Statystyka i rachunek prawdopodobieństwa

Wykład 8: Centralne twierdzenie graniczne

Mariusz Tarnopolski

Instytut Astronomii UMK

Twierdzenie 1. Jeśli zmienna losowa X ma wartość przeciętną $\mathbb{E}(X)$ oraz wariancję $V(X)$, to przy $\varepsilon > 0$ zachodzi

$$P\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{V(X)}{\varepsilon^2}$$

Twierdzenie 2. Jeśli zmienna losowa X ma wartość przeciętną $\mathbb{E}(X)$ oraz wariancję $V(X)$, to przy $\varepsilon > 0$ zachodzi

$$P\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{V(X)}{\varepsilon^2}$$

Dowód. Niech X ma rozkład dyskretny. Wtedy

$$\begin{aligned}
V(X) &= \sum_i \left[x_i - \mathbb{E}(X)\right]^2 p_i \\
&\geqslant \sum_{|x_i - \mathbb{E}(X)| \geqslant \varepsilon} \left[x_i - \mathbb{E}(X)\right]^2 p_i \\
&\geqslant \varepsilon^2 \sum_{|x_i - \mathbb{E}(X)| \geqslant \varepsilon} p_i \\
&= \varepsilon^2 P\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right)
\end{aligned}$$

Dowód dla zmiennej losowej ciągłej przebiega analogicznie. ■

Powyżej jest najbardziej ogólne sformułowanie twierdzenia Czebyszewa. Jego interpretacja staje się bardziej intuicyjna jeśli podstawimy $\varepsilon = k\sigma$, czyli wyrazimy nierówności w odniesieniu do wielokrotności (niekoniecznie całkowitej) odchylenia standardowego:

$$P\left(|X - \mathbb{E}(X)| \geqslant k\sigma\right) \leqslant \frac{V(X)}{(k\sigma)^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$$

To z kolei oznacza, że prawdopodobieństwo odchylenia realizacji zmiennej losowej od wartości oczekiwanej o co najmniej $k\sigma$ jest nie większe niż $1/k^2$.

Powyżej jest najbardziej ogólne sformułowanie twierdzenia Czebyszewa. Jego interpretacja staje się bardziej intuicyjna jeśli podstawimy $\varepsilon = k\sigma$, czyli wyrazimy nierówności w odniesieniu do wielokrotności (niekoniecznie całkowitej) odchylenia standardowego:

$$P\left(|X - \mathbb{E}(X)| \geqslant k\sigma\right) \leqslant \frac{V(X)}{(k\sigma)^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$$

To z kolei oznacza, że prawdopodobieństwo odchylenia realizacji zmiennej losowej od wartości oczekiwanej o co najmniej $k\sigma$ jest nie większe niż $1/k^2$. Np. co najmniej 75% wyników będzie leżeć (**bez względu na rozkład!**) w przedziale $[\mathbb{E}(X) - 2\sigma, \mathbb{E}(X) + 2\sigma]$ (bo $1 - 1/2^2 = 3/4$).

Powyżej jest najbardziej ogólne sformułowanie twierdzenia Czebyszewa. Jego interpretacja staje się bardziej intuicyjna jeśli podstawimy $\varepsilon = k\sigma$, czyli wyrazimy nierówności w odniesieniu do wielokrotności (niekoniecznie całkowitej) odchylenia standardowego:

$$P\left(|X - \mathbb{E}(X)| \geqslant k\sigma\right) \leqslant \frac{V(X)}{(k\sigma)^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$$

To z kolei oznacza, że prawdopodobieństwo odchylenia realizacji zmiennej losowej od wartości oczekiwanej o co najmniej $k\sigma$ jest nie większe niż $1/k^2$. Np. co najmniej 75% wyników będzie leżeć (**bez względu na rozkład!**) w przedziale $[\mathbb{E}(X) - 2\sigma, \mathbb{E}(X) + 2\sigma]$ (bo $1 - 1/2^2 = 3/4$). Nierówność Czebyszewa daje najgorszy możliwy scenariusz rozrzutu pomiarów wokół wartości oczekiwanej.

Powyżej jest najbardziej ogólne sformułowanie twierdzenia Czebyszewa. Jego interpretacja staje się bardziej intuicyjna jeśli podstawimy $\varepsilon = k\sigma$, czyli wyrazimy nierówności w odniesieniu do wielokrotności (niekoniecznie całkowitej) odchylenia standardowego:

$$P\left(|X - \mathbb{E}(X)| \geqslant k\sigma\right) \leqslant \frac{V(X)}{(k\sigma)^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$$

To z kolei oznacza, że prawdopodobieństwo odchylenia realizacji zmiennej losowej od wartości oczekiwanej o co najmniej $k\sigma$ jest nie większe niż $1/k^2$.
Np. co najmniej 75% wyników będzie leżeć (**bez względu na rozkład!**) w przedziale $[\mathbb{E}(X) - 2\sigma, \mathbb{E}(X) + 2\sigma]$ (bo $1 - 1/2^2 = 3/4$).
Nierówność Czebyszewa daje najgorszy możliwy scenariusz rozrzutu pomiarów wokół wartości oczekiwanej.
Dla porównania, dla rozkładu Gaussa $P\left(|X - \mu| \leqslant 2\sigma\right) \approx 95\%$, czyli $P\left(|X - \mu| \geqslant 2\sigma\right) \approx 5\% < 25\%$.

