Valószínűségszámítás gyakorlat
2020. november 24.

1. Legyen $X \sim \text{Geom}(p)$. Igazoljuk, hogy teljesül rá a diszkrét örökifjú tulajdonság, azaz

$$\mathbf{P}(X > k + l | X > l) = \mathbf{P}(X > k).$$

Igazoljuk, hogy ez jellemzi is a geometriai eloszlást!

Megoldás. Ha X geometriai eloszlású, akkor valamilyen p-re

$$\mathbf{P}(X = k) = (1-p)^{k-1}p.$$

Ezért

$$\begin{aligned}\mathbf{P}(X > k) &= \sum_{i=k+1}^{\infty} \mathbf{P}(X = i) = \sum_{i=k+1}^{\infty} (1-p)^{i-1}p \\ &= p(1-p)^k \sum_{j=0}^{\infty} (1-p)^j = p(1-p)^k \frac{1}{1-(1-p)} = (1-p)^k.\end{aligned}$$

Tehát

$$\begin{aligned}\mathbf{P}(X > k + l | X > l) &= \frac{\mathbf{P}(X > k + l)}{\mathbf{P}(X > l)} \\ &= \frac{(1-p)^{k+l}}{(1-p)^l} = (1-p)^k = \mathbf{P}(X > k),\end{aligned}$$

amint állítottuk.

Megfordítva, tegyük fel, hogy

$$\mathbf{P}(X > k + l | X > l) = \mathbf{P}(X > k)$$

teljesül minden k, l pozitív egész számokra. Átrendezve

$$\mathbf{P}(X > k + l) = \mathbf{P}(X > k)\mathbf{P}(X > l).$$

Innen indukcióval kapjuk, hogy

$$\mathbf{P}(X > n) = (\mathbf{P}(X > 1))^n,$$

ahonnan

$$\begin{aligned}\mathbf{P}(X = n) = \mathbf{P}(X > n-1) - \mathbf{P}(X > n) &= (\mathbf{P}(X > 1))^{n-1} - (\mathbf{P}(X > 1))^n \\ &= (\mathbf{P}(X > 1))^{n-1}(1 - \mathbf{P}(X > 1)),\end{aligned}$$

azaz X geometriai eloszlású $p = 1 - \mathbf{P}(X > 1) = \mathbf{P}(X = 1)$ paraméterrel.

$\square$

2. Diszkrét örökifjúból folytonosat. Legyen $X_n \sim \text{Geom}(\lambda/n)$. Határozzuk meg X_n/n határeloszlását, azaz adjuk meg a

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{X_n}{n} \le x\right)$$

határértéket minden $x \in \mathbb{R}$ esetén!

Megoldás. Legyen $x > 0$ tetszőleges. Ekkor

$$\begin{aligned}
\mathbf{P}\left(\frac{X_n}{n} \le x\right) &= 1 - \mathbf{P}(X_n > nx) = 1 - \mathbf{P}(X_n > [nx]) \\
&= 1 - \left(1 - \frac{\lambda}{n}\right)^{[nx]} \\
&= 1 - \left(1 - \frac{\lambda}{n}\right)^{\frac{n}{\lambda}\cdot\frac{\lambda[nx]}{n}} \\
&\to 1 - e^{-\lambda x},
\end{aligned}$$

ami éppen a exponenciális eloszlás eloszlásfüggvénye. □

3. Az FC Barcelona passzolási hatékonysága 2019. áprilisában $p = 0{,}85$ (azaz egy passz ekkora valószínűséggel sikeres). Adjuk meg annak a valószínűségét, hogy a Liverpool elleni 525 passzból legalább 460 sikeres.

