

Statystyka Matematyczna

Anna Janicka

wykład VI, 04.04.2016

WŁASNOŚCI ESTYMATORÓW, CZ. II

Plan na dzisiaj

1. Informacja Fishera
2. Nierówność informacyjna
3. Efektywność estymatora
4. Asymptotyczne własności estymatorów
 - zgodność
 - *asymptotyczna normalność*
 - *asymptotyczna efektywność*



Informacja Fishera

Jeśli model statystyczny z obs. $X_1, X_2, \dots, X_n$ i p-stwem f_θ spełnia warunki regularności, tzn.:

1. Θ jest przedziałem otwartym 1-wymiarowym.
2. Nośnik rozkładu $\{x: f_\theta(x) > 0\}$ nie zależy od θ .
3. Istnieje pochodna $\frac{df_\theta}{d\theta}$.

to można zdefiniować **Informację Fishera** zawartą w obserwacjach $X_1, X_2, \dots, X_n$:

$$I_n(\theta) = E_\theta \left(\frac{d}{d\theta} \ln f_\theta(X_1, X_2, \dots, X_n) \right)^2$$

nie zakładamy tu niezależności $X_1, X_2, \dots, X_n$



Informacja Fishera – co oznacza

- Miara tego, jak wiele może powiedzieć próba wielkości n (uśredniona) o wartości nieznanego parametru θ .
- Np. jeśli funkcja gęstości wokół θ jest płaska, to informacja zawarta w (jednej) obserwacji nie będzie pozwalała różnicować naszych przewidywań co do θ . Jeśli jednak funkcja gęstości wokół θ nie jest płaska, to info o wynikach wnosi wiele.



Informacja Fishera – cd.

Wzory dla różnych przypadków:

□ jeśli rozkład ciągły

$$I_n(\theta) = \int_X \left(\frac{\frac{df_\theta(x)}{d\theta}}{f_\theta(x)} \right)^2 f_\theta(x) dx$$

□ jeśli rozkład dyskretny

$$I_n(\theta) = \sum_{x \in X} \left(\frac{\frac{dP_\theta(x)}{d\theta}}{P_\theta(x)} \right)^2 P_\theta(x)$$

□ jeśli f_θ dwukrotnie różniczkowalna

$$I_n(\theta) = -E_\theta \left(\frac{d^2}{d\theta^2} \ln f_\theta(X_1, X_2, \dots, X_n) \right)$$



Informacja Fishera – cd. (2)

- Jeśli próba składa się z niezależnych zmiennych losowych o identycznych rozkładach, to wówczas z uwagi na multiplikatywność prawdopodobieństwa oraz własności logarytmu

$$I_n(\theta) = nI_1(\theta)$$

$I_1(\theta)$ jest informacją Fishera zawartą w pojedynczej obserwacji



Informacja Fishera – przykłady

□ Rozkład wykładniczy $\exp(\lambda)$

$$I_1(\lambda) = \dots = \frac{1}{\lambda^2}$$

□ Rozkład Poissona $\text{Poiss}(\theta)$

$$I_1(\theta) = \dots = \frac{1}{\theta}$$



Nierówność Informacyjna (Craméra-Rao)

Niech $X=(X_1, X_2, \dots, X_n)$ będą obserwacjami o łącznej gęstości $f_\theta(x)$, gdzie $\theta \in \Theta \subseteq \mathbb{R}$, oraz:

- $T(X)$ jest statystyką o skończonej wartości oczekiwanej $E_\theta T(X)=g(\theta)$
- Informacja Fishera jest dobrze określona, $I_n(\theta) \in (0, \infty)$
- Wszystkie gęstości f_θ mają ten sam nośnik
- Można zamieniać kolejność różniczkowania ($d/d\theta$) oraz całkowania $\int \dots dx$.

Wówczas, dla dowolnego θ :

$$\text{Var}_\theta T(X) \geq \frac{(g'(\theta))^2}{I_n(\theta)}$$



Nierówność informacyjna – implikacje

- W szczególności, funkcja ryzyka nieobciążonego estymatora (=wariancja) nie może być mniejsza od ustalonej funkcji n i θ .
- Jeśli ryzyko estymatora nieobciążonego jest równe dolnemu ograniczeniu nierówności, to jest to estymator ENMW.
- Jeśli $\hat{\theta}(X)$ jest nieobciążonym estymatorem θ , to

$$\text{Var}_{\theta} \hat{\theta}(X) \geq \frac{1}{I_n(\theta)}$$



Nierówność informacyjna – przykłady zastosowań

- W modelu Poissona, $\hat{\theta} = \bar{X}$ jest ENMW(θ) $Var_{\theta}(\bar{X}) = \frac{\theta}{n}$
- W modelu wykładniczym, $\bar{X}$ jest ENMW($1/\lambda$) $Var_{\lambda}(\bar{X}) = \frac{1}{n\lambda^2}$
- Niestety, nierówność Craméra-Rao nie zawsze jest optymalna: w modelu wykładniczym, $\hat{\lambda} = 1/\bar{X}$ jest obciążonym estymatorem λ . Nieobciążonym estymatorem jest $\tilde{\lambda} = \frac{n-1}{n\bar{X}}$, który jest też ENMW(λ), chociaż jego wariancja jest *większa*, niż by to wynikało z ograniczenia Craméra-Rao.



