A november 21-i szeminárium témája
Rövid összefoglaló

1. Egy államban, (nevezzük az egyszerűség kedvéért Floridának,) 5 000 000 választó választ két párt (hívjuk ezeket mondjuk republikánus és demokrata pártnak) jelöltje között. Tegyük fel, hogy a választók egymástól függetlenül $\frac{1}{2}$ valószínűséggel választják valamelyik párt jelöltjét. Mi annak a valószínűsége, hogy a két jelölt által összegyüjtött szavazatok különbsége nem haladja meg a háromszázat.

   Az egyik jelöltre leadott szavazatok száma közelítőleg normális eloszlású $\frac{1}{2}\cdot 5\,000\,000$ várható értékkel és $\frac{1}{4}\cdot 5\,000\,000$ szórásnégyzettel. Annak a valószínűsége, hogy a két jelöltre adott szavazatok különbsége kisebb mint 300 megegyezik annak a valószínűségével, hogy az egyik jelöltre adott szavazatok száma a szavazatok várható értékétől kevesebb mint 150-nel tér el, ez pedig körülbelül, $\Phi\left(\frac{150}{\sqrt{\frac{1}{4}\cdot 5000000}}\right)-\Phi\left(-\frac{150}{\sqrt{\frac{1}{4}\cdot 5000000}}\right)=2\Phi\left(\frac{6\sqrt{5}}{100}\right)-1\sim 2\Phi(0.124)-1\sim 0.1.$

   Megjegyzés: Érdemes megjegyezni, hogy igaz a centrális határeloszlás tétel következő lokális változata is a binomiális eloszlásra. Ha ξ binomiális eloszlású valószínűségi változó n és p paraméterekkel, akkor ξ várható értéke $mn=np$, szórásnégyzete $n\sigma^2=np(1-p)$, és $P(\xi=k)=P\left(\frac{\xi-nm}{\sqrt{n}\sigma}=\frac{k-nm}{\sqrt{n}\sigma}\right)\sim\frac{1}{\sqrt{n}\sigma}\varphi\left(\frac{k-nm}{\sqrt{n}\sigma}\right)=\frac{1}{\sqrt{2\pi n}\sigma}\exp\left(\frac{(k-nm)^2}{2n\sigma^2}\right)$, ahol $\varphi(x)=\frac{1}{2\pi}e^{-x^2/2}$, a standard normális eloszlás sűrűségfüggvénye. Az előbb tekintett esetben $n=5\,000\,000$, $p=\frac{1}{2}$, $-150\le k\le 150$, valószínűségeket kell kiszámítani, és k-ra összegezni. Ebben az esetben a normális sűrűségfüggvényt a nulla közelében kell vennünk, és a kívánt valószínűesegre jó becsés $\frac{1}{\frac{1}{2}\sqrt{2\pi\cdot 5000000}}\cdot 300\sim\frac{6}{10\sqrt{10\pi}}\sim\frac{1}{10}$.

   2. Megjegyzés: Valójában az első feladatban tárgyalt modell némileg irreális. Általában vannak csoportok, melyeknek azonos a véleményük. Tegyük fel, hogy a különböző csoportokban levő emberek véleménye független, de az egyes csoportokban levő emberek véleménye megegyezik. Kérdés: Hogyan befolyásolja ez a tény annak valószínűségét, hogy rendkívül szoros választási eredmény szülessen? Be fogjuk látni, hogy e tény figyelembevétele csökkenti a szoros választási eredmény valószínűségét.

   A fenti kérdés azzal függ össze, hogy ha különböző embereknek azonos a véleményük, akkor az egyik jelöltre leadott (véletlen, közel normális eloszlású) számának a szórása nő vagy csökken. A feladatot úgy is reprezentálhatjuk, hogy egyes csoportokból kijelölünk egy embert, az annyi szavazatot ad le, mint amennyi a csoport tagjainak a száma, a többi ember nem szavaz. Az, hogy ebben az esetben nem változik az egyik jelöltre leadott szavazatok várható értéke, de növekszik annak szórása, következik az alábbi feladat eredményéből.

2. Legyen $\xi_1, \ldots, \xi_n$ független, egyforma eloszlású valószínűségi változók sorozata, melyeknek nem ismerjük a várható értékét. Lássuk be, hogy e valószínűségi változók tetszőleges súlyozott átlaga, azaz a $T_n(\xi_1, \ldots, \xi_n) = \dfrac{\sum_{k=1}^{n} a_k \xi_k}{\sum_{k=1}^{n} a_k}$, $\sum_{k=1}^{n} a_k \neq 0$ a ξ_1 valószínűségi változó várható értékének torzítatlan becslése, melynek szórásnégyzete akkor a legkisebb, ha $a_1 = a_2 = \cdots = a_n$.

