

Verjetnost in statistika - zapiski s predavanj prof. Drnovška

Tomaž Poljanšek

študijsko leto 2022/23

Kazalo

Posledica 0.1. Če so $X_1, X_2 \dots X_n$ neodvisne slučajne spremenljivke, porazdeljene $N(0, 1)$, potem je $Y := X_1^2 + \dots + X_n^2$ porazdeljena po $\chi^2(n)$

Dokaz. Vemo, da je $X_i^2 \sim \chi^2(1)$ in da so $X_1^2 \dots X_n^2$ neodvisne spremenljivke. Potem je po trditvi + indukciji $Y \sim \chi^2(1 + \dots + 1) = \chi^2(n)$ ■

Oglejmo si transformacijo $f : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \rightarrow (u, v)$, ki preslika zvezno porazdeljen slučajni vektor (x, y) v zvezno porazdeljen slučajni vektor (u, v) , torej $U = u(x, y), V = v(x, y)$

Označimo še $A_{u,v} = (-\infty, u] \times (-\infty, v]$

Potem je

$$F_{(U,V)}(u, v) = \iint_{A_{u,v}} p_{(U,V)}(s, t) ds dt$$

Pot drugi strani pa je

$$F_{(U,V)}(u, v) = P((U, V) \in A_{u,v}) = P((X, Y) \in f^{-1}(A_{u,v})) = \iint_{f^{-1}(A_{u,v})} p_{(X,Y)}(x, y) dx dy$$

Privzemimo še, da je f zvezno odvedljiva bijekcija. Potem je $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (u, v) \rightarrow (x, y)$ tudi zvezno odvedljiva. Z zamenjavo spremenljivk $x = X(u, v), y = Y(u, v)$ v zadnjem intergalu dobimo

$$F_{(U,V)}(u, v) = \iint_{A_{u,v}} p_{(X,Y)}(x(s, t), y(s, t)) \cdot |J(s, t)| dx ds$$

kjer je

$$J(u, v) = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} (u, v)$$

Jacobijeva determinanta.

Zaradi ?? imamo torej $p_{(U,V)}(u, v) = p_{(X,Y)}(x(u, v), y(u, v)) |J(u, v)|$

Oglejmo si poseben primer

Primer. $U = X, V = v(x, y)$ oz $X = U, Y = y(u, v)$

Tedaj je $p_{(U,V)}(u, v) = p_{(X,Y)}(u, y(u, v)) \left| \frac{\partial y}{\partial v}(u, v) \right|$

Gostota spremenljivke V je $\int_{-\infty}^{\infty} p_{(X,Y)}(u, y(u, v)) \left| \frac{\partial y}{\partial v}(u, v) \right| dv = p_V(v)$

Pišimo $Z = V$, torej je $Y = y(x, z)$, saj je $U = X$

Potem prepisemo $p_Z(z) = \int_{-\infty}^{\infty} p_{(X,Y)}(u, y(x, z)) \left| \frac{\partial y}{\partial z}(x, z) \right| dx$

Primer.

1. $Z = X + Y \implies Y = Z - X$, torej je $y(x, z) = z - x, \frac{\partial y}{\partial z}(x, z) = 1$

$$p_{X+Y}(z) = \int_{-\infty}^{\infty} p_{(X,Y)}(x, z-x) \cdot 1 dx$$

2. $Z = X \cdot Y \implies Y = \frac{Z}{X}$, torej je $y(x, z) = \frac{z}{x}, \frac{\partial y}{\partial z}(x, z) = \frac{1}{x}$

$$p_{X \cdot Y}(z) = \int_{-\infty}^{\infty} p_{(X,Y)}(x, \frac{z}{x}) \frac{1}{|x|} dx$$

Če sta še X in Y neodvisni slučajni spremenljivki, potem je

$$p_{X \cdot Y}(z) = \int_{-\infty}^{\infty} p_X(x) \cdot p_Y\left(\frac{z}{x}\right) \cdot \frac{1}{|x|} dx$$

0.1 Matematično upanje oz. pričakovana vrednost

V primeru $X : \begin{pmatrix} x_1 & \cdots & x_n \\ p_1 & \cdots & p_n \end{pmatrix}$ je matematično upanje oz. pričakovana vrednost vsota $E(X) := \sum_{k=1}^n x_k \cdot p_k$

V posebnem primeru $p_1 = \cdots = p_n = \frac{1}{n}$ je $E(X) = \frac{1}{n} \sum_{k=1}^n x_k = \frac{x_1 + \cdots + x_n}{n}$ - povprečje števil $x_1 \cdots x_n$

expected value, expectation, mean value

Naj bo X diskretno porazdeljena slučajna spremenljivka z neskončno zalogo vrednosti:

$$X : \begin{pmatrix} x_1 & x_2 & x_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{pmatrix}$$

X ima matematično upanje oz. pričakovano vrednost, če je $\sum_{k=1}^{\infty} |x_k| p_k < \infty$
Tedadaj je matematično upanje definirano kot $E(X) = \sum_{k=1}^{\infty} x_k \cdot p_k$
Naj bo sedaj X zvezno porazdeljena slučajna spremenljivka z gostoto p_X .
Potem ima X matematično upanje, če je $\int_{-\infty}^{\infty} |x| \cdot p_X(x) dx < \infty$. Tedaje
je matematično upanje slučajne spremenljivke X enako $E(X) = \int_{-\infty}^{\infty} x \cdot p_X(x) dx$

Primer.

$$1. X \sim Ber(p) \text{ oz. } X : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix} q = 1 - p, E(X) = 0 \cdot q + 1 \cdot p = p$$

$$2. X \sim Poi(\lambda), \text{ torej } p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} k = 0, 1 \dots$$

$$E(X) = \sum_{k=0}^{\infty} k \cdot p_k = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

$$3. \text{ Enakomerna porazdelitev na } [a, b]$$

$$p(X) = \begin{cases} \frac{1}{b-a} & \text{če } a \leq x \leq b \\ 0 & \text{sicer} \end{cases}$$

$$E(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

$$4. X \sim N(\mu, \sigma)$$

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \frac{-\infty}{\infty} x \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

$$U = \frac{x - \mu}{\sigma} \implies du = \frac{dx}{\sigma}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu) e^{-\frac{1}{2}u^2} du = \frac{1}{\sqrt{2\pi}} \sigma \int_{-\infty}^{\infty} u \cdot e^{-\frac{1}{2}u^2} du + \mu \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \cdot e^{-\frac{1}{2}u^2} du = \mu \quad (1)$$

Ker je v predzadnjem koraku 1. funkcija (v integralu) liha, 2. pa je gostota porazdelitve $N(0, 1)$

$$5. \text{ Cauchyjeva porazdelitev } p(x) = \frac{1}{\pi(1+x^2)}$$

$$\text{Nima matematičnega upanja, saj je } \int_{-\infty}^{\infty} |x| \cdot \frac{1}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx = \frac{1}{\pi} \ln(1+x^2) \Big|_0^{\infty} = \infty$$

