Teorija slučajnih skupova

Ivanković, Daniela

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Daniela Ivanković

TEORIJA SLUčAJNIH SKUPOVA

Diplomski rad

Voditelj rada:
prof. dr. sc. Bojan Basrak

Zagreb, rujan 2021.

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. ____________________________

Mami, tati i sestri

Sadržaj

Uvod

Centralni pojam ovoga rada je *zatvoreni slučajni skup*. Povijesno, začeci njegova proučavanja motivirani su primjenama u ekonomiji. Zatvoreni slučajni skupovi pojavljuju se u modelima višestrukog izbora, igrama s višestrukim ekvilibrijima i povezani su s teorijom kooperativnih igara. Motiviran problemima iz područja morfologije, francuski matematičar Georges Matheron je 1975. godine u svojoj knjizi *Random sets and integral geometry* precizno definirao koncept zatvorenog slučajnog skupa korištenjem *topologije pogodaka i promašaja*. Matheronov rad je postao temelj daljnjeg istraživanja teorije i primjene zatvorenih slučajnih skupova.

Intuitivno, zatvoreni slučajni skup je prirodno poopćenje pojmova slučajne varijable i slučajnog vektora. Slučajni skup se također može promatrati kao neprecizna opservacija. Odnosno, umjesto proučavanja opažene vrijednosti x neke slučajne varijable X, pretpostavljamo da tu vrijednost nije moguće točno odrediti pa promatramo skup s koji gotovo sigurno sadrži opažanje x. Tada je s zapravo jedna realizacija slučajnog skupa S.

Novije primjene zatvorenih slučajnih skupova doprinijele su brzom razvoju teorije. Razvojem računala, zatvoreni slučajni skupovi počeli su se upotrebljavati u analizi slike te umjetnoj inteligenciji za pametno upravljanje i donošenje odluka. U statistici zatvorene slučajne skupove pronalazimo u pouzdanim intervalima i neparametarskim procjenama vjerojatnosnih funkcija gustoće. Značajna je i primjena u biostatistici i biomedicini za modeliranje rasta tumora.

Rad je strukturiran u tri poglavlja. U prvom poglavlju dajemo pregled osnovnih pojmova o vjerojatnosnom prostoru i slučajnim elementima koji će nam biti potrebni za razumijevanje novih koncepata. Predstavljeno je nekoliko primjera zatvorenih slučajnih skupova kako bi se dala motivacija za njihovo proučavanje. Teoriju slučajnih skupova uvodimo postepeno, prvo definirajući *konačne slučajne skupove* i analizirajući osnovne rezultate vezane uz njihove *funkcije distribucija* i *funkcije kapaciteta*. Prelaskom na zatvorene slučajne skupove definiramo *topologiju pogodaka i promašaja* te dajemo primjere zatvorenih slučajnih skupova. Definiramo funkciju kapaciteta te pokazujemo da ona jedinstveno određuje zakon razdiobe zatvorenog slučajnog skupa. Naposljetku, iskazujemo Choquet–Kendall–Matheronov teorem koji daje nužne i dovoljne uvjete da dana funkcija bude funkcija kapaciteta nekog zatvorenog slučajnog skupa.

U drugom dijelu rada definira se *Choquetov integral* potreban za integriranje neaditivnih skupovnih funkcija, kakve su funkcije kapaciteta zatvorenih slučajnih skupova. Takva integracija nam je potrebna kako bi definirali tri tipa konvergencije niza zatvorenih slučajnih skupova koji su zapravo generalizacija konvergencija slučajnih varijabli. Pokazuju se i ekvivalentni načini definiranja konvergencije te neki nužni i dovoljni uvjeti za konvergenciju.

U posljednjem poglavlju donosimo primjenu zatvorenih slučajnih skupova u proučavanju ekstremnih opažanja te pronalaženju adekvatnih modela za vrijednosti koje prelaze unaprijed određeni prag. Za pogodni odabir praga koristi se ME funkcija. Kada promatramo vrijednosti procjenitelja ME funkcije u opažanjima iz nekog slučajnog uzorka, dobivamo graf koji čini zatvoreni slučajni skup. Iskazujemo osnovni teorem konvergencije takvih zatvorenih skupova te na simuliranim i stvarnim podacima ilustriramo rezultat tog teorema.

Poglavlje 1

Slučajni skupovi

1.1 Vjerojatnosni prostor i slučajni elementi

Neka je $(\Omega, \mathcal{A}, \mathbb{P})$ vjerojatnosni prostor. *Slučajna varijabla* definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{A}, \mathbb{P})$ je preslikavanje $X\colon \Omega \to \mathbb{R}$ koje je $(\mathcal{A}, \mathcal{B}_{\mathbb{R}})$ – izmjerivo, odnosno za proizvoljni Borelov skup B je $X^{-1}(B) \in \mathcal{F}$. Definicija slučajne varijable može se poopćiti na izmjeriva preslikavanja koja poprimaju vrijednosti u $\mathbb{R}^n$, odnosno do *slučajnih vektora*. Po uzoru na slučajne varijable i vektore mogu se sasvim općenito definirati *slučajni elementi* na sljedeći način.

Definicija 1.1.1. *Neka je* $(\Omega, \mathcal{A}, \mathbb{P})$ *i* $(U, \mathcal{U})$ *izmjerivi prostor.* Slučajni element *je preslikavanje* $X\colon \Omega \to U$ *koje je izmjerivo u paru* σ*–algebri* $(\mathcal{A}, \mathcal{U})$*.*

Prilikom provođenja istraživanja, najčešće nije moguće ispitati cijelu populaciju i zato se odabire podskup populacije pomoću kojega ćemo donositi zaključke. Kako bi se moglo pretpostaviti da doneseni zaključci vrijede za cijelu populaciju, taj podskup mora biti reprezentativan. Stoga se odabire uzorak iz populacije U *na slučajan način*. To jest, provodi se slučajan eksperiment čiji ishod su podskupovi od U koji se zovu *slučajni skupovi*.

Slučajni skupovi se prirodno javljaju pri traženju pouzdanih intervala. Neka je $\{f(x, \theta) : x \in \mathcal{X} \subseteq \mathbb{R}^m, \theta \in \Theta \in \mathbb{R}^d\}$ parametrizirana familija funkcija gustoća. Neka je od interesa vrijednost statistike $\varphi(\theta)$ koja ovisi o nepoznatom parametru θ. Neka je $X_1, X_2, \ldots, X_n$ slučajan uzorak iz populacije s pravom vrijednošću parametra θ_0. Potrebno je pronaći slučajan skup $C(X_1, X_2, \ldots, X_n)$ koji sadrži $\varphi(\theta_0)$ s određenom vjerojatnošću.

Slično, ako je X n-dimenzionalni slučajni vektor s nepoznatom funkcijom gustoće f, tada nosač te funkcije gustoće, $S(f) = \{x \in \mathbb{R}^d : f(x) > 0\}$ također nije poznat. Ako se $S(f)$ želi procijeniti iz nekog slučajnog uzorka, procjenitelj je skup koji ovisi o tom slučajnom uzorku, dakle slučajni skup.

U gruboj analizi podataka nije moguće opaziti točne vrijednosti neke varijable već samo skupove unutar kojih se nalaze stvarne vrijednosti. Tada se opažaju realizacije slučajnog skupa S koje sadrže vrijednosti slučajne varijable od interesa X. Umjesto slučajnog uzorka $X_1, X_2, \ldots, X_n$ od X, promatra se i analizira slučajan uzorak $S_1, S_2, \ldots, S_n$ za slučajni skup S. Slučajan skup S naziva se poopćenje od X i vrijedi $\mathbb{P}(X \in S) = 1$.

Ovo su samo neki od primjera za koje je važno razviti teoriju o slučajnim skupovima i njihovim distribucijama. Slijedi formalna definicija *konačnih slučajnih skupova* te njihove glavne karakteristike i pridružene funkcije. Iz toga slijedi motivacija za definiciju *zatvorenog slučajnog skupa* te pojmova koje je potrebno definirati za razvoj teorije zatvorenih slučajnih skupova.

Definicija 1.1.2. Konačan slučajni skup *je preslikavanje* $X : \Omega \to 2^U$*, gdje je* 2^U *partitivni skup konačnog skupa* U*, takvo da je za svaki podskup* A *od* U*, skup* $X^{-1}(\{A\}) = \{\omega \in \Omega : X(\omega) = A\}$ $\mathcal{A}$*–izmjeriv.*

Napomena 1.1.3. *Slučajan skup* X *je slučajan element ako promatramo izmjeriv prostor* $(2^U, \mathcal{E})$*, gdje je* $\mathcal{E}$ *partitivni skup skupa* 2^U*. Odnosno,* $X^{-1}(\mathcal{E}) \subseteq \mathcal{A}$ *jer* $\forall \mathbb{A} \in \mathcal{E}$, $X^{-1}(\mathbb{A}) = \bigcup\limits_{A \in \mathbb{A}} X^{-1}(\{A\})$ *(a prebrojiva unija* $\mathcal{A}$*-izmjerivih skupova je ponovo* $\mathcal{A}$*-izmjeriv skup). Zato je uvjet izmjerivosti iz definicije nužan za definiranje zakona razdiobe od* X*,* $\mathbb{P}_X$*, na* $\mathcal{E}$ *pomoću vjerojatnosti na jednočlanim skupovima.*

Neka je $f : 2^U \to [0, 1]$ *definirana sa* $f(A) = \mathbb{P}(X = A)$*. Tada je* f vjerojatnosna funkcija gustoće *na* 2^U*, odnosno,* $f \geq 0$ *i* $\sum\limits_{A \subseteq U} f(A) = \mathbb{P}(X \in 2^U) = \mathbb{P}_X(2^U) = 1$*. Za* $\mathbb{A} \in \mathcal{E}$*,* $\mathbb{P}_X(\mathbb{A}) = \sum\limits_{A \in \mathbb{A}} f(A)$*.*

Funkcija distribucije konačnog slučajnog skupa X je funkcija $F : 2^U \to [0, 1]$ definirana sa $F(A) = \mathbb{P}(X \subseteq A)$. To je monotona funkcija jer za $A, B \subseteq U$, $A \subseteq B$ vrijedi

$$F(A) = \mathbb{P}(X \subseteq A) \leq \mathbb{P}(X \subseteq B) = F(B).$$

Funkcija F je i 2*-monotona*, odnosno za proizvoljne $A, B \subseteq U$ je

$$F(A \cup B) \geq F(A) + F(B) - F(A \cap B),$$

što slijedi iz skupovne jednakosti $A \cup B = \{A \setminus (A \cap B)\} \cup \{B \setminus (A \cap B)\} \cup \{A \cap B\}$ i monotonosti vjerojatnosti. Proizvoljna 2-monotona funkcija ne mora biti monotona jer za $A \subseteq B$ imamo

$$F(B) = F(A \cup (B \setminus A)) \geq F(A) + F(B \setminus A) - F(A \cap (B \setminus A)) = F(A) + F(B \setminus A) - F(\emptyset).$$

pa je za monotonost tada nužno da vrijedi $F(\emptyset) = 0$. Slučajan skup X za koji vrijedi $F(\emptyset) = \mathbb{P}(X = \emptyset) = 0$ zovemo *neprazan slučajni skup*. Slučajni skupovi koji se javljaju u primjenama većinom zadovoljavaju pretpostavku o nepraznosti. Ako je $f : 2^U \to [0, 1]$

funkcija gustoće konačnog slučajnog skupa, onda za funkciju distribucije tog slučajnog skupa F vrijedi

$$F(A) = \sum_{B\subseteq A} f(B), \quad \text{za } A \subseteq U. \tag{1.1}$$

Relacija $\mathbb{P}(X \subseteq A) = \mathbb{P}(X \cap A^c = \emptyset) = 1 - \mathbb{P}(X \cap A^c \neq \emptyset)$ motivira definiciju *funkcije kapaciteta* konačnog slučajnog skupa X na U:

$$T\colon 2^U \to [0,1], \quad T(A) = \mathbb{P}(X \cap A \neq \emptyset) = 1 - F(A^c).$$

Teorem 1.1.4. *Neka je X neprazan slučajan skup na konačnom skupu U i neka je $F\colon 2^U \to [0,1]$ njegova funkcija distribucije definirana sa $F(A) = \mathbb{P}(X \subseteq A)$. Tada F zadovoljava sljedeća svojstva:*

1. *$F(\emptyset) = 0$, $F(U) = 1$,*
2. *(k–monotonost): Za proizvoljan $k \geq 2$ i proizvoljne $A_1, A_2, \dots, A_k \in U$ je*

$$F\left(\bigcup_{j=1}^{k} A_j\right) \geq \sum_{\emptyset\neq I\subseteq\{1,2,\dots,k\}} (-1)^{|I|+1} F\left(\bigcap_{i\in I} A_i\right). \tag{1.2}$$

Dokaz.

