

Feltételes sűrűségfüggvény és várható érték (11. előadás)

Legyen az (X, Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye a f . Az X , illetve Y sűrűségfüggvénye:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Definíció

Az X valószínűségi változónak az $Y = y$ feltételre vonatkozó **feltételes sűrűségfüggvénye** adott y valós számra:

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}.$$

A $g(X)$ mennyiség **feltételes várható értéke** az $Y = y$ feltétel mellett:

$$\mathbb{E}(g(X)|Y = y) = \int_{-\infty}^{\infty} g(x) f_{X|Y=y}(x) dx.$$

Feltételes várható érték: példa

Legyenek X és Z független standard normális eloszlású valószínűségi változók. Ekkor az $(X, X + Z)$ együttes sűrűségfüggvénye:

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xy + y^2}{2}\right).$$

Az $X + Z$ valószínűségi változó normális eloszlású $m = 0$ várható értékkel és $\sigma = 2$ szórásnégyzettel, így

$$f_{X+Z}(y) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{y^2}{4}\right).$$

Ebből:

$$\begin{aligned}\mathbb{E}(X|X + Z = y) &= \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_{X+Z}(y)} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} x \exp\left(-\frac{2x^2 - 2xy + y^2/2}{2}\right) dx \\ &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{\pi}} \exp\left(-\frac{(x - y/2)^2}{2 \cdot 1/2}\right) dx = \frac{y}{2}.\end{aligned}$$

Egyenlőtlenségek

Állítás (Markov-egyenlőtlenség)

Legyen $t > 0$ tetszőleges pozitív szám, X pedig olyan véges várható értékű valószínűségi változó, mely csak nemnegatív értékeket vesz fel, vagyis melyre $X \geq 0$ teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Egyenlőtlenségek

Állítás (Markov-egyenlőtlenség)

Legyen $t > 0$ tetszőleges pozitív szám, X pedig olyan véges várható értékű valószínűségi változó, mely csak nemnegatív értékeket vesz fel, vagyis melyre $X \geq 0$ teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Állítás (Csebisev-egyenlőtlenség)

Legyen X véges szórású valószínűségi változó, $t > 0$ pozitív szám. Ekkor

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{D^2(X)}{t^2}.$$

Egyenlőtlenségek

Állítás (Markov-egyenlőtlenség)

Legyen $t > 0$ tetszőleges pozitív szám, X pedig olyan véges várható értékű valószínűségi változó, mely csak nemnegatív értékeket vesz fel, vagyis melyre $X \geq 0$ teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Állítás (Csebisev-egyenlőtlenség)

Legyen X véges szórású valószínűségi változó, $t > 0$ pozitív szám. Ekkor

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{D^2(X)}{t^2}.$$

Következmény

Legyen X véges szórású valószínűségi változó, $t > 0$ pozitív szám. Ekkor

$$\mathbb{P}(|X - \mathbb{E}(X)| < t) \geq 1 - \frac{D^2(X)}{t^2}.$$

A Csebisev-egyenlőtlenség alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb $0,01$ -gyel tér el p -től, tetszőleges p esetén legalább 95% legyen?

A Csebisev-egyenlőtlenség alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb $0,01$ -gyel tér el p -től, tetszőleges p esetén legalább 95% legyen?

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95$$

teljesüljön minden $0 \leq p \leq 1$ -re.

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

Mivel X binomiális eloszlású:

$$\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = p; \quad D\left(\frac{X}{n}\right) = \frac{1}{n} \sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

Mivel X binomiális eloszlású:

$$\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = p; \quad D\left(\frac{X}{n}\right) = \frac{1}{n} \sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

Csebisev-egyenlőtlenség alkalmazva az X/n valószínűségi változóra:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0,01\right) \leq \frac{D^2\left(\frac{X}{n}\right)}{0,01^2} = \frac{p(1-p)}{0,01^2 \cdot n} \leq \frac{1}{4 \cdot 0,01^2 \cdot n},$$

mivel $p(1-p) \leq 1/4$ a számtani-mértani közepek közötti egyenlőtlenség szerint.

