Maximum likelihood becslés

Mintát veszünk az X val.változóból: $x_1, x_2, \ldots, x_n$, és ez alapján szeretnénk kitalálni X eloszlását.

Könnyítés: Tudjuk, hogy X diszkrét, sőt, az eloszlását is ismerjük egy paraméter erejéig. [Avagy: adott egy 1-paraméteres eloszláscsalád.] Cél a paraméter ~~megtalálása~~ becslése.

Pl1: Tudjuk, hogy $X \sim B(p)$, csak p ismeretlen

[Mondjuk onnan tudjuk, hogy $n=100$-szor feldobtunk egy hamis érmét, és kijött, hogy $\underbrace{FF11F1FF\ldots F1F}_{60 \text{ db } F,\ 40 \text{ db } 1}$.

Legyen $X = \begin{cases} 1, & \text{ha fej} \\ 0, & \text{ha nem} \end{cases}$ így a minta $\underbrace{11001011\ldots 1}_{60 \text{ db } 1\text{-es},\ 40 \text{ db } 0}$]

Ennek a konkrét mintának a valószínűsége

$$\cancel{\mathbb{P}_p(x_1, x_2, \ldots, x_n)} = \underbrace{p \cdot p \cdot (1-p) \cdot (1-p) \cdot p \cdot (1-p) \cdot p \cdot p \cdot \ldots \cdot p}_{60 \text{ db } p,\ 40 \text{ db } (1-p)}$$

$= p^{60}(1-p)^{40}$ persze függ a p paramétertől

Pontosabban: $\mathbb{P}_p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

Nyilván: minél nagyobb ez a val-ség, annál kevésbé „meglepő" az adott minta.

<u>Def</u>: a <u>Likelihood-függvénynek</u> nevezzük a $\mathbb{P}_p(X_1=x_1, X_2=x_2, \dots, X_n=x_n)$ valószínűséget, $\boxed{\text{mint a } p \text{ függvényét}}$, miközben az $x_1, x_2 \dots x_n$ mintát <u>paraméternek</u> tekintjük:

$$L(p)=L(\underbrace{x_1, x_2, \dots, x_n}_{\text{paraméterek}}; \underset{\text{változó}}{p}) := \mathbb{P}(X_1=x_1, \dots, X_n=x_n).$$

Esetünkben: $L(x_1 \dots x_n; p) = p^{60}(1-p)^{40}$.

[<u>Megj</u>: Angolban likelihood ≈ valószínűség, de nem keverendő a probability = valószínűség szóval.
Magyarul lehetne mondjuk hihetőség-függvény ...]

<u>Def</u>: A p paraméter <u>maximum likelihood becslése</u> a likelihood-függvény (globális) <u>maximumhelye</u>.

[Vagyis az a $p=p_{ML}$, ahol $L(p)$ maximális, vagyis amely p esetén az adott minta a legkevésbé meglepő / leginkább hihető.]

<u>Keressük meg</u> ezt a p-t!

$L(p) = p^{60}(1-p)^{40} \longrightarrow \max$

Jó lenne deriválni p szerint, de egy (sok-tényezős) szorzat deriválása elég csúnya: összeget könnyebb deriválni, ezért

$\ell(p) := \ln L(p) = 60 \ln p + 40 \ln(1-p)$ a log-likelihood fv.

Ennek nyilván ugyanott van a maximuma [mert $L \to \ln L$ szig. mon. növ.]

vagyis $\ell(p) = 60 \ln p + 40 \ln(1-p) \to \max$,

ezt már szívesebben deriválom, és a

$$\ell'(p) = 60 \cdot \frac{1}{p} + 40 \cdot \frac{1}{1-p} \cdot (-1) = \frac{60}{p} - \frac{40}{1-p} = 0$$

egyenletet kell megoldani

$$\frac{60}{p} = \frac{40}{1-p} \qquad / \cdot p(1-p)$$

$$60(1-p) = 40p$$

$$60 - 60p = 40p \qquad / + 60p$$

$$60 = 100p \qquad / : 100$$

$$p = \frac{60}{100} = 0.6$$

Mivel $\ell(p)$-nek ez az egyetlen stacionárius helye (ahol $\ell'(p) = 0$), nyilván itt van az egyetlen maximum, vagyis $\boxed{p_{ML} = 0.6}$.

[Aki szeretné, ellenőrizheti a 2. deriváltat.]

