Rachunek prawdopodobieństwa i statystyka

1 Momenty

Zmienna losowa jest wystarczająco dokładnie opisana przez jej rozkład prawdopodobieństwa. Względy praktyczne dyktują jednak potrzebę znalezienia charakterystyk liczbowych rozkładu, ponieważ są to opisy krótkie i umożliwiające szybkie porównanie rozkładów.
Momentem zwykłym rzędu r ($r = 1, 2, \ldots$) zmiennej losowej X nazywamy

$$m_r = \sum_k x_k^r p_k$$

w przypadku zmiennej losowej skokowej oraz

$$m_r = \int_{-\infty}^{+\infty} x^r f(x)\ dx$$

w przypadku zmiennej losowej ciągłej.
Moment zwykły rzędu pierwszego nazywamy wartością przeciętną lub wartością oczekiwaną i oznaczamy symbolem $E(X)$, tj.

$$E(X) = \sum_k x_k p_k$$

w przypadku zmiennej losowej skokowej oraz

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\ dx$$

w przypadku zmiennej losowej ciągłej.
Momentem centralnym rzędu r ($r = 1, 2, \ldots$) zmiennej losowej X nazywamy

$$\mu_r = \sum_k (x_k - m_1)^r p_k$$

w przypadku zmiennej losowej skokowej oraz

$$\mu_r = \int_{-\infty}^{+\infty} (x - m_1)^r f(x)\ dx$$

w przypadku zmiennej losowej ciągłej.

Moment centralny rzędu drugiego nazywamy wariancją i oznaczamy symbolem $D^2(X)$, tj.

$$D^2(X) = \sum_k \left[x_k - E(X)\right]^2 p_k$$

w przypadku zmiennej losowej skokowej oraz

$$D^2(X) = \int_{-\infty}^{+\infty} \left[x - E(X)\right]^2 f(x)\, dx$$

w przypadku zmiennej losowej ciągłej.
Pierwiastek kwadratowy z wariancji nazywamy odchyleniem standardowym i oznaczamy symbolem $D(X)$.
Przykład Rzut kostką do gry. Rozkład zmiennej losowej X jest następujący

x_k	1	2	3	4	5	6
p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Obliczamy wartość oczekiwaną

$$E(X) = \sum_{k=1}^{6} k p_k = \frac{1}{6}(1+2+3+4+5+6) = 3,5$$

oraz wariancję

$$D^2(X) = \sum_{k=1}^{6}[k - E(X)]^2 p_k = \frac{1}{6}[(1-3,5)^2 + (2-3,5)^2 +$$

$$+(3-3,5)^2 + (4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2] = 2,92,$$

skąd odchylenie standardowe $D(X) = 1,71$.
Przykład Zmienna losowa X ma rozkład jednostajny o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x < a, \\ \frac{1}{b-a} & \text{dla } a \leqslant x \leqslant b, \\ 0 & \text{dla } x > b. \end{cases}$$

Wartość oczekiwana tego rozkładu jest równa

$$E(X) = \frac{1}{b-a}\int_a^b x\, dx = \frac{a+b}{2},$$

a wariancja

$$D^2(X) = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{1}{12}(b-a)^2,$$

skąd odchylenie standardowe $D(X) = \frac{\sqrt{3}}{6}(b-a)$.

Przykład Zmienna losowa X podlega rozkładowi Bernoulliego. Funkcja prawdopodobieństwa tej zmiennej dana jest wzorem

$$P(X=k)=\binom{n}{k}p^k q^{n-k}, \qquad q=1-p, \qquad 0<p<1, \qquad k=0,1,\ldots,n.$$

Obliczymy wartość oczekiwaną tej zmiennej

$$E(X)=\sum_{k=0}^{n} k\cdot P(X=k)=\sum_{k=0}^{n} k\cdot\frac{n!}{k!(n-k)!}p^k q^{n-k}=\sum_{k=1}^{n} k\cdot\frac{n!}{k!(n-k)!}p^k q^{n-k}=$$

$$=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}=np\sum_{j=0}^{n-1}\frac{(n-1)!}{j!(n-1-j)!}p^j q^{n-1-j}=np(p+q)^{n-1}=np.$$

Analogicznie można obliczyć wariancję i odchylenie standardowe tej zmiennej i otrzymać kolejno $D^2(X)=npq$, $D(X)=\sqrt{npq}$.

