

A november 23.-i gyakorlat témája

Felidézek néhány eredményt arról, hogy hogyan tudjuk kiszámítani független (és nemcsak független) valószínűségi változók összegének a várható értékét és szórásnégyzetét. Ezenkívül ismertetem a valószínűségszámítás egyik legfontosabb eredményét, a centrális határeloszlástételt, és néhány példát mutatok arra, hogy a megtanult módszerek hogyan teszik lehetővé a centrális határeloszlástétel alkalmazását fontos feladatok megoldásában.

Megfogalmazok először néhány eredményt, amelyek azt mondják ki, hogy a diszkrét valószínűségi változók összegének várható értékéről és szórásnégyzetéről szóló eredmények érvényesek az általános esetben is.

Tétel. Ha két ξ_1, ξ_2 valószínűségi változónak (amelyek ugyanazon a valószínűségi mezőn vannak definiálva) létezik várható értéke, c_1 és c_2 két valós szám, akkor a $c_1\xi_1 + c_2\xi_2$ kifejezésnek is létezik várható értéke, és

$$E(c_1\xi_1 + c_2\xi_2) = c_1E\xi_1 + c_2E\xi_2.$$

Fontos megjegyzés: A fenti eredményben nem tettük fel, hogy a tekintett valószínűségi változók függetlenek.

Tétel. Legyenek $\xi_j, 1 \leq j \leq k$, valószínűségi változók ugyanazon a valószínűségi mezőn, amelyekre teljesül $E\xi_j^2 < \infty$ feltétel. Ekkor

$$\text{Var} \left(\sum_{j=1}^k \xi_j \right) = \sum_{j=1}^k \text{Var} \xi_j + \sum_{\substack{1 \leq j, l \leq k \\ j \neq l}} \text{Cov}(\xi_j, \xi_l) = \sum_{j=1}^k \text{Var} \xi_j + 2 \sum_{j=1}^{k-1} \sum_{l=j+1}^k \text{Cov}(\xi_j, \xi_l).$$

Speciálisan, ha a ξ_j valószínűségi változók függetlenek, akkor

$$\text{Var} \left(\sum_{j=1}^k \xi_j \right) = \sum_{j=1}^k \text{Var} \xi_j.$$

Az alábbi eredmény szintén fontos a valószínűségszámításban, de időhiány miatt valószínűleg nem tudjuk tárgyalni, és lehet, hogy ugyancsak időhiány miatt az előadáson sem kerül sor ennek tárgyalására. A biztonság kedvéért mégis közlöm.

Tétel független valószínűségi változók összegének a sűrűségfüggvényéről. Legyen ξ és η két független valószínűségi változó $f(\cdot)$ és $g(\cdot)$ sűrűségfüggvénnyel. Ekkor a $\xi + \eta$ összegnek is létezik sűrűségfüggvénye, és az az

$$f * g(x) = \int_{-\infty}^{\infty} f(u)g(x-u) du, \quad -\infty < x < \infty$$

függvény.

Rátérek a valószínűségszámítás legfontosabb eredményének a centrális (központi) határeloszlástételének az ismertetésére. Ennek érdekében először bevezetem a következő fogalmat.

Egy valószínűségi változó normalizáltjának a fogalma. Legyen ξ olyan valószínűségi változó, amelyre $E\xi^2 < \infty$. A ξ valószínűségi változó normalizáltja a $\bar{\xi} = \frac{\xi - E\xi}{\sqrt{\text{Var } \xi}}$ valószínűségi változó, azaz a ξ valószínűségi változónak az a lineáris transzformáltja, amelynek várható értéke nulla, szórásnégyzete pedig 1.

Jegyezzük meg, hogy ha ξ_1, \dots, ξ_n független valószínűségi változók, amelyekre $E\xi_j^2 < \infty$ minden $1 \leq j \leq n$ indexre, akkor az $S = \sum_{j=1}^n \xi_j$ összeg normalizáltja az

$$\bar{S} = \frac{\sum_{j=1}^n \xi_j - \sum_{j=1}^n E\xi_j}{\sqrt{\sum_{j=1}^n \text{Var } \xi_j}}$$

kifejezés.

