# Distribucije teškog repa i primjena u reosiguranju

**Mandić, Dominik**

**SVEUČILIŠTE U ZAGREBU**

**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**

**MATEMATIČKI ODSJEK**

Dominik Mandić

# DISTRIBUCIJE TEŠKOG REPA I PRIMJENA U REOSIGURANJU

Diplomski rad

Voditelj rada:
prof.dr.sc.Zoran Vondraček

Zagreb, Srpanj, 2022

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

# Sadržaj

# Uvod

Činjenice poput te da 1% najbogatijih ljudi na svijetu čini ukupno 27% ukupnog bogatstva, ili pak da je na popularnoj društvenoj mreži Twitter čak 60% svih objava došlo od samo 2% najpopularnijih korisnika, na prvu prkose intuiciji kada na pojave promatramo kroz prizmu dobro nam poznate normalne distribucije. Ona je određena sa svoja 2 parametra, koja opisuju oko koje vrijednosti se njene realizacije najviše grupiraju i koliko oko nje variraju. Tako se formira klasična krivulja u obliku zvona, čiji repovi teže prema nuli kako se udaljavamo od očekivanja. Naprimjer, ako promatramo visinu ili težinu ljudi razumna je pretpostavka da su one otprilike tako distribuirane te velika većina ljudi nije viša od 2 metra - što je 20-ak cm više od svjetskog prosjeka.
To očito nije pojava u primjeru s početka, pa nas takvo razmišljanje može navesti na to da se radi o kakvoj anomaliji. Također, zbog ključne uloge centralnog graničnog teorema u teoriji vjerojatnosti i statistici, imamo dojam da bi svaka pojava trebala imati takvo ponašanje. Usprkos tome, to nipošto nije istina. Takve situacije nisu ni nove ni rijetke te se takvim pojavama već u prošlom stoljeću bavio talijanski ekonomist Vilfredo Pareto, konkretno promatrajući takozvani 80-20 problem. On se pojavljuje u brojnim aspektima industrije i poslovanja, od ekonomije pa do fizike i informacijskih znanosti. Činjenice koje se temelje na tom principu sugeriraju kako posrijedi nije normalna distribucija, već je pravi razlog možda neočekivanih podataka neka *distribucija teškog repa*, od kojih je jedna i dobila ime po Paretu, te će biti detaljnije opisana u nastavku rada. Zbog svoje strukture i svojstava, distribucije teškog repa često su razlog zabuna, ali i zanimljivih zapažanja koja su otkrivena u raznim poljima. To je i razlog zbog čega su na neki način tretirane kao nejasne, ali i kontroverzne, jer standardan statistički alat može dovesti do zabuna kada su one u pitanju.
Prirodno pitanje koje se postavlja u kontekstu distribucija teškog repa je što to zapravo znači, što su uopće distribucije teškog repa te kako težinu repa kvantificirati. Cilj rada biti će dati odgovor na ta pitanja, razjasniti ulogu takvih distribucija kao i navesti njene osnovne primjere te svojstva. Što se njhovih primjena tiče, fokus će biti na njihove primjene u reosiguranju, koje će također biti polje interesa ovoga rada. Da bismo to mogli, prvo treba uopće razjasniti taj pojam, kao i samu motivaciju kako osiguravajućih kuća prvog reda, tako i reosiguravajućih kuća. Prvo poglavlje bavit će se reosiguranjem kao takvim te

njegovim osnovnim oblicima dok će drugi dio biti rezerviran za opisivanje modela koji se koriste kod predviđanja velikih šteta u reosiguranju i počivaju na gore spomenutim distribucijama.

# Poglavlje 1

# Reosiguranje

## 1.1 Motivacija i osnovni pojmovi

Pitanje na koje je bitno odgovoriti je što je zapravo reosiguranje i kako se razlikuje od običnog osiguranja. *Reosiguranje* je sporazum u kojem se jedna strana (reosiguravatelj) obvezuje osigurati drugu stranu (cedenta) za unaprijed dogovorene dijelove svog preuzetog rizika osiguranja kojeg pokriva. Zauzvrat, cedent plaća takozvanu fiksnu premiju reosiguranja, koja ovisi o njihovom ugovoru. Budući da upravo početni osiguravatelj postaje onaj koji razmišlja o prevenciji i ublaženju potencijalnog gubitka imovine u budućnosti, koncept osiguranja je pomaknut jedan nivo iznad. Dok ovakvo gledanje na to sugerira kako su dva koncepta vrlo slična, tj. da je reosiguranje samo posebna vrsta osiguranja, činjenica je da postoje vidne razlike kako u veličini preuzetog rizika, tako i u količini podataka dostupnih za analizu.
Ugovori o reosiguranju mogu biti obvezni, koji se primjenjuju na sve rizike određene vrste (npr. požar, potres...), ili fakultativni. Fakultativni ugovori su koncipirani tako da reosiguravatelj može odbiti preuzeti određeni rizik, kao što i osigurani član ima pravo odstupiti. Glavne mane fakultativnih ugovora su veliko uloženo vrijeme u preuzimanje rizika i postojanje neizvjesnosti hoće li osiguratelj dobiti reosigurateljno pokriće. Dakako, prednost ovakvih ugovora je prvenstveno u fleksibilnosti da se metoda prilagođava svakom pojedinom slučaju preuzimanja rizika. Ipak, obavezni ugovori u praksi su dosta češći od fakultativnih. Postoji još i takozvano fakultativno-obligatorno reosiguranje, koje se naziva još i otvoreno pokriće (engl. opencover). Tim ugovorom osiguravatelj ima pravo izbora prijenosa rizika na reosiguratelja, a reosiguratelj je obvezan preuzeti rizike. Drugim riječima, ovaj oblik reosiguranje je opcionalan za osiguratelja, ali obligatoran za reosiguratelja koji je izložen riziku.
Kada pričamo o reosiguranju, izravna veza između reosiguravatelja i pojedinačnog ugovaratelja osiguranja ne postoji, a razlozi zbog čega je takav odnos u interesu svima su

brojni. Većinu njih možemo svesti na jednu poprilično generalniju činjenicu, a to je da se time izbjegava mogućnost prevelikog gubitka kapitala odjednom, inducirana "nasumičnom" štetom, tj disperzija rizika. Zato je potencijalni gubitak lakše zamijeniti determinističkom premijom koju osiguravatelj plaća unaprijed. Svakako je sigurno reći da reosiguranje omogućava osiguravateljima širenje svog poslovanja, što čini osiguranje dostupnijim i pristupačnijim, samim time potičući gospodarski rast i učinkovitost tržišta. Za bolje razumijevanje treba formalizirati pojmove, a zatim i navesti oblike reosiguranja, njihovu motivaciju te značajke.

**Definicija 1.1.1.** *Neka je* $(X_i\colon\ i \in \mathbb{N})$ *niz slučajnih varijabli koji predstavlja veličinu odštetnih zathjeva osiguravatelja prvog reda te neka je* $(N(t)\colon\ t \geq 0)$ *brojeći proces, gdje* $N(t)$ *predstavlja broj odštetnih zathjeva do vremenskog trenutka* $t > 0$ *(mjereno u godinama). Tada ukupni ili agregirani iznos štete u trenutku* $t$ *za osiguravatelja prvog reda iznosi*

$$S(t) = \begin{cases} \sum_{i=1}^{N(t)} X_i & \text{ako } N(t) > 0, \\ 0 & \text{ako } N(t) = 0 \end{cases}$$

Većina ne-životnih osiguranja odnosi se na trajanje od jedne godine, pa će nam od glavnog interesa biti slučajna suma $S(1)$.
U ugovorima, ukupni iznos štete podijeljen je na dva dijela,

$$S(t) = R(t) + D(t)$$

gdje je $D(t)$ iznos koji ostaje cedentu na pokriće, dok je $R(t)$ iznos kojeg plaća reosiguravatelj. Najčešće su ugovori reosiguranja definirani po individualnim rizicima $X_i$, pa pišemo $X_i = D_i + R_i$ (ili samo $X = R + D$ u slučaju da su jednako distribuirani). Kao što je već spomenuto, postoje različiti oblici reosiguranja, a osnovna podjela je na one s proporcionalnim dijeljenjem rizika i one s neproporcionalnim.

## 1.2 Vrste reosiguranja i njihova svojstva

### 1.2.1. Kvotno reosiguranje

Najosnovniji oblik je kvotno reosiguranje (engl. *Quota-share reinsurance - QS*), koje spada pod reosiguranje s potpuno proprocionalnim dijeljenjem rizika. Vrijedi

$$R = a \cdot X$$

$$R(t) = a \cdot S(t)$$

za *faktor proporcionalnosti* $0 < a < 1$. Očito je da je ovaj oblik vrlo jednostavan i praktičan, a osnovna ideja je da cedent ustupa reosiguravatelju određeni udio premije. Za funkciju distribucije $F_X$ slučajnih varijabli $X_i$ vrijedi

$$\mathbb{P}(R \leq x) = F_X(\frac{x}{a}), \ \ \mathbb{P}(D \leq x) = F_X(\frac{x}{1-a})$$

Za ukupni rizik vrijedi analogno,

$$\mathbb{P}(R(t) \leq x) = \mathbb{P}(S(t) \leq \frac{x}{a}), \ \ \mathbb{P}(D(t) \leq x) = \mathbb{P}(S(t) \leq \frac{x}{1-a}),$$

Što se $r$-tog momenta tiče, $r \in \mathbb{N}$, vrijedi

$$\mathbb{E}(R^r) = a^r\mathbb{E}(X^r), \ \ \mathbb{E}(R^r) = (1-a)^r\mathbb{E}(X^r),$$

a možemo i izvesti da su koeficijent varijacije i asimetrija (*skewness*) jednaki. Promotrimo vjerojatnost da je u trenutku $t > 0$ određeni kapital $v$ zajedno s dosad primljenim premijama $P(t)$ bio dovoljan za pokriti štetu $S(t)$,

$$\mathbb{P}(v + P(t) - S(t) > 0).$$

Nakon ulaska u *QS sporazum* s faktorom proporcionalnosti $a$ ta vjerojatnost se mijenja u

$$\mathbb{P}(v + (1-a)P(t) - (1-a)S(t) > 0) = \mathbb{P}(\frac{v}{1-a} + P(t) - S(t) > 0)$$

To sugerira da na *QS* reosiguranje možemo gledati kao na svojevrsno povećanje početnog kapitala. Jedna od mana ovog pristupa je ta što su sve štete osigurane jednakim udjelom, a ne samo one najveće - što često nije u primarnom interesu cedenta, uzimajući u obzir da one manje gubitke ionako može podnijeti i sam.

## 1.2.2 Reosiguranje viška svote

Ovaj oblik proporcionalnog reosiguranja poboljšava glavne nedostatke QS oblika,a idalje zadržava njegove glavne pozitivne strane.
Neka je $Q_i$ osigurani iznos štete $X_i$. Za fiksni *samopridržaj M* dano je

$$R_i = (1 - \frac{M}{Q_i})X_i \cdot 1_{\{Q_i > M\}},$$

$$D_i = X_i \cdot 1_{\{Q_i \leq M\}} + \frac{M}{Q_i}X_i \cdot 1_{\{Q_i > M\}},$$

pa onda i

$$R(t) = \sum_{i=1}^{N(t)} R_i, \quad D(t) = \sum_{i=1}^{N(t)} D_i$$

Drugim riječima, cedent homogenizira rizik i nema velikih fluktuacija šteta jer zadržava u svom portfelju samo dio rizika koji odgovara visini samopridržaja.
S $V_i = \frac{X_i}{Q_i}$ označavamo *stupanj gubitka* odštetnog zahtjeva $X_i$. U ovom obliku reosiguranja svaki odštetni zahtjev s osiguranim iznosom manjim od $M$ je u potpunosti pod pokrićem osiguravatelja prvog reda. Iz samog oblika reosiguranja na prvu vidimo da ono idalje zadržava svojstvo proporcionalnosti pa je i dogovor o premijama jasan. Na ovakav oblik reosiguranja cedent se najčešće odlučuje da bi u svakom scenariju mogao pokriti očekivani dio kojeg bi morao isplatiti. Procjena osiguranog iznosa štete može se vršiti na temelju takozvanog *PML-a*( probable maximum loss). PML je moguće opisati kao očekivani pretrpljen iznos štete u jednom događaju, pa je razumno da se raspodjela rizika na temelju PML-a vrši u onim vrstama osiguranja kod kojih ne postoji mogućnost nastupa totalne štete. Nužno je da PML bude procijenjen što realnije, a to je vrlo složen posao u praksi. Ovaj oblik najviše se koristi u osiguranju od požara, instalacijskih, električnih ili pomorskih nesreća. Mana mu je kompleksna administracija i visoki administrativni troškovi.

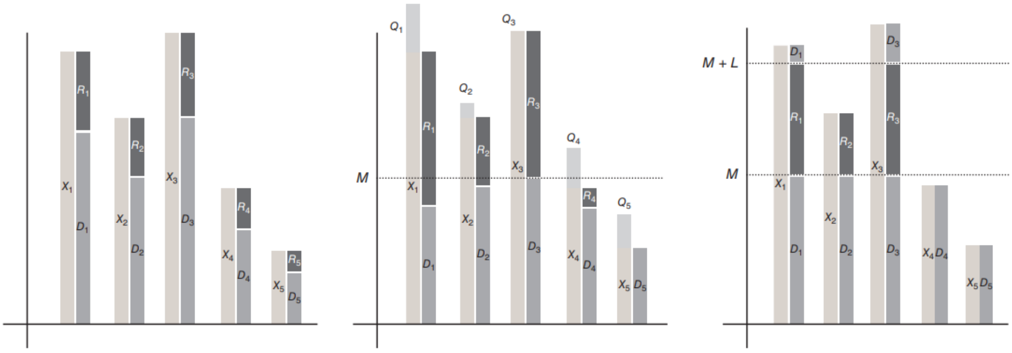
### 1.2.3 Reosiguranje viška šteta

Bitno je spomenuti i neke neproporcionalne oblike reosiguranja, čiji je najosnovniji oblik reosiguranje viška štete (*engleski excess-of-loss* - *XL*). Definiran je s

$$R(t) = \sum_{i=1}^{N(t)} (X_i - M)_+, \quad D(t) = \sum_{i=1}^{N(t)} min(X_i, M),$$

pri čemu $X_+$ označava samo nenegativni dio, a $M$ je fiksno zadan samopridržaj. Iz definicije slijedi da u ovom slučaju reosiguravatelj pokriva sve što prekorači fiksni $M$, a najčešće to vrijedi samo do određene granice $L$. Ugovor s reosiguranjem opisanim na ovaj način kraće se naziva '*X xs L*' ugovor i okarakterizian je *slojem* [M, L+M], dok omjer $\frac{M+L}{M}$ nazivamo *relativnom duljinom sloja*. Na Slici 1.1 je usporedba do sada opisana tri oblika reosiguranja. Što se analitičkih svojstava ovakvog oblika reosiguranja tiče, promotrimo najprije slučajnu varijablu definiranu s $\tilde{X} := \min((X - u)_+, v)$, za $u, v \geq 0$. Ako je $u = M$, $\tilde{X}$ predstavlja $R$, reosigurani dio štete u '*v xs M*' ugovoru. S druge strane, za $u = 0, v = M, \tilde{X}$ se odnosi na dio $D = \min\{X, M\}$ u '$\infty$ *xs M*' ugovoru, pa je intuitivno i svima od interesa proučavati njenu funkciju distirbucije. Ako je $F_X$ funkcija distribucije slučajne varijable $X$, onda vrijedi

$$F_{\tilde{X}}(z) = \begin{cases} F_X(u+z) & ako\ 0 \leq z \leq v, \\ 1 & ako\ v \leq z \end{cases}$$




