Statisztika

Nagy Gergely Gábor

BME SZIT

2012. május 3.

Vázlat

- Valószínűségszámítási alapok
- Becsléselmélet
- Hipotéziselmélet

Valószínűségi mező definíciója

$(\Omega, \mathcal{F}, P)$ hármas valószínűségi mező, ha

- Ω: alaphalmaz (elemi események)

Valószínűségi mező definíciója

$(\Omega, \mathcal{F}, P)$ hármas valószínűségi mező, ha

- Ω: alaphalmaz (elemi események)
- $\mathcal{F} \subseteq 2^\Omega$ σ-algebra (események, mérhető halmazok)
  - $\mathcal{F} \neq \emptyset$
  - $A \in \mathcal{F} \Longrightarrow \Omega \setminus A \in \mathcal{F}$
  - $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow A_1 \cup A_2 \cup \ldots \in \mathcal{F}$

Valószínűségi mező definíciója

$(\Omega, \mathcal{F}, P)$ hármas valószínűségi mező, ha

- Ω: alaphalmaz (elemi események)
- $\mathcal{F} \subseteq 2^\Omega$ σ-algebra (események, mérhető halmazok)
  - $\mathcal{F} \neq \emptyset$
  - $A \in \mathcal{F} \Longrightarrow \Omega \setminus A \in \mathcal{F}$
  - $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow A_1 \cup A_2 \cup \ldots \in \mathcal{F}$
- $P : \mathcal{F} \to [0,1]$ σ-additív (valószínűségfv)
  - $A_1, A_2, \ldots \in \mathcal{F}$ diszjunktak $\Longrightarrow P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$
  - $P(\Omega) = 1$

Diszkrét példa

$$\Omega = \{\text{fej, írás}\}$$

Diszkrét példa

$$\Omega = \{\text{fej, írás}\}$$
$$\mathcal{F} = 2^{\Omega} = \{\emptyset, \{\text{fej}\}, \{\text{írás}\}, \{\text{fej, írás}\}\}$$

Diszkrét példa

$$\Omega = \{\text{fej, írás}\}$$
$$\mathcal{F} = 2^{\Omega} = \{\emptyset, \{\text{fej}\}, \{\text{írás}\}, \{\text{fej, írás}\}\}$$
$$P(\emptyset) = 0$$
$$P(\{\text{fej}\}) = \frac{1}{2}$$
$$P(\{\text{írás}\}) = \frac{1}{2}$$
$$P(\{\text{fej, írás}\}) = 1$$

Valószínűségi változó definíciója (valós eset)

$X : \Omega \to \mathbb{R}$ valószínűségi változó, ha

- mérhető, azaz $\forall x \in \mathbb{R}$-re
  $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$

Valószínűségi változó definíciója (valós eset)

$X : \Omega \to \mathbb{R}$ valószínűségi változó, ha

- mérhető, azaz $\forall x \in \mathbb{R}$-re
  $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$

X eloszlása $Q = P \circ X^{-1}$

Valószínűségi változó definíciója (valós eset)

$X : \Omega \to \mathbb{R}$ valószínűségi változó, ha

- mérhető, azaz $\forall x \in \mathbb{R}$-re
  $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$

X eloszlása $Q = P \circ X^{-1}$

X eloszlásfüggvénye $F : \mathbb{R} \to [0,1]$, ha

- $\forall x \in \mathbb{R}$-re $F(x) = Q((-\infty, x]) = P(X \leq x)$

Valószínűségi változó definíciója (valós eset)

$X : \Omega \to \mathbb{R}$ valószínűségi változó, ha

- mérhető, azaz $\forall x \in \mathbb{R}$-re
  $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$

X eloszlása $Q = P \circ X^{-1}$

X eloszlásfüggvénye $F : \mathbb{R} \to [0,1]$, ha

- $\forall x \in \mathbb{R}$-re $F(x) = Q((-\infty, x]) = P(X \leq x)$

X sűrűségfüggvénye $f : \mathbb{R} \to [0, \infty)$, ha

- $\forall x \in \mathbb{R}$-re $F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$
- (Ha F deriválható, akkor $f = F'$)

Várható érték, szórásnégyzet

X várható értéke $E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$

Várható érték, szórásnégyzet

X várható értéke $E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$

h(X) várható értéke $E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) \mathrm{d}x$

Várható érték, szórásnégyzet

X várható értéke $E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$

h(X) várható értéke $E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) \mathrm{d}x$

X szórásnégyzete $D^2(X) = E((X - E(X))^2) = E(X^2) - E^2(X)$

Normális eloszlás

$N(\mu, \sigma^2)$ normális eloszlás

- sűrűségfüggvénye $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- eloszlásfüggvénye $F(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2\pi\sigma^2}}\right)\right)$
- várható értéke μ
- szórásnégyzete σ^2

Normális eloszlás

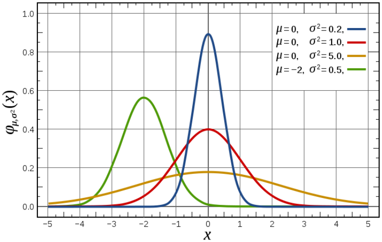
$N(\mu, \sigma^2)$ normális eloszlás

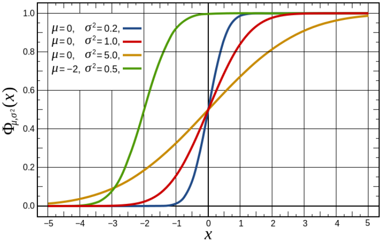
- sűrűségfüggvénye $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- eloszlásfüggvénye $F(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2\pi\sigma^2}}\right)\right)$
- várható értéke μ
- szórásnégyzete σ^2

$N(0,1)$ standard normális eloszlás

- sűrűségfüggvénye $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- eloszlásfüggvénye $F(x) = \Phi(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2\pi}}\right)\right)$
- várható értéke 0
- szórásnégyzete 1

Normális eloszlás





Centrális határeloszlás-tétel

Miért jó a normális eloszlás?

Centrális határeloszlás-tétel

Miért jó a normális eloszlás?

Legyenek $X_1, X_2, \ldots$ azonos eloszlású független valószínűségi változók μ várható értékkel és σ^2 véges szórásnégyzettel.

Legyen $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$ az első n változó átlaga.

Centrális határeloszlás-tétel

Miért jó a normális eloszlás?

Legyenek $X_1, X_2, \ldots$ azonos eloszlású független valószínűségi változók μ várható értékkel és σ^2 véges szórásnégyzettel.

Legyen $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$ az első n változó átlaga.

Centrális határeloszlás-tétel: $\sqrt{n}(S_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

Átrendezgetve: $S_n \approx N(\mu, \frac{\sigma^2}{n})$

Statisztikai mező definíciója

$(\Omega, \mathcal{F}, \mathcal{P})$ hármas statisztikai mező, ha

- $\forall P \in \mathcal{P}$-re $(\Omega, \mathcal{F}, P)$ valószínűségi mező

Statisztikai mező definíciója

$(\Omega, \mathcal{F}, \mathcal{P})$ hármas statisztikai mező, ha

- $\forall P \in \mathcal{P}$-re $(\Omega, \mathcal{F}, P)$ valószínűségi mező

Paraméteres alak: $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$, gyakran $\Theta \subseteq \mathbb{R}^p$

Statisztikai mező definíciója

$(\Omega, \mathcal{F}, \mathcal{P})$ hármas statisztikai mező, ha

- $\forall P \in \mathcal{P}$-re $(\Omega, \mathcal{F}, P)$ valószínűségi mező

Paraméteres alak: $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$, gyakran $\Theta \subseteq \mathbb{R}^p$

Mi most feltesszük, hogy minden P_ϑ-nak van f_ϑ sűrűségfüggvénye.

