**Rachunek prawdopodobieństwa MAT1332**
Wydział Matematyki, Matematyka Stosowana
Wykładowca: dr hab. A. Jurlewicz

# Wykład 3: Zmienna losowa.

## Schemat Bernoulliego:

Ciąg Bernoulliego zdarzeń losowych modeluje sytuację, w której:

1. Wykonujemy doświadczenie, w którym możliwe są dwa wyniki. Jeden z tych wyników nazywamy sukcesem, drugi porażką. Szansa na wynik „sukces" wynosi $p$.

2. Doświadczenie możemy powtarzać bez zmiany warunków, niezależnie.

Jest to ciąg $A_1$, $A_2, \dots$ niezależnych zdarzeń losowych takich, że $P(A_i) = p$ dla każdego $i$ dla pewnego $0 < p < 1$.
Jeśli zaszło zdarzenie $A_i$, to w $i$-tej próbie odnieśliśmy „sukces", a jeśli zaszło $A_i^c$, to w $i$-tej próbie odnieśliśmy „porażkę".

**Plan I:** Wykonamy $n$ takich doświadczeń i zliczymy ilość sukcesów. Oznaczmy ilość sukcesów przez $X$. Przed realizacją eksperymentu nie wiemy, jaka jest ta liczba. Wiemy natomiast, że możliwe wartości $X$ to $0, 1, 2, \dots, n$ oraz że dla $k$ z tego zbioru możliwych wartości prawdopodobieństwo, że w $n$ próbach otrzymamy dokładnie $k$ sukcesów wynosi

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

**Plan II:** Będziemy wykonywać kolejne doświadczenia tak długo aż pojawi się wynik „sukces". Oznaczmy przez $Y$ ilość wykonanych doświadczeń, czyli czas oczekiwania na pierwszy sukces. Przed realizacją eksperymentu nie wiemy, jaka jest ta ilość. Wiemy natomiast, że możliwe wartości $Y$ to $1, 2, \dots$ oraz że dla $k$ z tego zbioru możliwych wartości prawdopodobieństwo, że pierwszy sukces pojawi się w $k$-tej próbie wynosi

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

$X$ i $Y$ to przykłady zmiennych losowych. Po wykonaniu planów otrzymujemy konkretne liczby - realizacje tych zmiennych losowych.

Opracowanie: dr hab. Agnieszka Jurlewicz

## Zmienna losowa:

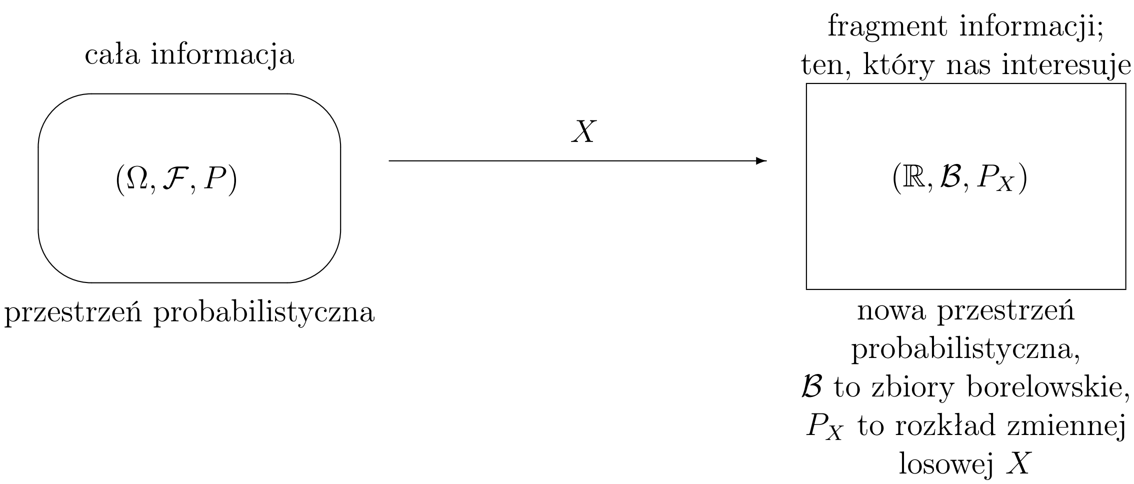
Definicja. **Zmienna losowa** $X$ to funkcja $X : \Omega \longrightarrow \mathbb{R}$, dla której dla dowolnego borelowskiego zbioru $B \subset \mathbb{R}$ mamy

$$\{\omega : X(\omega) \in B\} \in \mathcal{F},$$

Zbiór $\{\omega : X(\omega) \in B\}$ oznaczmy krótko $\{X \in B\}$. Dla zmiennej losowej $X$ zbiór $\{X \in B\}$ jest zdarzeniem losowym.