Slika 1.1: *Reinsurance: Actuarial and Statistical Aspects - QS lijevo(a = 0.3), reosiguranje viška u sredini, reosiguranje viška gubitka desno*

Što se tiče njenih $k$-tih momenata, imamo izraz

$$\tilde{\mu}_k = \mathbb{E}(\tilde{X}^k) = k\int_0^\infty (1 - F_{\tilde{X}}(z))\ z^{k-1}\,dz = k\int_0^\nu (1 - F_X(u+z))\ z^{k-1}\,dz$$

odakle lako iščitavamo izraze za očekivanja

$$\mathbb{E}(D) = \int_0^M (1 - F_X(z))\,dz,\ \ \mathbb{E}(R) = \int_M^\infty (1 - F_X(z))\,dz$$

dijelova koje pokrivaju cedent i reosiguravatelj u '$\infty$ *xs M*' ugovoru. Na Slici 1.2 je i grafički prikaz odnosa očekivanja ta dva iznosa u ovisnosti o $M$. U slučaju '$L$ *xs* $M$' ugovora vrijedi analogno

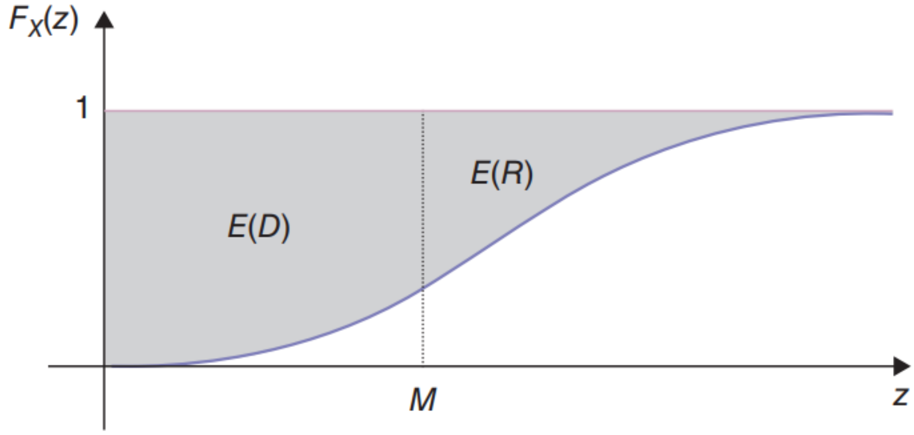
$$\mathbb{E}(R) = \int_M^{M+L} (1 - F_X(z))\,dz$$

## 1.2.4 Reosiguranje tehničkog rezultata

Reosiguranje tehničkog rezultata (*engl. stop-loss SL.*) je reosiguranje oblika

$$R(t) = (\sum_{i=1}^{N(t)} X_i - C)_+$$

i ono ograničava da ukupni gubitak osiguravatelja prvog reda pređe preko samopridržaja $C$ tako što obvezuje reosiguravatelja da pokrije višak. Iako ovaj oblik štiti osiguravatelja prvog reda od jako velikih rizika, neke probleme kao što su *problem moralne opasnosti*




Slika 1.2: Očekivanja iznosa kojeg plaća osiguravatelj i kojeg plaća reosiguravatelj u '$\infty$ *xs M'* ugovoru

dodatno naglašava, pa se često u praksi određuje i gornja granica $\mathcal{U}$. Kada je kod *XL* teško razlučiti indivudalne štete konkretno je koristan ovaj oblik, jer nije nužno moći procijeniti vjerojatnost svakog događaja zasebno. Unatoč tome, ovaj oblik ima svoje mane, s obzirom da je jako teško sastaviti ugovor koji obuhvaća sve moguće štete i štiti obje strane. Dodatno, uzima u obzir i manje štete za koje i nije cilj da nužno budu pokrivene pa je ovaj oblik rijedak i ne toliko učinkovit (osim među na neki način povezanim ustanovima ili društvima).

### 1.2.5 Reosiguranje velikih šteta

Neka su štete poredane po veličini na sljedeći način

$$X_{1,N(t)} \leq X_{2,N(t)} \leq ... \leq X_{N(t),N(t)}$$

Reosiguranje velikih šteta bazira se na tome da se reosiguravatelj obvezuje pokriti $r \geq 1$ najvećih šteta, tj

$$R(t) = \sum_{i=1}^{r} X_{N(t)-i+1,N(t)}$$

I ovaj oblik ima dodatne varijante i modifikacije, koje imaju svoje mane i prednosti, pa ovisno o situaciji biramo oblik koji je najprikladniji. Unatoč tome što komplicira analizu, moguće je i kombinirati oblike, a razlozi za to je najčešće neujednačenost i raznovrsnost mogućih šteti. Što se tiče fakultativnih ugovora, oni se najčešće pojavljuju u specifičnim situacijama kada je rizik individualne prirode sam po sebi i ne postoji dobar mehanizam za procjenu ukupnog rizika (npr gradnja nebodera, elektrana...) - a tada su automatski i premije veće nego kod standardnih oblika. Važno je još i dodati da iako je svaki oblik reosiguranja dizajniran za specifičnu situaciju, nijedan oblik ne može osiguratelju pružiti potpunu zaštitu od svih vrsta štete.

# Poglavlje 2

# Modeliranje veličine štete

U ovom poglavlju cilj je iznijeti koje modele koristimo za opisivanje kako individualnih, tako i ukupnih šteta, a veliku ulogu će odigrati distribucije spomenute na početku. Sada kada znamo što je reosiguranje i kakvi sve oblici postoje, najbitnije je dati odgovor na pitanje kako razlikovati male i velike štete, o čemu će naposlijetku u cijelosti i ovisiti ugovori između osiguravatelja prvog reda i reosiguravatelja.
S obzirom da su glavni razlog zašto koristiti opciju reosiguranja potencijalne upravo velike štete, vrlo je bitno definirati što u ovom kontekstu točno znači *veliko*. U svjetlu aktuarskih tema potrebno je naći efektivan način za razlikovati prosječne od velikih, a intuitivna misao vodilja je usporedba distribucija štete s eksponencijalnom distribucijom. Time bi eksponencijalna distribucija bila distribucija kojom ugrubo rečeno dijelimo one iznad i one ispod nje. Jasno je da kako god definirali velike štete, distribucije kojima ih opisujemo moraju biti sličnije onim s početka nego normalnoj. Takav način razmišljanja nas dovodi do više različitih klasa distribucija od kojih svaka ima zanimljiva svojstva i igra svoju ulogu u teoriji distribucija teškog repa.

## 2.1 Regularno varirajuće funkcije

**Definicija 2.1.1.** *Za funkciju $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ kažemo da je regularno varirajuća s indeksom $\rho \in \mathbb{R}$ ako vrijedi*

$$\lim_{x\to+\infty} \frac{f(yx)}{f(x)} = y^{\rho}$$

*za sve $y > 0$. Nadalje, za $p \leq 0$, distribucija $F$ je regularno varirajuća distribucija s indeksom $\rho$ ako je $\bar{F}(x) = 1 - F(x)$ regularno varirajuća funkcija s indeksom $\rho$. Slučajna varijabla $X$ je regularno varirajuća ako joj je pripadna distribucija regularno varirajuća.*

**Definicija 2.1.2.** *Funkcija distribucije F je invarijantna na skaliranje ako postoji $x_0 > 0$ i neprekidna, pozitivna funkcija g takva da*

$$\bar{F}(\lambda x) = g(\lambda)\bar{F}(x)$$

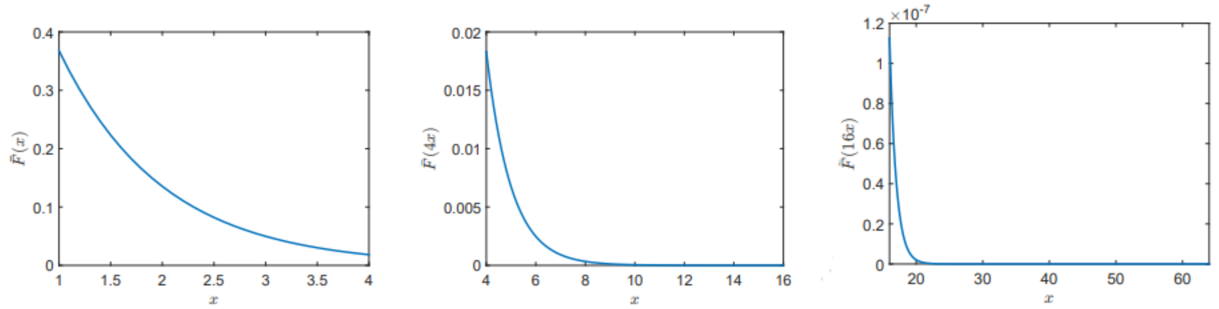
*za sve x, $\lambda$ za koje $\lambda x \geq x_0$*

Interpretacija ove definicije bi bila promatranje parametra $\lambda$ kao promjenu skale jedinica koje koristimo. Distribucije $F$ invarijantne na skaliranje imaju svojstvo da oblik $\bar{F}$ ostaje nepromijenjen do na multiplikativni faktor $g(\lambda)$ ako mjerne jedinice skaliramo s $\lambda$. Ovo svojstvo, unatoč elegantnoj definiciji, vrlo je osjetljivo, i ne vrijedi za mnogo poznatih distribucija.

**Primjer 1.** *Očito, eksponencijalna distribucija (s parametrom $\mu$) nije invarijantna na skaliranje jer vrijedi*

$$\bar{F}(\lambda x) = e^{-\mu\lambda x} = \bar{F}(x) \cdot e^{-\mu(\lambda-1)x}$$

*pa vidimo da ne postoji funkcija $g(\lambda)$ iz definicije. Isto se i vidi i iz slike.*




Slika 2.1: Eksponencijalna distribucija (s parametrom $\mu = 1$) nije asimptotski invarijantna na skaliranje jer krivulja izgleda potpuno različito na više skala

Uvjet iz definicije je naizgled vrlo snažan, pa nam je od interesa i svojstvo koje od funkcije zahtjeva da se na ovaj način ponaša samo asimptotski.

**Definicija 2.1.3.** *Funkcija distribucije F je asimptotski invarijantna na skaliranje ako postoji neprekidna, pozitivna funkcija g takva da za sve $\lambda > 0$*

$$\lim_{x\to+\infty} \frac{\bar{F}(\lambda x)}{\bar{F}(x)} = g(\lambda)$$

Vidimo da se zahtjevi razlikuju jedino u tome što asimptotska invarijantnost na skaliranje zahtjeva takvo ponašanje samo pri $x \to +\infty$, tj. kraće zapisano $\bar{F}(\lambda x) \sim g(\lambda)\bar{F}(x), x \to +\infty$.

**Teorem 2.1.4.** *Funkcija distirbucije F je asimptotski invarijantna na skaliranje ako i samo ako je regularno varirajuća.*

*Dokaz.* Iz definicija trivijalno slijedi da su regularno varirajuće funkcije asimptoski invarijantne na skaliranje, pa dokazujemo samo obratni smjer. Svojstvo asimptotske invarijantnosti na skaliranje implicira

$$\lim_{z \to +\infty} \frac{\bar{F}(xyz)}{\bar{F}(z)} = g(xy),$$

a isti izraz možemo i zapisati kao

$$\frac{\bar{F}(xyz)}{\bar{F}(z)} = \frac{\bar{F}(xyz)}{\bar{F}(xz)} \frac{\bar{F}(xz)}{\bar{F}(z)}$$

Primijetimo kako $\frac{\bar{F}(xyz)}{\bar{F}(xz)} \to g(y)$ i $\frac{\bar{F}(xz)}{\bar{F}(z)} \to g(x)$ ako $z \to \infty$, pa onda i cijeli izraz teži u $g(x)g(y)$. Jedine neprekidne i pozitivne funkcije koje zadovoljavaju $g(xy) = g(x)g(y)$ su oblika $g(x) = x^{\theta}$ za $\theta \in \mathbb{R}$, pa postoji $\theta$ takav da $g(x) = x^{\theta}$. Po definiciji je tada $\bar{F}(x)$ regularno varirajuća funkcija pa je i $F$ regularno varirajuća funkcija distribucije. □

**Definicija 2.1.5.** *Funkcija $L\colon \mathbb{R}_+ \to \mathbb{R}_+$ je sporo varirajuća ako $\lim_{x\to\infty} \frac{L(xy)}{L(x)} = 1$ za sve $y > 0$.*

Idući teorem daje vezu između sporo i regularno varirajućih funkcija.

**Teorem 2.1.6.** *Funkcija $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ je regularno varirajuća funkcija s indeksom $\rho$ ako i samo ako je $f(x) = x^{\rho}L(x)$ za neku sporo varirajuću funkciju $L$.*

*Dokaz.* Pretpostavimo prvo da je f regularno varirajuća te definirajmo $L(x) = \frac{f(x)}{x^{\rho}}$. Da bismo dokazali teorem, dovoljno je pokazati da je $L$ sporo varirajuća funkcija.

$$\lim_{x\to\infty} \frac{L(xy)}{L(x)} = \lim_{x\to\infty} \frac{f(xy)}{f(x)} \cdot \frac{x^{\rho}}{xy^{\rho}} = 1$$

Za obratni smjer, trebamo pokazati da je f regularno varirajuća s indeksom $\rho$ ako je $f(x) = x^{\rho}L(x)$ za sporo varirajuću funkciju $L$. Za $y > 0$,

$$\lim_{x\to\infty} \frac{f(xy)}{f(x)} = \lim_{x\to\infty} \frac{L(xy)}{L(x)} \cdot \frac{(xy)^{\rho}}{x^{\rho}} = y^{\rho},$$

pa po definiciji slijedi da je f regularno varirajuća. □

Kao što će se kasnije pokazati, distribucijama kakvim se bavimo lako se računaju momenti, uvjetne vjerojatnosti i konvolucije jer se svaka operacija svodi na integriranje ili deriviranje polinoma, dok je račun kod normalne ili pak lognormalne distribucije kompliciraniji. Osnovni rezultat koji se bavi svojstvima regulrano varirajućih funkcija što se integriranja i deriviranja tiče je Karamatin teorem. Za početak ćemo, zbog elegantnosti, diskutirati samo o teoremu koji se tiče integracije. Kada bi funkcija bila oblika $f(t) = t^\rho$, imali bismo

$$\int_0^x f(t)\,dt = \frac{x^{\rho+1}}{\rho+1} = \frac{xf(x)}{\rho+1}, \quad \rho > -1$$

$$\int_x^\infty f(t)\,dt = \frac{x^{\rho+1}}{-(\rho+1)} = \frac{xf(x)}{-(\rho+1)}, \quad \rho < -1.$$

Takoreći, Karamatin teorem kaže da se integral regularno varirajuće funkcije, barem što se repa tiče, ponaša polinomijalno.

