2. Estymacja największej wiarygodności (max-likelihood)

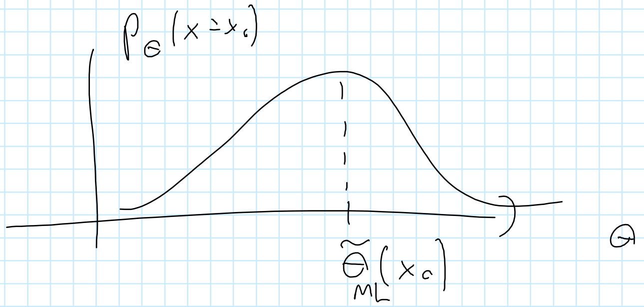
Co zrobić kiedy warunki wysycenia C-R nie spełniamy i nie potrafimy podać optymalnego estymatora.

2.1 Estymator największej wiarygodności

$$p_\theta(x)$$

jeśli mierzymy x_0 to jako estymator przyjmujemy taką wartość θ dla której $p_\theta(x)$ maksymalne

$$p_\theta(x=x_0) \text{ – likelihood function}$$



Intuicja: wartość parametru, dla którego obserwowane zdarzenie najbardziej prawdopodobne

Uwaga

Zazwyczaj wygodniej szukać max $\log p_\theta(x=x_0)$
a to do tego samego punktu bo log monotoniczny

Przykład

$N(\theta, \sigma^2)$

$$x_i \sim N(\theta, \sigma^2)$$

$$\frac{d}{d\theta} \log p_\theta(x) = 0 \qquad \text{- szukamy max}$$

$$\frac{1}{\sigma^2} \sum_i (x_i - \theta) = 0 \qquad \Rightarrow \qquad \tilde{\theta}_{ML}(x) = \frac{1}{N} \sum_i x_i$$

<u>Ogólna własność</u> - jeśli istnieje estymator efektywny - wysycający C-R to zawsze to ten sam, który znajdziemy max-likelihood

Pamiętamy że warunek wysycenia to:

$$\frac{d}{d\theta} \log p_\theta(x) = \lambda(\theta)\big(\underbrace{f(x)}_{\text{efektywny estymator}} - \theta \big)$$

czyli θ dla którego minimum to $\tilde{\theta}_{ML} = f(x)$ ■

W szczególności dla modeli liniowych

2.2. Asymptotyczna efektywność Max-likelihood

W ogólności nie ma gwarancji, że max-likelihood wysyci C-R ani, że jest najlepszym estymatorem

ale ... jeśli dużo niezależnych obserwacji to tak

<u>Twierdzenie</u>

$$x^{(N)} = [x_1, \ldots, x_N]$$

$$p_\theta^{(N)}(x^{(N)}) = p_\theta(x_1) \cdot \ldots \cdot p_\theta(x_N)$$

$\Big\}$ x_i może reprezentować wiele danych

$\Uparrow$

N powtórzeń
eksperymentu

wtedy w granicy $N \to \infty$, estymator
ML będzie miał rozkład:

$$\tilde{\Theta}_{ML} \overset{N\to\infty}{\sim} \mathcal{N}\left(\Theta_0, F^{(N)^{-1}}\right), \quad F^{(N)} = N \cdot F$$

czyli wysyca C-R, i jest nieobciążony

<u>Lemat:</u>
Niech $p_1(x)$, $p_2(x)$ - rozkłady prawdop.

$$\int p_1(x) \log \frac{p_1(x)}{p_2(x)} \geqslant 0$$

$$\parallel$$

$D(p_1 | p_2)$ - entropia względna

<u>Dowód lematu:</u>

funkcja $\log t$ - wklęsła tzn. $\log\left(\sum_i w_i t_i\right) \geq \sum_i w_i \log t_i$

$w_i \geq 0 \quad \sum w_i = 1$

$t_i = \frac{1}{p_1(x)} \qquad w_i = p_1(x)$

$$\sum_x \underbrace{p_1(x)}_{w_i} \log \underbrace{\frac{p_2(x)}{p_1(x)}}_{t_i} \leq \log \sum_x p_2(x) = 0$$

czyli $\quad \sum_x p_1(x) \log \frac{p_1(x)}{p_2(x)} \geq 0$ ■

Dowód twierdzenia

2. Założenia techniczne: • istnieją 1-sza, 2-ga pochodna $\log p_\theta(x)$

• $\langle \frac{\partial}{\partial \theta} \log p_\theta(x) \rangle = 0$

• asymptotyczna nieobciążoność

Niech $\tilde{\theta}$ - będzie naszym estymatorem, funkcja wiarygodności dla niego:

$$\frac{1}{N} \log p^{(N)}_{\tilde{\theta}}(x^{(N)}) = \frac{1}{N} \sum_i \log p_{\tilde{\theta}}(x_i)$$

z prawa wielkich liczb, to zbiega do średniej czyli

$$\frac{1}{N} \log p^{(N)}_{\tilde{\theta}}(x^{(N)}) \xrightarrow{N \to \infty} \int dx \; p_{\theta_0}(x) \log p_{\tilde{\theta}}(x)$$

