

A sztochasztika alapjai

MBNXK262

7. előadás: Normális eloszlás; várható érték

Kevei Péter

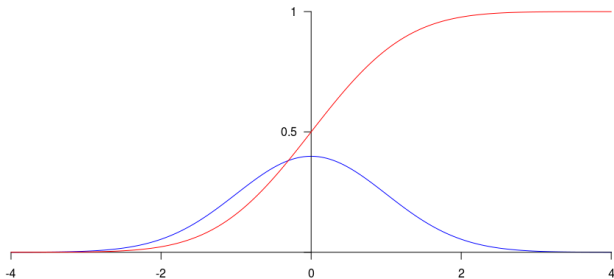
2024/25 tavasz

Normális eloszlás

ξ normális eloszlású μ és σ^2 paraméterekkel, $\xi \sim N(\mu, \sigma^2)$,
 $\mu \in \mathbb{R}, \sigma^2 > 0$, ha sűrűségfüggvénye

$$f_{\mu, \sigma}(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

$\mu = 0, \sigma^2 = 1$ standard normális



Standardizálás

Állítás

Ha $\xi \sim N(\mu, \sigma^2)$, akkor $(\xi - \mu)/\sigma \sim N(0, 1)$.

Várható érték, szórás

$\xi \sim N(\mu, \sigma^2)$ és $Z \sim N(0, 1)$, akkor

$$\begin{aligned}\mathbf{P}(|\xi - \mu| \leq \lambda\sigma) &= \mathbf{P}(|Z| \leq \lambda) \\ &= \Phi(\lambda) - \Phi(-\lambda) \\ &= 2\Phi(\lambda) - 1 \\ &= \begin{cases} 0,6827, & \lambda = 1, \\ 0,9545, & \lambda = 2, \\ 0,9973, & \lambda = 3, \\ 0,9999, & \lambda = 4. \end{cases}\end{aligned}$$

Tipikus példák

Ha egy véletlen változó sok kis véletlen hatás együttese, akkor az eloszlás általában jól közelíthető normális eloszlással. Például testtömeg, testmagasság, hossz, fejkörfogat, mérési hiba, felvételi pontszámok, IQ, tőzsdei árfolyam, éves csapadékmennyiség, . . .

Várható érték: ismételés

ξ diszkrét $x_1, x_2, \dots$ lehetséges értékekkel,

$$\mathbf{E}(\xi) = \sum_i x_i \mathbf{P}(\xi = x_i),$$

ξ folytonos $f(x)$ sűrűségfüggvénnyel,

$$\mathbf{E}(\xi) = \int_{-\infty}^{\infty} y f(y) dy,$$

Várható érték tulajdonságai

Állítás

ξ, η véletlen változók, $g : \mathbb{R} \rightarrow \mathbb{R}$, $h : \mathbb{R}^2 \rightarrow \mathbb{R}$. Ekkor

$$\mathbf{E}(g(\xi)) = \sum_{i=1}^{\infty} g(x_i) \mathbf{P}(\xi = x_i), \text{ ill. } \mathbf{E}(g(\xi)) = \int_{-\infty}^{\infty} g(y) f(y) dy,$$

ahol $f(x)$ az ξ sűrűségfüggvénye, és

$$\mathbf{E}(h(\xi, \eta)) = \sum_i \sum_j h(x_i, y_j) \mathbf{P}(\xi = x_i, \eta = y_j).$$

Várható érték tulajdonságai

Állítás

A várható érték lineáris, azaz tetszőleges $a, b \in \mathbb{R}$ állandókra

$$\mathbf{E}(a\xi + b) = a\mathbf{E}(\xi) + b.$$

Várható érték tulajdonságai

Állítás

Ha $a \leq \xi \leq b$, akkor $a \leq \mathbf{E}(\xi) \leq b$ tetszőleges $a, b \in \mathbb{R}$ számok esetén.