Innymi słowy, zmienna losowa to taka funkcja $X$ na zbiorze zdarzeń elementarnych o wartościach liczbowych, dla której można określić (teoretycznie) prawdopodobieństwo przyjmowania przez $X$ wartości z dowolnie wybranego zakresu $B$.

W rachunku prawdopodobieństwa interesuje nas **rozkład** zmiennej losowej, ewentualnie jej **charakterystyki liczbowe** (takie jak wartość średnia, wariancja i inne momenty, mediana i inne kwantyle, mody).




## Rozkład zmiennej losowej

Definicja. **Rozkład zmiennej losowej** $X$ to funkcja określona na zbiorach borelowskich w następujący sposób:

$$P_X(B) := P(X \in B) \ \text{ dla dowolnego borelowskiego zbioru } \ B.$$

$P_X$ to funkcja o własnościach prawdopodobieństwa dla przestrzeni stanów $\mathbb{R}$ i rodziny zdarzeń losowych $\mathcal{B}$ (rodzina zbiorów borelowskich).

Opracowanie: dr hab. Agnieszka Jurlewicz

## Typy zmiennych losowych:

1. **Zmienna losowa dyskretna** (in. o rozkładzie dyskretnym)
   to taka zmienna losowa $X$, której rozkład jest singularny względem miary Lebesgue'a i skupiony na zbiorze przeliczalnym (tzn. istnieje przeliczalny zbiór $B \subset \mathbb{R}$ taki, że $P_X(B) = 1$.)
   Inaczej mówiąc, jest to zmienna losowa, która przyjmuje z dodatnim prawdopodobieństwem jedynie skończoną lub nieskończoną przeliczalną liczbę różnych wartości.

2. **Zmienna losowa ciągła** (in. rozkładzie ciągłym)
   to taka zmienna losowa $X$, której rozkład jest absolutnie ciągły względem miary Lebesgue'a.
   Z twierdzenia Radona-Nikodyma dla takiej zmiennej losowej istnieje taka nieujemna funkcja $f(x)$, że dla każdego borelowskiego zbioru $B$

   $$P_X(B) = \int_B f(x)dx.$$

   Funkcja $f(x)$ zwana jest **gęstością** rozkładu $X$.

3. **Zmienna losowa osobliwa** (in. o rozkładzie osobliwym)
   to zmienna losowa, której rozkład jest singularny względem miary Lebesgue'a, skupiony na zbiorze nieprzeliczalnym (tzn. istnieje nieprzeliczalny zbiór $B \subset \mathbb{R}$ o zerowej mierze Lebesgue'a taki, że $P_X(B) = 1$, przy czym dla dowolnego zbioru przeliczalnego $A \subset \mathbb{R}$ mamy $P_X(A) < 1$) oraz zachodzi warunek:
   $P_X(\{x\}) = P(X = x) = 0$ dla każdego $x \in \mathbb{R}$.

4. Dowolna zmienna losowa albo jest jednego z tych trzech typów, albo ma
   **rozkład mieszany** składający się z rozkładów tych typów.

## Uwaga:

Budując model probabilistyczny danego zjawiska używamy zwykle pojęcia zmiennej losowej. Rzadko jednak definiujemy tę zmienną losową bezpośrednio. Zwykle określamy tylko jej rozkład (tak, jak opisano na następnych stronach). To uważa się zwykle za model probabilistyczny, na ogół nie precyzuje się przestrzeni probabilistycznej $(\Omega, \mathcal{F}, P)$ dla danego zjawiska.

Opracowanie: dr hab. Agnieszka Jurlewicz

## TECHNIKA OKREŚLANIA ROZKŁADU ZMIENNEJ LOSOWEJ $X$ ZA POMOCĄ DYSTRYBUANTY:

- Pełna informacja o rozkładzie zmiennej losowej $X$ zawarta jest w funkcji

$$F(x) = P(X < x), \quad x \in \mathbb{R},$$

nazywanej **dystrybuantą**.

