Matematički kolokvijum (Banja Luka)
MAT-KOL (Banja Luka)
XIV(1)(2008), 59 - 83

TAČNE I ASIMPTOTSKE RASPODJELE NEKIH UZORAČKIH STATISTIKA

Aleksandra Vasilić

Prirodno-matematički fakultet
Banja Luka, Mladena Stojanovića br. 2
E-mail: avasilic@blic.net

Uvod

Predmet razmatranja u matematičkoj statistici je *populacija* ili generalni skup elemenata. Kvantitativna ili kvalitativna karakteristika elemenata populacije zove se *obilježje*. Pod statističkim eksperimentom se podrazumijeva registrovanje vrijednosti obilježja X na nekom podskupu populacije, koji se naziva *uzorak*.

Bitno je da uzorak bude reprezentativan tj. elemente uzorka treba birati na slučajan način, tako da se neutrališe svaka moguća zavisnost između posmatranog obilježja i uzorka.

Osnovni cilj matematičke statistike je da se na osnovu statističkog eksperimenta dođe do zaključaka o nepoznatoj raspodjeli $F(x)$ obilježja X, radi prognoziranja budućih rezultata i eventualnog preduzimanja mjera da ti rezultati budu što povoljniji.

Neka se na populaciji statističkim eksperimentom bira element i registruje vrijednost obilježja. Rezultat eksperimenta je slučajna promjenljiva X. Ponavljajući eksperiment n puta složeni eksperiment se opisuje n- dimenzionalnom slučajnom promjenljivom $(X_1, X_2, ..., X_n)$ koja predstavlja slučajni uzorak. U daljem ćemo posmatrati takozvane *proste slučajne uzorke*.

Definicija 1 *Prost slučajan uzorak obima n, iz populacije na kojoj obilježje X ima funkciju raspodjele $F(x)$, je slučajni vektor $(X_1, X_2, ..., X_n)$, pri čemu su $X_1, X_2, ..., X_n$ nezavisne slučajne veličine sa istom funkcijom raspodjele $F(x)$ kao i obilježje X. Funkcija raspodjele slučajnog vektora $(X_1, X_2, ..., X_n)$ je:*

$$\begin{aligned} F(x_1, x_2, ..., x_n) &= P\{X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n\} \\ &= P\{X_1 \le x_1\}P\{X_2 \le x_2\}...P\{X_n \le x_n\} \\ &= \prod_{k=1}^{n} F(x_k) \end{aligned}$$

U realizovanom statističkom eksperimentu slučajni vektor $(X_1, X_2, ..., X_n)$ predstavlja n-torku brojeva $(x_1, x_2, ..., x_n)$ i zove se *realizacija uzorka.*

Definicija 2 *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ Borelova funkcija i $(X_1, X_2, ..., X_n)$ uzorak uzet iz raspodjele F. Tad se slučajna veličina $f(X_1, X_2, ..., X_n)$ zove statistika.*

Ovim radom opisane su tačne i asimptotske raspodjele sledećih statistika:

1. Empirijska funkcija raspodjele;
2. Uzorački momenti (specijalno: uzoračka sredina i uzoračka disperzija);
3. Statistike poretka (minimalna, maksimalna ili proizvoljnog reda);
4. Uzoračka medijana i kvantili;

1 Empirijska funkcija raspodjele

Empirijska funkcija raspodjele $F_n(x)$ predstavlja uzorački ekvivalent teorijskoj funkciji raspodjele $F(x) = P\{X \leq x\}$. Prije definisanja empirijske funkcije raspodjele potrebno je podsjetiti se na slučajnu veličinu *indikatora događaja*:

$$I(X \leq x) = \begin{cases} 1, & \text{za } X \leq x \\ 0, & \text{za } X > x \end{cases} \tag{1.1}$$

sa dvotačkastom raspodjelom *nula-jedan*:

$$I : \begin{pmatrix} 1 & 0 \\ F(x) & 1 - F(x) \end{pmatrix} \tag{1.2}$$

Definicija 3 *Neka je $(X_1, X_2, ..., X_n)$ prost slučajni uzorak iz raspodjele F. Slučajna veličina $F_n : \mathbb{R} \to \mathbb{R}$ data sa*

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I(X_k \leq x), \ -\infty < x < \infty \tag{1.3}$$

zove se empirijska funkcija raspodjele.

Kako su matematičko očekivanje i disperzija indikatora događaja dati sa:

$$EI(X_k \leq x) = F(x)$$

i

$$DI(X_k \leq x) = F(x)(1 - F(x)), \ k = 0, 1, ..., n$$

i kako slučajna veličina

$$nF_n(x) = \sum_{k=1}^{n} I(X_k \leq x) \tag{1.4}$$

predstavlja sumu od n nezavisnih jednako raspodijeljenih slučajnih veličina sa dvotačkastom nula-jedan raspodjelom, to slučajna veličina (1.4) ima binomnu $\mathcal{B}(n, F(x))$ raspodjelu.

Odatle slijedi da je za $k = 0, 1, ..., n$:

$$\begin{aligned} P\{nF_n(x) = k\} &= P\{F_n(x) = \frac{k}{n}\} \\ &= \binom{n}{k}(F(x))^k(1 - F(x))^{n-k} \end{aligned} \tag{1.5}$$

sa matematičkim očekivanjem:

$$E(F_n(x)) = F(x) \tag{1.6}$$

i disperzijom:

$$D(F_n(x)) = \frac{1}{n}F(x)(1 - F(x)) \tag{1.7}$$

Ako uzoračke vrijednosti $X_1, X_2, ..., X_n$ poredamo u neopadajući poredak

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

tada se niz $X_{(1)}, X_{(2)}, ..., X_{(n)}$ zove *varijacioni* niz.

Realizaciju $x_{(1)}, x_{(2)}, ..., x_{(n)}$ varijacionog niza posmatramo na realizovanom uzorku, te za nju analogno vrijedi

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}.$$

Empirijska funkcija raspodjele može se odrediti na sledeći način:

$$F_n(x) = \begin{cases} 0, & \text{ako je } x < x_{(1)} \\ \frac{k}{n}, & \text{ako je } x_{(k)} \leq x < x_{(k+1)},\ 1 \leq k \leq n-1 \\ 1, & \text{ako je } x \geq x_{(n)}. \end{cases} \tag{1.8}$$

Fundamentalan značaj empirijske funkcije raspodjele $F_n(x)$ je u tome što ona, u svakoj fiksiranoj tački x, dobro aproksimira nepoznatu teorijsku funkciju raspodjele $F(x)$ obilježja X, kad obim uzorka neograničeno raste.

Naime, na osnovu jakog Borelovog zakona velikih brojeva za svaku fiksiranu tačku x vrijedi skoro sigurna konvergencija

$$F_n(x) \xrightarrow{s.s} F(x),\ n \to \infty \tag{1.9}$$

Sledeća teorema, nazvana *centralnom teoremom matematičke statistike*, tvrdi da je gore navedena skoro sigurna konvergencija uniformna po x (vidi [6]).

Teorema 1.1 **(Glivenko-Kanteli)** *Neka je $\{F_n\}_{n\in\mathbb{N}}$ niz empirijskih funkcija raspodjele dobijenih na uzorku obima n, uzetog iz raspodjele F. Vrijedi jednakost:*

$$P\{\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0\} = 1 \tag{1.10}$$

Graničnu raspodjelu statistike maksimalnog odstupanja empirijske od teorijske funkcije raspodjele:

$$D_n \equiv D_n(X_1, X_2, ..., X_n) = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|, \ x \in \mathbb{R} \tag{1.11}$$

uz pretpostavku da je F neprekidna funkcija, odredio je A.N. Kolmogorov 1933. godine. Sledećom teoremom je pokazano da granična raspodjela statistike D_n ne zavisi od funkcije F.

