Slučajne varijable i funkcija razdiobe

Neka je $(\Omega, \mathscr{F}, \boldsymbol{P})$ vjerojatnosni prostor. Preslikavanje $X : \Omega \to \mathbf{R}$ nazivamo **slučajna varijabla** ako je za svaki $x \in \mathbf{R}$ skup

$$A_x = \{\omega \in \Omega : X(\omega) < x\}$$

događaj, dakle element algebre $\mathscr{F}$.

Skup $\{\omega \in \Omega : X(\omega) < x\}$ označavat ćemo kraće sa $\{X < x\}$.

Funkcija razdiobe slučajne varijable X je funkcija $F : \mathbf{R} \to [0, 1]$ definirana formulom

$$F(x) := \boldsymbol{P}(\{X < x\}). \tag{5.1}$$

Funkcija razdiobe

Teorem 5.1. *Neka je F funkcija razdiobe slučajne varijable X. Ona posjeduje svojstva:*

1° $\boldsymbol{P}(\{x_1 \leqslant X < x_2\}) = F(x_2) - F(x_1)$,

2° *F je neopadajuća:* $x_1 < x_2 \implies F(x_1) \leqslant F(x_2)$,

3° $\lim\limits_{x\to-\infty} F(x) = 0$, $\lim\limits_{x\to\infty} F(x) = 1$,

4° *F je neprekinuta slijeva:*

$$F(x-0) := \lim_{\varepsilon\downarrow 0} F(x-\varepsilon) = F(x), \quad \forall x \in \mathbf{R}.$$

Neprekinute slučajne varijable. Gustoća razdiobe

Za slučajnu varijablu X kažemo da je **neprekinuta** (kontinuirana) ako postoji nenegativna funkcija $f : \mathbf{R} \to \mathbf{R}$ takva da vrijedi

$$F(x) = \int_{-\infty}^{x} f(t)dt. \tag{5.2}$$

Funkcija f naziva se **gustoća razdiobe vjerojatnosti** slučajne varijable X. Ona nije nužno neprekinuta, no u točkama neprekinutosti od f vrijedi

$$f(x) = \frac{dF(x)}{dx}. \tag{5.3}$$

Na sreću odabrani broj. Jednolika razdioba

Kažemo da biramo *na sreću* broj unutar intervala $[a, b]$ ako je vjerojatnost da će on biti izabran unutar nekog podintervala proporcionalna duljini tog podintervala.

Za slučajnu varijablu koja uzima vrijednost ovako izabranog broja kažemo da ima **jednoliku** (**uniformnu**) razdiobu na intervalu $[a, b]$.

Jednolika razdioba, alternativna definicija, razdioba i gustoća

Za slučajnu varijablu X kažemo da je **jednoliko** (**uniformno**) distribuirana na intervalu $[a, b]$, ako je zadana funkcijom razdiobe odnosno funkcijom gustoće:

$$F(x) = \frac{x-a}{b-a}, \qquad a \leqslant x \leqslant b,$$

$$f(x) = \frac{1}{b-a}, \qquad a \leqslant x \leqslant b.$$

Pišemo $X \sim \mathscr{U}(a, b)$.

Definicija nezavisnosti

Kažemo da su slučajne varijable X i Y **nezavisne**, ukoliko za sve intervale A, B iz skupa $\mathbf{R}$ vrijedi

$$\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B).$$

Svojstva očekivanja i disperzije

Za sve slučajne varijable X, Y i realne brojeve s, t vrijedi

$$\boldsymbol{E}(sX + tY) = s\boldsymbol{E}(X) + t\boldsymbol{E}(Y)$$

(svojstvo **linearnosti** očekivanja). Za disperziju pak vrijedi

$$\boldsymbol{D}(sX) = s^2\boldsymbol{D}(X).$$

Ako su X i Y nezavisne, onda vrijedi

$$\boldsymbol{E}(XY) = \boldsymbol{E}(X)\boldsymbol{E}(Y),$$
$$\boldsymbol{D}(X+Y) = \boldsymbol{D}(X) + \boldsymbol{D}(Y).$$