Megoldás: $E\dfrac{\sum_{k=1}^{n} a_k \xi_k}{\sum_{k=1}^{n} a_k} = \dfrac{\sum_{k=1}^{n} a_k E\xi_1}{\sum_{k=1}^{n} a_k} = E\xi_1$, és ez jelenti azt, hogy a várható érték torzítatlan becslését írtuk fel. Másrészt

$$\operatorname{Var} \frac{\sum_{k=1}^{n} a_k \xi_k}{\sum_{k=1}^{n} a_k} = \frac{\sum_{k=1}^{n} a_k^2 E\xi_1}{\left(\sum_{k=1}^{n} a_k\right)} = \frac{\sum_{k=1}^{n} a_k^2}{\left(\sum_{k=1}^{n} a_k\right)^2} \operatorname{Var} \xi_1 .$$

Ezért, ha $a_1 = a_2 = \cdots = a_n$, akkor $\operatorname{Var} \dfrac{\sum_{k=1}^{n} a_k \xi_k}{\sum_{k=1}^{n} a_k} = \dfrac{\operatorname{Var} \xi_1}{n}$. Másrészt, az úgynevezett Schwarz egyenlőtlenség alapján $\sum_{k=1}^{n} a_k^2 \leq \dfrac{1}{n} \left(\sum_{k=1}^{n} a_k\right)^2$, ahonnan

$$\operatorname{Var} \frac{\sum_{k=1}^{n} a_k \xi_k}{\sum_{k=1}^{n} a_k} \leq \frac{\operatorname{Var} \xi_1}{n}$$

az általános esetben.

3. A múlt órán láttuk, hogy ha egy a $[\vartheta, \vartheta + 1]$ intervallumban egyenletes eloszlású valószínűségi változó várható értékét akarjuk becsülni úgy, hogy a legnagyobb és legkisebb megfigyelt mintaelem átlagát vesszük, akkor kisebb szórású torzítatlan becslést kapunk a várható értékre mint az összes valószínűségi változó átlaga. Miért nem mond ellent ez az eredmény a előző feladat eredményének.

4. Legyen birtokunkban 100 lámpa, melyek mindegyike egymástól független időtartamig működik, élettartamuk pedig exponenciális eloszlású. Azt a hipotézist akarjuk ellenőrizni, hogy a lámpák élettartama legalább 10 óra, azaz az exponenciális eloszlás paramétere kisebb vagy egyenlő mint $\lambda = \dfrac{1}{10}$. Egy termet bevilágítunk ezen

lámpák valamelyikével, majd amikor az kiégett új lámpát használunk fel. Megfigyeljük, hogy mennyi a lámpák összélettartama. Ennek alapján akarjuk eldönteni, hogy hipotézisünk helyes-e. Természetes úgy dönteni, hogy amennyiben a lámpák elég sokáig égnek, akkor elfogadjuk a hipotézist, ha pedig nem akkor elutasítjuk. Mekkora élettartam esetén fogadjuk el a hipotézist, ha azt akarjuk, hogy a hipotézis teljesülése esetén legalább 0.95 valószűséggel fogadjuk azt el. (A matematikai statisztika szokásos terminológiájával ezt a követelményt úgy fogalmazzák meg, hogy az elsőfajú hiba legyen kisebb mint 0.05.)

Jelölje ξ_j a j-ik lámpa éléttartamát, $1 \le j \le 100$, és adjunk jó becslést $P(\xi_1 + \cdots + \xi_{100} > x)$ valószínűségre ha az összegben független exponenciális eloszlású valószínűségi változók szerepelnek $\lambda = \frac{1}{10}$ paraméterrel. (Ez a hipotézis teljesülése esetén a a legkellemetlenebb eset.) Vezessük be az $\eta = \xi_1 + \cdots + \xi_{100}$ jelölést.

Kiszámoltuk, hogy jelen esetben $E\eta = mE\xi_1 = \dfrac{m}{\lambda} = 1000$, $\operatorname{Var}\eta = \dfrac{m}{\lambda^2} = 10000$ ($m = 100$ és $\lambda = \dfrac{1}{10}$ választással). Ezért a centrális határeloszlástétel szerint $\dfrac{\eta - E\eta}{\sqrt{\operatorname{Var}\eta}} = \dfrac{\eta - 1000}{100}$ jó közelítéssel standard normális eloszlású valószínűségi változó, és $P(\xi_1 + \cdots + \xi_{100} > x) = P\left(\dfrac{\eta - E\eta}{\sqrt{\operatorname{Var}\eta}} > x\right) \sim 1 - \Phi\left(\dfrac{x}{100} - 10\right)$. Ezért természetes az x szintet úgy választani, hogy $1 - \Phi\left(\dfrac{x}{100} - 10\right) = 0.95$, és ha az összélettartam nagyobb mint ez a szint akkor elfogadjuk a hipotézist. Innen $\Phi\left(10 - \dfrac{x}{100}\right) = 1 - \Phi\left(\dfrac{x}{100} - 10\right) = 0.95$. Mivel $\Phi(1.645) \sim 0.95$, ezért $x = 835.5$.