Przykład Zmienna losowa X ma rozkład normalny o parametrach m i σ, $\sigma>0$ o gęstości

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \qquad -\infty<x<+\infty.$$

Obliczymy wartość oczekiwaną tej zmiennej

$$E(X)=\int_{-\infty}^{+\infty}\frac{x}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\,dx.$$

Stosując podstawienie $\frac{x-m}{\sigma}=z$ otrzymujemy

$$E(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(m+\sigma z)\exp\left[-\frac{z^2}{2}\right]\,dz=\frac{m}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{z^2}{2}\right]\,dz+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z\exp\left[-\frac{z^2}{2}\right]\,dz.$$

Ponieważ

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{z^2}{2}\right]\,dz=1$$

oraz

$$\int_{-\infty}^{+\infty}z\exp\left[-\frac{z^2}{2}\right]\,dz=0,$$

więc ostatecznie otrzymujemy $E(X)=m$. Analogicznie można obliczyć wariancję $D^2(X)=\sigma^2$.

Moment zwykły rzędu drugiego m_2 może być traktowany jako wartość oczekiwana zmiennej losowej $Y=X^2$, czyli $m_2=E(X^2)$. Korzystając z tego możemy otrzymać relację

$$D^2(X)=E(X^2)-E^2(X).$$

Przykład Rzut kostką do gry. Rozkład zmiennej losowej X^2 jest następujący

x_k^2	1	4	9	16	25	36
p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Obliczamy wartość oczekiwaną

$$E(X^2) = \sum_{k=1}^{6} k^2 p_k = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = 15,17,$$

skąd mamy

$$D^2(X) = 15,17 - (3,5)^2 = 15,17 - 12,25 = 2,92.$$

Przykład Zmienna losowa X ma rozkład jednostajny o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x < a, \\ \frac{1}{b-a} & \text{dla } a \leqslant x \leqslant b, \\ 0 & \text{dla } x > b. \end{cases}$$

Wartość oczekiwana

$$E(X^2) = \frac{1}{b-a} \int_a^b x^2 \, dx = \frac{a^2 + ab + b^2}{3},$$

skąd

$$D^2(X) = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

2 Mediana

Medianą $Me(X)$ zmiennej losowej typu skokowego X nazywa się liczbę x spełniającą związki

$$P(X \leqslant x) \geqslant \frac{1}{2} \qquad \text{oraz} \qquad P(X \geqslant x) \geqslant \frac{1}{2}.$$

Przykład Rozkład zmiennej losowej X jest następujący

x_k	2	5	7	10
p_k	0,1	0,2	0,5	0,2

Mediana $Me(X) = 7$, ponieważ

$$P(X \leqslant 7) = 0,8 > 0,5 \qquad \text{oraz} \qquad P(X \geqslant 7) = 0,7 > 0,5.$$

Medianą $Me(X)$ zmiennej losowej typu ciągłego X o gęstości f i dystrybuancie F nazywa się liczbę x spełniającą równość

$$F(x) = \frac{1}{2}$$

lub

$$\int_{-\infty}^{x} f(t) \, dt = \frac{1}{2}.$$

Przykład Zmienna losowa X ma rozkład wykładniczy o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x < 0, \\ \lambda e^{-\lambda x} & \text{dla } x \geqslant 0. \end{cases}$$

