Uwaga 3.1.3 *Z powyższej definicji funkcji F wynika, że w każdym punkcie x_n F ma nieciągłość typu skok. Wartość tego skoku w punkcie x_n wynosi*

$$\lim_{x \to x_n^+} F(x) - F(x_n) = p_n.$$

Uwaga 3.1.4 *Funkcję, o której mowa w twierdzeniu 3.1.1 nazywamy dystrybuantą rozkładu prawdopodobieństwa d.*

Przykład 3.1.2 *Dla rozważanego wyżej rozkładu dwupunktowego dystrybuanta tego rozkładu ma postać*

$$F(x) = \begin{cases} 0, & \text{gdy} \quad x \leq 0 \\ q, & \text{gdy} \quad 0 < x \leq 1 \\ 1, & \text{gdy} \quad 1 < x. \end{cases}$$

Następne twierdzenie tłumaczy, że dystrybuanta danego rozkładu wyznaczona jest jednoznacznie. Zachodzi bowiem twierdzenie odwrotne do twierdzenia 3.1.1.

Twierdzenie 3.1.2 *Dla każdej funkcji $\widetilde{F}$: $\mathbf{R} \to \mathbf{R}$ o własnościach 1–5 twierdzenia 3.1.1 istnieje rozkład prawdopodobieństwa d taki, że jego dystrybuanta F jest równa funkcji $\widetilde{F}$.*

Powyższe twierdzenie pozwala nam przetłumaczyć pojęcie rozkładu na pojęcie dystrybuanty tego rozkładu. Mówiąc inaczej, rozkłady będziemy identyfikowali na podstawie znajomości ich dystrybuant.

Przykład 3.1.3 *Weźmy rozkład d zdefiniowany na zbiorze $\{-2, -1, 1, 3, 4\}$ dany wzorem*

$$d(-2) = 0,1, \;\; d(-1) = 0,2, \;\; d(1) = 0,3, \;\; d(3) = 0,2, \;\; d(4) = 0,2.$$

Na wstępie zauważmy, że dziedzina rozkładu d nie musi być zapisana tak jak w treści naszego przykładu, a więc z uwzględnieniem porządku rosnącego. Jeśli tak nie jest, to na wstępie należy to zrobić. Dalej wprowadźmy następujące oznaczenia:

$$x_1, x_2, x_3, x_4, x_5$$

będące skutkiem uporządkowania argumentów funkcji d. Wtedy

$$p_1 = 0,1, \;\; p_2 = 0,2, \;\; p_3 = 0,3, \;\; p_4 = 0,2, \;\; p_5 = 0,2.$$

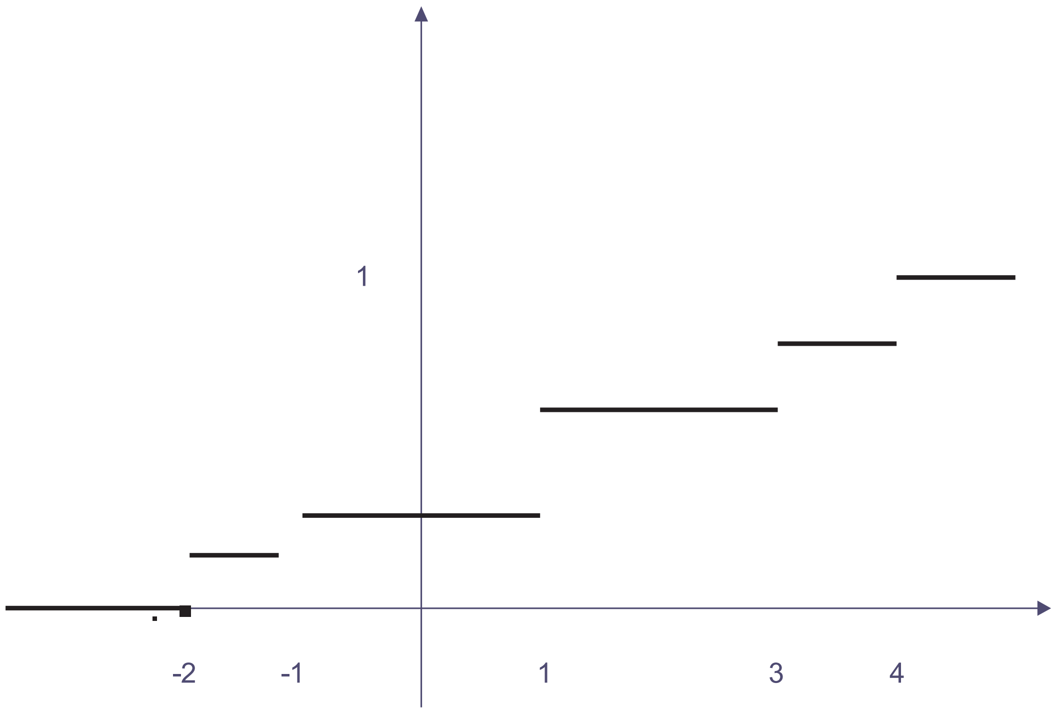
Definiujemy funkcję F: $\mathbf{R} \to \mathbf{R}$ wzorem

