KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

TAIKOMOSIOS MATEMATIKOS KATEDRA

Lina Kasperavičiūtė

MAKSIMUMŲ VIDURKIŲ ANALIZĖ

Magistro darbas

Vadovas
prof. dr. J. A. Aksomaitis

KAUNAS, 2008

KAUNO TECHNOLOGIJOS UNIVERSITETAS
FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS
TAIKOMOSIOS MATEMATIKOS KATEDRA

TVIRTINU
Katedros vedėjas
doc. dr. N. Listopadskis
2008 06 06

MAKSIMUMŲ VIDURKIŲ ANALIZĖ

Taikomosios matematikos magistro baigiamasis darbas

Vadovas
prof. dr. J. A. Aksomaitis
2008 06 03

Recenzentas
doc.dr. J.Venclovienė
2008 06 03

Atliko
FMMM 6 gr. stud.
L. Kasperavičiūtė
2008 05 23

KAUNAS, 2008

KVALIFIKACINĖ KOMISIJA

Pirmininkas: Leonas Saulis, profesorius (VGTU)

Sekretorius: Eimutis Valakevičius, docentas (KTU)

Nariai: Algimantas Jonas Aksomaitis, profesorius (KTU)

Vytautas Janilionis, docentas (KTU)

Vidmantas Povilas Pekarskas, profesorius (KTU)

Rimantas Rudzkis, profesorius (MII)

Zenonas Navickas, profesorius (KTU)

Arūnas Barauskas, UAB „Elsis" generalinio direktoriaus pavaduotojas

Kasperavičiūtė L., Analysis of Maxima Means: Master's work in applied mathematics / supervisor prof. dr. J. A. Aksomaitis; Department of Applied mathematics, Faculty of Fundamental Sciences, Kaunas University of Technology.- Kaunas, 2008. - 34p.

SUMMARY

In practice we often have to calculate extreme values of a sequence $X_1, X_2, ..., X_n$ of random variables. Extreme values are mathematical models of objects in many technical and economical sciences.

In this work maxima of independent and identically distributed random variables $X_1, X_2, ..., X_n$ with distribution function F are analyzed. We calculate maxima means for Pareto and Buro distributions and compare theoretical values with known estimates [2]. We use limit theorems for maxima means when the set size n is large and find the estimate of convergence rate for Pareto random variables. When the set size N is geometric random number maxima means for Buro random variables are calculated.

TURINYS

LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

ĮVADAS

Per pastaruosius 50 metų ekstremaliųjų reikšmių teorija tapo svarbia statistikos disciplina taikomuosiuose moksluose. Taip pat ji vis plačiau taikoma ir kitose disciplinose, pavyzdžiui, sprendžiant rizikos valdymo problemas susijusias su draudimu ir finansais, telekomunikacijų, ekonomikos, ekologijos, meteorologijos srityse. Taikant matematinius modelius dažnai tenka skaičiuoti ekstremaliųjų reikšmių skaitines charakteristikas. Taip pat sudėtingesniais atvejais, kai neįmanoma nustatyti ekstremumų pasiskirstymo dėsnio, skaitinės charakteristikos yra vienintelis būdas reiškiniui aprašyti.

Šiame darbe nagrinėjama paprastoji atsitiktinė imtis $X_1, X_2, ..., X_n$ iš Pareto ir Buro generalinės aibės. Skaičiuojami maksimumų vidurkiai, kai imties didumas n yra fiksuotas ir atsitiktinis, pasiskirstęs pagal geometrinį skirstinį.

Pareto skirstinys, pavadintas italų ekonomisto ir sociologo Vilfredo Pareto (1848-1923m.) garbei, yra laipsninio kitimo tikimybinis skirstinys. Jis buvo pastebėtas daugelyje gyvenimo sričių, tačiau pirmą kartą V. Paretas jį panaudojo aprašydamas turtų pasiskirstymą tarp asmenų. V. Paretas paaiškino, kad didesnė visuomenės turto dalis priklauso mažesnei visuomenės daliai. Ši idėja kartais vadinama Pareto principu arba „80-20 taisykle“, kuri sako kad 20 % populiacijos valdo 80 % turto. Įvairiose gyvenimo srityse ją galima interpretuoti įvairiai. Pvz.: „Įmonės parduodamos produkcijos asortimente yra 80 % prekių pavadinimų, kurie duoda 20 % apyvartos, ir 20 %, kurie duoda 80 % apyvartos“ arba „Iš 20 % klientų įmonė gauna 80 % pelno, o iš 80 % klientų – tik 20 % pelno“. Pastaraisiais metais Pareto tipo skirstiniai ($1 < \alpha < 2$) populiarūs finansiniuose modeliuose, kai finansų rinką apibrėžia ne Brauno judesio modeliai. Šis skirstinys naudojamas pajamų skirstiniui modeliuoti. Jis taip pat naudojamas orų sinoptikoje temperatūros ekstremumams nustatyti.

Darbo tikslai yra rasti tikslias maksimumo vidurkių reikšmes ir jas palyginti su žinomu įverčiu [2], išnagrinėti maksimumų vidurkių ribinį elgesį, kai n didelis bei rasti maksimumų vidurkių konvergavimo greičio įvertį, taip pat ištiri maksimumo vidurkį, kai imties didumas N yra atsitiktinis.

Darbo tema skaityti pranešimai šiose konferencijose: VI taikomosios matematikos studentų konferencija, Matematika ir matematikos dėstymas - 2007, Matematika ir matematikos dėstymas - 2008, VII taikomosios matematikos studentų konferencija. Pranešimų medžiaga pateikta priede.

1. BENDROJI DALIS

1.1 EKSTREMALIŲJŲ REIKŠMIŲ TEORIJA

Ekstremaliųjų reikšmių teorija yra tikimybių teorijos šaka, jungianti daug taikymų ir teorinių matematinių rezultatų. Todėl ši sritis yra įdomi tiek tikimybių teorijos specialistams, tiek inžinieriams bei statistikams. Daugelį metų šis mokslas buvo glaudžiai susijęs su E. J. Gumbelio, įdomios ir įvairiapusiškos asmenybės, veikla.

Ekstremaliųjų reikšmių teorija tiria atsitiktinių dydžių maksimumus ir minimumus. Tai yra ekstremalių reikšmių skirstinius, pozicines statistikas ir viršutines bei apatines ribų peržengimus – vadinamąsias skirstinio viršutinę ir apatinę uodegas. Ekstremaliųjų reikšmių teorija yra taikoma įvairiose srityse, kuriose dideli nuokrypiai yra reikšmingi. Galima paminėti tokias sritis, kaip aplinkosauga, meteorologija, finansai ir draudimas, medžiagų atsparumo tyrimas. Įprasti statistiniai metodai didelių nuokrypių situacijoms modeliuoti netinka. Būtent tokių situacijų prognozavimui priemones pateikia ekstremaliųjų reikšmių teorija.

Istorijoje ekstremaliųjų reikšmių problemų sprendimo pradžią jau galima įžvelgti N. Bernulio darbuose (1709). N. Bernulis tyrė vidutinį didžiausią nuotolį nuo atsitiktinai išsidėsčiusių n taškų iki fiksuoto ilgio t tiesės. Tačiau pati ekstremaliųjų reikšmių teorija pradėjo formuotis vėliau. Teorijos formavimąsi paskatino astronomų poreikis atrinkti patikimus stebėjimus, atmetant viršijančius tam tikro kriterijaus ribą. M. L. Fulerio (1914) ir A. Grifito (1920) ankstyvuosiuose šios srities darbuose tiriami tiek matematinės analizės metodai, tiek jų taikymai. Susistemintai bendrajai ekstremaliųjų reikšmių teorijai pradžią davė fon Bortkevičiaus darbas, aprašantis normaliųjų atsitiktinių dydžių imties intervalo skirstinį. Šis darbas svarbus tuo, kad čia pirmą kartą buvo pavartotas terminas – ,,maksimalios reikšmės pasiskirstymas". Kitais metais fon Mizesas (1923) apskaičiavo šio pasiskirstymo vidurkį, o Dodas (1923) apskaičiavo medianą ir tyrė kitokius nei normalusis pasiskirstymas atvejus. Didesnės reikšmės buvo R. M. Frešė (1927) darbas, kuriame aptartas maksimumo asimptotinis skirstinys. R. A. Fišeris ir L. H. C. Tipetas (1928) paskelbė atskiro šios problemos tyrimo rezultatus. Tačiau R. M. Frešė (1927) nurodė tik vieną galimą ribinį ekstremumo skirstinį, o R. A. Fišeris ir L. H. C. Tipetas (1928) parodė, kad ekstremumo ribinis skirstinys gali būti tik vieno iš trijų tipų. Fon Mizesas (1936) pateikė paprastai suformuluotas, pakankamas kiekvieno iš trijų, R. A. Fišerio ir L. H. C. Tipeto (1928) pateiktų, ekstremumų ribinių skirstinių silpno konvergavimo sąlygas. B.V. Gnedenko (1943) pateikė būtinas ir pakankamas ekstremumų konvergavimo sąlygas, tuo padėdamas tvirtą pagrindą tolimesniam ekstremaliųjų reikšmių teorijos vystymuisi.

D. Meizleris (1949), B. Markusas ir M. Pinskis (1969) (nežinodami apie D. Meizlerio rezultatus) ir L. de Hanas (1970) (1971) išplėtė B. V. Gnedenkos rezultatus. M. L. Junkoso (1949) darbas, išplečiantis B. V. Gnedenkos rezultatus nevienodai pasiskirsčiusiems nepriklausomiems atsitiktiniams dydžiams, teoriškai buvo svarbus, tačiau nepriimtas dėl savo menko pritaikomumo.

Praeito amžiaus 3-čiojo ir 4-tojo dešimtmečių teorinius darbus 4-ajame ir 5-ajame dešimtmečiuose papildė straipsniai apie ekstremaliųjų reikšmių teorijos taikymą. Tai taikymai žmogaus gyvenimo trukmės pasiskirstymui, radioaktyviai emisijai (E. J. Gumbelis (1937a,b)), medžiagų atsparumui tirti (E. H. V. Veiblas (1939)), potvynių analizei (E. J. Gumbelis (1941, 1944, 1945, 1949a), S. E. Rantsas ir H. C. Rigsas (1949)), seisminei analizei (J. M. Nordkvistas (1945)), liūčių tyrimui (K. N. Poteris (1949)) ir kita. Labai didelę įtaką taikymams turėjo E. J. Gumbelio darbai. Daugelis šių problemų nagrinėjama darbe ,,Ekstremumų statistikos" (E. J. Gumbelis (1958)).

E. J. Gumbelis buvo pirmasis atkreipęs inžinierių ir statistikų dėmesį į galimą formalios ekstremaliųjų reikšmių teorijos taikymą tam tikriems pasiskirstymams, kurie anksčiau buvo nagrinėjami empiriškai. Pirmosios problemos tokiu būdu pradėtos tirti JAV 1941 metais. Tai buvo susiję su meteorologiniais reiškiniais, tokiais kaip kasmetiniai potvyniai, kritulių maksimumas ir kita.

Daugelis medžiagų atsparumo statistinių modelių rėmėsi A. Grifito idėja, kur teigiama, kad neatitikimas tarp apskaičiuoto ir stebimo medžiagų atsparumo atsiranda dėl tam tikrų dinaminių procesų medžiagos viduje. Ryšį tarp medžiagų atsparumo ir ekstremaliųjų reikšmių pasiskirstymo dėsnių pirmasis pastebėjo F. T. Pirsas. Panašias idėjas savo darbuose taikė ir švedų fizikas bei inžinierius E. H. V. Veiblas (1939).

Vėliau šias problemas tyrė rusų fizikai – J. I. Frenkelis ir T. A. Kontorova (1943). Svarbus, bet nepripažintas, darbas buvo S. Kase'o (1953) straipsnis apie gumų stiprumo pasiskirstymo ekstremaliąsias reikšmes.

Jei būtų sudarytas visos su ekstremaliųjų reikšmių teorija susijusios literatūros sąrašas, jame būtų daugiau nei 1000 publikacijų. Tai rodo šios tikimybių teorijos šakos turtingumą, gyvybingumą bei taikymų gausą, tačiau lygiai taip pat koordinacijos tarp atskirų mokslininkų trūkumą, nes daug vienodų tyrimų buvo atlikta lygiagrečiai.

Lietuvoje ekstremaliųjų reikšmių teorija susidomėta palyginti neseniai. Šios teorijos pradininku Lietuvoje derėtų laikyti prof. J. A. Aksomaitį, kuris jau per dvidešimt metų dirba šioje srityje ir yra paskelbęs nemažai darbų. P. Gudynas (1990, 1992), L. Sakalauskas (1992), A. Jokimaitis, R. Vilkas taip pat yra paskelbę darbų ekstremaliųjų reikšmių tematika. Daug dėmesio atsitiktinių procesų ekstremumams yra skyręs R. Rudzkis (1985, 1986, 1987, 1989). Atsitiktinių procesų maksimumo vertinimui skirtas N. Kalinauskaitės darbas (1986). Nemažai Lietuvos matematikų yra tyrę atsitiktinių dydžių sumų maksimumus bei maksimumo įtaką atsitiktinių dydžių sumoms.

1.2 POZICINĖS STATISTIKOS IR JŲ MOMENTAI

Jeigu atsitiktiniai dydžiai $X_1,...,X_n$ išrikiuoti didėjimo tvarka ir užrašyti taip:

$$X_{(1)} \le ... \le X_{(n)},$$

tai $X_{(i)}$ vadinami *i*-tąja pozicine statistika $(i=1,...,n)$. Laikysime, kad dydžiai X_i yra nepriklausomi ir vienodai pasiskirstę, kurių kiekvienas turi skirstinio funkciją $F(x)$. $X_{(1)}$ ir $X_{(n)}$ vadinamos ekstremumais, atitinkamai minimumu ir maksimumu.

Dažnai literatūroje yra žymima:

$$W_n = \min(X_1,...,X_n),$$

$$Z_n = \max(X_1,...,X_n).$$

Tegu $F_{(r)}(x)\ (r=1,...,n)$ žymi r - tos pozicinės statistikos $X_{(r)}$ skirstinio funkciją. Tada didžiausios pozicinės statistikos $X_{(r)}$ skirstinio funkcija apibrėžiama taip:

$$F_{(n)}(x) = P\{X_{(n)} \le x\} = P\{visi\, X_i \le x\} = F^n(x). \tag{1.1}$$

Tiriamojoje darbo dalyje maksimumo skirstinio funkcijai naudosime žymėjimą $F_{Z_n}(x)$.