**Teorem 2.1.7** (Karamatin teorem)**.** *i) Za $\rho > -1$ i regularno varirajuću funkciju f s indeksom $\rho$ vrijedi $\int_0^x f(t)\,dt \sim \frac{xf(x)}{\rho+1}$*
*ii) Za $\rho < -1$ i regularno varirajuću funkciju f s indeksom $\rho$ vrijedi $\int_x^\infty f(t)\,dt \sim \frac{xf(x)}{-(\rho+1)}$*

Očekivano, regularno varirajuće funkcije ponašaju se asimptotski polinomijalno i s obzirom na derivacije. Točnije, ako $f(x) = x^\alpha$, tada $f'(x) = \alpha x^{\alpha-1}$, tj $\alpha f(x) = xf'(x)$. Idući teorem, poznat kao *teorem monotone gustoće* kaže da upravo ovakva vrsta veze vrijedi za nama zanimljive funkcije, uz zamjenu = s $\sim$, te neke manje tehničke dopune.

**Teorem 2.1.8** (Teorem monotone gustoće)**.** *Pretpostavimo da funkcija f ima derivaciju gotovo svugdje, da je derivacija integrabilna i da vrijedi $f(x) = f(0) + \int_0^x f'(t)dt$ za svaki x, to jest da je f apsolutno neprekidna. Ako je f regularno varirajuća funkcija s indeksom $\rho$ i ako je $f'$ monotona od neke točke $x_0$, tada vrijedi $xf'(x) \sim \rho f(x)$ za $x \to \infty$. Štoviše, ako je $\rho \neq 0$, tada je $|f'(x)|$ regularno varirajuća funkcija s indeksom $\rho - 1$.*

*Dokaz.* Radi jednostavnosti, pretpostavimo da je $f$ neopadajuća funkcija na $x \geq x_0$ (dokaz za nerastuću funkciju je sličan). Fiksirajući $a$ i $b$ takve da vrijedi $0 \leq a \leq b$ možemo zapisati

$$\int_{ax}^{bx} \frac{f'(t)}{f(x)} dt = \frac{f(bx) - f(ax)}{f(x)}$$

Za $x > \frac{x_0}{a}$, monotonost funkcije $f'$ implicira

$$\frac{f'(ax)x(b-a)}{f(x)} \leq \frac{f(bx) - f(ax)}{f(x)} \leq \frac{f'(bx)x(b-a)}{f(x)}.$$

Budući da znamo da je $f$ regularno varirajuća funkcija indeksa $\rho$, prva nejednakost je zapravo

$$\limsup_{x\to\infty} \frac{f'(ax)x}{f(x)} \leq \frac{b^\rho - a^\rho}{b-a}$$

Nadalje, puštajući $b \downarrow a$ s desne strane gornje nejednakosti i imajući na umu da je to zapravo derivacija funkcije $x^\rho$ u točki $x = a$ dobivamo

$$\limsup_{x\to\infty} \frac{f'(ax)x}{f(x)} \leq \rho a^{\rho-1}$$

Sličnim razmatranjem, korištenjem druge nejednakosti i puštanjem $a \uparrow b$ dobivamo

$$\liminf_{x\to\infty} \frac{f'(bx)x}{f(x)} \geq \rho b^{\rho-1}$$

Sada, ubrštavanjem $a = 1$ i $b = 1$ u gornje dvije nejednakosti zaključujemo $xf'(x) \sim \rho f(x)$. Dodatno, za $\rho \neq 0$ lako vidimo i da $f'(x) \sim \rho(\frac{f(x)}{x})$ povlači da je $|f'(x)|$ regularno varirajuća funkcija s indeksom $\rho - 1$. □

Jasno je da je ovaj teorem vrlo snažno i zanimljivo nam svojstvo regularno varirajućih funkcija. Vrijedi spomenuti još jedan bitan rezultat ove teorije kojim jako precizno možemo opisati regularno varirajuće funkcije, a to je *Karamatin teorem o reprezentaciji*.

**Teorem 2.1.9** (Karamatin teorem o reprezentaciji)**.** *Funkcija f je regularno varirajuća funkcija s indeksom $\rho$ ako i samo ako se može zapisati kao*

$$f(x) = c(x)exp\{\int_z^x \frac{\beta(t)}{t}dt\}, \quad \forall x \geq z$$

*za neki $z > 0$, gdje je $\lim_{x\to\infty} c(x) = c \in (0,\infty)$ i $\lim_{x\to\infty}\beta(x) = \rho$.*

Primijetimo da funkcija nije zapisana u obliku gdje je očito kako ima oblik potencije. No, ako tretiramo $\beta$ kao konstantu $\rho$, što ona u limesu i jest, tada je eksponent oblika $\rho\ln(x)$, pa vrijedi $e^{\rho\ln x} = x^\rho$.
Sada, koristeći prethodni teorem možemo izravno dokazati sljedeću lemu.

**Lema 2.1.10.** *Za sporo varirajuću funkciju $L : \mathbb{R}_+ \to \mathbb{R}_+$ vrijedi*

$$\lim_{x\to\infty} x^\rho L(x) = \begin{cases} 0 & \rho < 0 \\ \infty & \rho > 0. \end{cases}$$

Navodimo još neka važna tehnička svojstva ove klase, koja neće nužno vrijediti za sve ovdje bitne nam klase distribucija.

**Lema 2.1.11.** *Neka su X i Y međusobno nezavisne, regularno varirajuće slučajne varijable s indeksima $-\alpha_X$ i $-\alpha_Y$ redom. Tada vrijedi:*
*i) min $\{X, Y\}$ je regularno varirajuća slučajna varijabla s indeksom $-(\alpha_X + \alpha_Y)$*
*ii) max $\{X, Y\}$ je regularno varirajuća slučajna varijabla s indeksom $-min\{\alpha_X, \alpha_Y\}$*
*iii) $X + Y$ je regularno varirajuća slučajna varijabla s indeksom $-min\{\alpha_X, \alpha_Y\}$. Štoviše, vrijedi i $\mathbb{P}(X + Y > t) \sim \mathbb{P}(max\{X, Y\} > t)$*

Iako jednostavna i donekle intuitivna, ova svojstva pokazuju se vrlo moćna i korisna za dokazivanje značajnih rezultata. Dokazat ćemo sve tri tvrdnje, iako je tvrdnja iii) ključna za neke od rezultata prezentiranih u nastavku.

*Dokaz.* Koristimo se zapisom regularno varirajućih funkcija preko sporo varirajućih funkcija, kao u teoremu 2.1.2. Po njemu, postoje sporo varirajuće funkcije $L_X$ i $L_Y$ takve da $\mathbb{P}(X > t) = t^{-\alpha_X} L_X(t)$ i $\mathbb{P}(Y > t) = t^{-\alpha_Y} L_Y(t)$.
*i*) Onda, imamo:

$$\mathbb{P}(min\{X, Y\} > t) = \mathbb{P}(X > t)\mathbb{P}(Y > t) = t^{-(\alpha_X+\alpha_Y)} L_X(t)L_Y(t)$$

Iz definicije sporo varirajućih funkcija vidimo da je i njihov produkt sporo varirajuć, pa slijedi tvrdnja i).
*ii*) Budući da $\{max\{X, Y\} > t\} = \{X > t\} \cup \{Y > t\}$, pišemo

$$\mathbb{P}(\{max\{X, Y\} > t\}) = \mathbb{P}(X > t) + \mathbb{P}(Y > t) - \mathbb{P}(X > t)\mathbb{P}(Y > t).$$

Bez smanjenja općenitosti, pretpostavljamo da je ili $\alpha_X < \alpha_Y$, ili $\alpha_X = \alpha_Y$. U prvom slučaju iz gornje jednakosti slijedi $\mathbb{P}(\{max\{X, Y\} > t\}) \sim \mathbb{P}(X > t)$, odakle je $max\{X, Y\}$ regularno varirajuća s indeksom $\alpha_X$.
Ako je pak $\alpha_X = \alpha_Y$, imamo $\mathbb{P}(\{max\{X, Y\} > t\}) \sim \mathbb{P}(X > t) + \mathbb{P}(Y > t)$, tj $\mathbb{P}(\{max\{X, Y\} > t\}) \sim t^{-\alpha_X}(L_X(t) + L_Y(t))$. Opet, jer je i suma sporo varirajućih funkcija sporo varirajuća, zaključujemo ii).
*iii*) Prvi korak nam je odrediti gornju i donju među za vjerojatnost događaja $\{X + Y > t\}$. Najprije, primijetimo da događaj $\{X > t\} \cup \{Y > t\}$ povlači $\{X + Y > t\}$. To nam daje donju ogradu,

$$\mathbb{P}(X + Y > t) \geq \mathbb{P}(X > t) + \mathbb{P}(Y > t) - \mathbb{P}(X > t)\mathbb{P}(Y > t)$$

Fiksirajmo sada $\delta \in (0, 1/2)$.
Dokažimo prvo da događaj $\{X+Y > t\}$ implicira $\{X > (1-\delta)t\} \cup \{Y > (1-\delta)t\} \cup \{X > \delta t, Y > \delta t\}$. Koristeći obrat po kontrapoziciji, dokazujemo da $\{X \leq (1-\delta)t\} \cap \{Y \leq (1-\delta)t\} \cap \{\{X \leq \delta t\} \cup \{Y \leq \delta t\}\} \iff \{X + Y \leq t\}$. Možemo pretpostavit da vrijedi $\{X \leq \delta t\}$ (u suprotnom je dokaz analogan). Tada, uz činjenicu da je i $\{Y \leq (1 - \delta)t\}$, mora vrijediti i $\{X + Y \leq t\}$. Sada možemo dalje pisati

$$\mathbb{P}(X + Y > t) \leq \mathbb{P}(X > (1 - \delta)t) + \mathbb{P}(Y > (1 - \delta)t) + \mathbb{P}(X > \delta t)\mathbb{P}(Y > \delta t)$$

Sada ponovno razmatramo dva slučaja kao i u tvrdnji *ii*). Ako vrijedi $\alpha_X < \alpha_Y$, slijedi nam

$$\liminf_{t\to\infty} \frac{\mathbb{P}(X+Y>t)}{\mathbb{P}(X>t)} \geq 1.$$

Slično, slijedi i

$$\limsup_{t\to\infty} \frac{\mathbb{P}(X+Y>t)}{\mathbb{P}(X>t)} \leq \lim_{t\to\infty} \frac{\mathbb{P}(X>(1-\delta t))}{\mathbb{P}(X>t)} = (1-\delta)^{\alpha_X} \leq 1.$$

Iz gornje dvije nejednakosti zaključujemo da $\mathbb{P}(X+Y>t) \sim \mathbb{P}(X>t)$, $t \to \infty$. Iz toga slijedi da je $X+Y$ regularno varirajuća s indeksom $-\alpha_X$ te da vrijedi $\mathbb{P}(X+Y>t) \sim \mathbb{P}(max\{X,Y\}>t)$ (koristeći se dokazom *ii*).
Konačno, preostao je slučaj $\alpha_X = \alpha_Y$. Koristeći istu logiku kao i u prvom slučaju, može se pokazati

$$\mathbb{P}(X+Y>t) \sim \mathbb{P}(X>t) + \mathbb{P}(Y>t))$$

To naravno opet implicira da je $X+Y$ regularno varirajuća s indeksom $-\alpha_X$, te da $\mathbb{P}(X+Y>t) \sim \mathbb{P}(max\{X,Y\}>t)$. □

## 2.2 Subeksponencijalne distribucije

Od interesa nam je još jedna klasa distribucija. Najprije, važno je još jednom napomenuti kako je cilj rada predstaviti adekvatne modele za velike štete, a upravo to pitanje *veličine* je ono što nam se nameće. Neka su $X_1, X_2, \ldots, X_n$ sve štete koje se odnose na neku imovinu, neka je $S_n = \sum_{i=1}^{n} X_i$ zbroj ukupne štete te neka je $M_n = max\{X_1, X_2, \ldots, X_n\}$ najveća od njih. Jedna od mogućih interpretacija *velike štete* je sljedeća.
Prema primjeru iz početka, kada bismo promatrali zbroj visina nasumično odabranih 100 ljudi, ili pak zbroj objava na Twitteru istih, i kad bi nekim slučajem taj broj bio iznimno velik, različiti bi bili razlozi koji leže u pozadini tih pojava. U prvoj situaciji pretpostavili bismo kako su svi odabrani ljudi iznadprosječno visoki, dok bi se u drugoj situaciji nametalo postojanje osobe s iznimno velikim brojem objava. Prelaskom na kontekst reosiguranja i šteta, kada bismo znali da je broj smrti u npr. potresima velik, lako bismo zaključili da je postojao jedan jako smrtonosan potres, a ne jako puno manjih bez ikakvih posljedica. Navođeni time, možemo reći kako je šteta velika ako je zbroj svih šteta pretežno određen upravo njom. Matematička intepretacija ove pojave u literaturama još se naziva i princip katastrofe, a kaže

**Definicija 2.2.1** (Princip katastrofe)**.** *Za distribuciju $F$ kažemo da zadovoljava princip katastrofe ako vrijedi*

$$\mathbb{P}(S_n > t) \sim \mathbb{P}(M_n > t)$$

Ovo svojstvo je poprilično snažno, jer nam daje odgovor na pitanje radi čega točno je promatrana suma velika, a apriorni je razlog za to mogao biti bilokoji. Nalazimo ga u brojnim oblicima, a jedan od njih je i zapis preko uvjetnih vjerojatnosti.

$$\mathbb{P}(M_n > t \mid S_n > t) = \frac{\mathbb{P}(M_n > t, S_n > t)}{\mathbb{P}(S_n > t)} = \frac{\mathbb{P}(M_n > t)}{\mathbb{P}(S_n > t)} \to 1, \text{kako } t \to \infty.$$

Ovo naglašava činjenicu da ako vrijedi princip katastrofe, onda je vrlo vjerojatno da će maksimum od $n$ uzoraka biti veći od $t$ ako znamo da je zbroj uzoraka veći od $t$. Dakle, događaj da je suma neobično velika je, s vrlo velikom vjerojatnošću, rezultat upravo jednog velikog "katastrofalnog" događaja.
Naprimjer, već spomenute regularno varirajuće distribucije zadovoljavaju ovaj princip. Unatoč tome što ga velika klasa distribucija zadovoljava, ovaj princip neće vrijediti za baš sve distribucije spomenute u radu. Za razliku od principa katastrofe, *princip zavjere* usklađen je s našom intuicijom o tome kako se neočekivano veliki događaji ponašaju - kao kombinacija velikog broja čimbenika. To čini princip zavjere na neki način intuitivnijim od principa katastrofe za većinu ljudi. Međutim, manje je snažan jer u mnogo slučajeva ne pruža tako "jednostavno" objašnjenje za rijetke događaje. Kao u slučaju principa katastrofe, formaliziranje pojma principa zavjere je najprirodnije u smislu repa sume slučajnih varijabli. Konkretno, princip zavjere podrazumijeva da rep zbroja slučajnih varijabli dominira nad repom maksimalnog elementa u zbroju. Tako je zadovoljeno

$$\mathbb{P}(M_n > t \mid S_n > t) \to 0, \text{kako } t \to \infty.$$

Budući da je naš fokus na distribucijama teškog repa, prvenstveno ćemo se fokusirati na princip katastrofe. Njegova važnost dovela je do smislene definicije klasa distribucija nazvane *subeksponencijalne distribucije* koje odgovaraju točno onim distribucijama za koje vrijedi princip katastrofe, iako veza između te klase i definicije principa u početku neće biti očita. Najklasičnija definicija klase subeksponencijalnih distribucija je iduća.

