

Statystyka i rachunek prawdopodobieństwa

Wykład 3: Zmienne losowe

Mariusz Tarnopolski

Instytut Astronomii UMK

Statystyka ©2026



Pojęcie zmiennej losowej

Mając daną przestrzeń probabilistyczną $(\Omega, \mathcal{F}, P)$ wygodnie jest przejść do innej przestrzeni, w której prawdopodobieństwa są określone nie na zbiorach, a na liczbach.

Pojęcie zmiennej losowej

Mając daną przestrzeń probabilistyczną $(\Omega, \mathcal{F}, P)$ wygodnie jest przejść do innej przestrzeni, w której prawdopodobieństwa są określone nie na zbiorach, a na liczbach.

Ułatwieniem będzie wtedy wykonywanie operacji arytmetycznych na Ω a nie tylko mnogościowych. Pozwoli to też na wykorzystanie aparatu analizy matematycznej.

Pojęcie zmiennej losowej

Mając daną przestrzeń probabilistyczną $(\Omega, \mathcal{F}, P)$ wygodnie jest przejść do innej przestrzeni, w której prawdopodobieństwa są określone nie na zbiorach, a na liczbach.

Ułatwieniem będzie wtedy wykonywanie operacji arytmetycznych na Ω a nie tylko mnogościowych. Pozwoli to też na wykorzystanie aparatu analizy matematycznej.

Przyporządkujemy zatem **zdarzeniom** z przestrzeni Ω **liczby rzeczywiste** ze zbioru $A \subset \mathbb{R}$. Funkcję dokonującą takiego przekształcenia nazywamy **zmienną losową** $X : \Omega \rightarrow \mathbb{R}$. Przeciwobraz X jest zatem zdarzeniem: $X^{-1}(A) \in \mathcal{F}$.

Przykład. $X =$ wzrost człowieka $\in [55, 272]$ cm.

Pojęcie zmiennej losowej

Mając daną przestrzeń probabilistyczną $(\Omega, \mathcal{F}, P)$ wygodnie jest przejść do innej przestrzeni, w której prawdopodobieństwa są określone nie na zbiorach, a na liczbach.

Ułatwieniem będzie wtedy wykonywanie operacji arytmetycznych na Ω a nie tylko mnogościowych. Pozwoli to też na wykorzystanie aparatu analizy matematycznej.

Przyporządkujemy zatem **zdarzeniom** z przestrzeni Ω **liczby rzeczywiste** ze zbioru $A \subset \mathbb{R}$. Funkcję dokonującą takiego przekształcenia nazywamy **zmienną losową** $X : \Omega \rightarrow \mathbb{R}$. Przeciwobraz X jest zatem zdarzeniem: $X^{-1}(A) \in \mathcal{F}$.

Przykład. $X =$ wzrost człowieka $\in [55, 272]$ cm.

Zamiast więc obserwować wyniki $e : e \in \Omega$ zdarzenia losowego, obserwować będziemy występowanie wielkości $X(e)$ — liczby, będącej wartością jaką zmienna losowa (funkcja) przypisuje zdarzeniu e .

Pojęcie zmiennej losowej

Mając daną przestrzeń probabilistyczną $(\Omega, \mathcal{F}, P)$ wygodnie jest przejść do innej przestrzeni, w której prawdopodobieństwa są określone nie na zbiorach, a na liczbach.

Ułatwieniem będzie wtedy wykonywanie operacji arytmetycznych na Ω a nie tylko mnogościowych. Pozwoli to też na wykorzystanie aparatu analizy matematycznej.

Przyporządkujemy zatem **zdarzeniom** z przestrzeni Ω **liczby rzeczywiste** ze zbioru $A \subset \mathbb{R}$. Funkcję dokonującą takiego przekształcenia nazywamy **zmienną losową** $X : \Omega \rightarrow \mathbb{R}$. Przeciwobraz X jest zatem zdarzeniem: $X^{-1}(A) \in \mathcal{F}$.

Przykład. $X =$ wzrost człowieka $\in [55, 272]$ cm.

Zamiast więc obserwować wyniki $e : e \in \Omega$ zdarzenia losowego, obserwować będziemy występowanie wielkości $X(e)$ — liczby, będącej wartością jaką zmienna losowa (funkcja) przypisuje zdarzeniu e .

Podsumowując:

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\Omega_X, \mathcal{F}_X, P_X)$$

Realizacja zmiennej losowej — wartość zmiennej losowej faktycznie zaobserwowana. Np. zmienna losowa X przyjęła dla konkretnego człowieka wartość $x = 155$ cm. Możemy mówić o prawdopodobieństwie przyjęcia przez zmienną losową konkretnej wartości: $P(X = x)$, lub wartości z jakiegoś przedziału, np. $P(X > 155 \text{ cm})$.

Realizacja zmiennej losowej — wartość zmiennej losowej faktycznie zaobserwowana. Np. zmienna losowa X przyjęła dla konkretnego człowieka wartość $x = 155$ cm. Możemy mówić o prawdopodobieństwie przyjęcia przez zmienną losową konkretnej wartości: $P(X = x)$, lub wartości z jakiegoś przedziału, np. $P(X > 155 \text{ cm})$.

