Zvezne slučajne spremenljivke

Slučajna spremenljivka X je *zvezno porazdeljena*, če obstaja taka integrabilna funkcija p, imenovana *gostota verjetnosti*, da za vsak $x \in \mathbb{R}$ velja:

$$F(x) = P(X < x) = \int_{-\infty}^{x} p(t)dt$$

kjer $p(x) \geq 0$. To verjetnost si lahko predstavimo tudi grafično v koordinatnem sistemu, kjer na abscisno os nanašamo vrednosti slučajne spremenljivke, na ordinatno pa gostoto verjetnosti $p(x)$. Verjetnost je tedaj predstavljena kot ploščina pod krivuljo, ki jo določa $p(x)$. Velja

$$\int_{-\infty}^{\infty} p(x)dx = 1 \quad \text{in} \quad P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} p(t)dt$$

ter $p(x) = F'(x)$.

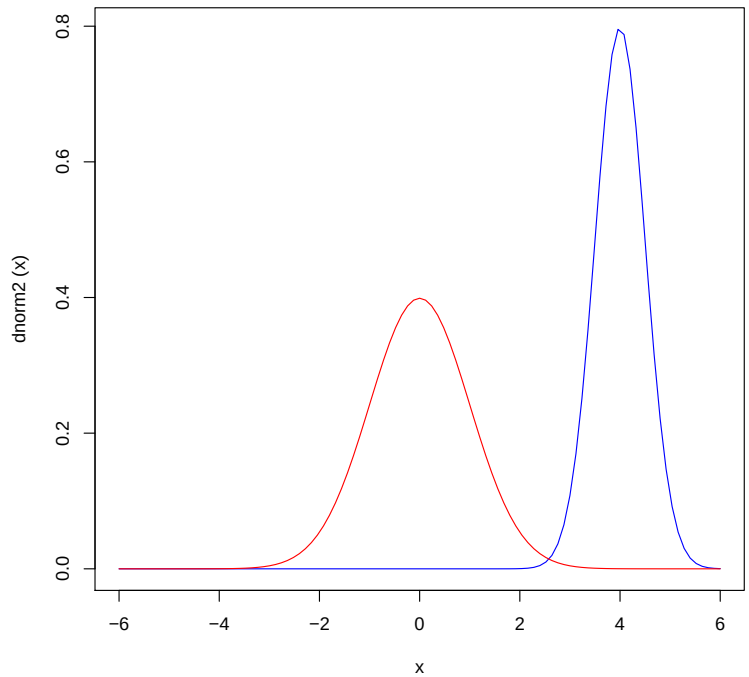
Enakomerna porazdelitev zvezne slučajne spremenljivke

Verjetnostna gostota enakomerno porazdeljene zvezne slučajne spremenljivke je:

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

Grafično si jo predstavljamo kot pravokotnik nad intervalom (a, b) višine $\frac{1}{b-a}$.

Normalna ali Gaussova porazdelitev




```
> d2 <- function(x){dnorm(x,mean=4,sd=0.5)}
> curve(d2,-6,6,col='blue')
> curve(dnorm,-6,6,col='red',add=TRUE)
```

Zaloga vrednosti *normalno porazdeljene* slučajne spremenljivke so vsa realna števila, gostota verjetnosti pa je:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Normalna porazdelitev je natanko določena z dvema parametroma: μ in σ. Če se slučajna spremenljivka X porazdeljuje normalno s parametroma μ in σ, zapišemo:

$$X : N(\mu, \sigma)$$

...Normalna porazdelitev

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}s^2} ds = \frac{1}{2} + \Phi(\frac{x-\mu}{\sigma})$$

$$P(x_1 \leq X < x_2) = \Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})$$

Porazdelitev $N(0,1)$ je *standardizirana normalna porazdelitev*.
Spremenljivko $X : \ N(\mu, \sigma)$ pretvorimo z $z = \frac{x-\mu}{\sigma}$ v standardizirano spremenljivko $Z : \ N(0,1)$.

Iz Laplaceovega obrazca izhaja $B(n,p) \approx N(np, \sqrt{npq})$.

Porazdelitev Poissonovega toka, eksponentna

Gostota *eksponentne porazdelitve* je enaka $p(x) = \lambda e^{-\lambda x}$, $x \geq 0$ porazdelitvena funkcija pa

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

Porazdelitev gama

Naj bosta $b, c > 0$. Tedaj ima *porazdelitev gama* $\Gamma(b, c)$ gostoto:

$$p(x) = \frac{c^b}{\Gamma(b)} x^{b-1} e^{-cx}, \ 0 < x$$

in $p(x) = 0$ za $x \leq 0$.

