A november 25.-i gyakorlat témája

Foglalkozzunk azzal, hogyan tudjuk kiszámítani független (és nemcsak független) valószínűségi változók összegének a várható értékét és szórásnégyzetét. Ezenkívül ismerjük meg a valószínűségszámítás egyik legfontosabb eredményét, a centrális határeloszlástételt, és lássunk példákat arra, hogy a megtanult módszerek hogyan teszik lehetővé a centrális határeloszlástétel alkalmazását fontos feladatok megoldásában.

Megfogalmazunk 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*

$$E\left(c_1\xi_1 + c_2\xi_2\right) = c_1 E\xi_1 + c_2 E\xi_2.$$

Tétel. *Legyenek $\xi_1, \dots, \xi_k$ független valószínűségi változók, $B_1, \dots, B_k$ a számegyenes Borel mérhető részhalmazai. Ekkor*

$$P(\xi_1 \in B_1, \dots, \xi_k \in B_k) = P(\xi_1 \in B_1) \cdots P(\xi_k \in B_k).$$

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

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

Speciálisan, ha a ξ_j valószínűségi változók függetlenek, akkor

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

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.

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.

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

   $H(x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = P(\xi < \sqrt{x}) - P(\xi \le \sqrt{x}) = F(\sqrt{x}) - F(-sqrtx + 0)$, ha $x \ge 0$, ahol $F(u+0) = \lim\limits_{h\to 0} F(u+h)$. (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 $h(x) = \frac{1}{2\sqrt{x}}\left(f(\sqrt{x}) + f(-\sqrt{x})\right)$, ha $x \ge 0$, $h(x) = 0$, ha $x < 0$.

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.*

2.) 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, $1 \le j \le 5\ 000\ 000$, 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. 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, $\le j \le 10\ 000$ 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, $1 \le j \le 10\ 000$ 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 centrális határeloszlástétel szerint $P\left(\frac{\sum\limits_{j=1}^{10000}(\xi_j-E\xi_j)}{\sqrt{10000\cdot\frac14}}>u\right)\sim$ $1-\Phi(u)$. Innen kapjuk, hogy az első valószínűség kiszámításához az $u=\pm\frac{100}{\frac12\cdot 100}=$ ± 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).

4.) 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 $P\left(2280\le\sum\limits_{j=1}^{1200}\xi_j\le 2500\right)$ valószínűséget kell jól megbecsülnünk. Vegyük észre, hogy $E\xi_j=\dfrac16(2+4+6)=2$, $\operatorname{Var}\xi_j=\dfrac16(4+16+36)-4=\dfrac{16}{3}$. Innen a centrális határeloszlástétel alapján

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

5.) Vegyünk egy olyan pénzdarabot, amely $\dfrac23$ valószínűséggel esik a fej és $\dfrac13$ 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 1830 küzé esik? Adjunk erre a valószínűségre jó közelítő becslést.

Megoldás: Az elvégzett dobások száma egy η negatív binomiális eloszlású valószínűségi változó $n=1200$ és $p=\dfrac23$ paraméterekkel, azaz $P(\eta=k+n)=\binom{n+k-1}{n-1}(1-p)^kp^n$, $p=\dfrac23$, 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, $2 \le j \le 1200$, 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 $n=1$, $p=\dfrac{2}{3}$ paraméterekkel, és minket a $P(1680 < \xi_1 + \cdots + \xi_{1200} < 1830)$ valószínűség érdekel. Megmutattuk korábban, hogy $E\xi_j = \dfrac{1}{p} = \dfrac{3}{2}$, $\operatorname{Var}\xi_j = \dfrac{1-p}{p^2} = \dfrac{3}{4}$. Ezért a centrális határeloszlástétel alapján $\eta = \xi_1 + \cdots + \xi_{1200}$ jelöléssel minket a $P\left(-4 < \dfrac{\eta - 1200E\xi_1}{\sqrt{1200\operatorname{Var}\xi_1}} < 1\right)$ valószínűség érdekel. A centrális határeloszlástétel alapján $P\left(-4 < \dfrac{\eta - 1200E\xi_1}{\sqrt{1200\operatorname{Var}\xi_1}} < 1\right) \sim \Phi(1) + \Phi(4) - 1 \sim \Phi(1)$.

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 vesztünk 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.