

Valószínűségszámítás

6. előadás

Arató Miklós

2021.03.16.

Kovariancia

$$\text{cov}(a \cdot X + b \cdot Y, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$$

$$\begin{aligned}\text{cov}(a \cdot X + b \cdot Y, Z) &= E(((a \cdot X + b \cdot Y) - E(a \cdot X + b \cdot Y))(Z - EZ)) = \\ &= E((a \cdot X - a \cdot EX)(Z - EZ)) + E((b \cdot Y - b \cdot EY)(Z - EZ)) = \\ &= \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)\end{aligned}$$

Kovariancia

$$\text{cov}(a \cdot X + b \cdot Y, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$$

$$\begin{aligned} \text{cov}(a \cdot X + b \cdot Y, Z) &= E(((a \cdot X + b \cdot Y) - E(a \cdot X + b \cdot Y))(Z - EZ)) = \\ &= E((a \cdot X - a \cdot EX)(Z - EZ)) + E((b \cdot Y - b \cdot EY)(Z - EZ)) = \\ &= \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z) \end{aligned}$$

$$D^2(X + Y) = D^2X + 2 \cdot \text{cov}(X, Y) + D^2Y$$

$$\begin{aligned} D^2(X + Y) &= E((X + Y - E(X + Y))^2) = \\ &= E((X - EX)^2) + 2E((X - EX)(Y - EY)) + E((Y - EY)^2) = \\ &= D^2X + 2 \cdot \text{cov}(X, Y) + D^2Y \end{aligned}$$

Konvergenciák

A ξ_n valószínűségi változó-sorozat **sztochasztikusan tart** ξ -hez, ha minden $0 < \varepsilon$ -ra $P(|\xi_n - \xi| \geq \varepsilon) \rightarrow 0$, $n \rightarrow \infty$ esetén. Jelölése: $\xi_n \Rightarrow \xi$ (vagy $\xi_n \xrightarrow{st.} \xi$).

$\xi_n \rightarrow \xi$ **eloszlásban**, ha $F_{\xi_n}(x) \xrightarrow{n \rightarrow \infty} F_{\xi}(x)$ az utóbbi minden folytonossági pontjában.

$\xi_n \rightarrow \xi$ **majdnem mindenütt**, ha $P(w : \xi_n(w) \rightarrow \xi(w)) = 1$. [1 valószínűségű konvergencia].

$\xi_n \rightarrow \xi$ **L^p -ben**, ha $E|\xi_n - \xi|^p \xrightarrow{n \rightarrow \infty} 0$.

Nagy számok erős törvénye

Nagy számok Cantelli-féle erős törvénye: ξ_1, ξ_2, \dots független, azonos eloszlású valószínűségi változók és $E|\xi_n - E\xi_n|^4$ véges.

Ekkor $\frac{\sum_{k=1}^n \xi_k}{n} \rightarrow E\xi_1$ 1 valószínűséggel.

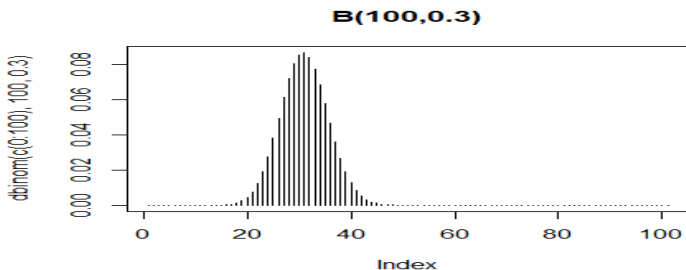
Megjegyzés: Elég, ha $E|\xi_j| < \infty$, ez a nagy számok Kolmogorov-féle erős törvénye

Binomiális eloszlás

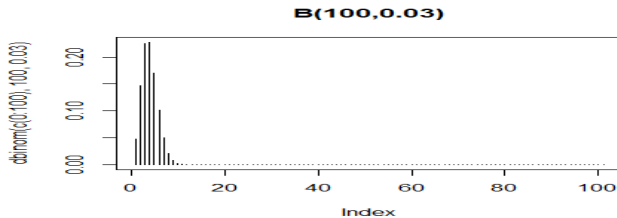
Legyenek η_i -k független, azonos eloszlású p -indikátorok, és