5. Ellenőrizni akarjuk, hogy egy pénzdarab szabályos-e. Feldobjuk a pénzdarabot 10 000 alkalommal. Milyen dobáseredmény esetén fogadjuk el a pénzdarabot szabályosnak, ha azt akarjuk, hogy a jó pénzdarabot legalább 0.95 valószínűséggel fogadjuk el jónak?

   Legyen ξ a fejdobások száma. Akkor fogadjuk el a pénzt szabályosnak, ha $|\xi - 5000| < k$, és a k számot úgy választjuk meg, hogy $P(|\xi - 5000| \le k) = 0.95$. $E\xi = 5000$, $\operatorname{Var}\xi = 2500$, ezért a centrális határeloszlástétel alapján azt követeljük meg, hogy $P(|\xi - 5000| \le k) = P\left(\dfrac{|\xi - E\xi|}{\sqrt{\operatorname{Var}\xi}} \le \dfrac{k}{50}\right) \sim 2\Phi\left(\dfrac{k}{50}\right) - 1$. Ezért a $2\Phi\left(\dfrac{k}{50}\right) - 1 = 0.95$ egyenletet kell megoldani. Innen $\Phi\left(\dfrac{k}{50}\right) = 0.975$, és mivel $\Phi(0.975) \sim 1.96$, $k = 98$.

6. Legyen ξ olyan valószínűségi változó, melynek eloszlásfüggvénye, az $F(x) = P(\xi < x)$ függvény, folytonos szigorúan monoton nő valamilyen $a \le x \le b$ intervallumban, mely intervallum végpontjai teljesítik az $F(a) = 0$ és $F(b) = 1$ feltételt. Ekkor az $\eta = F(\xi)$ valószínűségi változó a $[0, 1]$ intervallumban egyenletes eloszlású valószínűségi változó, azaz $P(\eta < x) = x$ minden $0 \le x \le 1$ számra. Megfordítva, ha η egyenletes eloszlású valószínűségi változó a $[0, 1]$ intervallumban, és az $F^{-1}(\cdot)$ függvény az $F(x)$ eloszlásfüggvény inverze, akkor az $F^{-1}(\eta)$ valószínűségi változó $F(x)$ eloszlású.

Megoldás: Legyen ξ F eloszlású valószínűségi változó. Ekkor $P(\eta < x) = P(F(\xi) < x) = P\left(\xi < F^{-1}(x)\right) = F^{-1}\left(F(x)\right) = x$ minden $0 \le x \le 1$ számra, és ez az azonosság az első bizonyítandó állítás. Másrészt ha ξ egyenletes eloszlású a $[0,1]$ intervallumban, akkor $P(F^{-1}(\eta) < x)) = P(\eta < F(x)) = F(x)$.

Megjegyzés: Ez az állítás általánosítható alkalmas módon arra az esetre, amikor az F eloszlásfüggvény nem feltétlenül szigorúan monoton nő. A feladat eredménye számunkra a következő miatt érdekes. Tipikus statisztikai feladat annak a vizsgálata, hogy egy n elemű minta $F(x)$ eloszlású-e, azaz független, egyforma eloszlású $\xi_1, \dots, \xi_n$ valószínűségi változók $F(x)$ eloszlásúak-e. Ez a kérdés ekvivalens azzal, hogy az $\eta_k = F(\xi_k)$ valószínűségi változók egyenletes eloszlásúak-e a $[0,1]$ intervallumban. Tehát elegendő annak a vizsgálatával foglalkozni, hogy egy minta egyenletes eloszlású-e a $[0,1]$ intervallumban.

További vizsgálandó kérdések:

a.) Megfigyelünk egy n elemű mintát, azaz n darab fggetlen egyforma eloszlású valószínűségi változót. Hogyan ellenőrizzük, hogy a minta elemei F eloszlásúak-e?

b.) Egy dobókockáról szeretnénk eldönteni, hogy szabályos-e. Feldobjuk sokszor egymás után. Azt várjuk, hogy ha a kocka szabályos, akkkor körülbelül egyforma sokszor esik minden oldalára. Egy effektiv dobássorozat esetén mikor természetes úgy tekinteni, hogy ez a tulajdonság teljesül?