Wykład III: Zmienne losowe

Alicja Janic

Politechnika Wrocławska *alicja.janic@pwr.edu.pl*

19 październik 2020

Zmienna Losowa

Definicja

Zmienną losową nazywamy każdą funkcję $X : \Omega \longrightarrow \mathrm{R}$ taką, że dla dowolnego $a \in \mathrm{R}$ zbiór

$$(X < a) = \{\omega \in \Omega : X(\omega) < a\}$$

jest zdarzeniem losowym, czyli $\{\omega \in \Omega : X(\omega) < a\} \in \mathcal{F}$ dla dowolnego $a \in \mathrm{R}$

Z własności rodziny $\mathcal{F}$ wynika, że zdarzeniami losowymi są też wszystkie zbiory postaci: $(X \leqslant a)$, $(X > a)$, $(X \geqslant a)$, $(a < X < b)$, $(a < X \leqslant b)$, $(a \leqslant X < b)$

Dystrybuanta zmiennej losowej

Definicja

Dystrybuantą zmiennej losowej $X : \Omega \to \mathrm{R}$ nazywamy funkcję $F_X : \mathrm{R} \to [0, 1]$ określoną wzorem:

$$F_X(x) = P(X < x)$$

Twierdzenie

Funkcja $F : \mathrm{R} \to \mathrm{R}$ *jest dystrybuantą pewnej zmiennej losowej wtedy i tylko wtedy, gdy:*

a) F jest niemalejąca

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

c) F jest lewostronnie ciągła

Przykłady

Przykład 1

Czy można dobrać stałe a, b tak, by funkcja $F(x)$ była dystrybuantą pewnej zmiennej losowej?

$$F(x) = \begin{cases} ae^x & \text{gdy} \quad x \leqslant -1 \\ 0.5 & \text{gdy} \quad -1 < x \leqslant 1 \\ b(2 - \frac{1}{x}) & gdy \quad x > 1 \end{cases}$$

Rozwiązanie

Sprawdźmy, czy istnieją stałe a, b, dla których $F(x)$ spełnia założenia twierdzenia

- $F(x)$ jest lewostronnie ciągła
- $\lim_{x\to-\infty} ae^x = 0$, więc $a \geqslant 0$
  $\lim_{x\to\infty} b(2-\frac{1}{x}) = 2b = 1$, więc $b = \frac{1}{2}$
- Aby $F(x)$ była niemalejąca musi zachodzić warunek $ae^{-1} \leqslant \frac{1}{2}$ oraz $b = \frac{1}{2}$

Zatem dla każdej pary liczb (a, b), gdzie $0 \leqslant a \leqslant \frac{e}{2}$ oraz $b = \frac{1}{2}$ funkcja $F(x)$ jest dystrybuantą pewnej zmiennej losowej. Tylko dla pary liczb $(\frac{e}{2}, \frac{1}{2})$, $F(x)$ jest ciągła na R

Własności dystrybuanty zmiennej losowej

Własności dystrybuanty

Niech $\lim_{x \to a^+} F(x) = F(a^+)$ Prawdziwe są następujące równości:

- $P(X \geqslant a) = 1 - F(a)$
- $P(a \leqslant X < b) = F(b) - F(a)$
- $P(X = a) = F(a^+) - F(a)$ (Stąd, jeżeli F jest ciągła w punkcie a, to $P(X = a) = 0$)
- $P(X \leqslant a) = F(a^+)$
- Jeżeli X jest typu ciągłego, to $P(X = a) = 0$ dla każdego $a \in \mathrm{R}$

Zmienna dyskretna

Definicja

Mówimy, że zmienna losowa X jest zmienną dyskretną jeżeli X przyjmuje skończenie lub co najwyżej przeliczalnie wiele wartości $x_i,\ i \in \mathrm{I}$ przy czym $P(X = x_i) = p(x_i) = p_i > 0$ oraz $\sum_{i \in \mathrm{I}} p_i = 1$

Rozkład prawdopodobieństwa zmiennej dyskretnej nazywa się funkcją prawdopodobieństwa i zapisuje w postaci

$$\{(x_i, p(x_i)) : i \in \mathrm{I}\}$$

Dystrybuanta zmiennej dyskretnej

Dystrybuanta

Dystrybuanta $F : \mathrm{R} \to [0, 1]$ zmiennej dyskretnej ma postać

$$F(x) = P(X < x) = \sum_{\{i\colon x_i < x\}} p_i$$

Jest to funkcja schodkowa, lewostronnie ciągła o skokach o wartości $p_i = p(x_i)$ w punktach $x_i, i \in \mathrm{I}$

