

Valószínűségszámítás

5. előadás

Arató Miklós

2023.03.30.

Tartalomjegyzék

- 1 Momentumok
 - Szórásnégyzet
 - Momentumok
- 2 Egyéb numerikus jellemzők
 - Medián
 - Kvantilisek
 - Móduszok
- 3 Kovariancia
- 4 Markov- és Csebisev-egyenlőtlenség

Szórásnégyzet

Definíció: $D^2\xi = E(\xi - E\xi)^2$ a ξ valószínűségi változó **szórásnégyzete**, ha az $E\xi$ létezik és véges.

Definíció: ξ **szórása** a szórásnégyzet négyzetgyöke, azaz $D\xi = \sqrt{D^2\xi}$.

1) A szórásnégyzet mindig nemnegatív.

2) $D^2\xi < \infty \Leftrightarrow E\xi^2 < \infty$

Ugyanis: $(\Leftarrow) |\xi| \leq 1 + |\xi|^2$ és $E\xi^2 < \infty$ miatt $E\xi < \infty$, így $(\xi - E\xi)^2 \leq 2(\xi^2 + E\xi^2)$, $(\Rightarrow) \xi^2 \leq 2((\xi - E\xi)^2 + (E\xi)^2)$.

3) $D^2\xi = E\xi^2 - (E\xi)^2$ $\therefore D^2\xi = E(\xi - E\xi)^2 = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E\xi^2 - 2E\xi E\xi + (E\xi)^2$.

4) Minden A valós számra $E(\xi - A)^2 \geq D^2\xi$:

$$E(\xi - A)^2 = E(\xi^2 - 2A\xi + A^2) =$$

$$E\xi^2 - (E\xi)^2 + (E\xi)^2 - 2A \cdot E\xi + A^2 = D^2\xi + (E\xi - A)^2.$$

Szórásnégyzet (folyt.)

$$5) D^2\xi = 0 \Leftrightarrow \xi = c :$$

$$\text{Ha } \xi = c, \text{ akkor } \xi - E\xi = c - c = 0 \Rightarrow E(\xi - E\xi)^2 = 0.$$

$$\text{Ha } E(\xi - E\xi)^2 = 0 \Rightarrow (\xi - E\xi)^2 = 0 \Rightarrow \xi = E\xi$$

$$6) D^2(a\xi + b) = a^2 D^2\xi :$$

$$E(a\xi + b - aE\xi - b)^2 = E(a^2(\xi - E\xi)^2) = a^2 E(\xi - E\xi)^2.$$

Szórásnégyzet (folyt.)

7) $\xi_1, \xi_2, \dots, \xi_n$ páronként függetlenek és

$$D^2\xi_1, \dots, D^2\xi_n < \infty \Rightarrow D^2(\xi_1 + \dots + \xi_n) = \sum_{i=1}^n D^2\xi_i \therefore$$

$$D^2(\xi_1 + \dots + \xi_n) = E\left(\sum_{i=1}^n (\xi_i - E\xi_i)\right)^2 =$$

$$E\left(\sum_{i=1}^n (\xi_i - E\xi_i)^2 + \sum_{i \neq j} (\xi_i - E\xi_i)(\xi_j - E\xi_j)\right) =$$

$$\sum_{i=1}^n \underbrace{E(\xi_i - E\xi_i)^2}_{D^2\xi_i} + \sum_{i \neq j} \underbrace{E((\xi_i - E\xi_i)(\xi_j - E\xi_j))}_{=E(\xi_i - E\xi_i) \cdot E(\xi_j - E\xi_j) = 0 \cdot 0}.$$

$$D^2\left(\sum_{i=1}^n c_i \xi_i\right) = \sum_{i=1}^n c_i^2 D^2\xi_i.$$

Szórásnégyzet példák

- 1) **A indikátorának** szórásnégyzete

$$D^2\chi_A = E\chi_A^2 - (E\chi_A)^2 = E\chi_A - (E\chi_A)^2 = P(A) \cdot (1 - P(A)).$$

- 2) $\xi \sim B(n, p)$ (**binomiális valószínűségi változó**)

X_1, \dots, X_n független p -ind. $\Rightarrow X_1 + \dots + X_n \sim B(n, p)$ és

$$D^2\xi = D^2(X_1 + \dots + X_n) = D^2X_1 + \dots + D^2X_n = n \cdot p \cdot (1 - p).$$

- 3) η λ -**exponenciális** valószínűségi változó,

$$E\eta = \frac{1}{\lambda}.$$

$$E\eta^2 = \int_0^\infty x^2 \cdot \lambda \cdot e^{-\lambda x} dx = [x^2(-e^{-\lambda x})]_0^\infty + \int_0^\infty 2x \cdot e^{-\lambda x} dx$$

az összeg első tagja 0, az integrál

$$\frac{2}{\lambda} \int_0^\infty x \cdot \lambda \cdot e^{-\lambda x} dx = \frac{2}{\lambda} E\eta \Rightarrow E\eta^2 = \frac{2}{\lambda^2}.$$

$$D^2\eta = E\eta^2 - (E\eta)^2 = \frac{1}{\lambda^2}.$$

Szórásnégyzet példák (folyt.)

