Rachunek ppb i statystyka

Estymatory największej wiarygodności (MLE estimators)

Zakładamy że dane są niezależne obserwacje zmiennej losowej X o tym samym rozkładzie. Funkcja gęstości zmiennej X ma postać $f(x;\theta)$, gdzie θ to parametr/parametry rozkładu. Można też uważać, że dane są niezależne zmienne $X_1, \ldots, X_n$ o tym samym rozkładzie. Prawdopodobieństwo zdarzenia można zatem zapisać jako funkcję wiarygodności L względem zmiennej θ

$$L(x;\theta) = P\left(X_1 = x_1, \ldots, X_n = x_n\right) = \prod_{k=1}^{n} f(x_k;\theta). \tag{1}$$

Zakładamy, że obserwowane wartości są typowe (najbardziej prawdopodobne). Chcemy zatem aby funkcja wiarygodności osiągała maksimum w pewnym punkcie $\hat{\theta}$. Oważ wyznaczona wartość $\hat{\theta}$ nazywana jest *estymatorem największej wiarygodności parametru* θ.
Uwaga: Bardzo często szukamy maksimum funkcji $\ln L(x;\theta)$, wyłącznie z powodów obliczeniowych; logarytm (naturalny) jako funkcja rosnąca daje tę samą odpowiedź.

Przykład:
P1: Rozważmy n niezależnych obserwacji z rozkładu $\mathrm{Exp}(\lambda)$. Prawdopodobieństwo zdarzenia $P\left(X_1 = x_1, \ldots X_n = x_n\right)$ jest równe – z racji niezależności – $L(\lambda) = \lambda^n \exp\left(-\sum_{k=1}^{n} x_i\right)$. Chcemy znaleźć wartość λ taką iż funkcja (λ) osiąga maksimum. [a]

Maksima funkcji $L(\lambda)$ oraz $\ln L(\lambda)$ znajdują się w tym samym punkcie $\hat{\lambda}$. Obliczając – i przyrównując do zera, – pochodną $\frac{\partial \ln L}{\partial \lambda}$ otrzymujemy równanie

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} + \left(\sum_{k=1}^{n} x_i\right) = \frac{n}{\lambda} - n \cdot \bar{x} = 0,$$

skąd wynika, że $\hat{\lambda} = {}^1\!/_{\bar{x}}$. Druga pochodna funkcji $\ln L(\lambda)$ równa $\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{n}{\lambda^2}$ jest ≤ 0 w każdym punkcie λ, czyli również dla wyznaczonej wcześniej wartości $\hat{\lambda}$, co dowodzi iż znaleźliśmy maksimum funkcji wiarygodności $\equiv$ estymator MLE $\hat{\lambda}$ parametru λ.

[a] Intuicja: to, co obserwujemy, jest najbardziej prawdopodobne.

Przykład:
P2: Rozważamy n niezależnych obserwacji z rozkładu $\mathrm{B}(n,p)$. Funkcja wiarygodności ma teraz postać

$$\begin{aligned} L(p) &= P\left(X_1 = x_1, \ldots, X_k = x_k\right) = \prod_{i=1}^{k} P\left(X_i = x_i\right) = \\ &= p^{\sum x_i}\ (1-p)^{nk - \sum x_i} \prod_{i=1}^{k} \binom{n}{x_i} = p^{k\bar{x}}\ (1-p)^{nk - k\bar{x}} \prod_{i=1}^{k} \binom{n}{x_i}. \end{aligned} \tag{2}$$

Logarytm funkcji wiarygodności ma zatem postać

$$\ln L(p) = \sum_{i=1}^{k} \ln\binom{n}{x_i} + k\bar{x}\ln p + k(n-\bar{x})\ln(1-p),$$

a jego pochodna to

$$\frac{\partial \ln L}{\partial p} = \frac{k\bar{x}}{p} - \frac{k(n-\bar{x})}{1-p} = 0.$$

Rozwiązując powyższe równanie otrzymujemy dla estymatora MLE wyrażenie $\hat{p} = \frac{\bar{x}}{n}$. Druga pochodna funkcji wiarygodności to

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{k\bar{x}}{p^2} - \frac{k(n-\bar{x})}{(1-p)^2} < 0,$$

dla $0 < \hat{p} < n$ zatem znaleźliśmy maksimum funkcji $L(p)$.

Jeżeli $\hat{p} = 0$, to $x_1 = \ldots = x_k = 0$. Funkcja wiarygodności (2) ma postać $L(p) = (1-p)^{nk}$ i osiąga maksimum dla $p = 0$. Podobnie, jeżeli $\hat{p} = n$, to $x_1 = \ldots = x_k = n$. Funkcja wiarygodności (2) ma w tym wypadku postać $L(p) = p^{nk}$ i osiąga maksimum dla $p = 1$.

Witold Karczewski