Ugyanez általánosan

legyen inkább f, mert a p többnyire jól jön másra

Def: Legyen $\mathbb{P}(X=x) = \overset{f}{\cancel{p}}(\theta;x)$, ahol $x \mapsto \overset{f}{\cancel{p}}(\theta;x)$

diszkrét valószínűségeloszlás $\theta \in \mathbb{R}$ paraméterrel.
ez nem is fontes

Legyen $\underline{x}=(x_1, x_2, \dots, x_n)$ minta X-ből. Ekkor a

$$\theta \overset{L(\theta)}{\longmapsto} L(\underbrace{x_1,\dots,x_n}_{\text{paraméter}};\underbrace{\theta}_{\text{változó}}) := \prod_{i=1}^{n} f(\theta;x_i) \quad \text{függvényt}$$

likelihood-függvénynek nevezzük, az $L(\theta)$ globális maximumhelyét pedig a θ paraméter maximum likelihood becslésének. Jele: θ_{ML}

[Hát persze: ha a paraméter θ (lenne), akkor a minta val.sége $\mathbb{P}(X_1=x_1, X_2=x_2, \dots X_n=x_n)=\prod_{i=1}^{n} f(\theta;x_i)$ (lenne).]

[Az előző példában $\theta=p$, $f(\theta,x)=\begin{cases}\theta, & \text{ha } x=1\\ 1-\theta, & \text{ha } x=0\end{cases}$,

és kivételesen hasznos az $f(\theta,x)=\theta^{x}(1-\theta)^{1-x}$ $(x=0,1)$

zárt alak, így $L(\underline{x},\theta)=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}$.]

Köv: A log-likelihood-függvény

$$\ell(\theta)=\ell(\underline{x};\theta):=\ln L(\theta)=\sum_{i=1}^{n}\ln f(\theta;x_i)$$ maximuma is θ_{ML}-nél van.

Pl: $n=60$ elemű mintát vettünk az $X \sim Geom(\theta)$ val. változóból, és az jött ki, hogy

$$(x_1, x_2 \ldots, x_n) = \underbrace{(1,1,3,1,2,1,\ldots,2)}_{\text{60 db szám, az összegük 100 lett.}}$$

Adjunk maximum likelihood becslést θ-ra!

Megoldás: Az eloszlás ($\theta = p$ a szokásos jelöléssel) $q = 1-p$

$$\mathbb{P}(X=k) = q^{k-1} p = p(1-p)^{k-1} \qquad k=1,2,\ldots$$

vagyis $f(\theta;x) = \mathbb{P}(X=x) = \theta(1-\theta)^{x-1} \qquad x=1,2,\ldots$

$$\ln f(\theta;x) = \ln\theta + (x-1)\ln(1-\theta)$$

$$\Rightarrow \ell(\underline{x};\theta) = \sum_{i=1}^{n} \ln f(\theta;x_i) = \sum_{i=1}^{n}\left[\ln\theta + (x_i-1)\ln(1-\theta)\right]$$

log-likelihood függvény

$$= n\ln\theta + \ln(1-\theta)\left[\sum_{i=1}^{n} x_i - n\right]$$

$$= \cancel{n\ln\theta - \left[n - \sum_{i=1}^{n} x_i\right]\ln(1-\theta)}$$

Keressük a ~~maximumát~~ maximumhelyét:

$$\ell'(\theta) = n\cdot\frac{1}{\theta} + \left[\sum_{i=1}^{n} x_i - n\right]\cdot\frac{1}{1-\theta}\cdot(-1) = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} x_i - n}{1-\theta} = 0$$

$$\frac{\sum_{i=1}^{n} x_i - n}{1-\theta} = \frac{n}{\theta} \qquad /\cdot\,\theta(1-\theta)$$

$$\theta\left(\sum_{i=1}^{n} x_i - n\right) = n - n\theta \qquad /+n\theta$$

$$\theta\sum_{i=1}^{n} x_i = n \Rightarrow \boxed{\theta_{ML} = \theta = \frac{n}{\sum_{i=1}^{n} x_i}}$$

$= \dfrac{1}{\bar{x}}$ hát persze

nyilván ez az egyetlen stac. hely

$\Downarrow$

maximumhely

A konkrét esetben $n=60$, $\sum_{i=1}^{n} x_i = 100 \Rightarrow \boxed{\theta_{ML} = \frac{60}{100} = 0.6}$

Hát persze:

- Az 1. példa arról szólt, hogy 100 dobásból 60 lett fej
- a 2. példa arról, hogy a 60-adik fej a 100-adik dobásra jött ki.

Folytonos eloszlások

Legyen X abszolút folytonos val. változó, sűrűségfüggvénye $f(\theta;x)$, ahol θ ismeretlen paraméter.

Pl: $f(\theta;x) = \begin{cases} \theta e^{-\theta x}, & \text{ha } x \geq 0 \\ 0, & \text{ha } x < 0 \end{cases}$ vagyis $X \sim Exp(\theta)$.

