Rozkłady ciągłe

Joanna Czarnowska[1]

[1]Uniwersytet Gdański
Instytut Informatyki
joanna.czarnowska@ug.edu.pl

Rozkład zmiennej losowej typu ciągłego – gęstość

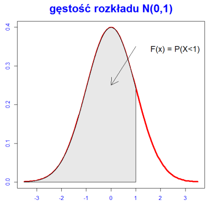
Rozkład zmiennej losowej, dla której dystrybuanta ma postać

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

gdzie $f : \mathbb{R} \longrightarrow \mathbb{R}$ jest funkcją nieujemną taką, że pole pomiędzy wykresem tej funkcji, a osią OX jest równe jeden

$$\forall_x \ f(x) \geqslant 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1,$$

nazywamy rozkładem *typu ciągłego* lub krótko *ciągłym*. Funkcję f nazywamy gęstością rozkładu.



Gęstość rozkładu normalnego standardowego:

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \ x \in \mathbb{R}.$

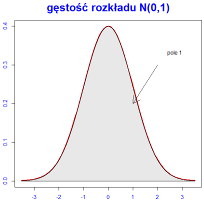
Dystrybuanta w punkcie jeden

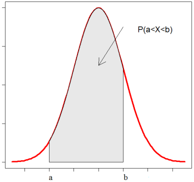
$F(1) = P(X \leqslant 1) = P(X < 1) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \ 0{,}84.$

„Czytanie” gęstości

Pole pod gęstością od punktu a do b to prawdopodobieństwo, że zmienna losowa X przyjmuje wartości z przedziału o takich końcach

$$P(a < X \leqslant b) = P(X \leqslant b) - P(X \leqslant a) = F(b) - F(a) = \int_a^b f(x).$$





Dystrybuanta zmiennej losowej ciągłej

Dla zmiennej losowej X ciągłej, zachodzi równość

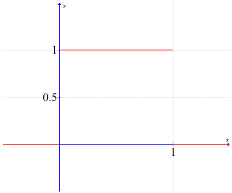
$$F(x) = P(X \leqslant x) = P(X < x) = \int_{-\infty}^{x} f(t)dt.$$

Stąd mamy

$$P(a < X \leqslant b) = P(a \leqslant X < b) = P(a < X < b) = P(a \leqslant X \leqslant b) = F(b) - F(a).$$

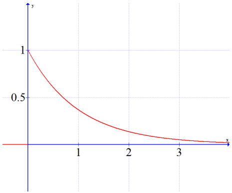
„Czytanie” gęstości – przykłady

Patrząc na wykresy gęstości odpowiedz na pytanie, dla którego z poniższych trzech rozkładów, prawdopodobieństwo przyjęcia wartości z przedziału $(1,2)$ jest największe?



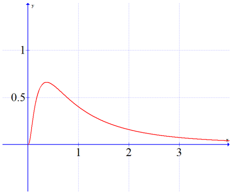
rozkład jednostajny U(0,1)

$f(x) = 1, \ x \in (0,1)$



rozkład wykładniczy Exp(1)

$f(x) = e^{-x}, \ x > 0$



rozkład log-normalny LN(0,1)

$f(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln x)^2}{2}}, \ x > 0$

„Czytanie” gęstości – przykłady

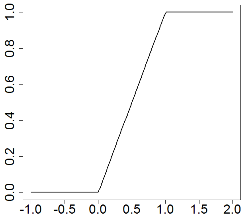
Obliczmy teraz prawdopodobieństwo $P(X \in (1,2))$, dla każdego z trzech powyższych rozkładów. W przypadku rozkładu jednostajnego odpowiedź jest oczywista – wynosi ono zero. Dla rozkładów wykładniczego i log-normalnego mamy odpowiednio

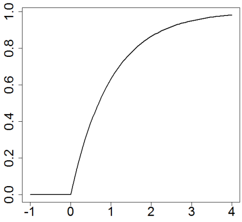
$$P(X \in (1,2)) = \int_1^2 e^{-x} dx = -e^{-x}\Big|_1^2 dx = e^{-1} - e^{-2} \approx 0{,}23$$

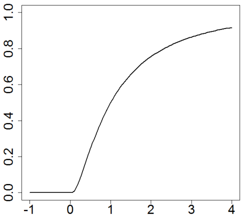
$$P(X \in (1,2)) = \int_1^2 \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln x)^2}{2}} dx \approx 0{,}26.$$

„Czytanie” dystrybuanty

Z definicji dystrybuanty wynika, że jest to funkcja niemalejąca, przyjmująca wartości z przedziału $[0, 1]$. Analiza wykresu nie jest tak oczywista jak w przypadku gęstości.







rozkład jednostajny U(0,1)

$F(x)=x, \;\; x \in (0, 1)$

rozkład wykładniczy Exp(1)

$F(x)=1-e^{-x}, \;\; x > 0$

rozkład log-normalny LN(0,1)

$F(x)=\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right), \;\; x > 0$

Uwaga. $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ – funkcja błędu (error function) – więcej o niej np. na https://en.wikipedia.org/wiki/Error_function

Parametry rozkładu

W nawiasach dla porównania podane są definicje parametrów dla zmiennej losowej dyskretnej X, przyjmującej wartośći $x_1, x_2, \ldots$.

Wartość oczekiwana

$$m = E(X) = \int_{-\infty}^{\infty} x f(x) dx \qquad (EX = \sum_n x_n P(X = x_n))$$

Wariancja

$$Var(X) = E(X - m)^2 = \int_{-\infty}^{\infty} (x - m)^2 f(x) dx \qquad (\mathrm{Var}(X) = \sum_n (x_n - m)^2 P(X = x_n))$$

k-ty moment zwykły

$$E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx \qquad (E(X^k) = \sum_n x_n^k P(X = x_n))$$

k-ty moment centralny

$$E(X - m)^k = \int_{-\infty}^{\infty} (x - m)^k f(x) dx \qquad (E(X - m)^k = \sum_n (x_n - m)^k P(X = x_n))$$

Przykłady rozkładów ciągłych

Zobacz: 02RozkładyCiagle.pdf