A december 12-i szeminárium témája
Rövid összefoglaló

Ezen a gyakorlaton hipotézisvizsgálattal és konfidencia intervallumok meghatározásával foglalkoztunk. Először néhány hipotézisvizsgálati problémát tárgyalunk, majd bevezetjük a hipotézisvizsgálat néhány fontos fogalmát.

1. Azt gondoljuk, (kissé hivatalosabb vagy ha úgy tetszik tudományosabb nyelven azt mondjuk, hogy az a hipotézisünk), hogy egy pénzdarab szabályos. Ellenőrizzük ezt a hipotézist.

   A természetes és gyakorlatban használt módszer abban áll, hogy feldobjuk a pénzdarabot sokszor, és ha a dobások körülbelül fele fej, akkor elfogadjuk a hipotézist, ha nem, akkor elutasítjuk. De mit jelent az, hogy körülbelül a fele? E kérdés tisztázásában a centrális határeloszlástétel nyújt számunkra segítséget.

   Dobjuk fel a pénzdarabot n alkalommal, és jelölje S_n a fejdobások számát. Ha hipotézisünk teljesül, akkor $ES_n = \frac{n}{2}$, $\operatorname{Var} S_n = \frac{n}{4}$, és a centrális határeloszlástétel alapján $P\left(\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} < x\right) \sim \Phi(x)$ nagy n számra, ahol $\Phi(\cdot)$ a standard normális eloszlásfüggvényt jelöli. A döntési eljárás, amit követünk a következő: Eldöntjük, hogy milyen biztonsággal akarjuk elérni azt, hogy a hipotézis teljesülése esetén a hipotézist elfogadjuk. Ha eldöntjuk ezt az u szintet (mondjuk, $u = 0.9$, $u = 0.95$ vagy $u = 0.99$, általában ilyen szintet szoktak előírni a gyakorlatban), akkor a centrális eloszlás segítségével választunk olyan k számot, melyre

   $$P\left(\frac{n}{2} - k \le S_n \le \frac{n}{2} + k\right) \sim u,$$

   és a hipotézist akkor fogadjuk el, ha a dobások száma $\frac{n}{2} - k$ és $\frac{n}{2} + k$ közé esik. Ellenkező esetben a hipotézist elutasítjuk. Ez a k szám a centrális határeloszlástétel alapján $k \sim \left[2\Phi^{-1}(2u-1) - 1\right]\frac{\sqrt{n}}{2}$. Természetesen ilyen módon nehezebben tudunk megkülönböztetni egy kissé szabálytalan pénzdarabot egy szabályostól mint egy nagyon szabálytalant.

2. Egy önkormányzat el akarja dönteni, hogy egy út megépítését támogatja-e a lakosság legalább két-harmada, mert csak ebben az esetben kívánja azt megépíteni. Tegyük fel, hogy az emberek valamilyen (ismeretlen) p valószínűséggel támogatják az út építését, és azt akarjuk eldönteni, hogy ez a p valószínűség meghaladja-e a $\frac{2}{3}$ számot. Ennek érdekében megkérdezzük sok ember véleményét. Hány igenlő vélemény esetén döntünk úgy, hogy megvan a kívánt támogatás?

   A természetes módszer ismét a centrális határeloszlástétel alkalmazása. Számítsuk ki annak a valószínűségét, hogy a feltétel teljesülése esetén n megkérdezett emberből k ember támogatja az út építését. Számítsuk ki ennek a valószínűségét abban az esetben, ha $p = \frac{2}{3}$. Ekkor a támogatók S_n számának várható értéke $ES_n = \frac{2}{3}n$, szórásnégyzete $\operatorname{Var} S_n = \dfrac{2n}{9}$, és $P\left(\dfrac{S_n - \frac{2n}{3}}{\sqrt{\frac{2n}{9}}} \geq x\right) \sim 1 - \Phi(x)$. Ha azt akarjuk, hogy annak valószínűsége, hogy $p = \dfrac{2}{3}$ esetén körülbelül u valószínűséggel döntsünk úgy, hogy megvan a szükséges támogatás, akkor $k \geq \dfrac{2n}{3} - \dfrac{\sqrt{2n}}{3}\Phi^{-1}(u)$ támogató esetén fogadjuk el azt, hogy megfelelő számú támogató van. Ha $p > \dfrac{2}{3}$, akkor még nagyobb valószínűséggel fogjuk elfogadni e döntési eljárás esetében, hogy megfelelő számú támogató van.

3. El akarjuk dönteni, hogy egy dobókocka szabályos-e, azaz igaz-e, hogy mindegyik oldalára egyforma $\dfrac{1}{6}$ valószínűséggel esik-e. A kockát feldobjuk n alkalommal, és ha körülbelül ugyanannyiszor esik mindegyik oldalára, akkor szabályosnak tekintjük a kockát, ha nem, akkor szabálytalannak. Megint felmerül a kérdés: Mikor tekinthetjük úgy, hogy a kocka ugyanannyiszor esik mindegyik oldalára?