4) $\xi \sim N(0, 1)$ (**normális** eloszlású valószínűségi változó)

$$E\xi = 0 \Rightarrow D^2\xi = E\xi^2 = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx =$$

$$\frac{1}{\sqrt{2\pi}} \cdot \left[x \cdot (-e^{-\frac{x^2}{2}}) \right]_{-\infty}^{\infty} + \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx}_{=1} = 1.$$

Általánosan $m + \sigma\xi \sim N(m, \sigma^2)$

$$D^2(m + \sigma\xi) = \sigma^2 \cdot D^2\xi = \sigma^2.$$

Szórásnégyzet példák (folyt.)

5) $\eta \sim \lambda$ -Poisson.

$$E\eta = \lambda.$$

$$E\eta^2 = \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k \cdot e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k \cdot e^{-\lambda}}{(k-1)!} =$$

$$\sum_{k=2}^{\infty} \frac{\lambda^k \cdot e^{-\lambda}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^k \cdot e^{-\lambda}}{(k-1)!} = (\lambda^2 \cdot 1) + (\lambda) \Rightarrow$$

$$D^2\eta = E\eta^2 - (E\eta)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Momentumok és centrális momentumok

Definíció: ξ k-adik momentuma: $E\xi^k$

Definíció: ξ k-adik abszolút momentuma: $E|\xi|^k$

Definíció: ξ k-adik centrális momentuma: $E(\xi - E\xi)^k$

Definíció: ξ k-adik abszolút centrális momentuma:
 $E|\xi - E\xi|^k$.

Medián

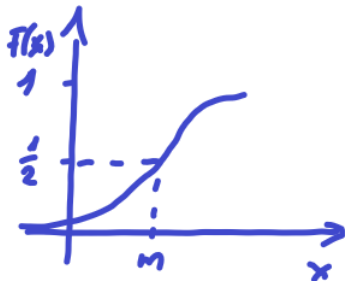
Tudjuk, hogy minden A valós számra $E(\xi - A)^2 \geq E(\xi - E\xi)^2$.
Mit mondhatunk $E|\xi - A|$ -ról?

Medián

Tudjuk, hogy minden A valós számra $E(\xi - A)^2 \geq E(\xi - E\xi)^2$.

Mit mondhatunk $E|\xi - A|$ -ről?

Keressük az $F(x) = \frac{1}{2}$ egyenlet megoldását:

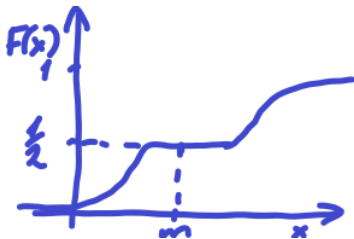


Amennyiben $\exists! x_0 : F(x_0) = \frac{1}{2}$, úgy a medián: $m = x_0$.

Medián (folyt.)

Amennyiben $\{x : F(x) = \frac{1}{2}\}$ intervallum, úgy a medián:

$$m = \frac{\inf\{x:F(x)=\frac{1}{2}\} + \max\{x:F(x)=\frac{1}{2}\}}{2}.$$

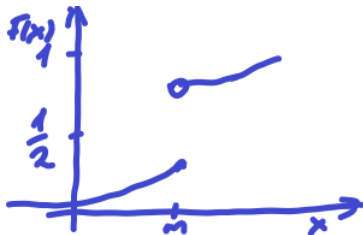


Az előző 2 esetben $P(\xi < m) = P(\xi \geq m) = \frac{1}{2}$.

Medián (folyt.)

Amennyiben

$\nexists x : F(x) = \frac{1}{2} \Rightarrow \exists x_0 : P(\xi < x_0) = F(x_0) < \frac{1}{2}, P(\xi \leq x_0) > \frac{1}{2}$,
 úgy a medián: $m = x_0$:



$$E|\xi - A| \geq E|\xi - m| \text{ (biz. nélkül)}$$

Medián (példa)

Amennyiben $x_1 < x_2 < \dots < x_n$ a lehetséges értékek és

$$P(\xi = x_1) = P(\xi = x_2) = \dots = P(\xi = x_n) = \frac{1}{n} \Rightarrow$$

páros n esetén $m = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$,

páratlan n esetén $m = x_{\frac{n+1}{2}}$

Kvantilis

p -adrendű kvantilis: medián általánosítása $\frac{1}{2}$ helyett $0 < p < 1$ -re.