Várható érték tulajdonságai

Állítás

$$\mathbf{E}(\xi + \eta) = \mathbf{E}(\xi) + \mathbf{E}(\eta).$$

Várható érték tulajdonságai

Állítás

Ha $\xi_1, \xi_2, \dots, \xi_n$ véletlen változók, akkor

$$\mathbf{E} \left(\sum_{i=1}^n \xi_i \right) = \sum_{i=1}^n \mathbf{E}(\xi_i).$$

Példa

Csodaország munka törvénykönyve szerint egy cég minden munkása fizetett szabadságot kap azokon a napokon, amikor legalább az egyiküknek születésnapja van. Ezen napok kivételével azonban az év minden napján mindenkinek dolgoznia kell. Minden munkás 1 TV-készüléket készít egy nap alatt. Hány alkalmazottat vegyen fel a cégtulajdonos, ha azt akarja, hogy a gyártott TV-készülékek számának a várható értéke maximális legyen?

Csodaország

Csodaország

Kupongyűjtő probléma

Egy N különböző elemből álló sokaságból visszatevésees mintát veszünk. S_r ennyi elemet kellett húznunk, hogy kapjunk r különbözőt.

Határozzuk meg $\mathbf{E}(S_r)$ várható értékét, majd adjunk $\mathbf{E}(S_N)$ -re kezelhető aszimptotikus egyenlőséget!

Kupongyűjtő probléma

Huffman-kód

$\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ véges halmaz, a *forrásábécé*.

Kód $f : \mathcal{X} \rightarrow \{\text{véges 0-1 sorozatok}\}$

Az f -hez tartozó lehetséges kódszavak $f(x_1), f(x_2), \dots, f(x_n)$.

Huffman-kód

$\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ véges halmaz, a *forrásábécé*.

Kód $f : \mathcal{X} \rightarrow \{\text{véges 0-1 sorozatok}\}$

Az f -hez tartozó lehetséges kódszavak $f(x_1), f(x_2), \dots, f(x_n)$.

Az f kód *prefix*, ha a lehetséges kódszavak közül egyik sem folytatása a másiknak. Jelölje $x \in \mathcal{X}$ esetén $|f(x)|$ a kódszó hosszát.

Példa

Legyen $\mathcal{X} = \{a, b, c\}$, és legyen $f_1(a) = 0$, $f_1(b) = 01$, $f_1(c) = 011$. Ekkor f_1 nem prefix kód, de könnyen látható, hogy egyértelműen dekódolható. Az $f_2(a) = 01$, $f_2(b) = 00$, $f_2(c) = 1$, kód prefix.

Legyen X egy véletlen betű, és eloszlása $\mathbf{P}(X = x_k) = p_k$, $k = 1, 2, \dots, n$. Tehát p_k a x_k betű gyakorisága az adott nyelvben.

Adott f kód esetén egy hosszú szövegben az egy karakterre eső átlagos kódszóhossz:

Feltehető, hogy $p_1 \geq p_2 \geq \dots \geq p_n$. Ha az f prefix kód optimális, akkor feltehető, hogy teljesülnek a következők:

(i) Hosszabb kódhoz ritkább betűk tartoznak, azaz

$$|f(x_1)| \leq |f(x_2)| \leq \dots \leq |f(x_n)|.$$

Feltehető, hogy $p_1 \geq p_2 \geq \dots \geq p_n$. Ha az f prefix kód optimális, akkor feltehető, hogy teljesülnek a következők:

- (i) Hosszabb kódhoz ritkább betűk tartoznak, azaz

$$|f(x_1)| \leq |f(x_2)| \leq \dots \leq |f(x_n)|.$$

- (ii) A két legkisebb valószínűséghez tartozó kód hossza egyenlő.