$$U_n = \sum_{i=1}^n \eta_i, \text{ ekkor } U_n \sim B(n, p).$$

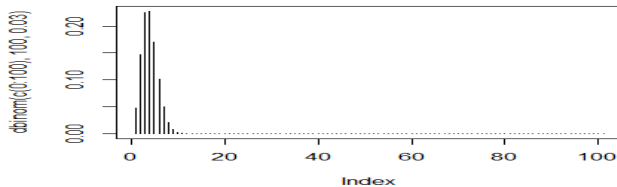
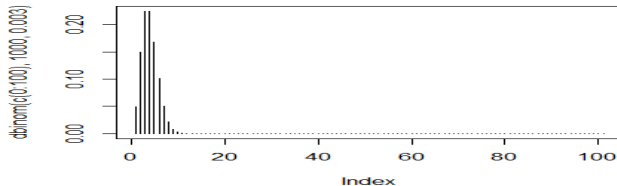
$$P(U_n = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$



Binomiális eloszlás



Binomiális eloszlás

B(100,0.03)**B(1000,0.003)**

Moivre-Laplace lokális tétel

$U_n \sim B(n, p)$ és legyen A rögzített szám, $q = 1 - p$ és $np - A\sqrt{n} \leq k \leq np + A\sqrt{n}$. Ekkor

$$\lim_{n \rightarrow \infty} \frac{P(U_n = k)}{\frac{1}{\sqrt{2\pi \cdot npq}} \cdot e^{-\frac{(k-np)^2}{2 \cdot npq}}} = 1.$$

Megjegyzés: $EU_n = np$ és $D^2U_n = npq$. Egy np várható értékű és npq szórásnégyzetű normális eloszlású valószínűségi változó sűrűségfüggvénye a k helyen pont:

$$\frac{1}{\sqrt{2\pi \cdot npq}} \cdot e^{-\frac{(k-np)^2}{2 \cdot npq}}$$

Moivre-Laplace bizonyítása

Stirling-formula egy alakja:

$$n! = \sqrt{2\pi n} \cdot n^n \cdot e^{-n} \cdot e^{R(n)}, \text{ ahol}$$

$$|R(n)| \leq \frac{1}{12n} \Rightarrow$$

$$\begin{aligned} P(U_n = k) &= \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} = \\ &= \frac{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}}{\sqrt{2\pi(n-k)} \cdot (n-k)^{n-k} \cdot e^{-(n-k)} \cdot \sqrt{2\pi k} \cdot k^k \cdot e^{-k}} \cdot e^{R(n)-R(n-k)-R(k)} \\ &\quad \cdot p^k \cdot (1-p)^{n-k} = \\ &= \frac{1}{\sqrt{2\pi n \cdot \underbrace{\frac{n-k}{n}}_{\rightarrow 1-p} \cdot \underbrace{\frac{k}{n}}_{\rightarrow p}}} \cdot \underbrace{\left[\frac{p}{\binom{k}{n}} \right]^k \cdot \left[\frac{1-p}{\binom{n-k}{n}} \right]^{n-k}}_{A_n} \cdot \underbrace{e^{R(n)-R(n-k)-R(k)}}_{\rightarrow 1}. \end{aligned}$$

Moivre-Laplace bizonyítása (folyt.)

$$A_n = ?$$

Legyen $k = np + x\sqrt{np(1-p)}$ és

$n - k = n(1-p) - x\sqrt{np(1-p)}$, ahol $|x| \leq B$.

$$\begin{aligned} \ln A_n &= -k \cdot \ln \left(\frac{k}{n} \cdot \frac{1}{p} \right) - (n-k) \cdot \ln \left(\frac{n-k}{n} \cdot \frac{1}{1-p} \right) = \\ &= -(np + x\sqrt{np(1-p)}) \cdot \ln \left(1 + x\sqrt{\frac{1-p}{np}} \right) - (n(1-p) \\ &\quad - x\sqrt{np(1-p)}) \cdot \ln \left(1 - x\sqrt{\frac{p}{n(1-p)}} \right). \end{aligned}$$

Moivre-Laplace bizonyítása (folyt.)

$$|y| < 1 \text{ esetén } \ln(1 + y) = y - \frac{y^2}{2} + O(y^3) \Rightarrow$$

$$\ln \left(1 + x \sqrt{\frac{1-p}{np}} \right) = x \sqrt{\frac{1-p}{np}} - \frac{x^2}{2} \cdot \frac{1-p}{np} + O(n^{-\frac{3}{2}})$$

