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, $\Phi\left(\dfrac{150}{\sqrt{\frac{1}{4}\cdot 5000000}}\right)-\Phi\left(-\dfrac{150}{\sqrt{\frac{1}{4}\cdot 5000000}}\right)=2\Phi\left(\dfrac{6\sqrt{5}}{100}\right)-1\sim 2\Phi(0.124)-1\sim 0.1.$

   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 $P(\xi=k)=P\left(\dfrac{\xi-nm}{\sqrt{n}\sigma}=\dfrac{k-nm}{\sqrt{n}\sigma}\right)\sim\dfrac{1}{\sqrt{n}\sigma}\varphi\left(\dfrac{k-nm}{\sqrt{n}\sigma}\right)=\dfrac{1}{\sqrt{2\pi n}\sigma}\exp\left(\dfrac{(k-nm)^2}{2n\sigma^2}\right)$, ahol $\varphi(x)=\dfrac{1}{2\pi}e^{-x^2/2}$, 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 $\dfrac{1}{\frac{1}{2}\sqrt{2\pi\cdot 5000000}}\cdot 300\sim\dfrac{6}{10\sqrt{10\pi}}\sim\dfrac{1}{10}$.

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

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

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

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

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

   az általános esetben.

3. Láttuk korábban, 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 ξ és η két független λ paraméterű exponenciális eloszlású valószínűségi változó. Számoljuk ki $\xi - \eta$ sűrűségfüggvényét.

   Megoldás: Írjuk fel először általánosan $\xi - \eta$ $h(x)$ sűrűségfüggvényét, ha ξ $f(x)$ és η $g(x)$ sűrűségfüggvénnyel rendelkezik. Ekkor $\xi - \eta = \xi + (-\eta)$, $-\eta$ sűrűségfüggvénye

$g^-(x) = g(-x)$, ezért $\xi - \eta$ sűrűségfüggvénye

$$h(x) = \int f(y)g^-(x-y)\,dy = \int f(y)g(y-x)\,dy.$$

Esetünkben $f(x) = g(x) = \lambda e^{-\lambda x}$, ha $x \geq 0$, és $f(x) = g(x) = 0$, ha $x \leq 0$. Ezért

$$f(y)g(y-x) = \begin{cases} \lambda^2 e^{-\lambda(2y-x)} & \text{ha } y \geq 0, \text{ és } y \geq x \\ 0 \quad \text{ha } y < 0, \text{ vagy } y < x \end{cases}.$$

Innen $x \geq 0$ esetében $h(x) = \displaystyle\int_x^\infty \lambda^2 e^{-\lambda(2y-x)}\,dy = \frac{\lambda}{2}e^{-x}$, $x < 0$ esetében $h(x) = \displaystyle\int_0^\infty \lambda^2 e^{-\lambda(2y-x)}\,dy = \frac{\lambda}{2}e^{x}$. E két eredményt úgyis összefoglalhatjuk, hogy $\xi - \eta$ $h(x)$ sűrűségfüggvénye $h(x) = \dfrac{\lambda}{2}e^{-\lambda|x|}$ minden $-\infty < x < \infty$ számra.