

LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

NEPARAMETRISKĀS STATISTIKAS METODES
EKSTREMĀLU NOTIKUMU MODELĒŠANĀ.

MAĢISTRA DARBS

Autors: **Inta Cīmure**

Stud. apl. ik08296

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2010

Anotācija

Ekstremālo vērtību sadalījumu astes indeksa aprēķināšanai labi pazīstams novērtējums ir tā sauktais Hila novērtējums. Darbs aplūko astes indeksa ticamības intervālus, izmantojot Hila novērtējumu gan ar parametriskajām, gan neparametriskajām metodēm.

Darba mērķis ir salīdzināt dažādas ticamību intervālu konstruēšanas metodes, izmantojot pārklājuma precizitātes analīzi, kā arī mēģinot to uzlabot ar literatūrā pazīstamām korekcijas metodēm. Katra metode tika implementēta izmantojot programmu R un pielietota uz reāliem un simulētiem datiem.

Darba galvenie rezultāti parāda, ka neparametriskās metodes strādā labi un ir salīdzināmas ar parametrisko metožu rezultātiem.

Atslēgas vārdi: Regulāri variējoša funkcija, ekstremālo vērtību sadalījumi, Hila novērtējums, parametriskā un neparametriskā ticamības funkcija, butstrapa metode, Bartleta korekcija.

Abstract

The so-called Hill estimator is well-known for the estimation of the tail index of heavy tailed distributions. In this master paper the confidence intervals for the tail index of a heavy tailed distribution using Hill estimation with several parametric and nonparametric methods are reviewed.

The main purpose of the paper is to compare the coverage accuracy of each method and to improve the coverage accuracy using some adjustments commonly described in literature. Each method has been implemented by R script, written for each, and has been applied on real and simulated data.

The main results of the real data application and simulation study indicates that non-parametrical methods work well and are comparable with the results of the parametrical methods.

Keywords: Regularly varying function, extreme value distributions, Hill estimator, parametrical and empirical likelihood, bootstrap, Bartlett correction

Saturs

Apzīmējumi	3
Ievads	5
1. Teorijas apskats	7
1.1. Regulāri variējošas (<i>RV</i>) funkcijas	7
1.2. Vispārīgais ekstremālo vērtību sadalījums	9
1.2.1. Ekstremālo vērtību teorijas pamati	9
1.2.2. Ekstremālo vērtību sadalījumi.	13
1.2.3. G_γ definīcijas apgabals	18
1.3. Ekscesa funkcija un Pareto sadalījums	22
1.4. Ekstremālo vērtību sakārtotās statistikas	23
1.4.1. vidējais k	26
1.5. Astes indeksa γ Hila novērtējums	26
1.5.1. Hila novērtējuma motivācija	26
1.5.2. Situācija, kad Hila novērtējums nestrādā	30
1.5.3. Optimālā $k = k(n)$ aprēķināšana	32
1.6. Hila novērtējuma ticamības intervālu novērtēšanas metodes	33
1.6.1. Normālā aproksimācija	33
1.6.2. Vislielākās ticamības metode	34
1.6.3. Neparimetriskā ticamības metode	35
1.6.4. Butstrapa metode	38
1.7. Ticamības intervālu pārklājuma precizitāte	39
1.7.1. Bartleta korekcija vislielākās ticamības funkcijai	40
1.7.2. Bartleta korekcija empīriskai ticamības funkcijai	42
2. Praktiskā daļa	45
2.1. Simulācijas	45

2.1.1. Pārklājuma precizitāte pie optimālā k	45
2.1.2. Pārklājuma precizitāte dažādiem k	47
2.2. Metožu pielietojums reāliem datiem	52
Secinājumi	56
Pateicības	57
Izmantotā literatūra un avoti	58
A Izveidoto programmu kods	61

Apzīmējumi

\xrightarrow{d}	konverģence pēc sadalījuma,
\xrightarrow{P}	konverģence pēc varbūtības,
γ	smagās astes indekss, pirmās kārtas parametrs,
ρ	otrās kārtas parametrs, kurš nosaka nosacījuma konverģences ātrumu,
Φ_γ	Frešē sadalījums,
Φ	Standarta normālais sadalījums,
Ψ_γ	Weibula sadalījums,
$1_{\{p\}}$	indikatora funkcija: vienāda ar 1, ja p ir patiess, citādi 0,
$a(t), b(t)$	normējošās konstantes,
D	definīcijas apgabals,
E_i	standarta eksponenciāli sadalīts gadījuma lielums,
$e(u)$	vidējā ekscesa funkcija,
f^{-1}	funkcijas f inversā funkcija,
F	teorētiskā sadalījuma funkcija,
F_n	empīriskā sadalījuma funkcija,
F_u	ekscesa sadalījuma funkcija,
G_γ	ekstremālo vērtību sadalījuma funkcija,
G_0	Gumbela sadalījums,
GP	vispārīgais Pareto sadalījums,
I_B	ticamības intervāls konstruēts ar butsrapa metodi,
I_E	ticamības intervāls konstruēts ar empīrisko ticamības metodi

I_{EB}	ticamības intervāls konstruēts ar empīrisko ticamības metodi un tam piemērota empīriskā Bartleta korekcija,
I_N	ticamības intervāls konstruēts ar normālās aproksimācijas metodi
I_P	ticamības intervāls konstruēts ar vislielākās ticamības metodi
I_{PB}	ticamības intervāls konstruēts ar vislielākās ticamības metodi un tam piemērota Bartleta korekcija,
$L(x)$	lēni variējoša funkcija
$L(\alpha, x_m)$	vislielākās ticamības funkcija parametriem α un x_m ,
$l(\alpha, x_m)$	logaritmiskā vislielākās ticamības funkcija parametriem α un x_m ,
M_n	izlases (X_1, \dots, X_n) maksimums,
$O(1)$	$a(x) = O(b(x))$, kad $x \rightarrow x_0$ nozīmē, ka $\limsup_{x \rightarrow x_0} a(x)/b(x) < \infty$,
$X_{n,n}$	maksimums no X_1, \dots, X_n ,
$X_{n-k,n}$	k -tā lielākā sakārtotā statistika,
\mathbb{R}^+	pozitīva reālo skaitļu kopa $[0, +\infty)$,
RV_α	regulāri variējošas funkcijas klase ar regulārās variācijas indeksu α ,
$R(\gamma)$	logaritmiskā vislielākās ticamības attiecība
U	inversā funkcija lielumam $1/(1 - F)$,

Ievads

Ekstremālo vērtību sadalījumi pieņem formu

$$G_\gamma(x) = \exp\{-(1 + \gamma x)^{-1/\gamma}\}$$

katram $1 + \gamma x > 0$. Parametru γ sauc par sadalījuma F astes indeksu. Ekstrēmālo vērtību teorijā šis parametrs ir viens no svarīgākajiem rīkiem un spēlē nopietnu lomu ļoti retu gadījuma lielumu prognozēšanā.

1975. gadā Hils publicēja rakstu [1], kur aprakstīja metodi, kā var novērtēt sadalījuma F astes indeksu, kas raksturojas caur indeksu no regulāri variējošām funkcijām. Mūsdienās šis novērtējums iespējams ir visplašāk izpētītais un ir ieguvis nosaukumu Hila novērtējums. Daudzi ir pētījuši novērtējuma asimptotisko uzvedību un detalizētu analīzi par Hila novērtējumu var atrast grāmatās, ko sarakstījuši Haan un Ferreira [2] un Embrechts [3].

Šajā darbā tiek apskatītas ticamības intervālu konstruēšanas metodes tādas kā normālā aproksimācija, parametriskā un empīriskā ticamības funkcija, butstrapa metode. Katra šī metode ir pielietota parametra γ ticamības intervālu aprēķināšanai. Barndorff un Cox 1984.gadā [4] parāda parametriskai ticamības metodei un Diccio 1991.gada [5] pierāda empīriskai ticamības metodei, ka abas metodes var tikt uzlabotas ar Bartleta korekciju. Līdz šim literatūrā Bartleta korekcija nav tikusi pielietota ekstrēmu sadalījumu gadījumā empīriskai ticamības funkcijai. Šajā darbā tiks analizēts, vai Bartleta korekciju var pielietot arī ekstrēmu sadalījumu gadījumā.

Šī darba mērķis ir noskaidrot kādas metodes var pielietot smagās astes indeksa ticamības intervālu novērtēšanai, ar kādu precizitāti šīs metodes strādā un vai ir iespējams šo precizitāti uzlabot.

Lai sasniegtu mērķi, tika izvirzīti vairāki darba uzdevumi:

1. iepazīties ar ekstremālo vērtību iespējamiem sadalījumiem un noskaidrot, kādi parametri raksturo šos sadalījumus, kā arī kādi ir visplašāk izmantotie parametru novērtējumi;
2. noskaidrot, kādas parametriskās un neparametriskās metodes var piemērot smagās astes indeksa ticamības intervālu novērtēšanai;
3. ar pārklājuma precizitātes simulāciju palīdzību noteikt, cik labi strādā katra metode;

4. pielietot apskatītās metodes reāliem datiem.

Darbā aprakstītās simulācijas un reālu datu analīze tika realizētas ar programmā R izveidotām programmām. Darbs sastāv no teorētiskā apskata un praktiskās daļas.

Teorijas apskats sākas ar 1.1. nodaļu, kurā tiek aprakstītas regulāri variējošas funkcijas. 1.2. nodaļā tiek parādīti, kādi ir iespējami ekstremālo vērtību sadalījumi, un 1.3. nodaļā - kā tiek iegūts Pareto sadalījums. Tā kā Hila novērtējums ietver sevī sakārtotās statistikas, tad 1.4. nodaļā īsi ir pastāstīts par sakārtotām statistikām un 1.5. nodaļā jau izklāstīts pats Hila novērtējums. 1.6. nodaļā ir aprakstītas ticamības intervālu novērtēšanas metodes, kā arī 1.7. nodaļā izklāstīta Bartleta korekcija parametriskai un empīriskai ticamības metodei.

Rezultāti ir apkopoti nodaļā 2., kur metodes tika pielietotas simulētiem un reāliem datiem. Reālie dati apraksta Dānijas ugunsgrēka riska apdrošināšanas atlīdzības laika periodā no 1980-1990.gadam. Secinājumos aprakstīti galvenie darbā iegūtie rezultāti. Pēdējā nodaļā ir ievietots izmantotās literatūras saraksts. Pielikumā ir ievietotas statistiskās paketes R [6] aplūkoto metožu izstrādātās programmas.

1. Teorijas apskats

1.1. Regulāri variējošas (*RV*) funkcijas

Ekstremālo vērtību teorijā svarīga loma ir regulāri variējošām funkcijām. Arī šajā darbā regulāri variējošas funkcijas jēdziens tiks izmantots ļoti bieži. Lai izprastu, kas ir regulāri variējoša funkcija, aplūkosim tuvāk šī termina skaidrojumu un piemērus. Galvenokārt informācija iegūta no grāmatas [2, 360–401 lpp.].

Motivācija. Viens veids kā izskaidrot terminu regulārā variācija ir definēt to kā funkcijas atvasinājumu bezgalībā. Kādai reāli mērojamai funkcijai $g(y)$ izsakām diferenciāli

$$\frac{g(y+h) - g(y)}{h},$$

kur $h \neq 0$. Tikai tagad aplūkojam nevis gadījumu, kad $h \rightarrow 0$, kādam fiksētam y , bet gan gadījumu, kad $y \rightarrow \infty$ un h ir fiksēts. Ja šāda robeža eksistē katram h ($h \neq 0$), tad apgalvosim, ka šī robeža nav atkarīga no h . Ja robežu apzīmēsim ar α , kad $y \rightarrow \infty$, tad definējam funkciju $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$, kur $f := e^{g(\log(t))}$.

Var redzēt, ka, ja mēs pieņemam, ka $y = \log t$, $h = \log x$

$$\lim_{y \rightarrow \infty} \frac{g(y+h) - g(y)}{h} = \lim_{t \rightarrow \infty} \frac{g(\log t + \log x) - g(\log t)}{\log x} = \alpha$$

Veicot dažus pārveidojumus:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{(g(\log t + \log x) - g(\log t))}{\log x} \log x &= \\ &= \lim_{t \rightarrow \infty} g(\log tx) - g(\log t) = \alpha \log x \end{aligned}$$

un ekvivalence nepazudīs, ja rakstīsim

$$\lim_{t \rightarrow \infty} e^{g(\log tx) - g(\log t)} = e^{\alpha \log x},$$

jo $\exp(x)$ ir monotoni augoša funkcija. Iegūstam, ka

$$\begin{aligned} \lim_{t \rightarrow \infty} e^{g(\log tx) - g(\log t)} &= \\ &= \lim_{t \rightarrow \infty} \frac{e^{g(\log tx)}}{e^{g(\log t)}} = \\ &= \lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha \end{aligned}$$

Definīcija 1. Lebeģa mērojamu funkciju $f(x) : \mathbb{R}^+ \rightarrow \mathbb{R}$ saucim par regulāri variējošu (bezgalībā), ja $\forall x > 0$ eksistē galīga robeža $\alpha \in \mathbb{R}$, kas nav vienāda ar nulli, un ja

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha. \quad (1.1.1)$$

Ja $\alpha = 0$, tad f sauc par lēni variējošu funkciju, kurai ir spēkā

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = 1.$$

α sauc par regulārās variācijas indeksu. Lieto pierakstu $f \in RV_\alpha$. Tas nozīmē katram reālam α , RV_α pieder funkciju klasei, kura regulāri variē bezgalībā ar regulāro variācijas indeksu α .

Piemērs 1. Katram $\alpha, \beta \in \mathbb{R}$ funkcijas x^α , $x^\alpha(\log x)^\beta$, $x^\alpha(\log \log x)^\beta$ pieder RV_α . Lēni variējošas funkcijas ir katram $\beta \in \mathbb{R}$, $L(x) = (\log x)^\beta$. Funkcijas $\sin x$, e^x nav regulāri variējošas funkcijas.

1. Ja $f(x) = x^\alpha$, tad

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = \lim_{t \rightarrow \infty} \frac{(tx)^\alpha}{t^\alpha} = x^\alpha.$$

2. Ja $f(x) = x^\alpha(\log x)^\beta$, tad var parādīt

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} &= \lim_{t \rightarrow \infty} \frac{(tx)^\alpha (\log tx)^\beta}{t^\alpha (\log t)^\beta} = \lim_{t \rightarrow \infty} \frac{x^\alpha \left(\log t \left(1 + \frac{\log x}{\log t} \right) \right)^\beta}{(\log t)^\beta} = \\ &= \lim_{t \rightarrow \infty} x^\alpha \left(1 + \frac{\log x}{\log t} \right)^\beta = x^\alpha. \end{aligned}$$

3. Ja $f(x) = (\log x)^\alpha$, tad robeža

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = \lim_{t \rightarrow \infty} \frac{(\log tx)^\alpha}{(\log t)^\alpha} = \lim_{t \rightarrow \infty} \left(\frac{\log x + \log t}{\log t} \right)^\alpha = \lim_{t \rightarrow \infty} \left(1 + \frac{\log x}{\log t} \right)^\alpha = 1.$$

4. Ja $f(x) = \sin x$, tad katram $x \geq 2$ izteiksmei $f(xt)/f(t)$ neeksistē robeža, kad $t \rightarrow \infty$.

Piezīme 1. Ja $\lim_{x \rightarrow \infty} L(x) = b \in (0, \infty)$, tad $L(x)$ ir lēni variējoša funkcija.

Piezīme 2. Regulāri variējoša funkcija pieņem formu $x^\alpha L(x)$, kur $\alpha \geq 0$ un $L(x)$ ir lēni variējoša funkcija (skatīt [7, 110.–116. lpp.]).

Pieminēsim dažas īpašības, kuras piemīt regulāri variējošām funkcijām. Īpašību pierādījumus skatīt ([2, 368–369 lpp.]).

1. Ja $f \in RV_\alpha$, tad $\frac{\log f(t)}{\log t} \rightarrow \alpha$, kad $t \rightarrow \infty$. Tas norāda, ka

$$\lim_{t \rightarrow \infty} f(t) = \begin{cases} 0, & \alpha < 0, \\ \infty, & \alpha > 0 \end{cases}$$

2. Ja $f_1 \in RV_{\alpha_1}$, $f_2 \in RV_{\alpha_2}$, tad $f_1 + f_2 \in RV_{\max(\alpha_1, \alpha_2)}$. Turklāt, ja $\lim_{t \rightarrow \infty} f_2(t) = \infty$, tad kompozīcija $f_1 \circ f_2 \in RV_{\alpha_1 \alpha_2}$.
3. Ja $f \in RV_\alpha$ ar $\alpha > 0$ ($\alpha < 0$), tad f asimptotiski ekvivalenta stingri augošai (dilstošai) diferencējamai funkcijai g , kuras atvasinājums $g' \in RV_{\alpha-1}$, ja $\alpha > 0$ un $-g' \in RV_{\alpha-1}$, ja $\alpha < 0$.
4. (Potera nevienādība (1942)) Pieņemam, ka $f \in RV_\alpha$. Ja patvaļīgiem $\delta_1, \delta_2 > 0$ eksistē $t_0 = t_0(\delta_1, \delta_2)$ tāds, ka katram $t \geq t_0$, $tx \geq t_0$ izpildās

$$(1 - \delta_1)x^\alpha \min(x^{\delta_2}, x^{-\delta_2}) < \frac{f(tx)}{f(t)} < (1 + \delta_1)x^\alpha \max(x^{\delta_2}, x^{-\delta_2}). \quad (1.1.2)$$

Piezīme 3. Izpildās arī pretējais, t.i. ja funkcijai f ir spēkā Potera nevienādība, tad $f \in RV_\alpha$

5. Ja $f \in RV_\alpha$, $\alpha > 0$ un f ir augoša, tad inversā funkcija $f^{-1} \in RV_{\frac{1}{\alpha}}$.

1.2. Vispārīgais ekstremālo vērtību sadalījums

1.2.1. Ekstremālo vērtību teorijas pamati

Pieņemsim, ka ir doti X_1, \dots, X_n neatkarīgi un vienādi sadalīti gadījuma lielumi. Ekstremālo vērtību teorija pēta kā robežgadījumā uzvedās izlašu ekstremālās vērtības $M_n := \max(X_1, X_2, \dots, X_n)$ vai $\min(X_1, X_2, \dots, X_n)$, kad $n \rightarrow \infty$. Un visbiežāk ekstremālo vērtību teorijā galvenais uzdevums ir noskaidrot, kādu robežsadalījumu var pieņemt izlašu maksimumi, ja izlases sastāv no neatkarīgiem un vienādi sadalītiem gadījuma lielumiem.

Ar F apzīmēsim teorētisko sadalījumu pēc kura ir sadalīti gadījuma lielumi X_1, \dots, X_n , un ar x^* apzīmēsim labo galapunktu, t.i.,

$$x^* := \sup\{x : F(x) < 1\}.$$

x^* var pieņemt arī bezgalīgu vērtību. Tad

$$\max(X_1, X_2, \dots, X_n) \xrightarrow{P} x^*, \quad n \rightarrow \infty,$$

kur \xrightarrow{P} nozīmē konvergenci pēc varbūtības, jo varbūtība

$$P(M_n \leq x) = P(\max(X_1, X_2, \dots, X_n) \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = F^n(x),$$

konverģē uz nulli, ja $x < x^*$ un uz viens, ja $x \geq x^*$. Šī varbūtība raksturo maksimuma M_n sadalījuma funkciju. Intuitīvi var saprast, ka M_n asimptotiskā uzvedība ir cieši saistīta ar sadalījuma funkcijas F labās astes labo galapunktu x^* .

Dziļāku ieskatu par maksimumu sadalījumu dot vājās konverģences rezultāts, kuru iegūst no centrētiem un normalizētiem maksimumiem. Tā ir viena no svarīgākajām tēmām ekstremālo vērtību teorijā. Lai iegūtu nedeģenerētu robežsadalījumu, gadījuma lielumus normē.

Pieņemsim, ka eksistē virkne konstanšu $a_n > 0$ un $b_n \in \mathbb{R}$ ($n = 1, 2, \dots$), tādās, ka

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n}$$

veido nedeģenerētu robežsadalījumu, kad $n \rightarrow \infty$, t.i.,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \tag{1.2.1}$$

katrā nepārtrauktā punktā x no G , un G ir nedeģenerēta sadalījuma funkcija. Robežizteiksmi (1.2.1) var pierakstīt arī savādāk, tāpēc alternatīvo izteiksmes pierakstu parādīsim nākošajā apakšnodaļā. 1.2.2. nodaļā tiks aprakstīti visi iespējamie sadalījumi G , kurus var iegūt kā robežu izteiksmē (1.2.1). Šādus sadalījumus G sauc par ekstremālo vērtību sadalījumiem. Katram iespējamam ekstremālo vērtību sadalījumam G tiks aprakstīti nepieciešamie un būtiskie nosacījumi. Šie nepieciešamie un būtiskie nosacījumi tiks uzlikti sadalījumiem F no (1.2.1), jo katrs G satur veselu klasi no dažādiem F . Šo klasi dažkārt sauc par G definīcijas apgabalu. 1.2.3. nodaļā tiks parādīti ekstremālo vērtību sadalījumi un to definīcijas apgabali.