Megoldás. Ez a legklasszikusabb példa a de Moivre–Laplace-tétel alkalmazására. Jelölje S a sikeres passzok számát. Mivel összesen 525 passz van, és mindegyik passz egymástól függetlenül $0,85$ valószínűséggel sikeres, ezért S eloszlása binomiális $n = 525$ és $p = 0{,}85$ paraméterekkel. A de Moivre–Laplace-tétel szerint

$$\frac{S - np}{\sqrt{np(1-p)}}$$

közelítőleg standard normális eloszlású, azaz

$$\mathbf{P}\left(a \le \frac{S - np}{\sqrt{np(1-p)}} \le b\right) \approx \Phi(b) - \Phi(a).$$

A pontos állítás az, hogy ha $n \to \infty$, akkor az $\approx$ helyett $=$ van. Ha csak felső korlát van, akkor $a = -\infty$ és $\Phi(-\infty) = 0$, ha csak alsó, akkor $b = \infty$ és $\Phi(\infty) = 1$. A feladat kérdése a $\mathbf{P}(S \ge 460)$ valószínűség. Egyszerű átalakítással elérjük, hogy $(S - np)/\sqrt{np(1-p)}$ jelenjen meg. Az $np = 525 \cdot 0{,}85 \approx 446$ és $\sqrt{np(1-p)} \approx 8{,}2$ értékeket beírva

$$\mathbf{P}(S \geq 460) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} \geq \frac{460 - np}{\sqrt{np(1-p)}}\right) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} \geq 1{,}7\right)$$
$$\approx 1 - \Phi(1{,}7) = 1 - 0{,}955 = 0{,}045.$$

Tehát a keresett valószínűsége 0,045. □

4. Magyarországon, és mindenütt a világon, több fiúgyermek születik, mint lány. Az újszülöttek 52%-a fiú, 48%-a lány. Nevezzük *lányos napoknak/heteknek* azokat a napokat/heteket, amikor több lány születik, mint fiú.

- Szegeden naponta 9 gyermek születik. Mennyi a pontos valószínűsége, hogy Szegeden egy adott nap lányos nap? Milyen eloszlású az egy héten bekövetkezett lányos napok száma? Várhatóan hány lányos nap van egy héten?
- Budapesten naponta 100 gyermek születik. Mennyi a közelítő (normális közelítés, de Moivre–Laplace-tétel) valószínűsége, hogy Budapesten egy adott nap lányos nap?
- Egész Magyarországon egy héten 2500 gyermek születik. Mennyi a lányos hét bekövetkezésének közelítő valószínűsége? Várhatóan hány lányos hét lesz 2020-ban? Adjuk meg annak a közelítő valószínűségét (Poisson-közelítés), hogy legalább 3 lányos hét lesz 2020-ban.

Megoldás. (a) Egy nap 9 gyermek születik. Annak a valószínűsége, hogy egy gyermek lány 0,48. Ezek az események egymástól függetlenek. Ezért, ha X jelöli a lányok számát, akkor X binomiális eloszlású $n = 9$ és $p = 0,48$ paraméterekkel. Ha X a lányok száma, akkor $9 - X$ a fiúké, és pontosan akkor születik több lány, ha $X \geq 5$. Ennek a valószínűsége

$$\mathbf{P}(X \geq 5) = \sum_{k=5}^{9} \mathbf{P}(X = k) = \sum_{k=5}^{9} \binom{9}{k} (0,48)^k \, (0,52)^{9-k} = 0,45.$$

Tehát annak a valószínűsége, hogy egy nap lányos nap, 0,45. A egyes napokon történt születések egymástól függetlenek, ezért az egy héten bekövetkezett lányos napok száma binomiális eloszlást követ, $n_1 = 7$ és $p_1 = 0,45$ paraméterekkel. A binomiális várható értéke $n_1 \cdot p_1 = 3,15$. Várhatóan 3,15 lányos nap van egy héten.

(b) Jelölje most X_B a Budapesten egy napon született lányok számát. Ekkor X_B binomiális $n = 100$ és $p = 0,48$ paraméterekkel. Akkor lesz több lány, ha $X_B \geq 51$. A de Moivre–Laplace-tétel szerint (most $np = 48$ és $\sqrt{np(1-p)} = 5$)

$$\begin{aligned}\mathbf{P}\left(X_B \geq 51\right) &= \mathbf{P}\left(\frac{X_B - np}{\sqrt{np(1-p)}} \geq \frac{51 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \mathbf{P}(Z > 0,6) = 1 - \Phi(0,6) = 0,27.\end{aligned}$$

Itt Z standard normális véletlen változó. Tehát közelítőleg 27% a lányos nap valószínűsége. A pontos valószínűség

$$\sum_{k=51}^{100} \binom{100}{k} (0,48)^k (0,52)^{100-k} = 0,30.$$

Egy héten várhatóan $0,27 \cdot 7 = 1,89$ lányos nap van.