Feltehető, hogy $p_1 \geq p_2 \geq \dots \geq p_n$. Ha az f prefix kód optimális, akkor feltehető, hogy teljesülnek a következők:

(i) Hosszabb kódhoz ritkább betűk tartoznak, azaz

$$|f(x_1)| \leq |f(x_2)| \leq \dots \leq |f(x_n)|.$$

(ii) A két legkisebb valószínűséghez tartozó kód hossza egyenlő.

(iii) $f(x_{n-1})$ és $f(x_n)$ csak az utolsó bitben térnek el.

Tétel

Tegyük fel, hogy az

$$\mathcal{X}' = \{x_1, \dots, x_{n-2}, y_{n-1}\}$$

($n - 1$) elemű forrásábécé és $p_1, \dots, p_{n-2}, p_{n-1} + p_n$ eloszlás esetén g egy optimális prefix kód. Ekkor az eredeti problémához tartozó optimális prefix kódot kapunk, ha az x_{n-1} , ill. x_n kódját úgy választjuk, hogy a $g(y_{n-1})$ kódszót kiegészítjük 0-val, ill. 1-gyel, a többi kódszót változatlanul hagyjuk.

Példa

Legyen $n = 6$, $\mathcal{X} = \{x_1, \dots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

(i) $x_5, x_6 \rightarrow x_{56}$, $p_{56} = 0.045$;

Példa

Legyen $n = 6$, $\mathcal{X} = \{x_1, \dots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

(i) $x_5, x_6 \rightarrow x_{56}$, $p_{56} = 0.045$;

(ii) x_1, x_{56} , $p_{156} = 0.177$;

Példa

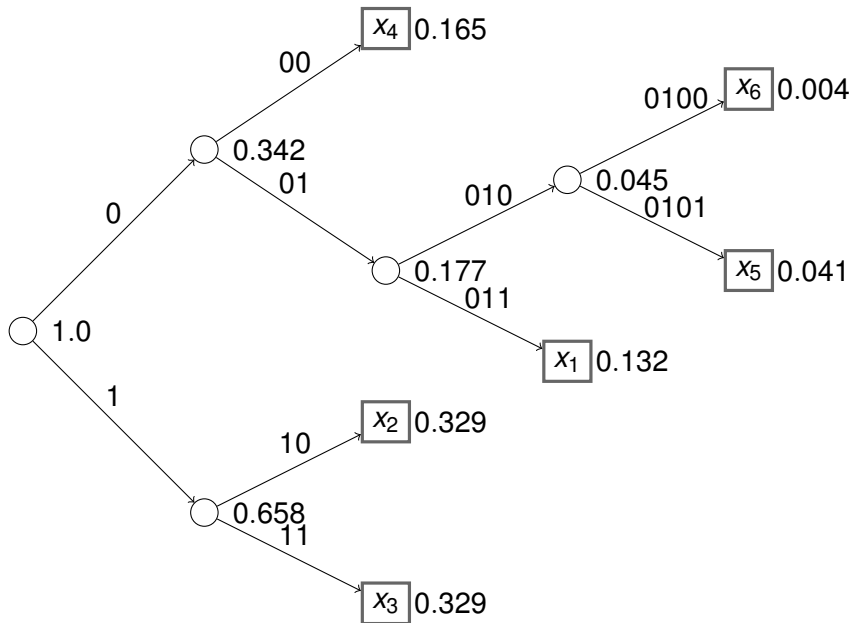
Legyen $n = 6$, $\mathcal{X} = \{x_1, \dots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

- (i) $x_5, x_6 \rightarrow x_{56}$, $p_{56} = 0.045$;
- (ii) x_1, x_{56} , $p_{156} = 0.177$;
- (iii) x_{156}, x_4 , $p_{1564} = 0.342$;

Példa

Legyen $n = 6$, $\mathcal{X} = \{x_1, \dots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

- (i) $x_5, x_6 \rightarrow x_{56}$, $p_{56} = 0.045$;
- (ii) x_1, x_{56} , $p_{156} = 0.177$;
- (iii) x_{156}, x_4 , $p_{1564} = 0.342$;
- (iv) x_2, x_3 , $p_{23} = 0.658$.



