Berry-Esseen tétel

~~A CH~~ Legyen $X_1, X_2, \dots, X_n$ független, azonos eloszlású, legyen $\mathbb{E}X_i = m$, $Var\, X_i = \sigma^2 < \infty$. ~~A CHT~~.

Legyen $S_n = X_1 + \dots + X_n$.

A CHT szerint

$$\mathbb{P}(S_n \leq k) \approx \phi\left(\frac{k - nm}{\sqrt{n}\sigma}\right),$$

ahol ϕ a standard normális eloszlásfüggvény.

<u>Kérdés</u> Jó, de mi az, hogy $\approx$?

Mekkora lehet a közelítés hibája?

<u>Válasz</u>: • <u>Általában</u> semmit nem tudunk mondani.

• Ha viszont teljesül +1 (nem túl erős) feltétel, [konkrétan: $\mathbb{E}(|X_i|^3) < \infty$], akkor van hibabecslés:

<u>Tétel</u> (Berry-Esseen tétel, 1941/42)

Legyen $X_1, \dots X_n$ független, azonos eloszlású.

Legyen $\mathbb{E}X_i = m$, $Var\, X_i = \sigma^2$, és legyen $\delta := \mathbb{E}(|X_i - m|^3) < \infty$

Legyen $S_n = X_1 + \dots + X_n$, és $k \in \mathbb{R}$. Ekkor $\mathbb{P}(S_n \leq k)$-ra a CHT becslés hibája

$$\text{hiba} = \left|\mathbb{P}(S_n \leq k) - \phi\left(\frac{k - nm}{\sqrt{n}\sigma}\right)\right| \leq \frac{C\delta}{\sigma^3 \sqrt{n}},$$

ahol C konstans.

$C = 0.4748$

Megj.: δ neve „centrált abszolút harmadik momentum"

(2/3)

(1)

- X_i-ből m kivonása a centrálás
- harmadik momentum = 3. hatvány várható értéke
- abszolút értéket veszünk a várható érték előtt:
  - → ez nagyon fontos
  - → ez kissé nehezíti a számolást

  } észnél kell lenni.

(2) Esseen a tételt 1942-ben $C = 7.59$-cel bizonyította.

A $C = 0.4748$ becslés 2011 óta ismert.

(3) Jó hír: hiba $\longrightarrow 0$, amint $n \to \infty$

(4) Rossz hír 1: A konvergencia lassú: $\sim \frac{1}{\sqrt{n}}$

Rossz hír 2: A hibára adott korlátban nem szerepel a K, így hiába választunk nagyon nagy K-t, ettől

- a tényleges $\mathbb{P}(S_n \leq K)$ valség ~~kicsi~~ nagy lesz,
- a $\phi\left(\frac{K-nm}{\sqrt{n}\sigma}\right)$ ~~becslés~~ közelítés is ~~kicsi~~ nagy lesz,

mégse lehetünk biztosak benne, hogy a ~~köze~~ közelítés jó.

P1: Feldobunk egy stabályos érmét $n = 10000$-ster.
Legyen S_n a dobott fejek stáma.
Perste $S_n = X_1 + X_2 + \dots + X_n$, ahol $X_1, X_2 \dots$ független
$X_i \sim B(\frac{1}{2})$, etért $\mathbb{E} S_n = 5000$,

$p = \frac{1}{2}$ $q = \frac{1}{2}$ | $\mathrm{Var}\, S_n = npq = 2500$

$\sigma^2 = \mathrm{Var}\, X_i = pq = \frac{1}{4}$ | $D\, S_n = \sqrt{\mathrm{Var}\, S_n} = 50 = \sqrt{n}\,\sigma$

Mit mondhatunk a $\mathbb{P}(S_n \leq K)$ valóstínűségről, ha

a.) $K = 4900$?

b.) $K = 4800$?

Válast:

a.) A CHT sterint $\mathbb{P}(S_n \leq K) \approx \phi\left(\frac{K - nm}{\sqrt{n}\,\sigma}\right) = \phi(-2) \approx 0.0228 \approx 2.3\%$

A Berry-Esseen tétel-hez kell $\delta = E(|X_i - m|^3)$.

Sterencsére $m = \frac{1}{2}$ és $X_i = 0$ vagy 1, így $X_i - m = -\frac{1}{2}$ vagy $\frac{1}{2}$,

etért $|X_i - m| = \frac{1}{2}$ mindig $\Rightarrow$ $\delta = \mathbb{E}\left[\left(\frac{1}{2}\right)^3\right] = \frac{1}{8}$

$\Rightarrow$ hiba $\leq \frac{C\delta}{\sigma^3 \sqrt{n}} = \frac{0.4748 \cdot \frac{1}{8}}{\frac{1}{8} \cdot 100} = 0.004748 \approx 0.5\%$

$\Rightarrow$ a tényleges $\mathbb{P}(S_n \leq K)$ valahol 1.8% és 3.8% között van.

b.) A CHT sterint $\mathbb{P}(S_n \leq K) \approx \phi\left(\frac{K - mn}{\sqrt{n}\,\sigma}\right) = \phi(-4) \approx 0.000032 = 0.0032\%$

De a hibáról most is csak annyit tudunk, hogy $\leq 0.5\%$

[A tényleges érték 0.0033%.]