Definicja: Rozkładem prawdopodobieństwa zmiennej X nazywamy prawdopodobieństwo przyjęcia przez zmienną X wartości ze zbioru $A \in \mathcal{F}_X$ określone jako

$$P_X(A) = P(X^{-1}(A))$$

Realizacja zmiennej losowej — wartość zmiennej losowej faktycznie zaobserwowana. Np. zmienna losowa X przyjęła dla konkretnego człowieka wartość $x = 155$ cm. Możemy mówić o prawdopodobieństwie przyjęcia przez zmienną losową konkretnej wartości: $P(X = x)$, lub wartości z jakiegoś przedziału, np. $P(X > 155 \text{ cm})$.

Definicja: Rozkładem prawdopodobieństwa zmiennej X nazywamy prawdopodobieństwo przyjęcia przez zmienną X wartości ze zbioru $A \in \mathcal{F}_X$ określone jako

$$P_X(A) = P(X^{-1}(A))$$

Zapis $P(X = x)$ czytamy „prawdopodobieństwo, że zmienna losowa X przyjmie wartość x ”.

Realizacja zmiennej losowej — wartość zmiennej losowej faktycznie zaobserwowana. Np. zmienna losowa X przyjęła dla konkretnego człowieka wartość $x = 155$ cm. Możemy mówić o prawdopodobieństwie przyjęcia przez zmienną losową konkretnej wartości: $P(X = x)$, lub wartości z jakiegoś przedziału, np. $P(X > 155 \text{ cm})$.

Definicja: Rozkładem prawdopodobieństwa zmiennej X nazywamy prawdopodobieństwo przyjęcia przez zmienną X wartości ze zbioru $A \in \mathcal{F}_X$ określone jako

$$P_X(A) = P(X^{-1}(A))$$

Zapis $P(X = x)$ czytamy „prawdopodobieństwo, że zmienna losowa X przyjmie wartość x ”. Np. orzeł $\rightarrow 0$, reszka $\rightarrow 1$, to $P(X = 0)$ jest prawdopodobieństwem wyrzucenia orła.

Ponadto, $P(a < X < b)$ oznacza prawdopodobieństwo, że zmienna losowa X przyjmie wartość z przedziału (a, b) — jest równe prawdopodobieństwu przeciwobrazu tego przedziału.

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Zmienna X określona na $\mathbb{R}$ jest typu dyskretnego jeśli istnieje skończony lub przeliczalny zbiór jej wartości $\Omega_X = \{x_1, x_2, \dots\}$, któremu można przypisać zestaw prawdopodobieństw p_i spełniających

$$P(X = x_i) = p_i \geq 0, \quad i \in \mathbb{N}_+$$

$$\sum_{i=1} p_i = 1$$

gdzie suma przebiega do n lub ∞ w zależności od tego, czy mamy do czynienia ze zbiorem Ω_X skończonym czy przeliczalnym. Z powyższych warunków wynika też, że $p_i \leq 1$.

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Zmienna X określona na $\mathbb{R}$ jest typu dyskretnego jeśli istnieje skończony lub przeliczalny zbiór jej wartości $\Omega_X = \{x_1, x_2, \dots\}$, któremu można przypisać zestaw prawdopodobieństw p_i spełniających

$$P(X = x_i) = p_i \geq 0, \quad i \in \mathbb{N}_+$$

$$\sum_{i=1} p_i = 1$$

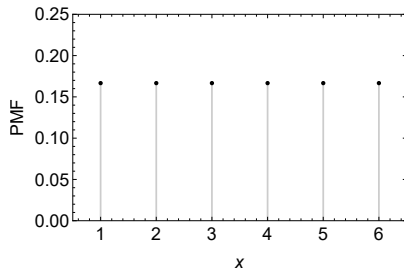
gdzie suma przebiega do n lub ∞ w zależności od tego, czy mamy do czynienia ze zbiorem Ω_X skończonym czy przeliczalnym. Z powyższych warunków wynika też, że $p_i \leq 1$.

Funkcję $P(x) = p_i$ nazywamy **funkcją masy prawdopodobieństwa** (*probability mass function*, PMF) lub (bez odróżniania od rozkładów zmiennych ciągłych; zob. slajd 16) **funkcją rozkładu prawdopodobieństwa**, **rozkładem prawdopodobieństwa**, **funkcją prawdopodobieństwa** lub po prostu **rozkładem**.

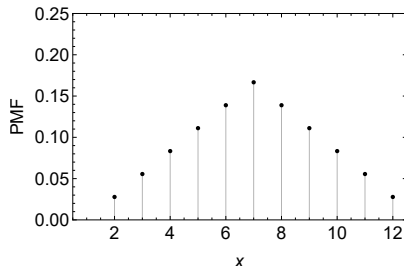
Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

- Rzut jedną kostką:
 $P_1(X = x) = \frac{1}{6}$
dla $x \in \{1, \dots, 6\}$
(rozkład jednostajny)



- Rzut dwiema kostkami:
 $P_2(X = x)$ dla $x \in \{2, \dots, 12\}$



Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Przykład. Pięciokrotny rzut monetą.

Jaki jest rozkład prawdopodobieństwa opisującego liczbę uzyskanych orłów?

X opisuje liczbę orłów; X może przyjąć wartości 0,1,2,3,4,5.

$$P(X = 0) = \sum_e P\{e : X(e) = 0\} = \frac{1}{2^5}$$

$$P(X = 1) = \sum_e P\{e : X(e) = 1\} = \frac{5}{2^5}$$

$$P(X = 2) = \dots$$

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Przykład. Pięciokrotny rzut monetą.

Jaki jest rozkład prawdopodobieństwa opisującego liczbę uzyskanych orłów?