$$F(x) = \sum_{x_i < x} p_i,$$

gdzie ostatnia suma równa jest zero, jeśli zbiór indeksów sumowania jest pusty. W takim razie dla $x \le x_1$, $F(x) = 0$. Jeśli weźmiemy dowolną liczbę $x \in (x_1, x_2]$, to z definicji funkcji F dostaniemy $F(x) = p_1$, bowiem sumujemy tylko po zbiorze $\{1\}$. Podobnie dla $x \in (x_2, x_3]$ otrzymamy $F(x) = p_1 + p_2$–teraz sumujemy po zbiorze $\{1, 2\}$. Powtarzając tę procedurę dla kolejnych przedziałów (jest ich skończenie wiele) otrzymamy wzór na dystrybuantę F rozkładu d

$$F(x) = \begin{cases} 0, & x \le x_1; \\ p_1, & x \in (x_1, x_2]; \\ p_1 + p_2, & x \in (x_2, x_3]; \\ p_1 + p_2 + p_3, & x \in (x_3, x_4]; \\ p_1 + p_2 + p_3 + p_4, & x \in (x_4, x_5]; \\ p_1 + p_2 + p_3 + p_4 + p_5 = 1, & x > x_5. \end{cases}$$

Poniższej przedstawiliśmy wykres skonstruowanej funkcji. Rozumiemy już dlaczego funkcje tego typu nazywa się funkcjami schodkowymi.



Rysunek 3.1: wykres dystrybuanty rozkładu d

Twierdzenie 3.1.2 mówi z kolei, że każda niemalejąca, lewostronnie ciągła funkcja schodkowa $\tilde{F}: \mathbf{R} \to [0,1]$ *taka, że jej granica w* $-\infty$ *jest równa zero, w* $+\infty$*–jeden, jednoznacznie wyznacza pewien rozkład prawdopodobieństwa. Podamy teraz przepis na skonstruowanie takiego rozkładu. Za dziedzinę funkcji* d *weźmiemy zbiór punktów nieciągłości funkcji* $\tilde{F}$*. Wcale nie tak trudno (szczegóły pominiemy) można pokazać, że zbiór ten ma postać*

$$\{x_n,\ n \in \mathcal{N}_o\},$$

gdzie konsekwentnie $\mathcal{N}_o$ *oznacza podzbiór zbioru liczb naturalnych. Zatem zbiór ten zawsze jest zbiorem przeliczalnym. Dla uproszszczenia dalszych rozważań przyjmijmy, że jest on skończony i uporządkujmy jego elementy rosnąco*

$$x_1 < x_2 < \ldots < x_{n-1} < x_n.$$

Ciąg liczb p_i *definiujemy następująco*

$$p_i = \lim_{x \to x_i^+} \tilde{F} - \tilde{F}(x_i)$$

dla $i = 1, 2, \ldots, n$*. Należy pokazać, że tak zdefiniowany ciąg ma własności opisane w definicji rozkładu. Jasne jest, że jest to ciąg liczb z przedziału* $[0,1]$*, bowiem* $\tilde{F}$ *przyjmuje swoje wartości w tym przedziale. Co więcej, są to liczby z przedziału obustronnie otwartego, bowiem w punktach* x_i *ma nieciągłość typu "skok". Uzasadnimy, że ciąg ten sumowalny jest do jedności. Na początek weźmy sumę dwóch pierwszych wyrazów tego ciągu, czyli*

$$p_1 + p_2 = \lim_{x \to x_1^+} \tilde{F} - \tilde{F}(x_1) + \lim_{x \to x_2^+} \tilde{F} - \tilde{F}(x_2).$$

