RACHUNEK PRAWDOPODOBIEŃSTWA

ZMIENNE LOSOWE

WIELOWYMIAROWE

Zmienne losowe wielowymiarowe

Rozważymy przypadek $n > 2$.

Załóżmy, że mamy n zmiennych losowych $X_1, X_2, \dots, X_n$ określonych na tej samej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, P)$.

Definicja.

Funkcję n-wymiarową $(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) : \Omega \to \mathcal{R}^n$ taką, że dla dowolnego $(x_1, x_2, \dots, x_n) \in \mathcal{R}^n$

$$\{\omega \in \Omega : X_1(\omega) \leqslant x_1, X_2(\omega) \leqslant x_2, \dots, X_n(\omega) \leqslant x_n\} \in \mathcal{F}$$

jest zdarzeniem losowym, nazywamy n-wymiarową zmienną losową.

Podobnie definiujemy dystrybuantę n-wymiarową

Definicja.

Funkcję n-wymiarową $F : \mathcal{R}^n \to [0,1]$ *zdefiniowaną*

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \dots, X_n \leqslant x_n)$$

dla dowolnych $(x_1, x_2, \dots, x_n) \in \mathcal{R}^n$ *nazywamy n-wymiarową dystrybuantą łączną zmiennej losowej* $(X_1, X_2, \dots, X_n)$.

Definicja.

Jeżeli zmienna losowa $(X_1, X_2, \ldots, X_n)$ *przyjmuje co najwyżej przeliczalną liczbę wartości, to zmienną nazywamy typu dyskretnego oraz funkcję*

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n),$$

funkcją prawdopodobieństwa rozkładu łącznego.

Własności funkcji prawdopodobieństwa

1. $p(x_1, \ldots, x_n) \geqslant 0, \ \forall (x_1, \ldots, x_n) \in R^n.$
2. $\sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_n \in \mathcal{X}_n} p(x_1, \ldots, x_n) = 1.$

Definicja.

Zmienną losową n-wymiarową $(X_1, X_2, \ldots, X_n)$ *nazywamy zmienną losową typu ciągłego, jeżeli istnieje taka nieujemna funkcja* $f(x_1, x_2, \ldots, x_n) \geqslant 0$, *że dla dowolnego zbioru* $A \subset \mathcal{R}^n$

$$P\{(X_1, X_2, \ldots, X_n) \in A\} = \int\limits_A f(x_1, x_2, \ldots, x_n) \mathrm{d}x_1 \, \mathrm{d}x_2 \cdots \mathrm{d}x_n.$$

W szczególności dla $(x_1, \ldots, x_n) \in R^n$ otrzymujemy postać dystrybuanty $F(x_1, \ldots, x_n)$

$$\begin{aligned} F(x_1, \ldots, x_n) &= P\{(X_1, \ldots, X_n) \in (-\infty, x_1] \times \cdots \times (-\infty, x_n]\} = \\ &= \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} f(u_1, \cdots u_n) \mathrm{d}u_1 \cdots \mathrm{d}u_n. \end{aligned}$$

Podobnie jak w przypadku dwuwymiarowym otrzymujemy

$$f(x_1,\ldots,x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1,\ldots,x_n),$$

dla wszystkich punktów, w których pochodna istnieje.

Własności gęstości wielowymiarowej

1. $f(x_1,\ldots,x_n) \geqslant 0, \ \forall (x_1,\ldots,x_n) \in R^n.$
2. $\int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(x_1,\ldots,x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n = 1.$

Jeżeli zmienna losowa $(X_1, X_2, \ldots, X_n)$ jest typu ciągłego, to wszystkie rozkłady brzegowe są również typu ciągłego. W szczególności, jednowymiarowe rozkłady brzegowe zmiennych losowych X_i, $1 \leqslant i \leqslant n$ otrzymujemy w następujący sposób:

$$f_{X_i}(x) = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n.$$

Podobnie jak poprzednio definiujemy niezależność $X_1, X_2, \ldots, X_n$.

Definicja.

Mówimy, że zmienne losowe $X_1, X_2, \ldots, X_n$ *są niezależne, jeżeli dla dowolnych* $x_1, x_2, \ldots, x_n \in \mathcal{R}$,

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \text{ dla dowolnego } (x_1, \ldots, x_n) \in \mathcal{R}^n,$$

gdzie $F_{X_i}(x_i) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$, $i = 1, \ldots, n$ są dystrybuantami brzegowymi jednowymiarowymi.

Innymi słowy, zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne jeśli dla dowolnego podzbioru $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ zmiennych losowych (każda para, trójka itd.), ma rozkład łączny, który jest produktem rozkładów brzegowych.