Przykłady

Przykład 2

Gracz rzuca symetryczna kostką do gry. Jeśli wyrzuci „piątkę”, wygrywa 10 zł. Jeśli wyrzuci liczbę podzielną przez 3, wygrywa 5 zł. W pozostałych przypadkach płaci 1 zł. Niech X oznacza wygraną gracza (przy czym przegrana 1 zł to inaczej wygrana - 1 zł). Znajdź i narysuj dystrybuantę zmiennej losowej X. Oblicz $P(X > 0)$

Rozwiązanie

- Rozkład X:
  X przyjmuje tylko trzy wartości: $10, 5$ i -1, przy czym
  $P(X = 10) = 1/6$, $P(X = 5) = 2/6 = 1/3$,
  $P(X = -1) = 1 - P(X = 10) - P(X = 5) = 1/2$
  Dokładniej:
  $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^{\Omega}$, P - prawdopodobieństwo klasyczne
  $X(5) = 10$, $X(3) = X(6) = 5$, $X(1) = X(2) = X(4) = -1$
  $P(X = 10) = P(\{5\}) = \frac{1}{6}$, $P(X = 5) = P(\{3, 6\}) = \frac{1}{3}$,
  $P(X = -1) = P(\{1, 2, 4\}) = \frac{1}{2}$

Rozwiązanie cd.

- Dystrybuanta:

$$\{\omega \in \Omega : \ X(\omega) < x\} = \begin{cases} \emptyset & x \leqslant -1 \\ \{X = -1\} = \{1,2,4\} & -1 < x \leqslant 5 \\ \{X = -1\} \cup \{X = 5\} = & \\ \{1,2,4,3,6\} & 5 < x \leqslant 10 \\ \Omega & x > 10 \end{cases}$$

$$F(x) = P(X < x) = \begin{cases} 0 & \text{dla} \quad x \leqslant -1 \\ 1/2 & \text{dla} \quad -1 < x \leqslant 5 \\ 5/6 & \text{dla} \quad 5 < x \leqslant 10 \\ 1 & \text{dla} \quad x > 10 \end{cases}$$

Rozwiązanie cd.

- Prawdopodobieństwo:

$$P(X > 0) = P(X = 5) + P(X = 10) = 0.5$$

$$P(X > 0) = 1 - \lim_{x \to 0^+} F(x) = 1 - 1/2 = 0.5$$

Wartość średnia

Definicja

Dla dyskretnej zmiennej losowej X o funkcji prawdopodobieństwa $p(\cdot)$ ($p_i = p(x_i)$, gdzie $i \in \mathrm{I}$) wartością średnią (oczekiwaną) X nazywamy liczbę

$$E(X) = \mu_X = \sum_{i \in \mathrm{I}} x_i p_i,$$

gdzie x_i, $i \in \mathrm{I}$ oznaczają wszystkie różne wartości zmiennej losowej

Wartość średnia $h(X)$

Twierdzenie

Niech X będzie dyskretną zmienna losową o wartościach x_i, $i \in \mathbf{I}$ i funkcji prawdopodobieństwa $p(\cdot)$ a h dowolną funkcją rzeczywistą. Wówczas dyskretna zmienna losowa $h(X)$ ma wartość średnią równą

$$E(h(X)) = \mu_{h(X)} = \sum_{i \in \mathbf{I}} h(x_i) p_i$$

Wniosek

$$E(aX + b) = aE(X) + b$$

Wariancja

Definicja

Wariancją *dyskretnej zmiennej losowej* X *o funkcji prawdopodobieństwa* $p(\cdot)$ $(p_i = p(x_i)$, *gdzie* $i \in \mathrm{I})$ *nazywamy liczbę*

$$Var(X) = \sigma_X^2 = \sum_{i \in \mathrm{I}} (x_i - \mu_X)^2 p_i$$

Odchylenie standardowe *definiuje się jako* $\sqrt{\sigma_X^2}$

Wariancja zmiennej losowej X jest wartością średnią kwadratu odchyłki wartości X od swojej wartości średniej i może być zapisana jako $E(X - EX)^2$