Teorema 1.2 *Ako je F neprekidna funkcija, tada je*

$$\begin{aligned} \lim_{n\to\infty} P\{\sqrt{n}D_n \le t\} &= \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2t^2} \\ &= 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2t^2}, \ t > 0 \end{aligned} \tag{1.12}$$

Teorema 1.2 koristi se za provjeravanje statističke hipoteze da obilježje X ima pretpostavljenu funkciju raspodjele.

Još jedna interesantna teorema govori o osobinama empirijskih funkcija raspodjela. To je teorema Smirnova iz 1944.godine, a koristi se za testiranje statističke hipoteze da su dva uzorka uzeta iz iste raspodjele.

Teorema 1.3 *(***Smirnov***) Neka su $F_{1n_1}(x)$ i $F_{2n_2}(x)$ empirijske funkcije raspodjela definisane na nezavisnim uzorcima $(X_{11}, X_{12}, ..., X_{1n_1})$ i $(X_{21}, X_{22}, ..., X_{2n_2})$, uzetih iz iste populacije na kojoj obilježje X ima raspodjelu F, i neka je statistika njihovih maksimalnih odstupanja*

$$D_{n_1n_2} = \sup_{x\in\mathbb{R}} |F_{1n_1}(x) - F_{1n_1}(x)|.$$

Ako je funkcija raspodjele F neprekidna, onda za svaki pozitivan broj t vrijedi:

$$\lim_{n_1,n_2\to\infty} P\{\sqrt{\frac{n_1n_2}{n_1+n_2}}D_{n_1n_2} \le t\} = K(t) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2t^2}. \tag{1.13}$$

2 Uzorački momenti

Posmatrajmo uzorak $(X_1, X_2, ..., X_n)$ uzet iz raspodjele F.

Definicija 4 *Uzoračka sredina je statistika koja predstavlja aritmetičku sredinu komponenti uzorka:*

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{2.14}$$

Statistika $\overline{X}_n$ predstavlja uzorački ekvivalent matematičkom očekivanju $E(X)$ kod teorijske raspodjele obilježja X.

Definicija 5 *Uzoračka disperzija je statistika*

$$\overline{S}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \tag{2.15}$$

gdje je sa $\overline{X}_n$ označena uzoračka sredina.

Statistika $\overline{S}_n^2$ predstavlja uzorački ekvivalent disperziji $D(X) = E(X - EX)^2$ kod teorijske raspodjele obilježja X.
Statistike $\overline{X}_n$ i $\overline{S}_n^2$ su specijalni slučajevi sledećih uzoračkih momenata.

Definicija 6 *Obični uzorački moment reda k je statistika*

$$A_{nk} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{2.16}$$

Statistika A_{nk} predstavlja uzorački ekvivalent teorijskom običnom momentu reda k definisanom sa $m_k = E(X^k)$.

Definicija 7 *Centralni uzorački moment reda k je statistika*

$$B_{nk} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k \tag{2.17}$$

Statistika B_{nk} predstavlja uzorački ekvivalent teorijskom centralnom momentu reda k definisanom sa $\mu_k = E(X - EX)^k$.
Na osnovu gornjih definicija je očigledno $A_{n1} = \overline{X}_n$ i $B_{n2} = \overline{S}_n^2$.
U vezi sa uzoračkim momentima razmotrićemo nekoliko pitanja:

1. Numeričke karakteristike uzoračkih momenata;
2. Tačna zajednička raspodjela statistika $\overline{X}_n$i $\overline{S}_n^2$ kod uzorka iz normalne raspodjele
3. Granične raspodjele uzoračkih momenata, pri $n \to \infty$.

2.1 Numeričke karakteristike uzoračkih momenata

Neka je X obilježje sa funkcijom raspodjele F, za koje vrijedi

$$E(X) = m,\ D(X) = \sigma^2,\ E(X^k) = m_k,\ E(X - m)^k = \mu_k$$

(vidi se da je $m_1 = m$ i $\mu_2 = \sigma^2$) i neka je $(X_1, X_2, ..., X_n)$ uzorak uzet iz raspodjele F.

Primjer 2.1 *Odrediti* $E(\overline{X}_n)$ *i* $D(\overline{X}_n)$.

Rješenje:

$$
\begin{aligned}
E(\overline{X}_n) &= E(\frac{1}{n}\sum_{k=1}^{n} X_k) \\
&= \frac{1}{n}\sum_{k=1}^{n} E(X_k) \\
&= \frac{1}{n} nm \\
&= m. \qquad (2.18)
\end{aligned}
$$

$$
\begin{aligned}
D(\overline{X}_n) &= D(\frac{1}{n}\sum_{k=1}^{n} X_k) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} D(X_k) \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n} \quad \blacklozenge \qquad (2.19)
\end{aligned}
$$

Primjer 2.2 *Odrediti* $E(\overline{S}_n^2)$ *i* $D(\overline{S}_n^2)$.

Rješenje:

$$
\begin{aligned}
E(\overline{S}_n^2) &= E(\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2) \\
&= \frac{1}{n}\sum_{k=1}^{n} E(X_k - \overline{X}_n)^2 \\
&= \frac{1}{n}\sum_{k=1}^{n} E(X_k)^2 - E(\overline{X}_n^2) \\
&= \frac{1}{n} n(\sigma^2 + m^2) - (\frac{\sigma^2}{n} + m^2) \\
&= \sigma^2 - \frac{\sigma^2}{n} \\
&= \frac{n-1}{n}\sigma^2 \qquad (2.20)
\end{aligned}
$$

Da bismo odredili $D(\overline{S}_n^2)$ *prvo nađimo* $E(\overline{S}_n^2)^2$.

$$
E(\overline{S}_n^2)^2 = E(\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2)^2
$$

$$
\begin{aligned}
&= E(\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X}_n^2)^2 \\
&= E(\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \frac{1}{n^2}(\sum_{k=1}^{n} X_k)^2)^2 \\
&= E(\frac{n-1}{n^2}\sum_{k=1}^{n} X_k^2 - \frac{2}{n^2}\sum_{k<j} X_k X_j)^2 \\
&= E(\frac{(n-1)^2}{n^4}\sum_{k=1}^{n} X_k^4 + \frac{2(n-1)^2+4}{n^4}\sum_{k<j} X_k^2 X_j^2) \\
&= \frac{(n-1)^2}{n^3}\mu_4 + \frac{2(n-1)^2+4}{n^3}(n-1)\mu_2 \qquad (2.21)
\end{aligned}
$$

Sada je

$$
\begin{aligned}
D(\overline{S}_n^2) &= E(\overline{S}_n^2)^2 - (E(\overline{S}_n^2))^2 \\
&= \frac{(n-1)^2}{n^3}(\mu_4 - \frac{n-3}{n-1}\mu_2^2) \ \blacklozenge \qquad (2.22)
\end{aligned}
$$

Primjer 2.3 *Odrediti $E(A_{nk})$ i $D(A_{nk})$.*
Rješenje:

$$
\begin{aligned}
E(A_{nk}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i^k) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) \\
&= \frac{1}{n} n m_k \\
&= m_k. \qquad (2.23)
\end{aligned}
$$

$$
\begin{aligned}
D(A_{nk}) &= D(\frac{1}{n}\sum_{i=1}^{n} X_i^k) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} D(X_i^k) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E(X_i^k - m_k)^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} (E(X_i^k)^2 - 2E(X_i^k)m_k + m_k^2) \\
&= \frac{1}{n^2} n(m_{2k} - m_k^2) \\
&= \frac{m_{2k} - m_k^2}{n} \ \blacklozenge \qquad (2.24)
\end{aligned}
$$

Zadatak 2.1 *Pokazati da je*

$$E(B_{nk}) = \mu_k$$

i

$$D(B_{nk}) = \frac{1}{n}(\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} - \mu_k^2 + k^2\mu_2\mu_{k-1}^2)$$

2.2 Zajednička raspodjela statistika $\overline{X}_n$i $\overline{S}_n^2$ kod uzorka iz normalne raspodjele

Za početak navedimo samo definicije dva tipa raspodjela bitnih za razumijevanje pomenute zajedničke raspodjele iz naslova.

