Univerzitet u Beogradu, Matematički fakultet

MASTER RAD

BETA-TIP RASPODELE I NJENA PRIMENA

Student:
Jelena Ljuboja 1057/2015

Mentor:
Prof. Dr. Slobodanka Janković

Beograd, 2018.

SADRŽAJ

Uvod

U ovom radu detaljno smo definisali i razradili beta-tip raspodela, kao i njenu primenu u različitim oblastima. Ono što je važno navesti u samom uvodu, je to da najveću primenu ovaj tip raspodela ima u ekonomiji i aktuarstvu. Samim tim, većina primera je iz tih dvaju polja.

Pre svega treba istaći da postoje dva tipa beta raspodela, pri čemu oba pripadaju poznatom Pirsonovom[1] sistemu. Oba tipa raspodela su suštinski povezana sa regularnom nekompletnom beta funkcijom (uopštenjem beta funkcije):

$$I_x(p,q) = \frac{1}{B(p,q)} \cdot \int_0^x u^{p-1} \cdot (1-u)^{q-1} du\,, 0 \leq x \leq 1, \qquad (1)$$

a time i sa samom nekompletnom beta funkcijom:

$$B_x(p,q) = \int_0^x u^{p-1} \cdot (1-u)^{q-1} du\,, 0 \leq x \leq 1, \qquad (2)$$

O ovim rezultatima je pisao Dutka[2], koji ih je pronašao u pismu Isaka Njutna Henriju Oldenbergu iz 1676. godine.

Sam rad se sastoji iz četiri poglavlja kroz koje smo obradili ovaj tip raspodele i njene specijalne slučajeve. Nazivi i kratko pojašnjenje samih poglavlja su sledeći:

1. Uopštena beta raspodela druge vrste: U ovom poglavlju smo definisali beta-tip raspodela druge vrste, upoznali se sa momentima raspodele, ključnim svojstvima, prezentacijom raspodele i slično;
2. Sing-Madala raspodela: Kao što samo ime poglavlja kaže, u njemu smo se bazirali na ovaj podtip beta raspodele druge vrste, tj. njen specijalni slučaj;
3. Fisk i Lomaks raspodela: U trećem poglavlju smo obradili ova dva tipa raspodele, njihove definicije, momente i slično. Uz to, naveli smo empirijske rezultate obe raspodele u prihodima i bogatstvu;
4. Uopštena beta raspodela prve vrste: Poslednje poglavlje smo posvetili beta raspodeli prve vrste, kao i njenim specijalnim slučajem – nama dobro poznatom standardnom beta raspodelom. Nakon osnovnih definicija, osvrnuli smo se na empirijske rezultate.

[1] Karl Pearson, engleski matematičar i statističar
[2] Jacques Dutka, američki matematičar

1 Uopštena beta raspodela druge vrste

Ključna raspodela u ovoj porodici je tzv. uopštena beta raspodela druge vrste (u daljem tekstu: GB2). Ono što je važno spomenuti je doprinos Mekdonalda[3], 1984. godine [3] i njegovih saradnika u razvoju pomenute GB2 raspodele, kao raspodele prihoda i ujedinjenju različitih istraživačkih aktivnosti u tesno povezanim naučnim poljima. Kao primer možemo navesti rad objavljen 1990. godine [8], koji je sačinilo više autora, a koji predstavlja kombinaciju dva rukopisa (jednog od Mekdonalda), a bavi se primenom GB2 familije raspodela u polju gubitaka u osiguranju. Rezultat ovog spisa je jedinstveni tretman primene GB2 familije raspodela.

Za izradu poglavlja korišćeni su sledeći izvori: [1], [2], [3], [4], [5], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20].

1.1 Definicija

Funkciju raspodele GB2 raspodele uvodimo koristeći alternativni izraz za regularnu nekompletnu beta funkciju, koja se dobija kada u formulu (1) uvedemo smenu:

$$u := \frac{t}{(1+t)} .$$

Prethodno pomenutom smenom dobijamo sledeću formulu:

$$I_z(p,q) = \frac{1}{B(p,q)} \cdot \int_0^z \frac{t^{p-1}}{(1+t)^{p+q}} dt \,, z > 0, \qquad (3)$$

Uvodeći dodatne parametre b i a za meru i oblik i koristeći smenu:

$$z := (\frac{x}{b})^a \,,$$

dobijamo sledeću funkciju raspodele:

$$F(x) = I_z(p,q), \text{ gde je } z = (\frac{x}{b})^a, x > 0 \,, \qquad (4)$$

čija je odgovarajuća gustina raspodele:

$$f(x) = \frac{a \cdot x^{a \cdot p - 1}}{b^{a \cdot p} \cdot B(p,q) \cdot [1 + (\frac{x}{b})^a]^{p+q}}, x > 0, \qquad (5)$$

Važno je navesti da su sva četiri parametra a, b, p i q pozitivna, od čega su a, p i q parametri oblika, dok je b parametar skaliranja. Ono što možemo na početku da primetimo iz same formule jeste da je funkcija gustine GB2 raspodele pravilno promenljiva u beskonačnosti sa indeksom $-aq - 1$ i u nuli sa indeksom $ap - 1$; tako da zapravo sva tri parametra oblika kontrolišu

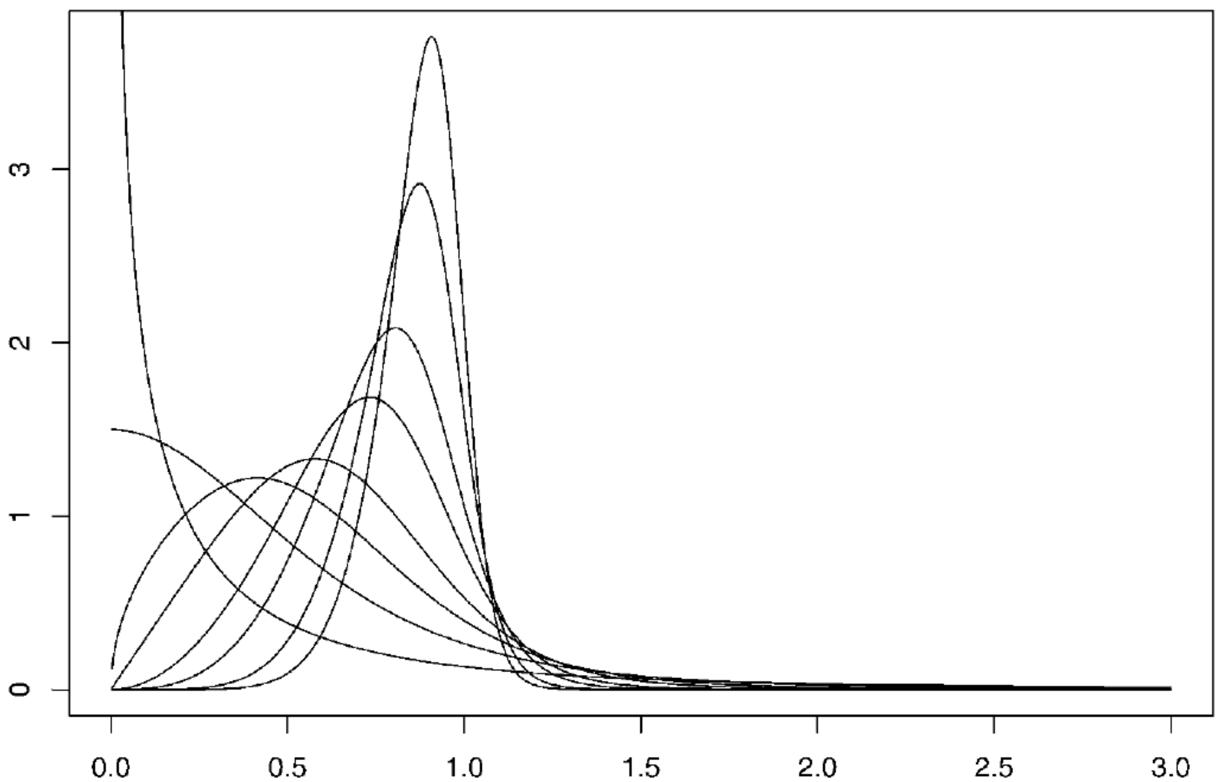
[3] James McDonald, američki matematičar, profesor ekonomije sa specijalizacijom u ekonometriji

ponašanje repa modela. Uprkos tome, ova tri parametra nisu jednaka po važnosti, a primer tome je raspodela $Y = logX$, čija gustina je:

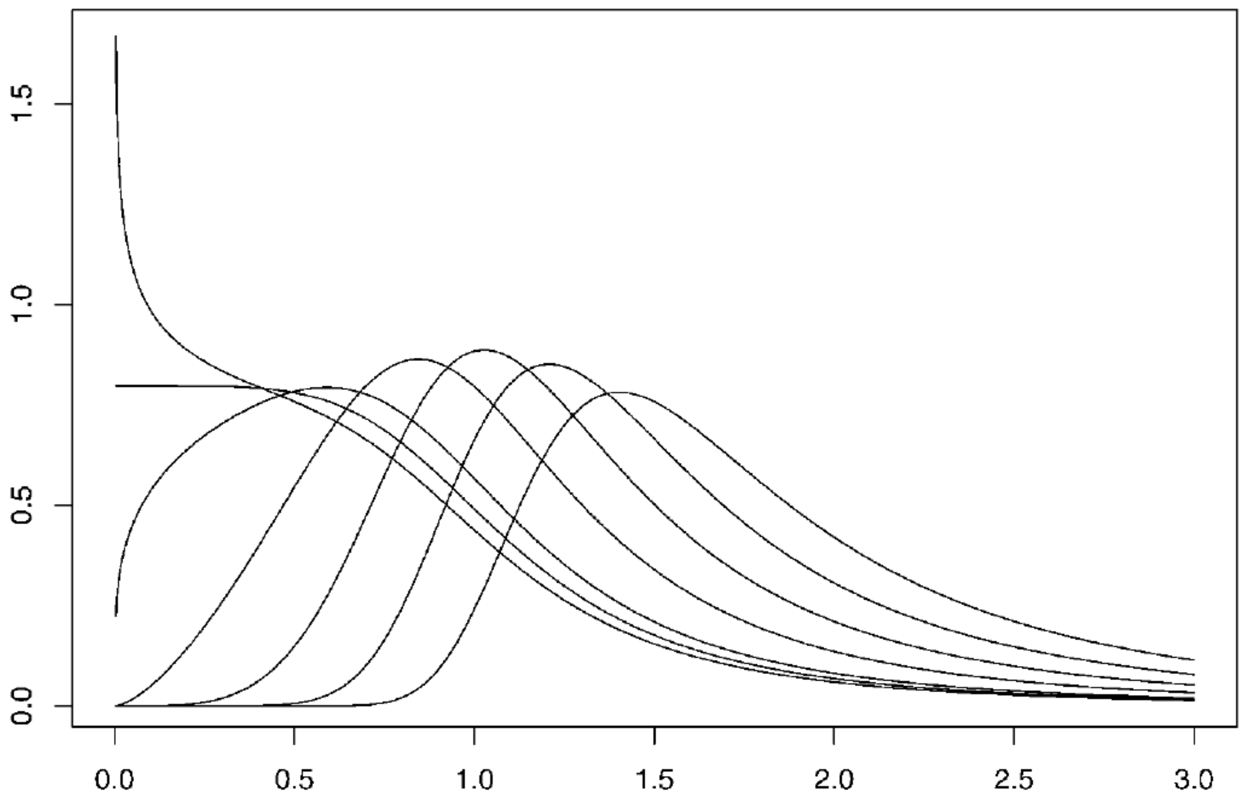
$$f(y) = \frac{a \cdot e^{a \cdot p \cdot (y - logb)}}{B(p,q) \cdot [1 + e^{a \cdot (y - logb)}]^{p+q}}, -\infty < y < \infty. \qquad (6)$$

U ovom slučaju, a postaje parametar skaliranja, dok su p i q i dalje parametri oblika.

Dodatno, iz formule (5) možemo primetiti da što je veća vrednost broja a, to su tanji repovi gustine, dok su vrednosti parametara p i q važne prilikom određivanja distorzije simetričnosti zvona na grafiku za raspodelu $logX$. Na slikama 1 i 2 u nastavku teksta, ilustrovali smo efekte promene sva tri parametra, s tim da su na oba grafika dva parametra konstantna, a preostali je promenljiv. Ono što se vidi iz samog grafika jeste da imamo značajnu varijaciju gustine za male vrednosti parametara a i p.



Slika 1: Grafik funkcije gustine raspodele GB2 za sledeće parametre: $p = 0.5$, $q = 2$ i $a = 1, 2, 3, 4, 6, 8, 12, 16$ (posmatrajući sa leva na desno).



Slika 2: Grafik funkcije gustine raspodele GB2 za sledeće parametre: $a = 5$, $q = 0.5$ i $p = 0.17, 0.25, 0.5, 1, 2, 4$ (posmatrajući sa leva na desno).

Kao raspodelu prihoda (5) je prvi put uveo Mekdonald 1984. godine [3], a nezavisno od toga je godinu dana ranije predstavljena kao model za raspodelu mere gubitka u aktuarstvu, od strane Ventera[4] [9], koji ju je tada nazvao transformisana beta raspodela. Deceniju ranije je razmatrana kao uopštenje dve raspodele, Dagum raspodele i Sing-Madala raspodele, o kojoj će biti reči u sledećem poglavlju [10]. Ova raspodela se takođe smatrala uopštenjem F raspodele i kao takva se pojavila u radovima iz 1980. godine [11], dok je 1983. godine nazivana Feler-Pareto raspodela [12].

Funkcija gustine Fišerove $F(m, n)$ raspodele je sledeća:

$$g(x) = \frac{\left(\frac{m}{n}\right)^{m/2}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \cdot \frac{x^{m/2-1}}{\left(1 + \frac{m}{n} \cdot x\right)^{m/2+n/2}}, x > 0.$$

Sama raspodela je dobila ime po engleskom statističaru i genetičaru Fišeru (Ronald Aylmer Fisher, 1890-1962), dok često koristimo naziv F raspodela.

[4] Gary Venter, američki aktuar i matematičar

Tokom osnovnih studija upoznali smo se sa primenom pomenute raspodele i to kod analize varijanse, testiranja hipoteza o jednakosti disperzija i slično.