Veiksim dažus pārveidojumus ar nosacījumu (1.2.1). Logaritmējam abas puses un iegūstam ekvivalentu izteiksmi

$$\lim_{n \rightarrow \infty} n \log F(a_n x + b_n) = \log G(x). \quad (1.2.2)$$

No vienādojuma (1.2.2) ir redzams, ka $F(a_n x + b_n) \rightarrow 1$, katram x . Aplūkojot robežu

$$\lim_{n \rightarrow \infty} \frac{-\log F(a_n x + b_n)}{1 - F(a_n x + b_n)} = 1,$$

var iegūt, ka izteiksme (1.2.2) ir ekvivalenta ar

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\log G(x),$$

kur $\log F(a_n x + b_n)$ tika aizstāts ar $1 - F(a_n x + b_n)$. Un, ja paņemam iegūtās sakarības apgriezto lielumu, iegūstam izteiksmi

$$\lim_{n \rightarrow \infty} \frac{1}{n(1 - F(a_n x + b_n))} = \frac{1}{-\log G(x)}. \quad (1.2.3)$$

Nosacījumu (1.2.3) izteiksim caur inverso funkciju. Pieņemsim, ka katrai nedilstošai funkcijai f inversā funkcija ir f^{-1} , kura ir nepārtraukta no kreisās puses, t.i.,

$$f^{-1}(x) := \inf\{y : f(y) \geq x\}.$$

Lemma 4. *Pieņemsim, ka f_n ir virkne no nedilstošām funkcijām un g ir nedilstoša funkcija. Pieņemsim, ka katrs x kādā vaļējā intervālā (a, b) tiek raksturots kā nepārtraukts punkts no g ,*

$$\lim_{n \rightarrow \infty} f_n(x) = g(x). \quad (1.2.4)$$

Pieņemam, ka f_n^{-1} , g^{-1} ir no kreisās puses nepārtrauktās inversās funkcijas funkcijām f_n un g . Tad katrs x intervālā $(g(a), g(b))$ ir nepārtraukts punkts g^{-1} un iegūst, ka

$$\lim_{n \rightarrow \infty} f_n^{-1}(x) = g^{-1}(x). \quad (1.2.5)$$

Lemmas pierādījums balstās uz to, ka tiek fiksēts $\varepsilon > 0$ un tiek pierādīts, ka katram n , $n_0 \in \mathbb{N}$, $n \geq n_0$ ir spēkā

$$f_n^{-1}(x) - \varepsilon \leq g^{-1}(x) \leq f_n^{-1}(x) + \varepsilon.$$

Tas ir iespējams, jo nepārtrauktie punkti x no g veido blīvu kopu. Lemmas pierādījumu var skatīt [2, 5 lpp.].

Lemmu 4 piemērosim sakarībai (1.2.3). Ar U apzīmēsim inverso funkciju no funkcijas $1/(1 - F)$, kura ir nepārtraukta no kreisās puses. Tas ir, ja mums ir doti X_1, X_2, \dots, X_n neatkarīgi un vienādi sadalīti gadījuma lielumi un $X_i \sim F$, tad lielumam

$$\frac{1}{1 - F(x)} = t$$

inversā funkcija ir

$$x = F^{-1} \left(1 - \frac{1}{t} \right) := U(t). \quad (1.2.6)$$

Atzīmēsim, ka $U(t)$ ir definēta visiem $t > 0$.

Savukārt $-1/\log G(x)$ inversā funkcija ir $G^{-1}(e^{-1/x})$. Un viegli redzēt, ka $1/(n(1 - F(a_n x + b_n)))$ inverso izteiksmi var izteikt kā

$$\begin{aligned} \frac{1}{n(1 - F(a_n x + b_n))} &= t, \\ F(a_n x + b_n) &= 1 + \frac{1}{nt}, \\ F^{-1}(F(a_n x + b_n)) &= a_n x + b_n = F^{-1} \left(1 + \frac{1}{nt} \right) = U(nt), \\ x &= \frac{U(nt) - b_n}{a_n}. \end{aligned}$$

Rezultātā katram $x > 0$ sakarību (1.2.3) ar inversām funkcijām var izteikt kā

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = G^{-1}(e^{-1/x}) =: D(x). \quad (1.2.7)$$

Tālāk izteiksim kopsavilkumu teorēmā, kurā formulēsim vairākus ekvivalentus apgalvojumus.

Teorēma 5. *Pieņemsim, ka dotas $a_n > 0, b_n \in \mathbb{R}$ konstanšu virknes un G ir nedeģenerējoša sadalījuma funkcija. Tad sekojoši apgalvojumi ir ekvivalenti:*

1. *Katram G nepārtrauktības punktam x*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x).$$

2. *Katram G nepārtrauktam punktam x , kur $0 < G(x) < 1$, $a(t) := a_{[t]}$, un $b(t) := b_{[t]}$ (kur $[t]$ ir veselā daļa no t)*

$$\lim_{n \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\log G(x).$$

3. *Katram $x > 0$ nepārtrauktam punktam no $D(x) = G^{-1}(e^{-1/x})$, $a(t) := a_{[t]}$, un $b(t) := b_{[t]}$*

$$\lim_{n \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x). \quad (1.2.8)$$

Pierādījumu skatīt [2, 6 lpp.].

1.2.2. Ekstremālo vērtību sadalījumi.

Tagad mēs atrodamies pozīcijā, kad varam identificēt ekstremālo vērtību sadalījumu klasi, kuru iespējams iegūt kā robežvērtību no sakarības (1.2.2). Teorēmā 6 noformulēsim visus iespējos $G(x)$ sadalījumus.

Teorēma 6. (*Fisher un Tippett (1928), Gnedenko (1943)*)

Vispārīgais ekstremālo vērtību sadalījums ir $G_\gamma(ax + b)$ ar $a > 0$, $b \in \mathbb{R}$, kur katram $1 + \gamma x > 0$,

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & \gamma \neq 0, \\ \exp(-e^{-x}), & \gamma = 0. \end{cases} \quad (1.2.9)$$

Definīcija 2. Parametrs γ no (1.2.9) tiek saukts par ekstremālo vērtību indeksu.

Pierādījums. Pierādīsim Teorēmu 6. Pieņemsim, ka mums ir dota robežfunkciju klase D no (1.2.8). Pieņemsim, ka vērtība 1 ir nepārtraukts punkts no D . Tad, atzīmēsim, ka nepārtrauktam punktam $x > 0$ spēkā,

$$\lim_{n \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D(x) - D(1) =: E(x). \quad (1.2.10)$$

Ņemsim $y > 0$ un rakstīsim

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}. \quad (1.2.11)$$

Mēs apgalvojam, ka eksistē robežas $\lim_{t \rightarrow \infty} (U(ty) - U(t))/a(t)$ un $\lim_{t \rightarrow \infty} a(ty)/a(t)$. Ja robežas neeksistē, tad eksistē tādi A_1, A_2, B_1, B_2 , ka $A_1 \neq A_2$ vai $B_1 \neq B_2$, kur B_i ir robeža no $(U(ty) - U(t))/a(t)$ un A_i ir robeža no $a(ty)/a(t)$, $i = 1, 2$, kad $t \rightarrow \infty$. No (1.2.11) atrodam, ka

$$E(xy) = E(x)A_i + B_i, \quad i = 1, 2, \quad (1.2.12)$$

katram nepārtrauktam punktam x no $E(\cdot)$. Kādam patvaļīgam x izvēlamies virkni nepārtrauktu punktu x_n tādus, ka $x_n \uparrow x$ ($n \rightarrow \infty$). Tad $E(x_n y) \rightarrow E(xy)$ un $E(x_n) \rightarrow E(x)$, ja E ir nepārtraukts no kreisās puses. (1.2.12) spēkā visiem pozitīviem x un y . Atņemot vienu no otras izteiksmes pie $i = 1, 2$, iegūstam

$$E(x)(A_1 - A_2) = B_2 - B_1$$

visiem $x > 0$. Tā kā E nedrīkst būt konstante (jo G ir nedeģenerējoša), jāiegūst, ka $A_1 = A_2$ un tāpēc arī $B_1 = B_2$. Secinājums:

$$A(y) := \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$$

eksistē katram $y > 0$, un katram $x, y > 0$,

$$E(xy) = E(x)A(y) + E(y).$$

Tādējādi ņemot $s := \log x$, $t := \log y$ ($x, y \neq 1$), un $H(x) := E(e^x)$, iegūstam

$$H(t + s) = H(s)A(e^t) + H(t), \quad (1.2.13)$$

ko varam pārrakstīt kā (jo $H(0) = 0$)

$$\frac{H(t + s) - H(t)}{s} = \frac{H(s) - H(0)}{s}A(e^t). \quad (1.2.14)$$

Šeit eksistē viens t , kur ir H -diferencējams (jo H ir monotona); tāpēc no (1.2.14) H ir diferencējama katrā punktā un

$$H'(t) = H'(0)A(e^t). \quad (1.2.15)$$

Apzīmējam $Q(t) := H(t)/H'(0)$. Atzīmējam, ka $H'(0) \neq 0$: H nevar būt konstante, jo G ir nedeģenerējoša. Tad $Q(0) = 0$, $Q'(0) = 1$. No (1.2.13) iegūstam

$$Q(t + s) - Q(t) = Q(s)A(e^t),$$

un no (1.2.15) iegūstam

$$Q(t + s) - Q(t) = Q(s)Q'(t). \quad (1.2.16)$$

Atņemot vienu no otras šīs izteiksmes ar t un s , iegūstam

$$Q(t) \frac{Q'(s) - 1}{s} = \frac{Q(s)}{s} (Q'(t) - 1),$$

tāpēc (ja $s \rightarrow 0$),

$$Q(t)Q''(0) = Q'(t) - 1.$$

Iegūstam, ka Q ir divreiz diferencējama, un

$$Q''(0)Q'(t) = Q''(t).$$

Tāpēc

$$(\log Q')'(t) = Q''(0) =: \gamma \in \mathbb{R}$$

katram t . No tā seko (atzīmēsim, ka $Q'(0) = 1$)

$$Q'(t) = e^{\gamma t}$$

un (jo $Q(0) = 0$)

$$Q(t) = \int_0^t e^{\gamma s} ds.$$

Tas nozīmē, ka

$$H(t) = H'(0) \frac{e^{\gamma t} - 1}{\gamma}$$

un

$$D(t) = D(1) + H'(0) \frac{t^\gamma - 1}{\gamma}.$$

Tāpēc

$$D^{-1}(x) = \left(1 + \gamma \frac{x - D(1)}{H'(0)}\right)^{1/\gamma}. \quad (1.2.17)$$

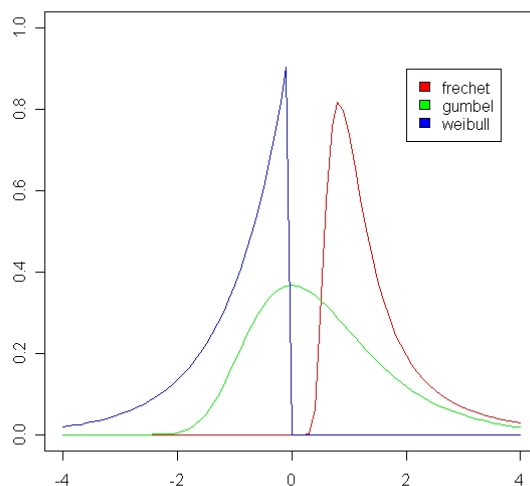
Ja izvēlamies, ka $D(x) = G^{-1}(e^{-1/x})$, tad iegūst

$$D^{-1}(x) = \frac{1}{-\log G(x)}. \quad (1.2.18)$$

No izteiksmēm (1.2.17) un (1.2.18) iegūstam teorēmas apgalvojumu.

Ja 1 nav D nepārtrauktības punkts, tad pierādījumā jāizmanto funkcija $U(tx_0)$ ar x_0 nepārtrauktu punktu no D . \square

Mūsdienās plaši izmanto Teorēmas 6 reprezentāciju, kuru aprakstīja von Mise (skatīt [8, 271–294 lpp.]) un Jenkinson (skatīt [9, 158–171 lpp.]). Teorēma 6 parāda, ka robežsadaliņuma funkciju var izteikt ar vienu parametru γ . Bez tam var redzēt, ka $G_\gamma(x)$ satur sadaliņumus ar diezgan atšķirīgām iezīmēm, piemēram bildē 1.1. ir parādītas $G_\gamma(x)$ blīvuma funkcijas pie dažādiem γ .



1.1. att.: $G_\gamma(x)$ blīvumu funkcijas pie $\gamma = 1$ (Frešē sadalījums), $\gamma = 0$ (Gumbela sadalījums) un $\gamma = -1$ (Veibula sadalījums).

Vispārīgajā ekstremālo vērtību sadalījumā papildus vēl var būt parametri $\mu \in \mathbb{R}$, kas raksturo novietojumu, un $\sigma > 0$, kas raksturo mērogu. Pie šiem parametriem $G_\gamma(x, \mu, \sigma)$ izsakās kā

$$G_\gamma(x; \mu, \sigma) = \exp \left\{ - \left[1 + \gamma \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\gamma} \right\}$$

Ja $\mu = 0$ un $\sigma = 1$, iegūstam standarta ekstremālo vērtību sadalījumu, kurš ir atkarīgs vienīgi no γ - sadalījuma formas parametra. Atkarībā no γ izšķir sekojošus sadalījumus:

1. Ja $\gamma > 0$, tad iegūstam Frešē sadalījumu

$$\Phi_\alpha(x; \mu, \sigma) = \begin{cases} 0 & , x \leq \mu \\ e^{-((x-\mu)/\sigma)^\alpha} & , x > \mu. \end{cases}$$

$G_\gamma(x) < 1$ katram x . Tas nozīmē, ka labais sadalījuma galapunkts ir bezgalība. Bez tam, kad $x \rightarrow \infty$, $1 - G_\gamma(x) \sim \gamma^{-1/\gamma} x^{-1/\gamma}$, tas nozīmē, ka sadalījumam piemīt smaga labā aste.

2. Ja $\gamma = 0$, tad iegūstam Gumbela sadalījumu

$$G_0(x; \mu, \sigma) = e^{-e^{-(x-\mu)/\sigma}} \quad \forall x \in \mathbb{R},$$

katram reālam x . Sadalījuma labais galapunkts ir bezgalība. Sadalījums ir ar vieglām astēm: $1 - G_0(x) \sim e^{-x}$, kad $x \rightarrow \infty$ un eksistē visi momenti.

3. Ja $\gamma < 0$, iegūstam Veibula sadalījumu

$$\Psi_\alpha(x; \mu, \sigma) = \begin{cases} e^{-(-(x-\mu)/\sigma)^\alpha} & , x < \mu \\ 1 & , x \geq \mu \end{cases}$$

Sadalījuma labais galapunkts ir $-1/\gamma$ un tam ir īsa aste, pārbaudot $1 - G_\gamma(-\gamma^{-1} - x) \sim (-\gamma x)^{-1/\gamma}$, kad $x \downarrow 0$.

Piemērs 2. Apskatīsim kādu sadalījumu pieņem maksimums no eksponenciāli sadalītiem gadījuma lielumiem. Pieņemsim, ka X_1, \dots, X_n ir virkne no neatkarīgiem un vienādi sadalītiem standarta eksponenciāliem gadījuma lielumiem. Maksimumu izsakām kā $M_n = \max(X_1, \dots, X_n)$. M_n statistikai atņemsim normējošu konstanti $\ln n$. Par normējošām konstantēm sīkāku izklāstu var rast grāmatā [3, 120–127 lpp.]. Aplūkojot iegūtā lieluma kumulatīvo sadalījuma funkciju

$$\begin{aligned} P(M_n - \ln n \leq x) &= P(X_1 \leq x + \ln n, \dots, X_n \leq x + \ln n) = \\ &= \prod_{j=1}^n P(X_j \leq x + \ln n) = \prod_{j=1}^n 1 - \exp(-x - \ln n) = \\ &= (1 - n^{-1}e^{-x})^n \xrightarrow{n \rightarrow \infty} \exp\{-e^{-x}\} = G_0(x; 0; 0) \end{aligned}$$

var redzēt, ka maksimumi no standarta eksponenciāli sadalītiem gadījuma lielumiem pieņem standatgumbela sadalījumu.

Piemērs 3. Apskatīsim kādu sadalījumu pieņem maksimums no standarta Koši sadalītiem gadījuma lielumiem. Pieņemsim, ka X_1, \dots, X_n ir virkne no neatkarīgiem un vienādi sadalītiem standarta Koši gadījuma lielumiem. Standarta Koši sadalījums ir nepārtraukts sadalījums ar blīvuma funkciju

$$f(x) = (\pi(1 + x^2))^{-1}, \quad x \in \mathbb{R}.$$

Kā arī pieņemsim faktu, ka standarta Koši sadalījuma F aste $\bar{F} := 1 - F \sim (\pi x)^{-1}$ (skatīt [3, 125 lpp.]). Maksimumu M_n sareizināsim ar π/n . Tad iegūsim, ka

$$\begin{aligned} P\left(\frac{\pi M_n}{n} \leq x\right) &= P\left(X_1 \leq \frac{nx}{\pi}, \dots, X_n \leq \frac{nx}{\pi}\right) = \\ &= \prod_{j=1}^n P\left(X_j \leq \frac{nx}{\pi}\right) = \prod_{j=1}^n \left(1 - \bar{F}\left(\frac{nx}{\pi}\right)\right) = \\ &= \left(1 - \frac{1}{nx} + o(1)\right)^n \xrightarrow{n \rightarrow \infty} \exp\{-x^{-1}\} = \Phi_1(x; 0; 1), \end{aligned}$$

tāpat maksimumi no standarta Koši sadalītiem gadījuma lielumiem pieņem Frešē sadalījumu.

Piemērs 4. Weibula sadalītu gadījuma lielumu var iegūt no vienmērīgā sadalījuma maksimumiem, kur $x \in (0, 1)$ un maksimumam M_n piemēro normēšanu $(M_n - 1)/n$.

Tagad mēs sakarībā (1.2.1) ieviesīsim izmaiņas, ņemot $G = G_\gamma$ kādam $\gamma \in \mathbb{R}$ un teiksim, ka sadalījuma funkcija F pieder G_γ definīcijas apgabalam. Piezīme: Tiek lietots pieraksts $F \in D(G_\gamma)$.

Teorēma 7. Katram $\gamma \in \mathbb{R}$ ir ekvivalenti sekojoši apgalvojumi:

1. Eksistē konstantes $a_n > 0$ un $b_n \in \mathbb{R}$ tādas, ka

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad (1.2.19)$$

katram x ar $1 + \gamma x > 0$.

2. Eksistē pozitīva funkcija a tāda, ka $\forall x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma}, \quad (1.2.20)$$

kur, ja $\gamma = 0$, tad robežu no labās puses interpretē kā $\log x$.

3. Eksistē pozitīva funkcija a tāda, ka

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \gamma x)^{-1/\gamma}, \quad (1.2.21)$$

visiem x , kuriem $1 + \gamma x > 0$.

4. Eksistē pozitīva funkcija f tāda, ka

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \quad (1.2.22)$$

katram x , kuram spēkā $1 + \gamma x > 0$, kur $x^* = \sup\{x : F(x) < 1\}$.

Teorēmas 7 pierādījumu skatīt [2, 10–11 lpp.].

1.2.3. G_γ definīcijas apgabals

Šajā apakšnodaļā aprakstīsim svarīgus nosacījumus sadalījumu funkcijām F , lai tās varētu nodrošināt, ka eksistē tādas normējošo konstanšu virknes $a_n > 0$ un $b_n \in \mathbb{R}$, ka

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) \quad (1.2.23)$$

kādam reālam γ un katram x . Šie nosacījumi pieprasa, lai eksistē pirmās un otrās kārtas atvasinājumi funkcijai F . Nākamā teorēma apgalvo svarīgu nosacījumu, kas nosaka pierību definīcijas apgabalam. Šo nosacījumu dēvē par von Mises nosacījumu. Teorēmas pierādījumu skatīt [2, 15–1111 lpp.]

Teorēma 8. *Pieņemsim, ka F ir sadalījuma funkcija un x^* ir labais galapunkts. Pieņemsim, ka eksistē $F''(x)$ un $F'(x) > 0$ katram x , kas atrodas pa kreisi no galapunkta x^* .*

Ja

$$\lim_{t \uparrow x^*} \left(\frac{1 - F}{F'} \right)'(t) = \gamma \quad (1.2.24)$$

vai ekvivalenti

$$\lim_{t \uparrow x^*} \frac{(1 - F(t))F''(t)}{(F'(t))^2} = -\gamma - 1, \quad (1.2.25)$$

tad F pieder G_γ definīcijas apgabalam.

