

Podstawy ekonometrii

dr inż. Zbigniew TARAPATA

Wykład 2

**Podstawowe rozkłady prawdopodobieństwa.
Statystyki i ich rozkłady.**

1. Zmienna losowa.

Niech dana będzie przestrzeń probabilistyczna $(\Omega, \mathcal{G}, P)$. Funkcję X określoną na przestrzeni zdarzeń elementarnych Ω o wartościach rzeczywistych oraz taką, że dla każdego $t \in \mathbb{R}$ zbiór

$$\{\omega \in \Omega: X(\omega) < t\}$$

jest zdarzeniem losowym (czyli należy do $\mathcal{G}$), będziemy nazywać zmienną losową. Wartość zmiennej losowej dla konkretnego zdarzenia elementarnego nazywa się *realizacją* tej zmiennej losowej.

Z definicji zmiennej losowej X , zbiór $\{\omega \in \Omega: X(\omega) < t\}$ dla dowolnego $t \in \mathbb{R}$ jest zdarzeniem w przestrzeni probabilistycznej $(\Omega, \mathcal{G}, P)$, wobec czego możemy obliczyć jego prawdopodobieństwo $P(\{\omega \in \Omega: X(\omega) < t\})$. Funkcję $F_X(t) = P(\{\omega \in \Omega: X(\omega) < t\})$ nazywamy dystrybuantą zmiennej losowej X . Będziemy opuszczać w opisie dystrybuanty symbol X pisząc zamiast $F_X(t)$ wyrażenie uproszczone $F(t)$.

Mówimy, że zmienna losowa X jest typu dyskretnego (skokowego), jeśli dla pewnego, skończonego lub nie, zbioru $W = \{x_1, x_2, \dots\}$ zachodzi

$$p_i = P(\{\omega \in \Omega: X(\omega) = x_i\}) > 0 \quad \text{dla } x_i \in W$$

oraz

$$\sum p_i = 1$$

gdzie sumowanie dotyczy wszystkich p_i .

Mówimy, że zmienna losowa X jest typu ciągłego, jeśli istnieje nieujemna funkcja f , określona i całkowalna do jedynki na całej osi taka, że dla każdego przedziału $[x_1, x_2]$ zachodzi

$$P(\{\omega : x_1 \leq X(\omega) \leq x_2\}) = \int_{x_1}^{x_2} f(x) dx$$

Funkcję f nazywamy gęstością rozkładu prawdopodobieństwa zmiennej losowej X . Łatwo pokazać, że

$$F(t) = \int_{-\infty}^t f(x) dx$$

2. Charakterystyki liczbowe zmiennej losowej.

Wartością oczekiwaną (przeciętną, średnią) zmiennej losowej X typu skokowego o zbiorze punktów skokowych $W=\{x_1, x_2, \dots\}$ i skokach $p_i = P(X=x_i)$ nazywamy liczbę EX określoną wzorem

$$EX = \sum_{x_i \in W} x_i p_i$$

pod warunkiem, że w przypadku przeliczalnej liczby punktów skokowych szereg po prawej stronie jest bezwzględnie zbieżny, to znaczy

$$\sum_{x_i \in W} |x_i| p_i < \infty$$

Wartością oczekiwaną (przeciętną, średnią) zmiennej losowej X typu ciągłego nazywamy liczbę EX określoną wzorem

$$EX = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} [1 - F(x)] dx$$

Jeśli zmienna losowa $Y = g(X)$, to

$$EY = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

Kwantylem rzędu p ($0 < p < 1$) zmiennej losowej X typu ciągłego o dystrybuancie F i gęstości f nazywamy każdą liczbę x_p spełniającą którykolwiek z następujących równoważnych warunków

- $F(x_p) = p$
- $P(X < x_p) = p$

- $\int_{-\infty}^{x_p} f(x)dx = p$

Wariancją zmiennej losowej X o wartości oczekiwanej EX nazywamy liczbę $\text{Var}X$ (D^2X , σ^2 , σ_x^2) określoną wzorem $\text{Var}X = E(X - EX)^2$

Z równości $E(X + c) = EX + c$ wynika, że

$$\text{Var}(X + c) = E[X + c - E(X + c)]^2 = E[X - EX]^2 = \text{Var}X$$

Wariancja jest zatem miara rozrzutu zmiennej losowej wokół jej wartości oczekiwanej.