Definicija 8 (χ^2- raspodjela) *Raspodjela* $\Gamma(\frac{1}{2}, \frac{n}{2})$ *određena gustinom*

$$f_n(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & \textit{ako je } x > 0 \\ 0, & \textit{ako je } x \leq 0 \end{cases} \tag{2.25}$$

zove se $\chi^2 -$ *raspodjela sa* n *stepeni slobode i označava sa* χ_n^2.

Metodom karakterističnih funkcija lako se dokakazuje da ako su $X_1, X_2, ..., X_n$ nezavisne slučajne veličine iz $\mathcal{N}(0,1)$ raspodjele, onda zbir njihovih kvadrata ima χ_n^2-raspodjelu, tj. važi jednakost:

$$\chi_n^2 = X_1^2 + X_2^2 + ... + X_n^2.$$

Definicija 9 (Studentova t raspodjela) *Neka su* $X_0, X_1, X_2, ..., X_n$ *nezavisne slučajne veličine sa* $\mathcal{N}(0,1)$ *raspodjelom. Raspodjela slučajne veličine*

$$t_n = \frac{X_0}{\sqrt{\frac{1}{n}(X_1^2 + X_2^2 + ... + X_n^2)}} = \frac{X_0}{\sqrt{\frac{1}{n}\chi_n^2}}$$

zove se Studentova t *raspodjela sa* n *stepeni slobode, i označava sa* t_n.

Gustina t_n raspodjele je određena formulom

$$f_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-(n+1)/2}, \ x \in \mathbb{R} \tag{2.26}$$

Poznavanje zajedničke raspodjele statistika $\overline{X}_n$ i $\overline{S}_n^2$, u slučaju kada obilježje X ima normalnu raspodjelu, ima veliku primjenu u matematičkoj statistici na primjer prilikom ocjenjivanja ili testiranja hipoteza o parametrima normalne raspodjele. Teoremu koja slijedi dokazao je Fišer još 1925. godine, a njen dokaz se može pronaći u većini udžbenika statistike (vidi npr.[9]).

Teorema 2.1 (Fišer) *Neka je uzorak* $(X_1, X_2, ..., X_n)$ *uzet iz normalne raspodjele* $\mathcal{N}(m, \sigma^2)$. *Tada vrijede tvrdnje:*

1. *Slučajne veličine* $\overline{X}_n$ *i* $\overline{S}_n^2$ *su nezavisne.*

2. *Slučajna veličina* $\frac{\overline{X}_n - m}{\sigma}\sqrt{n}$ *ima* $\mathcal{N}(0,1)$ *raspodjelu.*

3. *Slučajna veličina* $\frac{n\overline{S}_n^2}{\sigma^2}$ *ima* χ^2_{n-1} *raspodjelu.*

Koristeći prethodnu Teoremu 2.1 i Definiciju 9 dokazuje se još jedna vrlo važna i često korištena zajednička raspodjela statistika $\overline{X}_n$ i $\overline{S}_n^2$. Opisana je sledećim primjerom.

Primjer 2.4 *Neka je uzorak* $(X_1, X_2, ..., X_n)$ *uzet iz normalne raspodjele* $\mathcal{N}(m, \sigma^2)$. *Odrediti raspodjelu slučajne veličine* $\frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1}$.

Rješenje:

Ako je uzorak $(X_1, X_2, ..., X_n)$ *uzet iz normalne raspodjele* $\mathcal{N}(m, \sigma^2)$, *tada slučajna veličina* $\overline{X}_n$ *ima* $\mathcal{N}(m, \frac{\sigma^2}{n})$ *raspodjelu, odnosno kao što tvrdi prethodna teorema, slučajna veličina* $\frac{\overline{X}_n - m}{\sigma}\sqrt{n}$ *ima* $\mathcal{N}(0,1)$ *raspodjelu. Sem toga, prema tvrdnji prethodne teoreme, slučajna veličina* $\frac{n\overline{S}_n^2}{\sigma^2}$ *ima* χ^2_{n-1} *raspodjelu. Posmatrajmo količnik:*

$$t = \frac{\frac{\overline{X}_n - m}{\sigma}\sqrt{n}}{\sqrt{\frac{1}{n-1}\frac{n\overline{S}_n^2}{\sigma^2}}} = \frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1} \tag{2.27}$$

Iz Definicije 9 slijedi zaključak da slučajna veličina

$$t = \frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1} \tag{2.28}$$

ima t_{n-1} *raspodjelu.*

2.3 Granične raspodjele uzoračkih momenata

Za određivanje graničnih raspodjela gornjih statistika koristimo granične teoreme teorije vjerovatnoće. Za formulisanje ovih teorema potrebno je definisati *konvergencije nizova slučajnih veličina.*

2.3.1 Konvergencije nizova slučajnih vličina

Razlikujemo 4 osnovna tipa konvergencija (vidi [5]):
-konvergencija skoro sigurno
-konvergencija u srednjem reda p $(0 < p < \infty)$
-konvergencija u vjerovatnoći
-konvergencija u raspodjeli

Neka je $X_1, X_2, ...$ niz slučajnih veličina definisanih na istom prostoru vjerovatnoća, sa istom funkcijom raspodjele $F(x)$.

Definicija 10 *Niz slučajnih veličina* (X_n) ***konvergira skoro sigurno*** *ka slučajnoj veličini* X*, ako važi:*

$$P\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1 \tag{2.29}$$

Ova konvergencija se označava sa $X_n \overset{s.s.}{\to} X, n \to \infty$.

Definicija 11 *Niz slučajnih veličina* (X_n) ***konvergira u srednjem reda p*** *(za* $0 < p < \infty$*) ka slučajnoj veličini* X*, ako je* $E|X|^p < \infty$ *i vrijedi:*

$$\lim_{n\to\infty} E(|X_n - X|)^p = 0 \tag{2.30}$$

Ova konvergencija se označava sa $X_n \overset{L^p}{\to} X, n \to \infty$.

Specijalno, konvergencija u srednjem reda $p = 2$ *zove se* ***konvergencija u srednje-kvadratnom****, i označava sa* $X_n \overset{s.k.}{\to} X, n \to \infty$.

Definicija 12 *Niz slučajnih veličina* (X_n) ***konvergira u vjerovatnoći*** *ka slučajnoj veličini* X*, ako važi:*

$$(\forall \varepsilon > 0) \lim_{n\to\infty} P\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\} = 0 \tag{2.31}$$

Ova konvergencija se označava sa $X_n \overset{P}{\to} X, n \to \infty$.

Definicija 13 *Niz funkcija raspodjele* (F_n) *slabo konvergira ka funkciji raspodjele* F*, ako u svakoj tački neprekidnosti funkcije* F *vrijedi:*

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{2.32}$$

U tom slučaju se kaže da niz slučajnih veličina (X_n)*, sa odgovarajućim funkcijama raspodjele* $F_n(x) = P\{X_n \leq x\}$*,* ***konvergira u raspodjeli*** *ka slučajnoj veličini* X*, kojoj odgovara funkcija raspodjele* $F(x) = P\{X \leq x\}$*. Ova konvergencija se označava sa* $X_n \overset{D}{\to} X, n \to \infty$.

2.3.2 Kratak pregled graničnih teorema

Kaže se da za niz (X_n) slučajnih veličina važi **slabi zakon velikih brojeva** ako vrijedi:

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k) \overset{P}{\to} 0,\ n \to \infty, \tag{2.33}$$

a za niz (X_n) vrijedi **jaki zakon velikih brojeva** ako je:

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k) \overset{s.s.}{\to} 0,\ n \to \infty, \tag{2.34}$$

Sledeće teoreme daju potrebne, a neke i dovoljne uslove da za posmatrani niz slučajnih veličina (X_n) važi neki zakon velikih brojeva.

Teorema 2.2 ***(Čebišov)*** *Ako niz nezavisnih slučajnih veličina* (X_n) *ima konačne disperzije, ograničene istom konstantom, tj.*

$$(\exists C > 0)(\forall n \in \mathbb{N}) DX_n \leq C \tag{2.35}$$

tad za njega važi slabi zakon velikih brojeva (2.33).