Twierdzenie 3. Niech Y_n będzie ciągiem zmiennych losowych o rozkładzie dwumianowym:

$$P(Y_n = k) = \binom{n}{k} p^k q^{n-k}, \qquad k \in \{0, 1, 2, \ldots, n\}$$

Wówczas ciąg zmiennych losowych $Z_n = Y_n/n$ jest zbieżny (wg prawdopodobieństwa) do liczby p, tzn.

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P(|Z_n - p| > \varepsilon) = 0$$

lub równoważnie:

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P(|Z_n - p| < \varepsilon) = 1$$

co oznacza, że dla dowolnie małego $\varepsilon > 0$ i dla dostatecznie dużych n:

$$P(|Z_n - p| < \varepsilon) \approx 1$$

Dowód. Dla zmiennej losowej Y_n o rozkładzie dwumianowym zachodzi $\mathbb{E}(Y_n) = np$, $V(Y_n) = npq$, a zatem

$$\mathbb{E}(Z_n) = \frac{\mathbb{E}(Y_n)}{n} = p, \qquad V(Z_n) = \frac{V(Y_n)}{n^2} = \frac{pq}{n}$$

Przyjmując jakiekolwiek $\varepsilon > 0$ oraz korzystając z nierówności Czebyszewa otrzymujemy

$$P\left(|Z_n - p| \geqslant \varepsilon\right) \leqslant \frac{V(Z_n)}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

■

Prawo wielkich liczb Bernoulliego mówi, że można być niemal pewnym, że powtarzając doświadczenie losowe (na zmiennych Bernoulliego) wielką liczbę razy częstość pojawiania się sukcesu, k/n, będzie bliska teoretycznej wartości p.

Prawo wielkich liczb Bernoulliego mówi, że można być niemal pewnym, że powtarzając doświadczenie losowe (na zmiennych Bernoulliego) wielką liczbę razy częstość pojawiania się sukcesu, k/n, będzie bliska teoretycznej wartości p.
Prawo wielkich liczb można uznać za teoretyczny odpowiednik znanego faktu empirycznego — stabilności częstości zdarzeń losowych. Jednak stabilność częstości nie jest konsekwencją prawa wielkich liczb. Prawo to nie mówi bowiem nic o wynikach rzeczywiście wykonywanych doświadczeń losowych. Dotyczy ono nie samych doświadczeń, a ich probabilistycznych modeli.

Twierdzenie 4. Niech $X_1, X_2, X_3, \ldots$ będzie ciągiem niezależnych zmiennych losowych o tym samym (dowolnym) rozkładzie[1] i wartością oczekiwaną $m = \mathbb{E}(X)$.

Wówczas ciąg zmiennych losowych $Z_n = \frac{1}{n}\sum_i X_i$ jest zbieżny (wg prawdopodobieństwa) do liczby m, tj.

$$\lim_{n\to\infty} P\left(|Z_n - m| < \varepsilon\right) = 1$$

[1] iid — *independent and identically distributed*

Twierdzenie 5. Niech $X_1, X_2, X_3, \ldots$ będzie ciągiem niezależnych zmiennych losowych o tym samym (dowolnym) rozkładzie[1] i wartością oczekiwaną $m = \mathbb{E}(X)$.
Wówczas ciąg zmiennych losowych $Z_n = \frac{1}{n} \sum_i X_i$ jest zbieżny (wg prawdopodobieństwa) do liczby m, tj.

$$\lim_{n \to \infty} P\left(|Z_n - m| < \varepsilon\right) = 1$$

Dowód. Przy dodatkowym założeniu istnienia wariancji zmiennych X_i (tj. $V(X_i) < \infty$), dowód przebiega analogicznie jak dowód twierdzenia wielkich liczb Bernoulliego. Dowód bez tego warunku jest znacznie trudniejszy. ■

[1] iid — *independent and identically distributed*

Twierdzenie 6. Dla ciągu niezależnych zmiennych losowych $X_1, X_2, X_3, \ldots$ o tym samym rozkładzie oraz skończonych wartościach oczekiwanych $\mathbb{E}(X_i) = m$ i wariancjach $V(X_i) = \sigma^2$, ciąg dystrybuant $F_n(y)$ standaryzowanych sum tych zmiennych:

$$Y_n = \sum_{i=1}^{n} \frac{X_i - m}{\sqrt{n}\sigma} = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$$

jest zbieżny do dystrybuanty rozkładu normalnego:

$$\lim_{n\to\infty} F_n(y) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{y} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t = \Phi(y)$$

