Verjetnost in statistika - zapiski s predavanj prof. Drnovška

Tomaž Poljanšek

študijsko leto 2022/23

Kazalo

Denimo sedaj, da sta $p_{X,Y}$ in p_Y zvezni funkciji. Tedaj je $F_X(X \mid Y = y) = \frac{\frac{\partial}{\partial y}F_{(X,Y)}(x,y)}{F'_Y(y)} = \frac{1}{p_Y(y)}\int_{-\infty}^{x} p_{(X,Y)}(x,v)dv$
Če vpeljemo pogojno pogojno gostoto $p_X(x \mid Y = y) := \frac{p_{(X,Y)}(x,y)}{p_Y(y)}$, je torej

$$F_{(X,Y)}(x \mid Y = y) = \int_{-\infty}^{x} p_X(u \mid y)du$$

Pogojno matematično upanje slučajne spremenljivke X glede na dogodek $(Y = y)$ je

$$E(X \mid Y = y) := \int_{-\infty}^{\infty} x \cdot p_X(x|y)dx = \frac{1}{p_Y(y)} \cdot \int_{-\infty}^{\infty} xp_{(X,Y)}(x,y)dx$$

Vpeljimo regresijsko funkcijo $l(y) := E(X \mid Y = y)$, definirano na zalogi vrednosti slučajne spremenljivke Y. Tako dobimo novo slučajno spremenljivko $E(X \mid Y) := l(y)$: pogojno matematično upanje slučajne spremenljivke X glede na slucajno spremenljivko Y.
Kot v diskretnem primeru se pokaže enakost $E(E(X \mid Y)) = E(X)$

Primer. $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$
Robna gostota za Y je $N(\mu_y, \sigma_y)$
Zato je pogojna gostota

$$p_X(x \mid y) = \frac{p_{(X,Y)}(x,y)}{p_y(x)} \overset{\text{D.N.}}{=\cdots=} \frac{1}{\sigma_x\sqrt{(2\pi)(1-\rho^2)}} exp(-\frac{1}{2(1-\rho)^2}(\frac{x-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y})^2)$$

torej je $N(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y), \sigma_x\sqrt{1-\rho^2})$
Eksponent: $\frac{1}{2(1-\rho^2)}\sigma_x^2(x - (\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)))^2$
$\implies l(y) = E(X \mid Y = y) = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)$ - 1. parameter
$= \alpha + \beta y : \beta = \rho\frac{\sigma_x}{\sigma_y}, \alpha = \mu_x - \frac{\sigma_x}{\sigma_y} \cdot \mu_y$
Torej je $E(x \mid y) = \alpha + \beta y$

Primer. Meritev onesnaženosti zraka
Slučajna spremenljivka X meri koncentracijo ogljikovih delcev (v $\mu g/m^3$), Y pa koncentracijo ozona (v $\mu l/l = ppm$)
Podatki kažejo, da ima (X,Y) približno dvorazsežno normalno porazdelitev, $\mu_x = 10.7, \sigma_x^2 = 29, \mu_y = 0.1, \sigma_y^2 = 0.02, \rho = 0.72$
Koncentracija ozona je škodljiva zdravju, če je ≥ 0.3
Denimo, da naprava za merjenje ozona odpove, koncentracija škodljivih delcev je $X = 200$

a kolikšna je pričakovana koncentracija ozona?

b kolikšna je verjetnost, da je stopnja ozona zdravju skodljiva a

$$E(Y \mid X = x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x) = 0.1 + 0.72\sqrt{\frac{0.02}{29}}(20 - 10.7) \doteq 0.28$$

b Pogojna porazdelitev $Y \mid X = x$ je $N(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_x\sqrt{1 - \rho^2}) = N(0.28, 0.1)$

$$P(Y > 0.3 \mid X = 20) = 1 - P(Y \leq 0.3 \mid X = 20) = 1 - F_{N(0,1)}(\frac{0.3 - 0.28}{0.1}) \doteq 0.42$$

0.1 Višji momenti in vrstilne karakteristike

Definicija 0.1 (Momenti)**.** Naj bo $k \in \mathbb{N}$ in $a \in \mathbb{R}$. Moment reda k glede na točko a je $m_k(a) := E((X - a)^k)$ (če obstaja)

Za a obicajno vzamemo

1. $a = 0$: $z_k := m_k(0) = E(X^k)$ začetni moment reda k
2. $a = E(X)$: $m_k := m_k(E(X))$ cenralni moment reda k

Ocitno je $z_1 = E(X), m_2 = D(X)$

Trditev 0.2. Če $\exists m_n(a)$, potem obstajaj tudi moment $m_k(a)$ za vse $k < n$