Piezīme 9. *Izteiksmē (1.2.24) mēs iegūstam (1.2.23), ja izvēlamies $b_n = U(n)$ un $a_n = nU'(n) = 1/(nF'(b_n))$. Tālākās nodaļās mēs izmantosim sekojošos von Mises nosacījumus, kuri izmainīti, ieviešot lielumu U , kas definēts iepriekš (1.2.6).*

Sekas 10. *Nosacījums (1.2.24) ir ekvivalents sakarībai*

$$\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = \gamma - 1, \quad (1.2.26)$$

no kuras varam iegūt, ka izteiksme

$$\lim_{t \rightarrow \infty} \frac{U'(tx)}{U'(t)} = x^{\gamma-1} \quad (1.2.27)$$

ir lokāli vienmērīga intervālā $(0, \infty)$ un visbeidzot

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{tU'(t)} = \frac{x^\gamma - 1}{\gamma},$$

tāpēc pēc Teorēmas 7 var parādīt, ka $F \in D(G_\gamma)$.

Vienkāršāks nosacījums ir iespējams, ja pieņemam, ka $\gamma \neq 0$. Tālāk tiks dota teorēma, kad $\gamma > 0$, kuras pierādījumu skatīt [2, 17–18 lpp.].

Teorēma 11. *($\gamma > 0$) Pieņemsim, ka $x^* = \infty$ un eksistē F' . Ja*

$$\lim_{t \rightarrow \infty} \frac{tF'(t)}{1 - F(t)} = \frac{1}{\gamma} \quad (1.2.28)$$

kādam pozitīvam γ , tad F pieder G_γ definīcijas apgabalam.

Sekas 12. ($\gamma > 0$) Nosacījums (1.2.28) ir ekvivalents sakarībai

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(t)} = \gamma.$$

Tālāk tiks aprakstīti nepieciešamie un pietiekamie nosacījumi uz sadalījuma funkciju F , lai tā varētu izteikt G_γ definīcijas apgabalu. Kā arī tiks parādīts, ka iepriekš aprakstītie nosacījumi ir ļoti līdzīgi nepieciešamajiem un pietiekamajiem nosacījumiem.

Nosakot definīcijas apgabala nosacījumus, sāksim ar nosacījumu (1.2.8) no Teorēmas 5, kur $D(x) = (x^\gamma - 1)/\gamma$. Tas ir

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x) \quad (1.2.29)$$

katram $x > 0$, $\gamma \in \mathbb{R}$ un a ir pozitīva funkcija.

Teorēma 13. Sadalījuma funkcija F atrodas G_γ definīcijas apgabalā tad un tikai tad, ja

1. katram $\gamma > 0$: $x^* = \sup\{x : F(x) < 1\}$ ir galīgs un

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma} \quad (1.2.30)$$

katram $x > 0$. Tas nozīmē, ka funkcija $1 - F(x)$ ir regulāri variējoša bezgalībā ar variācijas indeksu $-1/\gamma$.

2. katram $\gamma < 0$: x^* ir galīgs un

$$\lim_{t \downarrow 0} \frac{1 - F(x^* - tx)}{1 - F(x^* - t)} = x^{-1/\gamma} \quad (1.2.31)$$

katram $x > 0$.

3. katram $\gamma = 0$: x^* var būt galīgs un bezgalīgs un

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x} \quad (1.2.32)$$

katram reālam x , kur f ir derīga pozitīva funkcija.

Teorēma 14. Sadalījuma funkcija F atrodas G_γ definīcijas apgabalā tad un tikai tad, ja katram $\gamma > 0$: $F(x) < 1$ katram x , $\int_1^\infty (1 - F(x))/x \, dx < \infty$, un

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \gamma; \quad (1.2.33)$$

Piezīme 15. Robeža (1.2.33) ir ekvivalenta izteiksmei

$$\lim_{t \rightarrow \infty} E(\log X - \log t | X > t) = \gamma.$$

Faktiski

$$\frac{\int_t^\infty 1 - F(x) \frac{dx}{x}}{1 - F(t)} = E(\log X - \log t | X > t),$$

jo

$$\int_t^\infty \log x - \log t dF(x) = \int_t^\infty 1 - F(x) \frac{dx}{x}.$$

Sakarība (1.2.33) būs par pamatu Hila novērtējuma konstruēšanai parametram γ (skatīt nodaļu 1.5.).

Tālāk parādīsim, kādām jābūt normējošām konstantēm $a_n > 0$, b_n , kuras izmantojām pamata robežsakarībā (1.2.23).

Sekas 16. Ja F atrodas G_γ definīcijas apgabalā, tad katram $\gamma > 0$:

$$\lim_{n \rightarrow \infty} F^n(a_n x) = \exp(-x^{-1/\gamma})$$

spēkā katram $x > 0$, ja $a_n := U(n)$.

Teorēma 17. Sadalījuma funkcija F atrodas ekstremālo vērtību sadalījuma funkcijas G_γ definīcijas apgabalā tad un tikai tad, ja kādai pozitīvai funkcijai f , kuram ir spēkā (1.2.29), izpildās $\lim_{t \rightarrow \infty} f(t)/t = \gamma$,

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \quad (1.2.34)$$

visiem x , kuriem spēkā $1 + \gamma x > 0$. Ja (1.2.34) spēkā kādai $f > 0$, tad tas ir spēkā arī pie $f(t) = \gamma t$, $\gamma > 0$. Bez tam, katram f , pie kura (1.2.34) izpildās, ir spēkā

$$\lim_{t \rightarrow \infty} f(t)/t = \gamma, \quad \gamma > 0. \quad (1.2.35)$$

Lemma 18. Pieņemsim, ka ir spēkā (1.2.29). Ja $\gamma > 0$, tad $\lim_{t \rightarrow \infty} U(t) = \infty$ un

$$\lim_{t \rightarrow \infty} \frac{U(t)}{a(t)} = \frac{1}{\gamma}. \quad (1.2.36)$$

Sekas 19. Katrai $\gamma > 0$ sakarība (1.2.29) ir ekvivalenta

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad x > 0. \quad (1.2.37)$$

Tas nozīmē, ka funkcija U ir regulāri variējoša bezgalībā ar variācijas indeksu γ .

Visu minēto teorēmu un lemmu pierādījumi ir atrodami Haan un Ferreira grāmatā (skatīt [2, 24–33 lpp.]).

1.3. Ekscesa funkcija un Pareto sadalījums

Lai definētu Pareto sadalījumu, no sākuma definēsim papildterminu.

Definīcija 3. (Ekscesa sadalījuma funkcija, Ekscesa vidējā funkcija) Pieņemsim, ka X ir gadījuma lielums no sadalījuma funkcijas F un dots labais galapunkts x^* . Katram fiksētam $t < x^*$,

$$F_t(x) = P(X - t \leq x | X > t), \quad x \geq 0, \quad (1.3.1)$$

sauc par ekscesa sadalījuma funkciju no gadījuma lieluma X (kas ņemts no sadalījuma funkcijas F), kurš pārsniedz vērtību t . Funkciju

$$e(t) = E(X - u | X > u)$$

sauc par ekscesa vidējo funkciju no X .

Ekscesa sadalījuma funkciju var raksturot kā nosacījuma sadalījuma funkciju. X ir vektors no gadījuma lielumiem, t interesējošā robežvērtība. Varbūtība $X - t$ tiek izteikta kā reizinājums no nosacītām varbūtībām un katra nosacītā varbūtība tiek izteikta kā notikums, kur gadījuma lielums X_i pārsniedz robežvērtību t .

Zinot, kas ir ekscesa funkcija, aplūkosit Teorēmas 7 apgalvojumu (1.2.22), kurā eksistē pozitīva, nedilstoša funkcija f tāda, ka

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \quad (1.3.2)$$

katram x , kuram spēkā $1 + \gamma x > 0$, un $x^* = \sup\{x : F(x) < 1\}$. Pieņemsim, ka X ir gadījuma lielums, sadalīts pēc sadalījuma F un $F \in D(G_\gamma)$ kādam reālam γ . Tad vienādojumu (1.3.2) varam pārformulēt ar ekscesa funkcijas palīdzību, kur katram $x \geq 0$, ja $\gamma \geq 0$, un katram $0 \leq x \leq -1/\gamma$, ja $\gamma < 0$, atbildīs

$$\lim_{t \uparrow x^*} P\left(\frac{X - t}{f(t)} > x \mid X > t\right) = (1 + \gamma x)^{-1/\gamma}. \quad (1.3.3)$$

Iegūstam, ka sadalījuma aproksimācija no $(X - t)/f(t)$ pie $X > t$ robežā ir vienāda ar

$$GD_\gamma(x) := 1 - (1 + \gamma x)^{-1/\gamma}, \quad (1.3.4)$$

kad $t \uparrow x^*$. Esam ieguvuši Pareto sadalījumu.

Definīcija 4. Par standarta Pareto sadalījumu sauc sadalījuma funkciju GD_γ , ja

$$GD_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - e^{-x}, & \gamma = 0, \end{cases} \quad (1.3.5)$$

kur

$$\begin{aligned} x &\geq 0, \text{ ja } \gamma \geq 0, \\ -\frac{1}{\gamma} &\leq x \leq 0, \text{ ja } \gamma < 0. \end{aligned}$$

Sakarība (1.3.2) citiem vārdiem sakot parāda, ka pie kādas zināma līmeņa t ($X > t$) sadalījuma funkciju var aproksimēt kā

$$1 - F(x) \approx (1 - F(t)) \left\{ 1 - GD_\gamma \left(\frac{x - t}{f(t)} \right) \right\}, \quad x > t,$$

kas veido astu sadalījumu parametrisku klasi. Mēs sagaidām, ka šī aproksimācija ir spēkā pie vidēja un liela skaita k sakārtotām statistikām.

Lai iegūtu vispārīgo Pareto sadalījumu, x vietā ievieto $(x - \nu)/\beta$, kur normējošās konstantes $\nu \in \mathbb{R}$ un $\beta > 0$. Un iegūstam, ka $GD_\gamma(x)$ ir vienāda ar

$$GD_{\gamma, \nu, \beta}(x) = 1 - \left(1 + \gamma \frac{x - \nu}{\beta} \right)^{-1/\gamma}$$

un

$$\begin{aligned} x &\geq 0, \text{ ja } \gamma \geq 0, \\ x &\in [0, -\beta/\gamma + \nu], \text{ ja } \gamma < 0, \nu > \beta/\gamma. \end{aligned}$$

Kopsavilkumā var secināt, ka ekstremālo vērtību sadalījumi G_γ apraksta robežsadalījumus, kurus iegūst no izlašu normalizētiem maksimumiem. Savukārt Pareto sadalījums raksturo robežsadalījumu no datiem, kas pārsniedz kādu interesējošu lielu vērtību t .

Pareto sadalījuma īpašības ir labi aprakstītas [3, 165.–166. lpp.].

1.4. Ekstremālo vērtību sakārtotās statistikas

Lai varētu nākamajā nodaļā aprēķināt Hila novērtējumu, nepieciešams zināt dažus faktus par sakārtotām statistikām. Ja mums ir doti gadījuma lielumi X_1, \dots, X_n , un mēs sakārtojam šos gadījuma lielumus augošā vai dilstošā secībā, tad iegūstam sakārtotās statistikas. Ja sakārtojam statistikas augošā secībā, tad sakārtotās statistikas parasti apzīmē ar

$$X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}.$$

Īsi aprakstīsim dažus faktus, kas saista sakārtotās statistikas ar ekstremālo vērtību teoriju.

Kā jau iepriekš parādījām, ja izpildās ekstremālo vērtību nosacījums

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \quad (1.4.1)$$

katram x , kur $1 + \gamma x > 0$, vai izpildās ekvivalents nosacījums

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \quad (1.4.2)$$

pie $x > 0$ un $\gamma \in \mathbb{R}$, tad iespējams no normalizētiem izlašu maksimumiem iegūt sadalījumu (1.2.1). Bet mēs varam pateikt vēl vairāk. Ja spēkā nosacījums (1.4.1) vai ekvivalentais nosacījums (1.4.2), tad mēs varam iegūt konverģenci arī citām ekstremālo vērtību statistikām.

Piemērs 5. (Sakārtotās statistikas un eksponenciāli sadalīti gadījuma lielumi)

Pieņemsim, ka E_1, E_2, \dots, E_n ir neatkarīgi un vienādi sadalīti pēc standarta eksponenciālā sadalījuma F . Ar

$$E_{1,n} \leq E_{2,n} \leq \dots \leq E_{n,n}$$

apzīmējam n sakārtotas statistikas. Kopējā blīvuma funkcija no sakārtotām statistikām izsakās kā

$$f_{E_{1,E_2,\dots,E_n}}(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Atzīmēsim faktu, ka no n vērtībām (E_1, E_2, \dots, E_n) var tikt izveidoti $n!$ dažādi sakārtojumi. Tas nozīmē, ka katra sakārtotā izlase $(E_{k,n})_{k=1,\dots,n}$ var tikt iegūta no $n!$ dažādām izlasēm. Tāpēc kopējo blīvuma funkciju no sakārtotās izlases varam pārrakstīt kā

$$\begin{aligned} f_{E_{1,E_2,\dots,E_n}}(x_1, \dots, x_n) &= n! \prod_{i=1}^n f(x_i) = \\ &= n! \prod_{i=1}^n e^{-x_i} = n! \exp \left\{ - \sum_{i=1}^n x_i \right\}, \quad 0 < x_1 < \dots < x_n. \end{aligned}$$

Tālāk ieviesīsim transformāciju T , kur

$$T(x_1, \dots, x_n) = (x_n - x_{n-1}, 2(x_{n-1} - x_{n-2}), \dots, nx_1) \quad 0 < x_1 < \dots < x_n.$$

Tad redzams, ka $(\partial T(x)/\partial x) = n!$ un

$$T^{-1}(x_1, \dots, x_n) = \left(\sum_{j=1}^n \frac{x_j}{n-j+1}, \sum_{j=1}^{n-1} \frac{x_j}{n-j+1}, \dots, \frac{x_1}{n} \right), \quad x_j > 0.$$

Tad iegūstam, ka blīvuma funkciju no $(E_{n,n} - E_{n-1,n}, 2(E_{n-1,n} - E_{n-2,n}), \dots, nE_{1,n})$ var izteikt formā

$$\begin{aligned} g(x_1, \dots, x_n) &= \frac{1}{n!} f_{E_{n,n}, \dots, E_{1,n}} \left(\sum_{j=1}^n \frac{x_j}{n-j+1}, \sum_{j=1}^{n-1} \frac{x_j}{n-j+1}, \dots, \frac{x_1}{n} \right) = \\ &= \exp \left\{ - \sum_{i=1}^n \sum_{j=1}^n \frac{x_j}{n-j+1} \right\} = \\ &= \exp \left\{ - \sum_{i=1}^n x_i \right\}. \end{aligned}$$

Tas parāda, ka

$$E_{n,n} - E_{n-1,n}, E_{n-1,n} - E_{n-2,n}, \dots, E_{1,n}$$

ir neatkarīgi un eksponenciāli sadalīti. $E_{n-i+1,n} - E_{n-i,n}$ pēc sadalījuma tiecas uz

$$(E_{n-i+1,n} - E_{n-i,n})_{i=1, \dots, n} \stackrel{d}{=} ((n-i+1)^{-1} E_i)_{i=1, \dots, n},$$

kur $E_{0,n} = 0$. No tā seko, ka sakārtotās statistikas no izlases ar eksponenciāli sadalītiem gadījuma lielumiem

$$(E_{n-i+1,n})_{i=1, \dots, n} \stackrel{d}{=} \left(\sum_{j=n-i+1}^n j^{-1} E_{n-j+1} \right)_{i=1, \dots, n}.$$

Tālāk izmantojam $E_{i,n}$ statistikas un pārrakstām formā, kuru ir parādījis Renuī (1953) [2], kur

$$(E_{1,n}, E_{2,n}, \dots, E_{k,n}) \stackrel{d}{=} \left(\frac{E_1^*}{n}, \frac{E_1^*}{n} + \frac{E_2^*}{n-1}, \dots, \frac{E_1^*}{n} + \frac{E_2^*}{n-1} + \dots + \frac{E_k^*}{n-k+1} \right)$$

un E_1^*, \dots, E_k^* ir neatkarīgi un vienādi sadalīti pēc standarta eksponenciālā sadalījuma.

Līdz ar to

$$n(E_{1,n}, E_{2,n}, \dots, E_{k,n}) \xrightarrow{d} (E_1^*, E_1^* + E_2^*, \dots, E_1^* + \dots + E_k^*). \quad (1.4.3)$$

Šo rezultātu vispārināsim un attiecināsim uz funkcijām, kuras pieder ekstremālo vērtību sadalījumu definīcijas apgabalam $D(G_\gamma)$.

Teorēma 20. *Pieņemam, ka X_1, X_2, \dots ir neatkarīgi un vienādi sadalīti pēc sadalījuma funkcijas F . Pieņemam, ka F pieder G_γ definīcijas apgabalam kādam $\gamma \in \mathbb{R}$. Ar $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ apzīmējam n sakārtotās statistikas. Tad eksistē normalizējošās konstantes $a_n > 0$ un $b_n \in \mathbb{R}$ no vienādojuma (1.2.1) un fiksēts $k \in \mathbb{N}$, ka*

$$\left(\frac{X_{n,n} - b_n}{a_n}, \frac{X_{n-1,n} - b_n}{a_n}, \dots, \frac{X_{n-k,n} - b_n}{a_n} \right)$$

konverģē uz sadalījumu

$$\left(\frac{(E_1^*)^{-\gamma} - 1}{\gamma}, \frac{(E_1^* + E_2^*)^{-\gamma} - 1}{\gamma}, \dots, \frac{(E_1^* + E_2^* + \dots + E_{k+1}^*)^{-\gamma} - 1}{\gamma} \right),$$

kur E_1^*, \dots, E_k^* ir neatkarīgi un vienādi sadalīti pēc standarta eksponenciālā sadalījuma.

Teorēmas pierādījums atrodas [2, 38. lpp.].

1.4.1. vidējais k

Iepriekš tika apskatītas $X_{n-k,n}$ sakārtotas statistikas, un rezultāts tika pasludināts pie nosacījuma, kad $n \rightarrow \infty$ un k ir fiksēts. Šajā gadījumā nedaudz mainīsim nosacījumu attiecībā uz k . Pieņemsim, ka k ir atkarīgs no apjoma n un rakstīsim $k = k(n) \rightarrow \infty$, kad $n \rightarrow \infty$, un $k(n)/n \rightarrow 0$. Šie nosacījumi nozīmē, ka k tiecas uz bezgalību, bet lēni. Tāpēc arī k iegūst terminu "vidējais k ".

1.5. Astes indeksa γ Hila novērtējums

Pieņemsim, ka X_1, \dots, X_n ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $F \in D(G_\gamma)$, $\gamma > 0$. Tāpēc F asti var izteikt kā regulāri variējošu funkciju $\bar{F}(x) = x^{-1/\gamma}L(x)$ kādam $\gamma > 0$ un $L(x)$ ir lēni variējoša funkcija. Jautājums ir, kā novērtēt γ .

Vienkāršs un plaši pielietots γ novērtējums ir Hila novērtējums.

Astes indeksa γ Hila novērtējums pieņem formu

$$\hat{\gamma} := \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}, \quad (1.5.1)$$

kur $k = k(n) \rightarrow \infty$ un $k/n \rightarrow 0$, kad $n \rightarrow \infty$.

1.5.1. Hila novērtējuma motivācija

Piemērs 6. (Pareto sadalījuma parametru vislielākās ticamības funkcijas novērtējums)

Pieņemsim, ka gadījuma lielums X ir sadalīts pēc Pareto sadalījuma, tad varbūtība, ka X pārsniedz kādu vērtību x ir vienāda ar

$$P(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^\alpha & \text{ja } x \geq x_m, \\ 1 & \text{ja } x < x_m. \end{cases}$$

Ar x_m apzīmē minimālo iespējamo vērtību no X un $\alpha := 1/\gamma > 0$. Pareto sadalījuma funkcija ir atkarīga no diviem parametriem $\alpha := 1/\gamma$, x_m un pieņem formu

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^\alpha & \text{ja } x \geq x_m, \\ 0 & \text{ja } x < x_m. \end{cases}$$

Pareto blīvuma funkcija izsakās kā

$$f_X(x) = \begin{cases} \alpha \frac{x_m^\alpha}{x^{\alpha+1}} & \text{ja } x > x_m, \\ 0 & \text{ja } x < x_m. \end{cases}$$

Vislielākās ticamības funkcija Pareto sadalījuma parametriem α , x_m pie dotas gadījumu izlases $X = (X_1, X_2, \dots, X_n)$ pieņem formu

$$L(\alpha, x_m) = \prod_{i=1}^n \alpha \frac{x_m^\alpha}{X_i^{\alpha+1}} = \alpha^n x_m^{n\alpha} \prod_{i=1}^n \frac{1}{X_i^{\alpha+1}}.$$

Tāpēc logaritmiskā vislielākās ticamības funkcija ir

$$\ell(\alpha, x_m) = n \ln \alpha + n\alpha \ln x_m - (\alpha + 1) \sum_{i=1}^n \ln x_i.$$

Tā kā redzams, ka $\ell(\alpha, x_m)$ monotoni pieaug, ja pieaug vērtība x_m . Tā kā dots sākumā, ka $x \geq x_m$, tad \hat{x}_m novērtējam, kā

$$\hat{x}_m = \min_i X_i.$$

Lai atrastu novērtējumu α , tad mēs atvasinām $\ell(\alpha, x_m)$ pēc α un iegūto izteiksmi pielīdzinām nullei. Tālāk var izteikt novērtējumu $\hat{\alpha}$.