X opisuje liczbę orłów; X może przyjąć wartości 0,1,2,3,4,5.

$$P(X = 0) = \sum_e P\{e : X(e) = 0\} = \frac{1}{2^5}$$

$$P(X = 1) = \sum_e P\{e : X(e) = 1\} = \frac{5}{2^5}$$

$$P(X = 2) = \dots$$

Jeśli interesuje nas bezwzględna różnica między liczbą wyrzuconych orłów i reszek, to można wprowadzić nową zmienną losową Y opisującą tę właśnie cechę i przyjmującą możliwe wartości: 1,3,5. Np. $P(Y = 3)$, czyli $\{4 \text{ orły}, 1 \text{ reszka}\}$ lub na odwrót:

$$P(Y = 3) = \sum_e P\{e : Y(e) = 3\} = \frac{10}{2^5}$$

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Przykład. Prawdopodobieństwo, że łucznik trafi do celu wynosi $p = 0.4$. Prawdopodobieństwo, że nie trafi, wynosi zatem $q = 1 - p = 0.6$. Niech X oznacza numer pierwszego podejścia, które zakończyło się trafieniem. Mamy $P(X = 1) = p$.

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Przykład. Prawdopodobieństwo, że łucznik trafi do celu wynosi $p = 0.4$.

Prawdopodobieństwo, że nie trafi, wynosi zatem $q = 1 - p = 0.6$.

Niech X oznacza numer pierwszego podejścia, które zakończyło się trafieniem. Mamy $P(X = 1) = p$.

Prawdopodobieństwo, że trafi za drugim razem oznacza, że za pierwszym chybił: $P(X = 2) = qp = 0.6 \cdot 0.4 = 0.24$.

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Przykład. Prawdopodobieństwo, że łucznik trafi do celu wynosi $p = 0.4$.

Prawdopodobieństwo, że nie trafi, wynosi zatem $q = 1 - p = 0.6$.

Niech X oznacza numer pierwszego podejścia, które zakończyło się trafieniem. Mamy $P(X = 1) = p$.

Prawdopodobieństwo, że trafi za drugim razem oznacza, że za pierwszym chybił: $P(X = 2) = qp = 0.6 \cdot 0.4 = 0.24$.

Prawdopodobieństwo, że **najpóźniej** czwarte zakończy się trafieniem jest równoważne ze zdarzeniem przeciwnym do zdarzenia, że nastąpiła seria 4 chybień:

$$P(X \leq 4) = 1 - q^4 = 0.8704$$

Rozkłady prawdopodobieństwa

Zmienne losowe dyskretne (skokowe)

Przykład. Prawdopodobieństwo, że łucznik trafi do celu wynosi $p = 0.4$.

Prawdopodobieństwo, że nie trafi, wynosi zatem $q = 1 - p = 0.6$.

Niech X oznacza numer pierwszego podejścia, które zakończyło się trafieniem. Mamy $P(X = 1) = p$.

Prawdopodobieństwo, że trafi za drugim razem oznacza, że za pierwszym chybił: $P(X = 2) = qp = 0.6 \cdot 0.4 = 0.24$.

Prawdopodobieństwo, że **najpóźniej** czwarte zakończy się trafieniem jest równoważne ze zdarzeniem przeciwnym do zdarzenia, że nastąpiła seria 4 chybień:

$$P(X \leq 4) = 1 - q^4 = 0.8704$$

Prawdopodobieństwo, że w pierwszych 4 próbach nie trafi:

$$P(X > 4) = 1 - P(X \leq 4) = 1 - 0.8704 = 0.1296 = q^4$$

Rozkłady prawdopodobieństwa

Rozkład geometryczny

Powyższe doświadczenie losowe: że r -ta próba zakończy się sukcesem, przy prawdopodobieństwie sukcesu p w pojedynczej próbie, określone jest rozkładem prawdopodobieństwa

$$P(X = r) = pq^{r-1}$$

Uogólniając przykład łucznika, mamy zatem:

- prawdopodobieństwo, że r pierwszych prób zakończy się porażką:

$$P(X > r) = q^r$$

- prawdopodobieństwo, że sukces pojawi się najpóźniej w r -tej próbie:

$$P(X \leq r) = 1 - q^r$$

Rozkłady prawdopodobieństwa

Rozkład Bernouliego (dwumianowy)

Powyższy przykład jest typowym zagadnieniem wielokrotnego losowania ze zwracaniem o prawd. sukcesu p oraz prawd. porażki $q = 1 - p$. Tutaj $\Omega = \{0, 1\}$, gdzie 1=sukces, 0=porażka.

Rozkłady prawdopodobieństwa

Rozkład Bernoulliego (dwumianowy)

Powyższy przykład jest typowym zagadnieniem wielokrotnego losowania ze zwracaniem o prawd. sukcesu p oraz prawd. porażki $q = 1 - p$. Tutaj $\Omega = \{0, 1\}$, gdzie 1=sukces, 0=porażka.

Przy n losowaniach przestrzenią zdarzeń jest zbiór n -wyrazowych ciągów zer i jedynek: $(i_1, i_2, \dots, i_n)$, $i_k = 0 \vee 1$, $k \in \{1, 2, \dots, n\}$.

Rozkłady prawdopodobieństwa

Rozkład Bernoulliego (dwumianowy)

Powyższy przykład jest typowym zagadnieniem wielokrotnego losowania ze zwracaniem o prawd. sukcesu p oraz prawd. porażki $q = 1 - p$. Tutaj $\Omega = \{0, 1\}$, gdzie 1=sukces, 0=porażka.