Wniosek

$$Var(aX + b) = a^2 Var(X)$$

Wariancja

Fakt

Dla dyskretnej zmiennej losowej X mamy

$$Var(X) = E(X^2) - (EX)^2$$

Moment rzędu k i moment centralny rzędu k

Momenty rzędu k

Dla $k = 1, 2, \ldots$ **moment** m_k **rzędu** k zmiennej losowej X jest zdefiniowany jako wartość średnia zmiennej losowej $Y = X^k$ a **moment centralny** μ_k **rzędu** k jako wartość średnia zmiennej losowej $Z = (X - EX)^k$

$$m_k = E(X^k), \quad \mu_k = E((X - EX)^k)$$

Zauważmy, że $m_1 = E(X)$ oraz $\mu_2 = Var(X)$, zatem wartość średnia jest pierwszym momentem X, a wariancja jego drugim momentem centralnym

Przykłady

Przykład 3

W wyższej szkole prywatnej uczy się 1000 studentów. Badamy zmienną losową X zdefiniowaną jako liczba podręczników przyniesionych na zajęcia przez losowo wybranego studenta w określonym dniu (przyjmujemy, że wszyscy studenci byli tego dnia w szkole). Załóżmy, że rozkład liczby podręczników wśród studentów wyglądał następująco:

Liczba podręczników	0	1	2	3	4	5
Liczba studentów	100	300	250	200	100	50

Obliczyć średnią liczbę podręczników przypadających na jednego studenta, czyli średnią wartość zmiennej losowej X oraz standardowe odchylenie zmiennej losowej X

Rozwiązanie

Funkcja prawdopodobieństwa zmiennej losowej X jest następująca:

x_i	0	1	2	3	4	5
$p(x_i)$	0,10	0,30	0,25	0,20	0,10	0,05

- Watość średnia:

$$\mu_X = \sum_{i=0}^{5} x_i \cdot p(x_i) = 2,05$$

- Standardowe odchylenie:

$$\sigma_X = \sqrt{\mu_{X^2} - (\mu_X)^2} = \sqrt{5,95 - (2,05)^2} = 1,322$$

Przykłady

Przykład 4

Wyznaczyć wartość oczekiwaną i wariancję zmiennej losowej $U = -2X^2 + 3$ jeżeli $E(X) = 2,\ Var(X) = 1,\ E(X^4) = 34.$

Rozkład dwupunktowy

Rozkład dwupunktowy

Zmienna losowa ma rozkład dwupunktowy, jeśli przyjmuje tylko dwie różne wartości x i y. Wówczas, jeśli oznaczymy prawdopodobieństwo przyjęcia wartości x przez p i y przez q, to mamy $q = 1 - p$

wartości	x	y
prawdopodobieństwa	p	q

Rozkład dwupunktowy

Rozkład zerojedynkowy

Zmienna losowa X ma rozkład zerojedynkowy z parametrem p, jeżeli przyjmuje tylko dwie wartości oznaczane przez 1 i 0 (nazywane odpowiednio **sukcesem** i **porażką**) oraz

$$P(X=1)=p, \quad P(X=0)=q=1-p, \quad \text{gdzie } p\in(0,1)$$

Typowe przykłady zmiennych o rozkładzie zerojedynkowym opisują jakość wyrobu (dobry, wadliwy) czy wynik gry (wygrana, przegrana)

Średnia i wariancja

$E(X)=0\cdot(1-p)+1\cdot p=p$
$Var(X)=(0-p)^2\cdot(1-p)+(1-p)^2\cdot p=(1-p)p$

Rozkład dwumianowy - Bernoulli'ego

Schemat dwumianowy

Niech X oznacza liczbę sukcesów w n powtórzeniach eksperymentu (próbach) ze zmienną losową o rozkładzie zerojedynkowym, przy czym

- liczba prób jest z góry ustalona i wynosi n
- każda próba kończy się jednym z dwóch możliwych wyników, które są takie same dla wszystkich prób
- wyniki prób nie zależą od siebie
- prawdopodobieństwo sukcesu p jest takie samo w każdej próbie

Rozkład dwumianowy - Bernoulli'ego

Rozkład dwumianowy

Zmienna losowa X ma rozkład dwumianowy z parametrami n i p, a fakt taki zapisujemy

$$X \sim \mathcal{B}in(n, p)$$

jeśli jej rozkład dany jest wzorem

$$P(X = k) = \binom{n}{k} \times p^k (1-p)^{n-k}, \quad \text{dla } k = 0, 1, ..., n$$

