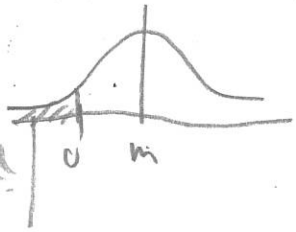
MAT. STATISZTIKA

Alapfeladat $\underbrace{X_1,\dots,X_n}_{\text{MINTA (n TERJEDELMŰ)}} \sim X$ AZONOS ELOSZLÁSÚ FGTL VÁLT

Ez konkrétan feltételezi: Pl. EMBEREK MAGASSÁGA $\approx N(m,\sigma)$ (mint sok FGTL VÁLT ÖSSZEGE)

Feladat: 1) m, σ becslése
2) TÉNYLEG NORM. A MINTA?



$X_1,\dots,X_n$
n VÉLETLENSZERŰEN VÁLASZTOTT EMBER MAGASSÁGA

$P(X<0)>0$
pedig 0 magas ember nincs

Paraméter A SZÓBAJÖHETŐ ELOSZLÁSOK INDEXELÉSE VALAMILYEN FUNKCIONÁL SZERINT

Pl. Emberek mag-nál $\mathbb{P}_{m,\sigma}$ $N(m,\sigma)$ normális

$$\mathbb{P}_{m,\sigma}(X<x)=\int_{s=-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-(s-m)^2/(2\sigma^2)}\,ds$$

Átlag $S_n(X)=(X_1+\dots+X_n)/n : \Omega\ni\omega\mapsto \frac{X_1(\omega)+\dots+X_n(\omega)}{n}$

EGY VÉLETLEN MÉRÉSI-SOROZAT

Kolmogorov Tétele Ha $\mathbb{E}(X)=\int_{x=-\infty}^{\infty} x\,dF_X(x)$ létezik,

Akkor $S_n(X)\xrightarrow{mb}\mathbb{E}(X)$ $(n\to\infty)$

VAGYIS $\mathbb{P}\{\omega: \frac{1}{n}[X_1(\omega)+\dots+X_n(\omega)]\not\to E(X)\}=0$

Empirikus eloszlásfgv-ek

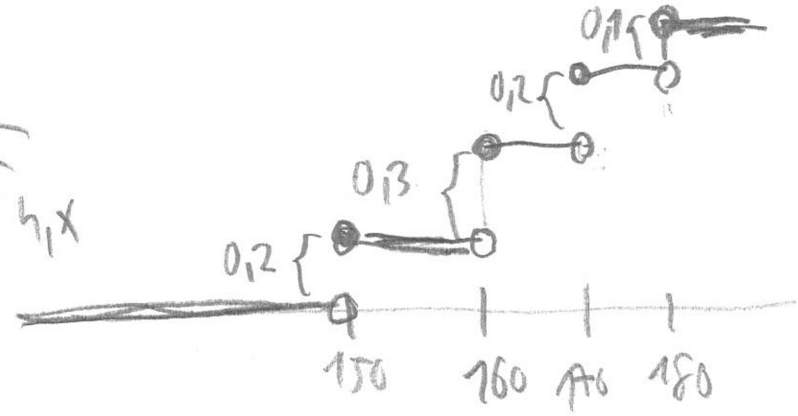
$\omega \in \Omega$ (egy véletlen mintavétel az esemény)

$X_1(\omega), \dots, X_n(\omega)$ konkrét mintaértékek (n db szám)

$$F_{n,X}(x) = F_n^{\omega}(x) = \frac{1}{n} \# \{ k : X_k(\omega) < x \}$$

Pl. egy mintavételnél 10 ember MAG: 150, 170, 150, 160, 170, 180, 180, 170, 160, 160

$$F^{\omega}_{n,X} = F_{n,X}$$



Minden mérésnél újabb fgv

KOLMOGOROV $\Rightarrow$ $F_{n,X}(x) \xrightarrow{m.b.} F_X(x)$ $\forall x \in \mathbb{R}$, azaz

$$P\{\omega : F^{\omega}_{n,X}(x) \to F_X(x) \ (n \to \infty)\} = 1 \quad \forall \text{ rögz } x \in \mathbb{R}$$

$$F_X(x) = \mathbb{P}(X < x) = \mathbb{P}\{\omega : X(\omega) < x\}$$

MAT STAT ALAPTÉTELE $\sup_{x \in \mathbb{R}} |F_{n,X}(x) - F_X(x)| \xrightarrow{m.b.} 0 \ (n \to \infty)$

