Monte Carlo Methods Homework 1

Bernice Feng (bf1318@nyu.edu)

Exercise 13. *Write a subroutine that takes N as an argument and generates a sample of the estimator $\overline{x}_N$ from the previous example (Python has a routine to generate N samples from the distribution exp(1)). Then write a routine that calls your subroutine to generate many copies of $\overline{x}_N$ and produces a histogram of the values of $\sqrt{N}\left(\overline{x}_N - \pi[x]\right)$. Produce this histogram for several values of N and show that for large N, the histograms approach the Gaussian density. A quantile–quantile (QQ) plot is a plot of the quantiles (i.e. the inverse of the cumulative distribution function) of two 1-dimensional distributions against one another. If the resulting curve is $y = x$, the distributions are the same. This is most often used when at least one of the distributions is empirical (i.e. a collection of samples) and you want to know how close those samples are to some specific distribution. Produce QQ plots to accompany your histograms.*

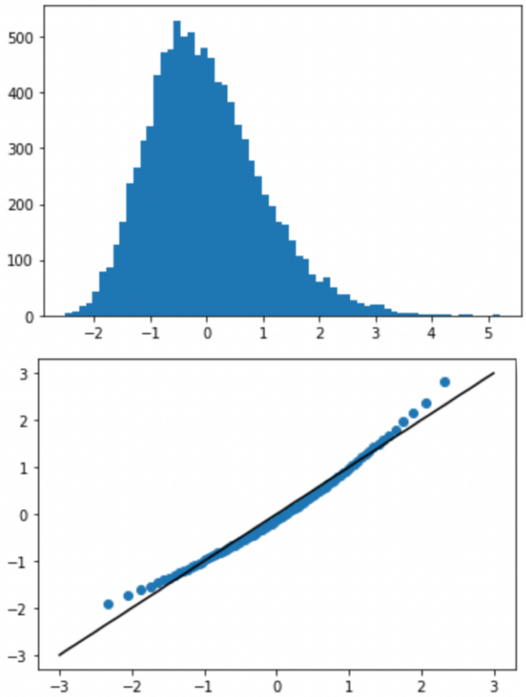
Next write a routine that constructs an estimate Q_N of the probability

$$p_N = \mathbf{P}\left[\overline{x}_N - 1 > 0.1\right]$$

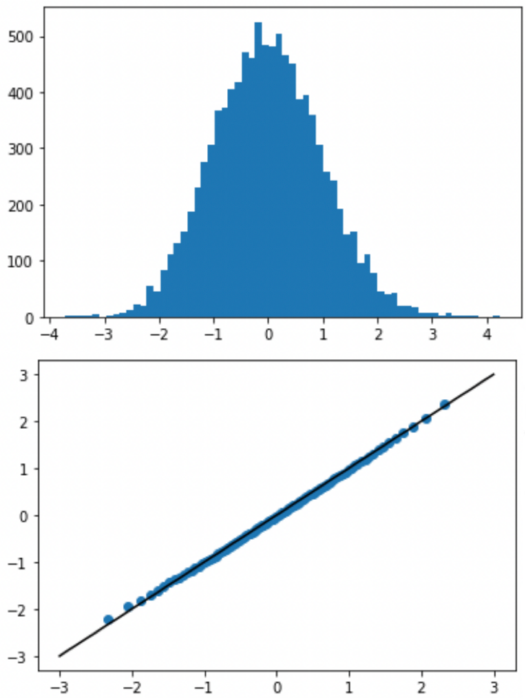
by generating many samples of $\overline{x}_N$ (remember that probabilities are expectations of indicators). To make your estimator less costly to evaluate, it may help to recall that the exp(λ) distribution is the same as the Gamma$(1, \lambda)$ distribution and that sums of gamma random variables, all of which have the same second parameter, is again a gamma random variable. Try to demonstrate the rate of decay we found in the last example. Estimating this quantity will require a huge number of samples of $\overline{x}_N$ as N increases. Write down a formula for the standard deviation of Q_N in terms of p_N. and compare it to p_N. Which, the standard deviation of Q_N or p_N, decays faster (you can answer this either by numerical test or by mathematical argument)?