Dokaz. (V zveznem primeru):

$$E((X-a)^k) = \int_{-\infty}^{\infty}(x-a)^k p_X(x)dx = \int a - 1^{a+1}(X-a)^k p_X(x)dx + \int_{(-\infty,a-1)\cup(a+1,\infty)}(x-a)^k p_X(x)d$$
$$\leq \int_{-\infty}^{\infty} p_X(x)dx + \int_{(-\infty,a-1)\cup(a+1,\infty)}(x - a)^k p_X(x)dx \leq$$
$$\leq 1 + E((X - a)^k) < \infty$$

■

Trditev 0.3. Če obstaja zacetni moment z_n, potem obstaja $m_n(a)$ glede na poljubno točko $a \in \mathbb{R}$

Dokaz.

$$E((X - a)^n) \leq E((|X| + |a|)^n) = \sum_{k=0}^{n}\binom{n}{k}E(a)^{n-k} \cdot E(|X|^k) < \infty$$

■

Centralne momente lahko izrazimo z začetnimi:

$$m_n(a) = E((X-a)^n) = \sum_{k=0}^{n} \binom{n}{k} (-a)^{n-k} E(X^k)$$

$$a = E(X) \implies m_k = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} z_1^{n-k} z_k$$

Asimetrija slučajne spremenljivke X je $A(X) := E(X_s^3) = E((\frac{X-E(X)}{\sigma_x})^3) = \frac{m_3}{m_2^{\frac{3}{2}}}$ $m_2 = \sigma^2 = D(X)$
$A(N(\mu, \sigma)) = 0$, ker $A(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 e^{-\frac{1}{2}x^2} dx$
Sploščenost (kurtozis) $K(X) := E(X_s^4) = \frac{m_4}{m_2^2}$
$K(N(\mu, \sigma)) = 3$
Ce momenti ne obstajajo (npr. že $E(X)$ ne), potem si lahko pomagamo z vrstilnimi karakteristikami

Definicija 0.4 (Mediana)**.** Mediana slučajne spremenljivke X je vsaka vrednost $x \in \mathbb{R}$, za katero velja $P(X \leq x) \leq \frac{1}{2}$ in $P(Y \geq x) \geq \frac{1}{2}(1 - P(X < x) = 1 - F(x-))$

Če je F porazdelitvena funkcija za X, je to ekvivalentno s pogojem $F(x-) \leq \frac{1}{2} \leq F(x)$
Če je X zvezno porazdeljena slučajna spremenljivka, dobimo $F(X) = \frac{1}{2}$ oz. $\int_{-\infty}^{\infty} p(t)dx = \frac{1}{2}$

Te vrednosti (lahko jih je več) označimo z $X_{\frac{1}{2}}$

Primer.

- $X \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{5} & \frac{4}{5} \end{pmatrix}$
  $x_{\frac{1}{2}} = 1, E(X) = \frac{4}{5}$
- $X : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{2}{4} \end{pmatrix}$
  Mediane so $[0, 1]$
-
- $X \sim N(0, 1)$
  $x_{\frac{1}{2}} = \mu = E(X)$

Definicija 0.5 (Kvantil)**.** Kvantil reda p ($p \in (0,1)$) je vsaka vrednost x_p, za katero velja $P(X \leq x_p) \geq p$ in $P(X \geq x_p) \geq 1-p$
Ekvivalentno je $F(x_p-) \leq p \leq F(x_p)$

Če je X zvezno porazdeljena, je pogoj $F(x_p) = p$ t.j. $\int_{-\infty}^{\infty} p(t)dt = p$

- Kvartili: $X_{\frac{1}{4}}, X_{\frac{2}{4}}, X_{\frac{3}{4}}$
- Percentili: $X_{\frac{1}{100}}, X_{\frac{2}{100}}, \cdots X_{\frac{99}{100}}$

Primer. Telesna višina odraslih moških

Definicija 0.6 ((Semiinter)kvartilni razmik)**.** $s := \frac{1}{2}(x_{\frac{3}{4}} - x_{\frac{1}{4}})$

je nadomestek (analog) za standardno deviacijo

Primer.

