1 Prawdopodobieństwo

Odcinek na prostej ma jednoznacznie rozumianą długość. Odwołując się do intuicji można bez trudu pojęcie długości rozszerzyć na zbiory będące rezultatem operacji sumowania i mnożenia wykonanych na odcinkach skończenie wiele razy. Można pójść jeszcze dalej i je uogólnić na sytuację przeliczalnej liczby tych operacji. Mówimy jednak wtedy nie o długości otrzymanych w ten sposób zbiorów, lecz o ich mierze.

Dla nas wyjściowe odcinki zawarte są w zbiorze $\Omega = [0,1]$, zbiory otrzymane w wyniku przeprowadzenia wspomnianych operacji są zdarzeniami, a ich miara jest prawdopodobieństwem. Jeśli zatem zdarzenie, czyli nasz zbiór zawarty w Ω, oznaczymy jako A, to jego miarę, czyli prawdopodobieństwo będziemy zapisywać jako $P\{A\}$. Jest oczywiste, że $0 \le P\{A\} \le 1$. Elementy zbioru Ω będziemy oznaczali jako ω.

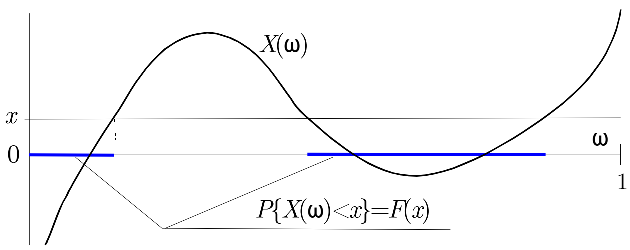
Warto przypomnieć, że

- $P\{A \cup B\} = P\{A\} + P\{B\}$, jeśli zdarzenia A i B są rozłączne,
- $P\{A \cap B\} = P\{A|B\}P\{B\}$,
- jeśli zdarzenia B_i są rozłączne i $\sum_i B_i = 1$, to $P\{A\} = \sum_i P\{A|B_i\}P\{B_i\}$, które to wyrażenie nazywa się wzorem Bayesa.

2 Zmienna losowa

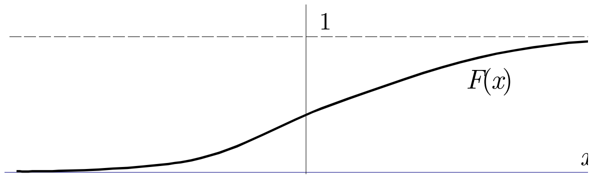
Zmienną losową $X(\omega)$ nazywamy funkcję określoną na zbiorze $\Omega = [0,1]$, rys 2.1. Dla prostoty będziemy ją zwykle oznaczać krótko jako X. Jej dystrybuanta $F(x)$, to

$$F(x) = P\{X < x\}.$$



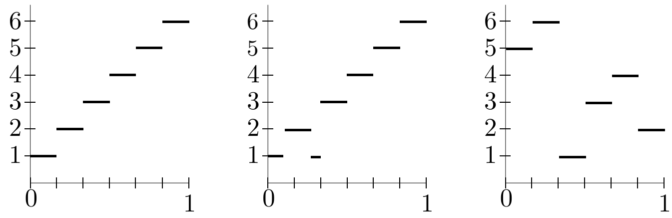
Rys. 2.1: Przykład zmiennej losowej $X(\omega)$.

Jest oczywiste, że $0 \le F(x) \le 1$, F jest funkcją niemalejącą oraz $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, patrz rys. 2.2.

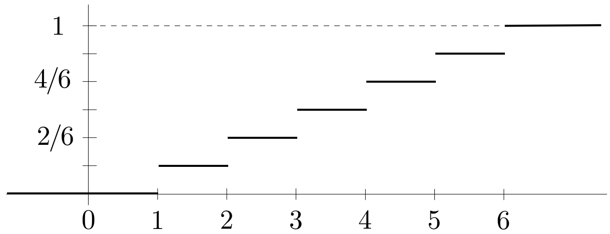


Rys. 2.2: Przykład dystrybuanty $F(x)$.

