Wykłady 5 i 6. Zmienne losowe typu ciągłego

dr Mariusz Grządziel

5,12 marca 2012

Pole trapezu krzywoliniowego
Przypomnienie: figurę ograniczoną przez:

- wykres funkcji $y = f(x)$, gdzie f jest funkcją ciągłą;
- proste $x = a$, $x = b$, $a < b$,
- oś OX (tj. prostą $y = 0$)

będziemy nazywać *trapezem krzywoliniowym* (odpowiadającym funkcji f oraz odcinkowi $[a, b]$).
Pole tej figury można przedstawić w postaci całki:

$$\int_a^b f(x)dx.$$

Pole „nieograniczonego" trapezu krzywoliniowego-całka niewłaściwa
Problem: jesteśmy zainteresowani polem figury ograniczonej: wykresem funkcji $f(x) = e^{-x}$ oraz prostymi $y = 0$, $x = 0$.
Pole tego obszaru można określić jako całkę:

$$\lim_{T\to\infty} \int_0^T f(x)dx.$$

Korzystając z faktu:

$$\int_0^T e^{-x}dx = -e^{-T} + e^0 = 1 - e^{-T}$$

znajdujemy, że granica ta jest równa 1.

Całka niewłaściwa z funkcji nieujemnej
Całkę niewłaściwą z funkcji nieujemnej f na półprostej $[a, \infty)$ można określić jako granicę:

$$\lim_{T\to\infty} \int_a^T f(t)dt$$

jeśli ona istnieje.
Analogicznie można całkę niewłaściwą z funkcji nieujemnej f na półprostej $(-\infty, b]$.
Całkę niewłaściwą z funkcji nieujemnej f na prostej $(-\infty, \infty)$ można określić jako granicę $\lim_{T\to\infty} \int_{-T}^T f(t)dt$, jeśli ona istnieje.

Zmienne losowe typu ciagłego

Definicja 1. *Mówimy, że zmienna losowa X jest typu ciągłego, jeśli istnieje nieujemna funkcja g (spełniająca pewne łagodne warunki- np. jest przedziałami ciagła) taka, że dla każdych $a < b$*

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

Rozkład jednostajny na odcinku $[0,1]$
Przykładem zmiennej losowej typu ciągłego jest rozkład jednostajny na odcinku $[0,1]$ (oznaczenie: $U(0,1)$). Jego funkcja gestości u dana jest wzorem

$$u(x) = \begin{cases} 1, & \text{jeśli } 0 \leqslant x \leqslant 1, \\ 0 & \text{jeśli } x < 0 \text{ lub } x > 1. \end{cases}$$

Rozkład ten może opisywać np. czas oczekiwania na autobus A, odjeżdżający do miejscowości B co godzinę, przez pasażera C; zakładamy, że C nie zna rozkładu jazdy dla tej linii i że przychodzi na przystanek w losowym momencie.

Rozkład jednostajny na odcinku $[0,1]$-przykład obliczeń
Czas oczekiwania na autobus- zmienna losowa $Y \sim U(0,1)$. Prawdopodobieństwo $P\big(\frac{1}{3} < Y < \frac{1}{2}\big)$ jest równe:

$$P(\frac{1}{3} < Y < \frac{1}{2}) = \int_{1/3}^{1/2} 1dx = \big[x\big]_{1/3}^{1/2} = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

Prawdopodobieństwa odpowiadające nierównościom ostrym i słabym
Dla zmiennej losowej X o rozkładzie typu ciągłego mamy:

$$P(a < X < b) = P(a \leqslant X < b) = P(a < X \leqslant b) = P(a \leqslant X \leqslant b).$$

Równość ta wynika z własności całki oznaczonej.

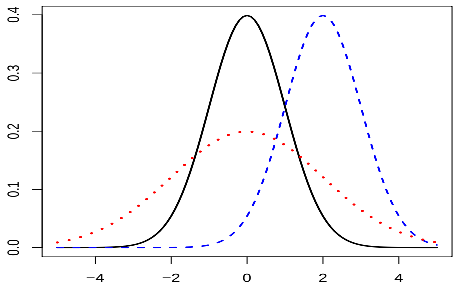
Rozkład normalny
Szczególnie ważnym w zastosowaniach jest rozkład normalny.

Definicja 2. *Mówimy, że zmienna losowa X ma rozkład normalny z parametrami μ i σ, gdzie $\mu \in \mathbb{R}$ i $\sigma > 0$, jeżeli gęstość jej rozkładu jest określona wzorem:*

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

Skrótowy zapis: $X \sim N(\mu,\sigma)$. Dla $\mu = 0$ i $\sigma = 1$ będziemy pisać zamiast $\phi_{\mu,\sigma}(x)$ krótko $\phi(x)$.

