajne veličine sa vrednostima, redom u separabilnim metričkim prostorima E i E', sa Borelovim σ-algebrama. Tada je (X, Y) slučajna veličina na $E \times E'$ sa proizvodom topologija i Borelovom σ-algebrom. Posebno, ako je $E = E'$ i ρ je metrika na E, tada je $\rho(X, Y)$ slučajna veličina.

Lema 4.3.14. *Neka je E separabilni metrički prostor sa metrikom ρ, i neka su X^ϕ i Y^ϕ, gde je $\phi \in \Phi$, skupovi slučajnih veličina definisani redom na prostorima verovatnoća $(\Omega_\phi, \mathrm{F}_\phi, P_\phi)$, sa vrednostima u E. Ako $L(X^\phi) \to^{ld} \Pi$, gde je Π jedna F-idempotentna verovatnoća na E, i $\rho(X^\phi, Y^\phi) \to^{P_\phi^{1/r_\phi}} 0$, kada $\phi \in \Phi$, tada $L(Y^\phi) \to^{ld} \Pi$.*

Dalje je data još jedna primena metrike.

Definicija 4.3.25. *Kaže se da skup $\{X^\phi, \phi \in \Phi\}$ slučajnih veličina na $(\Omega_\phi, \mathrm{F}_\phi, P_\phi)$ sa vrednostima u metričkom prostoru E sa metrikom ρ konvergira ka $z \in E$ super-eksponencijalno u verovatnoći, brzinom r_ϕ (ili super-eksponencijalno u verovatnoći ako se podrazumeva brzina) i piše se $X^\phi \to^{P_\phi^{1/r_\phi}} z$ ako $\lim_{\phi\in\Phi} P_\phi^{1/r_\phi}(\rho(X^\phi, z) > \epsilon) = 0$ za svako $\epsilon > 0$.*

Treba istaći da važi: $X^\phi \to^{P_\phi^{1/r_\phi}} z$ akko

$$\lim_{\phi\in\Phi} \Big(\int_{\Omega_\phi} (1 \wedge \rho(X_\phi, z))^{r_\phi} dP_\phi \Big)^{1/r_\phi} = 0.$$

Lema 4.3.15. *Neka je* $\{X^\phi, \phi \in \Phi\}$ *skup slučajnih veličina na redom prostorima verovatnoća* $(\Omega_\phi, \mathrm{F}_\phi, P_\phi)$ *sa vrednostima u metričkom prostoru* E*, sa metrikom* ρ*. Tada* $X^\phi \to^{P_\phi^{1/r_\phi}} z$ *akko* $L(X^\phi) \to^{ld} \boldsymbol{1}_z$*, gde* $\boldsymbol{1}_z$ *označava jediničnu masu u* z*.*

U sledećoj lemi se formuliše rezultat jezikom LD konvergencije u raspodeli i razmatra se združena LD konvergencija.

Lema 4.3.16. *Neka su* E *i* E' *separabilni metrički prostori i neka su* X^ϕ *i* Y^ϕ*, gde je* $\phi \in \Phi$*, skupovi slučajnih veličina na* $(\Omega_\phi, \mathrm{F}_\phi, P_\phi)$*, sa vrednostima u* E *i* E'*, redom. Neka su* X *i* Y *idempotentne veličine na idempotentnom prostoru verovatnoće* (Ω, Π) *sa vrednostima u* E *i* E' *redom, čije su idempotentne raspodele* τ*-glatke u odnosu na pridružene kolekcije zatvorenih skupova. Neka je* $E \times E'$ *opremljen proizvodom topologija.*

1. *Ako* $X^\phi \to^{ld} X$*,* $Y^\phi \to^{ld} Y$*,* X *i* Y *su nezavisne, i* X^ϕ *i* Y^ϕ *su nezavisne, tada* $(X^\phi, Y^\phi) \to^{ld} (X, Y)$*.*

2. *Ako* $X^\phi \to^{ld} X$ *i* $Y^\phi \to^{P_\phi^{1/r_\phi}} z$*, tada* $(X^\phi, Y^\phi) \to^{ld} (X, z)$*.*

Dokaz prethodne dve leme može se pogledati u Puhalskii, 2000.

