Valószínűségszámítás előadás, 8. hét, 2021. november 5.

Egyenlőtlenségek, nagy számok törvényei, konvolúció

1. Egyenlőtlenségek

Tegyük fel, hogy egy véletlenszerűen választott ember **p** valószínűséggel szavaz egy adott pártra – azonban p-t **nem ismerjük**.

Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya **legfeljebb** 1%**-kal** tér el p-től, **tetszőleges** p **esetén legalább** 95% **legyen**? Itt a hibát additívan értjük, tehát ha p az igazi valószínűség, akkor a becslésünk $p \pm 0,01$ lehet, és ennek a valószínűsége kell, hogy legalább 95% legyen.

Ennek megértéséhez ezekre van szükség:

- az arány átlagolás $\to$ milyen gyorsan csökken a szórás a mintaelemszám növelésével?
- ha a **szórás kicsi**, abból hogyan következik, hogy nagy valószínűséggel csak keveset tévedünk $\to$ ebben segít a **Markov– és a Csebisev-egyenlőtlenség**.

Többek között az **átlag** viselkedésének megértéséhez van szükség az alábbi egyenlőtlenségekre.

1.1. Állítás (Markov-egyenlőtlenség). *Legyen $t > 0$, és X* **nemnegatív, véges várható értékű** *valószínűségi változó, vagyis melyre $X \geq 0$ biztosan teljesül. Ekkor*

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

1.2. Állítás (Csebisev-egyenlőtlenség). *Legyen X* **véges szórású** *valószínűségi változó, $t > 0$ pozitív szám. Ekkor*

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

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

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

1.1. A Csebisev-egyenlőtlenség alkalmazása

Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya **legfeljebb** 1%**-kal** tér el p-től, **tetszőleges** p **esetén legalább** 95% **legyen**?

n megkérdezett, minden véletlenszerűen választott megkérdezett 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 – hiszen p-t nem ismerjük.

Mivel X binomiális eloszlású:

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

Csebisev-egyenlőtlenség alkalmazva az X/n valószínűségi változóra (utána beírjuk az előbb kiszámított szórást):

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

mivel $p(1-p)\leq\frac{(p+1-p)^2}{2^2}=1/4$ a számtani-mértani közepek közötti egyenlőtlenség szerint (a $p(1-p)=p-p^2$ függvény egy konkáv parabola, ami az $1/2$-re szimmetrikus). Az általános felső becslésre azért van szükség, mert p-t nem ismerjük, a feltételt viszont minden $0\leq p\leq 1$-re teljesítenünk kell.

A Csebisev-egyenlőtlenség szerint tehát

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

A feltétel, aminek teljesülnie kell:

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

azaz

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

Tehát ha n teljesíti az alábbi feltételt, akkor n embert **biztosan elég megkérdezni**,

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

hiszen ekkor

$$\mathbb{P}\left(\left|\frac{X}{n}-p\right|\geq 0,01\right)\leq\frac{1}{4\cdot 0,01^2\cdot n}\leq 0,05.$$

Ha $0,01$ helyett $0,005$-öt írnánk (a felét), $n\geq 200000$ (négyszer annyi) adódna. Általában is, ezzel a módszerrel azt kapjuk, hogy feleakkora megengedett (additív) hibához négyszer annyi, harmadakkorához kilencszerannyi stb. megfigyelés kell. Később más módszert is fogunk látni ennek a kérdésnek a megválaszolására, és további lehetőségek is elképzelhetők, de ez a jelenség a legtöbb módszernél megfigyelhető. Ha a 95% helyett írnánk például 99%-ot, akkor a $0,05$ helyére kerülne $0,01$, ötször annyi megkérdezésre van szükség – ez viszont a pontosabb számításoknál másképpen alakul majd.