Így az optimális kód:

x_1	x_2	x_3	x_4	x_5	x_6
011	10	11	00	0101	0100
0.132	0.329	0.329	0.165	0.041	0.004

A várható érték:

$$\begin{aligned} \mathbf{E}(f(X)) &= 0.132 \cdot 3 + 0.329 \cdot 2 + 0.329 \cdot 2 \\ &\quad + 0.165 \cdot 2 + 0.041 \cdot 4 + 0.004 \cdot 4 = 2.22 \end{aligned}$$

Entrópia

Az optimális várható kódhosszra teljesül, hogy

$$\sum_{k=1}^n p_k \log_2 \frac{1}{p_k} \leq \mathbf{E}(|f(\xi)|) < \sum_{k=1}^n p_k \log_2 \frac{1}{p_k} + 1.$$

Pl.: JPEG, MP3.

Momentumok

Definíció

Az ξ véletlen változó k -adik momentuma $\mathbf{E}(\xi^k)$, és k -adik centrális momentuma $\mathbf{E}[(\xi - \mathbf{E}\xi)^k]$, $k = 1, 2, \dots$

$$\mathbf{E}(\xi^k) = \begin{cases} \sum_i x_i^k \mathbf{P}(\xi = x_i), & \text{ha } \xi \text{ diszkrét,} \\ \int_{-\infty}^{\infty} x^k f(x) dx, & \text{ha } \xi \text{ folytonos.} \end{cases}$$

Szórás

Definíció

Az ξ véletlen változó szórása $\mathbf{D}(\xi) = \sqrt{\mathbf{E}(\xi - \mathbf{E}(\xi))^2}$.

Szórás tulajdonságai

Állítás

Tetszőleges ξ véletlen változó és a, b valós számok esetén

- (i) $\mathbf{D}^2(\xi) = \mathbf{E}(\xi^2) - (\mathbf{E}(\xi))^2$;
- (ii) $\mathbf{D}^2(a\xi + b) = a^2\mathbf{D}^2(\xi)$;
- (iii) $\mathbf{D}(\xi) = 0$ akkor és csak akkor, ha $\xi = \mathbf{E}(\xi)$, azaz ξ konstans véletlen változó.

Szórás tulajdonságai

Véletlen vektorváltozók

Definíció

$\xi = (\xi_1, \dots, \xi_n) : \Omega \rightarrow \mathbb{R}^n$ véletlen vektorváltozó, ha minden komponense véletlen változó. Eloszlásfüggvénye

$$F(x_1, \dots, x_n) = \mathbf{P}(\xi_1 < x_1, \dots, \xi_n < x_n).$$

$(\xi_1, \dots, \xi_n)$ véletlen vektorváltozó diszkrét, ha értékkészlete megszámlálható.

$\xi_i, i = 1, 2, \dots, n$, eloszlása a peremeloszlás, vagy marginális eloszlás

Trinomiális eloszlás

Szabályos dobókockával n -szer dobunk. ξ a hatosok, η egyesek száma

Függetlenség

Definíció

$\xi_1, \dots, \xi_n$ függetlenek, ha minden $x_1, \dots, x_n \in \mathbb{R}$ esetén

$$\mathbf{P}(\xi_1 < x_1, \dots, \xi_n < x_n) = \mathbf{P}(\xi_1 < x_1) \dots \mathbf{P}(\xi_n < x_n).$$

Állítás

$\xi_1, \dots, \xi_n$ *diszkrét véletlen változók úgy, hogy ξ_i lehetséges értékei $x_1^{(i)}, x_2^{(i)}, \dots, i = 1, 2, \dots, n$. Ekkor $\xi_1, \dots, \xi_n$ pontosan akkor függetlenek, ha*

$$\mathbf{P}(\xi_1 = x_{i_1}^{(1)}, \dots, \xi_n = x_{i_n}^{(n)}) = \mathbf{P}(\xi_1 = x_{i_1}^{(1)}) \dots \mathbf{P}(\xi_n = x_{i_n}^{(n)})$$

tetszőleges $i_1, \dots, i_n$ indexekre.