és

$$\ln \left(1 - x \sqrt{\frac{p}{n(1-p)}} \right) = -x \sqrt{\frac{p}{n(1-p)}} - \frac{x^2}{2} \cdot \frac{p}{n(1-p)} + O(n^{-\frac{3}{2}}).$$

Moivre-Laplace bizonyítása (folyt.)

$$\begin{aligned}
 \ln A_n &= -x\sqrt{np(1-p)} + \frac{x^2}{2}(1-p) + O\left(\frac{1}{\sqrt{n}}\right) - x^2(1-p) + O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right) \\
 &+ x\sqrt{np(1-p)} + \frac{x^2}{2}p + O\left(\frac{1}{\sqrt{n}}\right) - x^2p + O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right) = \\
 &= -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{n}}\right).
 \end{aligned}$$

Így $x = \frac{k-np}{\sqrt{np(1-p)}} \Rightarrow$ Tétel

Poisson közelítés

Állítás: Legyen $\lambda = np$, $\eta \sim \lambda$ -Poisson és $D \subseteq \mathbb{R}$, ekkor igaz a következő:

$$|P(U_n \in D) - P(\eta \in D)| \leq \frac{\lambda^2}{n}.$$

Megjegyzés: Rögzített λ -ra és $p = \frac{\lambda}{n}$ -re kapjuk, hogy

$$|P(U_n \in D) - P(\eta \in D)| \leq \frac{\lambda^2}{n}.$$

Poisson közelítés

Állítás: Legyen $\lambda = np$, $\eta \sim \lambda$ -Poisson és $D \subseteq \mathbb{R}$, ekkor igaz a következő:

$$|P(U_n \in D) - P(\eta \in D)| \leq \frac{\lambda^2}{n}.$$

Megjegyzés: Rögzített λ -ra és $p = \frac{\lambda}{n}$ -re kapjuk, hogy

$$|P(U_n \in D) - P(\eta \in D)| \leq \frac{\lambda^2}{n}.$$

Általánosabban is megnézhetjük. Legyenek $\eta_i \sim p_i$ -indikátorok függetlenek, $U_n = \sum_{i=1}^n \eta_i$, $\lambda = \sum_{i=1}^n p_i$, $\eta \sim \lambda$ -Poisson és

$$D \subseteq \mathbb{R} \Rightarrow |P(U_n \in D) - P(\eta \in D)| \leq \sum_{i=1}^n p_i^2$$

Poisson közelítés bizonyítása

Legyen $\Omega = [0, 1]^n$ geometriai valószínűségi mező,

$$\eta_i = \begin{cases} 0 & : w_i < 1 - p_i \\ 1 & : w_i \geq 1 - p_i \end{cases}, \text{ továbbá legyen}$$

$$Y_i = \begin{cases} 0 & : w_i < e^{-p_i} \\ k & : w_i \in [\pi_{k-1}, \pi_k] \end{cases}, \text{ ahol } \pi_k = \sum_{m=0}^k e^{-p_i} \cdot \frac{p_i^m}{m!}$$

$$(k = 0, 1, 2, \dots) \Rightarrow$$

$Y_i \sim p_i$ -Poisson és $\eta_i \sim p_i$ -indikátor.

Poisson közelítés bizonyítása (folyt.)

$\eta = Y_1 + \dots + Y_n \sim \lambda$ -Poisson és $1 - p_i \leq e^{-p_i} \Rightarrow$

$\eta_i = 1$ és $Y_i = 0$, ha $1 - p_i \leq w_i < e^{-p_i}$

$\eta_i = 1$ és $Y_i \geq 2$, ha $e^{-p_i} + p_i \cdot e^{-p_i} \leq w_i \Rightarrow$

$$P(\eta_i \neq Y_i) = (e^{-p_i} - (1 - p_i)) + 1 - e^{-p_i} - p_i \cdot e^{-p_i} = p_i \cdot (1 - e^{-p_i}) \leq p_i^2.$$

$$P(U_n \in D) = P(U_n \in D, U_n = \eta) + P(U_n \in D, U_n \neq \eta) = P(\eta \in D) - P(\eta \in D, U_n \neq \eta) + P(U_n \in D, U_n \neq \eta) \Rightarrow$$

$$|P(U_n \in D) - P(\eta \in D)| \leq |P(\eta \in D, U_n \neq \eta) - P(U_n \in D, U_n \neq \eta)| \leq P(U_n \neq \eta) \leq \sum_{i=1}^n p_i^2 \Rightarrow \text{Állítás}$$

Poisson közelítés (példa)

Legyenek $\eta_i \sim p_i, i = 1, \dots, 2500$ -indikátorok függetlenek,
 $p_i = 0.001, i = 1, \dots, 1000; p_i = 0.002, i = 1001, \dots, 2000; p_i =$
 $0.003, i = 2001, \dots, 2500; U_n = \sum_{i=1}^{2500} \eta_i, \lambda = \sum_{i=1}^{2500} p_i = 4.5, \eta \sim 4.5$ -
 Poisson.

