HW1 - Jacob Shkrob

Jacob Shkrob

September 2022

1 Introduction

Suppose we generate N samples $\{X_j\}_{j=1}^N$ from the standard Exponential distribution with rate 1, denoted as Exp(1). Below, we empirically demonstrate that the estimator $\bar{X}_N = \frac{1}{N}\sum_{j=1}^N X_j$ closely captures the mean 1 as $N \to \infty$. Furthermore, the histogram generated by repeatedly sampling $\bar{X}_N$ and computing $\sqrt{N}(\bar{X}_N - \pi[x])$, where $\pi \sim \text{Exp}(1)$ is approximately distributed like a standard Gaussian. Each estimator was computed 1000 times using a standard random generator in Python, over a varying number of samples N and a histogram of values was generated for each model. As we can see in Figure 1, the histograms approach the standard bell shape associated with the Normal distribution. Furthermore, in Figures 2 and 3, the QQ-plots, which show the observed quantiles plotted against the theoretical quantiles of the Normal distribution, are tight around the diagonal line as the number of samples we average over increases (which is predicted by the CLT).

To estimate the quantity p_N, where p_N is defined to be

$$p_N = \mathbb{P}\left(\bar{X}_N - 1 > 0.1\right),$$

we consider the estimator Q_N, defined to be

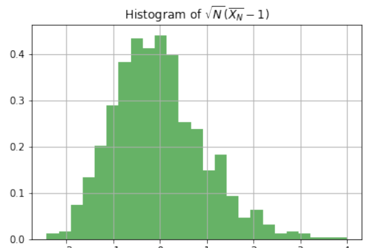
$$Q_N = Q_N^M = \frac{1}{M}\sum_{j=1}^M \mathbb{1}\{\bar{X}_N^{(j)} - 1 > 0.1\},$$

where $\bar{X}_N^{(j)}$ are i.i.d. random copies of $\bar{X}_N$. The random variable Q_N is therefore a $\text{Binom}(p_N, M)$ and variance

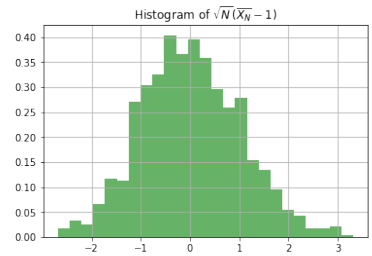
$$\mathbf{Var}\left(Q_N\right) = M(p_N(1 - p_N)).$$

From the previous exercise on the LDP for averages of Exp(1) random variables, we know that the tail probabilities for $\bar{X}_N$ scale like $-\epsilon + log(1+\epsilon)$. This implies that

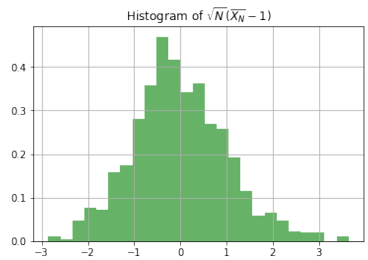
$$\begin{aligned} p_N = \mathbb{P}\left(\bar{X_N} - 1 > 0.1\right) &\asymp e^{-N\cdot 0.1 + N\log(1.1)} \\ &= e^{N(-0.1+\log(1.1)} \\ &= e^{-NK}, \end{aligned}$$



(a) Sample Size: N = 10



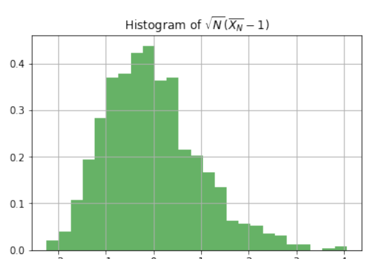
(b) Sample Size: N = 20



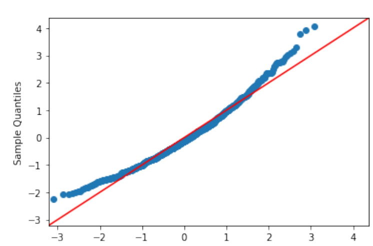
(c) Sample Size: N = 50

Figure 1: Histograms of 1000 iterations of $\sqrt{N}(\bar{X}_N - 1)$ with $N \in \{10, 30, 50\}$

for some $K > 0$. This implies that $\mathbf{Var}\,(Q_n)$ scales approximately like $M(e^{-NK} - e^{-2NK})$, which decays faster than p_N as $N \to \infty$ due to the dampening term $(1 - e^{-NK})$.

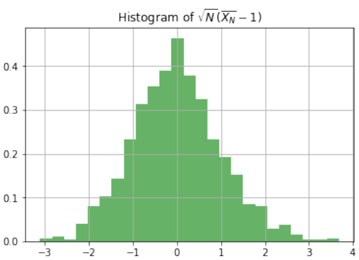


(a) Histogram

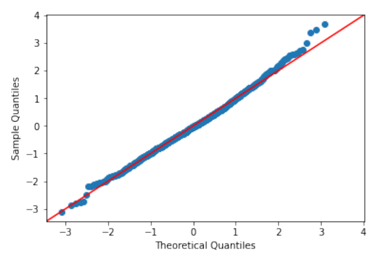


(b) QQ-Plot for $\mathcal{N}(0, 1)$

Figure 2: QQ-Plot and histogram for N = 10, iterations = 1000



(a) Histogram



(b) QQ-Plot for $\mathcal{N}(0, 1)$

Figure 3: QQ-Plot and histogram for N = 100, iterations = 1000