VII. gyakorlat

2010. XII. 08.

ZH példa

1. a, X: foglaltsági idő hossza.

Ekkor a teljes várható érték tétellel

0, 1, 2 új igény tud beérkezni

$$EX = \frac{1}{2}\cdot 1 + \frac{1}{4}\left(1 + EX\right) + \frac{1}{4}\left(1 + 2EX\right)$$

nincs új beérkezés

újrakezdődik az egész kiszolgálás, s olyan eloszlása van, mint X-nek

Ezt az egyenletet megoldva $EX = 4$

b,

$$g(z) = E(z^X)$$

$$g(z) = \frac{1}{2}z + \frac{1}{4}z\cdot g(z) + \frac{1}{4}z\cdot\left(g(z)\right)^2$$

csak 1 időegység a kiszolgálás

még bejön egy ugyanolyan eloszlású véletlen idő

Ezt aztán meg lehet oldani.

3. Az a kérdésre adott válasz függeni fog n-től.

$P(S_n > C)$ — Hoeffding.

~~összes videó~~

és ~~teljes sáv~~ ~~átlátott~~

intégrál

$S_n = 100$ videófolyam + n db videofolyam együttes sávszél.

$$= P\Big(S_n - ES_n \geq \underbrace{C - ES_n}_{t}\Big) \leq \exp\left(-\frac{t^2}{\sum_k (b_k - a_k)^2}\right)$$

$C = 145\,000$

$ES_n = 100 \cdot 1200 + n \cdot 250$

így $t = C - ES_n = 25000 - n \cdot 250 =$

$= 250(100 - n)$

$b_1 \to \quad \exp\left(-\frac{2t^2}{100 \cdot 400^2 + n \cdot 100^2}\right) = 10^{-8}$ kell megoldani.

(1.) A mérésből van 10 db minta:

$X_1, X_2, \dots, X_{10} \sim \text{BINOM}(10, p)$ függetlenek

↓ van egy $x_1, x_2, \dots, x_{10}$, ami a megvalósulás =

$= 2, 3, 2, \dots$ meg van adva.

Ezekből akarunk információt nyerni a p paraméterről.

Jó módszer: maximum likelihood-becslés.

A likelihood-fgv fog függeni p-től, de a mintaelemek konkrét megvalósulásától is:

~~$P_p($~~

$L_p(x_1, x_2, \dots, x_{10}) = P_p(X_1 = x_1, X_2 = x_2, \dots X_{10} = x_{10})$

arra gondolok, hogy X-el elamelyik p paraméterű binomiális.

Ha ez eloszlás $X_1, X_2, \dots, X_{10}$ binomiális eloszlás, akkor meg tudjuk mondani, hogy mi a valószínűsége, hogy ezek az értékek jönnek ki. Megnézzük, hogy melyiknél legnagyobb a valószínűsége, hogy ez jön ki → max. likelihood elv.

Mivel binomiális eloszlás van és függetlenek, ezért

$$P_p(X_1=x_1, X_2=x_2, \dots, X_{10}=x_{10}) = P(X_1=x_1)\cdot P(X_2=x_2)\cdots P(X_{10}=x_{10})$$

$$P(X_1=x_1) \underset{\uparrow}{=} \binom{10}{2}\cdot p^2 (1-p)^8 \quad \text{, mert } x_1=2$$

binomiális eloszlás

$X_1 \sim BINOM(10,p)$

$$P(X_i=x_i) = \binom{10}{x_i}\cdot p^2 (1-p)^{10-x_i}$$

Vagyis

$$= \binom{10}{x_1} p^{x_1}(1-p)^{10-x_1}\cdot\binom{10}{x_2} p^{x_2}(1-p)^{10-x_2}\cdots\binom{10}{x_{10}} p^{x_{10}}(1-p)^{10-x_{10}}$$

Ennek kell a maximuma.

↓ deriválás. De ez így csúnya, így nem ezt a szorzatot deriváljuk, hanem vesszük a logaritmusát (a logaritmus monoton fgv, így ennek és magának van a maximuma):

$$l_p(x_1,x_2,\dots,x_{10}) = \underset{\text{ez ln!}}{\log} L_p(x_1,\dots,x_{10}) =$$

$$= \log\binom{10}{x_1} + \ldots + \log\binom{10}{x_{10}} + (x_1 + x_2 + \ldots + x_{10}) \cdot \log p +$$

$$+ (100 - x_1 - x_2 - \ldots - x_{10}) \cdot \log(1-p)$$

úgy lett fel, hogy a különböző típusú tagolat csoportokból

$x_1, x_2, \ldots, x_{10}$ adottak, már csak p-től függ

Ezt már tudjuk maximálni → deriváljuk.