6. $1 - \frac{1}{2} + \frac{1}{3} - \dots$ je pogojno konvergentna vrsta, t.j. konvergira, a ne absolutno

$$X : \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}, \sum_{k=1}^{\infty} x_k \cdot p_k = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}$$

$$x_k \cdot p_k = \frac{(-1)^{k-1}}{k}$$

$$\sum_{k=1}^{\infty} p_k = 1$$

$$p_k := \frac{1}{2^k} \text{ npr. ker je vsota } 1$$

$$x_k := \frac{(-1)^{k-1}}{k} = 2^k$$

Ta porazdelitev nima matematičnega upanja, ker vrsta ne konvergira absolutno

Trditev 0.2. Naj bo $f : \mathbb{R} \rightarrow \mathbb{R}$ zvezna funkcija

- (a) Če je $X : \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$ potem je $E(f \circ X) \equiv E(f(X)) = \sum_{k=1}^{\infty} f(x_k) \cdot p_k$ (če le to matematično upanje obstaja)
- (b) Če je X zvezno porazdeljena z gostoto p_X , potem je $E(f \circ X) = \int_{-\infty}^{\infty} f(x) \cdot p_X(x) dx$

Dokaz. (samo (a)):

$$f \circ X : \begin{pmatrix} f(x_1) & f(x_2) & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$$

$$\text{npr če } f(x_1) = f(x_3) \implies \begin{pmatrix} f(x_1) & f(x_2) & \dots \\ p_1 + p_3 & p_2 & \dots \end{pmatrix}$$

$$(E(f \circ X) = \int_{-\infty}^{\infty} x \cdot p_{f(x)}(x) dx = \dots = \int_{-\infty}^{\infty} f(x) \cdot p_X(x) dx - \text{substitucija } y = f(x) \text{ v integralu}) \quad \blacksquare$$

Posledica 0.3. Slučajna spremenljivka X ima matematično upanje $\iff X$ ima matematično upanje. Tedaj velja $|E(X)| = E(|X|)$

Dokaz. (samo diskreten primer):

$$E(|X|) \stackrel{\text{trd. } f(x)=|x|}{=} \sum_i |x_i| \cdot p_i \geq \left| \sum_i x_i \cdot p_i \right| = |E(X)|$$

■

Posledica 0.4. Za $\forall a \in \mathbb{R}$ in vsako slučajno spremenljivko X z matematičnim upanjem velja $E(a \cdot X) = a \cdot E(X)$ (homogenost)

Dokaz. $f(x) = a \cdot x$, trditev (od prej) ■

Podobno kot zadnjo trditev se dokaže

Trditev 0.5. Naj bo $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ zvezna funkcija in (X, Y) slučajni vektor

- (a) Naj bo (X, Y) diskretno porazdeljen $p_{ij} := P(X = x_i, Y = y_j)$. Potem je $E(f(X, Y)) = \sum_i \sum_j f(x_i, y_j) \cdot p_{ij}$ (če le vrsta (oz. končna vsota) absolutno konvergira)
- (b) Naj bo (X, Y) zvezno porazdeljen z gostoto $p(X, Y)$. Potem je $E(f(X, Y)) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x, y) p_{(X,Y)}(x, y) dy$ (če le integral absolutno konvergira)

Posledica 0.6. Če slučajni spremenljivki X in Y imata matematično upanje, potem ga ima tudi $X + Y$ in velja $E(X + Y) = E(X) + E(Y)$ (aditivnost)

Dokaz. (samo zvezen primer):

$$\begin{aligned} E(X, Y) &\stackrel{f(x,y)=x+y}{=} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} (x+y) p_{(X,Y)}(x, y) dy = \\ &= \int_{-\infty}^{\infty} x dx \int_{-\infty}^{\infty} p_{(X,Y)}(x, y) dy + \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} p_{(X,Y)}(x, y) dx = \\ &= \int_{-\infty}^{\infty} x p_X(x) dx + \int_{-\infty}^{\infty} y p_Y(y) dy = E(X) + E(Y) \end{aligned}$$

■

Posledica 0.7. Za slučajne spremenljivke $X_1 \cdots X_n$, ki imajo matematično upanje, velja $E(a_1 X_1 + \cdots + a_n X_n) = a_1 E(X_1) + \cdots + a_n E(X_n)$ z $\forall a_1 \cdots a_n \in \mathbb{R}$

$$E(X + Y) = \int_{-\infty}^{\infty} x \cdot p_{X+Y}(x) dx \stackrel{?}{=} E(X) + E(Y) \text{ ni očitno iz tega}$$

Primer. 1. Če ima X matematično upanje, potem $E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0$

$$2. X_k \sim Ber(p), \text{ t.j. } X_k \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, q = 1 - p$$

$$X = X_1 + \cdots + X_n \implies E(X) = E(X_1) + \cdots + E(X_n) = n \cdot p$$

Posebej to (v 2. zgledu) velja v primeru, ko so $\{X_k\}_{i=1}^n$ neodvisne. To velja tudi za Bernoullijevo zaporedje ponovitev poskusa: opazujemo dogodek A s $P(A) = p$. X je frekvenca dogodka A v n ponovitvah poskusa. Potem je $X \sim \text{Bin}(n, p)$ in $X = X_1 + \dots + X_n$, kjer je $(X_k = 1)$ dogodek, da se A zgodi v k -ti ponovitvi poskusa, sicer je $(X_k = 0)$. Po zgornjem je $E(X) = n \cdot p$. Do tega lahko pridemo tudi direktno:

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \cdot p_k = \sum_{k=0}^n k \cdot \binom{n}{k} p^k q^{n-k} = \\ &= \sum_{k=1}^n k \cdot \frac{n}{k} \binom{n-1}{k-1} p^k q^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{n-k} \stackrel{j=k-1}{=} \\ &= np \left(\sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \right) = np(p+q)^{n-1} = np \end{aligned}$$

Trditev 0.8 (Cauchy-Schwartzova neenakost). Če obstajata $E(X^2)$ in $E(Y^2)$, potem obstaja tudi $E(X \cdot Y)$ in velja $E(|X \cdot Y|) \leq \sqrt{E(X^2) \cdot E(Y^2)}$. Enačaj velja samo v primeru $|Y| = \sqrt{\frac{E(Y^2)}{E(X^2)}} |X|$ z verjetnostjo 1

Dokaz. Ker za nenegativna realna števila velja neenakost

$$u \cdot v \leq \frac{1}{2}(u^2 + v^2) \iff (u - v)^2 \geq 0$$

za nenegativni slučajni spremenljivki U in V velja neenakost

$$U \cdot V \leq \frac{1}{2}(U^2 + V^2)$$

Enakost velja samo v točkah $\omega \in \Omega$, za katere je $U(\omega) = V(\omega)$

Če vstavimo $U = a \cdot |X|$ in $V = \frac{1}{a}|Y|$ za $a > 0$, dobimo $|X \cdot Y| \leq \frac{1}{2}(a^2 Y^2 + \frac{1}{a^2} X^2)$ in zato je

$$E(|X \cdot Y|) \leq \frac{1}{2}(a^2 E(X^2) + \frac{1}{a^2} E(Y^2)) \text{ za } \forall a > 0 \quad (2)$$

Če vstavimo $a^2 = \sqrt{\frac{E(Y^2)}{E(X^2)}}$ na desni strani dobimo

$$\frac{1}{2}(\sqrt{E(Y^2) + E(X^2)} + \sqrt{E(X^2 + E(Y^2))}) = \sqrt{E(X^2) + E(Y^2)}$$

