A november 5.-i előadáshoz kapcsolódó gyakorlat feladatai

1. Az előző gyakorlaton beláttuk, hogy ha $\xi_1,\dots,\xi_n$ független, a $\left[-\frac12,\frac12\right]$ intervallumban egyenletes eloszlású valószínűségi változók, $\xi_1^*=\min\limits_{1\le k\le n}\xi_k$, $\xi_n^*=\max\limits_{1\le k\le n}\xi_k$, akkor a (ξ_1^*,ξ_n^*) véletlen vektor eloszlásának van sűrűségfüggvénye, és az $f(x,y)=n(n-1)(y-x)^{n-2}$, ha $-\frac12\le x\le y\le\frac12$, és $f(x,y)=0$ egyébként. Számítsuk ki ennek az eredménynek a segítségével a $\frac{\xi_1^*+\xi_n^*}{2}$ valószínűségi változó várható értékét és szórásnégyzetét.

   Megoldás: Felírhatjuk az

$$
\begin{aligned}
E\frac{\xi_1^*+\xi_2^*}{2}&=\int\!\!\int\frac{x+y}{2}f(x,y)\,dx\,dy\\
&=n(n-1)\int\!\!\int_{\{(x,y)\colon\ \frac12\le x\le y\le\frac12\}}\frac{x+y}{2}(y-x)^{n-2}\,dx\,dy
\end{aligned}
$$

$$
\begin{aligned}
E\left(\frac{\xi_1^*+\xi_2^*}{2}\right)^2&=\int\!\!\int\left(\frac{x+y}{2}\right)^2f(x,y)\,dx\,dy\\
&=n(n-1)\int\!\!\int_{\{(x,y)\colon\ \frac12\le x\le y\le\frac12\}}\left(\frac{x+y}{2}\right)^2(y-x)^{n-2}\,dx\,dy.
\end{aligned}
$$

azonosságokat. A fenti integrálokat kell kiszámolnunk. Érdemes bevezetni az $u=y-x$, $v=y+x$ új változókat, és ezek segítségével kifejezni a vizsgálandó integrálokat. Vegyük észre, hogy a transzformáció Jacobiánja a $\begin{vmatrix}\frac{\partial u}{\partial x}, & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x}, & \frac{\partial v}{\partial y}\end{vmatrix}$ determináns (abszolut) értéke 2 (minden pontban), és a transzformáció hatására az $\{(x,y)\colon \frac12\le x\le y\le\frac12\}$ integrálási tartomány az $\{(u,v)\colon 0\le u\le 1, -(1-u)\le v\le 1-u\}$ tartományba képződik.

A fenti összefüggések segítségével a kiszámítandó integrálokat a következő módon írhatjuk fel:

$$
\begin{aligned}
E\frac{\xi_1^*+\xi_2^*}{2}&=\frac{n(n-1)}{2}\int_0^1\int_{-(1-u)}^{1-u}\frac{v}{2}u^{n-2}\,dv\,du\\
&=\frac{n(n-1)}{4}\int_0^1u^{n-2}\left(\int_{-(1-u)}^{1-u}v\,dv\right)du=0,
\end{aligned}
$$

mert a belső integrál nulla, és

$$
\begin{aligned}
E\left(\frac{\xi_1^*+\xi_2^*}{2}\right)^2&=\frac{n(n-1)}{2}\int_0^1\int_{-(1-u)}^{1-u}\left(\frac{v}{2}\right)^2u^{n-2}\,dv\,du\\
&=\frac{n(n-1)}{12}\int_0^1(1-u)^3u^{n-2}\,dv\,du
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{n(n-1)}{12}\left[\frac{u^{n-1}}{n-1} - 3\frac{u^n}{n} + 3\frac{u^{n+1}}{n+1} - \frac{u^{n+2}}{n+2}\right]_0^1 \\
&= \frac{n(n-1)}{12}\left(\frac{1}{n-1} - \frac{1}{n+2} - \frac{3}{n} + \frac{3}{n+1}\right) \\
&= \frac{n(n+1)(n+2-(n-1)) - 3(n-1)(n+2)}{12(n+1)(n+2)} \\
&= \frac{3n(n+1) - 3(n-1)(n+2)}{12(n+1)(n+2)} = \frac{1}{2(n+1)(n+2)}
\end{aligned}
$$

Tehát $E\frac{\xi_1^*+\xi_2^*}{2} = 0$ és $\operatorname{Var}\frac{\xi_1^*+\xi_2^*}{2} = \frac{1}{2(n+1)(n+2)}$.