Funkcija gustine Feler-Paretove raspodele je:

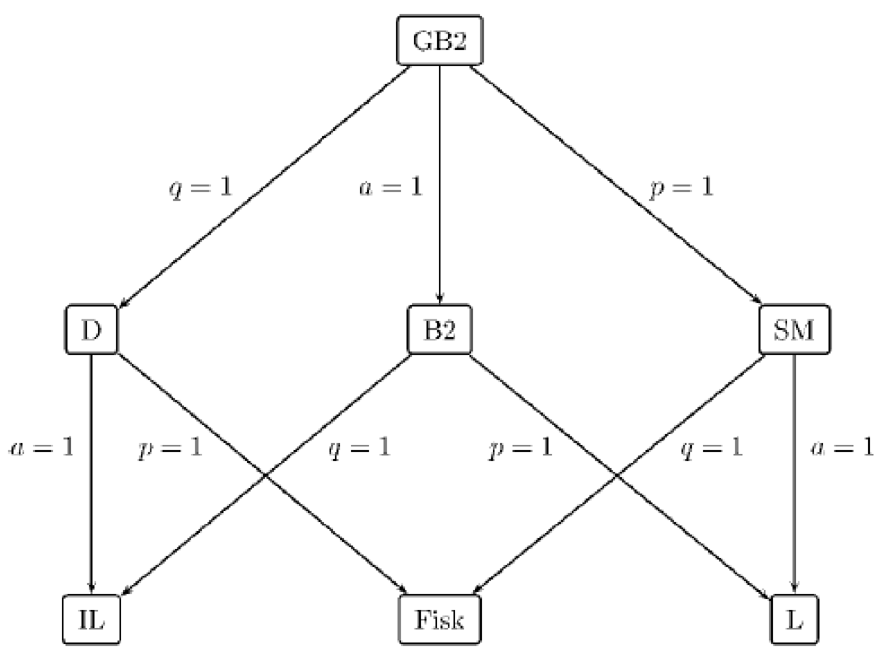
$$g(y) = \frac{y^{a-1} \cdot (1-y)^{b-1}}{B(a,b)}, 0 < y < 1; a, b > 0.$$

Dakle, možemo zaključiti da su razni naučnici dolazili do sličnih rezultata, koji su uopštenja beta-tip raspodele.

Imajući u vidu da GB2 raspodela objedinjuje veliki broj raspodela prihoda i gubitaka, ispod navodimo njene specijalne slučajeve:

- Sing-Madala raspodela, koja se dobija za $p = 1$,
- Dagum raspodela, koja se dobija za $q = 1$,
- Beta raspodela druge vrste (B2) za $a = 1$,
- Fisk (loglogistička) raspodela, koja se dobija za $p = q = 1$,
- Lomaks (ili Pareto II) raspodela, za $a = p = 1$.

U sledećim poglavljima smo se detaljnije bavili nekim od pomenutih specijalnih slučajeva, dok odnose među istim vidimo na slici 3:



Slika 3: Beta-tip raspodela i njeni međusobni odnosi: Uopštena beta-tip raspodela druge vrste (GB2), Dagum raspodela (D), beta raspodela druge vrste (B2), Sing-Madala raspodela (SM), inverzna Lomaks raspodela (IL), Fisk (loglogistička) raspodela (Fisk) i Lomaks raspodela (L).

Specijalni slučaj 1: Suprotno uopštenju koje smo do sada videli, ispod predstavljamo neke od specijalnih slučajeva ovog tipa raspodela. Kao što smo već naveli za $a = 1$, dobijamo beta raspodelu druge vrste (u daljem tekstu: B2) koja sa svojom funkcijom gustine pripada Pirsonovom sistemu raspodela, tzv. Pirsonova VI raspodela. Ona je kao raspodela prihoda predstavljena pre GB2 raspodele. Funkcija gustine B2 raspodele se dobija kada u formuli (5) zamenimo $a = 1$:

$$f(x) = \frac{x^{p-1}}{b^p \cdot B(p,q) \cdot [1 + x/b]^{p+q}}, x > 0. \tag{7}$$

Specijalni slučaj 2: Sledeći primer je raspodela koja se dobija izjednačenjem svih parametara sa jedinicom, tj. ukoliko u formuli (6) zamenimo $a = b = p = q = 1$. Tada dobijamo sledeće:

$$f(y) = \frac{e^y}{(1 + e^y)^2}, -\infty < y < \infty.$$

Dakle, dobili smo funkciju gustine standardne logističke raspodele. Konkretnije, raspodela čija je gustina prikazana formulom (6) jeste iskrivljena uopštena logistička raspodela, gde se simetrija postiže samo za $p = q$. Formula (6) se u literaturi može još naći pod imenom gustina eksponencijalne GB2 raspodele [4].

Za kraj prvog dela prvog poglavlja prikazali smo još da od GB2 možemo dobiti i uopštenu gama raspodelu ukoliko postavimo sledeće: $b = q^{1/a}\beta$, kada $q \to \infty$. Shodno tome, Gama, kao i Vejbulova raspodela su granični slučajevi GB2 raspodele, pošto su oba specijalni slučajevi uopštene gama raspodele.

Funkcija gustine Gama raspodele je:

$$f(x) = \frac{1}{\beta^p \cdot \Gamma(p)} \cdot x^{p-1} \cdot e^{-\frac{x}{\beta}}, \;\; x > 0.$$

Ovde je parametar skaliranja $\beta > 0$, dok je parametar oblika $p > 0$. Momenti raspodele postoje samo za k koje se nalazi u sledećem intervalu: $-p < k < \infty$. Momenti su prikazani sledećom jednakošću:

$$E\left(X^k\right) = \frac{\beta^k \cdot \Gamma(p + k)}{\Gamma(p)}.$$

Funkcija gustine Vejbulove raaspodele je:

$$f(x) = \frac{a}{\beta} \cdot \left(\frac{x}{\beta}\right)^{a-1} \cdot e^{-\left(\frac{x}{\beta}\right)^a}, \;\; x > 0,$$

gde su $a, \beta > 0$. Ovde je β parametar skaliranja, dok je a parametar oblika. Momenti postoje samo za k koje se nalazi u sledećem intervalu: $-a < k < \infty$ i dati su sledećom formulom:

$$E\left(X^k\right) = \Gamma(1 + \frac{k}{a}) \cdot \beta^k.$$

1.2 Momenti raspodele i druga osnovna svojstva

U formuli (4) GB2 smo uveli preko regularne nekompletne beta funkcije. Koristeći relaciju sa Gausovom hipergeometrijskom funkcijom 2F1, dobili smo sledeći izraz:

$$I_z(p,q) = \frac{z^p}{p \cdot B(p,q)} \cdot {}_2F_1(p, 1-q; p+1; z),\ \ 0 \le z \le 1, \tag{8}$$

gde je:

$$ {}_2F_1(a_1, a_2; b; x) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdot (a_2)_n}{(b)_n} \cdot \frac{x^n}{n!}, \quad |x| < 1. \tag{9}$$

Funkciju raspodele F možemo predstaviti u sledećem obliku:

$$F(x) = \frac{[\left(\frac{x}{b}\right)^a / (1+\left(\frac{x}{b}\right)^a)]^p}{p \cdot B(p,q)} \cdot {}_2F_1\left(p, 1-q; p+1; \left[\frac{\left(\frac{x}{b}\right)^a}{1+\left(\frac{x}{b}\right)^a}\right]\right),\ \ x > 0. \tag{10}$$

Moda (maksimalna vrednost verovatnoće) GB2 raspodele je u 0, ili se postiže u sledećoj tački:

$$x_{mode} = b \cdot \left(\frac{a \cdot p - 1}{a \cdot q + 1}\right)^{1/a} \text{, ako je } \ a \cdot p > 1. \tag{11}$$

Kao što smo primetili u prethodnom odeljku, funkcija gustine GB2 raspodele je pravilno promenljiva u beskonačnosti sa indeksom $-aq - 1$ i u nuli sa indeksom $ap - 1$. Ovo implicira da momenti postoje samo za k, koje zadovoljava sledeću nejednakost: $-ap < k < aq$. Momenti raspodele su sledeći:

$$E(X^k) = \frac{b^k \cdot B(p+\frac{k}{a}, q-\frac{k}{a})}{B(p,q)} = \frac{b^k \cdot \Gamma\left(p+\frac{k}{a}\right) \cdot \Gamma(q-\frac{k}{a})}{\Gamma(p) \cdot \Gamma(q)}. \tag{12}$$

Ukoliko poslednji izraz posmatramo kao funkciju po k, primetićemo da je jednak generišućoj funkciji momenata $m_Y(\cdot)$ raspodele $Y = logX$, imajući u vidu relaciju $m_Y(K) = E(e^{kY}) = E(X^k)$. Ovo nam je korisno ukoliko želimo da računamo momente $logX$ raspodele, koji su nam potrebni npr. za izvođenje Fišerove informacione matrice GB2 raspodele i njenih podfamilija.

Dalje, lako možemo videti da je GB2 raspodela zatvorenog tipa u sledećem smislu:

$$X \sim GB2(a,b,p,q) \Rightarrow X^r \sim GB2(\frac{a}{r}, b^r, p, q)\ ,\ \ r > o. \tag{13}$$

Dodatno, GB2 raspodela je zatvorenog tipa u odnosu na inverziju, tj. u sledećem smislu:

$$X \sim GB2(a,b,p,q) \Leftrightarrow \frac{1}{X} \sim GB2(a, \frac{1}{b}, q, p)\ ,\ \ r > o. \tag{14}$$

U formuli (5) je ponekad zgodno dopustiti da je $a < 0$ i zameniti a sa $|a|$ u brojiocu.

Stopa rizika (hazarda) GB2 raspodele može uzeti širok spektar. Stopa rizika za GB2 je sledeća:

$$r(x) = \frac{a \cdot p \cdot (q \cdot b)^q \cdot x^{a \cdot p - 1}}{(q \cdot b^a + x^a)^q \left[p \cdot B(p,q) \cdot (q \cdot b^a + x^a)^p - x^{a \cdot p} \cdot {}_2F_1[p, q-1; p+1; x^a/(q \cdot b^a) + x^a]\right]}.$$

1.3 Karakterizacija i reprezentacija

B2 raspodelu je moguće karakterizovati u smislu maksimalne entropije (mere neodređenosti pridružene slučajnoj promenljivoj) među svim raspodelama na intervalu $[0, \infty)$: ukoliko postoje oba $E[logX]$ i $E[\log(1+X)]$, onda je maksimalna entropija funkcije gustine upravo gustina B2 raspodele. Dakle, sama raspodela je označena kao geometrijska srednja vrednost od X i $X+1$.

Proširenje pristupa maksimalne entropije je razmatrano 1990. godine [13], i ono je je koristilo relativni maksimalni entropijski princip koji je doveo do nekoliko raspodela prihoda, kao što su GB2 i GB1 (koja je opisana u poglavlju 4) i Sing-Madala raspodele (opisane u sledećem poglavlju). Ovde je ideja da su primaoci prihoda pod uticajem rednih, kao i vrednosnih pozicija i zato subjektivno smanjuju svoje prihode tako što ih množe sa subjektivnim, ali ne sa individualnim, redukcionim faktorom $\phi[x, 1-F(x)]$, koji zavisi od stvarnog prihoda x, kao i sa statusom prihoda, koji se ovde meri proporcijom $1-F(x)$ od onih prihoda primaoca zarade koji zarađuju više od unapred određenog x. Ovo vodi do prilagođenog prihoda: $\xi(x) = x \cdot \phi[x, 1-F(x)]$. Gustina relativne entropije prihoda sa sada definiše pomoću funkcije gustine raspodele f i marginalnog subjektivnog prihoda $\frac{\partial \xi[x, F(x)]}{\partial x}$ i određuje se maksimizacijom integrala $\int f(x)dx$, čije su granice prva dva momenta od ξ, tj. $E\xi(X)$ i $E\xi^2(X)$, u odnosu na F. Nastala nelinearna diferencijalna jednačina je ponekad rešiva i pod odgovarajućim specifikacijama funkcije ϕ, dovodi do raspodele prihoda, koja je uopštenog beta tipa.

U nastavku teksta je prikazana veza dve uopštene raspodele gama tipa (u dalje tekstu: GG). Ukoliko imamo dve nezavisne raspodele koje pripadaju uopštenom gama tipu, tj. ako $X_1 \sim GG(a,1,p)$ i $X_2 \sim GG(a,1,q)$, pri čemu obe raspodele imaju isti parametar oblika a, onda važi sledeće:

$$\frac{X_1}{X_2} \sim GB2(a,1,p,q). \tag{15}$$

Prethodno pomenuta relacija (15) je možda poznatija u sledećem obliku:

$$\left(\frac{X_1}{X_2}\right)^{\frac{1}{a}} \sim GB2(a,1,p,q), \tag{16}$$

gde su raspodele X_1 i X_2 takođe nezavisne i pripadaju gama tipu raspodele, tj. $X_1 \sim Ga(1,p)$ i $X_2 \sim Ga(1,q)$, što je uopštenje dobro poznate relacije između standardne gama i B2 raspodele, tj:

$$\frac{X_1}{X_2} \sim B2(1,p,q). \tag{17}$$

Ove veze možemo koristiti u svrhu dobijanja slučajnih uzoraka iz GB2 raspodele: Postoji veliki broj generatora gama slučajnih brojeva, od kojih se putem stepene transformacije (eng. power transformation) mogu dobiti uopšteni gama uzorci. Drugim rečima, preko gore pomenute relacije, simulirajući nezavisne podatke iz dve uopštene gama raspodele sa potrebnim parametrima oblika, dolazimo do GB2 uzoraka.

Važi i sledeće: ako su slučajne veličine X_1 i X_2 nezavisne i imaju gama raspodelu, tj. $X_1 \sim Ga(1,p)$ i $X_2 \sim Ga(1,q)$, onda slučajna promenljiva:

$$W \coloneqq \frac{X_1}{X_1+X_2} \tag{18}$$

ima beta raspodelu I tipa sa parametrima p i q.

Posmatrajući raspodelu (18), definisaćemo B2 raspodelu preko sledećeg:

$$X \sim W^{-1} - 1,$$

Koja se naziva Paretova raspodela. Sada smo zaključili da GB2 raspodelu možemo prikazati na sledeći način, a koja se često Feler-Paretova raspodela [12]:

$$X \sim b \cdot [W^{-1} - 1]^{1/a}. \tag{19}$$

(15) se takođe može razumeti kao mešavina dveju raspodela, naime kao mešavina uopštenih gama raspodela sa inverznim uopštenim gama težinama. Funkcija gustine uopštene gama raspodele je:

$$f(x) = \frac{a}{\theta^{a\cdot p}\cdot\Gamma(p)} \cdot x^{a\cdot p-1} \cdot e^{-(\frac{x}{\theta})^a}.$$

Ukoliko parametar θ pripada inverznoj uopštenoj gama raspodeli (u daljem tekstu: InvGG), tj. $\theta \sim InvGG(a,b,q)$, onda važi sledeće:

$$GG(a,\theta,p) \bigwedge_\theta InvGG(a,b,q) = GB2(a,b,p,q). \tag{20}$$

Jasno je da je ovo samo ponovni iskaz stavke (15). Za beta raspodelu druge vrste, kada je $a = 1$, prethodni iskaz se pojednostavljuje na sledeće:

$$Ga(\theta,p) \bigwedge_\theta InvGa(b,q) = B2(b,p,q), \tag{21}$$

gde InvGa označava inverznu gama raspodelu.