Przy n losowaniach przestrzenią zdarzeń jest zbiór n -wyrazowych ciągów zer i jedynek: $(i_1, i_2, \dots, i_n)$, $i_k = 0 \vee 1$, $k \in \{1, 2, \dots, n\}$. Zdarzenie $A_{n,k}$ to pojawienie się takiego ciągu, w którym k razy zdarzył się sukces (bez względu na to w jakiej konfiguracji). Jest to zatem mnogościowa alternatywa wszystkich tych konfiguracji:

$$A_{n,k} = \bigcup_{\sum_{s=1}^n i_s = k} (i_1, i_2, \dots, i_n)$$

Rozkłady prawdopodobieństwa

Rozkład Bernoulliego (dwumianowy)

Powyższy przykład jest typowym zagadnieniem wielokrotnego losowania ze zwracaniem o prawd. sukcesu p oraz prawd. porażki $q = 1 - p$. Tutaj $\Omega = \{0, 1\}$, gdzie 1=sukces, 0=porażka.

Przy n losowaniach przestrzenią zdarzeń jest zbiór n -wyrazowych ciągów zer i jedynek: $(i_1, i_2, \dots, i_n)$, $i_k = 0 \vee 1$, $k \in \{1, 2, \dots, n\}$. Zdarzenie $A_{n,k}$ to pojawienie się takiego ciągu, w którym k razy zdarzył się sukces (bez względu na to w jakiej konfiguracji). Jest to zatem mnogościowa alternatywa wszystkich tych konfiguracji:

$$A_{n,k} = \bigcup_{\sum_{s=1}^n i_s = k} (i_1, i_2, \dots, i_n)$$

Przyjmując, że kolejne losowania są niezależne, mamy

$$P(i_1, i_2, \dots, i_n) = P(i_1)P(i_2) \cdots P(i_n) = p^k q^{n-k}$$

Rozkłady prawdopodobieństwa

Rozkład Bernouliego (dwumianowy)

Liczba interesujących nas zdarzeń to liczba wszystkich powyższych konfiguracji, czyli $\binom{n}{k}$, zatem

$$P(A_{n,k}) = P\left(\bigcup_{\sum_{s=1}^n i_s = k} (i_1, i_2, \dots, i_n)\right) = \binom{n}{k} p^k q^{n-k}$$

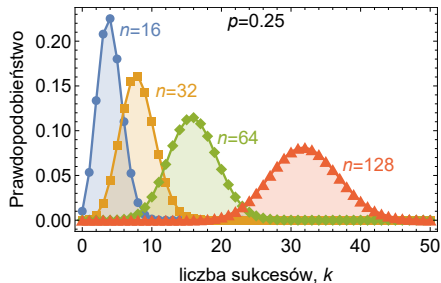
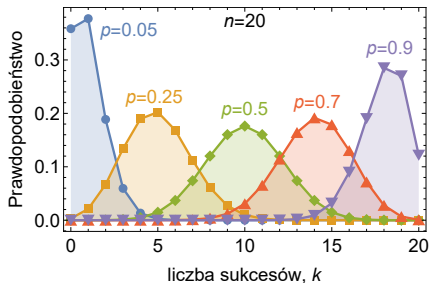
Ostatnie wyrażenie jest k -tym wyrazem rozwinięcia dwumianu¹, stąd rozkład Bernouliego nazywany jest też rozkładem dwumianowym.

¹ $(p + q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}$

Rozkład prawdopodobieństwa

Rozkład Bernouliego (dwumianowy)

$$P_{n,k} = \binom{n}{k} p^k q^{n-k}$$



Rozkłady prawdopodobieństwa

Rozkład hipergeometryczny

Przykład. Wybieramy n -krotnie kulę z urny, w której spośród wszystkich N kul jest B białych i $N - B$ czarnych. Losowania odbywają się bez zwracania. Ile wynosi prawdopodobieństwo, że kula biała pojawi się dokładnie $k \leq n$ razy?

Rozkłady prawdopodobieństwa

Rozkład hipergeometryczny

Przykład. Wybieramy n -krotnie kulę z urny, w której spośród wszystkich N kul jest B białych i $N - B$ czarnych. Losowania odbywają się bez zwracania. Ile wynosi prawdopodobieństwo, że kula biała pojawi się dokładnie $k \leq n$ razy?

Rozwiązanie. Kolejne losowania są **zależne**.

Skorzystamy z klasycznej definicji prawdopodobieństwa. Liczba sposobów wyciągnięcia n spośród N kul wynosi C_N^n . Liczba sposobów wyciągnięcia k kul białych ze wszystkich B wynosi C_B^k , zaś wylosowania pozostałych $n - k$ kul spośród $N - B$ kul czarnych wynosi C_{N-B}^{n-k} . Szukane prawdopodobieństwo ma więc rozkład:

$$P(X = k) = \frac{C_B^k C_{N-B}^{n-k}}{C_N^n}$$

Rozkłady prawdopodobieństwa

Rozkład Poissona

Często interesuje nas graniczny przypadek rozkładu dwumianowego przy $n \rightarrow \infty$. Wzięcie po prostu $\lim_{n \rightarrow \infty} \binom{n}{k} p^k q^{n-k}$ daje 0. Definiuje się zatem $\lambda = np$, która jest stała gdy $n \rightarrow \infty$, czyli p zmienia się tak by $\lambda \in \text{const.}$