Mamy
$$\int_{-\infty}^{x} f(t)\ dt = \lambda \int_{0}^{x} \mathrm{e}^{-\lambda t}\ dt = -\mathrm{e}^{-\lambda x} + 1 = \frac{1}{2},$$
skąd otrzymujemy
$$x = Me(X) = \frac{1}{\lambda} \ln 2.$$
Dystrybuanta rozkładu wykładniczego jest dana wzorem
$$F(x) = \begin{cases} 0 & \text{dla } x \leqslant 0, \\ 1 - \mathrm{e}^{-\lambda x} & \text{dla } x > 0. \end{cases}$$
Zatem warunek $F(x) = \frac{1}{2}$ prowadzi do równości
$$1 - \mathrm{e}^{-\lambda x} = \frac{1}{2},$$
skąd
$$x = Me(X) = \frac{1}{\lambda} \ln 2.$$

3 Kwartyle

Kwartylem pierwszym (dolnym) $Q_1(X) = Q_d(X)$ zmiennej losowej typu skokowego X nazywa się liczbę x spełniającą związki
$$P(X \leqslant x) \geqslant \frac{1}{4} \qquad \text{oraz} \qquad P(X \geqslant x) \geqslant \frac{3}{4}.$$
Kwartylem trzecim (górnym) $Q_3(X) = Q_g(X)$ zmiennej losowej typu skokowego X nazywa się liczbę x spełniającą związki
$$P(X \leqslant x) \geqslant \frac{3}{4} \qquad \text{oraz} \qquad P(X \geqslant x) \geqslant \frac{1}{4}.$$
Z powyższych definicji wynika, że $Q_2(X) = Me(X)$.

Przykład Rozkład zmiennej losowej X jest następujący

x_k	2	5	7	10
p_k	0,1	0,2	0,5	0,2

Kwartyl pierwszy $Q_1(X) = 5$, ponieważ
$$P(X \leqslant 5) = 0,3 > 0,25 \qquad \text{oraz} \qquad P(X \geqslant 5) = 0,9 > 0,75.$$
Kwartyl trzeci $Q_3(X) = 7$, ponieważ
$$P(X \leqslant 7) = 0,8 > 0,75 \qquad \text{oraz} \qquad P(X \geqslant 7) = 0,7 > 0,25.$$

Kwartylem pierwszym (dolnym) $Q_1(X) = Q_d(X)$ zmiennej losowej typu ciągłego X o gęstości f i dystrybuancie F nazywa się liczbę x spełniającą równość

$$F(x) = \frac{1}{4}$$

lub

$$\int_{-\infty}^{x} f(t)\ dt = \frac{1}{4}.$$

Kwartylem trzecim (górnym) $Q_3(X) = Q_g(X)$ zmiennej losowej typu ciągłego X o gęstości f i dystrybuancie F nazywa się liczbę x spełniającą równość

$$F(x) = \frac{3}{4}$$

lub

$$\int_{-\infty}^{x} f(t)\ dt = \frac{3}{4}.$$

Z powyższych definicji wynika, że $Q_2(X) = Me(X)$.

Przykład Zmienna losowa X ma rozkład wykładniczy o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x < 0, \\ \lambda \mathrm{e}^{-\lambda x} & \text{dla } x \geqslant 0. \end{cases}$$

Mamy

$$\int_{-\infty}^{x} f(t)\ dt = \lambda \int_{0}^{x} \mathrm{e}^{-\lambda t}\ dt = -\mathrm{e}^{-\lambda x} + 1 = \frac{1}{4},$$

skąd otrzymujemy

$$x = Q_1(X) = \frac{1}{\lambda} \ln \frac{4}{3}.$$

Dystrybuanta rozkładu wykładniczego jest dana wzorem

$$F(x) = \begin{cases} 0 & \text{dla } x \leqslant 0, \\ 1 - \mathrm{e}^{-\lambda x} & \text{dla } x > 0. \end{cases}$$

