

Sztochasztikus folyamatok 2. gyakorlat

1. Ez a feladat azt mutatja, hogy a konfidenciaintervallumok keresése, valamint a hipotézisvizsgálati próbák szorosan összefüggenek. Egy gépsor üdítőt tölt az üvegekbe, ξ_n mennyiséget az n -dikbe. Tudjuk, hogy az üdítő mennyisége m centiliter várható értékű és 15 centiliter szórású normális eloszlású (azaz $\xi_n \sim N(m, 15^2)$).

Megnéztük k üveg tartalmát, és ezek átlagos töltetét 105 centiliternek találtuk. (Azaz $\bar{\xi}_k = 105$.) Legyen $k = 25$.

Adjunk szimmetrikus konfidenciaintervallumot $1 - \varepsilon = 0.95$ szint mellett az m értékére! {Azaz olyan $[105 - x, 105 + x]$ alakú intervallumot, amelybe 0.95 valószínűséggel beleesik az m érték, vagyis $P(|m - \bar{\xi}_k| \leq x) = 0.95$.}

Keressünk alsó konfidenciahatárt is ugyanilyen szint mellett, azaz adjunk meg olyan x számot, amelyre

$$P(m - \bar{\xi}_k > x) = 0.95$$

teljesül.

Ha $k = 25$ és tudjuk, hogy a szabványos töltet 100 centiliter, akkor vajon 0.98-as (szignifikancia) szint mellett elfogadható-e az a hipotézis, hogy $m = 100$, azaz jól működik a gépsor?

2. Össze akarjuk hasonlítani a Lolka-kóla és a Bolka-kóla szénsavtartalmát. 10 üveg Lolka-kólában átlagosan 9.5 grammot találtunk, 15 üveg Bolka-kólában pedig átlagosan 11 grammot. A Lolka-kóla (korrigált) tapasztalati szórása 0.6 volt, a Bolka-kóláé 0.65. Tudva, hogy a szénsav mennyiségének szórása ugyanannyi mindkét esetben, elfogadható-e 95%-os szinten az a feltételezés, hogy a kétféle kóla szénsavtartalma ugyanannyi?
3. Egy csavartípus szabványos átmérője 10 mm. Megmértünk 15 darabot és az átlagos átmérőt 9.6 mm-nek találtuk, a tapasztalati szórás (korrigált) pedig 1.2 mm volt. Vajon 99%-os szinten elfogadható-e az a hipotézis, hogy a csavarok a szabványnak megfelelnek?
4. A korábbi tapasztalatok szerint 5 féle színben fordulnak elő egy madárfaj egyedei, 0.1; 0.3; 0.1; 0.2; 0.3 valószínűségekkel. Egy újonnan felfedezett populációban el szeretnénk dönteni, vajon ugyanilyen-e a színek eloszlása.

1000 madarat vizsgálunk meg, és a következő eredményeket kaptuk: 105, 292, 80, 221, 302. Döntsük el 90%-os szinten, vajon a feltételezett színeloszlást követik-e az új populáció tagjai !

5. Össze akarjuk vetni Bergengócia északi és a déli felének a személyneveit eredet szerint. Megvizsgáltunk 20 északi és 18 déli családnevet. Azt találtuk, hogy a háromféle eredetű névből (angol, francia, ősberegengóc) északon (8, 5, 7), délen (7, 3, 8) volt. Elfogadható-e 0.9 szinten az a hipotézis, hogy a családnevek eloszlása északon és délen ugyanaz ?
6. Öt autót vizsgálunk. A szervíz előtt mindegyik átlagosan 8 litert fogyasztott. A szervíz után 7.6; 7.4; 8; 7.7; 8.2 fogyasztási értékeket kaptunk. Döntsünk 95%-os szinten arról, hogy az új fogyasztás megegyezik-e a régivel ! Milyen ellenhipotézist célszerű felvenni ? (Ez t próba, de egymintás, hiszen a kétmintásnál a két minta független, márpedig itt ugyanazokat a kocsikat néztük, tehát a korábbi és a mostani mérési adatok NEM függetlenek. Az egyszerűség kedvéért kétoldali t -próbát végzünk. A fogyasztás változások a $\xi_1, \dots, \xi_5$ és azt nézzük meg, hogy az $E\xi_1 = 0$ nullhipotézis teljesül-e.)
7. Döntsünk 99%-os szinten arról, hogy az alább ismertetett minta származhat-e exponenciális eloszlásból !

intervallumok	0-1	1-2	2-3	3-4
gyakoriságok	37	24	26	13
átlagok	0.4	1.3	2.4	3.2

Megoldásvázlat: Tipikus illeszkedésvizsgálat. Először meg kellene becsülni az exponenciális eloszlás paraméterét, λ -t. A múlt félévben szerepelt, hogy ennek maximum-likelihood becslése $\hat{\lambda}_n = n/(\xi_1 + \dots + \xi_n)$ (azaz a mintaátlag reciproka), ami nem meglepő, hiszen $E\xi_1 = 1/\lambda$.

