Az április 25.-i gyakorlat feladatai

A valószínűségszámítás legfontosabb eredménye a centrális (központi) határeloszlástétel. Megbeszéljük ennek tartalmát és olyan feladatokat fogunk tekinteni, amelyekben ezt az eredményt alkalmazzuk.

A centrális határeloszlástétel tétel megfogalmazása előtt vezessünk be egy az irodalomban gyakran használt fogalmat.

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

Jegyezzük meg, hogy ha $\xi_1, \dots, \xi_n$ független valószínűségi változók, amelyekre $E\xi_j^2 < \infty$ minden $1 \le j \le n$ indexre, akkor az $S = \sum\limits_{j=1}^n \xi_j$ összeg normalizáltja az

$$\bar{S} = \frac{\sum\limits_{j=1}^n \xi_j - \sum\limits_{j=1}^n E\xi_j}{\sqrt{\sum\limits_{j=1}^n \operatorname{Var}\xi_j}}$$

kifejezés.

Centrális határeloszlástétel. *Legyenek $\xi_1, \xi_2, \dots,$ független, egyforma eloszlású valószínűségi változók, jelölje $S_n = \sum\limits_{j=1}^n \xi_j$, $n = 1, 2, \dots,$ e valószínűségi változók részletösszegeit. Ekkor az S_n valószínűségi változók $\dfrac{S_n - ES_n}{\sqrt{\operatorname{Var}S_n}} = \dfrac{\sum\limits_{j=1}^n \xi_j - \sum\limits_{j=1}^n E\xi_j}{\sqrt{\sum\limits_{j=1}^n \operatorname{Var}\xi_j}}$ normalizáltjai teljesítik a*

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

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

Megjegyzés: Az előbb kimondott tétel valójában speciális esete az általános centrális határeloszlástételnek, amely hasonló eredményt mond ki független, de nem feltétlenül azonos eloszlású valószínűségi változók normalizált részletösszegeire nagyon általános feltételek mellett.

Feladatok:

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

Megoldás: Vezessük be a következő ξ_j, $\le j \le 10\ 000$ valószínűségi változókat, amelyekre $\xi_j = 1$, ha a j-ik dobás eredménye fej, $\xi_j = 0$, ha a j-ik dobás eredménye írás. Ekkor ξ_j, $1 \le j \le 10\ 000$ független valószínűségi változók, $E\xi_j = \frac{1}{2}$, $\operatorname{Var} \xi_j = \frac{1}{4}$, és a $P\left(\left|\sum\limits_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 100\right)$ és $P\left(\left|\sum\limits_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 200\right)$ valószínűségekre kell becslést adnunk. A Csebisev egyenlőtlenség az első valószínűségre a

$$P\left(\left|\sum_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 100\right) \le \frac{10000 \cdot \operatorname{Var} \xi_1}{100^2} = \frac{1}{4},$$

a második valószínűségre pedig a

$$P\left(\left|\sum_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 200\right) \le \frac{10000 \cdot \operatorname{Var} \xi_1}{200^2} = \frac{1}{16}$$

első becslést adja.

A centrális határeloszlástétel szerint $P\left(\frac{\sum\limits_{j=1}^{10000} (\xi_j - E\xi_j)}{\sqrt{10000 \cdot \frac{1}{4}}} > u\right) \sim 1 - \Phi(u)$. 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). (A Csebisev egyenlőtlenség 0.25 illetve 0.0625 felső becslést adta ezekre a valószínűségekre.)

2.) A 2000. évben az Egyesült Államok Florida államában rendkívül szoros eredmény született. 5 000 000 választó 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 egymástól független ül $\frac{1}{2}$ valószínűséggel választották valamelyik párt jelöltjét. E feltevés teljesülése estén mi annak a valószínűsége, hogy a két jelölt által összegyüjtött szavazatok különbsége nem haladja meg a háromszázat.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 5\ 000\ 000$, 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 $S = \sum\limits_{j=1}^{5\ 000\ 000} \xi_j$ a demokrata, és $5\ 000\ 000 - S$ a republikánus jelöltre leadott szavazatok száma, és minket a $P(|2S - 5\ 000\ 000| \le 300)$

valószínűség nagysága érdekel. Vegyük észre, hogy a ξ_j valószínűségi változók függetlenek, $E\xi_j = \frac{1}{2}$, $\operatorname{Var}\xi_j = \frac{1}{4}$, ezért $ES_n = 2\ 500\ 000$, $\operatorname{Var} S_n = \frac{1}{4} \cdot 5\ 000\ 000$, 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

$$
\begin{aligned}
P(|2S - 5\ 000\ 000| \le 300) &= P\left(\left|\frac{S - ES}{\sqrt{\operatorname{Var} S}}\right| \le \frac{150}{\sqrt{\frac{1}{4}\cdot 5\ 000\ 000}}\right) \\
&\sim \Phi\left(\frac{150}{\sqrt{\frac{1}{4}\cdot 5000000}}\right) - \Phi\left(-\frac{150}{\sqrt{\frac{1}{4}\cdot 5000000}}\right) \\
&= 2\Phi\left(\frac{6\sqrt{5}}{100}\right) - 1 \sim 2\Phi(0.124) - 1 \sim 0.1.
\end{aligned}
$$