**Definicija 2.2.2.** *Distribucija $F$ je subeksponencijalna distribucija ako vrijedi da je za sve $n \geq 2$ nezavisne slučajne varijable $X_1, X_2, ..., X_n$ s funkcijom distribucije $F$ zadovoljeno*

$$\mathbb{P}(S_n > t) \sim n\mathbb{P}(X_1 > t)$$

To kompaktnije još zapisujemo i kao $\bar{F}(x)^{*n} \sim n\bar{F}(x)$. Ova klasična definicija ne daje odmah do znanja kako su subeksponencijalne distribucije povezane s principom katastrofe, međutim, jednostavnim računom to je lako uvidjeti. Konkretno, moguće je vidjeti da je $\mathbb{P}(X_1 > t)$ asimptotski jednako $\mathbb{P}(M_n > t)$:

$$\lim_{t\to\infty} \frac{\mathbb{P}(M_n > t)}{\mathbb{P}(X_1 > t)} = \lim_{t\to\infty} \frac{1 - (1 - \bar{F}(t))^n}{\bar{F}(t)} = \lim_{t\to\infty} \frac{1 - (1 - n\bar{F}(t) + \binom{n}{2}\bar{F}(t)^2 - ...)}{\bar{F}(t)} = n$$

Gornji račun govori da je rep maksimuma $n$ slučajnih varijabli proporcionalan $n$ puta repu pojedinačne slučajne varijable, pa vidimo da je subeksponencijalnost ekvivalentna principu katastrofe. Formalnije,

**Lema 2.2.3.** *Neka su $X_1, X_2, .., X_n$ nezavisne slučajne varijable s funkcijom distribucije $F$. Tada su iduće tvrdnje ekvivalentne:*
*i) $F$ je subeksponencijalna distribucija*
*ii) $F$ zadovoljava princip katastrofe, to jest* $\mathbb{P}(S_n > t) \sim \mathbb{P}(M_n > t)$ *za sve* $n \geq 2$.

Koristeći Lemu 2.1.11., jasno vidimo kako sve regularno varirajuće funkcije zadovoljavaju princip katastrofe, pa direktno iz prošle leme zaključujemo

**Teorem 2.2.4.** *Sve regularno varirajuće distribucije su subeksponencijalne.*

Posebno, ispostavlja se da nije ni potrebno da definicija subeksponencijalnosti vrijedi za sve $n \geq 2$,već ako vrijedi za $n = 2$ onda nužno vrijedi i za sve $n \geq 2$. Nadalje, ako vrijedi za neki $n \geq 2$, nužno vrijedi i za $n = 2$ i, posljedično, za sve $n \geq 2$. Takvo razmatranje je sažeto u idućoj lemi, koju navodimo bez dokaza.

**Lema 2.2.5.** *Neka su $X_1, X_2, \ldots, X_n$ nezavisne slučajne varijable s funkcijom distribucije $F$. Tada su iduće tvrdnje ekvivalentne:*
*i) $F$ je subeksponencijalna distribucija, to jest* $\mathbb{P}(S_n > t) \sim n\mathbb{P}(X_1 > t)$ *za sve* $n \geq 2$
*ii)* $\mathbb{P}(X_1 + X_2 > t) \sim 2\mathbb{P}(X_1 > t)$
*iii)* $\mathbb{P}(S_n > t) \sim n\mathbb{P}(X_1 > t)$ *za neki* $n \geq 2$

Iako nam prethodna lema podosta olakšava zadatak provjere subeksponencijalnosti, to i dalje nije lak posao. Ova klasa često je korištena kao kandidat za distribuciju kojom modeliramo veličine šteta, ali i u brojnim drugim vjerojatnostim kontekstima od kojih ćemo neke opisati kasnije. Glavni nedostaci subeksponencijalnih distribucija su ti što su definirane samo preko limesa (a to je u praksi teško provjeravati) i što nisu definirane parametarskim uvjetom. Do sada je subeksponencijalna klasa na neki način prkosila reprezentaciji pa je izazovan problem procijeniti dolazi li skup aktuarskih podataka iz subeksponencijalne distribucije ili ne. Nadalje, poznato je da skup subeksponencijalnih distribucija nije zatvoren ni pod konvolucijom ni pod konveksnim kombinacijama. Sve su to razlozi zašto u praksi izbjegavamo tu klasu kao takvu, već se radije odlučujemo za distribucije koje su subeksponencijalne, ali u isto vrijeme i sadrže dovoljno parametara. Među mnogim parametriziranim primjerima u klasi subeksponencijalnih distribucija spominjemo lognormalnu razdiobu, Pareto razdiobu i kasnije definirane distribucije Pareto tipa, kao i ne-normalne stabilne distribucije. U cilju nam je dati našu definiciju distribucija repa težeg i lakšeg od eksponencijalne, kao i distribucija dugog repa, a zatim i uspostaviti veze između svih navedenih klasa.

**Definicija 2.2.6.** *Za distribuciju F kažemo da je dugog repa ako za svaki $y \in \mathbb{R}$ vrijedi*

$$\lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1,$$

*to jest $\bar{F}(x+y) \sim \bar{F}(x)$. Za nenegativnu slučajnu varijablu kažemo da je dugog repa ako je njena odgovarajuća funkcija distribucije dugog repa.*

**Definicija 2.2.7.** *Kažemo da je distribucija F teškog repa (težeg od eksponencijalne) ako za svaki $\mu > 0$*

$$\limsup_{x\to\infty} \frac{1-F(x)}{e^{-\mu x}} = \limsup_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \infty$$

*dok su distribucije lakog repa (lakšeg od eksponencijalne) one za koje*

$$\limsup_{x\to\infty} \frac{1-F(x)}{e^{-\mu x}} = \limsup_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = 0$$

*kako $x \to \infty$.*

Definicija za slučajne varijable vrijedi svugdje analogno. Sada, od interesa nam je dokazati i iskazati teorem koji govori o odnosu subeksponencijalnih distribucija i distribucija teškog repa. Njegova važnost je ta što u dokazu koristi i klasu distribucija dugog repa, to jest preko nje uspostavlja odnos svih klasa.
Prvo, dokažimo jednu tehničku lemu.

**Lema 2.2.8.** *Neka je X slučajna varijabla. Ako vrijedi*

$$\liminf_{x\to\infty} \frac{-\ln \mathbb{P}(X > x)}{x} = 0,$$

*onda je X teškog repa.*

*Dokaz.* Neka je X slučajna varijabla s distribucijom $F$ koja zadovoljava izraz iz iskaza. To implicira da postoji strogo rastući niz $(x_k)_{k\geq 1}$ koji zadovoljava $\lim_{k\to\infty} x_k$, takav da je

$$\lim_{k\to\infty} -\frac{\ln(\bar{F}(x_k))}{x_k} = 0$$

Sada, za dani $\mu > 0$, postoji $k_0 \in \mathbb{N}$ takav da

$$\begin{aligned} &-\frac{\ln(\bar{F}(x_k))}{x_k} < \frac{\mu}{2} \quad \forall k > k_0 \\ \iff &\bar{F}(x_k) > e^{\frac{-\mu x_k}{2}} \quad \forall k > k_0 \\ \iff &\frac{\bar{F}(x_k)}{e^{-\mu x_k}} > e^{\frac{\mu x_k}{2}} \quad \forall k > k_0. \end{aligned}$$

Sada lako vidimo da je $\lim_{k\to\infty} \frac{\bar{F}(x_k)}{e^{-\mu x_k}} = \infty$. Budući da ovo vrijedi za svaki $\mu > 0$, po definiciji smo dokazali da je $F$ teškog repa. □

**Teorem 2.2.9.** *Subeksponencijalne distribucije su distribucije teškog repa.*

*Dokaz.* Najprije pokažimo da je subeksponencijalna distribucija $F$ distribucija dugog repa, tj. da zadovoljava ekvivalentnu definiciju

$$\lim_{t\to\infty} \frac{\bar{F}(t-y)}{\bar{F}(t)} = 1, \quad \forall y > 0$$

Neka su $X_1, X_2$ nezavisne slučajne varijable s distribucijom $F$. Prisjetimo se da $\bar{F}^{2*}(t) := \mathbb{P}(X_1 + X_2 > t)$. Budući da je $\mathbb{P}(X_1 + X_2 > t) = \mathbb{P}(X_1 > t) + \mathbb{P}(X_1 < t, X_1 + X_2 > t)$, vrijedi $\bar{F}^{2*}(t) = \bar{F}(t) + \int_0^t \bar{F}(t-u)dF(u)$. Također, iz definicije subeksponencijalnosti znamo da je $\lim_{t\to\infty} \frac{\bar{F}^{2*}(t)}{\bar{F}(t)} = 2$. Iz prethodnog razmatranja vidimo da vrijedi zapis

$$\frac{\bar{F}^{2*}(t)}{\bar{F}(t)} = \frac{\bar{F}(t) + \int_0^t \bar{F}(t-u)dF(u)}{\bar{F}(t)},$$

što nam daje

$$\lim_{t\to\infty} \int_0^t \frac{\bar{F}(t-u)}{\bar{F}(t)} dF(u) = 1.$$

Označimo funkciju pod gornjim limesom s $I(t)$. Fiksirajući $y > 0$, možemo ograničiti $I(t)$ na sljedeći način:

$$I(t) = \int_0^y \frac{\bar{F}(t-u)}{\bar{F}(t)} dF(u) + \int_y^t \frac{\bar{F}(t-u)}{\bar{F}(t)} dF(u) \geq \int_0^y dF(u) + \frac{\bar{F}(t-y)}{\bar{F}(t)} \int_y^t dF(u) =$$
$$= F(y) + \frac{\bar{F}(t-y)}{\bar{F}(t)}(F(t) - F(y)),$$

pa onda vrijedi

$$\frac{\bar{F}(t-y)}{\bar{F}(t)} \leq \frac{I(t) - F(y)}{F(t) - F(y)}.$$

Kako znamo da $\lim_{t\to\infty} I(t) = 1$, zaključujemo

$$\limsup_{t\to\infty} \frac{\bar{F}(t-y)}{\bar{F}(t)} \leq 1.$$

S druge strane, jer vrijedi $\bar{F}(t-y) \geq \bar{F}(t)$ imamo i

$$\liminf_{t\to\infty} \frac{\bar{F}(t-y)}{\bar{F}(t)} \geq 1.$$

Ovime je dokazano da $F$ zadovoljava definiciju s početka dokaza, tj da je $F$ distribucija dugog repa. Kada bismo dokazali da taj uvjet povlači da je $F$ i distribucija teškog repa, tvrdnja teorema bi bila dokazana u potpunosti.
Prethodna lema kaže da je dovoljno dokazati da je $\liminf_{x\to\infty} \frac{-ln\mathbb{P}(X>x)}{x} = 0$. Time vođeni, definirajmo $\Psi(t) = -\ln(\mathbb{P}(X > t))$. Lako se vidi da definicija dugog repa implicira da $lim_{t\to\infty}(\Psi(t) - \Psi(t-1)) = 0$. Zato, za $\epsilon > 0$ postoji $t_0 > 0$ takav da $\Psi(t) - \Psi(t-1) \leq \epsilon$ za sve $t \geq t_0$. Budući da je $\Psi$ neopadajuća funkcija, slijedi da za $t \geq t_0$,

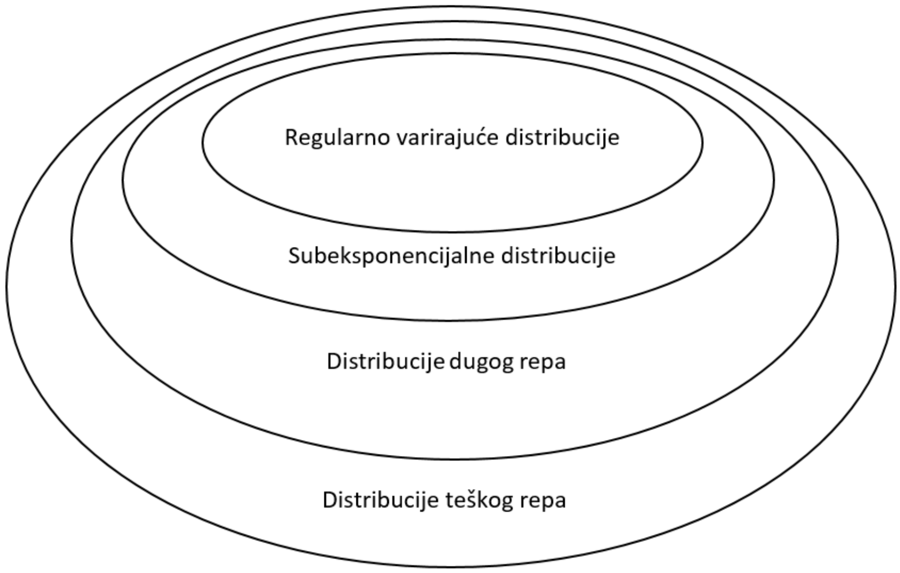
$$\Psi(t) \leq \Psi(t-1) + \epsilon \leq \Psi(t-2) + 2\epsilon \leq \cdots \leq \Psi(t_0) + \lceil t - t_0 \rceil \epsilon,$$

pa je $\limsup_{t\to\infty} \frac{\Psi(t)}{t} \leq \epsilon$. Puštanjem $\epsilon \downarrow 0$, dobijamo $\limsup_{t\to\infty} \frac{\Psi(t)}{t} \leq 0$. Kako znamo da je $\Psi(t) \geq 0$, zaključujemo

$$\lim_{t\to\infty} \frac{\Psi(t)}{t} = 0,$$

pa je po prethodnoj lemi dokazana tvrdnja. □

U dokazu prethodnog teorema uspostavljena je veza subeksponencijalnih distribucija i distribucija teškog repa preko distribucija dugog repa. Kompletan vizualni prikaz odnosa klasa distribucija nalazi se na Slici 2.2.




Slika 2.2: Međusobni odnos dosad navedenih klasa distribucija

## Primjeri i primjene

S obzirom na dan pregled subeksponencijalnih distribucija u gornjem dijelu, cilj nam je i dati primjer kako svojstva subeksponencijalnih distribucija mogu biti moćni analitički alati, omogućujući analizu u širokom spektru primjena.

**Slučajne sume**
Budući da klasa subeksponencijalnih distribucija služi kao formalizacija principa katastrofe, prirodno je da se njena česta primjena u okruženjima koji su u temelju povezani s nekim oblikom slučajne sume. U industriji se često srž analize oslanja na razumijevanje vrlo jednostavnog procesa – sume slučajnog broja nezavisnih jednako distribuiranih slučajnih varijabli. Primjer s početka poglavlja, to jest broj smrtnih slučajeva od potresa u određenoj godini razumno je modelirati na način da postoji nasumičan broj potresa u godini i da su brojevi smrti od svakog nezavisni i jednako distribuirani. Naravno, i distribucija broja potresa i broja smrtnih slučajeva u potresu su teškog repa. Sličan model bi mogao obuhvatiti i druge prirodne pojave, radi čega nam je u kontekstu reosiguranja ovo i posebno bitno.
Neka je $\{X_i\}_{\{i\geq 1\}}$ niz nezavisnih jednako distribuiranih slučajnih varijabli s očekivanjem $\mathbb{E}(X)$ te neka $N$ poprima vrijednosti u $\mathbb{N}$ i nezavisan je s $\{X_i\}_{\{i\geq 1\}}$. Cilj nam je opisati

$$S_N = \sum_{i=1}^{N} X_i.$$

U toj analizi od velikog značaja nam je Waldova jednakost.