Zatem warunek $F(x) = \frac{1}{4}$ prowadzi do równości

$$1 - \mathrm{e}^{-\lambda x} = \frac{1}{4},$$

skąd

$$x = Q_1(X) = \frac{1}{\lambda} \ln \frac{4}{3}.$$

Analogicznie

$$\int_{-\infty}^{x} f(t)\ dt = \lambda \int_{0}^{x} \mathrm{e}^{-\lambda t}\ dt = -\mathrm{e}^{-\lambda x} + 1 = \frac{3}{4},$$

skąd otrzymujemy

$$x = Q_3(X) = \frac{1}{\lambda} \ln 4.$$

Dystrybuanta rozkładu wykładniczego jest dana wzorem

$$F(x) = \begin{cases} 0 & \text{dla } x \leqslant 0, \\ 1 - \mathrm{e}^{-\lambda x} & \text{dla } x > 0. \end{cases}$$

Zatem warunek $F(x) = \frac{3}{4}$ prowadzi do równości

$$1 - \mathrm{e}^{-\lambda x} = \frac{3}{4},$$

skąd

$$x = Q_3(X) = \frac{1}{\lambda} \ln 4.$$

4 Dominanta (Moda)

Dominantą $Do(X)$ zmiennej losowej typu skokowego X nazywa się wartość zmiennej losowej, której odpowiada największe prawdopodobieństwo.

Przykład Rozkład zmiennej losowej X jest następujący

x_k	2	5	7	10
p_k	0,1	0,2	0,5	0,2

Z rozkładu zmiennej loswej wynika bezpośrednio, że $Do(X) = 7$.

Dominantą $Do(X)$ zmiennej losowej typu ciągłego X nazywa się wartość zmiennej losowej X, dla której gęstość przyjmuje maksimumu lokalne.

Przykład Zmienna losowa X ma rozkład o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x < 0, \\ 6(x - x^2) & \text{dla } 0 \leqslant x \leqslant 1, \\ 0 & \text{dla } x > 1. \end{cases}$$

Ponieważ

$$f'(x) = \begin{cases} 0 & \text{dla } x < 0, \\ 6(1 - 2x) & \text{dla } 0 < x < 1, \\ 0 & \text{dla } x > 1. \end{cases}$$

co oznacza, że $Do(X) = \frac{1}{2}$.

Uwaga Zachodzi wzór przybliżony (zwany wzorem Pearsona)

$$E(X) - Do(X) \approx 3\left[E(X) - Me(X)\right].$$

Przykład Zmienna losowa X ma rozkład o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x < 0, \\ 6(x - x^2) & \text{dla } 0 \leqslant x \leqslant 1, \\ 0 & \text{dla } x > 1. \end{cases}$$

Wiemy, że $Do(X) = \frac{1}{2}$. Obliczymy $E(X)$

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\ dx = \int_0^1 (6x^2 - 6x^3)\ dx = \left[2x^3 - \frac{3}{2}x^4\right]_0^1 = \frac{1}{2}.$$

Z wzoru Pearsona wynika, że $Me(X) \approx \frac{1}{2}$.
Wyznaczenie mediany z definicji prowadzi do równania

$$\int_0^x (6t - 6t^2)\ dt = 3x^2 - 2x^3 = \frac{1}{2},$$

którego jednym z rozwiązań jest $x = \frac{1}{2}$, czyli $Me(X) = \frac{1}{2}$.

Przykład Rozkład zmiennej losowej X jest następujący

x_k	2	5	7	10
p_k	0,1	0,2	0,5	0,2

Wiemy, że $Me(X) = Do(X) = 7$.
Z wzoru Pearsona wynika więc, że $E(X) \approx 7$.
Wyznaczenie wartości oczekiwanej z definicji prowadzi do

$$E(X) = 2 \cdot 0,1 + 5 \cdot 0,2 + 7 \cdot 0,5 + 10 \cdot 0,2 = 6,7.$$