1 Histograms and QQ Plots of $\sqrt{N}(\bar{x}_N - \pi[x])$

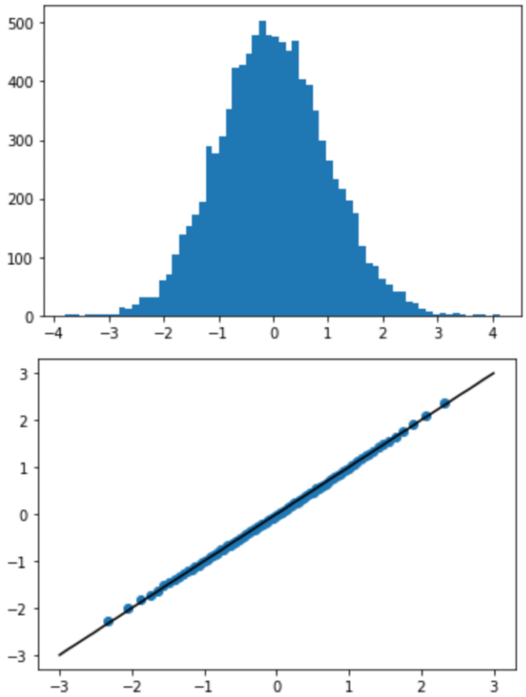
From the histograms and QQ plots of $\sqrt{N}(\bar{x}_N - \pi[x])$ for $N = 10^1$ [1], 10^2 [2], 10^3 [3], 10^4 [4], 10^5 [5], we can see that as N grows larger, the distribution of $\sqrt{N}(\bar{x}_n - \pi[x])$ converges to Gaussian distribution.



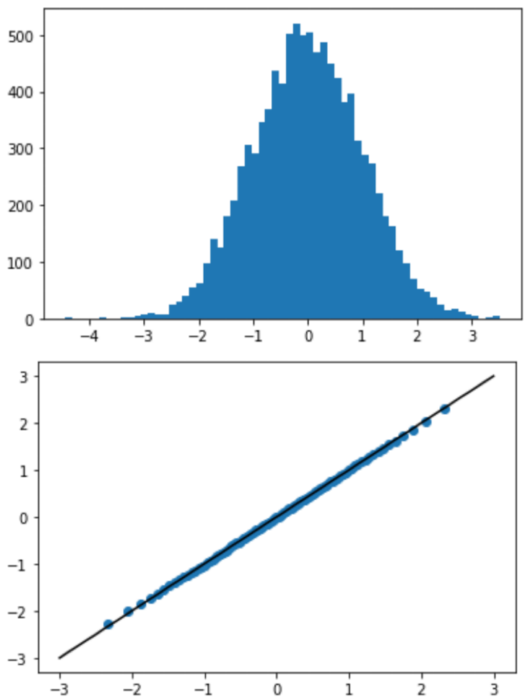
Figur 1: Sample size $N = 10^1$



Figur 2: Sample size $N = 10^2$



Figur 3: Sample size $N = 10^3$

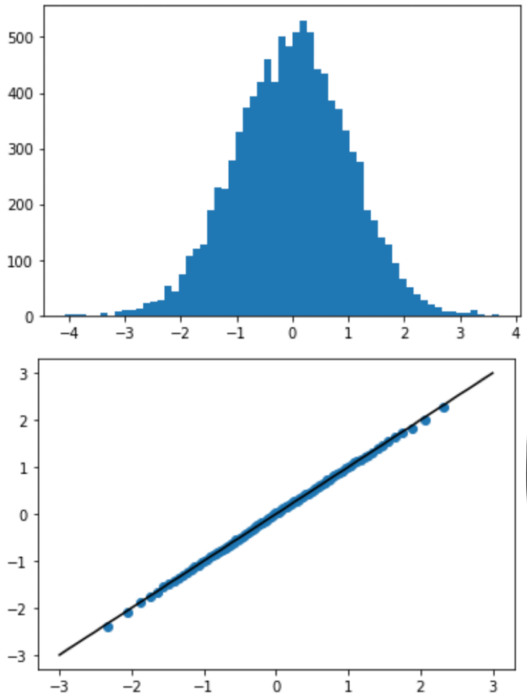


Figur 4: Sample size $N = 10^4$

2 Estimate Q_N of $p_N = P[\bar{x}_N - 1 > 0.1]$

According to the property of Gamma distribution, for exponential distribution $X^{(i)} \sim \exp(1)$, we have

$$\bar{x}_N = \frac{1}{N}\sum_{i=0}^{N-1} X^{(i)} 3= Y \sim \Gamma(N, \frac{1}{N}),$$



Figur 5: Sample size $N = 10^5$

Hence

$$p_N = P[\bar{x}_N - \pi[x] > 0.1] = P[\bar{x}_N - 1 > 0.1] = P[Y > 1.1].$$

An estimator Q_N for the probability p_N is:

$$Q_N = \frac{1}{M}\sum_{i=0}^{M-1} \mathbf{1}_{\{Y^{(i)}>1.1\}}$$

where $Y^{(i)}$ are i.i.d. random variables and M is a fixed sample size.