Treba reći da poslednji Odeljak 4.3. predstavlja uopštenje principa velikog odstupanja, koji je razmatran u Odeljku 4.1., posebno za Tihonovljev prostor. U disertaciji je podvučena paralela između odgovarajućih teorema u ovim odeljcima, takođe i za jednodimenzioni slučaj i za slučaj $\mathbb{R}^k$. Data je i ekstenzija teorije za slučajnu veličinu (X, Y), na proizvodu odgovarajućih topologija. U tom smislu, moguće je dati ekstenziju rezultata iz Odeljka 4.2., brzine konvergencije ocene kvantila, posebno na Tihonovljevom prostoru, i dobiti analogne rezultate. To je mogući pravac daljeg istraživanja.

Zaključak

Jedan od osnovnih ciljeva ove disertacije jeste ispitivanje brzine konvergencije direktno-simulirane ocene kvantila, posebno u slučajevima Pareto, uopštene Pareto i Gama raspodele. Kako su to funkcije raspodela često korišćene u praksi, iz razloga što dobro modeliraju mnoge realne događaje, jasna je jedna komponenta ove disertacije, a to je praktična primena dobijenih rezultata. Drugi cilj ove disertacije jeste da da jasan prikaz prvo, intervala poverenja za sredinu, disperziju i proporciju, kako za problem jednog tako i za problem dva uzorka, tako i prikaz ocena i intervala poverenja za indeks ekstremne vrednosti i visokih kvantila. Važnost ispitivanja ponašanja funkcije raspodele na krajevima, samim tim važnost određivanja što preciznije ocene i boljeg intervala poverenja za indeks ekstremne vrednosti i kvantila, leži u činjenici da je problem o prognoziranju teško predvidivih i nestabilnih događaja od izuzetne važnosti u mnogim oblastima primene, kao što su osiguranje, finansije, hidrologija i druge. Događaji iz skorašnje prošlosti, krah berzi i različiti slučajevi finansijskih turbulencija, za koje se pokazalo da se dešavaju mnogo frekfentnije nego što se očekivalo, ukazuju na važnost i aktuelnost teme ove disertacije. Moguća uopštenja i pravci u kojima bi moglo da krene dalje istraživanje, takođe su istaknuti u samoj disertaciji.

Literatura

[1] Abramovitch, L. and Singh, K.(1985). Edgeworth correctid Pivotal Statistics and the Bootstrap. *The Annals of Statistics 13*, 116–132.

[2] Abu-Shawiesh, M.O.A., Banik, S. and Golan Kibria, B.M.(2011). An simulation study on some confidence intervals for the population standard deviation. *SORT 35(2)*, 83–102.

[3] Agresti, A. and Caffo, B.(2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *Amer. Statist.54*, 280–288.

[4] Bain, L.J. and Engelhardt, M.(1991). *Statistical Analysis of Reliability and Life-Testing Models, 2nd. Edition.* Marcel and Dekker, New York.

[5] Barhan, A.M. and Jeyaratnam, S.(1999). Robust confidence interval for the variance. *J. Stat. Comput. Simul. 62(3)*, 189–205.

[6] Barndorff-Nielsen, O.E. and Cox, D.R.(1989). *Asymtotic Techniques for Use in Statistics.* New York: Chapman and Hall.

[7] Barrett, J. and Goldsmith, L.(1976). When is n sufficiently large? *American Statisticion 30*, 67–70.

[8] Beirlant, J. and Teugels, J.L.(1986). Asymtotics of Hills estimator. *Theory Probab. Appl. 31*, 463–469.

[9] Beirlant, J., Teugels, J.L. and Vyckier, P.(1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc. 91*, 1659–1667.

[10] Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G.(1999). Tail index estimation and an exponential regression model. *Extremes 2*, 177–200.

[11] Beran, R.J.(1987). Prepivoting to reduce level error of confidence sets. *Biometrika 74*, 457–468.

[12] Bhattacharya, R.N. and Ghosh, J.K.(1978). On the Validity of the Formal Edgeworth Exspansion. *The Annals of Statistics 6*, 434–451.

[13] Bickel, P.J.(1992). *Theoretical Comparison of Different Bootstrep-t Confidence Bounds.* Wiley, New York, 65–76.