Riemann-Stieltjesov integral, definicija

Kažemo da je g Riemann-Stieltjes integrabilna u odnosu na F, ako postoji limes integralnih suma, neovisno o izboru particije i točaka $\tilde{x}_i \in [x_{i-1}, x_i]$. Taj limes nazivamo **Riemann-Stieltjesov integral**, a označavamo na sljedeći način:

$$\int_a^b g(x)dF(x) := \lim_{\Delta\to 0} S(\mathscr{P}, g, F)$$

Riemann-Stieltjesov integral, način računanja

Ako F ima skokove iznosa p_i u točkama c_i, za njezin Riemann-Stieltjesov integral vrijedi

$$\int_a^b g(x)dF(x) = \sum_{i=1}^n g(c_i) \cdot p_i.$$

Ako je F neprekinuto diferencijabilna funkcija, onda vrijedi

$$\int_a^b g(x)dF(x) = \int_a^b g(x)F'(x)dx.$$

Prema tome, korištenjem Riemann-Stieltjesovog integrala mi ćemo istovremeno pokrivati obje važne klase slučajnih varijabli, diskretne i neprekinute slučajne varijable.

Tako, na primjer, očekivanje neke slučajne varijable možemo izraziti formulom

$$E(X) = \int_{-\infty}^{\infty} x\,dF(x),$$

a disperziju

$$D(X) = \int_{-\infty}^{\infty} x^2\,dF(x) - E(X)^2.$$

Integrali su Riemann-Stieltjesovi.

Transformacija funkcije gustoće

Neka je $Y = \psi(X)$. Ako je funkcija ψ rastuća ili padajuća funkcija, onda vrijedi

$$g(y) = f(x)\left|\frac{dx}{dy}\right|, \qquad y = \psi(x), \tag{5.1}$$

tj.

$$g(y) = f(\psi^{-1}(y))\left|\frac{d\psi^{-1}(y)}{dy}\right|.$$

Općenitije, ova formula vrijedi za svaku injektivnu funkciju ψ.

Eksponencijalna razdioba, definicija

Kažemo da slučajna varijabla X ima **eksponencijalnu razdiobu** s parametrom $\lambda > 0$ ako ona poprima pozitivne vrijednosti s gustoćom

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0. \tag{6.1}$$

Pišemo $X \sim \mathscr{E}(\lambda)$. Njezina je funkcija razdiobe

$$F(x) = 1 - e^{-\lambda x}, \qquad x > 0.$$

Karakterizacija eksponencijalne razdiobe

Teorem 6.2. *Neka za slučajnu varijablu X koja uzima samo pozitivne vrijednosti za sve pozitivne x i t vrijedi*

$$P(X < x + t \mid X > t) = P(X < x). \tag{6.2}$$

Tada X ima eksponencijalnu razdiobu.

Normalna razdioba, definicija

Slučajna varijabla X ima **normalnu razdiobu** s parametrima $a \in \mathbf{R}$ i $\sigma^2 > 0$ ako je X neprekinuta slučajna varijabla s gustoćom

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-a)^2}{2\sigma^2}\Big). \tag{6.1}$$

Pišemo $X \sim \mathscr{N}(a, \sigma^2)$.