1. Vrijedi $F(\emptyset) = \mathbb{P}(X \subseteq \emptyset) = \mathbb{P}(X = \emptyset) = 0$, gdje posljednja jednakost slijedi iz činjenice da je X neprazan slučajni skup. Također, $F(U) = \mathbb{P}(X \subseteq U) = 1$ jer je X slučajan skup definiran na skupu U.

2. Za $B \subseteq U$ neka je $J(B) = \{i \in \{1,2,\dots,k\} : B \subseteq A_i\}$. Vrijedi $B \subseteq \bigcap_{i\in J(B)} A_i$. Sada je lijeva strana u (1.2) jednaka

$$F\left(\bigcup_{j=1}^{k} A_j\right) \overset{(1.1)}{=} \sum_{B\subseteq \bigcup_{i=1}^{k} A_i} f(B) \geq \sum_{\substack{B\subseteq U,\\ J(B)\neq\emptyset}} f(B),$$

gdje posljednja nejednakost vrijedi zato što $J(B) \neq \emptyset$ povlači da postoji A_j takav da je $B \subseteq A_j$ pa je posebno i $B \subseteq \bigcup_{i=1}^{k} A_i$. Za svaki neprazan konačan skup S je $\sum_{B\subset S} (-1)^{|B|+1} = 1$ (jer strogih podskupova od S ima $2^{|S|}-1$) pa primjenom na $S = J(B)$ dobivamo

$$\sum_{\substack{B\subseteq U,\\ J(B)\neq\emptyset}} f(B) = \sum_{\substack{B\subseteq U,\\ J(B)\neq\emptyset}} \left[\sum_{\emptyset\neq I\subseteq J(B)} (-1)^{|I|+1}\right] f(B) = \sum_{\emptyset\neq I\subseteq U} (-1)^{|I|+1} \sum_{\substack{B\subseteq U,\\ J(B)\supseteq I}} f(B) =$$

$$= \sum_{\emptyset\neq I\subseteq\{1,2,\dots,k\}} (-1)^{|I|+1} \sum_{B\subseteq\bigcap_{i\in I} A_i} f(B) \overset{(1.1)}{=} \sum_{\emptyset\neq I\subseteq\{1,2,\dots,k\}} (-1)^{|I|+1} F\left(\bigcap_{i\in I} A_i\right).$$

□

Općenito, funkcija $F\colon 2^U \to [0,1]$ koja zadovoljava uvjete teorema 1.1.4 se zove *funkcija distribucije* na 2^U. Može se pokazati da funkcija distribucije jedinstveno definira neki slučajan skup na U (vidi [6]).

Skupovna funkcija $T\colon 2^U \to [0,1]$ je funkcija kapaciteta nekog konačnog slučajnog skupa na U ako zadovoljava

1. $T(\emptyset) = 0$ i $T(U) = 1$,

2. Za proizvoljan $k \geq 2$ i proizvoljne $A_1, A_2, \ldots, A_k \in U$ je

$$T\left(\bigcap_{j=1}^{k} A_j\right) \leq \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} T\left(\bigcup_{i \in I} A_i\right).$$

Iz relacije $T(A) = 1 - F(A^c)$ slijedi da i funkcija kapaciteta T također karakterizira zakon razdiobe konačnog slučajnog skupa X.

1.2 Zatvoreni slučajni skupovi

Zatvoreni slučajni skupovi su slučajni skupovi koji poprimaju vrijednosti u klasi zatvorenih podskupova od $\mathbb{R}^d$, odnosno općenitije, u klasi zatvorenih podskupova Hausdorffovog, lokalno kompaktnog topološkog prostora koji zadovoljava drugi aksiom prebrojivosti.

Slučajan vektor $X\colon \Omega \to \mathbb{R}^d$ može se promatrati kao slučajan skup $S : \Omega \to \mathcal{P}(\mathbb{R}^d)$ koji poprima samo vrijednosti jednočlanih skupova $\{X\}$. Jednočlani podskupovi od $\mathbb{R}^d$ su zatvoreni skupovi.

Za preciznu definiciju zatvorenih slučajnih skupova potrebna je dobro definirana σ-algebra na familiji svih zatvorenih podskupova od $\mathbb{R}^d$, $\mathcal{F} = \mathcal{F}(\mathbb{R}^d)$. Prirodni način konstrukcije σ-algebre $\sigma(\mathcal{F})$ na $\mathcal{F}$ je opskrbiti $\mathcal{F}$ nekom pogodnom topologijom i definirati $\sigma(\mathcal{F})$ kao Borelovu σ-algebru na tom topološkom prostoru.

Ako je U konačan skup, na U promatramo diskretnu topologiju u kojoj je svaki podskup od U otvoren skup, odnosno $\mathcal{U}(U) = \mathcal{P}(U)$. Tada je svaki konačan slučajni skup X ujedno i zatvoreni slučajan skup i njegov zakon razdiobe na σ-algebri $\mathcal{P}(2^U)$ na 2^U je jedinstveno određen funkcijom kapaciteta

$$T\colon 2^U \to [0,1], \quad T(A) = \mathbb{P}(X \cap A \neq \emptyset).$$

Vrijedi: $\{X \cap A \neq \emptyset\} = \{\omega : X(\omega) \cap A \neq \emptyset\} = X^{-1}(\mathcal{F}_A)$, gdje je $\mathcal{F}_A = \{B \subseteq U : B \cap A \neq \emptyset\}$. Dakle, za $U = \mathbb{R}^d$ i $X\colon \Omega \to \mathcal{F}$, podskupovi od $\mathcal{F}$ oblika $\mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \emptyset\}$ za proizvoljni $A \subseteq \mathbb{R}^d$ moraju pripadati σ-algebri $\sigma(\mathcal{F})$ generiranoj topologijom $\mathcal{T}$ na $\mathcal{F}$.

Tada i komplement tog skupa $\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \emptyset\}$ mora pripadati $\sigma(F)$. Skup $\mathcal{F}_A$ zovemo *skup pogodaka od A*, a skup $\mathcal{F}^A$ *skup promašaja od A*.

Neka je $\mathcal{F}$ familija zatvorenih podskupova od $\mathbb{R}^d$, $\mathcal{G}$ familija otvorenih podskupova od $\mathbb{R}^d$, a $\mathcal{K}$ familija kompaktnih podskupova od $\mathbb{R}^d$. Za bazu topologije $\mathcal{T}$ na $\mathcal{F}$ je moguće uzeti familiju skupova $\mathcal{F}_K$ za $K \in \mathcal{K}$. Topologija generirana tom bazom naziva se *topologija pogodaka i promašaja* od $\mathcal{F}$ i pripadna Borelova σ-algebra je $\mathcal{B}(\mathcal{F})$. Ekvivalentno, za bazu možemo uzeti i familiju skupova $\mathcal{F}_G$ za $G \in \mathcal{G}$. To vrijedi jer za svaki otvoreni skup $G \subseteq \mathbb{R}^d$ postoji rastući (u smislu inkluzije) niz kompaktnih skupova $(K_n)_{n\in\mathbb{N}}$ takvih da je $K_n \subset G$ za svaki $n \in \mathbb{N}$ te je $G = \bigcup_{n\in\mathbb{N}} K_n$ te također za svaki kompaktan skup $K \subseteq \mathbb{R}^d$ postoji padajući niz otvorenih skupova u $\mathbb{R}^d$, $(G_n)_{n\in\mathbb{N}}$ takvih da je $K = \bigcap_{n\in\mathbb{N}} G_n$. Tada je $\mathcal{F}_K = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{G_n}$ i $\mathcal{F}_G = \bigcup_{n\in\mathbb{N}} \mathcal{F}_{K_n}$.

Definicija 1.2.1. *Neka je* $(\Omega, \mathcal{A}, \mathbb{P})$ *vjerojatnosni prostor. Preslikavanje* $X\colon \Omega \to \mathcal{F}$ *je* zatvoren slučajni skup *ako je* X *izmjerivo u paru σ-algebri* $(\mathcal{A}, \mathcal{B}(\mathcal{F}))$*, odnosno ako je skup* $\{\omega : X(\omega) \cap K \neq \emptyset\}$ $\mathcal{A}$*-izmjeriv za svaki* $K \in \mathcal{K}$ *kompaktan podskup od* $\mathbb{R}^d$.

Ekvivalentno, $X\colon \Omega \to \mathcal{F}$ je zatvoren slučajan skup ako vrijedi jedno od sljedećeg:

1. $X^{-1}(\mathcal{F}_F) \in \mathcal{A}$, za svaki $F \in \mathcal{F}$,
2. $X^{-1}(\mathcal{F}_G) \in \mathcal{A}$, za svaki $G \in \mathcal{G}$,
3. $X^{-1}(\mathcal{F}_K) \in \mathcal{A}$, za svaki $K \in \mathcal{K}$.

Neka je $\mathcal{F}^A_{A_1,A_2,\dots,A_k} = \mathcal{F}^A \cap \mathcal{F}_{A_1} \cap \cdots \cap \mathcal{F}_{A_k}$. Vrijedi $\mathcal{F}_{A_1} \cup \mathcal{F}_{A_2} = \mathcal{F}_{A_1\cup A_2}$ jer

$$F \in \mathcal{F}_{A_1} \cup \mathcal{F}_{A_2} \Leftrightarrow F \cap A_1 \neq \emptyset \text{ ili } F \cap A_2 \neq \emptyset \Leftrightarrow F \cap (A_1 \cup A_2) \neq \emptyset \Leftrightarrow F \in \mathcal{F}_{A_1\cup A_2}.$$

Međutim, $\mathcal{F}_{A_1} \cap \mathcal{F}_{A_2}$ ne mora biti jednako $\mathcal{F}_{A_1\cap A_2}$. Primjerice, ako uzmemo $A_1 = \{1, 2, 3\}$, $A_2 = \{2, 3, 4\}$ i $X = \{1, 4\}$, tada je $X \in \mathcal{F}_{A_1} \cap \mathcal{F}_{A_2}$ jer $X \cap A_1 = \{1\} \neq \emptyset$ i $X \cap A_2 = \{4\} \neq \emptyset$, ali $X \notin \mathcal{F}_{A_1\cap A_2}$ jer $X \cap (A_1 \cap A_2) = X \cap \{2, 3\} = \emptyset$.

Primjer 1.2.2. *Primjeri zatvorenih slučajnih skupova:*

1. *Slučajni jednočlani skup je zatvoren slučajni skup. Neka je* $X\colon \Omega \to \mathcal{F}$, $X(\omega) = \{\xi(\omega)\}$. *Tada* X *zapravo svakom elementu* $\omega \in \Omega$ *pridruži jedan element* $\xi(\omega) \in \mathbb{R}^d$. *Tako definiran* X *doista zadovoljava uvjet iz definicije zatvorenog slučajnog skupa. Neka je* $K \in \mathcal{K}$ *proizvoljan. Tada je* $\{\omega : X(\omega) \cap K \neq \emptyset\} = \{\omega : \{\xi(\omega)\} \cap K \neq \emptyset\} = \{\omega : \xi(\omega) \in K\} = \{\xi \in K\}$ *što je* $\mathcal{A}$*-izmjeriv skup jer je* ξ *slučajna varijabla i* $K \in \mathcal{B}(\mathbb{B}^d)$.