Łatwo sprawdzić, wykorzystując wzór na dwumian Newtona, że nieujemne wartości $P(X = k)$ sumują się do 1:

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

Rozkład dwumianowy - Bernoulli'ego

Fakt

Zauważmy, że jeżeli X_i $(i = 1, 2, ..., n)$ są niezależnymi zmiennymi losowymi, z których każda ma rozkład $\mathcal{B}in(1, p)$ (zerojedynkowy), to zmienna losowa

$$Y = X_1 + X_2 + ... + X_n$$

opisuje łączną liczbę sukcesów w tych n próbach, czyli ma rozkład dwumianowy z parametrami n i p, tzn. $\mathcal{B}in(n, p)$

Średnia i wariancja dla $\mathcal{B}in(n, p)$

a) $E(Y) = E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n) = nE(X_1) = np$
b) $Var(Y) = Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n) = nVar(X_1) = np(1 - p)$

Najbardziej prawdopodobna wartość

Najbardziej prawdopodobna wartość

Wartość k_0, którą zmienna losowa dyskretna X przyjmuje z największym prawdopodobieństwem, nazywamy **najbardziej prawdopodobną wartością** X

Aby znaleźć k_0 rozwiązujemy: $P(X = k + 1) > P(X = k)$

Dla rozkładu $\mathcal{B}in(n, p)$

Jeżeli X ma rozkład $\mathcal{B}in(n, p)$, to

$$k_0 = [p(n + 1)]$$

Przykłady

Przykład 5

W pewnym biurze zainstalowano 10 drukarek. Każda z drukarek pracuje niezależnie średnio przez 12 minut w ciągu jednej godziny. a) Jakie jest prawdopodobieństwo, że w losowo wybranej chwili będzie włączonych co najmniej 1 drukarka? 9 drukarek? co najwyżej 9 drukarek? b) Jaka jest najbardziej prawdopodobna liczba drukarek włączonych w danej chwili?

Rozwiązanie

Jeżeli drukarki pracują niezależnie średnio przez 12 minut w ciągu jednej godziny, to zmienna losowa X oznaczająca liczbę drukarek włączonych w danym momencie ma rozkład Bernoulli'ego z parametrami $n = 10$ oraz $p = \frac{12}{60} = \frac{1}{5}$ Zatem prawdopodobieństwo, że w losowo wybranej chwili będzie włączona przynajmniej 1 drukarka wynosi

$P(X \geqslant 1) = 1 - P(X = 0) = 1 - \binom{10}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{10} = 1 - \left(\frac{4}{5}\right)^{10} = 0,8926$

Prawdopodobieństwo, że będzie włączonych 9 drukarek wynosi

$P(X = 9) = \binom{10}{9}\left(\frac{1}{5}\right)^9\left(\frac{4}{5}\right) = 8\left(\frac{1}{5}\right)^9 = 0,000004096$

Co najwyżej 9 drukarek wynosi

$P(X \leqslant 9) = 1 - P(X = 10) = 1 - \binom{10}{10}\left(\frac{1}{5}\right)^{10}\left(\frac{4}{5}\right)^0 = 1 - \left(\frac{1}{5}\right)^{10} = 0,9999998976$

Przykłady

Przykład 6

Co jest bardziej prawdopodobne: wygrać z równorzędnym przeciwnikiem nie mniej niż 3 partie z 4 partii, czy nie mniej niż 5 partii z 8 partii?

Rozwiązanie

Zmienna losowa X określająca liczbę wygranych spotkań w czterech partiach ma rozkład $\mathcal{B}in\left(4, \frac{1}{2}\right)$, a zmienna losowa Y określająca liczbę wygranych spotkań w ośmiu partiach ma rozkład $\mathcal{B}in\left(8, \frac{1}{2}\right)$ Otrzymujemy zatem

$$P(X \geqslant 3) = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = 5\left(\frac{1}{2}\right)^4 = \frac{5}{16} = \frac{80}{256}$$

$$P(Y \geqslant 5) =$$

$$\binom{8}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3+\binom{8}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2+\binom{8}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)+\binom{8}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0 = \frac{93}{256}$$