Minta

Minta: $X : \Omega \to \mathcal{X}$ mérhető leképezés ($\mathcal{X}$ a mintatér)

Minta

Minta: $X : \Omega \to \mathcal{X}$ mérhető leképezés ($\mathcal{X}$ a mintatér)

Spec. eset (n elemű független minta):

- $X = (X_1, \ldots, X_n)$, ahol $X_1, \ldots, X_n$ azonos eloszlású, független változók.

Statisztika

Statisztika: $T : \mathcal{X} \to \ldots$ mérhető függvény

Statisztika

Statisztika : $T : \mathcal{X} \to \ldots$ mérhető függvény

Például :

- Átlag : $T(X) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Statisztika

Statisztika: $T : \mathcal{X} \to \dots$ mérhető függvény

Például:

- Átlag: $T(X) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Tapasztalati szórásnégyzet: $T(X) = s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$

Statisztika

Statisztika: $T : \mathcal{X} \to \ldots$ mérhető függvény

Például:

- Átlag: $T(X) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Tapasztalati szórásnégyzet: $T(X) = s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$
- Korrigált tapasztalati szórásnégyzet: $T(X) = s_n^{*2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$

Statisztika

Statisztika: $T : \mathcal{X} \to \dots$ mérhető függvény

Például:

- Átlag: $T(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Tapasztalati szórásnégyzet: $T(X) = s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$
- Korrigált tapasztalati szórásnégyzet: $T(X) = s_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
- Tapasztalati eloszlás: $T(X) = Q_n^*$, ahol $Q_n^*(B) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in B)$

Statisztika

Statisztika: $T : \mathcal{X} \to \ldots$ mérhető függvény

Például:

- Átlag: $T(X) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Tapasztalati szórásnégyzet: $T(X) = s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$
- Korrigált tapasztalati szórásnégyzet:
  $T(X) = s_n^{*2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$
- Tapasztalati eloszlás: $T(X) = Q_n^*$, ahol
  $Q_n^*(B) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B)$
- Tapasztalati eloszlásfüggvény: $T(X) = F_n^*$, ahol
  $F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$

Glivenko-Cantelli tétel (a statisztika alaptétele)

Glivenko-Cantelli tétel: $P\left(\sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \to 0\right) = 1$, azaz a tapasztalati eloszlásfüggvény 1 valószínűséggel egyenletesen konvergál a valódihoz.

Becslés

$\varphi(P)$-t szeretnénk becsülni, ahol $\varphi : \mathcal{P} \to \mathbb{R}^k$ függvény.
Ha $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$, akkor $g(\vartheta) = \varphi(P_\vartheta)$-t kell becsülnünk.

Becslés

$\varphi(P)$-t szeretnénk becsülni, ahol $\varphi : \mathcal{P} \to \mathbb{R}^k$ függvény.
Ha $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$, akkor $g(\vartheta) = \varphi(P_\vartheta)$-t kell becsülnünk.

$g(\vartheta)$ becslése: $T : \mathcal{X} \to \mathbb{R}^k$

Becslés

$\varphi(P)$-t szeretnénk becsülni, ahol $\varphi : \mathcal{P} \to \mathbb{R}^k$ függvény.
Ha $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$, akkor $g(\vartheta) = \varphi(P_\vartheta)$-t kell becsülnünk.

$g(\vartheta)$ becslése: $T : \mathcal{X} \to \mathbb{R}^k$
T torzítatlan becslés, ha

- $\forall\vartheta \in \Theta$-ra $E_\vartheta(T(X)) = g(\vartheta)$

Becslés

$\varphi(P)$-t szeretnénk becsülni, ahol $\varphi : \mathcal{P} \to \mathbb{R}^k$ függvény.

Ha $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$, akkor $g(\vartheta) = \varphi(P_\vartheta)$-t kell becsülnünk.

$g(\vartheta)$ becslése: $T : \mathcal{X} \to \mathbb{R}^k$

T torzítatlan becslés, ha

- $\forall \vartheta \in \Theta$-ra $E_\vartheta(T(X)) = g(\vartheta)$

T torzítása $b_T(\vartheta) = E_\vartheta(T(X)) - g(\vartheta)$

Veszteség, rizikó

$w : \mathbb{R}^k \times \Theta \to \mathbb{R}$ veszteségfüggvény, ahol

- (jelentése: ha ϑ az igazi paraméter és $g(\vartheta)$-t T-vel becsüljük, akkor a veszteségünk $w(T, \vartheta)$)
- $w(g(\vartheta), \vartheta) = 0$
- $w(T, \vartheta) \geq 0$
- w első változójában konvex

Veszteség, rizikó

$w : \mathbb{R}^k \times \Theta \to \mathbb{R}$ veszteségfüggvény, ahol

- (jelentése: ha ϑ az igazi paraméter és $g(\vartheta)$-t T-vel becsüljük, akkor a veszteségünk $w(T, \vartheta)$)
- $w(g(\vartheta), \vartheta) = 0$
- $w(T, \vartheta) \geq 0$
- w első változójában konvex

T rizikófüggvénye $R_T(\vartheta) = E_\vartheta(w(T(X), \vartheta))$ (az átlagos veszteség)

Veszteség, rizikó

$w : \mathbb{R}^k \times \Theta \to \mathbb{R}$ veszteségfüggvény, ahol

- (jelentése: ha ϑ az igazi paraméter és $g(\vartheta)$-t T-vel becsüljük, akkor a veszteségünk $w(T, \vartheta)$)
- $w(g(\vartheta), \vartheta) = 0$
- $w(T, \vartheta) \geq 0$
- w első változójában konvex

T rizikófüggvénye $R_T(\vartheta) = E_\vartheta(w(T(X), \vartheta))$ (az átlagos veszteség)

Példa (négyzetes veszteségfüggvény):

- $w(T, \vartheta) = \|T - g(\vartheta)\|^2$

Veszteség, rizikó

$w : \mathbb{R}^k \times \Theta \to \mathbb{R}$ veszteségfüggvény, ahol

- (jelentése: ha ϑ az igazi paraméter és $g(\vartheta)$-t T-vel becsüljük, akkor a veszteségünk $w(T, \vartheta)$)
- $w(g(\vartheta), \vartheta) = 0$
- $w(T, \vartheta) \geq 0$
- w első változójában konvex

T rizikófüggvénye $R_T(\vartheta) = E_\vartheta(w(T(X), \vartheta))$ (az átlagos veszteség)

Példa (négyzetes veszteségfüggvény):

- $w(T, \vartheta) = \|T - g(\vartheta)\|^2$
- $R_T(\vartheta) = E_\vartheta(\|T - g(\vartheta)\|^2)$

Veszteség, rizikó

$w : \mathbb{R}^k \times \Theta \to \mathbb{R}$ veszteségfüggvény, ahol

- (jelentése: ha ϑ az igazi paraméter és $g(\vartheta)$-t T-vel becsüljük, akkor a veszteségünk $w(T, \vartheta)$)
- $w(g(\vartheta), \vartheta) = 0$
- $w(T, \vartheta) \geq 0$
- w első változójában konvex

T rizikófüggvénye $R_T(\vartheta) = E_\vartheta(w(T(X), \vartheta))$ (az átlagos veszteség)

Példa (négyzetes veszteségfüggvény):

- $w(T, \vartheta) = \|T - g(\vartheta)\|^2$
- $R_T(\vartheta) = E_\vartheta(\|T - g(\vartheta)\|^2)$
- $k = 1$ esetben $R_T(\vartheta) = E_\vartheta((T - g(\vartheta))^2) = D^2_\vartheta(T - g(\vartheta)) + + E^2_\vartheta(T - g(\vartheta)) = D^2_\vartheta(T) + b^2_T(\vartheta)$

Becslések összehasonlítása

T_1, T_2 becslések ugyanarra a $g(\vartheta)$-ra. Melyik a jobb?