A jelen esetben $37 + 24 + 26 + 13 = 100$ mintaelem van. Az átlagok intervallumonként vannak megadva, tehát a teljes átlag

$$\bar{\xi}_{100} = (0.4 \times 37 + 1.3 \times 24 + 2.4 \times 26 + 3.2 \times 13)/100,$$

ahonnan $\hat{\lambda} = \hat{\lambda}_{100} = 2/3$. A megfigyeléseket 4 csoportba soroljuk, azaz a következő 4 eseményt vesszük: $\{\xi_i \in [n, n+1)\}$, $n = 0, 1, 2$ illetve $\{\xi \in [3, \infty)\}$. Ez utóbbi azért szükséges, hogy teljes eseményrendszer szerepeljen.

A táblázat adataiból ismerjük a fenti 4 osztály megfigyelési gyakoriságait: $\nu_1 = 37, \nu_2 = 24, \nu_3 = 26, \nu_4 = 13$. A fenti 4 esemény elméleti

gyakoriságai:

$$np_i, i = 1, 2, 3, 4, n = 100,$$

$$p_1 = F(1) - F(0), p_2 = \dots, p_4 = F(\infty) - F(3) = 1 - F(3),$$

ahol $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$ az exponenciális eloszlásfüggvény. Kijön a p_i -kre: 0.51, 0.25, 0.12, 0.12. A próbastatisztika:

$$\chi^2 = \sum_{i=1}^4 \frac{(\nu_i - n_i p_i)^2}{n p_i} = \text{sok} \approx 20.$$

Most $4 - 1 = 3$ szab. fokú χ^2 -eloszlás táblázatát kell nézni, mert 4 csoportba osztottuk a megfigyeléseket. Itt $P_{H_0}(\chi^2 > c) = 0.01$ -ből $c = 11.34$, tehát a nullhipotézist el kell utasítani 99%-os szinten.

8. Két csaló hamis kockával játszik, mindegyik a sajátjával. Kipróbáltuk őket és az alábbi gyakoriságokat figyeltük meg:

	1	2	3	4	5	6
I. kocka	27	24	26	23	18	32
II. kocka	18	12	15	21	14	20

Döntsünk 95%-os szignifikanciaszinten arról, hogy vajon azonos-e a két kocka dobásértékeinek eloszlása !

Megoldásvázlat: Homogenitásvizsgálat, a képlet:

$$\chi^2 := \sum_{i=1}^6 \frac{\left(\frac{\nu_i}{n} - \frac{\mu_i}{m}\right)^2}{\frac{\nu_i}{n} + \frac{\mu_i}{m}},$$

ahol ν_i , μ_i az i -dobás gyakorisága az első ill. második kockánál (a táblázat első ill. második sorában álló értékek). $n = 150$ dobást végeztünk az első, $m = 100$ dobást a második kockával. Behelyettesítve 2.2 jön ki, a χ^2 szabadsági foka $6 - 1 = 5$, $P_{H_0}(\chi^2 > c) = 0.05$ -ből 11.07 adódik c -re. Mivel $2.2 < 11.07$, elfogadjuk H_0 -t.

9. Az alábbi táblázat egy évfolyam sztoch. foly. jegyeit mutatja, aszerinti bontásban, hogy a hallgatók sportoltak v. nem. 95%-os szignifikanciaszinten döntsünk arról, hogy vajon a sztoch. foly. jegy független-e a sportolástól !

	2	3	4	5
nem sportolnak	50	35	35	20
sportolnak	10	15	25	10

Megoldásvázlat: Függetlenségvizsgálat, az egyes szempontok szerinti osztályok száma $k = 2$ (sorok) ill. $l = 4$ (oszlopok). A statisztika:

$$T = n \sum_{i=1}^k \sum_{j=1}^l \frac{\left(\nu_{ij} - \frac{\nu_{i\cdot} \nu_{\cdot j}}{n}\right)^2}{\nu_{i\cdot} \nu_{\cdot j}},$$

ahol n a táblázat elemeinek összege, ν_{ij} -k a táblázatban álló számok, $\nu_{i\cdot}$ a sorösszegek, $\nu_{\cdot j}$ az oszlopösszegek.

Kijön 9.13, $P_{H_0}(\chi^2 > c) = 0.05$ megoldása $c = 7.81$ mivel a próba (aszimptotikus) eloszlása $(r - 1)(s - 1) = 3$ szab. fokú χ^2 -eloszlás. Tehát el kell utasítani a H_0 hipotézist miszerint a jegyek ill. a sportolás független lenne.

10. Kötelező házi feladat: tegyük fel, hogy egy tetraéder alakú dobókockával 100-szor dobtunk. Találj ki olyan gyakoriságokat (azaz ν_i , $i = 1, 2, 3, 4$ egészeket, melyekre $\sum_i \nu_i = 100$) amelyekre igaz, hogy 99%-os szinten elfogadható az a hipotézis, hogy a tetraéder szabályos. Adj meg olyan gyakoriságokat is, amelyekre ez a hipotézis nem fogadható el az adott szinten ! (Őrizkedjete a nagyon triviális példáktól, mint pl. mind a 100-szor 4-es lett. Két egyforma megoldás ne legyen.)