

Charakterystyki liczbowe zmiennych losowych: wartość oczekiwana i wariancja

dr Mariusz Grządziel

Wykłady 3 i 4; 11,18 marca 2014

Wartość oczekiwana zmiennej losowej dyskretnej

Definicja 1. Dla zmiennej losowej dyskretnej X wartość oczekiwana, jeśli istnieje, jest liczbą określoną wzorem

$$EX = \sum_i x_i p_i, \quad (1)$$

w którym sumowanie obejmuje wszystkie wartości zmiennej X ; jeśli suma (1) jest szeregiem nieskończonym, zakładamy, że jest on bezwzględnie zbieżny, to znaczy

$$\sum_i |x_i| p_i = g,$$

gdzie g jest pewną liczbą rzeczywistą.

Uwaga Wartość oczekiwana zmiennej losowej nazywana jest w literaturze także wartością średnią zmiennej losowej.

Wartość oczekiwana może być interpretowana jako "środek ciężkości" układu punktów materialnych $x_1, x_2, \dots$ o wagach $p_1, p_2, \dots$.

Własności wartości oczekiwanej

Łatwo widać, że wartość oczekiwana zmiennej losowej $Y = aX + b$ jest równa

$$E(aX + b) = aEX + b \text{ (wartość oczekiwana ma własność liniowości).}$$

Jeśli wartości oczekiwane zmiennych losowych X_1 i X_2 istnieją i są równe, odpowiednio, μ_1 i μ_2 , to

$$E(X_1 + X_2) = \mu_1 + \mu_2.$$

Wariancja zmiennej losowej

Definicja 2. Wariancję zmiennej losowej X określamy wzorem

$$\text{Var} X = E(X - \mu)^2,$$

gdzie $\mu = EX$.

Wariancja jest równa wartości oczekiwanej kwadratu odchylenia wartości zmiennej losowej od swojej wartości przeciętnej.

Uwaga: W definicji tej nie zakładamy, że zmienna losowa X jest dyskretna. Zakładamy natomiast istnienie wartości oczekiwanej $E(X - \mu)^2$.

Dla $a > 0$ mamy:

$$Var(aX + b) = a^2 Var(X).$$

Dla zmiennej dyskretnej X wariancja jest równa

$$Var X = \sum_i (x_i - \mu)^2 p_i.$$

Własności wariancji rozkładu

Widzimy, że wariancja jest tym większa, im większa jest średnia odległość punktów x_i od środka ciężkości μ - wartości oczekiwanej.

Jeśli wszystkie wartości x_i są sobie równe, wtedy wariancja jest równa zeru.

Odchylenie standardowe

Odchylenie standardowe zmiennej losowej X , oznaczane przez DX , definiujemy jako pierwiastek kwadratowy wariancji X . Odchylenie standardowe zmiennej losowej X często jest też oznaczane grecką literą σ .

Można pokazać, że dla $a \geq 0$:

$$D(aX + b) = aDX. \quad (2)$$

Wartość oczekiwana i wariancja zmiennej o rozkładzie dwumianowym

Niech X będzie zmienną losową o rozkładzie dwumianowym z parametrami $n \in \mathbb{N}$ i $p \in (0, 1)$. Można pokazać, że:

$$E(X) = np, \quad Var(X) = np(1 - p).$$

Wartość oczekiwana i wariancja dla zmiennych losowych typu ciągłego

Wartość oczekiwana zmiennej losowej typu ciągłego X określona jest wzorem

$$E(X) = \int_{-\infty}^{\infty} xg(x)dx, \quad (3)$$

przy założeniu, że całka (3) jest bezwzględnie zbieżna, to jest

$$\int_{-\infty}^{\infty} |x|g(x)dx = h,$$

gdzie h jest pewną liczbą rzeczywistą.

Wariancję zmiennej losowej X (typu ciągłego), przy założeniu, że ona istnieje, można wyrazić wzorem:

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 g(x)dx.$$

Wartość oczekiwana i wariancja zmiennej losowej o rozkładzie $U(0, 1)$

Niech $X \sim U(0, 1)$. Funkcja gęstości g jest równa 1 na $[0, 1]$; poza tym przedziałem jest równa 0. Mamy:

$$E(X) = \int_{-\infty}^{\infty} xg(x)dx = \int_0^1 xdx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

oraz

$$\begin{aligned} Var(X) &= \int_{-\infty}^{\infty} (x - \frac{1}{2})^2 g(x) dx = \\ &= \int_0^1 (x - \frac{1}{2})^2 dx = \left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}. \end{aligned}$$

Wartość oczekiwana i wariancja zmiennej losowej o rozkładzie normalnym

Można pokazać, że jeśli $X \sim N(\mu, \sigma)$, to

$$E(X) = \mu, \quad Var(X) = \sigma^2, \quad D(X) = \sigma.$$

Wymaga to obliczenia całek trochę bardziej skomplikowanych niż dla przypadku odpowiadającego $U(0, 1)$.