- $X \sim N(0,1)$
  $X_{\frac{1}{2}} = 0$
  $\int_{-\infty}^{\frac{1}{4}} p(t)dt = \frac{1}{4} \xRightarrow{\text{tabelca}} x_{\frac{1}{4}} \doteq -0.67$
  $\xRightarrow{\text{simetrija}} x_{\frac{3}{4}} \doteq 0.67 \implies s = 0.67, \sigma(x) = 1$

- X naj ima Cauchyjevo porazdelitev
  $p(x) = \frac{1}{\pi(1+x^2)}$
  $x_{\frac{1}{2}} = 0$
  Momenti ne obstajajo

$$\int_{-\infty}^{x_{\frac{1}{4}}} \frac{1}{\pi}\frac{1}{1+x^2}dx = \frac{1}{4}$$
$$\frac{1}{\pi} \arctan x|_{x=-\infty}^{x_{\frac{1}{4}}} = \frac{1}{4}$$
$$\frac{1}{\pi} arctan x_{\frac{1}{4}} + \frac{1}{2} = \frac{1}{4}$$
$$arctan x_{\frac{1}{4}} = \frac{1}{4} \implies x_{\frac{1}{4}} = -1$$
$$\xRightarrow{\text{simetrija}} x_{\frac{3}{4}} = 1, s = 1$$

0.2 Rodovne funkcije

Definicija 0.7. Naj bo X slučajna spremenljivka z vrednostmi v $\mathbb{N} \cup \{0\}$:
$p_k = P(X = k) k = 0, 1, 2 \cdots$ $p_k \geq 0, \sum_{k=0}^{\infty} = 1$
Rodovna funkcija skučajne spremenljivke X je

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + \cdots = \sum_{k=0}^{\infty} p_k \cdots s^k$$

za $\forall s \in \mathbb{R}$, za katere vrsta absolutno konvergira.

Očitno je $G_X(0) = p_0, G_X(1) = \sum_{k=0}^{\infty} p_k = 1$
Ker je $s^X : \begin{pmatrix} s^0 & s^1 & s^2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}$, je $G_X(s) = E(s^X)$
Za $s \in [-1, 1]$ velja $|p_k \cdot s^k| \leq P_k$ in $\sum_{k=0}^{\infty} p_k = 1$. Zato je vrsta konvergentna, če je $|s| \leq 1$. Torej je konvergenčni radij vrste vsaj 1

Primer.

- $X \sim geo(p)$, $p \in (0, 1)$

$$\begin{aligned} p_k &= P(X = k) = p \cdot q^{k-1} \; k = 1, 2, 3 \cdots \\ G_X(s) &= \sum_{k=1}^{\infty} p \cdot q^{k-1} s^k = ps \sum_{k=0}^{\infty} (qs)^{k-1} \\ &= ps \frac{1}{1 - qs} \end{aligned}$$

  konvergira, ko $|qs| < 1 \Leftrightarrow |s| < \frac{1}{|q|} =: R$

- $p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$$G_X(s) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} =$$

$$= e^{-\lambda} \cdot e^{\lambda s} = e^{\lambda(s-1)}$$

  $R = \infty \; \forall s \in \mathbb{R}$

Iz teorije Taylorjevih vrst sledi

Izrek 0.8 (O eniličnosti)**.** Naj imata X in Y rodovni funkciji G_X in G_Y. Potem je $G_X(s) = G_Y(s)$ za $\forall s \in [-1, 1] \leftrightarrow P(X = k) = P(Y = k)$ za vse $k = 0, 1, 2 \cdots$
Tedaj velja $P(X = k) = \frac{1}{k!} G_X^k(0)$

$G_X(s) = \sum_{k=0}^{\infty} p_k s^k$, $p_k = P(X = k)$
Naj ima rodovna funkcija G_X slučajne spremenljivke X konvergenčni radij R > 1. Potem za $\forall s \in (-R, R)$ velja $G_X^{'}(s) = \sum_{k=1}^{\infty} k \cdot p_k s^{k-1}$
Če postavimo $s = 1$, dobimo $G^{'}(1) = \sum_{k=1}^{\infty} k \cdot p_k = E(X)$

Izrek 0.9. Naj ima X rodovno funkcijo $G_X(s)$ in naj bo $n \in \mathbb{N}$. Potem je

$$G_X^n(1-) \equiv \lim_{s \nearrow 1} G_X^n(s) = E(X(X-1)(X-2)\cdots(X-N+1))$$

Dokaz. Za $\forall s \in [0, 1)$ je $G_X^n(s) = \sum_{k=n}^{\infty} k(k-1)(k-2)\cdots(k-n+1)p_k s^{k-n+1} =$

$$= E(X(X-1)(X-2)\cdots(X-n+1)\cdot s^{X-n})$$

Ko gre $s \uparrow 1$, z uporabo Abelove leme dobimo

$$\lim_{s \nearrow 1} G_X^n(s) = \lim_{s \nearrow 1} \sum_{k=n}^{\infty} k(k-1)\cdot(k-n+1) =$$

$$\overset{\text{Abelova lema}}{=} \sum_{k=n}^{\infty} lim_{s \nearrow 1} k(k-1)\cdot(k-n+1) = \sum_{k=n}^{\infty} k(k-1)\cdot(k-n+1)p_k = E(X(X-1)\cdots(X-n+1))$$