Amennyiben $\exists! x_0 : F(x_0) = p \Rightarrow Q_p = Q_p(\xi) = x_0$,

amennyiben $\{x : F(x) = p\}$ intervallum

$$\Rightarrow Q_p = \frac{\inf\{x:F(x)=p\} + \max\{x:F(x)=p\}}{2},$$

Amennyiben

$$\exists x_0 : P(\xi < x_0) = F(x_0) < p, P(\xi \leq x_0) > p \Rightarrow Q_p = x_0.$$

Nevezetes kvantilisek

$Q_{0.5}$: medián

$Q_{0.25}, Q_{0.75}$: alsó és felső kvartilis

$Q_{k/100}$: k -adik percentilis ($k = 1, 2, \dots, 99$)

"200 éves esemény"

Legyenek ξ_1, ξ_2, \dots független, azonos eloszlásúak,
 $P(\xi_i \geq Q_{99,5\%}) = 0,5\%$.

$$N := \min\{n : \xi_n \geq Q_{99,5\%}\}$$

$$\Rightarrow EN = 200$$

"200 éves esemény": félreérthető megfogalmazás

Módusz

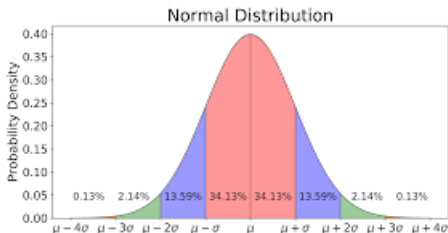
Diszkrét eloszlás esetén, ha $x_1 < x_2 < \dots$ a lehetséges értékek, $p_i = P(\xi = x_i)$ -k a valószínűségek, akkor x_j az **eloszlás módusza**, ha p_j a $\{p_i\}$ sorozat lokális maximumhelye.

Abszolút folytonos eloszlás esetén x az **eloszlás módusza**, ha a sűrűségfüggvény lokális maximumhelye.

unimodális eloszlás: 1 módusz van, bimodális eloszlás: 2 módusz van, ...

Módusz (példa)

$N(\mu, \sigma^2)$ esetén μ a várható érték, a medián és a módusz is.



Kovariancia és korreláció

Definíció: ξ és η valószínűségi változók **kovarianciája**

$$\text{cov}(\xi, \eta) = E\{(\xi - E\xi)(\eta - E\eta)\},$$

korrelációja $R(\xi, \eta) = \text{corr}(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{D\xi \cdot D\eta}.$

Tulajdonságok:

1) $\text{cov}(\xi, \eta) = E(\xi \cdot \eta) - E\xi \cdot E\eta.$

2) $|R(\xi, \eta)| \leq 1 :$

$$|\text{cov}(\xi, \eta)| = |E\{(\xi - E\xi)(\eta - E\eta)\}| \stackrel{\text{Cauchy}}{\leq} \sqrt{E(\xi - E\xi)^2 \cdot E(\eta - E\eta)^2} = D\xi \cdot D\eta.$$

Cauchy-egyenlőtlenség

Tétel: Tetszőleges véges várható értékű X, Y v.v.-ra
 $(E(XY))^2 \leq EX^2 EY^2$

Bizonyítás:

Tetszőleges valós λ -ra igaz, hogy $0 \leq E(X - \lambda Y)^2 \Rightarrow$
 $0 \leq \lambda^2 EY^2 - 2\lambda E(XY) + EX^2 \Rightarrow 4(E(XY))^2 - 4EX^2 EY^2 \leq 0.$

Kovariancia és korreláció tulajdonságok

3) $|R(\xi, \eta)| = 1 \Leftrightarrow$ létezik $a \neq 0$ és b , hogy $\xi = a\eta + b$.

Biz.: \Leftarrow

létezik ilyen a és $b \Rightarrow \xi - E\xi = a(\eta - E\eta)$, $D^2\xi = a^2 D^2\eta \Rightarrow$
 $D\xi = |a|D\eta$, $\text{cov}(\xi, \eta) = aE(\eta - E\eta)^2 = aD^2\eta \Rightarrow$

$$R = \frac{aD^2\eta}{|a|D\eta D\eta} = \frac{a}{|a|}$$

\Rightarrow

$$R(\xi, \eta) = 1$$

$$\xi' = \frac{\xi - E\xi}{D\xi}, \eta' = \frac{\eta - E\eta}{D\eta} \Rightarrow$$

$$E\xi' = E\eta' = 0 \text{ és } D^2\xi' = D^2\eta' = 1 \Rightarrow$$

$$E(\xi'\eta') = 1, E(\xi' - \eta')^2 = E\xi'^2 - 2E(\xi'\eta') + E\eta'^2 = 0 \Rightarrow$$

$$\xi' = \eta' \Rightarrow$$

ξ az η -nak lineáris transzformáltja

Kovariancia és korreláció tulajdonságok

4) ξ és η függetlenek $\Rightarrow R(\xi, \eta) = 0$

$$\text{cov}(\xi, \eta) = E(\xi\eta) - E\xi \cdot E\eta = E\xi \cdot E\eta - E\xi \cdot E\eta = 0.$$