Becslések összehasonlítása

T_1, T_2 becslések ugyanarra a $g(\vartheta)$-ra. Melyik a jobb?
T_1 jobb („nem rosszabb"), mint T_2, ha $\vartheta R_{T_1}(\vartheta) \leq R_{T_2}(\vartheta)$ fennáll minden ϑ-ra.

Becslések összehasonlítása

T_1, T_2 becslések ugyanarra a $g(\vartheta)$-ra. Melyik a jobb?
T_1 jobb („nem rosszabb"), mint T_2, ha $\vartheta R_{T_1}(\vartheta) \le R_{T_2}(\vartheta)$ fennáll minden ϑ-ra.
Legyen $\mathcal{D}$ becslések egy osztálya. $T \in \mathcal{D}$ optimális, ha minden $\mathcal{D}$-beli becslésnél jobb.

Becslések összehasonlítása

T_1, T_2 becslések ugyanarra a $g(\vartheta)$-ra. Melyik a jobb?
T_1 jobb („nem rosszabb"), mint T_2, ha $\vartheta R_{T_1}(\vartheta) \leq R_{T_2}(\vartheta)$ fennáll minden ϑ-ra.
Legyen $\mathcal{D}$ becslések egy osztálya. $T \in \mathcal{D}$ optimális, ha minden $\mathcal{D}$-beli becslésnél jobb.
Ha $\mathcal{D}$ a torzítatlan becslések osztálya, és w négyzetes veszteségfüggvény, akkor az optimálist hatásosnak hívjuk. Gyakran MVUE (minimum-variance unbiased estimator).

Tapasztalati becslés

Ha φ az eloszlásokon van értelmezve, $\varphi(Q)$-t szeretnénk becsülni, Q a valódi eloszlásfüggvény.
Tapasztalati becslés: $\varphi(Q)$-t $\varphi(Q_n^*)$-gal becsüljük.

Tapasztalati becslés

Ha φ az eloszlásokon van értelmezve, $\varphi(Q)$-t szeretnénk becsülni, Q a valódi eloszlásfüggvény.
Tapasztalati becslés: $\varphi(Q)$-t $\varphi(Q_n^*)$-gal becsüljük.
Például:

- Az átlag $(\overline{X})$ a várható érték (E) tapasztalati becslése.

Tapasztalati becslés

Ha φ az eloszlásokon van értelmezve, $\varphi(Q)$-t szeretnénk becsülni, Q a valódi eloszlásfüggvény.
Tapasztalati becslés: $\varphi(Q)$-t $\varphi(Q_n^*)$-gal becsüljük.
Például:

- Az átlag $(\overline{X})$ a várható érték (E) tapasztalati becslése.
- A tapasztalati szórásnégyzet (s_n^2) a szórásnégyzet (D^2) tapasztalati becslése.

Tapasztalati becslés

Ha φ az eloszlásokon van értelmezve, $\varphi(Q)$-t szeretnénk becsülni, Q a valódi eloszlásfüggvény.
Tapasztalati becslés: $\varphi(Q)$-t $\varphi(Q_n^*)$-gal becsüljük.
Például:

- Az átlag $(\overline{X})$ a várható érték (E) tapasztalati becslése.
- A tapasztalati szórásnégyzet (s_n^2) a szórásnégyzet (D^2) tapasztalati becslése.
- A tapasztalati eloszlásfüggvény (F_n^*) az eloszlásfüggvény (F) tapasztalati becslése.

Momentum-módszer

Legyen most $g(\vartheta) = \vartheta$, azaz magát a paramétert szeretnénk becsülni ($\vartheta \in \Theta \subseteq \mathbb{R}^p$)

Momentum-módszer

Legyen most $g(\vartheta) = \vartheta$, azaz magát a paramétert szeretnénk becsülni ($\vartheta \in \Theta \subseteq \mathbb{R}^p$)

Válasszunk egy olyan $f : \mathcal{X} \to \mathbb{R}^p$ függvényt, amire $h(\vartheta) = E_\vartheta(f(X_1))$ invertálható $\mathbb{R}^p \to \mathbb{R}^p$ leképezés.

Momentum-módszer

Legyen most $g(\vartheta) = \vartheta$, azaz magát a paramétert szeretnénk becsülni ($\vartheta \in \Theta \subseteq \mathbb{R}^p$)

Válasszunk egy olyan $f : \mathcal{X} \to \mathbb{R}^p$ függvényt, amire $h(\vartheta) = E_\vartheta(f(X_1))$ invertálható $\mathbb{R}^p \to \mathbb{R}^p$ leképezés.

$h(\vartheta)$-t n elemű mintából így becsülhetjük: $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$

Momentum-módszer

Legyen most $g(\vartheta) = \vartheta$, azaz magát a paramétert szeretnénk becsülni ($\vartheta \in \Theta \subseteq \mathbb{R}^p$)

Válasszunk egy olyan $f : \mathcal{X} \to \mathbb{R}^p$ függvényt, amire $h(\vartheta) = E_\vartheta(f(X_1))$ invertálható $\mathbb{R}^p \to \mathbb{R}^p$ leképezés.

$h(\vartheta)$-t n elemű mintából így becsülhetjük: $\frac{1}{n}\sum_{i=1}^n f(X_i)$

Ezalapján ϑ becslése: $T_n(X) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^n f(X_i)\right)$

Momentum-módszer

Legyen most $g(\vartheta) = \vartheta$, azaz magát a paramétert szeretnénk becsülni ($\vartheta \in \Theta \subseteq \mathbb{R}^p$)

Válasszunk egy olyan $f : \mathcal{X} \to \mathbb{R}^p$ függvényt, amire $h(\vartheta) = E_\vartheta(f(X_1))$ invertálható $\mathbb{R}^p \to \mathbb{R}^p$ leképezés.

$h(\vartheta)$-t n elemű mintából így becsülhetjük: $\frac{1}{n}\sum_{i=1}^n f(X_i)$

Ezalapján ϑ becslése: $T_n(X) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^n f(X_i)\right)$

Ezt a fajta becslést hívják momentum-módszernek. Miért?

Momentum-módszer

Legyen most $g(\vartheta) = \vartheta$, azaz magát a paramétert szeretnénk becsülni ($\vartheta \in \Theta \subseteq \mathbb{R}^p$)

Válasszunk egy olyan $f : \mathcal{X} \to \mathbb{R}^p$ függvényt, amire $h(\vartheta) = E_\vartheta(f(X_1))$ invertálható $\mathbb{R}^p \to \mathbb{R}^p$ leképezés.

$h(\vartheta)$-t n elemű mintából így becsülhetjük: $\frac{1}{n}\sum_{i=1}^n f(X_i)$

Ezalapján ϑ becslése: $T_n(X) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^n f(X_i)\right)$

Ezt a fajta becslést hívják momentum-módszernek. Miért?

