Valószínűségszámítás

4. előadás

Arató Miklós

2021.03.02.

Tartalomjegyzék

Emlékeztető: Diszkrét valószínűségi változók várható értéke

Definíció: $\xi \geq 0$ **diszkrét**, ekkor $E\xi = \sum_k x_k \cdot P(\xi = x_k)$.

Definíció: ξ **diszkrét**, ekkor $E\xi = E\xi^+ - E\xi^-$, ha $\min(E\xi^+, E\xi^-) < \infty$.

$E\xi$ véges $\Leftrightarrow \sum_k |x_k| \cdot P(\xi = x_k) < \infty$ Ekkor $E\xi = \sum_k x_k \cdot P(\xi = x_k)$

Feltételes várható érték

Definíció: $P(A) > 0$, ξ diszkrét, ekkor ξ **feltételes várható értéke** az A feltételre nézve:
$E(\xi|A) = \sum_k x_k \cdot P(\xi = x_k|A)$.

Teljes várható érték tétel: ξ diszkrét valószínűségi változó véges várható értékkel és $A_1, A_2, \dots$ teljes eseményrendszer, ahol $0 < P(A_i)$. Ekkor $E\xi = \sum_n E(\xi|A_n) \cdot P(A_n)$.

Meghatározás és tulajdonságok

Definíció: $\xi \geq 0$ **abszolút folytonos eloszlású**, ekkor $E\xi = \int_{\mathbb{R}} x \cdot f_\xi(x)\, dx$.

Definíció: ξ **abszolút folytonos eloszlású**, ekkor $E\xi = E\xi^+ - E\xi^-$, ha $\min(E\xi^+, E\xi^-) < \infty$.

$Eg(\xi) = \int_{\mathbb{R}} g(x) \cdot f_\xi(x)\, dx$

$\xi \geq 0 \Rightarrow E\xi = \int_0^\infty (1 - F_\xi(y))\, dy$

ξ, η függetlenek, $E|\xi|$, $E|\eta|$ végesek. $\Rightarrow E|\xi \cdot \eta|$ is véges és $E(\xi \cdot \eta) = E(\xi) \cdot E(\eta)$.

Abszolút folytonos eloszlású példák

1. $\xi \sim E(a,b)$ esetén $E\xi = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{a+b}{2}$.
2. λ-**exponenciális** eloszlás várható értéke $E\xi = \int_0^\infty e^{-\lambda y}\, dy = \frac{1}{\lambda}$.
3. $\xi \sim N(0,1)$ esetén $E\xi = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\, dx = 0$, hiszen a sűrűségfüggvény szimmetrikus, így az integrálban egy páratlan függvény szerepel (továbbá az integrál konvergens, mert elég nagy x-re $x \cdot e^{-\frac{x^2}{2}}$ felülről becsülhető az e^{-x} függvénnyel). Általánosan pedig $m + \sigma\xi \sim N(m, \sigma^2)$ esetén $E(m+\sigma\xi) = m + \sigma \cdot E\xi = m$.

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. $D^2\xi = E\xi^2 - (E\xi)^2$:, $D^2\xi = E(\xi - E\xi)^2 = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E\xi^2 - 2E\xi E\xi + (E\xi)^2$.
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) $D^2\xi = 0 \Leftrightarrow \xi = c$:
Ha $\xi = c$, akkor $\xi - E\xi = c - c = 0 \Rightarrow E(\xi - E\xi)^2 = 0.$
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) $\xi_1, \xi_2, \ldots, \xi_n$ páronként függetlenek és

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

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

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

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

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

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**)
$X_1,\ldots,X_n$ független p-ind. $\Rightarrow X_1+\ldots+X_n \sim B(n,p)$ és
$D^2\xi = D^2(X_1+\ldots+X_n) = D^2X_1+\ldots+D^2X_n = n\cdot p\cdot(1-p)$.

3. η λ-**exponenciális** valószínűségi változó,
$E\eta = \frac{1}{\lambda}$.
$E\eta^2 = \int_0^\infty x^2\cdot\lambda\cdot e^{-\lambda x}\,dx = \left[x^2(-e^{-\lambda x})\right]_0^\infty + \int_0^\infty 2x\cdot e^{-\lambda x}\,dx$
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: $E|\xi - E\xi|^k$

.

Kovariancia és korreláció

Definíció: ξ és η valószínűségi változók **kovarianciája**
$cov(\xi,\eta) = E\{(\xi - E\xi)(\eta - E\eta)\}$,
korrelációja $R(\xi,\eta) = corr(\xi,\eta) = \frac{cov(\xi,\eta)}{D\xi \cdot D\eta}$.