Panašiai gauname

$$F_{(1)}(x) = P\{X_{(1)} \le x\} = 1 - P\{X_{(1)} > x\} = 1 - P\{visi\, X_i > x\} = 1 - (1 - F(x))^n. \tag{1.2}$$

r - tos pozicinės statistikos skirstinio funkcija $F_{(r)}(x)$užrašoma taip:

$$F_{(r)}(x) = P\{X_{(r)} \le x\} = P\{bent\, vienas\, r\, iš\, X_i \le x\} = \sum_{i=r}^{n} C_n^i F^i(x)(1 - F(x))^{n-i}. \tag{1.3}$$

kadangi sumos narys yra binominė tikimybė, kurios konkretus *i* iš $X_1,...,X_n$ yra mažesnis arba lygus *x*. Kitaip (1.3) gali būti užrašoma

$$F_{(r)}(x) = F^r(x) \sum_{j=0}^{n-r} C_{r+j-1}^{r-1} (1 - F(x))^j,$$

kur dešinioji lygybės pusė yra tikimybių *r* iš $X_1, X_2,..., X_{r+j}$ suma, įskaitant ir X_{r+j}, yra mažesnė arba lygi *x*.

(1.3) perrašome taip:

$$F_r(x) = E_{F(x)}(n, r), \tag{1.4}$$

čia funkcija E randama iš lentelių(pavyzdžiui, Harvard Computation Laboratory, 1955, kur naudojamas žymėjimas $E(n, r, F(x))$. $F_{(r)}(x)$ dar galima užrašyti naudojantis sąryšiu tarp binominių sumų ir nepilnosios beta funkcijos:

$$F_{(r)}(x) = I_{F(x)}(r, n-r+1) \tag{1.5}$$

čia $I_p(a,b) = \dfrac{\int_0^p t^{a-1}(1-t)^{b-1}\,dt}{B(a,b)}$, $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$, $a > 0, b > 0$.

Taigi $F_{(r)}(x)$ taip pat gali būti apskaičiuota iš $I_p(a,b)$ lentelių (K. Pearson, 1934). X_r procentinės reikšmės gaunamos iš 16 *„Biometrika Tables“* lentelės (Pearson and Hartley, 1970), kurioje pateikiami nepilnosios beta funkcijos procentiniai taškai.

Reikia pažymėti, kad (1.1) – (1.5) rezultatai teisingi tiek diskrečiuoju, tiek tolydžiuoju atveju. Dabar tarsime, kad X_i yra tolydus, su tankio funkcija $f_r(x) = F_r'(x)$. Taigi (1.5) galime užrašyti:

$$f_r(x) = \frac{1}{B(r,n-r+1)}\frac{d}{dx}\int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,dt = \frac{1}{B(r,n-r+1)}F^{r-1}(x)[1-F(x)]^{n-r} f(x). \tag{1.6}$$

Šis rezultatas gali būti gaunamas ir kitu būdu. Įvykis $x < X_{(r)} \le x + \delta x$ aprašomas taip:

r - 1	1	n - r
x		x + δx

$X_i \le x$ visiems $r-1$ iš X_i, $x < X_i \le x + \delta x$ vienam X_i, ir $X_i > x + \delta x$ likusiems $n-r$ iš X_i. Tokių būdų skaičius, kai n stebėjimų suskaidome į tris intervalus, yra

$$\frac{n!}{(r-1)!1!(n-r)!} = \frac{1}{B(r,n-r+1)},$$

ir kiekvienas toks būdas turi tikimybę

$$F^{r-1}(x)[F(x+\delta x) - F(x)][1 - F(x+\delta x)]^{n-r}.$$

δx laikydami mažu turime:

$$P\{x < X_{(r)} \le x + \delta x\} = \frac{1}{B(r,n-r+1)} \cdot F^{r-1}(x) f(x)\delta x[1 - F(x+\delta x)]^{n-r} + O(\delta x)^2,$$

kur $O(\delta x)^2$ reiškia $(\delta x)^2$ eilės narį su stebėjimų $x < X_{(r)} \le x + \delta x$ tikimybėmis, kur intervale $(x, x+\delta x]$ yra daugiau nei vienas X_i. Abi puses padalijus iš δx, kai $\delta x \to 0$, vėl gauname (1.6).

Norėdami akcentuoti imties didumą n, $X_{(r)}$, $r = 1,\dots,n$ žymėsime $X_{r:n}$. Kai $F(x)$ yra tolydi, vidurkis

$$\mu_{r:n} = \int_{-\infty}^{\infty} x \cdot f_{r:n}(x)dx,$$

arba

$$\mu_{r:n} = C_{r,n} \int_{-\infty}^{\infty} xF^{r-1}(x)[1-F(x)]^{n-r} f(x)dx = \tag{1.7}$$

$$= C_{r,n} \int_0^1 F^{-1}(u)u^{r-1}(1-u)^{n-r} du, \tag{1.7`}$$

čia $C_{r,n} = \dfrac{n!}{(r-1)!(n-r)!}$ ir $u = F(x)$. Svarbu, kad (1.7') tinka bet kokiai $F(x)$, kadangi

$$E(X_{r:n}) = E[F^{-1}(U_{r:n})] = C_{r,n} \int_0^1 F^{-1}(u)u^{r-1}(1-u)^{n-r} du,$$

čia $U_{r:n}$ yra $\beta(r,\, n-r+1)$ beta atsitiktinis dydis su skirstinio funkcija $P(X \le x) = I_p(a,b)$,

$$I_p(a,b) = \frac{\int_0^p t^{a-1}(1-t)^{b-1} dt}{B(a,b)}, \; B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx, \; a > 0, b > 0.$$

Iš čia seka, kad

$$|\mu_{r:n}| \le C_{r,n} \int_0^1 |F^{-1}(u)| du = C_{r,n} E(|X|),$$

$\mu_{r:n}$ egzistuoja, jei egzistuoja $E(X)$. Atvirkštinis teiginys nebūtinai teisingas. Jei

$$E(X) = \int_0^1 F^{-1}(u)du$$

neegzistuoja, kai $u = 0$ ir $u = 1$, nepaisant to, $\mu_{r:n}$ gali egzistuoti konkrečioms (bet ne visoms) r reikšmėms. Pavyzdžiui, Koši skirstinio atveju $\mu_{r:n}$ egzistuoja tik tada, kai $r = 1$ arba $r = n$.

Tokiu būdu, jei egzistuoja $E[g(X)]$, kai $g(x)$ bet kokia x funkcija, egzistuoja ir $E[g(X_{r:n})]$. Atskiri atvejai, kai $g(x) = x^k$, $(x-\mu_{r:n})^k$ ir e^{tx} duoda atitinkamai $X_{r:n}$ k – tųjų eilių momentus, centrinius momentus ir momentus generuojančias funkcijas. k – tosios eilės momentą užrašome taip:

$$\mu_{r:n}^{(k)} = E(X_{r:n}^k) \tag{1.8}$$

Jei atsitiktinis dydis X įgyja neneigiamas reikšmes, jo vidurkis gali būti skaičiuojamas pagal formulę:

$$MX = a + \int_a^{\infty} (1 - F_X(x))dx, \; a \ge 0, \tag{1.9}$$

Tikrai, nes

$$MX = \int_a^\infty x dF_X(x) = -\int_a^\infty x \cdot d(1 - F_X(x)) = -(x(1 - F_X(x)))\Big|_a^\infty - \int_a^\infty (1 - F_X(x))dx = a + \int_a^\infty (1 - F_X(x))dx .$$

1.3 POZICINIŲ STATISTIKŲ MOMENTŲ ĮVERČIAI

Jei atsitiktinis dydis *X* turi baigtinę dispersiją, ekstremumų $X_{(n)}$ ir $X_{(1)}$ tikėtinos reikšmės negali būti kiek norima didelės, net jeigu *X* plotis yra neįvertintas. Ekstremumų atveju įvertis gali būti randamas konkrečiai skirstinių klasei. Geresni įverčiai gaunami simetriniams skirstiniams. Pozicinių statistikų atveju, skirtingai negu ekstremumų, rasti įverčiai nėra pasiekiami, tačiau gali būti pagerinti panaudojant apibendrintą Švarco nelygybę.

Nagrinėkime imties iš *n* dydžių didžiausios pozicinės statistikos su skirstinio funkcija $F(x)$ekstremalią reikšmę. Vietoj

$$E\left(X_{(n)}\right) = \int_{-\infty}^{\infty} nx[F(x)]^{n-1} dF(x)$$

patogiau naudoti kitą formą, gautą panaudojant keitinį $u = F(x)$:

$$E\left(X_{(n)}\right) = \int_0^1 nx(u)u^{n-1} du , \qquad \textbf{(1.10)}$$

čia $x(u)$ yra $F^{-1}(u)$. Laikysime, kad atsitiktinis dydis *X* turi vidurkį lygų 0 ir dispersiją lygią 1, t.y.

$$\int_0^1 x(u)du = 0, \quad \int_0^1 [x(u)]^2 du = 1 . \qquad \textbf{(1.11)}$$

Dėl šio pažymėjimo neprarandamas bendrumas, kad skirstinys turi antros eilės momentą. Iš to seka, kad $E\left(X_{(n)}\right)$ turi įvertį, nepriklausomai nuo $F(x)$ pavidalo. Remiantis variaciniu skaičiavimu galime rasti ekstremalią $x(u)$ reikšmę , kai žinoma stacionari (1.10) reikšmė ir tenkinamos (1.11) sąlygos. Tai galima padaryti pirmiausia radus besąlyginį ekstremumą

$$\int_0^1 \left(nxu^{n-1} - ax - \frac{1}{2}bx^2 \right) du$$

ir parinkus koeficientus *a* ir *b* taip, kad būtų tenkinamos (1.11) sąlygos. Stacionarus sprendinys gaunamas iš lygties

$$\frac{\partial}{\partial x}\left(nxu^{n-1} - ax - \frac{1}{2}bx^2 \right) = 0 .$$

tada

$$bx = nu^{n-1} - a ;$$

čia

$$\int_0^1 \left(nu^{n-1} - a\right)du = 0 \qquad \int_0^1 \left(nu^{n-1} - a\right)^2 du = b^2 .$$

Tokiu būdu

$$a = 1,\ \ b = \frac{n-1}{\sqrt{2n-1}},\quad x(u) = \frac{\sqrt{2n-1}\cdot\left(nu^{n-1} - 1\right)}{n-1}, \tag{1.12}$$

ir ekstremumui

$$E\left(X_{(n)}\right) = \frac{n\sqrt{2n-1}}{n-1}\int_0^1 u^{n-1}\left(nu^{n-1} - 1\right)du = \frac{n-1}{\sqrt{2n-1}}$$

Variacinis skaičiavimas naudingas sprendimo parinkimui, tačiau to neužtenka parodyti, kad (1.12) teikia $E\left(X_{(n)}\right)$ maksimumą, o ne prie stacionarią reikšmę. Naudojame Švarco nelygybę

$$\int fg \le \sqrt{\int f^2 du \cdot \int g^2 du} ,$$

čia $f = x$, $g = nu^{n-1} - 1$.

Iš čia gauname

$$E\left(X_{(n)}\right) \le \sqrt{1 \cdot \int_0^1 \left(n^2 u^{2n-2} - 2nu^{n-1} + 1\right)du}$$

taigi

$$E\left(X_{(n)}\right) \le \frac{n-1}{\sqrt{2n-1}} . \tag{1.13}$$

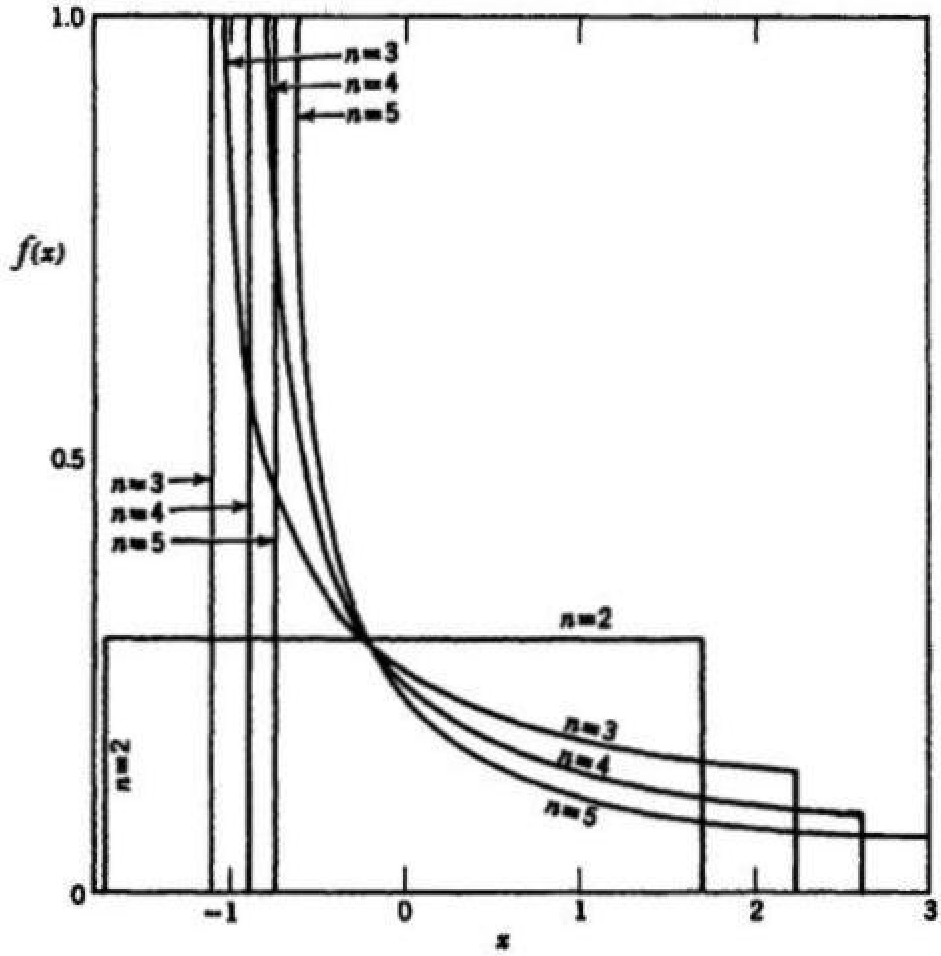
Lygybė čia pasiekiama, kai $x(u)$ randame iš (1.12). Ją panaudojus gauname

$$u = F(x) = \left(\frac{1 + bx}{n}\right)^{\frac{1}{n-1}},\ \text{kai}\ -\frac{\sqrt{2n-1}}{n-1} \le x \le \sqrt{2n-1} . \tag{1.14}$$

Atitinkamai skirstinio tankis lygus

$$f(x) = \frac{b}{n(n-1)}\left(\frac{1 + bx}{n}\right)^{\frac{1}{n-1} - 1},\ \text{kai}\quad -\frac{\sqrt{2n-1}}{n-1} \le x \le \sqrt{2n-1} ,$$

jo grafikas pavaizduotas **1.1** paveiksle



1.1 pav. Bet kokio skirstinio tankio funkcijos grafikai prie skirtingų *n* reikšmių

Tarkime, kad vidurkis ir dispersija lygūs μ ir σ^2, tada (1.13) perrašome taip:

$$E\left(X_{(n)}\right) \le \mu + \frac{(n-1)\sigma}{\sqrt{2n-1}}. \qquad \textbf{(1.15)}$$

Analogiškai galima įrodyti, kad

$$E\left(X_{(1)}\right) \ge \mu - \frac{(n-1)\sigma}{\sqrt{2n-1}}. \qquad \textbf{(1.15`)}$$

