Metody probabilistyczne
Pominięte dowody

13. Statystyka

16.01.2020

Twierdzenie (nierówność Craméra-Rao): *Niech $\widehat{\theta}$ będzie nieobciążonym estymatorem parametru θ wyznaczonym z próby $X_1, X_2, \ldots, X_n$. Zdefiniujmy funkcję informacji Fishera $I(\theta)$ jako:*

- *Dla rozkładów dyskretnych:*
$$I(\theta) \;=\; E\left(\left(\frac{\partial \ln p(X)}{\partial \theta}\right)^2\right),$$
gdzie $p(x)$ jest prawdopodobieństwem przyjęcia wartości x przez zmienną X.
- *Dla rozkładów ciągłych:*
$$I(\theta) \;=\; E\left(\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right),$$
gdzie $f(x)$ jest gęstością zmiennej X.

Wtedy:
$$D^2(\widehat{\theta}) \;\geqslant\; \frac{1}{nI(\theta)}$$

Dowód: Udowodnimy tę własność tylko dla rozkładów dyskretnych. Dowód dla rozkładów ciągłych jest identyczny, trzeba tylko zamienić $p(x)$ na $f(x)$ i wszystkie sumy na całki.

Niech $\mathcal{X}$ będzie zbiorem wartości przyjmowanych przez zmienną losową X. Zauważmy, że $p(x)$ będzie zależało w jakiś sposób od parametru θ, np. w rozkładzie Bernoulliego gdy $\theta = p$, $p(1) = p$, a $p(0) = 1-p$.

Dla dowolnego i, zdefiniujmy zmienną losową:
$$Y_i = \frac{\partial \ln p(X_i)}{\partial \theta}.$$

Zauważmy, że Y_i jest funkcją zmiennej X_i, tzn. $Y_i = g(X_i)$, gdzie $g(x) = \frac{\partial \ln p(x)}{\partial \theta}$. Zgodnie z zasadą różniczkowania funkcji złożonej, mamy:
$$\frac{\partial \ln p(x)}{\partial \theta} \;=\; \frac{1}{p(x)} \frac{\partial p(x)}{\partial \theta}$$

co daje (zauważając, że rozkład zmiennej X_i dany jest przez $p(x)$):
$$EY_i \;=\; \sum_{x \in \mathcal{X}} p(x) g(x) \;=\; \sum_{x \in \mathcal{X}} p(x) \frac{\partial \ln p(x)}{\partial \theta} \;=\; \sum_{x \in \mathcal{X}} \cancel{p(x)} \frac{1}{\cancel{p(x)}} \frac{\partial p(x)}{\partial \theta} \;=\; \sum_{x \in \mathcal{X}} \frac{\partial p(x)}{\partial \theta}.$$

Ale:
$$\sum_{x \in \mathcal{X}} \frac{\partial p(x)}{\partial \theta} \;\overset{(*)}{=}\; \frac{\partial}{\partial \theta} \Bigg(\underbrace{\sum_{x \in \mathcal{X}} p(x)}_{=1} \Bigg) \;=\; \frac{\partial}{\partial \theta} 1 \;=\; 0,$$

gdzie równość w $(*)$ wynika z faktu, że pochodna sumy to suma pochodnych, oraz użyliśmy warunku normalizacji prawdopodobieństwa $\sum_{x \in \mathcal{X}} p(x) = 1$. Tym samym pokazaliśmy $EY_i = 0$.

Zapiszmy teraz zmienne losowe łącznie jako wektor $\boldsymbol{X} = (X_1, \ldots, X_n)$, a ich rozkład łączny jako $p(\boldsymbol{x})$ dla $\boldsymbol{x} = (x_1, \ldots, x_n)$. Z niezależności zmiennych losowych:

$$p(\boldsymbol{x}) = p(x_1)p(x_2) \cdot \ldots \cdot p(x_n).$$

Zdefiniujemy zmienną $Y = h(\boldsymbol{X})$ będącą funkcją wektora $\boldsymbol{X}$ jako:

$$Y = \frac{\partial \ln p(\boldsymbol{X})}{\partial \theta}$$

Z niezależności wynika, że:

