1-2. STATISZTIKA GYAKORLAT: Rendezett minták

Az $X_1, \dots, X_n$ fae mintából nyert $X_1^* \leq \dots \leq X_n^*$ rendezett mintaelemek általában nem függetlenek és nem is azonos eloszlásúak. Bizonyos alapeloszlások esetén levezetjük eloszlásukat, ill. a differenciák eloszlását.

Olvassák el a tankönyvbeli anyagot X_k^* eloszlásfvéről ($F_{n;k}$) és sűrűségfv-éről, ha létezik ($f_{n;k}$). Egyenletes alapeloszlás esetén pontosan tudjuk karakterizálni X_k^* eloszlását (k, n-től függő beta-eo), azonban a differenciák eloszlásához már az együttes eloszlásokra is szükség van. Ehhez is találhatók formulák a könyvben, melyek kiterjeszthetők tetszőleges abs.folyt. eloszlásokra.

1. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{U}(0,1)$ fae. mintából nyert rendezett minta. Határozzuk meg az
$$U_1 := X_1^*, \quad U_k := X_k^* - X_{k-1}^* \quad (k = 2, \dots, n), \quad U_{n+1} := 1 - X_n^*$$
differenciák eloszlását!

2. (Social distancing) Mi a val.sége, hogy n ember úgy helyezkedik el véletlenszerűen egy 1km-es szakaszon, hogy bármely kettőjük közti távolság legalább d. Itt $0 < d \leq \frac{1}{n-1}$ adott konstans, pl. d =2m COVID idején.

3. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{E}xp(\lambda)$ fae. mintából nyert rendezett minta. Határozza meg az
$$U_1 := X_1^*, \quad U_k := X_k^* - X_{k-1}^* \quad (k = 2, \dots, n)$$
differenciák együttes és egyedi eloszlását is!

4. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{E}xp(\lambda)$ fae. mintából nyert rendezett minta. Határozza meg X_k^* sfv-ét ($k = 1, \dots, n$)!

5. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{E}xp(\lambda)$ fae. mintából nyert rendezett minta. Határozza meg X_k^* várható értékét és szórását ($k = 1, \dots, n$)!

6. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{U}(0,1)$ fae. mintából nyert rendezett minta. Határozza meg X_k^* tetszőleges momentumát!

7. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{U}(0,1)$ fae. mintából nyert rendezett minta. Határozza meg X_k^* varianciáját (szórásnégyzetét)! Hogyan változik ez k-val?

8. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{U}(0,1)$ fae. mintából nyert rendezett minta. Határozza meg X_i^* és X_j^* együttes eloszlását ($1 \leq i < j \leq n$)!

9. Legyen $X_1^* \leq \dots \leq X_n^*$ egy $\mathcal{U}(0,1)$ fae. mintából nyert rendezett minta. Határozza meg $X_i^*(1 - X_j^*)$ várható értékét ($1 \leq i < j \leq n$)!

10. Általánosítás az $\mathcal{U}(0,\theta)$ és $\mathcal{U}(-\theta,\theta)$ eloszlások rendezett mintáira.

Megoldások

1. Kihasználjuk, hogy (0,1)-en egyenletes eloszlásnál egy intervallumba esés val.sége nem más, mint az intervallum hossza. Így

$$G_{n;1}(y) = \mathbb{P}(U_1 < y) = 1 - \mathbb{P}(X_1^* \geq y) = 1 - \prod_{i=1}^{n} \mathbb{P}(X_i \geq y) = 1 - (1-y)^n.$$

$$G_{n;n+1}(y) = \mathbb{P}(U_{n+1} < y) = \mathbb{P}(1 - X_n^* < y) = \mathbb{P}(X_n^* \geq 1-y) = 1 - \prod_{i=1}^{n} \mathbb{P}(X_i < 1-y) = 1-(1-y)^n.$$

$k = 2, \ldots, n$ esetén:

$$G_{n;k}(y) = \mathbb{P}(U_k < y) = 1 - \mathbb{P}(X_k^* - X_{k-1}^* \geq y) = 1 - (1-y)^n,$$

mivel ilyenkor az összes független mintaelem egy y hosszú intervallumon kívül helyezkedik el.

Megjegyzés:

- Tehát a differenciák azonos eloszlásúak, de nem függetlenek. Közös sfv-ük:

$$g_{n;k}(y) = G'_{n,k}(y) = n(1-y)^{n-1}, \quad k = 1, \ldots, n+1.$$

  Hogy nem függetlenek, az következik abból, hogy a sfv-ek szorzata nem ugyanaz, mint az együttes sűrűség, melyet a 3. feladat utáni megjegyzésben láthatunk.
- A megoldás geometriai val.ségekkel (ld. Feller) és teljes indukcióval is kapható.
- $n \to \infty$ esetén, ha $y = \frac{z}{n}$, akkor

$$\lim_{n\to\infty} [1 - (1 - \frac{z}{n})^n] = 1 - e^{-z},$$

  ami az exponenciális eloszlásfv. Ez azt jelenti, hogy a differenciák n-szerese aszimptotikusan 1 paraméterű, míg a differenciák aszimptotikusan n paraméterű ($\frac{1}{n}$ várható értékű) exponenciális eloszlást követnek.