GB2 raspodela i njene podfamilije imaju i teorijsku opravdanost za prezentaciju prihoda koji proizilaze iz heterogene populacije primaoca prihoda, ili prema aktuarskoj terminologiji, kao prezentaciju šteta nastalih is heterogene populacije izloženosti. Ovaj argument je iskorišćen 1997. godine, kada je B2 raspodela korišćena kao raspodela za prihode samozaposlenih, gde je tvrđeno da su pojedinci heterogeni u pogledu preduzetničkih karakteristika [14].

Tabela 1: Prikaz dobijanja određene raspodele beta-tipa preko mešavine druge dve raspodele.

Raspodela beta-tipa	Strukturalna raspodela	Druga raspodela mešavine
$GB2(x;a,b,p,q)$	$GG(x;a,\theta,p)$	$InvGG(\theta;a,b,q)$
$B2(x;\ b,p,q)$	$Ga(x;\ \theta,p)$	$InvGa(\theta;b,q)$
$SM(x;a,b,q)$	$Wei(x;a,\theta)$	$InvGG(\theta;a,b,q)$
$Dagum(x;a,b,p)$	$GG(x;a,\theta,p)$	$InvWei(\theta;a,b)$
$Fisk(x;a,b)$	$Wei(x;a,\theta)$	$InvWei(\theta;a,b)$
$Lomaks(x;b,q)$	$Exp(x;\ \theta)$	$InvGa(\theta;b,p)$

1.4 Lorencova kriva i mere nejednakosti

Literatura o Lorencovoj krivoj, merama nejednakosti, kao i srodnim pojmovima je sa vremenom sve obimnija, pa bi bilo lako proširiti rad na stotine strana. Kako nam je u cilju da se najviše pozabavimo definicijom beta-tipa raspodela spomenuli smo samo osnovne pojmove vezane za Lorencovu krivu.

Lorencova kriva (kriva raspodele prihoda), koja je nazvana po ekonomisti Lorencu[5], tj. kriva koncentracije predstavlja grafički prikaz raspodele jednog obeležja nekog skupa jedinica, npr: skupa aktivnog stanovništva prema prihodu. Njeno tumačenje se olakšava ucrtavanjem dijagonale koja predstavlja idealnu raspodelu, pa što je Lorencova kriva udaljenija od nje, utoliko više imamo izraženu situaciju da manji broj ljudi prisvaja najveći deo dohotka. Nažalost, Lorencov pionirski rad nekoliko decenija stajao u engleskoj statističkoj literaturi sve dok ga nije oživeo Gastvirt[6] 1971. godine [15].

Matematička definicija empirijske Lorencove krive: Neka postoji uzorak od n osoba iz date populacije i neka je obeležje X_i prihod osobe $i, i = 1, 2, 3, \dots, n$, tako da važi $X_1 \le X_2 \le \cdots \le X_n$. Lorencova kriva uzorka predstavlja izlomljenu liniju koja redom spaja tačke $(h/n, L_h/L_n), h = 0, 1, 2, \dots, n$, gde je $L_0 = 0$ i $L_h = \sum_{i=1}^{h} X_i$ ukupni prihod h najsiromašnijih osoba. Otuda, Lorencova kriva ima za svoju apcisu kumulativni procenat stanovnika koji primaju dohodak, poređanih rastuće, dok za svoju ordinatu ima odgovarajuću proporciju primljenih dohodaka.

Ocena Lorencove krive sa $n + 1$ tačkom u diskretnom slučaju se definiše na sledeći način:

$$L\left(\frac{k}{n}\right) = \sum_{i=1}^{k} x_{i:n} \Bigg/ \sum_{i=1}^{n} x_{i:n}, \qquad k = 0, 1, \dots, n,$$

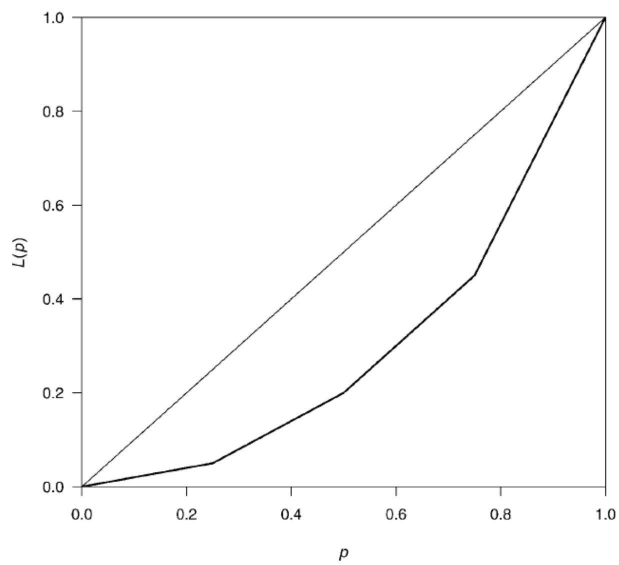
gde $x_{i:n}$ označava i-ti najmanji prihod, a neprekidna kriva $L(u), u \in [0, 1]$, je data sledećom formulom:

$$L(u) = \frac{1}{n \cdot \bar{x}} \cdot \left\{ \sum_{i=1}^{\lfloor un \rfloor} x_{i:n} + (un - \lfloor un \rfloor) \cdot x_{\lfloor un \rfloor + 1:n} \right\}, \qquad 0 \le u \le 1,$$

gde $\lfloor un \rfloor$ označava najveći ceo broj koji nije veći od un.

Na sledećoj slici grafički smo prikazali Lorencovu krivu za vektor prihoda x, gde je $x = (1, 3, 5, 11)$. Po definiciji, dijagonala jediničnog kvadrata odgovara Lorenzovoj krivoj društva u kojoj svi dobijaju isti prihod i na taj način služi kao referentni slučaj u odnosu na koji se mogu meriti stvarne raspodele prihoda.

[5] Max Otto Lorenz, američki ekonomista
[6] Joseph L. Gastwirth, američki statističar



Slika 4: Lorencova kriva za $x = (1, 3, 5, 11)$.

Odgovarajuća definicija za Lorencovu krivu za neprekidni tip raspodela se lako može naslutiti iz definicije krive za diskretne raspodele, i to kao niz standardizovanih empirijskih nepotpunih prvih momenata. Imajući u vidu da je $E(X) = \int_0^1 F^{-1} dt$, gde je funkcija kvantila F^{-1} definisana kao:

$$F^{-1}(t) = \sup\{x | F(x) \leq t\}, t \in [0, 1],$$

prethodnu formulu za $L(u)$ možemo izložiti na sledeći način:

$$L(u) = \frac{1}{E(X)} \cdot \int_0^u F^{-1}(t) dt, u \in [0, 1].$$

Iz ove formule direktno smo izvukli da Lorencova kriva ima sledeća svojstva:

- Lorencova kriva je neprekidna na intervalu $[0, 1]$, gde je $L(0) = 0$ i $L(1) = 1$,
- Lorencova kriva je konveksna i rastuća funkcija.

Sad se vraćamo na Lorencovu krivu GB2 raspodele, koja postoji samo ako je $a \cdot q > 1$. Ova kriva se ne može direktno dobiti iz Gastvirtove prezentacije, imajući u vidu da funkcija kvantila nije dostupna u zatvorenoj formi. Međutim, Lorencova kriva je dostupna u smislu raspodele prvih momenata tako što normalizovani nepotpuni momenti, to jest, funkcije gustine raspodela momenata višeg reda mogu da se izraze preko funkcija gustine GB2 raspodele sa različitim setom parametara. Naime, za $X \sim GB2(a, b, p, q)$ i:

$$F_{(k)}(x) = \frac{\int_0^x t^k \cdot f(t) dt}{E(X^k)}, 0 < x < \infty.$$

dobili smo sledeću jednakost:

$$F_{(k)}(x) = F\left(x, a, b, p + \frac{k}{a}, q - \frac{k}{a}\right), 0 < x < \infty. \tag{22}$$

Osobina koju smo sada prikazali od praktičnog je značaja u kompjuterskim programima koji se koriste za ocenu GB2 raspodele, a takođe se može koristiti za ocenu raspodele momenata višeg reda, menjajući samo vrednosti parametara. Konkretno, teoretske Lorencove krive GB2 raspodele se mogu dobiti crtanjem $F_{(1)}(x)$ u odnosu na $F(x), 0 < x < \infty$. Ova osobina specijalno važi samo za GB2 i B2 raspodele i ne proširuje se ni na jednu od podfamija o kojima govorimo u sledećim poglavljima.

Kao što je navedeno u prethodnom odeljku, GB2 raspodela ima teške repove, tako da postoji samo nekoliko momenata. Ovo implicira da mere nejednakosti, kao što su uopštene mere entropije postoje samo za parametre u određenom rasponu, što je veoma retko upotrebljivo u praksi.

Gore pomenuta relacija (22) dodatno pruža i jednostavan način da dođemo do indeksa nejednakosti, tačnije do $P = F(\mu) - F_{(1)}\mu$, gde je μ srednja vrednost od X.

Poslednje što prikazujemo u ovom odeljku je Đinijev koeficijent GB2 raspodele (najčešće korišćena mera disperzije), koji je jedan dugačak izraz koji uključuje uopštenu hipergeometrijsku funkciju ${}_2F_3$:

$$G = \frac{2 \cdot B(2 \cdot p + \frac{1}{a}, 2 \cdot q - \frac{1}{a})}{p \cdot B(p,q) \cdot B(p + \frac{1}{a}, q - \frac{1}{a})}$$
$$\cdot \left\{ \frac{1}{p} \cdot {}_2F_3 \left[1, p + q, 2 \cdot p + \frac{1}{a}; p + 1, 2 \cdot (p + q); 1 \right] - \frac{1}{p + \frac{1}{a}} \cdot {}_2F_3 \left[1, p + q, 2 \cdot p + \frac{1}{a}; p + \frac{1}{a} + 1, 2 \cdot (p + q); 1 \right] \right\}$$

Za B2 podfamiliju, prethodni izraz je znatno manje rogobatan i jednak je:

$$G = \frac{2 \cdot B(2 \cdot p, 2 \cdot q - 1)}{p \cdot B^2(p,q)}.$$

1.5 Ocena parametara

Postoji relativno malo rezultata koji se odnose na ocenu parametara za ovu nešto komplikovaniju raspodelu sa čak četiri parametra. Jedan od načina ocene parametara je metod maksimalne verodostojnosti. Metod maksimalne verodostojnosti uveden je u matematičku statistiku u drugoj deceniji dvadesetog veka. Ideja ovog metoda jeste da se za ocenu parametra izabere vrednost (tog parametra) pri kojoj je verovatnoća realizacije dobijenog uzorka najveća. Pokazalo se da ovaj metod daje ocene koje su asimptotski (tj. za veliki uzorak) efikasnije od ocena dobijenih na bilo koji drugi način. Kako znamo da je maksimum funkcije verodostojnosti (L) najlakše naći preko logaritma funkcije, pogledajmo sada taj logaritam za slučajni uzorak obima n:

$$logL = n \cdot \log\Gamma(p+q) + n \cdot loga + (a \cdot p - 1) \cdot \sum_{i=1}^{n} logx_i - n \cdot a \cdot p \cdot logb - n \cdot \log\Gamma(p)$$
$$- n \cdot log\Gamma(q) - (p+q) \cdot \sum_{i=1}^{n} log\left[1 + \left(\frac{x_i}{b}\right)^a\right].$$

Izvod funkcije $\log\Gamma(\cdot)$ označavamo sa sa $\psi(\cdot)$, te dobijamo sledeće četiri jednačine iz parcijalnih izvoda po parametrima a, b, p, q tim redosledom:

$$\frac{n}{a} + p \cdot \sum_{i=1}^{n} \log\left(\frac{x_i}{b}\right) = (p+q) \cdot \sum_{i=1}^{n} \log\left(\frac{x_i}{b}\right)\left[\left(\frac{b}{x_i}\right)^a + 1\right]^{-1}, \qquad (23)$$

$$n \cdot p = (p+q) \cdot \sum_{i=1}^{n}\left[1 + \left(\frac{b}{x_i}\right)^a\right]^{-1}, \qquad (24)$$

$$n\psi(p+q) + a \cdot \sum_{i=1}^{n} \log\left(\frac{x_i}{b}\right) = n \cdot \psi(p) + \sum_{i=1}^{n} log\left[1 + \left(\frac{x_i}{b}\right)^a\right], \qquad (25)$$

$$n\psi(p+q) = n\psi(q) + \sum_{i=1}^{n} log\left[1 + (\frac{x_i}{b})^a\right]. \qquad (26)$$

Iz prve dve jednačine primećujemo da su obe linearne po p i q, te se one mogu lako rešiti tako da daju ova dva parametra kao funkcije od a i b. Kada zamenimo nastale izraze u treću i četvrtu jednačinu, dobijamo dve nelinearne jednačine sa dve nepoznate, a što možemo rešiti nekom od poznatih iteracija.