Poseben primer je *porazdelitev hi-kvadrat* $\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$ z gostoto ($n \in \mathbb{N}$ je število prostostnih stopenj)

$$p(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ 0 < x$$

Cauchyeva porazdelitev

z gostoto

$$p(x) = \frac{a}{\pi}\frac{1}{1+a^2(x-b)^2}, \quad -\infty < x < \infty,\ a > 0$$

ima porazdelitveno funkcijo

$$F(x) = \frac{a}{\pi}\int_{-\infty}^{x}\frac{1}{1+a^2(x-b)^2}\mathrm{d}x = \frac{1}{\pi}\operatorname{arc\,tg}(a(x-b)) + \frac{1}{2}$$

Porazdelitve v R-ju

V R-ju so za delo s pomembnejšimi porazdelitvami na voljo funkcije:

`d`*ime* – gostota porazdelitve *ime* $p_{ime}(x)$

`p`*ime* – porazdelitvena funkcija *ime* $F_{ime}(q)$

`q`*ime* – obratna funkcija: $q = F_{ime}(p)$

`r`*ime* – slučajno zaporedje iz dane porazdelitve

Za *ime* lahko postavimo: `unif` – zvezna enakomerna, `binom` – binomska, `norm` – normalna, `exp` – eksponentna, `lnorm` – logaritmičnonormalna, `chisq` – porazdelitev $\chi^2, \ldots$

Opis posamezne funkcije in njenih parametrov dobimo z ukazom `help`. Na primer `help(rnorm)`.

Slučajni vektorji

Slučajni vektor je n-terica slučajnih spremenljivk $X = (X_1, X_2, X_3, \ldots, X_n)$.
Opišemo ga s porazdelitveno funkcijo ($x_i \in \mathbb{R}$)

$$F(x_1, x_2, x_3, \ldots, x_n) = P(X_1 < x_1, X_2 < x_2, X_3 < x_3, \ldots, X_n < x_n)$$

za katero velja:

$0 \leq F(x_1, x_2, x_3, \ldots, x_n) \leq 1$

Funkcija F je za vsako spremenljivko naraščajoča in od leve zvezna.

$F(-\infty, -\infty, -\infty, \ldots, -\infty) = 0$ in $F(\infty, \infty, \infty, \ldots, \infty) = 1$.

Funkciji $F_i(x_i) = F(\infty, \infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$ pravimo
robna porazdelitvena funkcija spremenljivke X_i.

Slučajni vektorji – primer

Naj bo $A(x,y) = \{(u,v) \in \mathbb{R}^2 : u < x \wedge v < y\}$ (levi spodnji kvadrant glede na (x,y)). Naj porazdelitvena funkcija opisuje verjetnost, da je slučajna točka (X,Y) v množici $A(x,y)$

$$F(x,y) = P(X < x, Y < y) = P((X,Y) \in A(x,y)).$$

Tedaj je verjetnost, da je slučajna točka (X,Y) v pravokotniku $[a,b) \times [c,d)$ enaka

$$P((X,Y) \in [a,b) \times [c,d)) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

Diskretne večrazsežne porazdelitve

Zaloga vrednosti je kvečjemu števna množica. Opišemo jo z *verjetnostno funkcijo* $p_{k_1,k_2,\ldots,k_n} = P(X_1 = x_{k_1}, X_2 = x_{k_2}, \ldots, X_n = x_{k_n})$.

Za $n = 2$, $X : \{x_1, x_2, \ldots, x_k\}$, $Y : \{y_1, y_2, \ldots, y_m\}$ in $P(X = x_i, Y = y_j)$, sestavimo *verjetnostno tabelo*:

$X \setminus Y$	y_1	y_2	$\ldots$	y_m	X
x_1	p_{11}	p_{12}	$\ldots$	p_{1m}	p_1
x_2	p_{21}	p_{22}	$\ldots$	p_{2m}	p_2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_k	p_{k1}	p_{k2}	$\ldots$	p_{km}	p_k
Y	q_1	q_2	$\ldots$	q_m	1

$p_i = P(X = x_i) = \sum_{j=1}^{m} p_{ij}$ in $q_j = P(Y = y_j) = \sum_{i=1}^{k} p_{ij}$

Diskretne večrazsežne porazdelitve – Polinomska

Polinomska porazdelitev $P(n; p_1, p_2, \ldots, p_r)$, $\sum p_i = 1$, $\sum k_i = n$ je določena s predpisom

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r) = \frac{n!}{k_1! k_2! \cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}.$$

Za $r = 2$ dobimo binomsko porazdelitev, tj. $B(n, p) = P(n; p, q)$.