Czyli bardziej prawdopodobne jest wygranie przynajmniej pięciu spośród ośmiu partii niż przynajmniej trzech spośród czterech partii

Twierdzenie Poissona

Twierdzenie

Jeżeli (X_n) *jest ciągiem zmiennych losowych o rozkładzie dwumianowym* $\mathcal{B}in(n, p_n)$, *przy czym* $lim_{n\to\infty} np_n = \lambda$, *to dla każdego* $k \in \mathrm{N} \cup \{0\}$ *zachodzi równość*

$$lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Rozkład Poissona

Mówimy, że zmienna losowa X ma rozkład Poissona z parametrem λ, $\lambda > 0$ (co zapisujemy $X \sim \mathcal{P}(\lambda)$), jeżeli

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad \text{dla} \;\; k = 0, 1, 2, \ldots$$

Rozkład Poissona

Rozkład Poissona

- Przybliżanie rozkładu Bernoulli'ego rozkładem Poissona jest stosowane w przypadku, gdy n jest duże ($n \geqslant 50$) a p - małe tak, by $p \leqslant 0.1$ $np = \lambda \leqslant 10$
- Najbardziej prawdopodobną wartością zmiennej lososwej o rozkładzie Poissona z parametrem λ jest $k_0 = [\lambda]$
- Jeżeli $X \sim \mathcal{P}(\lambda)$, to wartość średnia X jest równa wariancji X

$$E(X) = \lambda, \quad Var(X) = \lambda$$

Przykłady

Przykład 7

Wiadomo, że 1% produkowanych żarówek to braki. Obliczyć dokładnie i w przybliżeniu, prawdopodobieństwo, że:

a) wśród losowo wybranych 100 żarówek nie ma ani jednej wybrakowanej

b) wśród losowo wybranych 100 żarówek są co najmniej 2 wybrakowane

c) jaka jest minimalna liczba żarówek, które należy sprawdzić, by prawdopodobieństwo znalezienia złej żarówki było nie mniejsze niż 0,9

Rozwiązanie

Niech X oznacza liczbę wybrakowanych żarówek wśród 100 wylosowanych. Wówczas

$$X \sim \mathcal{B}in(100, 0.01) \quad \text{oraz} \quad X \approx \mathcal{P}(\lambda = 100 \cdot 0.01 = 1)$$

a) $P(X = 0) = \binom{100}{0} \left(\frac{1}{100}\right)^0 \left(\frac{99}{100}\right)^{100} \approx e^{-1} \cdot \frac{1^0}{0!} \approx 0.368$
b) $P(X \geqslant 2) = 1 - P(X = 0) - P(X = 1) \approx 1 - 2e^{-1} \approx 0.264$
c) Niech Y oznacza liczbę wybrakowanych żarówek wśród n wylosowanych. Y ma rozkład dwumianowy $X \sim \mathcal{B}in(n, 0.01)$
$P(Y \geqslant 1) = 1 - P(Y = 0) \geqslant 0.9 \implies P(Y = 0) \leqslant 0.1$
Zatem szukamy n takiego, że $P(Y = 0) = \binom{n}{0} \left(\frac{1}{100}\right)^0 \left(\frac{99}{100}\right)^n \leqslant 0.1$
Stąd $n \geqslant 230$

Przykłady

Przykład 8

Z miesięcznej obserwacji małego skrzyżowania wynika, że między godziną 11.00 a 12.00 pojawiają się tam średnio 4 ciężarówki o ładowności ponad 3,5 tony. Zakładając, że momenty ich pojawiania się w ustalonym dniu mogą być modelowane za pomocą procesu Poissona, obliczyć prawdopodobieństwo, że między 11.00 a 11.30 nie pojawi się żadna taka ciężarówka

Rozwiązanie

Oznaczając przez X liczbę ciężarówek na skrzyżowaniu między 11.00 a 12.00 tego dnia. Wiemy, że X ma rozkład Poissona z parametrem $\lambda_1 = (12 - 11)\lambda = \lambda$, gdzie λ jest intensywnością procesu Poissona przejazdu ciężarówek w ciągu dnia. Ponieważ wiemy również, że $E(X) = \lambda_1 = 4$, a stąd $\lambda = 4$. Rozważmy teraz liczbę Y analogicznych zdarzeń między 11.00 a 11.30. Oczywiście, odcinek czasu między 11.00 a 11.30 jest zawarty w całkowitym czasie obserwacji tego dnia, a zatem zdarzenia w tym czasie tworza również proces Poissona. Tak więc zmienna losowa Y ma rozkład Poissona z parametrem $\lambda_2 = \frac{1}{2}\lambda = \frac{1}{2}4 = 2$. Zatem