Efektywność

Efektywność nieobciążonego estymatora $\hat{g}(X)$ wielkości $g(\theta)$ to:

$$\text{ef}(\hat{g}) = \frac{(g'(\theta))^2}{\text{Var}_{\theta}(\hat{g}) \cdot I_n(\theta)}$$

Efektywność względna nieobciążonych estymatorów $\hat{g}_1(X)$ oraz $\hat{g}_2(X)$:

$$\text{ef}(\hat{g}_1, \hat{g}_2) = \frac{\text{Var}_{\theta}(\hat{g}_2)}{\text{Var}_{\theta}(\hat{g}_1)} = \frac{\text{ef}(\hat{g}_1)}{\text{ef}(\hat{g}_2)}$$



Efektywność a nierówność informacyjna

- Jeśli spełniona jest nierówność informacyjna, to $ef(\hat{g}) \leq 1$ dla każdego estymatora nieobciążonego.
- Jeśli $\hat{g} = ENMW(g)$, to możliwe jest, że $ef(\hat{g}) = 1$, ale możliwe też, że $ef(\hat{g}) < 1$.
- Jeśli $ef(\hat{g}) = 1$, to estymator jest **efektywny**.

efektywność w sensie Craméra-Rao



Efektywność – przykłady

- ❑ W modelu Poissona, $\hat{\theta} = \bar{X}$ jest efektywny.
- ❑ W modelu wykładniczym, $\bar{X}$ jest efektywnym estymatorem $1/\lambda$.
- ❑ W modelu wykładniczym, $\hat{\lambda} = \frac{n-1}{n\bar{X}}$ *nie jest* efektywnym estymatorem λ , chociaż jest ENMW(λ).
- ❑ Branie części próbki nie jest efektywne.
- ❑ W modelu jednostajnym $U(0, \theta)$ dla $ENW(\theta)$ „wychodzi” $ef > 1$ (bo nie są spełnione założenia nierówności informacyjnej)



Własności asymptotyczne estymatorów

- Twierdzenia graniczne opisujące zachowanie/ własności estymatorów gdy $n \rightarrow \infty$
- W praktyce: informacja, jak w *przybliżeniu* zachowują się estymatory dla dużych próbek
- Problem: zwykle brak odpowiedzi na pytanie, *jaka* próbka jest odpowiednio duża (żeby przybliżenie było odpowiednio dobre)



Zgodność

Niech $X_1, X_2, \dots, X_n, \dots$ będzie próbą IID, tzn. niezależnych zmiennych losowych o identycznych rozkładach. Niech $\hat{g}(X_1, X_2, \dots, X_n)$ będzie ciągiem estymatorów wielkości $g(\theta)$. Estymator $\hat{g}$ jest **zgodny**, jeśli dla każdego $\theta \in \Theta$, dla każdego $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P_{\theta}(|\hat{g}(X_1, X_2, \dots, X_n) - g(\theta)| \leq \varepsilon) = 1$$

(tzn. estymator $\hat{g}$ jest zbieżny do $g(\theta)$ według prawdopodobieństwa)



Mocna zgodność

Niech $X_1, X_2, \dots, X_n, \dots$ będzie próbą IID, tzn. niezależnych zmiennych losowych o identycznych rozkładach. Niech $\hat{g}(X_1, X_2, \dots, X_n)$ będzie ciągiem estymatorów wielkości $g(\theta)$. Estymator $\hat{g}$ jest **mocno zgodny**, jeśli dla każdego $\theta \in \Theta$:

$$P_{\theta} \left(\lim_{n \rightarrow \infty} \hat{g}(X_1, X_2, \dots, X_n) = g(\theta) \right) = 1$$

(tzn. estymator $\hat{g}$ jest zbieżny do $g(\theta)$ prawie na pewno)



Zgodność – uwaga

Z tw. Gliwienki-Cantelliego mamy, iż dystrybuanty empiryczne są zbieżne p.n. do dystrybuanty teoretycznej. A zatem od wszystkich sensownych estymatorów należy oczekiwać (mocnej) zgodności.

Zgodność = minimalne sensowne żądanie w stosunku do estymatora.



Zgodność – jak badać?