2. Legyen $\xi_1,\dots,\xi_n$ független, a $[\vartheta,\vartheta+1]$ intervallumban egyenletes eloszlású valószínűségi változók sorozata valamely (ismeretlen) ϑ paraméterrel. Legyen $\xi_1^* = \min\limits_{1\le k\le n}\xi_k$ $\xi_n^* = \max\limits_{1\le k\le n}\xi_k$. Mutassuk meg, hogy $E\frac{\xi_1^*+\xi_n^*}{2} = \vartheta+\frac12$, $\operatorname{Var}\frac{\xi_1^*+\xi_n^*}{2} = \frac{1}{2(n+1)(n+2)}$, $E\frac{\xi_1+\cdots+\xi_n}{n} = \vartheta+\frac12$ és $\operatorname{Var}\frac{\xi_1+\cdots+\xi_n}{n} = \frac{1}{12n}$.

Beszéljük meg a következő kérdést: Meg akarjuk becsülni az ismeretlen ϑ paramétert a (megfigyelt) $\xi_1,\dots,\xi_n$ valószínűségi változók segítségével. Melyik becslést tartják jobbnak:

a.) $\tilde\vartheta_n^{(1)} = \frac{\xi_1+\cdots+\xi_n}{n} - \frac12$

b.) $\tilde\vartheta_n^{(2)} = \frac{\xi_1^*+\xi_n^*}{2} - \frac12$

c.) Kombináljuk a két becslést és legyen $\tilde\vartheta_n^{(3)} = \frac{\tilde\vartheta_n^{(1)}+\tilde\vartheta_n^{(2)}}{2}$.

Megoldás: A $\bar\xi_j = \xi_j - (\vartheta+12)$ valószínűségi változók függetlenek, egyenletes eloszlásúak a $\left[-\frac12,\frac12\right]$ intervallumban, és $\min\limits_{1\le k\le n}\bar\xi_k = \xi_1^* - (\vartheta+12)$, $\max\limits_{1\le k\le n}\bar\xi_k = \xi_n^* - (\vartheta+12)$. Ezért áttérve a ξ_k valószínűségi változókról a $\bar\xi_k$ valószínűségi változókra a vizsgált mennyiségek várható értéke $\vartheta+\frac12$-vel csökken, a szórásnégyzetük pedig nem változik. Ezért elegendő a feladatot a $\bar\xi_k$ valószínűségi változókra bizonyítani. (Ezt úgy is megfogalmazhatjuk, hogy elég a $\vartheta=-\frac12$ esetet tekinteni.) Az első kifejezés várható értékét és szórásnégyzetét az első feladatban kiszámoltuk, és onnan látszik a kívánt állítás. A második kifejezésre felírhatjuk, hogy $E\frac{\bar\xi_1+\cdots+\bar\xi_n}{n} = E\bar\xi_1 = 0$ és $\operatorname{Var}\frac{\bar\xi_1+\cdots+\bar\xi_n}{n} = \frac{\operatorname{Var}\bar\xi_1}{n} = \frac{1}{12n}$, mert $\operatorname{Var}\bar\xi_1 = \int_{-1/2}^{1/2} x^2\,dx = \frac{1}{12}$.

Mind a három esetben a vizsgált kifejezés várható értéke ϑ, (ezt szakszóval úgy mondják, hogy torzítatlan becsléseket adtunk meg), ezért természetes a becslések közül azt választani, amelyiknek a legkisebb a szórásnégyzete. Láttuk, hogy ebben az értelemben a $\tilde\vartheta_n^{(2)}$ becslés jobb, mint a $\tilde\vartheta_n^{(1)}$ becslés. Azt is meg lehet mutatni, hogy a $\tilde\vartheta_n^{(2)}$ becslés a $\tilde\vartheta_n^{(3)}$ becslésnél is jobb. Most elégedjünk meg annak megmutatásával, hogy nagy n elemszámra ez a helyzet, sőt azt is állíthatjuk, hogy a becslés nagyságrendje jóbb.