Ispod smo predstavili Fišerovu informacionu matricu za GB2 raspodele. Ova familija je regularna (u smislu metode maksimalne verodostojnosti), tako da ćemo koristiti sledeći izraz:

$$I(\theta) = \left[-E(\frac{d^2\, logL}{d\theta_i d\theta_j})_{ij}\right] =: \begin{bmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ I_{31} & I_{32} & I_{33} & I_{34} \\ I_{41} & I_{42} & I_{43} & I_{44} \end{bmatrix}. \qquad (27)$$

Za $\theta = (a, b, p, q)^T$, iz jednačina (26) i (27) sledi da su drugi izvodi logaritma funkcije maksimalne verodostojnosti konstante u odnosu na p i q, tako da se $I_{33}, I_{34} = I_{43}$ i I_{44} lako dobijaju. Ispod smo prikazali neke od elemenata matrice $I(\theta)$:

$$I_{11} = \frac{1}{a^2 \cdot (1+p+q)} \cdot \left\{1 + p + q + p \cdot q \cdot \left[\psi'(q) + \psi'(p) + (\psi(q) - \psi(p) + \frac{p-q}{p \cdot q})^2 - \frac{p^2+q^2}{p^2 \cdot q^2}\right]\right\} \qquad (28)$$

$$I_{21} = I_{12} = \frac{p - q - p \cdot q \cdot [\psi(p) - \psi(q)]}{b \cdot (1+p+q)}, \qquad (29)$$

$$I_{22} = \frac{a^2 \cdot p \cdot q}{b^2 \cdot (1+p+q)}, \qquad (30)$$

$$I_{23} = I_{32} = \frac{a \cdot q}{b \cdot (p+q)}, \qquad (31)$$

$$I_{24} = I_{42} = -\frac{a \cdot p}{b \cdot (p+q)}, \qquad (32)$$

$$I_{31} = I_{13} = -\frac{q \cdot [\psi(p) - \psi(q)] - 1}{a \cdot (p+q)}, \qquad (33)$$

$$I_{14} = I_{41} = -\frac{p \cdot [\psi(q) - \psi(p)] - 1}{a \cdot (p+q)}, \qquad (34)$$

$$I_{33} = \psi'(p) - \psi'(p+q), \tag{35}$$

$$I_{34} = I_{43} = \psi'(p+q) \tag{36}$$

$$I_{44} = \psi'(q) - \psi'(p+q). \tag{37}$$

U sledećim poglavljima smo prikazali vrednosti elemenata ove matrice za neke od specijalnih slučajeva GB2 raspodele.

1.6 Primena

U ovom odeljku prvog poglavlja navodimo primenu uopštene beta raspodele druge vrste i to u oblasti prihoda i aktuarskih gubitaka, tokom poslednjih decenija prethodnog veka.

Prihodi i bogatstvo:

Vartia i Vartia su 1980. godine [16] iskoristili pomerenu B2 raspodelu sa četiri parametra (pod nazivom skalirana i pomerena F raspodela) kao model za raspodelu oporezovanog prihoda u Finskoj iz 1967. godine. Koristeći metodu maksimalne verodostojnosti i metodu momenata utvrdili su da se njihov model sistemski bolje uklapa od log-normalne raspodele sa dva ili tri parametra.

Mekdonald je 1984. godine [3] iskoristio GB2 i B2 raspodele za porodične prihode u SAD-u iz 1970, 1975. iz 1980. godine. GB2 raspodela nadmašuje 10 drugih raspodela (uglavnom beta i gama tipa), dok B2 raspodela bolje fituje model nego gama ili uopštena gama raspodela, ali je lošija od Sing-Madala raspodele i to za svaki od razmatranih skupova podataka.

Mekdonald je zajedno sa Batlerom, 1989. godine [17] ocenio parametre GB2 raspodele za porodične prihode domaćinstava među belim stanovništvom u SAD-u od 1948-1980 godine, koristeći metod maksimalne verodostojnosti na grupisanim podacima. Pronašli su da parametri oblika, $a \in (2.8, 8.4)$, $p \in (0.17, 0.6)$ i $q \in (0.4, 1.7)$ opadaju tokom vremena, kao i da se najveće vrednosti p i q odnose na podatke posle 1966. godine. Ovo pokazuje da su indeksi repa $a \cdot p$ i $a \cdot q$ otprilike konstantni kroz vreme, dok u centru raspodele postoje značajne promene.

Dvojica statističara, Čakravarti i Mađumder[7] su 1990. godine [18] razmatrali korišćenje GB2 raspodele prilikom modelovanja podataka o prihodima u SAD-u za 1960, 1969, 1980. i 1983. godinu. GB2 raspodela se tada pokazala na drugom mestu u poređenju sa njihovom raspodelom. Međutim, u članku postoje neki kontradiktorni rezultati, dok se u ponovnoj proceni njihovih zaključaka došlo do rezultata da Mađumder-Čakravarti raspodela predstavlja samo reparametrizaciju GB2 raspodele. Mekdonald je fitovao ovaj model za porodične prihode u SAD-u iz 1970, 1980, i 1990. godine koristeći dve različite metode ocene i alternativna grupisanja

[7] Satya R. Chakravarty i M. Majumder, indijski statističari

podataka. Članak Mađumdera i Čakravartija je dao važnu lekciju o fitovanju podataka o prihodima, a takođe nam je ukazao na opasnosti i prepreke kada podaci nisu adekvatno pregledani.

U studiji velike potražnje za specifičnim linijama automobila, 1993. godine [19] korišćena je GB2 raspodela za direktnu ocenu elastičnosti prihoda u odnosu na raspodelu prihoda stanovništva. Ispostavilo se da su ti rezultati u skladu sa onima u tradicionalnim ekonometrijskim studijama potražnje za automobilima.

Mekdonald je 1995. godine proučavao prihode porodica u SAD-u iz 1985. godine [4]. Od 11 raspodela koje su razmatrane, od čega najviše beta i gama tipa, GB2 se rangirala kao prva u smislu fitovanja.

1996. godine je fitovana GB2 raspodela (kao i njene mnoge podfamilije i specijalne slučajeve) pojedinačnim prihodima domaćinstava iz nemačkog socio-ekonomskog panela (SOEP) za godine 1984-1993 [20]. Poređenje sa neparametrijskim ocenama gustine pokazuje da su samo GB2 i, u manjoj meri, Sing-Madala raspodela zadovoljavajuće.

Britanski statističar Parker je 1997. godine [14] razmatrao troparametarsku B2 raspodelu prilikom proučavanja prihoda od samozapošljavanja u Velikoj Britaniji od 1976. do 1991. godine. Polazeći od dinamičnog ekonomskog modela prihoda od samozapošljavanja, utvrdio je da prihodi iznad određenog praga $\gamma < 0$ imaju B2 raspodelu sa dva parametra. On je ocenio parametre modifikovanom metodom momenata, po preporuci iz 1988. godine za modele koji sadrže parametar praga. Ovo je tada predloženo nakon što se ispostavilo da su prihodi postajali progresivno nejednako raspodeljeni, što je pripisano povećanju heterogenosti među samozaposlenima.

Gubici u aktuarstvu:

U aktuarskoj nauci, Kumins[8] je u radu iz 1990. godine [8] koristio GB2 familiju raspodela za modeliranje godišnjih gubitaka usled požara koje su doživeli veliki univerziteti. Ispostavlja se da nije potrebna potpuna fleksibilnost GB2, a jednoparametarski specijalni slučaj, tačnije inverzna eksponencijalna raspodela se pokazala bolje od GB2 raspodele sa četiri parametra. Isti autori su takođe razmatrali podatke o ozbiljnosti gubitaka od požara i fitovali GB2 raspodelu kako za grupisana tako i za pojedinačna posmatranja. U ovom slušaju se biraju specijalni i ograničavajući slučajevi GB2 raspodele sa tri parametra, kao što su Sing-Madala i inverzna uopštena gama raspodela.

Gore pomenuti rad je pružio važnu lekciju o nepotrebnom preteranom korišćenju parametara. Dakle, možemo zaključiti da je GB2 familija raspodela jedna zaista privlačna i fleksibilna familija, ali da uključuje četiri parametra. Iskustvo prikupljeno statističkim aplikacijama tokom poslednjih 100 godina nas uči da raspodele sa četiri parametra ponekad mogu biti preopšte, te nam ne dozvoljavaju da prodremo u suštinu materije.

[8] John D. Cummins, američki aktuar

2 Sing-Madala raspodela

Sing-Madala raspodela je uvedena 1975. godine i u boljoj formi 1976. godine [26, 27] je dobila posebnu pažnju u literaturi posvećenoj raspodelama prihoda. Iako je pod navedenim imenom otkrivena pre GB2 raspodele, bilo bi pogodno tretirati je kao poseban slučaj GB2, a razlog tome ćemo videti u nastavku teksta. U skorašnjoj ekonometrijskoj literaturi često se upoređuje sa Dagum raspodelom, najverovatnije imajući u vidu sličnost njihovih funkcija raspodele.

Za izradu poglavlja korišćeni su sledeći izvori: [1], [2], [3], [4], [5], [8], [12], [20], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [37], [38], [39].

2.1 Definicija

Kao što smo već naveli, Sing-Madala raspodela je specijalni slučaj GB2 raspodele i to za $p = 1$. Njena funkciju gustine je:

$$f(x) = \frac{a \cdot q \cdot x^{a-1}}{b^a \cdot [1 + (\frac{x}{b})^a]^{1+q}}, x > 0, \qquad (38)$$

gde su sva tri parametra a, b, q pozitivna. Ovde je b parametar skaliranja, dok su a, q parametri oblika, gde parametar q utiče samo na desni rep, a a utiče na oba repa raspodele. U nastavku teksta koristimo sledeću oznaku: $GB2(a, b, 1, q) \equiv SM(a, b, q)$.

Važno je napomenuti da je ova raspodela više puta nezavisno otkrivana u nekoliko slabo povezanih područja, shodno tome, poznata je pod različitim imenima. Deluje da je po prvi put razmatrana 1942. godine, od strane Bura [28], gde se pojavljuje kao dvanaesti primer rešenja diferencijalne jednačine koja definiše Burov sistem. Zbog toga se obično naziva Bur XII raspodela, ili samo Bur raspodela, kako je najčešće možemo naći u aktuarstvu. Prema drugim navodima, imamo da je ova raspodela predložena kao raspodela prihoda već 1958. godine [12] u neobjavljenom dokumentu koji je predstavljen na sastanku jednog ekonometrijskog društva. Takođe je poznata i kao Pareto (IV) raspodela, beta-P raspodela ili kao uopštena loglogistička raspodela.

Poseban slučaj gde je $a = q$ se ponekad naziva paralogistička raspodela. Dalji specijalni slučajevi poznati su kao Fisk ili loglogistička, za $q = 1$ i Lomaks raspodela za $a = 1$. Ove raspodele smo detaljnije posmatrali u narednom poglavlju.

Za razliku od GB2 raspodele, funkcija raspodele kod Sing-Madale je dostupna u zatvorenom obliku i data je prilično jednostavnim izrazom:

$$F(x) = 1 - \left[1 + \left(\frac{x}{b}\right)^a\right]^{-q}, x > 0, \qquad (39)$$

Odavde smo direktno uočili da je funkcija kvantila jednako jednostavna:

$$F^{-1}(u) = b \cdot \left[(1-u)^{-\frac{1}{q}} - 1\right]^{\frac{1}{a}}, \text{ za } 0 < u < 1. \tag{40}$$

Dakle, Sing-Madala raspodela je jedna od retkih raspodela za koju funkcija gustine i raspodele, kao i funkcija kvantila imaju jednostavne zatvorene forme. Ovo delom objašnjava činjenicu koju smo pomenuli na samom početku ovog poglavlja, a to je da je ova raspodela više puta nezvisno otkrivana u različitim kontekstima.

Sing i Madala su 1976. godine [27] izveli svoju raspodelu razmatrajući stopu rizika (hazarda) $r(x)$ prihoda. Podsetimo se, stopa rizika je trenutna stopa incidencije, a funkcija rizika je verovatnoća da ako se preživi do određenog trena, već u sledećem trenutku događaj će se desiti. Podsetimo se još i ppadajuće stope neuspeha (DFR, eng. decreasing failure rate) – ona opisuje fenomen u kojem se verovatnoća događaja u fiksnom vremenskom intervalu u budućnosti smanjuje tokom vremena.

Nakon ovoga su Sing i Madala predstavili stopu rizika za $z \coloneqq logx$, i to veličinu koju su nazvali proporcionalna stopa neuspeha (PFR, eng. proportional failure rate):

$$r^*(z) = \frac{dF(z)}{dz} \cdot \frac{1}{1-F(z)}.$$

Ona meri, za bilo koji prihod, šansu da se napreduje ka većem prihodu, u proporcionalnom smislu. Za klasičnu Paretovu raspodelu imamo da je $r^*(z) = \alpha$. Ovo je ujedno i nagib funkcije preživljavanja prihoda u Paretovom dijagramu. Ovo ukazuje na to da $r^*(z)$ mora biti asimptotski konstantna. Sing i Madala su dalje pretpostavili da $r^*(z)$ raste sa z, prvo sa rastućom, a zatim opadajućom brzinom. Kada smo definisali $y \coloneqq -\log(1-F)\,, y' > 0, y'' > 0$, dobili smo sledeću diferencijalnu jednačinu:

$$y'' = a \cdot y' \cdot (\alpha - y'),$$

gde je a konstanta. Prethodni izraz smo reorganizovali na sledeći način:

$$\frac{y''}{y'} + \frac{y''}{\alpha - y'} = a \cdot \alpha,$$

od koga smo integracijom došli do sledećeg:

$$logy' - \log(\alpha - y') = a \cdot \alpha \cdot z + c_1,$$

gde je c_1 konstanta integracije. Odavde smo dobili sledeće:

$$\frac{y'}{\alpha - y'} = e^{a \cdot \alpha \cdot z + c_1},$$

ili:

$$y' = \frac{\alpha \cdot e^{a \cdot \alpha \cdot z + c_1}}{1 + e^{a \cdot \alpha \cdot z + c_1}}.$$

Dalja integracija nam je dala:

$$logy = \frac{1}{a} \cdot \log(1 + e^{a \cdot \alpha \cdot z + c_1}) + c_2,$$

gde je c_2 takođe konstanta integracije. Kada smo y zamenili sa $-\log(1-F)$, $logx$ sa z i preuredili, dobili smo sledeće:

$$F(x) = 1 - \frac{c}{(b + x^{a\cdot\alpha})^{1/a}},$$

gde je $c = (-c_2 - c_1)/\alpha$ i $b = 1/e^{c_1}$. Iz graničnog uslova $F(0) = 0$ sledi da je $c = b^{1/a}$. Tako da smo prethodni izraz sveli na oblik prikazan niže:

$$F(x) = 1 - \frac{b^{1/a}}{(b + x^{a\cdot\alpha})^{1/a}},$$

ili:

$$F(x) = 1 - \frac{1}{(1+a_1 x^{a_2})^{a_3}}, \qquad (41)$$

gde je $a_1 = 1/b$, $a_2 = a\cdot\alpha$ i $a_3 = 1/a$, što je zapravo Sing-Madala funkcija raspodele, imajući u vidu $a_2 = a$, $a_3 = q$ i $a_1 = b^{-a_2}$.

Ovo izvođenje je pokazalo da se Sing-Madala raspodela karakteriše proporcionalnom stopom neuspeha – troparametarskom funkcijom u pogledu moći prihoda:

$$r^*(z) = \frac{a\cdot q\cdot e^{a\cdot z - logb}}{1 + e^{a\cdot z - logb}},$$

Sing i Madalin pristup je kritikovan od strane Kramera 1978. godine [29], koji je smatrao da analogija sa stopama neuspeha nije ubedljiva. Bez obzira na to, izgleda da su ove osobine pouzdanosti neke značajne osobine empirijskih raspodela prihoda.