Simetrinių skirstinių klasei nelygybę (1.13) galima patikslinti. Kadangi $F(x) = 1 - F(-x)$, gauname

$$E\left(X_{(n)}\right) = \int_0^{\infty} nx\left\{(F(x))^{n-1} - (1-F(x))^{n-1}\right\} dF(x) = \int_{1/2}^{1} nx(x)\left(u^{n-1} - (1-u)^{n-1}\right) du. \qquad \textbf{(1.16)}$$

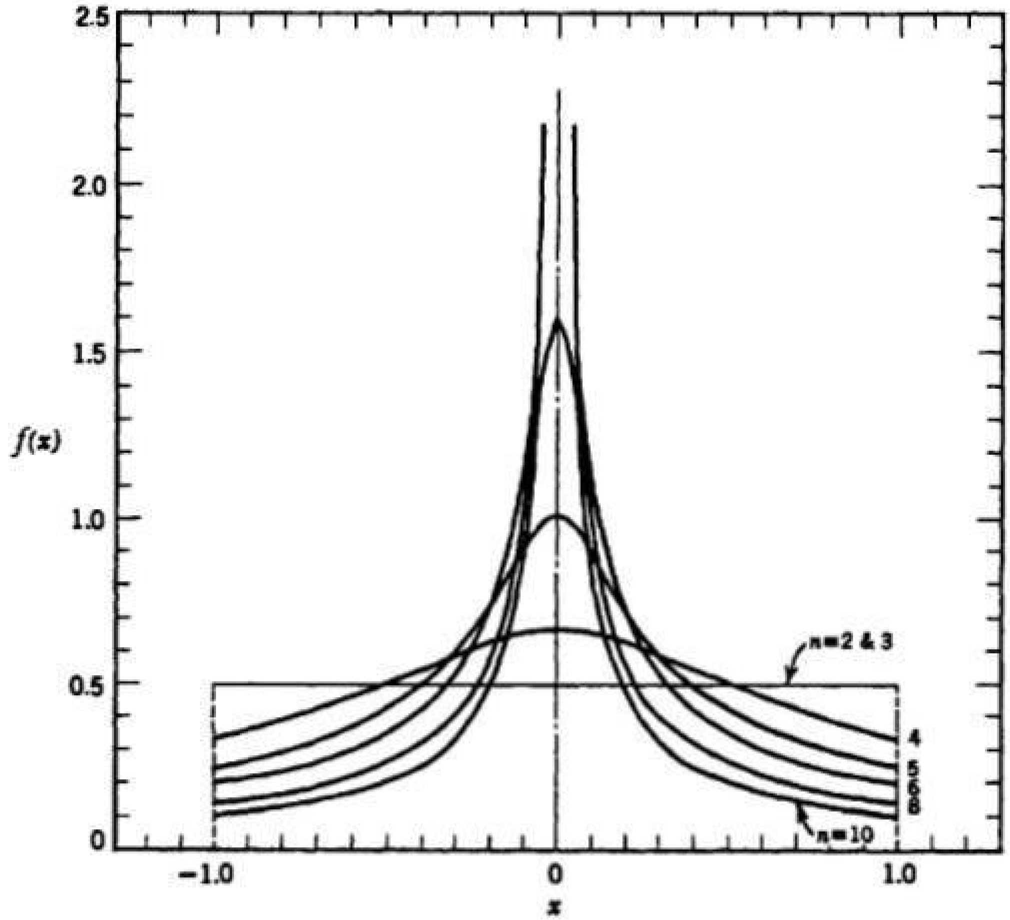
Tas pats priartėjimas, kaip ir anksčiau, panaudojant (1.16), priveda prie ekstremumo

$$cx(u) = u^{n-1} - (1-u)^{n-1}, \ \check{c}ia\ c = \sqrt{\frac{2\left(1 - \frac{1}{C_{2n-2}^{n-1}}\right)}{2n-1}} \qquad \textbf{(1.17)}$$

ir nelygybė įgyja formą

$$E\left(X_{(n)}\right) \leq \frac{1}{2} nc. \quad \textbf{(1.18)}$$

Iš (1.17) seka, kad x reikšmė yra intervale $\left(-\frac{1}{c}, \frac{1}{c}\right)$, t.y. *F(x)* vėl sukoncentruota baigtiniame intervale. Įdomu pažymėti, kad iš (1.14) ir (1.17) išplaukia, kad tolygusis skirstinys intervale $\left(-\sqrt{3}, \sqrt{3}\right)$ yra ekstremalus, kai n=2, o simetrinių skirstinių funkcijų klasėje taip pat, kai n =3.



1.2 pav. Simetrinio skirstinio tankio funkcijų grafikai prie skirtingų *n* reikšmių

1.2 paveiksle parodyti simetriniai ekstremalūs skirstinio tankiai prie skirtingų n reikšmių. Lentelėje 1.1 pateikti du viršutiniai $E\left(X_n\right)$ įverčiai įvairiems n ir palyginami su tikėtinomis standartinio normaliojo ir tolygiojo skirstinio reikšmėmis. Reikia pastebėti, kad mažiems n simetriniu atveju įvertis nedaug skiriasi nuo tiksliai apskaičiuotos reikšmės.

1.1 Lentelė

Dviejų viršutinių $M\left(X_{(n)}-\mu\right)/\sigma$ įverčių palyginimas su tiksliomis normaliojo ir tolygiojo skirstinių reikšmėmis

(Iš Moriguti (1951) darbo, Chartli ir Deivido (1954) darbo ir Tipeto (1925) darbo)

n	Viršutinis bet kurio skirstinio įvertis	Viršutinis simetrinio skirstinio įvertis	Normalusis skirstinys	Tolygusis skirstinys
2	0,5774	0,5774	0,5642	0,5774
3	0,8944	0,8660	0,8463	0,8660
4	1,1339	1,0420	1,0294	1,0392
5	1,3333	1,1701	1,1630	1,1547
6	1,5076	1,2767	1,2672	1,2372
7	1,6641	1,3721	1,3522	1,2990
8	1,8074	1,4604	1,4236	1,3472
9	1,9403	1,5434	1,4850	1,3856
10	2,0647	1,6222	1,5388	1,4171
12	2,2937	1,7693	1,6292	1,4656
15	2,5997	1,9696	1,7359	1,5155
20	3,0424	2,2645	1,8673	1,5671
50	4,9247	3,5533	2,2491	1,6641
100	7,0179	5,0125	2,5076	1,6978
1000	22,3439	15,8153	3,2414	1,7286

1.4 MAKSIMUMŲ RIBINĖS TEOREMOS

Tarkime, kad dydis $\omega(F)$ apibrėžtas taip:

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

Jis yra didžiausias skirstinio funkcijos $F(x)$ kraštinis taškas.

Klasikinė schema: dydžiai $X_1, X_2, ..., X_n$ nepriklausomi, vienodai pasiskirstę, normalizavimas yra tiesinis.

1 teorema: Tarkime, kad $\omega(F)=+\infty$. Jeigu egzistuoja $\alpha>0$, toks, kad $\forall x>0$

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

tada egzistuoja normalizavimo konstantos a_n, $b_n>0$, *čia* $n\geq 1$, su kuriomis

$$\lim_{n\to\infty} P\left(\frac{Z_n - a_n}{b_n} < x\right) = H_{1,\alpha}(x) = \begin{cases} e^{-x^{-\alpha}}, \ x > 0 \\ 0, \ x \le 0 \end{cases}.$$

Normalizavimo konstantas galima parinkti tokiu būdu:

$$a_n = 0; \quad b_n : \ 1 - F(b_n) = \frac{1}{n}.$$

2 teorema: Tarkime, kad $\omega(F) = +\infty$. Apibrėžiame skirstinio funkciją

$$F^*(x) = F\left(\omega(F) - \frac{1}{x}\right), \ x > 0.$$

Jei egzistuoja $\alpha > 0$, su kuriuo $F^*(x)$ tenkina 1 teoremos sąlygą:

$$\lim_{t\to+\infty} \frac{1 - F^*(tx)}{1 - F^*(t)} = x^{-\alpha},$$

tada egzistuoja normalizavimo konstantos $a_n, b_n > 0$, su kuriomis

$$\lim_{n\to\infty} P\left(\frac{Z_n - a_n}{b_n} < x\right) = H_{2,\alpha}(x) = \begin{cases} e^{-(-x)^{\alpha}}, \ x < 0 \\ 1, \ x \ge 0 \end{cases}.$$

Normalizavimo konstantas galima parinkti tokiu būdu:

$$a_n = \omega(F); \quad b_n : \ b_n^{'} : 1 - F(b_n^{'}) = \frac{1}{n}.$$
$$b_n = \omega(F) - b_n^{'}$$

3 teorema: Tarkime, baigtiniam a $\int_a^{\omega(F)} 1 - F(y)dy < \infty$. Apibrėžiame skirstinio funkciją

$$R(t) = \frac{\int_t^{\omega(F)} (1 - F(y))dy}{1 - F(t)}, \ \alpha(F) < t < \omega(F).$$

Jei $\lim_{t\to\omega(F)} \frac{1 - F(t + xR(t))}{1 - F(t)} = e^{-x}, \ x \in R$, tada $P\left(\frac{Z_n - a_n}{b_n}\right) = e^{-e^{-x}}, \ x \in R$.

Normalizavimo konstantas galima parinkti tokiu būdu:

$$a_n : \ 1 - F(a_n) = \frac{1}{n};$$
$$b_n : \ b_n = R(a_n).$$

2. TIRIAMOJI DALIS

Šioje dalyje spręsime tokias problemas:

- Pareto skirstinio atveju skaičiuosime maksimumo vidurkius, ir tiksliąsias reikšmes palyginsime su žinomais įverčiais.
- Didelių imčių atveju panaudosime maksimumų vidurkių ribinę teoremą ir konvergavimo greičių įverčius.
- Panašias problemas spręsime Buro skirstinio atveju. Bet to, šiuo atveju tirsime atsitiktinio didumo imčių maksimumų vidurkius.

2.1 PARETO ATSITIKTINIŲ DYDŽIŲ MAKSIMUMO VIDURKIAI

Atsitiktinio dydžio X skirstinį vadiname Pareto, jei jo skirstinio funkcija

$$F(x) = 1 - \frac{1}{x^{\alpha}}, x \geq 1, \alpha > 0\,, \tag{2.1}$$

tada tankis

$$p(x) = \frac{\alpha}{x^{\alpha+1}}\,.$$

Vidurkis ir dispersija yra tokie:

$$MX = \frac{\alpha}{\alpha - 1}, kai\ \alpha > 1, \quad DX = \frac{\alpha}{(\alpha-1)(\alpha-2)}, kai\ \alpha > 2\,.$$

Maksimumo skirstinio funkcija pagal (1.1) formulę yra

$$F_{Z_n}(x) = \left(1 - \frac{1}{x^{\alpha}}\right)^n .$$

Nagrinėkime statistiką $Z_n^k = \left(\max\left(X_1, X_2, ..., X_n\right)\right)^k$. Šios statistikos skirstinio funkcija

$$F_{Z_n^k}(x) = P\left(Z_n^k < x\right) = P\left(Z_n < x^{1/k}\right) = F^n\left(x^{1/k}\right) = \left(1 - \frac{1}{x^{\alpha/k}}\right)^n .$$

Skaičiuojame maksimumo k-tosios eilės momentą pagal (1.9) formulę:

$$MZ_n^k = 1 + \int_1^{\infty}\left(1 - \left(1 - \frac{1}{x^{\alpha/k}}\right)^n\right)dx = 1 + \int_1^{\infty}\left(1 - \sum_{i=0}^{n}(-1)^i C_n^i \frac{1}{x^{\alpha \cdot i/k}}\right)dx =$$

$$= 1 + \sum_{i=1}^{n}(-1)^{i+1} C_n^i \int_1^{\infty} \frac{1}{x^{\alpha \cdot i/k}}dx = 1 + \sum_{i=1}^{n}(-1)^{i+1} C_n^i \frac{1}{\alpha \cdot i/k - 1}\,.$$

Taigi

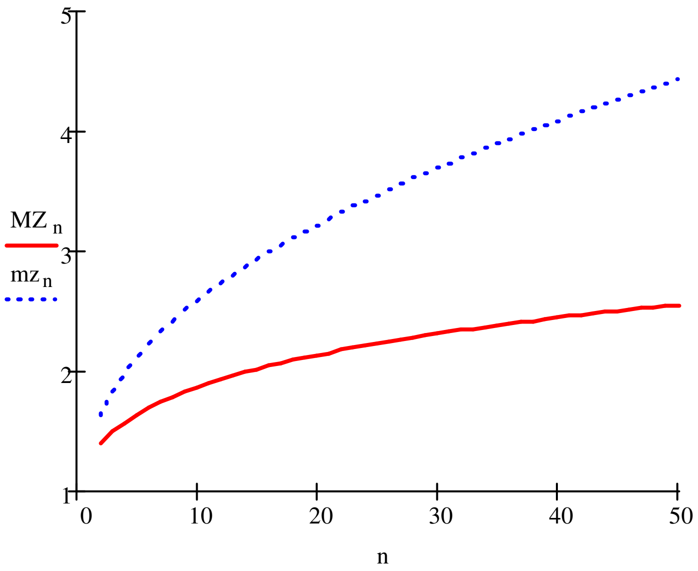
$$MZ_n^k = 1 + \sum_{i=1}^{n} (-1)^{i+1} C_n^i \frac{1}{\alpha \cdot i / k - 1}.$$

Kai k = 1, turime maksimumo vidurkį

$$MZ_n = 1 + \sum_{i=1}^{n} (-1)^{i+1} C_n^i \frac{1}{\alpha \cdot i - 1}.$$

Žinodami Pareto skirstinio vidurkį ir dispersiją galime apskaičiuoti maksimumo vidurkio įvertį (1.15) ir jį palyginti su tiksliai apskaičiuotomis maksimumo vidurkio reikšmėmis.

Įvertį pažymėkime $mz = MX + \frac{(n-1)\sqrt{DX}}{\sqrt{2n-1}}$.



2.1 pav. Maksimumo vidurkio reikšmių palyginimas su įverčio reikšmėmis, kai *α*=5

Kai n didėja, tikslių maksimumo vidurkio reikšmių ir įverčio skirtumas didėja, todėl Pareto skirstinio atveju šis įvertis yra grubus. Kuo parametras α yra didesnis, tuo santykinė paklaida yra mažesenė. Santykinės paklaidos prie skirtingų parametrų α yra pateiktos 5 priede.

Tikslaus vidurkio skaičiavimas yra pakankamai sudėtingas, todėl dažnai kai n dideli, tikslūs skirstiniai yra keičiami ribiniais skirstiniais. Tam naudojamos ribinės teoremos (1.5 skyrius), pagal kurias randamas ribinis skirstinys.