Dowód. Weźmy funkcję charakterystyczną zmiennej $(X_1 - m)/(\sigma\sqrt{n})$:

$$\varphi_1(t) = \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{X_1 - m}{\sigma\sqrt{n}}\right)\right]$$

Skoro funkcja charakterystyczna sumy niezależnych zmiennych losowych jest iloczynem funkcji charakterystycznych, to dla zmiennej losowej Y_n zachodzi

$$\varphi_n(t) = [\varphi_1(t)]^n$$

Rozwijając $\varphi_1(t)$ w szereg Maclaurina względem argumentu $t/\sqrt{n}$ otrzymujemy

$$\varphi_1(t) = \varphi_1(0) + t\varphi_1'(0) + \frac{t^2}{2}\varphi_1''(0) + \mathcal{O}\left(\frac{t^3}{n^{3/2}}\right)$$

gdzie $\mathcal{O}[(t/\sqrt{n})^k]$ oznacza wyrazy rzędu co najmniej $(t/\sqrt{n})^k$.

Z własności funkcji charakterystycznej mamy, że

$$\varphi_1'(0) = \mathrm{i}\mathbb{E}\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right) = 0$$

$$\varphi_1''(0) = \mathrm{i}^2\mathbb{E}\left[\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right)^2\right] = -\frac{1}{n}$$

ponieważ $\mathbb{E}\left[(X_1 - m)^2\right] = \sigma^2$. Zatem

$$\varphi_1(t) = 1 - \frac{t^2}{2n} + \mathcal{O}\left(\frac{t^3}{n^{3/2}}\right)$$

wobec czego

$$\varphi_n(t) = \left[1 - \frac{t^2}{2n} + \mathcal{O}\left(\frac{t^3}{n^{3/2}}\right)\right]^n$$

Z analizy matematycznej wiadomo, że

$$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = \mathrm{e}^a$$

więc, przyjmując $a = -t^2/2$, otrzymujemy

$$\lim_{n\to\infty}\varphi_n(t) = \mathrm{e}^{-\frac{t^2}{2}}$$

co jest funkcją charakterystyczną standardowego rozkładu normalnego; przy czym skorzystaliśmy z faktu, że $\mathcal{O}\left(\frac{1}{n}\right) \xrightarrow[n\to\infty]{} 0.$ ∎

Należy podkreślić, że zmienne losowe w CLT mogą mieć rozkład zarówno ciągły, jak i dyskretny. Z tezy CLT wynika, że dla dużych n (w praktyce — kilkunastu, jednak konkretna wartość zależy od wymaganej precyzji) obliczane prawdopodobieństwa można przybliżać przy pomocy dystrybuanty rozkładu normalnego:

$$P(a \leqslant Y_n \leqslant b) \approx \Phi(b) - \Phi(a)$$

Należy podkreślić, że zmienne losowe w CLT mogą mieć rozkład zarówno ciągły, jak i dyskretny. Z tezy CLT wynika, że dla dużych n (w praktyce — kilkunastu, jednak konkretna wartość zależy od wymaganej precyzji) obliczane prawdopodobieństwa można przybliżać przy pomocy dystrybuanty rozkładu normalnego:

$$P(a \leqslant Y_n \leqslant b) \approx \Phi(b) - \Phi(a)$$

Przykład. (Twierdzenie Moivre'a (1773 r.)-Laplace'a (1812 r.))
Dla ciągu niezależnych zmiennych losowych $X_1, X_2, X_3, \ldots$ o tym samym rozkładzie dwumianowym ($\mathbb{E}(X_i) = p$, $V(X_i) = pq$), ciąg dystrybuant $F_n(y)$ standaryzowanych sum tych zmiennych:

$$Y_n = \frac{\sum\limits_{i=1}^{n} (X_i - np)}{\sqrt{npq}}$$

jest zbieżny do dystrybuanty rozkładu normalnego, $\Phi(y)$.

Wniosek. Skoro zmienna $Y_n = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ zbiega do rozkładu $\mathcal{N}(0,1)$, to zmienna $\bar{X}_n$ zbiega do rozkładu $\mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$ — tzn. że odchylenie standardowe rozkładu średnich wynosi $\frac{\sigma}{\sqrt{n}}$, tj. niepewność (68% przedział ufności — zob. dalsze wykłady) wyznaczenia średniej $\bar{X}_n$ wynosi $\frac{\sigma}{\sqrt{n}}$.