   Analogno se pokaže da je konačan niz slučajnih vektora $Y = \{\xi_1, \xi_2, \dots, \xi_k\}$ *također zatvoren slučajan skup.*

2. *Slučajni segment* $X = [\xi, \eta]$ *generiran s dvije slučajne varijable* ξ *i* η *je zatvoren slučajan skup. Dovoljno je provjeriti da je za svaki otvoreni skup* $G \in \mathcal{G}$ *zadovoljeno da je* $X^{-1}(\mathcal{F}_G) \in \mathcal{A}$. *Budući da je ambijentni prostor* $\mathbb{R}$, *svaki otvoreni skup se može zapisati kao konačna ili prebrojiva unija disjunktnih otvorenih intervala, a* σ*-algebra* $\mathcal{A}$ *je zatvorena na prebrojive unije. Dakle, neka je* $G = \langle a, b\rangle$ *proizvoljan. Tada je*

$$\{X \cap G \neq \emptyset\} = \{[\xi, \eta] \cap \langle a, b\rangle \neq \emptyset\} = \\ \{\min\{\xi, \eta\} \leq a, \max\{\xi, \eta\} > a\} \cup \{\min\{\xi, \eta\} \in \langle a, b\rangle\}.$$

Oba ova skupa su $\mathcal{A}$*-izmjerivi jer su* ξ *i* η *slučajne varijable, pa je i njihova unija* $\mathcal{A}$*-izmjeriva.*

3. *Slučajni nivo skupovi. Neka je* $\varphi\colon \mathbb{R}^d \to [0, 1]$ *funkcija koja zadovoljava da je za svaki* $a \in \mathbb{R}$ *skup* $\{x \in \mathbb{R}^d : \varphi(x) \geq a\}$ *zatvoren podskup od* $\mathbb{R}^d$. *Neka je* $\alpha\colon \Omega \to [0, 1]$ *uniformno distribuirana slučajna varijabla na* $[0, 1]$. *Definiramo* $S : \Omega \to \mathcal{F}$, $S(\omega) = \{x \in \mathbb{R}^d : \varphi(x) \geq \alpha(\omega)\}$. *Tada je* S *zatvoreni slučajan skup na* $\mathbb{R}^d$.

4. *Neka su* $X, Y\colon (\Omega, \mathcal{A}) \to (\mathcal{F}, \mathcal{B}(\mathcal{F}))$ *zatvoreni slučajni skupovi. Tada je i njihov presjek* $X \cap Y\colon (\Omega, \mathcal{A}) \to (\mathcal{F}, \mathcal{B}(\mathcal{F}))$ *zatvoren slučajni skup. Treba pokazati da je* $X \cap Y$ *izmjeriv u paru* σ*-algebri* $(\mathcal{A}, \mathcal{B}(\mathcal{F}))$. *Budući da su* X *i* Y *zatvoreni slučajni skupovi, to znači da su* $X^{-1}(\mathcal{F}_\mathcal{K}) \in \mathcal{A}$ *te* $Y^{-1}(\mathcal{F}_\mathcal{K}) \in \mathcal{A}$ *za svaki* $K \in \mathcal{K}$. *Budući da je* $\mathcal{A}$ σ*-algebra, ona je zatvorena na presjeke pa je* $(X \cap Y)^{-1}(\mathcal{F}_\mathcal{K}) = X^{-1}(\mathcal{F}_\mathcal{K}) \cap Y^{-1}(\mathcal{F}_\mathcal{K}) \in \mathcal{A}$ *za svaki* $K \in \mathcal{K}$.

Primjer 1.2.3. *Neke slučajne varijable povezane sa zatvorenim slučajnim skupovima:*

1. *Ako je* $X\colon \Omega \to \mathcal{F} \subset \mathcal{P}(\mathbb{R}^d)$ *zatvoren slučajni skup, tada je za proizvoljan* $x \in \mathbb{R}^d$ *karakteristična funkcija* $\mathbf{1}_{x\in X}\colon \Omega \to \{0, 1\}$ *slučajna varijabla. Zaista,*

$$\{\mathbf{1}_{x\in X} = 1\} = \{x \in X\} = \{X \cap \{x\} \neq \emptyset\},$$

a posljednji skup je izmjeriv po definiciji zatvorenog slučajnog skupa (jer $\{x\}$ *je kompaktan skup u* $\mathbb{R}^d$*). Skup* $\{\mathbf{1}_{x\in X} = 0\}$ *je izmjeriv kao komplement izmjerivog skupa. Time je pokazano da je* $\mathbf{1}_{x\in X}$ *doista slučajna varijabla.*

2. *Norma zatvorenog slučajnog skupa* $X\colon \Omega \to \mathcal{F}$, *definirana kao supremum normi elemenata tog skupa,*

$$\|X\| = \sup\{\|x\| : x \in X\}$$

je slučajna varijabla koja poprima vrijednosti u $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$. *Neka je* $r > 0$ *proizvoljan. Potrebno je pokazati da je tada* $\{\|X\| < r\}$ *izmjeriv skup. Neka su* $K_n = \{x \in \mathbb{R}^d : r \leq \|x\| \leq n\}$ *kompaktni skupovi u* $\mathbb{R}^d$. *Tada vrijedi*

$$\{\|X\| < r\} = \{\|x\| < r, \forall x \in X\} = \{X \cap K_n = \emptyset,\ \textit{za } n \geq r\} = \bigcap_{n\geq r} \{X \cap K_n = \emptyset\},$$

što je izmjeriv skup kao presjek skupova $\{X \cap K_n = \emptyset\}$ koji su izmjerivi po deifniciji zatvorenog slučajnog skupa X.

3. *Neka je $X \colon \Omega \to \mathcal{F}$ zatvoren slučajan skup. Neka je $h_X \colon \mathbb{R}^d \to \overline{\mathbb{R}}$ njegova* funkcija nosača *definirana s*
$$h_X(u) = \sup_{x \in X} \langle x, u \rangle .$$
Tada je za fiksni $u \in \mathbb{R}^d$, $h_X \colon \Omega \to \overline{\mathbb{R}}$ slučajna varijabla.

4. *Za zatvoren slučajan skup $X \colon \Omega \to \mathcal{F}$ i metriku ρ na $\mathbb{R}^d$ udaljenost točke $a \in \mathbb{R}^d$ od X je dana s*
$$\rho(a, X) = \inf\{\rho(a, x) : x \in X\}.$$
Tada je za fiksni $a \in \mathbb{R}^d$, $\rho(a, X) \colon \Omega \to \mathbb{R}$ slučajna varijabla. Doista, za proizvoljni $r \in \mathbb{R}$
$$\{\rho(a, X) > r\} = \{X \cap \overline{K}(a, r) = \emptyset\},$$
gdje je $\overline{K}(a, r) = \{x \in \mathbb{R}^d : \rho(a, x) \leq r\}$ zatvorena kugla oko a radijusa r obzirom na metriku ρ. Skup $\{X \cap \overline{K}(a, r) = \emptyset\}$ je izmjeriv jer je X zatvoren slučajni skup, a $\overline{K}(a, r)$ kompaktan skup u $\mathbb{R}^d$.

5. *Neka je X zatvoreni slučajan skup na $\mathbb{R}^d$ i Λ_d Lebesgueova mjera na $\mathbb{B}(\mathbb{R}^d)$. Tada je $\Lambda_d(X)$ nenegativna slučajna varijabla. I općenitije, za μ lokalno konačnu mjeru na $\mathbb{B}(\mathbb{R}^d)$, $\mu(X)$ je slučajna varijabla.*

1.3 Funkcija kapaciteta

Zakon razdiobe slučajne varijable i slučajnog vektora je jedinstveno određen njihovom vjerojatnosnom funkcijom distribucije. Budući da su zatvoreni slučajni skupovi generalizacija slučajnih vektora, želimo na sličan način karakterizirati njihove zakone razdiobe.

Definicija 1.3.1. Funkcija kapaciteta *zatvorenog slučajnog skupa X je $T_X \colon \mathcal{K} \to \mathbb{R}$ definirana s*
$$T_X(K) = \mathbb{P}(X \cap K \neq \emptyset), \quad K \in \mathcal{K}.$$

Vrijedi $T_X(\emptyset) = 0$ jer $\emptyset \cap K = \emptyset$ za svaki $K \in \mathcal{K}$ pa je $\mathbb{P}(\emptyset \cap K \neq \emptyset) = 0$. Međutim,
$$T_X(\mathbb{R}^d) = \mathbb{P}(X \cap \mathbb{R}^d \neq \emptyset) = \mathbb{P}(X \neq \emptyset) \in [0, 1]$$
i ta vrijednost može biti strogo manja od 1.

Propozicija 1.3.2. *Funkcija kapaciteta jedinstveno određuje zakon razdiobe zatvorenog slučajnog skupa.*

Dokaz. Neka je $X \colon \Omega \to \mathcal{F}$ zatvoren slučajan skup i $\mathbb{P}_X$ njegov zakon razdiobe. σ-algebru $\mathcal{B}(\mathcal{F})$ generira π-sustav sastavljen od skupova oblika $\mathcal{F}^K_{K_1,K_2,\dots,K_m} = \mathcal{F}^K \cap \mathcal{F}_{K_1} \cap \cdots \cap \mathcal{F}_{K_m}$. Jedna inkluzija vrijedi jer je po definiciji $\mathcal{B}(\mathcal{F})$ generirana topologijom čija su baza skupovi oblika $\mathcal{F}_{K_1}$ pa u izrazu $\mathcal{F}^K_{K_1,K_2,\dots,K_m}$ uzmemo $m = 1$ i $K = \emptyset$. Obratna inkluzija vrijedi jer je $\mathcal{F}^K = (\mathcal{F}_K)^c$ i σ-algebra $\mathcal{B}(\mathcal{F})$ je zatvorena na presjeke. Dakle, dovoljno je pokazati da funkcija kapaciteta jedinstveno određuje vrijednosti $\mathbb{P}_X(\mathcal{F}^K_{K_1,K_2,\dots,K_m})$ za proizvoljne $m \in \mathbb{N}$ i $K, K_1, \dots, K_m \in \mathcal{K}$. Neka su $m = 1$ i $K, K_1 \in \mathcal{K}$ proizvoljni.

$$\begin{aligned}
\mathcal{F}^K_{K_1} = \mathcal{F}^K \cap \mathcal{F}_{K_1} &= \{F \in \mathcal{F} : F \cap K = \emptyset\} \cap \{F \in \mathcal{F} : F \cap K_1 \neq \emptyset\} \\
&= \{F \in \mathcal{F} : F \cap K = \emptyset \text{ i } F \cap K_1 \neq \emptyset\} \\
&= \{F \in \mathcal{F} : F \cap (K \cup K_1) \neq \emptyset\} \setminus \{F \in \mathcal{F} : F \cap K \neq \emptyset\}.
\end{aligned}$$

Tada je

$$\mathbb{P}_X(\mathcal{F}^K_{K_1}) = \mathbb{P}_X(\mathcal{F}_{K\cup K_1}) - \mathbb{P}_X(\mathcal{F}_K) = \mathbb{P}(X \cap (K \cup K_1) \neq \emptyset) - \mathbb{P}(X \cap K \neq \emptyset) = T_X(K \cup K_1) - T_X(K).$$

Neka su sada $m = 2$ i $K, K_1, K_2 \in \mathcal{K}$. Vrijedi

$$\begin{aligned}
\mathcal{F}^K_{K_1,K_2} &= \mathcal{F}^K \cap \mathcal{F}_{K_1} \cap \mathcal{F}_{K_2} \\
&= \{F \in \mathcal{F} : F \cap K = \emptyset\} \cap \{F \in \mathcal{F} : F \cap K_1 \neq \emptyset\} \cap \{F \in \mathcal{F} : F \cap K_2 \neq \emptyset\} \\
&= \{F \in \mathcal{F} : F \cap K_1 \neq \emptyset \text{ i } F \cap K = \emptyset\} \cap \{F \in \mathcal{F} : F \cap K = \emptyset \text{ i } F \cap K_2 = \emptyset\}^c \\
&= \left(\mathcal{F}_{K_1} \cap \mathcal{F}^K\right) \setminus \{F \in \mathcal{F} : F \cap (K \cup K_2) = \emptyset\} \\
&= \left(\mathcal{F}_{K_1} \cap \mathcal{F}^K\right) \setminus \mathcal{F}^{K \cup K_2} \\
&= \left(\mathcal{F}_{K_1} \cap \mathcal{F}^K\right) \setminus \left(\mathcal{F}_{K_1} \cap \mathcal{F}^{K \cup K_2}\right) \\
&= \mathcal{F}^K_{K_1} \setminus \mathcal{F}^{K \cup K_2}_{K_1},
\end{aligned}$$

pa je $\mathbb{P}_X(\mathcal{F}^K_{K_1,K_2}) = \mathbb{P}_X(\mathcal{F}^K_{K_1}) - \mathbb{P}_X(\mathcal{F}^{K\cup K_2}_{K_1})$, a za vjerojatnosti takvih skupova je pokazano da su jedinstveno određene funkcijom kapaciteta. Indukcijom se pokaže da su i vjerojatnosti oblika $\mathbb{P}_X(\mathcal{F}^K_{K_1,K_2,\dots,K_m})$ jedinstveno određene funkcijom kapaciteta. □

Propozicija 1.3.3. *Funkcija kapaciteta nepraznog zatvorenog slučajnog skupa X, T_X, je vjerojatnosna mjera ako i samo ako je X jednočlan slučajni skup.*

Dokaz. Neka je $X = \{\xi\}$ jednočlan zatvoren slučajan skup. Njegova funkcija kapaciteta tada glasi

$$T_X(K) = \mathbb{P}(\{\xi\} \cap K \neq \emptyset) = \mathbb{P}(\xi \in K) = F_\xi(K),$$

gdje je F_ξ funkcija distribucije slučajnog vektora ξ, dakle T_X je vjerojatnosna mjera.