∎

Posledica 0.10.

$$E(X) = G_X^{'}(1-)$$

$$D(X) = E(X^2) - (E(X))^2 = E(X(X-1)) + E(X) - (E(X))^2 = G_X^{(2)}(1-) + G_X^{(1)}(1) - (G_X^{(1)}(1-))^2$$

Izrek 0.11. Naj bosta X in Y neodvisni slučajni spremenljivki z rodovnima funkcijama G_X in G_Y. Potem je $G_{X+Y}(s) = G_X(s) \cdot G_Y(s)$ za $s \in [-1, 1]$

Dokaz. $G_{X+Y}(s) = E(s^{X+Y}) = E(s^X \cdot s^Y) \overset{\text{izrek}}{=} E(s^X) \cdot E(s^Y) = G_X(s) \cdot G_Y(s)$, saj sta s^X in s^Y neodvisni slučajni spremenljivki ∎

Posplo"sitev 0.12. Če so $X_1, X_2 \cdots X_n$ neodvisne slučajne spremenljivke, potem je za vse $s \in [-1, 1] G_{X_1 + \cdots + X_n}(s) = G_{X_1}(s) \cdot \cdots \cdot G_{X_n}(s)$.
Če so $X_1, X_2 \cdots X_n$ enako porazdeljene in neodvisne, potem je

$$G_{X_1 + \cdots + X_n}(s) = (G_X(s))^n$$

Izrek 0.13. Naj bodo za $\forall n \in \mathbb{N}$ slučajne spremenljivke $N, X_1, X_2 \cdots X_n$ neodvisne. Naj ima N rodovno funkcijo G_N, X_n pa rodovno funkcijo G_X. Potem ima slučajna spemenljivka $S := X_1 + X_2 + \cdots + X_n$ rodovno funkcijo enako $G_S = G_N \circ G_X$ oz. $G_S(s) = G_N(G_X(s))$ za $s \in [-1, 1]$

To je posplošitev formule dd: $P(N=n)=1, G_N(s)=1\cdot s^n = s^n$

Dokaz. Zaradi neodvisnosti imamo $P(S=k)=\sum_{n=0}^{\infty} P(S=k, N=n)=$

$$= \sum_{n=0}^{\infty} P(N=n, X_1+\dots+X_n=k) \overset{\text{neodvisnost}}{=} \sum_{n=0}^{\infty} P(N=n)\cdot P(X_1+\dots+X_n=k)$$

Zato je

$$G_S(s) = \sum_{k=0}^{\infty} P(S=k)\cdot s^k = \sum_{k=0}^{\infty}\sum_{n=1}^{\infty} P(N=n)\cdot P(X_1+\dots+X_n=k)\cdot s^k =$$

$$= \sum_{n=1}^{\infty} P(N=n)(\sum_{k=0}^{\infty} P(X_1+\dots+X_n=k)\cdot s^k) =$$

$$\overset{G_{X_1+\dots+X_n}(s)^{\text{neodvisnost}}(G_X(s)^n)}{=} \sum_{n=1}^{\infty} P(N=n)\cdot (G_X(s))^n = G_N(G_X(s))$$

za vse $s\in[-1,1]$ ■

Posledica 0.14. Pri predpostavkah iz izreka velja Waldova enakost:

$$E(S)=E(N)\cdot E(X)$$

Dokaz.

$$G_S(s) = G_N(G_X(s)) \forall s\in[-1,1] \tag{1}$$
$$E(S) = G_s^{'}(1-) = G_N^{'}(G_X(1-))\cdot G_X^{'}(1-) = E(N)\cdot E(X) \tag{2}$$

■

Primer. Kokoš, jajca, piščanci
N jajc, $N\sim Poi(\lambda)$
K je število piščancev
Definiramo $X_i=1$ dogodek, da se iz i-tega jajca izvali piščanec, sicer $X_i=0$.
Potem je $X_i : \begin{pmatrix} 0 & 1 \\ q & p\end{pmatrix}$, $q=1-p$ in X_i so neodvisne slučajne spremenljivke.
Očitno je $K = X_1+X_2+\dots+X_n$
Ker je $G_N(s)=e^{\lambda(s-1)}$ in $G_X(s)=q\cdot s^0+p\cdot s = q+ps$, je po izreku $G_K(s)=G_N(G_X(s))=e^{\lambda(q+ps-1)}=e^{\lambda(ps-p)}=e^{\lambda p(s-1)}\forall s\in[-1,1]$, zato je $K\sim Poi(\lambda p)$