Rozkład normalny— zastosowania
Wiele cech (zmiennych losowych) w życiu gospodarczym, w świecie przyrody ma rozkład zbliżony do normalnego.
Wynika to z tzw. centralnego twierdzenia granicznego, z którego wynika, że średnia $\frac{1}{n}(X_1 + X_2 + \ldots + X_n)$, gdzie $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, ma rozkład zbliżony do normalnego $N(\mu,\sigma)$ dla pewnych μ i σ. Dokładniejsze sformułowanie tego twierdzenia wymaga określenia wartości oczekiwanej i wariancji zmiennej losowej typu ciągłego.



Rysunek 1: Wykresy gęstości rozkładów normalnych: $N(0,1)$ (linia ciągła), $N(0,2)$ (linia „kropkowana"), $N(2,1)$ (linia „kreskowana").

Wartość oczekiwana i wariancja dla zmiennych losowych typu ciagłego
Wartość oczekiwana i wariancja zmiennej losowej X typu ciągłego zdefiniowana jest wzorami:

$$\mu = E(X) = \int_{-\infty}^{\infty} xg(x)dx, \tag{1}$$

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 g(x)dx, \tag{2}$$

gdzie g jest funkcją gęstości zmiennej X.

Wartość oczekiwana i wariancja zmiennej losowej o rozkładzie $U(0,1)$
Niech $X \sim U(0,1)$. Funkcja gęstosci g jest równa 1 na $[0,1]$; poza tym przedziałem jest równa 0. Mamy:

$$E(X) = \int_{-\infty}^{\infty} xg(x)dx = \int_0^1 xdx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

oraz

$$Var(X) = \int_{-\infty}^{\infty} (x-\frac{1}{2})^2 g(x)dx = \\ = \int_0^1 (x-\frac{1}{2})^2 dx = \left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}.$$

Wartość oczekiwana i wariancja zmiennej losowej o rozkładzie normalnym
Można pokazać, że jeśli $X \sim N(\mu, \sigma)$, to

$$E(X) = \mu, \quad Var(X) = \sigma^2, \quad D(X) = \sigma.$$

Wymaga to obliczenia całek troche bardziej skomplikowanych niż dla przypadku odpowiadającego $U(0,1)$.

Obliczanie prawdopodobieństw w rozkładzie normalnym-$N(0,1)$
Dla $a < b$ prawdopodobieństwo $P(a < X < b)$, gdzie $X \sim N(0,1)$ jest równe:

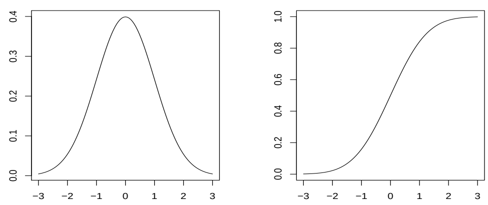
$$P(a < X < b) = \int_a^b \phi(x)dx = \Phi(b) - \Phi(a),$$

gdzie Φ jest określona przez:

$$\Phi(t) = \int_{-\infty}^t \phi(x)dx.$$

Funkcja Φ jest dystrybuantą rozkładu normalnego $N(0,1)$. Funkcji Φ nie da się wyrazić za pomocą skończnej liczby działań na podstawowych funkcjach elementarnych — stąd potrzeba sporządzania tablic statystycznych zawierających wartości funkcji Φ (można je znaleźć w prawie każdym podręczniku statystyki).

Własności funkcji Φ



Rysunek 2: Wykresy gęstości ϕ rozkładu normalnego (z lewej strony) $N(0,1)$ i dystrybuanty rozkładu normalnego Φ (z prawej strony)

Można pokazać, że $\Phi(0) = 0{,}5$ oraz $\Phi(t) = 1 - \Phi(-t)$ dla dowolnego t; stąd można się ograniczyć do tablicowania funkcji Φ dla $t \geqslant 0$.

Obliczanie prawdopodobieństw w rozkładzie normalnym-$N(\mu,\sigma)$
Można pokazać, że jeśli $X \sim N(\mu,\sigma)$, to

$$\frac{X-\mu}{\sigma} \sim N(0,1).$$

Wynika to z własności wartości oczekiwanej i dyspersji (omówionych podczas poprzedniego wykładu).
Stąd dla $a < b$ prawdopodobieństwo $P(a < X < b)$, $X \sim N(\mu,\sigma)$ jest równe:

$$P(a < X < b) = P\Big(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\Big) = \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi(\frac{a-\mu}{\sigma}).$$

Obliczanie prawdopodobieństw w rozkładzie normalnym— przykład
Niech X oznacza wzrost dorosłych mężczyzn w panstwie A; zakładamy, że $X \sim N(177, 10)$. Chcemy obliczyć: (a) $P(174 < X < 182)$, (b) $P(X > 182)$. Obliczenia dla (a):

$$\begin{aligned} P(174 < X < 182) &= \Phi\Big(\frac{182-177}{10}\Big) - \Phi\Big(\frac{174-177}{10}\Big) \\ &= \Phi(0{,}5) - \Phi(-0{,}3) = \Phi(0{,}5) - (1-\Phi(0{,}3)) = \\ &= \Phi(0{,}5) + \Phi(0{,}3) - 1 \approx 0{,}6915 + 0{,}6179 - 1 = 0{,}3094. \end{aligned}$$