**Lema 2.2.10** (Waldova jednakost)**.** *Vrijedi jednakost*

$$\mathbb{E}(S_N) = \mathbb{E}(\sum_{i=1}^{N} X_i) = \mathbb{E}(N)\mathbb{E}(X),$$

*s tim da obje strane mogu biti jednake* $+\infty$.

Waldova jednakost govori da, s obzirom na očekivanje, možemo zanemariti činjenicu da je $N$ slučajan. Da smo uzeli upravo $N$ za neku fiksnu konstantu $n \in \mathbb{N}$, tada vrijedi $\mathbb{E}[S_n] = n\mathbb{E}[X]$, a Waldova jednakost jednostavno zamjenjuje $n$ s $\mathbb{E}[N]$. Iako je Waldova jednakost vrlo koristan rezultat, nije uvijek dovoljno imati samo karakterizaciju očekivanja $S_N$, već često želimo znati i $Var(S_N)$, ili čak cijelu distribuciju od $S_N$. Srećom, nije teško generalizirati Waldovu jednakost. Na primjer, varijanca slučajne sume $S_N$ još uvijek ima zgodnu zatvorenu formulu

$$Var(S_N) = \mathbb{E}(N)Var(X) + (\mathbb{E}(X))^2 Var(N)$$

Moguće je ići i dalje od same varijance i izvesti Waldove nejednakosti za rep slučajnih suma. Međutim, rezultati o repu $S_N$ nisu tako općeniti kao Waldova jednakost i oslanjaju se na korištenje specifičnih svojstava distribucija. Zapravo, kao što pokazujemo u nastavku, rep nasumičnih zbroja može se ponašati različito ovisno o tome jesu li $X_i$ i (ili) $N$ teškog ili lakog repa. Ovdje klasa opisanih subeksponencijalnih distribucija pokazuje svoju važnost (kao i klasa regularno varirajućih distribucija).
Diskusiju počinjemo razmatranjem ponašanja repa distribucije. Prirodno bi bilo očekivati ponašanje poput ovoga viđenog u Waldovoj jednakosti, kao da je slučajna varijabla $N$ konstanta $n$. Kada bi to bio slučaj, i kad bi $X_i$ bile subeksponencijalne, vrijedilo bi

$$\mathbb{P}(\sum_{i=1}^{N} X_i > t) \sim \mathbb{E}(N)\mathbb{P}(X_1 > t)$$

Bilo bi vrlo zgodno kad bi gornja formula vrijedila i za slučajnu sumu, uz zamjenu $n$ sa $\mathbb{E}(N)$. Idući teorem kaže da je to istina ukoliko je $N$ lakog repa.

**Teorem 2.2.11.** *Neka je $X_1, X_2, ...$ beskonačan niz nezavisnih jednako distribuiranih subeksponencijalnih slučajnih varijabli te neka je $N$ slučajna varijabla lakog repa nezavisna od niza. Tada,*

$$\mathbb{P}(\sum_{i=1}^{N} X_i > t) \sim n\mathbb{P}(X_1 > t).$$

*Dokaz.* Ideja dokaza je da zapišemo našu slučajnu sumu tako da možemo iskoristiti definiciju subeksponencijalnih distribucija. Ako je distribucija $X$ dana s $F$, uvjetovanjem po $N$, imamo

$$\mathbb{P}(\sum_{i=1}^{N} X_i > t) = \sum_{i\in\mathbb{N}} \mathbb{P}(N = i)\bar{F}^{i*}(t)$$

Gledajući limes, zapisujemo ovo u obliku na koji možemo primijeniti $\lim_{t\to\infty} \frac{\bar{F}^{i*}(t)}{\bar{F}_1(t)} = i$:

$$\lim_{t\to\infty} \frac{\mathbb{P}(\sum_{i=1}^{N} X_i > t)}{\mathbb{P}(X_i > t)} = \lim_{t\to\infty} \sum_{i\in\mathbb{N}} \frac{\bar{F}^{i*}(t)}{\bar{F}_1(t)}\mathbb{P}(N = i).$$

Pretpostavkom da smijemo ući s limesom pod znak sume, dobivamo

$$\lim_{t\to\infty} \sum_{i\in\mathbb{N}} \frac{\bar{F}^{i*}(t)}{\bar{F}_1(t)}\mathbb{P}(N = i) = \sum_{i\in\mathbb{N}} i\mathbb{P}(N = i) = \mathbb{E}(N)$$

Preostaje opravdati zamjenu poretka sumacije i limesa. To radimo pozivanjem na teorem o dominiranoj konvergenciji, to jest pronalaskom niza $\beta_i$ koji je gornja međa od $\frac{\bar{F}^{i*}(t)}{\bar{F}_1(t)}P(N = i)$

i za kojeg vrijedi $\sum_{i\in\mathbb{N}}\beta_i < \infty$. Koristit ćemo tehnički rezultat da za svaki $\epsilon > 0$, postoji $K > 0$ takav da je $\frac{\bar{F}^{i*}(t)}{\bar{F}_1(t)} \leq K(1+\epsilon)^i$, čiji dokaz nije u domeni ovog rada. Uzimanjem $\beta_i = K(1+\epsilon)^i\mathbb{P}(N = i)$, imamo $\sum_{i\in\mathbb{N}}\beta_i = \mathbb{E}((1+\epsilon)^N)$. Ovo je za dovoljno mali $\epsilon$ konačno radi činjenice što je N distribucija lakog repa. □

Karakterizacija slučajne sume u prethodnom teoremu oslanja se na činjenicu da su $X_i$ teškog repa, a $N$ lakog. U tom slučaju imamo jednostavan oblik za rep slučajne sume koji slijedi intuiciju izvedenu iz Waldove jednakosti. Ono što sada ostaje jest razumjeti što se događa kada su stvari obrnute i distribucija $N$ je dominantna, tj. kada je $N$ teškog repa a $X_i$ lakog. Intuitivno je očekivati da će rep slučajne sume biti određen repom od $N$. Kao u prethodnom primjeru, razmatramo što bi se dogodilo da je $X_i$ deterministička vrijednost $x$. U tom slučaju, ponašanje sume je jednostavno:

$$\mathbb{P}(\sum_{i=1}^{N} x > t) = \mathbb{P}(Nx > t) = \mathbb{P}(N > \frac{t}{x})$$

Dakle, moglo bi se pretpostaviti da bi ova formula trebala nastaviti vrijediti i za rep slučajne sume, jednostavnom zamjenom x s $\mathbb{E}(X)$. Ta pretpostavka opet je točna. Posebno, kada je $X_i$ je lakog repa, a $N$ teškog (konkretno kada je regularno varirajuća), slučajnu sumu možemo okarakterizirati na sljedeći način.

**Teorem 2.2.12.** *Neka su $X_i$ slučajne varijable lakog repa i neka je $N$ regularno varirajuća s indeksom $-\alpha$ ($\alpha > 0$). Tada:*

$$\mathbb{P}(\sum_{i=1}^{N} X_i > t) \sim \mathbb{P}(N > \frac{t}{\mathbb{E}(X)})$$

Teorem nećemo dokazivati, ali važno je istaknuti da ne vrijedi generalna tvrdnja za distribucije teškog repa, dakle bitna nam je pretpostavka da je $N$ regularno varirajuća. Svojstvo koje je pak nužan uvjet za $\mathbb{P}(\sum_{i=1}^{N} X_i > t) \sim \mathbb{P}(N > \frac{t}{\mathbb{E}(X)})$ je takozvana *neosjetljivost na korjenovanje*, to jest

$$\mathbb{P}(N > x) \sim \mathbb{P}(N > x - \sqrt{x}).$$

## 2.3 Teorija ekstremnih vrijednosti

Teorija koja igra veliku ulogu u ovoj priči je i *teorija ekstremnih vrijednosti*, a odgovora na pitanja od interesa u vezi ekstremnih događaja. Alternativni način definiranja velikih šteta možemo započeti promatrajući ponašanje maksimuma, koje je također usko povezano s već spomenutim konceptom PML-a. Središnji rezultat ove teorije je *Fisher-Tippett-Gnedenko*

*teorem*, koji specificira oblik granične distribucije za maksimume slučajnih varijabli centrirane i normalizirane konstantama. Tri su familije mogućih graničnih distribucija, a poznate su kao *distribucije ekstremnih vrijednosti*. U ovom odjeljku formalno ćemo ih definirati i izvesti njihove maksimalne domene privlačnosti i prikladne konstante za centriranje i normalizaciju. Ista pitanja korisno je postaviti i za sume slučajnih varijabli, no rezultati nisu analogni. Centrirajuće i normalizirajuće konstante ima smisla odabrati kao određene kvantile funkcije distribucije koja je u pozadini, a za formalnije iskazati i eventualno doakazati glavne rezultate oslonit ćemo se na teoriju iz prošlog odjeljka.
Kroz cijelo poglavlje $X, X_1, X_2, ...$ bit će niz nezavisnih i nedegeneriranih jednako distribuiranih slučajnih varijabli sa zajedničkom funkcijom distribucije $F$. Promatrat ćemo i maksimum uzorka $M_n = max\{X_1, X_2, ...X_n\}$, a dobivene rezultate lako možemo prilagoditi i za minimum koristeći $min\{X_1, X_2, ...\} = -max\{-X_1, -X_2, ... - X_n\}$. Budući da se ekstremi oni događaji bliži gornjem kraju nosača distribucije $F$, asimptotsko ponašanje $M_n$ mora biti povezano s desnim repom od $F$, blizu njene desne krajnje točke. Označavamo s

$$x_F = sup\{x \in \mathbb{R} : F(x) < 1\}$$

desnu krajnju točku distribucije $F$. Odmah vidimo da za $x < x_F$,

$$\mathbb{P}(M_n \leq x) = F^n(x) \to 0, \ n \to, \infty$$

a ako je $x_F < \infty$, za $x \geq x_F$ vrijedi

$$\mathbb{P}(M_n \leq x) = F^n(x) = 1.$$

To znači da $M_n$ konvergira po distribuciji prema konstanti $x_F$, pa samim time i po vjerojatnosti. Budući da onda $M_n$ ima podniz koji konvergira gotovo sigurno, i da je neopadajuć, slijedi da i $M_n$ konvergira gotovo sigurno prema $x_F$. Međutim, taj nam rezultat nije toliko značajan, već ćemo kao što je najavljeno promatrati slabu konvergenciju centriranih i normiranih maksimuma. To pitanje je na neki način analogon centralnog graničnog teorema u teoriji ekstremnih vrijednosti. Zanima nas koje sve distribucije i odgovarajuće konstante zadovoljavaju identitet

$$max\{X_1, X_2, ...X_n\} \stackrel{d}{=} c_n X + d_n$$

To pitanje je ekvivalentno pitanju koje klase od $F$ su do na afine transformacije zatvorene na maksimum.

**Definicija 2.3.1.** *Slučajna varijabla X je maksimalno stabilna ako zadovoljava gornji identitet za nezavisne jednako distribuirane varijable $X_1, X_2, ...X_n$ i odgovarajuće konstante $c_n > 0, d_n \in \mathbb{R}$, za svaki $n \geq 2$.*

Odsada ćemo konstante $c_n, d_n$ zajedničkim imenom zvati normirajuće konstante. Budući da je gornja jednakost ekvivalentna s

$$\frac{M_n - d_n}{c_n} \stackrel{d}{=} X,$$

zaključak će biti da je svaka maksimalno stabilna distribucija granična distribucija za maksimum nezavisnih jednako distribuiranih slučajnih varijabli. Štoviše, maksimalno stabilne distribucije su jedini kandidati za granične distribucije normiranih maksimuma.

**Teorem 2.3.2.** *Klasa maksimalno stabilnih distribucija je ekvivalentna klasi svih mogućih graničnih distribucija (nedegeneriranih) za prikladno normirane maksimume nezavisnih jednako distribuiranih slučajnih varijabli.*

Tako dolazimo do osnovnog teorema, koji formalno kaže:

**Teorem 2.3.3** (Fisher-Tippett-Gnedenko)**.** *Neka je $\{X_n\}_{n\geq 1}$ niz nezavisnih jednako distribuiranih slučajnih varijabli. Ako postoje normirajuće konstante $c_n \geq 0, d_n \in \mathbb{R}$ i nedegenerirana funkcija distribucije $H$ takva da je zadovoljeno*

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H,$$

*tada $H$ pripada jednoj od tri moguće familije funkcije distribucija*

$$\begin{aligned}
\textit{Fréchet–Pareto}: \quad & \phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ exp\{-x^{-\alpha}\} & x > 0 \end{cases} \\
\textit{Weibull}: \quad & \psi_\alpha(x) = \begin{cases} 1 & x > 0 \\ exp\{-(-x)^{\alpha}\} & x \leq 0 \end{cases} \\
\textit{Gumbel}: \quad & \Lambda(x) = exp\{-exp\{-x\}\}, \ x \in \mathbb{R}
\end{aligned}$$

*za $\alpha > 0$.*

Granična distribucija o kojoj je riječ u gornjem teoremu je jedinstvena samo do na afine transformacije, to jest ako vrijedi $\lim_{n\to\infty} \mathbb{P}(\frac{M_n-d_n}{c_n} \leq x) = H(xc + d)$, tada je $H(x)$ također granična distribucija, ali za transformirane normirajuće konstante:
$\lim_{n\to\infty} \mathbb{P}(\frac{M_n-b_n}{a_n} \leq x) = H(x)$, gdje su $a_n = \frac{c_n}{c}, b_n = d_n - \frac{dc_n}{c}$.
Unatoč činjenici da se distribucije $\phi_\alpha, \psi_\alpha$ i $\Lambda$ poptuno razlikuju po tome što njima modeliramo, postoji jasna matematička veza među njima.

$$\begin{aligned}
& X \ \text{ ima funckiju distribucije } \ \phi_\alpha \\
\Leftrightarrow \ & \ln X^\alpha \ \text{ ima funckiju distribucije } \ \Lambda \\
\Leftrightarrow \ & -X^{-1} \text{ima funkciju distribucije } \ \psi_\alpha
\end{aligned}$$

Ukratko, *Fisher-Tippett-Gnedenko teorem* kaže da tri moguće vrste distribucija koje zadovoljavaju spomenutu relaciju odgovaraju:

- Fréchet–Pareto: $M_n \stackrel{d}{=} n^{\frac{1}{\alpha}} X$
- Weibull: $M_n \stackrel{d}{=} n^{\frac{-1}{\alpha}} X$
- Gumbel: $M_n \stackrel{d}{=} X + \ln n$

Ove tri familije distribucija zajedničkim imenom zovemo *distribucije ekstremnih vrijednosti*.

