Statystyka w analizie i planowaniu eksperymentu

Paweł, Błażej

28 marca 2012

Zmienna losowa wstęp

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

- odczyty na skali;
- końcowe parametry jakiegoś procesu technologicznego;
- liczba osób w kolejce.

Takie liczbowe charakterystyki, które przy powtórzeniach eksperymentu dają różne wartości, określa się mianem **zmiennych losowych**.

Zmienna losowa i jej dystrybuanta

Niech $(\Omega, \mathcal{F}, P)$ będzie dowolną przestrzenią probabilistyczną. Zmienną losową nazywamy dowolną funkcję X określoną na przestrzeni zdarzeń elementarnych Ω, o wartościach ze zbioru liczb rzeczywistych.

$$X : \Omega \rightarrow \mathbf{R}.$$

Uwaga

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

Przykład

Rozważmy losowanie produktów z partii w celu zbadania ich jakości. Określmy zmienną losową

$$X(\omega) = \begin{cases} 1, & \text{gdy wyrób jest wadliwy;} \\ 0, & \text{gdy wyrób jest dobry.} \end{cases}$$

Ta zmienna przyjmuje tylko dwie wartości i jest nazywana **zerojedynkową**.

Przykład

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

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

Prawdopodobieństwo przyjęcia przez zmienną losową wartości z danego zbioru

Zdarzeniami losowymi są również zbiory definiowane w sposób następujący

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

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

Prawdopodobieństwo $P(X \in A)$ przyjęcia przez zmienną 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

Definicja

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

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

nazywamy dystrybuantą zmiennej losowej X.

Jeżeli nie ma wątpliwości z jaką zmienną losową mamy doczynienia 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 $x \in \mathbf{R}$;
2. $\lim_{x \to -\infty} F(x) = 0$ oraz $\lim_{x \to +\infty} F(x) = 1$;
3. jest funkcją niemalejącą;
4. jest funkcją (conajmniej) prawostronnie ciągła $F(x_0 + 0) = F(x_0)$, $x \in \mathbf{R}$
5. $P(a \leqslant X < b) = F(b) - F(a)$

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 ∞.

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 skokowego

Gdy dane jest funkcja prawdopodobieństwa zmiennej losowej X, to prawdopodobieństwo przyjęcia przez tę 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

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

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 $P(3 \leqslant X < 4,5)$.

Zmienna losowa typu ciągłego

Mówimy, że zmienna losowa X jest typu ciagłego, jeżeli istnieje nieujemna funkcja f, określona i całkowalna do jedynki na całej osi taka, że dla każdego przedziału (x_1, x_2)

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

Występująca tu funkcja f jest nazywana **gęstością rozkładu prawdopodobieństwa** zmiennej losowej X, a jej wykres - **krzywą gęstości**

Własności funkcji gęstości

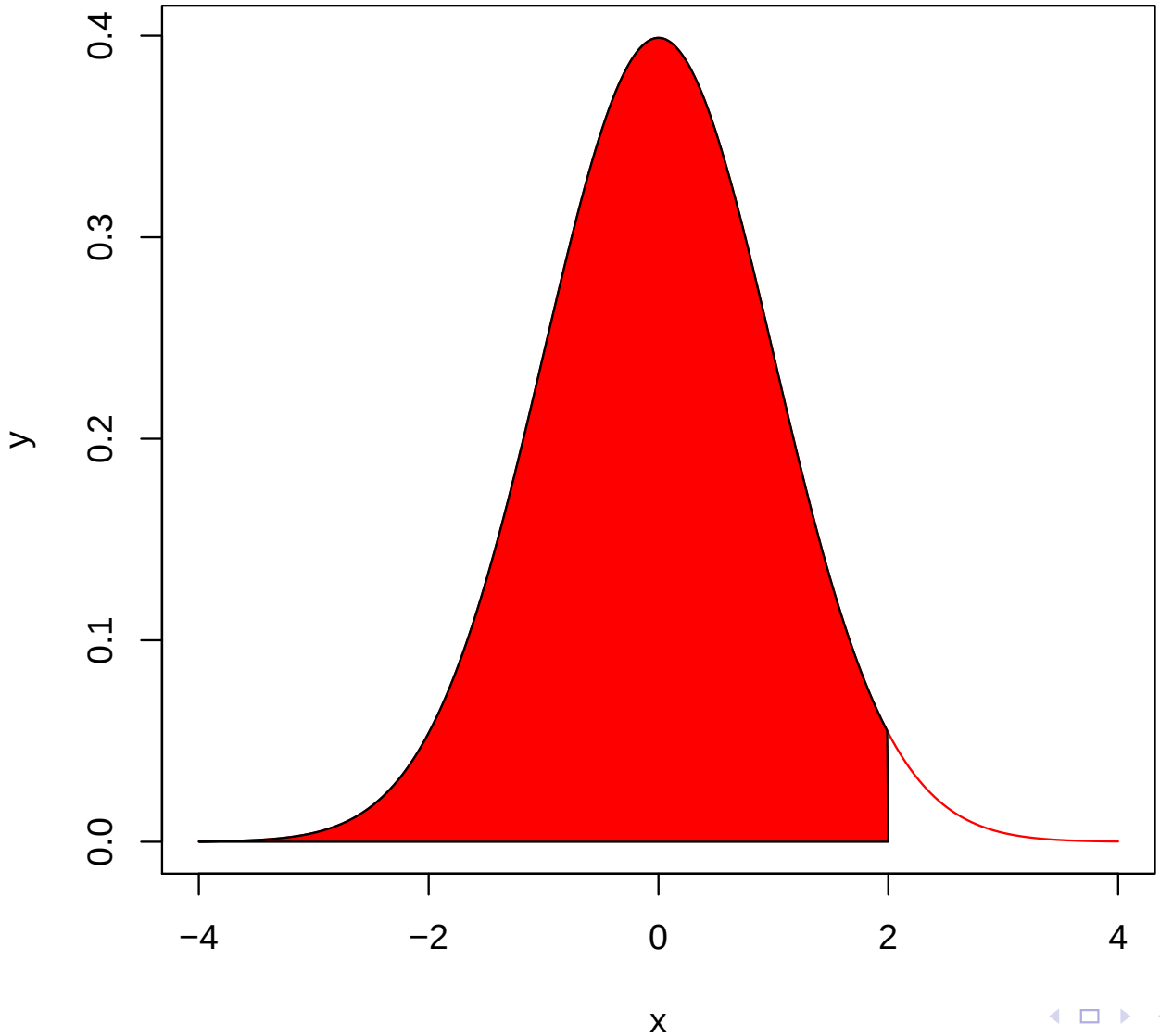
1. $f(x) \geqslant 0$ dla każdego x;
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Dystrybuanta zmiennej losowej typu ciągłego

