1. Sformułowanie problemu

$F(x)$ – dystrybuanta zmiennej losowej X

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

$x_1, x_2, ..., x_N$ – 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

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

Formalna postać estymatora

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

Powyższa propozycja funkjonuje dla dowolnego x, określa całą losową funkcję.

Wartość oczekiwana

Zdefiniujmy funkcję przynależności

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

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 $i = 1, 2, ..., N$

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

to

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

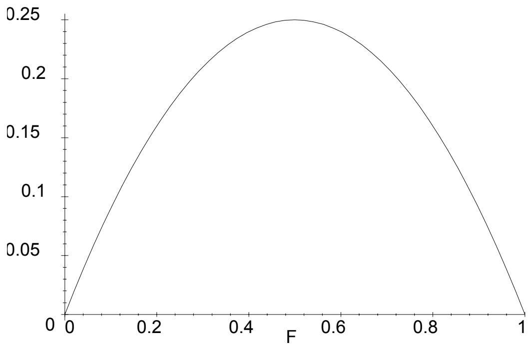
Wariancja

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

$$\begin{aligned} & varI(x_i) \text{ – taka sama dla wszystkich } i \\ varI(x_i) &= ? \\ EI(x_i) &= F(x) \\ varI(x_i) &= (1 - F(x))^2 F(x) + (0 - F(x))^2 (1 - F(x)) = \\ &= (1 - F(x))F(x)[1 - F(x) + F(x)] = F(x)(1 - F(x)) \end{aligned}$$

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

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



2. Nieparametryczna estymacja funkcji gęstości prawdopodobieństwa

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

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

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

Estymator jądrowy

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

$$
\begin{aligned}
&K() \text{ – funkcja jądra} \\
K(x) &\rightarrow 0 \text{ gdy } |x| \rightarrow \infty \\
\int K(x)dx &< \infty \\
K(x) &= K(-x)
\end{aligned}
$$

Warunki zbieżności

$$
\begin{aligned}
\lim_{N \to \infty} h(N) &= 0 \\
\lim_{N \to \infty} Nh(N) &= 0
\end{aligned}
$$