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + n \ln x_m - \sum_{i=1}^n \ln X_i = 0.$$

Vislielākās ticamības α novērtējums ir

$$\hat{\alpha} = \frac{n}{\sum_i (\ln X_i - \ln \hat{x}_m)}.$$

Tālāk ar $X_{1,n} \leq \dots \leq X_{n,n}$ apzīmējam X sakārtotās statistikas. Ja no izlases X mēs ņemam skaitā k vērtības, tad iegūstam, ka

$$\hat{x}_m = \min_i X_i = X_{n-k,n}.$$

Un α vietā ievietojam $1/\gamma$, tad novērtējums $\hat{\gamma}$ būs

$$\hat{\gamma} = \frac{1}{k} \sum_{i=1}^k \left(\ln X_{n-k+i} - \ln \hat{X}_{n-k,n} \right).$$

Iegūtais novērtējums ir tā sauktais Hila novērtējums.

Piemērs 7. (Hila novērtējums ar regulāri variējošās funkcijas tuvinājumu)

Hila novērtējums ir iegūts pārveidojot Teorēmu 13, kurā pieņem, ka $F \in D(G_\gamma)$, katram $\gamma > 0$ tad un tikai tad, ja

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

Tiek izmantota parciālā integrēšana, kur iegūstam, ka

$$\int_t^\infty (\ln x - \ln t) dF(x) = \int_t^\infty \frac{\bar{F}(x)}{x} dx,$$

un pēc Karamata teorēmas [3, 567. lpp.] seko, ka

$$\frac{1}{\bar{F}(t)} \int_t^\infty (\ln x - \ln t) dF(x) \xrightarrow{t \rightarrow \infty} \gamma. \quad (1.5.2)$$

Lai no (1.5.2) iegūtu novērtējumu $\hat{\gamma}$, veicam sekojošus soļus:

(a) F aizvietojam ar empīrisku sadalījuma funkciju

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

(b) t aizvietojam ar $t = X_{n-k,n}$ kādam $k = k(n)$. Mēs varam izvēlēties šādu t , jo zemāk uzrādīta lemma saka, ka $X_{n-k,n} \xrightarrow{g.d.} \infty$, kad $k = k(n) \rightarrow \infty$ un $k/n \rightarrow 0$.

Lemma 21. *Pieņemsim, ka doti Y_1, Y_2, \dots neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $1 - 1/y$, $y \geq 1$, un pieņemsim, ka $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$ ir n sakārtotās statistikas. Tad, ja $k = k(n)$,*

$$\lim_{n \rightarrow \infty} Y_{n-k,n} \xrightarrow{g.d.} \infty, \quad n \rightarrow \infty,$$

pie nosacījuma, ka $k(n) = o(n)$.

Lemmas pierādījumu skatīt [2, 70. lpp.].

Gala rezultātā mēs iegūstam Hila novērtējumu $\hat{\gamma}$, kurš tiek definēts kā

$$\hat{\gamma} := \frac{\int_{X_{n-k,n}}^\infty \log x - \log X_{n-k,n} dF_n(x)}{1 - F_n(X_{n-k,n})} = \frac{1}{k-1} \sum_{j=1}^{k-1} \ln X_{n-j,n} - \ln X_{n-k,n}.$$

Atzīmēsim, ka nav asimptotiski lielas nozīmes, vai mēs summu ņemam līdz k vai $k-1$.

Piemērs 8. (Hila novērtējums kā vidējās ekscesa funkcijas tuvinājums)

Ja pieņemam, ka X ir gadījuma lielums no sadalījuma funkcijas $F \in D(G_\gamma)$, $\gamma > 0$ un $X > 1$, tad (1.5.2) varam pārrakstīt, kā

$$E(\ln X - \ln t | \ln X > \ln t) \xrightarrow{t \rightarrow \infty} \gamma.$$

Apzīmējot $u = \ln t$ un $e^*(u)$ ir vidējā ekscesa funkcija no $\ln X$, tad iegūstam, ka

$$e^*(u) \xrightarrow{u \rightarrow \infty} \gamma.$$

Hila novērtējums ir interpretēts kā $e_n^*(\ln X_{n-k,n})$ - empīriskā vidējā ekscesa funkcija no $\ln X$, kura tiek aprēķināta pie robežvērtības $u = \ln X_{n-k,n}$.

Teorēma 22. Pieņemsim, ka doti X_1, X_2, \dots neatkarīgi un vienādi sadalīti gadījuma lielumi ar kopēju sadalījuma funkciju F . Pieņemsim, ka $F \in D(G_\gamma)$ ar $\gamma > 0$. Tad, ja $n \rightarrow \infty$, $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$, izpildās

$$\hat{\gamma} \xrightarrow{P} \gamma.$$

Pierādījums. Sekas 1.2.34, kur $F \in D(G_\gamma)$ ar $\gamma > 0$ norāda, ka robeža

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma$$

katram $x > 0$. Tas nozīmē, ka šo apgalvojumu varam izteikt ar regulāri variējošas funkcijas īpašību (1.1.2), kur visiem $X \geq 1$ un $t \geq t_0$, spēkā

$$(1 - \varepsilon)x^{\gamma - \varepsilon'} < \frac{U(tx)}{U(t)} < (1 + \varepsilon)x^{\gamma + \varepsilon'},$$

vai ekvivalenti,

$$\log(1 - \varepsilon) + (\gamma - \varepsilon') \log x < \log \left(\frac{U(tx)}{U(t)} \right) < \log(1 + \varepsilon) + (\gamma + \varepsilon') \log x. \quad (1.5.3)$$

Pieņemsim, ka doti Y_1, Y_2, \dots neatkarīgi un vienādi sadalīti gadījuma lielumi ar kopējo sadalījuma funkciju $1 - 1/y$, $y \geq 1$. Ievērosim, ka $U(Y_i) \stackrel{d}{=} X_i$, $i = 1, 2, \dots$

Ir svarīgi pierādīt rezultātu, ka $\hat{\gamma} := k^{-1} \sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n})$. Izmantojam (1.5.3), kur $t = Y_{n-k,n}$, $x = Y_{n-i,n}/Y_{n-k,n}$.

Lemma no [2, 69. lpp.] apgalvo, ka $Y_{n-k,n} \rightarrow \infty$ gandrīz droši, kad $n \rightarrow \infty$. Iegūstam, ka

$$\log(1 - \varepsilon) + (\gamma - \varepsilon') \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) < \log \left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})} \right) < \log(1 + \varepsilon) + (\gamma + \varepsilon') \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right),$$

kur $I = 0, 1, \dots, k-1$. Tāpēc

$$\begin{aligned} \log(1 - \varepsilon) + (\gamma - \varepsilon') \frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) &< \hat{\gamma} < \\ &< \log(1 + \varepsilon) + (\gamma + \varepsilon') \frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right). \end{aligned}$$

Tagad pietiek pierādīt

$$\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) \xrightarrow{P} 1.$$

Atzīmēsim, ka $\log Y_i$ ir sadalīts pēc standarta eksponenciālā sadalījuma. Un nepieciešamo rezultātu pierāda atsevišķa lemma no [2, 71. lpp.]. Lemma apgalvo, ja mums doti E, E_1, E_2, \dots neatkarīgi un vienādi sadalīti gadījuma lielumi, kas ir sadalīti pēc standarta eksponenciālā sadalījuma. Ar $E_{1,n} \leq E_{2,n} \leq \dots \leq E_{n,n}$ apzīmēsim sakārtotās n statistikas. Pieņemsim, ka f ir tāds, ka dispersija $\mathbb{D}f(E) < \infty$. Tad

$$\sqrt{k} \left(\frac{1}{k} \sum_{i=0}^{k-1} f(E_{n-i,n} - E_{n-k,n}) - Ef(E) \right)$$

ir neatkarīgs no $E_{n-k,n}$ un asimptotiski normāls ar vidējo vērtību nulle un dispersiju $f(E)$, kad $n \rightarrow \infty$, nodrošinot $k = k(n) \rightarrow \infty$ un $k/n \rightarrow 0$. \square

Tālāk aprakstīsim nosacījumus, kuri nodrošinās $\hat{\gamma}$ asimptotisko normalitāti.

Teorēma 23. *Pieņemsim, ka sadalījuma funkcija F apmierina nosacījumu, kur katram $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad (1.5.4)$$

vai ekvivalenti var šo nosacījumu uzdot ar

$$\lim_{t \rightarrow \infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma}}{A\left(\frac{1}{1-F(t)}\right)} = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\gamma\rho}, \quad (1.5.5)$$

kur $\gamma > 0$, $\rho \leq 0$, un A ir pozitīva vai negatīva funkcija un robeža $\lim_{t \rightarrow \infty} A(t) = 0$. Tad

$$\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right)$$

tiecas uz standarta normālo sadalījumu, ja $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$, $n \rightarrow \infty$ un

$$\lim_{n \rightarrow \infty} \sqrt{k}A(n/k) = \lambda < \infty. \quad (1.5.6)$$

1.5.2. Situācija, kad Hila novērtējums nestrādā

Pieņemsim, ka X_1, \dots, X_n ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $F \in D(G_\gamma)$, $\gamma > 0$. Tāpēc F asti var izteikt kā regulāri variējošu funkciju $\bar{F}(x) = x^{-1/\gamma}L(x)$ kādam $\gamma > 0$ un $L(x)$ ir lēni variējoša funkcija.

Ja lēni variējošā funkcija $L(x)$ no $\bar{F}(x) = x^{-1/\gamma}L(x)$ pietiekami ātri nekonverģē uz konstanti, varam iegūt nederīgu Hila novērtējumu parametram γ . Sīki tas ir aprakstīts [3, 191.–195. lpp.]. Šeit dosim tikai idejisku izklāstu. Pieņemam apgalvojumu, ka $\hat{\gamma}$ var izteikt kā

$$\hat{\gamma} = \frac{1}{k-1} \sum_{i=1}^{k-1} \ln F^{-1}(U_{n-i,n}) - \ln F^{-1}(U_{n-k,n}), \quad (1.5.7)$$

kur $U_{n-i,n} < \dots < U_{n,n}$ ir sakārtotās statistikas no vienmērīgā sadalījuma $U[0,1]$. Tiek noskaidrots, uz ko tiecas $\hat{\gamma}$ pēc sadalījuma katram fiksētam n . Regulārā variācija no \bar{F} parāda, ka

$$F^{-1}(y) = (1-y)^{-\gamma}L((1-y)^{-1}), \quad y \in (0,1),$$

kādam $L \in RV_0$. Un tālāk izmantojot (1.5.7) un standarta eksponenciālos gadījuma lielumus E_i (no piemēra 5), iegūst, ka $\hat{\gamma}_n$ ir vienāda ar

$$\hat{\gamma}_n = \beta_n^{(1)} + \beta_n^{(2)},$$

kur

$$\beta_n^{(1)} \stackrel{d}{=} -\frac{\gamma}{k-1} \sum_{i=n-k+1}^n \ln U_i \stackrel{d}{=} \frac{\gamma}{k-1} \sum_{i=n-k+1}^n E_i.$$

Pēc SLSL (stiprais lielo skaitļu likums) un CRT (centrālā robežteorēma) seko, ka pie neatkarīgiem un vienādi sadalītiem gadījuma lielumiem

$$\begin{aligned} \beta_n^{(1)} &\xrightarrow{P} \gamma, \\ \frac{\sqrt{k}}{1} (\beta_n^{(1)} - \gamma) &\xrightarrow{d} N(0, \gamma^2), \end{aligned}$$

ja $k = k(n) \rightarrow \infty$.

Savukārt $\beta_n^{(2)}$ caur lēni variējošās funkcijas reprezentāciju izsakās kā

$$\beta_n^{(2)} \sim C_n \frac{\beta_n^{(1)}}{\gamma} \xrightarrow{P} 0,$$

jo $C_n \xrightarrow{g.d.} 0$, kad $k = k(n) \rightarrow \infty$ un $k/n \rightarrow 0$.

Savukārt, ja mēs izvēlamies, ka

$$F^{-1}(y) = (1-y)^{-\gamma}(-\ln(1-y)), \quad y \in (0,1), \quad (1.5.8)$$

tad

$$\begin{aligned} \beta_n^{(2)} &= \dots \stackrel{d}{=} -(1+o(1)) \frac{1}{\ln n} \frac{1}{k-1} \sum_{i=n-k+1}^n \ln U_i \\ &\stackrel{d}{=} (1+o(1)) \frac{1}{\ln n} \frac{1}{k-1} \sum_{i=n-k+1}^n E_i = O((\ln n)^{-1}). \end{aligned}$$

Tas nozīmē, ka $\beta_n^{(2)} \xrightarrow{P} 0$ ar logaritmisku ātrumu. Līdz ar to, ja vēlēsimies konstruēt ticamības intervālus, izmantojot Hila novērtējumu, mums būs jāņem vērā $\beta_n^{(2)}$, piemērojot normējošās konstantes. Normējošās konstantes būs atkarīgas no (parasti nezināmās) lēni variējošās funkcijas L un iegūstam, ka mēs nevaram pielietot standarta centrālo robežteorēmu novērtējumam $\hat{\gamma}$.

Zemāk ir parādīti $\hat{\gamma}$ novērtējumi, ja izvēlamies F^{-1} no (1.5.8).

Piezīme 24. *Lai iegūtu labu novērtējumu, kad $\hat{\gamma} \rightarrow \gamma$, nepieciešama kopa no nepieciešamiem nosacījumiem uz \bar{F} un $k(n)$. Tāpēc tiek ieviesti nosacījumi, kas nodrošina $\hat{\gamma}$ asimptotisko normalitāti.*

Piezīme 25. *Hila novērtējumu vienmēr būtu ieteicams analizēt grafiski. Hila-attēlu veido*

$$\{(k, \hat{\gamma}_{k,n}) : k = 2, \dots, n\}.$$

Hila-attēls ir viens no instrumentiem kā noteikt optimālo k . Nākamajā apakšnodaļā aprakstīsim, kā izvēlēties optimālo k , lai aprēķinātu $\hat{\gamma}_{k,n}$.

1.5.3. Optimālā $k = k(n)$ aprēķināšana

Pieņemsim, ka X_1, \dots, X_n ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $F \in D(G_\gamma)$, $\gamma > 0$. Ja pieņemam, ka F asti varam izteikt kā

$$1 - F(x) = C_1 x^{-1/\gamma} + C_2 x^{-1/\gamma + \rho/\gamma} (1 + o(1)), \quad x \rightarrow \infty \quad (1.5.9)$$

pie $C_1 > 0$, $C_2 \neq 0$, $\gamma > 0$ un $\rho < 0$, tad ir spēkā Teorēmas 23 otrās kārtas pieņēmums (1.5.5). Parametrs ρ kontrolē $\hat{\gamma}$ asimptotiskās normalitātes konverģences ātrumu.

Ja mēs pieņemam, ka F aste pieņem formu (1.5.9), tad nosacījumā (1.5.5) funkcija $A(t)$ ir vienāda ar

$$A(t) = \rho \gamma^{-1} C_2 C_1^{\rho-1} t^\rho := C t^\rho. \quad (1.5.10)$$

Tā kā teorēma 23 saka, ja $\sqrt{k}A(n/k) \rightarrow \lambda$, tad

$$\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{d} \gamma N(0, 1) + \frac{\lambda}{1 - \rho}.$$

Līdz ar to iegūstam aproksimāciju

$$\hat{\gamma} - \gamma \stackrel{d}{\approx} \frac{\gamma N(0, 1)}{\sqrt{k}} + \frac{\lambda}{(1 - \rho)\sqrt{k}} \stackrel{d}{\approx} \frac{\gamma N(0, 1)}{\sqrt{k}} + \frac{A(n/k)}{(1 - \rho)}.$$

Un tagad vēlamies uzzināt pie kura k šī aproksimācija ir vislabākā. Tiek meklēts risinājums uzdevumam

$$\arg \min_k \left(\frac{\gamma^2}{k} + \frac{A^2(n/k)}{(1-\rho)^2} \right).$$

Tā kā pieņēmām, ka $A(t) = Ct^\rho$ un vienkāršākam pierakstam pieņemam, ka $t := n/k$, tad ir jāminimizē

$$\arg \min_{t>0} \left(\frac{t\gamma^2}{n} + \frac{C^2 t^{2\rho}}{(1-\rho)^2} \right).$$

Vienādojumu atvasinām pēc t un pielīdzinām nullei. Iegūstam, ka

$$k = \left[\left(\frac{\gamma^2(1-\rho)^2}{-2\rho C^2} \right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)} \right], \quad (1.5.11)$$

kur ar $[x]$ apzīmējam veselo daļu no x .

1.6. Hila novērtējuma ticamības intervālu novērtēšanas metodes

1.6.1. Normālā aproksimācija

Ticamības intervālu ar normālās aproksimācijas metodi parādīja Cheng un Peng 2001.gadā [10]. Šajā darbā tiek aplūkoti ticamības intervāli astes indeksa γ Hila novērtējumam, ka arī pie kādiem nosacījumiem tiek izvēlēts optimālais k un tiek sasniegta pārklājuma precizitāte $1 - \alpha$.

Pieņemsim, ka X_1, \dots, X_n ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $F(x)$, un pie $\gamma > 0$, $c \in \mathbb{R}$ F asti var izteikt kā regulāri variējošu funkciju

$$1 - F(x|\gamma, c) = cx^{-1/\gamma}$$

katram $x > 0$. Ar $X_{1,n} \leq \dots \leq X_{n,n}$ apzīmējam X_1, \dots, X_n sakārtotās statistikas. Tāpat kā pie optimālā k meklēšanas, pieņemam, ka F asti varam izteikt kā (1.5.9), t.i.,

$$1 - F(x) = C_1 x^{-1/\gamma} + C_2 x^{-1/\gamma + \rho/\gamma} (1 + o(1)), \quad x \rightarrow \infty \quad (1.6.1)$$

pie $C_1 > 0$, $C_2 \neq 0$, $\gamma > 0$ un $\rho < 0$. Tad ir spēkā Teorēmas 23 otrās kārtas pieņēmums (1.5.5).

Pieņēmums 1. Ja spēkā pieņēmums (1.5.9) un $k \rightarrow \infty$, $k/n \rightarrow 0$. Tad

$$\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \gamma^2),$$

tad un tikai tad, ja $k = o(n^{2(1-\rho)/(1+2(1-\rho))})$.

Ja pieņēmums 1 ir spēkā, tad astes indeksa γ Hila novērtējuma divpusējais ticamības intervāls, kas ir izteikts ar normālo aproksimāciju pie nozīmības līmeņa $1 - \alpha$ ir vienāds ar

$$I_N(1 - \alpha) = \left(\hat{\gamma}_n - \frac{z_\alpha \hat{\gamma}_n}{\sqrt{k}}, \hat{\gamma}_n + \frac{z_\alpha \hat{\gamma}_n}{\sqrt{k}} \right),$$

kur ar z_α apzīmē kvantili no $P(|N(0, 1)| \leq z_\alpha) = 1 - \alpha$.

1.6.2. Vislielākās ticamības metode

Ticamības intervālu konstruēšana Hila novērtējumam ar vislielākās ticamības metodi parādīja Lu un Peng 2002.gadā (skatīt [11, 340.–341. lpp.]). Pieņemsim, ka X_1, \dots, X_n ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $F(x|\gamma, c)$, kur $\gamma > 0$, $c \in \mathbb{R}$ un F asti var izteikt kā regulāri variējošu funkciju

$$1 - F(x|\gamma, c) = cx^{-1/\gamma},$$

kad $x \leq X_{n-k,n}$. Ar $X_{n-k,n}$ apzīmējam (X_i) sakārtotās statistikas. Vislielākās ticamības funkcija tiek definēta kā

$$L_P(\gamma, c) = \prod_{i=1}^n f(X_i|\gamma, c).$$

Tā kā mēs Hila novērtējuma novērtēšanai izmantojam tikai augšējās k statistikas, tad $L_1(\gamma, c)$ izsakās kā kopējā sadalījuma funkcija no $(X_{n-k,n}, \dots, X_{n,n})$ un tā pieņem formu

$$L_1(\gamma, c) = \frac{(n-k-1)!}{n!} \prod_{i=1}^{k+1} \left\{ \frac{c}{\gamma} X_{n-i+1,n}^{-1/\gamma-1} \right\} \left[1 - cX_{n-k,n}^{-1/\gamma} \right]^{n-k-1}.$$

Vislielākās ticamības metode novērtē parametru c , atrodot tādu \hat{c} , kas maksimizē $L_1(\gamma)$.