Obrnuto, neka je X proizvoljan zatvoren slučajan skup i neka je njegova funkcija kapaciteta T_X vjerojatnosna mjera. To znači da je posebno T_X aditivna funkcija pa za proizvoljni $K \in \mathcal{K}$ vrijedi

$$\begin{aligned} 1 = \mathbb{P}(X \cap \mathbb{R}^d \neq \emptyset) = \mathbb{P}(X \cap (K \cup K^c) \neq \emptyset) = T_X(K \cup K^c) &\overset{\text{adit.}}{=} \\ = T_X(K) + T_X(K^c) = \mathbb{P}(X \cap K \neq \emptyset) + \mathbb{P}(X \cap K^c \neq \emptyset) &= \\ = \mathbb{P}(X \cap K \neq \emptyset) + 1 - \mathbb{P}(X \cap K^c = \emptyset) = \mathbb{P}(X \cap K \neq \emptyset) + 1 - \mathbb{P}(X \subseteq K)&, \end{aligned}$$

iz čega slijedi da je $\mathbb{P}(X \cap K \neq \emptyset) = \mathbb{P}(X \subseteq K)$. Sada odaberimo K jednočlan skup za koji je $\mathbb{P}(X \cap K \neq \emptyset) > 0$ (takav K sigurno postoji jer je X neprazan). Tada vrijedi da je $\mathbb{P}(X \subseteq K) > 0$ samo ako je X jednočlan skup. □

Primjer 1.3.4. *Neka je $X = \langle -\infty, \xi]$ slučajan polupravac. Tada je*

$$T_X(K) = \mathbb{P}(\langle -\infty, \xi] \cap K \neq \emptyset) = \mathbb{P}(\inf K \leq \xi) = 1 - F_\xi(\inf K-),$$

gdje je $F_\xi(\inf K-)$ lijevi limes funkcije distribucije slučajne varijable ξ u točki $\inf K$.

Primjer 1.3.5. *Neka je X zatvoren slučajan skup čija funkcija kapaciteta T_X poprima samo vrijednosti* 0 *ili* 1*, odnosno $\forall K \in \mathcal{K}, \mathbb{P}(X \cap K \neq \emptyset) \in \{0, 1\}$. Prema [5] tada postoji $S \subseteq \mathbb{R}^d$ takav da je $X = S$, $\mathbb{P}$–g.s. Takav zatvoren slučajni skup X se zove* deterministički.

Primjer 1.3.6. *Neka je $f \colon \mathbb{R}^d \to [0, 1]$ neprekidna zdesna. Neka je zadana funkcija kapaciteta $T \colon \mathcal{K} \to \mathbb{R}$,*

$$T(K) = \sup_{x \in K} f(x).$$

Tada je zatvoren slučajan skup X čija je funkcija kapaciteta $T_X = T$ dan sa $X = \{x \in \mathbb{R}^d : f(x) \geq U\}$ gdje je U slučajna varijabla s uniformnom distribucijom na segmentu $[0, 1]$ *jer*

$$T_X(K) = \mathbb{P}(\{x \in \mathbb{R}^d : f(x) \geq U\} \cap K \neq \emptyset) = \mathbb{P}(\{x \in K : f(x) \geq U\} \neq \emptyset) = \\ \mathbb{P}(U \leq \sup_{x \in K} f(x)) \overset{U \sim U(0,1)}{=} \sup_{x \in K} f(x).$$

Neka je T funkcija kapaciteta nekog zatvorenog slučajnog skupa X na $\mathbb{R}^d$. Tada je T definirana na familiji kompaktnih skupova $\mathcal{K}$. Funkcija T se može proširiti do T^* definirane na cijelom $\mathcal{P}(\mathbb{R}^d)$. Prvo proširimo T na familiju svih otvorenih skupova u $\mathbb{R}^d$ sa

$$T^*(G) = \sup\{T(K) : K \subset G\} \tag{1.3}$$

i zatim za proizoljan $A \subseteq \mathbb{R}^d$ definiramo

$$T^*(A) = \inf\{T(G) : A \subset G\}. \tag{1.4}$$

Ovakvo proširenje je konzistentno jer vrijedi $T^*(K) = T(K)$ za proizvoljan kompaktan skup $K \in \mathcal{K}$.

1.4 Choquetov teorem

Želimo definirati svojstva funkcije T definirane na familiji kompaktnih skupova koja će osigurati da postoji zatvoren slučajan skup X kojem će funkcija T biti funkcija kapaciteta.

Teorem 1.4.1 (Choquet-Kendall-Matheron)**.** *Funkcija $T : \mathcal{K} \to [0,1]$ definirana na familiji kompaktnih skupova je funkcija kapaciteta zatvorenog slučajnog skupa X ako i samo ako vrijedi*

1. *T je neprekidna zdesna, odnosno za svaki padajući niz kompaktnih skupova $(K_n)_{n\in\mathbb{N}}$ koji konvergira prema kompaktnom skupu K, $K_n \searrow K$, vrijedi da $T(K_n) \searrow T(K)$,*
2. *T je potpuno alternirajuća funkcija, odnosno za proizvoljni $n \in \mathbb{N}$ i kompaktne skupove $K, K_1, \dots, K_n \in \mathcal{K}$ vrijedi da su sljedeći izrazi nepozitivni*

$$\begin{aligned} \Delta_{K_1} T(K) &= T(K) - T(K \cup K_1), \\ \Delta_{K_n} \cdots \Delta_{K_1} T(K) &= \Delta_{K_{n-1}} \cdots \Delta_{K_1} T(K) - \Delta_{K_{n-1}} \cdots \Delta_{K_1} T(K \cup K_n), \quad n \geq 2. \end{aligned}$$

Vjerojatnosni dokaz teorema 1.4.1 baziran na teoriji proširenja mjere s algebre na σ-algebru se može pronaći u [3].

Promotrimo razlike koje se javljaju u Choquetovom teoremu.

$$\begin{aligned} -\Delta_{K_1} T(K) = T(K \cup K_1) - T(K) &= \mathbb{P}(X \cap K = \emptyset, X \cap K_1 \neq \emptyset) \\ -\Delta_{K_n} \cdots \Delta_{K_1} T(K) &= \mathbb{P}(X \cap K = \emptyset, X \cap K_1 \neq \emptyset, \dots, X \cap K_n \neq \emptyset) \end{aligned}$$

pa je uvjet nepozitivnosti ekvivalentan uvjetu nenegativnosti odgovarajućih vjerojatnosti. Nepozitivnost razlike kada je $n = 1$ zapravo znači monotonost od T jer za $K_1 \subseteq K_2$ je

$$0 \geq T(K_1) - T(K_1 \cup K_2) = T(K_1) - T(K_2) \Longrightarrow T(K_2) \geq T(K_1).$$

Kada je $n = 2$ iz uvjeta nepozitivnosti dobivamo poopćenje svojstva konkavnosti funkcija.

$$\begin{aligned} 0 \geq \Delta_{K_2}\Delta_{K_1} T(K) &= \Delta_{K_1} T(K) - \Delta_{K_1} T(K \cup K_2) \\ &= T(K) - T(K \cup K_1) - T(K \cup K_2) + T(K \cup K_2 \cup K_1) \\ \Longrightarrow T(K) + T(K \cup K_2 \cup K_1) &\leq T(K \cup K_1) + T(K \cup K_2) \end{aligned}$$

pa ako uvrstimo $K = K_1 \cap K_2$ dobivamo

$$\begin{aligned} T(K_1 \cap K_2) + T((K_1 \cap K_2) \cup K_2 \cup K_1) &\leq T((K_1 \cap K_2) \cup K_1) + T((K_1 \cap K_2) \cup K_2) \\ &\Leftrightarrow T(K_1 \cap K_2) + T(K_2 \cup K_1) \leq T(K_1) + T(K_2). \end{aligned}$$

Choquetov teorem identificira funkcije kapaciteta zatvorenih slučajnih skupova kao neprekidne zdesna, potpuno alternirajuće funkcije definirane na familiji kompaktnih skupova s vrijednostima u $[0, 1]$ koje u praznom skupu poprimaju vrijednost 0. Ako proširimo skup mogućih vrijednosti koje funkcija poprima na $[0, +\infty\rangle$ tada, uz nepromijenjene ostale uvjete, dobivamo karakterizaciju svih lokalno konačnih mjera na $\mathcal{F}$, odnosno mjera na $\mathcal{F}$ koje poprimaju konačne vrijednosti na $\mathcal{F}_K$ za svaki $K \in \mathcal{K}$.

Za jedinstveno određivanje zakona razdiobe zatvorenog slučajnog skupa potrebno je definirati funkciju kapaciteta na svim kompaktnim skupovima K. Poželjno je što više reducirati familiju kompaktnih skupova $\mathcal{K}$ na kojoj je potrebno definirati funkciju kapaciteta kako bi i dalje imali osiguranu jedinstvenost. Jedna mogućnost je promatrati familiju konačnih unija zatvorenih krugova s centrima i radijusima koji su racionalni brojevi. Međutim, takva familija i dalje sadrži previše elemenata te se tražena razdioba može jedinstveno karakterizirati na puno manjoj familiji kompaktnih skupova.

Poglavlje 2

Konvergencije slučajnih skupova

2.1 Choquetov integral

Kao što je kod razvoja teorije slučajnih vektora jednostavnije rukovati funkcijama distribucija i korisno što više teorijskih rezultata iskazati preko njih, tako je pri proučavanju svojstava zatvorenih slučajnih skupova korisno koristiti pripadne funkcije kapaciteta. Budući da funkcija kapaciteta T zatvorenog slučajnog skupa X nije nužno aditivna skupovna funkcija kao što je bila funkcija distribucije, potrebno je definirati koncept integrala s obzirom na neaditivne skupovne funkcije (po uzoru na Lebesgueov integral s obzirom na aditivne skupovne funkcije, odnosno mjere). Takav integral nam je potreban, recimo, prilikom analize konvergencije po distribuciji zatvorenih slučajnih skupova, koju ćemo definirati kasnije u poglavlju. U sljedećem primjeru dana je jedna motivacija za definiranje integrala po neaditivnoj skupovnoj funkciji.

