

A november 7-i szeminárium témája

Rövid összefoglaló

Feladatok:

1. Legyenek $\xi_1, \dots, \xi_n$ független, egyforma eloszlású valószínűségi változók egyenletes eloszlással valamilyen $[\vartheta, \vartheta + 1]$ intervallumban, és legyen $\xi_1^* \leq \xi_2^* \leq \dots \leq \xi_n^*$ a belőlük készített rendezett minta. Őrizzük meg csak a ξ_1^* , és ξ_n^* valószínűségi változókat, és ha $\xi_1^* = x$, $\xi_n^* = y$ akkor készítsünk el egy az $[x, y]$ intervallumban $n - 2$ független, egyforma eloszlású valószínűségi változók által meghatározott $\eta_2^* \leq \eta_3^* \leq \dots \leq \eta_{n-1}^*$ rendezett mintát. Lássuk be az október 31-i gyakorlat harmadik feladatának segítségével, hogy a $\xi_1^* \leq \eta_2^* \leq \dots \leq \eta_{n-1}^* \leq \xi_n^*$ valószínűségi változók szintén egy rendezett mintát alkotnak az $[a, b]$ intervallumban.

Megjegyzés: Az η_j^* valószínűségi változók konstrukciójához nem volt szükségünk az $[a, b]$ intervallum végpontjainak az ismeretére. A matematikai statisztikában ez a tulajdonság valamint a feladat állítása felel meg annak a ténynek, hogy amennyiben az $[a, b]$ intervallum végpontjait akarjuk becsülni az $\xi_1^* \leq \xi_2^* \leq \dots \leq \xi_n^*$ rendezett minta segítségével, akkor a ξ_1^* és ξ_n^* ad meg minden értékes információt, ezt fejezik ki úgy, hogy ξ_1^* és ξ_n^* elégséges statisztikát ad.

Megoldás: Elég azt belátni, hogy tetszőleges (mérhető) $A \in R^2$ és $B \in R^{n-2}$ halmazra

$$P((\xi_1^*, \xi_2^*) \in A, (\xi_2^*, \dots, \xi_{n-1}^*) \in B) = P((\xi_1^*, \xi_2^*) \in A, (\eta_2^*, \dots, \eta_{n-1}^*) \in B)$$

Viszont a feltételes valószínűség definíciója alapján

$$\begin{aligned} &P((\xi_1^*, \xi_2^*) \in A, (\xi_2^*, \dots, \xi_{n-1}^*) \in B) \\ &= \int P((\xi_2^*, \dots, \xi_{n-1}^*) \in B | \xi_1^* = x, \xi_n^* = y) f(x, y) dx dy, \end{aligned}$$

ahol $f(\cdot, \cdot)$ a (ξ_1^*, ξ_n^*) vektor sűrűségfüggvénye. Másrészt

$$P((\xi_1^*, \xi_2^*) \in A, (\eta_2^*, \dots, \eta_{n-1}^*) \in B) = \int P_{x,y}(\eta_2^*, \dots, \eta_{n-1}^*) \in B) f(x, y) dx dy,$$

ahol $P_{x,y}((\eta_2^*, \dots, \eta_{n-1}^*) \in B)$ annak a valószínűsége, hogy $n - 2$ az $[x, y]$ intervallumban egyenletes eloszlású valószínűségi változó által meghatározott rendezett minta beleesik a B halmazba. Viszont az idézett feladat eredménye szerint

$$P_{x,y}((\eta_2^*, \dots, \eta_{n-1}^*) \in B) = P((\xi_2^*, \dots, \xi_{n-1}^*) \in B | \xi_1^* = x, \xi_n^* = y),$$

és innen következik a feladat állítása.

2. Legyenek $\xi_1, \dots, \xi_n$ független, egyforma eloszlású valószínűségi változók egyenletes eloszlással valamilyen $[\vartheta, \vartheta + 1]$ intervallumban, és legyen $\xi_1^* \leq \xi_2^* \leq \dots \leq \xi_n^*$

a belőlük készített rendezett minta. Mutassuk meg, hogy $E \frac{\xi_1^* + \xi_n^*}{2} = \vartheta + \frac{1}{2}$,
 $\text{Var} \frac{\xi_1^* + \xi_n^*}{2} = \frac{1}{2(n+1)(n+2)}$.

