Dostateczność

Przykład wstępny. Model zmiennej losowej X

$$(\{0,1\},\{D\left(\theta\right),\ \theta \in [0,1]\})$$

Celem jest wnioskowanie o parametrze θ na podstawie próby $X_1, X_2, \ldots, X_5$.
Rozważmy trzy próby:

Próba 1: $X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0$

Próba 2: $X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1$

Próba 3: $X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1$

Prawdopodobieństwa realizacji poszczególnych prób:

$$P_\theta\{\text{Próba 1}\} = \theta^3(1-\theta)^2$$

$$P_\theta\{\text{Próba 2}\} = \theta^3(1-\theta)^2$$

$$P_\theta\{\text{Próba 3}\} = \theta^3(1-\theta)^2$$

Istotna informacja: $X_1+X_2+X_3+X_4+X_5 (\stackrel{\text{ozn}}{=\!=} T)$

$$P_\theta\{X_1=x_1,X_2=x_2,X_3=x_3,X_4=x_4,X_5=x_5|T=t\}$$
$$= \begin{cases} 1/\binom{5}{t}, & \text{jeżeli } x_1+x_2+x_3+x_4+x_5=t, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Statystyka T jest dostateczna dla wnioskowania o parametrze θ.

. . . .

Definicja 5.1 Statystyka T jest dostateczna, jeżeli rozkład warunkowy $P_\theta\{\cdot|T=t\}$ nie jest zależny od θ dla wszystkich $\theta \in \Theta$. $\square$

Definicja 5.2 Statystyka T jest zupełna, jeżeli dla wszystkich $\theta \in \Theta$ zachodzi $E_\theta h(T) = 0$, to $h \equiv 0$. $\square$

Twierdzenie 5.1 *(O faktoryzacji) Statystyka T jest dostateczna dla rodziny rozkładów $\{P_\theta : \theta \in \Theta\}$ wtedy i tylko wtedy, gdy gęstość $p_\theta(x)$ może być przedstawiona w postaci*

$$p_\theta(x) = g_\theta\{T(x)\}h(x),$$

gdzie funkcja h nie zależy od θ. $\blacksquare$

Rodziny wykładnicze

Definicja 5.3 Rodzina rozkładów prawdopodobieństwa $\{P_\theta : \theta \in \Theta\}$ nazywa się rodziną wykładniczą, jeżeli każdy rozkład P_θ ma gęstość p_θ o postaci

$$p_\theta(x) = \exp\left\{\sum_{j=1}^{k} c_j(\theta)T_j(x) - b(\theta)\right\} \cdot h(x),$$

gdzie $T_1(x), T_2(x), \ldots, T_k(x)$ są liniowo niezależnymi funkcjami oraz $\{(c_1(\theta), c_2(\theta), \ldots, c_k(\theta)) : \theta \in \Theta\}$ jest pewnym k wymiarowym zbiorem w $\mathbb{R}^k$. $\square$

Twierdzenie 5.2 *Jeżeli* $\{P_\theta : \theta \in \Theta\}$ *oraz* $\Theta \subset \mathbb{R}^k$ *jest rodziną wykładniczą rozkładów z gęstościami*

$$p_\theta(x) = \exp\left\{\sum_{j=1}^{k} c_j(\theta)T_j(x) - b(\theta)\right\} \cdot h(x),$$

to $(T_1(x), T_2(x), \ldots, T_k(x))$ *jest* (k*–wymiarową) minimalną statystyką dostateczną zupełną.* $\blacksquare$