Wniosek. Skoro zmienna $Y_n = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ zbiega do rozkładu $\mathcal{N}(0,1)$, to zmienna $\bar{X}_n$ zbiega do rozkładu $\mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$ — tzn. że odchylenie standardowe rozkładu średnich wynosi $\frac{\sigma}{\sqrt{n}}$, tj. niepewność (68% przedział ufności — zob. dalsze wykłady) wyznaczenia średniej $\bar{X}_n$ wynosi $\frac{\sigma}{\sqrt{n}}$.

Podkreślmy:

Niepewność średniej obliczonej z n pomiarów wynosi $\frac{\sigma}{\sqrt{n}}$, gdzie σ to odchylenie standardowe rozkładu pomiarów.

Wniosek. Skoro zmienna $Y_n = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ zbiega do rozkładu $\mathcal{N}(0,1)$, to zmienna $\bar{X}_n$ zbiega do rozkładu $\mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$ — tzn. że odchylenie standardowe rozkładu średnich wynosi $\frac{\sigma}{\sqrt{n}}$, tj. niepewność (68% przedział ufności — zob. dalsze wykłady) wyznaczenia średniej $\bar{X}_n$ wynosi $\frac{\sigma}{\sqrt{n}}$.

Podkreślmy:

Niepewność średniej obliczonej z n pomiarów wynosi $\frac{\sigma}{\sqrt{n}}$,
gdzie σ to odchylenie standardowe rozkładu pomiarów.
Niepewność średniej $\neq$ odchylenie standardowe!
Niepewność średniej $=$ odchylenie standardowe podzielone przez $\sqrt{n}$.

Wniosek. Skoro zmienna $Y_n = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ zbiega do rozkładu $\mathcal{N}(0,1)$, to zmienna $\bar{X}_n$ zbiega do rozkładu $\mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$ — tzn. że odchylenie standardowe rozkładu średnich wynosi $\frac{\sigma}{\sqrt{n}}$, tj. niepewność (68% przedział ufności — zob. dalsze wykłady) wyznaczenia średniej $\bar{X}_n$ wynosi $\frac{\sigma}{\sqrt{n}}$.

Podkreślmy:

Niepewność średniej obliczonej z n pomiarów wynosi $\frac{\sigma}{\sqrt{n}}$,
gdzie σ to odchylenie standardowe rozkładu pomiarów.
Niepewność średniej $\neq$ odchylenie standardowe!
Niepewność średniej $=$ odchylenie standardowe podzielone przez $\sqrt{n}$.

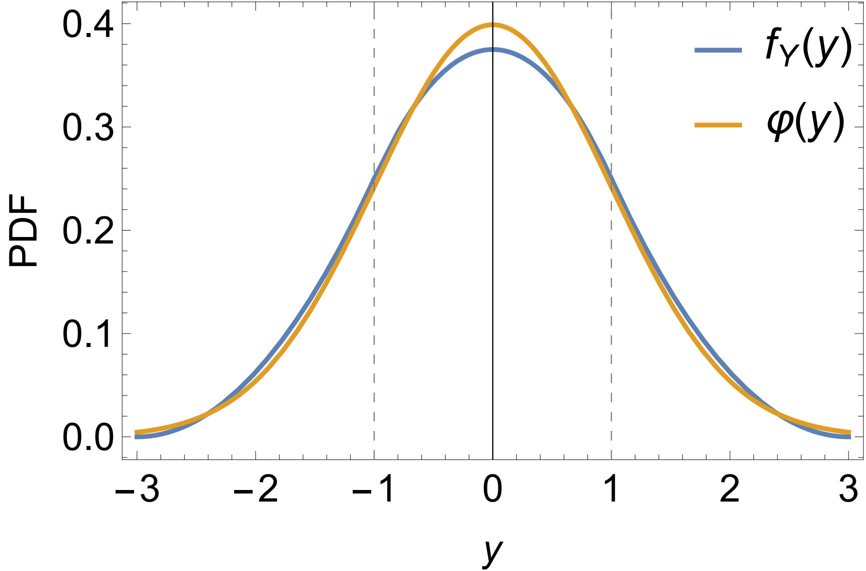
Licząc wprost: $V(\bar{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Przykład. Niech $X_i \sim \mathcal{U}(0,1)$; zdefiniujmy $Z = X_1 + X_2 + X_3$. Żmudne i nudne rachunki prowadzą do (tzw. rozkład Irwina-Halla dla sumy 3 zmiennych)

$$f_Z(z) = \begin{cases} \frac{1}{2}z^2 & 0 \leqslant z \leqslant 1 \\ \frac{1}{2}(-2z^2+6z-3) & 1 < z \leqslant 2 \\ \frac{1}{2}(z^2-6z+9) & 2 < z \leqslant 3 \\ 0 & z < 0 \wedge z > 3 \end{cases}$$

