Savić Jovana 2020/3423

Elektrotehnički fakultet u Beogradu
Matematička statistika 13M081MAST
Prof. dr Milan Merkle
Domaći zadatak 1

Zadatak 1. Dokazati da važi

$$\rho\left(X, aY + b\right) = \operatorname{sgn}(a)\rho\left(X, Y\right)$$

gde su a i b konstante, a X i Y proizvoljne slučajne promenljive.

Rešenje. Koeficijent korelacije između slučajnih promenljivih sa pozitivnim varijansama definiše se sa

$$\rho\left(X, Y\right) = \frac{\operatorname{Cov}\left(X, Y\right)}{\sqrt{\operatorname{Var}\left(X\right)}\sqrt{\operatorname{Var}\left(Y\right)}}.$$

U [1] su u teoremama 4.4. i 4.5. date sledeće osobine:

$$\begin{aligned}
&\operatorname{Var}\left(aX\right) = a^2 \operatorname{Var}\left(X\right), \\
&\operatorname{Var}\left(X + a\right) = \operatorname{Var}\left(X\right), \\
&\operatorname{Cov}\left(aX, bY\right) = ab \operatorname{Cov}\left(X, Y\right), \\
&\operatorname{Cov}\left(X + a, Y + b\right) = \operatorname{Cov}\left(X, Y\right)
\end{aligned}$$

gde su a i b realne konstante.

Polazeći od definicije i korišćenjem gore navedenih osobina dobijamo dokaz:

$$\begin{aligned}
\rho\left(X, aY + b\right) &= \frac{\operatorname{Cov}\left(X, aY + b\right)}{\sqrt{\operatorname{Var}\left(X\right)}\sqrt{\operatorname{Var}\left(aY + b\right)}} \\
&= \frac{a \operatorname{Cov}\left(X, Y\right)}{\sqrt{\operatorname{Var}\left(X\right)}\sqrt{a^2 \operatorname{Var}\left(Y\right)}} \\
&= \frac{a}{|a|}\rho\left(X, Y\right) \\
&= \operatorname{sgn}(a)\rho\left(X, Y\right)
\end{aligned}$$

□

Zadatak 2. Cena akcije jedne kompanije u Kardasiji n-tog radnog dana je X_n leka, $n = 0, 1, \ldots$ a priraštaji (razlika u ceni između danas i juče) su nezavisne slučajne promenljive

$$D_n = X_n - X_{n-1}, \quad n = 1, 2,$$

za koje se zna da imaju matematičko očekivanje 0 i varijansu 1/4. Cena akcija u početnom danu ($n = 0$) je bila 30 leka.

a) Naći matematičko očekivanje i varijansu za X_n u zavisnosti od n. **b)** Primenom nejednakosti Čebiševa naći donju granicu verovatnoće da cena akcije dana $n = 10$ bude veća od 25, a manja od 35 leka. **c)** Primenom centralne granične teoreme naći približnu vrednost verovatnoće da cena akcije dana $n = 100$ bude veća od 25, a manja od 35 leka.

Rešenje. **a)** Na osnovu

$$X_n = X_{n-1} + D_n, \quad X_0 = 30$$

nalazimo

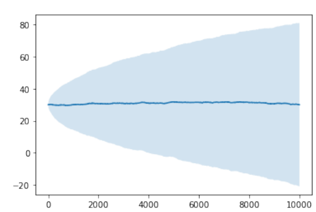
$$X_n = X_0 + \sum_{i=1}^{n} D_i.$$

Odavde je:

$$\begin{aligned} E\left[X_n\right] &= E\left[X_0 + \sum_{i=1}^{n} D_i\right] \\ &= X_0 + \sum_{i=1}^{n} E\left[D_i\right] = X_0 = 30. \end{aligned}$$

Na sličan način, korišćenjem činjenice da se varijanse nezavisnih promenljivih sabiraju, nalazimo i traženu varijansu

$$\begin{aligned} \mathrm{Var}(X_n) &= \mathrm{Var}\left(X_0 + \sum_{i=1}^{n} D_i\right) \\ &= \sum_{i=1}^{n} \mathrm{Var}(D_n) = \frac{n}{4}. \end{aligned}$$



Na slici pored je prikazan rezultat simulacije koji potvrđuje tačnost dobijenog rešenja. Svetlijom bojom je označena standardna devijacija, a tamnijom srednja vrednost.

b) Nejednakost Čebiševa tvrdi da ako postoji $\mathrm{Var}(X)$, tada je

$$P\left(|X - E\ X| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Potrebno je pronaći donju granicu verovatnoće događaja

$$(X_{10} > 25)\ \wedge\ (X_{10} < 35)\ \Leftrightarrow\ |X_{10} - 30| < 5\ \Leftrightarrow\ |X - E\ X_{10}| < 5.$$

Na osnovu Čebiševe nejednakosti imamo

$$P\left(|X_{10} - E\ X_{10}| \geq 5\right) \leq 0.1.$$

Odavde nalazimo donju granicu verovatnoće traženog događaja:

$$P\left(|X_{10} - 30| < 5\right) = 1 - P\left(|X_{10} - 30| \geq 5\right) \geq 0.9.$$

c) *(Centralna granična teorema)* Neka je X_i, $i = 1, 2, \ldots$ niz nezavisnih slučajnih promenljivih sa istom raspodelom, $E\ X = \mu$, $\mathrm{Var}\ X = \sigma^2$. Normirani zbir

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

ima $E\ Z_n = 0$, Var $Z_n = 1$.