**Primjer 2** (Maksimum eksponencijalnih slučajnih varijabli)**.** *Neka je* $\{X_n\}_{n\geq 1}$ *niz jednakod istribuiranih međusobno nezavisnih eksponencijalnih varijabli s parametrom* $\lambda = 1$*. Tada je* $\mathbb{P}(M_n - \ln n \leq x) = (\mathbb{P}(X \leq x + \ln n))^n = (1 - \frac{e^{-x}}{n})^n \to exp\{e^{-x}\} = \Lambda(x)$, $x \in \mathbb{R}$.

**Definicija 2.3.4** (Maksimalna domena privlačnosti)**.** *Za slučajnu varijablu X kažemo da pripada maksimalnoj domeni privlačnosti (maximum domain of attraction) distribucije ekstremnih vrijednosti H ako postoje normirajuće konstante* $c_n > 0, d_n \in \mathbb{R}$ *takve da vrijedi*

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H$$

*Još pišemo i* $X \in MDA(H)$.

**Lema 2.3.5** (Karakterizacija *MDA(H)*)**.** *Funkcija distribucije F pripada maksimalnoj domeni privlačnosti distribucije ekstremnih vrijednosti H s normirajućim konstantama* $c_n > 0$, $d_n \in \mathbb{R}$ *ako i samo ako* $\lim_{n\to\infty} \bar{F}(c_n x + d_n) = -\ln H(x), x \in \mathbb{R}$*. Kada je* $H(x) = 0$*, gornji limes interpretiramo kao* $\infty$.

Cilj nam je odrediti maksimalnu domenu privlačnosti distribucija ekstremnih vrijednosti koristeći uvedene pojmove. Uvjerit ćemo se da to nije teško za Weibullovu i Fréchet-Paretovu, dok je za Gumbelovu to teži zadatak. Uvodimo koncept iz kojeg nam slijedi jedna klasa ekvivalencije na skupu funkcija distribucija.

**Definicija 2.3.6** (Ekvivalencija repova)**.** *Dvije funkcije distribucije F i G su ekvivalentnog repa ako imaju istu desnu krajnju točku,* $x_F = x_G$*, i zadovoljeno je*

$$\lim_{x\uparrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c,$$

*za neku konstantu* $0 < c < \infty$.

Svaka maksimalna domena privlačnosti zatvorena je s obzirom na ekvivalenciju repova, to jest za $F$ i $G$ ekvivalentnog repa, $F \in MDA(H)$ ako i samo ako $G \in MDA(H)$. Čak štoviše, za dvije distribucije ekvivalentnog repa možemo uzeti iste normirajuće konstante. Taj rezultat pokazuje se vrlo koristan, budući da računanje istih može biti kompliciran postupak. Budući da je maksimum na neki način $M_n$ empirijska verzija $(1-\frac{1}{n})$-tog kvantila, ima smisla promatrati *kvantilnu funkciju*.

**Definicija 2.3.7** (Kvantilna funkcija). *Neka je F funkcija distribucije. Tada funkciju oblika*

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$$

*zovemo kvantilna funkcija od F.* $x_t = F^{\leftarrow}(t)$ *označava t-ti kvantil od F.*

Sljedeći niz tvrdnji i teorema, zbog svoje kompleknosti, a i cilja da naglasak bude na intuiciji i primjerima, biti će uglavnom dani bez dokaza, a odnosit će se na već najavljenu karakterizaciju maksimalnih domena atrakcija distribucija ekstremnih vrijednosti. Diskusiju o maksimalnoj domeni atrakcije od $\phi_\alpha$ za $\alpha > 0$, započinjemo razvojem u Taylorov red

$$1 - \phi_\alpha(x) = 1 - exp\{-x^{-\alpha}\} \sim x^{-\alpha}, x \to \infty$$

Može se pokazati da se maksimalna domena atrakcije od $\phi_\alpha$ sastoji od regularno varirajućih distribucija $F$ s indeksom $-\alpha$. Za $F \in MDA(H)$, $d_n$ može biti 0, a $c_n$ izražavamo kao

$$c_n = F^{\leftarrow}(1 - n^{-1}) = \inf\{x \in \mathbb{R} : F(x) \geq 1 - n^{-1}\} = \inf\{x \in \mathbb{R} : (\frac{1}{1 - F(x)}) \geq n\} =$$
$$= (\frac{1}{1 - F(x)})^{\leftarrow}(n).$$

Formalnije,

**Teorem 2.3.8.** *Funkcija distribucije F pripada maksimalnoj domeni atrakcije od* $\phi_\alpha, \alpha > 0$ *ako i samo ako* $\bar{F}(x) = x^{-\alpha}L(x)$ *za neku sporo varirajuću funkciju L. Dodatno, normirajuće konstante odabiremo kao u gornjem računu.*

Ukratko, vrijedi $F \in MDA(\phi_\alpha) \Leftrightarrow \bar{F}$ je regularno varirajuća s indeksom $-\alpha$. Važno je primijetiti kako ovaj teorem implicira da svaki $F \in MDA(\phi_\alpha)$ nema desnu krajnju točku, tj $x_F = \infty$. *Von Mises* je pronašao lako provjerljiv uvjet za provjeru maksimalne domene atrakcije preko funkcije gustoće distribucije.

**Lema 2.3.9** (Von Misesov uvjet za $\phi_\alpha$). *Neka je F apsolutno neprekidna funkcija distribucije s funkcijom gustoće f. Ako vrijedi*

$$\lim_{x\to\infty} \frac{xf(x)}{\bar{F}(x)} = \alpha > 0,$$

*tada je i* $F \in MDA(\phi_\alpha)$.

Idući rezultat daje nam uvid u strukturu skupa MDA($\phi_\alpha$), a pokazuje se i kao bitan rezultat pri računanju normirajućih konstanti.

**Teorem 2.3.10** (Zatvorenost MDA($\phi_\alpha$))**.** *Neka su $F$ i $G$ funkcije distribucije i neka je $F \in MDA(\phi_\alpha)$ s normirajućim konstantama $c_n > 0$, to jest*

$$\lim_{n\to\infty} F^n(c_n x) = \phi_\alpha(x),\ x > 0.$$

*Tada vrijedi*

$$\lim_{n\to\infty} G^n(c_n x) = \phi_\alpha(cx),\ x > 0$$

*za konstantu $c > 0$ ako i samo ako su $F$ i $G$ ekvivalentnog repa s*

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{\bar{G}(x)} = c^\alpha.$$

**Primjer 3.** *Ako promatramo distribucije oblika $\bar{F}(x) \sim Kx^{-\alpha}$, za $K,\alpha > 0$, očito je $\bar{F}$ regularno varirajuća s indeksom $-\alpha$, pa je i $F$ u MDA($\phi_\alpha$). Konstante $c_n$ odabiremo identičnim računom kao u gornjim rezultatima, pa vrijedi $c_n = (Kn)^{\frac{1}{\alpha}}$, to jest*

$$(Kn)^{\frac{-1}{\alpha}} M_n \xrightarrow{d} \phi_\alpha.$$

*Slično, i Cauchyjeva distirbucija se nalazi u MDA($\phi_\alpha$).*

Odgovarajući rezultati daju se izvesti i za $\psi_\alpha$. Najprije, bitno je istaknuti da sve funkcije distribucije iz MDA($\psi_\alpha$) imaju konačnu desnu krajnju točku $x_F$. Već otprije znamo da vrijedi

$$\psi_\alpha(-x^{-1}) = \phi_\alpha(x),\ x > 0.$$

Navodimo analogni teorem o karakterizaciji maksimalne domene privlačnosti za *Weibullov slučaj*.

**Teorem 2.3.11.** *Funkcija distribucije $F$ pripada maksimalnoj domeni atrakcije od $\psi_\alpha, \alpha > 0$, ako i samo ako $x_F < \infty$ i $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ za neku sporo varirajuću funkciju $L$. Dodatno, normirajuće konstante odabiremo kao $c_n = x_F - F^{\leftarrow}(1 - n^{-1})$, $d_n = x_F$. Tako imamo,*

$$\frac{M_n - x_F}{x_F - F^{\leftarrow}(1 - n^{-1})} \xrightarrow{d} \psi_\alpha.$$

Radi ograničenosti nosača funkcije distribucije s desna, *Weibullov slučaj* nije najpogodniji za modeliranje ekstremnih događaja u reosiguranju. Iako će svakako postojati gornja ograda za štetu, makar izuzeno visoka, ne treba nam dodatan parametar. Svakako, u primjeni se više fokusiramo na MDA($\Lambda$) i MDA($\phi_\alpha$) jer nam je u cilju dopustiti proizvoljno veliki iznos u uzorku. Što se tiče *Von Misesovog* uvjeta, formulacija je sljedeća:

**Lema 2.3.12** (Von Misesov uvjet za $\psi_\alpha$)**.** *Neka je $F$ apsolutno neprekidna funkcija distribucije s funkcijom gustoće $f$ pozitivnoj na nekom konačnom intervalu $(z, x_F)$. Ako vrijedi*

$$\lim_{x \uparrow x_F} \frac{(x_F - x) f(x)}{\bar{F}(x)} = \alpha > 0,$$

*tada je i $F \in MDA(\psi_\alpha)$.*

Analogna tvrdnja teoremu zatvorenosti za MDA($\psi_\alpha$) glasi:

**Teorem 2.3.13** (Zatvorenost MDA($\psi_\alpha$))**.** *Neka su $F$ i $G$ funkcije distribucije s jednakim desnim krajnjim točkama $x_F = x_G$ i neka je $F \in MDA(\psi_\alpha)$ s normirajućim konstantama $c_n > 0$, to jest*

$$\lim_{n \to \infty} F^n(c_n x + x_F) = \psi_\alpha(x), \ x < 0.$$

*Tada vrijedi*

$$\lim_{n \to \infty} G^n(c_n x + x_F) = \psi_\alpha(cx), \ x < 0,$$

*za konstantu $c > 0$ ako i samo ako su $F$ i $G$ ekvivalentnog repa s*

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c^{-\alpha}.$$

Možemo reći da MDA($\psi_\alpha$) pripadaju distribucije koje zadovoljavaju Von Misesov uvjet i one njima ekvivalentnog repa.

**Primjer 4** (Uniformna distribucija)**.** *Promatramo uniformnu distribuciju na* $[0, 1]$. *Očito je $x_F = 1$ i vrijedi $\bar{F}(1 - x^{-1}) = x^{-1}$, što je regularno varirajuća funkcija s indeksom -1. Po prethodnim teoremima, $c_n = \bar{F}(1 - n^{-1}) = n^{-1}$, pa zaključujemo da vrijedi $\frac{M_n - 1}{n} \xrightarrow{d} \psi_\alpha$.*

Kada je riječ o *Gumbelovom slučaju* već smo napomenuli kako je situacija složenija. MDA($\Lambda$) sadrži pregršt raznih funkcija distribucija za koje pojmovi regularne varijacije nisu sami po sebi dovoljni da ih u potpunosti opišu, pa definiramo novu klasu funkcija.

**Definicija 2.3.14** (Von Misesova funkcija)**.** *Neka je $F$ distribucija s krajnjom desnom točkom $x_F \leq \infty$. Pretpostavimo da postoji $z < x_F$ takav da se $F$ može prikazati u obliku*

$$\bar{F}(x) = c \cdot exp\{-\int_z^x \frac{1}{a(t)} dt\}, \ \ z < x < x_F,$$

*gdje je $c$ pozitivna konstanta i $a(\cdot)$ pozitivna, apsolutno neprekidna funkcija s gustoćom $a'$ tako da vrijedi $\lim_{x \uparrow x_F} a'(x) = 0$. Tada $F$ zovemo Von Misesovom funkcijom, a funkciju $a(\cdot)$ pomoćnom funkcijom od $F$.*

Prvo što možemo uočiti je da ovaj prikaz ima sličan oblik kao onaj iz *Karamatinog teorema o reprezentaciji*. Zamjenom $a(t) = \frac{t}{\beta(t)}$ za $\beta(t) \to \alpha \in (0, \infty)$, imamo isti izraz kao iz teorema, i $\bar{F}$ je regularno varirajuća s indeksom $-\alpha$.

**Primjer 5** (Eksponencijalna distribucija)**.** *Ovo je najjednostavniji oblik Von Misesove funkcije jer* $\bar{F}(x) = e^{-\lambda x}, \ \ x \geq 0, \ \ \lambda > 0$. *Pomoćna funkcija je* $a(x) = \lambda^{-1}$.

Ovakve funkcije ne daju sve moguće kandidate za MDA($\Lambda$), ali uz manje modifikacije možemo ju opisati u potpunosti. Zato dajemo sljedeću moguću karakterizaciju tog skupa.

**Teorem 2.3.15.** *Funkcija distribucije $F$ s desnom krajnjom točkom $x_F \leq \infty$ pripada maksimalnoj domeni privlačnosti od $\Lambda$ ako i samo ako postoji neki $z < x_F$ takav da se $F$ može prikazati u obliku*

$$\bar{F}(x) = c(x) exp\{-\int_z^x \frac{g(t)}{a(t)} dt\}, \ \ z < x < x_F,$$

*gdje su $c$, $g$ izmjerive funkcije koje zadovoljavaju $c(x) \to c > 0$, $g(x) \to 1$ kako $x \uparrow x_F$ i $a(x)$ je pozitivna, apsolutno neprekidna funkcija s gustoćom $a'$ takvom da vrijedi* $\lim_{x \uparrow x_F} a'(x) = 0$.
*Za normirajuće konstante uzimamo $d_n = F^{\leftarrow}(1 - n^{-1})$ i $c_n = a(d_n)$. Mogući kandidat za funkciju $a$ je*

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt, \ \ x < x_F.$$

Motivirani Von Miseovom funkcijom, funkciju $a$ zovemo *pomoćna funkcija*. Funkcija $a$ zapisana kao u prethodnom teoremu je zapravo $a(x) = \mathbb{E}(X - x | X > x)$, $x < x_F$, a zovemo ju *funkcija očekivanja viška*. Kod reosiguranja viška štete sa samopridržajem $u$, očekivanje svote koju plaća reosiguravatelj dano je s $a(u)\bar{F}(u)$.

