Wyklad 4 - cramer-rao

27 października 2014
09:17

II Klasyczna teoria estymacji

1. Estymacja parametru

(podejście klasyczne - nie Bayesowskie)

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

nieznany (ale określony) parametr(y)

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 informacje o θ

Przykład

$X = (x_1, \ldots, x_N)$

$x_i = \theta + w_i$ niezależne zmienne losowe

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

rozkład gaussowski, średnia, 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 $(x_1, \ldots, x_N)$

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

- $\tilde{\theta}(x) = x_1$ $\quad \langle \tilde{\theta} \rangle = \langle x_1 \rangle = \theta$ $\quad$ ok
- $\tilde{\theta}(x) = \frac{\sum_i x_i}{N}$ $\quad \langle \tilde{\theta} \rangle = \frac{1}{N} \sum_i \langle x_i \rangle = \theta$ $\quad$ 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 $\tilde{\theta} = \theta_0$ 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

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

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

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

$$= \frac{1}{N^2} N \sigma^2 = \frac{\sigma^2}{N}$$ jest MVU (dowód później)

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

Przykład: (MVU może nie istnieć)

$$x_1 = N(\theta, 1)$$

$$x_2 = \begin{cases} N(\theta, 1) & \theta \geq 0 \\ N(\theta, 2) & \theta < 0 \end{cases}$$

Dla $\theta \geq 0$ najlepszy estymator to $\tilde{\theta} = \frac{1}{2}(x_1 + x_2)$,

Dla $\theta < 0$ najlepszy estymator to $\tilde{\theta} = \frac{2}{3} x_1 + \frac{1}{3} x_2$,

Są nieobciążone ale nie ma jednego estymatora dającego minimalną wariancję dla wszystkich θ

Jak szukać optymalnych estymatorów?

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 rozkład $p_\theta(x)$ zmienia się z θ tym większa precyzja estymacji w ograniczeniu pojawia się $\frac{d}{d\theta} p_\theta(x)$

Założenia

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

w zasadzie będziemy potrzebować:

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

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

|| jeśli można zamienić kolejność całkowania i różniczkowania

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

Wyprowadzenie:

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

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

$$\overset{C-S}{\geq} \left|\int dx\,\left(\tilde\theta(x)-\theta\right)\frac{dp_\theta(x)}{dx}\right|^2 = \left|\underbrace{\int dx\,\hat\theta(x)\frac{dp_\theta(x)}{d\theta}}_{\overset{(*)}{=}\,1} - \theta\underbrace{\int dx\,\frac{dp_\theta(x)}{d\theta}}_{\overset{(**)}{=}\,0}\right|^2$$

$$\underbrace{\Delta^2\tilde\theta \geq \frac{1}{F}, \qquad F=\int dx\,\frac{1}{p_\theta(x)}\left(\frac{dp_\theta(x)}{d\theta}\right)^2 \quad \text{Informacja Fishera}}_{F \text{ w ogólności zależy od } \theta}$$

Uwaga:

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 $\Delta^2\hat\theta = \frac{1}{F}$ wiemy że jest optymalny. Mówimy wtedy $\tilde\theta$ jest „efficient".

Addytywność F

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

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

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

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

$$\delta^{-}\theta \geq N\cdot F$$

Przykład

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

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

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

Mając N realizacji:

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

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

efficient estymator

Wysycanie nierówności CR

Potrzeba na wysycenie nierówności C-S wysycam 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 $\lambda(\theta)$, $\hat{\theta}(x)$ takie że powyższe równanie spełnione 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 $\hat{\theta}(x) = \sum_i \frac{x_i}{N}$, $\lambda(\theta) = \frac{N}{\sigma^2}$

Ogólnie:

$$\log p_\theta(x) = \Lambda(\theta)\,\hat{\theta}(x) + a(\theta) + b(x), \quad a' = -\lambda\cdot\theta, \quad \Lambda' = \lambda$$

$$p_\theta(x) = e^{a(\theta)}\, e^{b(x)}\, e^{\Lambda(\theta)\hat{\theta}(x)}$$

Szczególny przypadek tzw. exponential family

$$p_\theta(x) = e^{a(\theta) + b(x) + c(\theta)d(x)}$$

gdzie żeby estymacja θ była efficient $\theta = -\frac{\frac{da}{d\theta}}{\frac{dc}{d\theta}}$

