Емпиријска функција расподеле

Нека је $X_1, .X_2, ..., X_n$ п.с.у. из популације на којој посматрамо обележје X. Природна оцена функције расподеле је

$$F_n(x) = \frac{\sum_{i=1}^{n} I\{X_i \leq x\}}{n}.$$

Лако се показује да $E(F_n(x)) = F(x)$ и $D(F_n(x)) = \frac{F(x)(1-F(x))}{n}$.

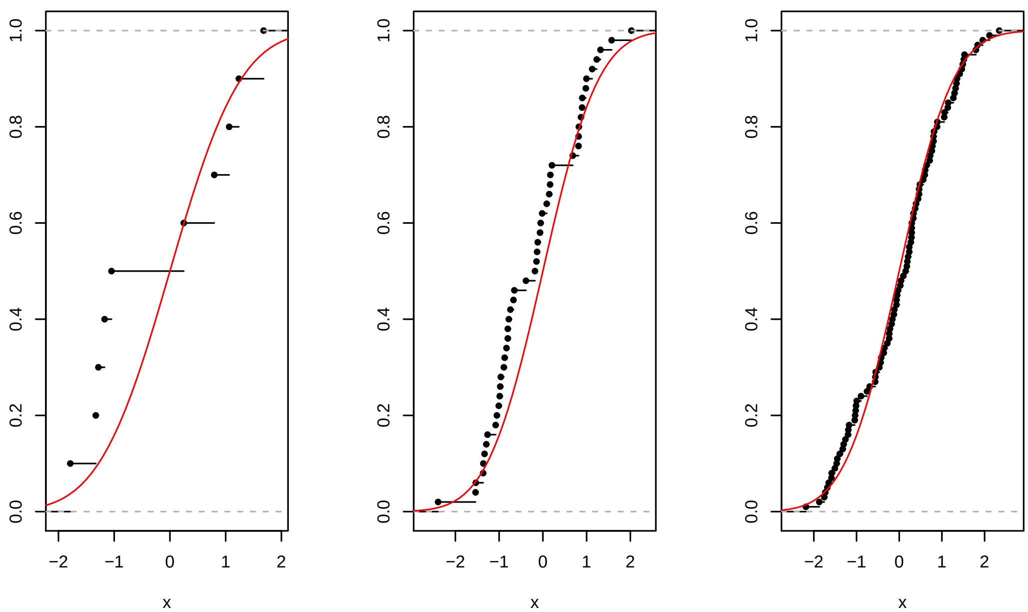
Одавде видимо да како расте обим узорка тако се и емпиријска функција расподеле "приближава" правој функцији расподеле F. Ово запажање садржано је у следећој теореми која је позната још и као централа теорема статистике.

Теорема 0.1 (Гливенко-Кантелијева теорема)**.** *Нека је $X_1, X_2, ..., X_n$ п.с.у. из популације са обележјем X са функцијом расподеле $F(x)$. Даље, нека је $F_n(x)$ одговарајућа емпиријска функција расподеле. Тада*

$$P\{\sup_x |F_n(x) - F(x)| \to 0, \text{ кад } n \to \infty\} = 1$$

На графику 1 приказане су емпиријске функције расподеле узорка из обележја са нормалном $\mathcal{N}(0, 1)$ расподелом. Црвеном лнијом приказана је одговарајућа функција расподеле.

Задатак 0.1. Како бисте оценили расподелу збира две једнако расподељене независне случајне величине ако на располагању имате п.с.у. $X_1, X_2, ..., X_n$?



Слика 1: Емпиријска функција расподеле на основу узорка обима $n = 10$ (лево), $n = 50$ (у средини) и $n = 100$ (десно)