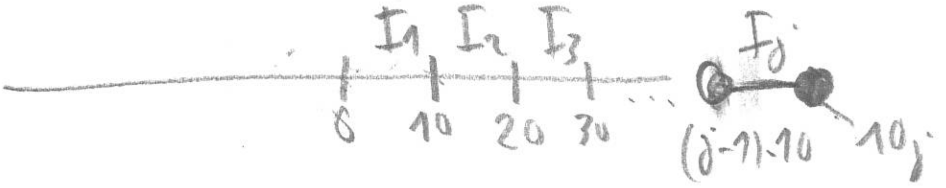
azaz $\mathbb{P}\{\omega : F^{\omega}_{n,X} \underset{\text{egyenletes konv}}{\rightrightarrows} F_X\} = 1$

Hisztogramok $\mathbb{R} = \ldots \cup I_{-1} \cup I_0 \cup I_1 \cup I_2 \cup \ldots$ $\mathbb{R}$ beosztása INTV-okra

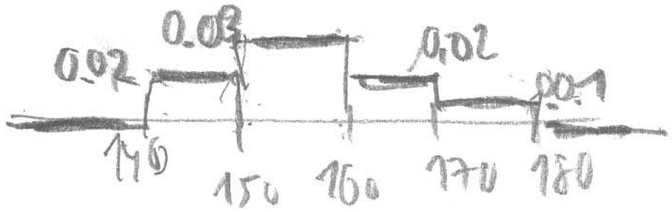
$$f^{J}_{n,X}(x) = f^{J,\omega}_{n,X}(x) = \left[\frac{\#\{k : X_k(\omega) \in I_j^n\}}{n\,|I_j|} \text{ HA } x \in I_j^n\right]$$

$$f^{\omega}_{n,X} = \sum_{j=-\infty}^{\infty} \frac{1}{n}\#\{k : X_k(\omega) \in I_j^n\} \cdot 1_{I_j} / |I_j^n|$$

Pl [életkor]



$f^{\omega}_{n,X}$



Tétel Ha $f_X : x \mapsto \frac{d}{dt}\Big|_{t=x} F_X(t) = F_X'(x)$ FOLYT, akkor

$F_X'(t) \to 0$ $(|t| \to \infty)$

$$\sup_{x\in\mathbb{R}} \left| f^{J_n}_{n,X}(x) - f_X(x) \right| \xrightarrow{m.b.} 0 \quad (n\to\infty) \quad \text{HA } \sup_j |I_j| \to 0$$

X sűrűségFGV-e

ALAPMENNYISÉGEK BECSLÉSEI

1) Várható érték: $\mathbb{E}(X) = \int_{x\in\mathbb{R}} x \, dF_X(x) = \int_{\omega\in\Omega} X(\omega)\, \mathbb{P}(d\omega)$

$$\overline{X}^{(n)} = \frac{1}{n}(X_1 + \ldots + X_n) \xrightarrow{m.b.} \mathbb{E}(X)$$

2) Szórásnégyzet: $\mathbb{D}(X)^2 = \mathbb{E}[X - \mathbb{E}(X)]^2 = \int_{\omega\in\Omega} [X(\omega) - \mathbb{E}(X)]^2 \, \mathbb{P}(d\omega)$

$$V_n(X) = \frac{1}{n} \sum_{k=1}^{n} [X_k - \overline{X}^{(n)}]^2 \xrightarrow{m.b.} \mathbb{D}^2(X)$$

Megjegyzés $\mathbb{E}(V_n(X)) \neq \mathbb{D}^2(X)$ HANEM $= \frac{n-1}{n}\mathbb{D}^2(X)$!!!