Tulajdonságok:

1. $cov(\xi,\eta) = E(\xi \cdot \eta) - E\xi \cdot E\eta.$
2. $|R(\xi,\eta)| \leq 1$:
   $|cov(\xi,\eta)| = |E\{(\xi - E\xi)(\eta - E\eta)\}| \overset{CSB}{\leq}$
   $\sqrt{E(\xi - E\xi)^2 \cdot E(\eta - E\eta)^2} = D\xi \cdot D\eta.$

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 $b \Rightarrow \xi - E\xi = a(\eta - E\eta)$, $D^2\xi = a^2 D^2\eta \Rightarrow$
   $D\xi = |a|D\eta$, $cov(\xi,\eta) = aE(\eta - E\eta)^2 = aD^2\eta \Rightarrow$.
   $R = \frac{aD^2\eta}{|a|D\eta D\eta} = \frac{a}{|a|}$
   $\Rightarrow$
   $R(\xi,\eta) = 1$
   $\xi' = \frac{\xi - E\xi}{D\xi}, \eta' = \frac{\eta - E\eta}{D\eta} \Rightarrow$
   $E\xi' = E\eta' = 0$ 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 $\Rightarrow R(\xi,\eta)=0$
   $cov(\xi,\eta)=E(\xi\eta)-E\xi\cdot E\eta=E\xi\cdot E\eta-E\xi\cdot E\eta=0.$

5. $\min\limits_{a,b} E(\xi-a\eta-b)^2=E(\xi-m_1-r\cdot\frac{\sigma_1}{\sigma_2}\cdot(\eta-m_2))^2=(1-r^2)\cdot\sigma_1^2,$
   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.:
   $E(\xi-a\eta-b)^2=E(\xi-m_1-a(\eta-m_2)+m_1-am_2-b)^2=$
   $=\sigma_1^2+a^2\sigma_2^2+(m_1-am_2-\underbrace{b}_{=m_1-am_2})^2-2a\cdot\underbrace{cov(\xi,\eta)}_{=r\cdot\sigma_1\cdot\sigma_2}=$
   $=\sigma_1^2+\underbrace{a^2\sigma_2^2-2a\cdot r\sigma_1\sigma_2}_{(a\sigma_2-r\sigma_1)^2-r^2\sigma_1^2}=(1-r^2)\cdot\sigma_1^2+(a\sigma_2-r\sigma_1)^2\Rightarrow$
   $a=r\cdot\frac{\sigma_1}{\sigma_2}$

Egyenlőtlenségek

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

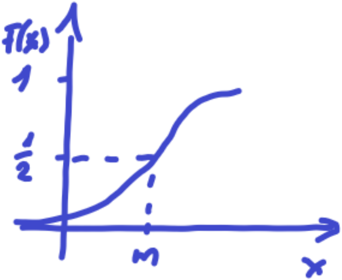
Tétel[Csebisev-egyenlőtlenség]: Ha ξ szórásnégyzete véges, azaz $D^2\xi < \infty$, valamint $0 \leq \lambda$, akkor teljesül a $P(|\xi - E\xi| \geq \lambda) \leq \frac{D^2\xi}{\lambda^2}$ egyenlőtlenség.
Biz.: A Markov-egyenlőtlenségből könnyen adódik, hiszen az $\eta := (\xi - E\xi)^2$ választással $P(\eta \geq \lambda^2) \leq \frac{E\eta}{\lambda^2} = \frac{D^2\xi}{\lambda^2}$.

Medián

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

Medián

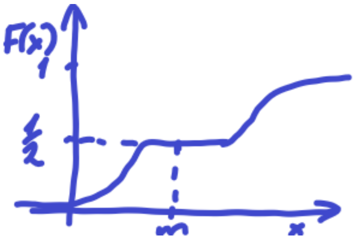
Tudjuk, hogy minden A valós számra $E(\xi - A)^2 \geq E(\xi - E\xi)^2$.
Mit mondhatunk $E|\xi - A|$-ról?
Keressük az $F(x) = \frac{1}{2}$ egyenlet megoldását:



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

Medián (folyt.)

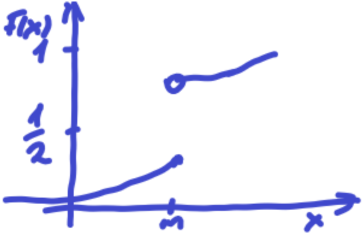
Amennyiben $\{x : F(x) = \frac{1}{2}\}$ intervallum, úgy a medián:
$m = \frac{\inf\{x:F(x)=\frac{1}{2}\}+\max\{x:F(x)=\frac{1}{2}\}}{2}$:



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: $m = x_0$:



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

Medián (példa)

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

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

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 ($k = 1, 2, \ldots, 99$)

"200 éves esemény"

Legyenek $\xi_1, \xi_2, \ldots$ 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