Rozkłady prawdopodobieństwa

Rozkładem prawdopodobieństwa zmiennej losowej X o wartościach w $\mathbb{R}^n$ nazywamy rozkład prawdopodobieństwa μ_X określony na $\mathcal{B}(\mathbb{R}^n)$ zależnością

$$\mu_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Zdefiniowana w ten sposób miara zawiera wszelkie interesujące z punktu widzenia probabilistycznego informacje o zmiennej losowej X: o jej zbiorze wartości i tym, jak prawdopodobieństwo jest na tym zbiorze rozłożone. Dokładniej, otrzymaliśmy nową przestrzeń probabilistyczną $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_X)$. Warto jeszcze raz podkreślić, że rozkład prawdopodobieństwa nie musi być kojarzony z konkretną zmienną losową — rozkładem prawdopodobieństwa nazwaliśmy dowolną unormowaną miarę μ na $\mathcal{B}(\mathbb{R}^n)$.

I uwaga dotycząca oznaczeń. $P(X^{-1}(B))$ można zapisywać tak:

$$P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}) = P(X \in B).$$

Ostatniej skrótowej notacji będziemy używać najczęściej. Fakt, że zmienna losowa X ma rozkład prawdopodobieństwa μ, będziemy zapisywać krótko $X \sim \mu$.

Przypomnijmy sobie teraz wcześniejszy przykład. Była tam mowa o czasie oczekiwania na autobus. Ustaliliśmy, że

$$\Omega = [0, \infty) \quad \text{i} \quad P(A) = \int_A f(x)\, dx, \quad A \in \mathcal{F}.$$

Przyjmując $\mathcal{F} = \mathcal{B}(\mathbb{R})$, otrzymamy sensowny model, jeśli tylko f jest mierzalna względem $\mathcal{B}(\mathbb{R})$. Zmienną losową X będzie sam czas oczekiwania, zatem $X(\omega) = \omega$, $\omega \in [0, \infty)$. W takim razie

$$\mu_X(A) = \int_A f(x)\, dx, \quad A \in \mathcal{B}(\mathbb{R}).$$

Powyższa zależność definiuje *gęstość rozkładu prawdopodobieństwa* (albo krócej *gęstość prawdopodobieństwa*). Ponieważ umiemy całkować względem miary Lebesgue'a po $\mathbb{R}^n$, warto przyjąć definicję ogólniejszą.

Jeśli μ jest rozkładem prawdopodobieństwa na $\mathbb{R}^n$ i dla pewnej funkcji $f : \mathbb{R}^n \to \mathbb{R}$ całkowalnej w sensie Lebesgue'a mamy

$$\mu(A) = \int_A f(x)\, dx, \quad A \in \mathcal{B}(\mathbb{R}^n),$$

to f nazywamy *gęstością rozkładu*.

Funkcję f przyjęto nazywać gęstością, gdyż jest ona odpowiednikiem gęstości masy w fizyce. Jeśli $f : \mathbb{R} \to \mathbb{R}$, to w punktach ciągłości funkcji f mamy

$$f(x) = \lim_{n\to\infty} \frac{\int_{-\infty}^{x+h} f(s)\,ds - \int_{-\infty}^{x} f(s)\,ds}{h} = \lim_{h\to 0} \frac{\mu([x, x+h))}{h}.$$

Rozkład, który ma gęstość, będziemy nazywać *rozkładem ciągłym.*

Zajmiemy się teraz rozkładami z przeciwnego bieguna — rozkładami dyskretnymi. Niech X będzie wynikiem rzutu symetryczną kostką. Wtedy

$$\mu_X(\{i\}) = P(X = i) = \frac{1}{6}, \quad \text{gdy } i = 1, 2, \ldots, 6.$$

Rozkład μ_X jest skupiony na zbiorze $\{1, 2, \ldots, 6\}$, czyli

$$\mu_X(A) = \mu_X(A \cap \{1, 2, \ldots, 6\}) = \sum_{i=1}^{6} \frac{1}{6} \mathbb{1}_A(i).$$

Rozkład μ na $\mathbb{R}^n$ nazywamy *dyskretnym*, jeśli istnieje zbiór przeliczalny $S \subset \mathbb{R}^n$, dla którego $\mu(S) = 1$.

Źródło:

JACEK JAKUBOWSKI, RAFAŁ SZTENCEL: *Wstęp do teorii prawdopodobieństwa. Wydanie IV.* Warszawa, Script, 2010, str. 63–65.