Biz: $n\mathbb{E}(V_n(X)) = n\mathbb{E}\left\{\left(X_1 - \frac{X_1+\dots+X_n}{n}\right)^2 + \dots + \left(X_n - \frac{X_1+\dots+X_n}{n}\right)^2\right\} =$

$$nV_n(X) = \left[X_1^2 - 2X_1\frac{X_1+\dots+X_n}{n} + \left(\frac{X_1+\dots+X_n}{n}\right)^2\right] + \dots +$$

$$+\left[X_n^2 - 2X_n\frac{X_1+\dots+X_n}{n} + \left(\frac{X_1+\dots+X_n}{n}\right)^2\right] =$$

$$= (X_1^2+\dots+X_n^2) - 2(X_1+\dots+X_n)\frac{X_1+\dots+X_n}{n} + n\left(\frac{X_1+\dots+X_n}{n}\right)^2 =$$

$$= (X_1^2+\dots+X_n^2) - \frac{1}{n}(X_1+\dots+X_n)^2 =$$

$$= \sum_{k=1}^{n} X_k^2 - \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \frac{2}{n}\sum_{1\le i<j\le n} X_i X_j =$$

$$= \left(1-\frac{1}{n}\right)\sum_{k=1}^{n} X_k^2 - \frac{2}{n}\sum_{1\le i<j\le n} X_i X_j$$

$$\mathbb{E}(V_n(X)) = \frac{1}{n}\left(1-\frac{1}{n}\right)\sum_{k=1}^{n}\underset{\text{Azonos eloszl.}}{\mathbb{E}(X^2)} - \frac{2}{n^2}\sum_{1\le i<j\le n}\underset{\text{FGTL}}{\mathbb{E}(X)\mathbb{E}(X)} =$$

$$= \left(1-\frac{1}{n}\right)\mathbb{E}(X^2) - \frac{2}{n^2}\binom{n}{2}\mathbb{E}(X)^2 =$$

$$= \frac{n-1}{n}\mathbb{E}(X^2) - \frac{n-1}{n}\mathbb{E}(X)^2 =$$

$$= \frac{n-1}{n}\left(\mathbb{E}(X^2) - \mathbb{E}(X)^2\right) = \frac{n-1}{n}\mathbb{D}(X)^2$$

Def Legyen $\mathcal{D}$ eloszláscsalád $\theta : \mathcal{D} \to \mathbb{R}$ paraméter

(pl. $\mathcal{D} = \{N(m,\sigma)$ normális eloszlások$\}$, $\theta = m : X \mapsto \mathbb{E}(X)$)

Egy $T : \mathbb{R}^n \to \mathbb{R}$ fgv TORZÍTATLAN BECSLÉSÉT ADJA θ-nak

ha $\mathbb{E}(T(X_1, \dots, X_n)) = \theta(X)$ minden $X \in \mathcal{D}$ vltóra

Lemma: $\bar{X}^n = \frac{1}{n}(X_1 + \dots + X_n)$ AKÁRMELYEN $\mathcal{D}$-ből TORZÍTATLAN BECSLÉSE $\mathbb{E}(X)$-nek (VÁRHATÓ ÉRT.)

Ugyanakkor $V_n = \frac{1}{n} \sum_{k=1}^{n} (X_k - \bar{X}^n)^2$ NEM TORZÍTATLAN $\mathbb{D}^2(X)$-re,

DE $V_n(X) = \frac{n-1}{n} \mathbb{D}(X)^2 \to \mathbb{D}(X)^2$ $(n \to \infty)$

Def $T_1, T_2, \dots$ becslőfgv-sorozat $T_n : \mathbb{R}^n \to \mathbb{R}$

ASZIMPTOTIKUSAN TORZÍTATLAN BECSLÉST AD θ-ra, ha

$T_n(X_1, \dots, X_n) \to \theta(X)$ minden $X (\in \mathcal{D})$-re esetén

Tehát: $V_1, V_2, \dots$ a szórásnégyzet ASZ. TORZ. BECSL.

HF $C_n = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x}^n)(y_k - \bar{y}^n)$ ASZ. TORZÍTATLAN

$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] =$

$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ -nak

ha $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$

MAX-LIKELIHOOD BECSLŐFGV-EK

<u>Adott</u> $\mathcal{D}$ eloszlás-család θ-val paraméterezve

Pontosan: $\mathcal{D} = \{F_\theta : \theta \in \underset{\mathbb{R}^N}{\overset{\cap}{G}}\}$, $F_\theta(x)$ eloszlásfgv

Pl. $\mathcal{D} = \{\text{normális eloszl.}\} = \{N(\underbrace{m,\sigma}_{F_{m,\sigma} \text{ eloszlásfgv}}) : m \in \mathbb{R}, \sigma > 0\}$

$$F_{m,\sigma}(x) = \int_{t=-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-m)^2/(2\sigma^2)}\, dt$$