Gyakran f olyan, hogy a koordinátái momentumok. Például $f(x) = (x^{k_1}, x^{k_2}, \ldots, x^{k_p})$.

Maximum-likelihood becslés

Likelihood-függvény: $L(\vartheta \mid x) = f_\vartheta(x)$
(n elemű független mintánál $f_\vartheta(x) = \prod_{i=1}^n f_\vartheta(x_i)$)

Maximum-likelihood becslés

Likelihood-függvény: $L(\vartheta \mid x) = f_\vartheta(x)$

(n elemű független mintánál $f_\vartheta(x) = \prod_{i=1}^{n} f_\vartheta(x_i)$)

ϑ ML-becslése az a $\hat{\vartheta} \in \Theta$ érték, amire a likelihood-függvény maximális: $f_{\hat{\vartheta}}(x) = \max_{\vartheta \in \Theta} f_\vartheta(x)$.

Maximum-likelihood becslés

Likelihood-függvény: $L(\vartheta \mid x) = f_\vartheta(x)$
(n elemű független mintánál $f_\vartheta(x) = \prod_{i=1}^n f_\vartheta(x_i)$)

ϑ ML-becslése az a $\hat{\vartheta} \in \Theta$ érték, amire a likelihood-függvény maximális: $f_{\hat{\vartheta}}(x) = \max_{\vartheta \in \Theta} f_\vartheta(x)$.

Nem biztos, hogy létezik, sem az hogy egyértelmű.

Maximum-likelihood becslés

Likelihood-függvény: $L(\vartheta \mid x) = f_\vartheta(x)$
(n elemű független mintánál $f_\vartheta(x) = \prod_{i=1}^n f_\vartheta(x_i)$)

ϑ ML-becslése az a $\hat{\vartheta} \in \Theta$ érték, amire a likelihood-függvény maximális: $f_{\hat{\vartheta}}(x) = \max_{\vartheta \in \Theta} f_\vartheta(x)$.

Nem biztos, hogy létezik, sem az hogy egyértelmű.

$g(\vartheta)$ ML-becslése $g(\hat{\vartheta})$.

Hipotézisvizsgálat

A paraméterteret két részre bontjuk: $\Theta = \Theta_0 \dot{\cup} \Theta_1$.

A (null)hipotézisünk (H_0) az, hogy $\vartheta \in \Theta_0$.
Az ellenhipotézisünk (H_1) pedig, hogy $\vartheta \in \Theta_1$.

Hipotézisvizsgálat

A paraméterteret két részre bontjuk: $\Theta = \Theta_0 \dot\cup \Theta_1$.

A (null)hipotézisünk (H_0) az, hogy $\vartheta \in \Theta_0$.
Az ellenhipotézisünk (H_1) pedig, hogy $\vartheta \in \Theta_1$.

Kérdés: lehet-e tényleg, hogy teljesül H_0?
Ez nem szimmetrikus, az a prekoncepciónk, hogy H_0 igaz, ezt elfogadhatjuk vagy elvethetjük.

Hipotézisvizsgálat

A paraméterteret két részre bontjuk: $\Theta = \Theta_0 \dot{\cup} \Theta_1$.

A (null)hipotézisünk (H_0) az, hogy $\vartheta \in \Theta_0$.
Az ellenhipotézisünk (H_1) pedig, hogy $\vartheta \in \Theta_1$.

Kérdés: lehet-e tényleg, hogy teljesül H_0?
Ez nem szimmetrikus, az a prekoncepciónk, hogy H_0 igaz, ezt elfogadhatjuk vagy elvethetjük.

Két féle hiba lehet:

- Elsőfajú hibát követünk el, ha elvetjük H_0-t, pedig igaz. Ezt nem akarjuk!

Hipotézisvizsgálat

A paraméterteret két részre bontjuk: $\Theta = \Theta_0 \dot{\cup} \Theta_1$.

A (null)hipotézisünk (H_0) az, hogy $\vartheta \in \Theta_0$.
Az ellenhipotézisünk (H_1) pedig, hogy $\vartheta \in \Theta_1$.

Kérdés: lehet-e tényleg, hogy teljesül H_0?
Ez nem szimmetrikus, az a prekoncepciónk, hogy H_0 igaz, ezt elfogadhatjuk vagy elvethetjük.

Két féle hiba lehet:

- Elsőfajú hibát követünk el, ha elvetjük H_0-t, pedig igaz. Ezt nem akarjuk!
- Másodfajú hibát követünk el, ha elfogadjuk H_0-t, pedig hamis. Ez nem annyira nagy gond.

Statisztikai próba

A mintateret két részre bontjuk: $\mathcal{X} = \mathcal{X}_0 \dot{\cup} \mathcal{X}_1$.

$\mathcal{X}_0$ az elfogadási és $\mathcal{X}_1$ az elvetési tartomány.

Statisztikai próba

A mintateret két részre bontjuk: $\mathcal{X} = \mathcal{X}_0 \dot{\cup} \mathcal{X}_1$.

$\mathcal{X}_0$ az elfogadási és $\mathcal{X}_1$ az elvetési tartomány.

Ha $X \in \mathcal{X}_0$, akkor elfogadjuk H_0-t, ha pedig $X \in \mathcal{X}_1$, akkor elvetjük.

Statisztikai próba

A mintateret két részre bontjuk: $\mathcal{X} = \mathcal{X}_0 \dot{\cup} \mathcal{X}_1$.

$\mathcal{X}_0$ az elfogadási és $\mathcal{X}_1$ az elvetési tartomány.

Ha $X \in \mathcal{X}_0$, akkor elfogadjuk H_0-t, ha pedig $X \in \mathcal{X}_1$, akkor elvetjük.

Gyakran van egy $T : \mathcal{X} \to \mathbb{R}$ próbastatisztika, $c \in \mathbb{R}$ kritikus érték, és ezekkel definiáljuk $\mathcal{X}_1$-et: $\mathcal{X}_1 = \{x \in \mathcal{X} \mid T(x) > c\}$

Próba jellemzői

Legyen $\psi(\vartheta) = P_\vartheta(\mathcal{X}_1)$, azaz az elvetés valószínűsége.

- Ha $\vartheta \in \Theta_0$, akkor $\psi(\vartheta)$ az elsőfajú hiba valószínűsége.
- Ha $\vartheta \in \Theta_1$, akkor $\psi(\vartheta)$ annak a valószínűsége, hogy helyesen vetjük el H_0-t. Ezt a próba erejének hívjuk ϑ-ban.

Próba jellemzői

Legyen $\psi(\vartheta) = P_\vartheta(\mathcal{X}_1)$, azaz az elvetés valószínűsége.

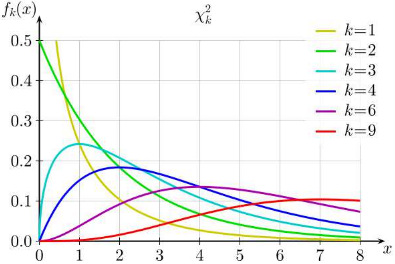
- Ha $\vartheta \in \Theta_0$, akkor $\psi(\vartheta)$ az elsőfajú hiba valószínűsége.
- Ha $\vartheta \in \Theta_1$, akkor $\psi(\vartheta)$ annak a valószínűsége, hogy helyesen vetjük el H_0-t. Ezt a próba erejének hívjuk ϑ-ban.