Primjer 2.1.1. *Za nenegativnu slučajnu varijablu X definiranu na vjerojatnosnom prostoru $(\Omega, \mathcal{A}, \mathbb{P})$, očekivanje od X možemo računati kao*

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)\, d\mathbb{P}(\omega) = \int_{0}^{+\infty} \mathbb{P}(X > t)\, dt = \int_{0}^{+\infty} \mathbb{P}(X \geq t)\, dt. \tag{2.1}$$

Neka je X proizvoljna slučajna varijabla od interesa. Pretpostavimo da u statističkoj analizi podataka nije moguće opažati sam X već neki skup S koji sadrži X gotovo sigurno. Dakle, $S : \Omega \to \mathcal{F}$ je zatvoreni slučajni skup na $\mathbb{R}$ takav da je $\mathbb{P}(X \in S) = 1$. Neka je $g\colon \mathbb{R} \to \mathbb{R}_+$ izmjeriva funkcija. Tada je

$$g_*(\omega) = \inf\{g(x) : x \in S(\omega)\} \leq g(X(\omega)) \leq \sup\{g(x) : x \in S(\omega)\} = g^*(\omega).$$

$g_*, g^*\colon \Omega \to \mathbb{R}$ su izmjerive funkcije jer za proizvoljan $t \in \mathbb{R}$ je

$$\{\omega : g_*(\omega) \geq t\} = \{\omega : \inf_{x \in S(\omega)} g(x) \geq t\} = \{\omega : g(x) \geq t, x \in S(\omega)\} = \{g(S) \geq t\}$$

što je izmjeriv skup jer su i S i g izmjeriva preslikavanja pa je i njihova kompozicija izmjerivo preslikavanje. Analogno se pokaže da je i g^* izmjeriva funkcija. Ta dva preslikavanja su i nenegativna pa možemo promatrati njihova očekivanja. Zbog monotonosti očekivanja vrijedi

$$\mathbb{E}(g_*) \leq \mathbb{E}(g(X)) \leq \mathbb{E}(g^*).$$

Neka su $F_*, F^* \colon \mathcal{B}(\mathbb{R}) \to [0,1]$ definirani s

$$F_*(B) = \mathbb{P}(S \subseteq B), \quad F^*(B) = \mathbb{P}(S \cap B \neq \emptyset).$$

Sada je

$$\begin{aligned}\mathbb{E}(g_*) &= \int_\Omega g_*(\omega)\, d\mathbb{P}(\omega) = \int_0^{+\infty} \mathbb{P}(g_* > t)\, dt = \int_0^{+\infty} \mathbb{P}(g_*^{-1}(\langle t, +\infty\rangle))\, dt = \\ &= \int_0^{+\infty} \mathbb{P}(S \subseteq g^{-1}(\langle t, +\infty\rangle))\, dt = \int_0^{+\infty} F_*(g^{-1}(\langle t, +\infty\rangle))\, dt = \int_0^{+\infty} F_*(x : g(x) > t)\, dt.\end{aligned}$$

I analogno

$$\begin{aligned}\mathbb{E}(g^*) &= \int_\Omega g^*(\omega)\, d\mathbb{P}(\omega) = \int_0^{+\infty} \mathbb{P}(g^* > t)\, dt = \int_0^{+\infty} \mathbb{P}(g^{*-1}(\langle t, +\infty\rangle))\, dt = \\ &= \int_0^{+\infty} \mathbb{P}(S \cap g^{-1}(\langle t, +\infty\rangle)) \neq \emptyset)\, dt = \int_0^{+\infty} F^*(g^{-1}(\langle t, +\infty\rangle))\, dt = \int_0^{+\infty} F^*(x : g(x) > t)\, dt.\end{aligned}$$

Preslikavanja F_* i F^* nisu vjerojatnosne mjere na $\mathcal{B}(\mathbb{R})$, ali se mogu koristiti za donošenje zaključaka o distribuciji slučajne varijable X na temelju zatvorenog slučajnog skupa S. Zato je potrebno definirati Choquetov integral *koji će omogućiti takve račune.*

Neka je $\mathbb{E}$ proizvoljan skup te neka je ϕ neaditivna mjera definirana na nekoj familiji podskupova $\mathcal{A}$ od $\mathbb{E}$ takvoj da je $\emptyset \in \mathcal{A}$. Odnosno, neka za ϕ vrijedi $\phi(\emptyset) = 0$ i ϕ je neopadajuća, odnosno za $A, B \in \mathcal{A}$ takve da je $A \subseteq B$ je $\phi(A) \leq \phi(B)$. Recimo, ϕ može biti funkcija kapaciteta nekog slučajnog skupa. Po uzoru na jednakost u (2.1) se definira Choquetov integral.

Definicija 2.1.2. *Neka je $f \colon \mathbb{E} \to \mathbb{R}_+$ i ϕ neaditivna mjera na $\mathbb{E}$.* Choquetov integral nenegativne funkcije f u odnosu na ϕ *se definira kao*

$$C_\phi(f) = \int f\, d\phi = \int_0^{+\infty} \phi(\{x : f(x) \geq t\})\, dt = \int_0^{+\infty} \phi(f \geq t)\, dt. \tag{2.2}$$

Integral u (2.2) je dobro definiran i postoji ako je primjerice f ograničena ili izmjeriva te ϕ konačna mjera, odnosno $\phi(\mathbb{E}) < +\infty$. Neka je sada $f\colon \mathbb{E} \to \mathbb{R}$ proizvoljna izmjeriva funkcija. Tada f možemo zapisati kao $f = f^+ - f^-$, gdje su $f^+, f^-\colon \mathbb{E} \to \mathbb{R}_+$ nenegativne funkcije koje zovemo pozitivni i negativni dio od f i koje se definiraju kao $f^+(x) = f(x)\mathbf{1}_{f\geq 0}(x)$ i $f^-(x) = -f(x)\mathbf{1}_{f<0}(x)$. Tada definiramo *Choquetov integral funkcije f u odnosu na ϕ* kao

$$C_\phi(f) = \int f\,d\phi = \int_0^{+\infty} \phi(f \geq t)\,dt + \int_{-\infty}^{0} (\phi(f \geq t) - \phi(\mathbb{E}))\,dt.$$

Motivacija za ovakvu definiciju slijedi iz Lebesgueovog integrala koji se za proizvoljnu izmjerivu funkciju f definira na sljedeći način

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu, \tag{2.3}$$

kada su oba integrala s desne strane u (2.3) konačna. Ako je $\mu(\mathbb{R}) < +\infty$, onda imamo

$$\int f^+\,d\mu = \int_0^{+\infty} \mu(f^+ \geq t)\,dt \overset{t>0}{=} \int_0^{+\infty} \mu(f \geq t)\,dt. \tag{2.4}$$

Za svaki $t \in \mathbb{E}$, $t > 0$, skupovi $\{f < -t\} = \{f^- > t\}$ i $\{f \geq -t\}$ tvore particiju skupa $\mathbb{E}$, pa zbog aditivnosti Lebesgueove mjere μ vrijedi $\mu(f^- > t) = \mu(\mathbb{E}) - \mu(f \geq -t)$. Iz aditivnosti Lebesgueovog integrala slijedi

$$\begin{aligned}\int f^-\,d\mu = \int_0^{+\infty} \mu(f^- > t)\,dt = \int_0^{+\infty} [\mu(\mathbb{E}) - \mu(f \geq -t)]\,dt = \\ \int_0^{-\infty} -[\mu(\mathbb{E}) - \mu(f \geq t)]\,dt = \int_{-\infty}^{0} [\mu(\mathbb{E}) - \mu(f \geq t)]\,dt,\end{aligned} \tag{2.5}$$

pa uz (2.4) i (2.5) izraz (2.3) postaje

$$\int f\,d\mu = \int_0^{+\infty} \mu(f \geq t)\,dt + \int_{-\infty}^{0} [\mu(f \geq t) - \mu(\mathbb{E})]\,dt,$$

što je analogon definicije Choquetovog integrala za proizvoljnu izmjerivu funkciju. Sada slijedi da ako je f izmjeriva i ako je ϕ mjera, tada Choquetov integral odgovara očekivanju.

Za proizvoljan $M \subset \mathbb{E}$ možemo restringirati Choquetov integral na M na sljedeći način

$$C_\phi(\mathbf{1}_M f) = \int_M f\, d\phi = \int_0^{+\infty} \phi(\{x \in M : f(x) \geq t\})\, dt =$$

$$= \int_0^{+\infty} \phi(\{f \geq t\} \cap M)\, dt + \int_{-\infty}^{0} [\phi(\{f \geq t\} \cap M) - \phi(M)]\, dt.$$

Primjer 2.1.3. *Neka je T_X funkcija kapaciteta nekog zatvorenog slučajnog skupa X na $\mathbb{R}^d$ i neka je $g\colon \mathbb{R}^d \to \mathbb{R}_+$ izmjeriva.*

1. *Funkcija T_X se proširuje na $\mathcal{P}(\mathbb{R}^d)$ pomoću formula* (1.3) *i* (1.4) *pa je posebno T_X dobro definirana na $\mathcal{B}(\mathbb{R}^d)$ zbog čega se može promatrati integral*

$$\int_0^{+\infty} T_X(x \in \mathbb{R}^d : g(x) \geq t)\, dt.$$

Tada je

$$\int g\, dT_X \overset{g \geq 0}{=} \int_0^{+\infty} T_X(g \geq t)\, dt = \int_0^{+\infty} \mathbb{P}(X \cap (g \geq t) \neq 0)\, dt =$$

$$= \int_0^{+\infty} \mathbb{P}(\sup_{x \in X} f(x) \geq t)\, dt = \mathbb{E}[\sup_{x \in X} f(x)] = \mathbb{E}[\sup f(X)].$$

2. *Neka je F_X dualna funkcija od T_X, odnosno*

$$F(A) = 1 - T(A^c) = 1 - \mathbb{P}(X \cap A^c \neq \emptyset) = \mathbb{P}(X \cap A^c = \emptyset) = \mathbb{P}(X \subseteq A).$$

Tada je

$$\int_0^{+\infty} F(g \geq t)\, dt = \int_0^{+\infty} [1 - T(g < t)]\, dt = \int_0^{+\infty} [1 - \mathbb{P}(X \cap \{g < t\} \neq \emptyset)]\, dt =$$

$$= \int_0^{+\infty} \mathbb{P}(X \cap \{g < t\} = \emptyset)\, dt = \int_0^{+\infty} \mathbb{P}(\{x \in X : g(x) < t\} = \emptyset)\, dt =$$

$$= \int_0^{+\infty} \mathbb{P}(\forall x \in X, g(x) \geq t)\, dt = \int_0^{+\infty} \mathbb{P}(\inf_{x \in X} g(x) \geq t)\, dt = \mathbb{E}[\inf_{x \in X} g(x)] = \mathbb{E}[\inf g(X)].$$

Choquetov integral je monoton i homogen je pri množenju s pozitivnim skalarom pa za $f \leq g$ i $\lambda > 0$, zato što je ϕ monotona i invarijantna na množenje pozitivnim skalarom kao neaditivna mjera, vrijedi:

$$C_\phi(f) = \int_0^{+\infty} \phi(f \geq t)\, dt + \int_{-\infty}^{0} (\phi(f \geq t) - \phi(\mathbb{E}))\, dt \leq$$
$$\leq \int_0^{+\infty} \phi(g \geq t)\, dt + \int_{-\infty}^{0} (\phi(g \geq t) - \phi(\mathbb{E}))\, dt = C_\phi(g),$$

$$C_\phi(\lambda f) = \int_0^{+\infty} \phi(\lambda f \geq t)\, dt + \int_{-\infty}^{0} (\phi(\lambda f \geq t) - \phi(\lambda\mathbb{E}))\, dt =$$
$$= \lambda \int_0^{+\infty} \phi(f \geq t)\, dt + \lambda \int_{-\infty}^{0} (\phi(f \geq t) - \phi(\mathbb{E}))\, dt = \lambda C_\phi(f).$$

Međutim, Choquetov integral nije općenito aditivan. Primjerice, neka su $f = \frac{1}{4}\mathbf{1}_A$ i $g = \frac{1}{2}\mathbf{1}_B$. Tada je

$$C_\phi(f) = \int_0^{+\infty} \phi\left(x \in A : \frac{1}{4} \geq t\right) dt = \int_0^{\frac{1}{4}} \phi(A)\, dt = \frac{1}{4}\phi(A)$$
$$C_\phi(g) = \int_0^{+\infty} \phi\left(x \in B : \frac{1}{2} \geq t\right) dt = \int_0^{\frac{1}{2}} \phi(B)\, dt = \frac{1}{2}\phi(B)$$
$$C_\phi(f+g) = \int_0^{+\infty} \phi\left(x : \frac{1}{4}\mathbf{1}_A(x) + \frac{1}{2}\mathbf{1}_B(x) \geq t\right) dt = \frac{1}{4}\phi(A \setminus B) + \frac{1}{2}\phi(B \setminus A) + \frac{3}{4}\phi(A \cap B).$$

Zbog toga što ϕ nije aditivna, za A i B koji nisu disjunktni može biti $C_\phi(f+g) \neq C_\phi(f) + C_\phi(g)$.