Przykład 2.1 *Na rys. 2.3 przedstawione są trzy zmienne losowe dotyczące rzutu kostką. Dla każdej z nich bowiem $P\{X = 1\} = \cdots = P\{X = 6\} = 1/6$. Rys. 2.4 przedstawia ich wspólną dystrybuantę.*



Rys. 2.3: Rzut kostką, przykładowe zmienne losowe.



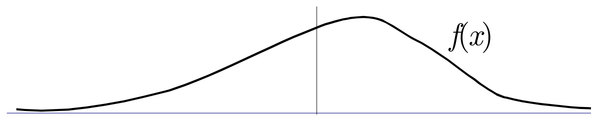
Rys. 2.4: Rzut kostką, dystrybuanta.

Gęstość prawdopodobieństwa $f(x)$, patrz rys. 2.5,

$$f(x) = \frac{d}{dx}F(x)$$

jest funkcją nieujemną i ponadto

$$\int_{-\infty}^{x} f(\xi)d\xi = F(x), \qquad \int_{-\infty}^{\infty} f(x)dx = 1.$$



Rys. 2.5: Przykładowa gęstość prawdopodobieństwa $f(x)$.

Jest więc oczywiste, że dla odcinka (a,b) ma miejsce równość $P\{X \in (a,b)\} = P\{X < b\} - P\{X < a\} = F(b) - F(a) = \int_a^b f(x)dx$. Na tej samej zasadzie

$$P\{X \in A\} = \int_A f(x)dx,$$

dla każdego zbioru A.

Wartość średnia zmiennej losowej X, to naturalnie

$$EX = \int_0^1 X(\omega)d\omega. \tag{2.1}$$

Aby ją wyliczyć, ustalmy punkty x_i na całej prostej w ten sposób, że $\ldots < x_{-2} < x_{-1} < x_0 < x_1 < x_2, \ldots$, przy czym $x_{i+1} - x_i = \Delta$. Dla małego Δ zatem

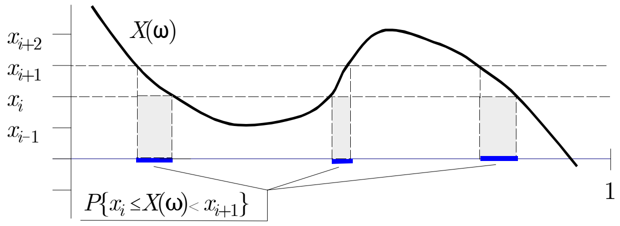
$$\begin{aligned}\int_0^1 X(\omega)d\omega &\approx \sum_i x_i P\{x_i \le X(\omega) < x_{i+1}\} \\ &= \sum_i x_i \int_{x_1}^{x_{i+1}} f(x)dx \approx \sum_i \int_{x_1}^{x_{i+1}} xf(x)dx \\ &= \int_{-\infty}^{\infty} xf(x)dx,\end{aligned}$$

patrz rys. 2.6. Pokazaliśmy w ten sposób, że

$$EX = \int xf(x)dx,$$

która to wielkość nazywa się wartością oczekiwaną. Jest oczywiste, że

$$E\{aX + bY\} = aEX + bEY.$$



Rys. 2.6: Suma pól zacienionych prostokątów, to $x_i P\{x_i \le X(\omega) < x_{i+1}\}$.

Momentem rzędu n nazywa się

$$EX^n = \int x^n f(x)dx.$$

Wariancję, oznaczaną jako $\operatorname{var} X$, bądź σ_X^2, bądź σ^2 w sytuacji jednoznacznej, definiuje się jako

$$\sigma^2 = \operatorname{var} X = E(X - EX)^2 = EX^2 - E^2X.$$

Warto przytoczyć w tym miejscu nierówność Schwarza

$$E\,|XY| \le \sqrt{EX^2EY^2}, \tag{2.2}$$

z której wynika, że $E^2|X| \le EX^2$.