- Zauważmy, że $F(b) = P(X < b) = P_X(B)$ dla $B = (-\infty, b)$. Z dystrybuanty możemy dostać informację o wartościach funkcji $P_X$ na innych zbiorach borelowskich. Dla dowolnych $a, b \in \mathbb{R}$ mamy:

$$\begin{aligned}
P(X < b) &= F(b) \\
P(X \leqslant b) &= \lim_{x \to b+} F(x) \\
P(X \geqslant b) &= 1 - F(b) \\
P(X > b) &= 1 - \lim_{x \to b+} F(x) \\
P(a \leqslant X < b) &= F(b) - F(a) \\
P(a < X < b) &= F(b) - \lim_{x \to a+} F(x) \\
P(a < X \leqslant b) &= \lim_{x \to b+} F(x) - \lim_{x \to a+} F(x) \\
P(a \leqslant X \leqslant b) &= \lim_{x \to b+} F(x) - F(a)
\end{aligned}$$

- Funkcja $F(x)$ jest dystrybuantą pewnej zmiennej losowej $X$ wtedy i tylko wtedy, gdy spełnia następujące warunki:

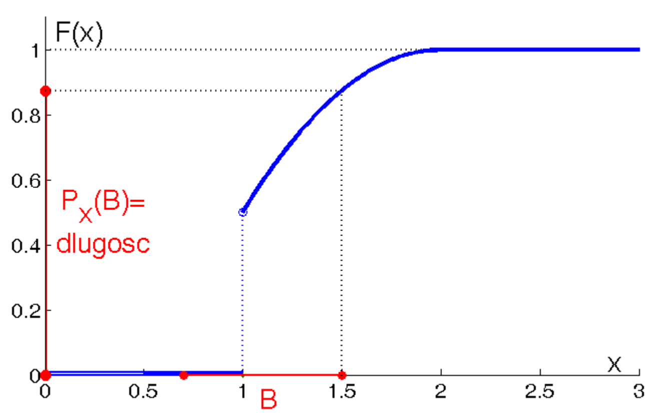
  (a) jest lewostronnie ciągła;

  (b) jest niemalejąca;

  (c) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

  Każda funkcja $F$, która spełnia powyższe warunki, ma probabilistyczną interpretację, reprezentację; może być używana w modelach w roli dystrybuanty.

- Przykładowy wykres dystrybuanty ($X$ - płaca losowo wybranego pracownika pewnej dużej grupy zawodowej w stosunku do płacy minimalnej):




Opracowanie: dr hab. Agnieszka Jurlewicz

## Technika określania rozkładu dyskretnego za pomocą ciągu wartości i prawdopodobieństw:

- Pełna informacja o rozkładzie **dyskretnej** zmiennej losowej $X$ zawarta jest także w **ciągu par wartości i prawdopodobieństw**, czyli ciągu $\{(x_n, p_n), n \in \mathbb{T} \subset \mathbb{N}\}$, gdzie $\{x_n, n \in \mathbb{T}\}$ to ciąg wszystkich wartości przyjmowanych przez $X$ z dodatnim prawdopodobieństwem, natomiast $p_n = P(X = x_n), \ \ n \in \mathbb{T}$.

- Dla dowolnego zbioru borelowskiego $B$ mamy $P(X \in B) = \sum\limits_{n \in \mathbb{T}_B} p_n$, gdzie $\mathbb{T}_B$ to zbiór tych $n$, dla których $x_n \in B$.

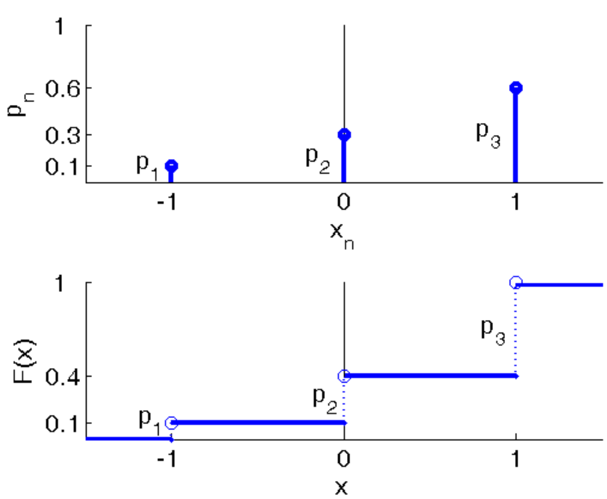
- W szczególności, dystrybuanta ma postać
$$F(x) = P(X < x) = \sum_{n \in \mathbb{T}(x)} p_n,$$
gdzie $\mathbb{T}(x)$ to zbiór tych $n$, dla których $x_n < x$.

