Slučajni vektori

Uvod

Zanima nas više slučajnih varijabli vezanih uz istu pojavu.

Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor i neka su $X, Y : \Omega \to \mathbb{R}$ dvije slučajne varijable. **Dvodimenzionalni slučajni vektor** je funkcija $(X, Y) : \Omega \to \mathbb{R}^2$, dana s $(X, Y)(\omega) = (X(\omega), Y(\omega))$.

Uočimo da za sliku od (X, Y) vrijedi $R(X, Y) \subseteq R(X) \times R(Y)$.

Shematski zapis

U slučaju da je

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix} \qquad \text{i} \qquad Y \sim \begin{pmatrix} y_1 & y_2 & \dots \\ q_1 & q_2 & \dots \end{pmatrix},$$

onda je (X, Y) **diskretni dvodimenzionalni slučajni vektor** i njegovu raspodjelu zapisujemo pomoću sheme:

$$(X, Y) \sim \begin{pmatrix} X \backslash Y & y_1 & y_2 & \dots & y_m \\ x_1 & p_{11} & p_{12} & \dots & p_{1m} \\ x_2 & p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & p_{n1} & p_{n2} & \dots & p_{nm} \end{pmatrix},$$

gdje su $R(X, Y) \subseteq R(X) \times R(Y) = \{(x_i, y_j) : x_i \in R(X), y_j \in R(Y)\}$
i

$$p_{ij} = \mathbb{P}((X, Y) = (x_i, y_j)) = \mathbb{P}(X = x_i, Y = y_j)$$

za $x_i \in R(X)$ i $y_j \in R(Y)$.

Funkcija vjerojatnosti i funkcija raspodjele

Funkcija vjerojatnosti $f : \mathbb{R}^2 \to \mathbb{R}$ od (X, Y) je dana s

$$f(x,y) = \begin{cases} p_{ij}, & x = x_i, y = y_j \\ 0, & \text{inače}, \end{cases}$$

a **funkcija raspodjele** $F : \mathbb{R}^2 \to [0,1]$ je dana s

$$F(x,y) = \mathbb{P}(X \leqslant x, Y \leqslant y) = \sum_{\substack{x_i \leqslant x,\, x_i \in R(X) \\ y_j \leqslant y,\, y_j \in R(Y)}} p_{ij}.$$

Očito, $p_{ij} = f(x_i, y_j)$ i $R(X,Y) = \{(x,y) \in \mathbb{R}^2 : f(x,y) > 0\}$.

Marginalne raspodjele

Uočimo

$$\sum_{y_j \in R(Y)} p_{ij} = \sum_{y_j \in R(Y)} \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) = p_i$$

i slično

$$\sum_{x_i \in R(X)} p_{ij} = \sum_{x_i \in R(X)} \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(Y = y_j) = q_j.$$

Dakle, ako zbrojimo sve elemente po y_j dobijemo raspodjelu od X i ako zbrojimo sve elemente po x_i dobijemo raspodjelu od Y. Te se raspodjele zovu **marginalne raspodjele** od (X, Y).

Slučajne varijable X i Y su nezavisne ako i samo ako je $p_{ij} = p_i q_j$, tj. $f(x, y) = f_X(x) f_Y(y)$.

Primjer

Bacamo dvije kocke. Neka je X zbroj brojeva koji su pali, a Y veći od brojeva koji su pali. Raspodjela od (X, Y) je dana s

$$\begin{pmatrix}
X\backslash Y & 1 & 2 & 3 & 4 & 5 & 6 & p_i \\
2 & \frac{1}{36} & 0 & 0 & 0 & 0 & 0 & \frac{1}{36} \\
3 & 0 & \frac{2}{36} & 0 & 0 & 0 & 0 & \frac{2}{36} \\
4 & 0 & \frac{1}{36} & \frac{2}{36} & 0 & 0 & 0 & \frac{3}{36} \\
5 & 0 & 0 & \frac{2}{36} & \frac{2}{36} & 0 & 0 & \frac{4}{36} \\
6 & 0 & 0 & \frac{1}{36} & \frac{2}{36} & \frac{2}{36} & 0 & \frac{5}{36} \\
7 & 0 & 0 & 0 & \frac{2}{36} & \frac{2}{36} & \frac{2}{36} & \frac{6}{36} \\
8 & 0 & 0 & 0 & \frac{1}{36} & \frac{2}{36} & \frac{2}{36} & \frac{5}{36} \\
9 & 0 & 0 & 0 & 0 & \frac{2}{36} & \frac{2}{36} & \frac{4}{36} \\
10 & 0 & 0 & 0 & 0 & \frac{1}{36} & \frac{2}{36} & \frac{3}{36} \\
11 & 0 & 0 & 0 & 0 & 0 & \frac{2}{36} & \frac{2}{36} \\
12 & 0 & 0 & 0 & 0 & 0 & \frac{1}{36} & \frac{1}{36} \\
q_j & \frac{1}{36} & \frac{3}{36} & \frac{5}{36} & \frac{7}{36} & \frac{9}{36} & \frac{11}{36} & 1
\end{pmatrix},$$

