

7. Gyakorlat

Normális eloszlás és a Centrális Határeloszlás Tétel

Normális eloszlás

Definíció (Normális eloszlás) Az $X \sim \mathcal{N}(\mu, \sigma^2)$ valószínűségi változó normális eloszlást követ $\mu \in \mathbb{R}$ várható értékkel és $\sigma > 0$ szórással, amennyiben a sűrűségfüggvénye

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\left(-\frac{1}{2} \cdot \frac{(x - \mu)^2}{\sigma^2}\right).$$

Definíció (Standard normális eloszlás) Az $X \sim \mathcal{N}(0, 1)$ eloszlást standard normális eloszlásnak nevezzük. Ekkor a sűrűségfüggvényét és eloszlásfüggvényét az alábbi módon jelöljük

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \quad \text{és} \quad \Phi(x) = \int_{-\infty}^x \varphi(s) ds.$$

Ekkor egy nagyon hasznos azonosság:

$$\Phi(-x) = 1 - \Phi(x).$$

Standardizálás

Állítás Amennyiben $X \sim \mathcal{N}(\mu, \sigma^2)$ normális eloszlást követ, akkor minden lineáris transzformáltja normális eloszlást követ. Vagyis

$$X + b \sim \mathcal{N}(\mu + b, \sigma^2) \quad \text{és} \quad aX \sim \mathcal{N}(a \cdot \mu, a^2 \cdot \sigma^2).$$

Következmény (Standardizálás) Legyen $X \sim \mathcal{N}(\mu, \sigma^2)$. Ekkor

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Centrális Határeloszlás Tétel (CHT)

Tétel (Centrális Határeloszlás Tétel) Legyenek $X_1, X_2, X_3, \dots$ független és azonos eloszlású valószínűségi változók $\mu \in \mathbb{R}$ várható értékkel és $\sigma > 0$ szórással. Ekkor minden $x \in \mathbb{R}$ esetén amennyiben $S_n = X_1 + X_2 + \dots + X_n$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \leq x\right) = \Phi(x).$$

Következmény A CHT két speciális esete:

- Amennyiben $X_n \sim \text{Bin}(n, p)$, akkor

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x).$$

Vagyis $\text{Bin}(n, p)$ "jól közelíthető" $\mathcal{N}(np, np(1-p))$ eloszlással.

Ez igaz, mert X_n eloszlása megegyezik $Z_1 + Z_2 + \dots + Z_n$ eloszlásával, ahol Z_j -k független azonos eloszlásúak: $\mathbb{P}(Z_j = 1) = p$ és $\mathbb{P}(Z_j = 0) = 1 - p$. Vagyis $Z_j \sim \text{Ber}(p)$.

- Amennyiben $Y_j \sim \text{Poi}(\lambda)$, független azonos eloszlásúak, akkor $S_n = Y_1 + Y_2 + \dots + Y_n \sim \text{Poi}(n\lambda)$. Ekkor

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\lambda}{\sqrt{n\lambda}} \leq x\right) = \Phi(x).$$

Vagyis $\text{Poi}(n\lambda)$ "jól közelíthető" $\mathcal{N}(n\lambda, n\lambda)$ eloszlással.

Feladatok

1. Az A4-et tanuló diákok végső százalékpontos eredményei normális eloszlásúnak tekinthetők 65 várható értékkel és 15 szórással.
 - (a) Mennyi a valószínűsége, hogy egy véletlenszerű diák középeket kap (azaz, a pontja 60 és 70 pont közé esik)?
 - (b) Az átment diákok (azok, akik legalább 50 pontot értek el) hány százaléka kap középeket?