[14] Boos, D. and Hughes-Oliver, J.(2000). How large does n have to be for z and t intervals? *American Statistician 54*, 121–128.

[15] Brown, L.D., Cai, T.T. and DasGupta, A.(2001). Interval estimation for a binomial proportion. *Statist. Sci. 16*, 101–133.

[16] Bucklev, J.A.(1990). *Large Deviations Techniques in Decision, Simulation and Estimation*, Wiley.

[17] Caeiro, F. and Gomes, M.I.(2006). A new class of estimators of a scale second order parameter. *Extremes 9*, 193–211.

[18] Caeiro, F., Gomes, M.I. and Pestana, D.(2005). Direct reduction of bias of the classical Hill estimator. *Revstat 3(2)*, 113–136.

[19] Caeiro, F. and Gomes, M.I.(2007). Semi-parametric second-order reduced-bias high quantile estimation. *Test 18*, 392–413.

[20] Caeiro, F. and Gomes, M.I.(2008). Minimum-variance reduced-bias tail index and high quantile estimation. *REVSTAT- Statistical Journal, Volume 6, Number 1*, 1–20.

[21] Caers, J., Beirlant, J. and Vynckier, P.(1998). Bootstrap confidence intervals for tail indices. *Comput. Stat. and Data Analysis 26*, 259–277.

[22] Castillo, E. and Hadi, A.S.(1997). Fitting the generalised Pareto distribution to data. *J. Amer. Statist. Assoc. 92*, 1604–1620.

[23] Chen, L.(1995). Testing the mean of skewed distributions. *Journal of the American Statistical Association 90*, 767–772.

[24] Cheng, S. and Pan, J.(1998). Asymtotic expansions of estimators for the tail index with applications. *Scand. J. Statist., 25*, 717–728.

[25] Cheng, S. and Peng, L.(2001). Confidence intervals for the tail index. *Bernouli 7*, 751–760.

[26] Coles, S. and Powell, E.(1996). Bayesian methods in extreme value modelling: a review and new development. *Internat. Statist. Rev. 64*, 119–136.

[27] Csörgo, S., Deheuvels, P. and Mason, D.(1985). Kernel estimates of the tail index of a distribution. *Ann. Statist. 13*, 1050–1077.

[28] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G.(2001). Using a bootsrap methods to chose the sample fraction in tail index estimation. *J. Multivariate Anal. 76*, 226–248.

[29] Davydov, Yu., Paulauskas, V. and Račkauskas, A.(2000). More on *P*-stable convex sets in Banach spaces. *J. Theoret. Probab. 13(1)*, 39–64.

[30] Davydov, Yu. and Paulauskas, V.(1999). On the estimation of the parameters of multivariate stable distributions. *Acta. Appl. Math. 58*, 107–124.

[31] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L.(1975). A moment estimator for the index of an extreme value distribution. *Ann. Statist. 17*, 1833–1855.

[32] Dekkers, A.L.M. and de Haan, L.(1989). On the estimation of the extreme-value index and large quantile estimation. *The Annals of Statistics 17(4)*, 1795–1832.

[33] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L.(1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics 17(4)*, 1833–1855.

[34] Dekkers, A.L.M. and de Haan, L.(1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal. 47(2)*, 172–173.

[35] Dembo, A. and Zeitouni, O.(1998). *Large Deviations Techniques and Applications*, 2nd edition, Springer-Verlang.

[36] Deuschel, J.D. and Strook, D.(1989). Large Deviations, *Academic Press.*

[37] Draisman, G., de Haan, L. and Peng, L.(1999). A Bootstrap-based method to achive optimality in estimating the extreme-value index. *Extremes 2(4)*, 367–404.

[38] Drees, H. and Kaufmann, E.(1998). Selecting the optimal sample fraction in univariate extreme value estimation. Stohastic Process. Appl. 75(2), 149–172.

[39] Drees, H., de Haan, L. and Resnick, S.(2000). How to make a Hill plot. *Ann. Statist. 28*, 254–274.

[40] Efron, B. and Tibshirani, R.J.(1993). *An Introduction to the Bootstrap.* New York: Chapman & Hall.