Ugyanis $E\left(\tilde\vartheta_n^{(1)}-\vartheta\right)^2$ nagyságrendje const. n^{-1}, míg a $E\left(\tilde\vartheta_n^{(2)}-\vartheta\right)^2$ kifejezés nagyságrendje const. n^{-2}, ha $n \to \infty$. Innen

$$E\left(\tilde{\vartheta}_n^{(3)} - \vartheta\right)^2 = \frac{1}{4}\left(E\left(\tilde{\vartheta}_n^{(1)} - \vartheta\right)^2 + E\left(\tilde{\vartheta}_n^{(2)} - \vartheta\right)^2 + 2E\left(\tilde{\vartheta}_n^{(1)} - \vartheta\right)\left(\tilde{\vartheta}_n^{(2)} - \vartheta\right)\right),$$

és a Cauchy–Schwarz egyenlőtlenség alapján

$$\left|E\left(\tilde{\vartheta}_n^{(1)} - \vartheta\right)\left(\tilde{\vartheta}_n^{(2)} - \vartheta\right)\right| \le E\left(\left(\tilde{\vartheta}_n^{(1)} - \vartheta\right)^2\left(E\tilde{\vartheta}_n^{(2)} - \vartheta\right)^2\right)^{1/2},$$

és ez a kifejezés $n^{-3/2}$ nagyságrendben tart nullához $n \to \infty$ esetén. Ezekből a becslésekből az következik, hogy az $E\left(\tilde{\vartheta}_n^{(3)} - \vartheta\right)^2$ kifejezés az $E\left(\tilde{\vartheta}_n^{(1)} - \vartheta\right)^2$ kifejezéshez hasonlóan n^{-1} nagyságrendben tart nullához $n \to \infty$ esetben. (Tehát lassabban, mint a b.) esetben.

A következő feladatban megtárgyalunk egy más módszert a rendezett minta legnagyobb és és legkisebb elemének együttes sűrűségfüggvényének kiszámítására.

3. Legyenek $\xi_1, \dots, \xi_n$ független egyforma eloszlású valószínűségi változók $f(x)$ sűrűségfüggvénnyel, $\xi_1^* < \xi_2^* < \dots < \xi_n^*$ a belőlük készített rendezett minta. Lássuk be, hogy a $(\xi_1^*, \dots, \xi_n^*)$ véletlen vektornak van sűrűségfüggvénye, és az a $g(x_1, \dots, x_n) = n! f(x_1) \cdots f(x_n)$ függvény, ha $x_1 < x_2 < \dots < x_n$, és $g(x_1, \dots, x_n) = 0$ egyébként.

   Lássuk be, hogy ha egy $(\eta_1, \dots, \eta_n)$ véletlen vektornak $g(x_1, \dots, x_n)$ van sűrűségfüggvénye, és $k \le n$, akkor az $(\eta_1, \dots, \eta_k)$ véletlen vektor sűrűségfüggvénye

$$h(x, \dots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n)\, dx_{k+1}\, dx_{k+2} \cdots dx_n.$$

   A $[-\frac{1}{2}, \frac{1}{2}]$ intervallumon egyenletes eloszlású független $\xi_1, \dots, \xi_n$ valószínűségi változókból elkészített $\xi_1^* < \xi_2^* < \dots < \xi_n^*$ rendezett mintára a (ξ_1^*, ξ_n^*) véletlen vektor sűrűségfüggvénye $f(x, y) = n(n-1)(y-x)^{n-2}$, ha $-\frac{1}{2} \le x \le y \le \frac{1}{2}$, és $f(x, y) = 0$ egyébként.

   Megoldás: Definiáljuk a következő $C \subset R^n$ halmazt: $C = \{(x_1, \dots, x_n) \colon x_1 \le x_2 \le \dots \le x_n\}$, azaz C azon pontokból áll, melynek koordinátái nagyság szerint rendezve vannak. Adva egy $B \subset C$ halmaz definiáljuk ennek B^{sz} szimmetrikus bővítését, mint $B^{sz} = \{(x_{\pi(1)}, \dots, x_{\pi(n)})\,(x_1, \dots, x_n \in B,\, (pi(1), \dots, \pi(n)) \in \Pi_n\}$, ahol Π_n jelöli az $\{1, \dots, n\}$ halmaz összes permutációját. Szemléletesen arról van szó, hogy a B halmaz $(x_1, \dots, x_n)$ pontjaiban vesszük a koordináták összes lehetséges permutációját, és az így előállítható pontok alkotják a B halmaz szimmetrizált bővítését. Ekkor $P((\xi_1^*, \dots, \xi_n^*) \in C) = 1$, és tetszőleges (mérhető) $A \subset R^n$ halmazra