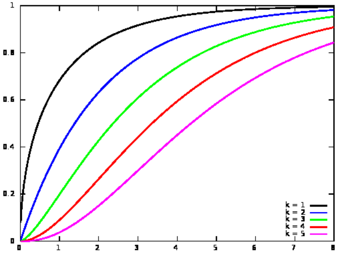
A próba terjedelme $\sup_{\vartheta \in \Theta_0} \psi(\vartheta)$, ez azt jelenti, hogy legfeljebb ekkora valószínűséggel hibázhatunk, azzal hogy H_0-t elvetjük. Sokszor megadják, hogy mekkora terjedelmű próbát kell alkalmaznunk ($\alpha = 0.05$ például).

χ^2-eloszlás

Legyenek $X_1, \ldots, X_q$ független, standard normális eloszlású valószínűségi változók.
Ekkor $Y = X_1^2 + \cdots + X_q^2$ eloszlását q szabadságfokú χ^2-eloszlásnak hívjuk (χ_q^2)

χ^2-eloszlás





Normális eloszlás mintái

Legyen $X_1, \ldots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból. Ilyenkor

- az átlag $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ eloszlású,

Normális eloszlás mintái

Legyen $X_1, \ldots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból. Ilyenkor

- az átlag $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ eloszlású,
- a korrigált tapasztalati szórásnégyzet pedig $s_n^{*2} \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$ eloszlású,

Normális eloszlás mintái

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból. Ilyenkor

- az átlag $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ eloszlású,
- a korrigált tapasztalati szórásnégyzet pedig $s_n^{*2} \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$ eloszlású,
- és ezek függetlenek (Fisher-Bartlett tétel)

Egymintás u-próba (z-test)

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol σ^2-et ismerjük, μ pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0 \colon \mu \leq \mu_0$, $H_1 \colon \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0 \colon \mu = \mu_0$, $H_1 \colon \mu \neq \mu_0$ (kétoldali ellenhipotézis).

Egymintás u-próba (z-test)

Legyen $X_1, \ldots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol σ^2-et ismerjük, μ pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0 \colon \mu \leq \mu_0$, $H_1 \colon \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0 \colon \mu = \mu_0$, $H_1 \colon \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}$$

Egymintás u-próba (z-test)

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol σ^2-et ismerjük, μ pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0: \mu \leq \mu_0$, $H_1: \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}$$

Ez $\mu = \mu_0$ esetben standard normális eloszlású.

Egymintás u-próba (z-test)

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol σ^2-et ismerjük, μ pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0 \colon \mu \leq \mu_0$, $H_1 \colon \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0 \colon \mu = \mu_0$, $H_1 \colon \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}$$

Ez $\mu = \mu_0$ esetben standard normális eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{x \mid T(x) > \Phi^{-1}(1-\alpha)\}$

Egymintás u-próba (z-test)

Legyen $X_1, \ldots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol σ^2-et ismerjük, μ pedig tetszőleges valós paraméter. Hipotéziseink:

- H_0: $\mu \leq \mu_0$, H_1: $\mu > \mu_0$ (egyoldali ellenhipotézis).
- H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}$$

Ez $\mu = \mu_0$ esetben standard normális eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{x \mid T(x) > \Phi^{-1}(1-\alpha)\}$
- Kétoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{x \mid |T(x)| > \Phi^{-1}(1-\frac{\alpha}{2})\}$

Kétmintás u-próba (z-test)

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból, ahol σ_1^2, σ_2^2-t ismerjük, μ_1, μ_2 pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0 \colon \mu_1 \leq \mu_2$, $H_1 \colon \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0 \colon \mu_1 = \mu_2$, $H_1 \colon \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

Kétmintás u-próba (z-test)

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból, ahol σ_1^2, σ_2^2-t ismerjük, μ_1, μ_2 pedig tetszőleges valós paraméter. Hipotéziseink:

- H_0: $\mu_1 \leq \mu_2$, H_1: $\mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x, y) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Kétmintás u-próba (z-test)

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból, ahol σ_1^2, σ_2^2-t ismerjük, μ_1, μ_2 pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0 : \mu_1 \leq \mu_2$, $H_1 : \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x, y) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ez $\mu_1 = \mu_2$ esetben standard normális eloszlású.

Kétmintás u-próba (z-test)

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból, ahol σ_1^2, σ_2^2-t ismerjük, μ_1, μ_2 pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0: \mu_1 \leq \mu_2$, $H_1: \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x,y) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ez $\mu_1 = \mu_2$ esetben standard normális eloszlású.

- Egyoldali ellenhipotézis esetén legyen
  $\mathcal{X}_1 = \{(x,y) \mid T(x,y) > \Phi^{-1}(1-\alpha)\}$

Kétmintás u-próba (z-test)

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból, ahol σ_1^2, σ_2^2-t ismerjük, μ_1, μ_2 pedig tetszőleges valós paraméter. Hipotéziseink:

- $H_0: \mu_1 \leq \mu_2$, $H_1: \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x, y) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

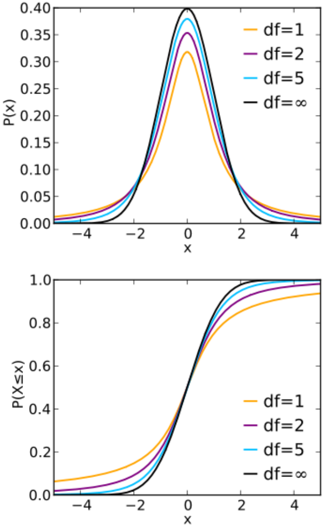
Ez $\mu_1 = \mu_2$ esetben standard normális eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x, y) \mid T(x, y) > \Phi^{-1}(1 - \alpha)\}$
- Kétoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x, y) \mid |T(x, y)| > \Phi^{-1}(1 - \frac{\alpha}{2})\}$

t-eloszlás

Legyen $X \sim N(0{,}1)$, $Y \sim \chi_q^2$ függetlenek.
Ekkor $Z = \frac{X}{\sqrt{\frac{Y}{n}}}$ eloszlását q szabadságfokú t-eloszlásnak hívjuk (t_q)

t-eloszlás


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
df=1
df=2
df=5
df=∞
1.0
0.8
0.6
0.4
0.2
0.0
P(X≤x)
−4
−2
0
2
4
x
df=1
df=2
df=5
df=∞

Egymintás t-próba

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol most se μ-t, se σ^2-et nem ismerjük. Hipotéziseink:

- $H_0\colon \mu \leq \mu_0$, $H_1\colon \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0\colon \mu = \mu_0$, $H_1\colon \mu \neq \mu_0$ (kétoldali ellenhipotézis).

Egymintás t-próba

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol most se μ-t, se σ^2-et nem ismerjük. Hipotéziseink:

- $H_0 \colon \mu \leq \mu_0$, $H_1 \colon \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0 \colon \mu = \mu_0$, $H_1 \colon \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{s_n^*}$$

Egymintás t-próba

Legyen $X_1, \ldots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol most se μ-t, se σ^2-et nem ismerjük. Hipotéziseink:

- $H_0: \mu \leq \mu_0$, $H_1: \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{s_n^*}$$

Erről a Fisher-Bartlett tétel alapján belátható, hogy $\mu = \mu_0$ esetben t_{n-1} eloszlású.