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

A Csebisev-egyenlőtlenség szerint

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0,01\right) \leq \frac{1}{4 \cdot 0,01^2 \cdot n}$$

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95,$$

azaz

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > 0,01\right) \leq 0,05.$$

Elég:

$$\frac{1}{4 \cdot 0,01^2 \cdot n} \leq 0,05 \quad \Leftrightarrow \quad n \geq \frac{1}{4 \cdot 0,01^2 \cdot 0,05} = 50000.$$

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

A Csebisev-egyenlőtlenség szerint

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0,01\right) \leq \frac{1}{4 \cdot 0,01^2 \cdot n}$$

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95,$$

azaz

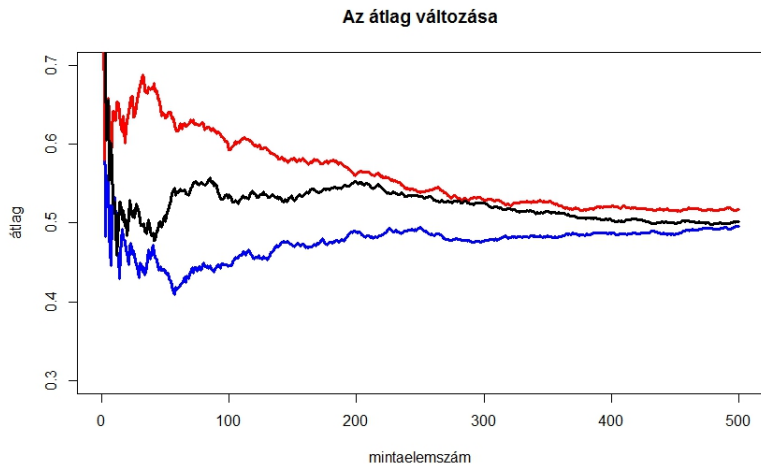
$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > 0,01\right) \leq 0,05.$$

Elég:

$$\frac{1}{4 \cdot 0,01^2 \cdot n} \leq 0,05 \quad \Leftrightarrow \quad n \geq \frac{1}{4 \cdot 0,01^2 \cdot 0,05} = 50000.$$

Ha 0,01 helyett 0,005-öt írnánk (a felét), $n \geq 200000$ (négyyszer annyi) adódna.

Az átlag konvergenciája



A $[0, 1]$ intervallumon egyenletes eloszlásból vett minta átlaga $n = 500$ -ig

Az átlag várható értéke

Állítás

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $m = \mathbb{E}(X_1) < \infty$. Ekkor

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = m.$$

Az átlag várható értéke

Állítás

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $m = \mathbb{E}(X_1) < \infty$. Ekkor

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = m.$$

Bizonyítás.

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n}\mathbb{E}(X_1 + \dots + X_n) = \frac{1}{n} \cdot nm = m.$$

Felhasználtuk a várható érték linearitását, és hogy csak eloszlástól függ:

- $\mathbb{E}(cX) = c\mathbb{E}(X)$, ha $c \in \mathbb{R}$;
- $\mathbb{E}(Y + Z) = \mathbb{E}(Y) + \mathbb{E}(Z)$;
- ha Y és Z eloszlása (azaz eloszlásfüggvényük) megegyezik, akkor $\mathbb{E}(Y) = \mathbb{E}(Z)$

Az átlag szórása

Állítás

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $\sigma = D(X_1) < \infty$. Ekkor

$$D(\bar{X}) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \sigma/\sqrt{n}.$$

Az átlag szórása

Állítás

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $\sigma = D(X_1) < \infty$. Ekkor

$$D(\bar{X}) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \sigma/\sqrt{n}.$$

Bizonyítás.

$$D(\bar{X}) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{D(X_1 + \dots + X_n)}{n} = \frac{\sqrt{n\sigma^2}}{n} = \frac{\sigma}{\sqrt{n}}.$$

Felhasználtuk a szórás alábbi tulajdonságait:

- $D(cX) = |c|D(X)$, ha $c \in \mathbb{R}$;
- $D^2(Y + Z) = D^2(Y) + D^2(Z)$, ha Y és Z függetlenek;
- ha Y és Z eloszlása megegyezik, akkor $D(Y) = D(Z)$

A nagy számok gyenge törvénye

Legyenek X_1, \dots, X_n független azonos eloszlású véges szórású valószínűségi változók. Legyen $m = \mathbb{E}(X_1)$ és $\sigma = D(X_1)$.

A korábbiak szerint

$$\mathbb{E}(\bar{X}) = m; \quad D^2(\bar{X}) = \frac{\sigma^2}{n}.$$

A nagy számok gyenge törvénye

Legyenek X_1, \dots, X_n független azonos eloszlású véges szórású valószínűségi változók. Legyen $m = \mathbb{E}(X_1)$ és $\sigma = D(X_1)$.

A korábbiak szerint

$$\mathbb{E}(\bar{X}) = m; \quad D^2(\bar{X}) = \frac{\sigma^2}{n}.$$

A Csebisev-egyenlőtlenség szerint minden $\varepsilon > 0$ -ra

$$\mathbb{P}(|\bar{X} - m| > \varepsilon) \leq \frac{D^2(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Tehát $\bar{X} \rightarrow m = \mathbb{E}(X_1)$ sztochasztikusan.