$p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$, $p_i = \mathbb{P}(X = x_i)$ i $q_j = \mathbb{P}(Y = y_j)$. $\square$

Neprekidni slučajni vektori – funkcija gustoće

Neka su $X, Y : \Omega \to \mathbb{R}$ neprekidne slučajne varijable. Dvodimenzionalni slučajni vektor $(X, Y) : \Omega \to \mathbb{R}^2$ je **neprekidni slučajni vektor** ako postoji (izmjeriva) funkcija $f : \mathbb{R}^2 \to \mathbb{R}$ za koju vrijedi:

(i) $f(x, y) \geqslant 0$

(ii) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$

(iii) $\displaystyle\mathbb{P}((X, Y) \in (-\infty, a] \times (-\infty, b]) = \int_{-\infty}^{a}\int_{-\infty}^{b} f(x, y)dxdy.$

Funkciju $f(x, y)$ zovemo **funkcija gustoće** od (X, Y).

Neprekidni slučajni vektori – funkcija raspodjele

Funkcija raspodjele $F : \mathbb{R}^2 \to [0,1]$ od (X,Y) je dana s

$$F(x,y) = \mathbb{P}(X \leqslant x, Y \leqslant y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t) ds dt.$$

Kada je funkcija gustoće neprekidna, vrijedi

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = \frac{\partial^2}{\partial y \partial x} F(x,y) = f(x,y).$$

Kao i u diskretnom slučaju imamo

$$\int_{-\infty}^{\infty} f(x,y) dy = f_X(x) \qquad \text{i} \qquad \int_{-\infty}^{\infty} f(x,y) dx = f_Y(y)$$

te $\lim_{y\to\infty} F(x,y) = F_X(x)$ i $\lim_{x\to\infty} F(x,y) = F_Y(y)$.
Integriranjem po y dobivamo raspodjelu od X, a integriranjem po x dobivamo raspodjelu od Y, tj. **marginalne raspodjele** od (X,Y).
Slučajne varijable X i Y su nezavisne ako i samo ako je $f(x,y) = f_X(x) f_Y(y)$ i $F(x,y) = F_X(x) F_Y(y)$.

Neprekidni slučajni vektori – primjer

Marko i Ivan dogovorili su sastanak na *Trgu bana Josipa Jelačića* u 12 h. Pretpostavimo da je njihov dolazak slučajan trenutak između 12 i 13 h. Neka slučajne varijable X i Y označavaju dolazak Marka i Ivana, redom.

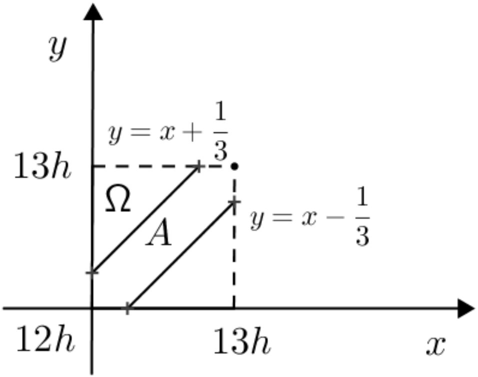
Pretpostavimo da je funkcija gustoće od (X, Y) dana s

$$f(x, y) = \begin{cases} 1, & x, y \in [12, 13] \\ 0, & \text{inače.} \end{cases}$$

Uočimo, $f_X(x) = \int_{12}^{13} f(x, y)dy = 1$ i $f_Y(y) = \int_{12}^{13} f(x, y)dx = 1$. Dakle, $X \sim U(12, 13)$ i $Y \sim U(12, 13)$. Nadalje, pretpostavimo da onaj koji stigne na dogovoreno mjesto prvi čeka onog drugog 20 min. Kolika je vjerojatnost da se sretnu?

Neprekidni slučajni vektori – primjer

Neka je $\Omega = [12, 13]^2$. Označimo s A događaj "Marko i Ivan su se sreli", tj. $A = \{\omega \in \Omega : |X(\omega) - Y(\omega)| \leqslant 1/3\}$.



Dakle, $\mathbb{P}(A) = \iint_A f(x, y)dxdy = \frac{5}{9}$.