2.2 Momenti raspodele i druga osnovna svojstva

U drugom odeljku ovog poglavlja pokazali smo momente Sing-Madala raspodele i neka od svojstava. k-ti momenat postoji za svako k u intervalu $-a < k < a\cdot q$ i imajući u vidu izraz (12) on iznosi:

$$E(X^k) = \frac{b^k \cdot B(1+\frac{k}{a}, q-\frac{k}{a})}{B(1,q)} = \frac{b^k\cdot\Gamma\left(1+\frac{k}{a}\right)\cdot\Gamma(q-\frac{k}{a})}{\Gamma(q)}. \qquad (42)$$

Ovde ćemo videti momente:

$$E(X) = \frac{b\cdot B(1+\frac{1}{a}, q-\frac{1}{a})}{B(1,q)} = \frac{b\cdot\Gamma\left(1+\frac{1}{a}\right)\cdot\Gamma(q-\frac{1}{a})}{\Gamma(q)}. \qquad (43)$$

i

$$D(X) = \frac{b^2\cdot\{\Gamma(q)\cdot\Gamma(1+2/a)\cdot\Gamma(q-2/a) - \Gamma^2(1+1/a)\cdot\Gamma^2(q-1/a)\}}{\Gamma^2(q)}. \qquad (44)$$

Odatle koeficijent devijacije iznosi (CV, eng. coefficient of variation):

$$CV = \frac{\sigma}{E(X)} = \sqrt{\frac{\Gamma(q)\cdot\Gamma(1+2/a)\cdot\Gamma(q-2/a)}{\Gamma^2(1+1/a)\cdot\Gamma^2(q-1/a)} - 1}, \tag{45}$$

a faktori oblika su:

$$\sqrt{\beta_1} = \frac{\Gamma^2(q)\cdot\lambda_3 - 3\cdot\Gamma(q)\cdot\lambda_2\cdot\lambda_1 + 2\cdot{\lambda_1}^3}{\left[\Gamma(q)\cdot\lambda_2 - {\lambda_1}^2\right]^{3/2}}, \tag{46}$$

$$\beta_2 = \frac{\Gamma^3(q)\cdot\lambda_4 - 4\cdot\Gamma^2(q)\cdot\lambda_3\cdot\lambda_1 + 6\cdot\Gamma(q)\cdot\lambda_2\cdot{\lambda_1}^2 - 3\cdot{\lambda_1}^4}{\left[\Gamma(q)\cdot\lambda_2 - {\lambda_1}^2\right]^2}, \tag{47}$$

gde imamo $\lambda_i = \Gamma(1+i/a)\cdot\Gamma(q-1/a)\,, i = 1,2,3,4$.

Iz (11) imamo da se moda Sing-Madala raspodele postiže u 0 ili se postiže u sledećoj tački:

$$x_{mode} = b\cdot\left(\frac{a-1}{a\cdot q+1}\right)^{1/a}, \text{ ako je } \ a > 1. \tag{48}$$

Dakle, vidimo da se moda smanjuje sa q, što odražava činjenicu da desni rep postaje lakši kako se q povećava.

Zatvoreni oblici funkcije raspodele i funkcije kvantila Sing-Madala raspodele takođe omogućavaju pogodne manipulacije sa karakteristikama statistika poretka. Ovde imamo:

$$E(X_{1:n}) = \frac{b\cdot\Gamma(1+1/a)\cdot\Gamma(n\cdot q-1/a)}{\Gamma(n\cdot q)}. \tag{49}$$

Sledeći iskaz smo dobili iz osobine zatvorenosti i izraza za momente:

$$X \sim SM(a,b,q) \Rightarrow X_{1:n} \sim SM(a,b,n\cdot q). \tag{50}$$

Koristeći rekurziju sada smo izračunali momente od $X_{k:n}, 2 \le k \le n$:

$$E\left(X_{i:n}^k\right) = i\cdot\binom{n}{i}\cdot b^k\cdot q\cdot\sum_{j=0}^{i-1}(-1)^j\cdot\binom{i-1}{j}\cdot B\left(q\cdot n - q\cdot i + q - \frac{k}{a} + q\cdot j, 1 + \frac{k}{a}\right).$$

Kako je funkcija raspodele dostupna u zatvorenoj formi, ovo je takođe slučaj i za stopu rizika (hazarda), koja je data sledećom funkcijom:

$$r(x) = \frac{a\cdot q\cdot x^{a-1}}{b^a\cdot\{1+(x/b)^a\}}, x > 0. \tag{51}$$

Opšti oblik ove funkcije zavisi od vrednosti parametra oblika a. Za sve $a > 0$ stopa rizika (hazarda) se polako smanjuje. Za $a > 1$ imamo unimodalnu funkciju, dok za $a \le 1$ se smanjuje za sve $x > 0$. Napomenuli bismo da je parametar q parametar skaliranja u (51) i ne određuje oblik funkcije.

2.3 Karakterizacija i reprezentacija

U statističkoj i ekonometrijskoj literaturi dostupno je nekoliko karakterizacija Sing-Madala raspodele. Većina njih se verovatno najbolje shvati ukoliko ih povežemo sa karakteristikama eksponencijalne raspodele, najpoznatije i najviše korišćene raspodele na skupu $\mathbb{R}_+$. Eksponencijalna raspodela se izuzetno dobro ponaša, jer su njena funkcija raspodele, funkcija kvantila, itd. sve dostupne u jednostavnom zatvorenom obliku. Samim tim nije iznenađujuće što postoji bogata literatura o karakterizaciji.

Za svaku monotonu funkciju $h(\cdot)$, karakterizacija od X je ekvivalentna karakterizaciji od $h(X)$. Stoga, ako se određena raspodela može povezati sa eksponencijalnom raspodelom, željeni predmet karakterizacije postaje dostupan. Sing-Madala raspodela je povezana sa eksponencijalnom raspodelom preko sledećeg:

$$X \sim SM(a,b,q) \Leftrightarrow log\left[1+\left(\frac{x}{b}\right)^a\right] \sim Exp(q), \qquad (52)$$

gde $Exp(q)$ označava eksponencijalnu raspodelu sa parametrom skaliranja q. Tako da je $h(\cdot)$ potrebna monotona transformacija, gde je:

$$h(x) = log\left[1+\left(\frac{X}{b}\right)^a\right], a > 0.$$

Sada su mnoge karakterizacije Sing-Madala raspodele dostupne primenom transformacije $h^{-1}(\cdot)$ na promenljive sa eksponencijalnom raspodelom.

Svojstvo odsustva sećanja je najpopularnija i intuitivno transparentna karakterizacija eksponencijalne raspodele. Jedan od njegovih mnogih ekvivalentnih izraza dat je u smislu sledeće funkcionalne jednačine:

$$P(X > x+y) = P(X > x) \cdot P(X > y). \qquad (53)$$

El-Saidi, Sing i Bartoluči su 1990. godine pokazali da sledeća funkcionalna jednačina karakteriše Sing-Madala raspodelu, među svim neprekidnim raspodelama na $[0,\infty)$ [30]:

$$P\left[1+\left(\frac{X}{b}\right)^a > x\cdot y\right] = \left[1+\left(\frac{X}{b}\right)^a > x\right]\cdot\left[1+\left(\frac{X}{b}\right)^a > y\right],$$

gde su $a, b > 0$ i $x, y > 1$.

Ovo svojstvo kao i svojstva koja smo pokazali u nastavku teksta proizilaze is svojstva odsustva sećanja (53) eksponencijalne raspodele. U ostatku ovog odeljka smo zanemarili parametar skaliranja b i razmotrili $h(x) = \log(1+x^a)$.

Dalja karakterizacija Sing-Madala raspodele je u vezi sa statistikama poretka. Pretpostavimo da je niz $X_1, \dots, X_n$ slučajni uzorak sa apsolutno neprekidnom raspodelom F i neka niz $X_{1:n}, \dots, X_{n:n}$ označava njihove statistike poretka. Al-Husaini je 1991. godine [31] pokazao da $X_i \sim SM(a,1,q)$, $i = 1, \dots, n$, ako i samo ako postoji $a > 0$ tako da su sledeće veličine nezavisne:

$$Z_1 = 1 + X_{1:n}^a, Z_2 = \frac{1 + X_{2:n}^a}{1 + X_{1:n}^a}, Z_3 = \frac{1 + X_{3:n}^a}{1 + X_{2:n}^a}, \dots, Z_n = \frac{1 + X_{n:n}^a}{1 + X_{n-1:n}^a}.$$

1996. godine je pokazano da je Sing-Madala raspodela karakterisana i sledećim svojstvom [32]:

$$E[\log(1 + X^a)\,|X > y] = \log(1 + y^a) + c, \tag{54}$$

za sve nenegativne vrednosti y i za određene $a > 0, c > 0$ među neprekidnim raspodelama na intervalu $[0, \infty)$ za koje je $E|h(X)| < \infty$. Ovo je u suštini ekvivalent uslova $E(X|X > y) = y + E(X)$, za sve nenegativne y, za koji je poznato da karakteriše eksponencijalnu raspodelu.

U nastavku smo prikazali tri povezana rezultata. Prvo, za $h(x) = \log(1 + x^a)$ sledeća rekurzivna relacija karakteriše $SM(a, 1, q)$ raspodelu:

$$E[h^k(X)\big|X > y] = h^k(y) + \tfrac{1}{q} \cdot k \cdot E[h^{k-1}(X)\big|X > y]. \tag{55}$$

Kada dodamo još jedan uslov da je $Eh^2(X) < \infty$, isto važi ukoliko je za sve y i neko fiksno c:

$$var[h(X)|X > y] = c^2. \tag{56}$$

Treće, sledeći uslov takođe karakteriše $SM(a, 1, q)$ raspodelu:

$$E[h^k(X_{i:n})] = E[h^k(X_{i-1:n})] + \tfrac{k}{q\cdot(n-i+1)} \cdot E[h^{k-1}(X_{i:n})]. \tag{57}$$

Opet, sva tri rezultata su povezana sa karakterizacijama eksponencijalne raspodele putem transformacije h. Karakterizacija (55) je verzija momenta višeg reda (54).

U delu teksta niže predstavili smo dodatne karakterizacije zasnovane na uslovnim očekivanjima statistika poretka. Naime, ukoliko je $a = q \cdot (n - i)$ nezavisno od x, onda sledeći iskaz karakteriše Sing-Madala raspodelu među neprekidnim raspodelama na intervalu $[0, \infty)$ za koje je $F(0) = 0$:

$$E\left[X_{i+1:n}^k|X_{i:n} = x\right] = \tfrac{a}{a-1} \cdot \left[x^k + \tfrac{1}{a}\right]. \tag{58}$$

Imajući u vidu da je:

$$E[X_{n:n}^k|X_{n-1:n} = x] = E[X^k|X > x] \text{ i } E[X_{1:n}^k|X_{2:n} = y] = E[X^k|X \le y],$$

ovo ćemo prikazati u sledećem obliku:

$$E[X^k|X \ge x] = \frac{q}{q-1} \cdot \left[x^k + \frac{1}{q}\right].$$

Ispod smo prikazali još jednu karakterizaciju:

$$E\left[X_{1:n}^k|X_{2:n} = x\right] = \tfrac{1}{q-1} - \tfrac{q}{q-1} \cdot \tfrac{1-F(x)}{F(x)} \cdot x^k. \tag{59}$$

Primetimo sada za kraj ovog odeljka, a posmatrajući (21) za GB2 raspodelu, da Sing-Madala raspodelu možemo smatrati složenom Vejbulovom raspodelom čiji parametar skaliranja ima inverznu uopštenu gama raspodelu:

$$Wei(a, \theta) \bigwedge_\theta InvGG(a, b, q) = SM(a, b, q). \tag{60}$$

2.4 Lorencova kriva i mere nejednakosti

Kao što smo već spomenuli, funkcija kvantila raspodele Sing-Madala dostupna je u zatvorenom obliku. Shodno tome, njen (normalizovani) integral, ili Lorencova kriva, takođe je dostupna u relativno jednostavnoj formi, tačnije:

$$L(u) = I_y\left(1 + \frac{1}{a}, q - \frac{1}{a}\right), 0 \leq u \leq 1, \tag{61}$$

gde je $y = 1 - (1-u)^{1/q}$. Podklasa, gde je $q = (a+1)/a$ je još analitički jednostavnija i daje Lorencovu krivu koja je simetrična:

$$L(u) = \left[1 - (1-u)^{a/(a+1)}\right]^{(a+1)/a}, 0 < u < 1. \tag{62}$$

Pomoću (62) uvodimo termin Lorencovo uređenje na skupu svih nenegativnih slučajnih promenljivih sa konačnim pozitivnim očekivanjem, oznakom $\geq_L$, koje se definiše sledećim:

$$X \leq_L Y \Leftrightarrow L_X(u) \geq L_Y(u), \text{ gde } u \in [0,1].$$

Za $X_i \sim SM(a_i, b_i, q_i), i = 1, 2$, neophodni i dovoljni uslovi za $X_1 \geq_L X_2$ su sledeći:

$$a_1 \leq a_2 \text{ i } a_1 \cdot q_1 \leq a_2 \cdot q_2. \tag{63}$$

Iz izraza (49) smo dobili uopštene Đinijeve koeficijente:

$$G_n = 1 - \frac{E(X_{1:n})}{E(X)} = 1 - \frac{\Gamma(n \cdot q - 1/a) \cdot \Gamma(q)}{\Gamma(n \cdot q) \cdot \Gamma(q - 1/a)},$$

za $n = 2, 3, 4, \ldots,$ gde za $n = 2$ smo dobili obični Đinijev koeficijent:

$$G = 1 - \frac{\Gamma(2q - 1/a) \cdot \Gamma(q)}{\Gamma(2q) \cdot \Gamma(q - 1/a)}. \tag{64}$$

Za kraj ovog poglavlja osvrnuli smo se i na prvostepenu stohastičku dominaciju (FSD), koja omogućava parcijalno uređenje slučajnih promenljivih. Kloner je 2000. godine [33] predstavio neophodne, kao i dovoljne uslove za pvostepenu stohastičku dominaciju u Sing-Madala familiji. Uslovi $a_1 \geq a_2$, $a_1 \cdot q_1 \leq a_2 \cdot q_2$ i $b_1 \geq b_2$ su dovoljni za $X_1 \geq_{FSD} X_2 \Leftrightarrow F_{X_1}(\mu) \leq F_{X_2}(\mu)$, za sve μ, dok su uslovi $a_1 \geq a_2$, $a_1 \cdot q_1 \leq a_2 \cdot q_2$ neophodni. Sada nam je interesantno da uporedimo ove uslove sa onima za Lorencovu uređenost (63): iako je uslov $a_1 \cdot q_1 \leq a_2 \cdot q_2$ takođe uslov i za $X_1 \geq_L X_2$, imamo da se drugi uslov $a_1 \geq a_2$ pojavljuje u obrnutom obliku od onog u (63). Razlog za ovo je taj što FSD opisuje uređenost po „veličini“, dok Lorencova opisuje uređenost po „varijabilnosti“. Naime, uslovi $a_1 \leq a_2$ i $a_1 \cdot q_1 \leq a_2 \cdot q_2$ znače da X_1 ima teži levi i teži desni rep i stoga je više promenljiva od X_2. Sa druge strane, uslovi $a_1 \geq a_2$, $a_1 \cdot q_1 \leq a_2 \cdot q_2$ znače da X_1 ima lakši levi i teži desni rep i stoga je stohastički veća od X_2.