Rozkłady prawdopodobieństwa

Rozkład Poissona

Często interesuje nas graniczny przypadek rozkładu dwumianowego przy $n \rightarrow \infty$. Wzięcie po prostu $\lim_{n \rightarrow \infty} \binom{n}{k} p^k q^{n-k}$ daje 0. Definiuje się zatem $\lambda = np$, która jest stała gdy $n \rightarrow \infty$, czyli p zmienia się tak by $\lambda \in \text{const}$. Wtedy

$$\begin{aligned} P(X = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

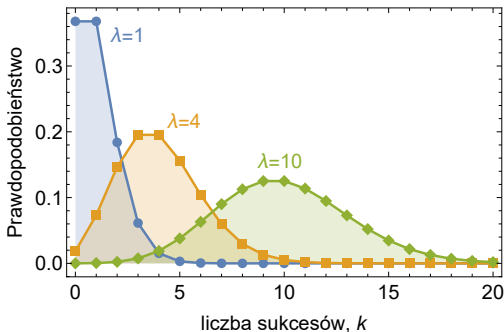
Rozkłady prawdopodobieństwa

Rozkład Poissona

Wtedy

$$\lim_{n \rightarrow \infty} P(X = k) = \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^k}{k!} e^{-\lambda}$$

co jest rozkładem Poissona—dobre przybliżenie rozkładu dwumianowego dla dużego n i małego p (stąd potoczna nazwa: rozkład rzadkich zdarzeń). W praktyce jest dobrym przybliżeniem dla $p \lesssim 0.2$ i $n \gtrsim 20$.

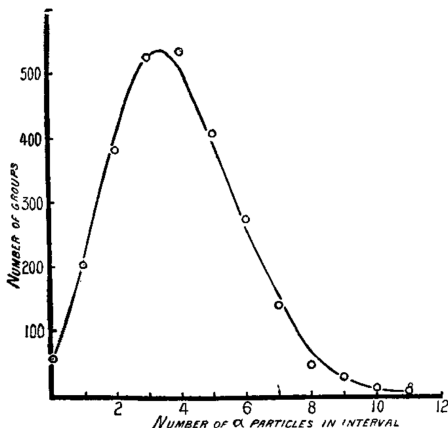


Rozkłady prawdopodobieństwa

Rozkład Poissona — eksperyment Rutherforda, Geigera, Batemana (1910 r.)

Obserwowano emisję cząstek α w $n = 2608$ przedziałach czasu trwających po 7.5 sek. W każdym przedziale czasu mamy do czynienia z dośw. losowym — emisją cząstki (cząstek) α lub jej brakiem.

W sumie zaobserwowane 10,097 rozpadów α . Zakładając niezależność dla poszczególnych N atomów, to zmienna opisująca liczbę $k \ll N$ sukcesów (rozpadów α) będzie dana rozkładem dwumianowym; przy małym prawd. sukcesu przybliżenie rozkładem Poissona jest bardzo dobre. 👍



Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Zmienna losowa X określona na $\mathbb{R}$ jest typu ciągłego gdy istnieje funkcja $f(x) \geq 0$ (**gęstość prawdopodobieństwa**) taka, że dla $\mathcal{A} \subset \mathbb{R}$ zachodzi

$$P(X \in \mathcal{A}) \equiv P(\mathcal{A}) = \int_{\mathcal{A}} f(x) dx$$

czyli np. dla przedziału $\mathcal{A} = (a, b)$:

$$P(a < X < b) = \int_a^b f(x) dx$$

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Zmienna losowa X określona na $\mathbb{R}$ jest typu ciągłego gdy istnieje funkcja $f(x) \geq 0$ (**gęstość prawdopodobieństwa**) taka, że dla $\mathcal{A} \subset \mathbb{R}$ zachodzi

$$P(X \in \mathcal{A}) \equiv P(\mathcal{A}) = \int_{\mathcal{A}} f(x) dx$$

czyli np. dla przedziału $\mathcal{A} = (a, b)$:

$$P(a < X < b) = \int_a^b f(x) dx$$

Powiemy też potocznie, że rozkład zmiennej ciągłej jest ciągły, zaś rozkład zmiennej dyskretnej jest dyskretny (por. slajd 4).

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Zmienna losowa X określona na $\mathbb{R}$ jest typu ciągłego gdy istnieje funkcja $f(x) \geq 0$ (**gęstość prawdopodobieństwa**) taka, że dla $\mathcal{A} \subset \mathbb{R}$ zachodzi

$$P(X \in \mathcal{A}) \equiv P(\mathcal{A}) = \int_{\mathcal{A}} f(x) dx$$

czyli np. dla przedziału $\mathcal{A} = (a, b)$:

$$P(a < X < b) = \int_a^b f(x) dx$$

Powiemy też potocznie, że rozkład zmiennej ciągłej jest ciągły, zaś rozkład zmiennej dyskretnej jest dyskretny (por. slajd 4).

