CLI in njegova zgodba pri vzorčenju

1. Opazujemo določeno lastnost (neke populacije velikosti N), ki jo spremlja spremenljivka X.

2. Radi bi ocenili pričakovano vrednost $E(X)$, pri čemer predpostavimo $D(X) < \infty$ (oznaki $\mu = E(X)$ in $\sigma^2 = D(X)$).

3. V ta namen si izberemo vzorec velikosti $n \ll N$ (nimamo namreč možnosti, da bi opravili meritve na celotni populaciji):

   (a) vzorec je vektor $(x_1, \ldots, x_n)$, sestavljen iz n meritev,

   (b) izberemo ga naključno, meritve in izbira pa so med seboj neodvisne,

   (c) vzorec mora biti dovolj velik (npr. vsaj $n \geq 30$).

4. Naj bo X_i $(1 \leq i \leq n)$ spremenljivka, ki spremlja i-to meritev. Lahko predpostavimo, da ima enako porazdelitev kot X, tj.
$$E(X_i) = \mu \quad \text{in} \quad D(X_i) = \sigma^2.$$

5. Iščemo dobro formulo oz. funkcijo $f(x_1, \ldots, x_n)$ (cenilko), za katero velja, da bo njena vrednost na danem vzorcu z dovolj veliko verjetnostjo blizu $E(X)$. V primeru $E(X)$ je to vzorčno povprečje (ki je tudi slučajna spremenljivka):
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

6. CLI nam zagotavlja, da se $\overline{X}$ porazdeljuje normalno, tj. $\overline{X} \sim N(\mu_{\overline{X}}, \sigma_{\overline{X}})$ (predpostavki: $n \geq 30$ in $\sigma < \infty$!). V praksi nas zanima, koliko sta parametra $\mu_{\overline{X}}, \sigma_{\overline{X}}$!

7. Iz lastnosti pričakovane vrednosti (linearnost) izračunamo $\mu_{\overline{X}} = E(\overline{X}) = \cdots = \mu$

8. Iz lastnosti odklona (linearnost za nekorelirane slučajne spremenljivke) izračunamo $\sigma^2_{\overline{X}} = D(\overline{X}) = \cdots = \sigma^2/n$. ("…" v točkah 7. in 8. je potrebno znati dopolniti.)

Povzemimo (točke 6-8): za vzorčno povprečje vemo naslednje
$$\overline{X} \sim N(\mu, \sigma/\sqrt{n})$$
in ne pozabimo, da za računanje verjetnosti uporabljamo tabelirano funkcijo napake od $N(0,1)$, zato vpeljemo še standardizirano slučajno spremenljivko $Z = (\overline{X} - \mu_{\overline{X}})/\sigma_{\overline{X}}$.

Mimogrede: kje smo zgoraj uporabili $D(X+Y) = D(X) + D(Y) + 2K(X,Y)$ (kar spominja na $(a+b)^2 = a^2 + 2ab + b^2$)?

To je bil Centralni limitni izrek (CLI) za μ. Kaj pa za delež π, odklon σ ali celo razliko pričakovanih vrednosti $\mu_1 - \mu_2$ ali deležev $\pi_1 - \pi_2$ (da ne govorimo o σ_1/σ_2). Odgovore na slednje vprašanje najdete v poglavju o intervalih zaupanja.