[41] Ferreira, A., de Haan, L. and Peng, L.(2003). On optimising the estimation of high quantiles of probability distribution. *Statistics 37(5)*, 401–434.

[42] Feldman, D. and Tucker, H.G.(1966). Estimation of non-unique quantiles. *Annals of Mathematical Statistics 37*, 451–457.

[43] Ferreira, D.F. and de Vries, C.(2004). Optimal confidence intervals of the tail index and high quantiles. *Tinbergen institute Discussion Paper.*

[44] Ferreira, D.F.(2011). A normal approximation to the F distribution. *Rev. Bras. Biom. Sao Paolo 29(2)*, 222–228.

[45] Feuerverger, A. and Hall, P.(1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist. 27*, 760–781.

[46] Fraga Alves, M.I., Gomes, M.I. and de Haan, L.(2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica 60(2)*, 193–213.

[47] Gayen, A.(1949). The distribution of Student t in random samples of any size drawn from non-normal universis. *Biometrica 36*, 353–369.

[48] Geluk, J.L. and de Haan,L.(1980). *Regular Variation, Exstension and Tauberion Theorems.* CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

[49] Geluk, J.L. and de Haan, L.(1987). *Regular Variation, Exstensions and Tauberian Theorems.* Math. Centre Tracts 40. Centre for Mathematics and Computer Science, Amserdam.

[50] Geluk, J.L. and Peng, L.(2000). An adaptive optimal estimate of the tail index for MA(1) time series. *Statist. Probab. Lett. 46*, 217–227.

[51] Gnedenko, B.V.(1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math., 44(6)*, 423–453.

[52] Gomes, M.I., Martines, M.J. and Neves, M.(2000). Alternatives to a semiparametric estimator of parameters of rare events - Jackknife methodology. *Extremes 3(3)*, 207–229.

[53] Gomes, M.I., de Haan, L. and Peng, L.(2002). Semi-parametric estimation of the second order parameter- asymtotic and finite sample behaviour. *Extremes 5(4)*, 387–414.

[54] Gomes, M.I. and Martins, M.J.(2002). Asymtotically unbiased estimators of the tail index based on external estimation of the second order parameter. *Extremes 5(1)*, 5–31.

[55] Gomes, M.I. and Oliveira, O.(2003). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events. *Comm. in Statist. Simulation and Computation 32(4)*, 1005–1028.

[56] Gomes, M.I., Caeiro, F. and Figueiredo, F.(2004a). Bias reduction of a extreme value index estimator trough an external estimation of the second order parameter. *Statistics 38(6)*, 497–510.

[57] Gomes, M.I., de Haan,L. and Henriques Rodrigues, L.(2004b). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. Royal Statistical Society B*, DOI:10.1111/j.1467-9869.2007.00620.x, 2007.

[58] Gomes, M.I. and Figueiredo, F.(2006). Bias reduction in risk modelling: semi-parametric quantile estimation. *Test 15(2)*, 375–396.

[59] Gomes, M.I., Martins, M.J. and Neves, M.(2007a). Improving second order reduced bias extreme value index estimation. *Revstat 5(2)*, 177–207.

[60] Gomes, M.I. and Pestana, D.(2007b). A sturdy reduced bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.*, 102, 477, 280–292.

[61] Grossman, M. and Norton, H.W.(1980). Approximate intrinsic bias in estimates of heritability based on variance component analysis. *J. Hered 71*, 295–297.

[62] de Haan, L.(1970). On Regular Variation and its Application to the Weak Convergence of Sample Extremes. *Mathematical Centre Tract 32*, Amsterdam.

[63] de Haan, L. and Resnick, S.(1980). Method to choose the sample fraction in tail index estimation. *J. Roy. Statist. Soc. Ser.B 42*, 83–87.

[64] de Hann, L. and Rootzén, H.(1983). On the estimation of high quantiles. *J. Statist. Planning and Inference 35*, 1–13.

[65] de Haan, L. and Stadtmüller, U.(1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A, 61*, 381–395.

[66] de Haan, L. and Peng, L.(1998). Comparision of tail index estimators. *Statistica Neerlandica 52*, 60–70.

[67] Hall, P.(1982). Improving the normal approximation when constructing one-side confidence intervals for binomial or Poisson parameters. *Biometrica 69*, 647–652.