Teorema 2.3 ***(Hinčin)*** *Ako je* (X_n) *niz nezavisnih i jednako raspodijeljenih slučajnih veličina sa konačnim matematičkim očekivanjem* $EX_n = m$ *tad za niz vrijedi slabi zakon velikih brojeva :*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} m, \ n \to \infty \tag{2.36}$$

Teorema 2.4 ***(Borel)*** *Neka je* S_n *broj uspjeha posmatranog događaja u* n *nezavisnih eksperimenata, pri čemu je vjerovatnoća uspjeha u svakom od eksperimenata jednaka* p. *Tada vrijedi:*

$$P\{\lim_{n\to\infty} \frac{S_n}{n} = p\} = 1. \tag{2.37}$$

Teorema 2.5 ***(1. teorema Kolmogorova)***

Ako je (X_n) *niz nezavisnih slučajnih veličina, koje imaju konačne disperzije, i koji zadovoljava uslov:*

$$\sum_{k=1}^{n} \frac{DX_n}{n^2} < \infty \tag{2.38}$$

tad za njega vrijedi jaki zakon velikih brojeva (2.34).

Teorema 2.6 ***(2. teorema Kolmogorova)***

Ako je (X_n) *niz nezavisnih slučajnih veličina sa istom raspodjelom i konačnim matematičkim očekivanjem* $E(X_n) = m$, *tad za njega vrijedi jaki zakon velikih brojeva (2.34):*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{s.s.} m, \ n \to \infty \tag{2.39}$$

Teorema 2.7 ***(Centralna granična teorema Lindeberg-Levija)*** *Neka je* (X_n) *niz nezavisnih slučajnih veličina sa istom raspodjelom, matematičkim očekivanjem* $E(X_n) = m$ *i konačnom disperzijom* $D(X_n) = \sigma^2 > 0$. *Ako je* $S_n = \sum_{k=1}^{n} X_k$ *onda za svako* $x \in \mathbb{R}$, *pri* $n \to \infty$ *vrijedi:*

$$P\{\frac{S_n - nm}{\sigma\sqrt{n}} \leq x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \tag{2.40}$$

Gornja teorema se može formulisati i ovako:

$$\frac{S_n - E(S_n)}{\sqrt{D(S_n)}} \xrightarrow{D} \mathcal{N}(0,1), \; n \to \infty \tag{2.41}$$

odnosno niz funkcija raspodjela rastućeg niza standardizovanih suma (S_n) nezavisnih slučajnih veličina konvergira ka normalnoj Gausovoj funkciji raspodjele. Više o graničnim teoremama može da se vidi u [5].

Primjenom gore navedenih slabih i jakih zakona velikih brojeva mogu se dokazati sledeće konvergencije statistika koje smo razmatrali na početku rada.

Primjer 2.5 *Dokazati da vrijedi:* $A_{nk} \xrightarrow{s.s.} m_k, \; n \to \infty$

Rješenje:

Ako je $(X_1, X_2, ..., X_n)$ *uzorak uzet iz raspodjele* F*, tada su sve* X_i *jednako raspodijeljene, nezavisne i sa istim matematičkim očekivanjem* $E(X_i) = m, \; i = 1, .., n$*. Prema teoremi Hinčina (2.36) za niz* (X_n) *vrijedi slabi zakon velikih brojeva:*

$$A_{nk} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} m_k, \; n \to \infty \tag{2.42}$$

Čak štaviše, prema drugoj teoremi Kolmogorova (2.39) za niz (X_n) *vrijedi i jaki zakon velikih brojeva:*

$$A_{nk} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{s.s.} m_k, \; n \to \infty \; \blacklozenge \tag{2.43}$$

Napomena 1 *Specijalan slučaj iz prethodnog primjera za* $k = 1$ *daje rezultat:*

$$A_{n1} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{s.s.} m, \; n \to \infty$$

ŏdnosno uzoračka sredina skoro sigurno konvergira ka matematičkom očekivanju obilježja X*, kad obim uzorka neograničeno raste.* $\blacklozenge$

Zadatak 2.2 *Dokazati da vrijedi:* $B_{nk} \xrightarrow{s.s.} \mu_k, \; n \to \infty$

Napomena 2 *Specijalan slučaj iz prethodnog zadatka za* $k = 2$ *daje rezultat:*

$$B_{n2} = \overline{S}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \xrightarrow{s.s.} \sigma^2, \; n \to \infty \; \blacklozenge$$

odnosno uzoračka disperzija skoro sigurno konvergira ka disperziji obilježja X*, kad obim uzorka neograničeno raste.* $\blacklozenge$

Primjenom centralne granične i ostalih graničnih teorema, u nekoliko sledećih primjera se dokazuju asimptotske raspodjele statistika A_{nk} i B_{nk}.

Primjer 2.6 *Dokazati da vrijedi:* $A_{nk} \sim \mathcal{N}(m_k, (m_{2k} - m_k^2)/n)$*, pri* $n \to \infty$

Rješenje:

Statistika $n \cdot A_{nk} = \sum_{i=1}^{n} X_i^k$ *predstavlja sumu n nezavisnih i jednako raspodijeljenih slučajnih veličina* X_i^k*, sa matematičkim očekivanjem* $E(X_i^k) = m_k$ *i disperzijom* $D(X_i^k) = m_{2k} - m_k^2$ *(vidi Primjer 2.3).*

Prema tome, normirana suma

$$\begin{aligned} \frac{n \cdot A_{nk} - E(n \cdot A_{nk})}{\sqrt{D(n \cdot A_{nk})}} &= \frac{\sum_{i=1}^{n} X_i^k - E(\sum_{i=1}^{n} X_i^k)}{\sqrt{D(\sum_{i=1}^{n} X_i^k)}} \\ &= \frac{n \cdot A_{nk} - n \cdot m_k}{\sqrt{n(m_{2k} - m_k^2)}} \end{aligned}$$

ima pri $n \to \infty$ *graničnu normalnu raspodjelu* $\mathcal{N}(0,1)$*, na osnovu centralne granične teoreme (Teorema 2.7).*

Odatle slijedi da statistika A_{nk} *ima asimptotski normalnu raspodjelu* $\mathcal{N}(m_k, (m_{2k} - m_k^2)/n)$ *pri* $n \to \infty$*, čime je tvrdnja dokazana* ⧫

Napomena 3 *Specijalan slučaj u prethodnom primjeru za* $k = 1$ *dovodi do niza jednakosti:*

$$\begin{aligned} \frac{n \cdot \overline{X}_n - E(n \cdot \overline{X}_n)}{\sqrt{D(n \cdot \overline{X}_n)}} &= \frac{n \cdot A_{n1} - E(n \cdot A_{n1})}{\sqrt{D(n \cdot A_{n1})}} \\ &= \frac{\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)}{\sqrt{D(\sum_{i=1}^{n} X_i)}} \\ &= \frac{n \cdot \overline{X}_n - n \cdot m}{\sqrt{n\sigma^2}} \end{aligned}$$

odnosno

$$\overline{X}_n \sim \mathcal{N}(m, \sigma^2/n), \ n \to \infty$$

kao i

$$\frac{\overline{X}_n - m}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1), \ n \to \infty$$

To znači da uzoračka sredina ima asimptotski normalnu raspodjelu $\mathcal{N}(m, \sigma^2/n)$ *pri* $n \to \infty$*, a asimptotska raspodjela normalizovane vrijednosti uzoračke sredine je normalna* $\mathcal{N}(0,1)$ *raspodjela* ⧫

Primjer 2.7 *Dokazati da vrijedi:*

$$B_{nk} \sim \mathcal{N}(\mu_k, (\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} - \mu_k^2 + k^2\mu_2\mu_{k-1}^2)/n), \ n \to \infty \tag{2.44}$$

Rješenje:

