Házi feladat V.

Matematikai statisztika
Elek Péter

Határidő: konzultáció eleje
Email: matstat.pads.2015@gmail.com

1. Feladat. Tegyük fel, hogy n elemű f.a.e. Bernoulli (azaz 0 vagy 1 értéket felvevő) mintánkban az egyesek száma k, a nullák száma pedig $n-k$ lett. Becsülni szeretnénk a p bekövetkezési valószínűséget. Tudjuk, hogy ennek ML- és MM-becslése k/n. Az órán bebizonyítottuk, hogy ha p a priori eloszlása Beta-eloszlás (a,b) paraméterekkel, akkor p a posteriori eloszlása is Beta-eloszlás $(k+a, n-k+b)$ paraméterekkel. Azt is tudjuk (bár ezt nem bizonyítottuk), hogy a (a,b) paraméterű Beta-eloszlás várható értéke $a/(a+b)$, varianciája pedig $ab/\left[(a+b)^2(a+b+1)\right]$. Így p Bayes-becslése az a posteriori eloszlás várható értéke, azaz $(k+a)/(n+a+b)$.

a. Egy dobókockával a hatos dobásának valószínűségét szeretnénk megbecsülni. Kétféle a priori eloszlással fogunk dolgozni. Az egyik az egyenletes eloszlás, amit az $a=b=1$ választással kapunk. A másik pedig egy olyan Beta-eloszlás, amelynek várható értéke 1/6, szórása pedig 0.1. (Ekkor tehát a priori valószínűnek tartjuk, hogy p 1/6 környékén van.) Határozzuk meg azt az (a,b) paraméterpárt, amely ilyen Beta-eloszlást eredményez!

b. Ábrázoljuk a kétféle a priori eloszlás sűrűségfüggvényét! (A Beta-eloszlással kapcsolatos függvények az R programnyelvben: dbeta, pbeta, qbeta, rbeta.)

c. Tegyük fel, hogy $n=10$ és $k=4$ (azaz tízből négyszer dobtunk hatost). Határozzuk meg p a posteriori eloszlását mindkét a priori eloszlás esetén! Ábrázoljuk az a posteriori eloszlások sűrűségfüggvényeit is egy ábrán! Határozzuk meg p Bayes-becslését mindkét a priori eloszlás esetén!

d. Ismételjük meg a megoldást akkor is, ha $n=1000$ és $k=400$. Ábrázoljuk egy ábrán a kétféle a posteriori sűrűségfüggvényt és határozzuk meg p Bayes-becsléseit! Értelmezzük az eredményeket!

2. Feladat. Egy n elemű f.a.e. mintát veszünk egy olyan sokaságból, amelynek szórása $\sigma=14$. Milyen nagynak kell n-nek lennie ahhoz, hogy a várható értékre vonatkozó 95%-os konfidenciaintervallum szélessége 3-nál kisebb legyen?

3. Feladat. Egy kórház meg kívánja becsülni a sürgősségi osztályra egy nap alatt felvett betegek várható számát. Véletlenszerűen kiválasztanak 16 napot, megmérik a felvett betegek számát és a következő eredményeket kapják: mintaátlag $\overline{x}=59,5$ és korrigált tapasztalati szórás $s^* = 6$. Adjunk 95%-os konfidenciaintervallumot a felvett betegek várható számára, feltételezve, hogy a napi felvételek száma normális eloszlást követ!

4. Feladat. Egy új stimuláns hatását vizsgálják a futókra úgy, hogy minden kísérlet során két, véletlenszerűen kiválasztott futó közül pontosan az egyiknek adnak a szerből, majd versenyeztetik a két futót. Összesen 72 kísérletet végeznek, és összeírják, hogy hányszor nyert a stimulánst kapó versenyző. A nullhipotézis szerint a stimulánsnak nincs hatása a nyerési valószínűségre, míg az alternatív hipotézis szerint elméletileg bármilyen irányú lehet a hatás.

a. Határozzuk meg azt a próbát, amellyel 5%-os szignifikanciaszinten dönteni lehet a kérdésről! Mennyi a próba ereje abban az esetben, ha a stimulált futó tényleges nyerési valószínűsége 0,6?

b. Ha a kísérletek során 48-szor nyert a stimulált futó, akkor elvetjük-e a nullhipotézist 5%-os szinten? Határozzuk meg a p-értéket!

5. Feladat. A következő adatok egy olyan véletlenített kísérletből származnak, amely azt vizsgálta, hogy egy, a munkanélkülieknek szóló képzési program hogyan hat a munkanélküliek elhelyezkedési gyorsaságára. Kétszáz munkanélküli vett részt a programban, ők átlagosan 25,4 hét alatt tudtak elhelyezkedni, 10 hét számított szórással. A kontrollcsoport szintén kétszáz emberből állt, ők átlagosan 28 hét alatt és 13 hét szórással tudtak elhelyezkedni.

a. Van-e elegendő bizonyíték arra, hogy a képzési program befolyásolta a munkanélküliség hosszát? Határozzuk meg a próba p-értékét!

b. Határozzuk meg a képzési program hatásának 95%-os konfidenciaintervallumát!