$$\begin{aligned} P((\xi_1^*, \dots, \xi_n^*) \in A) &= P((\xi_1^*, \dots, \xi_n^*) \in A \cap C) = P((\xi_1, \dots, \xi_n) \in (A \cap C)^{sz}) \\ &= \int_{A \cap C)^{sz}} f(x_1) f(x_2) \dots f(x_n)\, dx_1 \dots dx_n \end{aligned}$$

$$= \int_{A\cap C)} n! f(x_1) f(x_2) \dots f(x_n)\, dx_1 \dots dx_n$$
$$= \int_A g(x_1, x_2, \dots, x_n)\, dx_1 \dots dx_n,$$

és ez jelenti azt, hogy a $(\xi_1^*, \dots, \xi_n^*)$ véletlen vektor eloszlása a $g(x_1, x_2, \dots, x_n)$ függvény.

A feltétel teljsülése esetén minden $A \subset R^k$ halmazra

$$P((\eta_1, \dots, \eta_k) \in A) = P((\eta_1, \dots, \eta_k) \in A, (\eta_{k+1}, \dots, \eta_n) \in R^{n-k})$$
$$= \int \cdots \int_{A\times R^{n-k}} g(x_1, \dots, x_n)\, dx_1 \dots dx_k\, dx_{k+1} \dots dx_n$$
$$= \int \cdots \int_A \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n)\, dx_{k+1}\, dx_{k+2} \cdots dx_n \right) dx_1 \dots dx_k$$
$$= \int \cdots \int_A h(x, \dots, x_k)\, dx_1 \dots dx_k,$$

és ez volt a második bizonyítandó állítás.

Az előbbiek alapján $-\frac{1}{2} \le x \le y \le y \le \frac{1}{2}$ esetén a keresett sűrűségfüggvény

$$g(x, y) = \int_{x\le x_1<x_2<\cdots<x_{n-1}<y} n!\, dx_2 \dots dx_{n-1}$$
$$= n!\lambda\{(x_2, \dots, x_{n-1} \colon x \le x_1 < x_2 < \cdots < x_{n-1} < y\} = \frac{n!(y-x)^{n-2}}{(n-2)!}$$
$$= n(n-1)(y-x)^{n-2},$$

ahol $\lambda(A)$ jelöli az A halmaz térfogatát. A maradék halmazon a sűrűségfüggvény nulla, mert a rendezett minta $g(x, x_2, \dots, x_{n-1}, y)$ sűrűségfüggvénye, amit integrálni kell, nulla, ha $y < x$, $y > \frac{1}{2}$ vagy $x < -\frac{1}{2}$.

4. Legyen egy urnában 10 piros és 20 fehér golyó. Húzzunk ki 10 golyót visszatevés nélkül, és a második húzástól kezdve minden húzás alkalmával nyerjünk 2 forintot, ha a kihúzott golyó színe megegyezik az előző kihúzott golyó színével, és veszítsünk 1 forintot, ha ez a két szín különbözik. Számoljuk ki nyereményünk várható értékét és szórásnégyzét.

   Megoldás: Vezessük be a követező η_j, $1 \le j \le 9$, valószínűségi változókat: $\eta_j = 2$, ha a j-ik és $j+1$ húzás mindegyike fehér vagy mindegyike piros, és $\eta_j = -1$, ha a két húzás eredménye fehér, piros, vagy piros fehér. Ekkor minket a $\sum\limits_{j=1}^{9} \eta_j$ véletlen összeg várható értéke és szórásnégyzete érdekel. Ennek kiszámítása érdekében számoljuk ki először az $E\eta_j$, $\operatorname{Var}\eta_j$ és $\operatorname{Cov}(\eta_j, \eta_k)$ mennyiségeket. η_j értéke 2, ha piros, piros vagy fehér, fehér húzás történt, és ennek valószínűsége $\frac{10}{30}\frac{9}{29} + \frac{20}{30}1930 = \frac{47}{87}$, annak valószínűsége, hogy $\eta_j = -1$ $\frac{40}{87}$. Innen $E\eta_j = \frac{2\cdot 47-40}{87} = \frac{18}{29}$, és a keresett várható érték $9 \cdot \frac{18}{29} = \frac{162}{29}$. Hasonlóan $E\eta_j^2 = \frac{4\cdot 47+40}{87} = \frac{228}{87}$, $\operatorname{Var}\eta_j = E\eta_j^2 - (E\eta_j)^2$.