(c) Ez ugyanaz, mint az előbb csak $n = 2500$ és $p = 0,48$ paraméterekkel. Legyen X_M a Magyarországon egy héten született lányok száma. Akkor van lányos hét, ha $X_M \geq 1251$. Most $np = 1200$, $\sqrt{np(1-p)} = 25$, így

$$\begin{aligned}\mathbf{P}\left(X_M \geq 1251\right) &= \mathbf{P}\left(\frac{X_M - np}{\sqrt{np(1-p)}} \geq \frac{1251 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \mathbf{P}(Z > 2,04) = 1 - \Phi(2,04) = 0,02.\end{aligned}$$

Tehát a magyarországi lányos hét valószínűsége 0,02. Egy évben 52 hét van, így az egy évben levő lányos hetek száma binomiális eloszlású $n = 52$ és $p = 0,02$ paraméterekkel. Mivel p kicsi n pedig nagy, ezt közelíthetjük Poisson-eloszlással. Mivel a várható érték megegyezik a paraméterrel, ezért $\lambda = 52 \cdot 0,02 = 1,04$. Legyen tehát Y Poisson-eloszlású $\lambda = 1,04$ paraméterrel. Annak a valószínűsége, hogy legalább 3 lányos hét lesz egy évben

$$\mathbf{P}(Y \geq 3) = 1 - \mathbf{P}(Y \leq 2) = 1 - \left(e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2}{2} e^{-\lambda}\right) = 0,09.$$

$\square$

5. Chicago és Los Angeles között két vasútvonal van, melyek mindegyikén egy-egy vonat közlekedik. Mindkét vonat egyidőben indul, lényegében egyformán kényelmes és k személyes. Tegyük fel, hogy 1000 utas egymástól függetlenül $1/2 - 1/2$ valószínűséggel választ vonatot. Legalább mekkora legyen az ülőhelyek k száma, hogy 0,01-nél kisebb legyen annak a valószínűsége, hogy lesz olyan utas, akinek nem jut ülőhely?

Megoldás. Ez is de Moivre–Laplace. Jelölje S az A társasággal utazók számát. Ekkor a B-vel utazók száma $1000 - S$. Mivel $n = 1000$ ember egymástól függetlenül $1/2 - 1/2$ valószínűséggel dönt A ill. B mellett, ezért S binomiális eloszlású $n = 1000$ és $p = 1/2$ paraméterekkel. Legyen k az ülőhelyek száma (mindkét vonaton). Az, hogy lesz olyan, akinek nem jut hely, azt jelenti, hogy vagy az A társaságnál túl sokan vannak, vagy a B-nél. Azaz, vagy $S > k$ (A vonat betelik) vagy $1000 - S > k$ (B vonat betelik). A kérdés a legkisebb olyan k érték, melyre

$$\mathbf{P}(S > k \text{ vagy } S < 1000 - k) \leq 0{,}01.$$

Vegyük észre, hogy $k \geq 500$ kell legyen, és ekkor csak az egyik vonat telhet be, azaz a fenti valószínűségben szereplő két esemény egymást kizáró, ahonnan

$$\mathbf{P}(S > k \text{ vagy } S < 1000 - k) = \mathbf{P}(S > k) + \mathbf{P}(S < 1000 - k).$$

Használjuk a de Moivre–Laplace-tételt. Beírjuk az $np = 500$, $\sqrt{np(1-p)} = 5\sqrt{10} \approx 15{,}8$ értékeket. Ekkor

$$\mathbf{P}(S > k) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} > \frac{k - np}{\sqrt{np(1-p}}\right) \approx 1 - \Phi((k - 500)/15{,}8),$$

és ugyanígy (csak felhasználjuk, hogy $\Phi(-x) = 1 - \Phi(x)$ ha $x > 0$, és hogy $500 - k \leq 0$)

$$\begin{aligned}\mathbf{P}(S < 1000 - k) &= \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} < \frac{1000 - k - np}{\sqrt{np(1-p}}\right)\\ &\approx 1 - \Phi((k - 500)/15{,}8).\end{aligned}$$