□ Z definicji; można wówczas korzystać np. z nierówności Czebyszewa-Bienaymé i t.p.:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{E(X - E(X))^2}{\varepsilon^2}$$

Pamiętając, że ryzyko estymatora to

$$R(\theta, \hat{g}) = E_{\theta}(\hat{g}(X) - g(\theta))^2$$

dostajemy np. warunek wystarczający zgodności:

$$\lim_{n \rightarrow \infty} R(\theta, \hat{g}) = 0$$

□ Z PWL



Zgodność – przykłady

- Dla dowolnej rodziny rozkładów mających wartość oczekiwaną: średnia z próbki $\bar{X}_n$ jest zgodnym estymatorem wartości oczekiwanej $\mu(\theta) = E_\theta(X_1)$. Zbieżność wynika z MPWL.
- Dla dowolnej rodziny rozkładów mających wariancję: $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ i $\hat{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ są zgodnymi estymatorami wariancji $\sigma^2(\theta) = \text{Var}_\theta(X_1)$. Zbieżność wynika z MPWL



Zgodność – przykłady/własności

- Estymator może być nieobciążony i nie być zgodny; np. $T_n(X_1, X_2, \dots, X_n) = X_1$ jako estymator $\mu(\theta) = E_\theta(X_1)$.
- Estymator może być obciążony i być zgodny; np. obciążony estymator wariancji j.w., lub np. dowolny zgodny estymator nieobciążony $+ 1/n$.



Asymptotyczna normalność

Estymator $\hat{g}(X_1, X_2, \dots, X_n)$ wielkości $g(\theta)$ jest **asymptotycznie normalny**, jeśli dla każdego $\theta \in \Theta$ istnieje $\sigma^2(\theta)$ takie, że gdy $n \rightarrow \infty$

$$\sqrt{n}(\hat{g}(X_1, X_2, \dots, X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2(\theta))$$

Zbieżność wg. rozkładu, tzn. dla dowolnego a

$$\lim_{n \rightarrow \infty} P_{\theta} \left(\frac{\sqrt{n}}{\sigma(\theta)} (\hat{g}(X_1, X_2, \dots, X_n) - g(\theta)) \leq a \right) = \Phi(a)$$

lub równoważnie, rozkład $\hat{g}(X_1, X_2, \dots, X_n)$ jest dla dużych n zbliżony do rozkładu $N(g(\theta), \frac{\sigma^2}{n})$



Asymptotyczna normalność – co to znaczy

- Estymator asymptotycznie normalny jest zgodny (niekoniecznie mocno zgodny).
- Zawiera warunek *podobny* do nieobciążoności – wartość oczekiwana rozkładu asymptotycznego jest równa $g(\theta)$ (ale sam estymator *nie musi* być nieobciążony).
- Podobnie dla tzw. **wariancji asymptotycznej** definiowanej jako $\sigma^2(\theta)/n$ lub $\sigma^2(\theta)$ – jest to wariancja rozkładu asymptotycznego



Asymptotyczna normalność – czym nie jest

- Zazwyczaj dla estymatora asymptotycznie normalnego zachodzi:

$$E_{\theta} \hat{g}(X_1, X_2, \dots, X_n) \xrightarrow{n \rightarrow \infty} g(\theta)$$
$$n \operatorname{var} \hat{g}(X_1, X_2, \dots, X_n) \xrightarrow{n \rightarrow \infty} \sigma^2(\theta)$$

ale nie musi tak być, bo zbieżność wg rozkładu nie pociąga za sobą zbieżności wartości oczekiwanych ani wariancji



Asymptotyczna normalność – przykład

- Niech $X_1, X_2, \dots, X_n, \dots$ będzie próbą IID o średniej μ i wariancji σ^2 . Wówczas dla średniej z próby mamy z CTG

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$$

Asymptotyczna wariancja, $\sigma^2(\theta)/n$, jest tu tożsama z wariancją estymatora.



Asymptotyczna normalność – jak liczyć

W wielu przypadkach przydatny jest lemat:

Metoda Delta. Jeśli dla ciągu zmiennych losowych T_n mamy $\sqrt{n}(T_n - \mu) \xrightarrow{D} N(0, \sigma^2)$ gdy $n \rightarrow \infty$ oraz $h: \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją różniczkowalną w punkcie μ t.ż. $h'(\mu) \neq 0$, to

$$\sqrt{n}(h(T_n) - h(\mu)) \xrightarrow{D} N(0, \sigma^2 (h'(\mu))^2)$$

μ, σ^2 są funkcjami parametru θ

stosujemy zwykle wtedy, kiedy estymatory są funkcjami

statystyk T_n , których zbieżność łatwo wywnioskować z CTG

Asymptotyczna normalność – przykłady cd.