Mintát veszünk X-ből, és az jön ki, hogy

$\underline{x} = (0.9;\ 0.2;\ 4.3;\ 0.7;\ 1.1;\ 0.4;\ 1.4;\ 1.0)$

Ekkor persze a konkrét $\underline{x}$ minta valószínűsége nulla:

eleve $\mathbb{P}(X=0.9)=0$, de attól még, ha $f(\theta;x)$ nagy, akkor az $\{X=x\}$ esemény „kevésbé meglepő", mint ha $f(\theta;x)$ kicsi (adott θ paraméter mellett). Ezért

Def: Legyen most is $L(\underline{x};\theta) := \prod_{i=1}^{n} f(\theta;x_i)$ a likelihood-függvény, θ_{ML} pedig ennek maximumhelye.

Megj: Az $L(\underline{x};\theta) := \prod_{i=1}^{n} f(\theta; x_i)$ definíció első ránézésre ugyanaz a diszkrét és a folytonos esetben, de ez csak a szerencsétlen jelölés(em) miatt van:

- a folytonos esetben $f(x) = f(\theta; x)$ <u>sűrűségfüggvényt</u>,
- a diszkrét esetben pedig <u>diszkrét eloszlást</u>

jelöl.

A ~~fen~~ fenti példában

$$L(\underline{x};\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i}$$, mert szerencsére minden $x_i \geq 0$

$$\ell(\underline{x};\theta) = \ln L(\underline{x};\theta) = \sum_{i=1}^{n} \left[\ln\theta - \theta x_i \right] = n \ln\theta - \theta \sum_{i=1}^{n} x_i$$

Megj: $\ell(\underline{x};\theta) = \sum_{i=1}^{n} \ln f(\theta; x_i)$ ahol $\ln f(\theta; x_i) = \begin{cases} \ln\theta - \theta x, & \text{ha } x \geq 0 \\ \underline{\text{értelmetlen}}, & \text{ha } x < 0. \end{cases}$

Vagyis ha valamelyik x_i negatív (lenne),

akkor $L(\underline{x};\theta) = 0$, (minden θ-ra) (lenne),

$\ell(\underline{x};\theta)$ pedig értelmetlen.

<u>Hát persze:</u> az ilyen minta <u>lehetetlen</u>: ha ez jön ki, akkor <u>rossz a modell</u> (vagy rossz a valóság).

Keressük $\ell(\theta)$ ~~maximumát~~ maximumhelyét:

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} x_i = \frac{n}{\theta} \Rightarrow \boxed{\theta_{ML} = \theta = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}}$$

[Nem meglepő: $\mathbb{E}X = \frac{1}{\theta}$, és $\mathbb{E}X$ becslése $\overline{x}$.]

A konkrét esetben $n = 8$, $\sum_{i=1}^{n} x_i = 10.0 \Rightarrow \underline{\underline{\theta_{ML} = 0.8}}$.

Példa: normális eloszlás paramétereinek M-L becslése

Legyen $X \sim N(m, \sigma)$, vagyis sűrűségfüggvénye

$$f_{m,\sigma}(x) = f(m, \sigma; x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Mintát veszünk X-ből: $\underline{x} = (x_1, x_2, \ldots, x_n)$.

① Tfh $\underline{\sigma \text{ ismert}}$, mondjuk $\sigma = 1$, ~~így~~. Így $\theta := m$ az egyetlen paraméter. Adjunk rá M-L becslést!

Megoldás: $\ln f(\theta; x) \overset{m=\theta}{\underset{\sigma \text{ ismert}}{=\!=}} \underbrace{\ln \frac{1}{\sqrt{2\pi}\,\sigma}}_{\text{konstans}} - \frac{(x-\theta)^2}{2\sigma^2} = c - \frac{(x-\theta)^2}{2\sigma^2}$

$$\Rightarrow \ell(\theta) = \sum_{i=1}^{n} \ln f(\theta; x_i) = \sum_{i=1}^{n} \left[c - \frac{(x_i-\theta)^2}{2\sigma^2} \right] = nc - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta)^2$$

maximumhelyét keressük:

$$\ell'(\theta) = 0 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i - \theta)(-1) = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(x_i - \theta)\right] = 0 \quad / \cdot \sigma^2$$

$$\cancel{n\theta =} \quad \sum_{i=1}^{n} x_i - n\theta = 0$$

$$\Downarrow$$

$$\underline{\underline{m_{ML}}} = \left[\theta_{ML} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}\right] \quad \text{Hurrá.}$$

(2) Tfh $\underline{m \text{ ismert}}$, mondjuk $m = 0$, így $\theta = \sigma$ az egyetlen paraméter. Adjunk rá M.L. becslést!