$$\cancel{\frac{d}{dp}\left(l_p = 0 \Rightarrow\right)}$$

$$\frac{d}{dp} l_p(x_1, x_2, \ldots x_{10}) = 0$$

$$\Downarrow$$

$$\frac{x_1 + x_2 + \ldots + x_{10}}{p} - \frac{100 - (x_1 + x_2 + \ldots + x_{10})}{1-p} = 0$$

$$\bar{x} = \frac{\sum_{k=1}^{10} x_k}{10}$$

átlag

Ezzel a jelöléssel átírva

$$\frac{10 \cdot \bar{x}}{p} = \frac{100 - 10\bar{x}}{1-p} \qquad / \cdot p(1-p)\frac{1}{10}$$

$$\bar{x}(1-p) = (10 - \bar{x})p$$

$$\bar{x} - p\bar{x} = 10p - \bar{x}p$$

$$\bar{x} = 10p \rightarrow p = \frac{\bar{x}}{10}$$

ez lenne a max. likelihood becslés.

Ezek binomiális valószínűségi változók, a várható érték $n \cdot p = 10 \cdot p$

így $\frac{\bar{X}}{10} = p$ ✓

2.

b, Poisson max. likelihood becslése

$X_1, X_2, \ldots, X_n \sim Poi(\lambda)$ függetlenek.

↑ ez a λ az, amit tenni szeretnénk.

Van egy ~~X_1~~ $x_1, x_2, \ldots, x_n$ megvalósulás. A likelihood fogja azt mondani meg, hogy mi a valószínűsége annak, hogy ez bekövetkezett?

$$L_\lambda(x_1, x_2, \ldots, x_n) = \cancel{P_\lambda(X_1, \ldots, X_n)}$$
$$P_\lambda(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) =$$

$$= P_\lambda(X_1 = x_1) \cdot P_\lambda(X_2 = x_2) \ldots P_\lambda(X_n = x_n) =$$

$$= e^{-\lambda} \cdot \frac{\lambda^{x_1}}{x_1!} \cdot \ldots \cdot e^{-\lambda} \cdot \frac{\lambda^{x_n}}{x_n!}$$

↓

$$l_\lambda(x_1, \ldots, x_n) = \log L_\lambda(x_1, \ldots, x_n) =$$

$$= -\lambda \cdot n + (x_1 + \ldots + x_n) \cdot \log \lambda -$$

$$- \log(x_1! \ldots x_n!)$$

λ szerint deriválva:

$$\frac{d}{d\lambda} l_\lambda(x_1, \ldots, x_n) = -n + \frac{x_1 + \ldots + x_n}{\lambda} = 0$$

Ebből $A = \frac{x_1 + \dots x_n}{n} = \bar{x}$

(4.) Ez egy folytonos eset

$$f(x) = \frac{\Theta}{x^{\Theta+1}}, \text{ ha } x \geq 1$$

sűrűségfgv.

Adott egy minta és: 10 db megvalósulás $(x_1, x_2, \dots, x_{10})$ független azonos eloszlással $f(x)$ sűrűségfüggvénnyel.

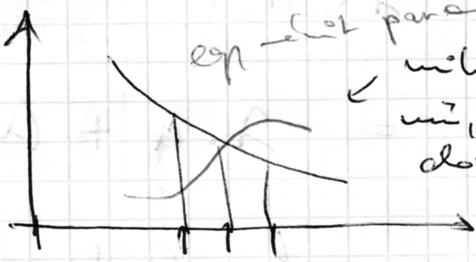
Θ paraméterezi a likelihood becslést:

$$L_\Theta(x_1, x_2, \dots, x_{10}) = f_\Theta(x_1) \cdot f_\Theta(x_2) \cdots f_\Theta(x_{10}) =$$

a sűrűségfüggvényt [illegible] kell ide betenni.

$$= \frac{\Theta}{x_1^{\Theta+1}} \cdot \frac{\Theta}{x_2^{\Theta+1}} \cdot \ldots \cdot \frac{\Theta}{x_{10}^{\Theta+1}}$$

A sűrűségfgv. önmagában hordozza, hogy milyen valószínűséggel egy adott pont környezetébe.



ebből

$$l_\Theta(x_1, x_2, \dots x_{10}) = \log L_\Theta(x_1, x_2, \dots x_{10})$$

$$= 10 \cdot \log\Theta - (\Theta+1) \cdot (\log x_1 + \dots + \log x_{10})$$

Több ilyen fgv-ünk van, ezt jelzi a Θ index.

Θ szerint deriválva

$$\frac{d}{d\Theta} l_\Theta(x_1, \dots x_{10}) = \frac{10}{\Theta} - (\log x_1 + \log x_2 + \dots + \log x_{10}) = 0$$

amiből

$$\Theta = \frac{10}{\log x_1 + \log x_2 + \dots + \log x_{10}}$$ és $x_1, x_2, \dots, x_{10}$ behelyettesíthetők.

