Valószínűségszámítás

10. előadás

Arató Miklós

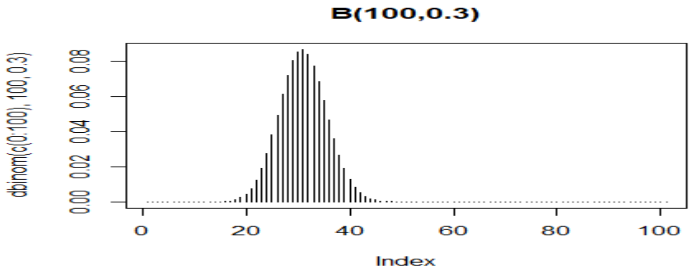
2022.04.25.

Tartalomjegyzék

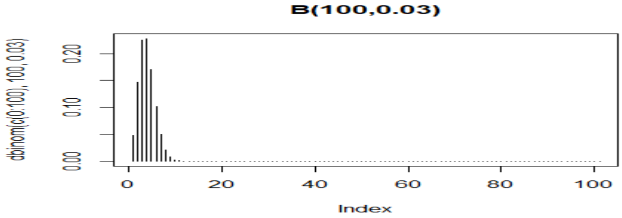
Binomiális eloszlás

Legyenek η_i-k független, azonos eloszlású p-indikátorok, és $U_n = \sum_{i=1}^{n} \eta_i$, ekkor $U_n \sim B(n, p)$.

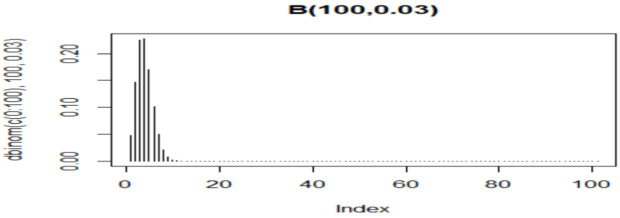
$$P(U_n = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

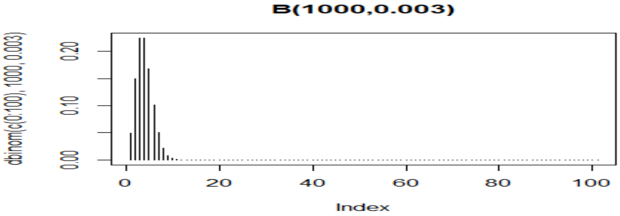


Binomiális eloszlás



Binomiális eloszlás





Moivre-Laplace lokális tétel

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

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

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

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

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

$$= \frac{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}}{\sqrt{2\pi(n-k)} \cdot (n-k)^{n-k} \cdot e^{-(n-k)} \cdot \sqrt{2\pi k} \cdot k^k \cdot e^{-k}} \cdot e^{R(n)-R(n-k)-R(}$$

$$\cdot p^k \cdot (1-p)^{n-k} =$$

$$= \frac{1}{\sqrt{2\pi n \cdot \underbrace{\frac{n-k}{n}}_{\to 1-p} \cdot \underbrace{\frac{k}{n}}_{\to p}}} \cdot \underbrace{\left[\frac{p}{\left(\frac{k}{n}\right)}\right]^k \cdot \left[\frac{1-p}{\left(\frac{n-k}{n}\right)}\right]^{n-k}}_{A_n} \cdot \underbrace{e^{R(n)-R(n-k)-R(k)}}_{\to 1}.$$

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 $|x| \le B$.

$$\ln A_n = -k \cdot \ln\left(\frac{k}{n} \cdot \frac{1}{p}\right) - (n-k) \cdot \ln\left(\frac{n-k}{n} \cdot \frac{1}{1-p}\right) =$$

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

$$-x\sqrt{np(1-p)}) \cdot \ln\left(1 - x\sqrt{\frac{p}{n(1-p)}}\right).$$

Moivre-Laplace bizonyítása (folyt.)

$|y| < 1$ 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.)

$$\ln A_n =$$

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

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

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

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,

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

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

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

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

Poisson közelítés bizonyítása (folyt.)

$\eta = Y_1 + \ldots + Y_n \sim \lambda$-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 $e^{-p_i} + p_i \cdot e^{-p_i} \leq w_i \Rightarrow$

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

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

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

CHT

Tétel [Centrális határeloszlástétel független, azonos eloszlású változókra]: Legyenek $\xi_1, \xi_2, \ldots$ független, azonos eloszlásúak, $m := E\xi_1$ és $0 < \sigma^2 = D^2\xi_i < \infty$. Ekkor minden $x \in \mathbb{R}$-re

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

Normális közelítés

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 koronavírusosak számát, 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 koronavírusos és 0 különben. Ekkor a

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \frac{M}{N}\right| \leq 0,01\right) \geq 0,95$$

egyenlőtlenségnek kell teljesülnie.

Példa 1. (folyt.)

A centrális határeloszlástétel szerint

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

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

azaz $\Phi\left(\frac{\sqrt{n}\cdot 0,01}{\sqrt{p(1-p)}}\right) \ge 0,975 = \Phi(1,96)$, tehát legyen $\frac{\sqrt{n}\cdot 0,01}{\sqrt{p(1-p)}} \ge 1,96$. Ezzel $\sqrt{n} \ge 1,96\cdot\frac{1}{0,01}\cdot\sqrt{p(1-p)}$, ha [1] $n \ge 10000$, tehát ha normális közelítéssel dolgozunk, kb. 10 000 embert kell megkérdezni.

[1] mivel 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: $e^{-n}\sum_{k=0}^{n-1}\frac{n^k}{k!}\xrightarrow{n\to\infty}$?

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:

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

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

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

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

CHT független változókra

Tétel [Centrális határeloszlástétel független változókra]: Legyenek $\xi_1, \xi_2, \ldots$ függetlenek. $m_k := E\xi_k$ és $0 < \sigma_k^2 = D^2\xi_k < \infty$, továbbá $S_n = \xi_1 + \ldots + \xi_n$, $D_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$. Teljesül az ún. Ljapunov-feltétel:

$$\frac{1}{D_n^{2+\delta}} \cdot \sum_{k=1}^{n} E|\xi_k - m_k|^{2+\delta} \xrightarrow{n\to\infty} 0,$$

valamely $0 < \delta$-ra. Ekkor minden $x \in \mathbb{R}$-re

$$P\left(\frac{S_n - ES_n}{D_n} < x\right) \to \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt$$

Közelítés hibája

Tétel [Berry-Esséen] Legyenek $\xi_1, \xi_2, \ldots$ független, azonos eloszlású valószínűségi változók, továbbá $E|\xi_1|^3 < \infty$ és $T_n := \frac{\sum\limits_{k=1}^{n} \xi_k - n \cdot m}{\sigma\sqrt{n}}$. Ekkor $\sup\limits_x |F_{T_n}(x) - \Phi(x)| \leq 0,474 \cdot \frac{E|\xi_1 - m|^3}{\sigma^3\sqrt{n}}$.

Tétel [Esséen] Legyenek $\xi_1, \xi_2, \ldots$ független valószínűségi változók, $E|\xi_k|^3 < \infty$, továbbá $S_n = \xi_1 + \ldots + \xi_n$, $D_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$,
$F_n(x) = P\left(\frac{S_n - ES_n}{D_n} < x\right)$, $L_n = \frac{\sum_{k=1}^{n} E|X_k - m_k|^3}{B_n^{-3/2}}$.
Ekkor $\sup\limits_x |F_n(x) - \Phi(x)| \leq 0,56 L_n$