Állítás

ξ, η független diszkrét véletlen változók.

$$\mathbf{E}(g_1(\xi)g_2(\eta)) = \mathbf{E}(g_1(\xi)) \mathbf{E}(g_2(\eta)).$$

Speciálisan, ha ξ és η függetlenek, akkor $\mathbf{E}(\xi\eta) = \mathbf{E}(\xi)\mathbf{E}(\eta)$.

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

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

Tulajdonságok

Állítás

Tetszőleges $\xi, \xi_1, \dots, \xi_n, \eta, \eta_1, \dots, \eta_m$ véletlen változók és a, b valós számok esetén igazak az alábbiak.

- (i) $\mathbf{Cov}(\xi, \xi) = \mathbf{D}^2(\xi);$
- (ii) $\mathbf{Cov}(\xi, \eta) = \mathbf{Cov}(\eta, \xi);$
- (iii) $\mathbf{Cov}(a(\xi + c), b(\eta + d)) = ab\mathbf{Cov}(\xi, \eta);$
- (iv) $\mathbf{Cov}\left(\sum_{i=1}^n \xi_i, \sum_{j=1}^m \eta_j\right) = \sum_{i=1}^n \sum_{j=1}^m \mathbf{Cov}(\xi_i, \eta_j);$
- (v) *ha ξ és η függetlenek, akkor $\mathbf{Cov}(\xi, \eta) = 0$.*

Állítás

(i) *Bunyakovszkij–Cauchy–Schwarz-egyenlőtlenség:*

$$|\mathbf{Cov}(\xi, \eta)| \leq \mathbf{D}(\xi)\mathbf{D}(\eta).$$

Innen adódik, hogy $\rho(\xi, \eta) \in [-1, 1]$;

(ii) *ha $\rho(\xi, \eta) = 1$, akkor*

$$\xi = \mathbf{E}(\xi) + \frac{\mathbf{D}(\xi)}{\mathbf{D}(\eta)}(\eta - \mathbf{E}(\eta));$$

(iii) *ha $\rho(\xi, \eta) = -1$, akkor*

$$\xi = \mathbf{E}(\xi) - \frac{\mathbf{D}(\xi)}{\mathbf{D}(\eta)}(\eta - \mathbf{E}(\eta)).$$

Teszt

Legyen $\xi \sim N(2, 9)$. Melyik változó eloszlása standard normális?

- ▶ $(\xi - 9)/4$
- ▶ $(\xi - 2)/9$
- ▶ $(\xi - 2)/3$
- ▶ $(\xi - 9)/2$

Teszt

A Panini kiadta a Matematikusok c. matricás albumját, melyben 200 híres matematikus képe szerepel. Várhatóan hány matricát kell vegyünk, mire kigyűlik a teljes album? (Segítség:

$\sqrt{200} \approx 14$, $\ln 200 \approx 5.3$, $200\pi \approx 628$)

- ▶ 200
- ▶ 1060
- ▶ 628
- ▶ 1400

Teszt

Melyik biztosan NEM normális eloszlású (pontosabban nem közelíthető jól azzal)?

- ▶ az euró árfolyama forintban 2025. május 2-án
- ▶ tűzesetek száma Szegeden májusban
- ▶ egy véletlenszerűen választott fűszál hossza a Rerrich Béla téren
- ▶ egy véletlenszerűen kiválasztott informatikus hallgató testtömege kg-ban

Teszt

A Huffman-kódban a két legkisebb valószínűségű betű kódja

- ▶ pontosan az első bitben tér el
- ▶ pontosan az utolsó bitben tér el
- ▶ nem egyforma hosszú
- ▶ megegyezik