Číselné charakteristiky diskrétneho rozdelenia

Nech náhodné veličiny X, Y nadobúdajú hodnoty

x = [-2 -1 0 1 2 3 4]

y = [3 4 5 6 7 8]

Použijeme Matlab na to, aby nám vyrobil pravdepodobnostnú funkciu vektora (X,Y). Z rozmerov vektorov x, y vidno, že tabuľka, ktorá by mala určovať pravdepodobnostnú funkciu, musí mať rozmery 7 x 6.

```
pf = rand(7, 6) ;
```

Aby pf mohla naozaj reprezentovať pravdepodobnostnú funkciu, musí byť súčet všetkých jej prvkov 1 (prečo?). Použijeme príkaz sum, ktorý v matici vykoná súčty prvkov v stĺpcoch. Pri vektoroch (stĺpcových aj riadkových) sčíta všetky prvky.

```
sf = sum(sum(pf))
pf = pf/sf ;
```

Overte, že súčet prvkov v sf je 1.

Poznámka: Maticu pf použijeme na ilustráciu výpočtov jednotlivých číselných charakteristík. Ide o náhodne vytvorené čísla, preto vzťah medzi X, Y určený maticou pf je iba formálny a nemusí zodpovedať žiadnej reálnej situácii.

Postupne budeme počítať číselné charakteristiky veličín X, Y – strednú hodnotu, varianciu, kovarianciu a korelačný koeficient.

Na tento výpočet nie je možné použiť matlabovské príkazy var, cov, corrcoef, pretože tie zodpovedajú situácii, ktorou sme sa zaoberali v 2. kapitole (v čom bola tá situácia iná?).

Stredná hodnota:

Stredná hodnota je daná vzorcom $E(X) = \Sigma_i x_i f(x_i)$.
Na jej výpočet budeme potrebovať vektor pravdepodobnostného rozdelenia fx(x) premennej X samostatne (bez ohľadu na jej vzťah ku Y). Ten získame z tabuľky f súčtom hodnôt v jej riadkoch:

```
fx = sum(f ')
```

Výpočet strednej hodnoty (možno si ho predstaviť ako skalárny súčin vektorov x a fx) v matlabe počítame ako súčin matíc (prvý vektor musí byť riadkový, druhý stĺpcový):

```
Ex=x*fx'
```

Podobne vypočítajte EY.

Variancia:

Vzorec pre varianciu je $var(X) = E(x^2) - (E(x))^2$, kde $E(x^2) = \Sigma_i x_i^2 f(x_i)$.
V matlabe:

varx = (x.^2)*fx' - Ex^2 .

Podobne vypočítajte vary.

Kovariancia:

Kovarianciu počítame podľa vzorca $cov(X,Y) = E(X.Y) - E(X).E(Y)$, kde

$E(X.Y) = \Sigma_{ij} x_i y_j f_{ij}$ = [1]

$= x_1y_1 f_{11} + x_1y_2 f_{12} + \dots + x_1y_n f_{1n} +$ $+ x_2y_1 f_{21} + x_2y_2 f_{22} + \dots + x_2y_n f_{2n} +$ $\vdots$ $+ x_my_1 f_{m1} + x_my_2 f_{m2} + \dots + x_my_n f_{mn} =$	$= x_1 (f_{11} y_1 + f_{12} y_2 + \dots + f_{1n} y_n) +$ $x_2 (f_{21} y_1 + f_{22} y_2 + \dots + f_{2n} y_n) +$ $\vdots$ $x_m (f_{m1} y_1 + f_{m2} y_2 + \dots + f_{mn} y_n) =$

$= [x_1, x_2, \dots , x_m] * [f_{11} y_1 + f_{12} y_2 + \dots + f_{1n} y_n ;$ $f_{21} y_1 + f_{22} y_2 + \dots + f_{2n} y_n ;$ $\vdots$ $f_{m1} y_1 + f_{m2} y_2 + \dots + f_{mn} y_n] =$	= x * (f*y‘) = x*f*y‘

Vďaka týmto zdĺhavým úpravám dostávame veľmi stručné vyjadrenie matlabe (využívame to, čo sme už vypočítali v predošlých bodoch:

covxy = x*f*y‘ - EX*EY

Korelačný koeficient:

Na výpočet korelačného koeficientu v matlabe stačí len dosadiť potrebné hodnoty (získané v predošlých bodoch) do vzorca:

kkxy = covxy / (varx*vary)^0.5 alebo
kkxy = covxy / sqrt(varx*vary)

[1] Čísla f_{ij} sú prvky matice pf.