Veza jedinične i opće normalne razdiobe

Jedinična i opća normalna razdioba mogu se dobiti jedna iz druge *linearnom transformacijom*:

$$X \sim \mathscr{N}(0,1) \implies a + \sigma X \sim \mathscr{N}(a, \sigma^2),$$

$$X \sim \mathscr{N}(a, \sigma^2) \implies \frac{X-a}{\sigma} \sim \mathscr{N}(0,1).$$

Računanje vjerojatnosti za jediničnu normalnu razdiobu

Za jediničnu normalnu razdiobu X vrijedi:

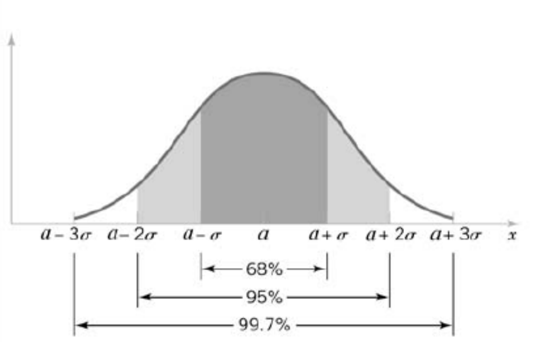
$$P(u_1 < X < u_2) = \Phi(u_2) - \Phi(u_1) = \frac{1}{2}\Big[\Phi^*(u_2) - \Phi^*(u_1)\Big]. \tag{6.5}$$

Posebice, u slučaju simetričnog intervala

$$P(|X| < u) = \frac{1}{2}\Big[\Phi^*(u) - \Phi^*(-u)\Big] = \Phi^*(u). \tag{6.6}$$

Pravilo 3σ

Neka je $X \sim \mathcal{N}(a, \sigma^2)$ Izračunajmo
$$P(|X-a| < k\sigma), \qquad k = 1, 2, 3.$$



Sl. 6.5.

Imamo
$$P(|X-a| < k\sigma) = P(-k\sigma < X-a < k\sigma) = P(-k < \widetilde{X} < k) = \Phi^*(k).$$
Vrijedi $\Phi^*(1) = 0.6827$, $\Phi^*(2) = 0.9545$, $\Phi^*(3) = 0.9973$. Dakle, s vjerojatnošću 99.73% (odnosno, praktički sigurno) normalna varijabla uzima vrijednosti unutar intervala $(a-3\sigma, a+3\sigma)$. Ta se činjenica naziva *pravilo tri sigma*.

Stabilnost normalne razdiobe

Teorem 6.3. *Neka su X_1 i X_2 nezavisne slučajne varijable s normalnim razdiobama*
$$X_1 \sim \mathcal{N}(a_1, \sigma_1^2), \quad X_2 \sim \mathcal{N}(a_2, \sigma_2^2)$$
i s_1, s_2 bilo koji realni brojevi. Tada vrijedi
$$s_1X_1 + s_2X_2 \sim \mathcal{N}(s_1a_1 + s_2a_2, s_1^2\sigma_1^2 + s_2^2\sigma_2^2).$$

Lokalni teorem Moivre-Laplacea

Teorem 6.4. *Vjerojatnosti realizacija binomne slučajne varijable $\mathcal{B}(n,p)$ mogu se aproksimirati pomoću funkcije gustoće normalne varijable $\mathcal{N}(np, npq)$:*
$$P(X=m) \approx P(m-\tfrac{1}{2} < Y < m+\tfrac{1}{2}) \approx f(m) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(m-np)^2}{2npq}}.$$

Teorem Moivre-Laplace

Teorem 6.5. *Neka je $X \sim \mathcal{B}(n,p)$. Za veliki n razdioba slučajne varijable $\dfrac{X-np}{\sqrt{npq}}$ može se aproksimirati jediničnom normalnom razdiobom:*
$$P\left(x_1 < \frac{\mathcal{B}(n,p)-np}{\sqrt{npq}} < x_2\right) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}u^2} du. \tag{6.7}$$
Pišemo $\mathcal{B}(n,p) \approx \mathcal{N}(np, npq)$, za dovoljno veliki n.