2.5 Ocena parametara

Sing i Madala su 1976. godine [27] ocenili parametre raspodele metodom regresije minimizirajući sledeću sumu:

$$\sum_{i=1}^{n}\left\{log[1-F(x_i)]+q\cdot log\left[1+\left(\frac{x_i}{b}\right)^a\right]\right\}^2, \tag{65}$$

tj, nelinearnom regresijom najmanjih kvadrata. Čak 21. godinu kasnije Stopa [34] je diskutovao o dodatnoj metodi regresije koristeći elastičnost $dlogF(x)/dlogx$ raspodele. Dobijene ocene se mogu koristiti, na primer, kao početne vrednosti u metodi maksimalne verodostojnosti.

Logaritam funkcije verodostojnosti za potpun slučajan uzorak obima n je sledeći:

$$logL = n\cdot\log q + n\cdot loga + (a-1)\cdot\sum_{i=1}^{n} logx_i - n\cdot a\cdot logb - (1+q)\cdot\sum_{i=1}^{n} log\left[1+\left(\frac{x_i}{b}\right)^a\right],$$

odakle smo dobili jednačine prikazane niže:

$$\frac{n}{a}+\sum_{i=1}^{n}\log\left(\frac{x_i}{b}\right) = (1+q)\cdot\sum_{i=1}^{n}\log\left(\frac{x_i}{b}\right)\left[\left(\frac{b}{x_i}\right)^a+1\right]^{-1}, \tag{66}$$

$$n = (1+q)\cdot\sum_{i=1}^{n}\left[1+\left(\frac{b}{x_i}\right)^a\right]^{-1}, \tag{67}$$

$$\frac{n}{q}=\sum_{i=1}^{n} log\left[1+\left(\frac{x_i}{b}\right)^a\right]. \tag{68}$$

Specijalno, iz informacione matrice GB2 raspodele (27), za $\theta = (a,b,q)^T$ smo dobili

$$I(\theta)=\left[-E\left(\frac{d^2\,logL}{d\theta_i d\theta_j}\right)_{ij}\right] =: \begin{bmatrix} I_{11} & I_{12} & I_{13}\\ I_{21} & I_{22} & I_{23}\\ I_{31} & I_{32} & I_{33}\end{bmatrix}. \tag{69}$$

gde su:

$$I_{11}=\frac{1}{a^2\cdot(2+q)}\cdot\{q\cdot[\,\psi'(q)+\psi'(1)+(\psi(q)-\psi(1)-1)^2]+2\cdot[\psi(q)-\psi(1)]\}, \tag{70}$$

$$I_{21}=I_{12}=\frac{1-q-q\cdot[\psi(q)-\psi(1)]}{b\cdot(2+q)}, \tag{71}$$

$$I_{22}=\frac{a^2\cdot q}{b^2\cdot(2+q)}, \tag{72}$$

$$I_{23}=I_{32}=-\frac{a}{b\cdot(1+q)}, \tag{73}$$

$$I_{31} = I_{13} = -\frac{\psi(q) - \psi(1) - 1}{a \cdot (1+q)}, \tag{74}$$

$$I_{33} = \frac{1}{q^2}. \tag{75}$$

Za I_{33} koristili smo identitet $\psi'(x) - \psi'(x+1) = x^{-2}$.

Šmitlajn je 1983. [35] pružio eksplicitni izraz inverzne Fišerove informacione matrice (koristeći drugu parametrizaciju), kao i same informacione matrice kada su podaci grupisani i/ili cenzurisani prvog tipa. Upoređivanje tih formula sa gore navedenim formulama omogućuje nam ocenu gubitka informacija usled dejstva grupisanja i / ili cenzure.

Poslednja stavka koju smo spomenuli kod ocene parametara je alternativa metodi maksimalne verodostojnosti. Kao alternativu, 1993. godine Šah i Gokle [36] su razmatrali ocenu parametara metodom maksimalnog proizvoda razmaka (MPS, eng. maximum product of spacings). Ovaj metod daje ocenu parametra θ maksimizacijom:

$$H = \frac{1}{n+1} \cdot \sum_{i=1}^{n+1} log\{F(x_i, \theta) - F(x_{i-1}, \theta)\},$$

$i = 1, 2, \dots, n+1$, gde je $x_0 = -\theta$ i $x_{n+1} = \theta$. Iz jedne simulacije koja koristi deset kombinacija parametara i devet uzoraka u rasponu od $n = 10$ do $n = 150$, oni su zaključili da je MPS metoda superiornija od metode maksimalne verodostojnosti, barem za male uzorke.

2.6 Primena

U ovom odeljku ćemo navesti primenu Sing-Madala raspodele, slično kao u prvom poglavlju, kroz dva ključna polja primene - prihodi i aktuarski gubici.

Prihodi i bogatstvo:

Sing i Madala su 1976. godine [27] uporedili svoj model sa rezultatima drugog modela iz 1974. godine koji je koristio gama raspodelu za porodične prihode u SAD-u u periodu 1960-1972 godine. Na osnovu poređenja su zaključili da njihov model pruža bolji fit od gama ili logormalne raspodele.

Važno je napomenuti da u poređenju sa lognormalnom, gama i beta tipom raspodele prve vrste za porodične prihode u SAD-u za 1960. godinu i za period 1969-1975 godine Sing-Madala raspodela se generalno bolje pokazala od ostalih pomenutih, pri čemu se u samo nekoliko slučajeva beta raspodela prvog tipa pokazala nešto boljom.

1983. godine Dagum[9] je fitovao Sing-Madala raspodelu na podatke iz 1978. godine [37], takođe o porodičnim prihodima u SAD-u i ovaj fit je nadmašio dvoparametarsku lognormalnu i gama raspodelu. Međutim, Dagum raspodela III i I tipa (sa četiri i tri parametra) se pokazala još boljom.

Za prihode u Japanu za period 1963-1971, SM raspodela je bolja od Fisk, beta, gama, logormalne i Pareto (II) raspodele [38]. Međutim, pokazano je da potpuna fleksibilnost Sing-Madala raspodele ipak nije potrebna i da dvoparametarska Fisk raspodela već obezbeđuje odgovarajući fit. Raspodela je takođe fitovana na podatke koji su prikupljeni u anketi o redistribuciji od strane 3 japanska statističara [24]. Razmatrane su četiri radne klase, kao i primarni i preraspodeljeni prihodi, a primenjeno je čak pet različitih tehnika ocene. U kasnijoj studiji koja koristi isti skup podataka, korišćen je metod maksimalne verodostojnosti na osnovu individualnih zapažanja [39]. Ovde je model Sing-Madala skoro uvek najbolji od poređenih šest različitih funkcija (uključujući uopštenu gama i Vejbulobu raspodelu).

Primena Sing-Madala raspodele na podatke o prihodima među nemačkim domaćinstvima u periodu 1984-1993 godine pokazuje da on jeste jedan od dva odgovarajuća modela. Dvoparametarski modeli, kao što su gama ili Vejbulova raspodela nisu bili prikladni za ove podatke [20].

Gubici u aktuarstvu:

U aktuarskoj literaturi, 1983. godine dvojica aktuara Hog i Klugman [25] su fitovali Sing-Madala raspodelu (tada poznatu pod imenom Burova raspodela) na 35 opservacija o gubicima usled uragana u SAD-u. U poređenju sa Vejbulovim i logormalnim modelima, Sing-Madala u ovom slučaju deluje da je prekomerno parametrizovana.

U već pomenutom radu Kuminsa iz 1990. godine [8] Sing-Madala raspodela je fitovana na podacima o zbirnim gubicima od požara. Iako se raspodela dobro pokazala, manje parametrizovani specijalni slučajevi GB2 raspodele, kao što je inverzna eksponencijalna raspodela, su se pokazali još boljim. Takođe su razmatrani podaci o ozbiljnosti gubitaka od požara i fitovali su GB2 raspodelu kako za grupisana, tako i za pojedinačna posmatranja. Ovde se Sing-Madala ne razlikuje od GB2 raspodele sa četiri parametra za pojedinačne opservacije, tako da je u ovom slučaju raspodela stvar izbora.

[9] Camilo Dagum, argentinski statističar i ekonometrista

3 Fisk i Lomaks raspodele

Treće poglavlje rada se sastoji iz dva odeljka, definicije raspodele Fiska i definicije i osnovnih svojstava raspodele Lomaksa. Kako su momenti i druga osnovna svojstva predstavljena u opštijoj formi u prethodnim poglavljima, u ovom poglavlju spominjemo samo one rezultate koji nisu prošireni na uopštenje raspodele.

Za izradu poglavlja korišćeni su sledeći izvori: [1], [3], [4], [12], [30], [40], [41], [42], [43].

3.1 Raspodela Fiska

Kao što smo već naveli u prvom odeljku prvog poglavlja, Fisk raspodela je specijalni slučaj GB2 raspodele i to za $p = q = 1$. Prvo što smo prikazali jeste funkcija raspodele i to koristeći izraz (39), tj. funkciju raspodele SM raspodele, a kako je ona specijalni slučaj od GB2, kada je $p = 1$, imamo:

$$F(x) = 1 - \left[1 + \left(\frac{x}{b}\right)^{a}\right]^{-1} = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-1}, x > 0, \qquad (76)$$

dok je funkcija gustine sledeća:

$$f(x) = \frac{a \cdot x^{a-1}}{b^{a} \cdot [1 + (\frac{x}{b})^{a}]^{2}}, x > 0, \qquad (77)$$

gde su parametri $a, b > 0$.

Fisk raspodela može se naći u literaturi i pod nazivima loglogistička raspodela ili Pareto (III) raspodela, ukoliko je dodat još jedan parametar lokacije. Termin loglogistička može se objasniti time što je raspodela od $logX$ logistička, kod koga je a parametar skaliranja, dok je $logb$ parametar lokacije.

Korisno svojstvo raspodele Fiska je to što ona dozvoljava nemonotone stope rizika (hazarda), tačnije:

$$h(x) = \frac{a \cdot x^{a-1}}{b^{a} \cdot [1 + (x/b)^{a}]}, x < 0, \qquad (78)$$

koje su opadajuće za $a \leq 1$, a inače unimodalne sa modom u $x = b \cdot (a-1)^{1/a}$. Među raspodelama o kojima smo govorili u ovom radu, Fisk raspodela je najjednostavnija raspodela sa ovom osobinom. Nasuprot Fisk raspodeli, popularna Vejbulova raspodela sa dva parametra dozvoljava samo monotone stope rizika.

Kako je b parametar skaliranja, a a je jedini parametar oblika, ne možemo očekivati veliku fleksibilnost u vezi sa merama nejednakosti. Kako bismo lakše naveli dovoljan i potreban uslov za Lorencovu uređenost u ovom slučaju, prvo smo naveli potrebne uslove kod GB2 raspodele. Klajber je 1999. godine [40] pokazao da za $X_i \sim GB2(a_i, b_i, p_i, q_i), i = 1, 2$, sledeći uslovi zajedno impliciraju $X_1 \geq_L X_2$:

$$a_1 \leq a_2, \; a_1 \cdot p_1 \leq a_2 \cdot p_2 \text{ i } a_1 \cdot q_1 \leq a_2 \cdot q_2 \,. \tag{79}$$

Kod raspodele Fiska, Lorencovo uređenje je linearno, specijalno za $X_i \sim Fisk(a_i, b_i), i = 1, 2,$ iz uslova (79) smo dobili sledeće:

$$X_1 \geq_L X_2 \; \Leftrightarrow \; a_1 \leq a_2, \tag{80}$$

gde je $a_i > 1, i = 1, 2, \dots$. Kod Fisk raspodele izraz za Đinijev koeficijent je još jednostavniji nego kod obične Paretove raspodele i jednak je:

$$G = \frac{1}{a}. \tag{81}$$

Među svim neprekidnim raspodelama na intervalu $[0, \infty)$ Fisk raspodela se karakteriše sledećom osobinom:

$$P\left[1 + \left(\frac{X}{b}\right)^a > x \cdot y\right] = P\left[1 + \left(\frac{X}{b}\right)^a > x\right] \cdot P\left[1 + \left(\frac{X}{b}\right)^a > y\right], \tag{82}$$

gde su $a, b > 0$ i $x, y > 1$ [41]. Prethodni izraz su 1990. godine uopštili El-Saidi, Sing i Bartoluči [30], a to je u direktnoj vezi sa svojstvom odsustva sećanja eksponencijalne raspodele, o čemu smo već pisali u prethodnom poglavlju.

Što se tiče ocene parametara, važno je da spomenemo da je Fišerova informaciona matrica Fisk modela dijagonalna.

Čen je 1997. godine [42] izveo tačne intervale poverenja i testove za parametar oblika a raspodele Fiska. Uočivši da je raspodela količnika $\xi = M_A / M_G$, tj. aritmetičke i geometrijske sredine od:

$$\left(\frac{X_{1:n}}{b}\right)^a, \left(\frac{X_{2:n}}{b}\right)^a, \dots, \left(\frac{X_{k:n}}{b}\right)^a$$

bez parametara, sa striktno rastućom funkcijom raspodele, dobio je tačke percentila za $3 \leq k \leq n \leq 30$ preko Monte Karlo simulacije. Imajući u vidu da je razmatran mali obim uzorka, ovi rezultati su možda korisniji u aktuarstvu, nego u oblasti raspodela prihoda.

