A november 2.-i gyakorlat témája

Tárgyaljuk meg néhány példán keresztül azt, hogy hogyan lehet kiszámítani olyan valószínűségi változók szórásnégyzetét, amelyek előállíthatóak mint egyszerű, de nem feltétlenül független valószínűségi változók összegei. Ebből a célból be kell vezetnünk két valószínűségi változó kovarianciafüggvényének a fogalmát. Ez azt méri, hogy milyen mértékben függ a két valószínűségi változó egymástól.

A kovarianciafüggvény definiciója. *Legyen ξ és η két valószínűségi változó, amelyek mindegyikének létezik szórásnégyzete, azaz $E\xi^2 < \infty$ és $E\eta^2 < \infty$. Ekkor a ξ és η kovarianciafüggvényét* $\operatorname{Cov}(\xi,\eta)$*-t a*

$$\operatorname{Cov}(\xi,\eta) = E\left[(\xi - E\xi)(\eta - E\eta)\right]$$

kifejezés definiálja.

Lemma.

$$\operatorname{Cov}(\xi,\eta) = E\xi\eta - E\xi E\eta.$$

Valóban,

$$\begin{aligned}\operatorname{Cov}(\xi,\eta) &= E\left[(\xi - E\xi)(\eta - E\eta)\right] = E\left[(\xi\eta - \eta E\xi - \xi E\eta + E\xi E\eta)\right] \\ &= E\xi\eta - E\eta E\xi - E\xi E\eta + E\xi E\eta = E\xi\eta - E\xi E\eta.\end{aligned}$$

1.) ξ és η két független valószínűségi változó, akkor $\operatorname{Cov}(\xi,\eta) = 0$.

Megoldás: Láttuk, hogy $\operatorname{Cov}(\xi,\eta) = E\xi\eta - E\xi E\eta$. Viszont, ha ξ és η független valószínűségi változók, akkor $E\xi\eta = E\xi E\eta$.

Az alábbi eredmény megadja, hogy hogyan tudjuk kifejezni valószínűségi változók összegének a szórásnégyzetét viszonylag egyszerű módon.

Tétel. *Ha $\xi_1, \dots, \xi_n$ valószínűségi változók ugyanazon a valószínűségi mezőn, amelyek mindegyikének létezik szórásnégyzete, akkor*

$$\begin{aligned}\operatorname{Var}\left(\sum_{j=1}^{n} \xi_j\right) &= \sum_{j=1}^{n} \operatorname{Var}\xi_j + 2\sum_{1\le j<k\le n} (E\xi_j\xi_k - E\xi_j E\xi_k) \\ &= \sum_{j=1}^{n} \operatorname{Var}\xi_j + 2\sum_{1\le j<k\le n} (\operatorname{Cov}(\xi_j,\xi_k) \\ &= \sum_{j=1}^{n} \operatorname{Var}\xi_j + \sum_{1\le j,k\le n,\ j\ne k} (\operatorname{Cov}(\xi_j,\xi_k).\end{aligned}$$

Bizonyítás.

$$\begin{aligned}
\operatorname{Var}\left(\sum_{j=1}^{n}\xi_j\right) &= E\left(\sum_{j=1}^{n}\xi_j\right)^2-\left(E\sum_{j=1}^{n}\xi_j\right)^2 = E\left(\sum_{j=1}^{m}\sum_{k=1}^{n}\xi_j\xi_k\right)-\left(\sum_{j=1}^{m}\sum_{k=1}^{n}E\xi_jE\xi_k\right)\\
&=\left(\sum_{j=1}^{m}\sum_{k=1}^{n}E\xi_j\xi_k\right)-\left(\sum_{j=1}^{m}\sum_{k=1}^{n}E\xi_jE\xi_k\right)\\
&=\sum_{j=1}^{n}\left(E\xi_j^2-(E\xi_j)^2\right)+2\sum_{1\le j<k\le n}(E\xi_j\xi_k-E\xi_jE\xi_k).
\end{aligned}$$

és $E\xi_j^2-(E\xi_j)^2=\operatorname{Var}\xi_j$, $E\xi_j\xi_k-E\xi_jE\xi_k=\operatorname{Cov}(\xi_j,\xi_k)$.

Megjegyzés: A fenti tételben két különböző, de ekvivalens módon fejeztük ki valószínűségi változók összegének a kovarianciafüggvényét. Az egyik módon az összegzés a második tagban az $1\le j<k\le n$ alakú tagokra történt, és az összeget beszoroztuk kettővel. A másik kifejezésben az összegezés az $1\le j,k\le n$, $j\neq k$, alakú számpárokra történt, és a 2 szorzó nem szerepelt. Ez analóg a következő egyszerű algebrai azonosságokkal: $\left(\sum\limits_{j=1}^{n}x_j\right)^2=\sum\limits_{j=1}^{n}x_j^2+2\sum\limits_{1\le j<k\le n}x_jx_k$, és $\left(\sum\limits_{j=1}^{n}x_j\right)^2=\sum\limits_{j=1}^{n}x_j^2+\sum\limits_{1\le j,k\le n,j\neq k}x_jx_k$.