Vegyük észre, hogy a szükséges mintaelemszám csökkenthető, ha a p-ről van valamilyen előzetes információnk. Ha például biztosan tudnánk, hogy $p\leq 0,1$, akkor ennél jobb becslést is használhatunk, hiszen ilyenkor $p(1-p)\leq 0,1\cdot 0,9=0,09$, és $n=18000$ embert elég megkérdezni. De ha semmilyen előzetes információt nem használunk p-ről, akkor $p=1/2$ is előfordulhat, akkor pedig $1/4$ jelenik meg a felső becslésben, és ezzel a módszerrel nem tudunk mást állítani, mint hogy $n=50000$ elég.

1.2. A Markov-egyenlőtlenség bizonyítása

Legyen X nemnegatív értékeket felvevő valószínűségi változó, melynek várható értéke létezik, és legyen $t>0$ pozitív szám. Tekintsük az alábbi Y valószínűségi változót:

$$Y=\begin{cases}t, & \text{ha } X\geq t;\\ 0, & \text{ha } X<t.\end{cases}$$

Mindkét esetben Y értéke legfeljebb X értéke (hiszen $X\geq 0$):

$$Y\leq X\qquad\Rightarrow\qquad\mathbb{E}(Y)\leq\mathbb{E}(X).$$

Másrészt, mivel Y értéke összesen kétféle lehet, a várható értékét kiszámítva:

$$\mathbb{E}(Y)=0\cdot\mathbb{P}(Y=0)+t\cdot\mathbb{P}(Y=t)=t\cdot\mathbb{P}(X\geq t)\leq\mathbb{E}(X).$$

Az utolsó egyenlőtlenség mindkét oldalát a t pozitív számmal osztva a Markov-egyenlőtlenséget kapjuk. □

1.3. A Csebisev-egyenlőtlenség bizonyítása

Legyen $Z = (X - \mathbb{E}(X))^2$. Ez a valószínűségi változó nemnegatív, ezért alkalmazható a Markov-egyenlőtlenség, a t^2 pozitív számmal és a Z valószínűségi változóval:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) = \mathbb{P}((X - \mathbb{E}(X))^2 \geq t^2) = \mathbb{P}(Z \geq t^2) \overset{\text{Markov}}{\leq} \frac{\mathbb{E}(Z)}{t^2} = \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{t^2} = \frac{D^2(X)}{t^2}$$

a szórásnégyzet definíciója alapján. □

2. Az átlag viselkedése független azonos eloszlású esetben

A Csebisev-egyenlőtlenséget az átlag viselkedésének leírására is használhatjuk.

A statisztikában alapvető kérdés, hogy ha

- **ugyanazt a mérést**
- sokszor, **egymástól függetlenül** megismételjük,
- majd a kapott eredményeket **átlagoljuk**,
- akkor az átlag, mint valószínűségi változó hogyan viselkedik

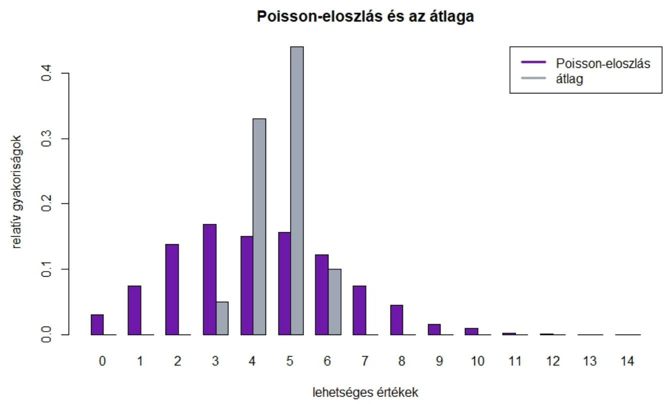
Vagyis: $X_1, X_2, \ldots, X_n$ **független**, **azonos eloszlású** valószínűségi változók, akkor mit mondhatunk az

$$\frac{X_1 + X_2 + \ldots + X_n}{n}$$

átlagról: mennyi a **várható értéke** és mennyi a **szórása**?

