

LLN

Pedro

2020-08-09

Law of Large Numbers (LLN) and Central Limit Theorem (CLT)

LLN states that if $X_1, X_2, X_3, \dots, X_n$ are a sequence of independent and identically distributed (iid) observations of size n drawn at random from a population with mean μ and variance σ^2 , as the number of random samples drawn increase, i.e., as $n \rightarrow \infty$, the mean $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$ of the sampling distribution approaches μ :

$$\bar{X} \sim \mu \quad (1)$$

The LLN is a statement about the **center of the sampling distribution**. The CLT is a statement about the **shape of the sampling distribution**. As $n \rightarrow \infty$,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \rightarrow \mathcal{N}(0, 1). \quad (1)$$

In other words, as the number of random samples drawn increase the sampling distribution approaches a normal distribution, with mean given by the LLN and standard deviation given by the CLT:

$$\bar{X} \rightarrow \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right). \quad (0)$$

Figure 1 displays an example of the entire process. The population distribution (top left) is gamma distributed with $\alpha = \beta = 1$ and therefore far from Normal. The population mean and standard deviation are given by $\mu = \alpha\beta = 1$ and $\sigma = \sqrt{\alpha\beta} = 1$ respectively. The top right and center left plots of the figure show two random samples from this population, each of size $n = 1000$. The sample means \bar{x}_i vary from sample to sample. The distribution of 1000 sample means is the **sampling distribution**. This sampling distribution appears on the center right plot. It is close to Normal (bottom left) with sampling standard deviation given by $\frac{\sigma}{\sqrt{n}} \simeq s = 0.14$, as expected due to the CLT since n is very large. The sampling mean is close to the population mean $\mu \simeq \bar{x} = 0.99$, also as expected due to the LLN. It takes approximately 400 random drawn samples for the LLN and CLT to converge (center bottom left and center bottom right).

For $n = 10$ (Figure 2), the sampling distribution is less Normal (bottom left). However, the LLN and CLT still converge after approximately 400 random samples being drawn. The sampling mean is close to the population mean: $\mu \simeq \bar{x} = 0.99$, and the sampling standard deviation is given by $\frac{\sigma}{\sqrt{n}} \simeq s = 0.32$.

Confidence intervals

Bootstrap

shows the bootstrap idea: we avoid the task of taking many samples from the population by instead taking many resamples from a single sample. The values of x from these resamples form the bootstrap distribution. We use the bootstrap distribution rather than theory to learn about the sampling distribution.