Megoldás: A $\xi_1^* - \vartheta - \frac{1}{2}$ és $\vartheta + \frac{1}{2} - \xi_n^*$ valószínűségi változók eloszlása, ezért várható értéke megegyezik. Ezért $E \frac{\xi_1^* + \xi_n^*}{2} = \vartheta + \frac{1}{2}$. 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 $-\frac{1}{2} \leq x \leq y \leq \frac{1}{2}$, $f(x, y) = 0$ egyébként, és a $\text{Var} \frac{\xi_1^* + \xi_n^*}{2} = \int \left(\frac{x+y}{2} \right)^2 f(x, y) dx dy$ integrált kell kiszámolnunk.

Alkalmazzuk a $v = x + y$, $u = y - x$ (lineáris transzformációt ennek az integrálnak a kiszámításánál. A transzformáció Jacobianja 2, és az (u, v) síkon a következő integrált kell kiszámítani.

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

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

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

Házi feladat

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

A centrális határeloszlástétel tárgyalása.

Meg lehet adnia az elégséges sőt alkalmas megfogalmazásban szükséges és elégséges feltételét annak, hogy a centrális határeloszlástétel teljesüljön. Ez a következő eredmény:

Centrális határeloszlástétel. Legyen ξ_k , $k = 1, 2, \dots$, független valószínűségi változók sorozata, melyre $E\xi_k = 0$, $\sigma_k^2 = E\xi_k^2 < \infty$, $k = 1, 2, \dots$, és az $s_n^2 = \sum_{k=1}^n \sigma_k^2$,

$n = 1, 2, \dots$, sorozat teljesíti a $\lim_{n \rightarrow \infty} s_n^2 = \infty$ feltételt. Ha ezenkívül a ξ_n , $n = 1, 2, \dots$, sorozat teljesíti a következő úgynevezett Lindeberg feltételt, mely azt követeli meg, hogy minden $\varepsilon > 0$ számra teljesüljön a

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n E\xi_k^2 I(\{|\xi_k| > \varepsilon s_n\}) = 0.$$

reláció, akkor az $\frac{S_n}{s_n}$ sorozat, ahol $S_n = \sum_{k=1}^n \xi_k$ eloszlásban konvergál a standard normális eloszlásfüggvényhez $n \rightarrow \infty$ esetén.

Igaz a fenti állítás megfordítása is: Ha a fenti jelöléseket használva a ξ_k , $E\xi_k = 0$, $E\xi_k^2 = \sigma_k^2$, $k = 1, 2, \dots$, független valószínűségi változókból álló sorozatra $s_n^2 \rightarrow \infty$, $\sup_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \rightarrow 0$, ha $n \rightarrow \infty$, és az $\frac{S_n}{s_n}$ sorozat eloszlásban konvergál a standard normális eloszlásfüggvényhez $n \rightarrow \infty$ esetén, akkor ez a sorozat teljesíti a Lindeberg feltételt.

Jegyezzük meg, hogy ebben a megfordításban nem csak azt követeltük meg, hogy a határeloszlástétel normális legyen, hanem azt is, hogy a "helyes szórással" rendelkezzen. A $\sup_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \rightarrow 0$, ha $n \rightarrow \infty$, feltétel megkövetelése természetes. Egyrészt, ez azt fejezi ki szemléletesen, hogy az egyes összeadandók hatása elhanyagolható, azaz a centrális határeloszlástétel sok kis eredmény összhatásaként áll elő, másrészt e feltételnek teljesülnie kell, mert következménye a Lindeberg feltételnek. Valóban, tetszőleges $\varepsilon > 0$ és $1 \leq k \leq n$ számokra

$$\begin{aligned} \frac{\sigma_k^2}{s_n^2} &= \frac{E\xi_k^2}{s_n^2} = \frac{E(\xi_k^2 I(|\xi_k| \leq \varepsilon s_n))}{s_n^2} + \frac{E(\xi_k^2 I(|\xi_k| > \varepsilon s_n))}{s_n^2} \\ &\leq \varepsilon^2 + \frac{1}{s_n^2} \sum_{k=1}^n E\xi_k^2 I(\{|\xi_k| > \varepsilon s_n\}) \leq 2\varepsilon^2, \end{aligned}$$

ha $n \geq n_0(\varepsilon)$.