Pareto skirstiniui, kurio skirstinio funkcija yra (2.1) taikysime vieną iš 1.5 skyriaus teoremų. Kadangi $\omega(F) = +\infty$, todėl tinka 1 arba 3 teorema. Tikriname 1 ribinės teoremos sąlygą: ar egzistuoja $\alpha > 0$, toks, kad $\forall x > 0$

$$\lim_{t \to +\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}.$$

Pareto skirstinio atveju gauname:

$$\lim_{t\to+\infty}\frac{1-\left(1-\dfrac{1}{(tx)^{\alpha}}\right)}{1-\left(1-\dfrac{1}{t^{\alpha}}\right)}=\lim_{t\to+\infty}\frac{\dfrac{1}{t^{\alpha}x^{\alpha}}}{\dfrac{1}{t^{\alpha}}}=\lim_{t\to+\infty}\frac{t^{\alpha}}{t^{\alpha}x^{\alpha}}=x^{-\alpha}.$$

Taigi egzistuoja toks $\alpha>0$, su kuriuo $\lim\limits_{t\to+\infty}\dfrac{1-F(tx)}{1-F(t)}=x^{-\alpha}$. Rasime normalizavimo konstantas $a_n, b_n>0$, su kuriomis $\lim\limits_{n\to\infty}P\left(\dfrac{Z_n-a_n}{b_n}<x\right)=H_{1,\alpha}(x)=\begin{cases}e^{-x^{-\alpha}},\ x>0\\ 0,\ x\le 0\end{cases}$.

Remiantis 1-ąją teorema, gauname

$$a_n=0\ ,\ b_n:\ 1-F(b_n)=\frac{1}{n}\Rightarrow 1-\left(1-\frac{1}{b_n^{\alpha}}\right)=\frac{1}{n}\Rightarrow b_n=n^{\frac{1}{\alpha}};$$

Pareto skirstinio atveju gauname:

$$\lim_{n\to\infty}P\left(\frac{Z_n}{n^{1/\alpha}}<x\right)=H_{1,\alpha}(x)=\begin{cases}e^{-x^{-\alpha}},\ x>0\\ 0,\ x\le 0\end{cases}$$

Taigi, kai n yra didelis, maksimumo skirstinį galime aproksimuoti ribiniu skirstiniu, t.y.

$$P\left(\frac{Z_n}{n^{1/\alpha}}<x\right)\approx e^{-x^{-\alpha}},\ x\ge 0.$$

Ribinis skirstinys vadinamas Freše skirstiniu. Ieškome jo k-tosios eilės momento (žym. MY_n^k):

$$MY_n^k=\alpha\int_0^{\infty}x^k\cdot x^{-\alpha-1}\cdot e^{-x^{-\alpha}}dx=\begin{bmatrix}x^{-\alpha}=t & x=t^{\frac{-1}{\alpha}}\\ dx=-\dfrac{1}{\alpha}\cdot t^{-\frac{1}{\alpha}-1} & \end{bmatrix}=\alpha\cdot\left(-\frac{1}{\alpha}\right)\int_{\infty}^{0}t^{-\frac{k}{\alpha}}\cdot t^{-\frac{1}{\alpha}(-\alpha-1)}\cdot e^{-t}\cdot t^{-\frac{1}{\alpha}-1}dt=$$

$$=\int_0^{\infty}t^{-\frac{k}{\alpha}+1+\frac{1}{\alpha}-\frac{1}{\alpha}-1}\cdot e^{-t}dt=\int_0^{\infty}t^{-\frac{k}{\alpha}}\cdot e^{-t}dt=\int_0^{\infty}t^{\frac{\alpha-k}{\alpha}-1}\cdot e^{-t}dt=\Gamma\left(\frac{\alpha-k}{\alpha}\right)=\Gamma\left(1-\frac{k}{\alpha}\right),\ k\le\alpha$$

čia

$$\Gamma(\tau)=\int_0^{\infty}t^{\tau-1}e^{-t}dt,\tau>0,$$

$$\Gamma(n)=\Gamma(n-1)!,\ n\in N.$$

Taigi

$$MY_n^k=\Gamma\left(1-\frac{k}{\alpha}\right),$$

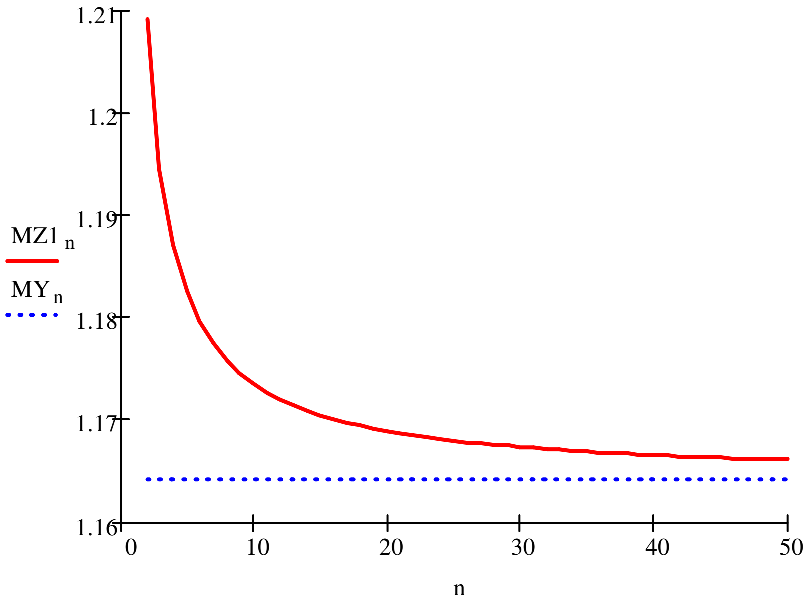
kai $k=1$, gauname Freše skirstinio vidurkį

$$MY_n = \Gamma\left(1-\frac{1}{\alpha}\right).$$

Tada normuotas maksimumo vidurkis, kai n didelis, yra:

$$\frac{MZ_n}{n^{1/\alpha}} \approx \Gamma\left(1-\frac{1}{\alpha}\right). \qquad \textbf{(2.1)}$$

$\frac{MZ_n}{n^{1/\alpha}}$ pažymėkime $MZ1_n$.



2.2 pav. Normuoto maksimumo vidurkio artėjimas prie ribinio skirstinio vidurkio, kai α=5

Keičiant tikslius skirstinius ribiniais skirstiniais svarbu įvertinti konvergavimo greitį $\Delta_n(x) = \left|F^n(xb_n + a_n) - H(x)\right|$.

Ieškosime konvergavimo greičio įverčio Pareto skirstinio atveju. Nagrinėkime skirtumą

$$\Delta_n(x) = \left|P(Z_n < xb_n + a_n) - e^{-x^{-\alpha}}\right| = \left|P(Z_n < xn^{1/\alpha}) - e^{-x^{-\alpha}}\right| = \left|F^n(xn^{1/\alpha}) - e^{-x^{-\alpha}}\right| = \left|\left(1-\frac{1}{(xn^{1/\alpha})^n}\right)^n - e^{-x^{-\alpha}}\right| =$$

$$= \left|\left(1-\frac{1}{nx^\alpha}\right)^n - e^{-x^{-\alpha}}\right| = \left|\exp\left\{\ln\left(1-\frac{1}{nx^\alpha}\right)^n\right\} - e^{-x^{-\alpha}}\right| = \left|\exp\left\{n\ln\left(1-\frac{1}{nx^\alpha}\right)\right\} - e^{-x^{-\alpha}}\right|$$

Pasinaudosime elementariosios funkcijos $\ln x$ reiškimu jos Makloreno eilute:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots, \; -1 \le x \le 1,$$

$$\Delta_n(x) = \left| \exp\left\{ n \left(\frac{-1}{nx^{\alpha}} - \frac{\left(\frac{-1}{nx^2}\right)^2}{2} + \frac{\left(\frac{-1}{nx^3}\right)^3}{3} - \ldots \right) \right\} - \exp\left\{ -x^{-\alpha} \right\} \right| =$$

$$\left| \exp\left\{ -x^{-\alpha} - \frac{1}{2nx^{2\alpha}} - \frac{1}{3n^2x^{3\alpha}} - \ldots \right\} - \exp\left\{ -x^{-\alpha} \right\} \right| \le \left| e^{-x^{-\alpha}} \left(\exp\left\{ -\frac{1}{2nx^{2\alpha}} - 1 \right\} \right) \right|$$

Pasinaudosime elementariosios funkcijos e^x reiškimu jos Makloreno eilute:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots, x \in R,$$

$$\Delta_n(x) \le \left| e^{-x^{-\alpha}} \left(1 - \frac{1}{2nx^{2\alpha}} + \frac{\left(\frac{-1}{2nx^{2\alpha}}\right)}{2!} + \ldots - 1 \right) \right| \le \left| -\frac{e^{-x^{-\alpha}} x^{-2\alpha}}{2n} \right| = \frac{e^{-x^{-\alpha}} x^{-2\alpha}}{2n}$$

Gavome netolygųjį konvergavimo greičio įvertį Pareto skirstiniui:

$$\Delta_n(x) \le \frac{e^{-x^{-\alpha}} x^{-2\alpha}}{2n}, \; x \ge 0. \qquad \textbf{(2.2)}$$

Dešiniąją nelygybės pusę pažymėkime $\Delta 1_n$.

Žinodami konvergavimo greičio netolygiuosius įverčius galime gauti vidurkių konvergavimo greičių įverčius.

Teorema 2.1. Tarkime, kad X_i yra Pareto atsitiktiniai dydžiai $(i = 1, 2, \ldots, n)$. Tada

$$\left| \frac{MZ_n}{n^{1/\alpha}} - \Gamma\left(1 - \frac{1}{\alpha}\right) \right| \le \frac{1}{2\alpha n} \Gamma\left(2 - \frac{1}{\alpha}\right), \; \alpha > 1.$$

Įrodymas. Naudojantis priešingo įvykio tikimybės apibrėžimu, skirtumą $\Delta_n(x)$ galime užrašyti taip:

$$\Delta_n(x) = \left| 1 - P(Z_n > xb_n + a_n) - e^{-x^{-\alpha}} \right| = \left| -P(Z_n > xn^{1/\alpha}) + 1 - e^{-x^{-\alpha}} \right| =$$

$$\left| -\left(1 - P(Z_n < xn^{1/\alpha})\right) + 1 - e^{-x^{-\alpha}} \right| = \left| \left(1 - e^{-x^{-\alpha}}\right) - \left(1 - P(Z_n < xn^{1/\alpha})\right) \right| =$$

$$= \left| \left(1 - P(Z_n < xn^{1/\alpha}{}_n)\right) - \left(1 - e^{-x^{-\alpha}}\right) \right|$$

(2.2) galime užrašyti taip:

$$\left| \left(1 - P(Z_n < xn^{1/\alpha})\right) - \left(1 - e^{-x^{-\gamma}}\right) \right| \le \Delta 1_n(x) \qquad \textbf{(2.3)}$$

Integruokime abi nelygybės (2.3) puses rėžiuose nuo 0 iki ∞:

$$\int_0^\infty \left|\left(1-P\left(Z_n < xn^{1/\alpha}\right)\right)-\left(1-e^{-x^{-\gamma}}\right)\right|dx \le \int_0^\infty \Delta 1_n(x)dx;$$

$$\left|\int_0^\infty \left(1-P\left(Z_n < xn^{1/\alpha}\right)\right)dx-\int_0^\infty\left(1-e^{-x^{-\gamma}}\right)dx\right| \le \int_0^\infty \Delta 1_n(x)dx;$$

$$\left|\frac{MZ_n}{n^{1/\alpha}}-\Gamma\left(1-\frac{1}{\gamma}\right)\right| \le \int_0^\infty \Delta 1_n(x)dx.$$

Skaičiuojame įverčio $\Delta 1_n$ integralą:

$$\int_0^\infty \Delta 1_n(x)dx = \int_0^\infty \frac{x^{-2\alpha}e^{-x^{-\alpha}}}{2n}dx = \frac{1}{2n}\int_0^\infty x^{-2\alpha}e^{-x^{-\alpha}}dx =$$

$$= \frac{1}{2n}\begin{bmatrix} x^{-\alpha}=t; \quad tx^{\alpha}=1; \; x=\left(\frac{1}{t}\right)^{\frac{1}{\alpha}}; \quad -\alpha x^{-\alpha-1}dx=dt; \\ \frac{1}{x^{\alpha}}=t; \quad x^{\alpha}=\frac{1}{t}; \quad dt=dx^{-\alpha}; \quad dx=\frac{dt}{-\alpha x^{-\alpha-1}}=\frac{dt\cdot x\cdot x^{\alpha}}{-\alpha}\end{bmatrix} =$$

$$= \frac{1}{2n}\left(-\int_{-\infty}^{0} t^2e^{-t}\frac{dt\left(\frac{1}{t}\right)^{\frac{1}{\alpha}}\frac{1}{t}}{-\alpha}\right) = \frac{1}{2\alpha n}\int_0^\infty t^2e^{-t}\left(t^{-1}\right)^{\frac{1}{\alpha}+1}dt = \frac{1}{2\alpha n}\int_0^\infty t^{2-\frac{1}{\alpha}-1}e^{-t}dt = \frac{1}{2\alpha n}\Gamma\left(2-\frac{1}{\alpha}\right).$$

Gautą integralo išraišką pažymėkime $\Delta 2_n(x)$.

Taigi

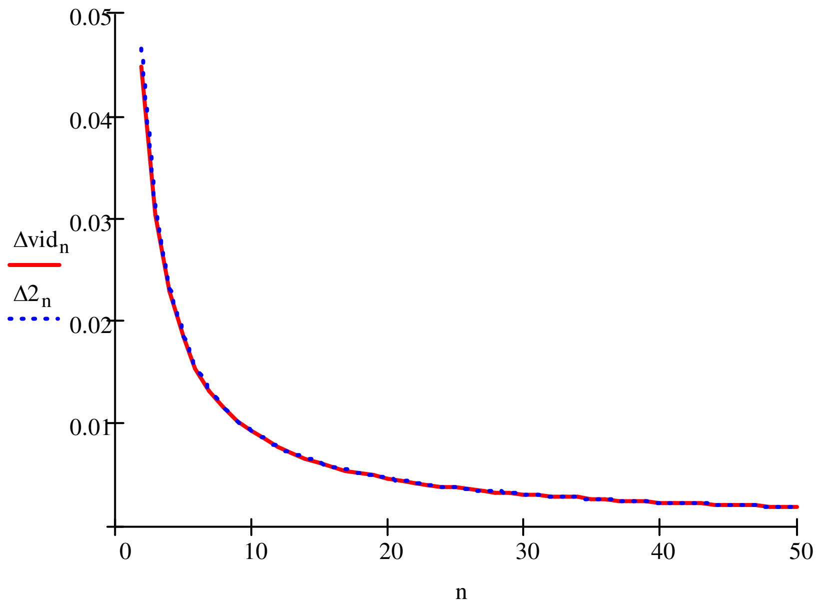
$$\Delta 2_n(x) = \frac{1}{2\alpha n}\Gamma\left(2-\frac{1}{\alpha}\right).$$

Tokiu būdu

$$\left|\frac{MZ_n}{n^{1/\alpha}}-\Gamma\left(1-\frac{1}{\gamma}\right)\right| \le \frac{1}{2\alpha n}\Gamma\left(2-\frac{1}{\alpha}\right), \; \alpha>1. \tag{2.4}$$

Teorema įrodyta.