Nierówność Czebyszewa to

$$P\{|X| > \varepsilon\} \le \frac{1}{\varepsilon}E|X|, \tag{2.3}$$

dla każdego $\varepsilon > 0$. Jeśli zatem $EX = 0$, to $EX^2 = \sigma^2$, skąd wynika

$$P\{|X| > \varepsilon\} \le \frac{1}{\varepsilon}E|X| \le \frac{1}{\varepsilon}\sqrt{EX^2} \le \frac{1}{\varepsilon}\sigma.$$

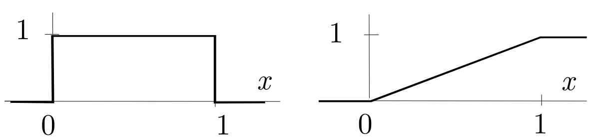
3 Rozkłady prawdopodobieństwa

3.1 Rozkład równomierny

Rozkład równomierny ma gęstość, patrz rys. 3.1, i dystrybuantę jak poniżej:

$$f(x) = \begin{cases} 1, & \text{dla } 0 \le x \le 1 \\ 0, & \text{dla pozostałych } x, \end{cases}$$

$$F(x) = \begin{cases} 0, & \text{dla } x < 0 \\ x, & \text{dla } 0 \le x \le 1 \\ 1, & \text{dla } 1 < x. \end{cases}$$

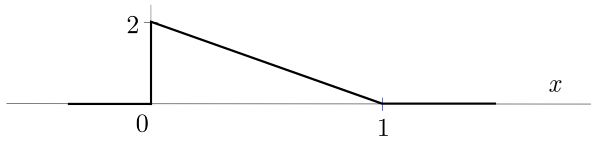


Rys. 3.1: Gęstość i dystrubuanta rozkładu równomiernego.

3.2 Rozkład trójkątny

Rozkład trójkątny ma gęstość, rys. 3.2,

$$f(x) = \begin{cases} 0, & \text{dla } x < 0 \\ 2(1-x), & \text{dla } 0 \le x \le 1 \\ 0, & \text{dla } 1 < x. \end{cases}$$



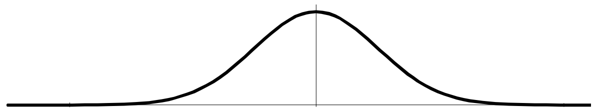
Rys. 3.2: Gęstość rozkładu trójkątnego.

3.3 Rozkład normalny

Rozkład normalny, rys. 3.3, czyli Gaussa, ma gęstość

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

gdzie $\mu = EX$ oraz $\sigma = \operatorname{var} X$. Dystrybuanty nie można wyrazić w postaci analitycznej.



Rys. 3.3: Gęstość rozkładu normalnego, $\mu = 0$.

3.4 Rozkład wykładniczy

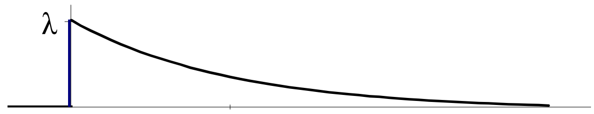
Gęstość rozkładu wykładniczego, to:

$$f(x) = \begin{cases} 0, & \text{dla } x < 0 \\ \lambda e^{-\lambda x}, & \text{dla } 0 \le x, \end{cases}$$

Jej dystrybuantą jest

$$F(x) = \begin{cases} 0, & \text{dla } x < 0 \\ 1 - e^{-\lambda x}, & \text{dla } 0 \le x. \end{cases}$$

Dla rozkładu tego $EX = 1/\lambda$ oraz $\operatorname{var} X = 1/\lambda^2$.



Rys. 3.4: Gęstość rozkładu wykładniczego.