**Teorem 2.3.16** (Zatvorenost MDA($\Lambda$) na ekvivalenciju repova)**.** *Neka su $F$ i $G$ funkcije distribucije s jednakim desnim krajnjim točkama $x_F = x_G$ i neka je $F \in MDA(\Lambda)$ s normirajućim konstantama $c_n > 0$ i $d_n \in \mathbb{R}$, to jest*

$$\lim_{n \to \infty} F^n(c_n x + d_n) = \Lambda(x), x \in \mathbb{R}.$$

*Tada,*

$$\lim_{n \to \infty} G^n(c_n x + d_n) = \Lambda(x + b), x \in \mathbb{R}$$

*za konstantu $b > 0$ ako i samo ako su $F$ i $G$ ekvivalentnog repa s*

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = e^b.$$

Možemo zaključiti kako se MDA($\Lambda$) sastoji od *Von Misesovih* funkcija i distribucija njima ekvivalentnog repa.
Diskusija iz ovog odjeljka pokazuje da se MDA($\Lambda$) sastoji od velikog broja distribucija čiji se repovi itekako razlikuju. Neke od distribucija koje tu spadaju su normalna i lognormalna.
Ukratko, prethodni rezultati nam govore da su distribucije ekstremnih vrijednosti jedini vrijedeći limesi za afino transformirane maksimume nezavisnih jednako distribuiranih varijabli. Bilo bi korisno uvesti prikaz te tri familije pomoću jednog parametra $\xi$, a to i radimo na način koji je danas prihvaćen kao standardni prikaz distribucija ekstremnih vrijednosti.

$$H_\xi(x) = \begin{cases} exp\{-(1+\xi x)^{\frac{-1}{\xi}}\} & \xi \neq 0 \\ exp\{-exp\{-x\}\} & \xi = 0 \end{cases}$$

za $1+\xi x > 0$, pri čemu je $x > -\xi^{-1}$ za $\xi > 0$, $x < -\xi^{-1}$ za $\xi < 0$ i $x \in \mathbb{R}$ za $\xi = 0$. $H_\xi$ se zove standardna generalizirana distribucija ekstremnih vrijednosti($GEV$) s indeksom ekstremne vrijednosti $\xi$ (EVI). Ima smisla uvesti i skaliranu verziju, dodavanjem parametara $\sigma$ i $\mu$, to jest supstituciju $x = \frac{x-\mu}{\sigma}$, sa $\mu \in \mathbb{R}$ i $\sigma > 0$. Tada je $H_{\xi,\mu,\sigma}$ $GEV$ - jedinstven i prikladan jednoparametarski prikaz za opisane tri familije distribucija.
Neka je sada *kvantilna funkcija repa* $U(t) = F^{\leftarrow}(1-t^{-1})$, $t > 0$. Tada su za $\xi \in \mathbb{R}$ iduće tri tvrdnje ekvivalentne.

1. $F \in MDA(H_\xi)$

2. Postoji pozitivna, izmjeriva funkcija $a(\cdot)$, takva da za $1 + \xi x > 0$ vrijedi

$$\lim_{x \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1+\xi x)^{\frac{-1}{\xi}} & \xi \neq 0 \\ e^{-x} & \xi = 0 \end{cases}$$

3. Za $x, y > 0$, $y \neq 1$,

$$\lim_{s \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \xi \neq 0 \\ \frac{\ln x}{\ln y} & \xi = 0 \end{cases}$$

Izraz u tvrdnji 2. ima i svoju vjerojatnosnu interpretaciju kao

$$\lim_{u \to x_F} \mathbb{P}(\frac{X-u}{a(u)} > x | X > u),$$

pa je ovime dan izraz granične distribucije za skalirani višak preko $u$. To je motivacija za sljedeću klasu distribucija.

**Definicija 2.3.17** (Generalizirana Paretova distribucija(GPD))**.** *Definiramo funkciju distribucije $G_\xi$ kao*

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{\frac{-1}{\xi}} & \xi \neq 0 \\ 1 - e^{-x} & \xi = 0, \end{cases}$$

*gdje je $x \geq 0$ ako $\xi \geq 0$, a $0 \leq x \leq \frac{-1}{\xi}$ ako $\xi < 0$. $G_\xi$ nazivamo standardnom generaliziranom Paretovom distribucijom, koju poznatom transformacijom $x = \frac{x-\mu}{\sigma}$ za $\mu \in \mathbb{R}$ i $\sigma > 0$ pretvaramo u $G_{\xi,\mu,\sigma}$(GPD).*

Kada je $\xi > 0$ za distribucije iz tvrdnje 2., riječ je o distribucijama *Paretovog tipa* za koje je $1 - F(x) \sim x^{\frac{-1}{\xi}} L(x)$, kada $x \uparrow \infty$ za sporo varirajuću $L$. Drugim riječima, $F$ je regularno varirajuća s indeksom $\frac{-1}{\xi}$. Za $\xi < 0$, riječ je o distribucijama s repom lakšim od distribucija Paretovog tipa. U slučaju $\xi = 0$ može postojati krajnja točka ali i ne mora (nosač beskonačan). Ukratko, uvedene GEV distribucije su zapravo granične distribucije za normalizirane maksimume, a GPD su granične distribucije skaliranih višaka preko nekog praga.
U kontekstu modeliranja velikih šteta, za prvi pokušaj njihovog modeliranja parametriziranom distribucijom zaslužan je Benckert. On je pretpostavio da distribucija veličine štete počinje kao distribucija Paretovog tipa, ali da je odsječena u točki osigurane sume tako da ostatak mase distribucije bude koncentrirano u njoj. Kasnije modifikacije su pokazale da je Paretova distribucija uistinu pogodna za modele velikih šteta pa su oduvijek popularne u osiguranjima od požara, oluja i odgovornosti.
Postoje i drugi načini definicije veličine štete osim uspoređivanja s eksponencijalnom distribucijom, ispitivanja utjecaja maksimalne štete na ukupnu sumu i promatranja ponašanja maksimuma u limesu. Neki od njih su na primjer analiza konvergencije izraza $\frac{S_n}{M_n}$ u nedegerirani limes ili pak očekivanja $\mathbb{E}(\frac{S_N - n\mu}{M_n})$ u konstantu $c$. Potonji pristup opet nas vodi do distribucija Paretovog tipa gdje je $\xi \in (0, 1) \bigcup (1, 2)$.
Bitno je još naglasiti da postoji razlika između velike štete i odstupanja. Dok je velika šteta legitimni član uzorka, odstupanjem se smatra neregularna vanjska vrijednost. Uvijek su mogući i događaji koji su potpuno neočekivani s obzirom na sve povijesne podatke - zato su korisne metode analize ekstremnih vrijednosti (EVA). Kada se oni pak dogode, treba biti spreman na prilagodbe i potencijalnu promjenu statističkih modela.

# Poglavlje 3

# Najpoznatiji primjeri

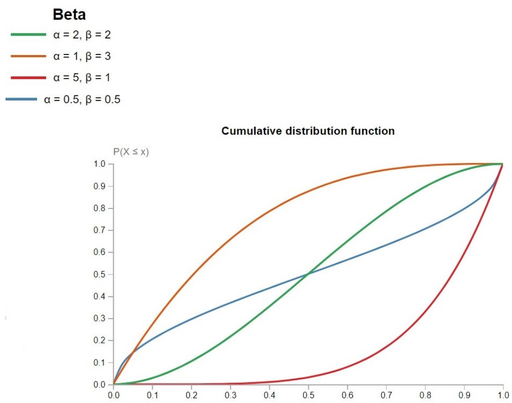
Preostaje navesti klasične primjere distribucija za modeliranje veličine štete koje se najčešće spominju u aktuarskoj literaturi. Neki od njih su jednostavni, a neki malo razrađenije varijacije, često dobivene nekom od transformacija iz osnovnog modela. Već je spomenuto kako su, radi važnosti ekstremnih događaja, u reosiguranju posljedično važne i distribucije ekstremnih vrijednosti. Njihov jednoparametraski zapis s parametrom $\xi$ je

$$H_\xi(x) = \begin{cases} exp\{-(1+\xi x)^{\frac{-1}{\xi}}\} & \xi \neq 0 \\ exp\{-exp\{-x\}\} & \xi = 0 \end{cases}$$

za $1 + \xi x > 0$, pri čemu je $x > -\xi^{-1}$ za $\xi > 0$, $x < -\xi^{-1}$ za $\xi < 0$ i $x \in \mathbb{R}$ za $\xi = 0$.
Neke od spomenutih transformacija mogu potpuno promijeniti težinu repa distribucije, pa u nastavku odjeljka dajemo primjere, s naglaskom na one težih repova.

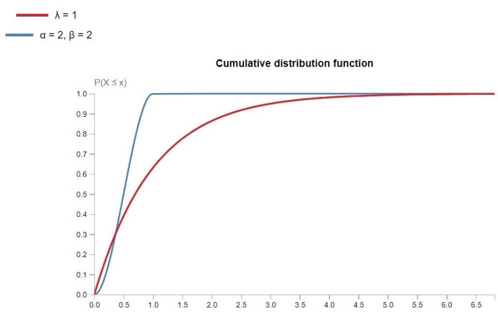
## 3.1 Laki repovi

- $\xi < 0$
  Klasični primjer distribucije s $\xi < 0$ je *Beta distribucija*, kojoj je $x_F = 1$ i
  $$F(x) = \frac{1}{B(p,q)} \int_0^x u^{p-1}(1-u)^{q-1} du,$$
  gdje je $B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, s indeksom ekstremne vrijednosti $\frac{-1}{q}$. Specijalni slučaj kada vrijedi $p = q = 1$ odgovara uniformnoj distribuciji na $(0,1)$. Tako je ovo prikladan model za stupanj gubitka. Transformacijom slučajne varijable *Pareto tipa* $Y$ pomoću $X = x_F - \frac{1}{Y}$, dobivamo slučajnu varijablu $X$ čija je funkcija distribucije gore spomenuta beta distribucija s indeksom ekstremne vrijednoti $-\xi$:
  $$1 - F(x) = \mathbb{P}(Y > \frac{1}{x_F - x}) = (x_F - x)^{\frac{1}{\xi}} L(\frac{1}{x_F - x}), \quad x < x_F,$$




Slika 3.1: Prikaz beta distribucija s različitim parametrima $\alpha$ i $\beta$

.




Slika 3.2: Usporedba beta distribucije ($\alpha = 2, \beta = 2$) s eksponencijalnom($\lambda = 1$)

.

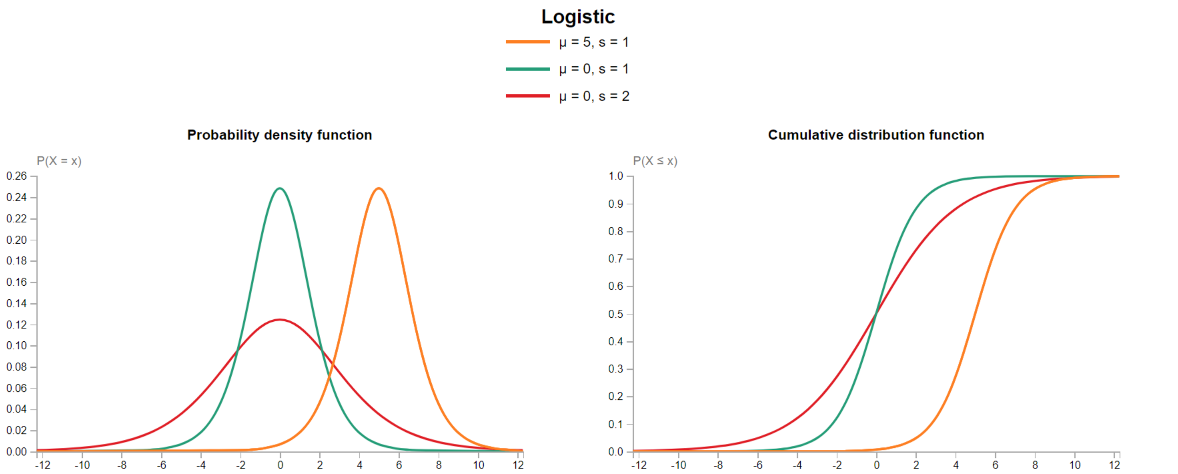
gdje je $L$ sporo varirajuća funkcija. Drugi način kako možemo doći do lakog repa s konačnom krajnjom točkom preko transformacije je uvjetovanjem na $W < T$ za neku vrijednost $T$,

$$X \stackrel{d}{=} W | W < T.$$

Ta operacija zove se rezanje odozgo, a za fiksan $T$ može se pokazati da je $X$ lakog repa s indeksom $\xi = -1$. Pri modeliranju velikih šteta prikladno je promatrati dovoljno veliki $T$, po mogućnosti jednak osiguranoj sumi.

- $\xi = 0$
  Weibullova distribucija razlikuje se od već spominjane Weibullove familije distribucija, a njena funkcija distribucije je $F(x) = 1 - e^{-(\frac{x}{\lambda})^k}$. Ona je lakog repa samo za

$k > 1$, jer $\limsup_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \limsup_{x\to\infty} e^{\mu x - (\frac{x}{\lambda})^k}$.
Specijalni slučaj Weibullove distribucije je *Rayleighjeva distribucija*, za koju vrijedi $k = 2$. Za Weibullovu distribuciju kvantilna funkcija iznosi $F^{\leftarrow}(t) = (\lambda^k \ln t)^{\frac{1}{k}}$.
Transformacijom eksponencijalne možemo doći i do *logističke distribucije*, kojoj je funkcija distribucija oblika $F(x) = 1 - \frac{1}{1+e^{\frac{x-\mu}{s}}}$, $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $s > 0$. Da se primijetiti kako disitribucija živi na skupu realnih brojeva, a mogu se promatrati i jednostrane verzije, od kojih je na primjer jedna oblika $F(x) = 1 - 2(1 + e^x)^{-1}$.

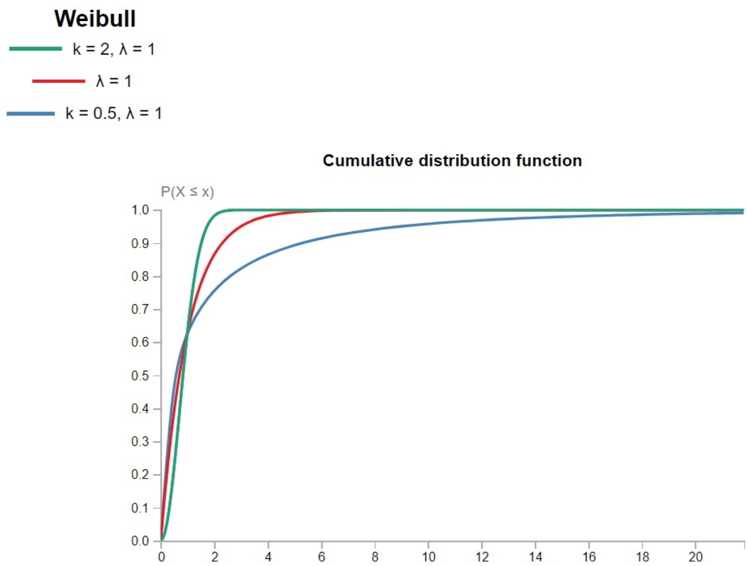



Slika 3.3: Logistička distribucija s različitom kombinacijom parametara

*Gama distribucija* također igra svoju ulogu u modeliranju šteta. Njena funkcija distribucije je $F(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\lambda x} e^{-u} u^{\alpha-1} du$. Za cjelobrojne vrijednosti $\alpha$, to odgovara distribuciji zbroja nezavisnih eksponencijalnih slučajnih varijabli (*Erlangova distribucija*), dok je za $\alpha = \frac{n}{2}$, $n \in \mathbb{Z}$, riječ o *hi kvadrat distribuciji*. Kod modeliranja šteta, normalna distribucija ne može biti optimalno rješenje zbog činjenice da su veličine štete uvijek pozitivne vrijednosti. Zato se opredjeljujemo na *jednostranu normalnu distribuciju*, kojoj je funkcija gustoće oblika $f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2}$, $x > 0$.

## 3.2 Teški repovi

- $\xi = 0$
  Gore već spomenuta Weibulova distribucija je teškog repa (a tada upada i u klasu subeksponencijalnih distribucija) za $0 < k < 1$.