<u>Adott</u> $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ vv-sorozat

$X_1(\omega), X_2(\omega), \ldots$ egy <u>KONKRÉT</u> MÉRÉS-SOROZAT

<u>Feltevés</u> $X_1, X_2, \ldots$ FGTL AZONOS ELOSZL. $X_n \sim X$

$\underline{F_X \in \mathcal{D}}$ tehát $F_X = F_\theta$ valamely θ-ra

<u>Kérdés</u> $\theta = ???$ $X_1(\omega), X_2(\omega), \ldots$ ALAPJÁN

<u>Max-likelihood</u>: Egy $X_1(\omega), \ldots, X_n(\omega)$ eredmény MILYEN θ MELLETT A LEGVALÓSZÍNŰBB?

└ bizonytalan fogalom(!)

Diszkrét eset X_θ értékei $\in \{y_1, y_2, \dots, y_M\}$ $\forall \theta \in \Theta$

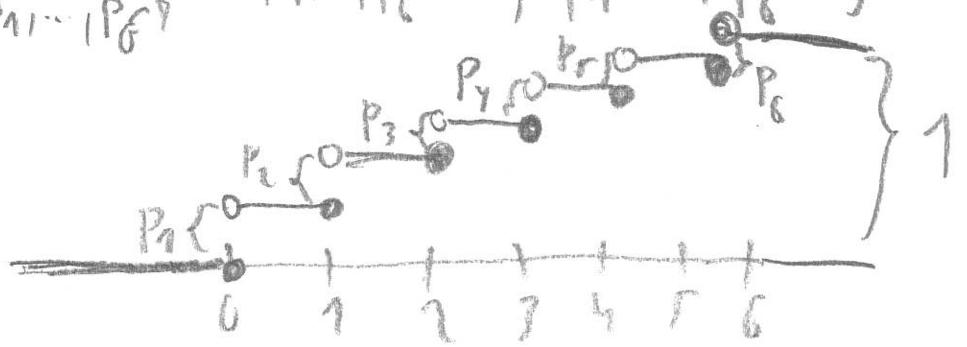
$[(X_\theta = y_j)$ valség-e$] = p_{j,\theta}$ $(j = 1, 2, \dots, M)$ $p_{1,\theta} + \dots + p_{M,\theta} = 1$

Példa Kockadobás $X_k = [k.$ dobás eredménye$]$

$\{y_1, \dots, y_M\} = \{1, 2, 3, 4, 5, 6\}$

$\mathcal{S} = \{F_{(p_1, \dots, p_6)} : p_1, \dots, p_6 \geq 0,\ p_1 + \dots + p_6 = 1\}$

$F_{p_1, \dots, p_6} =$



Amikor dobunk a kockával, hogy az $[1$-es valség$] = p_1, \dots$ stb.

Kísérletsorozat $X_1(\omega), X_2(\omega), X_3(\omega), \dots$

Pl.: 4 4 2 6, ...

HA $X_1, X_2, \dots \sim X_\theta$ eloszlású, AKKOR

$[(X_1 = x_1(\omega), X_2 = x_2(\omega), \dots, X_n = x_n(\omega))$ valség-e$] =$

$= p_{x_1(\omega),\theta} \cdot p_{x_2(\omega),\theta} \cdots p_{x_n(\omega),\theta}$ ÁLTALÁBAN IS

Pl.: $\theta = (\frac{1}{9}, \frac{1}{3}, \frac{1}{3}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9})$ esetén

a 4, 4, 2, 6 DOBÁSSOROZAT $\frac{1}{9} \cdot \frac{1}{9} \cdot \frac{1}{3} \cdot \frac{1}{6}$

Technikai feltevés (MAT LEHETSÉGES!)