3.5 Rozkład Laplace'a

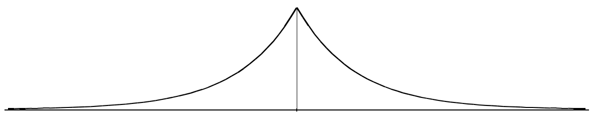
Gęstość rozkładu Laplace'a dana jest wzorem:

$$f(x) = \frac{1}{2}e^{-|x|},$$

patrz rys. 3.5. Jej dystrybuantą jest

$$F(x) = \begin{cases} \frac{1}{2}e^{x} & \text{dla } x < 0 \\ 1 - \frac{1}{2}e^{-x}, & \text{dla } x \ge 0. \end{cases}$$

Dla rozkładu tego $EX = 0$ oraz $\operatorname{var} X = 2$.



Rys. 3.5: Gęstość rozkładu Laplace'a.

3.6 Rozkład Cauchy'ego

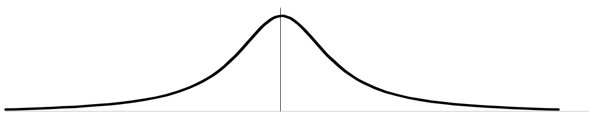
Jeśli zmienna losowa X ma gęstość

$$f(x) = \frac{\sigma}{\pi(x^2 + \sigma^2)},$$

patrz rys. 3.6, to mówimy, że ma ona rozkład Cauchy'ego. Jej dystrybuantą jest

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{\sigma}\right).$$

Rozkład ten nie ma żadnego skończonego momentu, tzn. $EX^n = \infty$, $n = 1, 2, \ldots$.



Rys. 3.6: Gęstość rozkładu Cauchy'ego.

Jeśli niezależne zmienne losowe X i Y mają rozkład normalny, to X/Y ma rozkład Cauchy'ego.

3.7 Rozkład logistyczny

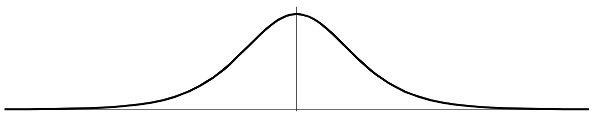
Jeśli zmienna losowa X ma gęstość

$$f(x) = \frac{1}{2 + e^x + e^{-x}} = \frac{e^{-x}}{(1 + e^{-x})^2},$$

rys. 3.7, to mówimy, że ma ona rozkład logistyczny. Jej dystrybuantą jest

$$F(x) = \frac{1}{1 + e^{-x}}.$$

Dla rozkładu tego $EX = 0$, $\operatorname{var} X = \pi^2/3$. Rozkład ten służy do opisywania zjawisk zużywania się części, czy procesów, w których zachodzi efekt nasycenia.



Rys. 3.7: Gęstość rozkładu logistycznego.

3.8 Rozkład Pareto

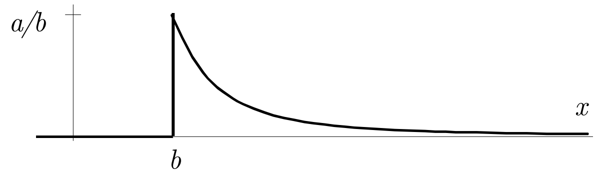
Jeśli zmienna losowa X ma gęstość, patrz rys. 3.8,

$$f(x) = \begin{cases} 0, & \text{dla } x < b \\ \dfrac{ab^a}{x^{a+1}}, & \text{dla } b \le x, \end{cases}$$

gdzie $b > 0$, to mówimy, że ma ona rozkład Pareto. Jej dystrybuantą jest

$$F(x) = \begin{cases} 0, & \text{dla } x < b \\ 1 - \left(\dfrac{b}{x}\right)^a, & \text{dla } b \le x. \end{cases}$$

Dla rozkładu tego $EX = \infty$ dla $a < 1$ oraz $ab/(a-1)$ dla $1 \le a$. Ponadto, dla $1 \le a$, $\operatorname{var} X = ab^2/(a-2)(a-1)^2$. Rozkład Pareto stosowany jest do przedstawienia np. rozkładu dochodów w społeczeństwie.