prawdziwa wartość

z lematu

$$\leq \int dx \; p_{\theta_0}(x) \log p_{\theta_0}(x)$$

czyli maximum $\frac{1}{N} \log p^{(N)}_{\tilde{\theta}}(x^{(N)})$ dla $N \to \infty$

będzie odpowiadało $\tilde{\theta}_{ML} = \theta_0$

- asymptotyczna nieobciążoność $\tilde{\theta}_{ML}$

• asymptotyczna efektywność

Twierdzenie o wart. średniej $\Bigg|$ $\Bigg\{$ dla ustalenia uwagi: $\theta_0 < \tilde{\theta}$

$$\frac{\partial \log p^{(N)}_\theta(x)}{\partial \theta}\Bigg|_{\theta=\tilde{\theta}} - \frac{\partial \log p^{(N)}_\theta(x)}{\partial \theta}\Bigg|_{\theta=\theta_0} = \frac{\partial^2 \log p^{(N)}(x)}{\partial \theta^2}\Bigg|$$

$$\frac{\frac{\partial \log p^{(N)}}{\partial \theta}\Big|_{\theta=\hat\theta} - \frac{\partial}{\partial \theta}\Big|_{\theta=\theta_0}}{\hat\theta - \theta_0} = \frac{\partial^2 \log p^{(N)}(x)}{\partial \theta^2}\Big|_{\theta=\bar\theta}$$

gdzie $\theta_0 < \bar\theta < \hat\theta$

Jeśli $\tilde\theta = \tilde\theta_{ML}$ to $\frac{\partial \log p_\theta^{(N)}(x)}{\partial \theta}\Big|_{\theta=\tilde\theta_{ML}} = 0$

(*) $$\frac{\partial \log p_\theta^{(N)}(x)}{\partial \theta}\Big|_{\theta=\theta_0} = \frac{\partial^2 \log p_\theta^{(N)}(x)}{\partial \theta^2}\Big|_{\theta=\bar\theta} (\theta_0 - \tilde\theta_{ML})$$

Weźmy:

$$\frac{\partial^2 \log p_\theta^{(N)}(x)}{\partial \theta^2}\Big|_{\theta=\bar\theta} = N \frac{1}{N} \sum_i \frac{\partial^2 \log p_\theta(x_i)}{\partial \theta^2}\Big|_{\theta=\bar\theta}$$

Wiemy że dla $N \to \infty$ $\tilde\theta_{ML} \to \theta$ a więc i $\bar\theta$

$$= N \frac{1}{N} \sum \frac{\partial^2 \log p_\theta(x_i)}{\partial \theta^2}\Big|_{\theta=\theta_0} \xrightarrow{N} N \left\langle \frac{\partial^2 \log p_\theta(x)}{\partial \theta^2}\Big|_{\theta=\theta_0} \right\rangle$$

$$= -NF(\theta_0)$$

Z drugiej

$$\frac{\partial \log p_\theta^{(N)}(x)}{\partial \theta}\Big|_{\theta=\theta_0} = \sqrt{N} \underbrace{\frac{1}{\sqrt{N}} \sum_i \frac{\partial \log p_\theta(x_i)}{\partial \theta}\Big|_{\theta=\theta_0}}_{\xi}$$

ξ - zmienna losowa będąca suma niezależnych zmiennych losowych

$\langle \xi \rangle = 0$

$$\langle \xi \rangle = 0$$ [illegible]

$$\langle \xi^2 \rangle = \left\langle \frac{1}{N} \left(\sum_i \frac{\partial \log p_\theta(x_i)}{\partial \theta} \right)^2 \right\rangle =$$

$$= \frac{1}{N} \sum_i \left\langle \left(\frac{\partial \log p_\theta(x_i)}{\partial x_i} \right)^2 \Big|_{\theta=\theta_0} \right\rangle = F(\theta_0)$$

$$\xi \overset{N\to\infty}{\sim} \mathcal{N}(0, F(\theta_0)^{-1})$$

Czyli:

$$\tilde{\theta}_{ML} - \theta_0 = \frac{\left.\frac{\partial \log p_\theta^{(N)}(x)}{\partial \theta}\right|_{\theta=\theta_0}}{\left.\frac{\partial^2 \log p_\theta^{(N)}(x)}{\partial \theta^2}\right|_{\theta=\bar{\theta}}} \overset{N\to\infty}{\sim} \mathcal{N}\left(0, \frac{N F(\theta_0)}{N^2 F^2(\theta_0)}\right)$$

$$\tilde{\theta}_{ML} \overset{N\to\infty}{\sim} \mathcal{N}\left(\theta_0, \frac{1}{NF}\right)$$

Uwaga: z góry nie wiadomo jak duży N trzeba by wziąć ale zazwyczaj coś < 1000 wystarcza.

Przykład Max-likelihood dla estymacji φ jeśli $p_0 = \sin^2\frac{\varphi}{2}$ $p_1 = \cos^2\frac{\varphi}{2}$ ale jakie N odstęp od C-R mniejszy niż 1%