Číselné charakteristiky diskrétneho rozdelenia

Cieľom tohto textu je ukázať prehľadným spôsobom postup, ako v matlabe počítať jednotlivé číselné charakteristiky diskrétneho rozdelenia a naznačiť, ako nám matlab môže „vymyslieť" podobné príklady. Neuvádzame číselné výsledky, čo však neznamená, že nie sú dôležité.

Nech náhodné veličiny X, Y nadobúdajú hodnoty

```
>> x=-5:2:21;  y=3:40;
```

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 12 x 13.

```
>> f=min(abs(x'*(1./y)),abs((1./x)'*y));
```

Aby f mohla naozaj reprezentovať pravdepodobnostnú funkciu, musia byť všetky jej prvky nezáporné a súčet všetkých jej prvkov musí byť 1 (prečo?).

```
>> f=f/sum(sum(f));
```

Postupne budeme počítať číselné charakteristiky veličín X, Y a (X,Y) – strednú hodnotu, varianciu, kovarianciu a korelačný koeficient.
Na tento výpočet **nie je možné** použiť matlabovské príkazy var, cov, corrcoef.

Stredná hodnota:

Stredná hodnota je daná vzorcom $E(x) = \Sigma_i x_i f(x_i)$. Na výpočet budeme potrebovať vektor pravdepodobnostného rozdelenia f(x). Ten získame z tabuľky f súčtom hodnôt v jej riadkoch:

```
fx = sum(f, 2) ;
```

Výpočet strednej hodnoty v matlabe počítame ako skalárny súčin:

```
EX = x*fx ;
```

Podobne počítame EY:

```
>> fy = sum(f, 1)' ;
>> EY = y*fy ;
```

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 vypočítame vary:

```
>> vary = (y.^2)*fy - EY^2
```

Kovariancia:

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

$$E(X.Y) = \Sigma_{ij}\ f_{ij}\, x_i y_j \quad =$$

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

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

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

```
covxy = x*f*y' - EX*EY
```

Korelačný koeficient:

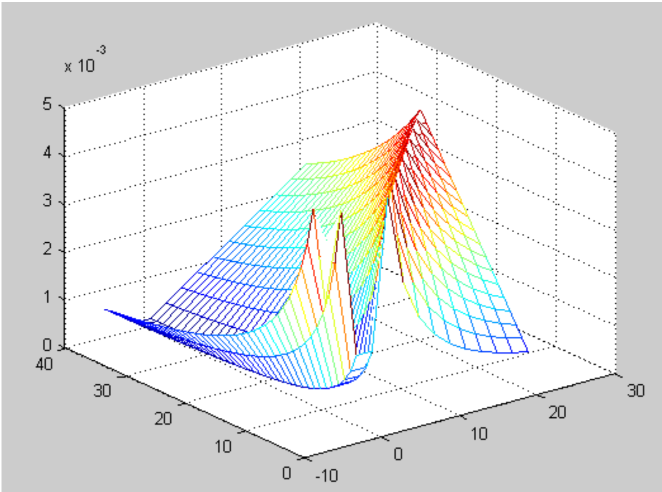
Na výpočet korelačného koeficientu v matlabe stačí len dosadiť potrebné hodnoty (získané v predošlých bodoch) do vzorca:

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

Kreslenie:

Pri daných údajoch môže byť zaujímavé nakresliť v rovine jednotlivé body [x,y] a im prislúchajúcu „pravdepodobnostnú váhu“ f(x,y):

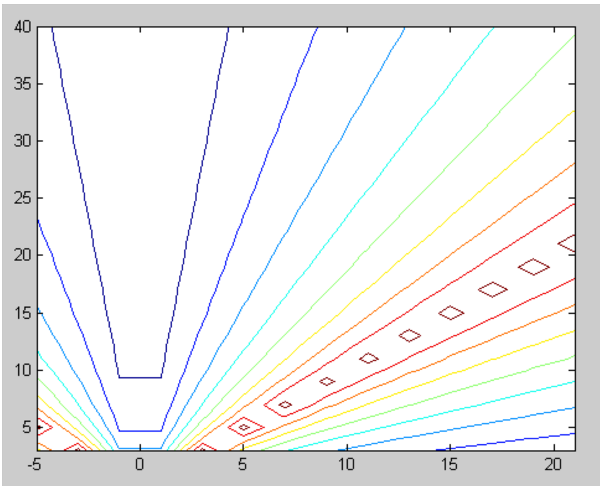
```
>> mesh(x,y,f')
```



Upozornenie: príkaz *mesh* aj nasledujúci *contour* potrebuje vstup premenných x, y v opačnom poradí alebo transponovanú maticu f.

Skúsme sa na daný obrázok pozrieť zhora – vzdáme sa tretej dimenzie:

```
>> contour(x,y,f'), hold on
```



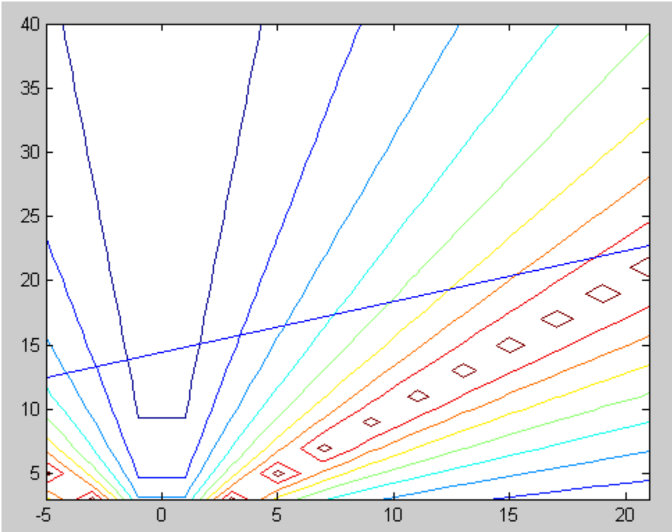
(Dobre si všimnite vzťah medzi trojrozmerným nákresom a pohľadom zhora.)

Pristúpme teraz k nákresu regresnej priamky, aj keď korelačný koeficient nie je zrovna najväčší. Vzorec sa nelíši od toho, ktorý sme používali doteraz:

y = EY + (covxy/varx)*(x-EX)

Na kreslenie nám stačia dva krajné body:

```
>> hold on, xk=[-5 21]; yk=EY+(covxy/varx)*(xk-EX); plot(xk,yk)
```



<u>Úloha:</u> Navrhnite zaujímavejšiu maticu f, prípadne aj vektory x, y, aby boli výsledné obrázky výraznejšie.