Megoldás Legyen X egy A4-et hallgató diák százalékos eredménye. Ekkor a feladat szerint

$$X \sim \mathcal{N}(65, 15^2).$$

- (a) A keresett valószínűség (mivel X folytonos) standardizálás után

$$\begin{aligned} \mathbb{P}(60 \leq X \leq 70) &= \mathbb{P}\left(\frac{60-65}{15} \leq \frac{X-60}{15} \leq \frac{70-65}{15}\right) = \mathbb{P}\left(-\frac{1}{3} \leq \underbrace{\frac{X-60}{15}}_{\sim \mathcal{N}(0,1)} \leq \frac{1}{3}\right) \\ &= \Phi\left(\frac{1}{3}\right) - \Phi\left(-\frac{1}{3}\right) = \Phi\left(\frac{1}{3}\right) - \left(1 - \Phi\left(\frac{1}{3}\right)\right) \\ &= 2\Phi\left(\frac{1}{3}\right) - 1 \approx 2\Phi(0.33) - 1 \approx 2 \cdot 0.6293 - 1 \approx 0.2586 \end{aligned}$$

- (b) A keresett valószínűség

$$\mathbb{P}(60 \leq X \leq 70 \mid X \geq 50) = \frac{\mathbb{P}(60 \leq X \leq 70, X \geq 50)}{\mathbb{P}(X \geq 50)} = \frac{\mathbb{P}(60 \leq X \leq 70)}{\mathbb{P}(X \geq 50)}.$$

A számlálót már kiszámoltuk. A nevező standardizálás után

$$\mathbb{P}(X \geq 50) = \mathbb{P}\left(\frac{X-60}{12} \geq \frac{50-65}{15}\right) = \mathbb{P}\left(\frac{X-60}{15} \geq 1\right) = 1 - \Phi(-1) = \Phi(1) \approx 0.8413.$$

Innen

$$\mathbb{P}(60 \leq X \leq 70 \mid X \geq 50) \approx \frac{0.2586}{0.8413} \approx 0.3074.$$

Vagyis körülbelül az átment diákok 30.74%-ka kap középeket.

2. Legyen A normális eloszlású valószínűségi változó 2 várható értékekkel és 3 szórással. Mennyi a valószínűsége, hogy lesz az $A \cdot x^2 + 2x + 1$ egyenletnek valós gyöke?

Megoldás Pontosan akkor lesz az egyenletnek valós gyöke, amennyiben a diszkriminánsa

$$D = 2^2 - 4 \cdot A \cdot 1 = 4 - 4 \cdot A$$

nemnegatív. Másfelől mivel D lineáris transzformáltja A -nak, így normális eloszlást követ. Továbbá

$$\mathbb{E}(D) = \mathbb{E}(4 - 4A) = 4 - 4\mathbb{E}(A) = 4 - 4 \cdot 2 = -4,$$

$$\mathbb{D}^2(D) = \mathbb{D}^2(4 - 4A) = 4^2 \mathbb{D}^2(A) = 16 \cdot 3^2 = 144.$$

Vagyis azt kaptuk, hogy

$$D \sim \mathcal{N}(-4, 144).$$

Ekkor a keresett valószínűség standardizálás után

$$\mathbb{P}(D \geq 0) = \mathbb{P}\left(\frac{D+4}{\sqrt{144}} \geq \frac{4}{\sqrt{144}}\right) = \mathbb{P}\left(\frac{D+4}{12} \geq \frac{1}{3}\right) = 1 - \Phi\left(\frac{1}{3}\right) \approx 1 - \Phi(0.33) = 1 - 0.6293 \approx 0.3707.$$

3. Egy BME-VIK évfolyamon 500 diák hallgat egy tárgyat. A vizsgadolgozat előtt konzultációt szerveznek. Előzetes felmérések szerint a hallgatók külön-külön, egymástól függetlenül 0,25 valószínűséggel jönnek el a konzultációra. Hány fős terem kell ahhoz, hogy a konzultációra érkező hallgatók 90%-os biztonsággal mind elférjenek a teremben?