Emlékeztetek arra is, hogy a $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $-\infty < x < \infty$, függvény sűrűségfüggvény, integrálja pedig a $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$, $-\infty < x < \infty$ függvény eloszlásfüggvény. A $\varphi(x)$ függvényt a standard normális sűrűségfüggvénynek, $\Phi(x)$ függvényt pedig a standard normális eloszlásfüggvénynek nevezik. Az, hogy $\varphi(x)$ sűrűségfüggvény azt jelenti, hogy egyrészt teljesül a (nyilvánvaló) $\varphi(x) \geq 0$ reláció minden $x > 0$ számra, másrészt $\int_{-\infty}^{\infty} \varphi(u) du = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 1$. Ez az azonosság nem nyilvánvaló állítás.

Centrális határeloszlástétel. Legyenek ξ_1, ξ_2, \dots , független, egyforma eloszlású valószínűségi változók, jelölje $S_n = \sum_{j=1}^n \xi_j$, $n = 1, 2, \dots$, e valószínűségi változók részlet-

összegeit. Ekkor az S_n valószínűségi változók $\frac{S_n - ES_n}{\sqrt{\text{Var } S_n}} = \frac{\sum_{j=1}^n \xi_j - \sum_{j=1}^n E\xi_j}{\sqrt{\sum_{j=1}^n \text{Var } \xi_j}}$ normalizáltjai

teljesítik a

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - ES_n}{\sqrt{\text{Var } S_n}} < x\right) = \Phi(x), \quad \text{minden } -\infty < x < \infty \text{ számra}$$

relációt, ahol $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$, a standard normális eloszlásfüggvény.

Fontos megjegyzés: Mivel a $\varphi(x)$ sűrűségfüggvény páros függvény, ezért $\Phi(-x) = \int_{-\infty}^{-x} \varphi(u) du = \int_{\infty}^x \varphi(u) du = 1 - \int_{-\infty}^x \varphi(u) du = 1 - \Phi(x)$. Ez azt jelenti, hogy a normális eloszlásfüggvény értékét elegendő pozitív számokra megadni. Ezért a normális eloszlásfüggvény táblázatokban a normális eloszlásfüggvény értékét nem adják meg negatív számokra.

Feladatok:

1. Jelölje a ξ valószínűségi változó eloszlásfüggvényét $F(x)$, és tegyük fel, hogy a ξ valószínűségi változónak létezik $f(x)$ sűrűségfüggvénye is. Legyen $a > 0$ és b két valós szám. Számoljuk ki az $\eta = a\xi + b$ valószínűségi változó $G(x)$ eloszlás és $g(x)$ sűrűségfüggvényét. Számoljuk ki a $\zeta = \xi^2$ valószínűségi változó $H(x)$ eloszlás és $h(x)$ sűrűségfüggvényét.

Megoldás: $G(x) = P(\eta < x) = P(a\xi + b < x) = P(\xi < \frac{x-b}{a}) = F(\frac{x-b}{a})$ minden $-\infty < x < \infty$ számra. A $g(x)$ sűrűségfüggvény a $G(x)$ eloszlásfüggvény deriváltja, azaz $g(x) = \frac{dG(x)}{dx} = \frac{dF(\frac{x-b}{a})}{dx} = \frac{1}{a} f(\frac{x-b}{a})$.

$H(x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = P(\xi < \sqrt{x}) - P(\xi \leq \sqrt{x}) = F(\sqrt{x}) - F(-\sqrt{x} + 0)$, ha $x \geq 0$, ahol $F(u + 0) = \lim_{h \rightarrow 0} F(u + h)$. (Az $F(u + 0)$ érték bevezetésére nincs szükség, ha az F eloszlásfüggvénynek van sűrűségfüggvénye, vagy általánosabban $F(x)$ folytonos függvény.) Továbbá, $H(x) = 0$, ha $x < 0$, mert $P(\xi^2 < x) = 0$ $x < 0$ esetén. Végül a $H(x)$ függvény deriválásának segítségével kapjuk, hogy $h(x) = \frac{1}{2\sqrt{x}} (f(\sqrt{x}) + f(-\sqrt{x}))$, ha $x \geq 0$, $h(x) = 0$, ha $x < 0$.

A következő feladat megfogalmazása előtt vezessük be az exponenciális eloszlásfüggvény fogalmát.