Azonos eloszlás: $\mathbb{P}(X_j \in A) = \mathbb{P}(X_1 \in A)$ tetszőleges j-re és $A \subseteq \mathbb{R}$ „megfelelő” halmazra, vagy: X_j és X_1 eloszlásfüggvénye megegyezik tetszőleges j-re.

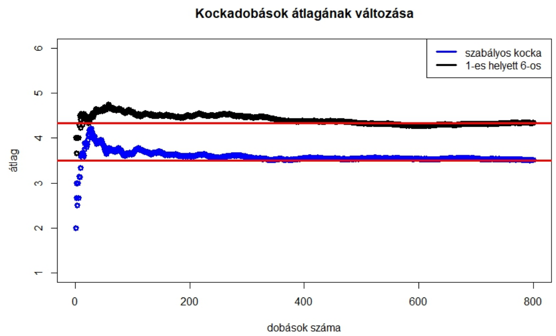
azonos eloszlás $\Rightarrow$ **azonos várható érték**, **azonos szórás**



1. ábra. Poisson-eloszlások átlaga

Az 1. ábrán 1000 darab $\lambda = 5$ paraméterű Poisson-eloszlású valószínűségi változó, illetve 100 darab, tíz független, $\lambda = 5$ paraméterű Poisson-eloszlású valószínűségi változó átlagaként előálló megfigyelés hisztogramja látható. A tíz megfigyelés átlagának eloszlása különbözik az eredeti eloszlástól, nem Poisson-eloszlást kapunk. Az ábra alapján **átlagolásnál a várható érték nem változik**, ez mindkét esetben 5, **a szórás csökken**. Pontosabban: az átlagolás után kapott új eloszlás várható értéke közel van az 5-höz, de a szórása kisebb, mint a kiindulásként használt Poisson-eloszlás szórása.

Sőt: a 2. ábrán két dobókocka esetében tekintettünk $800 - 800$ dobást, és k függvényében ábrázoltuk az első k dobás átlagát. Az egyik esetben a kockán $1-6$-ig láthatók a számok (ezek átlaga, így egy dobás várható értéke $3,5$, a másik esetben 1-es helyett is hatos van, ilyenkor egy dobás várható értéke $4,33$.



2. ábra. A dobások átlagának változása a dobások számának növelésével

2.1. Az átlag várható értéke

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

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

Bizonyítás.

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

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 $\mathbb{E}(Y) = \mathbb{E}(Z)$

2.2. Az átlag szórása

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

$$D(\overline{X}) = D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{D(X_1)}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}.$$

Bizonyítás.

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

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 $D(Y) = D(Z)$

3. A nagy számok törvényei

A fenti ábra azt sugallják, hogy a mintaátlag független azonos eloszlású valószínűségi változók esetén a várható értékhez konvergál. Kérdés, mit is jelent az, hogy valószínűségi változók sorozata konvergál.

Valószínűségi változók sorozatának (mint amilyen az átlagok sorozata) **többféle értelemben** definiálhatjuk a határértékét.

3.1. Definíció. *A $Z_1, Z_2, \ldots,$ valószínűségi változókból álló sorozat* **sztochasztikusan konvergál** *az Z valószínűségi változóhoz, ha minden $\varepsilon > 0$-ra*

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \to 0$$

teljesül $n \to \infty$ esetén.

3.2. Definíció. *A $Z_1, Z_2, \ldots,$ valószínűségi változókból álló sorozat* **1 valószínűséggel** *konvergál az Z valószínűségi változóhoz, ha*

$$\mathbb{P}(Z_n \to Z) = \mathbb{P}(\omega \in \Omega : Z_n(\omega) \to Z(\omega) \ n \to \infty \text{ esetén}) = 1.$$

Megjegyzés. 1 valószínűséggel konvergál a sorozat $\Rightarrow$ sztochasztikusan is, de fordítva nem feltétlenül: van olyan sorozat, ami sztochasztikusan 0-hoz konvergál, de 1 valószínűséggel nem konvergens.

