Statystyka w analizie i planowaniu eksperymentu

Paweł Błażej

10 marca 2014

Zmienna losowa - wstęp

Przeprowadzane w praktyce badania i eksperymenty mają bardzo różnorodny charakter, niemniej jednak wiążą się one z rejestracją jakiś sygnałów (danych). Mogą to być na przykład:

1. odczyty na skali;

Zmienna losowa - wstęp

Przeprowadzane w praktyce badania i eksperymenty mają bardzo różnorodny charakter, niemniej jednak wiążą się one z rejestracją jakiś sygnałów (danych). Mogą to być na przykład:

1. odczyty na skali;
2. końcowe parametry jakiegoś procesu technologicznego;

Zmienna losowa - wstęp

Przeprowadzane w praktyce badania i eksperymenty mają bardzo różnorodny charakter, niemniej jednak wiążą się one z rejestracją jakiś sygnałów (danych). Mogą to być na przykład:

1. odczyty na skali;
2. końcowe parametry jakiegoś procesu technologicznego;
3. liczba osób w kolejce.

Klasyfikacja zmiennych Stevensa

1. Zmienne jakościowe (nazywane również kategorycznymi, czynnikowymi), to zmienne przyjmujące określoną liczbę wartości (najczęściej nieliczbowych):
    - binarne, np. płeć (kobieta/mężczyzna);
    - nominalne, np. marka samochodu;
    - porządkowe, np. wykształcenie (podstawowe, średnie, wyższe).

Klasyfikacja zmiennych Stevensa

1. Zmienne jakościowe (nazywane również kategorycznymi, czynnikowymi), to zmienne przyjmujące określoną liczbę wartości (najczęściej nieliczbowych):
   - binarne, np. płeć (kobieta/mężczyzna);
   - nominalne, np. marka samochodu;
   - porządkowe, np. wykształcenie (podstawowe, średnie, wyższe).
2. Zmienne ilościowe, opisują ilość. Wyróżnia się skale:
   - licznikową (liczebność wystąpień pewnego zjawiska, opisywana przez liczby naturalne) np. liczba lat nauki;
   - przedziałową (interwałową) skala, w której zmienna może przyjmować dowolne wartości z określonego przedziału.

Zmienna losowa i jej dystrybuanta

Tej definicji nie trzeba uczyć się na pamięć!

Zmienną losową nazywamy funkcję X określoną na przestrzeni zdarzeń elementarnych Ω, o wartościach ze zbioru liczb rzeczywistych $\mathbf{R}$.

Uwaga

Zmienne losowe oznaczamy dużymi literami X, Y, Z a ich konkretne wartości małymi literami x, y, z.

Przykład - Rzut kostką

Niech $\Omega = \{\omega_1, \dots, \omega_6\}$, gdzie ω_i, $i = 1, 2, \dots, 6$ oznacza zdarzenie elementarne polegające na wyrzuceniu i-oczek. Określmy zmienną losową

$$X(\omega_i) = i, \;\; i = 1, 2, \dots, 6.$$

Prawdopodobieństwo przyjęcia przez zmienną losową wartości z danego zbioru

Zdarzeniami losowymi są również zbiory definiowane w następujący sposób:

$$\{\omega \ : \ X(\omega) \in A\}$$

gdzie A jest dowolnym podzbiorem $\mathbf{R}$.

Prawdopodobieństwo przyjęcia przez zmienną losową wartości z danego zbioru

Prawdopodobieństwo $P(X \in A)$ przyjęcia przez zmienna losową X wartości ze zbioru A gdzie $A \subset \mathbf{R}$, określamy następującą równością:

$$P(X \in A) = P(\{\omega : \ X(\omega) \in A\}).$$

Dystrybuanta zmiennej losowej i jej własności

Dystrybuanta zmiennej losowej X

Funkcję F_X określoną na zbiorze $\mathbf{R} = (-\infty, +\infty)$ liczb rzeczywistych wzorem:

$$F_X(x) = P(X \leqslant x), \ \ x \in \mathbf{R}$$

nazywamy dystrybuantą zmiennej losowej X.