Ez egy realisztikus modellje a fizetéseknek.

(7.) Hipotézisvizsgálat

↓ vannak valószínűségi változóink: $X_1, X_2, \dots, X_{25} \sim N(m, \sigma^2)$, fgtlen. ahol σ^2 ismert.

~~Van~~

$\sigma = 50$

Megvalósulások:

$(x_1, x_2, \dots, x_{25})$ $\quad \bar{x} = 986$

Van 2 hipotézisünk:

$H_0 : m = 1000$

$H_1 : m \neq 1000$ (ellen hipotézis)

Statisztikai tesztet végzünk: feltételezzük az ártatlanságot: 1000g-ot akar beletenni, csak nem sikerül.

$$\frac{X_1 + X_2 + \dots + X_{25}}{25} \sim N\left(m, \frac{\sigma^2}{25}\right)$$ ← ez az ő átlaguk.

$$\frac{m + m + \dots + m}{25} = m$$

a szórásnégyzetek adódnak össze

$$\frac{25\sigma^2}{25^2} = \frac{\sigma^2}{25}$$

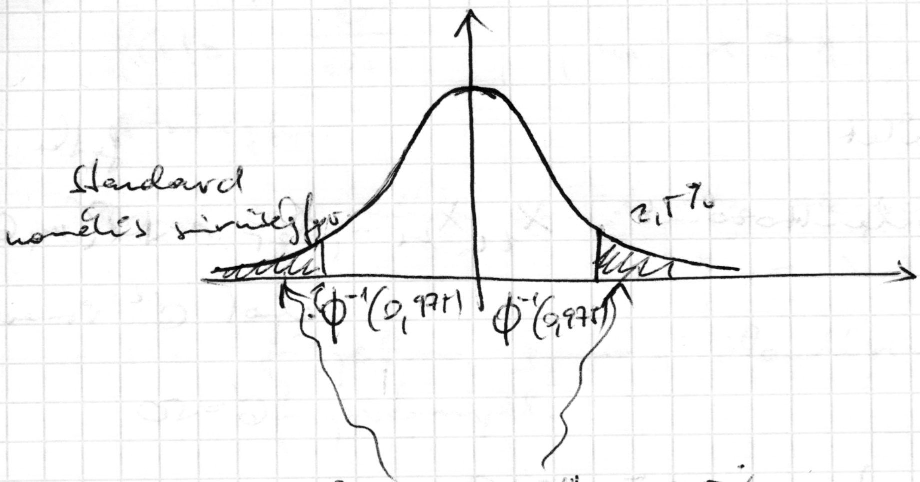
$$\sigma' = \frac{\sigma}{\sqrt{25}}$$

Standardizáljuk a normális eloszlású valószínűségi változót: ártatlanságot feltételezzük

$$U = \frac{\frac{x_1 + x_2 + \dots + x_{25}}{25} - 1000}{50} \cdot \sqrt{25} \sim N(0,1)$$ ez az U-próba.

$$U = \frac{\bar{x} - 1000}{50} \cdot \sqrt{25}$$

← Standard normális egy megvalósulása.



Ha nagyon valószínűtlen, akkor H_0-t elvetjük.

És a feladatban 95%-os volt az, így ez a két levágott szélen együtt 5%, vagyis egy-egy részt 2,5%.

$\Phi^{-1}(0,975)$-öt kell venni, mert ettől balra van 2,5% fölött van.

$\Phi^{-1}(0,975) = 1,96$.

Kiszámoljuk az U-statisztikát, s ezzel összehasonlítjuk.

így ~~U = 2 = 14~~ $U = \frac{986-1000}{50} \cdot 5 = -1,4$

$|U| = 1,4 < 1,96$, vagyis nem tudjuk elvetni a hipotézist, s elfogadjuk.

b, Ha $\sigma = 20$, akkor

$U = -\frac{14}{20} \cdot 5 = -\frac{14}{4} = -3,5 \to |U| > 1,96$, így ezt elvetjük ($H_0$-t).

Ha elutasítjuk, akkor tudunk valamit bizonyítani.

(8.)

mérési eredmények: 7,7 8,1 7,7 7,5 7,0

H_0: várható érték = 7

H_1: várható érték ≠ 7 (még kétoldali ellenhipotézis).

t-próbát végzünk, ~~ez~~ és kétoldali és egymintás.

$$\bar{x} = \frac{7,7+8,1+7,7+7,5+7,0}{5} = 7,6$$

A korrigált empirikus szórásnégyzetet használjuk:

$$(s_n^*)^2 = \frac{(7,7-7,6)^2+(8,1-7,6)^2+(7,7-7,6)^2+(7,5-7,6)^2+(7,0-7,6)^2}{5-1}$$

ebből $s_n^* = 0,4$

A t-statisztika:

$$t = \frac{\bar{x}-7}{s_n^*}\cdot\underbrace{\sqrt{5}}_{\sqrt{n}} = \frac{7,6-7}{0,4}\cdot\sqrt{5} =$$

$= 3,35$

Itt a t-eloszlás táblázatot kell használni. Ennek van egy szabadsági foka: ez $n-1$.