Azaz, keressük azt a legkisebb k értéket melyre

$$2\left[1 - \Phi((k - 500)/15{,}8)\right] \leq 0{,}01.$$

Átrendezve,

$$\Phi((k - 500)/15{,}8) \geq 0{,}995.$$

Kikeressük a normális eloszlás táblázatából, hogy hol veszi fel Φ a $0{,}995$ értéket. Ez $2{,}57$, azaz

$$\frac{k - 500}{15{,}8} \geq 2{,}57,$$

amit átrendezve

$$k \geq 540{,}6.$$

Mivel k egész, ezért 541 a legkisebb olyan k, amire a feltétel teljesül. □

6. Budapesten meg akarják állapítani a dohányosok p arányát. Ehhez kiválasztanak n egyént úgy, hogy minden választásnál mindenki ugyanakkora valószínűséggel kerül kiválasztásra, és csak ezek közt nézik meg a dohányosok k számát. Legalább mekkora legyen az n, hogy a kapott $p' = k/n$ arány legalább 0,95 valószínűséggel legfeljebb 0,005 hibával közelítse a valódi p arányt, akármi is $p \in (0,1)$?

Megoldás. Ez már érdekesebb feladat, ugyanis semmi nincs megadva. Világos, hogy a feladat nagyon fontos, ugyanis a közvéleménykutatásokhoz pontosan ilyen típusú kérdést kell feltenni. Van valami ismeretlen valószínűség p, ami azt mutatja meg, hogy az emberek ilyen aránya szavazna az A párt jelöltjére. Nem tudjuk mi a p, de erről szeretnénk valamit mondani. Hány embert kell megkérdezni, hogy valami okosat mondhassunk?

Jelölje S a dohányzók számát a megkérdezettek között. Ekkor S binomiális eloszlású véletlen változó n (meghatározandó, de ismert) és p (ismeretlen) paraméterekkel. Világos, hogy az ismeretlen p értékre az S/n becslést adjuk. Elég összetett a kérdés, kicsit el kell rajta gondolkodni. A becslés hibája $|S/n - p|$. Azt akarjuk, hogy ez nagy valószínűséggel (0,95) kicsi legyen (0,005-nél kisebb), azaz olyan n értéket keresünk, amire

$$\mathbf{P}\left(\left|\frac{S}{n} - p\right| < 0{,}005\right) \geq 0{,}95.$$

(Annak a valószínűsége, hogy a hiba 0,005-nél kisebb, legalább 0,95.) A neheze megvan. Használjuk a de Moivre–Laplace-tételt. Eszerint

$$\mathbf{P}\left(a \leq \frac{S - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a).$$

Tehát be kell erőltetni az $(S - np)/\sqrt{np(1-p)}$ kifejezést. Tegyük meg:

$$\begin{aligned}
\mathbf{P}\left(\left|\frac{S}{n} - p\right| < 0{,}005\right) &= \mathbf{P}\left(\left|\frac{S - np}{n}\right| < 0{,}005\right) \\
&= \mathbf{P}\left(\left|\frac{S - np}{\sqrt{np(1-p)}}\right| < 0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&\approx \Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= 2\Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1.
\end{aligned}$$

Annyit használtunk, hogy $|x| \leq a$ pontosan akkor, ha $-a < x < a$, és hogy $\Phi(-x) = 1 - \Phi(x)$. Ezek szerint az kell, hogy

$$2\Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0{,}95,$$

azaz

$$\Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0{,}975.$$

A táblázatból kikeresve azt kapjuk, hogy

$$0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}} \geq 1{,}96.$$

Átrendezve

$$n \geq 392^2 \cdot p(1-p). \qquad (*)$$

Na de nem ismerjük p értékét. Úgy kell n-et választani, hogy a fenti egyenlőtlenség minden p-re igaz legyen. Tehát válasszuk p-t úgy, hogy a jobb oldal maximális legyen. Ez $p = 1/2$-nél van, értéke $1/4$. Tehát, ha

$$n \geq 396^2\frac{1}{4} = 38416,$$

akkor $(*)$ teljesül minden $p \in [0,1]$ esetén. $\square$