Jeżeli nie ma wątpliwości z jaką zmienną losową mamy do czynienia wtedy dystrybuantę oznaczamy przez F.

Przykład cd.

Załóżmy, że prawdopodobieństwo wylosowania wadliwego towaru wynosi $0 < p < 1$. Wówczas dystrybuanta zmiennej losowej X przyjmuje postać

$$F(x) = \begin{cases} 0, & \text{dla} \quad x < 0 \\ 1 - p & \text{dla} \quad 0 \leqslant x < 1 \\ 1 & \text{dla} \quad x \geqslant 1 \end{cases}$$

Własności dystrybuanty

1. $0 \leqslant F(x) \leqslant 1$ dla każdego $x \in \mathbf{R}$;

Własności dystrybuanty

1. $0 \leqslant F(x) \leqslant 1$ dla każdego $x \in \mathbf{R}$;
2. $\lim_{x \to -\infty} F(x) = 0$ oraz $\lim_{x \to +\infty} F(x) = 1$;

Własności dystrybuanty

1. $0 \leqslant F(x) \leqslant 1$ dla każdego $x \in \mathbf{R}$;
2. $\lim_{x \to -\infty} F(x) = 0$ oraz $\lim_{x \to +\infty} F(x) = 1$;
3. jest funkcją niemalejącą;

Własności dystrybuanty

1. $0 \leqslant F(x) \leqslant 1$ dla każdego $x \in \mathbf{R}$;
2. $\lim_{x \to -\infty} F(x) = 0$ oraz $\lim_{x \to +\infty} F(x) = 1$;
3. jest funkcją niemalejącą;
4. $P(a < X \leqslant b) = F(b) - F(a)$

Dwa ważne typy zmiennych losowych

Zmienna losowa typu dyskretnego

Mówimy, że zmienna losowa X jest typu skokowego (dyskretnego) jeżeli istnieje skończony albo przeliczalny zbiór $W_x = \{x_1, x_2, \ldots, \ldots\}$ jej wartości taki, że

$$P(X = x_i) = p_i, \;\; i \in \mathbf{N},$$

$$\sum_{i=1} p_i = 1,$$

gdzie górna granica sumowania wynosi n albo ∞

Dwa ważne typy zmiennych losowych

Funkcja prawdopodobieństwa

Funkcję p przyjmującą wartości $p(x_i) = P(X = x_i)$ oznaczaną często przez p_i nazywamy funkcją prawdopodobieństwa zmiennej losowej X.

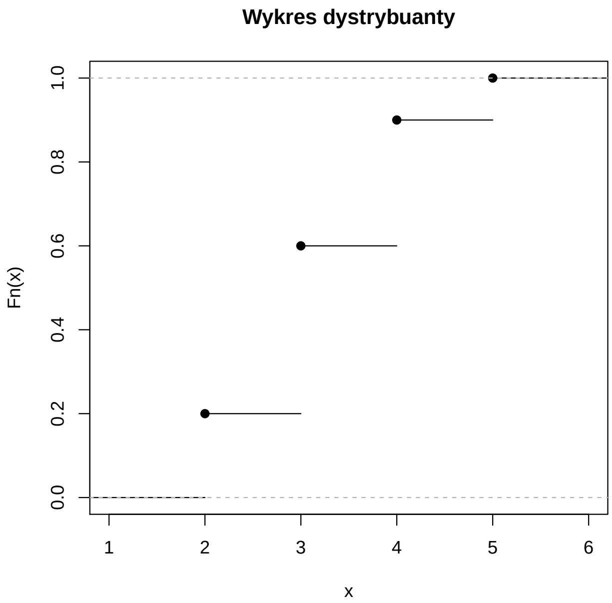
Dystrybuanta zmiennej losowej typu dyskretnego

