

Statystyka Matematyczna

Anna Janicka

wykład V, 21.03.2016

WŁASNOŚCI ESTYMATORÓW, CZ. I

Plan na dzisiaj

1. Podstawowe własności estymatorów:

- obciążenie estymatora
- estymatory nieobciążone

2. Mierniki jakości: porównywanie estymatorów

- ryzyko estymatora
- nieporównywalne estymatory
- estymator nieobciążony o minimalnej wariancji

3. Informacja Fishera



Własności estymatorów

- Czy nie robimy za dużych błędów? Czy estymujemy to, co chcemy?
- $\hat{\theta}$ ma przybliżać θ .
Ogólniej: $\hat{g}(X)$ ma przybliżać $g(\theta)$.
- Chcemy: mały błąd przybliżenia. Ale
 - błąd jest zmienną losową (dane są losowe)
→ możemy badać wartość oczekiwaną błędu
 - błąd zależy od nieznanego θ .
→ *trudno...*



Obciążenie estymatora

Jeśli statystyka $\hat{\theta}(X)$ jest estymatorem θ :
obciążenie estymatora to

$$b(\theta) = E_{\theta}(\hat{\theta}(X) - \theta) = E_{\theta}\hat{\theta}(X) - \theta$$

Jeśli statystyka $\hat{g}(X)$ jest estymatorem $g(\theta)$:
obciążenie estymatora to

$$b(\theta) = E_{\theta}(\hat{g}(X) - g(\theta)) = E_{\theta}\hat{g}(X) - g(\theta)$$

Estymator $\hat{\theta}(X)$ / $\hat{g}(X)$ jest **nieobciążony**,
jeśli $b(\theta) = 0$ dla $\forall \theta \in \Theta$



Przypomnienie z wykładu 3: model normalny

Model normalny: $X_1, X_2, \dots, X_n$ są próbką z rozkładu $N(\mu, \sigma^2)$. μ, σ nieznane.

Tw. W modelu normalnym, statystyki $\bar{X}$ i S^2 są niezależnymi zmiennymi losowymi, t. że

$$\bar{X} \sim N(\mu, \sigma^2/n)$$
$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

W szczególności:

$$E_{\mu, \sigma} \bar{X} = \mu, \text{ oraz } \text{Var}_{\mu, \sigma} \bar{X} = \sigma^2/n$$

$$E_{\mu, \sigma} S^2 = \sigma^2, \text{ oraz } \text{Var}_{\mu, \sigma} S^2 = 2\sigma^4/(n-1)$$



Obciążenie estymatora – Przykład 1

W modelu ~~normalnym~~: dowolnym, o nieznanym μ :

□ Estymator $\hat{\mu} = \bar{X}$ jest nieobciążonym estymatorem μ :

$$E_{\mu,\sigma} \hat{\mu}(X) = E_{\mu,\sigma} \bar{X} = E_{\mu,\sigma} \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} n\mu = \mu$$

□ Podobnie, $\hat{\mu}_1 = X_1$ jest nieobciążonym estymatorem μ : $E_{\mu,\sigma} \hat{\mu}_1(X) = E_{\mu,\sigma} X_1 = \mu$

□ Estymator $\hat{\mu}_2 = 5$ jest obciążony:

$$E_{\mu,\sigma} \hat{\mu}_2(X) = E_{\mu,\sigma} 5 = 5 \neq \mu \quad \text{np. dla } \mu = 2$$

obciążenie: $b(\mu) = 5 - \mu$



Obciążenie estymatora – Przykład 1 cd.

□ Estymator $\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ jest *obciążonym* estymatorem σ^2 :

$$\begin{aligned} E_{\mu, \sigma} \hat{S}^2(X) &= E_{\mu, \sigma} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} E_{\mu, \sigma} (\sum X_i^2 - n\bar{X}^2) \\ &= \frac{1}{n} (n(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)) = \sigma^2 - \sigma^2/n \neq \sigma^2 \end{aligned}$$

□ Zaś estymator $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ jest *nieobciążonym* estymatorem σ^2 :

$$\begin{aligned} E_{\mu, \sigma} S^2(X) &= E_{\mu, \sigma} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} E_{\mu, \sigma} (\sum X_i^2 - n\bar{X}^2) \\ &= \frac{1}{n-1} (n(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)) = \frac{1}{n-1} (\sigma^2(n-1)) = \sigma^2 \end{aligned}$$



Obciążenie estymatora – Przykład 1 cd (2)

Obciążenie estymatora $\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

to

$$b(\sigma) = -\frac{\sigma^2}{n}$$

gdy $n \rightarrow \infty$, obciążenie dąży do 0, a więc ten estymator też jest OK dla dużych prób



Asymptotyczna nieobciążoność

- Estymator $\hat{g}(X)$ parametru $g(\theta)$ jest **asymptotycznie nieobciążony**, jeśli

$$\forall \theta \in \Theta : \lim_{n \rightarrow \infty} b(\theta) = 0$$



Jak porównywać estymatory?

- ❑ Chcemy minimalizować błąd estymatora; estymator, który mniej się myli jest *lepszy*.
- ❑ Błąd może być na + albo na -, stąd standardowo patrzy się na kwadrat błędu (średni kwadrat odchylenia estymatora od wartości estymowanej)



Ryzyko estymatora (ryzyko kwadratowe)

Jeśli statystyka $\hat{\theta}(X)$ jest estymatorem θ :
ryzyko estymatora $\hat{\theta}(X)$ to funkcja

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

Jeśli statystyka $\hat{g}(X)$ jest estymatorem $g(\theta)$:
ryzyko estymatora $\hat{g}(X)$ to funkcja

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

My rozważać będziemy tylko ryzyko kwadratowe (błąd średniokwadratowy). Możliwe też inne (np. z modułem)



Własności ryzyka kwadratowego

Mamy:

$$R(\theta, \hat{g}) = b^2(\theta) + \text{Var}(\hat{g})$$

A zatem dla estymatorów nieobciążonych
ryzyko to wariancja estymatora



Ryzyko estymatora – Przykład 1

Model: $X_1, X_2, \dots, X_n$ są próbką z rozkładu o średniej μ , wariancji σ^2 . μ, σ nieznane.

