12. gyakorlat

Matematika A4
Vetier András kurzusa

2009. május 8.

1. Egy ampermérővel az alábbi áramerősségeket mértük: 15,2; 15,7; 14,6; 16;8; 13,9 A. Adjon olyan konfidencia intervallumot, amely

   a) 0.8

   b) 0.9

   c) 0.99

   valószínűséggel az áramerősség várható értékét magában foglalja, ha az áramerősség szórása 2 A?

1. Általános regresszió

Adott (X, Y) kétdimenziós valószínűségi változó az együttes eloszlásával (például folytonos esetben sűrűségfüggvényével). X-et megfigyelve Y-t szeretnénk közelíteni egy $k(X)$ alakú tippelő függvénnyel. A közelítés azt jelenti, hogy az elkövetett $(Y - k(X))^2$ négyzetes hiba átlagát szeretnénk minimalizálni. Pontosabban azt a k függvényt keressük, amire

$$E\left((Y - k(X))^2\right)$$

minimális. Az órán tanult tétel kimondja, hogy ebben az esetben a megoldás a feltételes várható érték:

$$k(x) = E(Y \mid X = x).$$

Ha pedig az elkövetett abszolút hibát, azaz $E(|Y - k(X)|)$ -t szeretnénk minimalizálni, akkor a lehető legjobb tippelés az Y feltételes mediánja. A feltételes sűrűségfüggvény

$$f_{2|1}(y|x) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,v)dv}$$

amelynek az első esetben a várható értékét kell kiszámolni, a második esetben a feltételes eloszlásfüggvényt ($F_{2|1}(y|x) = \int_{-\infty}^{y} f_{2|1}(v|x)dv$) kell $\frac{1}{2}$-del egyenlővé tenni és belőle y-t mint x függvényét kifejezni.

Feladatok:

2. A Duna holnaputáni budapesti vízállását akarjuk becsülni a mai bécsi vízállásból. Bár a két vízállás közt szoros kapcsolat van, azért pontosan nem lehet megmondani a vízállást, mindkettőt egy-egy valószínűségi változó írja le. Tegyük fel, hogy mindkét vízállást egy 0 és 1 közti számmal tudunk jellemezni, melynek legyen az együttes eloszlásfüggvénye
   $f(x,y) = \frac{6}{5}(x + (y-1)^2)$ ahol $0 < x < 1$ és $0 < y < 1$.

   a) Határozzuk meg a budapesti vízállás eloszlását a bécsi ismeretében, azaz mi a feltételes sűrűségfüggvény?

b) Mi annak a valószínűsége, hogy Budapesten alacsonynak nevezhető (azaz 0 és $1/2$ közé esik) a vízállás, ha Bécsben x volt? Mennyi ez $x = 1/3$-ra?

c) Ha már ismerjük a bécsi vízállást, mire tippelünk a budapestire, ha a lehető legkisebb négyzetes hibát akarjuk elkövetni?

d) És ha az abszolút hibát akarjuk minimalizálni?

3. $X = RND1, Y = RND_1 \cdot RND_2$ eloszás esetén láttuk, hogy az együttes sűrűségfüggvény $f(x, y) = 1/x$, ha $0 < y < x < 1$ háromszögön vagyunk. Mi a regressziós görbe, ha X-et ismerve Y-t tippelve az abszolút hibát, illetve ha a négyzetes hibát szeretnénk minimalizálni?

4. $X = RND_1 * RND_2, Y = RND_1$ eloszás esetén láttuk, hogy az együttes sűrűségfüggvény $f(x, y) = 1/y$, ha $0 < x < y < 1$ háromszögön vagyunk. Mi a regressziós görbe, ha X-et ismerve Y-t tippelve az abszolút hibát, illetve ha a négyzetes hibát szeretnénk minimalizálni?

5. Az egységkörlapon választunk egyenletes eloszlás szerint egy (X, Y) pontot. Az X koordinata ismeretében hogyan kozelítené $|Y|$-t, feltéve, hogy a hiba abszolútértéknégyzetét szertné minimalizálni?

6. Többpártrendszer esetén az egyes pártokra leadott szavazatok százalékos aránya valószínűségi változó. A Zöldek az összes szavazatok X, a Demokraták az összes szavazatok Y hányadát kapják, együttes eloszlásuk $h(x, y) = 24xy$, ha $0 < x, 0 < y, x + y < 1$.
Ha a Demokraták az összes szavazatok 40%-át kaptak, mire tippelünk, mennyit kaptak a zöldek?

2. Lineáris regresszió

A gyakorlatban gyakran nem tudjuk az együttes sűrűségfüggvényt meghatározni. Ilyenkor könnyebb lehet a várható értéket, a szórást, a kovarianciát kiszámolni. Ezek segítségével már meg tudjuk mondani, hogy melyik az a lineáris függvény, amivel tippelve a hiba négyzetének a várható értéke a legkisebb.

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

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 együttható. Ha $|R(X, Y)| = 1$, akkor a két valószínűségi változó közt lineáris függés van, ha 0, akkor nincs függés (azaz a két valószínűségi változó korrelálatlan, erről még lesz szó).
A kovariancia definíciója:

$$\operatorname{cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

Kibontva a zárójeleket:

$$\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y)$$

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, azaz két valószínűségi változóra ez így néz ki:

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

Feladatok:

7. 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?

8. * Ugyanaz a probléma, mint az előző feladatban, de most a tippelő függvényünk csak $c\sqrt{y}$ alakú lehet?
   Segitseg: itt a
   $m(c) = E((X - c\sqrt{Y})^2) = E(X^2) + c^2E(Y) - 2cE(X\sqrt{Y}) = E(X^2) + E(Y)\left(c - \frac{E(X\sqrt{Y})}{E(Y)}\right)^2 - \frac{E(X\sqrt{Y})^2}{E(Y)}$
   függvényt kell minimalizálni, ahol c változhat.

9. 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?

10. Az (X, Y) kovarianciamátrixa $\begin{pmatrix} 8 & 4 \\ 4 & 2 \end{pmatrix}$ Van-e lineáris kapcsolat X és Y között?

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

12. 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!

13. * Legyenek X és Y két véges szórású valószínűségi változó. Legyen $A = X + Y$, $B = X - Y$. Bizonyítsa be, hogyha tudjuk, hogy B-nek A-ra vonatkozó regressziós egyenese konstans, akkor X és Y szórása egyenlő!

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

    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?