16.06.2021

Zadanie Na podstawie losowej próby szacujemy procent dorosłych, którzy potrafią pisać i czytać. Wiadomo na pewno, że jest to powyżej 90% dorosłej populacji. Chcemy, aby błąd był mniejszy niż 0.01 z p-stwem 0.9. Ile osób musi liczyć próba?

CTG: $\{X_n\}$ i.i.d., $\underset{m}{\overset{\shortparallel}{EX_1}} < \infty$, $\underset{\sigma^2}{\overset{\shortparallel}{\mathrm{Var}\, X_1}} < \infty$.

$$\frac{X_1 + \ldots + X_n - nm}{\sqrt{n}\,\sigma} = \frac{S_n - ES_n}{\sqrt{\mathrm{Var}\, S_n}} \xrightarrow{d} \mathcal{N}(0,1)$$

Rozwiązanie zadania

Chcemy znaleźć $p \approx \dfrac{\text{osoby, które potrafią pisać}}{\text{wielkość próbki}}$

n – l. testowanych osób

$$X_i = \begin{cases} 1 & \text{i-ta osoba czyta} \\ 0 & \text{w p.w.} \end{cases}$$

$S_n = X_1 + \ldots + X_n$ — l. osób, które czytają.

$$p \approx \frac{S_n}{n}, \ \mathbb{P}\left[\left|\frac{S_n}{n} - p\right| \leq 0{,}01\right] \geq \frac{9}{10}$$

S_n ma rozkład $Bin(n,p)$

$\mathbb{E}X_1 = p$, $VarX_1 = p(1-p)$. Chcemy skorzystać z CTG.

$$\mathbb{P}\left[\underbrace{\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right|}_{\text{Z CTG } \sim \mathcal{N}(0,1)} \leq \overbrace{\frac{0{,}01\sqrt{n}}{\sqrt{p(1-p)}}}^{x_0}\right] \geq \frac{9}{10}$$

tu się też generuje pewien błąd

Szukamy x_0 t.że

$$\frac{9}{10} \leq \mathbb{P}\left[|\underset{\sim \mathcal{N}(0,1)}{Y}| \leq x_0\right] = \int_{-x_0}^{x_0} \varphi(x)\,dx = \Phi(x_0) - \Phi(-x_0)$$

$$= \Phi(x_0) - 1 + \Phi(x_0) = 2\Phi(x_0) - 1$$

Czyli $\Phi(x_0) \geq 0.95 \Rightarrow x_0 \geq 1.65$.

Zatem $\frac{\sqrt{n}}{100\sqrt{p(1-p)}} \geq 1.65$

$$\sqrt{n} \geq 165\sqrt{p(1-p)}$$

Wiemy, że $p \geq 0.9$. Wtedy $p(1-p) \leq 0.09$.

Czyli $\sqrt{n} \geq 0.3 \cdot 165 = 49.5$

Zatem $n \geq 49.5^2 = 2450.25$

Czyli $n \geq 2451$.

<u>Tw. 11.2</u> (Berry–Essen)

Jeżeli $\{X_n\}$ i.i.d., $\mathbb{E}|X_1|^3$, $\mathbb{E}X_1 = 0$, $\sigma^2 = \mathrm{Var}\, X_1$, to

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left[\frac{X_1 + \ldots + X_n}{\sqrt{n}\,\sigma} < t \right] - \Phi(t) \right| \leq \frac{C \cdot \mathbb{E}|X_1|^3}{\sigma^3 \sqrt{n}}$$

gdzie $\Phi(t)$ jest dystrybuantą rozkł. normalnego $\mathcal{N}(0,1)$, a C jest ustaloną stałą (niezależną od rozkładu X_1)

Przykład Niech u_n będzie ciągiem Bernoulliego, tj. $\mathbb{P}[u_n = \pm 1] = 1/2$. Wówczas dla parzystych wartości n:

$$\left|\mathbb{P}\left[\frac{u_1+\ldots+u_n}{\sqrt{n}\sqrt{\mathrm{Var}\,u_n}} < 0\right] - \Phi(0)\right| = \left|\mathbb{P}[u_1+\ldots+u_n<0] - 1/2\right|$$

$$= \left|1/2\,\mathbb{P}[u_1+\ldots+u_n<0] + 1/2\,\mathbb{P}[u_1+\ldots+u_n>0] - 1/2\right|$$

$$= \left|1/2\left(1-\mathbb{P}[u_1+\ldots+u_n=0]\right) - 1/2\right|$$

$$= 1/2\,\mathbb{P}[u_1+\ldots+u_n=0]$$

$$= 1/2 \cdot \frac{\binom{n}{n/2}}{2^n} \overset{\text{stirling}}{\sim} 1/2\frac{1}{\sqrt{\pi n/2}} = \frac{1}{\sqrt{2\pi}\cdot\sqrt{n}}$$