Torej je

$$E(|X \cdot Y|) \leq \sqrt{E(X^2) \cdot E(Y^2)}$$

Enakost v neenakosti velja $\iff a|X| = \frac{1}{a}|Y|$, torej $|Y| = a^2|X| = \frac{E(Y^2)}{E(X^2)}|X|$ z verjetnostjo 1 ■

Posledica 0.9. Če obstaja $E(X^2)$, potem obstaja $E(X)$ in velja $(E(X))^2 \leq E(X^2)$

Dokaz. $Y = 1$, t.j. $Y : \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow$

$$E(|X \cdot 1|) \leq \sqrt{E(X^2) \cdot 1^2} \quad (E(|X|))^2 \leq E(X^2)$$

■

Trditev 0.10. Naj bosta X in Y neodvisni slučajni spremenljivki, ki imata matematični upanji. Potem ima matematično upanje tudi $X \cdot Y$ in velja $E(X \cdot Y) = E(X) \cdot E(Y)$

Dokaz. (samo zvezem primer):

$$E(X \cdot Y) \stackrel{\text{trd}}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot p_{(X,Y)}(x,y) dx dy \stackrel{\text{neodvisnost}}{=} \iint_{\mathbb{R}} x \cdot y \cdot p_X(x) \cdot p_Y(y) dx dy = \int_{-\infty}^{\infty} x p_X(x) dx \cdot \int_{-\infty}^{\infty} y p_Y(y) dy = E(X) \cdot E(Y)$$

■

Definicija 0.11 (Nekoreliranost). Slučajni spremenljivki X in Y sta nekorelirani, če velja $E(X \cdot Y) = E(X) \cdot E(Y)$, sicer sta korelirani.

Po trditvi iz neodvisnosti sledi nekoreliranost. Obratno pa ne velja:

Primer.

$$U = \begin{pmatrix} 0 & \frac{\pi}{2} & \pi \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

$$X = \cos(U) : \begin{pmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

$$Y = \sin(U) : \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

$$E(X) = 0, E(Y) = \frac{1}{3}$$

$$X \cdot Y = \sin(U) \cdot \cos(U) = 0 \Rightarrow E(X \cdot Y) = 0 \Rightarrow X \text{ in } Y \text{ sta nekorelirani slučajni spremenljivki}$$

X \ Y	0	1	Σ
-1	$\frac{1}{3}$	0	$\frac{1}{3}$
0	0	$\frac{1}{3}$	$\frac{1}{3}$
-1	$\frac{1}{3}$	0	$\frac{1}{3}$
Σ	$\frac{2}{3}$	$\frac{1}{3}$	1

$$\Rightarrow \text{nista neodvisni, npr}$$

$$\frac{1}{3} = P(X = 1, Y = 0) \neq P(X = 1) \cdot P(Y = 0) = \frac{1}{3} \cdot \frac{2}{3}$$

Trditev 0.12. $X : \begin{pmatrix} x_1 & x_2 \\ p_1 & p_2 \end{pmatrix}, Y : \begin{pmatrix} y_1 & y_2 \\ q_1 & q_2 \end{pmatrix}$

Potem sta X in Y neodvisni \iff nekorelirani
 $\iff E(X \cdot Y) = E(X) \cdot E(Y)$

0.2 Disperzija, kovarianco in korelacijski koeficient

Definicija 0.13 (Disperzija). Naj obstaja $E(X^2)$. Disperzija oz. varianca slučajne spremenljivke X je $D(X) \equiv \text{var}(X) := E((X - E(X))^2)$

Disperzija meri razpršenost slučajne spremenljivke X okoli $E(X)$

Ker je $E((X - E(X))^2) = E(X^2 - 2E(X)X + (E(X))^2) = E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2$, je $D(X) = E(X^2) - (E(X))^2$

Lastnosti disperzije:

- $D(X) \geq 0$ in $D(X) = 0 \iff P(X = E(X)) = 1$, t.j. X je izrojena slučajna spremenljivka
- $D(a \cdot X) = a^2 D(X)$ $a \in \mathbb{R}$
- $\forall a \in \mathbb{R}$ velja: $E((X - a)^2) \geq D(X)$. Enakost velja le v primeru $a = E(X)$

Dokaz.

$$E((X - a)^2) = E(X^2 - 2aX + a^2) = E(X^2) - 2E(X)a + a^2 = (a - E(X))^2 + E(X^2) - (E(X))^2$$

Enakost velja samo za $a = E(X)$ ■

Definicija 0.14 (Standardna deviacija). Standardna deviacija ali standardni odklon slučajne spremenljivke X je $\sigma(X) := \sqrt{D(X)}$

Zanjo velja $\sigma(aX) = |a| \cdot \sigma(X)$ za $\forall a \in \mathbb{R}$

Primeri nekaterih $E(X)$ in $D(X)$

1. enakomerna diskretna porazdelitev: $\begin{pmatrix} x_1 & \cdots & x_n \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$

$$E(X) = \frac{x_1 + \cdots + x_n}{n}, D(X) = E(X^2) - (E(X))^2 = \frac{x_1^2 + \cdots + x_n^2}{n} - \left(\frac{x_1 + \cdots + x_n}{n}\right)^2$$

2. Binomska porazdelitev $\text{Bin}(n, p)$, $n \in \mathbb{N}$, $p \in (0, 1)$, $q = 1 - p$

$$E(X) = n \cdot p, D(X) = npq, \sigma(X) = \sqrt{npq}$$

3. Poissonova porazdelitev $Poi(\lambda), \lambda > 0$

$$E(X) = \lambda, D(X) = \lambda$$

4. Geometrijska porazdelitev $geo(p), p \in (0, 1), q = 1 - p$

$$E(X) = \frac{1}{p}, D(X) = \frac{q}{p^2}$$

5. Pascalova porazdelitev $Pas(m, p), m \in \mathbb{N}, p \in (0, 1)$

$$E(X) = \frac{m}{p}, D(X) = \frac{mq}{p^2}$$

6. Enakomerna zvezna porazdelitev Ed na $[a, b]$

$$E(X) = \frac{a+b}{2}, D(X) = \frac{(b-a)^2}{12}$$

7. Normalna porazdelitev $N(\mu, \sigma)$

$$E(X) = \mu, D(X) = \sigma^2, \sigma(X) = \sigma$$

8. Porazdelitev gama $\gamma(b, c)$

$$E(X) = \frac{b}{c}, D(X) = \frac{b}{c^2}$$

9. Porazdelitev $\chi^2(n) = \gamma(\frac{n}{2}, \frac{1}{2})$

$$E(X) = n, D(X) = 2n$$

10. Eksponentna porazdelitev $Exp(\lambda), \lambda > 0 = \gamma(1, \lambda)$

$$E(X) = \frac{1}{\lambda}, D(X) = \frac{1}{\lambda^2}, \sigma(X) = \frac{1}{\lambda}$$