$$P(Y = 0) = e^{-2} \cdot \frac{2^0}{0!} \approx 0.135$$

Przykłady

Przykład 9

Liczba samochodów, które ulegają wypadkowi w ciągu jednej dnia w danym mieście ma rozkład Poissona. Średnia liczba uszkodzonych samochodów wynosi 6. Jaka jest najbardziej prawdopodobna ilość uszkodzonych samochodów? Ile miejsc należy przygotować na stacjach obsługi, by z prawdopodobieństwem 0, 95 było wolne miejsce dla uszkodzonego samochodu? Zakładamy, że każdy uszkodzony samochodów wymaga naprawy na stacji

Rozwiązanie

Najbardziej prawdopodobną liczbą uszkodzonych samochodów jest 6. Korzystając z tablic rozkładu Poissona dla $\lambda = 6$ wyznaczamy liczbę n, taką że

$$P(X < n) > 0,95 \quad \text{lub} \quad P(X \geqslant n) < 1 - 0,95$$

Otrzymujemy $n = 11$

Zmienna ciągła

Definicja

Zmienną losową X nazywamy ciągłą zmienną losową, jeśli dla pewnej nieujemnej funkcji f i takich dowolnych liczb a i b, że $-\infty \leqslant a < b \leqslant \infty$ zachodzi równość

$$P(a \leqslant X < b) = \int_a^b f(s)ds$$

Funkcję f nazywamy **gęstością prawdopodobieństwa**

Zauważmy, że przyjmując w powyższej równości $a = -\infty$, otrzymujemy, że dystrybuanta $F_X(\cdot)$ spełnia równość

$$F_X(b) = \int_{-\infty}^b f(s)ds$$

Stwierdzenia, fakty i twierdzenia

Stwierdzenie

Dla ciągłej zmiennej losowej X o dystrybuancie F zachodzi

$$P(a < X < b) = P(a < X \leqslant b) = P(a \leqslant X < b) = P(a \leqslant X \leqslant b) = \\ = F(b) - F(a)$$

Fakt

Wiemy z analizy, że funkcja F, dla zmiennych losowych typu ciągłego, jest ciągła. Ponadto jest ona różniczkowalna we wszystkich punktach ciągłości funkcji f i w punktach tych zachodzi równość

$$F'(x) = f(x)$$

Stwierdzenia, fakty i twierdzenia

Twierdzenie

Funkcja f jest gęstością pewnej zmiennej losowej wtedy i tylko wtedy, gdy:

- $f(x) \geqslant 0$ *dla każdego* $x \in R$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Przykłady

Przykład 10

Niech gęstość f pewnej zmiennej losowej X wynosi

$$f(x) = \begin{cases} \frac{2}{3} + x^2 & \text{dla } 0 < x < 1 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Znajdź dystrybuantę F zmiennej losowej X oraz prawdopodobieństwo $P(X > 0.5)$

Wartość średnia i kwantyl rzędu p

Definicja

Wartością średnią ciągłej zmiennej losowej o gęstości f nazywamy wielkość

$$E(X) = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$$

Definicja

Kwantylem rzędu p dla $0 < p < 1$ ciągłej zmiennej losowej jest dowolny taki punkt x_p, że

$$F(x_p) = p$$

W szczególności, mediana jest określona jako kwantyl rzędu 0.5

Wartość średnia $h(X)$

Twierdzenie

Niech X będzie zmienną losową o rozkładzie ciągłym i h dowolną funkcją określoną na zbiorze wartości X. Wówczas dla zmiennej losowej $Y = h(X)$ mamy

$$E(Y) = E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx$$

Wniosek

W szczególności

$$E(aX + b) = aE(X) + b$$

Wariancja

Definicja

Wariancją *ciągłej zmiennej losowej o gęstości f nazywamy wielkość*

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty}(x - EX)^2 f(x)dx$$

Odchylenie standardowe *definiuje się jako* $\sigma_X = \sqrt{Var(X)}$

Wariancja zmiennej losowej X jest wartością średnią kwadratu odchyłki wartości X od swojej wartości średniej i może być zapisana jako $E(X - EX)^2$

