Matematika B4
XII. gyakorlat

2005. november 30., december 2.

1. Lineáris regresszió

A gyakorlatban ritkábban tudjuk az együttes sűrűségfüggvényt becsülni, könnyebb a kovarianciát, várható értéket és a szórást kiszámolni, és felmerül a kérdés, hogy ezek ismeretében mit tudunk mondani. Itt jön be a **lineáris regresszió**. Y-t közelítjük X-szel egy egyenes segítségével:

$$y - E(Y) = \frac{\operatorname{cov}(X,Y)}{\sigma^2(X)}(x - E(X)).$$

Ilyenkor az $E\left((Y - k(X))^2\right)$ kifejezésben olyan k függvényeket keresünk, amelyek linárisak.
Ez egy kicsit másképp:

$$\frac{y - E(Y)}{\sigma(Y)} = R(X,Y)\frac{x - E(X)}{\sigma(X)}.$$

ahol $R(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}$ az X és a Y korrelációs hányadosa. Ha $|R(X,Y)| = 1$, akkor a két valószínűségi változó közt lineáris függés van, ha 0, akkor nincs függés.
A kovariancia definíciója:

$$cov(X,Y) = E(XY) - E(X)E(Y)$$

A kovariancia mátrixban az $i.$ oszlop $j.$ sorában az $i.$ és a $j.$ valószínűségi változó kovarianciája áll, vagyis ez egy szimmetrikus mátrix, melynek főátlójában pedig épp a szórásnégyzetek helyezkednek el, aza két valváltozóra ez így néz ki:

$$\begin{pmatrix} \sigma^2(X) & cov(X,Y) \\ cov(X,Y) & \sigma^2(Y) \end{pmatrix}$$

Feladatok:

1. Egy kétdimenziós valószínűségi változó sűrűségfüggvénye $\frac{1}{6}xy$ $(0 < x < 2, x < y < 2x)$. Milyen $k(y)$ függvénnyel érdemes a második koordinátából az elsőt tippelni, ha az a célunk, hogy a tippelésnél elkövetett hiba négyzetének átlagos értéke sok kísérlet esetén minél kisebb legyen,
   a) ha feltesszük, hogy $k(y)$ lineáris,
   b) ha $k(y)$ tetszőleges valós lehet?
2. Ugyanaz a probléma, mint az előző feladatban, de most a tippelő függvényünk csak $c\sqrt{y}$ alakú lehet?
   Segitseg: itt a
   $m(c) = E((X - c\sqrt{Y})^2) = E(X^2) + c^2E(Y) - 2cE(X\sqrt{Y}) = E(X^2) + E(Y)(c^2 - \frac{E(X\sqrt{Y})}{E(Y)})^2 - \frac{E(X\sqrt{Y})^2}{E(Y)}$
   függvényt kell minimalizálni, ahol c változhat.

3. X és Y együttes sűrűségfüggvénye $h(x, y) = 60xy^2$, ha $0 \leq x \leq 1$, $0 \leq y \leq 1 - x$. Határozzuk meg a kovarianciájukat!
   Tegyük fel, hogy a második koordinátát tudjuk megfigyelni és az elsőt ezen megfigyelt adattól függően becsüljük az $x = \frac{2}{3}(1 - y)$ képlet alapján. Van-e ennél jobb módszer, ha négyzetes eltérés hibáját akarjuk minimalizálni?

4. Az (X, Y) kovarianciamátrixa $\binom{8 \quad 4}{4 \quad 2}$ Van-e linearis kapcsolat X és Y között?

5. Statisztikai adatok alapján annak a valószínűsége, hogy ikerszületéskor mindkét gyerek fiú, 0.32, annak a valószínűssge, hogy mindkét gyermek lány, 0.28. Annak a valószínűsége, hogy az első iker fiú és a masodik lány ugyanannyi, mint fordítva. Jelölje X illetve Y az első, illetve a masodik gyerek nemét, legyen a felvett értékük fiú esten 1, lány esetén 0. Számítsuk ki az X és a Y korrelációs együtthatóját! Hogyan tippelnénk Y ismeretében X-re lineáris függvénnyel, ha a tippeles átlagos hibáját akarjuk minimalizálni?

6. Legyen (X, Y) egyenletes eloszlású a $(0, 0)$, $(1, 0)$, $(0, 2)$ pontok által meghatározott háromszögön. Számítsuk ki Y-nak X-ra vonatkozó regressziós függvényét!

7. Legyenek X és Y két véges szórasú valószínűségi változó. Legyen $A = X + Y$, $B = X - Y$ Bizonyítsa be, hogyha tudjuk, hogy B-nek A-ra vonatkozó regressziós egyenese konstans, akkor a X es Y szórasa egyenlő!

8. Magyarországon a 18 év feletti férfiak testmagasságának átlagos érteke 178 cm, szórása 10 cm. nőknel ugyanezek az adatok:166 cm, és 8 cm. Focimeccseken a drukkerek 10%-a nő, a többiek férfiak. Mindkét nem testmagasságának eloszlását normalis eloszlásúnak véve:

   a) Mi annak a valószínűsége, hogy egy 170 cm-nel alacsonyabb szurkoló nő?
   b) Adja meg x függvényében annak a valószínűségét, hogy egy x cm magas drukker férfi!
   c) Hogyan tippeljunk a szurkolók testmagasságából a nemükre, ha a celunk az, hogy a lehető legnagyobb valószínűséggel helyesen tippeljünk?