Wyklad 4 - cramer-rao

27 października 2014 09:17

II Klasyczna teoria estymacji

1. Estymacja parametru

(podejście klasyczne - nie Bayesowskie)

$$\theta \xrightarrow{p_\theta(x)} x$$

θ: nieznany (ale określony) parametr (y)

x: obserwowane zdarzenie (a)

$p_\theta(x)$ - rodzina rozkładów prawdopodobieństwa opisująca zmienną(e) losową x

Chcemy wywnioskować θ na podstawie x

$\tilde{\theta}(x)$ - estymator, ma nam dać informację o θ

Przykład

$$x = (x_1, \dots, x_N)$$

$$x_i = \theta + w_i$$

niezależne zmienne losowe

$$w_i \sim N(0, \sigma^2)$$

N: rozkład gaussowski, 0: średnia, σ^2: wariancja

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

$$p_\theta(x) = p_\theta(x_1) \cdot \ldots \cdot p_\theta(x_N)$$

$$p_\Theta(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\Theta)^2}{2\sigma^2}}$$

Mierzymy $(x_1, \dots, x_N)$

Jaki byłby „dobry" estymator Θ

- $\tilde{\Theta}(x) = x_1$
- $\tilde{\Theta}(x) = \frac{\sum_i x_i}{N}$

1.1 Optymalne estymatory

Nieobciążoność estymatora

$$\langle \tilde{\Theta} \rangle = \Theta \quad \text{tzn.} \quad \forall_\Theta \int \tilde{\Theta}(x)\, p_\Theta(x)\, dx = \Theta$$

średnio daje prawdziwą wartość

- $\tilde{\Theta}(x) = x_1$ $\qquad \langle \tilde{\Theta} \rangle = \langle x_1 \rangle = \Theta$ ok
- $\tilde{\Theta}(x) = \frac{\sum_i x_i}{N}$ $\qquad \langle \tilde{\Theta} \rangle = \frac{1}{N} \sum_i \langle x_i \rangle = \Theta$ OK

Optymalność estymatora

Wariancja estymatora

$$\Delta^2 \tilde{\Theta} = \langle (\tilde{\Theta} - \Theta)^2 \rangle = \int (\tilde{\Theta}(x) - \Theta)^2 p_\Theta(x)\, dx$$

Chcemy aby $\Delta^2 \tilde{\Theta}$ była jak najmniejsza

Uwaga: można napisać estymator, który będzie dobry dla pewnych wartości θ a gorszy dla innych.

trywialny przykład $\hat{\theta} = \theta_0$ nie zależne od danych – to jest przykład estymatora obciążonego

Zazwyczaj bardziej interesujący jest:

"Minimum variance unbiased estimator" (MVU)

Nieobciążony estymator z minimalną wariancją – minimalna wariancja dla wszystkich θ

Przykład

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

$$\tilde{\theta}(x_1, \dots, x_N) = \frac{1}{N}\sum_i x_i$$

$$s^2_{\tilde{\theta}} = \left\langle \left(\frac{1}{N}\sum_i x_i - \theta\right)^2 \right\rangle = \frac{1}{N^2}\left\langle (x_1 - \theta + x_2 - \theta + \dots x_N - \theta)^2 \right\rangle =$$

$$= \frac{1}{N^2} N \sigma^2 = \frac{\sigma^2}{N}$$

jest MVU (dowód później)

gdyby wziąć $\tilde{\theta}(x_1 \dots x_N) = x_1$ $\quad s^2_{\hat{\theta}} = \sigma^2$

1.2 Ograniczenie Cramera-Rao (CR)

Chcemy mieć ograniczenie na najlepszy możliwy estymator nieobciążony

wtedy jak znajdziemy estymator wysycający ograniczenie wiemy że jest optymalny

Intuicja: $p_\theta(x)$ im bardziej rozmyte $p_\theta(x)$

zmienia się z θ tym większa precyzja estymacji

w ograniczeniu pojawia się $\frac{d}{d\theta} p_\theta(x)$

Założenia

$$\langle\tilde{\Theta}\rangle = \int dx\, \hat{\theta}(x)\, p_\theta(x) = \theta$$ – nieobciążoność

w zasadzie będziemy potrzebować:

$$(*)\quad \frac{d\langle\tilde{\theta}\rangle}{d\theta} = \int dx\, \tilde{\theta}(x)\, \frac{dp_\theta(x)}{d\theta} = 1$$ – lokalna nieobciążoność

$$(**)\quad \int dx\, \frac{dp_\theta(x)}{d\theta} = 0$$ – warunek regularności

$\parallel$ jeśli można zamienić kolejność całkowania i różniczkowania

$$\frac{d}{d\theta}\int dx\, p_\theta(x) = \frac{d}{d\theta} 1 = 0$$ spełniony

Wyprowadzenie:

$$\underbrace{\int dx\, p_\theta(x)\left(\tilde{\theta}(x)-\theta\right)^2}_{\Delta^2\tilde{\theta}} \cdot \underbrace{\int dx\, \frac{1}{p_\theta(x)}\left(\frac{dp_\theta(x)}{d\theta}\right)^2}_{F}$$

$$= \int dx\left[\sqrt{p_\theta(x)}\left(\tilde{\theta}(x)-\theta\right)\right]^2 \cdot \int dx\left(\frac{1}{\sqrt{p_\theta(x)}}\frac{dp_\theta(x)}{d\theta}\right)^2$$

$$\overset{C-S}{\geq} \left(\int dx\, \left(\tilde{\theta}(x)-\theta\right) \frac{dp_\theta(x)}{dx}\right)^2 = \Bigg(\underbrace{\int dx\, \hat{\theta}(x)\frac{dp_\theta(x)}{d\theta}}_{\parallel\ (*)} - \theta\underbrace{\int dx\, \frac{dp_\theta(x)}{d\theta}}_{\parallel\ (**)}\Bigg)^2$$

$$\overset{(*)}{=} 1 \qquad \overset{(**)}{=} 0$$

$$\boxed{\Delta^2 \tilde{\theta} \geq \frac{1}{F}, \qquad F = \underbrace{\int dx \, \frac{1}{p_\theta(x)} \left(\frac{dp_\theta(x)}{d\theta} \right)^2}_{\text{informacja Fishera}}}$$

F w ogólności zależy od θ

<u>Uwaga:</u>

Możemy spotkać inne równoważne postacie F

$$F = \left\langle \left(\frac{d}{d\theta} \log p_\theta(x) \right)^2 \right\rangle = \int dx \, p_\theta(x) \, \frac{1}{p_\theta(x)^2} \left(\frac{dp_\theta(x)}{d\theta} \right)^2 = \int \frac{1}{p_\theta(x)} \left(\frac{dp_\theta(x)}{d\theta} \right)^2$$

$$F = - \left\langle \frac{d^2}{d\theta^2} \log p_\theta(x) \right\rangle$$

jawnie widać addytywność

Jeśli znajdziemy estymator nieobciążony dla którego $\Delta^2 \tilde{\theta} = \frac{1}{F}$ wiemy że jest optymalny. Mówimy wtedy $\tilde{\theta}$ jest „efficient".

<u>Addytywność F</u>

Jeśli $p_\theta^{(12)}(x_1, x_2) = p_\theta^{(1)}(x_1) \cdot p_\theta^{(2)}(x_2)$ to wtedy

$$F^{(12)} = F^{(1)} + F^{(2)}$$

Wniosek: N niezależnych realizacji zmiennej losowej X: $F^{(N)} = N \cdot F$

$$s^2 \tilde{\theta} \geq \frac{1}{N \cdot F}$$

Przykład

$$x_i \sim N(\theta, \sigma^2) \qquad i = 1, \dots, N$$

$$p_\theta(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}$$

$$F = \left\langle \left(\frac{d}{d\theta} \log p_\theta(x_i)\right)^2 \right\rangle = \int dx\, p_\theta(x) \left(\frac{x_i-\theta}{\sigma^2}\right)^2 = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Mając N realizacji:

$$s^2 \tilde{\theta} \geq \frac{1}{NF} = \frac{\sigma^2}{N}$$

Czyli tyle ile wychodzi z estymatora $\tilde{\theta} = \frac{\sum_i x_i}{N}$,

efficient estymator

Wysycanie nierówności CR

Potrzebne na wysycanie nierówności C-S wysycana iff:

$$\lambda(\theta) \sqrt{p_\theta(x)} \left(\hat{\theta}(x) - \theta\right) = \frac{1}{\sqrt{p_\theta(x)}} \frac{dp_\theta(x)}{d\theta}$$