Polazi se od razlaganja

$$\begin{aligned}
B_{nk} &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^k \\
&= \frac{1}{n}\sum_{i=1}^{n}[\binom{k}{0}X_i^k-\binom{k}{1}X_i^{k-1}\overline{X}_n+\binom{k}{2}X_i^{k-2}\overline{X}_n^2-\binom{k}{3}X_i^{k-3}\overline{X}_n^3+...+(-1)^k\overline{X}_n^k] \\
&= A_{nk}-\binom{k}{1}A_{n(k-1)}\overline{X}_n+\binom{k}{2}A_{n(k-2)}\overline{X}_n^2-\binom{k}{3}A_{n(k-3)}\overline{X}_n^3+......+(-1)^k\overline{X}_n^k
\end{aligned} \tag{2.45}$$

Uvedimo normirane promjenljive C_k na sledeći način :

$$C_k=\frac{A_{nk}-m_k}{\sigma_k}\sqrt{n} \tag{2.46}$$

gdje je

$$m_k=EX^k$$

i

$$\begin{aligned}
\sigma_k^2 &= DX^k=E(X^k-m_k)^2 \\
&= E(X^{2k}-2X^km_k+m_k^2) \\
&= m_{2k}-2m_k^2+m_k^2 \\
&= m_{2k}-m_k^2
\end{aligned}$$

odnosno

$$\sigma_k^2=E(n(A_{nk}-m_k)^2)=m_{2k}-m_k^2$$

Tada je $E(C_k)=0$ i

$$D(C_k)=E(\frac{A_{nk}-m_k}{\sigma_k}\sqrt{n})^2=\frac{1}{\sigma_k^2}E(n(A_{nk}-m_k)^2)=\frac{\sigma_k^2}{\sigma_k^2}=1$$

a mješoviti momenti su

$$\begin{aligned}
E(C_kC_j) &= E(\frac{A_{nk}-m_k}{\sigma_k}\sqrt{n}\cdot\frac{A_{nj}-m_j}{\sigma_j}\sqrt{n}) \\
&= \frac{n}{\sigma_k\sigma_j}E[(A_{nk}-m_k)(A_{nj}-m_j)] \\
&= \frac{1}{\sigma_k\sigma_j}E[n(A_{nk}-m_k)(A_{nj}-m_j)] \\
&= \frac{\sigma_{kj}}{\sigma_k\sigma_j}
\end{aligned}$$

gdje su

$$\sigma_{kj}=E[n(A_{nk}-m_k)(A_{nj}-m_j)]=m_{k+j}-m_k\cdot m_j,\ k\neq j$$

mješoviti momenti drugog reda.

Iz prethodnih rezultata slijedi da

$$C_k = \frac{A_{nk} - m_k}{\sigma_k}\sqrt{n} \sim \mathcal{N}(0,1),\ n \to \infty$$

Dalje se uvrštavanjem izraza za momente B_{nk} *(iz 2.45) i* $\mu_k = E(X-m)^k$ *nakon obimnih transformacija dobija sledeća jednakost:*

$$\sqrt{n}(B_{nk} - \mu_k) = \sigma_k C_k - k\sigma_1\mu_{k-1}C_1 + \frac{L}{\sqrt{n}} \tag{2.47}$$

pri čemu za veličinu L *dobijenu u gornjim transformacijama, kao ostatak od izdvojene linearne kombinacije veličina* C_k *i* C_1*, vrijedi konvergencija* $\frac{L}{\sqrt{n}} \xrightarrow{P} 0,\ n \to \infty$. *Kako* C_1 *i* C_k *imaju zajedničku asimptotski normalnu raspodjelu, a linearna kombinacija normalno raspodijeljenih slučajnih veličina takođe ima normalnu raspodjelu, to iz (2.47) slijedi dokaz tražene asimptotske normalne raspodjele za statistiku* B_{nk} *iz (2.44)*♦

Napomena 4 *Specijalan slučaj u prethodnom primjeru za* $k = 2$ *dovodi do rezultata:*

$$B_{n2} \sim \mathcal{N}(\mu_2, (\mu_4 - 4\mu_1\mu_3 - \mu_2^2 + 4\mu_2\mu_1^2)/n)$$

tj.

$$\overline{S}_n^2 \sim \mathcal{N}(\sigma^2, (\mu_4 - \sigma^2)/n),\ \ n \to \infty$$

To znači da statistika $\overline{S}_n^2$ ima asimptotski normalnu raspodjelu $\mathcal{N}(\sigma^2, (\mu_4 - \sigma^2)/n)$, pri $n \to \infty$♦

Primjer 2.8 *Dokazati da zajednička raspodjela ma kog skupa promjenljivih* $\sqrt{n}(A_{nk} - m_k)$ *teži normalnoj raspodjeli sa nultim matematičkim očekivanjem i drugim momentima:*

$$\sigma_{kk} = \sigma_k^2 = E[n(A_{nk} - m_k)^2] = m_{2k} - m_k^2 \tag{2.48}$$

$$\sigma_{kj} = E[n(A_{nk} - m_k)(A_{nj} - m_j)] = m_{k+j} - m_k \cdot m_j \tag{2.49}$$

Rješenje:

Primjenom CGT za dvodimenzionalne raspodjele (vidjeti [4]) na par statistika $nA_{nk} = \sum_{i=1}^{n} X_i^k$ *i* $nA_{nj} = \sum_{i=1}^{n} X_i^j$ *pokazuje se da zajednička raspodjela slučajnih veličina* $\sqrt{n}(A_{nk} - m_k)$ *i* $\sqrt{n}(A_{nj} - m_j)$ *konvergira nekoj dvodimenzionalnoj normalnoj raspodjeli, pri* $n \to \infty$.

Primjenjujući centralnu graničnu teoremu za višedimenzionalne raspodjele (vidi npr. [10]) pokazuje se da npr. r *slučajnih veličina*

$$\sqrt{n}(A_{nk} - m_k),\ \sqrt{n}(A_{nj} - m_j), ..., \sqrt{n}(A_{nq} - m_q)$$

ima asimptotski r *-dimenzionalnu normalnu raspodjelu sa nultim matematičkim očekivanjem i drugim momentima određenim relacijama (2.48) i (2.49)*♦

3 Statistike poretka

Članovi varijacionog niza $X_{(1)}, X_{(2)}, ..., X_{(n)}$ zovu se *statistike poretka (ranga)*. Pri tom ako je $X_j = X_{(k)}$ onda je rang $X_j = k$.

Ekstremne statistike poretka su *statistika maksimuma i minimuma*:

$$X_{(n)} = M_n \quad = \quad \max\{X_1, X_2, ..., X_n\} \tag{3.50}$$

$$X_{(1)} = m_n \quad = \quad \min\{X_1, X_2, ..., X_n\} \tag{3.51}$$

kao i *širina (raspon) uzorka*:

$$R_{(n)} = X_{(n)} - X_{(1)}$$

3.1 Raspodjele statistika poretka

Neka je obilježje X apsolutno-naprekidnog tipa sa funkcijom raspodjele $F(x)$ i gustinom $f(x)$.