Két természetes kérdés:

- a.) Tudunk-e példát adni arra, hogy a centrális határeloszlástétel nem teljesül, és mi van az ilyen példák hátterében?
 - b.) Milyen egyszerűen ellenőrizhető, hasznos feltételeket tudunk adni arra, hogy a Lindeberg feltétel, ezért a centrális határeloszlástétel teljesüljön?
- a.) Legyenek ξ_k , $k = 1, 2, \dots$, független valószínűségi változók, $P(\xi_k = k) = P(\xi_k = -k) = \frac{1}{2k^2}$, $P(\xi_k = 0) = 1 - \frac{1}{k^2}$. Ekkor $E\xi_k = 0$, $E\xi_k^2 = 1$, $s_n^2 = n$. Viszont a Borel–Cantelli lemma alapján, mivel $\sum_{k=1}^{\infty} P(\xi_k \neq 0) < \infty$ egy valószínűséggel csak véges sok k indexre teljesül az, hogy $\xi_k(\omega) \neq 0$, ezért $\sup_{1 \leq n < \infty} |S_n(\omega)| < \infty$,

és $\frac{S_n(\omega)}{s_n} \rightarrow 0$ egy valószínűséggel. Lehet rafináltabb példákat is konstruálni.

(Például tekinthetjük független η_k , $k = 1, 2, \dots$, valószínűségi változók sorozatát, melyekre $\eta_{2k} = \xi_k$, és az η_{2k-1} , $k = 1, 2, \dots$ valószínűségi változók minden egyéb valószínűségi változótól független standard normális eloszlású valószínűségi változók. Ekkor a $\frac{1}{\sqrt{n}} \sum_{k=1}^n \xi_k$ véletlen összegek eloszlásban tartanak egy nulla várható értékű, és $\frac{1}{2}$, azaz a vártnál kisebb szórásnégyzetű normális valószínűségi változóhoz.) De minden ellenpélda hátterében az van, hogy bizonyos valószínűségi változók kis valószínűséggel rendkívül nagy értékeket vesznek fel. Ezek a rendkívüli értékek csak nagyon kis mértékben befolyásolják a normalizált összeg eloszlását, de nagyon befolyásolják az összeg szórásnégyzetét.

- b.) A következő célunk az, hogy a Lindeberg feltételre, tehát a centrális határeloszlástételre jól ellenőrizhető és elég általános feltételt adjunk. Ennek érdekében a következő állítást fogjuk felidézni.

Legyen ξ_n , $n = 1, 2, \dots$, független valószínűségi változók sorozata, melyre $E\xi_n = 0$, $\sigma_n^2 = E\xi_n^2 < \infty$, $n = 1, 2, \dots$, $\lim_{n \rightarrow \infty} s_n^2 = \infty$, ahol $s_n^2 = \sum_{k=1}^n \sigma_k^2$. Ez a sorozat teljesíti a Lindeberg feltételt, ha a következő tulajdonságok egyike teljesül.

- a.) $E|\xi_k|^{2+\alpha} < \infty$, minden $k = 1, 2, \dots$ számra valamilyen $\alpha > 0$ konstanssal, és

$$\lim_{n \rightarrow \infty} \frac{\left(\sum_{k=1}^n E|\xi_k|^{2+\alpha} \right)^{2/(2+\alpha)}}{s_n^2} = 0. \text{ Speciálisan, ez a feltétel teljesül akkor, ha } E\xi_k^2 \geq K \text{ valamilyen } K > 0 \text{ számmal minden } k = 1, 2, \dots \text{ indexre, és ezenkívül érvényes az } E \lim_{k \rightarrow \infty} k^{-\alpha/2} E|\xi_k|^{2+\alpha} = 0 \text{ reláció.}$$

- b.) A független ξ_n , $n = 1, 2, \dots$, valószínűségi változók egyforma eloszlásúak. (Tehát független egyforma eloszlású valószínűségi változók sorozata is teljesíti a Lindeberg feltételt, ezért a rájuk vonatkozó centrális határeloszlástétel speciális esete a centrális határeloszlástétel Lindeberg feltételből következő alakjának.)