Exponenciális eloszlásfüggvény definíciója. Egy ξ valószínűségi változó exponenciális eloszlású λ paraméterrel, $\lambda > 0$, ha eloszlásfüggvénye, $F(x) = P(\xi < x) = 1 - e^{-\lambda x}$, ha $x \geq 0$, és $F(x) = P(\xi < x) = 0$, ha $x < 0$. Ezzel ekvivalens jellemzés: Egy valószínűségi változó exponenciális eloszlású $\lambda > 0$ paraméterrel, ha létezik $f(u)$ sűrűségfüggvénye, és az $f(u) = \lambda e^{-\lambda u}$, ha $u \geq 0$, $f(u) = 0$, ha $u \leq 0$ alakú.

- 2.) Számoljuk ki egy λ paraméterű ξ exponenciális eloszlású valószínűségi változó várható értékét és szórásnégyzetét.

Megoldás: Parciális integrálással kapjuk, hogy

$$E\xi = \int_0^{\infty} u \lambda e^{-\lambda u} du = \frac{1}{\lambda} \int_0^{\infty} u e^{-u} du = \frac{1}{\lambda} \left([-u e^{-u}]_0^{\infty} + \int_0^{\infty} e^{-u} du \right) = \frac{1}{\lambda},$$

$$\begin{aligned} E\xi^2 &= \int_0^{\infty} u^2 \lambda e^{-\lambda u} du = \frac{1}{\lambda^2} \int_0^{\infty} u^2 \lambda e^{-u} du \\ &= \frac{1}{\lambda^2} \left([u^2 e^{-u}]_0^{\infty} + \int_0^{\infty} 2u e^{-u} du \right) = \frac{2}{\lambda^2}. \end{aligned}$$

$$\text{Ezért } \text{Var } \xi = E\xi^2 - (E\xi)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Az alábbi feladatokban azt tárgyaljuk, hogyan tudjuk a centrális határeloszlástételt felhasználni minket érdeklő valószínűségek kiszámolására.

- 3.) A 2000. évben az Egyesült Államok Florida államában az elnökválasztáson rendkívül szoros eredmény született. 5 000 000 választó választott két párt, a republikánus és demokrata párt jelöltjei között. A két jelölt által szerzett szavazatok száma (egy adott időpontbeli felmérés szerint) mindössze 300 volt. Tegyük fel, hogy a választók egymástól függetlenül $\frac{1}{2}$ valószínűséggel választották valamelyik párt jelöltjét. E feltevés teljesülése estén mi annak a valószínűsége, hogy a két jelölt által összegyűjtött szavazatok különbsége nem haladja meg a háromszázat.

Megoldás: Vezessük be a következő ξ_j , $1 \leq j \leq 5\,000\,000$, valószínűségi változókat: $\xi_j = 1$, ha a j -ik választó a demokrata, $\xi_j = 0$, ha a j -ik választó a republikánus jelöltre szavaz. Ekkor $S = \sum_{j=1}^{5\,000\,000} \xi_j$ a demokrata, és $5\,000\,000 - S$ a republikánus jelöltre leadott szavazatok száma, és minket a $P(|2S - 5\,000\,000| \leq 300)$ valószínűség nagysága érdekel. Vegyük észre, hogy a ξ_j valószínűségi változók függetlenek, $E\xi_j = \frac{1}{2}$, $\text{Var } \xi_j = \frac{1}{4}$, ezért $ES_n = 2\,500\,000$, $\text{Var } S_n = \frac{1}{4} \cdot 5\,000\,000$, és a $P\left(\left|\frac{S_n - 2\,500\,000}{\sqrt{\frac{1}{4} \cdot 5\,000\,000}}\right| < x\right)$ valószínűségek kiszámítására alkalmazhatjuk a centrális határeloszlástételt. Ennek alapján

$$\begin{aligned} P(|2S - 5\,000\,000| \leq 300) &= P\left(\left|\frac{S - ES}{\sqrt{\text{Var } S}}\right| \leq \frac{150}{\sqrt{\frac{1}{4} \cdot 5\,000\,000}}\right) \\ &\sim \Phi\left(\frac{150}{\sqrt{\frac{1}{4} \cdot 5\,000\,000}}\right) - \Phi\left(-\frac{150}{\sqrt{\frac{1}{4} \cdot 5\,000\,000}}\right) \\ &= 2\Phi\left(\frac{6\sqrt{5}}{100}\right) - 1 \sim 2\Phi(0.124) - 1 \sim 0.1. \end{aligned}$$

4. Tekintsünk egy szabályos pénzdarab 10 000 egymás utáni (független) feldobásából származó fej-írás sorozatot. Adjunk becslést a centrális határeloszlástétel segítségével annak valószínűségére, hogy a fej-dobások számának eltérése a várt 5000 számtól legalább 100-zal, illetve legalább 200-zal eltér.