Gdy dane jest funkcja prawdopodobieństwa zmiennej losowej X, to prawdopodobieństwo przyjęcia przez tą zmienną wartości ze zbioru A jest określone równością:

$$P(X \in A) = \sum_{x_i \in A} p_i$$

W szczególności dla dowolnego przedziału (a, b) zachodzi

$$P(-\infty < X < b) = \sum_{-\infty < x_i < b} p_i$$


Wykres dystrybuanty
Fn(x)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
x

Wybrane charakterystyki rozkładów - przypadek dyskretny

1. wartość średnia (parametr położenia)

$$\mu = \sum_{i=1}^{k} x_i p_i;$$

2. wariancja (parametr skali)

$$\sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 p_i.$$

3. skośność

$$\gamma = \sum_{i=1}^{k} (x_i - \mu)^3 p_i / \sigma^3$$

4. kurtoza

$$\eta = \sum_{i=1}^{k} (x_i - \mu)^4 p_i / \sigma^4$$

Ważne rozkłady typu dyskretnego

1. Rozkład Bernoulliego (rzut monetą);

Ważne rozkłady typu dyskretnego

1. Rozkład Bernoulliego (rzut monetą);
2. Rozkład dwumianowy (rzut wieloma monetami);

Ważne rozkłady typu dyskretnego

1. Rozkład Bernoulliego (rzut monetą);
2. Rozkład dwumianowy (rzut wieloma monetami);
3. Rozkład Poissona (liczba sygnałów);

Ważne rozkłady typu dyskretnego

1. Rozkład Bernoulliego (rzut monetą);
2. Rozkład dwumianowy (rzut wieloma monetami);
3. Rozkład Poissona (liczba sygnałów);
4. Rozkład hipergeometryczny (ryby w stawie).

Przykłady rozkładów dyskretnych

Rozkład Bernoulliego

Zmienna losowa X ma rozkład Bernoulliego, jeżeli jej funkcja prawdopodobieństwa jest postaci

x_i	0	1
p_i	q	p

$$\mu = p$$

$$\sigma^2 = pq$$

Rozkład dwumianowy

Zmienna losowa X ma rozkład dwumianowy z parametrami (n, p), $n \in \mathbf{N}$, $0 < p < 1$, jeżeli jej funkcja prawdopodobieństwa jest postaci

$$P(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mu = np$$

$$\sigma^2 = npq$$

Zadanie

Rzucamy 6 razy symetryczną monetą oblicz prawdopodobieństwo wyrzucenia conajmniej jednego „orła”

Rozkład hipergeometryczny

Zmienna losowa X ma rozkład hipergeometryczny z parametrami (N, M, n), gdzie N, M, n to liczby naturalne oraz $M, n \leqslant N$, jeżeli jej funkcja prawdopodobieństwa jest postaci

$$P(k; N, M, n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$\mu = np$$

$$\sigma^2 = npq\frac{N-n}{N-1}$$

Zadanie

W stawie jest $N = 100$ ryb odławiamy $M = 40$ z nich znakujemy je i z powrotem wrzucamy do stawu. Następnie łowimy $n = 20$ sztuk. Jakie jest prawdopodobieństwo, że wsród nich będzie dokładnie k oznakowanych?

Rozkład Poissona

Zmienna losowa X ma rozkład Poissona z parametrem λ, gdzie $\lambda > 0$, jeżeli jej funkcja prawdopodobieństwa jest postaci

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$

Rozkład Poissona jest związany z sytuacją zliczania zdarzeń losowych określonego typu w pewnym odcinku czasu. Może to być na przykład zliczenie ilości kolejnych klientów pojawiających się w kasie w banku, ilość samochodów przejeżdżających punkt kontrolny.

Twierdzenie Poissona

Jeżeli $n \to \infty$, $p_n \to 0$, $np_n \to \lambda$, to

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}$$

Zadanie

Udowodnij twierdzenie Poissona

Twierdzenie Poissona zastosowania

Prawdopodobieństwo trafienia „szóstki ”w Toto-Lotku jest równe $1/\binom{49}{6} = 1/139883816$. Ilu szóstek należy się spodziewać w każdym tygodniu, jeżeli grający wypełniają kupony niezależnie od siebie i całkowicie losowo, kuponów jest $n = 10^7$?