   Legyen $U_1(n), \dots, U_6(n)$ a megfelelő oldalakra eső dobások száma, és tekintsük a $T_n = \sum\limits_{j=1}^{6} \dfrac{(U_j(n) - \frac{n}{6})^2}{\frac{n}{6}}$ kifejezést. Be lehet látni, hogy az T_n valószínűségi változó $G_n(x) = P(T_n < x)$ eloszlásfüggvénye konvergál az úgynevezett 5-paraméterű χ-négyzet eloszláshoz, azaz a $G^{(5)}(x) = P\left(\sum\limits_{j=1}^{5} \xi_j^2 < x\right)$ eloszlásfüggvényhez, ahol ξ_j, $j = 1, \dots, 5$ független standard normális eloszlású valószínűségi változók. Ez a következő eljárást sugallja. Ha azt akarjuk, hogy egy szabályos dobókockát u körülbelül u valószínűséggel fogadjunk el szabályosnak, akkor számítsuk ki a dobáseredmények segítségével számítsuk ki a $T_n = \sum\limits_{j=1}^{6} \dfrac{(U_j(n) - \frac{n}{6})^2}{\frac{n}{6}}$ kifejezést. Ezután nézzük meg egy táblázatban, hogy az 5 paraméterű $G^{(5)}(\cdot)$ χ-négyzet eloszlásfüggvény, mely x számra teljesíti a $G^{(5)}(x) = u$ egyenletet, és fogadjuk el a kockát szabályosnak, ha $T_n \leq x$, és tekintsük szabálytalannak, ha $T_n > x$.

   A fenti módszerben alkalmazott határeloszlás hátterében is a centrális határeloszlástétel, pontosabban annak több-dimenziós változata áll. Be lehet látni, hogy ha $n \to \infty$ akkor az

$$\left(\frac{(U_1(n) - EU_1(n)}{\sqrt{n}}, \dots, \frac{U_6(n) - EU_6(n)}{\sqrt{n}}\right)$$

   véletlen vektorok eloszlásban konvergálnak egy hat-dimenziós normális eloszlású véletlen vektorhoz, melynek elemei nulla várható értékűek, kovariancia struktúrájuk és eloszlásuk pontosan leírható. A T_n valószínűségi változók határeloszlása meghatározható e hat-dimenziós normális eloszlású véletlen vektor koordinátáinak alkalmas kvadratikus formájának az eloszlásaként. Ezután némi lineáris algebrai ismeret felhasználásával be lehet bizonyítani a kívánt eredményt. Megjegyezzük, hogy az, hogy 5 (és nem 6) szabadságfokú χ-négyzet eloszlású valószínűségi változó jelent meg a határeloszlásban azzal függ össze, hogy bár az $(U_1(n), \dots, U_6(n))$ vektornak hat koordinátája van, de mivel $\sum\limits_{j=1}^{6} U_j(n) = n$, e hat koordináta közül csupán öt "független".

Végül megjegyezzük, hogy az előbb idézett eredménynek igaz a következő általánosítása. Tegyük fel, hogy van rögzített számú k urnánk. Egymás után bedobunk ezekbe egymástól függetlenül n golyót, úgy hogy mindegyik golyó a j-ik urnába p_j valószínűséggel esik, $1 \le j \le k$, $\sum\limits_{j=1}^{k} p_j = 1$. Jelölje $U_j(n)$ a j-ik urnába eső golyók számát, és tekintsük a $T_k(n) = \sum\limits_{j=1}^{k} \frac{(U_j(n) - np_j)^2}{np_j}$ valószínűségi változókat, és ezek eloszlását, ha $n \to \infty$. Be lehet látni, hogy a $T_k(n)$ valószínűségi változóknak létezik határeloszlásuk, ha $n \to \infty$, és ez a $k-1$ paraméterű χ-négyzet eloszlás, melynek eloszlásfüggvénye a $G^{(k-1)}(x) = P\left(\sum\limits_{j=1}^{k-1} \xi_j^2 < x\right)$ eloszlásfüggvény, ahol ξ_j, $j = 1, \dots, k-1$ független standard normális eloszlású valószínűségi változók. Ha $k = 6$, $p_j = \frac{1}{6}$, $1 \le j \le 6$, akkor a fenti eredmény speciális esetként tartalmazza a kockadobás vizsgálatában használt eredményt.