3. 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, $1 \le j \le 1200$, 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 $P\left(2280 \le \sum_{j=1}^{1200} \xi_j \le 2500\right)$ valószínűséget kell jól megbecsülnünk. Vegyük észre, hogy $E\xi_j = \frac{1}{6}(2+4+6) = 2$, $\operatorname{Var}\xi_j = \frac{1}{6}(4+16+36) - 4 = \frac{16}{3}$. Innen a centrális határeloszlástétel alapján

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

4. 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 kiégett új lámpát használunk fel. Adjunk jó becslést arra, hogy a lámpák összélettartama legalább 1150 óra.

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

Kiszámoltuk, hogy jelen esetben $E\eta = mE\xi_1 = \frac{m}{\lambda} = 1000$, $\operatorname{Var}\eta = \frac{m}{\lambda^2} = 10000$ ($m = 100$ és $\lambda = \frac{1}{10}$ választással). Ezért a centrális határeloszlástétel szerint $\frac{\eta - E\eta}{\sqrt{\operatorname{Var}\eta}} = \frac{\eta - 1000}{100}$ jó közelítéssel standard normális eloszlású valószínűségi változó, és $P(\xi_1 + \cdots + \xi_{100} > 1150) = P\left(\frac{\eta - E\eta}{\sqrt{\operatorname{Var}\eta}} > 1.5\right) \sim 1 - \Phi(1.5)$.

5. Péter és Pál a következő játékot játssza. Mind a ketten feldobnak egy szabályos dobókockát. Ha Pál dobott nagyobbat, akkor ő nyer három forintot Pétertől, ha Péter dobása nagyobb, akkor Pál fizet Péternek három forintot. Ha a dobások egyenlőek, akkor senki sem fizet a másiknak. Ezt a játékot játsszák 3000 alkalommal. Adjunk jó közelítő becslést a centrális határeloszlástétel és a mellékelt normális eloszlástáblázat segítségével arra, hogy Péter legalább 90 forintot nyer.

Megoldás: Legyen ξ_j Péter nyereménye a j-ik játékban. Ekkor a ξ_j valószínűségi változók függetlenek, Péter össznyereménye $S = \sum\limits_{j=1}^{3000} \xi_j$, és $P(\xi_j = 3) = \frac{5}{12}$, $P(\xi_j = -3) = \frac{5}{12}$, $P(\xi_j = 0) = \frac{1}{6}$. (Annak valószínűségét írtuk fel, hogy Péter vagy Pált dob nagyobbat, illetve hogy a két dobás egyenlő. Innen $E\xi_j = 0$, $\operatorname{Var}\xi_j = E\xi_j^2 = \frac{45}{6}$, $ES = 0$, $\operatorname{Var} S = 15 \cdot 1500 = 150^2$. Innen annak valószínűsége, hogy Péter legalább 90 forintot nyer

$$P(S \ge 90) = P\left(\frac{S - ES}{\sqrt{\operatorname{Var} S}} \ge \frac{90}{150}\right) = 1 - P\left(\frac{S - ES}{\sqrt{\operatorname{Var} S}} < 0.6\right) \sim 1 - \Phi(0.6) \sim 0.274$$

Független valószínűségi változók összegének a sűrűségfüggvénye.

Tárgyaljuk a következő feladatot is. Legyen ξ és η két független valószínűségi változó, amelyeknek létezik $f(\cdot)$ és $g(\cdot)$ sűrűségfüggvénye. Be lehet látni, hogy a $\xi+\eta$ összegnek is létezik $h(\cdot)$ sűrűségfüggvénye, és azt meg lehet adni azt a formulát, amelynek segítségével a $\xi+\eta$ összeg sűrűségfüggvényét. Ennek érdekében bevezetjük két (sűrűség)függvény konvoluciójának a fogalmát.

