Rozkłady probabilistyczne

Mgr inż. Szymon Łukasik

szymonl@pk.edu.pl

Gęstość rozkładu prawdopodobieństwa

Dotąd rozważane było prawdopodobieństwo w jego klasycznej definicji (zdarzenia sprzyjające, przestrzeń zdarzeń itp.). Poznaliście także Państwo metody generowania liczb losowych.

Stwierdziliśmy, że idealny generator powinien generować liczby:

a) niezależne od siebie (to w przypadku generatora kongruencyjnego okazało się nieprawdą)
b) z równomiernie rozłożonym prawdopodobieństwem

Co to jest rozkład prawdopodobieństwa?

Rozkład prawdopodobieństwa opisuje prawdopodobieństwa związane z wszystkimi możliwymi wartościami jakie może przyjąć zmienna losowa.

No dobrze – jak to ująć funkcyjnie?

Poprzez **gęstość rozkładu prawdopodobieństwa**:

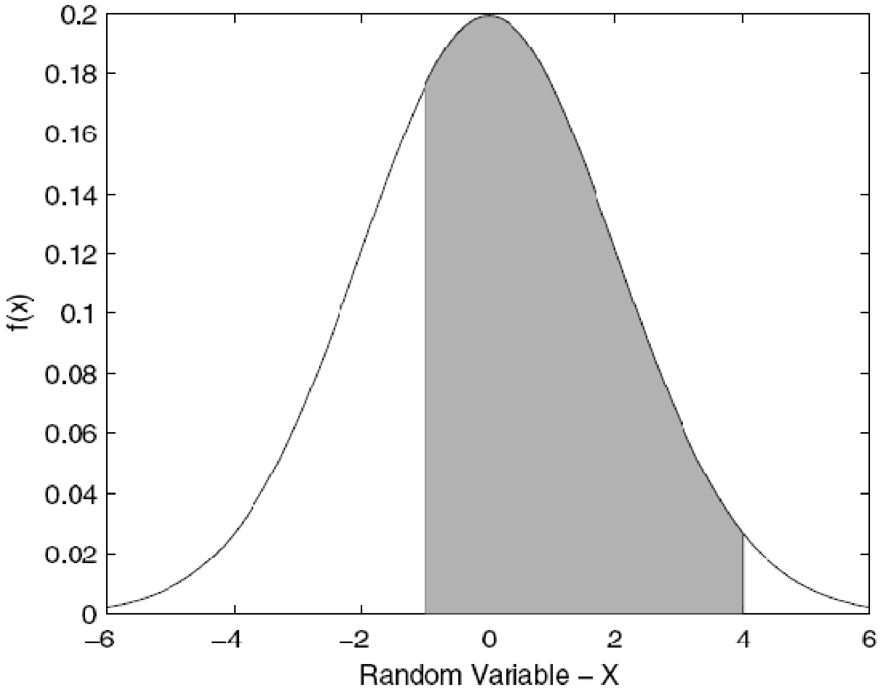
- zmienna ciągła – funkcja f(x) gdzie *x* oznacza element próby zmiennej losowej *X*

$$f : x \to R^{+}$$

$$P(a \le X \le b) = \int_a^b f(x)dx$$

oczywiste jest, że jeśli wartości zmiennej są określone na przedziale np. $(-\infty, \infty)$ to wartość powyższej całki powinna być równa 1.

Ilustracja:



- zmienna o wartościach dyskretnych – analogicznie: jeśli może przyjmować wartości $x_1, x_2, \ldots x_k$ to:

$$f(x_i) = P(X = x_i), \quad i = 1, 2, \ldots,$$

Dodatkowo definiuje się następujące charakterystyki funkcyjne:

- **dystrybuanta F(x)**

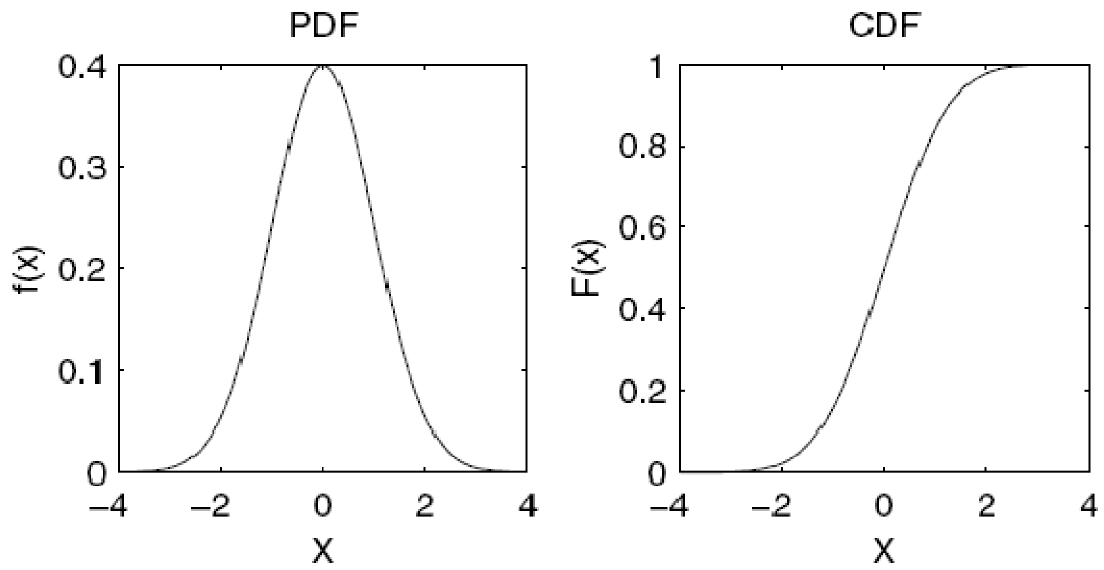
- zmienna ciągła

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

- zmienna dyskretna

$$F(a) = \sum_{x_i \le a} f(x_i), \quad i = 1, 2, \ldots$$

Ilustracja gęstości rozkładu i dystrybuanty:



- **kwantyl**

rzędu r to wartość dla której *F(q)=r*. Kwantyl rzędu ½ to tzw. mediana.

Dla określonego rozkładu istotne znaczenie mają: wartość oczekiwana (czyli miara oczekiwanej „centralnej tendencji" rozkładu) oraz wariancja (miara rozrzutu rozkładu względem wartości oczekiwanej).

Dla ciągłej zmiennej losowej mają one postać:

$$\mu = E[X] = \int_{-\infty}^{x} xf(x)dx$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

inaczej:

$$V(x) = E[X^2] - \mu^2 = E[X^2] - (E[X])^2$$

A dla dyskretnej:

$$\mu = E[X] = \sum_{i=1}^{\infty} x_i f(x_i)$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i)$$

Odchylenie standardowe – pierwiastek z wariancji.

Najczęściej spotykane postacie rozkładów prawdopodobieństwa

a) rozkład jednostajny

$$f(x;a,b) = \frac{1}{b-a}, \quad a < x < b$$

Wartość oczekiwana:

$$E[X] = \frac{a+b}{2}$$

Wariancja:

$$V(X) = \frac{(b-a)^2}{12}$$

Dystrybuanta:

$$F(x) = \begin{cases} 0; & x \le a \\ \dfrac{x-a}{b-a}; & a < x < b \\ 1; & x \ge b. \end{cases}$$

Funkcje w MATLABie do obliczenia f(x) i F(x):

unifpdf, unifcdf (obie wymagają podania : nośnika i dwóch parametrów a, b)

b) rozkład normalny

$$f(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- wartość oczekiwana i wariancja są tu widoczne gołym okiem (nieuzbrojonym po zęby). Oznaczenie – czytelne $X \sim N(\mu,\sigma^2)$.

Własności – maksimum w μ, symetryczna względem μ, gdy x zmierza do $-\infty, \infty$ to f(x) zmierza do zera.

Funkcje do wyznaczenia f(x) i F(x) – analogicznie: *normpdf* i *normcdf* (podajemy x, wartość oczekiwaną i odchylenie standardowe).

Istnieje także śliczna funkcja do wyznaczania prawdopodobieństwa, że zmienna losowa przyjmuje wartości z przedziału [a,b]: *normspec([a,b], mu, sigma*).

c) rozkład dwumianowy (dyskretny)

$$f(x;n,p)=P(X=x)=\binom{n}{x}p^x(1-p)^{n-x};\quad x=0,1,....,n$$

p – prawdopodobieństwo pojedynczego sukcesu

n – ilość eksperymentów

X – ilość sukcesów w tychże n-próbach

Wartość oczekiwana:

$$E[X]=np$$

Wariancja:

$$V(X)=np(1-p)$$

Funkcje: *binopdf, binocdf* (parametry *x, n, p*)

d) rozkład wykładniczy

$$f(x;\lambda) = \lambda e^{-\lambda x}; \quad x \geq 0; \lambda > 0$$

lub:

$$f(x;\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}; \quad x \geq 0; \mu > 0$$

Rozkład ten może zostać użyty do modelowania czasu pomiędzy zdarzeniami niezależnymi. Parametr lambda to wtedy częstotliwość zdarzeń przypadająca na okres obserwacji.

Funkcje - *exppdf(x,mu), expcdf(x,mu)*

d) rozkład chi kwadrat

$$f(x;\nu) = \frac{1}{\Gamma(\nu/2)} \left(\frac{1}{2}\right)^{\nu/2} x^{\nu/2-1} e^{-\frac{1}{2}x}; \quad x \geq 0$$

v – stopnie swobody

Dlaczego się o nim uczymy ? Bo stosowany jest w testach statystycznych.

Funkcje: *chi2pdf, chi2cdf* (trzeba podać x i ilość stopni swobody).