Normuoto ir ribinio maksimumų vidurkių skirtumo modulį $\left|\frac{MZ_n}{n^{1/\alpha}}-\Gamma\left(1-\frac{1}{\gamma}\right)\right|$ pažymėkime Δvid_n ir jį palyginkime su konvergavimo greičio įverčiu $\Delta 2_n$. Iš grafiko matyti, kad Δvid_n reikšmės artėja prie $\Delta 2_n$ reikšmių, kai n didėja.

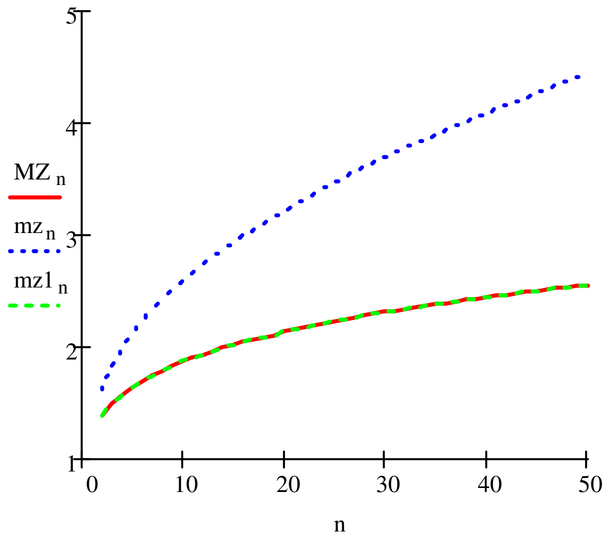


2.3 Normuoto ir ribinio maksimumų vidurkių skirtumo palyginimas su gautu įverčiu, kai *α*=5

Iš (2.4) nelygybės išreiškę MZ_n gauname:

$$1 \le MZ_n \le n^{1/\alpha}\left(\frac{1}{2\alpha n}\Gamma\left(2-\frac{1}{\alpha}\right)+\Gamma\left(1-\frac{1}{\alpha}\right)\right),\ \alpha \ge 1\,.$$

Dešiniąją nelygybės pusę pažymėkime $mz1_n$ ir šį įvertį palyginkime su įverčiu mz_n.



2.4 pav. Įverčių *mz* ir *mz1* palyginimas su tiksliomis maksimumo vidurkio reikšmėmis

Matome, kad įvertis *mz1* daug tiksliau įvertina maksimumo vidurkį, nei įvertis *mz*. Absoliutinės ir santykinės šių įverčių paklaidos pateiktos priede (7 priedas).

2.2 BURO ATSITIKTINIŲ DYDŽIŲ MAKSIMUMO VIDURKIS

Atsitiktinio dydžio X skirstinį vadiname Buro skirstiniu, jei jo skirstinio funkcija yra

$$F(x) = 1 - \left(\frac{1}{1+x^\alpha}\right)^\tau, x \geq 0,\ \alpha > 0, \tau > 0.$$

Yra žinomi k-tosios eilės Buro skirstinio momentai [5]:

$$MX^n = \frac{\Gamma\left(\tau - \frac{k}{\alpha}\right)\Gamma\left(\tau + \frac{k}{\alpha}\right)}{\Gamma(\tau)}.$$

Tirsime atvejį, kai $\tau = 1$. Tada

$$F(x) = 1 - \frac{1}{1+x^\alpha}, x \geq 0,\ \alpha > 0.$$

Vidurkis ir dispersija yra tokie:

$$MX = \Gamma\left(1 - \frac{1}{\alpha}\right)\Gamma\left(1 + \frac{1}{\alpha}\right),\ \alpha > 1,$$

$$DX = \Gamma\left(1 - \frac{2}{\alpha}\right)\Gamma\left(1 + \frac{2}{\alpha}\right),\ \alpha > 2.$$

Maksimumo skirstinio funkcija

$$F_{Z_n}(x) = \left(1 - \frac{1}{1+x^\alpha}\right)^n.$$

Naudojantis (1.9) formule, skaičiuojame maksimumo vidurkį:

$$MZ_n = \int_0^\infty \left(1 - \left(1 - \frac{1}{1+x^\alpha}\right)^n\right)dx = \int_0^\infty \left(1 - \sum_{i=0}^{n} (-1)^i C_n^i \left(\frac{1}{1+x^\alpha}\right)^k\right)dx =$$

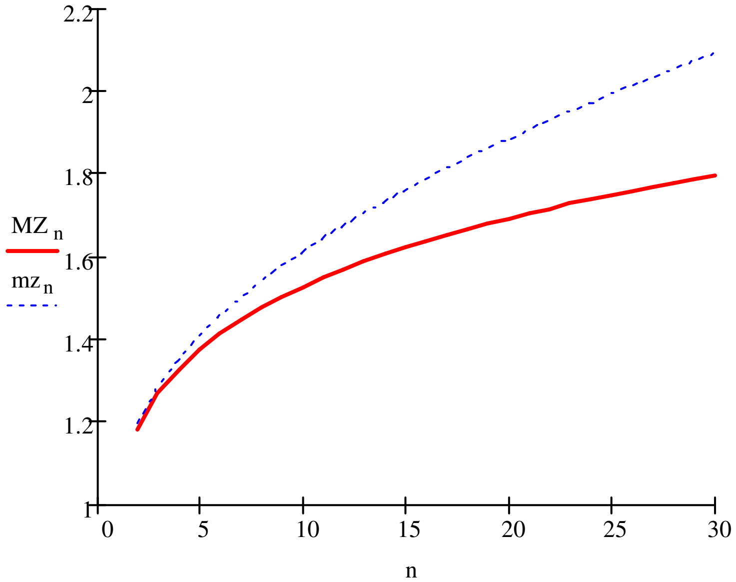
$$= \sum_{i=0}^{\infty} (-1)^{i+1} C_n^i \int_0^\infty \frac{1}{\left(1+x^\alpha\right)^k} dx = \sum_{i=0}^{\infty} (-1)^{i+1} C_n^i \frac{\Gamma\left(i - \frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\alpha\Gamma(i)},$$

čia $C_n^i = \frac{n!}{i!(n-i)!}$, ir gama funkcija

$$\Gamma(\tau) = \int_0^\infty t^{\tau-1} e^{-t} dt, \tau > 0,$$

$$\Gamma(n) = \Gamma(n-1)!,\ n \in N.$$

Apskaičiuotas maksimumo vidurkio reikšmes lyginame su įverčiu $mz = MX + \frac{(n-1)\sqrt{DX}}{\sqrt{2n-1}}$.



2.5 Maksimumo vidurkio reikšmių palyginimas su įverčio reikšmėmis, kai α=7

Ribinės teoremos taikymas. Tikriname pirmosios ribinės teoremos sąlygą: ar egzistuoja $\alpha > 0$, toks, kad $\forall x > 0$

$$\lim_{t \to +\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \text{ kai } \omega(F) = +\infty .$$

Gauname, kad

$$\lim_{t \to +\infty} \frac{1 - \left(1 - \dfrac{1}{1 + (tx)^{\alpha}}\right)}{1 - \left(1 - \dfrac{1}{1 + t^{\alpha}}\right)} = \lim_{t \to +\infty} \frac{1 + t^{\alpha}}{1 + t^{\alpha} x^{\alpha}} = x^{-\alpha} .$$

Teoremos sąlyga yra tenkinama, taigi

$$\lim_{n \to \infty} P\left(\frac{Z_n - a_n}{b_n} < x\right) = H_{1,\alpha}(x) = \begin{cases} e^{-x^{-\alpha}}, & x > 0 \\ 0, & x \le 0 \end{cases} .$$

Normalizavimo konstantas galima parinkti tokiu būdu:

$$a_n = 0; \quad b_n : \; 1 - F(b_n) = \frac{1}{n} .$$

$$1 - \left(1 - \frac{1}{1 + b_n^{\alpha}}\right) = \frac{1}{n} \Rightarrow 1 + b_n^{\alpha} = n \Rightarrow b_n = (n-1)^{\frac{1}{\alpha}} \approx n^{\frac{1}{\alpha}} .$$

Pritaikius ribinę teoremą Buro skirstiniui gauname

$$\lim_{n\to\infty} P\left(\frac{Z_n}{n^{\frac{1}{\alpha}}} < x\right) = H_{1,\alpha}(x) = \begin{cases} e^{-x^{-\alpha}}, & x > 0 \\ 0, & x \le 0 \end{cases}.$$

Taigi, kai n yra didelis, maksimumo skirstinį galime aproksimuoti ribiniu skirstiniu, t.y.

$$P\left(\frac{Z_n}{n^{\frac{1}{\alpha}}} < x\right) \approx e^{-x^{-\alpha}},\ x \ge 0$$

Ribinis skirstinys $H_{1,\alpha}(x) = e^{-x^{-\alpha}}$ vadinamas Freše skirstiniu. Jo k-tosios eilės momentas yra

$$MY_n^k = \Gamma\left(1 - \frac{k}{\alpha}\right),$$

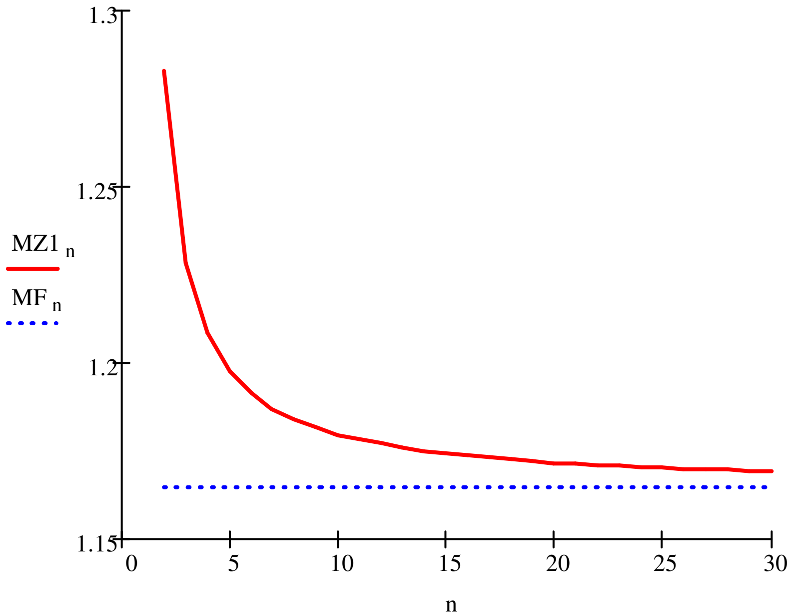
kai $k = 1$ gauname Freše skirstinio vidurkį

$$MY = \Gamma\left(1 - \frac{1}{\alpha}\right).$$

Normuotas Buro skirstinio maksimumo vidurkis, kai n dideli, yra

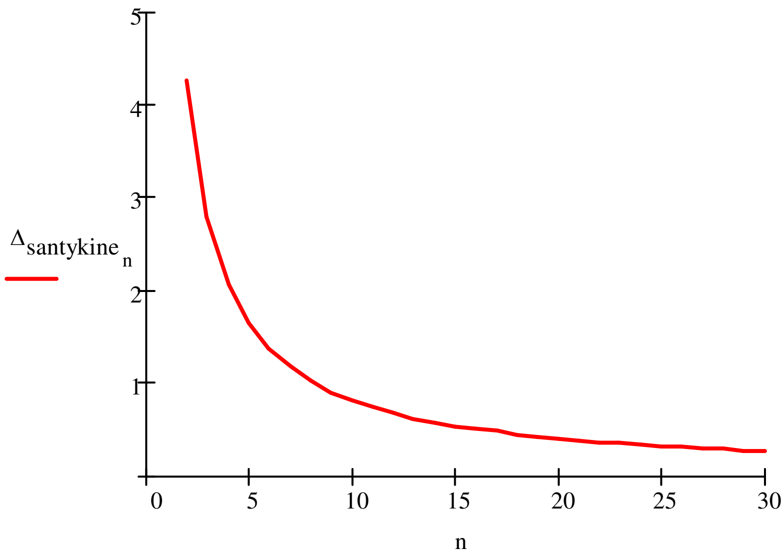
$$\frac{MZ_n}{n^{\frac{1}{\alpha}}} \approx \Gamma\left(1 - \frac{1}{\alpha}\right).$$

$\frac{MZ_n}{n^{\frac{1}{\alpha}}}$ pažymėkime $MZ1_n$. Žinodami tikslias MZ_n reikšmes, galime palyginti normuotą maksimumo vidurkį su ribinio Freše skirstinio vidurkiu, t.y.



2.6 Normuoto maksimumo vidurkio artėjimas į ribinio skirstinio vidurkį, kai α=7

Keičiant tikslius skirstinius ribiniais atsiranda paklaidos. Didėjant imties didumui *n* santykinė paklaida mažėja, vadinasi, normuotą Buro maksimumo skirstinio vidurkį galime aproksimuoti ribinio skirstinio Freše vidurkiu, kai *n* didelis.



2.7 Santykinės paklaidos grafikas didėjant *n*

Didėjant parametrui α, skirtumas tarp normuotų Buro maksimumų vidurkių ir Freše skirstinio vidurkių mažėja. Vidurkių reikšmės ir santykinės paklaidos prie skirtingų α reikšmių pateiktos 6 priede.

2.3 ATSITIKTINIO KOMPONENTŲ SKAIČIAUS MAKSIMUMO VIDURKIAI

Nagrinėkime struktūrą $Z_N = \max(X_1, X_2, .., X_N)$, čia X_i, $i = \overline{1, N}$, pasiskirstęs pagal Buro skirstinį, su skirstinio funkcija

$$F(x) = 1 - \frac{1}{1 + x^{\alpha}}, x \geq 0, \ \alpha > 0,$$

o atsitiktinis dydis N pasiskirstęs pagal geometrinį skirstinį ir nepriklauso nuo visų X_i, $i = \overline{1, N}$, t.y.

$$P(N = k) = pq^{k-1}, \ k \geq 1, \ p + q = 1, \ p > 0.$$

Teorema 2.2. Jeigu nepriklausomų atsitiktinių dydžių X_i, $i \geq 1$ skirstiniai yra Buro, o N – geometrinis, tai

$$MZ_N = p^{-1/\alpha} MX ,$$

čia X – Buro atsitiktinis dydis.