A fenti eredmény segítségével viszonylag egyszerűen meg tudjuk oldani a következő két feladatot.

2.) Egy szabályos dobókockát feldobunk 101 alkalommal. Arra vagyunk kiváncsiak, hogy hány olyan dobásidőpont volt, (a 2., 3., és így tovább a 101. dobásidőpontok között), amikor a dobás értéke (szigorúan) nagyobb volt, mint a megelőző dobás értéke. Számítsuk ki az ilyen dobásidőpontok számának a várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_k valószínűségi változókat, $1\le k\le 100$: $\xi_k=1$, ha a $k+1$-ik kockadobás értéke nagyobb, mint a k-ik kockadobás értéke, $\xi_k=0$, ha a $k+1$-ik kockadobás értéke kisebb vagy egyenlő, mint a k-ik kockadobás értéke. Ekkor minket a $\xi=\sum\limits_{k=1}^{100}\xi_k$ valószínűségi változó várható értéke és szórásnégyzete érdekel. Viszont $E\xi=\sum\limits_{k=1}^{100}E\xi_k$, $E\xi_k=P(\xi_k=1)$, és $P(\xi_k=1)=\frac{1}{36}(5+4+3+2+1)=\frac{5}{12}$, (Ha a k-ik dobás értéke 1, akkor a $k+1$-ik dobás 5 értéket vehet fel, ha 2 akkor 4-et, ha 3 akkor 3-at stb.) Ezért a keresett várható érték $\frac{500}{12}=\frac{125}{3}$.

A ξ valószínűségi változó szórásnégyzetét kiszámolhatjuk a fenti tétel segítségével. E tétel alkalmazása során ki kell számolnunk az $\operatorname{Var}\xi_k$ szórásnégyzeteket, továbbá a $\operatorname{Cov}(\xi_j,\xi_k)$, $j\neq k$, kovarianciákat. $\operatorname{Var}\xi_k=E\xi_k^2-(E\xi_k)^2=P(\xi_k=1)-P(\xi_k=1)^2=\frac{5}{12}-\frac{25}{144}=\frac{35}{144}$. Vegyük észre, hogy amennyiben $|k-j|\ge 2$, akkor ξ_j és ξ_k független valószínűségi változók, ezért $\operatorname{Cov}(\xi_j,\xi_k)=0$ ebben az esetben.

A $\mathrm{Cov}(\xi_k, \xi_{k+1}) = E\xi_k\xi_{k+1} - E\xi_k E\xi_{k+1} = P(\xi_k = \xi_{k+1} = 1) - \frac{25}{144}$ azonosság teljesül. A $P(\xi_k = \xi_{k+1} = 1)$ valószínűség kiszámolásához azt kell összeszámolni, hogy hány három hosszúságú szigorúan monoton növekvő sorozat van, amelynek tagjai értéküket 1 és 6 között veszik fel. Ezt kézzel is össze lehetne számolni, de az első két gyakorlaton megbeszéltük, hogy $\binom{6}{3} = 20$ ez a szám. (A hat számból kiválasztunk három különbözőt, és azokat nagyság szerint sorba rakjuk.) Ezért $E\xi_k\xi_{k+1} = \frac{20}{216} = \frac{5}{54}$, $\mathrm{Cov}(\xi_k, \xi_{k+1}) = \frac{20}{216} - \frac{25}{144} = \frac{-35}{432}$. Innen $\mathrm{Var}\,\xi = \sum\limits_{k=1}^{100} \mathrm{Var}\,\xi_k + 2\sum\limits_{k=1}^{99} \mathrm{Cov}(\xi_k, \xi_{k+1}) = \frac{500}{12} - \frac{3465}{432} = \frac{14535}{432}$.