Dystrybuanta F zmiennej losowej typu ciągłego wyraża się wzorem

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

Interpretacja graficzna dystrybuanty typu ciągłego



Zadanie

Niech gęstość f ciągłej zmiennej losowej X wynosi:

$$f(x) = \begin{cases} \frac{2}{3} + x^2 & 0 \leqslant x \leqslant 1 \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Znaleźć dystrybuantę F zmiennej losowej X oraz prawdopodobieństwo $P(X > 0.5)$.

Charakterystyki liczbowe zmiennych losowych

Dystrybuanta zmmiennej losowej daje pełny probabilistyczny opis, jednak z powodu zbytniej szczegółowości może on być mało czytelny. W praktyce wygodniej jest posługiwać się kilkoma charakterystykami liczbowymi. Do najważniejszych charakterystyk należą **miary położenia** i **miary rozrzutu**.

przykład miary położenia Wartość oczekiwana

Wartością oczekiwaną zmiennej losowej X typu dyskretnego o zbiorze punktów skoku $W = \{x_1, x_2, \ldots, x_n\}$ i skokach $p_i = P(X = x_i)$, nazywamy liczbę EX określoną wzorem

$$EX = \sum_{x_i \in W} x_i p_i$$

Przykład - trudny

Średnia arytmetyczna. Załóżmy, że zmienna losowa X ma skończony zbiór punktów skoku $W = \{x_1, x_2, \ldots, x_n\}$ oraz, że wszystkie skoki jej funkcji prawdopodobieństwa wynoszą $\frac{1}{n}$. Wówczas

$$EX = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Przykład - łatwy

Rzut kostką symetryczną

$$EX = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

Wartość oczekiwana

Wartością oczekiwaną zmiennej losowej X typu ciągłego o gęstości f nazywamy liczbę EX określoną wzorem

$$EX = \int_{-\infty}^{+\infty} xf(x)dx$$

Przykład

Niech f będzie funkcją gęstości zmiennej losowej określoną następującym wzorem:

$$f(x) = \begin{cases} 1 & 0 \leqslant x \leqslant 1 \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Wtedy

$$EX = \int_0^1 x dx = \frac{1}{2} x^2 /_0^1 = \frac{1}{2}$$

Uwagi

1. Jeżeli zmienna losowa $Y = g(X)$, to $EY = \sum_{x_i \in W} g(x_i)p_i$ $(EY = \int_{-\infty}^{+\infty} g(x)f(x)dx)$;
2. Istnieją zmienne losowe dla których wartość oczekiwana nie istnieje;

Własności wartości oczekiwanej

1. $E(C) = C$, gdzie C oznacza pewną stałą
2. $E(aX) = aE(X)$;
3. $E(aX + b) = aE(X) + b$;
4. $E(X + Y) = E(X) + E(Y)$;
5. $E(X \cdot Y) = E(X)E(Y)$, gdy X i Y są niezależne

Zadanie

W urnie są trzy czarne kule i jedna biała. Losujemy kolejno kule bez zwracania, aż do momentu wylosowania kuli białej. Niech X oznacza liczbę wyciągniętych kul. Oblicz EX.