Ako je funkcija raspodjele $k-$te donje statistike ranga $X_{(k)}$, $1 \leq k \leq n$ određena sa $F_{X_{(k)}}(x) = P\{X_{(k)} \leq x\}$, tada se gustina statistike $X_{(k)}$ može odrediti na sledeći način (vidi [1] i [3]):

$$\begin{aligned}
f_{X_{(k)}}(x) &= F'_{X_{(k)}}(x) = \frac{\partial}{\partial x} P\{X_{(k)} \leq x\} \\
&= \frac{\partial}{\partial x} P\{\text{najmanje k od n slučajnih veličina } X_s \text{ je} \leq x\} \\
&= \frac{\partial}{\partial x} P\{> \text{ k uspjeha u n pokušaja}\} \\
&= \frac{\partial}{\partial x} \sum_{j=k}^{n} \binom{n}{j} P\{X_1 \leq x\}^j (1 - P(X_1 \leq x))^{n-j} \\
&= \frac{\partial}{\partial x} \sum_{j=k}^{n} \binom{n}{j} F^j(x) \cdot (1 - F(x))^{n-j} \\
&= \sum_{j=k}^{n} \binom{n}{j} (jF^{j-1}(x) f(x) (1 - F(x))^{n-j} + \\
&\quad +F^j(x)(n-j)(1-F(x))^{n-j-1}(-f(x))) \\
&= nf(x)(\sum_{j=k-1}^{n-1} \binom{n-1}{j} F^j(x)(1-F(x))^{(n-1)-j} - \\
&\quad \sum_{j=k}^{n} \binom{n-1}{j} F^j(x)(1 - F^{(n-1)-j}(x))
\end{aligned}$$

Kako se svi članovi u gornjim sumama poništavaju, osim prvog u prvoj sumi i poslednjeg u drugoj sumi, to imamo:

$$= \quad nf(x)[\binom{n-1}{k-1} F^{k-1}(x)(1-F(x))^{(n-1)-(k-1)} -$$

$$
\begin{aligned}
& \underbrace{-\binom{n-1}{n} F^n(x)(1-F(x))^{(n-1)-n}]}_{=\ 0} \\
= \;& n f(x)\binom{n-1}{k-1} F^{k-1}(x)(1-F(x))^{n-k} \\
= \;& \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x)(1-F(x))^{n-k} f(x) \qquad (3.52)
\end{aligned}
$$

Specijalni slučajevi:

1) Za $\boldsymbol{k = n}$ funkcija raspodjele statistike maksimuma je:

$$
\begin{aligned}
F_{X_{(n)}}(x) &= P\{X_{(n)} \le x\} \\
&= P\{\max\{X_1, X_2, ..., X_n\} \le x\} \\
&= P\{X_1 \le x, X_2 \le x, ..., X_n \le x\} \\
&= P(\bigcap_{i=1}^{n}\{X_i \le x\}) \\
&= \prod_{i=1}^{n} P\{X_i \le x\} \text{ (iz uslova nezavisnosti } X_i) \\
&= \prod_{i=1}^{n} F_{X_{(i)}}(x) \\
&= F^n(x)
\end{aligned}
$$

te je gustina za $X_{(n)}$ određena sa :

$$
f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = nF^{n-1}(x)f(x) \qquad (3.53)
$$

što se može dobiti i iz (3.52) za $k = n$.

2) Za $\boldsymbol{k = 1}$ funkcija raspodjele statistike minimuma je:

$$
\begin{aligned}
F_{X_{(1)}}(x) &= P\{X_{(1)} \le x\} \\
&= P\{\min\{X_1, X_2, ..., X_n\} \le x\} \\
&= 1 - P\{\min\{X_1, X_2, ..., X_n\} > x\} \\
&= 1 - P\{X_1 > x, X_2 > x, ..., X_n > x\} \\
&= 1 - P(\bigcap_{i=1}^{n}\{X_i > x\}) \\
&= 1 - \prod_{i=1}^{n} P\{X_i > x\} \text{ (iz uslova nezavisnosti } X_i) \\
&= 1 - \prod_{i=1}^{n}(1 - P\{X_i \le x\})
\end{aligned}
$$

$$\begin{aligned} &= 1 - \prod_{i=1}^{n}(1 - F_{X_{(i)}}(x)) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

te je gustina za $X_{(1)}$ određena sa :

$$\begin{aligned} f_{X_{(1)}}(x) &= F'_{X_{(1)}}(x) \\ &= -n(1 - F(x))^{n-1}(-f(x)) \\ &= nf(x)(1 - F(x))^{n-1} \end{aligned} \tag{3.54}$$

što se može dobiti i iz (3.52) za $k = 1$.

Zajednička gustina raspodjele za slučajni vektor svih statistika ranga na uzorku $(X_{(1)}, X_{(2)}, ..., X_{(n)})$ je:

$$f_{(X_{(1)}, X_{(2)}, ..., X_{(n)})}(x_1, x_2, ..., x_n) = n! \prod_{k=1}^{n} f(x_i) \tag{3.55}$$

za $x_1 < x_2 < ... < x_n$, za što se dokaz može vidjeti u [10].

Iz ove gustine mogu da se dobiju marginalne gustine za bilo koju statistiku ranga, kao i gustine za bilo koji podskup statistika ranga.

Na primjer, gustina podvektora $(X_{(k_1)}, X_{(k_2)}, ..., X_{(k_r)})$, gdje je $1 \leq k_1 < k_2 < ... < k_r \leq n,\ \ 1 < r \leq n$, slučajnog vektora $(X_{(1)}, X_{(2)}, ..., X_{(n)})$, dobija se integracijom gustine iz (3.55) po preostalim promjenljivim:

$$\begin{aligned} & f_{(X_{(k_1)}, X_{(k_2)}, ..., X_{(k_r)})}(x_{k_1}, x_{k_2}, ..., x_{k_r}) \\ = \ & n! F(x_{k_1})^{k_1 - 1} \cdot [F(x_{k_2}) - F(x_{k_1})]^{k_2 - k_1 - 1} \cdot \\ & \cdot ... [F(x_{k_r}) - F(x_{k_{r-1}})]^{k_r - k_{r-1} - 1} \cdot \\ & \cdot [1 - F(x_{k_r})]^{n - k_r} \cdot f(x_{k_1}) f(x_{k_2}) \cdot ... \cdot f(x_{k_r}) / [(k_1 - 1) \cdot \\ & \cdot (k_2 - k_1 - 1) \cdot ... \cdot (k_r - k_{r-1} - 1) \cdot (n - k_r)] \end{aligned} \tag{3.56}$$

za $x_{k_1} < x_{k_2} < ... < x_{k_n}$.

Specijalan slučaj u (3.56) je npr. marginalna gustina za statistiku $X_{(k)},\ 1 \leq k \leq n$:

$$f_{X_{(k)}}(x_k) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x_k)(1 - F(x_k))^{n-k} f(x) \tag{3.57}$$

što je u skladu sa dobijenim rezultatom u (3.52).

3.2 Asimptotske raspodjele statistika poretka

Razmotrimo asimptotske ponašanje raspodjela ekstremnih statistika ranga, maksimuma $M_n = X_{(n)}$ i minimuma $m_1 = X_{(1)}$, pri $n \to \infty$ za niz $\{X_i\}$, $i = \overline{1, n}$ nezavisnih slučajnih veličina sa istom funkcijom raspodjele $F(x)$. Ovom problematikom se bavi teorija ekstremnih vrijednosti.

Pokazano je da je raspodjela maksimuma data sa

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

Niže navedeni rezultati odnose se na maksimume, ali se jednostavnim transformacijama mogu prilagoditi i minimalnim vrijednostima pojava koje razmatramo, jer vrijedi:

$$P\{m_n \leq x\} = 1 - (1 - F(x))^n$$

i

$$min\{X_1, X_2, ..., X_n\} = -max\{-X_1, -X_2, ..., -X_n\}$$

Pri ispitivanju asimptotskog ponašanja maksimuma, osnovno pitanje je da li postoje nizovi realnih konstanti $a_n > 0$ i $b_n \in$ R, za koje vrijedi

$$\lim_{n\to\infty} P\{\frac{M_n - b_n}{a_n} \leq x\} = F^n(a_n x + b_n) = G(x) \tag{3.58}$$

za svaku tačku neprekidnosti x neke nedegenerisane funkcije raspodjele $G(x)$. Ako takvi nizovi postoje zovu se normirajuće konstante, a funkcija $G(x)$ je granična raspodjela linearno normiranog maksimuma M_n.

Definicija 14 *Funkcija raspodjele $F(x)$ pripada oblasti privlačenja nedegenerisane funkcije raspodjele $G(x)$ ako postoje nizovi realnih brojeva $a_n > 0$ i b_n, $n \in$ N, tako da vrijedi:*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{3.59}$$

za svaku tačku neprekidnosti funkcije $G(x)$.

Ovu pripadnost označavamo sa $F \in D(G)$.

