Večrazsežne porazdelitve

Diskretne večrazsežne porazdelitve

Funkcija gostote verjetnosti:

$$f_{XY}(x,y) = P(X = x, Y = y)$$

$$0 \leq f_{XY}(x,y) \leq 1$$

$$\sum_x \sum_y f_{XY}(x,y) = 1$$

Robne porazdelitve:

$$f_X(x) = P(X = x) = \sum_{R_x} f_{XY}(x,y)$$

$$f_Y(y) = P(Y = y) = \sum_{R_y} f_{XY}(x,y)$$

Pogojne porazdelitve:

$$f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$f_{Y|x}(y) = P(Y = y | X = x)$$

$$0 \leq f_{Y|x}(y) \leq 1$$

$$\sum_{R_x} f_{Y|x}(y) = 1$$

Neodvisnost spremenljivk X in Y:

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

$$f_{Y|x}(y) = f_Y(y)$$

$$f_{X|y}(x) = f_X(x)$$

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

Matematično upanje:

$$E(X) = \sum_x x f_X(x) = \sum_R x f_{XY}(x,y)$$

$$E(Y|x) = \sum_{R_x} y f_{Y|x}(y)$$

Zvezne večrazsežne porazdelitve

Funkcija gostote verjetnosti:

$$\int\int_R f_{XY}(x,y)\,dx\,dy = P([X,Y]\in R)$$

$$0 \le f_{XY}(x,y) \le 1$$

$$P(X=x, Y=y) = 0$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$

Robne porazdelitve:

$$f_X(x) = \int_{R_x} f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_{R_y} f_{XY}(x,y)\,dx$$

Pogojne porazdelitve:

$$f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$\int_B f_{Y|x}(y)\,dy = P(Y\in B|X=x)$$

$$0 \le f_{Y|x}(y) \le 1$$

$$\int_{R_x} f_{Y|x}(y)\,dy = 1$$

Neodvisnost spremenljivk X in Y:

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

$$f_{Y|x}(y) = f_Y(y)$$

$$f_{X|y}(x) = f_X(x)$$

$$P(X\in A, Y\in B) = P(X\in A)P(Y\in B)$$

Matematično upanje:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\,dx\,dy$$

$$E(Y|x) = \int_{R_x} y f_{Y|x}(y)\,dy$$

VAJA 7: Večrazsežne verjetnostne porazdelitve

1. Najprej si oglejmo dve izmed najbolj znanih večrazsežnih verjetnostnih porazdelitev.

   - **Multinomska porazdelitev**, $(X_1, \ldots, X_k) \sim \text{Multinom}(n, \mathbf{p})$, je posplošitev binomske porazdelitve, kjer je $n \geq 1$ število paroma neodvisnih poskusov, katerih izide lahko razvrstimo v enega izmed k razredov z verjetnostmi $p_1, \ldots, p_k$ (velja $p_1 + \cdots + p_k = 1$), slučajne spremenljivke $X_1, \ldots, X_k$ pa označujejo število poskusov, katerih izide razvrstimo v razrede $1, \ldots, k$ (velja $x_1 + \cdots + x_k = n$). Označili smo $\mathbf{p} = (p_1, \ldots, p_k)$.

     V programu R so na voljo naslednji ukazi:

     – `dmultinom(x,prob=p)` za funkcijo gostote verjetnosti,

     $$P(X_1 = x_1, \ldots, X_k = x_k) = f(x_1, \ldots, x_k),$$

     kjer sta $\mathbf{x}$=$c(x_1, \ldots, x_k)$ in $\mathbf{p}$=$c(p_1, \ldots, p_k)$,

     – `rmultinom(l,size=n,prob=p)` za funkcijo naključnega generiranja izidov l serij s po n neodvisnimi poskusi, kjer je $\mathbf{p}$=$c(p_1, \ldots, p_k)$.

   - **Dvorazsežna normalna porazdelitev**, $(X, Y) \sim \text{Bivariate}(\mu, \sigma, \rho)$, je posplošitev enorazsežne normalne porazdelitve na primer dveh normalno porazdeljenih spremenljivk, ki sta odvisni – njun korelacijski koeficient je $\rho \in [-1, 1]$. Pri tem sta $\mu = (\mu_X, \mu_Y) \in \mathbb{R}^2$ vektor povprečnih vrednosti, $\sigma = (\sigma_X, \sigma_Y) \in \mathbb{R}^2$ pa vektor standardnih odklonov spremenljivk X in Y.