Iegūstam funkciju $L_2(\gamma)$

$$\begin{aligned} L_2(\gamma) &= \arg \max_c L_1(\gamma, c) = \\ &= \left(\frac{(n-k-1)!}{n!} \right) \left[1 - \frac{k+1}{n} \right]^{n-k-1} \prod_{i=1}^{k+1} \left\{ \frac{k+1}{n\gamma} X_{n-k,n}^{1/\gamma} X_{n-i+1,n}^{-1/\gamma-1} \right\}. \end{aligned}$$

Tālāk funkciju $L_2(\gamma)$ maksimizējam pēc γ un iegūstam, ka

$$\hat{\gamma}_n = \arg \max_{\gamma} L_2(\gamma).$$

Iegūstam, ka

$$\hat{\gamma}_n = \frac{1}{k+1} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}}.$$

Logaritmiskā vislielākās ticamības attiecības testa statistika [12], pareizināta ar -2 , ir

$$R_P(\gamma) = -2 \log \frac{L_2(\gamma)}{L_2(\hat{\gamma}_n)} = 2(k+1) \left\{ \frac{\hat{\gamma}_n}{\gamma} - 1 - \log \frac{\hat{\gamma}_n}{\gamma} \right\}.$$

Tālāk sekojošā teorēma ir balstīta uz Hila novērtējuma asimptotisko normalitāti.

Teorēma 26. *Pieņemsim, ka kādam $\rho < 0$ un $k \rightarrow \infty$, $k/n \rightarrow 0$, kad $n \rightarrow \infty$, ir spēkā Hila novērtējuma otrās kārtas nosacījums*

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(t) - x^\rho}{A(t)} = x^\rho \frac{x^\rho - 1}{\rho},$$

kur $U(t)$ ir inversā funkcija no $1/(1 - F(t))$. Un funkcija $A(t) \rightarrow 0$, kad $t \rightarrow \infty$. Tad $R(\gamma)$ asimptotiski tiecas uz χ_1^2 sadalījumu, ja $\sqrt{k}A(n/k) \rightarrow 0$ pie $n \rightarrow \infty$.

No šīs teorēmas tiek konstruēts Hila novērtējuma ticamības intervāls ar nozīmības robežu $1 - \alpha$. Astes indeksa ticamības intervālu atrod no

$$I_P(\alpha) = \{\gamma : R(\gamma) \leq c_\alpha\},$$

kur c_α atbilst $P(\chi_1^2 \leq c_\alpha) = 1 - \alpha$. Iegūsim astes indeksa γ ticamības intervālu, kurš asimptotiski atbilst pārklājuma precizitātei $1 - \alpha$.

1.6.3. Neparimetriskā ticamības metode

Ticamības intervālu konstruēšana ar neparimetrisko vislielākās ticamības metodi pirmais iepazīstināja Owen (1988, 1990) skatīt [13]. Savukārt šo metodi Lu un Peng (2002) izmanto, lai konstruētu ticamības intervālus astes indeksa γ Hila novērtējumam [11, 339.–340. lpp.].

Pieņemsim, ka X_1, \dots, X_n ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju $F(\gamma)$. Tālāk, definējam Y kā naturālo logaritmu attiecību no $n - i + 1$ un $n - i$ -tās sakārtotās statistikas, ko var pierakstīt: $Y_i = i \log(X_{n,n-i+1}/X_{n,n-i})$ katram $i = 1, \dots, k$. Var parādīt (skatīt [14, 812.–815. lpp.]) ka katram fiksētam k , kad $n \rightarrow \infty$

$$\{Y_i, 1 \leq i \leq k\} \rightarrow \{\gamma w_i, 1 \leq i \leq k\},$$

kur $\{w_i, 1 \leq i \leq k\}$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar eksponenciālo sadalījuma funkciju un vidējo vērtību 1. Citiem vārdiem var teikt, ka Hila novērtējums $\hat{\gamma}_n$ asimptotiski ir izlases no gadījuma lielumiem $\{w_i, 1 \leq i \leq k\}$ vidējā vērtība, kur k ir fiksēts.

Mūsu uzdevums ir uzbūvēt ticamības intervālus Hila novērtējumam $\hat{\gamma}$.

Pieņemsim, ka $p = (p_1, \dots, p_k)$ ir punktveida masas varbūtību vektors, kurus iegūst no pētāmās datu kopas punktiem $\{Y_i, 1 \leq i \leq k\}$ un pieņemsim, ka $\sum_{i=1}^k p_i = 1$ un $p_i \geq 0$. Tad empīriskā ticamības funkcija parametram γ ir

$$L_E(\gamma) = \sup_p \prod_{i=1}^k p_i, \quad (1.6.2)$$

ar ierobežojumu $\sum_{i=1}^k p_i Y_i = \gamma$ [15]. Attiecīgi logaritmisko empīriskās ticamības attiecību definējam ar $R_E(\gamma) = -2 \log(k^k L_E(\gamma))$. Izteiksmes (1.6.2) maksimumu var atrisināt ar Lagranža reizinātāju metodi.

Funkciju $L_E(\gamma) = \prod_{i=1}^k p_i$ maksimizē pie ierobežojumiem

$$p_i \geq 0, \quad \sum_{i=1}^k p_i = 1, \quad \sum_{i=1}^k p_i (Y_i - \gamma) = 0$$

Dotam γ eksistē viens vienīgs atrisinājums, ja 0 ir izliektajā čaulā, kura sastāv no punktiem $(Y_1 - \gamma), \dots, (Y_n - \gamma)$.

$$H = \sum_{i=1}^k \log(p_i) + \lambda_0 \left(1 - \sum_{i=1}^k p_i\right) - n\lambda \sum_{i=1}^k p_i (Y_i - \gamma), \quad (1.6.3)$$

kur λ_0, λ ir Lagranža reizinātāji. Atvasinot izteiksmi (1.6.3) pēc p_i , iegūstam, ka

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \lambda_0 - k\lambda(Y_i - \gamma) = 0.$$

Tālāk redzams, ka

$$\begin{aligned} \sum_{i=1}^k p_i \frac{\partial H}{\partial p_i} &= \sum_{i=1}^k \frac{p_i}{p_i} - \sum_{i=1}^k p_i \lambda_0 - \sum_{i=1}^k k p_i \lambda (Y_i - \gamma) = 0 \\ \sum_{i=1}^k p_i \frac{\partial H}{\partial p_i} &= k - \lambda_0 = 0 \Rightarrow \lambda_0 = k \\ \frac{1}{p_i} &= k + k\lambda(Y_i - \gamma) \end{aligned}$$

un iegūstam, ka

$$p_i = \frac{1}{k} \frac{1}{1 + \lambda(Y_i - \gamma)},$$

Ņemot vērā ierobežojumu $\sum_{i=1}^k p_i (Y_i - \gamma) = 0$ un tajā ievietojot p_i , iegūstam, ka

$$\sum_{i=1}^k \frac{Y_i - \gamma}{1 + \lambda(Y_i - \gamma)} = 0. \quad (1.6.4)$$

No šīs izteiksmes var izteikt λ katram uzdotam γ . Empīriskā ticamības funkcija parametram γ tagad tiek definēta kā

$$L_E(\gamma) = \prod_{i=1}^k \left\{ \frac{1}{k} \frac{1}{1 + \lambda(\gamma)(Y_i - \gamma)} \right\}.$$

Logaritmiskā empīriskās ticamības funkcijas attiecība tiek definēta sekojoši:

$$R_E(\gamma) = 2 \sum_{i=1}^k \log\{1 + \lambda(\gamma)(Y_i - \gamma)\}.$$

Teorēma 27. *Ja ar $U(x)$ apzīmējam inverso funkciju lielumam $1/(1 - F(x))$ un pieņemam, ka eksistē funkcija $A(t) \rightarrow 0$, kad $t \rightarrow \infty$, ka*

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},$$

ir spēkā kādam $\rho < 0$, un $k \rightarrow \infty$, $k/n \rightarrow 0$, kad $n \rightarrow \infty$. Tad $R_E(\gamma)$ asimptotiski tiecas uz χ_1^2 sadalījumu, ja $\sqrt{k}A(n/k) \rightarrow 0$, kad $n \rightarrow \infty$.

Teorēmas pierādījums ir parādīts [11].

Teorēma 27 ir Vilksa (Wilks) Teorēmas neparametriskā versija. Vilksa Teorēma nosaka, ka logaritmiskā vislielākās ticamības attiecība $-2 \ln \lambda_n \sim \chi_r^2$. Teorēma ļauj γ novērtējumam konstruēt ticamības intervālus ar nozīmības līmeni α . Logaritmiskā attiecības testa statistikai ir χ_1^2 sadalījums un empīriskās ticamības konverģences apgabals

$$I_E(1 - \alpha) = \{\gamma : R_E(\gamma) \leq c_{1-\alpha}\},$$

asimptotiski ir ar uzdotu līmeni $c_{1-\alpha}$, kur $c_{1-\alpha}$ ir izvēlēts kā $P(\chi_1^2 \leq c_{1-\alpha}) = 1 - \alpha$. No tā tiks iegūts ticamības intervāls parametram γ ar asimptotisku $1 - \alpha$ pārklājuma precizitāti, kuru aprakstīsim zemāk sekās.

Sekas 28. *Ja*

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(x) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},$$

ir spēkā kādam $\rho < 0$, un $k \rightarrow \infty$, $k/n \rightarrow 0$, kad $n \rightarrow \infty$ un $\sqrt{k}A(n/k) \rightarrow 0$, kad $n \rightarrow \infty$.

Tad, ja $n \rightarrow \infty$

$$P(\gamma \in I_E(1 - \alpha)) = 1 - \alpha + o(1).$$

1.6.4. Butstrapa metode

Par butstraba ticamības intervālu konstruēšanu smagās astes indeksam γ informācija iegūta no Guillou [16], Caers un Van Duck [17], un Qi [18].

Pieņemsim, ka mums ir dota izlase (X_1, \dots, X_n) no neatkarīgiem un vienādi sadalītiem gadījuma lielumiem sadalītiem pēc nezināmās sadalījuma funkcijas F . Ar $X_{1,n} \leq \dots \leq X_{n,n}$ apzīmējam izlases $\{X\}$ sakārtotās statistikas. Mērķis ir noteikt, kā uzvedās sadalījuma aste, tas ir, kādu sadalījumu pieņem lielākās sakārtotās statistikas.

Pieņemam, ka izlašu maksimumi $X_{n,n}$ veido Frešē ekstremālo vērtību sadalījuma klasi, tas ir, astes variācijas indekss $\gamma > 0$. Sadalījuma aste ir regulāri variējoša funkcija

$$1 - F(x) = x^{-1/\gamma} L(x), \quad x > 0,$$

kur $L(x)$ ir lēni variējoša funkcija. Parametra γ novērtēšanai izmantojam Hila novērtējumu, kur izmantojam k vidējās statistikas un

$$\bar{\gamma}_n(k) = \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1} - \log X_{n-k,n}.$$

Lai gan butstrapa metodi kā neparametrisko metodi plaši izmanto daudzās jomās, ekstremālo vērtību teorijā tai ir ierobežotas iespējas. Butstraps ar pilno izlases pārlasi nedarbojas ekstremālo vērtību teorijā. Hall [19] 1990.gadā norādīja, ka pilnās pārlases butstrapa statistikām nav novirzes, ja statistiku var izteikt kā datu lineāru kombināciju. Tāpēc pilnās pārlases butstrapa statistika var būt nederīga, ja īstā statistika ir ar ievērojamu novirzi. Kā risinājums tiek piedāvāts izmantot butstrapa izlases bez atkārtojumiem un tāpēc butstrapa izlases apjomu ņemt mazāku par sākotnējās izlases apjomu n . Tā kā mēs izmantosim k vidējās statistikas, tad Qi [18] norāda, ka šajā gadījumā butstraps strādā tikai tad, ja tiek ņemtas izlases bez atkārtojumiem.

Ticamības intervālu konstruēšanā butstrapa metode ar nepilno pārlasi un bez atkārtojumiem ietver sevī sekojošu algoritmu:

1. Uzģenerē vai iegūst datu izlasi $\{X\}$ ar apjomu n . Aprēķina izlases Hila novērtējumu pie uzdota uzdota līmeņa k .
2. Ģenerē skaitā B butstrapa izlases $\{X^*\}$. Izlases $\{X^*\}$ apjoms $n_1 = \sqrt{n}$. Pieņēmumu par to, ka jāņem šāds butstrapotās izlases apjoms n_1 , neviens nav pierādījis, bet tikai simulāciju ceļā radies rezultāts, ko parāda savā rakstā Caers un Van Duck [17].

3. Katrai izlasei aprēķina butstrapoto Hila novērtējumu $\hat{\gamma}^*$ pie līmeņa k_1 . k_1 ir saistīts ar k caur sakarību

$$k = k_1 \left(\frac{n}{n_1} \right)^\beta,$$

kur $\beta = 2/3$. Šādu β vērtību pieņem Hall [19] un tā ir iegūta no aprēķiniem, kur tika piemeklēts k_{opt} pie pieņēmumiem, ka $\gamma > 0$ un $\{X\}$ sadalījuma aste konverģē uz Pareto sadalījumu. Pie līmeņa k_{opt} novērtējums $\hat{\gamma}$ ir ar vismazāko vidējo kvadrātisko kļūdu.

4. Sakārtojam iegūtos butstrapotos novērtējumus $\hat{\gamma}^*$ augošā secībā un ar butstrapa procentīļu intervāliem nosakām ticamības intervālu parametram γ pie nozīmības līmeņa α ,

$$I_B = (\hat{\gamma}_{\alpha/2}^*; \hat{\gamma}_{1-\alpha/2}^*).$$

1.7. Ticamības intervālu pārklājuma precizitāte

Lai noteiktu, kura no ticamības intervālu konstruēšanas metodēm ir labāka, tiek noteikta pārklājuma precizitāte. Mūsu mērķis ir ne tikai noteikt metožu pārklājuma precizitāti, bet ieviest arī korekcijas, kuras var pārklājuma precizitāti uzlabot.

Iepriekšējās nodaļās tika aprakstītas vairākas metodes, kuras aprēķina Hila novērtējuma ticamības intervālus.

Normālās aproksimācijas ticamības intervāls

$$I_N(\alpha, k) = [(1 - k^{-1/2} z_{\alpha/2}) \hat{\gamma}, (1 + k^{-1/2} z_{\alpha/2}) \hat{\gamma}],$$

balstās uz vislielākās ticamības novērtējuma normālā sadalījuma ierobežošanu. Ar α apzīmējam kļūdas līmeni un ticamības intervāla pārklājuma precizitāte atbilst $1 - \alpha$. z_α ir kvantile no standarta normālā sadalījuma Φ , kur $\Phi(z_\alpha) = 1 - \alpha$.

Vislielākās ticamības attiecības statistikas ticamības intervāls

$$I_P(\alpha, k) = \{0 < \gamma < \infty : R_P(\gamma, k) \leq c_\alpha\},$$

kur c_α atbilst $P(\chi_1^2 \leq c_\alpha) = 1 - \alpha$, balstās uz to, ka logaritmiskais vislielākās ticamības attiecības statistika ir sadalīta pēc χ_1^2 sadalījuma.

Empīriskais ticamības funkcijas attiecības ticamības intervāls ir

$$I_E(\alpha, k) = \{\gamma : R_E(\gamma, k) \leq c_\alpha\},$$

ir ar uzdotu līmeni c_α , kur $P(\chi_1^2 \leq c_\alpha) = 1 - \alpha$.

Ja ticamības intervālu konstruēšanai izmanto parametrisko vai neparametrisko vislielākās ticamības attiecību statistiku, tad $R_n(\gamma, k)$ un $l(\gamma, k)$ asimptotiskā sadalījuma konverģences ātrums ir $O(k^{-1/2})$, kur k ir sakārtoto statistiku skaits. Uzlabot konverģences ātrumu var ar Bartleta korekciju, kur konverģences kļūda tiek samazināta līdz $O(k^{-3/2})$.

Vislielākās ticamības attiecības statistika ar Bartleta korekciju dod ticamības intervālu

$$I_{PB}(\alpha, k) = \{0 < \gamma < \infty : R_P(\gamma, k)/(1 + (6k)^{-1}) \leq c_\alpha\}.$$

Vislielākās ticamības attiecības statistika tika izdalīta ar tās asimptotisko vidējo vērtību, ko arī sauc par Bartleta korekcija.

Empīriskais ticamības funkcijas attiecības ticamības statistika ar Bartleta korekciju ir

$$I_{EB}(\alpha, k) = \{\gamma : R_E(\gamma, k)/(1 + a/k) \leq c_\alpha\},$$

kur a ir

$$a = \left(-\frac{1}{3} \frac{\mu_3^2}{\mu_2^3} + \frac{1}{2} \frac{\mu_4}{\mu_2^2} \right),$$

un $\mu_j = E(X - \bar{X})^j$.

Tālāk nodaļās tiks sīkāk izklāstīta Bartleta korekcija un tās pielietojums parametriskajai un neparametriskajai ticamības funkcijai, kuras savukārt ir pielāgotas Hila novērtējuma ticamības intervālu aprēķināšanai.

1.7.1. Bartleta korekcija vislielākās ticamības funkcijai

Vispārīgu ieskatu kā Bartlet korekciju var piemērot vislielākās ticamības funkcijai aprakstīja Barndorff un Cox 1984.gadā (skatīt [4]). McCullagh un Cox 1986.gadā parādīja, ka Bartlet korekciju var izteikt ar logaritmisko attiecību pirmo divu atvasinājumu kumulantu invariantajām kombinācijām (skatīt [20]). Savukārt Cordeiro 1995.gadā publicē vienkāršāku un pievilcīgāku Bartlet korekcijas izteiksmi viena parametra eksponenciālo modeļu klasei (skatīt [21]). Haeusler un Segers 2007.gadā (skatīt [22]) sīkāk izanalizēja Hila novērtējuma ticamības intervālu, ja to aprēķina ar vislielākās ticamības funkciju (ar un bez Bartleta korekcijas).

Kā jau agrāk tika parādīts, Hila novērtējumu var dabiskā veidā iegūt kā ticamības funkcijas novērtējumu no Pareto sadalījuma.

Logaritmiskā vislielākās ticamības attiecības testa statistika, pareizināta ar -2 , ir

$$R_P(\gamma, k) = 2\{l_P(\hat{\gamma}, k) - l_P(\gamma, k)\}.$$

Logaritmiskā vislielākās ticamības funkcija izsakās kā

$$\begin{aligned} l_P(\gamma, k) &= -k \ln \gamma + \frac{k}{\gamma} \ln X_{n-k,n} - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^k \ln X_{n-k+i} = \\ &= -k \left(\frac{\hat{\gamma}}{\gamma} + \ln \gamma + \sum_{i=1}^k \ln X_i \right), \end{aligned}$$

kur

$$\hat{\gamma} = \frac{1}{k} \sum_{i=1}^k k \left(\ln X_{n-k+i} - \ln \hat{X}_{n-k,n} \right).$$

Tad iegūstam, ka

$$R_P(\gamma, k) = 2k \left(\frac{\hat{\gamma}_n(k)}{\gamma} - 1 - \ln \frac{\gamma_n(k)}{\gamma} \right).$$

Pēc Bartleta pieņēmuma, ja spēkā

$$E(R_P(\gamma, k)) = 1 + \frac{b}{k} + O(k^{-3/2}),$$

tad iegūstam koriģēto logaritmisko vislielākās ticamības funkciju

$$R'_P(\gamma, k) = w_n(\gamma, k)(1 + b/k)^{-1}.$$

$R'_P(\gamma, k)$ konverģēs uz χ_1^2 sadalījumu ar kļūdu $O(n^{-3/2})$.

$E(R_P(\gamma, k))$ sauc par Bartlet korekciju. Cordeiro savā darbā ([21]) pierāda, ka $E(R_P(\gamma, k))$ eksponenciālo sadalījumu klasē iegūst vienkāršu izteiksmi

$$E(R_P(\gamma, k)) = 1 + \frac{\rho(\gamma)}{12k}.$$

$\rho(\gamma)$ pieņem formu

$$\rho(\gamma) := -\frac{4\beta'^2\alpha''^2 + \alpha'\beta'\alpha''\beta'' - 5\alpha'^2\beta''^2 - 3\alpha'\beta'^2\alpha''' + 3\alpha'^2\beta'\beta'''}{(\alpha'\beta')^3}.$$

Ja aplūkojam Pareto sadalījumu un spēkā nosacījumi, ka $\gamma > 0$, $k > 0$ un k ir zināms, tad, lai novērtētu Hila novērtējuma Bartlet korekciju, funkcijas $\alpha(\gamma)$, $\beta(\gamma)$ ir (skatīt [21]).

$$\begin{aligned} \alpha(\gamma) &:= \gamma + 1, \\ \xi(\gamma) &:= \frac{1}{\gamma k^\gamma}, \\ \beta(\gamma) &:= \frac{\xi'}{\xi \alpha'}. \end{aligned}$$

Piezīme 29. Var parādīt, ja $\alpha(\gamma) = c_1\gamma + c_2$ pieņem lineāru formu un $\xi(\gamma) = c_3/(\gamma c_4^\gamma)$, kur c_1, c_2, c_3, c_4 ir konstantes, tad $\rho(\gamma) = 2$.