Megoldás Legyen X azon diákok száma, akik elmennek a konzultációra. Ekkor a feladat szövege szerint

$$X \sim \text{Bin}(500, 0.25).$$

Legyen T a terem befogadóképessége. Ekkor azt a T -t keressük, amire

$$0.9 \leq \mathbb{P}(X \leq T).$$

Ekkor alkalmazva a CHT-t egy kó közelítés

$$\mathbb{P}(X \leq T) = \mathbb{P}\left(\frac{X - 500 \cdot 0.25}{\sqrt{500 \cdot 0.25 \cdot 0.75}} \leq \frac{T - 500 \cdot 0.25}{\sqrt{500 \cdot 0.25 \cdot 0.75}}\right) \approx \Phi\left(\frac{T - 125}{9.6825}\right).$$

Vagyis nekünk kell

$$\begin{aligned} 0.9 &\leq \Phi\left(\frac{T - 125}{9.6825}\right) \\ \Phi^{-1}(0.9) &\leq \frac{T - 125}{9.6825} \\ 1.28 &\leq \frac{T - 125}{9.6825} \\ 137.39 &\leq T. \end{aligned}$$

Így egy legalább 138 fős terem megfelelő lesz.

4. Egy gyár kétfajta érmét gyárt: egy igazságosat, és egy hamisat ami 55% eséllyel mutat fejet. Van egy ilyen érménk, de nem tudjuk igazságos-e vagy pedig hamis. Ennek eldöntésére a következő statisztikai tesztet hajtjuk végre: feldobjuk az érmét 1000-szer, ha legalább 525-ször fejet mutat, akkor hamisnak nyilvánítjuk, ha 525-nél kevesebb fej lesz a dobások között, akkor az érmét igazságosnak tekintjük. Mennyi a valószínűsége, hogy a tesztünk téved abban az esetben, ha az érme igazságos volt? És ha hamis volt?

Megoldás Legyen X a fejek száma az igazságos érmével. Ekkor a feladat szövege szerint

$$X \sim \text{Bin}\left(1000, \frac{1}{2}\right).$$

Ekkor az érmét az alábbi valószínűséggel nyilvánítom tévesen hamisnak

$$\mathbb{P}(X \geq 525).$$

Ekkor a CHT alapján egy jó becslés

$$\mathbb{P}(X \geq 525) = \mathbb{P}\left(\frac{X - 1000 \cdot \frac{1}{2}}{\sqrt{1000 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \geq \underbrace{\frac{525 - 1000 \cdot \frac{1}{2}}{\sqrt{1000 \cdot \frac{1}{2} \cdot \frac{1}{2}}}}_{\approx 1.58}\right) \approx 1 - \Phi(1.58) \approx 1 - 0.9429 \approx 0.0571.$$

Legyen Y a fejek száma a hamis érmével. Ekkor a feladat szövege szerint

$$Y \sim \text{Bin}\left(1000, \frac{11}{20}\right).$$

Ekkor az érmét az alábbi valószínűséggel nyilvánítom tévesen igazságosnak

$$\mathbb{P}(Y < 525).$$

Ekkor a CHT alapján egy jó becslés

$$\mathbb{P}(Y < 525) = \mathbb{P}\left(\frac{Y - 1000 \cdot \frac{11}{20}}{\sqrt{1000 \cdot \frac{11}{20} \cdot \frac{9}{20}}} < \underbrace{\frac{525 - 1000 \cdot \frac{11}{20}}{\sqrt{1000 \cdot \frac{11}{20} \cdot \frac{9}{20}}}}_{\approx -1.59}\right) \approx \Phi(-1.59) = 1 - \Phi(1.59) \approx 1 - 0.9441 \approx 0.0559.$$

5. Egy részfeladat elkészítéséhez szükséges idő mennyisége $\text{Uni}(1, 3)$ eloszlású, ahol az egyes részfeladatokhoz szükséges idők mennyiségei (együttesen) függetlenek. A részfeladatokat egymás után végezzük. Ezer részfeladat esetén mi a valószínűsége, hogy kevesebb, mint 1984 időegység alatt elkészülünk?