     Funkcijo gostote verjetnosti bomo sami definirali:

```
bivariate = function(x,y){
        t = 1 / (2 * pi * sig1 * sig2 * sqrt(1 - rho^2))
        z = (x - mu1)^2 / sig1^2 - (2 * rho * (x - mu1)*(y - mu2))
        /(sig1 * sig2) + (y - mu2)^2 / sig2^2
        fxy = t * exp((-z / (2 *(1 - rho^2))))
        return(fxy)
}
```

     Vrednost $f_{XY}(x, y)$ funkcije gostote verjetnosti dvorazsežne normalne porazdelitve sedaj dobimo s klicem `bivariate(x,y)`.

     Podobno lahko definiramo tudi ostale funkcije (gostota verjetnosti, porazdelitvena funkcija, funkcija kvantilov, ...), ki opisujejo izbrano porazdelitev. Splošna oblika definicije funkcije v programskem paketu R:

```
ime = function(x){
        ...
        rezultat = ...
        return(rezultat)
}
```

     Po potrebi uporabimo pogojne stavke in/ali zanke:

```
if (pogoj1) {
        naredi ...
        }
        else if (pogoj2) {
        naredi ...
        }
        else {
        naredi ...
        }
for (i in n:m) {
        naredi ...
}
while (pogoj) {
        naredi ...
}
```

2. V grafičnem vmesniku programskega paketa R odprite novo skriptno datoteko.

3. Eden izmed parametrov, ki vplivajo na oceno kvalitete tehnične podpore, ki jo nudi proizvajalec računalnikov, je hitrost odziva osebja v tehnični podpori na telefonske klice. Po preteklem sledenju hitrosti odzivov na telefonske klice smo ugotovili, da se osebje v tehnični podpori na 70 % klicev odzove po dveh ali manj (1-2) zvonjenjih, na 25 % klicev odgovori po treh ali štirih (3-4) zvonjenjih, za preostale klice pa rabi pet ali več (≥ 5) zvonjenj.

   Na tehnično podporo bomo poklicali 10-krat. Predpostavimo, da so klici paroma neodvisni.

   a.) Kakšna je verjetnost, da bo osem klicev odgovorjenih po dveh ali manj zvonjenjih, en klic po treh ali štirih zvonjenjih in en klic šele po pet ali več zvonjenjih?
   b.) Kakšna je verjetnost, da bo vseh 10 klicev odgovorjenih po štirih ali manj zvonjenjih?
   c.) Kakšno je pričakovano število klicev, na katere se bo tehnično osebje odzvalo po štirih ali manj zvonjenjih?
   d.) Določite pogojno porazdelitev števila klicev, ki zahtevajo pet ali več zvonjenj, če veste, da osem klicev zahteva največ dve zvonjenji.
   e.) Kakšno je pogojno pričakovano število klicev, na katere se bo tehnično osebje odzvalo po pet ali več zvonjenjih, če veste, da osem klicev zahteva največ dve zvonjenji?
   f.) Ali sta število klicev, ki zahtevajo dve ali manj zvonjenji, in število klicev, ki zahtevajo pet ali več zvonjenj, neodvisni slučajni spremenljivki?

   R: a.) 0.06485401 b.) 0.5987369 c.) 9.5 d.) - e.) 0.3333333 f.) odvisni

4. Slučajni vektor (X, Y) je porazdeljen po zakonu za dvorazsežno normalno porazdelitev s parametri $\mu_X = 0$, $\sigma_X = 0.5$, $\mu_Y = 0.5$, $\sigma_Y = 2$ in $\rho = 0.5$.

   a.) V programskem okolju R definirajte funkcijo `bivariate(x,y)` gostote verjetnosti $f_{XY}(x, y)$ dvorazsežne normalne porazdelitve.

b.) Narišite graf funkcije gostote verjetnosti $f_{XY}(x, y)$ dvorazsežne normalne porazdelitve.

Pri risanju si pomagajte z R ukazoma `outter(`x`,`y`,ime-funkcije)`, ki izbrano funkcijo uporabi na vseh točkah kartezičnega produkta seznamov x in y, in `persp(x,y,z,...)`, ki nariše ploskev v prostoru, tj. graf funkcije $z(x, y)$.

c.) Izračunajte $P(X = 0, Y = 0)$ in $P(0 < X < 1, 0 < Y < 2)$.

d.) Določite robno porazdelitev $f_X(x)$ in narišite njen graf.

R: a.) - b.) - c.) 0, 0.1995153 d.) -

5. Med nalogami k poglavju 5.3 v knjigi *Applied Statistics and Probability for Engineers* izberite poljubno zvezno dvorazsežno porazdelitev.

   - V programskem okolju R definirajte funkcijo gostote verjetnosti `dmyfun(x,y)` izbrane porazdelitve.
   - Narišite graf funkcije gostote verjetnosti.
   - Funkcijo `dmyfun(x,y)` uporabite za izračun verjetnosti.

6. Vsebino skriptne datoteke shranite pod imenom ‘vaja7.r’.