

**12.**

**Nieparametryczna estymacja dystrybuanty zmiennych losowych**

# 1. Sformułowanie problemu

$F(x)$  – dystrybuanta zmiennej losowej  $X$

$$F(x) = P(X \leq x)$$

$x_1, x_2, \dots, x_N$  – wylosowane (niezależnie) liczby, obserwacje (tzw. realizacje zmiennej losowej  $X$ )

Mając daną (ustaloną, konkretną) wartość  $x$  oraz dysponując ciągiem  $N$  losowych obserwacji  $\{x_i\}_{i=1}^N$  należy wyestymować wartość  $F(x)$ .

## Przykład

Wyznaczyć odsetek osób dorosłych, których wzrost nie przekracza 150 cm.

Losujemy z populację  $N = 1000$  osobową próbę i każdej z osób mierzymy wzrost  $x_i$  [cm].

Proponowane oszacowanie poszukiwanego prawdopodobieństwa

$$\widehat{F}_N(x = 150\text{cm}) = \frac{\text{liczba osób o wzroście} \leq 150 \text{ cm}}{1000 \text{ (tj. liczba wszystkich badanych osób)}}$$

## Formalna postać estymatora

$$\widehat{F}_N(x) = \frac{\#\{x_i : x_i \leq x\}}{N}$$

Powyższa propozycja funkcjonuje dla dowolnego  $x$ , określa całą losową funkcję.

## Wartość oczekiwana

Zdefiniujmy funkcję przynależności

$$I(x_i) = \begin{cases} 1, & \text{gdy } x_i \leq x \\ 0, & \text{gdy } x_i > x \end{cases}$$

Estymator przyjmuje postać

$$\widehat{F}_N(x) = \frac{\sum_{i=1}^N I(x_i)}{N}$$

a zatem

$$E\widehat{F}_N(x) = \frac{1}{N}E \sum_{i=1}^N I(x_i) = \frac{1}{N} \sum_{i=1}^N EI(x_i)$$

a ponieważ dla każdego  $i = 1, 2, \dots, N$

$$EI(x_i) = 1 \cdot P(x_i \leq x) + 0 \cdot P(x_i > x) = F(x)$$

to

$$E\widehat{F}_N(x) = \frac{1}{N} \cdot NF(x) = F(x) \quad (\text{estymator nieobciążony})$$

## Wariancja

$$\text{var}\widehat{F}_N(x) = \text{var}\frac{\sum_{i=1}^N I(x_i)}{N} = \frac{1}{N^2}\text{var}\sum_{i=1}^N I(x_i) = \frac{1}{N^2}\sum_{i=1}^N \text{var}I(x_i)$$

$\text{var}I(x_i)$  – taka sama dla wszystkich  $i$

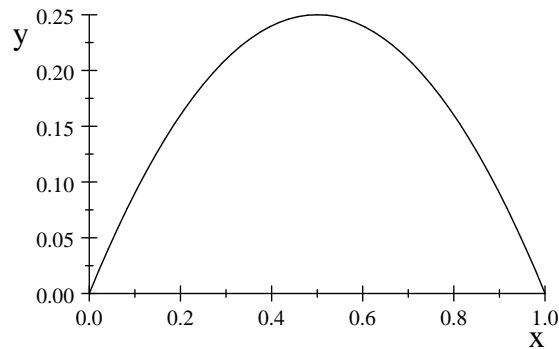
$$\text{var}I(x_i) = ?$$

$$EI(x_i) = F(x)$$

$$\begin{aligned}\text{var}I(x_i) &= (1 - F(x))^2 F(x) + (0 - F(x))^2 (1 - F(x)) = \\ &= (1 - F(x))F(x)[1 - F(x) + F(x)] = F(x)(1 - F(x))\end{aligned}$$

$$\text{var}\widehat{F}_N(x) = \frac{1}{N^2}NF(x)(1 - F(x))$$

$$\text{var}\widehat{F}_N(x) = \frac{F(x)(1 - F(x))}{N}$$



## 2. Nieparametryczna estymacja funkcji gęstości prawdopodobieństwa

$$F'(x) = f(x)$$

$$\hat{f}_N(x) = \frac{\hat{F}_N(x + \frac{h}{2}) - \hat{F}_N(x - \frac{h}{2})}{h}$$

$$\hat{f}_N(x) = \frac{1}{Nh} \sum_{i=1}^N I_{(x-\frac{h}{2}, x+\frac{h}{2})}\{x_i\}$$

Estymator jądrowy

$$\hat{f}_N(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x_i - x}{h}\right)$$

$K()$  – funkcja jądra

$$K(x) \rightarrow 0 \text{ gdy } |x| \rightarrow \infty$$

$$\int K(x)dx < \infty$$

$$K(x) = K(-x)$$

## Warunki zbieżności

$$\lim_{N \rightarrow \infty} h(N) = 0$$

$$\lim_{N \rightarrow \infty} Nh(N) = \infty$$