Pierādījums. Ja $\alpha(\gamma) = c_1\gamma + c_2$, tad $\alpha' = c_1$ un pārējie atvasinājumi ir vienādi ar nulli. Iegūstam, ka $\rho(\gamma)$ vienkāršojas līdz

$$\rho(\gamma) = -\frac{-5\beta'^2 + 3\beta'\beta'''}{c_1\beta'^3}.$$

Tālāk izvēršam $\beta(\gamma)$.

$$\begin{aligned}\beta(\gamma) &= \frac{\xi'}{\xi\alpha'} = \left(-\frac{c_3 \log c_4}{c_4^\gamma \gamma} - \frac{c_3}{c_4^\gamma \gamma^2} \right) \frac{\gamma c_4^\gamma}{c_1 c_3} = \\ &= -\frac{\log k}{c_1} - \frac{1}{\gamma c_1}, \\ \beta' &= \frac{1}{\gamma^2 c_1}, \quad \beta'' = -\frac{2}{\gamma^3 c_1}, \quad \beta''' = \frac{6}{\gamma^4 c_1}.\end{aligned}$$

Līdz ar to iegūstam, ka

$$\begin{aligned}\rho(\gamma) &= -\frac{-5\beta'^2 + 3\beta'\beta'''}{c_1\beta'^3} = \\ &= -\frac{-5\left(-\frac{2}{\gamma^3 c_1}\right)^2 + 3\left(\frac{1}{\gamma^2 c_1}\right)\frac{6}{\gamma^4 c_1}}{c_1\frac{1}{\gamma^2 c_1}^3} = \\ &= \frac{\frac{20}{\gamma^6 c_1^2} - \frac{18}{\gamma^6 c_1^2}}{\frac{c_1}{\gamma^6 c_1^3}} = 2.\end{aligned}$$

□

No piezīmes 29 redzams, ka Bartlet korekcija Hila novērtējumam ir $E(R_P(\gamma, k)) = 1 + 1/(6k)$ un koriģētā logaritmiskā vislielākās ticamības funkcija ir

$$R'_P(\gamma, k) = R_P(\gamma, k)(1 + 1/(6k))^{-1}.$$

1.7.2. Bartleta korekcija empīriskai ticamības funkcijai

Pirmais, kurš piemēroja Bartleta korekciju empīriskai ticamības funkcijai, bija Diccio (1991) (skatīt [5]). Šajā darbā tiks analizēts, vai Bartleta korekciju var pielietot arī ekstrēmu sadalījumu gadījumā. Līdz šim literatūrā Bartleta korekcija nav tikusi pielietota ekstrēmu sadalījumu gadījumos.

Savā darbā Diccio apraksta kā tiek aprēķināta Bartleta korekcija q dimensionālam parametram $\theta = (\theta_1, \dots, \theta_q)$ no r -dimensiju ($q \leq r$) sadalījumam F_0 , kuram eksistē vidējā vērtība μ_0 un nesingulāra kovariāciju matrica Σ_0 . Svarīgs nosacījums ir arī tas,

ka parametrs θ ir jāspēj izteikt kā funkciju no vidējās vērtības μ_0 un iegūto novērtējumu apzīmējam ar $\theta_0 = \theta(\mu_0)$. Tad ir iegūts gludās funkcijas modelis, kuru aprakstīja Bhattacharya un Ghosh [23], un savā darbā pielieto Diciccio.

Tā kā šajā darbā ir jānovērtē tikai viens parametrs γ , tad mūsu gadījumā $q = 1$. Lielumu r nosaka tas, cik dažādas vidējās vērtības ir nepieciešams pielietot, lai varētu caur šīm vidējām vērtībām izteikt mūs interesējošo parametru γ . Mūsu gadījumā $r = 1$. To var pamatot ar sekojošu algoritmu.

Ja mums ir dota izlase no gadījuma lielumiem X_1, X_2, \dots, X_n , kuri ir sadalīti pēc nezināmā sadalījuma F_0 . Tālāk mēs sakārtojam izlasi augošā secībā un sakārtotās statistikas apzīmējam ar

$$X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}.$$

Tā kā mūs interesē izpētīt tikai sadalījuma asti, tad izmantojam tikai pēdējās k sakārtotās statistikas. Veidojam jaunus gadījuma lielumus Y_i , kur

$$Y_i = i(\log X_{n-i+1,n} - \log X_{n-i,n}), \quad i = 1, \dots, k.$$

Apskatot izlases vidējo vērtību no lielumiem Y_i , tad iegūsim, ka

$$\begin{aligned} \bar{Y} &= \frac{1}{k} \sum_{i=1}^k Y_i = \\ &= \frac{1}{k} \sum_{i=1}^k i(\log X_{n-i+1,n} - \log X_{n-i,n}) = \\ &= \frac{1}{k} (\log X_{n,n} - \log X_{n-1,n} + 2 \log X_{n-1,n} - \dots + k \log X_{n-k+1,n} - k \log X_{n-k,n}) = \\ &= \frac{1}{k} (\log X_{n+1,n} + \log X_{n+2,n} + \dots + \log X_{n+k,n} - k \log X_{n-k,n}) = \\ &= \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X_{n-i+1}}{X_{n-k,n}} \right) = \hat{\gamma}. \end{aligned}$$

Redzams, ka Y_i vidējā vērtība ir γ novērtējums, kurš pazīstams kā Hila novērtējumu. Līdz ar to $r = 1$, jo nepieciešama tikai viena vidējā vērtība, lai novērtētu gadījuma lielumu.

Izmantosim gadījuma izlasi Y_1, \dots, Y_k un tās vidējo vērtību \bar{Y} . Mūsu interesējošais parametrs var tikt izteikts kā $\hat{\gamma} = \bar{Y}$ un līdz ar to $q = r = 1$.

Logaritmētā empīriskās ticamības attiecība ir

$$R_E = -2 \log \{L_E(\theta_0)/L_E(\hat{\theta})\} = -2 \log \{k^k L_E(\theta_0)\}.$$

Tika iepriekš parādīts, ka

$$P(R_E \leq c_\alpha) = P(\chi_1^2 \leq c_\alpha) + O(1).$$

Analogi Diccio rakstam ir jāspēj parādīt, ka Bartleta korekcija Hī-kvadrāta aproksimāciju var uzlabot līdz kļūdai $O(k^{-2})$. Diccio savā tehniskajā atskaitē ([24]) Bartleta korekciju definē caur lielumu $E(nR^T R)$ un apgalvo, ka

$$P(R_E/E(nR^T R) \leq c_\alpha) = P(\chi_1^2 \leq c_\alpha) + O(k^{-2}).$$

Līdzīgi kā parametriskajā gadījumā, arī neparametriskai ticamības attiecībai, Bartleta korekcija tiek izteikta kā vidējā vērtība. Šai gadījumā izmanto $R_E = nR^T R + O_p(k^{-3/2})$. Šādu izvirzījumu iegūst no pieņēmuma, ka pastāv Edgeworth izvirzījums, jo eksistē F_0 sadalījuma momenti un $\theta(\cdot)$ ir gluda funkcija.

Ieviesīsim nelielu atkāpi, lai aprakstītu, kas ir Edgeworth izvirzījums [25].

Definīcija 5. Par statistikas $T_n = \sqrt{n}(\bar{x} - \mu)/\sigma$ Edgeworth izvirzījumu sauc

$$P(T_n \leq x) = \Phi(x) + n^{-1/2}P_1(x)\phi(x) + n^{-1}P_2(x)\phi(x) + \dots + n^{-j/2}P_j(x)\phi(x) + o(n^{-j/2}),$$

kur $\Phi(x)$ ir standarta normalā sadalījuma funkcija, $\phi(x)$ ir standarta normālā sadalījuma blīvuma funkcija un $P_j(x)$ ir polinomi, kas izsakās caur Hermita polinomiem un aproksimētās sadalījuma funkcijas kumulantiem.

Statistikai T_n Edgeworth izvirzījuma pirmie polinomi izsakās formā

$$P_1(x) = \frac{1}{6} \frac{E(x - \mu)^3}{\sigma^3} (1 - x^2).$$

Ja apzīmējam ar $C_1 = \frac{1}{6} \frac{E(x - \mu)^3}{\sigma^3}$, tad

$$P_2(x) = C_1^5 (10x^3 - 15x - x^5) + \frac{1}{24} \left(\frac{E(x - \mu)^4}{\sigma^4} - 3 \right) (3x - x^3).$$

Jāatrod izvirzījums $E(nR^T R) = 1 - ak^{-1}$. Ja aplūkojam gadījumu, kad $q = r = 1$ un mūs interesē noteikt izlases vidējo vērtību, tad Diccio apgalvo, ka

$$\{E(R^T R)\}^{-1} = 1 - k^{-1} \left(-\frac{1}{3} \frac{\mu_3^2}{\mu_2^3} + \frac{1}{2} \frac{\mu_4}{\mu_2^2} \right) + O(k^{-2}),$$

kur $\mu_j = E(Y - \bar{Y})^j$.

2. Praktiskā daļa

Praktiskā daļa ir izpildīta statistikas paketē R. R paketē tika izmantotas paketes `evd` [26], kura ļauj ģenerēt ekstremālo vērtību sadalījumu, un `SMPacticals` [27], kurā bija iebūvēti Dānijas ugunsgrēku risku apdrošināšanas dati.

2.1. Simulācijas

2.1.1. Pārklājuma precizitāte pie optimālā k .

Lai noskaidrotu, cik labi strādā neparametriskās metodes, salīdzinot ar parametriskām metodēm, konstruēsim γ ticamības intervālus un noteiksim pārklājuma precizitāti pie optimālā k .

Lai vērtība k būtu pietiekami liela, gadījuma izlases tiek ģenerētas ar apjomu $n = 400; 1000; 5000$ no sekojošiem sadalījumiem:

1. $F(x) = \exp(x^{-1/\gamma})$, kur $\gamma = 1/2$ un 2 ;
2. $F^-(x) = (1/2)(1-x)^{-\gamma}(1+(1-x)^{-\rho})$, ($0 < x < 1$), kur ar F^- apzīmējam F inverso funkciju un $(\gamma, \rho) = (1/2, -1), (1/2, -0.1)$.

Atzīmēsim, ka pirmajā gadījumā sadalījuma funkcijai ir negatīva otrās kārtas regulārā variācija, t.i., $A(t) < 0$, kura ir definēta ar (1.5.10). Savukārt otram sadalījumam ir pozitīva otrās kārtas regulārā variācija, t.i., $A(t) > 0$. Simulāciju skaits ir $N = 500$ reizes.

Empīriskās pārklājuma precizitātes ticamības intervāliem $I_E(0.95), I_P(0.95), I_N(0.95), I_{EB}(0.95), I_{PB}(0.95)$ tiek noteiktas pie optimālā k , kuru var aprēķināt pēc nodaļā 1.5.3. aprakstītā algoritma.

Lai noteiktu, kāds optimālais k ir funkcijai $1 - F(x) = \exp(-x^{-1/\gamma})$, šo funkciju izvirzām Teilora rindā un ņemam pirmos divus rindas locekļus,

$$1 - F(x) = 1 - \exp(-x^{-1/\gamma}) \stackrel{TR}{\approx} 1 - 1 + x^{-1/\gamma} - 1/2(-x^{-1/\gamma})^2 = x^{-1/\gamma} - 1/2x^{-2/\gamma}.$$

No kurienes iegūstam, ka izteiksmei

$$1 - F(x) = C_1 x^{-1/\gamma} + C_2 x^{-1/\gamma + \rho/\gamma},$$

kur atbilstoši $C_1 = 1$, $C_2 = -1/2$, $\rho = -1$. Tad attiecīgi pie $n = 400$, 1000 , 5000 optimālais k_{opt} no formulas (1.5.11) ir 43, 79 un 232.

Pārējām funkcijām k_{opt} tiek aprēķināti līdzīgi.

2.1. tabulā ir attēloti rezultāti, kas parāda, kādas pārklājuma precizitātes veido ticamības intervāli no abām sadalījuma funkcijām pie k_{opt} .

2.1. tabula: Pārklājuma precizitāte pie k_{opt} .

	$\exp(x^{-2})$	$\exp(x^{-1/2})$	$\frac{1}{2}(1-x)^{-1/2}(2-x)$	$\frac{1}{2}(1-x)^{-1/2}(1+(1-x)^{0.1})$	
$n = 400$	E	0.94	0.942	0.856	0.886
	P	0.946	0.95	0.88	0.912
	N	0.946	0.946	0.816	0.852
	EB	0.952	0.946	0.86	0.89
	PB	0.952	0.956	0.888	0.916
$n = 1000$	E	0.926	0.936	0.856	0.888
	P	0.938	0.944	0.888	0.884
	N	0.946	0.952	0.824	0.84
	EB	0.93	0.944	0.856	0.888
	PB	0.944	0.95	0.898	0.896
$n = 5000$	E	0.946	0.942	0.89	0.77
	P	0.948	0.946	0.89	0.78
	N	0.952	0.946	0.85	0.72
	EB	0.952	0.95	0.89	0.77
	PB	0.954	0.954	0.92	0.8

2.1. tabula parāda, ka neparametriskās metodes strādā labi, jo pilnīgi visas metodes tiecas uz 95% nozīmības līmeni, ja sadalījums atbilst prasībām.

Empīriskā ticamības funkcija (E) dod mazāku ticamības pārklājumu, kad izvēlēta funkcija ir laba (šai gadījuma Frešē sadalījums), un visvairāk pietuvinās Normālās aproksimācijas (N) metodei, ja dots sadalījums ar $A(t) > 0$.

Bartleta korekcija empīriskai ticamības funkcijai (EB) labāk koriģē pārklājuma precizitāti, salīdzinot ar Bartleta korekciju parametriskai ticamības funkcijai (PB). Bet redzams, ka pati par sevi parametriskā ticamības funkcija (P) dod labus pārklājuma rezultātus.

Normālās aproksimācijas metode (N) ļoti labi nosaka pārklājuma precizitāti gadījumā, ja mēs varam nodrošināt novērtējumu normalitāti, pretējā gadījumā nevar sasniegt labus rezultātus.

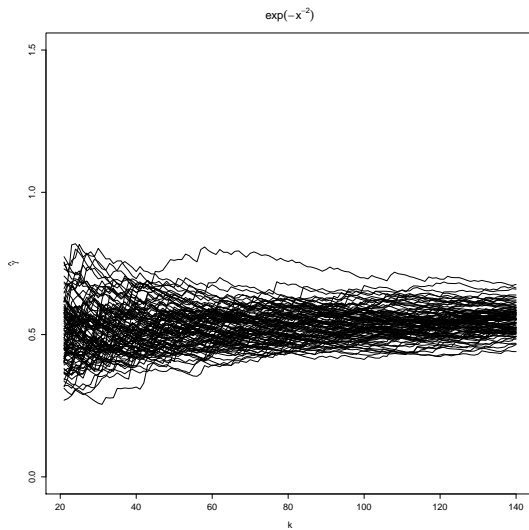
Pēdējā stabiņā redzams, ka, ja izlases apjoms samazinās, tad pārklājuma precizitāte palielinās un tuvojas uz 1. Līdz ar to var secināt, ka nevar balstīties uz vienu noteiktu k un nepieciešams izstrādāt precīzāku k_{opt} aprēķināšanas algoritmu.

2.1.2. Pārklājuma precizitāte dažādiem k .

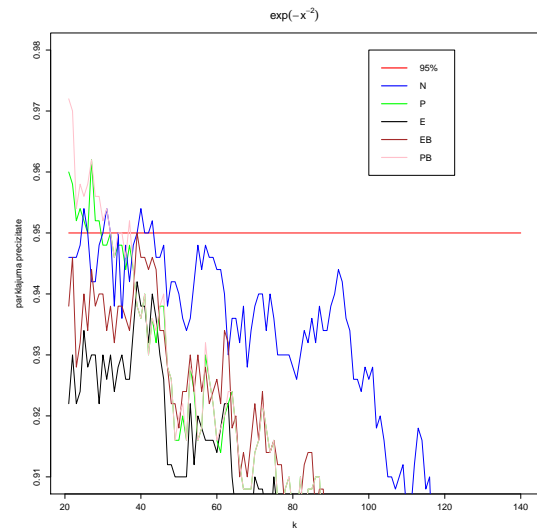
Tā kā ir ļoti grūti pieņemt lēmumu, pie kura fiksēta k ir jānovērtē Hila novērtējums un jānosaka ticamības intervāls, daudzos literatūras apskatos iesaka veikt analīzi veselai kopai no dažādiem k .

Simulācijām izmantosim tos pašus sadalījumus ar tādiem pašiem izvēlētiem γ un ρ , kuri tika aprakstīti iepriekšējā apakšnodaļā. Rezultātus attēlosim grafiski.

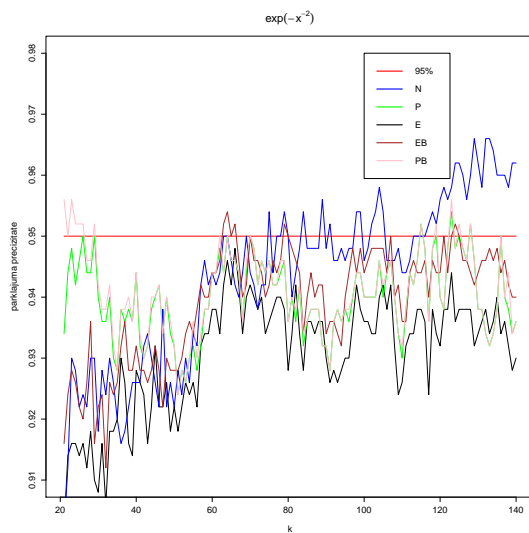
2.1. attēlā novērtē $\hat{\gamma}$ novērtējumus sadalījuma funkcijai $F(x) = \exp(-x^{-2})$. **2.1.a** attēlā ir parādīti, kādus $\hat{\gamma}$ novērtējumus iegūst no simulētiem datiem. **2.1.b**, **2.1.c** un **2.1.d** attēlā ir konstruēta pārklājuma precizitāte no ticamības intervāliem pie izlases apjoma $n = 400, 1000$ un 5000 . No **2.1.a** attēla var redzēt, ka $\hat{\gamma}$ novērtējumi nav ar lielu izkliedi un tie svārstās ap īsto parametru $\gamma = 1/2$. Tāpēc tālāk var redzēt, ka ir jāpaņem salīdzinoši mazs $k(n)$, lai pārklājuma precizitāte būtu tuva 95%. Jau iepriekš parādījām, ka šai sadalījuma funkcijai par optimālo k izvēlas $k = 43$ pie $n = 400$, $k = 79$, ja $n = 1000$ un $k = 232$, ja $n = 5000$. Simulāciju rezultāti rāda, ka labu pārklājuma precizitāti iegūst, ja $k \in (30, 40)$, ja $n = 400$, $k \in (60, 80)$, ja $n = 1000$, un $k \in (75, 90)$, ja $n = 5000$. Tāpēc labi noder apskatīties pārklājuma precizitāti nevis pie viena fiksēta k , bet aplūkot situāciju kopumā.



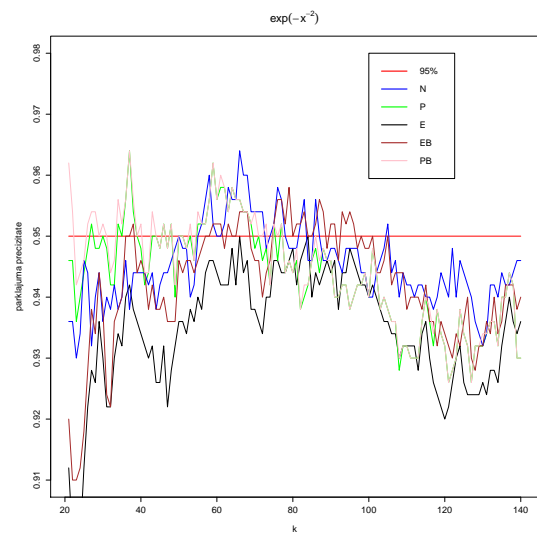
(a) $\hat{\gamma}$ novērtējumi



(b) $n = 400$



(c) $n = 1000$

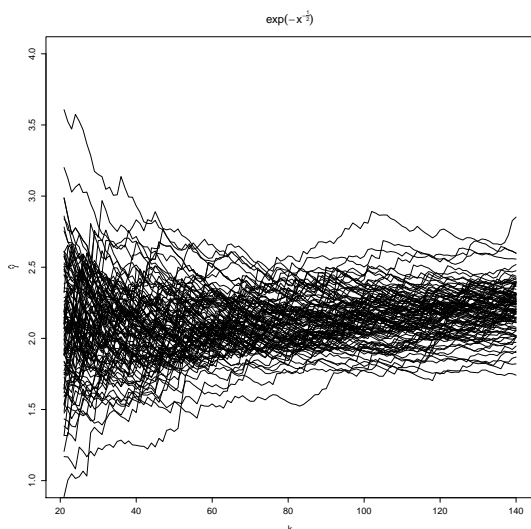


(d) $n = 5000$

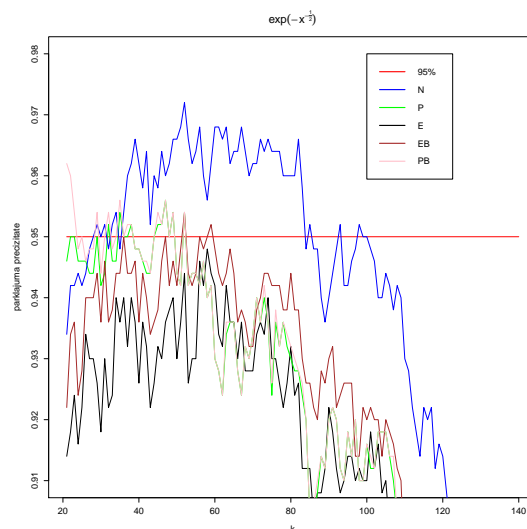
2.1. att.: $F(x) = \exp(-x^{-2})$ funkcijas $\hat{\gamma}$ novērtējumi un pārklājuma precizitātes.