4. Egy gyárban készített lámpa élettartama exponenciális eloszlású valószínűségi változó valamilyen λ paraméterrel, azaz sűrűségfüggvénye $\lambda(x) = \lambda e^{-\lambda x}$, ha $x \ge 0$, $f_\lambda(x) = 0$, ha $x < 0$. Az a feltételezésünk, hogy a lámpák élettartamának a várható értéke legalább 100 óra, azaz $\lambda \le \frac{1}{100}$. (Egy λ paraméterű exponenciális eloszlású valószínűségi várható értéke $\frac{1}{\lambda}$, szórásnégyzete pedig $\frac{1}{\lambda^2}$.) Megvizsgáljuk n lámpa élettartamát, és ennek alapján ellenőrizni akarjuk hipotézisünk helyességét.

   Természetes a hipotézis helyességét a következő módon ellenőrizni. Ha a lámpák összélettartama elég nagy, akkor elfogadjuk a hipotézist, ha nem akkor elutasítjuk. A megfelelő szintet, azt hogy a lámpák mekkora összélettartama esetén fogadjuk el a hipotézist a centrális határeloszlástétel segítségével számíthatjuk ki. Jelölje a lámpák élettartamát ξ_j, $1 \le j \le n$, $S_n = \sum\limits_{k=1}^{n} \xi_j$. A centrális határeloszlástétel szerint $P\left(\frac{S_n - \frac{n}{\lambda}}{\frac{\sqrt{n}}{\lambda}} < x\right) \sim \Phi(x)$. Határozzuk el milyen p valószínűséggel engedjük meg azt, hogy a hipotézis helyessége esetén mégis elutasítjuk azt. A legkellemetlenebb helyes eset az, ha a lámpák élettartama pontosan 100 óra, azaz $\lambda = \frac{1}{100}$. Ebben az esetben a centrális határeloszlástétel segítségével meghatározzuk azt az $x = x(p, n, \lambda)$ számot, $\lambda = \frac{1}{100}$, melyre igaz a követekező állítás: Annak valószínűsége, hogy a lámpák élettartama nagyobb mint x a p számmal egyenlő, ha $\lambda = \frac{1}{100}$. Ha a lámpák összélettartama nagyobb mint ez az x szám akkor elfogadjuk, ellenkező esetben elutasítjuk a hipotézist.

Fogalmazzuk meg a hipotézis-vizsgálat néhány fontos fogalmát, ismertessük a szokásos szóhasználatot. A hipotézisvizsgálatban valamilyen feltevést szoktunk ellenőrizni. Ez a feltevés az, hogy a megfigyelt események, vagy véletlen számok valamillyen valószínűségeloszlást követnek. Azt a hipotézist, melyet ellenőrizni akarunk nevezik null-hipotézisnek. Vannak olyan feladatok, melyekben a null-hipotézis egyetlen eloszlásból áll, ilyen például a fentebb tekintett példák közül az első és a harmadik. Ha ez az eset, akkor azt mondjuk, hogy a null-hipotézis egyszerű. Ha a null hipotézis több lehetséges eloszlást tartalmaz (ilyen például az általunk tekintett m]'asodik és negyedik feladat, ahol tetszőleges $p \geq \dfrac{2}{3}$ valószínűséget, vagy 100 óránal nagyobb élettartamot elfogadunk), akkor összetett hipotézisről beszélünk. Azokat a lehetőségeket, melyek lehetőségét ki akarjuk zárni ellenhipotézisnek nevezik.

A hipotézisvizsgálatban a megfigyelések alapján olyan döntést hozunk, hogy elfogadjuk vagy elutasítjuk-e a null hipotézis teljesülését. Ha a null-hipotézis teljesül, mi mégis úgy döntünk, hogy az nem igaz akkor elsőfajú hibáról beszélünk. Ha a null-hipotézis nem teljesülése esetén mi mégis elfogadjuk azt, akkor másodfajú hibáról beszélünk. Ha az elsőfajú hibát kívánjuk csökkenteni, akkor minél több esetben kell elfogadni, ha a másodfajú hibát akarjuk csökkenteni, akkor minél több esetben kell elutasítani a null-hipotézis teljesülését. Gyakorlatban általában úgy szokott megjelenni a probléma, hogy előírjuk mekkora elsőfajú hibát engedünk meg, és e kikötés mellett a másodfajú hibát próbáljuk minél kisebbé tenni. Ha a null-hipotézis összetett, akkor az elsőfajú hiba előírása azt jelenti, hogy a null-hipotézisban megengedett bármely lehetőség bekövetkezése esetén legfeljebb egy előírt valószínűséggel utasítsuk el a null-hipotézis teljesülését.

A konfidencia intervallum (megbizhatósági intervallum) meghatározása a becsléselméletben megadott becslés pontosságát méri. A formális definició megfogalmazása előtt tekintsük a következő példát.