Megoldás Legyen $X_1, X_2, X_3 \dots$ az adott sorszámú részfeladat elvégzéséhez szükséges idő. Ekkor a feladat szerint független azonos $\text{Uni}(1, 3)$ eloszlásúak. Ekkor a várható értékük

$$\mu = \mathbb{P}(X_j) = \frac{1 + 3}{2} = 2,$$

továbbá a szórásuk

$$\sigma = \mathbb{D}(X_j) = \frac{3 - 1}{\sqrt{12}} = \frac{\sqrt{3}}{3}.$$

Ekkor az 1 000 részfeladat elvégzéséhez szükséges idő

$$S_{1000} = X_1 + X_2 + \dots + X_{1000}.$$

Ekkor a keresett valószínűsége a CHT miatt egy jó közelítés

$$\begin{aligned} \mathbb{P}(S_{1000} < 1984) &= \mathbb{P}\left(\frac{S_{1000} - 1000 \cdot 2}{\sqrt{1000 \cdot \frac{\sqrt{3}}{3}}} < \underbrace{\frac{1984 - 1000 \cdot 2}{\sqrt{1000 \cdot \frac{\sqrt{3}}{3}}}}_{\approx -0.88}\right) \\ &\approx \Phi(-0.88) = 1 - \Phi(0.88) = 1 - 0.8106 = 0.1894. \end{aligned}$$

6. Egy kockát folyamatosan feldobunk addig, amíg a dobások összege meg nem haladja a 300-at. Becsüljük meg annak a valószínűségét, hogy legalább 80 dobásra van ehhez szükségünk!

Megoldás Legyenek $X_1, X_2, X_3 \dots$ a megfelelő sorszámú kockadobás eredményei. Ekkor függetlenek és azonos eloszlásúak a feladat szerint. Továbbá a várható értékük

$$\mu = \mathbb{E}(X_j) = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{7}{3}.$$

Továbbá a második momentumuk

$$\mathbb{E}(X_j^2) = \sum_{k=1}^6 k^2 \frac{1}{6} = \frac{91}{6}.$$

amiből a szórásuk

$$\sigma = \sqrt{\frac{91}{6} - \left(\frac{7}{3}\right)^2} = \frac{\sqrt{479}}{6}.$$

Az első 80 dobás összege

$$S_{80} = X_1 + X_2 + \dots + X_{80}.$$

Ekkor a keresett valószínűség egy jó közelítése a CHT miatt

$$\begin{aligned} \mathbb{P}(S_{80} > 300) &= 1 - \mathbb{P}(S_{80} \leq 300) = 1 - \mathbb{P}\left(\frac{S_{80} - 80 \cdot \frac{7}{3}}{\sqrt{80 \cdot \frac{\sqrt{479}}{6}}} \leq \underbrace{\frac{300 - 80 \cdot \frac{7}{3}}{\sqrt{80 \cdot \frac{\sqrt{479}}{6}}}}_{\approx 0.02}\right) \\ &\approx 1 - \Phi(0.02) \approx 1 - 0.508 = 0.492. \end{aligned}$$

7. Mennyi a valószínűsége annak, hogy 50 db független azonos eloszlású valószínűségi változó összege a $[0, 30]$ intervallumba esik, ha egy ilyen valószínűségi változó eloszlása a $[0, 1]$ intervallumon
- (a) egyenletes?
- (b) a sűrűségfüggvénye $f(x) = 2x$?

Megoldás Legyenek a valószínűségi változóink $X_1, X_2, \dots, X_{50}$, továbbá az összegük

$$S_{50} = X_1 + X_2 + \dots + X_{50}.$$

Ekkor a keresett valószínűség, mivel $X_1, X_2, \dots, X_{50}$ nemnegatívak

$$\mathbb{P}(S_{50} \leq 30).$$

- (a) Ekkor az összegben szereplő tagok várható értéke

$$\mu = \mathbb{E}(X_j) = \frac{1+0}{2} = \frac{1}{2},$$

továbbá a szórása

$$\sigma = \mathbb{D}(X_j) = \frac{1-0}{\sqrt{12}} = \frac{1}{\sqrt{12}}.$$

Ekkor a keresett valószínűség egy jó közelítése a CHT alapján

$$\mathbb{P}(S_{50} \leq 30) = \mathbb{P}\left(\frac{S_{50} - 50 \cdot \frac{1}{2}}{\sqrt{50} \cdot \frac{1}{\sqrt{12}}} \leq \underbrace{\frac{30 - 50 \cdot \frac{1}{2}}{\sqrt{50} \cdot \frac{1}{\sqrt{12}}}}_{\approx 2.45}\right) \approx \Phi(2.45) \approx 0.9929.$$