Preverimo, da je $D(X) = \sigma^2$ za $X \sim N(\mu, \sigma)$

$$D(X) = E((X - E(X))^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

$$t = \frac{x - \mu}{\sigma} \implies x - \mu = \sigma t, dx = \sigma dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt =$$

$$u = t, dv = t \cdot e^{-\frac{1}{2}t^2}$$

$$du = dt, v = -e^{-\frac{1}{2}t^2}$$

$$\frac{\sigma^2}{\sqrt{2\pi}} (-te^{-\frac{1}{2}t^2} |_{-\infty}^{\infty}) + \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \frac{\sigma^2}{\sqrt{2\pi}} (0 + \sqrt{2\pi}) = \sigma^2$$

Definicija 0.15 (Kovarianca). Kovarianca slučajnih spremenljivk $K(X, Y) \equiv \text{Cov}(X, Y) := E((X - E(X))(Y - E(Y)))$

Ker je

$$E((X - E(X))(Y - E(Y))) = E(XY - E(Y)X - E(X)Y + E(X)E(Y)) = E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

je $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

Lastnosti:

1. $K(X, X) = D(X)$
2. $K(X, Y) = 0 \iff X$ in Y sta neodvisni
3. K je simetrična in bilinearna funkcija:
 - $K(X, Y) = K(Y, X)$
 - $K(aX + bY, Z) = aK(X, Z) + bK(Y, Z) \forall a, b \in \mathbb{R}$
4. Če obstajata $D(X)$ in $D(Y)$, potem obstaja tudi $K(X, Y)$. Tedaj velja $|K(X, Y)| \leq \sqrt{D(X) \cdot D(Y)} = \sigma(X) \cdot \sigma(Y)$
 To sledi iz Cauchy-Schwartzove neenakosti ($|E(U \cdot V)| \leq \sqrt{E(U^2) \cdot E(V^2)}$) za slučajni spremenljivki $X - E(X)$ in $Y - E(Y)$. Enačaja v neenakosti velja $\iff Y - E(Y) \pm \frac{\sigma(Y)}{\sigma(X)}(X - E(X))$ z verjetnostjo 1
5. Če X in Y imata disperziji, potem jo ima tudi $X + Y$ in velja $D(X + Y) = D(X) + D(Y) + 2K(X, Y)$
 če sta X in Y nekorelirani (posebej neodvisni), potem je $D(X + Y) = D(X) + D(Y)$

Dokaz. Sledi iz enakosti

$$(X + Y - E(X + Y))^2 = ((X - E(X)) + (Y - E(Y)))^2 = (X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))$$

$$D(X + Y) = E((X - E(X))^2) + E((Y - E(Y))^2) + E(2(X - E(X))(Y - E(Y))) = D(X) + D(Y)$$

■

6. Posplošitev: $D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n) + 2 \sum_{i < j} K(X_i, X_j)$
 Če so $X_1 \dots X_n$ paroma nekorelirani (posebej neodvisni), potem je
 $D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n)$

Primer. $\text{Bin}(n, p)$ je vsota $X = X_1 + \dots + X_n$, kjer je $X_i \sim \text{Ber}(p)$, t.j.

$X_i \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, ki so neodvisne

Zato je $D(X) = D(X_1 + \dots + X_n) = n \cdot D(X_1) = n \cdot p \cdot q$, saj je $D(X_n) = E(X_n^2) - (E(X_n))^2 = p - p^2 = pq$

Definicija 0.16 (Standardizacija slučajne spremenljivke). Standardizacija slučajne spremenljivke X je slučajna spremenljivka $X_s = \frac{X - E(X)}{\sigma(X)}$

Zanjo velja:

- $E(X_s) = 0$
- $D(X_s) = \frac{1}{\sigma(X)^2} \cdot D(X - E(X)) = \frac{1}{\sigma(X)^2} D(X) = 1$

Primer.

$$X \sim N(\mu, \sigma) \implies X_s = \frac{X - E(X)}{\sigma(X)} = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Definicija 0.17 (Korelacijski koeficient). Korelacijski koeficient slučajnih spremenljivk X in Y je

$$r(X, Y) = \frac{K(X, Y)}{\sigma(X)\sigma(Y)} = \frac{E((X - E(X))(Y - E(Y)))}{\sigma(X)\sigma(Y)} = E(X_s \cdot Y_s)$$

Lastnosti:

1. $r(X, Y) = 0 \iff X$ in Y sta nekorelirani
2. $r(X, Y) \in [-1, 1]$, kar sledi iz lastnosti (4) za kovarianco
3. • $r(X, Y) = 1 \iff Y = \frac{\sigma(Y)}{\sigma(X)}(X - E(X)) + E(Y)$ z verjetnostjo 1

- $r(X, Y) = -1 \iff Y = -\frac{\sigma(Y)}{\sigma(X)}(X - E(X)) + E(Y)$ z verjetnostjo 1

Tedaj imamo linearno zvezo med X in Y

Primer.

$$(X, Y) \sim N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho) \quad \mu_x, \mu_y \in \mathbb{R}, \sigma_x, \sigma_y \in [0, \infty], \rho \in [-1, 1]$$

Trdimo, da je $r(X, Y) = \rho$

$$(X_s, Y_s) \sim N(0, 0, 1, 1, \rho)$$

$$r(X, Y) = E(X_s \cdot Y_s) = \frac{1}{2\pi\sqrt{1-\rho^2}} \iint_{\mathbb{R}} xy e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy' + y'^2)} dx dy$$

$$\begin{aligned} x^2 - 2\rho xy + y^2 &= (x - \rho y)^2 + (1 - \rho^2)y^2 \\ &= \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2(1-\rho^2)}(x-\rho y)^2} dx = \\ &= E(N(\rho y, \sqrt{1-\rho^2})), \text{ ker je } p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \\ &= \rho \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy = \\ &= \left(\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy = D(N(0, 1)) = 1\right) \implies = \rho \end{aligned}$$

Torej sta X in Y nekorelirani $\overset{\text{v splošnem}}{\iff} \rho = 0 \overset{\text{ta primer}}{\iff} X, Y$ neodvisni

Kakšna je gostota, če je ρ blizu 1? $\rho \uparrow 1 : \rho \downarrow -1 :$

gostota je skoraj skoncentrirana na neki premici, torej med X in Y obstaja skoraj linearna zveza

0.3 Pogojna porazdelitev in pogojno matematično upanje

Izberimo si dogodek B s $P(B) > 0$

Definicija 0.18. Pogojna porazdelitvena funkcija slučajne spremenljivke X glede na B je $F_X(X | B) := P(X \leq x | B) = \frac{P(X \leq x \wedge B)}{P(B)}$

Ima enake lastnosti kot porazdelitvena funkcija

A Diskreten primer

Naj bo (X, Y) diskretno porazdeljen slučajni vektor z verjetnostno funkcijo $p_{ij} = P(X = x_i, Y = y_j), i, j = 1, 2, \dots$

Za pogoj B vzemimo $B = (Y = y_j)$ pri nekem j , torej $q_j = P(Y = y_j)$. Potem je pogojna porazdelitvena funkcija slučajne spremenljivke X glede na $Y = y_j$ $F_X(X | Y = y_j) := \frac{P(X \leq x | Y = y_j)}{P(Y = y_j)} = \frac{1}{q_j} \sum_{j: x_j \leq x} p_{ij}$