> Adva egy pénzdarab el akarjuk dönteni, milyen valószínűséggel esik a pénzdarab a fej oldalra. Ennek érdekében feldobjuk a pénzdarabot n alkalommal, és a fejdobások számának relativ gyakoriságával becsüljük a fejdobás valószínűséget. Szeretnénk meghatározni, hogy milyen pontosságú ez a becslés. Ennek érdekében határozzunk meg egy olyan a megfigyelések alapján meghatározható viszonylag rövid intervallumot, melyre igaz az, hogy annak valószínűsége, hogy az (ismeretlen) paraméter ebbe a megfigyelésektől, tehát a véletlentől függő intervallumba esik legalább 0.9. Ez úgy értendő, hogy bármi volt is az ismeretlen $0 \leq p \leq 1$ szám, a megszerkesztett (és közvetett módon a p paramétertől függő) intervallum tartalmazza ezt a p számot.
>
> Vezessük be a következő jelöléseket. Legyen $\xi_j = 1$, ha a j-ik dobás eredménye fej, $\xi_j = 0$, ha a j-ik dobás eredménye írás, $S_n = \sum\limits_{j=1}^{n} \xi_j$. Ekkor ξ_j, $j = 1, \dots, n$, független, egyforma eloszlású valószínűségi változók, $P(\xi_j = 1) = 1 - P(\xi_j = 0) = p$, (ahol p, $0 \leq p \leq 1$, az ismeretlen, becsülendő paraméter). Ezért $E\xi_j = p$, $\operatorname{Var} \xi_j = p(1-p)$, és a centrális határeloszástétel szerint $P\left(\dfrac{S_n - np}{\sqrt{np(1-p)}} < x\right) \to$

$\Phi(x)$, ha $n \to \infty$. Továbbá a nagy számok törvénye szerint $\dfrac{S_n}{n} \Rightarrow p$. Ezért a centrális határeloszlástétel helyett felírhatjuk a következő, számunkra kényelmesebb összefüggést: $P\left(\dfrac{S_n - np}{\sqrt{S_n\left(1-\frac{S_n}{n}\right)}} < x\right) \to \Phi(x)$, ha $n \to \infty$, ahonnan azt kapjuk, hogy $P\left(-x < \dfrac{np - S_n}{\sqrt{S_n\left(1-\frac{S_n}{n}\right)}} < x\right) \sim 2\Phi(x)-1$, ha n elég nagy szám, minden $x >$ 0 számra. Innen $P\left(\dfrac{S_n}{n} - \dfrac{x}{\sqrt{n}}\sqrt{\dfrac{S_n}{n}\left(1-\dfrac{S_n}{n}\right)} < p < \dfrac{S_n}{n} + \dfrac{x}{\sqrt{n}}\sqrt{\dfrac{S_n}{n}\left(1-\dfrac{S_n}{n}\right)}\right)$ $\sim 2\Phi(x) - 1$. Mivel a $2\Phi(x) - 1 = 0.9$ egyenlet megoldása körülbelül $x = 1.65$, ez azt jelenti, hogy

$$P\left(p \in \left[\frac{S_n}{n} - \frac{1.65}{\sqrt{n}}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}, \frac{S_n}{n} + \frac{1.65}{\sqrt{n}}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}\right]\right) \sim 0.9.$$

Ilyen módon megadtunk a p paraméter $\dfrac{S_n}{n}$ becslése körül egy intervallumot, mely nagy n számú pénzdobás esetén rövid, és annak valószínűsége, hogy az igazi p paraméter ebbe a (véletlen) intervallumba esik körülbelül 0.9. Ezzel az intervallummal tulajdonképpen azt mérjük, hogy milyen pontos a becslésünk.

A fenti feladat példát mutatott a konfidenciaintervallum jelentésére. Ezután megfogalmazzuk a konfidenciaintervallum általános definicióját.

Konfindenciaintervallum definiciója. *Legyen adva eloszlásfüggvényeknek egy λ paraméterrel paraméterezett $F_\lambda(\cdot)$ családja, valamint egy F_λ eloszlású $\xi_1, \dots, \xi_n$ minta, azaz független F_λ eloszlású valószínűségi változók sorozata. Rögzítsünk valamilyen p, $0 < p < 1$, számot. Azt mondjuk, hogy egy a $\xi_1, \dots, \xi_n$ mintától függő $A = A(\xi_1, \dots, \xi_n)$, $B = B(\xi_1, \dots, \xi_n)$ végpontú $[A, B]$ intervallum p szintű konfidenciaintervallum, ha minden λ paraméter esetén $P_\lambda(\lambda \in [A, B]) \geq p$. (Itt a λ index a valószínűségben arra utal, hogy a valószínűség az F_λ eloszlástól függ.*