2.2. attēlā apskatīti $\hat{\gamma}$ novērtējumi sadalījuma funkcijai $F(x) = \exp(-x^{-1/2})$. 2.2.a attēlā ir parādīti, kādus $\hat{\gamma}$ novērtējumus iegūst no simulētiem datiem. 2.2.b, 2.2.c un 2.2.d attēlā ir konstruēta pārklājuma precizitāte no ticamības intervāliem pie izlases apjoma $n = 400, 1000$ un 5000 . No 2.2.a attēla var redzēt, ka $\hat{\gamma}$ novērtējumiem palielinās izkliede un īstais parametrs ir $\gamma = 2$. Simulāciju rezultāti rāda, ka labu pārklājuma precizitāti iegūst, ja $k \in (30, 38)$, ja $n = 400$. Bet empīriskā ticamības metode (E) labākos rezultātus dod pie $k \in (55, 60)$. Ja $n = 1000$, tad $k \in (70, 85)$ un pie $n = 5000$ nevar skaidri izdalīt k intervālu, kur visas metodes uzrādītu maksimāli precīzus rezultātus. Šai gadījumā

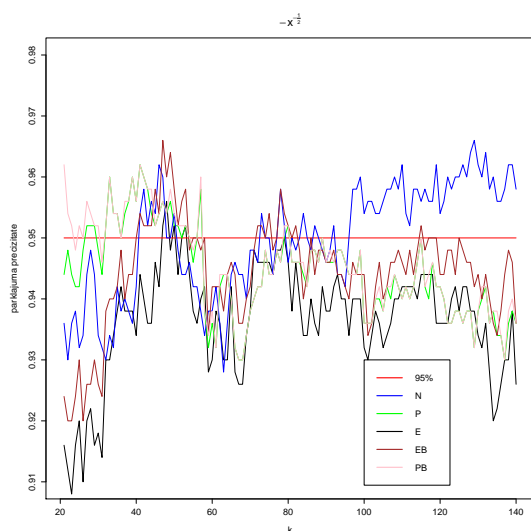
iespējams ir jāņem lielākas k vērtības. Var redzēt, ka katrai metodei var būt savs k apgabals, kur tā dod labu rezultātu.



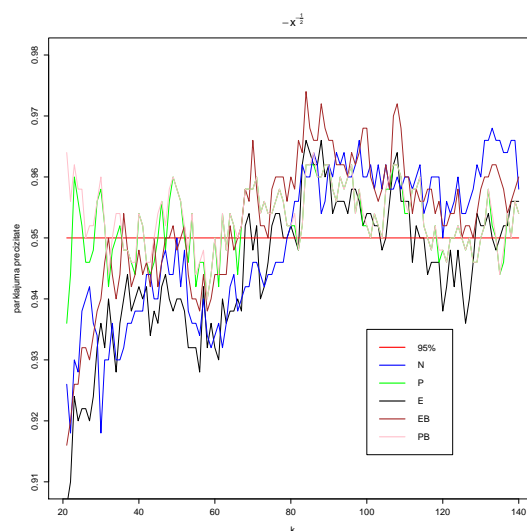
(a) $\hat{\gamma}$ novērtējumi



(b) $n = 400$



(c) $n = 1000$

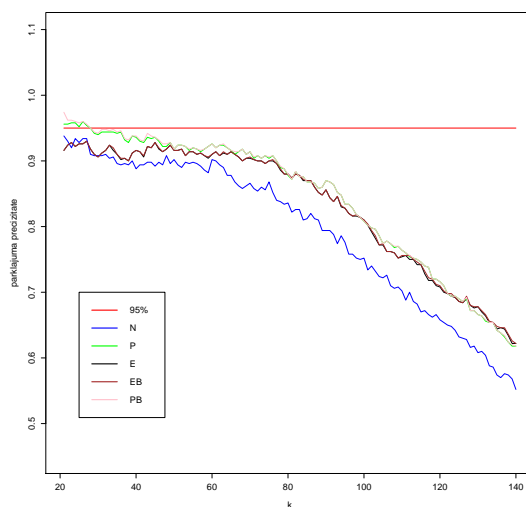
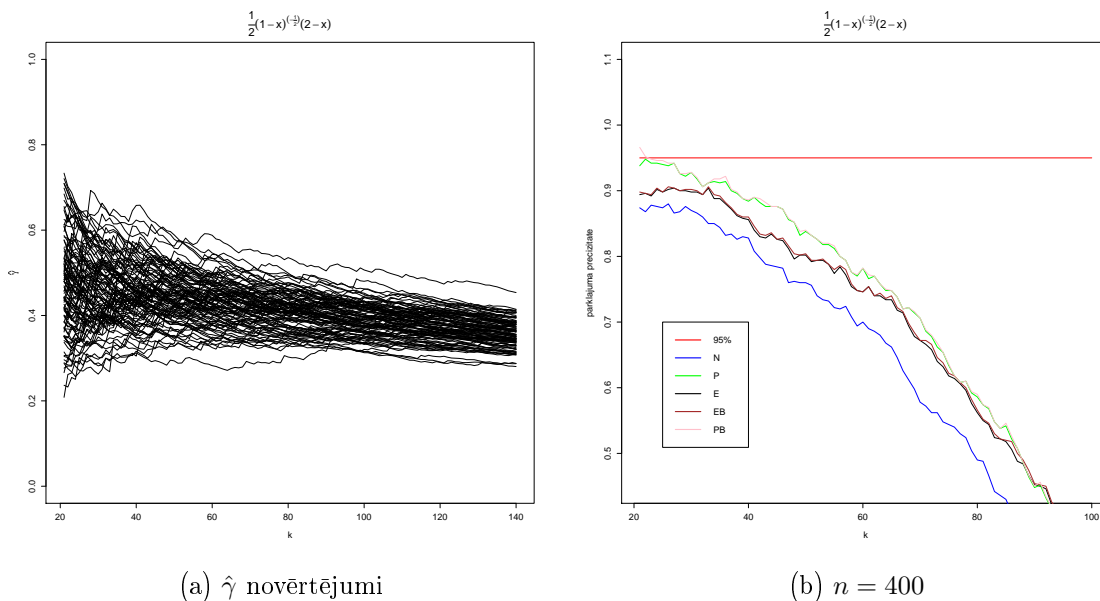


(d) $n = 5000$

2.2. att.: $F(x) = \exp(-x^{-1/2})$ funkcijas $\hat{\gamma}$ novērtējumi un pārklājuma precizitāte.

2.3. attēlā novērtē $\hat{\gamma}$ novērtējumus sadalījuma funkcijai $F(x)$ kuras inversā funkcija izsakās kā $F^{-1}(x) = 1/2(1-x)^{-1/\gamma}(2-x)$. 2.3.a attēlā ir parādīti, kādus $\hat{\gamma}$ novērtējumus iegūst no simulētiem datiem. 2.3.b un 2.3.c attēlā ir konstruēta pārklājuma precizitāte no ticamības intervāliem pie izlases apjoma $n = 400, 1000$. No 2.3.a attēla var redzēt, ka $\hat{\gamma}$ novērtējumi nav ar lielu izkliedi un bet tie diezgan ātri novirzās no īstā parametra vērtības $\gamma = 1/2$. Tāpēc tālāk var redzēt, ka ir jāņem salīdzinoši mazs $k(n)$, lai pārklājuma

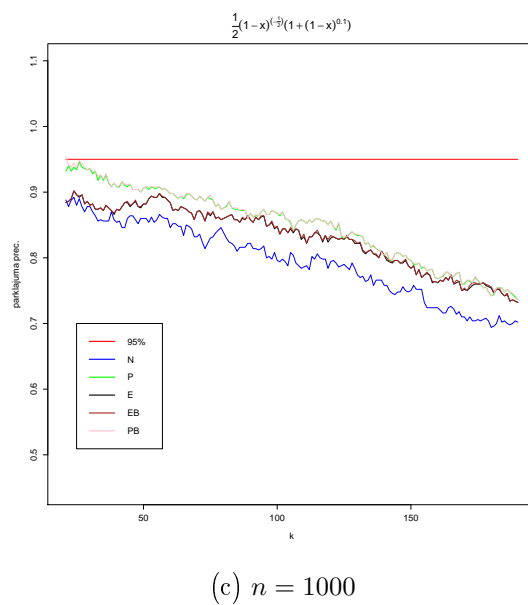
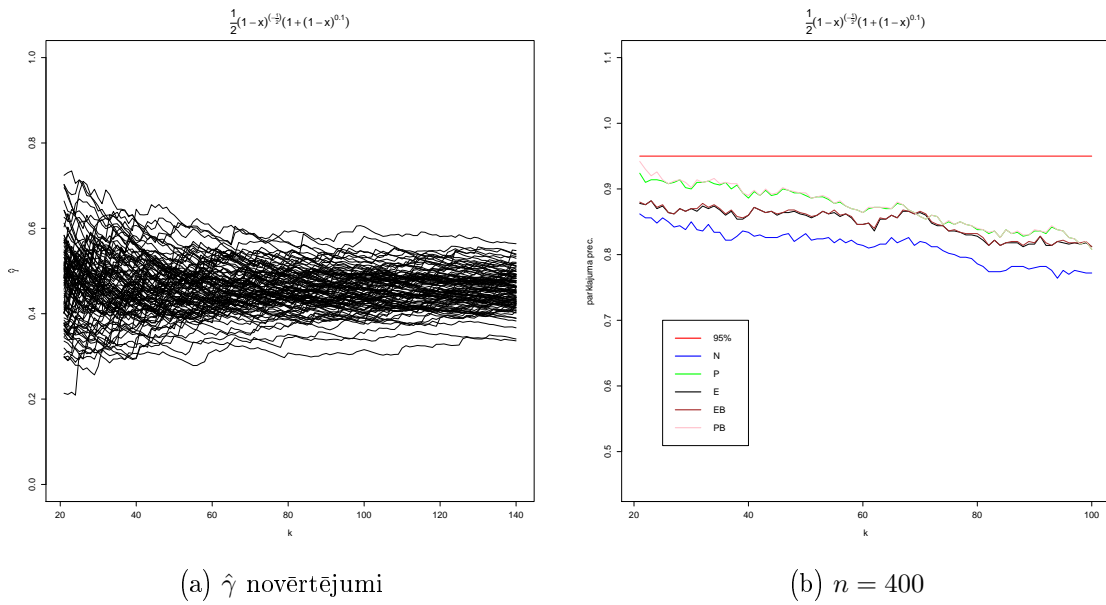
precizitāte būtu vislielākā. Pārklājuma precizitāte ne tuvu nesasniedz 95% līmeni un tā samazinās, palielinoties k vērtībai, jo šīs funkcijas otrās kārtas regulārās variācijas funkcija $A(t)$ ir pozitīva.



2.3. att.: $F^{-}(x) = \frac{1}{2}(1-x)^{-1/2}(2-x)$ parametra γ novērtējumi un pārklājuma precizitātes.

2.4. attēlā novērtē $\hat{\gamma}$ novērtējumus sadalījuma funkcijai $F(x)$, kuras inversā funkcija izsakās kā $F^{-}(x) = 1/2(1-x)^{-2}(1+(1-x)^{0.1})$. 2.4.a attēlā ir parādīti, kādus $\hat{\gamma}$ novērtējumus iegūst no simulētiem datiem. 2.4.b un 2.4.c attēlā ir konstruēta pārklājuma precizitāte no ticamības intervāliem pie izlases apjoma $n = 400, 1000$. No 2.4.a attēla var redzēt, ka $\hat{\gamma}$ novērtējumi nav ar lielu izkliedi un ir mazāk nobīdīti no īstās parametra

vērtības $\gamma = 1/2$. Tendence samazināties pārklājuma precizitātei saglabājas, palielinoties k vērtībai. Bet šī tendence nav tik izteikta, jo šai gadījumā ρ vērtība ir 0,1 un tā ir paņemta mazāka, nekā iepriekšējā gadījumā, kur $\rho = -1$.

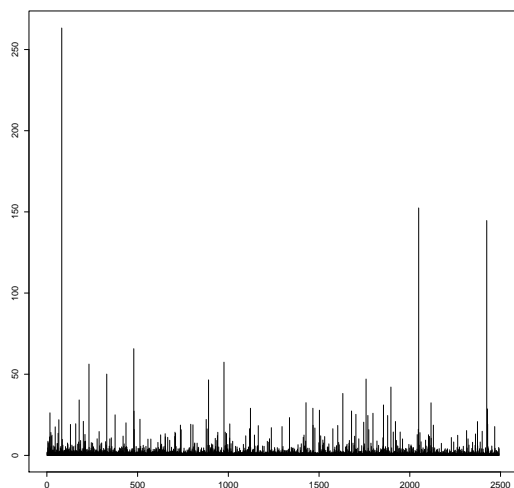


2.4. att.: $F(x) = \frac{1}{2}(1-x)^{-1/2}(1+(1-x)^{0.1})$ $\hat{\gamma}$ novērtējumi un pārklājuma precizitāte.

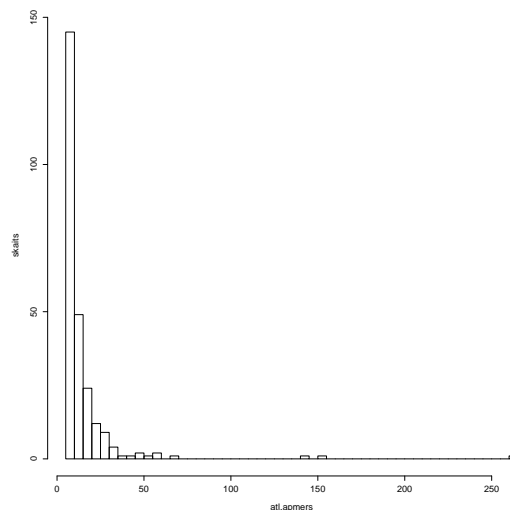
2.2. Metožu pielietojums reāliem datiem

Resnick savā darbā [28] norāda, ka Aleksandrs McNeils pētot Dānijas ugunsgrēku riska apdrošināšanas atlīdzību apmērus, parādīja lielisku piemēru kā ekstremālo vērtību vērtību teorija var tikt pielietota reāliem datiem. Mēs arī izmantosim šos datus, lai izpētītu, kādus ticamības intervālus izveido vislielākās ticamības metodes, normālās aproksimācijas metode un butstrapa metode.

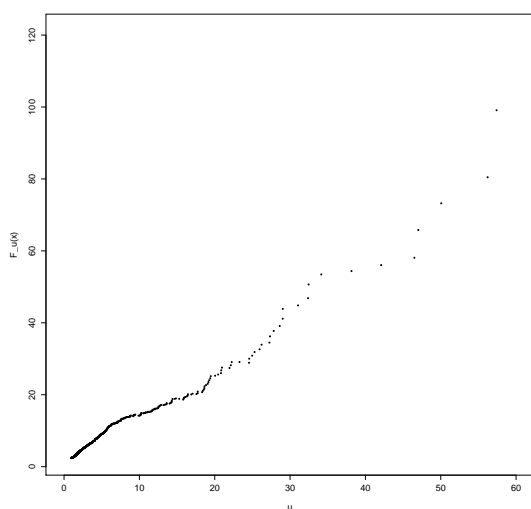
Dati apraksta ugunsgrēku riska zaudējuma apmēru laika periodā no 1980-1990.gadam ieskaitot. Zaudējuma apmēri veidojās no nodegušās ēkas, mantas vērtības un uzņēmējdarbības saglabāšanas pasākumiem. Kopējais atlīdzību skaits $n = 2156$, kuri pārsniedz vienu miljonu Dānijas kronu. 2.5.a attēlā ir redzama laikrinda no Dānijas datiem, kur redzams, ka ļoti lielas atlīdzības notiek reti, ne katru gadu. 2.5.b attēlā parādīta histogramma no atlīdzībām, kuras pārsniedz 5 miljonus Dānijas kronu. Var ieraudzīt, ka tās dažas lielās atlīdzības veido smagu asti. 2.5.c attēlā ir parādīta vidējā ekscesa funkcija. 2.5.d attēlā parādīts Dānijas datu Q-Q plots pret standarta eksponenciālām kvantilēm. Dānijas datiem aste ir smagāka par eksponenciālo sadalījumu un tas ļauj mums domāt, ka dati ir ar smagu asti un atbilst ekstremālo vērtību sadalījumam.



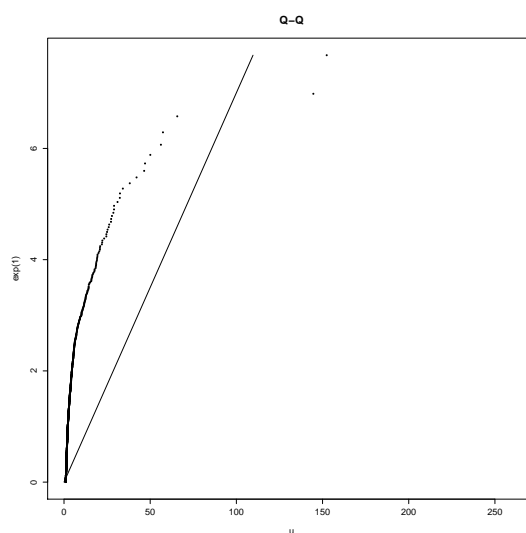
(a) Laikrinda



(b) Lielo atlīdzību histogramma



(c) $F_u(x)$

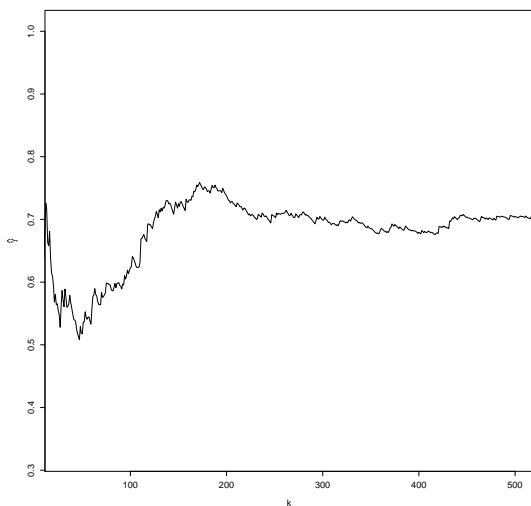


(d) Q-Q plot pret $\exp(0)$

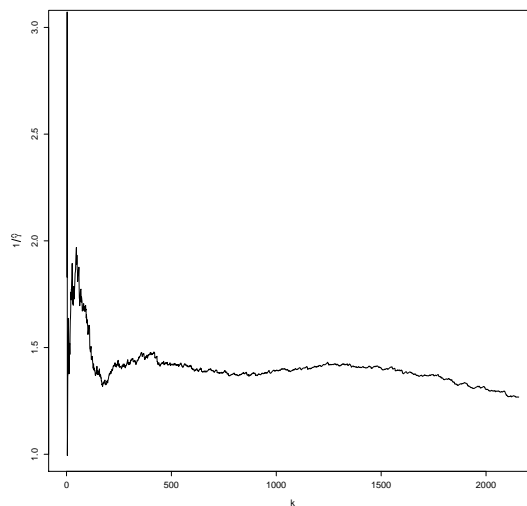
2.5. att.: Dānijas ugunsgrēku riska apdrošināšanas dati.

2.6.a attēlā parādīts $\hat{\gamma}$ novērtējums atkarībā no sakārtoto statistiku skaita k . 2.6.b attēlā parādīts Hila novērtējums $1/\hat{\gamma}$ atkarībā no k . Var redzēt, ka $\hat{\gamma}$ novērtējums pie vidējam sakārtotām statistikām k ir aptuveni 0.7. 2.6.c attēlā attēloti $\hat{\gamma}$ ticamības intervāli. Ticamības intervāli ir zilā, sarkanā un zaļā krāsā, kur tie tika uzkonstruēti atbilstoši ar normālās aproksimācijas, parametriskās un empīriskās ticamības metodēm. Ticamības intervāli tika konstruēti pie nozīmības līmeņa $\alpha = 95\%$ un dažādiem izvēlētiem k . Ieguvām, ka visšaurākos ticamības intervālus dod normālā aproksimācija. Visplašākos ticamības intervālus dod empīriskā ticamības metode. 2.6.d attēlā ir parādīti tie paši ticamības

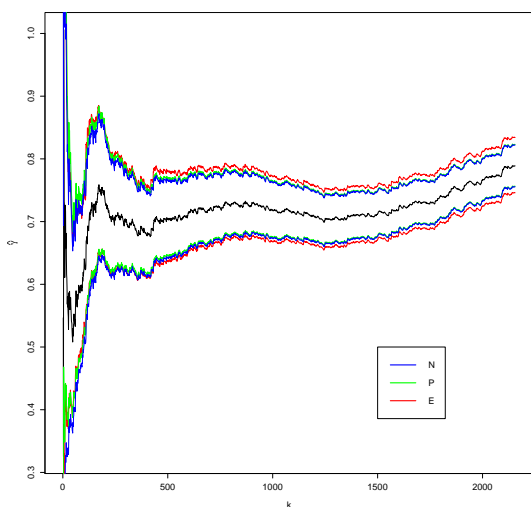
intervāli, tikai k paņemts intervālā $(30, 250)$. Klāt vel tika pievienots ticamības intervāls ar butstrapa metodi (līnija brūnā krāsā). Var redzēt, ka, empīriskā ticamības funkcija ir labāka par butstrapa metodi. Butstrapa ticamības intervāli ir salīdzinoši ļoti plaši un pieļauj lielu novirzi no novērtējuma, kur citas metodes to nepieļauj.