$t_4(0,975)$ - öt nézzük
szabadságfok, $1-\frac{0,05}{2}$

$t_4(0,975) = 2,776$

$|3,31| > 2,776$, így elutasítjuk H_0-t.

(9.) Egymintás t-próba, mert a változás lesz releváns.

diéta előtt			
diéta után			
változás	110	-20	...

$n = 10$

← csak ez érdekel minket

Ezért végzünk egy egyoldali, egymintás t-próbát.
Egyoldali, mert csak azt kérdeztük, hogy javult-e, azt nem, hogy romlott-e.

H_0: a változás várható értéke ≤ 0

H_1: a változás várható értéke > 0.

Akkor tudunk állítani bármit is, ha H_0-t el tudjuk utasítani.

A változást X-szel jelöljük:

$\overline{X} = 50$

$s_n^* = 63,78$

A próbastatisztika:

$$t = \frac{50-0}{63,78} \cdot \sqrt{10} \approx 2,48$$

Ezt hasonlítjuk össze

mert egyoldali próba

$$t_9(0,95) = 1,833$$

2,48 > 1,833 , így el tudjuk utasítani H_0-t.

(10.) Ez már kétmintás (legtöbbször nem is ugyanannyi az elemszámunk). Különböző emberek végeztek különbö-ző kérdőíveket töltött.

fizetés (A) : $\bar{x}_A = 1$ $\quad s_A^* = 0,83$ $\quad n_A = 6$

munkakörülmény (B) : $\bar{x}_B = 1,9$ $\quad s_B^* = 0,98$ $\quad n_B = 10$

a, és b, is kétoldali, de egymintás

c, Kétmintás, egyoldali t-próba.

$$H_0: EX_A \geq EX_B$$

$$H_1: EX_A < EX_B$$

A t-próba:

$$t = \frac{\bar{x}_B - \bar{x}_A}{\sqrt{\frac{(n_A-1)\cdot(s_A^*)^2 + (n_B-1)(s_B^*)^2}{n_A+n_B-2}}} \cdot \sqrt{\frac{n_A \cdot n_B}{n_A+n_B}}$$

$= 1,88$

Megnézzük, hogy melyik t-eloszlás melyik kritikus értékével kell számolnunk:

$$t_{5+9}(0,95) = 1,76$$

(0,95: egyoldali)

Mivel $1,88 > 1,76$, így el lehet utasítani.

(12.) Illeszkedésvizsgálat

$N = 90$ mintaelemek

H_0: a minta egyenletes eloszlásból származik

H_1: a minta nem egyenletes eloszlásból származik

Ez név χ^2 próba, ami $\forall$ egyoldali.

amennyi lenne a várható értéke, mert 90 mintaelem van és 10 elemű egy halmaz, így egy halmazba 9 db-nak kellene esnie.

at egyenletes eloszlás miatt ezek állandók. más eloszlásnál nem biztos!

$$\chi^2 = \sum_{i=0}^{9} \frac{(N_i - 9)^2}{9} = $$

$$= \frac{(35-9)^2}{9} + \frac{(4-9)^2}{9} + \dots = 126{,}2$$

0	1	...	9
N_0	N_1		N_9

χ^2-hez is van egy eloszlás.

mindig egyoldali

$$\chi^2_9(0{,}95) = 16{,}9$$

szabadsági fok. Ez a cellák száma -1

Mivel ~~16,9~~ $126{,}2 > 16{,}9$, ezért elutasítjuk.

(14.) Homogenitásvizsgálat

~~E~~ Ezt is a χ^2 eloszlásra ~~vissza~~ vezetjük vissza.

Kontroll	N_1	N_2	N_3
Balesetes	M_1	M_2	M_3

$n = 399$

$m = 333$

H_0: a kontroll és a balesetes eloszlása ugyanaz a fenti kategóriák szerint (bukósisakok száma)

H_1: nem —||—

A próbastatisztika:

$$\chi^2 = n \cdot m \cdot \sum_{i=1}^{3} \frac{\left(\frac{N_i}{n} - \frac{M_i}{m}\right)^2}{N_i + M_i} = \ldots -8{,}77$$

$\frac{N_i}{n}$: relatív gyakoriság (a kontrollcsoportból mellére hányad viselt ilyen bukósisakot

$\frac{M_i}{m}$: a sérült csoportban hányan viseltek ilyet

$\chi^2_{2}(0{,}95) = 5{,}99$

mert 3 oszlop van

$|-8{,}77| > 5{,}99$, ezért elutasítjuk H_0-t.