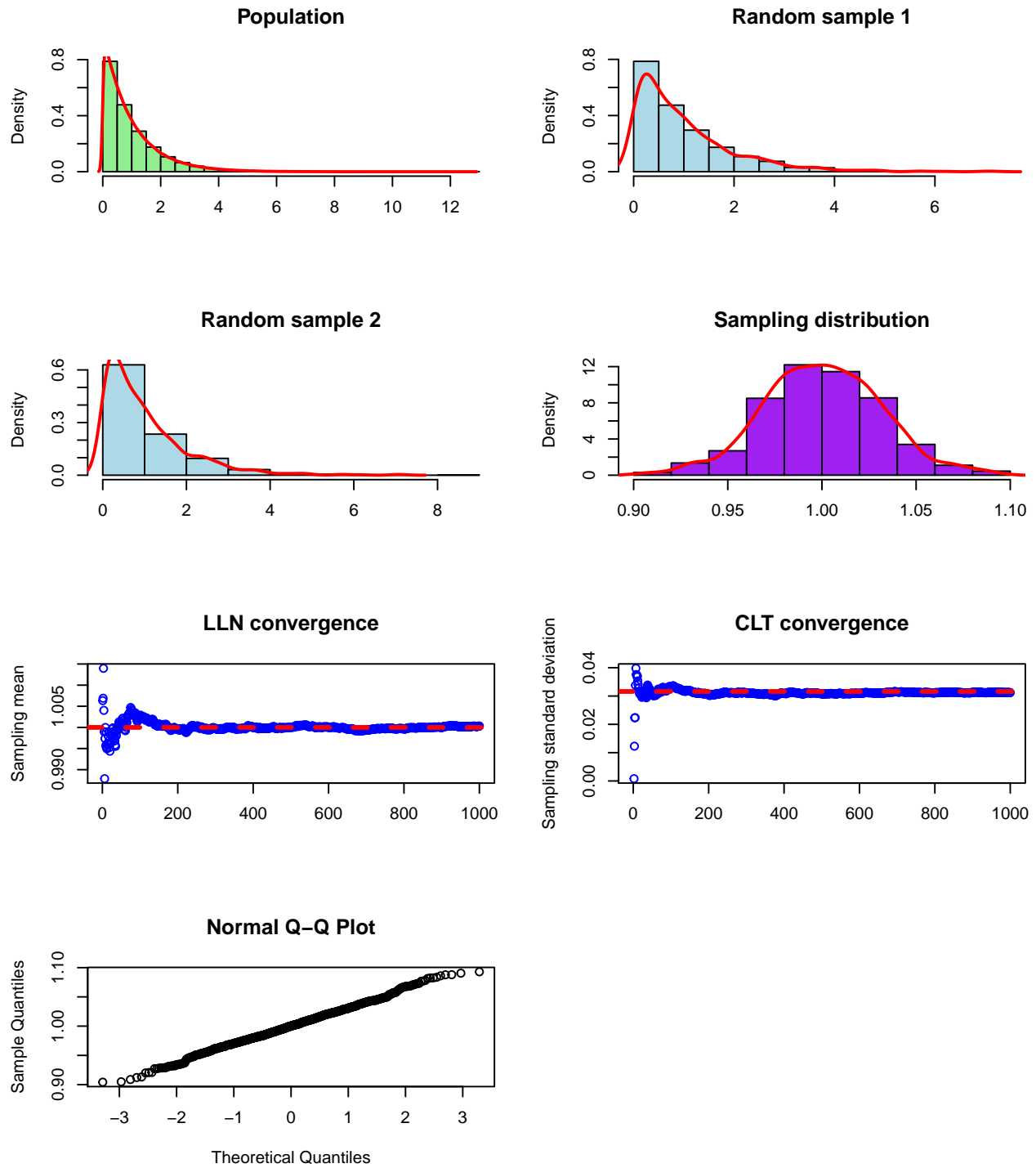


Figure 1: LLN and CLT (population parameters: $\alpha = 1, \beta = 1, \mu = 1$; random sample: $n = 50, \bar{x}_1 = 1.01, \bar{x}_2 = 0.98$; sampling distribution: $\bar{x} = 0.99, s = 0.03$)

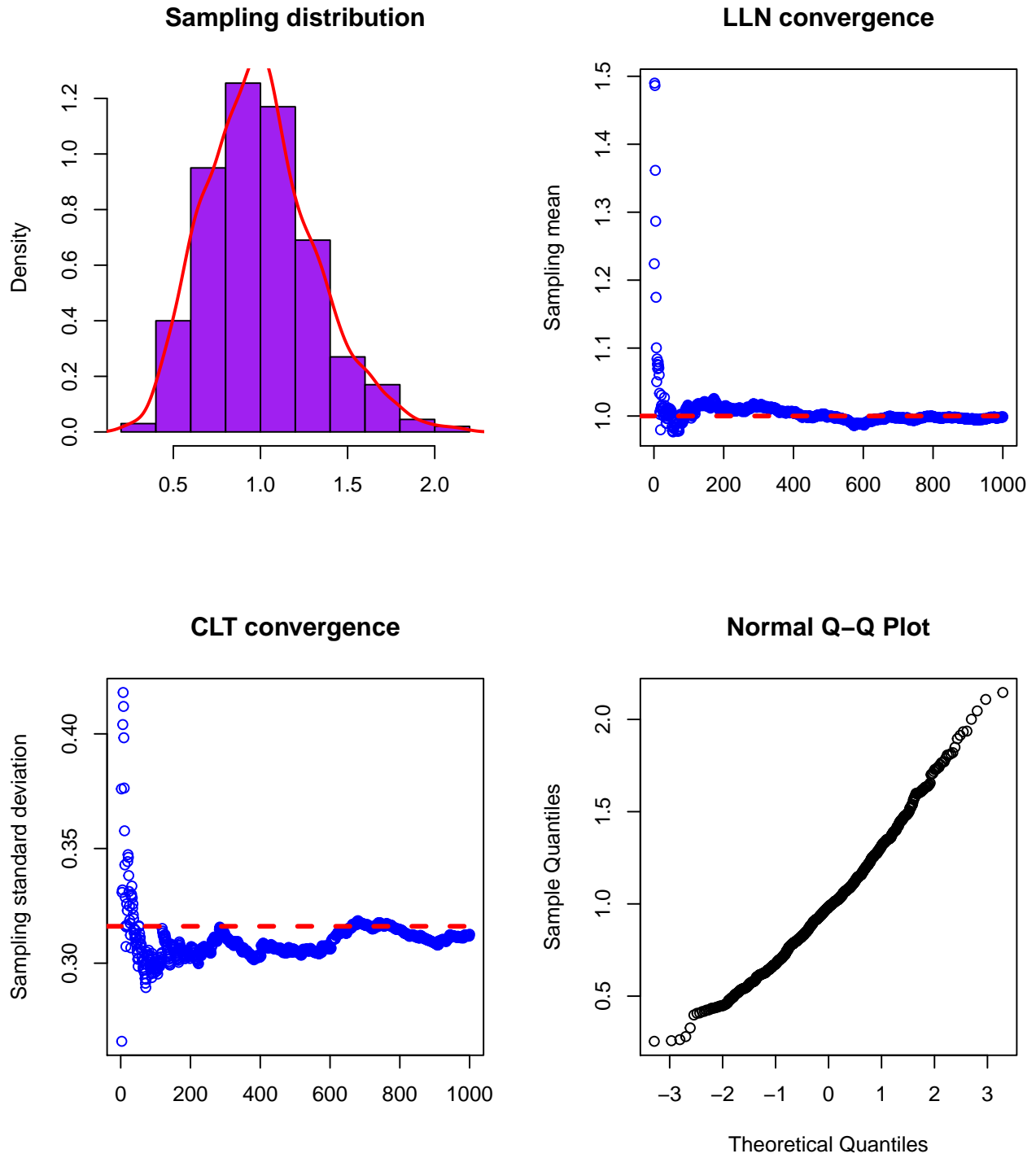


Figure 2: LLN and CLT (population parameters: $\alpha = 1, \beta = 1, \mu = 1$; random sample: $n = 10, \bar{x}_1 = 1.52, \bar{x}_2 = 0.70$; sampling distribution: $\bar{x} = 0.98, s = 0.32$)

The one samples stands in for the population, and the resampling of random samples stand if for the random samples. The bootstrap distribution of sample means stand in for the sampling distribution.

Confidence intervals revisited