Niz Z_n konvergira u raspodeli ka $Z \sim \mathcal{N}(0,1)$, to jest

$$\lim_{n\to\infty} P(Z_n \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \mathrm{d}t$$

za svako $x \in \mathbb{R}$.

Definišemo Z_{100} kao

$$\begin{aligned} Z_{100} &= \frac{D_1 + \cdots + D_{100} - 100\mu_D}{10\sigma_D} \\ &= \frac{D_1 + \ldots D_{100}}{5} \\ &= \frac{X_{100} - X_0}{5}. \end{aligned}$$

Na osnovu CTG je $Z_{100} \sim \mathcal{N}(0,1)$.

Potrebno je pronaći verovatnoću događaja

$$25 < X_{100} < 35 \ \Leftrightarrow 25 < 5Z_{100} + 30 < 35 \ \Leftrightarrow -1 < Z_{100} < 1.$$

Imamo

$$\begin{aligned} P(-1 < Z_{100} < 1) &= P(Z_{100} < 1) - P(Z_{100} < -1) \\ &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6828 \end{aligned}$$

□

Zadatak 3. Broj osoba koje posete neku izložbu u toku dana je Poissonova slučajna promenljiva sa parametrom λ. Verovatnoća da je posetilac muškarac je p, a verovatnoća da je žena je $1-p$. Neka je X, odnosno Y , broj muškaraca, odnosno žena koje posete izložbu u toku dana. Odrediti zakon raspodele slučajnog vektora (X,Y), a zatim naći raspodele za X i Y . Da li su X i Y nezavisne slučajne promenljive?

Rešenje. Broj muškaraca koji poseti izložbu u toku jednog dana u kome je ukupan broj posetilaca n je dat binomnom raspodelom

$$f_X(k,n) = P(X = k \mid Z = n) = \binom{n}{k} p^k (1-p)^{n-k},$$

kao i broj žena

$$f_Y(k,n) = P(Y = k \mid Z = n) = \binom{n}{k} (1-p)^k p^{n-k}.$$

Broj posetilaca u toku dana Z je takođe slučajna promenljiva koja ima Poisson-ovu raspodelu

$$f_Z(n,\lambda) = P(Z = n) = \frac{\lambda^n e^{-\lambda}}{n!}.$$

Na osnovu zakona totalne verovatnoće imamo:

$$f_X(k) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \ \frac{\lambda^n e^{-\lambda}}{n!}$$

i

$$f_Y(k) = \sum_{n=k}^{\infty} \binom{n}{k} (1-p)^k p^{n-k} \ \frac{\lambda^n e^{-\lambda}}{n!}$$

Prvi izraz sređujemo na sledeći način:

$$\begin{aligned} f_X(k) &= \sum_{n=k}^{\infty} \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \ \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \frac{p^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}\lambda^n}{(n-k)!} \end{aligned}$$

Uvođenjem smene $m = n - k$ dobijamo:

$$\begin{aligned} f_X(k) &= \frac{p^k e^{-\lambda}}{k!} \sum_{m=0}^{\infty} \frac{(1-p)^m \lambda^{m+k}}{m!} \\ &= \frac{(p\lambda)^k e^{-\lambda}}{k!} \sum_{m=0}^{\infty} \frac{((1-p)\lambda)^m}{m!} \\ &= \frac{(p\lambda)^k e^{-\lambda}}{k!} e^{\lambda(1-p)} \\ &= \frac{(p\lambda)^k e^{-p\lambda}}{k!} \end{aligned}$$

Odavde vidimo da je $X \sim \mathrm{Pois}(p\lambda)$, i analogno $Y \sim \mathrm{Pois}((1-p)\lambda)$.

Ove dve slučajne promenljive su nezavisne. Ovo dokazujemo na sledeći način:

$$\begin{aligned} P(X = k,\ Y = l) &= P(X = k\ ,\ Z = k + l) \\ &= P(X = k \mid Z = k + l) P(Z = k + l) \\ &= \binom{k+l}{k} p^k (1-p)^l \ \frac{\lambda^{k+l} e^{-\lambda}}{(k+l)!} \\ &= \frac{(k+l)!}{k!\ l!}\ p^k (1-p)^l \ \frac{\lambda^{k+l} e^{-\lambda(1-p+p)}}{(k+l)!} \\ &= \frac{(p\lambda)^k e^{-\lambda p}}{k!} \ \frac{((1-p)\lambda)^l e^{-\lambda(1-p)}}{l!} \\ &= P(X = k)\ P(Y = l). \end{aligned}$$

Ovo je ujedno i zakon raspodele slučajnog vektora (X, Y).