2. Kedvező esetben az $X_1^* \leq \cdots \leq X_n^*$ rendezett $\mathcal{U}(0,1)$ fae. mintából nyert mintában a szomszédosak közt legalább d távolság van. Pl. az első ember kivételével mindenki visz egy d hosszúságú dárdát és azzal távol tartja az előtte levőt. Ez azt jelenti, hogy $X_1, \ldots, X_n$ egymástól függetlenül egy $1-(n-1)d$ hosszú szakaszon oszlik el, melynek val.sége $[1-(n-1)d]^n$.

3. Egy analitikus és egy logikai megoldást is ismertetünk.

- $(X_1^*, \ldots, X_n^*)$ együttes eloszlása (ld. tk.):

$$f(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = n! \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, \quad 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n.$$

Tudnunk kell, hogy a $t : \mathbf{X} \to \mathbf{Y}$, **régi→új**, $\mathbb{R}^n \to \mathbb{R}^n$ transzformáció invertálható t esetén $\mathbf{X}$ sfv-ét ($f(\mathbf{x})$) $\mathbf{Y}$ sfv-ébe ($g(\mathbf{y})$) viszi át:

$$g(\mathbf{y}) = f(t^{-1}(\mathbf{y})) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|, \tag{1}$$

ahol az utolsó tényező az inverz Jacobi mátrix nem-0 determinánsának abszolút értéke. Az inverz Jacobi mátrix ij eleme $\frac{\partial x_i}{\partial y_j}$.
Esetünkben az inverz transzformációt a következő koordinátafv-ek írják le:

$$\begin{aligned} X_1^* &= U_1 \\ X_2^* &= U_1 + U_2 \\ &\vdots \\ X_n^* &= U_1 + U_2 + \cdots + U_n. \end{aligned}$$

Az inverz Jacobi transzformáció mátrixa:

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix},$$

melynek determinánsa 1. Így $U_1, \dots, U_n$ együttes sfv-e:

$$g(y_1, \dots, y_n) = n!\lambda^n e^{-\lambda \sum_{k=1}^n (n-k+1)y_k} \cdot 1 = \prod_{k=1}^n [(n-k+1)\lambda e^{-\lambda(n-k+1)y_k}].$$

Ebből látható, hogy $U_1, \dots, U_n$ független, exponenciális eloszlásúak csökkenő paraméterrel: $U_k \sim \mathcal{E}xp((n-k+1)\lambda)$, $k = 1, \dots, n$.

- A másik mo az exp. eo örökifjúságát használja.

$$G_{n;1}(y) = \mathbb{P}(U_1 < y) = 1 - \mathbb{P}(X_1^* \geq y) = 1 - \prod_{i=1}^n (1 - F(y) = 1 - [(1-(1-e^{-\lambda y})]^n = 1 - e^{-n\lambda y},$$

azaz $U_1 \sim \mathcal{E}xp(n\lambda)$.
Gondoljuk azt, hogy 0 időtől n örökifjú élettartamú izzólámpát üzemeltetünk párhuzamosan. Ekkor $U_1 = X_1^*$ a leghamarabb kiégő izzó élettartama. A többi $n-1$ izzó ezután függetlenül üzemel ugyanolyan esélyekkel. A közülük leghamarabb kiégő élettartama ettől számítva

$$U_2 = X_2^* - X_1^* \sim \mathcal{E}xp((n-1)\lambda)$$

a fentiek szerint és független U_1-től, s.í.t, a k-adiknak kiégő izzó élettartama

$$U_k = X_k^* - X_{k-1}^* \sim \mathcal{E}xp((n-k+1)\lambda),$$

és független $U_1, \dots U_{k-1}$-től, $k = 2, \dots, n$.

Megjegyzés:

- Tehát a differenciák függetlenek, de nem azonos eloszlásúak.

- A fenti (1) képlettel, az 1. feladatbeli $\mathcal{U}(0,1)$ rendezett mintából nyert differenciák együttes sfv-e:

$$g(y_1, \dots, y_n) = n! \cdot 1 \cdot 1 = n!,$$

ami konstans, ellentétben az 1. feladat megjegyzésében látható $g_{n;k}$ sfv-ek szorzatával. Tehát a differenciák ott nem függetlenek (már csak azért sem, mert összegük 1, de az első n sem független).