Musi zachodzić unormowanie:

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

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Dystrybuanta $F(x)$ (*cumulative distribution function*, CDF) zmiennej losowej X to funkcja na $\mathbb{R}$, która każdemu $x \in \mathbb{R}$ przypisuje prawdopodobieństwo, że X przyjmie wartość mniejszą niż x :

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

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Dystrybuanta $F(x)$ (*cumulative distribution function*, CDF) zmiennej losowej X to funkcja na $\mathbb{R}$, która każdemu $x \in \mathbb{R}$ przypisuje prawdopodobieństwo, że X przyjmie wartość mniejszą niż x :

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

Mamy zatem związek między f a F :

$$\frac{dF(x)}{dx} = f(x)$$

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Dystrybuanta $F(x)$ (*cumulative distribution function*, CDF) zmiennej losowej X to funkcja na $\mathbb{R}$, która każdemu $x \in \mathbb{R}$ przypisuje prawdopodobieństwo, że X przyjmie wartość mniejszą niż x :

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

Mamy zatem związek między f a F :

$$\frac{dF(x)}{dx} = f(x)$$

Własności dystrybuanty:

- $F(x)$ jest monotonicznie niemalejąca
- $0 \leq F(x) \leq 1$; $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$
- $P(X \geq x) = 1 - F(x)$ oraz $P(a \leq X < b) = F(b) - F(a)$

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Dystrybuanta $F(x)$ (*cumulative distribution function*, CDF) zmiennej losowej X to funkcja na $\mathbb{R}$, która każdemu $x \in \mathbb{R}$ przypisuje prawdopodobieństwo, że X przyjmie wartość mniejszą niż x :

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

Mamy zatem związek między f a F :

$$\frac{dF(x)}{dx} = f(x)$$

Własności dystrybuanty:

- $F(x)$ jest monotonicznie niemalejąca
- $0 \leq F(x) \leq 1$; $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$
- $P(X \geq x) = 1 - F(x)$ oraz $P(a \leq X < b) = F(b) - F(a)$

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Dystrybuanta $F(x)$ (*cumulative distribution function*, CDF) zmiennej losowej X to funkcja na $\mathbb{R}$, która każdemu $x \in \mathbb{R}$ przypisuje prawdopodobieństwo, że X przyjmie wartość mniejszą niż x :

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

Mamy zatem związek między f a F :

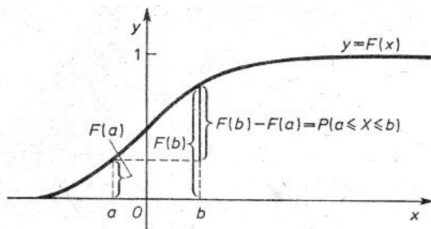
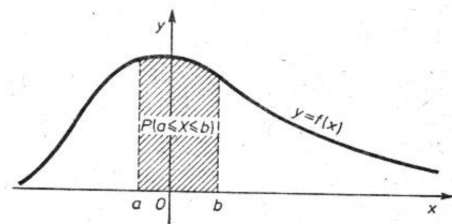
$$\frac{dF(x)}{dx} = f(x)$$

Własności dystrybuanty:

- $F(x)$ jest monotonicznie niemalejąca
- $0 \leq F(x) \leq 1$; $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$
- $P(X \geq x) = 1 - F(x)$ oraz $P(a \leq X < b) = F(b) - F(a)$

Rozkłady prawdopodobieństwa

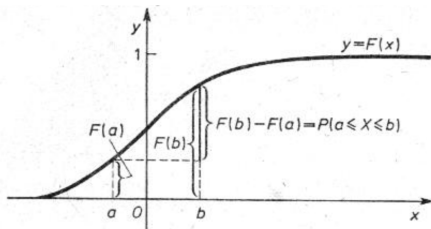
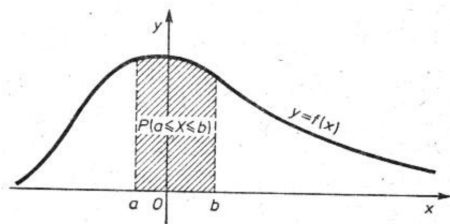
Rozkłady ciągłej zmiennej losowej



Gęstość prawdopodobieństwa i dystrybuanta zmiennej losowej

Rozkłady prawdopodobieństwa

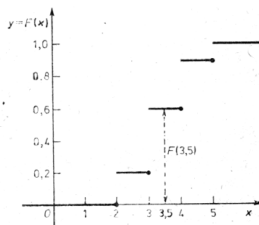
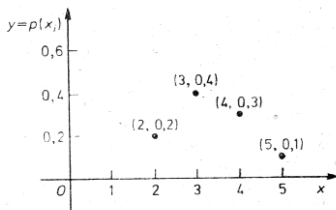
Rozkłady ciągłej zmiennej losowej



Gęstość prawdopodobieństwa i dystrybuanta zmiennej losowej

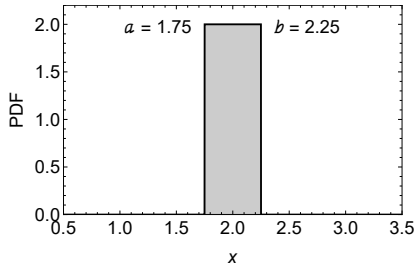
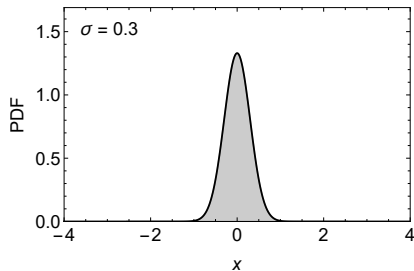
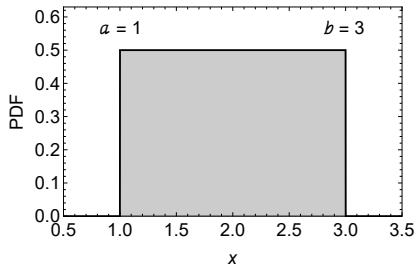
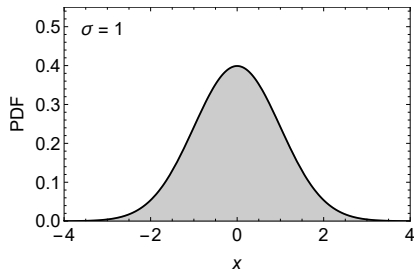
Dla zmiennych losowych dyskretnych, dystrybuanta to

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



Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej



Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Caveat emptor!