Z założenia na przedziale $(x_1, x_2]$ *funkcja* $\tilde{F}$ *jest stała i lewostronnie ciągła, zatem*

$$\lim_{x \to x_1^+} \tilde{F} = \lim_{x \to x_2^-} \tilde{F} = \tilde{F}(x_2).$$

Dlatego

$$p_1 + p_2 = \lim_{x \to x_2^+} \tilde{F} - \tilde{F}(x_1).$$

Ale x_1 *jest najmniejszym elementem zbioru punktów nieciągłości funkcji* $\tilde{F}$*, więc jako schodkowa dla* $x < x_1$ *musi być stała. Ponieważ jej granica w* $-\infty$ *jest równa zero, z lewostronnej ciągłości* $\tilde{F}(x_1) = 0$*. Wobec tego*

$$p_1 + p_2 = \lim_{x \to x_2^+} \tilde{F}.$$

Powtarzając to rozumowanie dla dowolnego $k \leq n$ *otrzymamy*

$$p_1 + p_2 + \ldots + p_k = \lim_{x \to x_k^+} \tilde{F}.$$

W szczególności, dla $k = n$ dostaniemy

$$\sum_{i=1}^{n} p_i = \lim_{x \to x_n^+} \tilde{F}.$$

Ale x_n jest największym punktem, w którym $\tilde{F}$ jest nieciągła, więc jako funkcja schodkowa dla $x > x_n$ musi być stała. W takim razie musi przyjmować wartość 1 dla $x > x_n$, ponieważ jej granica w $+\infty$ jest równa jeden. Dowodzi to, że ciąg (p_i) sumowalny jest do jedności. Pozwala to nam zdefiniować rozkład d

$$d(x_i) = p_i, \ i = 1, 2, \ldots, n$$

Pozostała ostatnia kwestia. Jak ma się dystrybuanta rozkładu do przestrzeni probabilistycznej i dlaczego w nazwie rozkładu pojawia się przymiotnik *prawdopodobieństwa*? Na to pytanie odpowiada trzecie twierdzenie.

Twierdzenie 3.1.3 *Dla każdej dystrybuanty schodkowej F istnieje przestrzeń probabilistyczna (Ω, Σ, P) oraz odwzorowanie $\mathbf{X} : \Omega \longrightarrow \mathbf{R}$, o następujących własnościach:*

1. $$\{\omega \in \Omega : \mathbf{X}(\omega) < x\}$$
   jest zdarzeniem dla każdej liczby rzeczywistej x

2. $$F(x) = P(\{\omega \in \Omega : \mathbf{X}(\omega) < x\}),$$
   dla $x \in \mathbf{R}$

Uwaga 3.1.5 *Funkcję $\mathbf{X}$, o której mówi powyższe twierdzenie, nazywamy zmienną losową, natomiast dystrybuantę F jej rozkładem. Ze względu na typ rozważanego rozkładu w tym rozdziale mówimy, że mamy zmienną losową typu dyskretnego.*

Uwaga 3.1.6 *Niech F i $\mathbf{X}$ będą jak w twierdzeniu 3.1.3. Wtedy*

$$P(\{\omega \in \Omega : \mathbf{X}(\omega) \le x_n\}) = \lim_{x \to x_n+} F,$$

gdzie x_n jest dowolnym argumentem funkcji rozkładu d odpowiadającego dystrybuancie F. Z drugiej strony dla każdego rzeczywistego x

$$\{\omega \in \Omega\colon \mathbf{X}(\omega) = x\} = \{\omega \in \Omega\colon \mathbf{X}(\omega) \le x\} \setminus \{\omega \in \Omega\colon \mathbf{X}(\omega) < x\}.$$