Megoldás: Vezessük be a következő ξ_j , $1 \leq j \leq 10\,000$ valószínűségi változókat: $\xi_j = 1$, ha a j -ik dobás eredménye fej, $\xi_j = 0$, ha a j -ik dobás eredménye írás. Ekkor ξ_j , $1 \leq j \leq 10\,000$ független valószínűségi változók, $E\xi_j = \frac{1}{2}$, $\text{Var } \xi_j = \frac{1}{4}$, és a $P\left(\left|\sum_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 100\right)$ és $P\left(\left|\sum_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 200\right)$ valószínűségekre kell

becslést adnunk. A centrális határeloszlástétel szerint $P\left(\frac{\sum_{j=1}^{10000} (\xi_j - E\xi_j)}{\sqrt{10000 \cdot \frac{1}{4}}} > u\right) \sim 1 -$

$\Phi(u)$. Innen kapjuk, hogy az első valószínűség kiszámításához az $u = \pm \frac{100}{\frac{1}{2} \cdot 100} = \pm 2$ értékeket kell tekinteni, és a vizsgált valószínűség közelítőleg $(1 - \Phi(2)) + \Phi(-2) = 2(1 - \Phi(2)) \sim 2(1 - 0.97720) = 0.0456$. A második valószínűség hasonlóan körülbelül $2(1 - \Phi(4)) \sim 0$, (az első 4 tizedesjegy 0).

- 5.) Egy szabályos dobókockát feldobunk 1200 alkalommal egymástól függetlenül, és összeadjuk a páros értékű dobások eredményét. Adjunk jó közelítő becslést a centrális határeloszlástétel és egy normális eloszlástáblázat segítségével arra, hogy ez az összeg 2280 és 2500 közé esik.

Megoldás: Vezessük be a következő ξ_j , $1 \leq j \leq 1200$, valószínűségi változókat: $\xi_j = 2$, ha a j -ik dobás eredménye 2, $\xi_j = 4$, ha a j -ik dobás eredménye 4, $\xi_j = 6$, ha a j -ik dobás eredménye 6, $\xi_j = 0$, ha a j -ik dobás eredménye 1, 3 vagy 5. Ekkor a $P\left(2280 \leq \sum_{j=1}^{1200} \xi_j \leq 2500\right)$ valószínűséget kell jól megbecsülnünk. Vegyük észre, hogy $E\xi_j = \frac{1}{6}(2 + 4 + 6) = 2$, $\text{Var} \xi_j = \frac{1}{6}(4 + 16 + 36) - 4 = \frac{16}{3}$. Innen a centrális határeloszlástétel alapján

$$P\left(2280 \leq \sum_{j=1}^{1200} \xi_j \leq 2500\right) = P\left(\frac{-120}{\sqrt{1200 \cdot \frac{16}{3}}} \leq \frac{\sum_{j=1}^{1200} \xi_j - \sum_{j=1}^{1200} E\xi_j}{\sqrt{\sum_{j=1}^{1200} \text{Var} \xi_j}} \leq \frac{100}{\sqrt{1200 \cdot \frac{16}{3}}}\right) \\ \sim \Phi(1.25) - \Phi(-1.5) = 0.8944 + 0.9322 - 1 = 0.8266.$$

- 6.) Legyen birtokunkban 100 lámpa, amelyek mindegyike egymástól független időtartamig működik, élettartamuk pedig exponenciális eloszlású $\lambda = \frac{1}{10}$ paraméterrel. (A lámpák élettartamának exponenciális eloszlása természetes feltételezés.) Egy termet bevilágítunk ezen lámpák valamelyikével, majd amikor az kiegészített új lámpát használunk fel. Adjunk jó becslést arra, hogy a lámpák összélettartama legalább 1150 óra.

Megoldás: Jelölje ξ_j a j -ik lámpa élettartamát, $1 \leq j \leq 100$. Ekkor a $P(\xi_1 + \dots + \xi_{100} > 1150)$ valószínűségre kell jó becslést adnunk, ahol az összegben független exponenciális eloszlású valószínűségi változók szerepelnek $\lambda = \frac{1}{10}$ paraméterrel. Vezessük be az $\eta = \xi_1 + \dots + \xi_{100}$ jelölést.