□ Ryzyko estymatora $\hat{\mu} = \bar{X}$ (nieobc.):

$$R(\mu, \sigma, \bar{X}) = E_{\mu, \sigma} (\bar{X} - \mu)^2 = \text{Var}_{\mu, \sigma} \bar{X} = \frac{\sigma^2}{n}$$

□ Ryzyko estymatora $\hat{\mu}_1 = X_1$ (nieobc.):

$$R(\mu, \sigma, X_1) = E_{\mu, \sigma} (X_1 - \mu)^2 = \text{Var}_{\mu, \sigma} X_1 = \sigma^2$$

□ Ryzyko estymatora $\hat{\mu}_2 = 5$ (obc.):

$$R(\mu, \sigma, 5) = E_{\mu, \sigma} (5 - \mu)^2 = (5 - \mu)^2$$



Ryzyko estymatora – Przykład 1 cd.

□ Ryzyko estymatora $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

$$R(\mu, \sigma, S^2) = E_{\mu, \sigma} (S^2 - \sigma^2)^2 = \text{Var}_{\mu, \sigma} S^2 = \frac{2\sigma^4}{n-1}$$

□ Ryzyko estymatora $\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

$$\begin{aligned} R(\mu, \sigma, \hat{S}^2) &= E_{\mu, \sigma} (\hat{S}^2 - \sigma^2)^2 = b^2(\sigma) + \text{Var}_{\mu, \sigma} \hat{S}^2 \\ &= \frac{\sigma^4}{n^2} + \frac{(n-1)^2}{n^2} \frac{2\sigma^4}{n-1} = \frac{2n-1}{n^2} \sigma^4 \end{aligned}$$

w modelu dowolnym:

analogicznie, tylko inne wzory

$$R(\mu, \sigma, S^2) > R(\mu, \sigma, \hat{S}^2)$$

Ryzyko i obciążenie estymatora – Przykład 2.

Model Poissona: $X_1, X_2, \dots, X_n$ są próbką z rozkładu Poissona z nieznanym parametrem θ .

$$\hat{\theta}_{NW} = \dots = \bar{X}$$

$$b(\theta) = 0$$

$$R(\theta, \bar{X}) = \text{Var}_{\theta} \bar{X} = \text{Var}_{\theta} \frac{1}{n} \sum_{i=1}^n X_i = \frac{\theta}{n}$$



Porównywanie estymatorów

Estymator $\hat{g}_1(X)$ jest **lepszy** od $\hat{g}_2(X)$, jeśli

$$\forall \theta \in \Theta \quad R(\theta, \hat{g}_1) \leq R(\theta, \hat{g}_2)$$

oraz

$$\exists \theta \in \Theta \quad R(\theta, \hat{g}_1) < R(\theta, \hat{g}_2)$$

a zatem: jeden estymator jest lepszy od drugiego tylko jeśli cały wykres funkcji ryzyka jednego leży pod wykresem drugiego; jeśli wykresy się krzyżują, to estymatory są **nieporównywalne**



Porównywanie estymatorów – cd.

Z uwagi na to, że bardzo wiele estymatorów jest nieporównywalnych, porównywanie *byle czego* nie ma sensu; trzeba się ograniczyć do pewnej klasy estymatorów.

Jeśli porównujemy dwa estymatory nieobciążone, to lepszy będzie ten, który ma mniejszą wariancję.



Porównanie estymatorów – Przykład 1.

W modelu dowolnym:

- Spośród estymatorów $\hat{\mu} = \bar{X}$ oraz $\hat{\mu}_1 = X_1$ lepszy jest (dla $n > 1$) estymator $\hat{\mu}$
- Estymatory $\hat{\mu} = \bar{X}$ oraz $\hat{\mu}_2 = 5$ są nieporównywalne, tak samo jak estymatory $\hat{\mu}_1 = X_1$ oraz $\hat{\mu}_2 = 5$
- Spośród estymatorów S^2 oraz $\hat{S}^2$ lepszy jest $\hat{S}^2$



Estymator nieobciążony o minimalnej wariancji

Ograniczamy poszukiwania/porównania do estymatorów nieobciążonych. W tej klasie często da się znaleźć estymator najlepszy:

Statystyka $g^*(X)$ jest **estymatorem nieobciążonym o minimalnej wariancji** dla $g(\theta)$, jeśli

- $g^*(X)$ jest nieobciążonym estymatorem $g(\theta)$,
- dla każdego nieobciążonego estymatora $\hat{g}(X)$

$$Var_{\theta} g^*(X) \leq Var_{\theta} \hat{g}(X) \text{ dla } \theta \in \Theta$$



Jak stwierdzić, że estymator ma minimalną wariancję?

- W ogólności nie da się dowolnie zmniejszać wariancji estymatorów nieobciążonych – dla wielu modeli probabilistycznych istnieje pewna granica zmniejszania wariancji. Zależy ona od konkretnego rozkładu i od wielkości próby.



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 – przykład

□ Rozkład wykładniczy $\exp(\lambda)$

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





WARSAW UNIVERSITY
Faculty of Economic Sciences