Zmienna ustandaryzowana $Y = \dfrac{Z - {}^3/_2}{{}^1/_2}$ ma z kolei rozkład

$$f_Y(y) = \begin{cases} \frac{1}{16}(y^2+6y+9) & -3 \leqslant y \leqslant -1 \\ \frac{1}{8}(-y^2+3) & -1 < y \leqslant 1 \\ \frac{1}{16}(y^2-6y+9) & 1 < y \leqslant 3 \\ 0 & |y| > 3 \end{cases}$$


0.4
0.3
0.2
0.1
0.0
PDF
−3
−2
−1
0
1
2
3
y
$f_Y(y)$
$\varphi(y)$

Twierdzenia graniczne

Centralne twierdzenie graniczne (*central limit theorem, CLT*), twierdzenie Lindeberga-Fellera (Lindeberg 1922 r., Feller 1935 r.)

Twierdzenie Lindeberga-Lévy'ego ma swoje ograniczenia — w szczególności, mówi explicite o **iid zmiennych losowych**. W wielu praktycznych sytuacjach mamy do czynienia z sumą niezależnych zmiennych losowych **o różnych rozkładach**. Wtedy zachodzi

Twierdzenie 7. Niech będzie ciąg $X_1, X_2, X_3, \ldots$ niezależnych zmiennych losowych takich, że

$$S_n = \sum_{i=1}^{n} X_i, \qquad \sigma_n^2 = \sum_{i=1}^{n} V(X_i)$$

Ustandaryzujmy zmienną S_n:

$$Y_n = \frac{S_n - \mathbb{E}(S_n)}{\sigma_n}$$

Jeśli spełniony jest warunek Lindeberga:

$$\forall \varepsilon > 0 : \quad \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mathbb{E}(X_i))^2 \mathbb{1}_{|X_i - \mathbb{E}(X_i)| > \varepsilon \sigma_n} \right] \xrightarrow[n \to \infty]{} 0$$

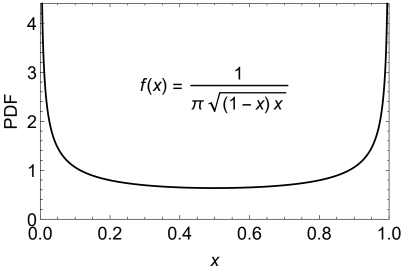
to Y_n ma rozkład $\mathcal{N}(0,1)$.
Ten warunek sprawdza, czy żadna pojedyncza zmienna losowa nie fluktuuje na tyle silnie by dominować w sumie będącej przedmiotem zainteresowania CLT (tj. czy rzadkie duże odchylenia nie są **zbyt** duże w porównaniu z typową skalą fluktuacji).

Funkcja $\mathbb{1}_\omega$ to tzw. funkcja wskaźnikowa:

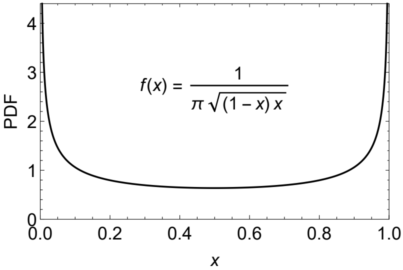
$$\mathbb{1}_\omega = \begin{cases} 1 & \text{jeśli warunek } \omega \text{ jest spełniony} \\ 0 & \text{jeśli warunek } \omega \text{ nie jest spełniony} \end{cases}$$

Np. $\mathbb{1}_{1<2} = 1$, ale $\mathbb{1}_{2>3} = 0$.

CLT — przykład

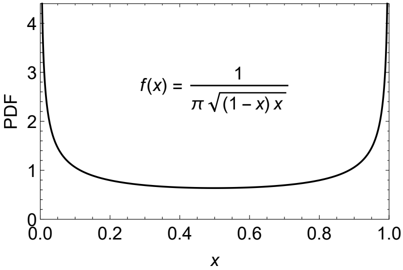


CLT — przykład

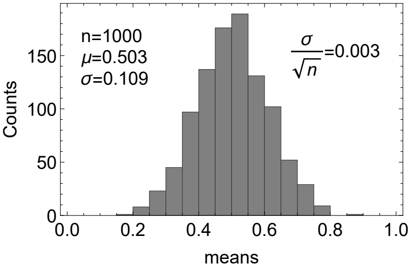


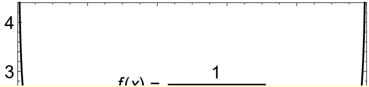
Losujemy 10 wartości:
{0.12, 0.95, 0.77, 0.97, 0.03, 0.32, 0.6, 0.87, 0.84, 0.06}.
Ich średnia to 0.553.
Powtarzamy 1000 razy, otrzymując rozkład średnich:

CLT — przykład

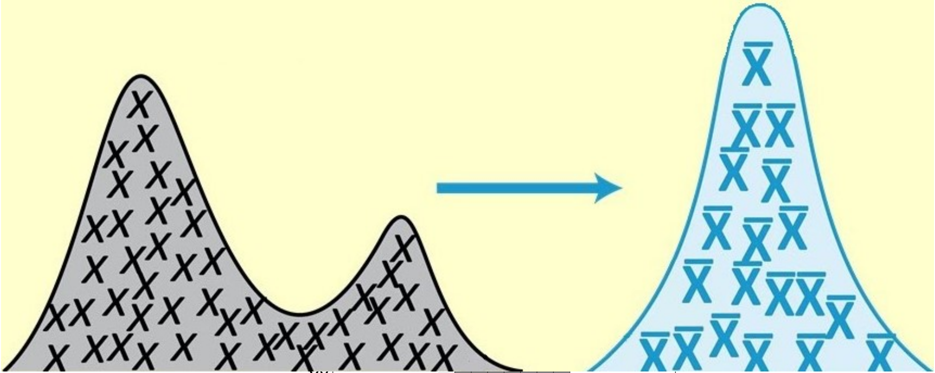


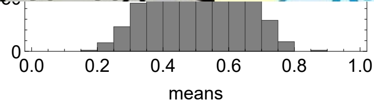
Losujemy 10 wartości:
{0.12, 0.95, 0.77, 0.97, 0.03, 0.32, 0.6, 0.87, 0.84, 0.06}.
Ich średnia to 0.553.
Powtarzamy 1000 razy, otrzymując rozkład średnich:





Losujemy 10 wartości:
{0.12, 0.95, 0.77, 0.97, 0.03,





Kiedy CLT nie działa

- CLT wymaga by wariancja była skończona.

  Przykład. Niech zmienne X_i mają rozkład Cauchy'ego, który nie ma określonych ani wartości oczekiwanej, ani wariancji. Funkcja charakterystyczna to $\varphi_{X_i}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$ też ma funkcję charakterystyczną $\varphi_{\bar{X}_n}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n$ również ma rozkład Cauchy'ego a nie Gaussa — nawet gdy $n \to \infty$.

- Należy podkreślić, że CLT nie mówi „dużo obserwacji ⇒ rozkład normalny" — mówi, że suma (średnia) zmiennych losowych zmierza do rozkładu normalnego.

-
  - Prawa wielkich liczb mówią co się dzieje z wartością oczekiwaną — $\bar{X}_n \to m$
  - CLT mówią, jakie są wokół niej fluktuacje statystyczne — $\sqrt{n}\frac{\bar{X}_n - m}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Kiedy CLT nie działa

- CLT wymaga by wariancja była skończona.

  Przykład. Niech zmienne X_i mają rozkład Cauchy'ego, który nie ma określonych ani wartości oczekiwanej, ani wariancji. Funkcja charakterystyczna to $\varphi_{X_i}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$ też ma funkcję charakterystyczną $\varphi_{\bar{X}_n}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n$ również ma rozkład Cauchy'ego a nie Gaussa — nawet gdy $n \to \infty$.

- Należy podkreślić, że CLT nie mówi „dużo obserwacji $\Rightarrow$ rozkład normalny" — mówi, że suma (średnia) zmiennych losowych zmierza do rozkładu normalnego.

-
  - Prawa wielkich liczb mówią co się dzieje z wartością oczekiwaną — $\bar{X}_n \to m$
  - CLT mówią, jakie są wokół niej fluktuacje statystyczne — $\sqrt{n}\frac{\bar{X}_n - m}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Kiedy CLT nie działa

- CLT wymaga by wariancja była skończona.

  Przykład. Niech zmienne X_i mają rozkład Cauchy'ego, który nie ma określonych ani wartości oczekiwanej, ani wariancji. Funkcja charakterystyczna to $\varphi_{X_i}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$ też ma funkcję charakterystyczną $\varphi_{\bar{X}_n}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n$ również ma rozkład Cauchy'ego a nie Gaussa — nawet gdy $n \to \infty$.

- Należy podkreślić, że CLT nie mówi „dużo obserwacji $\Rightarrow$ rozkład normalny" — mówi, że suma (średnia) zmiennych losowych zmierza do rozkładu normalnego.

-
  - Prawa wielkich liczb mówią co się dzieje z wartością oczekiwaną — $\bar{X}_n \to m$
  - CLT mówią, jakie są wokół niej fluktuacje statystyczne — $\sqrt{n}\frac{\bar{X}_n - m}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Kiedy CLT nie działa

- CLT wymaga by wariancja była skończona.

  Przykład. Niech zmienne X_i mają rozkład Cauchy'ego, który nie ma określonych ani wartości oczekiwanej, ani wariancji. Funkcja charakterystyczna to $\varphi_{X_i}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$ też ma funkcję charakterystyczną $\varphi_{\bar{X}_n}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n$ również ma rozkład Cauchy'ego a nie Gaussa — nawet gdy $n \to \infty$.