Gustoća razdiobe. Neprekinuti slučajni vektori

Za slučajan vektor $(X_1, \dots, X_n)$ kažemo da je **neprekinut** ako postoji funkcija $f : \mathbf{R}^n \to [0, \infty\rangle$ takva da za sve $x_1, x_2, \dots, x_n$ vrijedi
$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \dots, u_n) du_1 \dots du_n. \tag{7.2}$$
Funkciju f nazivamo **gustoća razdiobe** slučajnog vektora. Tamo gdje je F diferencijabilna, vrijedi
$$f(x_1, \dots, x_n) := \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n} \tag{7.3}$$

Računanje vjerojatnosti realizacije slučajnog vektora

Za svaki izmjerivi skup $G \subset \mathbf{R}^n$ vrijedi
$$P((X_1, \dots, X_n) \in G) = \int \cdots \int_G f(x_1, \dots, x_n) dx_1 \cdots dx_n. \tag{7.4}$$

Za dvodimenzionalni slučajni vektor (X,Y) koristit ćemo jednostavnije oznake. Za gustoću vrijedi

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

marginalne razdiobe:

$$F_X(x) := F(x,\infty) = \int_{-\infty}^{x} dx \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{x} f_X(x)dx,$$
$$F_Y(y) := F(\infty,y) = \int_{-\infty}^{y} dy \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{y} f_Y(y)dy,$$

marginalne gustoće:

$$f_X(x) := \int_{-\infty}^{\infty} f(x,y)dy,$$
$$f_Y(y) := \int_{-\infty}^{\infty} f(x,y)dx.$$

Kriterij nezavisnosti za neprekinute slučajne vektore

Teorem 7.1. *Komponente $X_1,\dots,X_n$ neprekinutog slučajnog vektora $(X_1,\dots,X_n)$ su nezavisne onda i samo onda ako vrijedi*

$$f(x_1,\dots,x_n) = f_1(x_1)\cdot\ldots\cdot f_n(x_n), \qquad \forall (x_1,\dots,x_n)\in \mathbf{R}^n. \tag{7.6}$$

Svojstva očekivanja

Teorem 7.2. *Za svake dvije slučajne varijable $X,Y:\Omega\to\mathbf{R}$ vrijedi*

$$E(X+Y) = E(X) + E(Y). \tag{7.7}$$

Ako su X i Y nezavisne, onda je

$$E(XY) = E(X)\cdot E(Y). \tag{7.8}$$

Disperzija zbroja

Teorem 7.4. *Ako su $X_1,X_2,\dots,X_n$ nekorelirane slučajne varijable, tada vrijedi*

$$D(X_1+X_2+\ldots+X_n) = D(X_1)+D(X_2)+\ldots+D(X_n).$$

Uvjetna gustoća

Neka je $f(x,y)$ gustoća razdiobe slučajnog vektora (X,Y). Ako je poznata realizacija $Y=y$ varijable Y, tada se **uvjetna gustoća** varijable X uz uvjet $Y=y$ definira formulom

$$f_{X|Y=y}(x) := \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}. \tag{7.1}$$

Uglavnom pišemo jednostavnije $f(x \mid y)$, umjesto $f_{X|Y=y}(x)$.

Marginalne i uvjetne gustoće

Marginalne gustoće možemo računati formulama

$$f_X(x) = \int_{-\infty}^{\infty} f(x \mid y) f_Y(y)dy,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(y \mid x) f_X(x)dx.$$

Uvjetna očekivanja i vjerojatnosti

Očekivanje varijable X koja ovisi o realizacijama varijable Y:

$$E(X) = \int_{-\infty}^{\infty} E(X \mid Y{=}y) f_Y(y)dy. \tag{7.2}$$

Vjerojatnost događaja koji ovisi o realizacijama slučajne varijable X:

$$P(A) = \int_{-\infty}^{\infty} P(A \mid X{=}x) f_X(x)dx. \tag{7.3}$$

Gustoća funkcije slučajnog vektora

Gustoća slučajne varijable $Z=\psi(X,Y)$ dobiva se formulom

$$g_Z(z) = \int_{-\infty}^{\infty} f(x,y)\left|\frac{\partial y}{\partial z}\right| dx, \tag{8.4}$$

gdje je f gustoća vektora (X,Y).