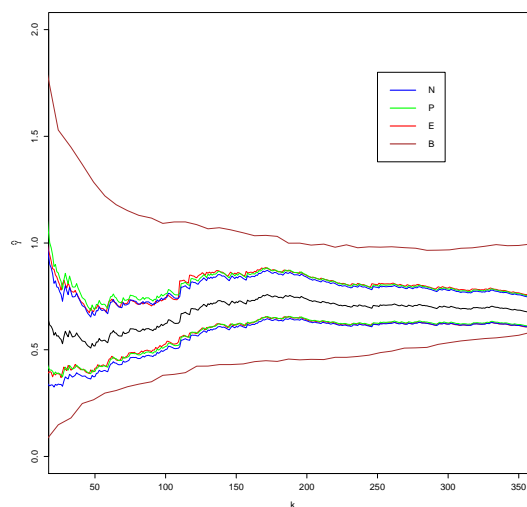
(a) $\hat{\gamma}$ novērtējumi



(b) $1/\hat{\gamma}$ novērtējumi



(c) Ticamības intervāls $k \in (0, 2300)$

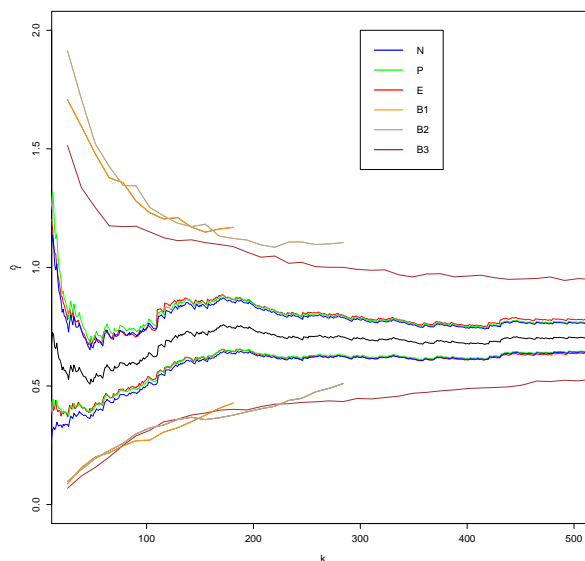


(d) Ticamības intervāls $k \in (30, 250)$

2.6. att.: Dānijas ugunsgrēku riska apdrošināšanas dati.

Empīriskā un parametriskā Bartleta korekcija netika attēlos attēlota, jo tās gandrīz neatšķirās no parametriskās un empīriskās ticamības metodes rezultātiem. Savukārt tika tuvāk izpētīta butstrapa metode bez atkārtojumiem. 2.7. attēlā ir attēloti visi tie paši ticamības intervāli, kas iepriekš (E -empīriskā ticamības metode, P -parametriskā ticamības

metode, N -normālā aproksimācija), bet papildus ir attēloti ticamības intervāli ar butstrapa metodi, kur atšķiras apakšizlases apjoms n_1 . $B1$ ir veidots no izlasēm apjomā 27, $B2$ ir ar $n_1 = 33$ un $B3$ ir ar apjomu $n_1 = 139$. Var redzēt, jo lielāku izvēlas apakšizlases n_1 apjomu, jo lielākam k intervālam spējam noteikt ticamības intervālu. Kā arī, jo lielāks ir n_1 , jo šaurāks ticamības intervāls ir pie mazākiem k un paliek platāks pie lielākiem k . Tomēr tāpat redzams, ka ticamības intervāls ar butstrapa metodi paliek visplašākais no pārējām apskatītajām metodēm.



2.7. att.: Attēloti butstrapa ticamības intervāli $B1$; $B2$; $B3$ atkarībā no apakšizlases apjoma, kur $B1$; $B2$; $B3$ ir veidoti no apakšizlasēm ar apjomu $n_1 = 27$; 33 un 139.

Secinājumi

Darbā tika apskatītas vairākas metodes, kas ļauj novērtēt smagās astes indeksa Hila novērtējuma γ ticamības intervālus. Tika aplūkota parametriskā un empīriskā ticamības funkcijas metode, bootstrapa metode un normālās aproksimācijas metode. Simulējot dažādu sadalījumu datus un izmantojot pārklājuma precizitāti, tika analizēta metožu efektivitāte.

Simulāciju rezultāti liecina, ja izlašu apjomi ir mazi, tad parametriskās metodes strādā labāk par empīrisko ticamības metodi. Bet, ja empīriskai ticamības metodei tiek pielāgota Bartleta korekcija, tad šis novērtējums ir salīdzināms ar parametriskām metodēm un var dot pat labākus rezultātus par normālās aproksimācijas metodi.

No neparametriskām metodēm empīriskā ticamības funkcija dod šaurākus ticamības intervālus nekā bootstrapa metode. Bootstrapa metode konstruē visplašākos ticamības intervālus. Jo lielāku izvēlas bootstrapotās izlases apjomu, jo lielākus ticamības intervālus iegūst. Lai cik ir plašas bootstrapa metodes iespējas, var redzēt, ka ekstremālo vērtību teorijā šī metode nestrādā tik labi, kā arī bootstrapa metodei tiek papildus pieņemti nosacījumi, ka izlases tiek ģenerētas bez atkārtojumiem un tapēc arī bootstrapa izlases apjoms ir mazāks par sākotnējās izlases apjomu.

Lai iegūtu stabilu 95% pārklājuma precizitāti pēc iespējas lielākā k intervālā, otrās kārtas regulāri variējošai funkcijai $A(t)$ jābūt ar tādu pašu zīmi, kā parametram ρ . Šajā darbā mēs aplūkojam sadalījumus, kuriem $\rho < 0$. Kā arī pašam parametram ρ ir jābūt mazam.

Pēdējais secinājums ir tāds, ka reāli dzīvē ir ļoti grūti vai praktiski neiespējami novērtēt $A(t)$, ρ un sadalījumu, pēc kuriem ir sadalīti dati. Tāpēc šajā gadījumā liela priekšrocība ir neparametriskajām metodēm, kuras balstās vienīgi uz γ novērtējumu.

Pateicības

Pateicos rakstu autoriem, kuri deva norādes, kur meklēt informāciju un studentiem, kuri palīdzēja dabūt nepieciešamos rakstus. Pateicos arī ģimenei un savam darba vadītājam par ziedoto laiku un morālo atbalstu, bez kura darbs nebūtu tapis.

Izmantotā literatūra un avoti

- [1] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.
- [2] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.
- [3] Paul Embrechts, Claudia Klupperberg, and Thomas Mikosh. *Modelling Extremal Events for Insurance and Finance*. Springer, 2006.
- [4] O.E. Bandorff-Nielsen and D.R. Cox. Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society*, 46(3):483–495, 1984.
- [5] Thomas Diccio, Peter Hall, and Joseph Romano. Empirical likelihood is bartlett-correctable. *The Annals of Statistics*, 19(2):1053–1061, June 1991.
- [6] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0.
- [7] J. Galambos and E. Seneta. Regularly varying sequences. *Proceedings of the American Mathematical Society*, 41(1):110–116, November 1973.
- [8] R. von Mises. *Selected Papers of Richard von Mises*. American Mathematical Society, 1964.
- [9] A.F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171, 1955.
- [10] Shihong Cheng and Liang Peng. Confidence intervals for the tail index. *Bernoulli*, 7(5):751–760, 2001.

- [11] Jye-Chyi Lu and Liang Peng. Likelihood based confidence intervals for the tail index. *Extremes*, 5(4):337–352, December 2002.
- [12] <http://www-users.math.umd.edu/~evs/s701/WilksThm.pdf>.
- [13] Art Owen. *Empirical Likelihood*. Chapman & Hall/CRC, 2001.
- [14] Ishay Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73:812–815, 1978.
- [15] Art Owen. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120, 1990.
- [16] Armelle Guillou. Bootstrap confidence intervals for the pareto index. *Communications in statistics - theory and methods*, 29(1):211–226, 2000.
- [17] Jef Caers and Jozef Van Dyck. Nonparametric tail estimation using a double bootstrap method. *Computational Statistics and Data Analysis*, 29:191–212, 1999.
- [18] Yongcheng Qi. Bootstrap and empirical likelihood methods in extremes. *Extremes*, 11:81–97, 2008.
- [19] Peter Hall. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32:177–203, 1990.
- [20] P. McCullagh and D.R. Cox. Invariants and likelihood ratio statistics. *The Annals of Statistics*, 14:1419–1430, 1986.
- [21] Gauss Cordeiro, F. Cribari-Neto, Silvia Ferrari, and Miguel Uribe. Improved score tests for one-parameter exponential family models. Econometrics 9508001, Econ-WPA, August 1995.
- [22] Erich Haeusler and Johan Segers. Assessing confidence intervals for the tail index by edgeworth expansions for the hill estimator. *Bernoulli*, 13:175–194, 2007.
- [23] R. N. Bhattacharya and J. K. Ghosh. On the validity of the formal edgeworth expansion. *The Annals of Statistics*, 6(2):434–451, 1978.

- [24] Thomas Diccio, Peter Hall, and Joseph Romano. Bartlett adjustment for empirical likelihood. Technical Report 298, Department of Statistics Stanford University, August 1988.
- [25] Wikipedia. Edgeworth series — wikipedia, the free encyclopedia, 2010. [Online; accessed 1-June-2010].
- [26] A. G. Stephenson. evd: Extreme value distributions. *R News*, 2(2):0, June 2002.
- [27] Anthony Davison. *SMPracticals: Practicals for use with Davison (2003) Statistical Models*, 2006. R package version 1.1-1.
- [28] Sidney I. Resnick. Discussion of the danish data on large fire insurance losses. *Astin Bulletin*, 27:139–151, 1997.

A Izveidoto programmu kods

```
library(evd)
# F(x)=exp(-x^(-1/gamma)), x>0
library(evd)
N<-1000 # simulaciju skaits
n1<-1000 # izlases apjoms
n2<-20
n3<-120 # k apgabals (n2+1,n2+n3)
a<-1
b<-2
s<-2
X1<-sapply(1:N,function(i) sort(rgev(n1,loc=a,scale=b,shape=s)) )
# Hila estimator
gamma<-function(k,X)
{
n<-length(X)
logX<-sapply(1:n,function(i)log(X[i]))
a<-sapply(1:k,function(i)logX[n-i+1])
1/k*sum(a)-logX[n-k]
}
gammavektors1<-function(n3,X)
{
sapply(1:n3,function(i)gamma((i+n2),X))
}
gammamatrixa<-sapply(1:N, function(j) gammavektors1(n3,X1[,j]))
```

```

k<-seq(n2+1,n2+n3,by=1)
for( i in 1:N)
{
points(k,gammamatrix[,i],type="l")
}
# empirical likelihood method
g<-seq(1.7,2.3,by=0.01)
Y2<-function(k,X)
{
sapply(1:(k-1), function(i) i*log(X[k-i+1]/X[k-i]))
}
Ymatrix<-sapply(1:N, function(j) Y2(n1,X1[,j]))
Ymatrix[,1]
lam.fun<-function(lamda,G,Y)
{
sum((Y-G)/(1+lamda*(Y-G)))
}
eR<-function(n,go,y)
{
gal1<-(1-n^{-1})/(go-max(y))
gal2<-(1-n^{-1})/(go-min(y))
lambda<-uniroot(lam.fun,c(gal1,gal2),tol=0.00001,G=go,Y=y)$root
2*sum(log(1+lambda*(y-go)))
}
ex<-function(n,y,hillest,gammamatrix[,j])
{
solis<-0.1
vertiba<-qchisq(0.95,1)
xx<-gammamatrix[,j]
l1<-eR(n,xx,y)-vertiba
l2<-eR(n,(xx-solis),y)-vertiba
while(l1*l2>0)

```

```

{
xx<-xx-solis
l1<-l2
l2<-eR(n,(xx-solis),y)-vertiba
}
ex1s<-uniroot(function(x) eR(n,x,y)-vertiba,c((xx-solis),xx))$root
xx<-gammamatrixaij
l1<-eR(n,xx,y)-vertiba
l2<-eR(n,(xx+solis),y)-vertiba
while(l1*l2>0)
{
xx<-xx+solis
l1<-l2
l2<-eR(n,(xx+solis),y)-vertiba
}
ex2s<-uniroot(function(x) eR(n,x,y)-vertiba,c(xx,(xx+solis)))$root
(ex1s-hillest)*(hillest-ex2s)
}
statvektors<-function(j)
{
sapply(1:n3,function(i)ex((n2+i),Ymatrica[1:(n2+i),j],2,gammamatrixa[i,j]))
}
statvektors(3)
statmatrica<-sapply(1:N,function(j)statvektors(j))
galavertiba<-sapply(1:n3,function(i)
  length(statmatrica[i,][statmatrica[i,]>0])/N)
k<-seq((n2+1),(n2+n3),by=1)
plot(k,galavertiba,type="l",ylim=c(0.91,0.98),ylab="coverage probability",
main="n=10000,N=1000")
points(k,0.95+0*k,type="l",col="red")
#parametric likelihood
pR<-function(i,g,N)

```

```

{
2*(n2+i)*(gammamatrixa[i,N]/g-1-log(gammamatrixa[i,N]/g))
}
px<-function(n,N,hillest,gammamatrixa[i,j])
{
solis<-0.1
vertiba<-qchisq(0.95,1)
xx<-gammamatrixa[i,j]
l1<-pR(n,xx,N)-vertiba
l2<-pR(n,(xx-solis),N)-vertiba
while(l1*l2>0)
{
xx<-xx-solis
l1<-l2
l2<-pR(n,(xx-solis),N)-vertiba
}
ex1s<-uniroot(function(x) pR(n,x,N)-vertiba,c((xx-solis),xx))$root
xx<-gammamatrixa[i,j]
l1<-pR(n,xx,N)-vertiba
l2<-pR(n,(xx+solis),N)-vertiba
while(l1*l2>0)
{
xx<-xx+solis
l1<-l2
l2<-pR(n,(xx+solis),N)-vertiba
}
ex2s<-uniroot(function(x) pR(n,x,N)-vertiba,c(xx,(xx+solis)))$root
(ex1s-hillest)*(hillest-ex2s)
}
statvektors1<-function(j)
{
sapply(1:n3,function(i)px(i,j,2,gammamatrixa[i,j]))
}

```

```

}
statmatrica1<-sapply(1:N,function(j)statvektors1(j))
galavertiba1<-sapply(1:n3,function(i) length(statmatrica1[i,]
[statmatrica1[i,]>0])/N)
points(k,galavertiba1,type="l",col="green")
# ticamibas intervalu konstrueshana ar normal approximation
n<-n3
alpha<-0.05
statvektors2<-function(j)
{
nx1<-sapply(1:n3,function(i) gammamatrica[i,j]-qnorm(1-alpha/2)*
  gammamatrica[i,j]/sqrt((n2+i)))
nx2<-sapply(1:n3,function(i) gammamatrica[i,j]+qnorm(1-alpha/2)*
  gammamatrica[i,j]/sqrt((n2+i)))
sapply(1:n3,function(i) (nx2[i]-2)*(2-nx1[i]))
}
statmatrica2<-sapply(1:N,function(j)statvektors2(j))
galavertiba2<-sapply(1:n3,function(i) length(statmatrica2[i,]
[statmatrica2[i,]>0])/N)
points(k,galavertiba2,type="l",col="blue")
legend(100,0.98,c("95%","Normal","Parametric","Empirical"),
  lwd=2,col=c("red","blue","green","black"))
# Bartlett correction for empirical likelihood
BartletX<-Ymatrica
meanX<-sapply(1:N,function(i) mean(BartletX[,i]))
mik<-function(k,meanx,bartletx)
{
n<-length(bartletx)
mk<-sapply(1:n,function(i) (bartletx[i]-meanx)^k)
mean(mk)
}
mi2<-sapply(1:N, function(j) mik(2,meanX[j],BartletX[,j]))

```

```

mi3<-sapply(1:N, function(j) mik(3,meanX[j],BartletX[,j]))
mi4<-sapply(1:N, function(j) mik(4,meanX[j],BartletX[,j]))
a<-sapply(1:N, function(i) 1/2*mi4[i]/(mi2[i])^2-1/3*(mi3[i])^2/(mi2[i])^3)
#jaunie ticambas intervali (1+a/n)*qchisq(0.95,1)
n<-length(BartletX[,1])
jaunieticint<-sapply(1:N,function(i) (1+a[i]/n)*qchisq(0.95,1))
#piemerojam empirical likelihood
exbartlet<-function(n,y,hillest,gammamatricaij,jaunticrob)
{
solis<-0.1
xx<-gammamatricaij
l1<-eR(n,xx,y)-jaunticrob
l2<-eR(n,(xx-solis),y)-jaunticrob
while(l1*l2>0)
{
xx<-xx-solis
l1<-l2
l2<-eR(n,(xx-solis),y)-jaunticrob
}
ex1s<-uniroot(function(x) eR(n,x,y)-jaunticrob,c((xx-solis),xx))$root
xx<-gammamatricaij
l1<-eR(n,xx,y)-jaunticrob
l2<-eR(n,(xx+solis),y)-jaunticrob
while(l1*l2>0)
{
xx<-xx+solis
l1<-l2
l2<-eR(n,(xx+solis),y)-jaunticrob
}
ex2s<-uniroot(function(x) eR(n,x,y)-jaunticrob,c(xx,(xx+solis)))$root
(ex1s-hillest)*(hillest-ex2s)
}

```

```

statvektorsbartlet<-function(j)
{
sapply(1:n3,function(i)exbartlet((n2+i),Ymatrica[1:(n2+i),j],2,
gammamatrica[i,j],jaunieticint[j]))
}
statmatricabartlet<-sapply(1:N,function(j)statvektorsbartlet(j))
galavertibabartlet<-sapply(1:n3,function(i)
  length(statmatricabartlet[i,]
[statmatricabartlet[i,]>0])/N)
galavertibabartlet
points(k,galavertibabartlet,type="l",col="brown")
#Bartleta korekcija parametriskai ticamibas metodei
pbx1<-sapply(1:length(px1),function(i) px1[i]*(1+1/(6*k[i])))
pbx2<-sapply(1:length(px2),function(i) px2[i]*(1+1/(6*k[i])))
points(k,pbx1,type="l",col="brown")
points(k,pbx2,type="l",col="brown")
#butstraps
B<-1000
n1<-sqrt(length(X))
res<-lapply(1:B,function(i) sort(sample(X,n1,replace=FALSE)))
#nosakam istos novertejumus katram atseviskam kbn
kbn<-seq(2,n1,by=1)
kk<-sapply(1:(n1-1), function(i) round(kbn[i]*(n/n1)^(2/3)))
logX<-log(X)
gamma<-function(k,logX)
{
n<-length(logX)
a<-sapply(1:k,function(i)logX[n-i+1])
1/k*sum(a)-logX[n-k]
}
Hnk<-sapply(1:(n1-1),function(i)gamma(kk[i],logX))
Hnk

```

```

# nosakam butstrapotos novertejumus
logY<-sapply(1:B,function(i) log(res[[i]]))
statvektors<-function(k)
{
sapply(1:B,function(i) gamma(k,logY[,i]))
}
statvektors(kbn[1])
statmatrica<-sapply(1:(n1-2),function(j) statvektors(kbn[j]))
# statmatrica - katra kolonna atbilst savam kbn
# procentilu intervali
sort.statmatrica<-sapply(1:(n1-2),function(i) sort(statmatrica[,i]))
alfa<-0.05
T1.p<-sapply(1:(n1-2),function(i) sort.statmatrica[B*(alfa/2),i])
T2.p<-sapply(1:(n1-2),function(i) sort.statmatrica[B*(1-alfa/2),i])
T1.p;T2.p
points(kk[1:length(T1.p)],T1.p,type="l",col="brown")
points(kk[1:length(T1.p)],T2.p,type="l",col="brown")

```

Maģistra darbs “Neparametriskās statistikas metodes ekstremālu notikumu modelēšanā.” izstrādāts LU Fizikas un Matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Inta Cīmure

(paraksts)

(datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

(paraksts)

(datums)

Recenzents: prof. Dr.math. Gaida Petere

(paraksts)

(datums)

Darbs iesniegts Matemātikas nodaļā _____

(datums)

(darbu pieņēma)

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

_____ prot. Nr. _____, vērtējums _____

(datums)

Komisijas sekretārs/-e: _____

(Vārds, Uzvārds)

(paraksts)