Megoldás: $\ln f(\theta; x) \underset{m \text{ ismert}}{\overset{\sigma = \theta}{=\!=}} \underbrace{\ln \frac{1}{\sqrt{2\pi}}}_{\text{konstans}} - \ln\theta - \frac{(x-m)^2}{2\theta^2} = c - \ln\theta - \frac{(x-m)^2}{2\theta^2}$

$$\Rightarrow \ell(\theta) = \sum_{i=1}^{n} \ln f(\theta; x_i) = nc - n\ln\theta - \frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - m)^2$$

$$\ell'(\theta) = 0 - \frac{n}{\theta} - \frac{1}{2} \sum_{i=1}^{n} (x_i - m)^2 \left(\frac{-2}{\theta^3}\right) = 0 \quad / + \frac{n}{\theta}$$

$$\left[\text{mert } \left(\frac{1}{\theta^2}\right)' = \left(\theta^{-2}\right)' = -2\theta^{-3} = \frac{-2}{\theta^3}\right]$$

$$\frac{n}{\theta} = \frac{\sum_{i=1}^{n} (x_i - m)^2}{\theta^3} \quad / \cdot \theta^3 / n$$

$$\theta^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2 \Rightarrow \boxed{\sigma_{ML} = \theta_{ML} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2}}$$

Hurrá.

2b Legyen most is m ismert, de $\theta = \sigma^2 \neq \sigma$ az egyetlen paraméter. Most

$$\ell(\theta) = nc - n \ln\sqrt{\theta} - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-m)^2,$$

ami ugyanaz, mint az előbb, csak $\theta^2 \to \theta$ átjelöléssel

$\Rightarrow$ a maximumhelye is „ugyanott" van:

$$\boxed{\left(\sigma^2\right)_{ML} = \left(\sigma_{ML}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-m)^2}$$ Hurrá.

Ez éppen az m-től való négyzetes eltérés átlaga, szuperül megfelel a $\sigma^2 = \mathbb{E}((X-m)^2)$ definíciónak.

3 Sem m-et, sem σ-t nem ismerjük, így legyen a paraméter $\boxed{\theta = (m, \sigma^2) = (m, v)}$ vektor.
[várható érték és variancia].

Adjunk rá(juk) ML becslést!

Megoldás: $f(\theta; x) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-m)^2}{2v}}$

$$\ln f(\theta;x) = -\ln(\sqrt{2\pi v}) - \frac{(x-m)^2}{2v} = \underbrace{-\ln\sqrt{2\pi}}_{\text{konstans}} - \frac{1}{2}\ln v - \frac{(x-m)^2}{2v}$$

$$\Rightarrow \ell(\theta) = \ell(m,v) = \sum_{i=1}^{n}\left[c - \frac{1}{2}\ln v - \frac{(x_i-m)^2}{2v}\right] = nc - \frac{n}{2}\ln v - \frac{1}{2v}\sum_{i=1}^{n}(x_i-m)^2$$

Vagyis

11/11

$$\boxed{\ell(\theta) = \ell(m,v) = \text{konstans} - \frac{n}{2}\ln v - \frac{1}{2v}\sum_{i=1}^{n}(x_i-m)^2} \qquad \boxed{\ell : \mathbb{R}^2 \to \mathbb{R}}$$

még mindig ugyanaz, mint az előző 2 esetben, csak most kétváltozós fv-ként keressük a maximumhelyét.

Ehhez legyen $\begin{cases} \dfrac{\partial \ell}{\partial m} = 0 \\ \dfrac{\partial \ell}{\partial v} = 0 \end{cases}$ egyszerre: ettől stacionárius hely az (m,v) pont. $\in \mathbb{R}^2$

$$\frac{\partial \ell}{\partial m} = -\frac{1}{2v}\sum_{i=1}^{n} 2(x_i-m)(-1) = \frac{1}{v}\sum_{i=1}^{n}(x_i-m) = \frac{1}{v}\left[\sum_{i=1}^{n} x_i - nm\right]$$

$$\frac{\partial \ell}{\partial v} = -\frac{n}{2}\frac{1}{v} + \frac{1}{2v^2}\sum_{i=1}^{n}(x_i-m)^2 = \frac{1}{2v^2}\left[\sum_{i=1}^{n}(x_i-m)^2 - v\cdot n\right]$$

Vagyis meg kell oldani az alábbi egyenletrendszert:

$$\begin{cases} \cancel{\frac{1}{v}}\left[\sum_{i=1}^{n} x_i - nm\right] = 0 \xRightarrow{\text{könnyű}} m = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \\ \cancel{\frac{1}{2v^2}}\left[\sum_{i=1}^{n}(x_i-m)^2 - n\cdot v\right] = 0 \xRightarrow{\text{könnyű}} v = \frac{1}{n}\sum_{i=1}^{n}(x_i-m)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \end{cases}$$

EGYBEN:

$$\boxed{(m,v)_{ML} = (\bar{x}, S_n^2)}$$

átlag — tapasztalati szórásnégyzet