Įrodymas. Panaudoję pilnosios tikimybės formulę gauname:

$$P(Z_N \leq x) = \sum_{k=1}^{\infty} F^k(x) P(N = k) = \sum_{k=1}^{\infty} F^k(x) p (1-p)^{k-1} = \frac{pF(x)}{1 - (1-p)F(x)}.$$

Buro skirstinio atveju

$$P(Z_N \leq x) = \frac{p\left(1 - \frac{1}{1 + x^{\alpha}}\right)}{1 - (1-p)\left(1 - \frac{1}{1 + x^{\alpha}}\right)}.$$

Imdami normavimo konstantas $a = 0$ ir $b = p^{-1/\alpha}$ gauname:

$$P\left(\frac{Z_N - a}{b} < x\right) = \frac{p\left(1 - \frac{1}{1 + x^{\alpha} p^{-1}}\right)}{1 - (1-p)\left(1 - \frac{1}{1 + x^{\alpha} p^{-1}}\right)} = \frac{x^{\alpha}}{1 + x^{\alpha}} = 1 - \frac{1}{1 + x^{\alpha}}.$$

Tokiu būdu

$$P\left(\frac{Z_N - a}{b} < x\right) = F(x)$$

Ir Buro skirstinys yra geometriškai maks – stabilus

$$P\left(Z_N p^{1/\alpha} < x\right) = 1 - \frac{1}{1 + x^{\alpha}}.$$

Tada

$$MZ_N = p^{-1/\alpha} MX ,$$

Čia Buro atsitiktinio dydžio X vidurkis

$$MX = \Gamma\left(1-\frac{1}{\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right), \ \alpha > 1.$$

Teorema įrodyta.

IŠVADOS

1. Įvertis $MZ_n \le \mu + \dfrac{(n-1)\sigma}{\sqrt{2n-1}}$ Pareto skirstiniams yra netikslus, tačiau yra skirstinių, kuriems galioja lygybės ženklas.
2. Kai n dideli, Pareto dydžių maksimumo vidurkius galima keisti Freše skirstinio vidurkiais ir įvertinti paklaidas.
3. Buro atsitiktinių dydžių imtims įvertis $MZ_n \le \mu + \dfrac{(n-1)\sigma}{\sqrt{2n-1}}$ yra taip pat netikslus. Imant atsitiktinį komponenčių skaičių maksimumo vidurkiai išreiškiami Buro dydžio vidurkiu ir parametru p.

LITERATŪRA

1. Aksomaitis A. Tikimybių teorija ir statistika. Kaunas: Technologija, 2000, 344 psl.
2. David H. A., Nagraja H. N. Order Statictics. Wiley – Interscience, 2003, 458 psl.
3. Галамбош Я. Асимптотическая теория экстремальных порядковых статистик. - Москва: Наука, 1984
4. Embrechets P., Kluppelbrg C., Mikosch T. Modelling Extremal Events for Insuarance and Finance, Springer, 1997, 648 psl.
5. Robert V. Hogg, Stuart A. Klygman, Loss Distribution, John Wiley & Sons, 1984, 235 psl.
6. Matematikos terminų žodynas, Mokslo ir enciklopedijų leidykla, Vilnius, 1984, 726 psl.

1 PRIEDAS PRANEŠIMO „EKTREMUMŲ VIDURKIO ĮVERČIAI“ MEDŽIAGA

VI taikomosios matematikos studentų konferencija

Įvadas

Tarkime, kad atsitiktinio dydžio skirstinys yra Pareto:

$$F(x) = 1 - \frac{1}{x^{\alpha}},\ x \geq 1.$$

Sudarome struktūrą

$$Z_n = \max(X_1, ..., X_n).$$

Skaičiuosime MZ_n ir jį palyginsime su vidurkio įverčiu, pateiktu [1] monografijoje (66psl.).

Rezultatų formuluotė, pagrindimas ir išvados

$$P(Z_n < x) = F^n(x) = \left(1 - \frac{1}{x^{\alpha}}\right)^n;$$

Tada:

$$MZ_n = 1 + \int_1^{\infty}\left(1 - \left(1 - \frac{1}{x^{\alpha}}\right)^n\right)dx = 1 + \int_1^{\infty}\left(1 - \sum_{k=0}^{n}(-1)^k C_n^k \frac{1}{x^{\alpha k}}\right)dx =$$

$$= 1 + \int_1^{\infty}\left(1 - 1 - \sum_{k=1}^{n}(-1)^k C_n^k \frac{1}{x^{\alpha k}}\right)dx = 1 + \int_1^{\infty}\sum_{k=1}^{n}(-1)^{k+1} C_n^k \frac{1}{x^{\alpha k}}\,dx = 1 + \sum_{k=1}^{n}(-1)^{k+1} C_n^k \int_1^{\infty}\frac{1}{x^{\alpha k}}\,dx =$$

$$= 1 + \sum_{k=1}^{n}(-1)^{k+1} C_n^k \left.\frac{x^{-\alpha k+1}}{-\alpha k - 1}\right|_1^{\infty} = 1 + \sum_{k=1}^{n}(-1)^{k+1} C_n^k \frac{1}{\alpha k - 1}.$$

Taigi

$$MZ_n = 1 + \sum_{k=1}^{n}(-1)^{k+1} C_n^k \frac{1}{\alpha \cdot k - 1}.$$

Monografijoje [1] yra pateiktas ir įrodytas ryšys: jei egzistuoja $DX = \sigma^2$, tai

$$MZ_n \leq m + \frac{(n-1)\sigma}{\sqrt{(2n-1)}}. \qquad \textbf{(1)}$$

Pareto dydžio atveju:

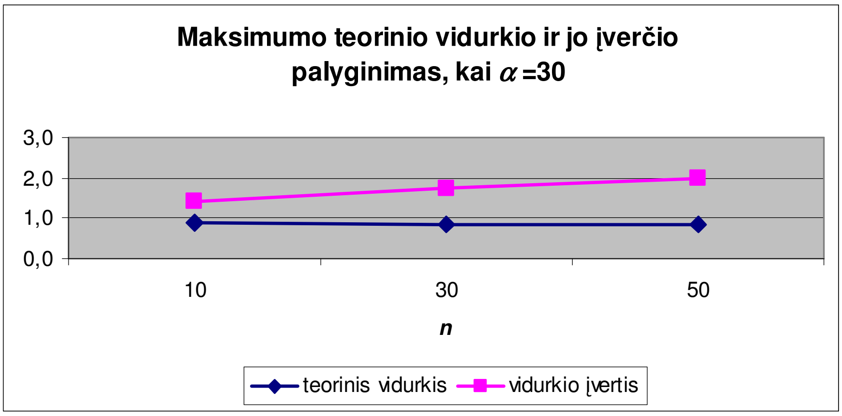
$$m = \frac{\alpha}{\alpha - 1}, kai\ \alpha > 1,\quad \sigma^2 = \frac{\alpha}{(\alpha-1)(\alpha-2)}, kai\ \alpha > 2.$$

Pateiksime lentelę, kurioje apskaičiuotos maksimumo vidurkių reikšmės prie skirtingų n bei α reikšmių, bei apskaičiuotas jų įvertis pagal **(1)** formulę.

Maksimumo teorinio vidurkio ir jo įverčio palyginimas kintant n bei α reikšmėms

α \ n	10		30		50	
	MZ_n	įvertis	MZ_n	įvertis	MZ_n	įvertis
7	0,4542	2,1640	0,19879	2,9904	0,0640	3,5455
10	0,6487	1,8806	0,49621	2,51815	0,4184	2,9464
15	0,7800	1,6642	0,68998	2,15531	0,6452	2,4852
20	0,8400	1,5519	0,77638	1,96564	0,7451	2,2436
25	0,8743	1,4811	0,82518	1,84514	0,8013	2,0897
30	0,8965	1,4314	0,85652	1,76018	0,8371	1,9811

Iš lentelės duomenų matome, kad kuo didesnė α reikšmė, tuo skirtumas tarp teorinio vidurkio ir jo įverčio yra mažesnis. Didėjant n reikšmei, skirtumas tarp teorinio vidurkio ir jo įverčio didėja.



Taigi, galime padaryti išvadą, jog **(1)** formulė nėra labai tiksli. Mūsų nagrinėjamu Pareto skirstinio atveju lygybė nėra pasiekiama, tačiau monografijoje [1] yra įrodyta, kad ji pasiekiama tolygaus skirstinio atveju, kai n=2.

Literatūra:

1. Дэйвид Г. Порядковые статистики. – Москва: Наука, 1979. 336 с.

2. Aksomaitis A. Tikimybių teorija ir statistika. Kaunas: Technologija, 2000, 344 psl.

2 PRIEDAS PRANEŠIMO „THE ASYMPTOTICAL ANALYSIS OF STOCHASTIC MAXIMUM MOMENTS" MEDŽIAGA

Konferencija "Matematika ir matematikos dėstymas – 2007"

1. Introduction. Let $(X_1, X_2, \ldots, X_n)$ be a sequence of independent and identically distributed random variables with Pareto distribution function

$$F(x) = 1 - \frac{1}{x^{\gamma}},\ x \ge 1,\ \gamma > 0\,.$$

Let us define a structure

$$Z_n = max(X_1, X_2, \ldots, X_n).$$

Distribution function of maxima

$$F_{Z_n}(x) = \left(1 - \frac{1}{x^{\gamma}}\right)^n.$$

There are many works written about this structure. For example [1], [2].

Maxima mean:

$$MZ_n = 1 + \int_1^{\infty}\left(1 - \left(1 - \frac{1}{x^{\gamma}}\right)^n\right)dx = 1 + \int_1^{\infty}\left(1 - \sum_{k=0}^{n}(-1)^k C_n^k \frac{1}{x^{\gamma k}}\right)dx =$$

$$= 1 + \int_1^{\infty}\sum_{k=1}^{n}(-1)^{k+1} C_n^k \frac{1}{x^{\gamma k}}\,dx = 1 + \sum_{k=1}^{n}(-1)^{k+1} C_n^k \frac{1}{\gamma k - 1}\,.$$

We got

$$MZ_n = 1 + \sum_{k=1}^{n}(-1)^{k+1} C_n^k \frac{1}{\gamma \cdot k - 1}\,.$$

2. Limited distribution function and moments.

We verify condition of maxima limit theorem [1]:

$$\lim_{t\to+\infty}\frac{1 - \left(1 - \dfrac{1}{(tx)^{\gamma}}\right)}{1 - \left(1 - \dfrac{1}{t^{\gamma}}\right)} = \lim_{t\to+\infty}\frac{\dfrac{1}{t^{\gamma}x^{\gamma}}}{\dfrac{1}{t^{\gamma}}} = \lim_{t\to+\infty}\frac{t^{\gamma}}{t^{\gamma}x^{\gamma}} = x^{-\gamma}\,.$$

There is such $\gamma>0$, with which $\lim\limits_{t\to+\infty}\dfrac{1-F(tx)}{1-F(t)} = x^{-\gamma}$. We will find centralizing and normalizing constants $a_n, b_n > 0$, here $n \ge 1$, with which $\lim\limits_{n\to\infty} P\left(\dfrac{Z_n - a_n}{b_n} < x\right) = H_{1,\gamma}(x) = \begin{cases} e^{-x^{-\gamma}},\ x > 0 \\ 0,\ x \le 0 \end{cases}.$

We choose following constants in such way[1]: $a_n = 0$, $b_n = n^{\frac{1}{\gamma}}$.

We get that $\lim\limits_{n\to\infty} P\left(\dfrac{Z_n}{n^{\frac{1}{\gamma}}} < x\right) = H_{1,\gamma}(x) = \begin{cases} e^{-x^{-\gamma}}, & x > 0 \\ 0, & x \le 0 \end{cases}$.

Limited distribution function is called Fechet distribution function. We calculate moments of limited distribution function:

$$MX^k = \gamma \int_0^\infty x^k \cdot x^{-\gamma-1} \cdot e^{-x^{-\gamma}} dx = \begin{bmatrix} x^{-\gamma} = t \\ x = t^{\frac{-1}{\gamma}} \\ dx = -\dfrac{1}{\gamma} \cdot t^{-\frac{1}{\gamma}-1} \end{bmatrix} = \gamma \cdot \left(-\frac{1}{\gamma}\right) \int_0^\infty t^{-\frac{k}{\gamma}} \cdot t^{-\frac{1}{\gamma}(-\gamma-1)} \cdot e^{-t} \cdot t^{-\frac{1}{\gamma}-1} dt =$$

$$= \int_0^\infty t^{-\frac{k}{\gamma}+1+\frac{1}{\gamma}-\frac{1}{\gamma}-1} \cdot e^{-t} dt = \int_0^\infty t^{-\frac{k}{\gamma}} \cdot e^{-t} dt = \int_0^\infty t^{\frac{\gamma-k}{\gamma}-1} \cdot e^{-t} dt = \Gamma\left(\frac{\gamma-k}{\gamma}\right),\ k \le \gamma\ ;$$

here $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

When k=1, we have mean of Frechet distribution:

$$MX = \Gamma\left(\frac{\gamma-1}{\gamma}\right) = \Gamma\left(1-\frac{1}{\gamma}\right),\ \gamma > 1 .$$

3. The Analysis of Convergence Rate. The asymptotical analysis results that Pareto maxima distribution can be approximated by Frechet distribution, when n is large.

Having nonuniform estimates of convergence rate [3]

$$\left| P(Z_n < xb_n + a_n) - e^{-x^{-\gamma}} \right| \le \Delta_n(x) ,$$

We get convergence rate of means:

$$\left| \frac{MZ_n}{b_n} - \Gamma\left(1-\frac{1}{\gamma}\right) \right| \le \int_0^\infty \Delta_n(x) dx .$$

Proof:

$$\left|1-P(Z_n > xb_n)-e^{-x^{-\gamma}}\right| \le \Delta_n(x);$$
$$\left|(1-P(Z_n < xb_n))-\left(1-e^{-x^{-\gamma}}\right)\right| \le \Delta_n(x);$$
$$\int_0^\infty \left|(1-P(Z_n < xb_n))-\left(1-e^{-x^{-\gamma}}\right)\right| dx \le \int_0^\infty \Delta_n(x)dx;$$
$$\int_0^\infty (1-P(Z_n < xb_n))dx-\int_0^\infty \left(1-e^{-x^{-\gamma}}\right)dx \le \int_0^\infty \Delta_n(x)dx;$$
$$\left|\frac{MZ_n}{b_n}-\Gamma\left(1-\frac{1}{\gamma}\right)\right| \le \int_0^\infty \Delta_n(x)dx.$$

[2] results that

$$\Delta_n(x) \le \frac{z_n^2(x)e^{-z_n(x)}}{2}\cdot\frac{1}{n-1}, x \ge 0, \text{ here } z_n(x)=n(1-F(xb_n+a_n)).$$

When distribution function is Pareto distribution function we have

$$z_n(x)=n\left(1-\left(1-\frac{1}{\left(xn^{\frac{1}{\gamma}}\right)^{\gamma}}\right)\right)=n\frac{1}{x^\gamma n}=x^{-\gamma}.$$