Wniosek

$$Var(aX + b) = a^2 Var(X)$$

Wariancja

Fakt

Dla dowolnej ciągłej zmiennej losowej X mamy również

$$Var(X) = E(X^2) - (EX)^2$$

Rozkład normalny

Gęstość rozkładu normalnego

Zmienna losowa X ma rozkład normalny z parametrami m, σ ($m \in$ R, $\sigma > 0$), jeżeli jej gęstość ma postać

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-m)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

Piszemy

$$X \sim \mathcal{N}(m, \sigma)$$

Rozkład normalny jest najważniejszym rozkładem w teorii prawdopodobieństwa. Został wprowadzony w XVIIIw. przez Gaussa i Laplace'a

Rozkład normalny

Dystrybuanta rozkładu normalnego

Dystrybuanty rozkładu normalnego $\mathcal{N}(m, \sigma)$, czyli funkcji

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(t-m)^2}{2\sigma^2}} dt$$

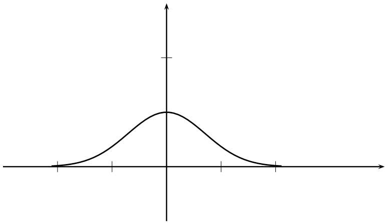
nie można wyrazić przez funkcje elementarne. Wartości dystrybuanty rozkładu $\mathcal{N}(0, 1)$, czyli funkcji

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

podane są w tablicach

Rozkład normalny

Wykres gęstości rozkładu $N(0,1)$ jest następujący



Dystrybuanta rozkładu standardowego normalnego

Z symetrii wykresu gęstości względem osi Oy otrzymujemy wygodną w obliczeniach równość:

$$\Phi(-x) = 1 - \Phi(x)$$

Rozkład normalny

Okazuje się, że wartości dystrybuanty dowolnego rozkładu $\mathcal{N}(m, \sigma)$ można obliczyć, znając wartości funkcji $\Phi(x)$

Standaryzacja rozkładu normalnego

Jeżeli $X \sim \mathcal{N}(m, \sigma)$, to zmienna losowa $Z = \frac{X-m}{\sigma}$ ma rozkład $\mathcal{N}(0, 1)$ oraz

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right)$$

Ostatni fakt daje następujący, często wykorzystywany wzór

$$P(a < X < b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)$$

Rozkład normalny

Prawo trzech sigm

Korzystając ze standaryzacji i z tablic rozkładu $\mathcal{N}(0,1)$ łatwo sprawdzić, że gdy X ma rozkład $\mathcal{N}(m,\sigma)$, to

$$P(m-3\sigma < X < m+3\sigma) = P\left(-3 < \frac{X-m}{\sigma} < 3\right) = 2\Phi(3) - 1 = \\ = 2\cdot 0.9987 - 1 \geqslant 0.997$$

Oznacza to, że wartości zmiennej X z prawdopodobieństwem bliskim 1 zawarte są w przedziale $(m-3\sigma, m+3\sigma)$

Wartość średnia i standardowe odchylenie

Stwierdzenie

Jeśli X jest zmienną losowa o rozkładzie normalnym $\mathcal{N}(m, \sigma)$, to wartość średnia X jest równa m, a odchylenie standardowe równe σ:

$$\mu_X = E(X) = m \quad \text{oraz} \quad \sigma_X = \sqrt{Var(X)} = \sigma$$

Jeżeli $Z \sim \mathcal{N}(0, 1)$, to

$$\mu_Z = E(Z) = 0 \quad \text{oraz} \quad \sigma_Z^2 = Var(Z) = 1$$

Zatem

$$E(X) = E(\sigma Z + m) = \sigma E(Z) + m = m$$
$$Var(X) = Var(\sigma Z + m) = \sigma^2 Var(Z) = \sigma^2$$

Przykłady

Przykład 11

Średnica metalowych kulek produkowanych przez automat jest zmienną losową X o rozkładzie $\mathcal{N}(0.5, 0.04)$. Za zgodne z normą uznaje się kulki o średnicy z przedziału $[0.41, 0.59]$. Obliczyć prawdopodobieństwo, że wybrana losowo z produkcji kulka spełnia wymagania normy. Jaka jest najbardziej prawdopodobna liczba kulek spełniajacych wymagania normy wśród 1000 kulek?