- (b) Ekkor az összegben szereplő tagok várható értéke

$$\mu = \mathbb{E}(X_j) = \int_0^1 x \cdot 2x \, dx = \frac{2}{3}.$$

A második momentumuk

$$\mathbb{E}(X_j^2) = \int_0^1 x^2 \cdot 2x \, dx = \frac{1}{2},$$

amiből a szórásuk

$$\sigma = \mathbb{D}(X_j) = \sqrt{\frac{1}{2} - \left(\frac{2}{3}\right)^2} = \frac{1}{\sqrt{18}}.$$

Ekkor a keresett valószínűség egy jó közelítése a CHT alapján

$$\mathbb{P}(S_{50} \leq 30) = \mathbb{P}\left(\frac{S_{50} - 50 \cdot \frac{2}{3}}{\sqrt{50} \cdot \frac{1}{\sqrt{18}}} \leq \frac{30 - 50 \cdot \frac{2}{3}}{\sqrt{50} \cdot \frac{1}{\sqrt{18}}}\right) = \mathbb{P}\left(\frac{S_{50} - 50 \cdot \frac{2}{3}}{\sqrt{50} \cdot \frac{1}{\sqrt{18}}} \leq -2\right) \approx \Phi(-2) \approx 0.0228.$$

8. Egy kereskedelmi rádió 100 zeneszámot ismételtet úgy, hogy a következő számot mindig a korábbiaktól függetlenül választja, azonos valószínűséggel a 100 lehetséges szám közül (így ugyanaz a szám akár egymás után többször is előfordulhat). A 100 szám közül az egyik Angéla kedvenc száma. Angéla bekapcsolja a rádiót, és addig hallgatja, amíg a kedvenc számát századszorra végig nem hallgatta, aztán kikapcsolja a rádiót. Közelítsük alkalmas approximációval annak a valószínűségét, hogy Angéla a rádió kikapcsolása előtt legalább 10 200 számot hallgat végig.

Megoldás Legyen $X_1, X_2, \dots, X_{100}$ a kedvenc zeneszám meghallgatása között lejátszott zeneszámoknak a száma (beleértve azt is amikor lejátszák végül a kedvenc zeneszámot). Ekkor a feladat szerint függetlenek és azonos eloszlásúak

$$X_j \sim \text{Geo}\left(\frac{1}{100}\right).$$

Ekkor az összesen végighallgatott zeneszámok száma

$$S_{100} = X_1 + X_2 + \dots + X_{100}.$$

A keresett valószínűség

$$\mathbb{P}(S_{100} \geq 10200).$$

Ekkor az összegben szereplő tagok várható értéke

$$\mu = \mathbb{E}(X_j) = \frac{1}{\frac{1}{100}} = 100,$$

továbbá a szórásuk

$$\sigma = \mathbb{D}(X_j) = \sqrt{\frac{1 - \frac{1}{100}}{\left(\frac{1}{100}\right)^2}} = 30\sqrt{11}.$$

Ekkor a keresett valószínűség egy jó közelítése a CHT alapján

$$\mathbb{P}(S_{100} \geq 10200) = \mathbb{P}\left(\frac{S_{100} - 100 \cdot 100}{\sqrt{100} \cdot 30\sqrt{11}} \geq \underbrace{\frac{10200 - 100 \cdot 100}{\sqrt{100} \cdot 30\sqrt{11}}}_{\approx 0.2}\right) \approx 1 - \Phi(0.2) \approx 1 - 0.508 = 0.492.$$