Wariancja, odchylenie standardowe

Wariancją zmiennej losowej X o wartości oczekiwanej EX nazywamy liczbę $VarX$ (inne oznaczenia: D^2X,σ^2, σ_X^2,μ_2), określoną wzorem

$$VarX = E(X - EX)^2.$$

Odchyleniem standardowym zmiennej losowej X nazywamy liczbę DX (inne oznaczenia σ, σ_X), określoną wzorem

$$DX = \sqrt{VarX}.$$

Własności wariancji

1. $D^2(C) = 0$;
2. $D^2(aX) = a^2 D^2(X)$;
3. $D^2(X + b) = D^2(X)$;
4. $D^2(X) = E(X^2) + (E(X))^2$.

Zadanie

Zmienna losowa ma wartość oczekiwaną μ i wariancję σ^2 Obliczyć dla jakich wartości a i b zmienna losowa $aX + b$ ma wartość oczekiwaną równą zero i wariancję równą jeden?

Uwaga

Zmienną losową $Z = (X - \mu)/\sigma$, gdzie $EX = \mu$ i $VarX = \sigma$ nazywamy zmienną losową standaryzowaną.

Kwantyle

Kwantylem rzędu p $(0 < p < 1)$ zmiennej losowej X typu ciągłego o dystrybuancie F i gęstości f nazywamy każdą liczbę x_p, spełniającą którykolwiek z następujących równoważnych warunków

1. $F(x_p) = p$
2. $P(X < x_p) = p$
3. $\int_{-\infty}^{x_p} f(x)dx = p$

Przykłady rozkładów dyskretnych

Rozkład równomierny

Mówimy, że zmienna losowa X ma rozkład równomierny, jeżeli jej funkcja prawdopodobieństwa jest postaci

x_i	x_1	x_2	$\dots$	x_n
p_i	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$D^2(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))^2$$

Rozkład jednopunktowy

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

x_i	x_1
p_i	1

$$E(X) = x_1$$

$$D^2(X) = 0$$

Rozkład zerojedynkowy

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

x_i	0	1
p_i	q	p

$$E(X) = p$$

$$D^2(X) = pq$$

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

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

$$E(X) = np$$

$$D^2(X) = npq$$

Zadanie 10

Pewien matematyk nosi w kieszeniach (lewej i prawej) po jednym pudełku zapałek. Ilekroć chce zapalić papierosa sięga do wybranej losowo kieszeni. Jaka jest szansa, że gdy po raz pierwszy wyciągnie puste pudełko w drugim będzie k zapałek? $(k = 1, 2, \ldots, m)$ gdzie m jest liczbą zapałek w pełnym pudełku. Zakładamy, że w chwili początkowej matematyk ma dwa pełne pudełka.

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

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

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 ujemny dwumianowy

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

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

$$E(X) = \frac{k}{p}$$

$$D^2(X) = \frac{kp}{p^2}$$

Paradoks Petersburski

Piotr i Paweł grają w następującą grę: Paweł rzuca monetą do momentu gdy pojawi się pierwszy orzeł a Piotr wypłaca mu $X = 1$, gdy orzeł pojawia się w pierwszym rzucie, $X = 2$ gdy orzeł pojawia się w drugim rzucie i ogólnie $X = 2^{i-1}$ zł, gdy orzeł pojawi się dopiero w i- tym rzucie. Oblicz wartość średnią EX.

Rozkład Poissona

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

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

$$E(X) = \lambda$$

$$D^2(X) = \lambda$$

Rozkład ten 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.

Zadanie

Z miesięcznej obserwacji małego skrzyżowania w Kaliszu wynika, że między godziną 11 : 00 a 12 : 00 pojawiają się tam średnio 4 ciężarówki o ładowności powyżej 3.5 tony. Zakładając, że momenty ich pojawiania się w ustalonym dniu mogą być modelowane za pomocą procesu Poissona obliczmy prawdopodobieństwo, że między 11 : 00 a 11 : 30 nie pojawi się żadna ciężarówka.

Twierdzenie Poissona

Jeżeli liczba doświadczeń n w rozkładzie dwumianowym jest duża, wtedy obliczenie prawdopodobieństwa danej liczby sukcesów staje się kłopotliwa. Klasyczne twierdzenie Poissona dostarcza prostego przybliżenia, które ma rozsądną dokładność, gdy prawdopodobieństwo sukcesu p jest małe a iloczyn np umiarkowany.

Twierdzenie Poissona

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

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

Zadanie

Udowodnij twierdzenie Poissona

Twierdzenie Poissona zastosowania

1. Prawdopodobieństwo trafienia „szóstki ”w Toto-Lotku jest równe $1/\binom{49}{6} = 1/139883816$. Ilu szóstek należy się spodziewać w każdym tygodniu, jeśli grający wypełniają kupony niezależnie od siebie i całkowicie losowo, kuponów jest $n = 10^7$?
2. W Warszawie na Ursynowie ginie średnio 7 samochodów tygodniowo. Jaka jest szansa, że jutro będzie dzień bez kradzieży, przy założeniu stałej intensywności działania złodziei.