Obliczenia dla (b) można przeprowadzić w analogiczny sposób, korzystając z równości:

$$P(X > 182) = 1 - P(X < 182) = 1 - \Phi\Big(\frac{182-175}{10}\Big) = 1 - \Phi(0{,}5).$$

Centralne Twierdzenie Graniczne

Twierdzenie 1. *Jeżeli $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, $E(X_1) = \mu$, $D(X_1) = \sigma$, to zmienna losowa*

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

gdzie $\bar{X} = \frac{1}{n}\big(X_1 + X_2 + \ldots + X_n\big)$, ma w przybliżeniu standardowy rozkład normalny $N(0,1)$, tj.

$$P\Big(a \leqslant \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant b\Big) \to \Phi(b) - \Phi(a)$$

dla dowolnych a i b, $a < b$

Innymi słowy, rozkład $\bar{X}$ jest w przybliżeniu równy rozkładowi normalnemu $N(\mu, \sigma/\sqrt{n})$.

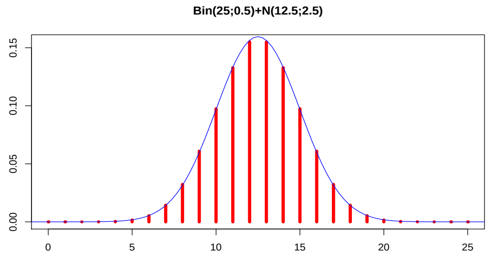
Przykład
Rzucamy monetą 25 razy. Niech Y oznacza liczbę orłów. X ma rozkład $Bin(25; 0{,}5)$. Zmienna Y może być przedstawiona jako suma 25 niezależnych zmiennych losowych $X_1, \ldots, X_{25}$ o rozkładzie $Bin(1; 0{,}5)$. Korzystając z Centralnego Twierdzenia Granicznego: $\bar{X} \sim N(p, \sigma/\sqrt{n})$, gdzie $n = 25, p = 0{,}5, \sigma = \sqrt{p(1-p)} = 0{,}5$, czyli

$$\bar{X} \text{ ma w przybliżeniu rozkład } N(0{,}5; 0{,}5/5)$$

więc

$$Y \text{ ma w przybliżeniu rozkład } N(12{,}5; 2{,}5)$$

Cechy o rozkładzie normalnym w życiu gospodarczym i w przyrodzie
Centralne twierdzenie graniczne: sumy wyników niezależnych eksperymentów losowych mają rozkład normalny (w przybliżeniu).
Można oczekiwać, że wielkości takie jak: suma wydatków dziesięciu kolejnych klientów w sklepie; wynik ankiety określającej preferencje osób ankietowanych itd. będą miały rozkład zbliżony do normalnego.



Rysunek 3: Rozkład dwumianowy $Bin(25; 0{,}5)$ i odpowiadający mu rozkład normalny $N(12{,}5; 2{,}5)$

Inne rozkłady ciągłe

Dowolna funkcja g spełniająca warunki: (i) dziedziną funkcji g jest zbiór liczb rzeczywistych $\mathbb{R}$; (ii) $g(x) \geqslant 0$ dla $x \in \mathbb{R}$; (iii) $\int_{-\infty}^{\infty} g(x) = 1$. jest funkcją gestością pewnej zmiennej losowej; Poza rozkładem normalnym i rozkładem jednostajnym $U(0, 1)$ do opisu cech w życiu gospodarczym i naukach przyrodniczych stosuje się wiele innych rozkładów prawdopodobieństwa (wykładniczy, t-Studenta itd.). Również histogram probabilistyczny spełnia warunki (i)–(iii)- rozkład odpowiadający histogramowi probabilistycznemu jest niekiedy określany mianem histogramu empirycznego.

Lektura uzupełniająca

T. Bednarski, Elementy matematyki w naukach ekonomicznych. Oficyna ekonomiczna. Kraków 2004, str. 234–244.
Koronacki, J., Mielniczuk, J. Statystyka dla studentów kierunków technicznych i przyrodniczych. WNT. Warszawa 2001, s. 111–118.
A. Walanus, Krzywa dzwonowa. http://www.statsoft.pl/czytelnia/statdlaka/dzwon.html
Tablica z wartościami funkcji Φ, dystrybuanty rozkładu normalnego $N(0, 1)$, http://www.math.unb.ca/ knight/utility/NormTble.htm; Probability Content from $-\infty$ to Z odpowiada wartościom $\Phi(Z), \quad Z = 0; 0{,}01; \ldots$