- Należy podkreślić, że CLT nie mówi „dużo obserwacji ⇒ rozkład normalny" — mówi, że suma (średnia) zmiennych losowych zmierza do rozkładu normalnego.

-
  - Prawa wielkich liczb mówią co się dzieje z wartością oczekiwaną — $\bar{X}_n \to m$
  - CLT mówią, jakie są wokół niej fluktuacje statystyczne — $\sqrt{n}\frac{\bar{X}_n - m}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Kiedy CLT nie działa

- CLT wymaga by wariancja była skończona.

  Przykład. Niech zmienne X_i mają rozkład Cauchy'ego, który nie ma określonych ani wartości oczekiwanej, ani wariancji. Funkcja charakterystyczna to $\varphi_{X_i}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$ też ma funkcję charakterystyczną $\varphi_{\bar{X}_n}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n$ również ma rozkład Cauchy'ego a nie Gaussa — nawet gdy $n \to \infty$.

- Należy podkreślić, że CLT nie mówi „dużo obserwacji ⇒ rozkład normalny" — mówi, że suma (średnia) zmiennych losowych zmierza do rozkładu normalnego.

-
  - Prawa wielkich liczb mówią co się dzieje z wartością oczekiwaną — $\bar{X}_n \to m$
  - CLT mówią, jakie są wokół niej fluktuacje statystyczne — $\sqrt{n}\frac{\bar{X}_n - m}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Kiedy CLT nie działa

- CLT wymaga by wariancja była skończona.

  Przykład. Niech zmienne X_i mają rozkład Cauchy'ego, który nie ma określonych ani wartości oczekiwanej, ani wariancji. Funkcja charakterystyczna to $\varphi_{X_i}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$ też ma funkcję charakterystyczną $\varphi_{\bar{X}_n}(t) = \mathrm{e}^{-|t|}$, zatem $\bar{X}_n$ również ma rozkład Cauchy'ego a nie Gaussa — nawet gdy $n \to \infty$.

- Należy podkreślić, że CLT nie mówi „dużo obserwacji ⇒ rozkład normalny" — mówi, że suma (średnia) zmiennych losowych zmierza do rozkładu normalnego.
-
  - Prawa wielkich liczb mówią co się dzieje z wartością oczekiwaną — $\bar{X}_n \to m$
  - CLT mówią, jakie są wokół niej fluktuacje statystyczne — $\sqrt{n}\dfrac{\bar{X}_n - m}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Twierdzenie 8. Niech $X_1, X_2, X_3, \ldots, X_n$ bedą iid zmiennymi losowymi takimi, że $\mathbb{E}(X_i) = 0$, $V(X_i) = \sigma^2 < \infty$, $\mathbb{E}(|X_i|^3) = \rho < \infty$.
Wtedy ustandaryzowana zmienna

$$S_n = \frac{\bar{X}\sqrt{n}}{\sigma} = \sum_{i=1}^{n} \frac{X_i}{\sigma\sqrt{n}}$$

o dystrybuancie $F_n(x)$ spełnia

$$\sup_x |F_n(x) - \Phi(x)| \leqslant \frac{C\rho}{\sigma^3\sqrt{n}}$$

gdzie C jest pewną stałą rzędu jedności[2].

[2]Najlepsze obecnie oszacowanie na C to $C \lesssim 0.4748$.

Wniosek.

- Standaryzowana suma n zmiennych losowych zmierza do rozkładu normalnego w tempie $1/\sqrt{n}$ — stosunkowo wolno, a jednak dostatecznie szybko by już dla niewielkich n dostrzec podobieństwo do rozkładu Gaussa (por. rozkład Irwina-Halla ze slajdu 15).
- Tempo zbieżności jest zależne od ρ: cięższe ogony $\Rightarrow$ większe ρ.
- W dowodzie CLT wystarczyło, że reszty $\mathcal{O}(t^3/n^{3/2})$ asymptotycznie znikały; tw. Berry'ego-Esseena określa jak szybko znikają (stąd pojawienie się ρ).
- CLT mówi „dokąd zmierzamy";
  Berry-Esseen mówi „jak szybko tam dojdziemy"...

Wniosek.

- Standaryzowana suma n zmiennych losowych zmierza do rozkładu normalnego w tempie $1/\sqrt{n}$ — stosunkowo wolno, a jednak dostatecznie szybko by już dla niewielkich n dostrzec podobieństwo do rozkładu Gaussa (por. rozkład Irwina-Halla ze slajdu 15).
- Tempo zbieżności jest zależne od ρ: cięższe ogony $\Rightarrow$ większe ρ.
- W dowodzie CLT wystarczyło, że reszty $\mathcal{O}(t^3/n^{3/2})$ asymptotycznie znikały; tw. Berry'ego-Esseena określa jak szybko znikają (stąd pojawienie się ρ).
- CLT mówi „dokąd zmierzamy";
  Berry-Esseen mówi „jak szybko tam dojdziemy"...