Ekkor $|P(U_n \in D) - P(\eta \in D)| \leq P(U_n \neq \eta) \leq \sum_{i=1}^{2500} p_i^2 = 0.0095$

Tétel [Centrális határeloszlástétel független, azonos eloszlású változókra]: Legyenek

ξ_1, ξ_2, \dots független, azonos eloszlásúak, $m := E\xi_1$ és

$0 < \sigma^2 = D^2\xi_i < \infty$. Ekkor minden $x \in \mathbb{R}$ -re

$$P\left(\frac{\sum_{i=1}^n \xi_i - n \cdot m}{\sigma \cdot \sqrt{n}} < x\right) \rightarrow \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$$

Példa 1.

Egy országban meg szeretnék becsülni a koronavírussal fertőzöttek arányát úgy, hogy legalább 0,95 valószínűséggel legfeljebb 1%-ot tévedjünk.

Példa 1.

Egy országban meg szeretnék becsülni a koronavírussal fertőzöttek arányát úgy, hogy legalább 0,95 valószínűséggel legfeljebb 1%-ot tévedjünk. Jelölje N az ország lakosai, M a koronavírussal szavazók, n pedig a megvizsgáltak számát, ekkor $p := \frac{M}{N}$ -et akarjuk jól közelíteni. Legyen továbbá X_i értéke 1, ha az i -edik megvizsgált koronavírussal fertőzött 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.

Példa 1. (folyt.)

A centrális határeloszlástétel szerint

$$P \left(\frac{-\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \leq \frac{\sum_{i=1}^n x_i - np}{\sqrt{n}\sqrt{p(1-p)}} \leq \frac{\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \right) \sim$$

$$\sim \Phi \left(\frac{\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \right) - \Phi \left(\frac{-\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \right) = 2 \cdot \Phi \left(\frac{\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \right) - 1 \geq 0,95$$

azaz $\Phi \left(\frac{\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \right) \geq 0,975 = \Phi(1,96)$, tehát legyen

$\frac{\sqrt{n} \cdot 0,01}{\sqrt{p(1-p)}} \geq 1,96$. Ezzel $\sqrt{n} \geq 1,96 \cdot \frac{1}{0,01} \cdot \sqrt{p(1-p)}$, ha ¹
 $n \geq 10000$, tehát ha normális közelítéssel dolgozunk, kb. 10 000
 embert kell megkérdezni.

¹mivel a $\sqrt{p(1-p)}$ nem lehet nagyobb 0,5-nél, így a jobb oldal felülről
 becsülhető 98-cal

Példa 2.

Mihez tart a következő: $e^{-n} \sum_{k=0}^{n-1} \frac{n^k}{k!} \xrightarrow{n \rightarrow \infty} ?$

Ha végtelenig összegeznénk, nyilván 1 lenne az összeg, de most $n - 1$ -ig megyünk!

Példa 2. (folyt.)

Legyenek $\eta_i \sim 1$ -Poisson függetlenek, ezekre teljesülnek a centrális határeloszlástétel feltételei, azaz felírható:

$$P\left(\frac{\sum_{i=1}^n \eta_i - n}{1 \cdot \sqrt{n}} < x\right) \rightarrow \Phi(x), \text{ ahol } \sum_{i=1}^n \eta_i \sim n\text{-Poisson, így}$$

$$P\left(\sum_{i=1}^n \eta_i < n\right) = e^{-n} \cdot \sum_{k=0}^{n-1} \frac{n^k}{k!}.$$

$$\text{Ekkor speciálisan } x = 0\text{-ra } P\left(\frac{\sum_{i=1}^n \eta_i - n}{\sqrt{n}} < 0\right) \rightarrow \Phi(0) = \frac{1}{2}.$$

Tehát a keresett összeg $\frac{1}{2}$ -hez tart.