Kiszámoltuk, hogy jelen esetben $E\eta = mE\xi_1 = \frac{m}{\lambda} = 1000$, $\text{Var} \eta = \frac{m}{\lambda^2} = 10000$ ($m = 100$ és $\lambda = \frac{1}{10}$ választással). Ezért a centrális határeloszlástétel szerint $\frac{\eta - E\eta}{\sqrt{\text{Var} \eta}} = \frac{\eta - 1000}{100}$ jó közelítéssel standard normális eloszlású valószínűségi változó,

és $P(\xi_1 + \dots + \xi_{100} > 1150) = P\left(\frac{\eta - E\eta}{\sqrt{\text{Var} \eta}} > 1.5\right) \sim 1 - \Phi(1.5)$.

- 7.) Vegyünk egy olyan pénzdarabot, amely $\frac{2}{3}$ valószínűséggel esik a fej és $\frac{1}{3}$ valószínűséggel az írás oldalra. Ezt a pénzdarabot annyiszor dobjuk fel, ahány dobás után megjelenik 1200 fej dobás. Mi annak a valószínűsége, hogy az elvégzett dobások száma 1680 és 1830 közé esik? Adjunk erre a valószínűségre jó közelítő becslést.

Megoldás: Az elvégzett dobások száma egy η negatív binomiális eloszlású valószínűségi változó $n = 1200$ és $p = \frac{2}{3}$ paraméterekkel, azaz $P(\eta = k + n) = \binom{n+k-1}{n-1} (1-p)^k p^n$, $p = \frac{2}{3}$, és $n = 1200$ paraméterrel. Egy ilyen valószínűségi változónak ki lehet számolni a pontos eloszlását, azaz azt, hogy milyen értéket milyen valószínűséggel vesz fel. Elvileg, ez lehetőséget ad a kívánt valószínűség kiszámítására egy bonyolult összeg kiszámításának a segítségével. Ennél hasznosabb becslést tudunk kapni a következő érvelés segítségével, amely a kívánt valószínűséget jó pontossággal kiszámítja a centrális határeloszlástétel segítségével.

Jelölje ξ_j , $2 \leq j \leq 1200$, a $j - 1$ -ik és j -ik fejdobás közötti dobások számát (a j -ik fejdobást beleszámítjuk a $j - 1$ -iket viszont nem számítjuk bele e dobások közé), és legyen ξ_1 az első fejdobásig (ezt is beleszámítva) elvégzett dobások száma. Ekkor a ξ_j valószínűségi változók függetlenek, negatív binomiális eloszlásúak $n = 1$, $p = \frac{2}{3}$ paraméterekkel, és minket a $P(1680 < \xi_1 + \dots + \xi_{1200} < 1830)$ valószínűség érdekel. Megmutattuk korábban, hogy $E\xi_j = \frac{1}{p} = \frac{3}{2}$, $\text{Var } \xi_j = \frac{1-p}{p^2} = \frac{3}{4}$. Ezért a centrális határeloszlástétel alapján $\eta = \xi_1 + \dots + \xi_{1200}$ jelöléssel minket a

$$P\left(-4 < \frac{\eta - 1200E\xi_1}{\sqrt{1200\text{Var } \xi_1}} < 1\right)$$

valószínűség érdekel. A centrális határeloszlástétel alapján

$$P\left(-4 < \frac{\eta - 1200E\xi_1}{\sqrt{1200\text{Var } \xi_1}} < 1\right) \sim \Phi(1) + \Phi(4) - 1 \sim \Phi(1).$$

Házi feladat

Egy szabályos dobókockát és egy szabályos érmét feldobunk 3300 alkalommal egymástól függetlenül. (Az érme és kockadobások eredményei is függetlenek egymástól.) Ha a kockadobás eredménye páros és az érme a fej oldalra esett, akkor annyi forintot nyerünk, amennyi a kockadobás eredménye. Ha az érme az írás oldalra esett vagy a kockadobás eredménye páratlan szám, akkor nem nyerünk, és nem is veszünk semmit. Mi a valószínűsége annak, hogy az össznyereményünk 3190 és 3520 forint közé esik? Adjunk erre jó közelítő becslést a centrális határeloszlástétel és egy normális eloszlástáblázat segítségével.