3. Tekintsünk egy szabályos pénzdarab 10 000 egymás utáni (független) feldobásából származó fej-írás sorozatot. Adjunk becslést a Chebishev egyenlőtlenség segítségével annak valószínűségére, hogy a fej-dobások számának eltérése a várt 5000 számtól legalább 100-zal, illetve legalább 200-zal eltér! Milyen becslést ad ezekre a valószínűségekre a centrális határeloszlástétel?

Megoldás: Vezessük be a következő ξ_j , $1 \leq j \leq 10\,000$ valószínűségi változókat, melyekre $\xi_j = 1$, ha a j -ik dobás eredménye fej, $\xi_j = 0$, ha a j -ik dobás eredménye írás. Ekkor ξ_j , $1 \leq j \leq 10\,000$ független valószínűségi változók, $E\xi_j = \frac{1}{2}$, $\text{Var } \xi_j = \frac{1}{4}$, és a $P\left(\left|\sum_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 100\right)$ és $P\left(\left|\sum_{j=1}^{10000} (\xi_j - E\xi_j)\right| > 200\right)$ valószínűségekre kell

becslét adnunk. A Chebishev egyenlőtlenség az első valószínűségre a $\frac{10000 \cdot \frac{1}{4}}{100^2} = \frac{1}{4}$, a második valószínűségre pedig a $\frac{10000 \cdot \frac{1}{4}}{200^2} = \frac{1}{16}$ felső becslést adja.

A centrális határeloszlástétel szerint $P \left(\frac{\sum_{j=1}^{10000} (\xi_j - E\xi_j)}{\sqrt{10000 \cdot \frac{1}{4}}} > u \right) \sim 1 - \Phi(u)$. Innen

kapjuk, hogy az első valószínűség kiszámításához az $u = \pm \frac{100}{\frac{1}{2} \cdot 100} = \pm 2$ 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. Számoljuk ki egy $(1, p)$ paraméterű ξ negatív binomiális eloszlású valószínűségi változó várható értékét és szórásnégyzetét. (Azaz $P(\xi = k) = (1 - p)p^{k-1}$, $k = 1, 2, \dots$).

Ha ξ $(1, p)$ paraméterű negatív binomiális eloszlású valószínűségi változó, akkor

$$E\xi = \sum_{k=1}^{\infty} k(1-p)p^{k-1} \quad E\xi^2 = \sum_{k=1}^{\infty} k^2(1-p)p^{k-1}$$

Ezen összegek egyik lehetséges kiszámolása: $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$, ha $|x| < 1$. Két egymás utáni deriválással kapjuk, hogy $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$ és $\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$. Innen $\sum_{k=1}^{\infty} kp^{k-1} = \frac{1}{(1-p)^2}$, $\sum_{k=1}^{\infty} k^2p^{k-1} = \sum_{k=1}^{\infty} kp^{k-1} + p \sum_{k=2}^{\infty} k(k-1)p^{k-2} = \frac{1}{(1-p)^2} + \frac{2p}{(1-p)^3}$, és $E\xi = \frac{1}{1-p}$, $E\xi^2 = \frac{1}{1-p} + \frac{2p}{(1-p)^2}$. Ezért $\text{Var } \xi = E\xi^2 - (E\xi)^2 = \frac{p}{(1-p)^2}$.

5. Vegyünk egy olyan pénzdarabot, mely $\frac{2}{3}$ valószínűséggel esik a fej és $\frac{1}{3}$ valószínűséggel az írás oldalra. Ezt a pénzdarabot annyiszor dobjuk fel, ameddig megjelenik 1200 fej dobás. Mi annak a valószínűsége, hogy az elvégzett dobások száma 1680 és 1830 közé esik? Adjunk erre a valószínűségre jó közelítő becslést.