0.3 Momentno rodovna funkcija

Definicija 0.15 (Momentno rodovna funkcija)**.** Momentno rodovna funkcija je $M_X(t) = E(e^{tX})$ za $t \in \mathbb{R}$, za katere obstaja matematično upanje

V primeru zvezne porazdelitve je $M_X(t) = \int_{-\infty}^{\infty} e^{tx} p_X(x) dx$
To je Laplaceova transformacija funkcije p_X
V diskretnem primeru $X : \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$ je $M_X(t) = \sum_i e^{tx} p_i$

V posebnem primeru, ko ima X nenegative celoštevilske vrednosti, je $M_X(t) = \sum_{i=0}^{\infty} e^{it} p_i =$

$$= \sum_{i=0}^{\infty} p_i (e^t)^i = G_X(e^t) \ (M_X(t) = E((e^t)^X) = G_X(e^t))$$

$$G_X(s) = E(s^X)$$

Očitno je $M_X(0) = E(e^0) = E(1) = 1$

Primer.

$$X \sim N(0,1)$$
$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx \cdot e^{-\frac{t^2}{2}} =$$
$$= e^{\frac{t^2}{2}} \forall t \in \mathbb{R}$$

ker je $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}}$ gostota za $N(0,1)$

Izrek 0.16. Naj bo $M_X(t) < \infty$ (obstaja, $< \infty$ zato, ker je $e^t > 0$) za $\forall t \in (-\delta, \delta)$ pri nekem $\delta > 0$. Potem je porazdelitev za X natanko določena z M_X, vsi začetni momenti obstajajo, $z_k = E(X^k) = M_X^k(0)$ za $\forall k \in \mathbb{N}$ in velja $M_X(t) = \sum_{k=0}^{\infty} \frac{z_k}{k!} t^k$ za $\forall t \in (-\delta, \delta)$

Dokaz. (bistvo)

$$M_X(t) = E(e^{t \cdot X}) = E(\sum_{k=0}^{\infty} t^k \frac{x^k}{k!}) =$$
$$\sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k = \sum_{k=0}^{\infty} \frac{z^k}{k!} t^k$$

■

Trditev 0.17. $M_{aX+b}(t) = e^{bt} M_X(at), a \neq 0, b \in R$

Dokaz. $M_{aX+b}(t) = E(e^{t(aX+b)}) = E(e^{(at)X} \cdot e^{bt}) = e^{bt} M_X(at)$ ■

Izrek 0.18. Če sta X in Y neodvisni slučajni spremenljivki, potem je $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

Dokaz. $M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{t^X} \cdot e^{tY}) \overset{e^{tX},\, e^{tY} \text{ neodvisni}}{=}$
$= E(e^{t^X}) \cdot E(e^{tY}) = M_X(t) \cdot M_Y(t)$ ■

Trditev 0.19. Naj bosta X in Y neodvisni slučajni spremenljivki in $X \sim N(\mu_x, \sigma_x), Y \sim N(\mu_y, \sigma_y)$. Potem je $X + Y \sim N(\mu_x + \mu_y, \sqrt{\sigma_x^2 + \sigma_y^2})$

Dokaz. Ker je

$$U := \frac{X - \mu_x}{\sigma_x} = \frac{X - E(X)}{\sigma(X)} \sim N(0,1)$$

(standardizacija), je

$$X = \sigma_x \cdot U + \mu_x$$

in zato je

$$M_X(t) = e^{\mu_x t} \cdot M_U(\sigma_x t)$$

po zadnji trditvi. Potem je

$$M_U(t) = e^{\frac{t^2}{2}}$$

je

$$M_X(t) = e^{\mu_x t} \cdot e^{\frac{\sigma_x^2 t^2}{2}} = e^{\frac{\sigma_x^2 t^2}{2} + \mu_x t} \; \forall t \; in \mathbb{R}$$

za Y velja podobno. Po zadnjem izreku je

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\frac{\sigma_x^2 t^2}{2} + \mu_x t} \cdot e^{\frac{\sigma_y^2 t^2}{2} + \mu_y t} =$$

$$= e^{\frac{(\sigma_x^2 + \sigma_y^2) t^2}{2} + (\mu_x + \mu_y) t}$$

Po izreku je

$$X + Y \sim N(\mu_x + \mu_y, \sqrt{\sigma_x^2 + \sigma_y^2})$$

■

Opomba. Če bi vedeli, da je $X + Y$ porazdeljena normalno, bi "samo" izračunali parametra

Primer.

$$X \sim N(0,1), M_X(t) = e^{\frac{t^2}{2}} = \sum_{k=0}^{\infty} \frac{(\frac{t^2}{2})^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{2^k \cdot k!} t^{2k}$$