Egymintás t-próba

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol most se μ-t, se σ^2-et nem ismerjük. Hipotéziseink:

- $H_0: \mu \le \mu_0$, $H_1: \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{s_n^*}$$

Erről a Fisher-Bartlett tétel alapján belátható, hogy $\mu = \mu_0$ esetben t_{n-1} eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{x \mid T(x) > F^{-1}_{t_{n-1}}(1-\alpha)\}$

Egymintás t-próba

Legyen $X_1, \dots X_n$ n elemű minta az $N(\mu, \sigma^2)$ eloszlásból, ahol most se μ-t, se σ^2-et nem ismerjük. Hipotéziseink:

- $H_0: \mu \leq \mu_0$, $H_1: \mu > \mu_0$ (egyoldali ellenhipotézis).
- $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ (kétoldali ellenhipotézis).

$$T(x) = \sqrt{n}\frac{\overline{X} - \mu_0}{s_n^*}$$

Erről a Fisher-Bartlett tétel alapján belátható, hogy $\mu = \mu_0$ esetben t_{n-1} eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{x \mid T(x) > F_{t_{n-1}}^{-1}(1-\alpha)\}$
- Kétoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{x \mid |T(x)| > F_{t_{n-1}}^{-1}(1-\frac{\alpha}{2})\}$

Kétmintás t-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \mu_1 \leq \mu_2$, $H_1\colon \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0\colon \mu_1 = \mu_2$, $H_1\colon \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

Kétmintás t-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \mu_1 \leq \mu_2$, $H_1\colon \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0\colon \mu_1 = \mu_2$, $H_1\colon \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x,y) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n_1-1)s_X^{*2} + (n_2-1)s_Y^{*2}}{n_1+n_2-2}}}$$

Kétmintás t-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \mu_1 \leq \mu_2$, $H_1\colon \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0\colon \mu_1 = \mu_2$, $H_1\colon \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x,y) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n_1-1)s_X^{*2} + (n_2-1)s_Y^{*2}}{n_1+n_2-2}}}$$

Erről hosszadalmasan, de belátható, hogy $\mu_1 = \mu_2$ esetben $t_{n_1+n_2-2}$ eloszlású.

Kétmintás t-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \mu_1 \leq \mu_2$, $H_1\colon \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0\colon \mu_1 = \mu_2$, $H_1\colon \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x,y) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n_1-1)s_X^{*2} + (n_2-1)s_Y^{*2}}{n_1+n_2-2}}}$$

Erről hosszadalmasan, de belátható, hogy $\mu_1 = \mu_2$ esetben $t_{n_1+n_2-2}$ eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x,y) \mid T(x,y) > F^{-1}_{t_{n_1+n_2-2}}(1-\alpha)\}$

Kétmintás t-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \mu_1 \leq \mu_2$, $H_1\colon \mu_1 > \mu_2$ (egyoldali ellenhipotézis).
- $H_0\colon \mu_1 = \mu_2$, $H_1\colon \mu_1 \neq \mu_2$ (kétoldali ellenhipotézis).

$$T(x,y) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n_1-1)s_X^{*2}+(n_2-1)s_Y^{*2}}{n_1+n_2-2}}}$$

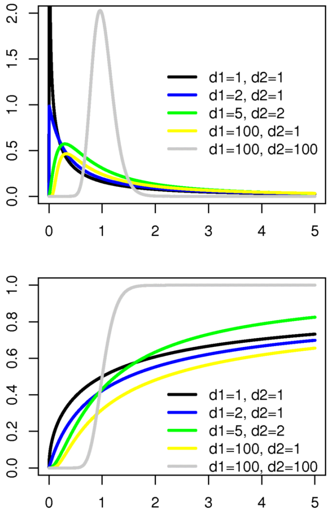
Erről hosszadalmasan, de belátható, hogy $\mu_1 = \mu_2$ esetben $t_{n_1+n_2-2}$ eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x,y) \mid T(x,y) > F^{-1}_{t_{n_1+n_2-2}}(1-\alpha)\}$
- Kétoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x,y) \mid |T(x,y)| > F^{-1}_{t_{n_1+n_2-2}}(1-\frac{\alpha}{2})\}$

F-eloszlás

Legyen $X \sim \chi^2_n$, $Y \sim \chi^2_m$ függetlenek.
Ekkor $Z = \frac{mX}{nY}$ eloszlását n, m szabadságfokú F-eloszlásnak hívjuk $(F_{n,m})$

F-eloszlás


2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
d1=1, d2=1
d1=2, d2=1
d1=5, d2=2
d1=100, d2=1
d1=100, d2=100
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
d1=1, d2=1
d1=2, d2=1
d1=5, d2=2
d1=100, d2=1
d1=100, d2=100

F-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \sigma_1^2 \leq \sigma_2^2$, $H_1\colon \sigma_1^2 > \sigma_2^2$ (egyoldali ellenhipotézis).
- $H_0\colon \sigma_1^2 = \sigma_2^2$, $H_1\colon \sigma_1^2 \neq \sigma_2^2$ (egyoldali ellenhipotézis).

F-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból. Hipotéziseink:

- $H_0: \sigma_1^2 \leq \sigma_2^2$, $H_1: \sigma_1^2 > \sigma_2^2$ (egyoldali ellenhipotézis).
- $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$ (egyoldali ellenhipotézis).

$$T(x) = \frac{s_X^{*2}}{s_Y^{*2}}$$

F-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból. Hipotéziseink:

- H_0: $\sigma_1^2 \leq \sigma_2^2$, H_1: $\sigma_1^2 > \sigma_2^2$ (egyoldali ellenhipotézis).
- H_0: $\sigma_1^2 = \sigma_2^2$, H_1: $\sigma_1^2 \neq \sigma_2^2$ (egyoldali ellenhipotézis).

$$T(x) = \frac{s_X^{*2}}{s_Y^{*2}}$$

Ez $\mu = \mu_0$ esetben F_{n_1-1, n_2-1} eloszlású.

F-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból. Hipotéziseink:

- H_0: $\sigma_1^2 \leq \sigma_2^2$, H_1: $\sigma_1^2 > \sigma_2^2$ (egyoldali ellenhipotézis).
- H_0: $\sigma_1^2 = \sigma_2^2$, H_1: $\sigma_1^2 \neq \sigma_2^2$ (egyoldali ellenhipotézis).

$$T(x) = \frac{s_X^{*2}}{s_Y^{*2}}$$

Ez $\mu = \mu_0$ esetben F_{n_1-1,n_2-1} eloszlású.

- Egyoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x, y) \mid T(x, y) > F^{-1}_{F_{n_1-1,n_2-1}}(1-\alpha)\}$

F-próba

Legyen X n_1 elemű minta az $N(\mu_1, \sigma_1^2)$ eloszlásból, Y n_2 elemű minta az $N(\mu_2, \sigma_2^2)$ eloszlásból. Hipotéziseink:

- $H_0\colon \sigma_1^2 \leq \sigma_2^2$, $H_1\colon \sigma_1^2 > \sigma_2^2$ (egyoldali ellenhipotézis).
- $H_0\colon \sigma_1^2 = \sigma_2^2$, $H_1\colon \sigma_1^2 \neq \sigma_2^2$ (egyoldali ellenhipotézis).

$$T(x) = \frac{s_X^{*2}}{s_Y^{*2}}$$

Ez $\mu = \mu_0$ esetben F_{n_1-1,n_2-1} eloszlású.