Če vpeljemo pogojno verjetnostno funkcije $p_{i|j} = P(X = x_i | Y = y_j) = \frac{p_{ij}}{q_j}$, $F_X(X | Y = y_j) = \sum_{i: x_i \leq x} p_{i|j}$

Pogojno matematično upanje slučajne spremenljivke X glede na $Y = y_j$ je matematično upanje te porazdelitve:

$$E(X | Y = y_j) := \sum_i x_i \cdot p_{i|j} = \frac{1}{q_j} \sum_i x_i \cdot p_{ij}$$

Regresijska funkcija $\ell(y_j) = E(X | Y = y_j)$, ki je definirana na zalogi vrednoti slučajne spremenljivke Y

Definirajmo novo slučajno spremenljivko $E(X | Y) = \ell(y)$, ki ji rečemo pogojno matematično upanje slučajne spremenljivke X glede na slučajno spremenljivko Y

$$\text{Ta ima shemo } E(X | Y) = \begin{pmatrix} \ell(y_1) & \ell(y_2) & \dots \\ q_1 & q_2 & \dots \end{pmatrix} = \begin{pmatrix} E(X | Y = y_1) & \dots \\ q_1 & \dots \end{pmatrix}$$

Zanjo velja

$$E(X | Y) = \sum_j \ell(y_j) \cdot q_j = \sum_j \sum_i x_i \cdot p_{ij} = \sum_i x_i (\sum_j p_{ij}) = \sum_i x_i \cdot p_i = E(X)$$

kjer je $p_i = P(X = x_i)$

Kaj dobimo, če sta X in Y neodvisni slučajni spremenljivki?

Tedaj je $p_{i|j} = \frac{p_{ij}}{q_j} = \frac{p_i \cdot q_j}{q_j} = p_i$ in $\ell(y_j) = E(E(X | Y = y_j)) = \sum_i x_i \cdot p_{i|j} = \sum_i x_i \cdot p_i = E(X)$, torej je regresijska funkcija kar konstanta $E(X)$ oz. je $E(X | Y)$ izrojena slučajna spremenljivka z vrednostjo $E(X)$

Primer. Kokoš znese N jajc, kjer je $N \sim \text{Poi}(\lambda)$ z $\lambda > 0$. Iz vsakega jajca se z verjetnostjo $p \in (0, 1)$ izvali piščanec, neodvisno od drugih jajc. Naj bo K število piščancev. Določimo $E(K | N)$, $E(K)$ in $E(N | K)$

$$P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda} \quad n = 0, 1, 2, \dots$$

$$P(K = k | N = n) = \binom{n}{k} p^k q^{n-k} \quad k = 0, 1, \dots, n$$

$$\ell(n) = E(K \mid N = n) = E(\text{Bin}(n, p)) = n \cdot p$$

torej je $E(K \mid N) = \ell(n) = p \cdot N$

$$E(K \mid N) = \begin{pmatrix} p \cdot 0 & p \cdot 1 & p \cdot 2 & \dots \\ P(N=0) & P(N=1) & P(N=2) & \dots \end{pmatrix}$$

$$E(K) = E(E(K \mid N)) = E(p \cdot N) = p \cdot E(N) = p \cdot \lambda$$

$$\begin{aligned} P(K = k) &= \sum_{n=k}^{\infty} P(K = k \mid N = n) \cdot P(N = n) = \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k q^{n-k} \cdot \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \frac{1}{k!} e^{-\lambda} p^k \lambda^k \sum_{n=k}^{\infty} \frac{(qk)^{n-k}}{(n-k)!} = \frac{(p\lambda)^k}{k!} e^{-\lambda} e^{q\lambda} = \frac{(p\lambda)^k}{k!} e^{-p\lambda} \quad k = 0, 1, \dots, n \end{aligned}$$

Torej je $K \sim \text{Poi}(p \cdot \lambda)$

$$\begin{aligned} P(N = n \mid K = k) &= \frac{P(N = n, K = k)}{P(K = k)} = \frac{P(K = k \mid N = n) \cdot P(N = n)}{P(K = k)} = \\ &= \frac{n! p^k q^{n-k}}{k!(n-k)!} \cdot \frac{\lambda^n e^{-\lambda}}{n!} \cdot \frac{pk! e^{p\lambda}}{(p\lambda)^k} = \frac{(q\lambda)^{n-k}}{(n-k)!} \cdot e^{-q\lambda} \quad n = k, k+1, \dots \end{aligned}$$

To je za k premaknjena Poissonova porazdelitev: $k + \text{Poi}(q\lambda)$

Potem je $\psi(k) = E(N \mid K = k) = E(k + \text{Poi}(q\lambda)) = k + q \cdot \lambda$ in zato

je $E(N \mid K) = \psi(k) = k \cdot q + \lambda$

Preizkus: $E(E(N \mid K)) = E(k + q \cdot \lambda) = p\lambda + q\lambda = \lambda = E(N)$ (ok)

Regresijsko premico je vpeljal Golten (1822-1911)

B Zvezni primer

Naj bo (X, Y) zvezno porazdeljen slučajni vektor z gostoto $p_{(X,Y)}(x, y)$.

Vzemimo $B = (y < Y \leq y + k)$ za nek $y \in \mathbb{R}, k > 0$.

Potem je $F_X(X \mid y < Y \leq y + k) = P(x \leq x \mid y < Y \leq y + k) = \frac{P(X \leq x, y < Y \leq y + k)}{P(y < Y \leq y + k)} = \frac{F_{(X,Y)}(x, y+k) - F_{(X,Y)}(x, y)}{F_Y(y+k) - F_Y(y)}$

Pogojna porazdelitvena funkcija slučajne spremenljivke X glede na dogodek $(Y = y)$ je limita, če obstaja:

$$F_X(x \mid Y = y) = \lim_{h \downarrow 0} F_X(x \mid y < Y \leq y+h) = \lim_{h \downarrow 0} \frac{F_{(X,Y)}(x, y+h) - F_{(X,Y)}(x, y)}{F_Y(y+h) - F_Y(y)}$$

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 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

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

Vpeljimo regresijsko funkcijo $l(y) := E(X | Y = y)$, definirano na zalogi vrednosti slučajne spremenljivke Y . Tako dobimo novo slučajno spremenljivko $E(X | Y) := l(y)$: pogojno matematično upanje slučajne spremenljivke X glede na slučajno spremenljivko Y .