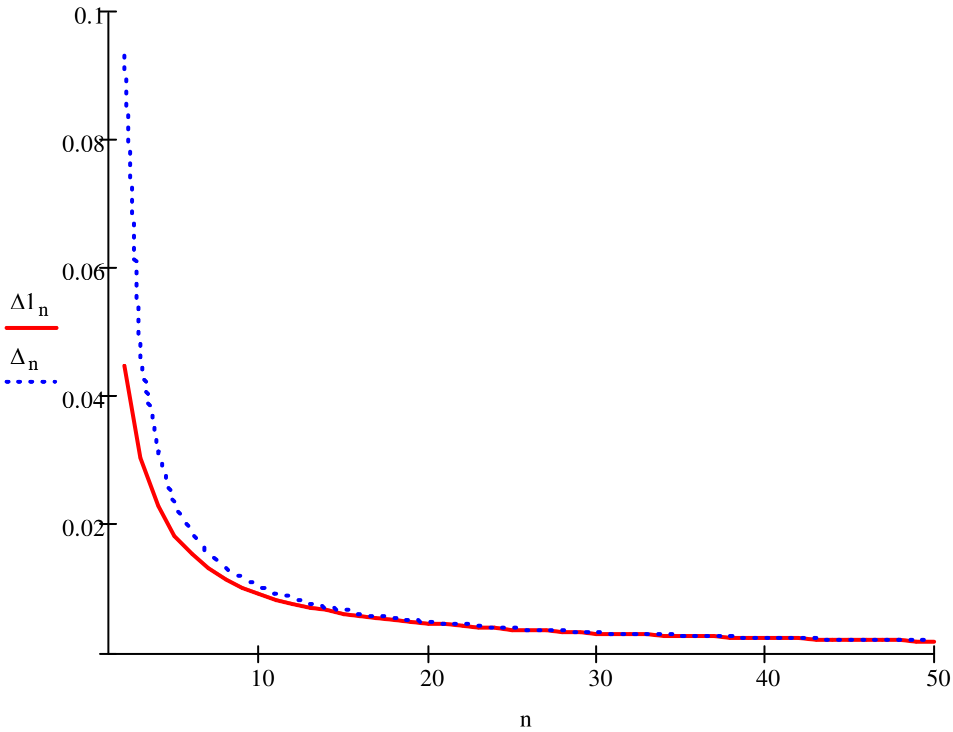
and

$$\Delta_n(x) \le \frac{x^{-2\gamma}e^{-x^{-\gamma}}}{2}\cdot\frac{1}{n-1}, x \ge 0;$$

By calculating the integral of $\Delta_n(x)$ we get:

$$\int_0^\infty \Delta_n(x)dx=\frac{1}{2(n-1)}\int_0^\infty x^{-2\gamma}e^{-x^{-\gamma}}dx=\frac{1}{2\gamma(n-1)}\int_0^\infty t^2e^{-t}\left(t^{-1}\right)^{\frac{1}{\gamma}+1}dt=$$
$$=\frac{1}{2\gamma(n-1)}\Gamma\left(2-\frac{1}{\gamma}\right)$$

Let us mark the difference $\left|\frac{MZ_n}{b_n}-\Gamma\left(1-\frac{1}{\gamma}\right)\right|$ $\Delta 1_n$, and $\int_0^\infty \Delta_n(x)dx$ mark Δ_n. Picture 1 is provided for comparison of $\Delta 1_n$ and Δ_n



Picture 1

References

1. Galambos J., The Asymptotic Theory of Extreme Order Statistics, Wiley, New York, 1978
2. Leadbetter M. R. (1983). Extremes and related properties of random sequences and processes. New York.
3. Aksomaitis A. "Atsitiktinių dydžių maksimumų vidurkiai". Lietuvos matematikų draugijos XXXV konferencija: pranešimų tezės. Vilnius, 1994, p. 236-237.

Stochastinių maksimumų momentų asimptotiniai tyrimai

L. Kasperavičiūtė, A.Aksomaitis

Straipsnyje gauti maksimumų vidurkių konvergavimo greičių įverčiai Pareto skirstinio atveju.

3 PRIEDAS PRANEŠIMO „APIE EKSTREMUMŲ MOMENTUS“ MEDŽIAGA

Konferencija “Matematika ir matematikos dėstymas – 2008“

1. Įvadas. Nagrinėkime atsitiktinę imtį *(X*[1]*, X*[2]*, …, X*[n]*)* iš generalinės aibės su Pareto skirstinio funkcija

$$F(x) = 1 - \frac{1}{x^{\alpha}},\ x \geq 1,\ \alpha > 0.$$

Apibrėžiame struktūrą

$$Z_n = \max\left(X_1, X_2, ..., X_n\right),$$

bei struktūrą

$$Z_n^k = \left(\max\left(X_1, X_2, ..., X_n\right)\right)^k.$$

Maksimumo skirstinio funkcija

$$F_{Z_n}(x) = \left(1 - \frac{1}{x^{\alpha}}\right)^n.$$

Skaičiuojame k-tosios eilės maksimumo momentą:

$$MZ_n^k = 1 + \int_1^{\infty}\left(1 - \left(1 - \frac{1}{x^{\alpha/k}}\right)^n\right)dx = 1 + \int_1^{\infty}\left(1 - \sum_{i=0}^{n}(-1)^i C_n^i \frac{1}{x^{\alpha\cdot i/k}}\right)dx =$$

$$= 1 + \sum_{i=1}^{n}(-1)^{i+1} C_n^i \int_1^{\infty}\frac{1}{x^{\alpha\cdot i/k}}dx = 1 + \sum_{i=1}^{n}(-1)^{i+1} C_n^i \frac{1}{\alpha\cdot i/k - 1}.$$

Taigi

$$MZ_n^k = 1 + \sum_{i=1}^{n}(-1)^{i+1} C_n^i \frac{1}{\alpha\cdot i/k - 1}.$$

Kai $k = 1$, turime maksimumo vidurkį

$$MZ_n = 1 + \sum_{i=1}^{n}(-1)^{i+1} C_n^i \frac{1}{\alpha\cdot i - 1}.$$

2. Maksimumo vidurkis ir jo įvertis.

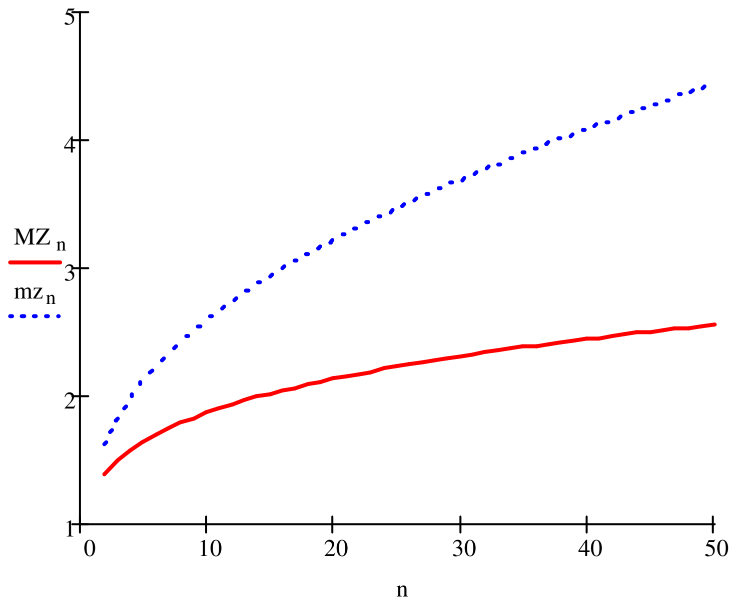
Monografijoje [1] yra pateiktas ir įrodytas ryšys: jei egzistuoja $DX = \sigma^2$, tai

$$MZ_n \leq m + \frac{(n-1)\sigma}{\sqrt{(2n-1)}}.$$

Žinodami Pareto skirstinio vidurkį ir dispersiją

$$m = \frac{\alpha}{\alpha - 1}, kai\ \alpha > 1,\quad \sigma^2 = \frac{\alpha}{(\alpha-1)(\alpha-2)}, kai\ \alpha > 2,$$

galime palyginti teorines maksimumo vidurkio reikšmes su įverčio reikšmėmis (1 pav.)



1 pav. Teorinio vidurkio ir jo įverčio reikšmių palyginimas

3. Ribinis skirstinys. Turimai Pareto skirstinio funkcijai pritaikysime ribinę teoremą, pateiktą monografijoje [2]:

Tarkime, kad $\omega(F) = +\infty$. Jeigu egzistuoja $\gamma > 0$, toks, kad $\forall x > 0$

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

tada egzistuoja normalizavimo konstantos $a_n, b_n > 0$, *čia* $n \geq 1$, su kuriomis

$$\lim_{n \to \infty} P\left(\frac{Z_n - a_n}{b_n} < x \right) = H_{1,\gamma}(x) = \begin{cases} e^{-x^{-\gamma}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Normalizavimo konstantas galima parinkti tokiu būdu:

$$a_n = 0; \quad b_n : \; 1 - F(b_n) = \frac{1}{n}.$$

Pareto skirstinio atveju gauname:

$$\lim_{n \to \infty} P\left(\frac{Z_n^k - a_n}{b_n} < x \right) = H_{1, \frac{\alpha}{k}}(x) = \begin{cases} e^{-x^{-\frac{\alpha}{k}}}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

normalizavimo konstantos $a_n = 0; \quad b_n = n^{\frac{k}{\alpha}}$.

Kai n yra didelis, maksimumo skirstinį galime aproksimuoti ribiniu skirstiniu, t.y.

$$P\left(\frac{Z_n^k}{n^{k/\alpha}} < x \right) \to e^{-x^{-\alpha/k}}, \; x \geq 0, \; n \to \infty.$$

Tada normuotas k-tosios eilės momentas:

$$\frac{MZ_n^k}{n^{k/\alpha}} \approx \int_0^\infty x de^{-x^{-\alpha/k}}, \ kai \ n \to \infty .$$

Ieškome ribinio skirstinio k-tosios eilės momento (žym. MF_n^k):

$$MF_n^k = \int_0^\infty x de^{-x^{-\alpha/k}} = \left[\frac{\alpha}{k} = \gamma\right] = \int_0^\infty x\gamma\, x^{-\gamma-1} e^{-x^{-\gamma}} dx = \gamma \int_0^\infty x^{-\gamma} e^{-x^{-\gamma}} dx =$$

$$= \begin{bmatrix} x^{-\gamma} = t; & dx = -\frac{1}{\gamma} t^{-\frac{1}{\gamma}-1} \\ x = t^{-1/\gamma}; & \end{bmatrix} = \gamma \int_0^\infty t e^{-t} \frac{1}{\gamma} t^{-\frac{1}{\gamma}-1} dt = \int_0^\infty e^{-t} t^{1-\frac{1}{\gamma}-1} dt = \Gamma\left(1 - \frac{1}{\gamma}\right) = \Gamma\left(1 - \frac{k}{\alpha}\right),$$

čia $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$

Taigi

$$MF_n^k = \Gamma\left(1 - \frac{k}{\alpha}\right).$$

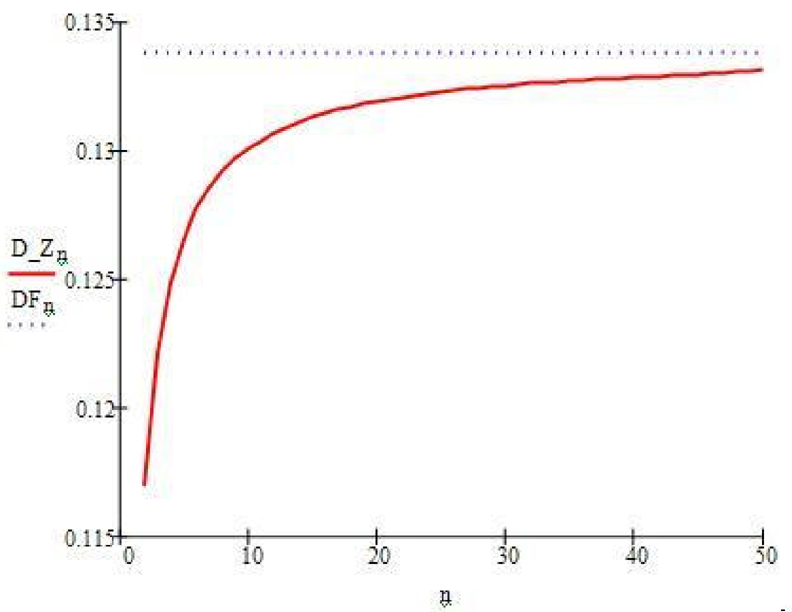
Kai $k = 1$, turime

$$\frac{MZ_n}{n^{1/\alpha}} \approx \Gamma\left(1 - \frac{1}{\alpha}\right)$$

Maksimumo dispersija yra

$$DZ_n = MZ_n^2 - (MZ_n)^2 .$$

Žinodami pirmosios ir antrosios eilės normuotus maksimumo momentus bei ribinių skirstinių momentus, galime palyginti ir dispersijas (2 pav.)



2 pav. Normuotų ir ribinių maksimumo dispersijų palyginimas

Literatūra

4. Дэйвид Г. Порядковые статистики. – Москва: Наука, 1979. 336 с
5. Galambos J., The Asymptotic Theory of Extreme Order Statistics, Wiley, New York, 1978
6. Aksomaitis A. Tikimybių teorija ir statistika. Kaunas: Technologija, 2000, 344 psl.

EXTREMA MOMENTS

L. Kasperavičiūtė, A. Aksomaitis

The analysis of the maxima moments when the distribution funcion is Pareto are presented.

4 PRIEDAS PRANEŠIMO „STOCHASTINIŲ MAKSIMUMŲ MOMENTAI“ MEDŽIAGA

VII taikomosios matematikos studentų konferencija

Nagrinėkime atsitiktinę imtį $(X_1, X_2, \ldots, X_n)$ iš generalinės aibės su Pareto skirstinio funkcija

$$F(x) = 1 - \frac{1}{x^\alpha},\ x \ge 1,\ \alpha > 0$$

Apibrėžiame struktūrą

$$Z_n^k = \left(\max\left(X_1, X_2, \ldots, X_n\right)\right)^k,$$

Skaičiuojame k-tosios eilės maksimumo momentą:

$$MZ_n^k = 1 + \int_1^\infty \left(1 - \left(1 - \frac{1}{x^{\alpha/k}}\right)^n\right) dx = 1 + \int_1^\infty \left(1 - \sum_{i=0}^n (-1)^i C_n^i \frac{1}{x^{\alpha \cdot i/k}}\right) dx =$$

$$= 1 + \sum_{i=1}^n (-1)^{i+1} C_n^i \int_1^\infty \frac{1}{x^{\alpha \cdot i/k}} dx = 1 + \sum_{i=1}^n (-1)^{i+1} C_n^i \frac{1}{\alpha \cdot i/k - 1}.$$

Kai $k = 1$, turime maksimumo vidurkį

$$MZ_n = 1 + \sum_{i=1}^n (-1)^{i+1} C_n^i \frac{1}{\alpha \cdot i - 1}.$$

Turimai Pareto skirstinio funkcijai pritaikysime ribinę teoremą, pateiktą monografijoje [1]: Tarkime, kad $\omega(F) = +\infty$. Jeigu egzistuoja $\gamma > 0$, toks, kad $\forall x > 0$

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

tada egzistuoja normalizavimo konstantos $a_n,\ b_n > 0$, čia $n \ge 1$, su kuriomis

$$\lim_{n \to \infty} P\left(\frac{Z_n - a_n}{b_n} < x\right) = H_{1,\gamma}(x) = \begin{cases} e^{-x^{-\gamma}},\ x > 0 \\ 0,\ x \le 0 \end{cases}.$$

Normalizavimo konstantas galima parinkti tokiu būdu:

$$a_n = 0;\quad b_n:\ 1 - F(b_n) = \frac{1}{n}.$$

Pareto skirstinio atveju gauname:

$$\lim_{n \to \infty} P\left(\frac{Z_n^k - a_n}{b_n} < x\right) = H_{1,\frac{\alpha}{k}}(x) = \begin{cases} e^{-x^{-\frac{\alpha}{k}}},\ x > 0 \\ 0,\ x \le 0 \end{cases},$$

normalizavimo konstantos $a_n = 0;\quad b_n = n^{\frac{k}{\alpha}}$.