Rys. 3.8: Gęstość rozkładu Pareto.

3.9 Rozkłady dyskretne

3.9.1 Rozkład zero-jedynkowy

O rozkładzie zero-jedynkowym mówimy, jeśli zmienna losowa przyjmuje dwie wartości: 0 lub 1. Dla przykładu, dla dowolnej zmiennej losowej X i dowolnego zdarzenia A zdefiniujemy zmienną losową

$$\xi = \begin{cases} 1, & \text{jeśli } X \in A \\ 0, & \text{jeśli } X \notin A. \end{cases}$$

Jeśli $X \in A$, tzn. $\xi = 1$, to mówimy o sukcesie. Ma ona rozkład zero–jedynkowy, w którym $P\{\xi = 1\} = P\{X \in A\}$ i jest prawdopodobieństwem sukcesu.

3.9.2 Rozkład dwumianowy

Procesem Bernouliego nazywa się ciąg niezależnych zmiennych losowych $\xi, \xi_1, \xi_2, \xi_3, \cdots$, z których każda ma rozkład jak w § 3.9.1. Zmienna

$$S_n = \xi_1 + \xi_2 + \cdots + \xi_n$$

ma rozkład nazywany dwumianowym. Jest ona liczbą sukcesów w procesie Bernouliego o długości n. Można wykazać, że

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

gdzie $p = P\{\xi = 1\}$ jest prawdopodobieństwem sukcesu.

3.9.3 Rozkład geometryczny

Zmienna losowa X przyjmująca wartości ze zbioru $\{1, 2, \cdots\}$ ma rozkład geometryczny z parametrem p, jeśli

$$P\{X = k\} = (1-p)^{k-1} p.$$

Jak można wyliczyć $EX = 1/p$. Powyższym wzorem wyraża się prawdopodobieństwo tego, że pierwszy sukces w procesie Bernouliego nastąpi w próbie k.

4 Korelacja, niezależność, regresja

Dla pary zmiennych losowych (X, Y) dystrubuantą jest $F(x, y) = P\{X < x, Y < y\}$. Łączna gęstość tej pary to

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

Kowariancja pomiędzy nimi to:

$$\operatorname{cov}(X,Y) = E(X - EX)(Y - EY).$$

Współczynnik korelacji definiuje się jako:

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Ponieważ z nierówności Schwartza (2.2) wynika, że

$$|\operatorname{cov}(X,Y)| \le \sqrt{E(X-EX)^2}\sqrt{E(Y-EY)^2} = \sigma_X\sigma_Y,$$

zatem $-1 \le \rho \le 1$. Jeśli $\operatorname{cov}(X,Y) = 0$, to także $\rho = 0$. W sytuacji takiej mówimy, że zmienne losowe X i Y nie są skorelowane.

Jeśli

$$f(x,y) = f_x(x)f_y(y).$$

gdzie f_x i f_y są odpowiednio gęstościami zmiennych X i Y, to mówimy, że zmienne te są niezależne. Dla takich zmiennych $\operatorname{cov}(X,Y) = \rho = 0$. Jeśli zatem zmienne losowe są niezależne, to są także nieskorelowane. Ponadto

$$E\{\varphi(X)\psi(Y)\} = E\{\varphi(X)\}E\{\psi(Y)\}.$$

dla dowolnych funkcji φ, ψ oraz

$$\operatorname{var}(X+Y) = \operatorname{var}X + \operatorname{var}Y. \tag{4.1}$$