Po drugi strani je $M_X(t) = \sum_{j=0}^{\infty} \frac{z_j}{j!} t^j \ \forall t \in \mathbb{R}$
Primerjamo koeficiente:

- lihi koeficienti: $z_{2k-1} = 0 \ k \in \mathbb{N}$
- sodi koeficienti:

$$\frac{z_{2k}}{(2k)!} = \frac{1}{k!2^k} \implies z_{2k} = \frac{(2k)!}{k!2^k} =$$

$$= \frac{1 \cdot 2 \cdot 3 \cdot \dots \cdot (2k)}{2 \cdot 4 \cdot 5 \cdot \dots \cdot (2k)} = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2k-1) = (2k-1)!! \ k \in \mathbb{N}$$

0.4 Šibki in krepki zakon velikih števil

Definicija 0.20 (Verjetnostna konvergenca)**.** Zaporedje slučajnih spremenljivk $\{X_n\}_{n\in\mathbb{N}}$ verjetnostno konvergira proti skučajni spremenljivki X, če za $\forall \epsilon > 0$ velja $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$
oz. $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$

Definicija 0.21 (Skoraj gotova konvergenca)**.** Zaporedje slučajnih spremenljivk $\{X_n\}_{n\in\mathbb{N}}$ skoraj gotovo konvergira proti skučajni spremenljivki X, če velja P(p $\lim_{n\to\infty} X_n = X) = 1$
Tukaj je $(\lim_{n\to\infty} X_n = X) = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} =$

$$= \{\omega \in \Omega : \forall k (\in \mathbb{N}) \exists m \in \mathbb{N} \forall n \geq m : |X_n(\omega) - X(\omega)| < \frac{1}{k}\} =$$

$$= \{\cap_{k\in\mathbb{N}} \cup_{m\in\mathbb{N}} \cap_{n\geq m} \omega \in \Omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\} \tag{3}$$

Opomba. Števne unije in preseki $\implies$ smo v $\sigma-$algebri, torej je to res dogodek

Trditev 0.22. Če $X_n \xrightarrow{n\to\infty} X$ skoraj gotovo, potem za $\forall \epsilon > 0 \lim_{n\to\infty} P(|X_n - X| < \epsilon$ za $n \geq m) = 1$

Dokaz. Označimo $c_m := (|X_n - X| < \epsilon$ za $n \geq m) = \cap_{n=m}^{\infty} (|x_n - X| < \epsilon)$.
Potem je $c_1 \subseteq c_2 \subseteq \cdots$

je c_m za $\epsilon = \frac{1}{k}$ in $(\lim_{n\to\infty} X_n = X) \subseteq \cup_{n=1}^{\infty} c_m$ (presek)
Torej je $1 = P(\lim_{n\to\infty} X_n = X) \subseteq= (\cup_{m=1}^{\infty} c_m) = \lim_{m\to\infty} P(c_m)$
Od tod sledi $\lim_{m\to\infty} P(c_m) = 1$ ∎

Posledica 0.23. Če $X_n \xrightarrow{n\to\infty} X$ skoraj gotovo, potem $X_n \xrightarrow{n\to\infty} X$ verjetnostno konvergira.

Dokaz. Izberemo $\epsilon > 0$. Potem velja

$$P(|X_n - X| < \epsilon \text{ za } \forall n \geq m) \leq P(|X_m - X| < \epsilon)$$

Če uporabimo trditev, dobimo $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$ (leva stran). ∎

Opomba. Obratna implikacija ne velja

Definicija 0.24. Naj bo $X_1, X_2, X_3 \cdots$ zaporedje slučajnih spremenljivk, ki imajo matematično upanje. Definirajmo $Y_n = \frac{S_n - E(S_n)}{n} = \frac{X_1 + \cdots + X_n}{n} - \frac{E(X_1) + \cdots + E(X_n)}{n}$
Potem je $E(Y_n) = 0$
Za $\{Y_n\}_{n\in\mathbb{N}}$ velja šibki zakon velikih števil (ŠZVŠ), kadar $Y_n \overset{n\to\infty}{\to} 0$ verjetnostno, torej za $\forall \epsilon > 0 \lim_{n\to\infty}(|y| < \epsilon) = 1 = \lim_{n\to\infty}(|\frac{S_n - E(S_n)}{n}| < \epsilon)$ Za $\{Y_n\}_{n\in\mathbb{N}}$ velja krepki zakon velikih števil (KZVŠ), kadar $Y_n \overset{n\to\infty}{\to} 0$ skoraj gotovo, torej $P(\lim_{n\to\infty} \frac{S_n - E(S_n)}{n} = 0) = 1$
Če velja KVZŠ, potem velja ŠVZŠ

