A sztochasztika alapjai
MBNXK262

7. előadás: Példák, lineáris regresszió, nevezetes eloszlások

Kevei Péter

2022/23 tavasz

Kovariancia, korreláció

Definíció

Az ξ és η véletlen változók *kovarianciája*

$$\mathbf{Cov}(\xi,\eta) = \mathbf{E}[(\xi - \mathbf{E}(\xi))(\eta - \mathbf{E}(\eta))],$$

korrelációja

$$\rho(\xi,\eta) = \frac{\mathbf{Cov}(\xi,\eta)}{\mathbf{D}(\xi)\mathbf{D}(\eta)}.$$

Kovariancia

Példa

Egy szabályos dobókockával n-szer dobunk. Jelölje ξ a hatosok, η egyesek számát! Adjuk meg a várható értéket, szórást, kovarianciát, korrelációt!

Legyen $I_i = 1$, ha az i-edik dobás hatos, különben 0, $J_i = 1$, ha az i-edik dobás egyes, különben 0, $i = 1, 2, \ldots, n$. Nyilván

$$\xi = \sum_{i=1}^{n} I_i \quad \text{és} \quad \eta = \sum_{i=1}^{n} J_i.$$

Példa

3, külsőre egyforma érmével a fejdobás valószínűsége $1/4, 2/4, 3/4$. Véletlenszerűen választunk egy érmét, és azzal kétszer dobunk. Legyen η a választott érmével dobva a fej valószínűsége, ξ a dobott fejek száma. Adjuk meg az együttes eloszlást!

Lineáris regresszió

(ξ, η) véletlen vektorváltozó. Az η változót tekintem *függő* változónak, ennek az értékére szeretnék következtetni a ξ *független változó* értékéből. Vagyis ismert ξ esetén szeretném megmondani η-t. Keressük azokat az a, b valós számokat, melyre a $\eta - (a\xi + b)$ változó kicsi. A kicsiséget négyzetes hibában mérve, keressük az

$$h(a, b) = \mathbf{E}\left[(\eta - (a\xi + b))^2\right]$$

függvény minimumhelyét, azaz a legjobb a, b választást.

Lineáris regresszió

$$\mathbf{E}\left[(\eta - (a\xi + b))^2\right] =$$

Lineáris regresszió

Beszúró rendezés

Kapunk egy $U = (a_1, a_2, \ldots, a_n)$ valós számokból álló vektort. Cél: U elemeinek sorbarendezése növekvő sorrendben, azaz az elemek azt a permutációját, melyre $a'_1 \leq a'_2 \leq \ldots \leq a'_n$.

Beszúró rendezés

A már rendezett i elem közé a megfelelő helyre beszúrja az a_{i+1} elemet. Ezt megcsinálja $i = 1$-től $i = (n-1)$-ig.
Példa: $U = (4, 2, 3, 1)$.

Milyen gyors?

Legrosszabb eset:

Legjobb eset:

Átlagos eset

- Az átlagos esetet egy véletlen bemeneten a összehasonlítások számának várható értéke adja.
- Tegyük fel, hogy az U elemeinek $n!$ lehetséges permutációja egyformán valószínű. Azaz

$$\Omega = \{1, 2, \ldots, n \quad \text{permutációi}\},$$

$\mathcal{A} = 2^{\Omega}$, $\mathbf{P}(A) = |A|/n!$.

- $\xi : \Omega \to \mathbb{R}$ a szükséges összehasonlítások száma
- Keressük: $\mathbf{E}(\xi) = ?$

Átlagos eset

η_i az *i*-edik lépésben végzett összehasonlítások számát.

$$\xi = \eta_1 + \eta_2 + \ldots + \eta_{n-1}.$$

Átlagos eset

Nevezetes eloszlások: Bernoulli-eloszlás

ξ véletlen változó *p paraméterű Bernoulli-eloszlású*, $\xi \sim \text{Bernoulli}(p)$, $p \in [0, 1]$, ha lehetséges értékei $0, 1$, és $\mathbf{P}(\xi = 1) = p = 1 - \mathbf{P}(\xi = 0)$.

Nevezetes eloszlások: Binomiális eloszlás

Az ξ véletlen változó (n, p) paraméterű binomiális eloszlású, $\xi \sim \mathrm{Bin}(n, p)$, $n \in \{1, 2, \ldots\}$, $p \in [0, 1]$, ha lehetséges értékei $0, 1, \ldots, n$, és $\mathbf{P}(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$.

Binomiális eloszlás

Bernoulli és binomiális

Állítás

Legyenek $I_1, \ldots, I_n$ *független, Bernoulli*(p) *eloszlású véletlen változók. Ekkor* $\sum_{i=1}^{n} I_i$ *binomiális eloszlású* (n, p) *paraméterekkel.*

Nevezetes eloszlások: Poisson-eloszlás

Az ξ véletlen változó λ *paraméterű Poisson-eloszlású*, $\xi \sim$ Poisson(λ), $\lambda \geq 0$, ha ξ lehetséges értékei $0, 1, 2, \ldots$, és

$$\mathbf{P}(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$

Poisson és binomiális

Poisson-eloszlás a binomiális eloszlás határeloszlásaként áll elő. Legyen $p = p_n = \lambda/n$, valamely $\lambda > 0$ számra. Ha $\xi_n \sim \mathrm{Bin}(n, p_n)$, akkor némi számolás után

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}(\xi_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{\lambda^k}{k!} e^{-\lambda}.\end{aligned}$$

Ezek alapján azt látjuk, hogy akkor lép fel Poisson-eloszlás, ha egy kis valószínűségű eseményt sokszor „ismételünk”:

- téves telefonhívások száma;
- autóbalesetek száma;
- nyomdahubák száma egy oldalon;
- földrengések száma;
- csillagok száma egy adott térrészben;
- mazsolák száma a pudingban.

Az első példa Ladislaus Bortkiewicz (1868–1931) orosz közgazdásztól (statisztikus) származik: halálos lórúgások száma egy év alatt a porosz hadseregben (20 évig figyelt 14 lovas ezredet). 1898: A kis számok törvénye (Bortkiewicz-eloszlás).

Geometriai eloszlás

Az ξ véletlen változó *p paraméterű geometriai eloszlású*, $\xi \sim \text{Geo}(p)$, ha a lehetséges értékek $1, 2, \ldots$ és

$$\mathbf{P}(\xi = k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots.$$

Tipikus példa: addig ismétlünk egy kísérletet, amíg a vizsgált A esemény be nem következik.
A geometriai eloszlás a diszkrét örökifjú eloszlás, hiszen ha $k, \ell \in \mathbb{N}$, akkor

$$\begin{aligned}
\mathbf{P}(\xi > k+\ell \mid \xi > k) &= \frac{\mathbf{P}(\xi > k+\ell)}{\mathbf{P}(\xi > k)} \\
&= \frac{q^{k+\ell}}{q^k} = q^\ell = \mathbf{P}(\xi > \ell).
\end{aligned}$$

Kupongyűjtő probléma

Példa

Egy N különböző elemből álló sokaságból visszatevéses mintát veszünk. Jelölje S_r azt a véletlen számot, ahány elemet kellett húznunk, hogy kapjunk r különböző elemet. Határozzuk meg S_r várható értékét, szórását!

Kupongyűjtő probléma

Legyen $\xi_k = S_{k+1} - S_k$ változót, $S_0 = 0$.