Wniosek.

- Standaryzowana suma n zmiennych losowych zmierza do rozkładu normalnego w tempie $1/\sqrt{n}$ — stosunkowo wolno, a jednak dostatecznie szybko by już dla niewielkich n dostrzec podobieństwo do rozkładu Gaussa (por. rozkład Irwina-Halla ze slajdu 15).
- Tempo zbieżności jest zależne od ρ: cięższe ogony $\Rightarrow$ większe ρ.
- W dowodzie CLT wystarczyło, że reszty $\mathcal{O}(t^3/n^{3/2})$ asymptotycznie znikały; tw. Berry'ego-Esseena określa jak szybko znikają (stąd pojawienie się ρ).
- CLT mówi „dokąd zmierzamy”;
  Berry-Esseen mówi „jak szybko tam dojdziemy”...

Wniosek.

- Standaryzowana suma n zmiennych losowych zmierza do rozkładu normalnego w tempie $1/\sqrt{n}$ — stosunkowo wolno, a jednak dostatecznie szybko by już dla niewielkich n dostrzec podobieństwo do rozkładu Gaussa (por. rozkład Irwina-Halla ze slajdu 15).
- Tempo zbieżności jest zależne od ρ: cięższe ogony $\Rightarrow$ większe ρ.
- W dowodzie CLT wystarczyło, że reszty $\mathcal{O}(t^3/n^{3/2})$ asymptotycznie znikały; tw. Berry'ego-Esseena określa jak szybko znikają (stąd pojawienie się ρ).
- CLT mówi „dokąd zmierzamy";
  Berry-Esseen mówi „jak szybko tam dojdziemy"...

Warunek Lapunowa

mówi „czy w ogóle możemy wyruszyć”

- Jeśli istnieje $\delta > 0$ takie, że

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[\left|X_i - \mathbb{E}(X_i)\right|^{2+\delta}\right] \xrightarrow[n\to\infty]{} 0$$

  to zachodzi CLT.

- Ten warunek sprawdza, czy rozkłady zmiennych losowych nie mają zbyt ciężkich ogonów — tzn. czy wyższe momenty odchyleń są na tyle małe w porównaniu z wariancją sumy, aby duże fluktuacje pojedynczych zmiennych nie zaburzały zachowania całej sumy. Stąd skalowanie przez $\sigma_n^{2+\delta}$ — warunek patrzy na wyższe (niż wariancja) momenty statystyczne, które odpowiadają za zachowania rozkładów w ogonach.
- Dla $\delta = 1$ dostajemy $\mathbb{E}(|X_i|^3) < \infty$, czyli dokładnie taki sam warunek jak w tw. Berry'ego-Esseena.

Warunek Lapunowa

mówi „czy w ogóle możemy wyruszyć"

- Jeśli istnieje $\delta > 0$ takie, że

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - \mathbb{E}(X_i)|^{2+\delta}\right] \xrightarrow[n\to\infty]{} 0$$

  to zachodzi CLT.
- Ten warunek sprawdza, czy rozkłady zmiennych losowych nie mają zbyt ciężkich ogonów — tzn. czy wyższe momenty odchyleń są na tyle małe w porównaniu z wariancją sumy, aby duże fluktuacje pojedynczych zmiennych nie zaburzały zachowania całej sumy. Stąd skalowanie przez $\sigma_n^{2+\delta}$ — warunek patrzy na wyższe (niż wariancja) momenty statystyczne, które odpowiadają za zachowania rozkładów w ogonach.
- Dla $\delta = 1$ dostajemy $\mathbb{E}(|X_i|^3) < \infty$, czyli dokładnie taki sam warunek jak w tw. Berry'ego-Esseena.

Warunek Lapunowa

mówi „czy w ogóle możemy wyruszyć"

- Jeśli istnieje $\delta > 0$ takie, że

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[\left|X_i - \mathbb{E}(X_i)\right|^{2+\delta}\right] \xrightarrow[n\to\infty]{} 0$$

  to zachodzi CLT.

- Ten warunek sprawdza, czy rozkłady zmiennych losowych nie mają zbyt ciężkich ogonów — tzn. czy wyższe momenty odchyleń są na tyle małe w porównaniu z wariancją sumy, aby duże fluktuacje pojedynczych zmiennych nie zaburzały zachowania całej sumy. Stąd skalowanie przez $\sigma_n^{2+\delta}$ — warunek patrzy na wyższe (niż wariancja) momenty statystyczne, które odpowiadają za zachowania rozkładów w ogonach.
- Dla $\delta = 1$ dostajemy $\mathbb{E}(|X_i|^3) < \infty$, czyli dokładnie taki sam warunek jak w tw. Berry'ego-Esseena.