3.2 Raspodela Lomaksa (Pareto II raspodela)

Sledeći dvoparametarski slučaj GB2 raspodele je raspodela Lomaksa, koja se dobija za $a = p = 1$. U ovom odeljku definisali smo funkcije raspodele i gustine, videli čemu je jednaka stopa hazarda i naveli uslove za Lomaksovo uređenje. Raspodela je dobila naziv po statističaru Lomaksu, koji ju je predstavio 1954. godine [43]. Funkcija raspodele je:

$$F(x) = 1 - \left[1 + \frac{x}{b}\right]^{-q}, x > 0, \tag{83}$$

dok je funkcija gustine:

$$f(x) = \frac{q}{b} \cdot \left(1 + \frac{x}{b}\right)^{-q-1}, x > 0. \tag{84}$$

Dakle, to je zapravo Sing-Madala raspodela gde je $a = 1$. Raspodela Lomaksa je možda poznatija kao Paretova (II) raspodela, kako ju je 1983. godine nazvao Arnold [12]. Odnos između raspodele Lomaksa i obične Paretove raspodele možemo prikazati na sledeći način:

$$X \sim Lomax(b, q) \Leftrightarrow X + b \sim Par(b, q).$$

Stopa rizika je prikazana niže:

$$r(x) = \frac{q}{b + x}, x > 0, \tag{85}$$

što je striktno opadajuća funkcija za sve dopuštene vrednosti parametara.

Za kraj ovog poglavlja osvrnuli smo se na Lorencovo uređenje, kao i u prethodnim poglavljima. Slično kao u slučaju raspodele Fiska, Lorencovo uređenje je linearno, tačnije za $X_i \sim Lomax(b_i, q_i), i = 1, 2,$ važi sledeće:

$$q_1 \leq q_2 \Leftrightarrow X_1 \geq_L X_2, \tag{86}$$

imajući u vidu da je $q_i > 1$. Đinijev koeficijent je takođe jednostavnog oblika:

$$G = \frac{q}{2 \cdot q - 1}. \tag{87}$$

4 Uopštena beta raspodela prve vrste

Poslednji deo rada se sastoji iz dva odeljka, definicija i svojstva uopštene beta raspodele prve vrste (u daljem tekstu: GB1) i njena primena. Najpoznatiji primer ove raspodele je njen specijalni slučaj, standardna beta raspodela (u daljem tekstu: B1). GB1 raspodelu je prvi predstavio Mekdonald 1984. godine [3] kao raspodelu prihoda, dok je u iste svrhe nešto više od decenije korišćena B1.

Za izradu poglavlja korišćeni su sledeći izvori: [1], [2], [3], [4], [7], [21], [22].

4.1 Definicija i svojstva

GB1 raspodela je definisana preko svoje funkcije gustine:

$$f(x) = \frac{a \cdot x^{a \cdot p-1} \cdot [1-(x/b)^a]^{q-1}}{b^{a \cdot p} \cdot B(p,q)}, 0 \leq x \leq b, \tag{88}$$

gde su sva četiri parametra a, b, p, q pozitivna. Ovde je b parametar skaliranja, dok su a, p, q parametri oblika. Kada je $a = 1$, dobijamo beta raspodelu sa tri parametra:

$$f(x) = \frac{x^{p-1} \cdot [1-x/b]^{q-1}}{b^{p} \cdot B(p,q)}, 0 \leq x \leq b. \tag{89}$$

Ispod smo prikazali relaciju između GB2 i GB1 raspodele, koja je zapravo uopštenje nama dobro poznate veze između B1 i B2 raspodele:

$$X \sim GB2(a,b,p,q) \Rightarrow \left(\frac{X^a}{1+X^a}\right)^{\frac{1}{a}} \sim GB1(a,b,p,q).$$

Funkcije raspodele GB1 i B1 se ne mogu izraziti preko elementarnih funkcija. Međutim, imajući u vidu (8), one su dostupne preko Gausove hipergeometrijske funkcije $_2F_1$ i to u sledećem obliku (Mekdonald, 1984. godina, izvor 3):

$$F(x) = \frac{\left(\frac{x}{b}\right)^a}{p \cdot B(p,q)} \cdot {}_2F_1\left(p, 1-q; p+1; \left(\frac{x}{b}\right)^a\right), \qquad x > 0. \tag{90}$$

Po analogiji, kao kod GB2 raspodele, prethodno pomenute raspodele možemo napisati kao regularnu nekompletnu beta funkciju, u slučaju GB1:

$$F(x) = I_z(p,q), \text{ gde je } z = \left(\frac{x}{b}\right)^a, x > 0\ , \tag{91}$$

Momenti GB1 raspodele postoje za k koje se nalazi u intervalu $-a \cdot p < k < \infty$ i jednaki su sledećem:

$$E(X^k) = \frac{b^k \cdot B(p + \frac{k}{a}, q)}{B(p,q)} = \frac{b^k \cdot \Gamma\left(p + \frac{k}{a}\right) \cdot \Gamma(p+q)}{\Gamma(p + q + \frac{k}{a}) \cdot \Gamma(p)}. \tag{92}$$

Lorencovo uređenje unutar GB1 familije raspodela proučavao je statističar Vilfling [21], koji je obezbedio četiri seta zadovoljavajućih uslova. Napominjući da je gustina GB1 raspodele pravilno promenljiva u 0 sa indeksom $ap - 1$, može se zaključiti da je uslov $a_1 \cdot p_1 \leq a_2 \cdot p_2$ neophodan za $X_1 \geq_L X_2$.

Kompletna karakterizacija Lorencovog uređenja trenutno ne postoji. Ono što možemo prikazati jesu sledeće dve relacije:

- Ako $X \sim GB2(a,b,p,q)$ i $Y \sim GB1(a,b,p,q)$, onda važi $X \geq_L Y$;
- Ako $X \sim GB1(a,b,p,q)$ i $Y \sim GG(\tilde{a},\tilde{b},\tilde{p})$ – gde je GG oznaka za uopštenu gama raspodelu, gde je $a \geq \tilde{a}$ i $a \cdot p \geq \tilde{a} \cdot \tilde{p}$, onda $Y \geq_L X$.

1984. godine Mekdonald [3] je dao Đinijev koeficijent GB1 raspodele kao jedan dugačak izraz koji uključuje uopštenu hipergeometrijsku funkciju 4F3. U nastavku teksta smo prikazali čemu je jednak ovaj koeficijet za B1 podfamiliju raspodela, jer je sam izraz manje glomazan:

$$G = \frac{2 \cdot B(p+q, 1/2) \cdot B(p + 1/2\,, 1/2)}{\pi \cdot \mathrm{B(q}, 1/2)} = \frac{\Gamma(p+q) \cdot \Gamma(p + 1/2) \cdot \Gamma(q + 1/2)}{\sqrt{\pi}\Gamma(p + q + 1/2) \cdot \Gamma(p+1) \cdot \Gamma(q)}.$$

Prva raspodela koja obuhvata i GB1 i GB2 familiju je predložena 1995. godine [4]. Ova raspodela se zove uopštena beta raspodela sa pet parametara i sledeća funkcija predstavlja njenu funkciju gustine:

$$f(x) = \frac{|a| \cdot x^{a \cdot p - 1} \cdot [1 + (1-c) \cdot (x/b)^a]^{q-1}}{b^{a \cdot p} \cdot B(p,q) \cdot [1 + c \cdot (x/b)^a]^{p+q}}, \tag{93}$$

gde je $0 < x^a < b^a/(1-c)$ i svi parametri a, b, c, p, q su pozitivni. Ponovo imamo da je b parametar skaliranja, dok su a, p, q parametri oblika. Novi parametar $c \in [0, 1]$ omogućava gladak prelaz između specijalnih slučajeva $c = 0$, GB1 raspodele i $c = 1$, GB2 raspodele. Momenti ove uopštene raspodele postoje za svako k ukoliko je $c < 1$ ili za $k < a \cdot q$ ako je $c = 1$. Momente smo prikazali sledećom formulom:

$$E(X^k) = \frac{b^k \cdot B(p + \frac{k}{a}, q)}{B(p,q)} {}_2\mathrm{F}_1\left(p + \frac{k}{a}, \frac{k}{a}; p + q + \frac{k}{a}; c]\right) \cdot$$

Za kraj ovog odeljka napominjemo da su i sami autori ovog uopštenja ostali skeptični u pogledu korisnosti i samog značenja ove raspodele, te je ostala kao interesantno teorijsko uopštenje.

4.2 Primena

U drugom odeljku poslednje teme osvrćemo se na primenu uopštene beta raspodele prve vrste slično kao i u prethodnim poglavljima, s tim da smo ovde prikazali i potencijalnu primenu u muzici.

Prihodi:

U uticajnom radu Turova iz 1970. godine [22] B1 raspodela je razmatrana kao raspodela prihoda domaćinstava (porodica i osoba koje nisu u srodstvu) u SAD-u za period 1949-1966, koristeći meru vrednosti dolara u odnosu na 1959. godinu. (Ocena parametara nije data u ovom radu, već u radu iz 1984. godine od strane Mekdonalda [3]). Turov je takođe proučavao uticaj različitih makroekonomskih faktora na parametre raspodele, putem regresionih tehnika. Konkretno, njegova istraživanja se bave pitanjem da li je ekonomski rast povezan sa egalitarizmom, a takođe i pretpostavlja da inflacija može dovesti do izjednačenije raspodele za belce. Međutim, 1984. godine Mekdonald [3] je izrazio određene sumnje u pogledu rezultata Turova. On je istakao da nijedna od ocenjenih gustina nije $\cap -$oblika (kao što se očekuje), a da se Đinijevi koeficijenti razlikuju od ocene za oko 30%, čime je zaključio da te razlike naglašavaju probleme ocene.

1979. godine je Mekdonald koristio B1 raspodelu za aproksimaciju porodičnih prihoda u SAD-u za 1960. godinu i od 1969. do 1975. godine [23]. Koristeći tri različite metode ocene, ispostavilo se da je ova raspodela pogodnije od gama i log-normalne raspodele, ali da je inferiornija od Sing-Madala raspodele, koji ima isti broj parametara.

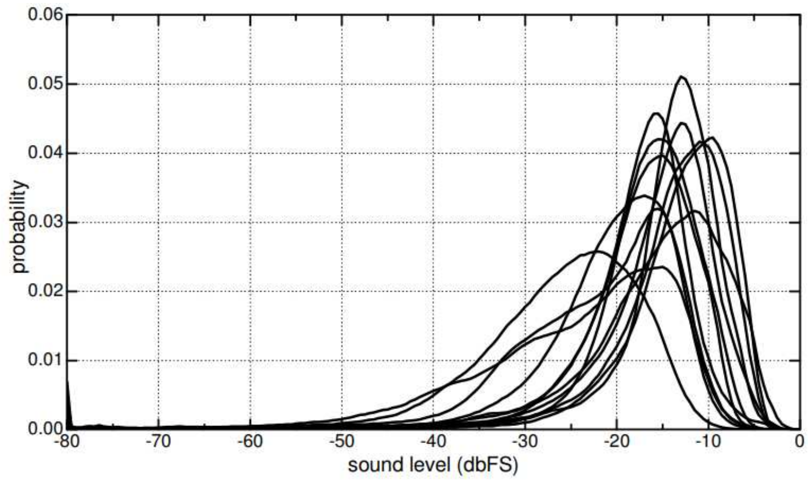
Do sličnog rezultata kao u prehodnom slučaju su 1988. godine došli japanski statističari, koristreći anketu iz 1975. godine [24]. Oni su koristili B1 raspodelu da bi fitovali podatke o prihodima u Japanu, u grupisanom obliku. Razmatrane su četiri radne klase, kao i primarni i preraspodeljeni prihodi, a primenjeno je čak pet različitih tehnika ocene. Pomenuta raspodela daje značajno bolji fit od uopštene gama i log-normalne raspodele, ali je i dalje Sing-Madala raspodela superiornija.

Nemački statističari su 1996. godine ocenjivali B1 i GB1 raspodelu, koristeći podatke o prihodima domaćinstava, prikupljenih preko nemačkog socio-ekonomskog panela (SOEP) za period 1984-1993 godine [20]. Oni su primetili da se ocena metodom maksimalne verodostojnosti za GB1 pokazala prilično teškom. Takođe, oba modela su imala tendenciju da potcenjuju srednju vrednost za ove skupove podataka.

Kao zaključak ovog odeljka, možemo reći da sve ove studije ukazuju na to da postoji nekoliko raspodela podržanih na neograničenom domenu koje pružaju znatno bolji fit od GB1 ili B1 raspodele.

Muzika:

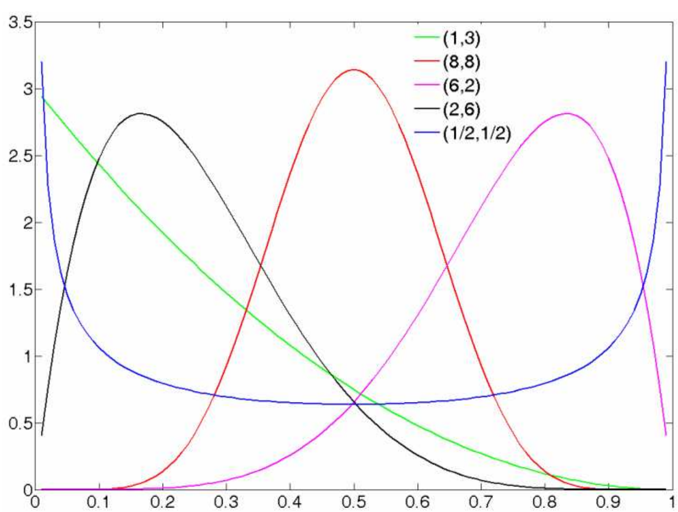
Poznavanje dinamičkih osobina audio signala ima posebno značajnu ulogu u mnogim inženjerskim oblastima koje se bave audiotehnikom. Istraživanje, sprovedeno na Elektrotehničkom fakultetu u Beogradu, pokazalo je jasno vidljivu razliku u obliku empirijskih raspodela intenziteta muzičkih signala između 12 različitih muzičkih žanrova [7]. Muzički žanrovi koji su bili predmet istraživanja su: klasična muzika, džez i bluz kao jedinstven žanr, rok, hard rok, pop, elektro pop, hip-hop i rep zajedno, haus, tehno, srpski folk, liturgijska i filmska muzika. Grafička predstava navedenih rezultata je prikazana na slici 4.



Slika 4: Usrednjene raspodele nivoa muzičkih signala za svih dvanaest žanrova.