Kot v diskretnem primeru se pokaže enakost $E(E(X | Y)) = E(X)$

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

$$p_X(x | y) = \frac{p_{(X,Y)}(x, y)}{p_Y(y)} \stackrel{\text{D.N.}}{=} \frac{1}{\sigma_x \sqrt{(2\pi)(1 - \rho^2)}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left(\frac{x - \mu_x}{\sigma_x} - \rho \frac{y - \mu_y}{\sigma_y}\right)^2\right)$$

torej je $N(\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x \sqrt{1 - \rho^2})$

Eksponent: $\frac{1}{2(1 - \rho^2)} \sigma_x^2 (x - (\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y)))^2$

$\implies l(y) = E(X | Y = y) = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y)$ - 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 $E(x | y) = \alpha + \beta y$

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 škodljiva

a

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

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

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

0.4 Višji momenti in vrstilne karakteristike

Definicija 0.19 (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 običajno vzamemo

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

Očitno je $z_1 = E(X)$, $m_2 = D(X)$

Trditev 0.20. Če $\exists m_n(a)$, potem obstajajo tudi momenti $m_k(a)$ za vse $k < n$

Dokaz. (V zveznem primeru):

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

■

Trditev 0.21. Če obstaja začetni moment z_n , potem obstajajo $m_n(a)$ 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:

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

Asimetrija slučajne spremenljivke X je $A(X) := E(X_s^3) = E\left(\left(\frac{X-E(X)}{\sigma_x}\right)^3\right) = \frac{m_3}{m_2^{3/2}} m_2 = \sigma^2 = D(X)$

$A(N(\mu, \sigma)) = 0$, 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 0.22 (Mediana). Mediana slučajne spremenljivke X je vsaka vrednost $x \in \mathbb{R}$, za katero velja $P(X \leq x) \leq \frac{1}{2}$ in $P(X \geq x) \leq \frac{1}{2}$ (in $P(X < x) = 1 - F(x-)$)

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 0.23 (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: $X_{\frac{1}{100}}, X_{\frac{2}{100}}, \dots, X_{\frac{99}{100}}$

Primer. Telesna višina odraslih moških

Definicija 0.24 ((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

$$\int_{-\infty}^{x_{\frac{1}{4}}} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{4}$$

$$\frac{1}{\pi} \arctan x \Big|_{x=-\infty}^{x_{\frac{1}{4}}} = \frac{1}{4}$$

$$\frac{1}{\pi} \arctan x_{\frac{1}{4}} + \frac{1}{2} = \frac{1}{4}$$

$$\arctan x_{\frac{1}{4}} = \frac{1}{4} \implies x_{\frac{1}{4}} = -1$$

$$\xrightarrow{\text{simetrija}} x_{\frac{3}{4}} = 1, s = 1$$

0.5 Rodovne funkcije

Definicija 0.25. Naj bo X slučajna spremenljivka z vrednostmi v $\mathbb{N} \cup \{0\}$:

$$p_k = P(X = k) \quad k = 0, 1, 2, \dots \quad p_k \geq 0, \sum_{k=0}^{\infty} p_k = 1$$

Rodovna funkcija slučajne spremenljivke X je

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

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 $s^X : \begin{pmatrix} s^0 & s^1 & s^2 & \dots \\ p_0 & p_1 & p_2 & \dots \end{pmatrix}$, 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.

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

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

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

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

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = \\ &= e^{-\lambda} \cdot e^{\lambda s} = e^{\lambda(s-1)} \end{aligned}$$

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

Iz teorije Taylorjevih vrst sledi

Izrek 0.26 (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 $k = 0, 1, 2, \dots$

Tedaj velja $P(X = k) = \frac{1}{k!} G_X^k(0)$

$$G_X(s) = \sum_{k=0}^{\infty} p_k s^k, \quad p_k = P(X = k)$$

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 0.27. 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

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

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

■

Posledica 0.28.

$$E(X) = G_X'(1-)$$

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

Izrek 0.29. Naj bosta X in Y neodvisni slučajni spremenljivki z rodovnicama 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. $G_{X+Y}(s) = E(s^{X+Y}) = E(s^X \cdot s^Y) \stackrel{\text{izrek}}{=} E(s^X) \cdot E(s^Y) = G_X(s) \cdot G_Y(s)$, saj sta s^X in s^Y neodvisni slučajni spremenljivki ■

Posplošitev 0.30. Če so X_1, X_2, \dots, X_n neodvisne slučajne spremenljivke, potem je za vse $s \in [-1, 1]$ $G_{X_1+\dots+X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s)$. Če so X_1, X_2, \dots, X_n enako porazdeljene in neodvisne, potem je

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

Izrek 0.31. Naj bodo za $\forall n \in \mathbb{N}$ slučajne spremenljivke N, X_1, X_2, \dots, X_n neodvisne. Naj ima N rodovno funkcijo G_N , X_n pa rodovno funkcijo G_X . Potem ima slučajna spremenljivka $S := X_1 + X_2 + \dots + X_n$ 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 $P(S = k) = \sum_{n=0}^{\infty} P(S = k, N = n) =$

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

Zato je

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

$$\begin{aligned}
&= \sum_{n=1}^{\infty} P(N=n) \left(\sum_{k=0}^{\infty} P(X_1 + \dots + X_n = k) \cdot s^k \right) = \\
&= \sum_{n=1}^{\infty} P(N=n) \cdot (G_X(s))^n = G_N(G_X(s))
\end{aligned}$$

za vse $s \in [-1, 1]$ ■

Posledica 0.32. Pri predpostavkah iz izreka velja Waldova enakost:

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

Dokaz.

$$G_S(s) = G_N(G_X(s)) \forall s \in [-1, 1] \quad (3)$$

$$E(S) = G'_S(1-) = G'_N(G_X(1-)) \cdot G'_X(1-) = E(N) \cdot E(X) \quad (4)$$
■

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 $K \sim Poi(\lambda p)$

0.6 Momentno rodovna funkcija

Definicija 0.33 (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 $X : \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$ je $M_X(t) = \sum_i e^{tx} p_i$

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

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

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

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

Primer.

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

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

Izrek 0.34. 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)

$$\begin{aligned} M_X(t) &= E(e^{t \cdot X}) = E\left(\sum_{k=0}^{\infty} t^k \frac{X^k}{k!}\right) = \\ &= \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k = \sum_{k=0}^{\infty} \frac{z_k}{k!} t^k \end{aligned}$$

■

Trditev 0.35. $M_{aX+b}(t) = e^{bt} M_X(at)$, $a \neq 0, b \in \mathbb{R}$

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

■

Izrek 0.36. Če sta X in Y neodvisni slučajni spremenljivki, potem je $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

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

■

Trditev 0.37. 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

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

za Y velja podobno. Po zadnjem izreku je

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

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 $M_X(t) = \sum_{j=0}^{\infty} \frac{z_j}{j!} t^j \quad \forall t \in \mathbb{R}$

Primerjamo koeficiente:

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

$$\begin{aligned} \frac{z_{2k}}{(2k)!} &= \frac{1}{k! 2^k} \implies z_{2k} = \frac{(2k)!}{k! 2^k} = \\ &= \frac{1 \cdot 2 \cdot 3 \cdot \dots \cdot (2k)}{2 \cdot 4 \cdot 5 \cdot \dots \cdot (2k)} = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2k-1) = (2k-1)!! \quad k \in \mathbb{N} \end{aligned}$$

0.7 Šibki in krepki zakon velikih števil

Definicija 0.38 (Verjetnostna konvergenca). Zaporedje slučajnih spremenljivk $\{X_n\}_{n \in \mathbb{N}}$ verjetnostno konvergira proti slučajni spremenljivki X , če za $\forall \epsilon > 0$ velja $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$ oz. $\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$

Definicija 0.39 (Skoraj gotova konvergenca). Zaporedje slučajnih spremenljivk $\{X_n\}_{n \in \mathbb{N}}$ skoraj gotovo konvergira proti slučajni spremenljivki X , če velja $P(\lim_{n \rightarrow \infty} X_n = X) = 1$

Tukaj je $(\lim_{n \rightarrow \infty} X_n = X) = \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} =$

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

$$= \{\cap_{k \in \mathbb{N}} \cup_{m \in \mathbb{N}} \cap_{n \geq m} \omega \in \Omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\} \quad (5)$$

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

Trditev 0.40. Če $X_n \xrightarrow{n \rightarrow \infty} X$ skoraj gotovo, potem za $\forall \epsilon > 0 \lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon \text{ za } n \geq m) = 1$

Dokaz. Označimo $c_m := (|X_n - X| < \epsilon \text{ za } n \geq m) = \cap_{n=m}^{\infty} (|x_n - X| < \epsilon)$.