Funkcja

$$f(y|x) = \frac{f(x,y)}{f_x(x)}$$

nazywa się warunkową gęstością Y przy warunku $X = x$, natomiast

$$E\{Y|X=x\} = \int y f(y|x)dy$$

jest warunkową wartością oczekiwaną Y przy tym samym warunku, czyli funkcją regresji. Zatem

$$EY = \int E\{Y|X=x\}f_x(x)dx.$$

Jeśli X jest natury dyskretnej i przyjmuje wartości $x_1, x_2, \ldots, x_n$, to

$$f(y) = \sum_{i=1}^{n} p_i f_i(y),$$

gdzie $p_i = P\{X = x_i\}$, natomiast $f_i(y)$ jest warunkową gęstością Y pod warunkiem, że $X = x_i$. W sytuacji takiej

$$E\{Y|X=x_i\} = \int y f_i(y)dy$$

oraz

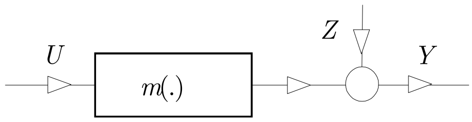
$$EY = \sum_{i=1}^{n} p_i E\{Y|X=x_i\}.$$

Uwaga 4.1 *Z istnienia korelacji pomiędzy zmiennymi nie wynika, że jedna z nich jest przyczyną, a druga skutkiem.*

Ćwiczenie 4.1 *Jeśli $Y = aX$ i $EX = 0$, to $\operatorname{cov}(X,Y) = a\sigma_X^2$, skąd wynika, że $\rho = a/|a| = \operatorname{sign}(a)$ (z uwagi na to, że $\sigma > 0$). Oznacza to, że $\rho = \pm 1$ w zależności od znaku a.*

Ćwiczenie 4.2 *Sprawdzić prawdziwość wzoru (4.1).*

Przykład 4.1 *W systemie statycznym, rys. 4.1, $Y = m(U) + Z$. Sygnały U oraz Z są niezależne. Ponadto $EZ = 0$. Zatem $m(u) = E\{Y|U=u\}$, co oznacza, że wykrywanie charakterystyki m na podstawie obserwacji wejścia U i wyjścia Y jest równoznaczne estymacji funkcji regresji.*



Rys. 4.1: System statyczny o charakterystyce $m(x)$.

5 Zbieżność probabilistyczna

Niech teraz

$$X, X_1, X_2, X_3, \ldots \tag{5.1}$$

będzie ciągiem zmiennych losowych. Probabilistyka operuje trzema rodzajami zbieżności: według średniej, według prawdopodobieństwa i z prawdopodobieństwem 1. Omówimy je kolejno.

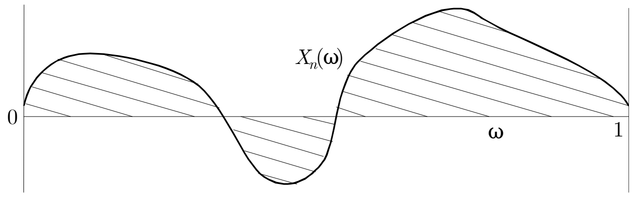
Jeśli

$$\lim_{n\to\infty} E\,|X_n(\omega) - g| = 0,$$

to mówimy, że ciąg jest zbieżny do granicy g według średniej pierwszego rzędu, patrz rys. 5.1, jeśli

$$\lim_{n\to\infty} E\,(X_n(\omega) - g)^2 = 0,$$

to według średniej drugiego rzędu, czyli w sensie średniokwadratowym.

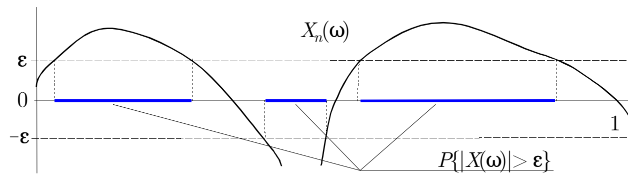


Rys. 5.1: Zbieżność do zera według średniej pierwszego rzędu. Powierzchnia zakreskowanego pola maleje do zera, gdy $n \to \infty$.