Sa slike 4 se jasno vidi da pojedini žanrovi imaju vrlo razvučen dinamički opseg intenziteta signala, dok drugi teže da ga koncentrišu u relativno uskom opsegu. Pitanje koje se tada postavilo i koje je bilo predmet istraživanja rada je analiza mogućnosti da se empirijski dobijeni rezultati modeluju preko egzaktnih matematičkih raspodela. U ovom smislu, beta raspodela je tip raspodele koja je, po svojoj prirodi, fleksibilna, te stoga naročito pogodna za ovakav tip analize s obzirom da u zavisnosti od svojih parametara može uzimati vrlo različite oblike gustine raspodele.

Na slici 5, na kojoj su prikazani neki tipični oblici beta raspodela, ilustrovana je izuzetna prilagodljivost beta raspodele, samim tim i razlog zbog čega beta raspodela može biti jako korisna pri raznim simulacijama.



Slika 5: Izgled karakterističnih beta raspodela sa dva parametra - $B(p,q)$.

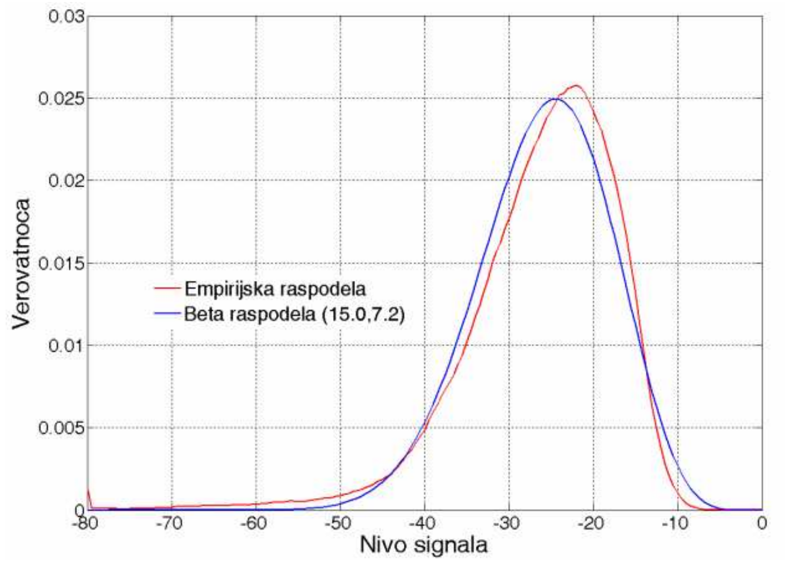
Kao što se sa slike 5 može videti, kada je $p > q$, beta raspodela uzima oblik (kriva označena roze linijom), koji već blago počinje da podseća na empirijsku raspodelu nivoa muzičkih signala prikazanih na slici 4.

Podaci koji su obrađivani u tom istraživanju predstavljaju podatke koji su dobijeni merenjem intenziteta različitih muzičkih signala (žanrova) pri čemu je trajanje pojedinačnog žanra bilo oko 2 sata. Izmereni intenziteti muzičkih signala se kreću u opsegu od -80dB do 0dB sa korakom od 0.5dB i za svaku vrednost intenziteta je izmerena pripadajuća verovatnoća pojavljivanja. Na osnovu tih vrednosti nacrtani su grafici prikazani na slici 4. Bitno je napomenuti da navedeni grafici sa slike 4 ne predstavljaju klasično matematički definisanu gustinu raspodele 12 muzičkih žanrova, već empirijsku raspodelu nivoa signala tih žanrova, koje su definisane po tačno određenim vrednostima (intervalima) nivoa muzičkih signala.

Za testiranje hipoteze o tome da li se pomoću neke beta raspodele sa fiksiranim parametrima može ili ne može modelirati određeni muzički žanr korišćen je hi kvadrat test, kao jedan od testova te vrste sa najširom oblašću primene. U toku istraživanja testirani su podaci vezani za svih 12 žanrova, na beta raspodele čiji se parametri kreću u opsegu od 1 do 40. Potencijalne, neodbačene, beta raspodele koncentrisale su se u tačno određenim intervalima parametara, koji određuju te raspodele. Navedeni rezultati su prikazani u tabeli 2.

Tabela 2: Granične vrednosti intervala parametara beta raspodela koje su prošle hi kvadrat test.

Tip muzičkog žanra	Početne vrednosti parametara p i q		Krajnje vrednosti parametara p i q	
Klasična muzika	13.0	6.2	16.5	7.9
Džez i bluz	20.6	6.9	28.0	9.3
Hard rok	31.7	6.8	38.5	8.2
Pop	25.0	6.9	32.1	8.8
Rok	25.9	7.6	33.4	9.8
Elektro pop	22.5	5.5	25.0	6.1
Hip-hop i rep	12.0	2.9	17.5	4.1
Haus	20.5	4.3	25.6	5.3
Tehno	22	4.3	29.8	5.8
Narodna	27.1	7.0	40.0	10.4
Filmska	7.9	3.4	10.1	4.2
Liturgijska	11.0	4.1	16.3	6.1



Slika 6: Ilustracija modelovanja empirijske krive raspodele klasične muzike beta raspodelom (15.0,7.2).

Zaključak

Ovaj rad predstavlja uvod u beta-tip raspodela i njenu primenu. On je baziran na već ustanovljenim definicijama, kao i svojstvima, koja čine neophodnu osnovu za dalje upoznavanje i razumevanje ove oblasti, koja je predmet izučavanja mnogih naučnika. Sama poenta rada je bila da se upoznamo sa uopštenjem beta raspodele korišćene tokom osnovnih studija, kao i da približimo njen značaj kroz primenu.

U prvom poglavlju smo se zadržali na definiciji uopštene beta raspodele druge vrste. Najveći deo primene smo zaključili da se odnosi na modelovanje prihoda, kao i gubitaka u aktuarstvu.

U drugom poglavlju smo definisali specijalni slučaj raspodele iz prvog poglavlja, tačnije Sing-Madala raspodele. Po uzoru na prvo poglavlje i ovo smo razradili kroz 6 odeljaka, zajedno sa definicijom Lorencove krive kao i ocenom parametara, između ostalog.

Treće poglavlje se odnosi na dodatna dva specijalna slučaja uopštene beta raspodele druge vrste, tačnije Fiks i Lomaks raspodelu.

Poslednje poglavlje rada je bazirano na definiciji i primeni uopštene beta raspodele prve vrste. U okviru ovog odeljka spomenuli smo i nama dobro poznatu standardnu beta raspodelu, koju smo izučavali na osnovnim studijama. U okviru primera u ovom odeljku izašli smo iz okvira ekononomije i aktuarstva i predstavili primenu u muzici.

Literatura

[1] Statistical Size Distributions in Economics and Actuarial Sciences, Wiley Series in probability and statistics; Kleiber C., Kotz S. (2003);

[2] Life Distributions, Structure of Nonparametric, Semiparametric and Parametric Families; Marshall A.W., Olkin I. (2007);

[3] Some generalized functions for the size distribution of income; Econometrica, 52, 647–663; McDonald, J. B. (1984);

[4] A generalization of the beta distribution with applications, Journal of Econometrics; McDonald, J. B., and Xu, Y. J. (1995);

[5] Hazard rates and generalized beta distributions, IEEE Transactions on Reliability; McDonald, J. B., and Richards, D. O. (1987);

[6] Comparing coefficients of variation in income distributions; Journal of the Indian Society of Agricultural Statistics; Singh, M., and Singh, R. P. (1992);

[7] Analiza mogućnosti modelovanja raspodela nivoa muzičkih signala različitih žanrova preko beta raspodela; 17. Telekomunikacioni forum TELFOR 2009; Srbija, Beograd, novembar 24.-26., 2009; 1001-1004; Ognjen Radović, Milan Merkle (2009);

[8] Applications of the GB2 family of distributions in modeling insurance loss processes. Insurance: Mathematics & Economics, 9, 257–272; Cummins, J. D., Dionne, G., McDonald, J. B., and Pritchett, B. M. (1990);

[9] Transformed beta and gamma distributions and aggregate losses; Proceedings of the Casualty Actuarial Society, 70, 156–193; Venter, G. (1983);

[10] Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. Water Resources Research, 10, 223–226; Mielke, P. W., and Johnson, E. S (1974);

[11] The Statistical Analysis of Failure Time Data, New York: John Wiley; Kalbfleisch, J. D., and Prentice, R. L. (1980);

[12] Pareto Distributions. Fairland, MD: International Co-operative Publishing House; Arnold, B. C. (1983);

[13] A maximum relative entropy principle for distribution of personal income with derivations of several known income distributions. Communications in Statistics—Theory and Methods, 19, 1003–1036; Leipnik, R. B. (1990);

[14] The distribution of self-employment income in the United Kingdom, 1976–1991. Economic Journal, 107, 455–466; Parker, S. C. (1997);

[15] A general definition of the Lorenz curve. Econometrica, 39, 1037–1039; Gastwirth, J. L. (1971);

[16] Description of the income distribution by the scaled F distribution model. In: N. A. Klevmarken and J. A. Lybeck (eds.): The Statics and Dynamics of Income. Clevedon, UK: Tieto, pp. 23–36; Vartia, P. L. I., and Vartia, Y. O. (1980);

[17] Using incomplete moments to measure inequality. Journal of Econometrics, 42, 109–119. Butler, R. J., and McDonald, J. B. (1989);

[18] Ethical Social Index Numbers. Berlin and New York: Springer. Chakravarty, S. R. (1990);

[19] Estimating aggregate automotive income elasticities from the population income-share elasticity. Journal of Business and Economic Statistics, 11, 209–214. Bordley, R. F., and McDonald, J. B. (1993);

[20] Evaluating parametric income distribution models. Allgemeines Statistisches Archiv, 80, 285–298. Brachmann, K., Stich, A., and Trede, M. (1996);

[21] A sufficient condition for Lorenz ordering. Sankhya , B58, 62–69. Wilfling, B. (1996a);

[22] Analyzing the American income distribution. American Economic Review (Papers and Proceedings), 48, 261–269. Thurow, L. (1970);

[23] Functional forms, estimation techniques, and the distribution of income. Econometrica, 47, 1513–1526. McDonald, J. B., and Ransom, M. (1979a);

[24] Statistical inference of functional form for income distribution. Economic Studies Quarterly, 39, 14–40. Atoda, N., Suruga, T., and Tachibanaki, T. (1988);

[25] On the estimation of long-tailed skewed distributions with actuarial applications. Journal of Econometrics, 23, 91–102. Hogg, R. V., and Klugman, S. A. (1983);

[26] A stochastic process for income distribution and tests for income distribution functions. ASA Proceedings of the Business and Economic Statistics Section, 551–553. Singh, S. K., and Maddala, G. S. (1975);

[27] A function for the size distribution of incomes. Econometrica, 44, 963–970. Singh, S. K., and Maddala, G. S. (1976);

[28] Cumulative frequency functions. Annals of Mathematical Statistics, 13, 215–232. Burr, I. W. (1942);

[29] A function for the size distribution of incomes: Comment. Econometrica, 46, 459–460. Cramer, J. S. (1978);

[30] A note on a characterization of the generalized log-logistic distribution. Environmetrics, 1, 337–342. El-Saidi, M. A., Singh, K. P., and Bartolucci, A. A. (1990);

[31] A characterization of the Burr type XII distribution. Applied Mathematics Letters, 4, 59–61. Al-Hussaini, E. K. (1991);

[32] Some remarks on a characterization of the generalized log-logistic distribution. Environmetrics, 7, 277–281. Ghitany, M. E; 1996;

[33] The first-order stochastic dominance ordering of the Singh–Maddala distribution. Economics Letters, 69, 123–128. Klonner, S. (2000);

[34] Explicit estimators for income distributions. In: C. Dagum and A. Lemmi (eds.): Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty. Greenwich, CT: JAI Press, pp. 393–405. Stoppa, G. (1995);

[35] Some sampling properties of a model for income distribution. Journal of Business and Economic Statistics, 1, 147–153. Schmittlein, D. C. (1983);

[36] On maximum product of spacings (MPS) estimation for Burr XII distributions. Communications in Statistics—Simulation, 22, 615–641. Shah, A., and Gokhale, D. V. (1993);

[37] Income distribution models. In: S. Kotz, N. L. Johnson, and C. Read (eds.): Encyclopedia of Statistical Sciences, Vol. 4. New York: John Wiley, pp. 27–34. Dagum, C. (1983);

[38] Functional forms of income distributions: The case of yearly income groups in the "Annual Report on the Family Income and Expenditure Survey" (in Japanese). Economic Studies Quarterly, 33, 79–85. Suruga, T. (1982);

[39] Estimations of income distribution parameters for individual observations by maximum likelihood method. Journal of the Japan Statistical Society, 27, 191–203. Tachibanaki, T., Suruga, T., and Atoda, N. (1997);

[40] On the Lorenz order within parametric families of income distributions. Sankhya , B61, 514–517. Kleiber, C. (1999a);

[41] Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data. Canadian Journal of Statistics, 16, 223–236. Shoukri, M. M., Mian, I. U. H., and Tracy, D. S. (1988);

[42] Exact confidence interval for the shape parameter of a log-logistic distribution. Journal of Statistical Computation and Simulation, 56, 193–211. Chen, Z. (1997);

[43] Business failures: Another example of the analysis of failure data. Journal of the American Statistical Association, 49, 847–852. Lomax, K. S. (1954).

Kratka biografija autora

Jelena Ljuboja je rođena 29. juna 1993. godine u Banja Luci, Bosna i Hercegovina. Osnovnu školu „Desanka Maksimović“ završila je u Beogradu, a zatim Osmu beogradsku gimnaziju, prirodno-matematički smer.

Matematički fakultet Univerziteta u Beogradu upisuje 2011. godine, smer: Statistika, aktuarska i finansijska matematika, a diplomira 22. septembra 2015. godine.

Od 1. oktobra 2015. godine do 23. oktobra 2017. godine bila je zaposlena u firmi PwC Srbija, članici multinacionalne mreže firmi PricewaterhouseCoopers, poznatijoj kao PwC, na poziciji saradnika za obezbeđenje od rizika u oblasti informacionih tehnologija. U tom periodu radi i na polju analize velikog seta podataka (tzv. big data), na osmišljavanju i izradi kodova pomoću kojih se podaci mogu transformisati i izvući tražene informacije. Deo rada u pomenutoj firmi je provela na stručnom usavršavanju u Hrvatskoj.

Od 23. oktobra 2017. godine zaposlena je u telekomunikacionoj kompaniji Vip mobile, članici A1 Telekom Austria Grupe, na poziciji internog revizora u oblasti informacionih tehnologija. Pomenuta pozicija je klaster, te pokriva Srbiju i Sloveniju (članicu grupe pod nazivom A1 Slovenija).