Jeśli $-\frac{\frac{da}{d\theta}}{\frac{dc}{d\theta}} \neq \theta$ wtedy nie istnieje estymator θ

który jest efficient ale istnieje dla funkcji od parametru $g(\theta) = -\frac{\frac{da}{d\theta}}{\frac{dc}{}}$ ⌂

⌂ estymacja θ^2 jeśli $x_i = N(\theta, \sigma^2)$

1.3 Wieloparametrowe ograniczenie C-R

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

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

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

macierz kowariancji $\qquad$ macierz Fishera

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

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)

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

W szczególności:

$$\Delta^2\theta_i \geqslant (\mathbb{F}^{-1})_{ii}$$ $\left\{\right.$ dla macierzy dodatnio określonych wyrazy diagonalne

$\Delta \omega_i \geq (F^{-1})_{ii}$ } wyrazy diagonalne macierzy

F – macierz symetryczna, $F \geq 0$

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

mniejsze

$$\begin{cases} 1 = e_i^T \sqrt{F}\sqrt{F^{-1}}\, e_i \leq e_i^T F e_i \; e_i^T F^{-1} e_i \\ (F^{-1})_{ii} \geq \frac{1}{F_{ii}} \end{cases}$$

Wysycenie:

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

1.4 Modele liniowe

$$\vec{x} = H \cdot \vec{\theta} + \vec{w}$$

$\vec{x} = [x_1, \dots, x_N]$

$\vec{\theta} = [\theta_1, \dots \theta_K]$

$\vec{w} = [w_1, \dots w_N]$

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

Jeśli $N > K$ i H macierz rzędu K to

$$\vec{x} \sim N(H\vec{\theta}, \sigma^2 \cdot \mathbb{1})$$

– wieloparametrowy gauss o średniej $H\vec{\theta}$ i macierzy kowariancji $\sigma^2 \mathbb{1}$

$$p(\vec{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{-\frac{1}{2\sigma^2}(\vec{x} - H\vec{\theta})^T(\vec{x} - H\vec{\theta})}$$

Warunek na wysycenie:

$$\vec{\nabla} \log p(\vec{x}) = -\frac{1}{2\sigma^2} \vec{\nabla}_\theta \Big(\vec{x}^T x + \vec{\theta}^T H^T H \theta \underbrace{- \vec{\theta}^T H^T \vec{x} - x^T H}_{-2x^T H \vec{\theta}}$$

$$= -\frac{1}{2\sigma^2}\left(-2H^T\vec{x} + 2H^T H\theta\right) = \frac{1}{\sigma^2}\left(H^T\vec{x} - H^T H\vec{\theta}\right)$$

$$\begin{cases} \vec{\nabla}_\theta (a^T\theta) = a & \left\{ \frac{\partial}{\partial\theta_j}\sum_i a_i\theta_i = a_j \right. \\ \vec{\nabla}_\theta (\theta^T A\theta) = 2A\theta & \left\{ \frac{\partial}{\partial\theta_k}\left(\sum_{ij}\theta_i A_{ij}\theta_j = \sum_i \theta_i A_{ik} + \sum_k A_{ki}\theta_i = 2A\right.\right. \end{cases}$$

macierz sym

Jeśli $H^T H$ odwracalne to:

$$\vec{\nabla}\log p(x) = \frac{1}{\sigma^2}(H^T H)\left(\underbrace{(H^T H)^{-1}H^T\vec{x}}_{\tilde{\vec{\theta}}} - \vec{\theta}\right)$$

estymator $\tilde{\vec{\theta}} = (H^T H)^{-1}H^T\vec{x}$

Macierz Fishera: $F = \frac{H^T H}{\sigma^2}$

- Dla ogólniejszego szumu gaussowskiego

$\vec{w} \sim N(0, C)$

macierz kowariancji

$F = H^T C^{-1} H \qquad \tilde{\vec{\theta}} = (H^T C^{-1} H)^{-1} H^T C^{-1}\vec{x}$

1.5 Estymacja funkcji od wielu parametrów

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

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