(Sűrűség)függvények konvoluciójának a definiciója. *Legyen $f(\cdot)$ és $g(\cdot)$ két sűrűségfüggvény a számegyenesen, általánosabban integrálható függvények, azaz tegyük fel, hogy $\int_{-\infty}^{\infty} |f(u)|\,du < \infty$ és $\int_{-\infty}^{\infty} |g(u)|\,du < \infty$. Az $f(\cdot)$ és $g(\cdot)$ függvények $f * g(\cdot)$ konvoluciója az*

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

függvény.

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.

5.) Legyen ξ és η két független, a $\left[-\frac{1}{2},\frac{1}{2}\right]$ intervallumban egyenletes eloszlású valószínűségi változó, azaz legyen ξ és η sűrűségfüggvénye $f(x)=1$, ha $-\frac{1}{2}\le x\le\frac{1}{2}$, és $f(x)=0$ egyébként. Számoljuk ki $\xi+\eta$ sűrűségfüggvényét.

Megoldás: A $\xi+\eta$ valószínűségi változó sűrűségfüggvénye a $g(x)=\int f(y)f(x-y)\,dy$ függvény, ahol $f(x)$ a $\left[-\frac{1}{2},\frac{1}{2}\right]$ intervallumban egyenletes eloszlás sűrűségfüggvénye. Ezért $f(y)f(x-y)=1$, ha $-\frac{1}{2}\le y\le\frac{1}{2}$, és $-\frac{1}{2}\le x-y\le\frac{1}{2}$, azaz $-\frac{1}{2}+x\le y\le\frac{1}{2}+x$, és nulla egyébként. Ez azt jelenti, hogy a $\xi+\eta$ összeg $g(x)$ sűrűségfüggvénye az x pontban megegyezik a $\left[-\frac{1}{2},\frac{1}{2}\right]\cap\left[-\frac{1}{2}+x,\le\frac{1}{2}+x\right]$ intervallum hosszával. Ha $|x|>1$, akkor a fenti metszet üres, ezért ebben az esetben $g(x)=$. Ha $0\le x\le 1$, akkor ez a metszet a $\left[-\frac{1}{2}+x,\frac{1}{2}\right]$ intervallum, és ennek hossza $1-x$, azaz ebben az esetben $g(x)=1-x$. Ha $-1\le x\le 0$, akkor ez a metszet a $\left[-\frac{1}{2},\frac{1}{2}+x\right]$ intervallum amelynek hossza $1+x=1-|x|$, azaz $g(x)=1+x=1-|x|$ ebben az esetben. Ez azt jelenti, hogy $g(x)=1-|x|$, ha $|x|\le 1$, és $g(x)=0$, ha $x>1$.

Megadunk egy másik geometriai érvelésen alapuló megoldást is, amelyik a korábban tárgyalt geometriai érvelésen alapul.

Számítsuk ki először a $\xi+\eta$ valószínűségi változó $G(x)$ eloszlásfüggvényét. Definiáljuk a $K=\left[-\frac{1}{2},\frac{1}{2}\right]\times\left[-\frac{1}{2},\frac{1}{2}\right]$ négyzetet, és jelölje λ a Lebesgue mértéket, azaz a területet a síkon. Ekkor a sík tetszőleges $A\subset R^2$ mérhető részhalmazára igaz az, hogy $P((\xi,\eta)\in A)=\lambda(A\cap K)$. Speciálisan, $G(x)=P(\xi+\eta<x)=\lambda(K\cap\{(u,v)\colon u+v<x\})$. Ha $x\le -1$, akkor $G(x)=0$, ha $-1\le x\le 0$, akkor $G(x)$ a $\left(-\frac{1}{2},-\frac{1}{2}\right)$, $\left(-\frac{1}{2},\frac{1}{2}+x\right)$ és $\left(\frac{1}{2}+x,-\frac{1}{2}\right)$ pontok által meghatározott háromszög területe $\frac{1}{2}(1+x)^2$. Hasonlóan, ha $x\ge 1$, akkor $G(x)=1$. Ha $0\le x\le 1$, akkor a $G(x)$ eloszlásfüggvény megegyezik annak a poligonnak területével, amelyet úgy kapunk, hogy a K négyzetből kihagyjuk a $\left(\frac{1}{2},\frac{1}{2}\right)$, $\left(\frac{1}{2},-\frac{1}{2}+x\right)$ és $\left(-\frac{1}{2}+x,\frac{1}{2}\right)$ pontok által meghatározott háromszöget. Ezért $G(x)=1-\frac{1}{2}(1-x)^2$ ebben az esetben. A $G(x)$ függvényt deriválva kapjuk, hogy $g(x)=0$, ha $|x|\le 1$, $g(x)=1+x$, ha $-1\le x\le 0$, és $g(x)=1-x$, ha $0\le x\le 1$.