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

Jeśli znajdziemy $\lambda(\theta)$, $\hat{\theta}(x)$ t.że powyższe równanie spełniane wysycamy C-R

Przykład

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

$$p_\theta(x) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}$$

$$\log p_\theta(x) = -\frac{N}{2}\log 2\pi\sigma^2 - \sum_{i=1}^{N} \frac{(x_i-\theta)^2}{2\sigma^2}$$

$$\frac{d}{d\theta}\log p_\theta(x) = \sum_i \frac{(x_i-\theta)}{\sigma^2} = \lambda(\theta)\left(\hat{\theta}(x)-\theta\right)$$

jeśli weźmiemy $\hat{\theta}(x) = \sum_i \frac{x_i}{N}$, $\lambda(\theta) = \frac{N}{\sigma^2}$

1.3 Wieloparametrowe ograniczenie C-R

$$p_{\vec{\theta}}(x) \qquad \vec{\theta} = (\theta_1, \dots, \theta_k) =: \theta$$

$$\vec{x} = (x_1, \dots, x_N) =: x$$

$$\boxed{C_{\tilde{\theta}} \geq \mathbb{F}^{-1}} \qquad \left\{ C_\theta - \mathbb{F}_\theta^{-1} \geq 0 \right.$$

macierz kowariancji — macierz Fishera

$$(C_{\tilde{\theta}})_{ij} = \int dx\, p_\theta(x)\left(\tilde{\theta}_i(x)-\theta_i\right)\left(\tilde{\theta}_j(x)-\theta_j\right)$$

gdzie $(\mathbb{F})_{ij} = \int dx \frac{1}{p_{\vec{\theta}}(x)} \frac{\partial p_\theta(x)}{\partial\theta_i}\frac{\partial p_\theta(x)}{\partial\theta_j}$ — Macierz Fishera (symetryczna dodatnio określona)

$$\left\{ \mathbb{F}_{ij} = \left\langle -\frac{\partial^2 \log p_\theta(x)}{\partial\theta_i\partial\theta_j}\right\rangle \right.$$

$$\} \ F_{ij} = \langle - \frac{\partial^2 \ldots}{\partial\theta_i \partial\theta_j} \rangle$$

W realizacji ..:

$$\Delta^2 \theta_i \geq (F^{-1})_{ii}$$

{ M. macierzy dodatnio określonej wyrazy diagonalne nieujemne

$$F - \text{macierz symetryczna}, \ F \geq 0$$

Wyprowadzenie:

$$w^T C w \cdot v^T F v = \int dx\, p(x)\, w_i (\tilde{\theta}_i - \theta_i)(\tilde{\theta}_j - \theta_j) w_j$$

$$\cdot \int dx \frac{1}{p(x)} v_i \frac{\partial p(x)}{\partial \theta_i} \frac{\partial p(x)}{\partial \theta_j} v_j \geq \left(\int dx\, w_i (\tilde{\theta}_i - \theta_i) \frac{\partial p(x)}{\partial \theta_i} v_i \right)^2$$

$$= (w^T v)^2 \quad \text{bierzemy } w = Fv$$

$$v^T F C F v \cdot v^T F v \geq (v^T F v)^2$$

$$v^T F C F v \geq v^T F v$$

$$F C F \geq F \implies C \geq F^{-1} \quad \blacksquare$$

Uwaga:

$$\Delta^2 \tilde{\theta}_i \geq (F^{-1})_{ii} \geq (F_{ii})^{-1}$$

niewspólne

$$\left\{ \begin{array}{l} 1 = e_i^T \sqrt{F} \sqrt{F^{-1}} e_i \leq e_i^T F e_i \ \ e_i^T F^{-1} e_i \\ (F^{-1})_{ii} \geq \frac{1}{F_{ii}} \end{array} \right.$$

Wysycone:

$$\vec{\nabla} \ln p_\theta(x) = F\left(\vec{g}(x) - \vec{\theta}\right)$$

1.5 Estymacja funkcji od wielu parametrów

$g(\vec{\theta})$ — g również może być wieloparametrowe ale dla uproszczenia niech skalarna

$$\Delta^2 \tilde{g} \geq \vec{\nabla} g \cdot F^{-1} \left(\vec{\nabla} g\right)^T$$