- Egyoldali ellenhipotézis esetén legyen
  $\mathcal{X}_1 = \{(x,y) \mid T(x,y) > F^{-1}_{F_{n_1-1,n_2-1}}(1-\alpha)\}$
- Kétoldali ellenhipotézis esetén legyen $\mathcal{X}_1 = \{(x,y) \mid T(x,y) >$
  $> F^{-1}_{F_{n_1-1,n_2-1}}(1-\frac{\alpha}{2}) \vee T(x,y) < F^{-1}_{F_{n_1-1,n_2-1}}(\frac{\alpha}{2})\}$

Illeszkedés-vizsgálat

Legyen X n elemű minta tetszőleges eloszlásból. Ezeket s osztályba soroljuk, az i-edik osztályba N_i került. Hipotéziseink:

- H_0: az osztályok valószínűségei $p_1, \ldots, p_s$, H_1: nem

Illeszkedés-vizsgálat

Legyen X n elemű minta tetszőleges eloszlásból. Ezeket s osztályba soroljuk, az i-edik osztályba N_i került. Hipotéziseink:

- H_0: az osztályok valószínűségei $p_1, \ldots, p_s$, H_1: nem

$$T(x) = \sum_{i=1}^{s} \frac{(N_i - np_i)^2}{np_i}$$

Illeszkedés-vizsgálat

Legyen X n elemű minta tetszőleges eloszlásból. Ezeket s osztályba soroljuk, az i-edik osztályba N_i került. Hipotéziseink:

- H_0: az osztályok valószínűségei $p_1, \ldots, p_s$, H_1: nem

$$T(x) = \sum_{i=1}^{s} \frac{(N_i - np_i)^2}{np_i}$$

Belátható, hogy ha H_0 teljesül és $n \to \infty$, akkor $T \to \chi^2_{s-1}$.

Illeszkedés-vizsgálat

Legyen X n elemű minta tetszőleges eloszlásból. Ezeket s osztályba soroljuk, az i-edik osztályba N_i került. Hipotéziseink:

- H_0: az osztályok valószínűségei $p_1, \ldots, p_s$, H_1: nem

$$T(x) = \sum_{i=1}^{s} \frac{(N_i - np_i)^2}{np_i}$$

Belátható, hogy ha H_0 teljesül és $n \to \infty$, akkor $T \to \chi^2_{s-1}$.

- $\mathcal{X}_1 = \{x \mid T(x) > F^{-1}_{\chi^2_{s-1}}(1-\alpha)\}$

Illeszkedés-vizsgálat

Legyen X n elemű minta tetszőleges eloszlásból. Ezeket s osztályba soroljuk, az i-edik osztályba N_i került. Hipotéziseink:

- H_0: az osztályok valószínűségei $p_1, \ldots, p_s$, H_1: nem

$$T(x) = \sum_{i=1}^{s} \frac{(N_i - np_i)^2}{np_i}$$

Belátható, hogy ha H_0 teljesül és $n \to \infty$, akkor $T \to \chi^2_{s-1}$.

- $\mathcal{X}_1 = \{x \mid T(x) > F^{-1}_{\chi^2_{s-1}}(1-\alpha)\}$

Ez egy nagymintás próba, meg szokták követelni, hogy minden osztályba legalább $4-6$ megfigyelés legyen. (Osztályokat összevonhatjuk például, ha ez nem teljesül.)

Becsléses illeszkedés-vizsgálat

A feladat ugyanaz, mint az előbb, csak most p_i-ket nem tudjuk konkrétan, hanem azok is függnek egy $\gamma \in \Gamma \subseteq \mathbb{R}^t$ dimenziós paramétertől.

Becsléses illeszkedés-vizsgálat

A feladat ugyanaz, mint az előbb, csak most p_i-ket nem tudjuk konkrétan, hanem azok is függnek egy $\gamma \in \Gamma \subseteq \mathbb{R}^t$ dimenziós paramétertől.
Ha $t < s - 1$, valamint p_i-k elég szépen függnek γ-tól ($\gamma \mapsto p(\gamma)$ folytonosan differenciálható és a Jacobi-mátrixának rangja t), akkor működik a következő:

Becsléses illeszkedés-vizsgálat

A feladat ugyanaz, mint az előbb, csak most p_i-ket nem tudjuk konkrétan, hanem azok is függnek egy $\gamma \in \Gamma \subseteq \mathbb{R}^t$ dimenziós paramétertől.
Ha $t < s - 1$, valamint p_i-k elég szépen függnek γ-tól ($\gamma \mapsto p(\gamma)$ folytonosan differenciálható és a Jacobi-mátrixának rangja t), akkor működik a következő:
Vegyük γ ML-becslését ($\hat{\gamma}$), és végezzük el az előző dián leírt $T(x)$ statisztikát a $p_i(\hat{\gamma})$ valószínűségekkel.

Becsléses illeszkedés-vizsgálat

A feladat ugyanaz, mint az előbb, csak most p_i-ket nem tudjuk konkrétan, hanem azok is függnek egy $\gamma \in \Gamma \subseteq \mathbb{R}^t$ dimenziós paramétertől.
Ha $t < s - 1$, valamint p_i-k elég szépen függnek γ-tól ($\gamma \mapsto p(\gamma)$ folytonosan differenciálható és a Jacobi-mátrixának rangja t), akkor működik a következő:
Vegyük γ ML-becslését ($\hat{\gamma}$), és végezzük el az előző dián leírt $T(x)$ statisztikát a $p_i(\hat{\gamma})$ valószínűségekkel.
Ekkor ha igaz a nullhipotézisünk, vagyis ha p_i-k tényleg $p_i(\gamma)$ alakúak, akkor $T \to \chi^2_{s-t-1}$, tehát végezhetünk ismét χ^2-próbát.

Homogenitás-vizsgálat

Most legyen két független mintánk n és m megfigyeléssel, ugyanabba az s osztályba soroljuk őket. Az osztályok valószínűségei p_i, q_i ; N_i, M_i megfigyelés került beléjük. Hipotéziseink:

- H_0: az osztályok valószínűségeire $p_1 = q_1, \ldots, p_s = q_s$, H_1: nem

Homogenitás-vizsgálat

Most legyen két független mintánk n és m megfigyeléssel, ugyanabba az s osztályba soroljuk őket. Az osztályok valószínűségei p_i, q_i; N_i, M_i megfigyelés került beléjük. Hipotéziseink:

- H_0: az osztályok valószínűségeire $p_1 = q_1, \ldots, p_s = q_s$, H_1: nem

$$T = mn \sum_{i=1}^{s} \frac{\left(\frac{N_i}{n} - \frac{M_i}{m}\right)^2}{N_i + M_i}$$

Homogenitás-vizsgálat

Most legyen két független mintánk n és m megfigyeléssel, ugyanabba az s osztályba soroljuk őket. Az osztályok valószínűségei p_i, q_i; N_i, M_i megfigyelés került beléjük. Hipotéziseink:

- H_0: az osztályok valószínűségeire $p_1 = q_1, \dots, p_s = q_s$, H_1: nem

$$T = mn \sum_{i=1}^{s} \frac{\left(\frac{N_i}{n} - \frac{M_i}{m}\right)^2}{N_i + M_i}$$

Belátható, hogy H_0 esetén $T \to \chi^2_{s-1}$, erre végezzünk χ^2-próbát.