Pitanje koje se nametalo je koje funkcije $G(x)$ se mogu pojaviti kao granične funkcije raspodjele linearno normiranog maksimuma niza od n nezavisnih slučajnih veličina, sa istom funkcijom raspodjele $F(x)$, pri $n \to \infty$? Dokazano je da je funkcija $G(x)$ istog tipa [1] kao funkcije iz parametarske familije:

Generalizovane raspodjela ekstremnih vrijednosti (GREV)

$$G_\gamma(x) = \begin{cases} e^{-(1+\gamma x)^{-\frac{1}{\gamma}}}, & 1 + \gamma x > 0, \gamma \neq 0 \\ e^{-e^{-x}}, & x \in \mathbb{R}, \qquad \gamma = 0 \end{cases} \tag{3.60}$$

$$\lim_{\gamma\to 0} G_\gamma(x) = G_0(x) \,\text{(Gumbelova raspodjela)}$$

Parametar γ zove se **indeks raspodjele ekstremnih vrijednosti**. Jednoparametarska familija $G_\gamma(x)$, za različite vrijednosti γ, svodi se na jednu od tri familije, koje zavise od **parametra oblika** $\alpha > 0$:
Za $\gamma = 0$ dobija se Gumbelova raspodjela (slika 1).
Za $\gamma > 0$ dobija se Frešeova raspodjela (smjena $\gamma = 1/\alpha$, slika 2).

[1] Funkcije raspodjele G_1 i G_2 su *istog tipa* ako postoje konstante $a > 0$ i b, takve da za svaki realan x vrijedi $G_2(x) = G_1(ax + b)$

Za $\gamma < 0$ dobija se Vejbulova raspodjela (smjena $\gamma = -1/\alpha < 0$, slika 3). Ove raspodjele definišu se funkcijama raspodjela:

1.Gumbelova rapodjela

$$G_0(x) = e^{-e^{-x}}, \ -\infty < x < \infty \tag{3.61}$$

2. Frešeova raspodjela ($\alpha > 0$)

$$G_{1,\alpha}(x) = \begin{cases} 0 & , \text{za } x < 0 \\ e^{-x^{-\alpha}} & , \text{za } x \geq 0 \end{cases} \tag{3.62}$$

3. Vejbulova raspodjela ($\alpha > 0$)

$$G_{2,\alpha}(x) = \begin{cases} e^{-(-x)^{\alpha}} & , \text{za } x < 0 \\ 1 & , \text{za } x \geq 0 \end{cases} \tag{3.63}$$

Više o osobinama raspodjela ekstremnih vrijednosti može da se vidi u [8]. U vezi sa raspodjelama $G_\gamma(x)$ je još jedna familija raspodjela ekstremnih vrijednosti

$$H_\gamma(x) = 1 + \log G_\gamma(x)$$

koja se definiše kao:

Generalizovana Paretova raspodjela (GPR)

$$H_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-\frac{1}{\gamma}}, & 1 + \gamma x > 0, \gamma \neq 0 \\ 1 - e^{-x}, & x \in \mathbb{R}, \qquad \gamma = 0 \end{cases} \tag{3.64}$$

$$\lim_{\gamma \to 0} H_\gamma(x) = H_0(x) \text{ (eksponencijalna raspodjela)}$$

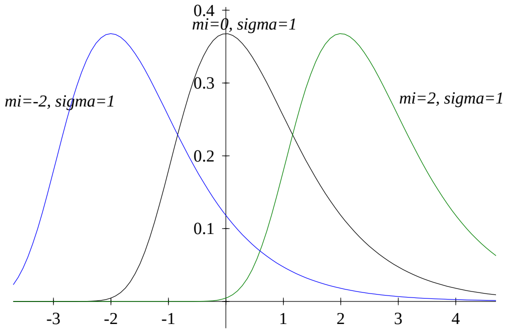
Korištenje familije raspodjela GPR je sugerisano u radu Balkema i de Haan (1974), kao i u radu Pickands III (1975), koji su pokazali da je F u oblasti privlačenja $D(G_\gamma)$, ako i samo ako je rep raspodjele F u određenom smislu 'blizak' repu raspodjele H_γ. Više o klasi ovih raspodjela ekstremnih vrijednosti može da se vidi u [7].

Familija $H_\gamma(x)$ za različite vrijednosti γ svodi se na jednu od tri familije:
Za $\gamma = 0$ dobija se eksponencijalna raspodjela (slika 4).
Za $\gamma > 0$ dobija se Paretova raspodjela (smjena $\gamma = 1/\alpha$, slika 5).
Za $\gamma < 0$ dobija se Beta raspodjela (smjena $\gamma = -1/\alpha < 0$, slika 6).

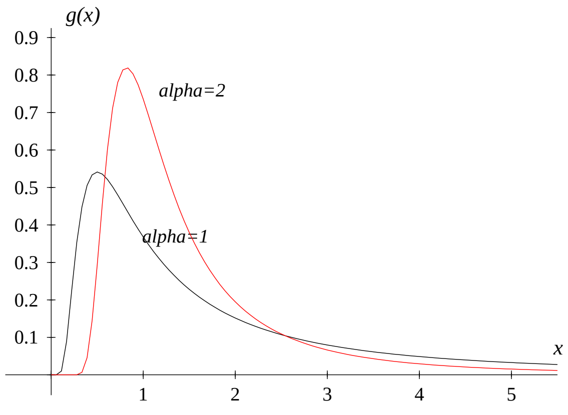
γ - parametrizacija navedenih familija raspodjela može se uopštiti uvođenjem parametra *položaja* μ i *razmjere* $\sigma > 0$ pri čemu se dobijaju kompletna GREV, odnosno kompletna GPR:

$$G_{\gamma,\mu,\sigma}(x) = G_\gamma(\frac{x-\mu}{\sigma}) \tag{3.65}$$

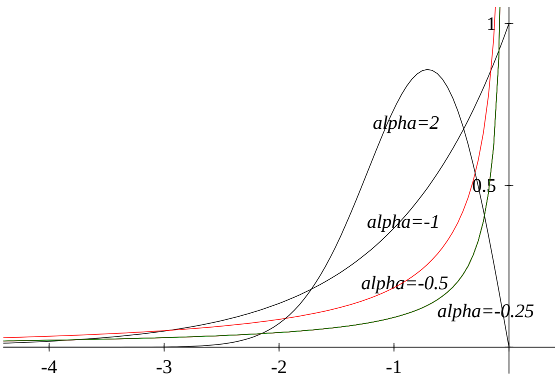
$$H_{\gamma,\mu,\sigma}(x) = H_\gamma(\frac{x-\mu}{\sigma}) \tag{3.66}$$



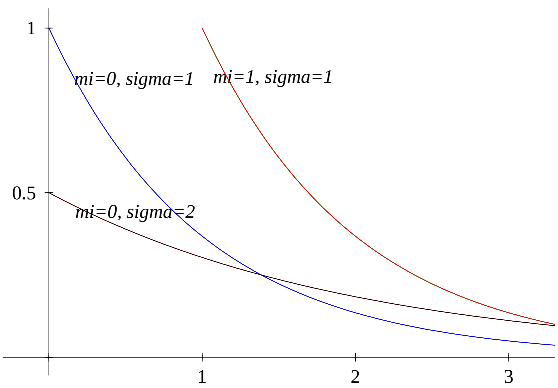
Slika 1: *Grafici gustine* $g_{0,\mu,\sigma}(x) = \frac{1}{\sigma}\exp(-\frac{x-\mu}{\sigma}) \cdot G_{0,\mu,\sigma}(x)$ *Gumbelove raspodjele, za* $\mu = -2, 0, 2$ *i* $\sigma = 1$*.*