2.2 Konvergencije nizova zatvorenih slučajnih skupova

Neka je $(\Omega, \mathcal{A}, \mathbb{P})$ vjerojatnosni prostor na kojem je definiran niz zatvorenih slučajnih skupova $(X_n)_{n\in\mathbb{N}}$ te zatvoreni slučajni skup X.

Konvergencija gotovo sigurno

Definicija 2.2.1. *Niz zatvorenih slučajnih skupova* $(X_n)_{n\in\mathbb{N}}$ konvergira gotovo sigurno *prema zatvorenom slučajnom skupu X ako je*

$$\mathbb{P}(\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)) = 1.$$

Oznaka: $X_n \xrightarrow{g.s.} X$.

Konvergenciju skupova $(X_n(\omega))_{n\in\mathbb{N}}$ iz definicije promatramo kao konvergenciju u smislu topologije pogodaka i promašaja $\mathcal{T}$. To jest, niz skupova $(X_n(\omega))_{n\in\mathbb{N}}$ konvergira prema skupu $X(\omega)$ ako za svaku okolinu V od $X(\omega)$ u $\mathcal{T}$, postoji $n(V) \in \mathbb{N}$ takav da je za svaki $n \geq n(V)$, $X_n(\omega) \in V$.

Konvergencija po vjerojatnosti

Neka je $d\colon \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ metrika na $\mathcal{F}$ kompatibilna s topologijom $\mathcal{T}$. Tada niz zatvorenih slučajnih skupova $(X_n)_{n\in\mathbb{N}}$ *konvergira po vjerojatnosti* prema zatvorenom slučajnom skupu X, u oznaci $X_n \xrightarrow{\mathbb{P}} X$, ako vrijedi

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0.$$

Ovaj tip konvergencije se za zatvorene slučajne skupove može opisati na ekvivalentan, ali operativniji način.

Prisjetimo se da za niz slučajnih varijabli $(Y_n)_{n\in\mathbb{N}}$ vrijedi da konvergencija tog niza gotovo sigurno prema slučajnoj varijabli Y povlači konvergenciju po vjerojatnosti, odnosno da za svaki $\varepsilon > 0$ vrijedi $\lim\limits_{n\to+\infty} \mathbb{P}(|Y_n - Y| > \varepsilon) = 0$. Sada na $\mathbb{R}$ vrijedi

$$\begin{aligned}\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\} &= \{\omega : Y(\omega) > Y_n(\omega) + \varepsilon\} \cup \{\omega : Y_n(\omega) > Y(\omega) + \varepsilon\} = \\ &= \{\omega : Y(\omega) \notin \overline{K}(Y_n(\omega), \varepsilon)\} \cup \{\omega : Y_n(\omega) \notin \overline{K}(Y(\omega), \varepsilon)\},\end{aligned}$$

gdje je za $x \in \mathbb{R}$, $\overline{K}(x, r) = \{y \in \mathbb{R} : |x - y| \leq r\}$. Sada promatramo metrički prostor $(\mathbb{R}^d, \rho)$ i na njemu definirane zatvorene slučajne skupove $X, (X_n)_{n\in\mathbb{N}}$. Za $A \subseteq \mathbb{R}^d$ i $\varepsilon > 0$ označimo s

$$A^\varepsilon = \{x \in \mathbb{R}^d : \rho(x, A) \leq \varepsilon\},$$

gdje je $\rho(x, A) = \inf\limits_{y\in A} \rho(x, y)$. Promatramo skupove oblika

$$E_{n,\varepsilon} = (X \setminus X_n^\varepsilon) \cup (X_n \setminus X^\varepsilon).$$

Može se pokazati (vidi [7]) da je ekvivalentno

1. $X_n \xrightarrow{\text{g.s}} X$,
2. Za svaki $\varepsilon > 0$, $E_{n,\varepsilon} \xrightarrow{\text{g.s}} \emptyset$,
3. Za svaki kompaktan skup $K \in \mathcal{K}$, $E_{n,\varepsilon} \cap K = \emptyset$ g.s, za dovoljno velike n.

Štoviše, ako $E_{n,\varepsilon} \xrightarrow{\text{g.s}} \emptyset$, tada za svaki $K \in \mathcal{K}$ vrijedi da je $\mathbb{P}(E_{n,\varepsilon} \cap K \neq \emptyset) \xrightarrow{n \to +\infty} 0$. Iz toga slijedi da je sljedeća definicija konvergencije po vjerojatnosti ekvivalentna početnoj:

Definicija 2.2.2. *Niz zatvorenih slučajnih skupova* $(X_n)_{n\in\mathbb{N}}$ konvergira po vjerojatnosti *prema zatvorenom slučajnom skupu* X *ako za svaki* $\varepsilon > 0$ *i za svaki kompaktni skup* $K \in \mathcal{K}$ *vrijedi*

$$\lim_{n\to+\infty} \mathbb{P}([(X \setminus X_n^{\varepsilon}) \cup (X_n \setminus X^{\varepsilon})] \cap K \neq \emptyset) = 0.$$

Konvergencija po distribuciji

Neka su $(T_n)_{n\in\mathbb{N}}$ i T funkcije kapaciteta niza zatvorenih slučajnih skupova $(X_n)_{n\in\mathbb{N}}$ i zatvorenog slučajnog skupa X, respektivno. Označimo s $(\mathbb{P}_n)_{n\in\mathbb{N}}$ i $\mathbb{P}_X$ zakone razdioba zatvorenih slučajnih skupova $(X_n)_{n\in\mathbb{N}}$ i X redom. Generalizirajući poznatu definiciju konvergencije po distribuciji slučajnih varijabli i vektora može se definirati da niz zatvorenih slučajnih skupova $(X_n)_{n\in\mathbb{N}}$ *konvergira po distribuciji* prema zatvorenom slučajnom skupu X, u oznaci $X_n \xrightarrow{D} X$, ako za svaki $A \in \mathcal{B}(\mathcal{F})$ takav da je $\mathbb{P}_X(\partial A) = 0$ vrijedi da $\mathbb{P}_n(A) \longrightarrow \mathbb{P}_X(A)$. Pri čemu je ∂A rub skupa A u odnosu na topologiju pogodaka i promašaja na $\mathcal{F}$. Iz *Portmanteau teorema* tada slijedi da je konvergencija po distribuciji ekvivalentna bilo kojoj od sljedećih tvrdnji

1. Za proizvoljnu realnu, neprekidnu, ograničenu funkciju f definiranu na $\mathcal{F}$ vrijedi $\int_{\mathcal{F}} f(F)\, d\mathbb{P}_n(F) \longrightarrow \int_{\mathcal{F}} f(F)\, d\mathbb{P}_X(F)$,
2. Za proizvoljnu realnu, neprekidnu, ograničenu funkciju f definiranu na $\mathcal{F}$ vrijedi $\mathbb{E}\,[f(X_n)] \longrightarrow \mathbb{E}\,[f(X)]$,
3. Za proizvoljan otvoren skup $G \in \mathcal{G}$ vrijedi $\mathbb{P}_X(G) \leq \liminf_{n\to\infty} \mathbb{P}_n(G)$.

Konvergencija mjera na $\mathcal{B}(\mathcal{F})$ iz 1 naziva se *slaba konvergencija* vjerojatnosnih mjera. Dakle, kaže se da niz $(\mathbb{P}_n)_{n\in\mathbb{N}}$ konvergira slabo prema $\mathbb{P}_X$, u oznaci $\mathbb{P}_n \xrightarrow{w} \mathbb{P}_X$, ako $X_n \xrightarrow{D} X$. Tada govorimo i o *slaboj konvergenciji* niza zatvorenih slučajnih skupova. Ponekad je jednostavnije rukovati s funkcijama kapaciteta zatvorenih slučajnih skupova pa je korisno iskazati da je definicija slabe konvergencije ekvivalentna uvjetu

$$\int_{\mathbb{R}^d} f(u)\, dT_n(u) \longrightarrow \int_{\mathbb{R}^d} f(u)\, dT(u) \tag{2.6}$$

za proizvoljnu nenegativnu, neprekidnu funkciju f s kompaktnim nosačem na $\mathbb{R}^d$. Za limese u (2.6) potreban nam je Choquetov integral s obzirom na to da funkcije kapaciteta nisu aditivne. Dokaz ove tvrdnje može se pronaći u [6].

Želimo konvergenciju po distribuciji niza zatvorenih slučajnih skupova karakterizirati pomoću njihovih funkcija kapaciteta s kojima je lakše računati. Prostor $\mathcal{F}$ je kompaktan u topologiji pogodaka i promašaja pa vrijedi varijanta Hellyjevog teorema.

Teorem 2.2.3 (Hellyjev teorem za zatvorene slučajne skupove)**.** *Proizvoljna familija zatvorenih slučajnih skupova je* relativno slabo kompaktna*, odnosno svaki niz zatvorenih slučajnih skupova ima slabo konvergentan podniz.*

Iz Hellyjevog teorema slijedi glavni rezultat, dokazan u [5]:

Teorem 2.2.4. *Niz zatvorenih slučajnih skupova $(X_n)_{n\in\mathbb{N}}$ konvergira slabo prema zatvorenom slučajnom skupu X ako i samo ako $T_n(K) \longrightarrow T(K)$ za svaki kompaktan skup K za koji vrijedi $T(K) = T(\mathrm{Int}(K))$, gdje je $\mathrm{Int}(K)$ interior skupa K.*

Poglavlje 3

Zatvoreni slučajni skupovi u primjeni

3.1 Konvergencija niza ME plotova

ME plotovi (eng. *mean excess plots*) su alat koji se često koristi u analizi ekstremnih vrijednosti, aktuarskoj matematici i procjeni rizika. Primjerice, u hidrologiji, može nas zanimati kolika je vjerojatnost da razina vode u rijeci ili moru bude viša od brane ili nasipa koji štite okolno područje od poplave. U znanosti o okolišu važno je znati granične razine koncentracija štetnih tvari u zraku ili vodi te vjerojatnosti da razina onečišćenja bude veća od dozvoljene. To omogućava donošenje potrebnih mjera kako bi se očuvala kvaliteta života ljudi u blizini. Matematički, ono što nas zanima je distribucija prekoračenja preko dopuštene vrijednosti, odnosno praga (eng. *threshold*), u za neku slučajnu varijablu X čija je funkcija distribucije F:

$$F_u(x) = \mathbb{P}(X - u \leq x \mid X > u).$$

Ova distribucija se pokazuje važnom pri analizi ekstremnih vrijednosti pomoću POT modela (eng. *peaks over threshold*) koje se bazira na analizi vrijednosti koje prelaze neki unaprijed određeni prag te pronalaženju modela koji će dobro opisivati te ekstremne vrijednosti. Iskazat ćemo teorem koji omogućava da se u praksi koristi familija generaliziranih Pareto razdioba najprikladnija za opisivanje statističkih svojstava ekstrema.

Kaže se da slučajna varijabla X ima *generaliziranu Pareto razdiobu s parametrima ξ i β*, odnosno $X \sim GPD(\xi, \beta)$, ako je njena funkcija distribucije oblika

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

pri čemu su $\beta > 0$ te $x \geq 0$ ako je $\xi \geq 0$ i $0 \leq x \leq -\frac{\beta}{\xi}$ za $\xi < 0$. Generalizirana Pareto razdioba se često koristi za modeliranje repova drugih razdiobi.