Očekivanje funkcije neprekidnog slučajnog vektora

Neka je (X, Y) dvodimenzionalan slučajni vektor i $g : \mathbb{R}^2 \to \mathbb{R}$ neka funkcija. Tada je $g(X, Y)$ slučajna varijabla. Zanima nas prosjek $g(X, Y)$. Pokazuje se da to možemo izravno izračunati iz zajedničke raspodjele. Ako donja suma i integral postoje, onda imamo

$$\mathbb{E}(g(X, Y)) = \sum_{x_i \in R(X)} \sum_{y_j \in R(Y)} g(x_i, y_j) p_{ij}$$

za diskretan slučajni vektor te

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

za neprekidan slučajni vektor. U slučaju diskretnih slučajnih vektora funkcija $g(x, y)$ može biti proizvoljna.

Očekivanje funkcije neprekidnog slučajnog vektora – primjer

Poludjeli stroj za proizvodnju vaza na slučajan način proizvodi vaze radijusa $R \sim U(7.5, 12.5)$ i visine $H \sim U(25, 35)$, gdje su obje dimenzije izražene u centimetrima. Pretpostavimo da su R i H nezavisne. Odredimo očekivani volumen slučajno odabrane vaze.

Označimo s V volumen slučajno proizvedene vaze. Imamo

$$
\begin{aligned}
\mathbb{E}(V) &= \mathbb{E}(R^2 \pi H) = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r^2 \pi h f_H(h) f_R(r) dr dh \\
&= \pi \int_{25}^{35} \int_{7.5}^{12.5} r^2 h \cdot \frac{1}{10} \cdot \frac{1}{5} dr dh = \\
&= \frac{\pi}{50} \int_{25}^{35} h\, dh \int_{7.5}^{12.5} r^2 dr = 9621.127 \text{ cm}^3. \square
\end{aligned}
$$

Kovarijanca slučajnih varijabli

Za slučajne varijable X i Y vrijedi $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. Ako su X i Y još i nezavisne, onda je $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Što ako X i Y nisu nezavisne? Imamo

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}((X+Y-\mathbb{E}(X+Y))^2) \\
&= \mathbb{E}(((X-\mathbb{E}(X)) + (Y-\mathbb{E}(Y)))^2) \\
&= \mathbb{E}((X-\mathbb{E}(X))^2 + (Y-\mathbb{E}(Y))^2 + 2(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))).
\end{aligned}$$

Izraz $\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$, u oznaci $\mathrm{Cov}(X,Y)$, nazivamo **kovarijancom** od X i Y. Uočimo

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y)) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Koreliranost slučajnih varijabli

U slučaju nezavisnosti $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, iz čega slijedi da je $\mathrm{Cov}(X, Y) = 0$.

- Ako je $\mathrm{Cov}(X, Y) > 0$ kažemo da su X i Y **pozitivno korelirane**,
- za $\mathrm{Cov}(X, Y) < 0$ kažemo da su X i Y **negativno korelirane**,
- a za $\mathrm{Cov}(X, Y) = 0$ kažemo da su X i Y **nekorelirane**.

Pozitivna koreliranost znači da X i Y imaju sklonost odstupanja od očekivanja u istu stranu, a u slučaju negativne koreliranosti imaju sklonost odstupanja na različite strane od očekivanja. Napomenimo da nekoreliranost ne znači nezavisnost.

Nekoreliranost ne povlači nezavisnost

Neka je $X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$. Tada imamo $X^2 \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$ i $X^3 \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$. Dakle, $\mathrm{Cov}(X, X^2) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2) = 0$ pa su X i X^2 nekorelirane.

S druge strane, vrijedi

$$\mathbb{P}(X = 0, X^2 = 0) = \mathbb{P}(X = 0) = \frac{1}{3} \neq \mathbb{P}(X = 0)\mathbb{P}(X^2 = 0) = \frac{1}{9}$$

pa X i X^2 nisu nezavisne.

Koeficijent korelacije slučajnih varijabli

Definiramo

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Broj $\rho(X, Y)$ nazivamo **koeficijent korelacije**.
Vrijedi $\rho(rX + s, tY + u) = \rho(X, Y)$.

Uočimo da $\rho(X, Y)$ poprima vrijednosti između -1 i 1, a značenje predznaka je isto kao kod kovarijance.

Stupanj linearne zavisnosti

Koeficijentom korelacije zapravo mjerimo stupanj linearne zavisnosti dviju slučajnih varijabli.

(i) $|\rho(X,Y)| = 1$ ako i samo ako postoje $a, b \in \mathbb{R}$ t.d. $Y = aX + b$ te $\sigma(X) > 0$, $\sigma(Y) > 0$ (pa nužno $a \neq 0$)

(ii) manji $|\rho(X,Y)|$ znači "manji stupanj linearne zavisnosti" od X i Y

(iii) $\rho(X,Y) = 0$ znači da su $X - \mathbb{E}(X)$ i $Y - \mathbb{E}(Y)$ "ortogonalne".

Dijagrami raspršenja za različite koeficijente korelacije