A nagy számok törvénye

Tétel (A nagy számok gyenge törvénye)

Legyenek X_1, X_2, \dots olyan valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel, hogy $D(X_1) < \infty$. Ekkor minden $\varepsilon > 0$ esetén

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| > \varepsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

azaz $\bar{X}_n \rightarrow \mathbb{E}(X_1)$ sztochasztikusan.

A nagy számok törvénye

Tétel (A nagy számok gyenge törvénye)

Legyenek X_1, X_2, \dots olyan valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel, hogy $D(X_1) < \infty$. Ekkor minden $\varepsilon > 0$ esetén

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| > \varepsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

azaz $\bar{X}_n \rightarrow \mathbb{E}(X_1)$ sztochasztikusan.

Tétel (A nagy számok erős törvénye)

Legyenek X_1, X_2, \dots valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel még, hogy $m = \mathbb{E}(X_1) < \infty$. Ekkor

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mathbb{E}(X_1) = m$$

teljesül 1 valószínűséggel $n \rightarrow \infty$ esetén.

A második esetben gyengébb feltevésből erősebb állítás következik.

Házi feladat december 11-ig

Legyenek $X_1, X_2, \dots, X_{10000}$ független azonos eloszlású valószínűségi változók. Az eloszlás tetszőlegesen választható, de ne normális eloszlás legyen. Legyen $m = \mathbb{E}(X_1)$ az X_1 várható értéke és $\sigma = D(X_1)$ az X_1 szórása.

Tetszőleges $j = 0, 1, 2, \dots, 99$ esetén legyen $Y_j = \frac{X_{100j+1} + X_{100j+2} + \dots + X_{100j+100} - 100m}{10\sigma}$, azaz a mintaelemeket százasaival csoportosítva tekintjük azt a mennyiséget, ami a centrális határeloszlástételben szerepel.

Készítsünk hisztogramot az $X_1, X_2, \dots, X_{10000}$ mintából, illetve az Y_0, Y_1, \dots, Y_{99} mintából (két külön ábrán).

Házi feladat november 27-ig: megoldás

Számítógép segítségével (például R-ben) generáljunk 1000 elemű mintát (1000 független valószínűségi változót) normális eloszlásból (a várható érték legyen tetszőlegesen választott szám 2 és 6 között, a szórás pedig szintén tetszőlegesen választott 1 és 3 között).

- 1 Készítsünk hisztogramot az első 10, 100, 1000 mintaelemből (három különböző ábrán).
- 2 Mennyi annak valószínűsége, hogy egy mintaelem értéke 2 és 6 közé esik?
- 3 Ábrázoljuk n függvényében azt, hogy a mintaelemek hányadrésze esik 2 és 6 közé, ha az első n mintaelemet tekintjük ($n = 1, 2, \dots, 1000$).

(2) Legyen $m = 5$ és $\sigma = 2$. Ekkor az R segítségével

$$\begin{aligned}\mathbb{P}(2 \leq X_1 \leq 6) &= \Phi\left(\frac{6-5}{2}\right) - \Phi\left(\frac{2-5}{2}\right) = \Phi(0,5) - \Phi(-1,5) = \\ &= \text{pnorm}(0.5) - \text{pnorm}(-1.5) = 0,624.\end{aligned}$$

Házi feladat november 27-ig: megoldás

```
> minta<-rnorm(1000, m=5, sd=2)

> hist(minta[1:10], col="#5a8ee2", main="Tízelemű N(5, 4) minta hisztogram-  
ja", xlab="értékek", ylab="gyakoriságok")

> hist(minta[1:100], col="#5a8ee2", main="Százelemű N(5, 4) minta hisztog-  
ramja", xlab="értékek", ylab="gyakoriságok")

> hist(minta, col="#5a8ee2", main="Ezerelemű N(5, 4) minta hisztogramja",  
xlab="értékek", ylab="gyakoriságok")

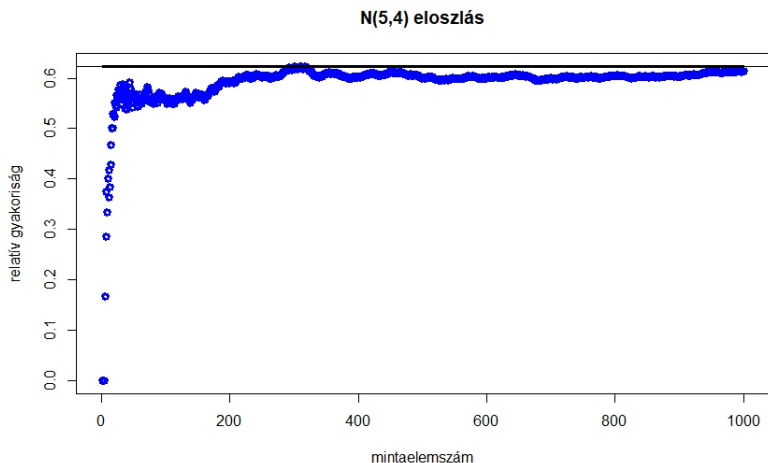
> relgyak <-function(n){sum(minta[1:n]>=2 & minta[1:n]<=6)/n}

> plot(sapply(1:1000, relgyak), col="blue", lwd="3", main="N(5,4) eloszlás",  
xlab="mintaelemszám", ylab="relatív gyakoriság")

> val=rep(0.624, 1000)

> lines(val, lwd="3")
```