- W modelu wykładniczym: $ENW(\lambda) = \frac{1}{X}$
Z CTG mamy

$$\sqrt{n}(\bar{X} - \frac{1}{\lambda}) \xrightarrow{D} N(0, \frac{1}{\lambda^2})$$

więc z Lematu Delta mamy dla $h(t)=1/t$:

$$\sqrt{n}(\frac{1}{\bar{X}} - \lambda) \xrightarrow{D} N(0, \frac{1}{\lambda^2} \cdot (-\frac{1}{(1/\lambda)^2})^2)$$

a więc $\frac{1}{\bar{X}}$ jest asymptotycznie normalnym
(co za tym idzie: zgodnym) estymatorem λ .



Asymptotyczna efektywność

Dla asymptotycznie normalnego estymatora $\hat{g}(X_1, X_2, \dots, X_n)$ wielkości $g(\theta)$ możemy określić **asymptotyczną efektywność** jako

$$\text{as.ef}(\hat{g}) = \frac{(g'(\theta))^2 n}{\sigma^2(\theta) \cdot I_n(\theta)},$$

gdzie $\sigma^2(\theta)/n$ jest wariancją asymptotyczną, tj. mamy gdy $n \rightarrow \infty$

$$\sqrt{n}(\hat{g}(X_1, X_2, \dots, X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2(\theta))$$



modyfikacja „zwykłej” efektywności do przypadku granicznego, z wariancją asymptotyczną zamiast zwykłej

$$\text{as.ef}(\hat{g}) = \frac{(g'(\theta))^2}{\sigma^2(\theta) \cdot I_1(\theta)}$$

Asymptotyczna efektywność względna

Asymptotyczna efektywność względna dla
asymptotycznie normalnych estymatorów
 $\hat{g}_1(X)$ i $\hat{g}_2(X)$

$$\text{as.ef}(\hat{g}_1, \hat{g}_2) = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)} = \frac{\text{as.ef}(\hat{g}_1)}{\text{as.ef}(\hat{g}_2)}$$

Uwaga. Estymator mniej (asymptotycznie) efektywny może mieć inne cechy, które sprawiają, że ma przewagę nad estymatorem bardziej (asymptotycznie) efektywnym.



Asymptotyczna efektywność względna – Przykład 1

Estymacja prawdopodobieństwa braku szkód
w modelu Poissona: $\text{Poiss}(\theta)$ (typowy model
aktuarialny dla liczby szkód w poszczególnych
latach dla polisy, albo dla grupy polis)

$g(\theta) = e^{-\theta} = P(X_i = 0)$ – pr-stwo braku szkód

$$\hat{g}_1(X) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i=0\}} \quad \hat{g}_2(X) = e^{-\bar{X}} = ENW(g(\theta))$$

$$\sqrt{n}(\hat{g}_1 - g(\theta)) \xrightarrow{D} N(0, e^{-\theta}(1 - e^{-\theta})) \quad \text{z CTG dla schematu Bernoulliego}$$

$$\sqrt{n}(\hat{g}_2 - g(\theta)) \xrightarrow{D} N(0, \theta \cdot (-e^{-\theta})^2) \quad \text{z Lematu Delta}$$

$$\rightarrow \text{as.ef}(\hat{g}_1, \hat{g}_2) = \frac{\theta e^{-2\theta}}{e^{-\theta}(1 - e^{-\theta})} = \frac{\theta}{e^{\theta} - 1} < 1$$



Asymptotyczna efektywność względna – Przykłady 2: lepsza średnia czy mediana z próbki?

To zależy od rozkładu!

a) model normalny $N(\mu, \sigma^2)$:

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2) \quad \text{as.ef}(\hat{m}, \bar{X}) = \frac{2}{\pi} < 1$$

$$\sqrt{n}(\hat{m} - \mu) \xrightarrow{D} N(0, \frac{\pi\sigma^2}{2})$$

b) model Laplace'a $\text{Lapl}(\mu, \lambda)$

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \frac{2}{\lambda^2}) \quad \text{as.ef}(\hat{m}, \bar{X}) = 2 > 1$$

$$\sqrt{n}(\hat{m} - \mu) \xrightarrow{D} N(0, \frac{1}{\lambda^2})$$

c) niektóre rozkłady nie mają średniej...

Twierdzenie: Dla próbki z rozkładu ciągłego o gęstości $f(x)$, mediana próbkowa jest estymatorem asymptotycznie normalnym dla mediany m (o ile gęstość jest ciągła i $\neq 0$ w punkcie m):

$$\sqrt{n}(\hat{m} - m) \xrightarrow{D} N(0, \frac{1}{4(f(m))^2})$$





WARSAW UNIVERSITY
Faculty of Economic Sciences