[68] Hall, P.(1982). On some simple estimates of an exponent of regular veriation. *J. R. Statist. Soc. 44(1)*, 37–42.

[69] Hall, P. and Welsh, A.H.(1985). Adaptive estimates of parameters of regular variation. *Ann. Statist. 13*, 331–341.

[70] Hall, P.(1988) Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist. 16*, 927–985.

[71] Hall, P.(1990). Using the bootstrap to estimate means squared error and select soothing parameter in nonparametric problems. *J. Multivariate Anal. 32*, 177–203.

[72] Hall, P.(1992a). *The Bootstrap and Edgeworth Expansion.* New York: Springer.

[73] Hall, P.(1992b). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B 54*, 221–228.

[74] Hill, B.M.(1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist. 3*, 1163–1174.

[75] Hoaglin, D.(1985). Summarizing shape numerically: the g-and-h distributions. In Hoaglin, D. et al. (eds), *Exploring Data, Tables, Trends, and Shapes*, 461–511. New York: John Wiley & Sons.

[76] Hosking, J.R.M. and Wallis, J.R.(1987). parameter and quantile estimation for the generalized Pareto distribution. *Technometrics 29*, 339–349.

[77] Hsing, T.(1991). On tail index estimation using dependent data. *Ann. Statist. Vol. 19*, 1547- 1569

[78] Jonhson, N., Kotz, S. and Balakrishnan, N.(1995). *Distributions in Statistics: Continuous Univariate Distributions.* John Wiley, New York.

[79] Jonhson, N.(1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association 73*, 536–547.

[80] Jovanović, M. and Rajić, V.(2013). *Estimation of* $P[X < Y]$ *for Gamma exponential model.* Jugoslav Journal of Operations Research 23.

[81] Kittani, H. and Zghoul, A.(2010). A robust confidence for the population standard deviation. *J. Appl. Stat. Sci. 18(2)*, 281–290.

[82] Lehmann, E.L.(1966). Some concepts of dependence. *Annals of Mathematical Statistics 37*, 1137–1153.

[83] Liu, R.Y. and Singh, K.(1987). On a partial correction by the bootstrap. *Ann. Statist. 15*, 1713–1718.

[84] Matthys, G. and Beirlant, J.(2000a). Adaptive threshold selection in tail index estimation. In Extremes and Integrated Risk Management (Edited by P.Embrechts), *Risk Books, London*, 37–49.

[85] Matthys, G. and Beirlant, J.(2000b). Extreme quantile estimation for heavy-tailed distributions. *Preprint, Center for Statistics, University of Leuven.*

[86] Matthys, G. and Beirlant, J.(2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica 13*, 853–880.

[87] Matthys, G., Delafosse, M., Guillou, A. and Beirlant, J.(2004). Estimating catastrophic quantile levels for heavy-tailed distributions. *Insurance: Mathematics and Economics 34*, 517–537.

[88] Mladenović, P.(2002). *Ekstremne vrednosti slučajnih nizova.* Matematički fakultet, Beograd, ISBN 86-7589-027-3.

[89] Mladenović, P. and Piterbarg, V.(2008). On estimation of the exponent of regular variation using a sample with missing observations. *Statist. Prob. Letters. Vol. 78*, 327-335.

[90] Mörters, P.(2008). *Large deviation theory and applications.* Schoolar article.

[91] Nelsen, R.B.(1999). *An Introduction to Copulas*, Springer, New-York.

[92] Patel, J.K., Kapadia, C.H. and Owen, D.B.(1976). *Handbook of Statistical Distribution.* Marcel Dekker, New York.

[93] Pearson, K.(1897). Mathematical contributions to the theory of evoluation. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. R. Soc. Lond. 60*, 489–498.

[94] Pearson, P.(1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably supposed to have arisen from random sampling. *Phil. Mag. Ser. 5*, 157.

[95] Peng, L.(1998). Asymtotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett. 38*, 107–115.

[96] Peng, L. and Qi, Y.(2006). Confidence regions for high quantiles of heavy tailed distribution. *The Annals of Statistics, 34(4)*, 1964–1986.