Niezależność zmiennych losowych

Definicja 3. Mówimy, że zmienne losowe X i Y są niezależne, jeżeli

$$P(X \in [a, b] \wedge Y \in [c, d]) = P(X \in [a, b]) \times P(Y \in [c, d])$$

dla dowolnych przedziałów $[a, b]$ i $[c, d]$.

Intuicyjnie: niezależne zmienne losowe odpowiadają realizacje liczbowe niezależnych zmiennych losowych.

Centralne Twierdzenie Graniczne

Twierdzenie 1. Jeżeli $X_1, X_2, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, $E(X_1) = \mu$, $D(X_1) = \sigma$, to zmienna losowa

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

gdzie $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$, ma w przybliżeniu standardowy rozkład normalny $N(0, 1)$, tj.

$$P\left(a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) \rightarrow \Phi(b) - \Phi(a)$$

dla dowolnych a i b , $a < b$

Innymi słowy, rozkład $\bar{X}$ jest w przybliżeniu równy rozkładowi normalnemu $N(\mu, \sigma/\sqrt{n})$.

Niech X będzie zmienną losową o rozkładzie dwumianowym z parametrami $n \in \mathbb{N}$ i $p \in (0, 1)$.

Z Centralnego Twierdzenia Granicznego wynika, że $\bar{X}$ ma w przybliżeniu rozkład

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

a X ma w przybliżeniu rozkład

$$N(np, \sqrt{np(1-p)}).$$

Momenty zmiennych losowych

Moment rzędu k rozkładu zmiennej losowej X , gdzie k jest liczbą naturalną, definiujemy wzorem

$$M_k = E(X^k).$$

Oznaczmy przez μ wartość oczekiwaną zmiennej losowej X . Moment centralny rzędu k rozkładu zmiennej losowej X , gdzie k jest liczbą naturalną większą lub równą niż dwa, definiujemy wzorem

$$m_k = E(X - \mu)^k.$$

Momenty próbkowe

Niech $X_1, X_2, \dots, X_n$ będą niezależnymi zmiennymi losowymi o wspólnym rozkładzie.

Niech $x_1, \dots, x_n$ będą realizacjami zmiennych losowych (odpowiednio) $X_1, \dots, X_n$. Dobrym oszacowaniem momentu rzędu k wspólnego rozkładu zmiennych losowych $X_j, j = 1, 2, \dots, n$, jest moment próbkowy rzędu k

$$\frac{1}{n} \sum_{j=1}^n x_j^k.$$

Dobrym oszacowaniem centralnego momentu rzędu k wspólnego rozkładu zmiennych losowych $X_j, j = 1, 2, \dots, n$, jest centralny moment próbkowy rzędu k

$$\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^k,$$

gdzie

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j.$$

Dla $k = 2$ zamiast centralnego momentu próbkowego jako oszacowanie wariancji rozkładu (czyli momentu centralnego rzędu dwa) wspólnego rozkładu zmiennych losowych $X_j, j = 1, 2, \dots, n$, stosowane jest również oszacowanie s^2 określone wzorem

$$s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2. \quad (4)$$

Zastosowanie do estymacji parametrów

Założmy, że $X_1, X_2, \dots, X_n$ są niezależnymi zmiennymi losowymi o wspólnym rozkładzie normalnym z parametrami μ i $\sigma > 0$. Niech $x_1, \dots, x_n$ będą realizacjami zmiennych losowych (odpowiednio) $X_1, \dots, X_n$.

Dobrym oszacowaniem parametru μ jest $\bar{x}$.

Za oszacowanie σ można przyjąć pierwiastek z wariancji próbkowej, zdefiniowanej wzorem (4), lub pierwiastek z momentu centralnego rzędu 2 dla danych $x_1, x_2, \dots, x_n$.

Lektura uzupełniająca

T. Bednarski, Elementy matematyki w naukach ekonomicznych. Oficyna ekonomiczna. Kraków 2004, str. 228–234.

Koronacki, J., Mielniczuk, J. Statystyka dla studentów kierunków technicznych i przyrodniczych. WNT. Warszawa 2001, podrozdział 2.2.1, str. 94–110.