Potem je $c_1 \subseteq c_2 \subseteq \dots$

je c_m za $\epsilon = \frac{1}{k}$ in $(\lim_{n \rightarrow \infty} X_n = X) \subseteq \cup_{n=1}^{\infty} c_m$ (preseki)

Torej je $1 = P(\lim_{n \rightarrow \infty} X_n = X) \subseteq (\cup_{m=1}^{\infty} c_m) = \lim_{m \rightarrow \infty} P(c_m)$

Od tod sledi $\lim_{m \rightarrow \infty} P(c_m) = 1$ ■

Posledica 0.41. Če $X_n \xrightarrow{n \rightarrow \infty} X$ skoraj gotovo, potem $X_n \xrightarrow{n \rightarrow \infty} X$ 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 $\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$ (leva stran). ■

Opomba. Obratna implikacija ne velja

Definicija 0.42. Naj bo $X_1, X_2, X_3 \dots$ zaporedje slučajnih spremenljivk, ki imajo matematično upanje. Definirajmo $Y_n = \frac{S_n - E(S_n)}{n} = \frac{X_1 + \dots + X_n}{n} - \frac{E(X_1) + \dots + E(X_n)}{n}$

Potem je $E(Y_n) = 0$

Za $\{Y_n\}_{n \in \mathbb{N}}$ velja šibki zakon velikih števil (ŠZVŠ), kadar $Y_n \xrightarrow{n \rightarrow \infty} 0$ verjetnostno, torej za $\forall \epsilon > 0 \lim_{n \rightarrow \infty} (|y| < \epsilon) = 1 = \lim_{n \rightarrow \infty} (|\frac{S_n - E(S_n)}{n}| < \epsilon)$ Za $\{Y_n\}_{n \in \mathbb{N}}$ velja krepki zakon velikih števil (KZVŠ), kadar $Y_n \xrightarrow{n \rightarrow \infty} 0$ skoraj gotovo, torej $P(\lim_{n \rightarrow \infty} \frac{S_n - E(S_n)}{n} = 0) = 1$

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 $Y_n = \frac{X_1 + \dots + X_n}{n} - \frac{7}{2}$

Ali konvergira $\frac{X_1 + \dots + X_n}{n} \xrightarrow{n \rightarrow \infty} \frac{7}{2}$ skoraj gotovo? (Da)

Izrek 0.43.

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

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

$$|a| \int_{\{|x| \geq a\}} p_x(x) dx = a \cdot P(|X| \geq a)$$

b

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

■

Izrek 0.44 (Markov). Če za zaporedje slučajnih spremenljivk $\{X_n\}_{n \in \mathbb{N}}$ velja $\frac{D(S_n)}{n^2} \xrightarrow{n \rightarrow \infty} 0$, potem velja ŠZVŠ. Tukaj je $S_n := X_1 + \dots + X_n$

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

$$P\left(\frac{|S_n - E(S_n)|}{n} \geq \epsilon\right) \leq \frac{P(S_n)}{n^2 \epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Če vzamemo $Y_n = \frac{|S_n - E(S_n)|}{n}$, je $P(|Y_n| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0$

oz. $P(|Y_n| < \epsilon) \xrightarrow{n \rightarrow \infty} 1$

Zato $Y_n \xrightarrow{n \rightarrow \infty} 0$ verjetnostno, torej velja ŠZVŠ za zaporedje $\{X_n\}_{n \in \mathbb{N}}$ ■

Posledica 0.45 (Izrek Čebišev). Če so X_1, X_2, \dots, X_n paroma nekorelirane slučajne spremenljivke in $\sup_{n \in \mathbb{N}} D(X_n) < \infty$, potem za $\{X_n\}_{n \in \mathbb{N}}$ velja ŠVZŠ

Dokaz. Ker je $D(S_n) = D(X_1) + \dots + D(X_n) \leq n \cdot c$, je $\frac{D(S_n)}{n^2} \leq \frac{n \cdot c}{n^2} = \frac{c}{n} \xrightarrow{n \rightarrow \infty} 0$, 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: $P\left(\frac{|S_n - E(S_n)|}{n} \geq \epsilon\right) \xrightarrow{n \rightarrow \infty} 0$

$$\implies P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

S_n je frekvenca dogodka, $\frac{S_n}{n}$ je relativna frekvenca, $\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \xrightarrow{n \rightarrow \infty} p$ verjetnostno

To je Bernoulijev zakon velikih števil iz 1713

Izrek 0.46 (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. $P(\lim_{n \rightarrow \infty} \frac{S_n - E(S_n)}{n} = 0) = 1$. 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. $\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \xrightarrow{n \rightarrow \infty} p$ skoraj gotovo. To splošuje Bernoulijev zakon

0.8 Centralni limitni izrek

Definicija 0.47. Naj bo $\{X_n\}_{n \in \mathbb{N}}$ zaporedje slučajnih spremenljivk s končnimi disperzijami. Definiramo $S_n := X_1 + \dots + X_n$ 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 $F_{Z_n}(x) = P(Z_n \leq x) \xrightarrow{n \rightarrow \infty} F_{N(0,1)} \forall x \in \mathbb{R}$, t.j.

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

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

Izrek 0.48 (Centralni limitni izrek (CLI, osnovna verzija)). Naj bodo $X_1, X_2 \dots$ neodvisne in enako porazdeljene slučajne spremenljivke. Potem zanje velja centralni limitni zakon, t.j.

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

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

Izrek 0.49 (O zveznosti rodovne funkcije). Naj za zaporedje $\{Z_n\}_{n \in \mathbb{N}}$ slučajnih spremenljivk velja:

$M_{Z_n}(t) \rightarrow M_{N(0,1)}(t) = e^{\frac{t^2}{2}}$ za vse $t \in (-\delta, \delta)$ pri nekem $\delta > 0$
Potem $F_{Z_n}(x) \rightarrow F_{N(0,1)}(x)$ 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 $M_U(t) = 1 + tE(U_n) + \frac{t^2}{2!}E(U_n^2) + o(t^2) = 1 + \frac{t^2}{2}\sigma^2 + o(t^2)$ ($\lim_{n \rightarrow \infty} \frac{o(n)}{n} = 0$)

Ker je $D(S_n) \stackrel{\text{neodvisne}}{=} D(X_1) + \dots + D(X_n) = n \cdot \sigma^2$ in $E(S_n) = n \cdot \mu = E(X_1) + \dots + E(X_n)$, je $Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{1}{\sigma\sqrt{n}} (\sum_{i=0}^n U_i)$