3.1. A nagy számok gyenge törvénye

3.1. Tétel (A nagy számok gyenge törvénye). *Legyenek $X_1, X_2, \ldots$ olyan valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel, hogy $D(X_1) < \infty$. Ekkor minden $\varepsilon > 0$ esetén*

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}(X_1)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

azaz $\overline{X}_n \to \mathbb{E}(X_1)$ sztochasztikusan. Itt $\overline{X}_n = (X_1 + \ldots + X_n)/n$ az átlagot jelöli.

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

A korábbiak szerint

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

Ebből az is következik, hogy az $\overline{X}$ valószínűségi változó szórása véges, alkalmazható a Csebisev-egyenlőtlenség.

A Csebisev-egyenlőtlenség (1.2. állítás) szerint minden $\varepsilon > 0$-ra

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

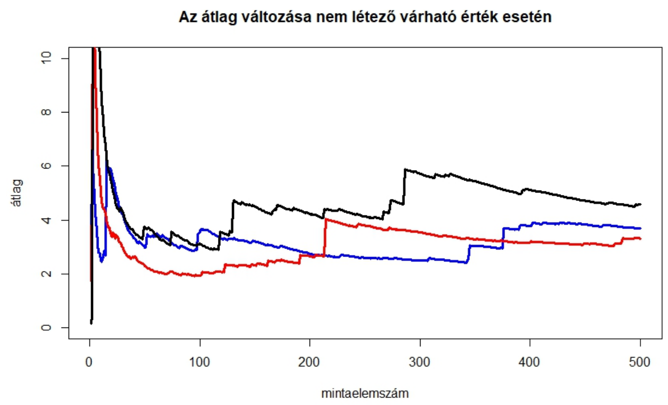
Tehát $\overline{X} \to m = \mathbb{E}(X_1)$ sztochasztikusan. □

3.2. Tétel (A nagy számok erős törvénye). *Legyenek $X_1, X_2, \ldots$ valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel még, hogy $m = \mathbb{E}(X_1) < \infty$. Ekkor*

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1) = m$$

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

A második esetben gyengébb feltevésből erősebb állítás következik. Ennek bizonyításához jóval összetettebb matematikai módszerekre van szükség.



3. ábra. Az átlag változása egy olyan esetben, amikor nem létezik a várható érték (nincs abszolút konvergencia a definícióban)

3.1. Következmény. *Legyen A egy tetszőleges esemény. Végezzünk n független kísérletet, és legyen $\mathbb{I}_j$ annak indikátora, hogy a j. próbálkozásnál bekövetkezett-e ez az A esemény. Ekkor*

$$\frac{\mathbb{I}_1 + \ldots + \mathbb{I}_n}{n} \to \mathbb{P}(A)$$

teljesül 1 *valószínűséggel $n \to \infty$ esetén, azaz a relatív gyakoriság tart az esemény valószínűségéhez, ha a kísérletek függetlenek, és a számuk végtelenhez tart.*

Ez könnyen következik az előzőből, hiszen $X_j = \mathbb{I}_j$ független valószínűségi változók, azonos eloszlásúak (mindegyik $\mathbb{P}(A)$ valószínűséggel 1, különben 0), véges várható értékűek, hiszen

$$\mathbb{E}(X_j) = \mathbb{E}(\mathbb{I}_j) = \mathbb{P}(A) \cdot 1 + (1 - \mathbb{P}(A)) \cdot 0 = \mathbb{P}(A) < \infty,$$

és egyben ez lesz a limesz is.

4. Konvolúció

Bizonyos esetekben pontosan meg is határozhatjuk független valószínűségi változók összegének eloszlását, nem csak az átlag viselkedését.