4. A tanult trnszformációs képlettel X_k^* sfv-e a (0,1)-en egyenletes k-adik rendezett mintaelem beta-sfv-ébe helyettesítve az exponenciális elozlás- és sfv-t kapható:

$$f_{n;k}(x) = n\binom{n-1}{k-1}(1-e^{-\lambda x})^{k-1} \cdot (e^{-\lambda x})^{n-k} \cdot \lambda e^{-\lambda x},$$

$k = 1, \dots, n$, ha $x > 0$ és 0 különben.

5. Mivel $X_k^* = \sum_{i=1}^k U_i$ függetlenek összege:

$$\mathbb{E}(X_k^*) = \sum_{i=1}^k \frac{1}{(n-k+1)\lambda} = \frac{1}{\lambda}\sum_{i=1}^k \frac{1}{n-k+1}, \quad k = 1, \dots, n$$

és

$$\mathbb{D}^2(X_k^*) = \sum_{i=1}^k \frac{1}{(n-k+1)^2\lambda^2} = \frac{1}{\lambda^2}\sum_{i=1}^k \frac{1}{(n-k+1)^2}, \quad k = 1, \dots, n.$$

$\mathbb{D}(X_k^*)$ ennek gyöke.

Megjegyzés: $n \to \infty$ esetén az $\mathbb{E}(X_n^*)$-t definiáló sor nem konvergens ($\ln n$ nagyságrendű), míg a $\mathbb{D}^2(X_n^*)$-t definiáló sor konvergens (π-től függő összeggel).

6. Mivel $X_k^* \sim \mathcal{B}eta(k, n-k+1)$, ezért

$$\mathbb{E}(X_k^*)^s = n\binom{n-1}{k-1}\int_0^1 u^s u^{k-1}(1-u)^{n-k}\,du = n\binom{n-1}{k-1}\int_0^1 u^{s+k-1}(1-u)^{(n+s)-(s+k)}\,du$$

$$= \frac{n\binom{n-1}{k-1}}{(n+s)\binom{n+s-1}{s+k-1}} = \frac{\binom{s+k-1}{s}}{\binom{n+s}{s}}, \qquad k = 1, \dots, n; \quad s = 1, 2, \dots.$$

7. Következésképp

$$\mathbb{E}(X_k^*) = \frac{k}{n+1}$$

$$\mathbb{E}(X_k^*)^2 = \frac{k(k+1)}{(n+1)(n+2)}$$

$$\mathbb{D}^2(X_k^*) = \mathbb{E}(X_k^*)^2 - \mathbb{E}^2(X_k^*) = \frac{k(n-k+1)}{(n+1)^2(n+2)} \qquad (k = 1, \dots, n).$$

Ebből látható, hogy

$$\mathbb{D}^2(X_k^*) = \mathbb{D}^2(X_{n-k+1}^*), \quad 1 \le k \le \frac{n}{2},$$

és a szimmetrikus szórásnégyzetek növekednek k-val (a fenti számláló folytonos megfelelője az $x(n-x)$ fv, melynek maximuma $\frac{n}{2}$-ben van).

8. Legyen $X_1^* \le \dots \le X_n^*$ egy $\mathcal{U}(0,1)$ fae. mintából nyert rendezett minta. $1 \le i < j \le n$ indexekre X_i^* és X_j^* együttes sfv-e az $r=2$, $k_1 = i$, $k_2 = j$ szereposztásban:

$$g_{n;i,j}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} x^{i-1}(y-x)^{j-i-1}(1-y)^{n-j}, \quad 0 \le x \le y \le 1.$$

9. Ezzel

$$\mathbb{E}(X_i^*(1-X_j^*)) = \int_0^1 \int_0^y x(1-y) g_{n;i,j}(x,y)\,dx\,dy =$$
$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_0^1 \int_0^y x^i (y-x)^{j-i-1}(1-y)^{n-j+1}\,dx\,dy =$$
$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \Big/ \frac{(n+2)!}{i!(j-i-1)!(n-j+1)!} \int_0^1 \int_0^y g_{n+2;i+1,j+1}(x,y)\,dx\,dy =$$
$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \cdot \frac{i!(j-i-1)!(n-j+1)!}{(n+2)!} = \frac{i(n-j+1)}{(n+1)(n+2)}.$$

A poén az volt, hogy a második sorbeli sűrűség egy $g_{n+2;i+1,j+1}(x,y)$ sűrűség ‘érdemi része’ volt, a normáló tényező nélkül. Így integrálja (a tartón) a normáló tényező reciproka.

Ebből $\mathbb{E}(X_i^* X_j^*)$, továbbá X_i^* és X_j^* kovarianciája is kiszámolható.

10. A mondottak alkalmazhatók az $X \sim \mathcal{U}(0,\theta)$ eloszlásra az $Y = \frac{X}{\theta} \sim \mathcal{U}(0,1)$ és az $X \sim \mathcal{U}(-\theta,\theta)$ eloszlásra az $Y = \frac{X+\theta}{2\theta} \sim \mathcal{U}(0,1)$ lineáris transzformációkkal.