$$\frac{\partial \ln p(\boldsymbol{x})}{\partial \theta} = \frac{\partial \ln (p(x_1) \cdot \ldots \cdot p(x_n))}{\partial \theta} = \frac{\partial (\ln p(x_1) + \ldots + \ln p(x_n))}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln p(x_i)}{\partial \theta},$$

co oznacza, ze $Y = \sum_{i=1}^n Y_i$. Ponieważ udowodniliśmy już, że $EY_i = 0$, mamy więc również:

$$EY = 0 \tag{1}$$

Zgodnie z nierównością Cauchy'ego-Schwarza (udowodnioną na wykładzie o wielowymiarowych zmiennych losowych) dla dowolnych zmiennych losowych X i Y zachodzi:

$$C(X, Y)^2 \leqslant D^2(X)D^2(Y).$$

Wykorzystamy tą nierówność podstawiając $X = \widehat{\theta}$:

$$C(\widehat{\theta}, Y)^2 \leqslant D^2(\widehat{\theta})D^2(Y), \qquad \text{czyli } D^2(\widehat{\theta}) \geqslant \frac{C(\widehat{\theta}, Y)^2}{D^2(Y)}. \tag{2}$$

Wykażemy teraz, że $C(\widehat{\theta}, Y)^2 = 1$ oraz $D^2(Y) = nI(\theta)$, co zakończy dowód nierówności Craméra-Rao. Zaczniemy od licznika. Ze wzoru skróconego mnożenia dla kowariancji $C(X, Y) = E(XY) - (EX)(EY)$ mamy:

$$C(\widehat{\theta}, Y) = E\big(\widehat{\theta}Y\big) - (E\widehat{\theta})(\underbrace{EY}_{=0}) = E\big(\widehat{\theta}Y\big),$$

gdzie wykorzystaliśmy (1). Ale:

$$E\big(\widehat{\theta}Y\big) = \sum_{\boldsymbol{x} \in \mathcal{X}^n} \widehat{\theta}(\boldsymbol{x}) \frac{\partial \ln p(\boldsymbol{x})}{\partial \theta} p(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{X}^n} \widehat{\theta}(\boldsymbol{x}) \frac{1}{p(\boldsymbol{x})} \frac{\partial p(\boldsymbol{x})}{\partial \theta} p(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{X}^n} \widehat{\theta}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial \theta}.$$

Zauważmy jednak, że $\widehat{\theta}$ nie zależy w żaden sposób od θ (jest wyłącznie funkcją realizacji próby $\boldsymbol{x}$), stąd możemy potraktować $\widehat{\theta}$ jako stałą ze względu na θ i wystawić pochodną przed sumę:

$$\sum_{\boldsymbol{x} \in \mathcal{X}^n} \widehat{\theta}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial \theta} = \frac{\partial}{\partial \theta} \left(\sum_{\boldsymbol{x} \in \mathcal{X}^n} \widehat{\theta}(\boldsymbol{x}) p(\boldsymbol{x}) \right) = \frac{\partial}{\partial \theta} E\widehat{\theta} = \frac{\partial}{\partial \theta} \theta = 1,$$

gdzie użyliśmy faktu, że estymator $\widehat{\theta}$ jest nieobciążony, tzn. $E\widehat{\theta} = \theta$. Tym samym pokazaliśmy, że $C(\widehat{\theta}, Y) = 1$, a więc również $C(\widehat{\theta}, Y)^2 = 1$, czyli licznik w (2) jest równy 1.

Teraz zajmiemy się mianownikiem w (2). Mamy:

$$D^2(Y) = D^2(Y_1 + \ldots + Y_n) = \sum_{i=1}^{n} D^2(Y_i), \tag{3}$$

ponieważ wszystkie zmienne $Y_i = g(X_i)$ są niezależne (co wynika z niezależności zmiennych $X_1, \ldots, X_n$). Na koniec zauważmy, że ze wzoru skróconego mnożenia dla wariancji

$$D^2(Y_i) = E(Y_i^2) - (\underbrace{EY_i}_{=0})^2 = E\left(\left(\frac{\partial \ln p(X_i)}{\partial \theta} \right)^2 \right) = I(\theta),$$

a więc z (3) wynika, że:

$$D^2(Y) = nI(\theta),$$

czyli mianownik w (2) jest równy $nI(\theta)$. To kończy dowód.