4.1. Állítás. *Legyenek X és Y független, nemnegatív egész értékű valószínűségi változók. Ekkor az $X+Y$ valószínűségi változó eloszlását az alábbi módon határozhatjuk meg:*

$$\mathbb{P}(X + Y = k) = \sum_{l=0}^{k} \mathbb{P}(X = l)\mathbb{P}(Y = k - l) \qquad (k \geq 0).$$

Bizonyítás. Diszjunkt eseményekre való szétbontással, illetve a függetlenség definíciójának felhasználásával:

$$\mathbb{P}(X + Y = k) = \sum_{l=0}^{k} \mathbb{P}(X = l, Y = k - l) = \sum_{l=0}^{k} \mathbb{P}(X = l)\mathbb{P}(Y = k - l).$$

□

4.1. Poisson-eloszlások konvolúciója

4.2. Állítás. *Legyenek X és Y független Poisson-eloszlású valószínűségi változók, X paramétere λ, az Y paramétere μ. Ekkor az $X + Y$ valószínűségi változó is Poisson-eloszlású, paramétere $\lambda + \mu$, várható értéke és szórásnégyzete is $\lambda + \mu$.*

Bizonyítás. Legyen $k \geq 0$ tetszőleges. Ekkor a Poisson-eloszlás definíciója alapján

$$\begin{aligned}\mathbb{P}(X+Y=k) &= \sum_{l=0}^{k} \mathbb{P}(X=l)\mathbb{P}(Y=k-l) = \sum_{l=0}^{k} \frac{\lambda^l}{l!}e^{-\lambda} \cdot \frac{\mu^{k-l}}{(k-l)!}e^{-\mu} = \\ &= e^{-(\lambda+\mu)}\frac{1}{k!}\sum_{l=0}^{k}\frac{k!}{l!(k-l)!}\lambda^l\mu^{k-l} = \\ &= e^{-(\lambda+\mu)}\frac{1}{k!}\sum_{l=0}^{k}\binom{k}{l}\lambda^l\mu^{k-l} = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!},\end{aligned}$$

ahol az utolsó lépésben a binomiális tételt használtuk.

4.2. Nevezetes eloszlások összege

- X, Y független Poisson-eloszlásúak λ_1 és λ_2 paraméterrel $\Rightarrow$ $X + Y$ Poisson-eloszlású $\lambda_1 + \lambda_2$ paraméterrel;
- X, Y független binomiális eloszlásúak, n_1, illetve n_2 renddel, és azonos p paraméterrel $\Rightarrow$ $X + Y$ binomiális eloszlású $n_1 + n_2$ renddel és p paraméterrel, hiszen ez olyan, mintha összesen $n_1 + n_2$ független kísérletet tekintenénk.

Házi feladat november 12., péntek, 12:15-ig Sorsoljunk két, egymástól független 1000 elemű mintát, az első legyen egyenletes eloszlású a $[2, 4]$, a második a $[3, 5]$ intervallumon. A két minta legyen $X_1, X_2, \ldots, X_{1000}$, illetve $Y_1, \ldots, Y_{1000}$.

a) Készítsünk hisztogramot a két minta összege alapján, az $X_1 + Y_1, X_2 + Y_2, \ldots, X_{1000} + Y_{1000}$ megfigyelésből.

b) Ábrázoljuk k függvényében ugyanazon az ábrán az alábbi három mennyiséget:

$$\frac{X_1 + \ldots + X_k}{k}; \qquad \frac{Y_1 + \ldots + Y_k}{k}; \qquad \frac{X_1 + \ldots + X_k + Y_1 + \ldots + Y_k}{2k}.$$

Ez alapján mit gondonánk, mi lesz a várható értéke az X_1, Y_1, illetve $X_1 + Y_1$ valószínűségi változóknak? (Erre pontos definíció még nincs, mert ezek nem diszkrét valószínűségi változó, de az eddig a várható értékről tanult állítások nagy része ebben az általánosabb esetben is igaz lesz, ez alapján lehet tippelni.)