W szczególności dla rozkładów dyskretnych mamy równoważny warunek, funkcja prawdopodobieństwa zmiennej losowej $X_1, X_2, \ldots, X_n$ jest produktem jednowymiarowych funkcji prawdopodobieństwa

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i)$$

Dla zmiennych losowych typu ciągłego mamy podobny warunek dla gęstości

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Twierdzenie.
Jeżeli $X_1, X_2, \ldots, X_n$ są zmiennymi losowymi, dla których istnieją kowariancje $Cov(X_i, X_j)$ dla dowolnych i i j, oraz $a_1, a_2, \ldots, a_n$ są dowolnymi rzeczywistymi stałymi, to

$$E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$$

oraz

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} Cov(X_i, X_j).$$

Jeżeli dodatkowo założymy, że $X_1, X_2, \ldots, X_n$ są parami niezależne, to

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i).$$

Przykład

Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o jednakowych rozkładach z wartością oczekiwaną $E(X_i) = \mu$ oraz wariancją $V(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$.

Zdefiniujmy średnią arytmetyczną zmiennych losowych $X_1, X_2, \ldots, X_n$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Wówczas

$$E(\bar{X}) = \mu \quad \text{oraz} \quad V(\bar{X}) = \frac{\sigma^2}{n}.$$

Jest to prosta konsekwencja poprzedniego twierdzenia.

Z punktu widzenia modeli probabilistycznych oraz statystycznych niezwykle ważną jest umiejętność wyznaczania rozkładów sum niezależnych zmiennych losowych o jednakowych rozkładach. Podamy teraz bez dowodu kilka twierdzeń o dodawaniu dla wcześniej poznanych rozkładów.

Twierdzenie o dodawaniu dla niezależnych zmiennych losowych o jednakowych rozkładach Bernoulliego $B(p)$.
Jeżeli $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o jednakowych rozkładach Bernoulliego $B(p)$, gdzie $p \in (0, 1)$, to zmienna losowa

$$S_n = \sum_{i=1}^{n} X_i \sim Bin(n, p)$$

ma rozkład dwumianowy z parametrami n i p.

Z tego twierdzenia wynika, że dowolna zmienna losowa o rozkładzie dwumianowym $Bin(n, p)$ może być przedstawiona jako suma n niezależnych zmiennych losowych o jednakowych rozkładach Bernoulliego $B(p)$.

Prawdziwe jest zatem następujące twierdzenie

Twierdzenie o dodawaniu rozkładów dwumianowych.

Jeżeli $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o rozkładach dwumianowych $X_i \sim Bin(n_i, p)$ z tym samym prawdopodobieństwem sukcesu $p \in (0, 1)$, to

$$S_n = \sum_{i=1}^{n} X_i \sim Bin\left(\sum_{i=1}^{n} n_i, p\right)$$

ma rozkład dwumianowy z parametrami $\sum_{i=1}^{n} n_i$ i p.

Twierdzenie o dodawaniu rozkładów Poissona.

Jeżeli $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o rozkładach Poissona $P(\lambda_i)$, gdzie $\lambda_i > 0$, dla $i = 1, 2, \ldots, n$, to

$$S_n = \sum_{i=1}^{n} X_i \sim P\left(\sum_{i=1}^{n} \lambda_i\right)$$

ma rozkład Poissona z parametrem $\lambda = \sum\limits_{i=1}^{n} \lambda_i$.

Przykład. *W wyniku przeprowadzonych badań, ustalono, że w pewnej fabryce wytwarzającej sprzęt RTV, liczba wyprodukowanych w ciągu dnia niesprawnych odbiorników telewizyjnych ma rozkład Poissona o średniej 2. Podobnie, liczba niesprawnych zestawów stereo ma również rozkład Poissona o wartości oczekiwanej 3. Wyznaczyć prawdopodobieństwo zdarzenia, że w ciągu 2 dni, liczba niesprawnych produktów nie przekroczy 5.*

Rozwiązanie.

Niech X_1 i X_2 oznaczają liczbę niesprawnych zestawów stereofonicznych wyprodukowanych odpowiednio pierwszego i drugiego dnia.

Podobnie, niech X_3 i X_4 będą liczbami niesprawnych odbiorników TV, wyprodukowanych pierwszego i drugiego dnia, odpowiednio.

Załóżmy, że wszystkie zmienne są niezależne.

Wówczas

$$S = X_1 + X_2 + X_3 + X_4$$

ma rozkład Poissona z parametrem $\lambda = 2 + 2 + 3 + 3 = 10$.
Otrzymujemy zatem

$$P(S \leqslant 5) = F(5; 10) = 0,067086.$$

Twierdzenie o dodawaniu dla zmiennych normalnych.
Jeżeli $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o rozkładach normalnych $N(\mu_i, \sigma_i)$, odpowiednio,
gdzie $\mu_i \in \mathcal{R}$, $\sigma_i > 0$, $i = 1, 2, \ldots, n$, to zmienna losowa

$$S_n = \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$$

ma rozkład normalny z parametrami

$$\mu = \sum_{i=1}^{n} \mu_i \quad \text{oraz} \quad \sigma = \sqrt{\sum_{i=1}^{n} \sigma_i^2}.$$