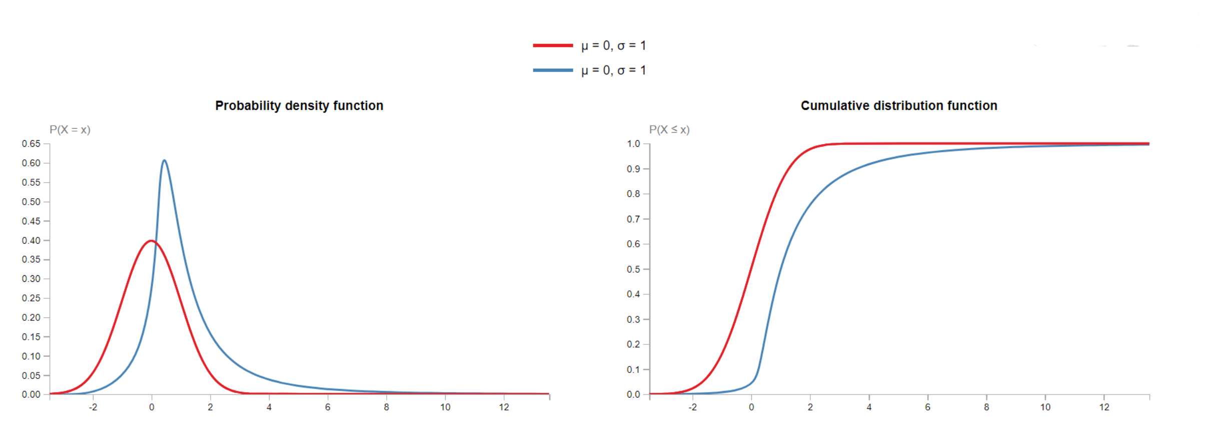

Slika 3.4: Usporedba dvaju Weibullovih funkcija distribucija(s parametrima $\lambda = 1$, $k = 2, 0.5$ s eksponencijalnom(parametar $\lambda = 1$).

Druga Benktanderova distribucija je još jedan primjer distribucije teškog repa s $\xi = 0$, a njena funkcija distribucija je oblika $F(x) = 1 - cax^{b-1}e^{\frac{-a}{b}x^b}$, $x > 0$, gdje su $a$ i $c$ pozitivne konstante, a $0 < b < 1$.
Iako normalna distribucija spada pod distribucije lakog repa, njene transformacije mogu biti dosta reprezentativni primjeri distribucija teškog repa. Jedna od njih je *log - normalna* distribucija, koja glasi

$$F(x) = \Phi(\frac{\ln(x) - \mu}{\sigma}),$$

gdje je $\mu \in \mathbb{R}$, $\sigma > 0$. Može se pokazati da ova distribucija pripada klasi subeksponencijalnih distribucija, a njen rep je teži od Weibullove. Svi momenti su konačni i dani su izrazom

$$\mathbb{E}(X^n) = e^{n\mu + \frac{n^2\sigma^2}{2}}.$$




Slika 3.5: Usporedba mormalne i lognormalne distribucije s parametrima $\mu = 0$, $\sigma = 1$.

*Benckert* je još 1962. godine predložio korištenje lognormalne distribucije za modeliranje podataka industrijskih i neindustrijskih požara, a Ferrara je distribuciju prilagodio stvarnim podacima o požarima. Još jedna od primjena ove distribucije su štete nastale olujama i štetama stakla.

- $\xi > 0$
  Najpoularnija distribucija teškog repa za modeliranje velikih šteta je takozvana *Paretova distribucija* i njene transformirane varijante. Najjednostavniji oblik je *stroga Paretova*, dana za $\alpha > 0$

$$F(x) = 1 - (\frac{x}{x_0})^{-\alpha},\ x > x_0 > 0.$$

Lako se vidi da je teškog repa iz činjenice da je $limsup_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \frac{x_0}{x}^{\alpha} = \infty$. Stroga Paretova distribucija također je subeksponencijalana za sve vrijednosti $\alpha > 0$. Osim toga, lako je moguće uspostaviti vezu između eksponencijancije distribucije i nje:

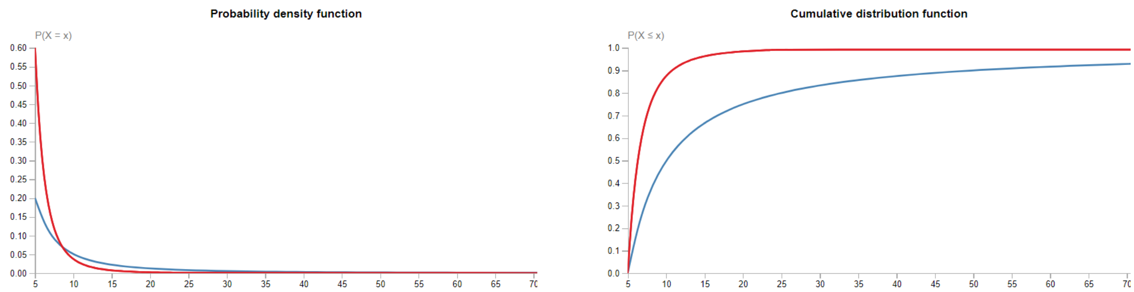
$$Y \sim Exp(\alpha) \Longleftrightarrow x_0 e^Y \sim Pareto(x_0, \alpha)$$

Činjenicu da je ovo distribucija definirana samo od pozitivne vrijednosti $x_0$ nadalje ne smatramo značajnim problemom jer se ionako uglavnom koristi za modeliranje jako velikih šteta.
Paretova distribucija ima još jedno bitno svojstvo zbog kojeg je bitna u primjeni, a to je *svojstvo zaboravljanja*. Za bilo koji $M > x_0$ vrijedi

$$\mathbb{P}(X > x|X > M) = \begin{cases} (\frac{x}{M})^{-\alpha} & x > M \\ 1 & inače, \end{cases}$$

to jest uvjetni višak je opet Pareto-distribuiran s parametrima $(\alpha, M)$ što nam je od posebnog interesa kod *XL* ugovora reosiguranja. Što se tiče momenata, vrijedi

$$\mathbb{E}(X^n) = \begin{cases} \frac{\alpha x_0^n}{\alpha - n} & n < \alpha \\ \infty & n \geq \alpha. \end{cases}$$




Slika 3.6: Stroga Paretova distribucija s parametrima $x_0 = 5$, $\alpha = 3$(crvena), $\alpha = 1$(plava).

Još jedna od varijacija Paretove distribucije je *pomaknuta Paretova distribucija*, određena s dva parametra

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

Njen nosač je cijeli pozitivni dio apscise (nema ograničenja na neki $x_0$). U aktuarskoj literaturi često se može pronaći pod nazivom *US-Paretova distribucija*.
Dodatno, uvodeći još jedan parametar $\tau$, imamo *Burrovu distribuciju*, čija je funkcija distribucije sljedeća:

$$F(x) = 1 - \beta^{\alpha}(\beta + x^{\tau})^{-\alpha} \quad x > 0,$$

dok joj je pripadna kvantilna funkcija repa jednaka

$$U(t) = \beta^{\frac{1}{\tau}}(t^{\frac{1}{t}} - 1)^{\frac{1}{\tau}}.$$

Kao i normalna, i gama distribucija nakon prikladne transformacije postaje teškog repa. Ističemo *log-gama distribuciju* koja je oblika

$$F(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\lambda \ln x} e^{-u} u^{\alpha - 1} du \quad x > 0.$$

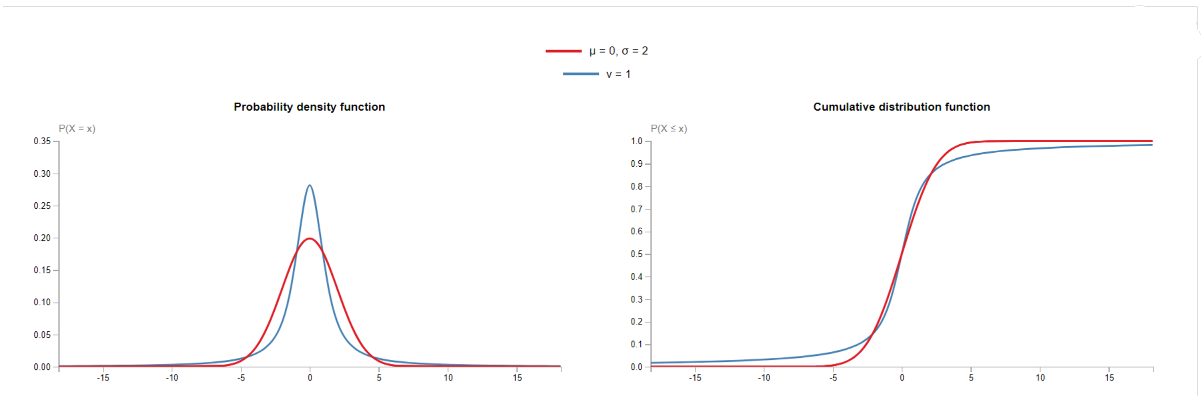
Ova distribucija pripada klasi distribucija Paretovog tipa s $\xi = \frac{1}{\lambda}$ i pomoćnom funkcijom $a(t) \sim \frac{1}{\lambda} U(t), t \to \infty$.

*T-distribucija* također nudi mogućnosti za modele velikih šteta.'Preklapanjem' dvostrane t-distribucije prema pozitivnoj polovici dobivamo familiju kandidata za distribucije velikih šteta koje nazivamo jednostranim t-distribucijama. Njihova gustoća dana je s

$$f(x) = \frac{2\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} \quad x > 0,$$

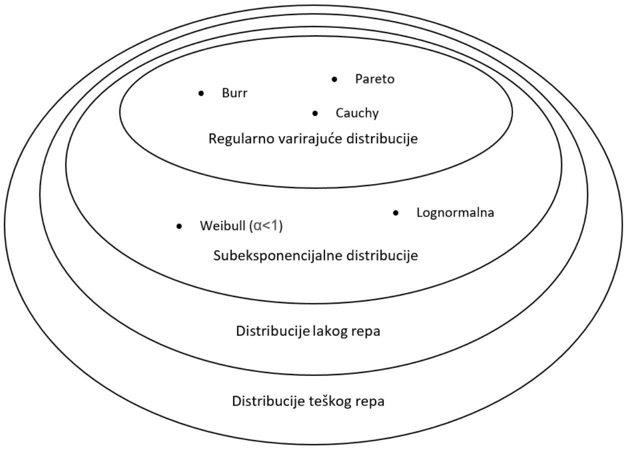
te također spadaju u klasu distribucija Paretovog tipa. I na ovu distribuciju moguće je primijeniti logaritamske i normalizirajuće transformacije, a jedna od rezultirajućih distribucija je i već poznata *Cauchyjeva distribucija* s funkcijom gustoće

$$f(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \quad x > 0.$$




Slika 3.7: Usporedba normalne distribucije s parametrima $\mu = 0$, $\sigma = 2$(crvena) i t-distribucije za $n = 1$.

Još neke popularne distribucije teškog repa su i *Wakebyjeva*, *log-Pearson III.*, *Fréchetova* i *prva Benktanderova distribucija*. Kompletno stanje navedenih distribucija i odnosi s otprije definiranim klasama dani su na Slici 3.8.




Slika 3.8: Distribucije u pripadnim klasama

# Bibliografija


[1] H. Albrecher, J. Beirlant, J. L.Teugels, Reinsurance: Actuarial and Statistical Aspects, John Wiley Sons Ltd, 2017.

[2] R. M. Cooke, D. Nieboer, Heavy-Tailed Distributions: Data, Diagnostics and New Developments, 2011. https://media.rff.org/archive/files/sharepoint/WorkImages/Download/RFF-DP-11-19.pdf,

[3] P. Embrechts C. Klüppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin, 1997.

[4] T Mikosch, Regular Variation Subexponentiality and Their Applications in Probability Theory, University of Groningen, https://www.eurandom.tue.nl/reports/1999/013-report.pdf 1999.

[5] J. Nair, A. Weirman, B. Zwart, The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation, https://adamwierman.com/wp-content/uploads/2021/05/book-05-11.pdf 2022.

[6] N. Sarapa. Teorija vjerojatnosti. Školska knjiga, 1986.

[7] T. Sekelj, Tehnike smanjenja rizika kod neživotnih osiguranja, završni rad na Poslijediplomskom specijalističkom studiju aktuarske matematike, Zagreb, lipanj 2018.

[8] Z. Vondraček, Slučajni procesi, Prirodoslovno-matematički fakultet, 2010.

# Sažetak

Cilj ovog rada bio je opisati modele za veličinu šteta u osiguranju i odgovarajuće vjerojatnosne distribucije, pri čemu najveću ulogu igraju distribucije teškog repa. U prvom dijelu govori se o konceptu reosiguranja, što je to i koje su situacije u kojima dolazi do potrebe za istim. Također, navode se osnovni oblici proporcionalnog i neproporcionalnog reosiguranja, kao i njihove prednosti i mane. Nastavak rada govori o različitim interpretacijama veličine štete pa u tom kontekstu definiramo neke klase distribucija kao što su regularno varirajuće, subeksponencijalne i distribucije dugog repa te njihova osnovna svojstva i primjere. Iskazan je i osnovni teorem teorije ekstremnih vrijednosti, takozvani Fisher-Tippett-Gnedenko teorem, koji karakterizira oblik granične distribucije za maksimume slučajnih varijabli i definira distribucije ekstremnih vrijednosti. Za sva tri slučaja izvedene su maksimalne domene privlačnosti te je definirana generalizirana Paretova distribucija koja igra veliku ulogu u teoriji ekstremnih vrijednosti. Za kraj, navedeni su primjeri najbitnijih distribucija koje se koriste u modeliranju šteta u reosiguranju s naglaskom na one teškog repa(Pareto, lognormalna, Weibull($\alpha < 1$)...) te vizualno prikazane njihove funkcije distribucija i gustoće u usporedbi s ostalima.

# Summary

The aim of this thesis was to describe models for claim sizes in insurance and appropriate probability distributions, with heavy tail distributions playing the largest role. The first chapter talks about the concept of reinsurance, what it is and what are the situations where it is necessary. Also, the basic forms of proportional and disproportionate reinsurance are listed, as well as their advantages and disadvantages. The sequel of the thesis discusses different approaches on what we might righteously call a large claim and how one can perhaps distinguish it from others, so in this context we define distribution classes such as regularly varying, subexponential and long tail distributions with their basic properties and examples. The so-called Fisher-Tippett-Gnedenko theorem, the basic result of the extreme-value theory is introduced, which characterizes the limit distribution for maximum of random variables. For all three cases the maximum domains of attraction are derived and the generalized Pareto distribution that plays a major role in the theory of extreme values is defined. Finally, examples of the most important distributions used in modelling reinsurance claims with an emphasis on one heavy tail are given (Pareto, lognormal, Weibull($\alpha < 1$)...), and their distribution and density functions are visually shown compared to other distributions.

# Životopis

Rođen sam 24. ožujka 1998. godine u Splitu i završio Osnovnu školu "Manuš" u Splitu 2012. godine. Nakon toga upisujem Prirodoslovno-matematičku gimnaziju u Splitu tokom čijeg sam pohađanjanja trenirao plivanje i aktivno sudjelovao na natjecanjima. Nakon mature upisujem preddiplomski studij Matematike na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu. Završetkom preddiplomskog studija, 2019. godine upisujem Diplomski studij Matematička statistika na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu.