Valószínűségszámítás és statisztika

10. gyakorlat

2014. 11. 27.

Elmélet

[X és Y kovarianciája:] $\text{Cov}(X,Y) = E[(X - EX)(Y - EY)]$. Köv.: $\text{Cov}(X,Y) = E(XY) - EX \cdot EY$.
Elnevezés: ha $\text{Cov}(X,Y) = 0$, akkor azt mondjuk, hogy X és Y *korrelálatlanok*.
Ha X és Y függetlenek egymástól, akkor korrelálatlanok is.
Ha X és Y korrelálatlanok, akkor ebből *nem* következik, hogy függetlenek is!

[A kovariancia tulajdonságai:]
Legyenek $X, Y, X_1, ..., X_n$ valószínűségi változók, $a, b \in \mathbb{R}$. Ekkor

$\text{Cov}(X,X) = D^2X$	$\text{Cov}(X,Y) = \text{Cov}(Y,X)$	$D^2(X+Y) = D^2X + D^2Y + 2\text{Cov}(X,Y)$
$\text{Cov}(aX,bY) = ab\text{Cov}(X,Y)$	$\text{Cov}(X,a) = 0$	$D^2(\sum\limits_{i=1}^n X_i) = \sum\limits_{i=1}^n D^2X_i + 2\sum\limits_{1\le i<j\le n}\text{Cov}(X_i,X_j)$

[Korreláció:] X és Y korrelációja: $R(X,Y) = \frac{\text{Cov}(X,Y)}{DX \cdot DY}$.
A korreláció két valószínűségi változó lineáris kapcsolatát méri:

$R > 0 \Rightarrow$ pozitív a kapcsolat		$R^2 \sim 1 \Rightarrow$ erős a kapcsolat
$R < 0 \Rightarrow$ negatív a kapcsolat		$R^2 \sim 0.5 \Rightarrow$ közepes a kapcsolat
		$R^2 \sim 0 \Rightarrow$ gyenge a kapcsolat

[Feltételes várható érték:] Legyenek X és Y valószínűségi változók. Y-nak X-re vonatkozó feltételes várható értéke $-$ E$(Y|X)$ $-$ precíz definiálására nem vállalkozunk, úgy gondoljunk rá, mint egy valószínűségi változóra; és ha X egy adott értéket vesz fel $-$ E$(Y|X=x)$ $-$, akkor mint konkrét számra, melyet diszkrét esetben a feltételes valószínűség segítségével számíthatunk ki.

[Nadarajah-módszer:] a feltételes várható érték becslése $(X_1,Y_1), ..., (X_n,Y_n)$ minta alapján: $E(Y|X=x) \approx \frac{\sum\limits_i Y_i k\left(\frac{x-X_i}{h_n}\right)}{\sum\limits_i k\left(\frac{x-X_i}{h_n}\right)}$.

Legyen g mérhető függvény. Ekkor $E[g(X)|X] = g(X)$. Ha X, Y függetlenek $\Rightarrow E(Y|X) = EY$

[Feladat:] Y val. változót szeretnénk közelíteni X val. változó tetszőleges függvénye segítségével:

$$\text{E}[Y - f(X)]^2 \longrightarrow \min_f \qquad \rightsquigarrow \text{Megoldása: } f_{opt} = \text{E}(Y|X)$$

[Feladat:] Y val. változót szeretnénk közelíteni X val. változó lineáris függvénye segítségével:

$$\text{E}[Y - (aX+b)]^2 \longrightarrow \min_{a,b} \quad \rightsquigarrow \text{Megoldása:} \quad a_{opt} = \frac{Cov(X,Y)}{D^2(X)}; \quad b_{opt} = EY - a_{opt}EX$$

[Feladat:] (*lineáris modell*): Adottak $(x_1,y_1), \ldots, (x_n,y_n)$ pontok, ezekre szeretnénk egyenest illeszteni (neve: *regressziós egyenes*) a legkisebb négyzetek módszerével.
A modell: $Y_i = aX_i + b + \varepsilon_i$, ahol E$\varepsilon_i = 0$ és D$^2\varepsilon_i = \sigma^2 < \infty$ $(i = 1, \ldots, n)$
Megoldás: $\hat{a} = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2}$, $\hat{b} = \overline{y} - \hat{a}\overline{x}$
Reziduumok: $\hat{\varepsilon}_i = y_i - \hat{a}x_i - \hat{b}$ (i=1,..., n)
Reziduális négyzetösszeg: RNÖ=$\sum\hat{\varepsilon}_i^2 = \sum(y_i - \overline{y})^2 - \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2}$; $\hat{\sigma}^2 = \frac{\text{RNÖ}}{n-2}$

[Markov-egyenlőtlenség:] Legyen $g : \mathbb{R} \to \mathbb{R}$ monoton növő függvény, $X \geq 0$ val.változó, $\varepsilon > 0$ tetsz.
Ekkor P$(X \geq \varepsilon) \leq \frac{E[g(X)]}{g(\varepsilon)}$.
Spec., ha $g(x) = x \Rightarrow$ P$(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$

[Csebisev-egyenlőtlenség:] P$(|X - EX| \geq \varepsilon) \leq \frac{D^2(X)}{\varepsilon^2}$.

Legyen $X_1, X_2, \ldots$ valószínűségi változók sorozata.
[1 valószínűségű konvergencia:] $X_n \stackrel{n\to\infty}{\longrightarrow} X$ *1 valószínűséggel*, ha $P(\{\omega : X_n(\omega) \stackrel{n\to\infty}{\to} X(\omega)\}) = 1$.
[Gyenge konvergencia:] $X_n \stackrel{n\to\infty}{\longrightarrow} X$ *gyengén*, ha az eloszlásfüggvényeikre $F_n(x) \stackrel{n\to\infty}{\longrightarrow} F(x)$ F minden folytonossági pontjában.
[Nagy számok törvénye (NSZT):] Legyenek $X_1, X_2, \ldots$ i.i.d. val. változók, E$X_1 = m < \infty$.
Ekkor $\frac{X_1+\ldots+X_n}{n} \stackrel{n\to\infty}{\to} m$ 1 valószínűséggel.
[Centrális határeloszlás tétel (CHT):] Legyenek $X_1, X_2, \ldots$ i.i.d. val. változók, E$X_1 = m$, $D^2(X_1) = \sigma^2 < \infty$.
Ekkor $\frac{X_1+\ldots+X_n-nm}{\sqrt{n}\sigma} \stackrel{n\to\infty}{\to}$ N(0,1) gyengén, azaz $P\left(\frac{X_1+\ldots+X_n-nm}{\sqrt{n}\sigma} < x\right) \stackrel{n\to\infty}{\to} \Phi(x)$.

Személyes adatok: Név: Bondici László E-mail: bondici@cs.elte.hu Honlap: www.cs.elte.hu/$\sim$bondici

Feladatok

1. (9-es feladatsor 5/b feladata) Határozzuk meg az ismeretlen paraméter maximum likelihood (ML) becslését, ha a minta $E(a, 1)$ (a az ismeretlen paraméter) eloszlású.

2. Az X és Y valószínűségi változók együttes eloszlását a következő táblázat mutatja:

Y \ X	0	1	2
0	3/27	3/27	4/27
1	2/27	2/27	4/27
2	1/27	1/27	7/27

   Határozzuk meg X és Y eloszlását, várható értékét, szórásnégyzetét és korrelációjukat.

3. Legyen X és Y független, azonos eloszlású. Tegyük fel azt is, hogy véges szórásúak. Számítsuk ki $R(X, aX + bY)$.

4. Legyen X a hatosok száma 6 kockadobásból, Y pedig $X + Z$, ahol Z további 6 kockadobásból a hatosok száma. Mi lesz Y legkisebb négyzetes közelítése X segítségével, ha

   a) X lineáris függvényével közelítünk;

   b) X tetszőleges függvényével közelítünk?

5. Legyenek adottak a következő (x, y) párok:

x	0	1	6	5	3
y	4	3	0	1	2

   a) Határozzuk meg és ábrázoljuk is az $aX + b$ alakú regessziós egyenest.

   b) Számoljuk ki a reziduálisokat és becsüljük meg a hiba-szórásnégyzetet.

   c) Adjuk meg az $E(Y|X = 2.5)$ feltételes várható érték közelítését a Nadarajah-féle módszerrel, ha a magfüggvény $k(x) = 1/4$, ha $-2 < x < 2$, és 0 különben, illetve $h = 1$.

6. Legalább hány embert kell megkérdezni egy közvélemény-kutatásnál, ha egy adott párt támogatottságát legalább 95%-os valószínűséggel 0.01–nál kisebb eltéréssel szeretnénk megbecsülni? Számoljunk egyrészt a Csebisev-egyenlőtlenséggel, másrészt a normális eloszlással.

7. Legyen X_n n-paraméterű Poisson-eloszlású. Mihez tart $P(X_n < n - \sqrt{n})$, ha $n \to \infty$?

8. Legyen X_n (n, p_n)-paraméterű binomiális eloszlású. Mihez tart

   a) $P(X_n < np - \sqrt{n})$, ha $p_n = p$ és $n \to \infty$?

   b) $P(X_n < 2)$, ha $p_n = \frac{1}{n}$ és $n \to \infty$?

Házi feladatok

1. a) Egy 52 lapos francia kártyacsomagból húzunk 2 lapot visszatevés nélkül. Legyen X a kőrök, Y pedig az ászok száma. Adjuk meg X és Y korrelációs együtthatóját. Függetlenek-e ezek a változók?

   b) Dömötör rulettezik a kaszinóban. Minden egyes körben 10 petákot tesz pirosra. 100 játék után 300 peták a vesztesége. Jogos-e a gyanúja, hogy svindlizik a krupié? (A rulettkeréken összesen 37 mező van 0-tól 36-ig számozva. Ezek közül egy (a 0 jelű) zöld, a fennmaradó 36-ból pedig 18 piros és 18 fekete.)

2. Véletlenszerűen választunk egy szót az alábbi mondatból: EGY TEVE LEGEL A KERTBEN. A feladatunk az, hogy kitaláljuk a szó hosszát úgy, hogy a tényleges és a tippelt szóhossz közötti eltérés négyzetének várható értéke minimális legyen.

   a) Mit tippelünk, ha semmi információ nem áll rendelkezésünkre?

   b) Hogyan tippelünk, ha valaki megsúgta a szóban szereplő „e” betűk számát?

   c) Hogyan tippeljünk, ha az „e” betűk számának lineáris függvényét használhatjuk?

3. a) Legyen X_n n-paraméterű Poisson-eloszlású. Mihez tart $P(X_n < n)$, ha $n \to \infty$?

   b) Legyen X_n (n, p_n)-paraméterű binomiális eloszlású. Mihez tart $\frac{X_n}{n}$, ha $p_n = p$ és $n \to \infty$?