Zadanie

Dana jest funkcja prawdopodobieństwa zmiennej losowej X

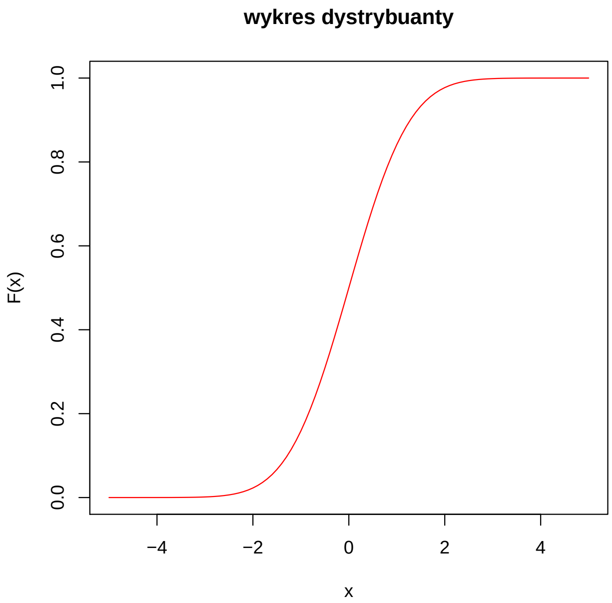
x_i	2	3	4	5
p_i	0,2	0,4	0,3	0,1

wyznacz:

1. funkcję dystrybuanty i jej wykres;
2. prawdopodobieństwo $P(X < 3,5)$ korzystając z wykresu dystrybuanty;
3. prawdopodobieństwo $P(3 < X \leqslant 4,5)$.

Zmienne losowe typu ciągłego

Zmienne losowe typu ciągłego mogą przyjmować nieskończenie wiele wartości (np. temperatura powietrza, waga ciała, prędkość samochodu w zadanej chwili czasu)

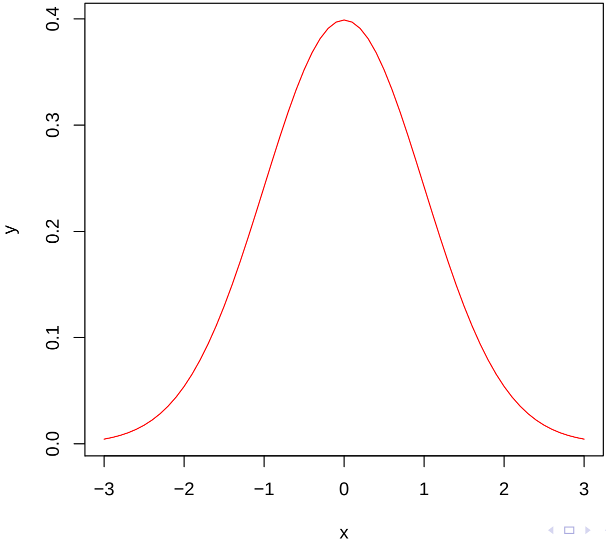

wykres dystrybuanty
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
x

Zmienna losowa typu ciągłego

Funkcja gęstości zmiennej losowej X typu ciągłego, to nieujemna funkcja f, taka, że dla każdego przedziału (x_1, x_2)

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

Gęstość zmiennej losowej typu ciągłego



Własności funkcji gęstości

1. $f(x) \geqslant 0$ dla każdego x;
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Wybrane rozkłady typu ciągłego