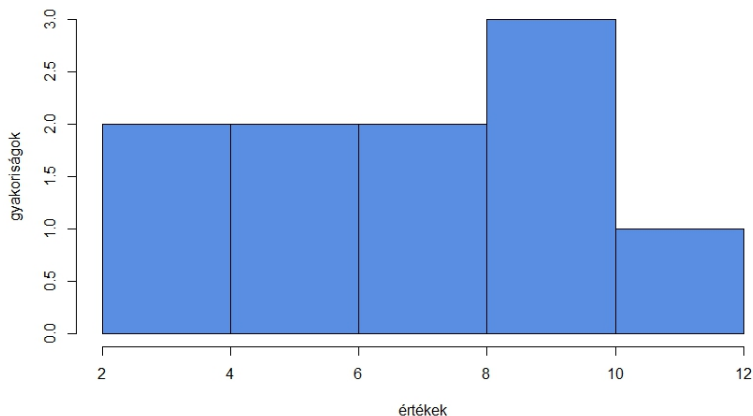
Házi feladat november 27-ig: megoldás



A 2 és 6 közötti mintaelemek relatív gyakorisága és a $\mathbb{P}(2 \leq X \leq 6)$ valószínűség.

Házi feladat november 27-ig: megoldás

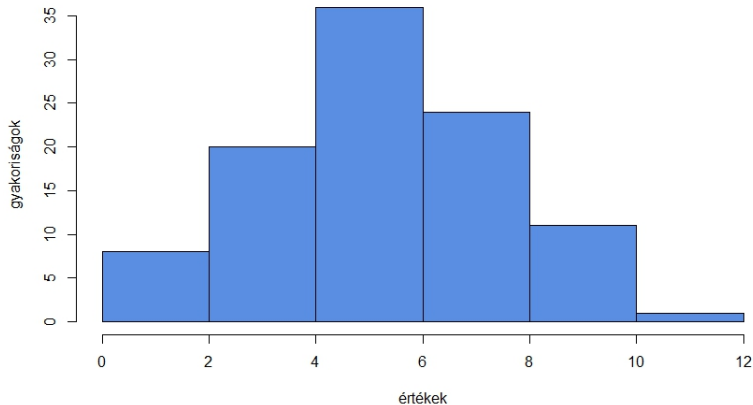
Tízelemű $N(5, 4)$ minta hisztogramja



Tízelemű $N(2, 4)$ normális eloszlású minta hisztogramja

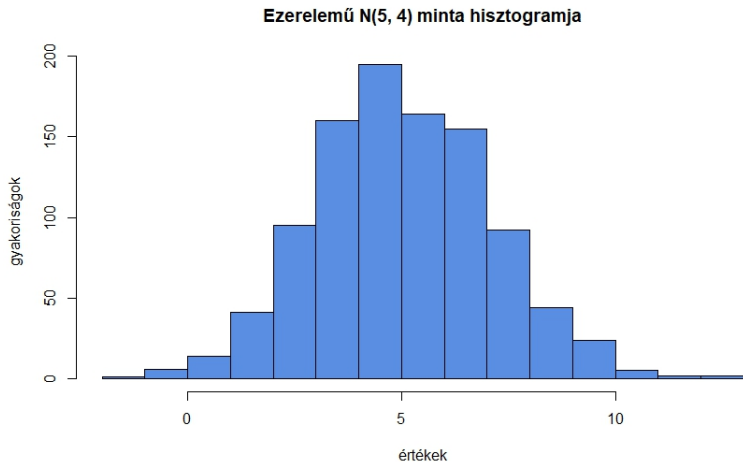
Házi feladat november 27-ig: megoldás

Százelemű $N(5, 4)$ minta hisztogramja



Százelemű $N(2, 4)$ normális eloszlású minta hisztogramja

Házi feladat november 27-ig: megoldás



Ezerlemű $N(2, 4)$ normális eloszlású minta hisztogramja