Odchylenie standardowe zmiennej losowej X oznaczamy je przez DX (σ , σ_x) określone jest formułą

$$DX = \sqrt{\text{Var}X}$$

Zauważmy, że $\text{Var}X = EX^2 - (EX)^2$

Uogólnienie:

- momentem rzędu r względem stałej c zmiennej losowej X nazywamy liczbę $E(X - c)^r$;
- momenty względem stałej $c=0$ nazywamy momentami zwykłymi;
- momenty względem stałej $c=EX$ nazywamy momentami centralnymi.

3.Podstawowe rozkłady prawdopodobieństwa.

a. Zmienne losowe skokowe

- **Rozkład dwumianowy**

Niech zmienna losowa X typu skokowego przyjmuje wartości i ze zbioru $W=\{1,2,...,n\}$ z prawdopodobieństwem

$$p_i = \binom{n}{i} p^i (1-p)^{n-i}$$

gdzie $0 < p < 1$ jest ustalonym parametrem. Mówimy wtedy, że zmienna losowa ma rozkład dwumianowy (Bernoulliego)
 $EX=np$, $VarX=np(1-p)$

- **Rozkład geometryczny**

Niech zmienna losowa X typu skokowego przyjmuje wartości i ze zbioru $W=\{1,2,...\}$ z prawdopodobieństwem

$$p_i = p(1-p)^{i-1}$$

gdzie $0 < p < 1$ jest ustalonym parametrem. Mówimy wtedy, że zmienna losowa ma rozkład geometryczny
 $EX=1/p$, $VarX=(1-p)/p^2$

- **Rozkład Poissona**

Mówimy, że zmienna losowa X typu skokowego ma rozkład Poissona z parametrem $\lambda > 0$ jeśli przyjmuje ona wartości ze zbioru $W=\{0,1,2,...\}$ z prawdopodobieństwem

$$p_i = e^{-\lambda} \frac{\lambda^i}{i!}$$

mamy dla niej $EX=\lambda$, $VarX=\lambda$

UWAGA!

Jeżeli $n \geq 30$ i p jest małe, to

$$\binom{n}{i} p^i (1-p)^{n-i} \approx \frac{\lambda^i}{i!} e^{-\lambda}, \text{ gdzie } \lambda = n \cdot p$$

b. Zmienne losowe ciągłe

- **Rozkład normalny**

Niech X będzie zmienną losową typu ciągłego o gęstości prawdopodobieństwa postaci

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

gdzie μ oraz σ są parametrami.

Mówimy wtedy, że X ma rozkład normalny (gaussowski)

$N(\mu, \sigma)$. Prosta $x=\mu$ jest osią symetrii wykresu funkcji gęstości f .

Łatwo pokazać, że

$EX=\mu$ oraz $VarX=\sigma^2$ i $DX=\sigma$

- **Rozkład wykładniczy**

Mówimy, że zmienna losowa X typu ciągłego ma rozkład wykładniczy, jeżeli dystrybuanta ma postać

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

natomiast gęstość f wyraża się wzorem

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

wtedy $EX=\lambda^{-1}$ oraz $VarX=\lambda^{-2}$ i $DX=\lambda^{-1}$

- **Rozkład t-Studenta**

mówimy, że zmienna losowa X typu ciągłego ma rozkład t-studenta o n -stopniach swobody, jeśli jej gęstość wyraża się wzorem

$$f(x) = \frac{\Gamma([n+1]/2)}{\sqrt{n\pi} \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

wykres gęstości zmiennej losowej t-Studenta jest symetryczny względem osi rzędnych i gdy $n \rightarrow \infty$ gęstość f dąży do gęstości rozkładu normalnego $N(0,1)$, co wykorzystuje się w praktyce dla $n \geq 30$ do aproksymacji rozkładu t-Studenta rozkładem normalnym

- **Rozkład χ^2 (chi-kwadrat)**

Mówimy, że zmienna losowa X typu ciągłego ma rozkład χ^2 o n -stopniach swobody, jeśli jej gęstość wyraża się wzorem

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Dystrybuanta zmiennej losowej $(X - n) / \sqrt{2n}$, gdzie X ma rozkład χ^2 o n -stopniach swobody, dąży przy $n \rightarrow \infty$ do dystrybuanty rozkładu normalnego $N(0,1)$. Wykorzystuje się to w praktyce dla $n \geq 50$ do aproksymacji kwantyli rozkładu χ^2 kwantylami rozkładu normalnego.