A november 14-i szeminárium témája
Rövid összefoglaló

1. Legyenek $\xi_1, \dots, \xi_n$ független, a $[\vartheta, \vartheta+1]$ intervallumban egyenletes eloszlású valószínűségi változók valamilyen (ismeretlen) ϑ számmal. Számoljuk ki a $\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k - \frac{1}{2}$ valószínűségi változó várható értékét és szórásnégyzetét.

   Megoldás: $E\xi_k = \int_\vartheta^{\vartheta+1} x\,dx = \vartheta + \frac{1}{2}$, $E\xi_k^2 = \int_\vartheta^{\vartheta+1} x^2\,dx = \vartheta^2 + \vartheta + \frac{1}{3} = \left(\vartheta + \frac{1}{2}\right)^2 + \frac{1}{12}$. Innen $E\xi_k = \vartheta + \frac{1}{2}$, $\operatorname{Var}\xi_k = \frac{1}{12}$. Ezért $E\left(\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k - \frac{1}{2}\right) = \vartheta$, és $\operatorname{Var}\left(\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k - \frac{1}{2}\right) = \frac{1}{12n}$.

   Legyenek $\xi_1, \dots, \xi_n$ független, a $[\vartheta, \vartheta+1]$ intervallumban egyenletes eloszlású valószínűségi változók valamilyen (ismeretlen) ϑ számmal. Rendezzük a megfigyelt számokat monoton növekvő $\xi_1^* \le \xi_2^* \le \dots \le \xi_n^*$ sorrendbe, (ezt hívják a statisztikában rendezett mintának), és tekintsük a $\frac{\xi_1^* + \xi_n^*}{2} - \frac{1}{2}$ kifejezést. A következő két feladat célja annak vizsgálata, hogy mennyi ennek a kifejezésnak a várható értéke és szórásnégyzete, azaz ez a kifejezés milyen jól becssüli meg az ismeretlen ϑ paramétert.

2. Lássuk be az előző jelölésekkel, hogy a $(\xi_1^*, \xi_2^*, \dots, \xi_n^*)$ véletlen vektor sűrűségfüggvénye $f(y_1, \dots, y_n) = n!$, ha $\vartheta \le y_1 \le \dots \le y_n \le \vartheta + 1$, és $f(y_1, \dots, y_n) = 0$ egyébként. A (ξ_1^*, ξ_n^*) (azaz a legkisebb és legnagyobb megfigyelt érték együttes) sűrűségfüggvénye $g(x, y) = n(n-1)(y-x)^{n-2}$, ha $\vartheta \le x \le y \le \vartheta + 1$, és $g(x, y) = 0$ egyébként.

3. Lássuk be az előző jelölésekkel, hogy
$$E\left(\frac{\xi_1^* + \xi_n^*}{2} - \frac{1}{2}\right) = \vartheta, \quad \text{és} \quad \operatorname{Var}\left(\frac{\xi_1^* + \xi_n^*}{2} - \frac{1}{2}\right) = \frac{1}{2(n+1)(n+2)}.$$

   A második feladat megoldása: Az $f(y_1, \dots, y_n)$ sűrűségfüggvény a
$$B = \{(y_1, \dots, y_n) \colon \vartheta \le y_1 \le \dots \le y_n \le \vartheta + 1\}$$
   halmazra van koncentrálva, mert a komplementer halmaz valószínűsége nulla. Viszont, ha $A \subset B$, akkor
$$P((\xi_1^*, \dots, \xi_n^*) \in A) = n! P((\xi_1, \dots, \xi_n) \in A) = \int_A n!\,dx_1 \dots dx_n,$$
   ahonnan következik, hogy a $(\xi_1^*, \dots, \xi_n^*)$ vektor sűrűségfüggvénye az $f(y_1, \dots, y_n) = n!$ függvény a B halmazon. Továbbá,
$$g(x, y) = \int f(x, x_2, x_3, \cdots, x_{n-1}, y)\,dx_2 \dots dx_{n-1} = n!\operatorname{Vol}(B_{x,y}^{(n)}),$$