Jeśli, dla każdego $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|X_n(\omega) - g| > \varepsilon\} = 0,$$

to mówimy, że ciąg jest zbieżny do granicy g według prawdopodobieństwa, rys. 5.2.



Rys. 5.2: Zbieżność do zera według prawdopodobieństwa. Miara zbioru (prawdopodobieństwo zdarzenia) zaznaczonego grubą linią (składającego się z punktów ω, w których $|X(\omega)| > \varepsilon$) maleje do zera, gdy $n \to \infty$.

O zbieżności z prawdopodobieństwem 1 mówimy, gdy

$$P\{\lim_{n\to\infty} X_n(\omega) = g\} = 1.$$

Dla pewnych ω ciąg liczbowy $X(\omega)$ jest zbieżny go g, dla innych nie. Zbieżność z prawdopodobieństwem 1 oznacza, że zbiór punktów ω (czyli zdarzenie), w których zbieżność ta zachodzi ma prawdopodobieństwo 1.

Posumowując, w zbieżności według średniej chodzi o pole powierzchni pod zmienną losową $X_n(\omega)$, które maleje do zera. W zbieżności według prawdopodobieństwa o prawdopodobieństwo (czyli miarę) zbioru punktów ω, dla których $|X_n(\omega) - g| > \varepsilon$, które maleje do zera. W zbieżności z prawdopodobieństwem 1 chodzi natomiast o zbieżność ciągu liczbowego $X_n(\omega)$, o miarę (prawdopodobieństwo) zbioru punktów ω, w których ciąg ten zbiega się do g, czyli o miarę punktów, w których zachodzi $\lim_{n\to\infty} X_n(\omega) = g$.

Na mocy nierówności Czebyszewa (2.3), zbieżność według średniej implikuje zbieżność według prawdopodobieństwa, bowiem:

$$P\{|X_n - g| > \varepsilon\} \le \frac{1}{\varepsilon}\sqrt{E(X_n - g)^2}.$$

Jeśli natomiast $X_n \to a$ i $Y_n \to b$ według prawdopodobieństwa (z prawdopodobieństwem 1), to $X_n/Y_n \to a/b$ według prawdopodobieństwa (z prawdopodobieństwem 1).

6 Prawa wielkich liczb

Załóżmy teraz że w ciągu (5.1) zmienne losowe są niezależne i mają jednakowy rozkład. Niech teraz

$$S_n = X_1 + X_2 + \cdots + X_n$$

Jest rzeczą oczywistą, że $E\{S_n/n\} = EX$. Zbieżność S_n/n do EX jest przedmiotem tzw. praw wielkich liczb. Czynią one różne założenia dotyczące zmiennej losowej X. My poprzestaniemy na spostrzeżeniu, że

$$\mathrm{var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\,\mathrm{var}(X).$$

Wynika stąd, że S_n/n zbiega się do granicy EX w sensie średniokwadratowym, tzn.

$$\lim_{n\to\infty} E\left(\frac{S_n}{n} - EX\right)^2 = 0, \tag{6.1}$$

bowiem $\lim_{n\to\infty} \mathrm{var}[S_n/n] = 0$. Można powiedzieć, że średnia empiryczna S_n/n zbiega się do średniej teoretycznej EX.

Prawa wielkich liczb są fundamentem metod statystyki matematycznej i wszelkiej działalności polegającej na analizie danych pomiarowych. Dzięki nim możemy skutecznie wnioskować na podstawie danych eksperymentalnych.

Przykład 6.1 *Dla procesu Bernouliego jak w § 3.9.2*

$$\lim_{n\to\infty} E\left(\frac{S_n}{n} - p\right)^2 = 0,$$

co oznacza, że empiryczna częstość sukcesów S_n/n zbiega się do p. Może ona zatem służyć jako estymator prawdopodobieństwa sukcesu p.