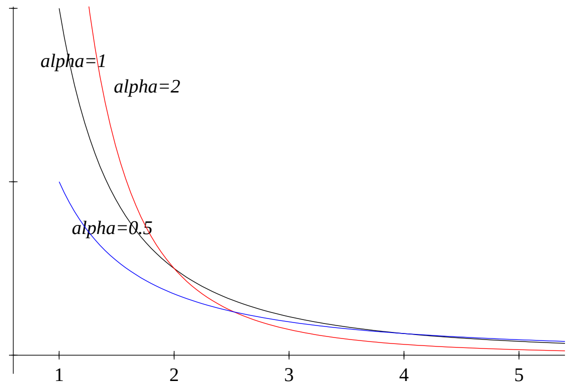
Slika 2: *Grafici gustine* $g_{1,\alpha}(x) = \alpha x^{-(1+\alpha)} e^{-x^{-\alpha}}$ *Frešeove raspodjele, za* $\alpha \in \{1, 2\}$, $\mu = 0$, $\sigma = 1$*.*



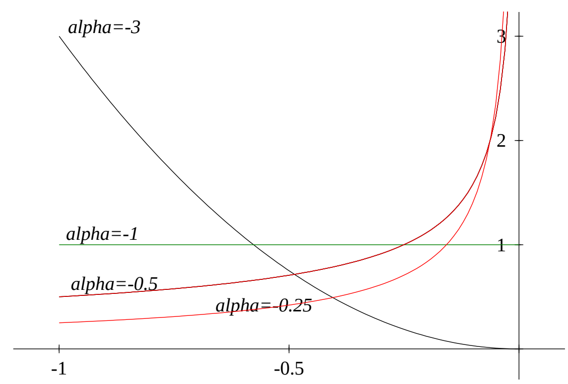
Slika 3: *Grafici gustine* $g_{2,\alpha}(x) = \alpha(-x)^{\alpha-1}e^{-(-x)^{\alpha}}$ *Vejbulove raspodjele, za* $\alpha \in \{-0.25, -0.5, -1, -2\}$, $\mu = 0$, $\sigma = 1$.



Slika 4: *Grafici gustina* $h_{0,\mu,\sigma}(x) = 1 - e^{-\frac{x-\mu}{\sigma}}$ *eksponencijalne raspodjele sa različitim parametrima mjesta i razmjere.*



Slika 5: *Grafici gustine* $h_{1,\alpha,\mu,\sigma}(x) = \frac{1}{\sigma}(1 + \frac{x-\mu}{\sigma\alpha})^{-\alpha-1}$ *Paretove raspodjele, za* $\alpha \in \{0.5, 1, 2\}$ *i* $\mu = 0$, $\sigma = 1$.



Slika 6: *Grafici gustine* $h_{2,\alpha,\mu,\sigma}(x) = \frac{1}{\sigma}(1 - \frac{x-\mu}{\sigma\alpha})^{\alpha-1}$ *Beta raspodjele, za* $\alpha \in \{-0.25, -0.5, -1, -3\}$ *i* $\mu = 0$, $\sigma = 1$.

Ne ograničavajući se na ekstremne statistike poretka tj. razmatrajući u opštem slučaju k-tu donju statitiku ranga $X_{(k)}$ i k-tu gornju statistiku ranga $X_{(n-k+1)}$ u varijacionom nizu, dobija se rezultat da statistike

$$Y_{(k)} = nF_{X_{(k)}}$$

i

$$Z_{(k)} = n(1 - F_{X_{(n-k+1)}})$$

imaju asimptotski pri $n \to \infty$ gama raspodjelu određenu gustinom

$$g(x; p, q) = \frac{q^p}{\Gamma(p)} x^{p-1} e^{-qx},\ x > 0 \tag{3.67}$$

u kojoj su parametri $q = 1,\ p = k$, te je

$$g(x; 1, k) = \frac{1}{\Gamma(k)} x^{k-1} e^{-x},\ x > 0 \tag{3.68}$$

4 Uzoračka medijana i kvantili

Neka je $(X_1, X_2, ..., X_n)$ uzorak uzet iz populacije sa funkcijom raspodjele $F(x)$ i gustinom $f(x)$. Neka je $0 < p < 1$ i neka je x_p kvantil reda p raspodjele obilježja X, tj. $F(x_p) = p$. Pretpostavimo da je funkcija gustine $f(x)$ neprekidna i pozitivna u x_p.

Pomoću statistika ranga $X_{(1)}, X_{(2)}, ..., X_{(n)}$ definiše se ***uzorački kvantil*** K_p reda p kao statistika ranga $X_{(k)}$ kod koje je $k = [np] + 1$.

Raspodjela uzoračkih kvantila za obilježja apsolutno-neprekidnog tipa može da se dobije iz (3.57).

Asimptotski, pri $n \to \infty$, uzorački kvantil K_p ima graničnu normalnu raspodjelu:

$$K_p \to \mathcal{N}(x_p, p(1-p)n(f(x_p))^2),\ n \to \infty \tag{4.69}$$

tj.

$$\sqrt{n}(K_p - x_p) \to \mathcal{N}(0, p(1-p)(f(x_p))^2),\ n \to \infty \tag{4.70}$$

Uzoračka medijana predstavlja specijalan slučaj uzoračkih kvantila K_p, $0 < p < 1$, za $p = \frac{1}{2}$.

Pomoću statistika ranga $X_{(1)}, X_{(2)}, ..., X_{(n)}$ uzoračka medijana definiše se na sledeći način:

$$Me = \begin{cases} X_{(k+1)} & , \text{za } n = 2k+1 \\ \frac{X_{(k)} + X_{(k+1)}}{2} & , \text{za } n = 2k. \end{cases} \tag{4.71}$$

Raspodjela uzoračke medijane za obilježja apsolutno-neprekidnog tipa može takođe da se dobije iz (3.57).

Asimptotski, pri $n \to \infty$, medijana ima graničnu normalnu raspodjelu:

$$Me(x) \sim \mathcal{N}(x_{\frac{1}{2}}, \frac{1}{4} n f^2(x_{\frac{1}{2}})) \tag{4.72}$$

gdje je $x_{\frac{1}{2}}$ medijana za odgovarajuću teorijsku raspodjelu obilježja, tj. $F(x_{\frac{1}{2}}) = \frac{1}{2}$. Asimptotsku raspodjelu uzoračkih kvantila proučavao je još Laplas prije skoro 200 godina. O novijim rezultatima vezanim za ovu raspodjelu može se vidjeti u [11].

NAPOMENA: O nekim karakteristikama raspodjela pomenutih u ovom radu i njihovim graficima gustina, više može da se sazna iz Atlasa raspodela (vidi [2]).

Literatura

[1] **Arnold, B.C.**(1988): Bounds on the expected maximum, *Comm. Statist. Theory and Methods* 17(7), 2135-2150

[2] **Djorić, D.,Jevremović, V.,Mališić, J.,Nikolić-Đorić, V.**(2007): Atlas raspodela, *Građevinski fakultet*, Beograd

[3] **Caraux, G.,Gascuel,O.**(1992): Bounds on distribution functions of order statistics for dependent variates, *Statistics and Probability Letters 14* North-Holland, 103-105

[4] **Cramer, H.**(1946): Mathematical Methods of statistics, *Princeton*

[5] **Ivanović, B.** (1977): Teorija verovatnoće, *Naučna knjiga*, Beograd, 268-284

[6] **Ivković, Z.**(1989): Teorija verovatnoća sa matematičkom statistikom, *Naučna knjiga*, Beograd, 141-143

[7] **Marohn, F.** (2000): Testing Extreme Value Models, *Extremes*, 3:4, 363-384

[8] **Mladenović, P.**(2002): Ekstremne vrednosti slučajnih nizova, *Matematički fakultet*, Beograd, 53-92

[9] **Mladenović, P.**(1995): Verovatnoća i statistika, *VESTA - Matematički fakultet*, Beograd, 240-242

[10] **Stojanović, M. S.**(1980): Matematička statistika, *Naučna knjiga*, Beograd, 54-59

[11] **Thomas S. Ferguson**: Asymptotic Joint Distribution of Sample Mean and a Sample Quantile, *UCLA*

[12] http://home.jesus.ox.ac.uk/∼clifford/a5/chap2/node8.html

[13] http://en.wikipedia.org/wiki/Order-statistic