Pickands–Balkema–De Haan teorem [4] kaže da je funkcija distribucije ekstrema F_u, uz određene blage uvjete, asimptotski ekvivalentna funkciji distribucije $G_{\xi,\beta}$ generalizirane Pareto razdiobe kako se prag u približava desnoj krajnjoj točki x^F distribucije F koja se definira kao

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

Uvjeti na funkciju F_u su dani u teoremu 3.1.1 za slučaj $\xi > 0$, dok su za ostale vrijednosti ξ uvjeti vrlo slični. Ovaj teorem pokazuje da je generalizirana Pareto razdioba prirodan izbor za modeliranje ekstremnih vrijednosti jer dobrim odabirom praga u možemo postići dovoljno dobru aproksimaciju funkcije distribucije ekstrema F_u funkcijom distribucije generalizirane Pareto razdiobe. Za pogodni odabir praga u te za ocjenu kvalitete modela dobivenog pomoću generalizirane Pareto razdiobe koristi se ME funkcija (eng. *mean excess function*). *ME funkcija slučajne varijable* X definira se kao

$$M(u) = \mathbb{E}\left[X - u \mid X > u\right],$$

uz pretpostavku da je $\mathbb{E}[X_+] < \infty$. Ako je $X_1, X_2, \ldots, X_n$ slučajan uzorak iz razdiobe s funkcijom distribucije F, tada je procjenitelj ME funkcije dan s

$$\hat{M}(u) = \frac{\sum_{i=1}^{n} (X_i - u)\mathbf{1}_{\{X_i > u\}}}{\sum_{i=1}^{n} \mathbf{1}_{\{X_i > u\}}}, \quad u \geq 0.$$

Može se pokazati [4, 1] da za slučajnu varijablu $X \sim GPD(\xi,\beta)$ vrijedi da je $\mathbb{E}[X] < \infty$ ako i samo ako je $\xi < 1$ i u tom slučaju je ME funkcija od X linearna po u i oblika

$$M(u) = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi}u,$$

za $0 \leq u < \infty$ ako je $0 \leq \xi < 1$, odnosno $0 \leq u \leq -\frac{\beta}{\xi}$ ako je $\xi < 0$. Također, linearnost funkcije $M(u)$ karakterizira klasu generaliziranih Paretovih razdioba. Dakle, jedan način provjere podliježu li neki podaci generaliziranom Pareto modelu je promatrati njihov ME plot $\left\{\left(X_{(k)}, \hat{M}(X_{(k)})\right) : k = 1, \ldots, n\right\}$ i koliko je on blizak grafu neke linearne funkcije. Ovdje je $X_{(1)}, \ldots, X_{(n)}$ uređajna statistika uzorka $X_1, \ldots, X_n$ u nerastućem poretku. ME plot je konačan skup slučajnih točaka u $\mathbb{R}^2$ što ga čini izmjerivim i slučajnim zatvorenim skupom. Promatrat ćemo konvergenciju ME plotova u topologiji pogodaka i promašaja, definiranoj u odjeljku Zatvoreni slučajni skupovi.

Neka je $\bar{F}(x) = 1 - F(x)$. Za funkciju $U\colon \langle 0, \infty\rangle \to \mathbb{R}_+$ kažemo da je regularno varirajuća s indeksom ρ, odnosno $U \in RV_\rho$, ako je

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

U [2] je pokazano da vrijede sljedeća dva teorema.

Teorem 3.1.1. *Neka je $\xi > 0$. Za funkciju distribucije F je ekvivalentno:*

1. *$\bar{F} \in RV_{-\frac{1}{\xi}}$, odnosno*
$$\forall t > 0, \quad \lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\frac{1}{\xi}}.$$

2. *F je u maksimalnoj domeni privlačnosti Frechetove razdiobe s parametrom $\frac{1}{\xi}$, odnosno*
$$\lim_{n\to\infty} F^n(c_n x) = \exp\left\{-x^{\frac{1}{\xi}}\right\}, \quad \text{za svaki } x > 0,$$
gdje je $c_n = F^{-1}(1 - n^{-1})$.

3. *Postoji pozitivna izmjeriva funkcija $\beta(u)$ takva da vrijedi*
$$\lim_{n\to\infty} \sup_{x\geq u} \left|F_u(x) - G_{\xi,\beta(u)}(x)\right| = 0.$$

Teorem 3.1.1 daje nužan i dovoljan uvjet za egzistenciju funkcije $\beta(u)$ te posljedično za egzistenciju generalizirane Paretove razdiobe čija funkcija distribucije dovoljno dobro aproksimira funkciju distribucije ekstrema podataka F_u.

Teorem 3.1.2. *Neka je $(X_n)_{n\in\mathbb{N}}$ niz nezavisnih, jednakodistribuiranih slučajnih varijabli s funkcijom distribucije F te neka su $\bar{F} \in RV_{-\frac{1}{\xi}}$ i $0 < \xi < 1$. Tada u familiji zatvorenih skupova $\mathcal{F}$, s topologijom pogodaka i promašaja, vrijedi*

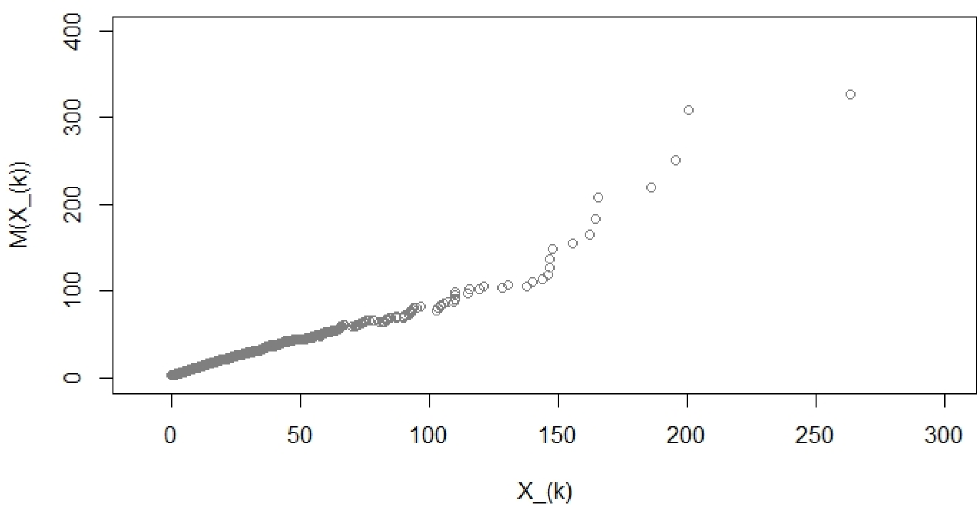
$$\mathcal{S}_k := \frac{1}{X_{(k)}} \left\{\left(X_{(i)}, \hat{M}(X_{(i)})\right) : i = 2, \ldots, k\right\} \xrightarrow{\mathbb{P}} \mathcal{S} := \left\{\left(t, \frac{\xi}{1-\xi}t\right) : t \geq 1\right\}. \tag{3.1}$$

Sada lako možemo naći procjenu $\hat{\xi}$ parametra ξ generalizirane Pareto razdiobe. Metodom najmanjih kvadrata ili metodom maksimuma funkcije vjerodostojnosti linearni dio ME plota aproksimiramo afinom funkcijom s koeficijentom smjera $\hat{b}$. Tada je $\hat{\xi} = \frac{\hat{b}}{1+\hat{b}}$. Slični rezultati konvergencije ME plotova u familiji zatvorenih skupova $\mathcal{F}$ vrijede i za slučajeve kada je $\xi < 0$ (tada je limes pravac s negativnim koeficijentom smjera) i $\xi = 0$ (limes je pravac paralelan s x-osi, s ordinatom jednakom 1). U slučaju $\xi > 1$ limes je slučajna krivulja. Rezultati se mogu pronaći u [2].

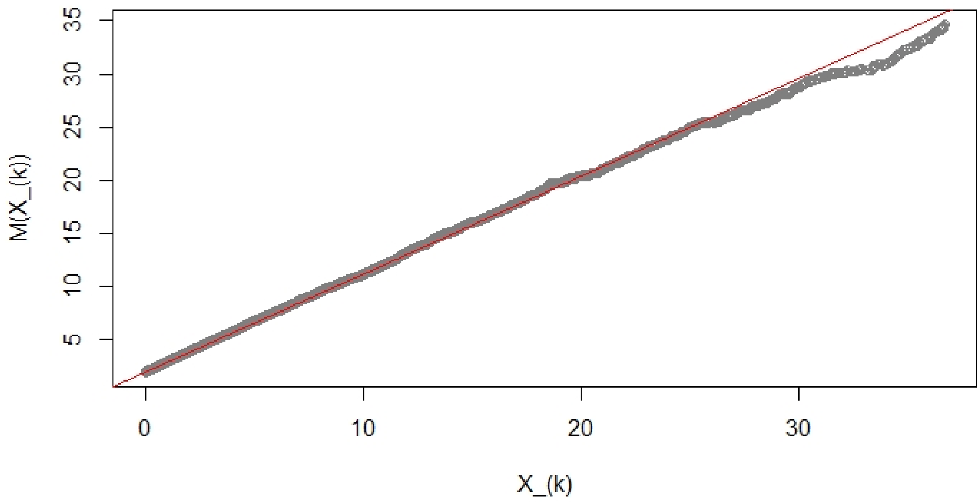
Ilustrirajmo teorem 3.1.2 o konvergenciji ME plotova na simuliranim te na stvarnim skupovima podataka.

Generalizirana Pareto razdioba

Simulirajmo u programskom jeziku R slučajni uzorak duljine $n = 100000$ iz generalizirane Pareto distribucije s parametrima $\xi = 0.5$ i $\beta = 1$ te pokažimo da je ME plot tih podataka približno linearan s koeficijentom smjera $\frac{\xi}{1-\xi} = \frac{0.5}{1-0.5} = 1$. Na slici 3.1 vidimo ME plot slučajnog uzorka duljine 100000 iz $GPD(0.5, 1)$ razdiobe uređene u nerastućem poretku te vrijednosti procjenitelja $\hat{M}$ ME funkcije u tim točkama. Može se vidjeti da je početni dio grafa približno linearan, dok s većim vrijednostima opažanja slučajne varijable X graf postaje sve više raspršen. Razlog tomu je što pri računanju vrijednosti empirijske ME funkcije u velikim vrijednostima X računamo aritmetičku sredinu malog broja elemenata (samo nekoliko opažanja je veće od trenutnog opažanja koje promatramo) pa je tada veća varijabilnost dobivenih vrijednosti. Na slici 3.2 vidimo prikaz dijela ME plota gdje smo odbacili početnih, odnosno najvećih, 250 opažanja. Sada smo metodom najmanjih kvadrata tim podacima pridružili pravac (označen crvenom bojom). Koeficijent smjera tog pravca je 0.92 (dok je očekivana vrijednost koeficijenta smjera 1), iz čega dobivamo da je procijenjena vrijednost parametra ξ generalizirane Pareto distribucije jednaka $\hat{\xi} = 0.48$, pri čemu je stvarna vrijednost jednaka 0.5.