Zvezne večrazsežne porazdelitve

Slučajni vektor $X = (X_1, X_2, X_3, \ldots, X_n)$ je *zvezno porazdeljen*, če obstaja integrabilna funkcija (*gostota verjetnosti*) $p(x_1, x_2, \ldots, x_n) \geq 0$ z lastnostjo

$$F(x_1, x_2, x_3, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_n} p(t_1, t_2, \ldots, t_n) dt_1 dt_2 \ldots dt_n$$

$F(\infty, \infty, \infty, \ldots, \infty) = 1.$

Zvezne dvorazsežne porazdelitve

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(u,v)dudv$$

$$P((X,Y) \in [a,b) \times [c,d)) = \int_{a}^{b} \int_{c}^{d} p(u,v)dudv$$

Kjer je p zvezna je

$$\frac{\partial F}{\partial x} = \int_{-\infty}^{y} p(x,v)dv \quad \text{in} \quad \frac{\partial^2 F}{\partial x \partial y} = p(x,y)$$

Robni verjetnostni gostoti sta

$$p_X(x) = F'_X(x) = \int_{-\infty}^{\infty} p(x,y)dy$$

$$p_Y(y) = F'_Y(y) = \int_{-\infty}^{\infty} p(x,y)dx$$

Večrazsežna normalna porazdelitev

V dveh razsežnostih $N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ ima gostoto

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}.$$

V splošnem pa jo zapišemo v matrični obliki

$$p(\mathbf{x}) = \sqrt{\frac{\det A}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T A(\mathbf{x}-\mu)}$$

kjer je A simetrična pozitivno definitna matrika.

Vse robne porazdelitve so normalne.

Neodvisnost slučajnih spremenljivk

Podobno kot pri dogodkih:

Slučajne spremenljivke $X_1, X_2, X_3, \ldots, X_n$ so med seboj *neodvisne*, če za poljubne vrednosti $x_1, x_2, x_3, \ldots, x_n \in \mathbb{R}$ velja

$$F(x_1, x_2, x_3, \ldots, x_n) = F_1(x_1) \cdot F_2(x_2) \cdot F_3(x_3) \cdots F_n(x_n)$$

kjer je F porazdelitvena funkcija vektorja, F_i pa so porazdelitvene funkcije njegovih komponent.

Če sta

$$X : \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \text{ in } Y : \begin{pmatrix} y_1 & y_2 & \cdots \\ q_1 & q_2 & \cdots \end{pmatrix}$$

diskretni slučajni spremenljivki in p_{ij} verjetnostna funkcija slučajnega vektorja (X, Y), potem sta X in Y neodvisni natanko takrat, ko je $p_{ij} = p_i q_j$ za vsak par i, j.

...Neodvisnost slučajnih spremenljivk

Če sta X in Y zvezno porazdeljeni slučajni spremenljivki z gostotama p_X in p_Y ter je p gostota zvezno porazdeljenega slučajnega vektorja (X, Y), potem sta X in Y neodvisni natanko takrat, ko za vsak par x, y velja $p(x, y) = p_X(x) \cdot p_Y(y)$.

Primer: Naj bo dvorazsežni vektor (X, Y) normalno porazdeljen po $N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. Če je $\rho = 0$ je

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}((\frac{x-\mu_x}{\sigma_x})^2 + (\frac{y-\mu_y}{\sigma_y})^2)} = p_X(x) \cdot p_Y(y)$$

Torej sta komponenti X in Y neodvisni.

...Neodvisnost slučajnih spremenljivk

Zvezno porazdeljeni slučajni spremenljivki X in Y sta neodvisni natanko takrat, ko lahko gostoto verjetnosti slučajnega vektorja (X, Y) zapišemo v obliki $p(x, y) = f(x) \cdot g(y)$.

Naj bosta zvezno porazdeljeni slučajni spremenljivki X in Y tudi neodvisni ter A in B poljubni (Borelovi) podmnožici v $\mathbb{R}$. Potem sta neodvisna tudi dogodka $X \in A$ in $Y \in B$.

Trditev velja tudi za diskretni slučajni spremenljivki X in Y.

Pogosto pokažemo odvisnost spremenljivk X in Y tako, da najdemo množici A in B, za kateri je

$$P(X \in A, Y \in B) \neq P(X \in A) \cdot P(Y \in B)$$