$\square$

Zadatak 4. Na osnovu nezavisnog uzorka $X_1, X_2, \dots, X_n$ iz normalne raspodele sa nepoznatim matematičkim očekivanjem μ i poznatom varijansom σ^2 dobijena je aritmetička sredina uzorka $\hat{\mu}$.

a) Koju raspodelu ima slučajna promenljiva $T = \sqrt{n}\frac{\hat{\mu}-\mu}{\sigma}$ (da li nam treba CGT?)? Zatim izraziti $P\left(|\hat{\mu}-\mu| < \varepsilon\right)$ preko funkcije raspodele za T.

b) Za $\sigma = 0.2$ koliko treba da bude najmanje n da bi verovatnoća događaja $|\hat{\mu}-\mu| < 0.1$ bila veća od 0.9?

Rešenje. **a)** Promenljivu T možemo da transformišemo kao

$$T = \sqrt{n}\frac{\hat{\mu}-\mu}{\sigma} = \frac{n(\hat{\mu}-\mu)}{\sigma\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Na osnovu CTG slučajna promenljiva T konvergira u raspodeli ka $T \sim \mathcal{N}(0,1)$. Imamo da je

$$|\hat{\mu}-\mu| < \varepsilon \iff |T| < \varepsilon\frac{\sqrt{n}}{\sigma}.$$

Na osnovu ovoga nalazimo približnu vrednost tražene verovatnoće

$$P\left(|\hat{\mu}-\mu| < \varepsilon\right) \approx 2\Phi\left(\varepsilon\frac{\sqrt{n}}{\sigma}\right) - 1.$$

b) Na osnovu nejednakosti Čebiševa je

$$P\left(|\hat{\mu}-\mu| < \varepsilon\right) = 1 - P\left(|\hat{\mu}-\mu| > \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

gde smo iskoristili

$$\mathrm{Var}(\hat{\mu_n}) = \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Iz

$$1 - \frac{0.2^2}{0.1^2\ n} > 0.9$$

nalazimo

$$n > 40.$$

Ako koristimo gore izvedenu aproksimaciju verovatnoće imamo

$$2\Phi\left(\varepsilon\frac{\sqrt{n}}{\sigma}\right) - 1 > 0.9.$$

Odavde je

$$\varepsilon\frac{\sqrt{n}}{\sigma} \geq K_{0.95}$$
$$n \geq \frac{\sigma^2}{\varepsilon^2}K_{0.95}^2$$
$$n \geq 10.7584$$

□

Zadatak 5. Slučajna promenljiva X ima Bin$(100, \frac{1}{2})$. raspodelu. Pokazati da se X može predstaviti kao zbir 100 nezavisnih slučajnih promenljivih sa istom raspodelom, sa detaljnim obrazloženjem. Zatim primenom centralne granične teoreme pokazati da se raspodela za X može aproksimirati normalnom raspodelom sa parametrima koje treba odrediti. Primenom izvedene aproksimacije naći verovatnoću $P(X = 60)$.

Rešenje. Ako su $X_1, \ldots, X_n$ nezavisne slučajne promenljive takve da je $P(X_i = 1) = p$ i $P(X_i = 0) = 1 - p$, njihova karakteristična funkcija je

$$\phi_{X_i}(t) = (1 - p) + pe^{it}.$$

Pošto su ove promenljive nezavisne, karakteristična funkcija njihovog zbira je jednaka proizvodu njihovih karakterističnih funkcija

$$\phi_{X_1+\cdots+X_n} = \left((1 - p) + pe^{it}\right)^n.$$

Dobijena karakteristična funkcija predstavlja karakterističnu funkciju binomne raspodele. Prema tome, binomna raspodela se može predstaviti kao zbir nezavisnih slučajnih promenljivih sa istom raspodelom.

(Moivre-Laplace-ova teorema) Neka je Y_n slučajna promenljiva sa binomnom raspodelom sa parametrima n i p. Tada za svako $x \in R$ važi da je

$$\lim_{n\to\infty} P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \mathrm{d}t.$$

Iznad je prikazano da se binomna raspodela sa parametrima n i p može predstaviti sumom n slučajnih promenljivih sa *Bernoulli*-jevom raspodelom sa verovatnoćom uspeha p. Za ove slučajne promenljive je $EX_i = p$ i $\mathrm{Var}(X_i) = p(1 - p)$. Zamenom ovih vrednosti u CTG dobijamo:

$$Z_n = \frac{Y_n - np}{\sqrt{p(1-p)n}}$$

gde Z_n konvergira u raspodeli ka $Z \sim \mathcal{N}(0, 1)$.

Aproksimacijom binomne raspodele normalnom raspodelom nalazimo $P(X = 60)$ kao

$$\begin{aligned} P(X = 60) &= P(59.5 \leq X \leq 60.5) = P(59.5 \leq 5Z_n + 50 \leq 60.5) \\ &= P(1.9 \leq Z_n \leq 2.1) = 0.0109. \end{aligned}$$

□

Literatura

[1] Milan Merkle. *Verovatnoća i statistika za inženjere i studente tehnike*. Akademska misao, 2010.