Kai n yra didelis, maksimumo skirstinį galime aproksimuoti ribiniu skirstiniu, t.y.

$$P\left(\frac{Z_n^k}{n^{k/\alpha}} < x\right) \to e^{-x^{-\alpha/k}},\ x \ge 0,\ n \to \infty.$$

Kai $k = 1$, yra žinomas pakankamai geras konvergavimo greičio įvertis [2]

$$\left|P\left(Z_n < x \cdot n^{1/\alpha}\right) - e^{-x^{-\alpha}}\right| \le \Delta_n(x),$$

Čia $\Delta_n(x) \le \frac{z_n^2(x) e^{-z_n(x)}}{2} \cdot \frac{1}{n-1},\ x \ge 0,\ z_n(x) = n\left(1 - F\left(x n^{1/\alpha}\right)\right).$

Naudojantis šiuo įverčiu galime rasti vidurkio konvergavimo greičio įvertį

$$\left|\frac{MZ_n}{n^{1/\alpha}} - \Gamma\left(1 - \frac{1}{\alpha}\right)\right| \le \int_0^\infty \Delta_n(x) dx.$$

Įrodymas:

$$\left|P\left(Z_n < x n^{1/\alpha}\right) - e^{-x^{-\alpha}}\right| \le \Delta_n(x);$$

$$\left|1 - P\left(Z_n > x n^{1/\alpha}\right) - e^{-x^{-\alpha}}\right| \le \Delta_n(x);$$

$$\left|- P\left(Z_n > x n^{1/\alpha}\right) + 1 - e^{-x^{-\alpha}}\right| \le \Delta_n(x);$$

$$\left|-\left(1 - P\left(Z_n < x n^{1/\alpha}\right)\right) + 1 - e^{-x^{-\alpha}}\right| \le \Delta_n(x);$$

$$\left|\left(1 - e^{-x^{-\alpha}}\right) - \left(1 - P\left(Z_n < x n^{1/\alpha}\right)\right)\right| \le \Delta_n(x);$$

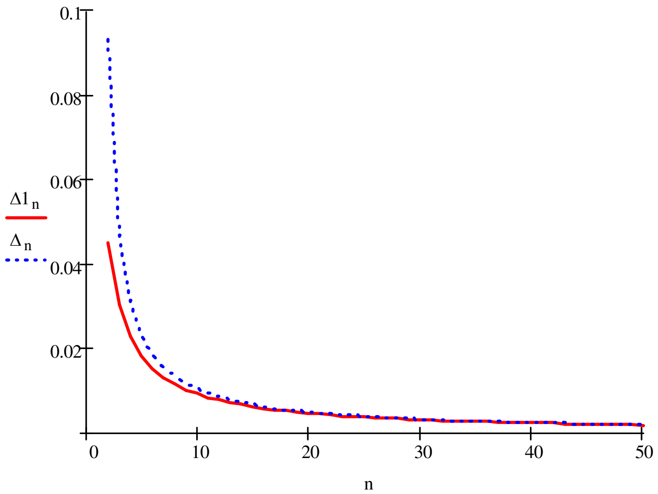
$$\left|\left(1 - P\left(Z_n < x n^{1/\alpha}\right)\right) - \left(1 - e^{-x^{-\alpha}}\right)\right| \le \Delta_n(x);$$

$$\int_0^\infty \left|\left(1 - P\left(Z_n < x n^{1/\alpha}\right)\right) - \left(1 - e^{-x^{-\alpha}}\right)\right| dx \le \int_0^\infty \Delta_n(x)dx;$$

$$\int_0^\infty \left(1 - P\left(Z_n < x n^{1/\alpha}\right)\right) dx - \int_0^\infty \left(1 - e^{-x^{-\alpha}}\right) dx \le \int_0^\infty \Delta_n(x)dx;$$

$$\left|\frac{MZ_n}{n^{1/\alpha}} - \Gamma\left(1 - \frac{1}{\alpha}\right)\right| \le \int_0^\infty \Delta_n(x)dx.$$

Žinodami tikslias MZ_n reikšmes, ribinio skirstinio vidurkį ir apskaičiavę paklaidos integralą, rezultatus pavaizduojame grafiškai (kairiąją nelygybės pusę žymime $\Delta 1_n$, o dešiniąją Δ_n).



1 pav. Normuoto ir ribinio vidurkio skirtumo palyginimas su gautu įverčiu

Literatūra:

1. Galambos J., The Asymptotic Theory of Extreme Order Statistics, Wiley, New York, 1978
2. Aksomaitis A. "Atsitiktinių dydžių maksimumų vidurkiai". Lietuvos matematikų draugijos XXXV konferencija: pranešimų tezės. Vilnius, 1994, p. 236-237.
3. Aksomaitis A. Tikimybių teorija ir statistika. Kaunas: Technologija, 2000, 344 psl.

5 PRIEDAS PARETO DYDŽIŲ MAKSIMUMŲ VIDURKIŲ IR JŲ ĮVERČIŲ PALYGINIMAS PRIE SKIRTINGŲ *α* REIKŠMIŲ

	$\alpha = 3$			$\alpha = 5$			$\alpha = 7$		
n	MZ_n	mz_n	Santykinė paklaida, %	MZ_n	mz_n	Santyki-nė paklaida, %	MZ_n	mz_n	Santyki-nė paklaida, %
2	1,8	2,20711	22,62	1,38889	1,62268	16,83	1,25641	1,44555	15,05
3	2,025	2,59545	28,17	1,4881	1,82735	22,80	1,31923	1,59872	21,19
4	2,20909	2,88873	30,77	1,56642	1,98193	26,53	1,36809	1,71439	25,31
5	2,36688	3,13299	32,37	1,63168	2,11066	29,35	1,40833	1,81073	28,57
6	2,50611	3,34637	33,53	1,68795	2,22312	31,71	1,44268	1,89489	31,34
7	2,63142	3,5381	34,46	1,73759	2,32417	33,76	1,47273	1,9705	33,80
8	2,74583	3,71359	35,25	1,78215	2,41667	35,60	1,49951	2,03972	36,03
9	2,85144	3,87635	35,94	1,82265	2,50245	37,30	1,5237	2,10391	38,08
10	2,94976	4,02878	36,58	1,85985	2,58278	38,87	1,54578	2,16403	40,00
11	3,04194	4,17261	37,17	1,89429	2,65859	40,35	1,56612	2,22076	41,80
12	3,12885	4,30915	37,72	1,9264	2,73055	41,74	1,58499	2,27461	43,51
13	3,21119	4,43939	38,25	1,9565	2,79919	43,07	1,6026	2,32598	45,14
14	3,28951	4,56413	38,75	1,98485	2,86494	44,34	1,61912	2,37518	46,70
15	3,36427	4,68401	39,23	2,01167	2,92812	45,56	1,63469	2,42246	48,19
16	3,43586	4,79956	39,69	2,03714	2,98902	46,73	1,64942	2,46803	49,63
17	3,50457	4,91121	40,14	2,06139	3,04787	47,85	1,66339	2,51207	51,02
18	3,5707	5,01933	40,57	2,08455	3,10485	48,95	1,6767	2,55471	52,37
19	3,63446	5,12424	40,99	2,10673	3,16014	50,00	1,6894	2,59609	53,67
20	3,69606	5,22621	41,40	2,12801	3,21388	51,03	1,70156	2,6363	54,93

6 PRIEDAS BURO ATSITIKTINIŲ DYDŽIŲ NORMUOTO MAKSIMUMO VIDURKIO PALYGINIMAS SU RIBINIO SKIRSTINIO VIDURKIU KINTANT PARAMETRUI α

	$\alpha = 3$			$\alpha = 7$			$\alpha = 12$		
n	$MZ1_n$	MY_n	Santykinė paklaida, %	$MZ1_n$	MY_n	Santykinė paklaida, %	$MZ1_n$	MY_n	Santykinė paklaida, %
2	1,6123	1,3541	16,01	1,1821	1,1058	6,46	1,0958	1,0555	3,67
3	1,4929	1,3541	9,30	1,1472	1,1058	3,61	1,0774	1,0555	2,03
4	1,4491	1,3541	6,56	1,1342	1,1058	2,50	1,0705	1,0555	1,40
5	1,4263	1,3541	5,06	1,1274	1,1058	1,92	1,0670	1,0555	1,07
6	1,4123	1,3541	4,12	1,1232	1,1058	1,55	1,0648	1,0555	0,86
7	1,4029	1,3541	3,48	1,1204	1,1058	1,30	1,0633	1,0555	0,73
8	1,3961	1,3541	3,01	1,1183	1,1058	1,12	1,0622	1,0555	0,63
9	1,3910	1,3541	2,65	1,1168	1,1058	0,99	1,0614	1,0555	0,55
10	1,3869	1,3541	2,37	1,1156	1,1058	0,88	1,0607	1,0555	0,49
11	1,3837	1,3541	2,14	1,1146	1,1058	0,80	1,0602	1,0555	0,44
12	1,3811	1,3541	1,95	1,1138	1,1058	0,73	1,0598	1,0555	0,40
13	1,3788	1,3541	1,79	1,1132	1,1058	0,67	1,0595	1,0555	0,37
14	1,3770	1,3541	1,66	1,1126	1,1058	0,62	1,0592	1,0555	0,34
15	1,3754	1,3541	1,54	1,1121	1,1058	0,57	1,0589	1,0555	0,32
16	1,3740	1,3541	1,44	1,1117	1,1058	0,54	1,0587	1,0555	0,30
17	1,3727	1,3541	1,36	1,1113	1,1058	0,50	1,0585	1,0555	0,28
18	1,3716	1,3541	1,28	1,1110	1,1058	0,47	1,0583	1,0555	0,26
19	1,3707	1,3541	1,21	1,1107	1,1058	0,45	1,0582	1,0555	0,25
20	1,3698	1,3541	1,15	1,1105	1,1058	0,42	1,0580	1,0555	0,23

7 PRIEDAS

n	MZ_n	mz_n	$mz1_n$	absoliutinė mz_n paklaida	santykinė mz_n paklaida, %	absoliutinė $mz1_n$ paklaida	santykinė $mz1_n$ paklaida, %
2	1,38889	1,62268	1,39084	0,23379	16,83282	0,00195	0,14067
3	1,48810	1,82735	1,48899	0,33926	22,79794	0,00090	0,06027
4	1,56642	1,98193	1,56693	0,41551	26,52610	0,00052	0,03310
5	1,63168	2,11066	1,63202	0,47898	29,35493	0,00034	0,02084
6	1,68795	2,22312	1,68819	0,53518	31,70567	0,00024	0,01430
7	1,73759	2,32417	1,73777	0,58658	33,75809	0,00018	0,01042
8	1,78215	2,41667	1,78229	0,63452	35,60419	0,00014	0,00792
9	1,82265	2,50245	1,82276	0,67980	37,29720	0,00011	0,00622
10	1,85985	2,58278	1,85994	0,72294	38,87077	0,00009	0,00502
11	1,89429	2,65859	1,89437	0,76430	40,34763	0,00008	0,00413
12	1,92640	2,73055	1,92646	0,80415	41,74397	0,00007	0,00346
13	1,95650	2,79919	1,95655	0,84270	43,07176	0,00006	0,00294
14	1,98485	2,86494	1,98490	0,88009	44,34020	0,00005	0,00253
15	2,01167	2,92812	2,01172	0,91645	45,55651	0,00004	0,00220
16	2,03714	2,98902	2,03718	0,95188	46,72651	0,00004	0,00193
17	2,06139	3,04787	2,06142	0,98648	47,85497	0,00004	0,00171
18	2,08455	3,10485	2,08458	1,02030	48,94585	0,00003	0,00152
19	2,10673	3,16014	2,10676	1,05342	50,00252	0,00003	0,00136
20	2,12801	3,21388	2,12803	1,08588	51,02785	0,00003	0,00123
21	2,14847	3,26619	2,14849	1,11773	52,02431	0,00002	0,00111
22	2,16818	3,31719	2,16820	1,14901	52,99405	0,00002	0,00101
23	2,18720	3,36695	2,18722	1,17975	53,93898	0,00002	0,00093
24	2,20558	3,41558	2,20560	1,21000	54,86076	0,00002	0,00085
25	2,22337	3,46313	2,22338	1,23977	55,76089	0,00002	0,00078
26	2,24060	3,50969	2,24062	1,26909	56,64069	0,00002	0,00072
27	2,25732	3,55531	2,25734	1,29799	57,50138	0,00002	0,00067
28	2,27356	3,60005	2,27358	1,32649	58,34403	0,00001	0,00062
29	2,28935	3,64395	2,28936	1,35460	59,16962	0,00001	0,00058
30	2,30471	3,68706	2,30473	1,38235	59,97906	0,00001	0,00054
31	2,31968	3,72942	2,31969	1,40974	60,77317	0,00001	0,00051
32	2,33427	3,77108	2,33428	1,43681	61,55269	0,00001	0,00048
33	2,34850	3,81205	2,34851	1,46355	62,31833	0,00001	0,00045
34	2,36240	3,85238	2,36241	1,48998	63,07071	0,00001	0,00042
35	2,37598	3,89210	2,37599	1,51612	63,81043	0,00001	0,00040
36	2,38925	3,93122	2,38926	1,54198	64,53804	0,00001	0,00038
37	2,40223	3,96979	2,40224	1,56756	65,25405	0,00001	0,00036
38	2,41495	4,00782	2,41495	1,59287	65,95894	0,00001	0,00034
39	2,42739	4,04533	2,42740	1,61793	66,65313	0,00001	0,00032
40	2,43959	4,08234	2,43960	1,64275	67,33706	0,00001	0,00031
41	2,45155	4,11888	2,45156	1,66733	68,01113	0,00001	0,00030
42	2,46328	4,15495	2,46329	1,69167	68,67566	0,00001	0,00028
43	2,47479	4,19059	2,47480	1,71580	69,33123	0,00001	0,00039
44	2,48609	4,22580	2,48610	1,73971	69,97789	0,00001	0,00046
45	2,49718	4,26059	2,49720	1,76341	70,61619	0,00002	0,00064
46	2,50807	4,29499	2,50810	1,78692	71,24703	0,00003	0,00128
47	2,51880	4,32901	2,51882	1,81020	71,86768	0,00002	0,00063
48	2,52931	4,36265	2,52936	1,83334	72,48402	0,00005	0,00197
49	2,53967	4,39594	2,53972	1,85627	73,09111	0,00006	0,00227
50	2,54986	4,42887	2,54992	1,87901	73,69068	0,00006	0,00242