Gęstość prawdopodobieństwa f to nie jest prawdopodobieństwo P :

- $0 \leq P \leq 1$
- $f \geq 0$

Różnica jest taka jak między gęstością materii a masą: przecałkowanie gęstości po objętości daje masę. Gęstość większa od 1 (j.a.) może dać masę mniejszą niż 1 (j.a.).

Przecałkowanie f daje P —samo f może być arbitralnie duże, w szczególności większe niż 1 (w całej dziedzinie lub pewnym jej przedziale).

Rozkłady prawdopodobieństwa

Rozkłady ciągłej zmiennej losowej

Caveat emptor!

Gęstość prawdopodobieństwa f to nie jest prawdopodobieństwo P :

- $0 \leq P \leq 1$
- $f \geq 0$

Różnica jest taka jak między gęstością materii a masą: przecałkowanie gęstości po objętości daje masę. Gęstość większa od 1 (j.a.) może dać masę mniejszą niż 1 (j.a.).

Przecałkowanie f daje P —samo f może być arbitralnie duże, w szczególności większe niż 1 (w całej dziedzinie lub pewnym jej przedziale).

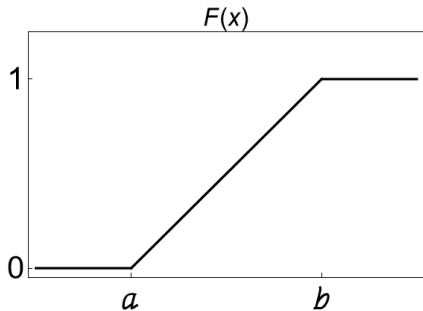
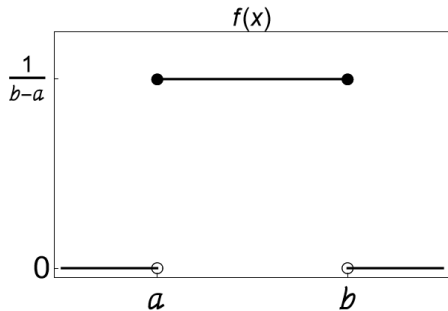
Przykładem zmiennej losowej ciągłej jest wzrost człowieka (por. slajd 2)—w praktyce mierzony z pewną dokładnością, jednak łatwo przyjąć, że w istocie może przyjąć wszystkie wartości (w ramach pewnego przedziału, np. [55, 272] cm—mówimy tu o skończonym *nośniku*).

Rozkłady prawdopodobieństwa

Rozkład jednostajny, *uniform distribution*, $\mathcal{U}(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{dla } a \leq x \leq b \\ 0 & \text{dla } x < a \vee x > b \end{cases}$$

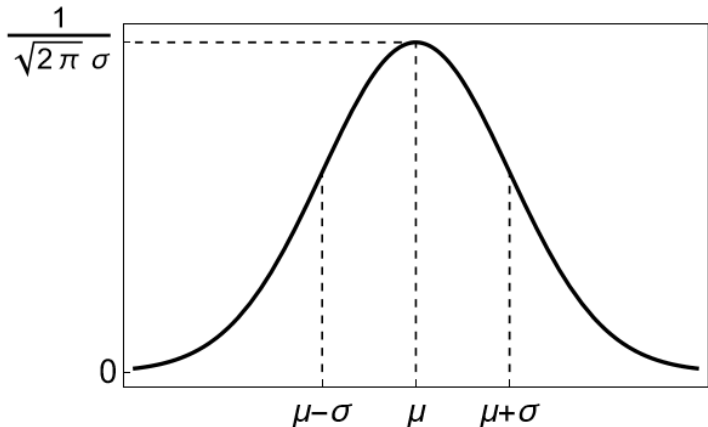
Równoważny zapis: $f(x) = \frac{1}{b-a} \mathbb{1}_{x \in [a, b]}$, gdzie $\mathbb{1}_\omega = 1$ jeśli warunek ω jest spełniony i $\mathbb{1}_\omega = 0$ w przeciwnym wypadku ($\mathbb{1}$ — *funkcja wskaźnikowa*).



Rozkłady prawdopodobieństwa

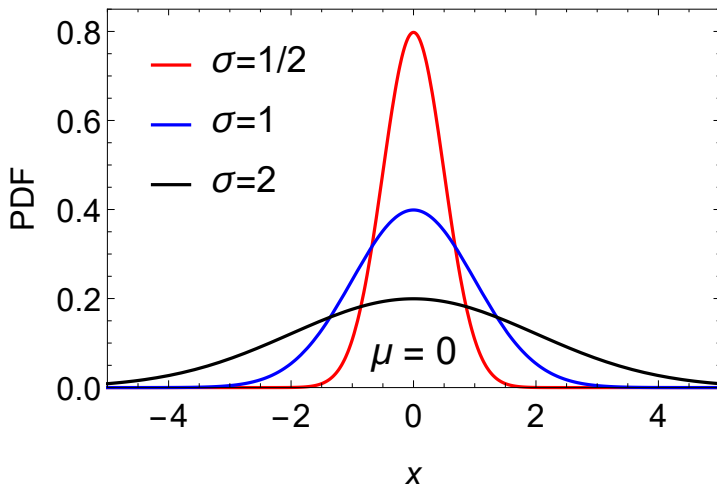
Rozkład Gaussa (normalny)

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$



Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)



Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

Rozkład Gaussa zależy od dwóch parametrów:

μ —średnia (określa położenie);

σ —odchylenie standardowe (określa dyspersję).