Potem je $M_{Z_n}(t) = E(e^{tZ_n}) = E(e^{\frac{t}{\sigma\sqrt{n}}(U_1 + \dots + U_n)}) = E(e^{\frac{t}{\sigma\sqrt{n}}U_1}) \dots E(e^{\frac{t}{\sigma\sqrt{n}}U_n}) = \stackrel{\text{enaki}}{=} (M_U(\frac{t}{\sigma\sqrt{n}}))^n = (1 + \frac{t^2}{2n} + o(\frac{1}{n}))^n$
 $\xrightarrow{n \rightarrow \infty \equiv o(\frac{1}{n} \rightarrow 0)} e^{\frac{t^2}{2}}$

Lema 0.50. Če $X_n \rightarrow X$, potem $(1 + \frac{X_n}{n})^n \xrightarrow{n \rightarrow \infty} e^x$

Po prejšnjem izreku: $F_{Z_n}(x) \xrightarrow{n \rightarrow \infty} F_{N(0,1)}(x)$

$\epsilon > 0 : x - \epsilon \leq x_n \leq x + \epsilon$ za dovolj velik n

$$\implies \left(1 + \frac{x - \epsilon}{n}\right)^n \leq \left(1 + \frac{x_n}{n}\right)^n \leq \left(1 + \frac{x + \epsilon}{n}\right)^n$$

$$\implies \left(1 + \frac{x - \epsilon}{n}\right)^n \rightarrow e^{x-\epsilon}$$

$$\implies \left(1 + \frac{x_n}{n}\right)^n \rightarrow e^x$$

$$\implies \left(1 + \frac{x + \epsilon}{n}\right)^n \rightarrow e^{x+\epsilon}$$

■

V splošnem se CLI dokaže s pomočjo karakterističnih funkcij:

naj bo X slučajna spremenljivka, $\ell_X(t) := E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX))$ $t \in \mathbb{R}$

za razliko od momentno rodovnih funkcij karakteristične funkcije vedno od-
stajajo

v zveznem primeru je $\int_{-\infty}^{\infty} e^{itx} p(x) dx$ - Fourierova transformacija funkcije $p_X(x)$

$X_1, X_2 \dots X_n$ neodvisne, enako porazdeljene

$$\mu := E(X_n), \sigma := \sigma(X_n)$$

$$E(S_n) \stackrel{\text{neodvisnost}}{=} E(X_1) + \dots + E(X_n) = n\mu$$

$$D(S_n) \stackrel{\text{neodvisnost}}{=} D(X_1) + \dots + D(X_n) = n\sigma^2$$

$X_1, X_2 \dots X_n$ neodvisne slučajne spremenljivke

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\bar{Z}_n := \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \implies Z_n = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Po CLI za velike n velja $Z_n \approx N(0, 1)$, zato je $\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$ oz. $S_n \approx N(n\mu, \sigma\sqrt{n})$

Če so X_1, X_2, \dots porazdeljene normalno $N(\mu, \sigma)$, potem je $Z_n \sim N(0, 1)$, torej $F_{Z_n}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$

Primer. Laplaceova formula je poseben primer CLI:

$X_n : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, $X_n = 1$ je dogodek, da se dogodek A (s $P(A) = p$) zgodi v n-ti ponovitvi poskusa, sicer je $X_n = 0$

$E(X_n) = p$, $S_n = X_1 + \dots + X_n$ frekvenca dogodka A v prvih n ponovitvah

$S_n \sim \text{Bin}(n, p)$, $E(S_n) = np$, $D(S_n) = npq$, ker je $D(X_1) = pq$

$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{npq}} \stackrel{\text{CLI}}{\approx} N(0, 1)$, če je n velik

$$\begin{aligned} P(S_n \leq X) &= P\left(\frac{S_n - np}{\sqrt{npq}} \leq \frac{X - np}{\sqrt{npq}}\right) \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x - np}{\sqrt{npq}}} e^{-\frac{t^2}{2}} dt = \\ &= \frac{1}{2} + \Phi\left(\frac{x - np}{\sqrt{npq}}\right) \end{aligned}$$

kjer je

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

verjetnostni integral

$$\begin{aligned} P(\alpha < S_n \leq \beta) &= \\ &= P(S_n \leq \beta) - P(S_n \leq \alpha) \approx \\ &\approx \frac{1}{2} + \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \frac{1}{2} - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right) = \\ &= \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right) \end{aligned}$$

Laplaceova aproksimacijska formula

Primer. Teža vrečke kostanja je porazdeljena približno normalno, saj je vsota tež posameznih kostanjev, ki so neodvisne, enako porazdeljene slučajne spremenljivke

$X_n \dots$ teža n-tega kostanja, $S_n = X_1 + \dots + X_n \approx$ normalno - aditiven efekt

Primer.

$$p_{X_n}(x) = \begin{cases} \frac{1}{2}; x \in [-1, 1] \\ 0 \text{ sicer} \end{cases}$$
$$E(X_1) = 0, D(X_1) = \frac{(b-a)^2}{12} = \frac{1}{3}$$
$$S_1 = X_1, Z_1 = \frac{X_1 - E(X_1)}{\sigma(X_1)} = \frac{X_1}{\sqrt{\frac{1}{3}}} = x_1\sqrt{3}$$
$$S_2 = X_1 + X_2, Z_2 = \frac{S_2 - E(S_2)}{\sigma(S_2)} = \frac{X_1 + X_2 - E(X_1 + X_2)}{\sigma(X_1 + X_2)}$$
$$S_3 = X_1 + X_2 + X_3, Z_3 = \frac{S_3 - E(S_3)}{\sigma(S_3)}$$

1 Statistika

1.1 Osnovni pojmi

Kot vedo statistiko razdelimo na:

1. opisno statistiko: zbiranje, razvrščanje, prikazovanje podatkov, računanje osnovnih količin
2. analitično statistiko: upraba podatkov pri sklepanju glede zakonitosti danega področja

Definicija 1.1 (Populacija). Populacija je končna ali neskončna množica elementov, pri katerih merimo ali opazujemo neko količino

Primer.

- (a) kontrole kvalitete: populacija je množica (serija) izdelka, npr. dnevna proizvodnja, merimo lastnosti izdelkov, npr. življensko dobo
- (b) testiranje seb: populacija je množica vseh zaposlenih v državi, merimo npr. starost, višino place ...

Matematični pogled: na verjetnostnem prostoru (Ω, \mathcal{F}) imamo slučajno spremenljivko X .

Praviloma ne moremo izmeriti cele populacije, ampak meritve opravimo na relativno majhnem delu populacije, na vzorcu. Le-ta mora biti reprezentativen, izbran nepristransko in dovolj velik.

Matematični pogled: vzorec velikosti n je slučajni vektor $(x_1 \cdots x_n)$, kjer so

komponente enako porazdeljene kot slučajna spremenljivka X in med seboj neodvisne.

Vrednost tega slučajnega vektorja pri enem naboru n meritev je realizacija vzorca: $(x_1 \cdots x_n)$: to so konkretni podatki, ki jih analiziramo. Pri opisni statistiki predstavimo in obdelamo te podatke.

Iz teh vzorčnih podatkov želimo oceniti nekatere lastnosti populacije, kot sta:

1. sredina populacije μ , t.i. matematično upanje slučajne spremenljivke X
2. povprečni odklon σ od sredine populacije, t.i. Standardna deviacija slučajne spremenljivke X