Slika 3.1: ME plot slučajnog uzorka duljine 100000 iz $GPD(0.5, 1)$ razdiobe



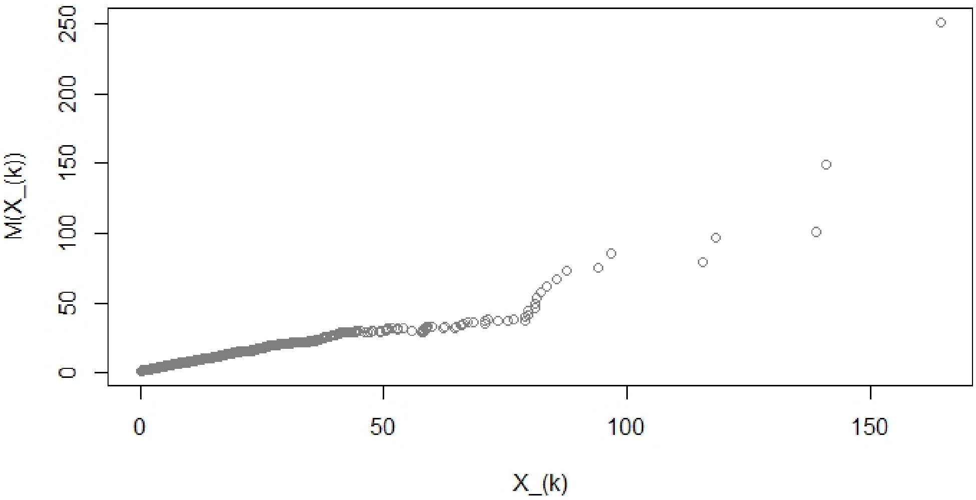
Slika 3.2: Dio ME plota slučajnog uzorka $X_{(250)}, \ldots, X_{(100000)}$ iz $GPD(0.5, 1)$ razdiobe te pravac prilagođen tim podacima metodom najmanjih kvadrata

F razdioba

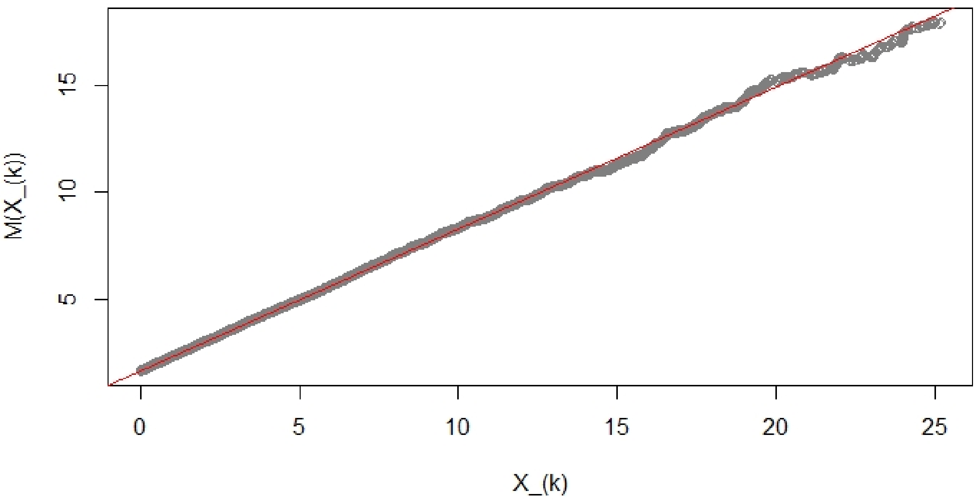
Ilustrirajmo rezultat teorema na F razdiobi koja nema linearan ME plot u teoriji. Ponovno simulirajmo uzorak od $n = 100000$ slučajnih brojeva iz $F(2, 5)$ razdiobe. Na slici 3.3 ponovno vidimo da je početak grafa sličan linearnoj funkciji, dok vrijednosti procijenjene ME funkcije za velike vrijednosti varijable X imaju veću varijabilnost. Stoga, prije aproksimacije linearnom funkcijom ponovno iz prikaza uklonimo 250 najvećih opažanja. Na slici 3.2 vidimo koliko dobro pravac s koeficijentom smjera 0.66 opisuje generirane podatke.

ME plot stvarnih podataka

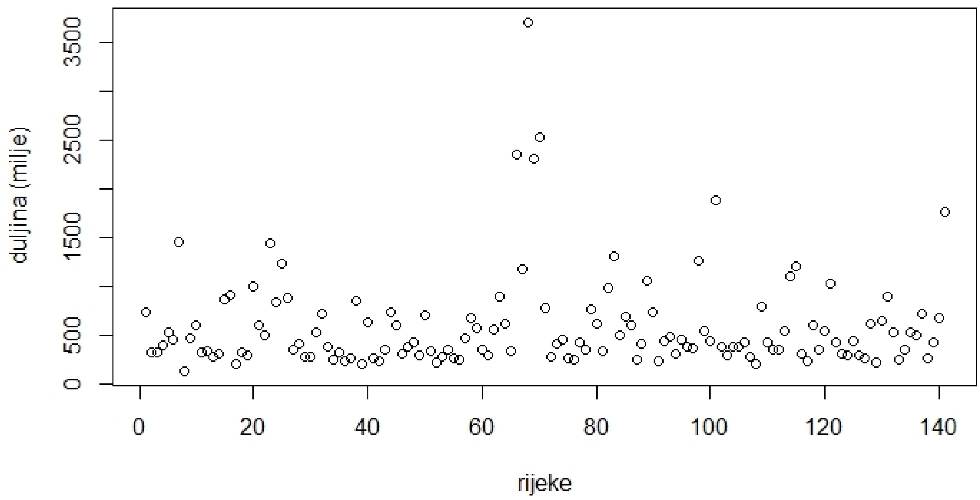
Konačno, pokažimo kako se teorem 3.1.2 ponaša na stvarnim podacima. Koristimo skup podataka *rivers* koji se može pronaći u programskom jeziku R. U njemu su zapisane duljine, u miljama, 141 najveće rijeke Sjeverne Amerike. Podaci su prikazani na slici 3.5, dok na slici 3.6 vidimo pravac prilagođen tim podacima.



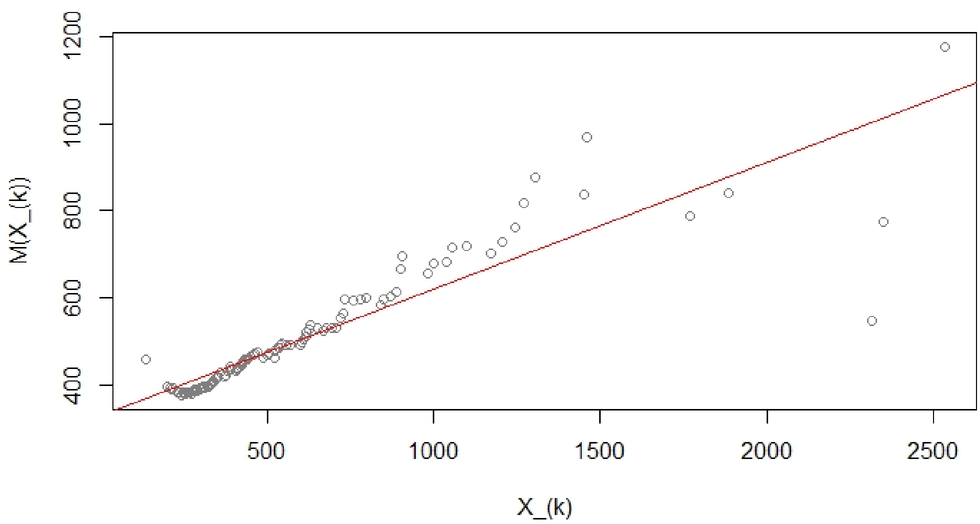
Slika 3.3: ME plot slučajnog uzorka duljine 100000 iz $F(2,5)$ razdiobe



Slika 3.4: Dio ME plota slučajnog uzorka $X_{(250)}, \ldots, X_{(100000)}$ iz $F(2,5)$ razdiobe te pravac prilagođen tim podacima metodom najmanjih kvadrata



Slika 3.5: Duljine (u miljama) 141 velike rijeke u Sjevernoj Americi



Slika 3.6: ME plot duljina (u miljama) rijeka u Sjevernoj Americi te pravac prilagođen tim podacima metodom najmanjih kvadrata

Bibliografija

[1] P. Embrechts, C. Klüppelberg i T. Mikosch, *Modelling Extremal Events*, Springer-Verlag Berlin Heidelberg, 1997.

[2] S. Ghosh i S. Resnick, *A discussion on mean excess plots*, Stochastic Processes and their Applications **120** (2010), br. 8, 1492–1517.

[3] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York, 1975.

[4] A. McNeil, R. Frey i P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools*, sv. 101, 2005.

[5] I. Molchanov, *Theory of Random Sets*, Springer, London, 2005.

[6] H. T. Nguyen, *An Introduction to Random Sets*, Chapman and Hall/CRC, 2006.

[7] G. Salinetti i J.B. Wets, *On the convergence of closed–valued measurable multifunctions*, Transactions of the American Math. Society, 1981.

Sažetak

Cilj ovog diplomskog rada bio je dati motivaciju za uvođenje pojma zatvorenog slučajnog skupa. Budući da je zatvoreni slučajni skup prirodno poopćenje pojma slučajnog vektora, na početku smo dali pregled već poznatih rezultata. Nakon definicije pojma zatvorenog slučajnog skupa pokazali smo neke jednostavne primjere. Uveli smo funkcije distribucije i kapaciteta i pokazali njihova osnovna svojstva. Iskazali smo Choquetov teorem koji daje nužne i dovoljne uvjete da bi funkcija bila funkcija kapaciteta nekog zatvorenog slučajnog skupa. Definirali smo Choquetov integral koji omogućuje integriranje neaditivnih funkcija. Funkcije kapaciteta zatvorenih slučajnih skupova su neaditivne pa Choquetov integral koristimo pri definiranju tri tipa konvergencije zatvorenih slučajnih skupova. Naposljetku, promotrili smo što se događa s konvergencijom ME plotova, odnosno grafova funkcije očekivanog viška. ME plotovi se koriste u analizi i modeliranju opaženih ekstremnih vrijednosti. Na simuliranim podacima iz generalizirane Pareto razdiobe i F razdiobe te na stvarnim podacima potvrdili smo teorijski rezultat o konvergenciji ME plotova prema linearnoj funkciji.

Summary

The aim of this master thesis was to provide the motivation for the introduction of a random closed set. Since a random closed set is a natural generalization of the concept of a random vector, at the beginning we provided an overview of some well known results. After defining the concept of a random closed set, we presented some simple examples. We have introduced the distribution and capacity functions and demonstrated their basic properties. We have presented the Choquet's theorem which gives the necessary and sufficient conditions for a function to be a capacity function of a random closed set. We have defined the Choquet integral which allows the integration of the nonadditive functions. Capacity functionals are nonadditive so we use Choquet integral to define the three modes of convergence of a series of random closed sets. Lastly, we observed what happens with the convergence of the ME plots, that is, the mean excess function plots. ME plots are used in the analysis and modeling of extreme value observations. On the simulated data from the generalized Pareto distribution and the F distribution and on the real–world data, we illustrated the theorem of the convergence of ME plots to a linear function.

Životopis

Rođena sam 22. studenog 1997. godine u Kutini. Tamo 2004. godine upisujem Osnovnu školu Zvonimira Franka u kojoj se, tijekom viših razreda, uz nastavnicu Manuelu Koch, razvija moja ljubav prema matematici i želja da upišem Prirodoslovno–matematički fakultet. Nakon toga, 2012. godine, upisujem matematičku gimnaziju u Srednjoj školi Tina Ujevića u Kutini. Tijekom srednjoškolskog obrazovanja sudjelovala sam na natjecanjima iz Matematike, Fizike, Kemije i Talijanskog jezika.

Preddiplomski studij Matematike na Prirodoslovno–matematičkom fakultetu u Zagrebu upisujem 2016. godine, a diplomski studij Matematičke statistike 2019. godine. Tijekom studiranja držala sam demonstrature iz kolegija Programiranje 1 i 2, Algebarske strukture, Obične diferencijalne jednadžbe, Statistika i Metode matematičke fizike. Prve dvije godine studija bila sam aktivni član studentske udruge eSTUDENT unutar koje sam organizirala studentsko natjecanje u dubinskoj analizi velikih količina podataka, strojnog učenja i umjetne inteligencije Mozgalo, kasnije preimenovano u LUMEN Data science. Na kraju druge godine studiranja, na temelju programerskih vještina stečenih tijekom studija, zapošljavam se u zagrebačkoj programerskoj tvrtki Microblink na poziciji C++ razvojnog inženjera. Na kraju oba studijska programa bila sam nagrađena nagradom Matematičkog odsjeka za izniman uspjeh na studiju.