Az elvégzett dobások száma negatív binomiális eloszlású valószínűségi változó $n = 1200$ $p = \frac{1}{3}$ paraméterekkel, és 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, mely a kívánt valószínűséget jó pontossággal kiszámítja a centrális határeloszlástétel segítségével.

Jelölje ξ_j , $2 \leq j \leq 1200$ a $j - 1$ -ik és j -ik fejdobás közötti dobások számát (a j -ik fejdobást beleszámítjuk a $j - 1$ -iket viszont nem számítjuk bele a 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 $n = 1$, $p = \frac{1}{3}$ paraméterekkel, és minket a $P(1680 < \xi_1 + \dots + \xi_{1200} < 1830)$ valószínűség érdekel. Megmutattuk a 4. feladatban), hogy $E\xi_j = \frac{1}{1-p} = \frac{3}{2}$, $\text{Var } \xi_j = \frac{p}{(1-p)^2} = \frac{3}{4}$. Ezért a centrális határeloszlástétel alapján $\eta = \xi_1 + \dots + \xi_{1200}$ jelöléssel minket a $P\left(-4 < \frac{\eta - 1200E\xi_1}{\sqrt{1200\text{Var } \xi_1}} < 1\right)$ valószínűség érdekel. A centrális határeloszlástétel alapján $P\left(-4 < \frac{\eta - 1200E\xi_1}{\sqrt{1200\text{Var } \xi_1}} < 1\right) \sim \Phi(1) + \Phi(4) - 1 \sim \Phi(1)$.

Házi feladat.

Egy szabályos dobókockát feldobunk 2700 alkalommal egymástól függetlenül, és összeszámoljuk a páros értékű dobások eredményét. Adjunk jó közelítő becslést a centrális határeloszlástétel és egy normális eloszlástáblázat segítségével arra, hogy ez az összeg 420 és 720 közé esik.

Segítség: Vezessük be a $\xi_j = 2$, 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, $1 \leq j \leq 2700$. Ekkor minket a $P\left(420 < \sum_{j=1}^{2700} \xi_j < 720\right)$ valószínűség érdekel. Számítsuk ki a ξ_j valószínűségi változók várható értékét és szórásnégyzetét, és alkalmazzuk a centrális határeloszlástételt.

6. Legyen birtokunkban 100 lámpa, melyek mindegyike egymástól független időtartamig működik, élettartamuk pedig exponenciális eloszlású $\lambda = \frac{1}{10}$ paraméterrel. (A lámpák élettartamának exponenciális eloszlása természetes feltételezés.) Egy termet bevilágítunk ezen lámpák valamelyikével, majd amikor az kiegétt új lámpát használunk fel. Adjunk jó becslést arra, hogy a lámpák összeélettartama legalább 1150 óra.

Jelölje ξ_j a j -ik lámpa élettartamát, $1 \leq j \leq 100$. Ekkor a $P(\xi_1 + \dots + \xi_{100} > 1150)$ valószínűségre kell jó becslést adnunk, ahol az összegben független exponenciális eloszlású valószínűségi változók szerepelnek $\lambda = \frac{1}{10}$ paraméterrel. Vezessük be az $\eta = \xi_1 + \dots + \xi_{100}$ jelölést.

Kiszámoltuk, hogy jelen esetben $E\eta = mE\xi_1 = \frac{m}{\lambda} = 1000$, $\text{Var } \eta = \frac{m}{\lambda^2} = 10000$ ($m = 100$ és $\lambda = \frac{1}{10}$ választással). Ezért a centrális határeloszlástétel szerint $\frac{\eta - E\eta}{\sqrt{\text{Var } \eta}} = \frac{\eta - 1000}{100}$ jó közelítéssel standard normális eloszlású valószínűségi változó, és $P(\xi_1 + \dots + \xi_{100} > 1150) = P\left(\frac{\eta - E\eta}{\sqrt{\text{Var } \eta}} > 1.5\right) \sim 1 - \Phi(1.5)$.