  Zmienna losowa ma rozkład dyskretny wtedy i tylko wtedy, gdy jej dystrybuanta jest funkcją schodkową. Schodki są w punktach $x_1, x_2, \ldots$ i mają wysokości odpowiednio $p_1, p_2, \ldots$

- Ciąg $\{(x_n, p_n), n \in \mathbb{T} \subset \mathbb{N}\}$ określa pewien rozkład dyskretny zmiennej losowej $X$ tak, że $p_n = P(X = x_n)$, wtedy i tylko wtedy, gdy spełnia następujące warunki:
  (a) $\{x_n, n \in \mathbb{T}\}$ to ciąg różnowartościowy;
  (b) $p_n \geqslant 0$ dla każdego $n \in \mathbb{T}$;
  (c) $\sum\limits_{n \in \mathbb{T}} p_n = 1$.

  Każdy ciąg spełniający te warunki ma probabilistyczną interpretację, reprezentację; może być używany w modelach do definiowania rozkładu dyskretnego.

- Przykładowy wykres ($X$ - zachowanie dziewczyny, gdy jej chłopak spóźnia sie na randkę, opisane liczbowo: $X = -1$ - gniewa się ; $X = 0$ - nie zauważa ; $X = 1$ - cieszy się, że wreszcie przyszedł ):




Opracowanie: dr hab. Agnieszka Jurlewicz

**Technika określania rozkładu ciągłej zmiennej losowej $X$ za pomocą gęstości:**

- Pełna informacja o **ciągłym** rozkładzie zmiennej losowej $X$ zawarta jest również w **gęstości** $f(x)$ rozkładu $X$.

- Z gęstości możemy dostać informację o wartościach funkcji $P_X$ na dowolnych zbiorach borelowskich, gdyż tak jak w definicji rozkładu ciągłego, mamy
$P_X(B) = \int_B f(x)dx$. W szczególności, dla dowolnych $a, b \in \mathbb{R}$ mamy:

$$\begin{aligned} P(X < b) = P(X \leqslant b) &= \int_{-\infty}^{b} f(x)dx \\ P(X \geqslant b) = P(X > b) &= \int_{b}^{\infty} f(x)dx \\ P(a \leqslant X < b) = P(a < X < b) &= P(a < X \leqslant b) = P(a \leqslant X \leqslant b) = \int_{a}^{b} f(x)dx \end{aligned}$$

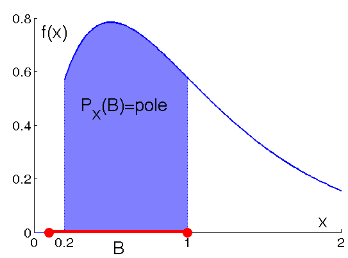
- Funkcja $f(x)$ jest gęstością pewnego rozkładu ciągłego wtedy i tylko wtedy, gdy spełnia następujące warunki:

  (a) $f(x) \geqslant 0$ dla każdego $x \in \mathbb{R}$;

  (b) $\int_{-\infty}^{\infty} f(x)dx = 1$.

  Każda funkcja $f$ spełniająca te warunki ma probabilistyczną interpretację, reprezentację; może być używana w modelach w roli gęstości rozkładu ciągłego.

- Przykładowy wykres ($X$ - czas pracy pewnego urzadzenia do pierwszej awarii):




- Dystrybuanta $F(x)$ zmiennej losowej $X$ równa jest całce

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

Wynika stąd, że dystrybuanta rozkładu ciągłego musi być funkcją ciągłą. Nie jest to jednak warunek wystarczający. Można pokazać, że:

**Twierdzenie.** Jeżeli dystrybuanta $F(x)$ jest funkcją ciągłą i różniczkowalną w prawie każdym punkcie (nie ma pochodnej jedynie w skończonej liczbie punktów), to jest to dystrybuanta rozkładu ciągłego o gęstości $f(x)$ postaci

$$f(x) = \begin{cases} F'(x) & \text{dla tych } x, \text{ dla których pochodna istnieje} \\ 0 & \text{poza tym.} \end{cases}$$

Opracowanie: dr hab. Agnieszka Jurlewicz