Punkty $x = \mu \pm \sigma$ są punktami przegięcia krzywej.

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

Rozkład Gaussa zależy od dwóch parametrów:

μ —średnia (określa położenie);

σ —odchylenie standardowe (określa dyspersję).

Punkty $x = \mu \pm \sigma$ są punktami przegięcia krzywej.

Prawdopodobieństwo, że realizacja zmiennej losowej X nie będzie się różnić od μ o więcej niż jedno odchylenie standardowe (1σ) wynosi

$$P(|X - \mu| \leq \sigma) \approx 0.683$$

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

Rozkład Gaussa zależy od dwóch parametrów:

μ —średnia (określa położenie);

σ —odchylenie standardowe (określa dyspersję).

Punkty $x = \mu \pm \sigma$ są punktami przegięcia krzywej.

Prawdopodobieństwo, że realizacja zmiennej losowej X nie będzie się różnić od μ o więcej niż jedno odchylenie standardowe (1σ) wynosi

$$P(|X - \mu| \leq \sigma) \approx 0.683$$

Mamy też:

$$P(|X - \mu| \leq 2\sigma) \approx 0.954$$

$$P(|X - \mu| \leq 3\sigma) \approx 0.997$$

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

Rozkład Gaussa zależy od dwóch parametrów:

μ —średnia (określa położenie);

σ —odchylenie standardowe (określa dyspersję).

Punkty $x = \mu \pm \sigma$ są punktami przegięcia krzywej.

Prawdopodobieństwo, że realizacja zmiennej losowej X nie będzie się różnić od μ o więcej niż jedno odchylenie standardowe (1σ) wynosi

$$P(|X - \mu| \leq \sigma) \approx 0.683$$

Mamy też:

$$P(|X - \mu| \leq 2\sigma) \approx 0.954$$

$$P(|X - \mu| \leq 3\sigma) \approx 0.997$$

Reguła „trzy sigma”: wielkości oddalone od średniej (wartości oczekiwanej) o więcej niż 3σ są mało prawdopodobne—zaobserwowanie takiej wielkości może być statystycznie istotne (ale może być, z $P \approx 0.003$, zwykłą fluktuacją; zdarzeniem o małym prawdopodobieństwie, ale nie niemożliwym).

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

- $f(x; 0, \sigma) = f(-x; 0, \sigma)$ oraz $f(x; \mu, \sigma) > 0$ dla $-\infty < x < +\infty$.
- Można **ustandaryzować**: $x \rightarrow \frac{x-\mu}{\sigma}$; mamy wtedy $\mathcal{N}(0, 1)$, którego PDF, $f(x; 0, 1)$, ozn. jako $\phi(x)$ — to jego wartości są tablicowane.
- $\phi(x)$ unormowany, zatem $f(x; \mu, \sigma) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$ — też unormowany:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1$$

- Dystrybuanta

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$$

gdzie $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ to tzw. funkcja błędu.

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

- $f(x; 0, \sigma) = f(-x; 0, \sigma)$ oraz $f(x; \mu, \sigma) > 0$ dla $-\infty < x < +\infty$.
- Można **ustandaryzować**: $x \rightarrow \frac{x-\mu}{\sigma}$; mamy wtedy $\mathcal{N}(0, 1)$, którego PDF, $f(x; 0, 1)$, ozn. jako $\phi(x)$ — to jego wartości są tablicowane.
- $\phi(x)$ unormowany, zatem $f(x; \mu, \sigma) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$ — też unormowany:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1$$

- Dystrybuanta

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$$

gdzie $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ to tzw. funkcja błędu.

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

- $f(x; 0, \sigma) = f(-x; 0, \sigma)$ oraz $f(x; \mu, \sigma) > 0$ dla $-\infty < x < +\infty$.
- Można **ustandaryzować**: $x \rightarrow \frac{x-\mu}{\sigma}$; mamy wtedy $\mathcal{N}(0, 1)$, którego PDF, $f(x; 0, 1)$, ozn. jako $\phi(x)$ — to jego wartości są tablicowane.
- $\phi(x)$ unormowany, zatem $f(x; \mu, \sigma) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$ — też unormowany:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1$$

- Dystrybuanta

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$$

gdzie $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ to tzw. funkcja błędu.

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

- $f(x; 0, \sigma) = f(-x; 0, \sigma)$ oraz $f(x; \mu, \sigma) > 0$ dla $-\infty < x < +\infty$.
- Można **ustandaryzować**: $x \rightarrow \frac{x-\mu}{\sigma}$; mamy wtedy $\mathcal{N}(0, 1)$, którego PDF, $f(x; 0, 1)$, ozn. jako $\phi(x)$ — to jego wartości są tablicowane.
- $\phi(x)$ unormowany, zatem $f(x; \mu, \sigma) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$ — też unormowany:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1$$

- Dystrybuanta

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$$

gdzie $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ to tzw. funkcja błędu.

Rozkłady prawdopodobieństwa

Rozkład Gaussa (normalny)