6. Feladat. a. Egy szabályos kocka feldobása esetén az 1, 2,. . . , 6 számok mindegyike 1/6 valószínűséggel következik be. Számítsuk ki a dobott szám várható értékét és szórását! (Megjegyzés: ez már egy korábbi feladatban általánosabban szerepelt.)

b. Egy dobókockát $n = 105$ alkalommal dobtunk fel, és a dobások átlagának $\overline{x} = 4$ adódott. Van-e elegendő bizonyíték arra, hogy a kocka nem szabályos? Határozzuk meg a tesztstatisztika p-értékét!

7. Feladat. Legyen az X valószínűségi változó egyenletes eloszlású a $[0, B]$ intervallumon. Egy n elemű f.a.e. mintát veszünk ebből az eloszlásból: $x_1, x_2, \ldots, x_n$. A mintaelemek maximumát ($M = \max\{x_1, x_2, \ldots, x_n\}$) használjuk annak tesztelésére, hogy a $B = 1$ nullhipotézis teljesül-e, azzal az alternatív hipotézissel szemben, hogy $B > 1$. (Megjegyzés: M eloszlásfüggvényét már tudjuk: $\Pr(M \leq x) = (x/B)^n$ ha $0 \leq x \leq B$.)

a. Döntési szabályunk: elutasítjuk a $H_0 : B = 1$ nullhipotézist a $H_1 : B > 1$ alternatív hipotézissel szemben, ha $M \geq k$, ahol k megfelelő konstans. Határozzuk meg $\alpha = 5\%$ szignifikanciaszint esetén a k értéket!

b. Egy $n = 4$ elemű mintában a maximális mintaelem 0.9-nek adódott: $M = 0.9$. Elutasítjuk H_0-t H_1-gyel szemben 5%-os szinten? Mennyi a tesztstatisztika p-értéke?

c. Elutasítjuk H_0-t, ha $M = 1.05$ a mintában? Mennyi a p-érték ebben az esetben?

d. Legyen továbbra is $n = 4$. Határozzuk meg a konstruált próba erőfüggvényét és a másodfajú hiba valószínűségét (azaz az erőt és a másodfajú hiba valószínűségét minden $B > 1$ esetén)! Ábrázoljuk az erőfüggvényt B függvényeként! Mi a függvény határértéke, ha $B \to 1$ illetve ha $B \to \infty$?

8. Feladat. Legyen X normális eloszlású valószínűségi változó $\mu = 1$ várható értékkel és $\sigma = 2$ szórással, és $x_1, x_2, \ldots, x_n$ f.a.e. minta. A minta alapján a $H_0 : \mu = 1$ nullhipotézist szeretnénk tesztelni a $H_1 : \mu \neq 1$ alternatív hipotézissel szemben $\alpha = 5\%$ szignifikanciaszinten. Tehát most biztosan tudjuk, hogy H_0 teljesül, de mégis tesztelünk, mert különféle próbák tulajdonságait szeretnénk szimulációval vizsgálni. Pontosabban azt vizsgáljuk, hogy mi történik, ha kismintás esetben mégis a nagymintás közelítést használjuk a próba elkészítéséhez: $\frac{\overline{x}-\mu}{s^*/\sqrt{n}} \overset{A}{\sim} N(0,1)$.

a. Legyen $n = 5$. Határozzuk meg a próba elutasítási tartományát (kritikus értékeit) a nagymintás közelítés alapján!

b. Generáljunk $r = 1000$ alkalommal $n = 5$ elemű mintákat az eloszlásból, és végezzük el a próbát minden egyes minta esetén. Határozzuk meg a próba empirikus elutasítási arányát (azaz azon minták arányát, amelyekre elutasítottuk H_0-t). Mennyire tér ez el az 5%-os "névleges" szignifikanciaszinttől?

c. Ismételjük meg ugyanezt $n = 30$ esetén, és határozzuk meg az empirikus elutasítási arányt. Értelmezzük az eredményeket!

d. A fenti elutasítási arányok valójában empirikus szignifikanciaszintek, azaz annak a valószínűségnek a közelítései, hogy elutasítjuk H_0-t, pedig igaz. Számítsuk ki elméleti úton – a használt tesztstatisztikák tényleges elméleti eloszlásai alapján – az elvégzett próbák tényleges szignifikanciaszintjeit, és hasonlítsuk össze azokat az empirikus szignifikanciaszintekkel!