VAN OLYAN Ω ESEMÉNYTÉR $\mathcal{A} \subset \{\Omega$ részhalmazai$\}$ melyek események

és $\forall \theta \in G$-re van $\mathbb{P}_\theta$ VALÓSZ. $\mathcal{A} \to [0,1]$, úgy hogy

$$F_\theta(x) = \mathbb{P}_\theta\{\omega \in \Omega : X_k(\omega) < x\} \quad \forall k$$

$\mathbb{P}_\theta$ jelentése: $\mathbb{P}_\theta(A) = [$Az A esemény VALÓSZ-E, HA θ A VALÓDI PARAMÉTER$]$

$$\mathbb{P}_\theta(\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}) =$$

$$= (\mathbb{P}_\theta\{x_1\}) \cdot (\mathbb{P}_\theta\{x_2\}) \cdots (\mathbb{P}_\theta\{x_n\})$$

Ezt kell MAXIMALIZÁLNI θ-RA

Példa: KOCKADOBÁSOK: 4, 4, 2, 6, 5, 3, 3, 4, 1, 4

$$\mathbb{P}_{(p_1, \ldots, p_6)}(X_1 = 4, X_2 = 4, \ldots, X_9 = 1, X_{10} = 4) =$$

$$= p_4 \cdot p_4 \cdot p_2 \cdot p_6 \cdot p_5 \cdot p_3 \cdot p_3 \cdot p_4 \cdot p_1 \cdot p_4 =$$

$$= p_1 \cdot p_2 \cdot p_3^2 \cdot p_4^4 \cdot p_5 \cdot p_6 \longrightarrow \underset{\substack{p_1, \ldots, p_6 \geq 0 \\ p_1 + \cdots + p_6 = 1}}{\text{MAX}}$$

Def LIKELIHOOD FGV (n-ik):

$$L_{\theta,n}: (x_1,\dots,x_n) \mapsto (\mathbb{P}_\theta\{x_1\})\cdots(\mathbb{P}_\theta\{x_n\})$$ DISZKRÉT ESET

$$\ell_{\theta,n}: (x_1,\dots,x_n) \mapsto [f_\theta(x_1)]\cdots[f_\theta(x_n)]$$ FOLYT ESET

$$F_\theta' = f_\theta$$

θ MAX-LIKELIHOOD BECSLÉSEI AZ $\{F_\theta : \theta \in G\}$ ELOSZLÁSOKNÁL

1) $\operatorname{argmax} L_{\theta,n}(x_1,\dots,x_n)$ (DISZKR.)

AZ A θ ÉRTÉK, AHOL MAXIMÁLIS

2) $\operatorname{argmax} \ell_{\theta,n}(x_1,\dots,x_n)$

HA A KÍSÉRLETEK EREDMÉNYEI $x_1, x_2, \dots,$

Példa: KOCKADOBÁS 4, 4, 2, 6, 5, 3, 3, 4, 1, 4

$$L_\theta = p_1 p_2 p_3^2 p_4^4 p_5 p_6 \to MAX \quad p_1,\dots,p_6 \geq 0, \; p_1+\dots+p_6=1$$

$$\nabla L_\theta = \mu \cdot \nabla(p_1+\dots+p_6-1)$$

$$p_2 p_3^2 p_4^4 p_5 p_6 = \mu \cdot 1 \qquad p_1 p_2 p_3^2 (4 p_4^3) p_5 p_6 = \mu$$

$$p_1 p_3^2 p_4^4 p_5 p_6 = \mu \qquad p_1 p_2 p_3^2 p_4^4 p_6 = \mu$$

$$p_1 p_2 (2p_3) p_4^4 p_5 p_6 = \mu \qquad p_1 p_2 p_3^2 p_4^4 p_5 = \mu$$

SZIMMETRIA $\Rightarrow$ VEHETŐ $p_1 = p_2 = p_5 = p_6$

$$\mu = p^3 p_3^2 p_4^4 = 2 p^4 p_3 p_4^4 = 4 p^4 p_3^2 p_4^3 \quad \Big/ : (p^3 \cdot p_3 \cdot p_4^3)$$

4 EGYENLETBEN

$$\frac{\mu}{p^3 p_3^2 p_4^3} = \Big| \; p_3 p_4 = 2 p p_4 = 4 p p_3$$

$\underline{q_i = \log_2 p_i}$ $\quad q_3 + q_4 = 1 + q + q_4 = 2 + q + q_3$

$$q_3 = 1 + q, \; q_4 = 2 + q$$

$p = 2^q = p_1 = p_2 = p_5, p_6 \quad p_3 = 2^{1+q} = 2 \cdot p \quad p_4 = 2^{2+q} = 4p$

$p + p + 2p + 4p + p + p = 1 \quad p = \frac{1}{10}$

$$\underline{\underline{\theta = \left(\frac{1}{10}, \frac{1}{10}, \frac{2}{10}, \frac{4}{10}, \frac{1}{10}, \frac{1}{10}\right)}}$$