5) $\min_{a,b} E(\xi - a\eta - b)^2 = E(\xi - m_1 - r \cdot \frac{\sigma_1}{\sigma_2} \cdot (\eta - m_2))^2 = (1 - r^2) \cdot \sigma_1^2,$

ahol $m_1 = E\xi$, $m_2 = E\eta$, $\sigma_1^2 = D^2\xi$, $\sigma_2^2 = D^2\eta$ és
 $r = R(\xi, \eta)$.

Biz.:

$$\begin{aligned} E(\xi - a\eta - b)^2 &= E(\xi - m_1 - a(\eta - m_2) + m_1 - am_2 - b)^2 = \\ &= \sigma_1^2 + a^2\sigma_2^2 + \underbrace{(m_1 - am_2 - b)^2}_{=m_1 - am_2} - 2a \cdot \underbrace{\text{cov}(\xi, \eta)}_{=r \cdot \sigma_1 \cdot \sigma_2} = \\ &= \sigma_1^2 + \underbrace{a^2\sigma_2^2 - 2a \cdot r\sigma_1\sigma_2}_{(a\sigma_2 - r\sigma_1)^2 - r^2\sigma_1^2} = (1 - r^2) \cdot \sigma_1^2 + (a\sigma_2 - r\sigma_1)^2 \Rightarrow \end{aligned}$$

$$a = r \cdot \frac{\sigma_1}{\sigma_2}$$

Egyenlőtlenségek

Tétel: [Markov-egyenlőtlenség]: Legyen ξ nemnegatív valószínűségi változó, amelynek létezik az $E\xi$ várható értéke, továbbá legyen c pozitív szám. Ekkor $P(\xi \geq c) \leq \frac{E\xi}{c}$.

Bizonyítás: Csak diszkrétre.

$$E\xi = \sum_k x_k P(\xi = x_k) \geq \sum_{k: x_k \geq c} x_k P(\xi = x_k) \geq \sum_{k: x_k \geq c} c P(\xi = x_k) = P(\xi \geq c)$$

Tétel[Csebisev-egyenlőtlenség]: Ha ξ szórásnégyzete véges, azaz $D^2\xi < \infty$, valamint $0 \leq \lambda$, akkor teljesül a

$$P(|\xi - E\xi| \geq \lambda) \leq \frac{D^2\xi}{\lambda^2} \text{ egyenlőtlenség.}$$

Bizonyítás: A Markov-egyenlőtlenségből könnyen adódik, hiszen az $\eta := (\xi - E\xi)^2$ választással $P(\eta \geq \lambda^2) \leq \frac{E\eta}{\lambda^2} = \frac{D^2\xi}{\lambda^2}$.

Példa

Egy párt szavazótáborát szeretnénk megbecsülni úgy, hogy legalább 0,95 valószínűséggel legfeljebb 1%-ot tévedjünk. Jelölje N az összes ember, M a kérdéses pártra szavazók, n pedig a megkérdezettek számát, ekkor $p := \frac{M}{N}$ -et akarjuk jól közelíteni. Legyen továbbá x_i értéke 1, ha az i -edik megkérdezett az adott pártra szavaz és 0 különben. Ekkor a

$$P\left(\left|\frac{x_1 + \dots + x_n}{n} - \frac{M}{N}\right| \leq 0,01\right) \geq 0,95$$

egyenlőtlenségnek kell teljesülnie, ami pontosan akkor igaz, ha $P\left(\left|\frac{x_1 + \dots + x_n}{n} - p\right| > 0,01\right) \leq 0,05$. A Csebisev-egyenlőtlenség

alapján $P\left(\left|\frac{x_1 + \dots + x_n}{n} - p\right| > 0,01\right) \leq \frac{D^2\left(\frac{x_1 + \dots + x_n}{n}\right)}{0,01^2}$, ahol

$$\frac{\frac{1}{n^2} D^2(\sum x_i)}{0,01^2} = \frac{\frac{1}{n^2} \cdot n \cdot p(1-p)}{0,01^2} = \frac{10000 \cdot p(1-p)}{n} \leq \frac{10000 \cdot \frac{1}{4}}{n} \leq \frac{5}{100}, \text{ tehát}$$

$n \geq 50000$ ember választása biztosan elegendő. 