Empirical Distribution Function

Exercise 1. We first consider univariate quantitative data where $x_1, \dots, x_n$ are n real observed values. We consider in the following examples variables obtained from the dataset 'diamonds' from the package 'ggplot2' in R.

Example 1. *The first following dataset gives the width of the twelve first diamonds.*
Width (mm) : 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 4.28 3.90

Example 2. *The following dataset gives the quality of the cut of the twelve first diamonds with the following correspondence :* 1 *meaning ideal,* 2 *premium,* 3 *very good,* 4 *good and* 5 *fair.*
Quality of the cut : 1 2 4 2 4 3 3 3 5 3 4 1

1. Compute the empirical mean and the median of the data set of Example 1.
2. Draw the empirical cumulative distribution function of the datasets of Examples 1 and 2.

Exercise 2. Empirical cumulative distribution function Let X_i be i.i.d. observations with c.d.f. F and $X_{1:n} = (X_1, \dots, X_n)$.

1. Show that for all $\alpha \in (0, 1)$

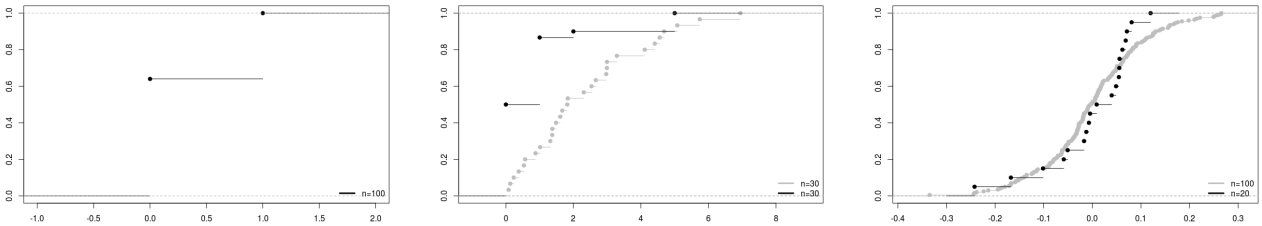
$$X_\alpha(n) = \inf\{t \in \mathbb{R},\, \widehat{F}_{X_{1:n}}(t) \geqslant \alpha\} =: \widehat{F}^{-1}_{X_{1:n}}(\alpha),$$

   where $\widehat{F}^{-1}_{X_{1:n}}$ is the generalized inverse of the empirical cumulative distribution function.
2. Fix $t \in \mathbb{R}$, what is the distribution of $n\widehat{F}_{X_{1:n}}(t)$? Can you complete the following limits :

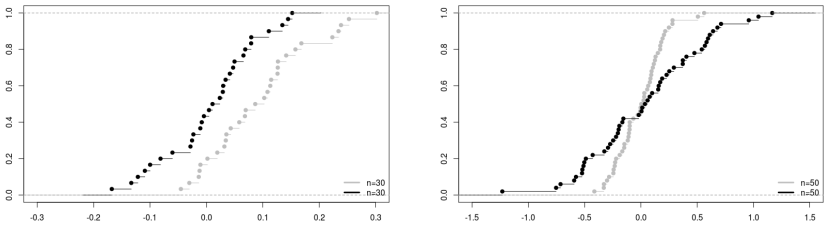
$$\widehat{F}_{X_{1:n}}(t) \xrightarrow[n\to\infty]{F-\text{proba}} ?? \quad \text{and} \quad \sqrt{n}\left(\widehat{F}_{X_{1:n}}(t) - ??\right) \xrightarrow[n\to\infty]{F-\text{dist.}} \mathcal{N}(0, ??) \,?$$

Exercise 3. Description of data and ecdf

1. Figure 1a represents the ecdf of some sample of size 100. Deduce the characteristics of the distribution of the sample and propose a distribution that is likely to have generated the data.
2. Each sub-figure 1(b-d) represents the ecdfs of two samples. For each sub-figure, compare the characteristics of the distributions of each sample.



(a) (b) (c)



(d) (e)

FIGURE 1 – ECDFs for some samples