1. Rozkład normalny;
2. Rozkład Studenta;
3. Rozkład χ^2;
4. Rozkład F.

Przykłady rozkładów ciągłych

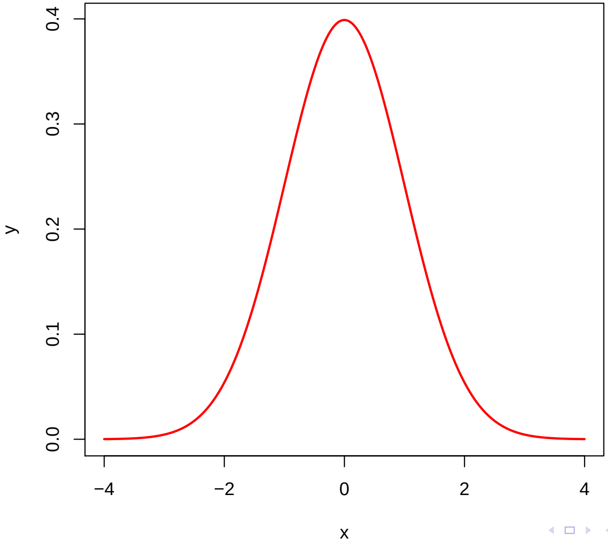
Rozkład normalny

Niech X będzie ciągłą zmienną losową o gęstości prawdopodobieństwa

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

gdzie $\mu \in \mathbf{R}$ i $\sigma > 0$ są danymi parametrami. Mówimy wtedy, że X ma *rozkład normalny* co oznaczamy $N(\mu, \sigma)$.

Funkcja gęstości rozkładu normalnego $N(\mu, \sigma)$.



Przykłady rozkładów ciągłych

Rozkład t-Studenta

Zmienna losowa X ma rozkład t-Studenta o n stopniach swobody, jeżeli jej gęstość f wyraża się wzorem

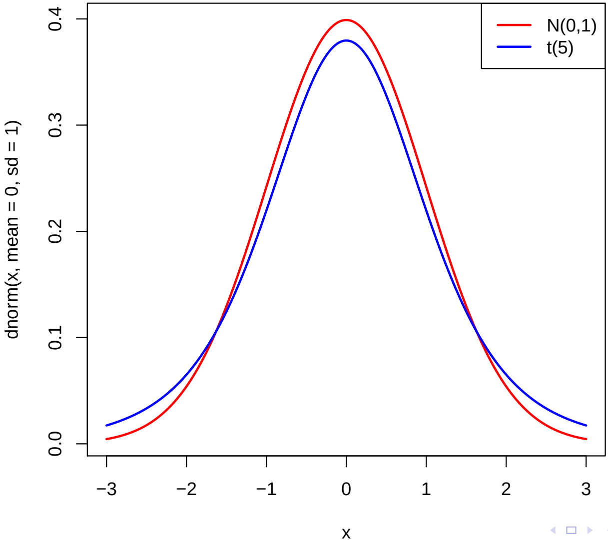
$$f(x) = \frac{\Gamma([n+1]/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+\frac{x^2}{n})^{-(n+1)/2}, x \in \mathbf{R}, n \in \mathbf{N}.$$

gdzie $\Gamma(n) = \int_0^{+\infty} x^{n-1}e^{-x}dx,\ \ n > 0$

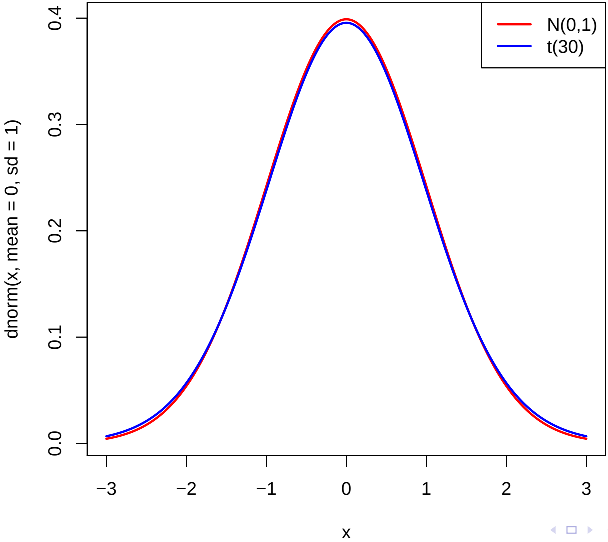
Uwaga

Gdy $n \to \infty$, gęstość f „dąży” do gęstości rozkładu normalnego $N(0,1)$, co wykorzystuje się w praktyce dla $n \geqslant 30$ do przybliżania rozkładu Studenta rozkładem normalnym.

Wykresy gęstości rozkładu Studenta i rozkładu normalnego



Wykresy gęstości rozkładu Studenta i rozkładu normalnego



Rozkład χ^2

Zmienna losowa X ma rozkład χ^2 o n stopniach swobody, jeżeli jej gęstość f wyraża się wzorem

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

Wykresy gęstości rozkładu χ^2 o $n = 2, 3, 4$ stopniach swobody