Rozwiązanie

Szukane prawdopodobieństwo obliczymy dokonując standaryzacji zmiennej losowej X oraz wykorzystamy tablice rozkładu $\mathcal{N}(0, 1)$

$$P(0.41 < X < 0.59) = P\left(\frac{0.41-0.5}{0.04} < \frac{X-0.5}{0.04} < \frac{0.59-0.5}{0.04}\right) =$$
$$= \Phi(2.25) - \Phi(-2.25) = 2\Phi(2.25) - 1 = 0.9756$$

Zmienna losowa Y określająca liczbę kulek spełniających wymagania normy, wśród 1000 kulek wyprodukowanych ma rozkład $\mathcal{B}in(1000, 0.9756)$ i jej wartością najbardziej prawdopodobną jest $[1001 \cdot 0.9756] = 976$

Rozkład jednostajny

Rozkład jednostajny na odcinku $[a, b]$

Zmienna losowa X ma rozkład jednostajny na odcinku $[a, b]$, jeżeli jej gęstość jest postaci

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{gdy} \quad x \in [a, b] \\ 0 & \text{gdy} \quad x \in \mathrm{R} \setminus [a, b] \end{cases}$$

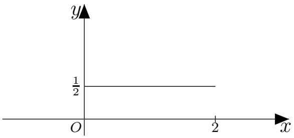
Piszemy

$$X \sim \mathcal{U}(a, b)$$

Rozkład jednostajny

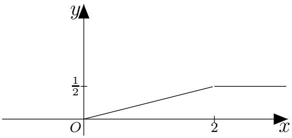
Dystrybuanta

$$F(x) = \begin{cases} 0 & \text{gdy} \quad x \leqslant a \\ \frac{x-a}{b-a} & \text{gdy} \quad a < x \leqslant b \\ 1 & \text{gdy} \quad x > b \end{cases}$$



Rys. 1.03.

Gęstość rozkładu jednostajnego na przedziale [0,2]



Rys. 1.03.

Dystrybuanta rozkładu jednostajnego na przedziale [0,2]

Przykłady

Przykład 12

Z przystanku autobusy odjeżdżają co 10 minut. Zakładamy, że rozkład T czasu przybycia pasażera na przystanek jest zmienną losową o rozkładzie jednostajnym. Obliczyć prawdopodobieństwo, że pasażer będzie czekał co najmniej 4 minuty, mniej niż 3 minuty

Rozwiązanie

Zauważmy, że $T \sim \mathcal{U}(0, 10)$. Zatem

$$P(T < 3) = F_T(3) = \frac{3}{10}$$

oraz

$$P(T \geqslant 4) = 1 - P(T < 4) = 1 - F_T(4) = 1 - \frac{4}{10} = \frac{6}{10}$$

Przykłady

Przykład 13

Automat produkuje kulki metalowe o średnicy X będącej zmienną losową o gęstości

$$f(x) = \begin{cases} 5 & \text{gdy} \quad x \in [0.4, 0.6] \\ 0 & \text{gdy} \quad x \in \mathrm{R} \setminus [0.4, 0.6] \end{cases}$$

Za zgodne z normą uznaje się kulki o średnicy z przedziału $[0.41, 0.59]$. Obliczyć prawdopodobieństwo, że wybrana losowo z produkcji kulka spełnia wymagania normy. Jaka jest najbardziej prawdopodobna liczba kulek spełniających wymagania normy wśród 1000 kulek?

Rozkład wykładniczy

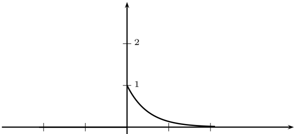
Rozkład wykładniczy z parametrem λ

Zmienna losowa X ma rozkład wykładniczy z parametrem λ, $\lambda > 0$, jeżeli jej gęstość ma postać

$$f(x) = \begin{cases} 0 & \text{gdy} \quad x \leqslant 0 \\ \lambda e^{-\lambda x} & \text{gdy} \quad x > 0 \end{cases}$$

Piszemy

$$X \sim \mathcal{E}xp(\lambda)$$



Rozkład wykładniczy

Dystrybuanta

Wówczas dystrybuanta jest postaci

$$F(x) = \begin{cases} 0 & \text{gdy} \quad x \leqslant 0 \\ 1 - e^{-\lambda x} & \text{gdy} \quad x > 0 \end{cases}$$