Függetlenség-vizsgálat

Van n megfigyelésünk, két szempont szerint osztályozzuk ezeket. Az első szerint s, a második szerint t osztályba osztjuk. N_{ij} elem került az (i, j) osztályba, az osztály valószínűsége p_{ij}.
Jelölés: $N_{i.} = \sum_{j=1}^{t} N_{ij}$, $N_{.j} = \sum_{i=1}^{s} N_{ij}$
Hipotéziseink:

- H_0: a két szempont független, azaz $\forall i, j$-re $p_{ij} = p_{i.}p_{.j}$, H_1: nem

Függetlenség-vizsgálat

Van n megfigyelésünk, két szempont szerint osztályozzuk ezeket. Az első szerint s, a második szerint t osztályba osztjuk. N_{ij} elem került az (i, j) osztályba, az osztály valószínűsége p_{ij}.
Jelölés: $N_{i.} = \sum_{j=1}^{t} N_{ij}$, $N_{.j} = \sum_{i=1}^{s} N_{ij}$
Hipotéziseink:

- H_0: a két szempont független, azaz $\forall i, j$-re $p_{ij} = p_{i.}p_{.j}$, H_1: nem

$$T = n\left(\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{N_{ij}^2}{N_{i.}N_{.j}} - 1\right)$$

Függetlenség-vizsgálat

Van n megfigyelésünk, két szempont szerint osztályozzuk ezeket. Az első szerint s, a második szerint t osztályba osztjuk. N_{ij} elem került az (i, j) osztályba, az osztály valószínűsége p_{ij}.
Jelölés: $N_{i.} = \sum_{j=1}^{t} N_{ij}$, $N_{.j} = \sum_{i=1}^{s} N_{ij}$
Hipotéziseink:

- H_0: a két szempont független, azaz $\forall i, j$-re $p_{ij} = p_{i.}p_{.j}$, H_1: nem

$$T = n\left(\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{N_{ij}^2}{N_{i.}N_{.j}} - 1\right)$$

Belátható, hogy H_0 esetén $T \to \chi^2_{(s-1)(t-1)}$, erre végezzünk χ^2-próbát.

Példa

1875-1894 között 10 porosz hadtestnél a lórúgás általi halálesetek száma (1 adat/hadtest/év):

halálesetek száma	0	1	2	3	4
gyakoriság	109	65	22	3	1

Példa

1875-1894 között 10 porosz hadtestnél a lórúgás általi halálesetek száma (1 adat/hadtest/év):

halálesetek száma	0	1	2	3	4
gyakoriság	109	65	22	3	1

Hipotézis: A halálesetek száma egy év alatt egy hadtestben Poisson-eloszlású.
Hogyan kell ellenőrizni?

Példa

1875-1894 között 10 porosz hadtestnél a lórúgás általi halálesetek száma (1 adat/hadtest/év):

halálesetek száma	0	1	2	3	4
gyakoriság	109	65	22	3	1

Hipotézis: A halálesetek száma egy év alatt egy hadtestben Poisson-eloszlású.
Hogyan kell ellenőrizni?
Becsléses illeszkedésvizsgálat

Példa

Utolsó két oszlopot összevonjuk, mert túl kicsik:

halálesetek száma	0	1	2	≥ 3
gyakoriság	109	65	22	4

Példa

Utolsó két oszlopot összevonjuk, mert túl kicsik:

halálesetek száma	0	1	2	≥ 3
gyakoriság	109	65	22	4

Tehát azt szeretnénk belátni, hogy $\mathrm{Poisson}(\lambda)$ az eloszlásunk.

Példa

Utolsó két oszlopot összevonjuk, mert túl kicsik:

halálesetek száma	0	1	2	≥ 3
gyakoriság	109	65	22	4

Tehát azt szeretnénk belátni, hogy $\mathrm{Poisson}(\lambda)$ az eloszlásunk. Ehhez először λ-ra kell ML-becslést adnunk ($\hat{\lambda}$), majd illeszkedéses χ^2-próbát kell végeznünk, és remélhetőleg χ^2_2 eloszlást kapunk.

Poisson eloszlás paraméterének ML-becslése

Emlékeztetőül, ha $X \sim \mathrm{Poisson}(\lambda)$, akkor:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Poisson eloszlás paraméterének ML-becslése

Emlékeztetőül, ha $X \sim \mathrm{Poisson}(\lambda)$, akkor:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Ha n elemű mintánk van, akkor a likelihood-függvényünk (aminek a maximumát keressük):

$$L(\lambda \mid x) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!}$$

Poisson eloszlás paraméterének ML-becslése

Emlékeztetőül, ha $X \sim \mathrm{Poisson}(\lambda)$, akkor:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Ha n elemű mintánk van, akkor a likelihood-függvényünk (aminek a maximumát keressük):

$$L(\lambda \mid x) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!}$$

Logaritmusa:

$$l(\lambda \mid x) = -n\lambda + \ln \lambda \sum_{i=1}^{n} x_i + c(x) = -n\lambda + n\overline{x} \ln \lambda + c(x)$$

Poisson eloszlás paraméterének ML-becslése

Emlékeztetőül, ha $X \sim \text{Poisson}(\lambda)$, akkor:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Ha n elemű mintánk van, akkor a likelihood-függvényünk (aminek a maximumát keressük):

$$L(\lambda \mid x) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!}$$

Logaritmusa:

$$l(\lambda \mid x) = -n\lambda + \ln \lambda \sum_{i=1}^{n} x_i + c(x) = -n\lambda + n\overline{x} \ln \lambda + c(x)$$

Deriváljuk λ szerint:

$$-n + \frac{n\overline{x}}{\hat{\lambda}} = 0 \Longrightarrow \hat{\lambda} = \overline{x}$$

Példa

Kiszámoljuk az átlagot:

$$\hat{\lambda} = \overline{X} = \frac{109 \cdot 0 + 65 \cdot 1 + 22 \cdot 2 + 3 \cdot 3 + 4 \cdot 1}{200} = \frac{122}{200} = 0.61$$

Példa

Kiszámoljuk az átlagot:

$$\hat{\lambda} = \overline{X} = \frac{109 \cdot 0 + 65 \cdot 1 + 22 \cdot 2 + 3 \cdot 3 + 4 \cdot 1}{200} = \frac{122}{200} = 0.61$$

halálesetek száma (i)	0	1	2	≥ 3	összeg
gyakoriság (N_i)	109	65	22	4	200
valószínűség (p_i)	0.54	0.33	0.10	0.03	1.00
várt gyakoriság (np_i)	108	66	20	6	200
eltérés $(N_i - np_i)$	1	-1	2	-2	0
$\frac{(N_i - np_i)^2}{np_i}$	0.01	0.02	0.20	0.67	0.90

Példa

Kiszámoljuk az átlagot:

$$\hat{\lambda} = \overline{X} = \frac{109 \cdot 0 + 65 \cdot 1 + 22 \cdot 2 + 3 \cdot 3 + 4 \cdot 1}{200} = \frac{122}{200} = 0.61$$

halálesetek száma (i)	0	1	2	≥ 3	összeg
gyakoriság (N_i)	109	65	22	4	200
valószínűség (p_i)	0.54	0.33	0.10	0.03	1.00
várt gyakoriság (np_i)	108	66	20	6	200
eltérés ($N_i - np_i$)	1	-1	2	-2	0
$\frac{(N_i - np_i)^2}{np_i}$	0.01	0.02	0.20	0.67	0.90

$T(X) = 0.90$ jött ki, azt tudjuk, hogy ha a hipotézisünk igaz, akkor $T \approx \chi^2_2$. Mit jelent ez?

Köszönöm a figyelmet!