[97] Pickands, J.(1975). Statistical inference using extreme order statistics. *Ann. Statist. 3*, 119–131.

[98] Peng, L.(1998). Asymtotically unbiased estimator for the extreme-value index. *Statistics and Probability Letters 38(2)*, 107–115.

[99] Puhalskii, A.(2000). *Large deviations and idempotent probability.* Chapman & Hall/CRC monographs and surveys in pure and applied mathematics: 119.

[100] Rajić, V. and Stanojević, J.(2011). Confidence intervals for the difference of two proportions applied in incurance property. *Proc. XXXVIII Symp. Oprat. Res. SYMOPIS 2011*, 758–760, ISBN: 978-86-403-1168-7.

[101] Satterthwaite, F.E.(1941). Synthesis of variance. *Psychometrica 6*, 309–316.

[102] Satterthwaite, F.E.(1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin 6*, 110–114.

[103] Smith, R.L.(1987). Estimating tails of probability distribution. *Ann. Statist. 15*, 1174–1207.

[104] Snedecor, W.(1934). Calculation and Interpretation of Analysis of Varianse and Covariance. Ames. Ia., Collegiate Press Inc.

[105] Sprott, D.A. and Farewell, V.T.(1993). The difference between two normal means. *The American Statistician 47*, 126–128.

[106] Stanojević, J.(2014). On estimation of high quantiles for certain classes of distributions. *Yujor, DOI: 10.2298/YJOR130606013S*.

[107] Staurt, A. and Ord, K.(1998). *Kendalls Advanced Theory of Statistics. Volume 1: Distribution Theory*, Edward Arnold, London.

[108] Student.(1908). The probable error of a mean. *Biometrica 6*, 1–25.

[109] Varadhan, S.R.S.(1984). *Large Deviations and Applications.* SIAM.

[110] Weissman, I.(1978). Estimation of parameters and large quantiles based on the k largest observation. *J. Amer. Statist. Assoc. 73*, 812–815.

[111] Wellner, J.A.(1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. verw. Gebiete 45*, 73–88.

[112] Xing Jin and Michael C.Fu (2002). A Large Deviations Analysis of Quantile Estimation with Application to Value at Risk. *MIT Operations Research Center Seminar Series.*

[113] Zhou, X.H. and Dinh, P.(2005). Nonparametric confidence intervals for the one- and two-sample problems. *Biostatistics 6*, 187–200.

[114] Zhou, X.H. and Gao, S.(1997). Confidence Intervals for the Lognormal Mean. *Statistics in Medicine 16*, 783–790.

[115] Zhou, X.H. and Gao, S.(2000). One-sided confidence intervals for means of positively skewed distributions. *American Statistician 54*, 100–104.

[116] Zhou, X.H., Gao, S. and Hui, S.(1997). Methods for comparing the means of two independent log-normal samples. *Biometrics 53*, 1129–1135.

[117] Zhou, X.H., Tsao, M. and Qin, G.(2004). New intervals for the difference between two independent binomial proportions. *Journal of Statistical Planning and Inference 123*, 97–115.

[118] Ćojbašić Rajić, V. and Stanojević, J. (2013). Confidence interval for the ratio of two variances, *Journal of Applied Statistics 40 (10)*, 2181–2187.

[119] Ćojbašić, V. and Tomović, A.(2007). Nonparametric confidence intervals for population variance of one sample and difference of variances of two samples. *Computational Statistics & Data Analysis 51*, 5562–5578.

Biografija autora

Jelena Stanojević rodjenja je 16.05.1978. godine u Leskovcu. Radi na Ekonomskom fakultetu Univerziteta u Beogradu u zvanju asistenta.

Obrazovanje: Osnovnu školu "Filip Kljajić Fića" završila je u Beogradu, kao i Matematičku gimnaziju u Beogradu, sa odličnim uspehom. Diplomirala je na Matematičkom fakultetu, Univerziteta u Beogradu, 2002. godine, sa prosečnom ocenom 9.71. Zvanje magistra stekla je na Matematičkom fakultetu, Univerziteta u Beogradu, 2007. godine, na temu: Ocena visokih kvantila raspodela verovatnoća i primene u analizi rizika, pod rukovodstvom Prof. dr Pavla Mladenovića. Doktorske studije na Katedri za verovatnoću i statistiku upisala je 2011. godine.