Primer. Mečemo kocko, X_k je # pik v k-tem metu. Potem je $E(X_k) = \frac{7}{2}$ in $Y_n = \frac{X_1 + \cdots + X_n}{n} - \frac{7}{2}$
Ali konvergira $\frac{X_1 + \cdots + X_n}{n} \overset{n\to\infty}{\to} \frac{7}{2}$ skoraj gotovo? (Da)

Izrek 0.25.

a Neenakost Markova: če slučajna spremenljivka X ima matematično upanje, potem je $P(|X| \geq a) \leq \frac{E(|X|)}{a}$ za $\forall a > 0$

b Neenakost Čebiševa: če slučajna spremenljivka X ima disperzijo, potem je $P(|X - E(X)| \geq a \cdot \sigma(x)) \leq \frac{1}{a^2}$ za $\forall a > 0$ (pomembno za $a \geq 1$, ker je verjetnost ≤ 1)
oz. če pišemo $\epsilon = a \cdot \sigma(x) \implies P(|X - E(X)| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$ za $\forall \epsilon > 0$

Dokaz. (samo zvezni primer)

a

$$E(X) = \int_{-\infty}^{\infty} |x| p_x(x) dx \geq \int_{\{x:|x|\geq a\}} |x| p_x(x) dx \geq$$
$$|a| \int_{\{x:|x|\geq a\}} p_x(x) dx = a \cdot P(|X| \geq a)$$

b

$$P((X - E(X)) \geq \epsilon) = P((X - E(X))^2 \geq \epsilon^2) \overset{\text{(a) za X-E(X)}}{\leq} \frac{E((X - E(X))^2)}{\epsilon^2} = \frac{D(X)}{\epsilon^2}$$

■

Izrek 0.26 (Markov)**.** Če za zaporedje slučajnih spremenljivk $\{X_n\}_{n\in\mathbb{N}}$ velja $\frac{D(S_n)}{n^2} \overset{n\to\infty}{\to} 0$, potem velja ŠZVŠ. Tukaj je $S_n := X_1 + \cdots + X_n$

Dokaz. V neenakosti Čebiševa vzamemo $X = \frac{S_n}{n}$

$$P(\frac{|S_n - E(S_n)|}{n} \geq \epsilon) \leq \frac{P(S_n)}{n^2 \epsilon^2} \overset{n\to\infty}{\to} 0$$

Če vzamemo $Y_n = \frac{|S_n - E(S_n)|}{n}$, je $P(|Y_n| \geq \epsilon) \overset{n\to\infty}{\to} 0$
oz. $P(|Y_n| < \epsilon) \overset{n\to\infty}{\to} 1$
Zato $Y_n \overset{n\to\infty}{\to} 0$ verjetnostno, torej velja ŠZVŠ za zaporedje $\{X_n\}_{n\in\mathbb{N}}$ ■

Posledica 0.27 (Izrek Čebišev)**.** Če so $X_1, X_2 \cdots X_n$ paroma nekorelirane slučajne spremenljivke in $\sup_{n\in\mathbb{N}} D(X_n) < \infty$, potem za $\{X_n\}_{n\in\infty}$ velja ŠVZŠ

Dokaz. Ker je $D(S_n) = D(X_1) + \cdots + D(X_n) \leq n \cdot c$, je $\frac{D(S_n)}{n^2} \leq \frac{n \cdot c}{n^2} = \frac{c}{n} \overset{n\to\infty}{\to} 0$, zato po izreku Markova velja ŠZVŠ ■

Primer. $X_n : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$ neodvisne slučajne spremenljivke, $D(X_n) = pq, E(X_n) = p, E(S_n) = n \cdot p$
Po izreku Čebiševa velja ŠZVŠ: $P(\frac{|S_n - E(S_n)|}{n} \geq \epsilon) \overset{n\to\infty}{\to} 0$

$$\implies P(|\frac{S_n}{n} - p| \geq \epsilon) \overset{n\to\infty}{\to} 0$$

S_n je frekvenca dogodka, $\frac{S_n}{n}$ je relativna frekvenca, $\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \overset{n\to\infty}{\to} p$ verjetnostno
To je Bernoulijev zakon velikih števil iz 1713

Izrek 0.28 (Kolmogorov)**.** Če za neodvisne slučajne spremenljivke $\{X_n\}_{n\in\mathbb{N}}$ velja $\sum_{n=1}^{\infty} \frac{D_n}{n^2} < \infty$, potem velja KZVŠ, t.j. $P(\lim_{n\to\infty} \frac{S-n-E(S_n)}{n} = 0) = 1$. Posebej je pogoj za vrsto izpolnjen, če je $\sup_n D(X_n) < \infty$

