

Rachunek Prawdopodobieństwa

Anna Janicka

wykład XIII, 8.01.2019

PRAWA WIELKICH LICZB – CD.

CENTRALNE TWIERDZENIE GRANICZNE

Plan na dzisiaj

23. Prawa Wielkich Liczb

- *Słabe PWL*
- Mocne PWL

24. Centralne Twierdzenie Graniczne

- Twierdzenie de Moivre'a-Laplace'a



Słabe prawa wielkich liczb (SPWL) – przypomnienie

Zbieżność **według prawdopodobieństwa** dla
sum (średnich) dużej liczby zmiennych
losowych

1. Słabe PWL dla schematu Bernoulliego

*Założmy, że $X_1, X_2, \dots$ są niezależne i mają rozkład
 $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0)$, $n = 1, 2, \dots$. Wówczas
(S_n/n) zbiega według prawdopodobieństwa do p (tzn. zmiennej
losowej stałej, równej p); innymi słowy, dla każdego $\varepsilon > 0$,
 $\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} - p \right| > \varepsilon \right) = 0$.*



Słabe prawa wielkich liczb – cd. przypomnienie

2. Słabe PWL dla zmiennych nieskorelowanych

Założmy, że $X_1, X_2, \dots$ jest ciągiem nieskorelowanych zmiennych losowych o wspólnie ograniczonej wariancji. Wówczas $(X_n)_{n \geq 1}$ spełnia słabe prawo wielkich liczb:

$(S_n - \mathbb{E}S_n)/n \xrightarrow{\mathbb{P}} 0$, tzn. dla dowolnego $\varepsilon > 0$ mamy
$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n - \mathbb{E}S_n}{n} \right| > \varepsilon \right) = 0.$$

- zmienne nie muszą być takie same!
- jeśli są takie same (o średniej m), to teza:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} - m \right| > \varepsilon \right) = 0$$



Mocne prawa wielkich liczb (MPWL)

Zbieżność prawie na pewno

3. Mocne PWL dla schematu Bernoulliego

Założmy, że $X_1, X_2, \dots$ są niezależne i mają rozkład $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0)$, $n = 1, 2, \dots$

Wówczas (S_n/n) zbiega prawie na pewno do p (tzn. zmiennej losowej stałej, równej p); innymi słowy, istnieje zdarzenie Ω' pełnej miary takie, że jeśli $\omega \in \Omega'$, to

$$\lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = p.$$


Mocne prawa wielkich liczb – cd.

4. Mocne PWL Kołmogorowa

Założmy, że $X_1, X_2, \dots$ jest ciągiem niezależnych całkownalnych zmiennych losowych o tym samym rozkładzie.

Wówczas $\frac{S_n}{n} \xrightarrow{p.n.} \mathbb{E}X_1$.

- wada: nie znamy tempa zbieżności...
- zastosowania: mnóstwo, w szczególności weryfikacja modelu probabilistycznego (np. paradoks kawalera de Méré), metody Monte Carlo obliczania całek, ...



Zastosowania MPWL

5. Średnia i wariancja z próby są dobre:
jeśli ciąg zmiennych spełnia założenia
MPWL (tzn. IID), to:

$$\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p.n.} \mathbb{E}X_1$$

$$\frac{1}{n} \sum_{k=1}^n (X_k - \overline{X})^2 \xrightarrow{p.n.} \text{Var}X_1$$



Zastosowania MPWL – cd.

6. Jeśli rozważymy dystrybuanty empiryczne związane z próbką liczebności N

$$F_N(t) = \frac{1_{\{X_1 \leq t\}} + 1_{\{X_2 \leq t\}} + \dots + 1_{\{X_N \leq t\}}}{N}$$

to zachodzi **Tw. Gliwenki-Cantelliego**:

Założmy, że $X_1, X_2, \dots$ są niezależne i mają ten sam rozkład o dystrybuancie F . Wówczas

$$\sup_{t \in \mathbb{R}} |F_N(t) - F(t)| \xrightarrow[N \rightarrow \infty]{p.n.} 0.$$



Centralne Twierdzenie Graniczne (CTG)

1. Twierdzenie (CTG):

Założmy, że $X_1, X_2, \dots$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, takim, że $\mathbb{E}X_1^2 < \infty$.

Oznaczmy $m = \mathbb{E}X_1$, $\sigma^2 = \text{Var } X_1$.

Wówczas dla dowolnego $t \in \mathbb{R}$,

$$\mathbb{P} \left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma \sqrt{n}} \leq t \right) \xrightarrow{n \rightarrow \infty} \Phi(t).$$

←
dystrybuanta
rozkładu
normalnego
standardowego

2. Dodatkowe własności:

$$\mathbb{P} \left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma \sqrt{n}} \geq s \right) \xrightarrow{n \rightarrow \infty} 1 - \Phi(s), \text{ oraz}$$

$$\mathbb{P} \left(s \leq \frac{X_1 + X_2 + \dots + X_n - nm}{\sigma \sqrt{n}} \leq t \right) \xrightarrow{n \rightarrow \infty} \Phi(t) - \Phi(s).$$

Centralne twierdzenie graniczne – cd.

3. CTG m.in. opisuje błąd MPWL:

$$\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \leq t$$

$$\Leftrightarrow$$

$$\frac{X_1 + X_2 + \dots + X_n}{n} - m \leq \frac{t\sigma}{\sqrt{n}}$$



Centralne twierdzenie graniczne – cd. (2)

4. Twierdzenie de Moivre'a – Laplace'a (CTG dla schematu Bernoulliego)

Założmy, że $X_1, X_2, \dots$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie

$$\mathbb{P}(X_n = 1) = p, \quad \mathbb{P}(X_n = 0) = q, \quad p + q = 1.$$

Wówczas dla $s < t$,

$$\mathbb{P} \left(s \leq \frac{X_1 + X_2 + \dots + X_n - np}{\sqrt{npq}} \leq t \right) \xrightarrow{n \rightarrow \infty} \Phi(t) - \Phi(s)$$

- każda nierówność po lewej stronie może być ostra (i tu, i w CTG)
-



Centralne twierdzenie graniczne – cd. (3)

5. Twierdzenie Berry-Esséena (Oszacowanie błędu CTG)

Założmy, że (X_n) jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie takim, że $\mathbb{E}|X_1|^3 < \infty$. Oznaczmy $m = \mathbb{E}X_1$ i $\sigma = \sqrt{\text{Var}X_1}$. Wówczas

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \leq t \right) - \Phi(t) \right| \leq C \frac{\mathbb{E}|X_1 - \mathbb{E}X_1|^3}{\sigma^3\sqrt{n}},$$

gdzie $C \in [1/\sqrt{2\pi}; 0,77]$.



Centralne twierdzenie graniczne – cd. (4)

6. Przykłady:

- Liczba dziewczynek i chłopców
- Ilu studentów przyjąć?
- Sumowanie błędów
- Przedziały ufności





WARSAW UNIVERSITY
Faculty of Economic Sciences