Iskustvo u nastavi: Asistent je na Ekonomskom fakultetu, Univerziteta u Beogradu od 2003. godine. Kursevi koje drži su: Matematika, Matematika 2 i Teorijska statistika.

Učešće na projektima: Od 2005-2010 bila je član projekta Ministarstva prosvete i Matematičkog instituta u Beogradu, projekat pod nazivom: O geometriji i topologiji multiplikativnih i integrabilnih dinamičkih sistema. Od 2010 član je projekta Ministarstva prosvete i Matematičkog instituta u Beogradu, projekat pod nazivom: Geometrija i topologija mnogostrukosti, klasična mehanika i integrabilni dinamički sistemi.

Oblasti naučnog interesovanja: Verovatnoća i statistika, Aktuarska i finansijska matematika, Teorija ekstremnih vrednosti.

Spisak naučih radova

Radovi u časopisima i serijskim publikacijama

1. Ćojbašić Rajić, V. and Stanojević, J., *Confidence interval for the ratio of two variances*, Journal of Applied Statistics 40 (10), ISSN 0266-4763, pp. 2181–2187, 2013.

2. Stanojević, J., *On estimation of high quantiles for certain classes of distribution*, Yujor- Yugoslav Journal of Operations Research, doi: 10.2298/YJOR130606013S, 2014.

3. Rajić, V. and Stanojević, J., *A random coefficient moving average model*, International Journal of Pure and Applied Matematical Sciences, ISSN 0972-9828, Volume 5, Number 1-2, pp. 97–102, 2011.

4. Stanojević, J., *Some possible interpretations of coefficients of kurtosis and skewness*, Proc. XXXV Symp. Oprat. Res. SYMOPIS 2008, ISBN: 978-86-7395-248-2, pp. 623–626, 2008.

5. Stanojević, J., *Distributions with heavy tails and relationship between some of subclasses*, Proc. XXXVI Symp. Oprat. Res. SYMOPIS 2009, ISBN: 978-86-80953-43-4, pp. 623-626, 2009.

6. Rajić, V. and Stanojević, J., *The mix of the exponential distribution*, Proc. XXXVII Symp. Oprat. Res. SYMOPIS 2010, ISBN 978-86-335-0299, pp. 739–742, 2010.

7. Rajić, V. and Stanojević, J., *Confidence intervals for the difference of two proportions applied in incurance property*, Proc. XXXVIII Symp. Oprat. Res. SYMOPIS 2011, ISBN: 978-86-403-1168-7, pp. 758–760, 2011.

8. Stanojević, J. and Levajković, T., *Levy proceses and Penjers recursion*, Proc. XXXIX Symp. Oprat. Res. SYMOPIS 2012, ISBN: 978-86-7488-086-9, pp. 101–104, 2012.

9. Stanojević, J. and Levajković, T., *On the Cramer-Lundberg model with stochastic premia and the Panjers recursion*, Proceedings of the XIII Conference on Mathematics and its Applications, ISSN: 1224-6069, pp. 294-302, 2012.

10. Levajković, T. and Stanojević, J., *Brownian motion and heat equation*, Proc. XL Symp. Oprat. Res. SYMOPIS 2013, ISBN 978-86-7680-286-9, pp. 863–868, 2013.

Прилог 1.

Изјава о ауторству

Потписани-а Станојевић Јелена

број индекса 2046/2011

Изјављујем

да је докторска дисертација под насловом

Статистички проблеми оцењивања количника дисперзија и високих квантила расподела

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, мај, 2015

Станојевић Јелена

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Јелена Станојевић

Број индекса 2046/2011

Студијски програм Математика

Наслов рада Статистички проблеми оцењивања количника дисперзија и високих квантила расподела

Ментор Проф. др Павле Младеновић

Потписани/а Станојевић Јелена

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, мај, 2015

Станојевић Јелена

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

СТАТИСТИЧКИ ПРОБЛЕМИ ОЦЕЊИВАЊА КОЛИЧНИКА ДИСПЕРЗИЈА И ВИСОКИХ КВАНТИЛА РАСПОДЕЛА

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, МАЈ, 2015

Стојановић Јелена