ahol $B_{x,y}^{(n)} = \{(y_2,\dots,y_{n-1})\colon x \le y_2 \le \cdots \le y_{n-1} \le y_n\}$, ahonnan némi számolással $\mathrm{Vol}\,(B_{x,y}^{(n)}) = \dfrac{1}{(n-2)!}(x-y)^{n-2}$. Innen következik a feladat állítása. Egy lehetséges indoklás: Adva a $\{2,\dots,n-1\}$ számok egy tetszőleges τ permutációja legyen $B_{x,y}^{(n)}(\tau) = \{(y_2,\dots,y_{n-1})\colon x \le y_{\tau(2)} \le \cdots \le y_{\tau(n-1)}\}$. Ekkor $\mathrm{Vol}\,(B_{x,y}^{(n)}) = \mathrm{Vol}\,(B_{x,y}^{(n)}(\tau))$ minden τ permutációra, $(n-2)!$ ilyen permutáció van, a $B_{x,y}^{(n)}(\tau)$ halmazok a határuk kivételével diszjunktak, és úniójuk egy $(y-x)^{n-2}$ térfogatú téglatest. Hogyan lehet a 2. feladat eredményét "látni"?

3. A $\xi_1^* - \vartheta - \frac{1}{2}$ és $\vartheta + \frac{1}{2} - \xi_n^*$ valószínűségi változók eloszlása, ezért várható értéke megegyezik. Ezért $E\dfrac{\xi_1^*+\xi_n^*}{2} = \vartheta + \dfrac{1}{2}$. A szórásnégyzet kiszámolásakor feltehetjük, hogy $\vartheta = -\frac{1}{2}$. Ekkor a (ξ_1^*,ξ_n^*) vektor sűrűségfüggvénye, $f(x,y) = n(n-1)(y-x)^{n-2}$, ha $-\frac{1}{2} \le x \le y \le \frac{1}{2}$, $f(x,y) = 0$ egyébként, és a $\mathrm{Var}\,\dfrac{\xi_1^*+\xi_n^*}{2} = \displaystyle\int \left(\frac{x+y}{2}\right)^2 f(x,y)\,dx\,dy$ integrált kell kiszámolnunk.

Alkalmazzuk a $v = x+y$ az x változó helyett rögzített y változó mellett. Vegyük észre, hogy azon a tartományon, ahol az integrandus nem zéró, ott az $f(v,y) = \dfrac{n(n-1)}{4}v^2(2v-y)^{n-2}$ alakú. Alkalmazva az $u = 2v-y$ (lineáris) transzformációt rögzített v váltzóra a kiszámítandó integrált egyszerűbb alakra tudjuk hozni. Az (u,v) síkon a következő integrált kell kiszámítani.

$$\begin{aligned}\mathrm{Var}\,\frac{\xi_1^*+\xi_n^*}{2} &= \int_B \frac{n(n+1)}{8} v^2 u^{n-2}\,du\,dv = \int_0^1 \frac{n(n+1)}{8} u^{n-2}\left(\int_{u-1}^{1-u} v^2\,dv\right)du \\ &= \frac{n(n-1)}{12}\int_0^1 u^{n-2}(1-u)^3\,du,\end{aligned}$$

ahol a B halmaz az $(0,1)$, $(0,-1)$ és $(1,0)$ csúcsok által meghatározott háromszög. Innen

$$\begin{aligned}\mathrm{Var}\,\frac{\xi_1^*+\xi_n^*}{2} &= \frac{n(n-1)}{12}\left[\frac{1}{n-1} - \frac{3}{n} + \frac{3}{n+1} - \frac{1}{n+2}\right] \\ &= \frac{n(n-1)}{12}\left(\frac{3}{(n-1)(n+2)} - \frac{3}{n(n+1)}\right) \\ &= \frac{n(n+1)-(n-1)(n+2)}{4(n+1)(n+2)} = \frac{1}{2(n+1)(n+2)}.\end{aligned}$$

Házi feladat

Legyenek $\xi_1,\dots,\xi_n$ független, egyforma eloszlású valószínűségi változók egyenletes eloszlással valamilyen $[\vartheta,\vartheta+1]$ intervallumban, és legyen $\xi_1^* \le \xi_2^* \le \cdots \le \xi_n^*$ a belőlük készített rendezett minta. Számítsuk ki ξ_n^* várható értékét és szórásnégyzetét.