3.) Legyen egy urnában 20 piros és 30 fehér golyó. Húzzunk ki 20 golyót visszatevés nélkül. Számoljuk ki a kihúzott piros golyók számának várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 20$, valószínűségi változókat: $\xi_j(\omega) = 1$, ha a j-ik húzás eredménye piros, $\xi_j(\omega) = 0$, ha a j-ik húzás eredménye fehér. Ekkor a $\xi = \sum\limits_{j=1}^{20} \xi_j$ összeg várható értékét és szórásnégyzetét kell kiszámolnunk. Továbbá $E\xi_j = E\xi_1 = \frac{2}{5}$, $\mathrm{Var}\,\xi_j = \mathrm{Var}\,\xi_1 = \frac{2}{5} - \frac{4}{25} = \frac{6}{25}$, $E\xi_j\xi_k - E\xi_j E\xi_k = E\xi_1\xi_2 - E\xi_1 E\xi_2 = \frac{2}{5} \cdot \frac{19}{49} - \frac{4}{25} = -\frac{6}{1225}$, ha $j \ne k$. Innen a 20 dobásban kihúzott piros golyók számának várható értéke $E\xi = \sum\limits_{j=1}^{20} E\xi_j = 20 \cdot \frac{2}{5} = 8$ és szórásnégyzete $\mathrm{Var}\,\xi = \sum\limits_{j=1}^{20} \mathrm{Var}\,\xi_j + 2\sum\limits_{j=1}^{19}\sum\limits_{k=j+1}^{20} \mathrm{Cov}(\xi_j, \xi_k) = 20 \cdot \frac{6}{25} - 380 \cdot \frac{6}{1225} = \frac{144}{49}$.

Házi feladat:

Egy szabályos pénzdarabot feldobunk egymás után 101-szer. Számítsuk ki az egymást követő fej-fej dobások számának a várható értékét és szórásnégyzetét.

Végül tekintsük azt, hogy hogyan lehet kiszámolni független, diszkrét eloszlású valószínűségi változók összegének az eloszlását. Megmutatjuk, hogy két független Poisson eloszlású valószínűségi változónak az összege is Poisson eloszlású, és az összeg paramétere megegyezik az összeadandók paraméterének az összegével. Először definiáljuk magát a Poisson eloszlást, és lássuk be, hogy az valóban eloszlás.

Poisson eloszlás definiciója. *Egy ξ valószínűségi változó λ paraméterű Poisson eloszlású, ha ξ nem negatív egész értékeket vesz fel, és*

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

4.) A fent definiált Poisson eloszlás, valóban eloszás, azaz $P(\xi = k) \ge 0$ minden egész számra, és $\sum\limits_{k=0}^{\infty} P(\xi = k) = 1$.

Megoldás: A $P(\xi = k) \geq 0$ reláció nyilvánvaló. Másrészt,

$$\sum_{k=0}^{\infty} P(\xi = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

5.) Mutassuk meg, hogy ha ξ és η két független, λ illetve μ paraméterű Poisson eloszlású valószínűségi változó, akkor $\xi + \eta$ $\lambda + \mu$ paraméterű Poisson eloszlású valószínűségi változó.

Megoldás:

$$\begin{aligned} P(\xi + \eta = k) &= \sum_{j=0}^{k} P(\xi = j, \eta = k - j) = \sum_{j=0}^{k} P(\xi = j) P(\eta = k - j) \\ &= \sum_{j=0}^{k} \frac{\lambda^j}{j!} e^{-\lambda} \frac{\mu^{k-j}}{(k-j)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} \lambda^j \mu^{(k-j)} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k. \end{aligned}$$

6.) Legyen ξ és η két független binomiális elsoszlású valószínűségi változó (n, p) illetve (m, p) paraméterekkel, azaz legyen $P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \leq k \leq n$, és $P(\xi = k) = 0$ egyébként, $P(\eta = k) = \binom{m}{k} p^k (1-p)^{m-k}$, $0 \leq k \leq m$, és $P(\xi = k) = 0$ egyébként. Lássuk be, hogy $\xi + \eta$ binomiális eloszlású $(n+m, p)$ paraméterrel, azaz legyen $P(\xi + \eta = k) = \binom{n+m}{k} p^k (1-p)^{n+m-k}$, $0 \leq k \leq n + m$, és $P(\xi = k) = 0$ egyébként.

Megoldás: Legyen $0 \leq k \leq n + m$. Ekkor

$$\begin{aligned} P(\xi + \eta = k) &= \sum_{j=0}^{n} P(\xi = j, \eta = k - j) = \sum_{j=0}^{n} P(\xi = j) P(\eta = k - j) \\ &= \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{k-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-(k-j)} \\ &= p^k (1-p)^{n+m-k} \sum_{j=0}^{n} \binom{n}{j} \binom{m}{k-j} = p^k (1-p)^{n+m-k} \binom{n+m}{k}, \end{aligned}$$

mert, mint azt az első gyakorlaton megbeszéltük érvényes a $\sum\limits_{j=0}^{n} \binom{n}{j} \binom{m}{k-j} \binom{n+m}{k}$. Minden más k számra teljesül a $P(\xi + \eta = k) = 0$ reláció.

Házi feladat (nem kötelező)

Bizonyítsuk be a fenti azonosságot a binomiális eloszlás szemléletes tartalmát felhasználva, azaz konstruáljunk két független (n, p) és (m, p) paraméterű binomiális eloszlású valószínűségi változót, amelyek összegéről látszik, hogy binomiális eloszlású $(n + m, p)$ paraméterrel.