Az Cov (η_j, η_k) mennyiségek kiszámolásánál különböztessük meg a $k = j+1$ és $k > j+1$ eseteket. Első esetben az $E\eta_j\eta_{j+1}$ mennyiség kiszámolásánal három egymást követő húzás eredményét kell figyelembe venni. $\eta_j\eta_k$ értéke 4, ha fehér, fehér, fehér, vagy piros, piros, piros húzás történik, és ennek valószínűsége $\frac{1\cdot 9\cdot 8+2\cdot 19\cdot 18}{3\cdot 29\cdot 28}$, -2, ha fehér, piros, piros vagy fehér, fehér, piros vagy piros, piros, fehér vagy piros, fehér, fehér húzássorozat történik, és ennek valószínűsége $2\cdot\frac{2\cdot 19\cdot\cdot 10+9\cdot\cdot 20}{3\cdot 29\cdot 28}$, és -1, ha piros, fehér, piros vagy fehér, piros, fehér húzás történik, aminek valószínűsége $\frac{9\cdot 20+2\cdot 19\cdot 10}{3\cdot 29\cdot 28}$. Innen $E\eta_j\eta_{j+1} = 4\cdot\left(\frac{1\cdot 9\cdot 8+2\cdot 19\cdot 18}{3\cdot 29\cdot 28} - \frac{2\cdot 19\cdot\cdot 10+9\cdot\cdot 20}{3\cdot 29\cdot 28}\right) + \frac{9\cdot 20+2\cdot 19\cdot 10}{3\cdot 29\cdot 28}$. Továbbá, Cov $(\eta_j, \eta_{j+1}) = E\eta_j\eta_{j+1} - E\eta_1^2$.

Azt az esetet, amikor a $k > j+1$ kissé nehezebb áttekinteni, mert ekkor 16 különböző húzássorozatot kell figyelembe venni. Némileg egyszerűsíti az áttekintést, ha észrevesszük, hogy az egyes húzássorozatok valószínűsége csak attól függ, hogy hány fehér golyót tartalmaznak. Ezenkívül még azt is nézve, hogy mennyi a különböző esetekben az $\eta_j\eta_k$ változó értéke azt kapjuk, hogy $k > j+1$ esetben $E\eta_j\eta_k = 4\cdot\frac{2\cdot 19\cdot 18\cdot 17}{3\cdot 29\cdot 28\cdot 27} - 8\cdot\frac{2\cdot 19\cdot 18\cdot 10}{3\cdot 29\cdot 28\cdot 27}(8+4)\frac{2\cdot 19\cdot 10\cdot 9}{3\cdot 29\cdot 28\cdot 27} - 8\cdot\frac{9\cdot 8\cdot 20}{3\cdot 29\cdot 28\cdot 27} + 4\cdot\frac{9\cdot 8\cdot 7}{3\cdot 29\cdot 28\cdot 27}$, és a $Cov(\eta_j, \eta_k) = E\eta_j E\eta_k - E\eta_j^2$ érvényes ebben az esetben is. Végül felírjuk, hogyan lehet a fenti képleteket felhasználva felírni a keresett szórásnégyzetet. Vegyük észre, hogy mindegyik Cov (η_j, eta_{j+1}) értéke megegyezik, és ugyanez mondható el a Var η_j illetve Cov $(\eta_j, \eta_k$ kifejezésekre, ha $k > j+1$. Figyelembe véve, hogy melyik tagból hányat kell venni kapjuk, hogy Var $\left(\sum_{j=1}^{9}\eta_j\right) = 9\text{Var}\,\eta_1 + 18\text{Cov}\,(\eta_1, \eta_2) + 54\text{Cov}\,(\eta_1, \eta_3)$.

Házi feladat:

Legyen adva két urna, mindkettőben 10 piros és 20 fehér golyó. Húzzunk ki 10 golyót mind a két urnából, az egyikből visszatevéssel, a másikból visszatevés nélkül. Minden húzáspár alkalmával nyerjünk 2 forintot, ha a két különböző urnából kihúzott golyók színe megegyezik, és veszítsünk 1 forintot, ha ez a két szín különbözik. Számoljuk ki nyereményünk várható értékét és szórásnégyzét.