TW. 11.3 (Lindeberg – Feller)

Niech $\{X_{n,k}\}$ będzie schematem serii spełniającym:

- $\mathbb{E}X_{n,k} = 0$ dla $n \geq 1$, $k \leq n$
- $\lim_{n\to\infty} s_n^2 = \sigma^2 > 0$, dla $s_n^2 := \mathrm{Var}\sum_{k=1}^{n} X_{n,k} = \sum \mathrm{Var}\,X_{n,k}$

• (warunek Lindeberga) dla każdego $r > 0$

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}|>r\}}\right] = 0$$

Wówczas $\sum_{k=1}^{n} X_{n,k} \xrightarrow{d} \mathcal{N}(0,\sigma^2)$

Przykład Liczba cykli w losowej permutacji:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 9 & 6 & 8 & 2 & 1 & 5 & 4 & 7 \end{pmatrix} = (1\,3\,6)(2\,9\,7\,5)(4\,8)$$

Problem: losujemy permutację $\Pi_n \in S_n$.

C_n – l. cykli tej permutacji.

Cel: $\frac{C_n - \log n}{\sqrt{\log n}} \xrightarrow{d} \mathcal{N}(0,1)$

Definiujemy $Z_{n,k} = \begin{cases} 1 & \text{jeżeli po } k\text{-tej liczbie w rozkł. na cykle pojawia się ")"} \\ 0 & \text{w p. w.} \end{cases}$

Zauważmy, że $C_n = \sum_{k=1}^{n} Z_{n,k}$

Lemat: dla ustalonego n $Z_{n,k}$ są nzl.

oraz $\mathbb{P}[Z_{n,k}=1] = \frac{1}{n-k+1}$.

Sprawdzimy warunek Lindeberga.

Zdef. $Y_{n,k} = Z_{n,n-k+1}$. Znowu $Y_{n,k}$ są nzl.

oraz $\mathbb{P}[Y_{n,k}=1] = 1/k$.

$\mathbb{E}Y_{n,k} = 1/k$, $\mathrm{Var}\, Y_{n,k} = 1/k - 1/k^2$

$$\mathbb{E}C_n = \sum_{k=1}^{n} \mathbb{E}Y_{n,k} = \sum_{k=1}^{n} 1/k \sim \log n$$

$\mathrm{Var}\, C_n \sim \log n$

Niech $X_{n,k} = \frac{Y_{n,k} - 1/k}{\sqrt{\log n}}$. $\{X_{n,k}\}$ - schemat serii.

1) $\mathbb{E}X_{n,k} = 0$

2) $s_n^2 = \sum_{k=1}^{n} \mathrm{Var}\, X_{n,k} = \sum_{k=1}^{n} \mathrm{Var}\left(\frac{Y_{n,k}}{\sqrt{\log n}}\right) = \frac{1}{\log n} \sum_{k=1}^{n} 1/k - 1/k^2$

$$\longrightarrow 1 = \sigma^2$$

3) $\sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}|>r\}}\right]$

$$= \sum \mathbb{E}\left[X_{n,k}^2 \mathbb{1}_{\left\{\frac{Y_{n,k}-1/k}{\sqrt{\log n}} > r\right\}}\right] \xrightarrow{n\to\infty} 0$$

$\frac{2}{\sqrt{\log n}} > r$

$\frac{2}{\sqrt{\log n}} \to 0$ dla dużych n

ten zbiór pusty

Z CTG L-F $\sum_{k=1}^{n} X_{n,k} \xrightarrow{d} \mathcal{N}(0,1)$

Czyli

$$\frac{C_n - \log n}{\sqrt{\log n}} + \frac{\log n - \sum 1/k}{\sqrt{\log n}} = \frac{\sum Y_{n,k} - \sum 1/k}{\sqrt{\log n}}$$

$\frac{C_n - \log n}{\sqrt{\log n}} \xrightarrow{d} \mathcal{N}(0,1)$, $\quad \frac{\log n - \sum 1/k}{\sqrt{\log n}} \to 0$, $\quad \frac{\sum Y_{n,k} - \sum 1/k}{\sqrt{\log n}} \xrightarrow{d} \mathcal{N}(0,1)$

KONIEC