Wyklad 5 - max-likelihood

2 listopada 2014
14:33

2. Estymacja największej wiarygodności (max-likelihood)

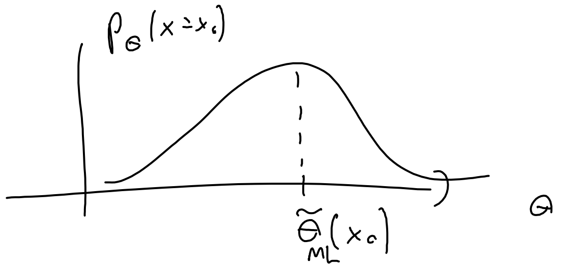
Co zrobić kiedy warunek wysycenia C-R nie spełniony i nie potrafimy podać optymalnego estymatora..

2.1 Estymator największej wiarygodności

$$p_\theta(x)$$

jeśli mamy 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 da ten sam punkt bo log monotoniczny

Przykład

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

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

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

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

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

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

Pamiętamy że warunek wysycenia to:

$$\frac{d}{d\theta} \log p_\theta(x) = \lambda(\theta) \left(f(x) - \theta \right)$$

efektywny estymator

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

Twierdzenie

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

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

x_i może reprezentować wiele danych

N powtórzonych eksperymentów

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

$$\tilde{\theta} \overset{N\to\infty}{\sim} \mathcal{N}\left(\theta, F^{(N)^{-1}}\right) \qquad F^{(N)} = N \cdot F$$

$$\Theta_{ML} \sim N(\Theta_0 + \ \ /\ ,\ F = N \cdot F$$

czyli wysyca C-R, i jest nieobciążony

<u>Lemat:</u>

Niech $p_1(x), p_2(x)$ - rozkłady prawdop.

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

$$\parallel$$

$$D(p_1 | p_2) \text{ - entropia względna}$$

<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 $\sum_x p_1(x) \log \frac{p_1(x)}{p_2(x)} \geq 0$ ∎

<u>Dowód twierdzenia</u>

Założenia techniczne: • istnieją 1-sza, 2-ga pochodne $\log p_\Theta(x)$

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

• asymptotyczna nieobciążoność

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

$$\frac{1}{N} \log p_{\tilde{\Theta}}^{(N)}(X^{(N)}) = \frac{1}{N} \sum_i \log p_{\tilde{\Theta}}(x_i)$$

z prawa wielkich liczb . to zbiega do średniej czy z prawa wielkich liczb, to zbiega do średniej czyli

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

prawdziwa wartość

z Lematu

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

Czyli maximum $\frac{1}{N}\log p_{\tilde\theta}^{(N)}(x^{(N)})$ dla $N\to\infty$

będzie odpowiadało $\tilde\theta_{ML}=\theta_0$

– asymptotyczna nieobciążoność $\tilde\theta_{ML}$

• asymptotyczna efektywność

Twierdzenie o wart. średniej | dla ustalenia uwagi $\theta_0<\tilde\theta$

$$\frac{\left.\frac{\partial\log p_\theta^{(N)}(x)}{\partial\theta}\right|_{\theta=\tilde\theta}-\left.\frac{\partial\log p_\theta^{(N)}(x)}{\partial\theta}\right|_{\theta=\theta_0}}{\tilde\theta-\theta_0}=\left.\frac{\partial^2\log p^{(N)}(x)}{\partial\theta^2}\right|_{\theta=\bar\theta}$$

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

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

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

Weźmy:

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

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 kolei

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

ξ - zmienna losowa będąca sumą niezależnych zmiennych losowych

$$\langle \xi \rangle = 0$$

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

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

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

Czyli:

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

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

Uwaga: z góry nie wiadomo jak duże N trzebaby wziąć ale zwykle coś < 100 wystarcza.

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