Primer. $X_n : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$ neodvisne slučajne spremenljivke, $D(X_n) = pq$
Po izreku Kolmogorova velja KVZŠ, t.j. $\frac{S_n}{n} = \frac{X_1+\cdots+X_n}{n} \overset{n\to\infty}{\to} p$ skoraj gotovo. To posplošuje Bernoullijev zakon

0.5 Centralni limitni izrek

Definicija 0.29. Naj bo $\{X_n\}_{n\in\mathbb{N}}$ zaporedje slučajnih spremenljivk s končnimi disperzijami. Definiramo $S_n := X_1 + \cdots + X_n$ in standardizirajmo: $Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)}$, torej $E(Z_n) = 0, D(Z_n) = 1$
Za $\{X_n\}_{n\in\mathbb{N}}$ velja centralni limitni izrek, če je $F_{Z_n}(x) = P(Z_n \leq x) \overset{n\to\infty}{\to} F_{N(0,1)} \forall x \in \mathbb{R}$, t.j.

$$P(\frac{S_n - E(S_n)}{\sigma(S_n)} \leq x) \overset{n\to\infty}{\to} \frac{1}{2\pi} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dx \text{ za } \forall x \in \mathbb{R}$$

Pracimo, da $\{Z_n\}_{n\in\mathbb{N}}$ po porazdelitvi konvergira proti standardizirani normalni porazdelitvi.

Izrek 0.30 (Centralni limitni izrek (CLI, osnovna verzija))**.** Naj bodo $X_1, X_2 \cdots$ neodvisne in enako porazdeljene slučajne spremenljivke. Potem zanje velja centralni limitni zakon, t.j

$$P(\frac{S_n - E(S_n)}{\sigma(S_n)} \leq x) \overset{n\to\infty}{\to} \int_{-\infty}^{x} e^{\frac{t^2}{2}} dx \text{ za } \forall x \in \mathbb{R}$$

Dokazal je Ljapunov (1900), s tem je posplošil Laplaceov izrek iz leta 1812. V dokazu bomo uporabili

Izrek 0.31 (O zveznosti rodovne funkcije)**.** Naj za zaporedje $\{Z_n\}_{n\in\mathbb{N}}$ slučajnih spremenljivk velja:
$M_{Z_n}(t) \to M_{N(0,1)}(t) = e^{\frac{t^2}{2}}$ za vse $t \in (-\delta, \delta)$ pri nekem $\delta > 0$
Potem $F_{Z_n}(x) \to F_{N(0,1)}(x)$ za $\forall x \in \mathbb{R}$

Dokaz. CLI v primeru, ko X_n imajo momentno rodovno funkcijo $M_X(t) = E(e^{tX_n})$ na neki okolici točke 0
Naj bo $E(X_n) = \mu, D(X_n) = \sigma^2$ in $U_n := X_n - \mu = X_n - E(X_n)$. Torej je $E(U_n) = 0$ in $D(U_n) = \sigma^2$ ter $M_U(t) = 1 + tE(U_n) + \frac{t^2}{2!}E(U_n^2) + o(t^2) =$

$= 1 + \frac{t^2}{2}\sigma^2 + o(t^2)$ $(\lim_{n\to\infty} \frac{o(n)}{n} = 0)$
Ker je $D(S_n) \overset{\text{neodvisne}}{=} D(X_1) + \cdots + D(X_n) = n \cdot \sigma^2$ in $E(S_n) = n \cdot \mu = E(X_1) + \cdots + E(X_n)$, je $Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} =$
$= \frac{1}{\sigma\sqrt{n}} \left(\sum_{n=0}^{n} U_i\right)$
Potem je $M_{Z_n}(t) = E(e^{tZ_n}) = E(e^{\frac{t}{\sigma\sqrt{n}}(U_1+\cdots+U_n)}) = E(e^{\frac{t}{\sigma\sqrt{n}}U_1}) \cdot \cdots \cdot E(e^{\frac{t}{\sigma\sqrt{n}}U_n}) =$
$\overset{\text{enaki}}{=} (M_U(\frac{t}{\sigma\sqrt{n}}))^n = (1 + \frac{t^2}{2n} + o(\frac{1}{n}))^n$
$\overset{n\to\infty \equiv o(\frac{1}{n}\to 0)}{\longrightarrow} e^{\frac{t^2}{2}}$

Lema 0.32. Če $X_n \to X$, potem $(1 + \frac{X_n}{n})^n \overset{n\to\infty}{\to} e^x$

Po prejšnjem izreku: $F_{Z_n}(x) \overset{n\to\infty}{\to} F_{N(0,1)}(x)$ ■