3 Rate of Decay

The rate of decay in previous example is that

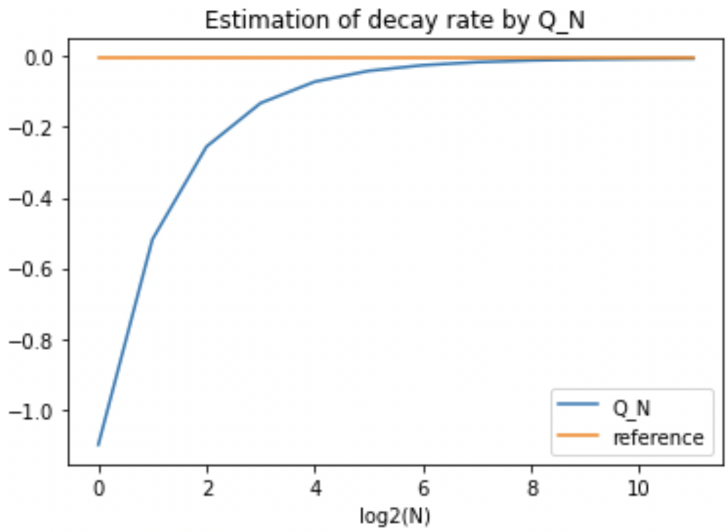
$$\lim_{N\to\infty} \frac{1}{N} \log P[x_N - 1 > \varepsilon] = -\varepsilon + \log(1+\varepsilon)$$

which we refer to as reference.

The plot of the rate of decay of Q_N compared with the reference is shown in figure [6]. Here, I chose $M = 10^6$, which works for $N \leq 2^{12}$, but will explode for $N \geq 2^{13}$.

The formula for the standard deviation of Q_N in terms of p_N is

$$\sigma(Q_N) = \sqrt{Var(Q_N)} = \sqrt{Var\left(\frac{1}{M}\sum_{i=0}^{M-1} 1_{\{Y_{(i)}>1.1\}}\right)} = \sqrt{\frac{Var\left(1_{\{Y_{(i)}>1.1\}}\right)}{M}} = \sqrt{\frac{p_N(1-p_N)}{M}}$$



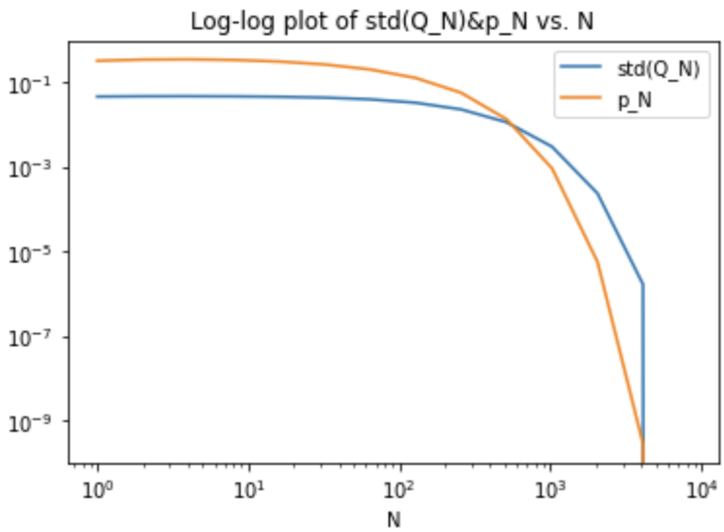
Figur 6: Demonstration of Rate of Decay by Q_N

We compare the standard deviation of Q_N with p_N, when $p_N \to 0$

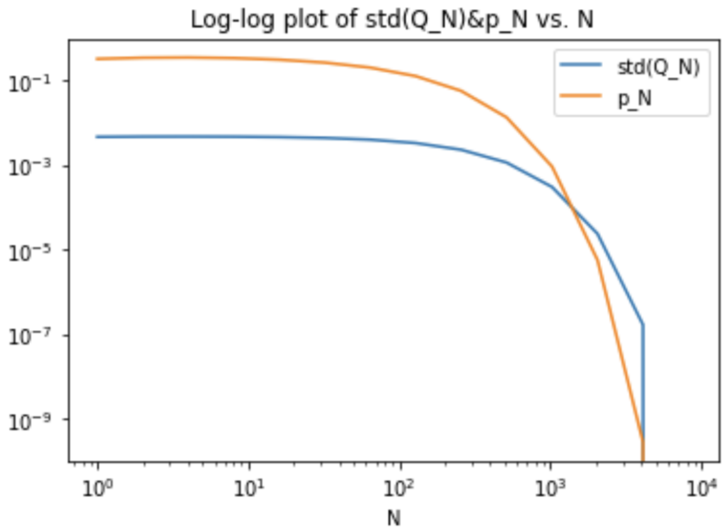
$$\frac{\sigma(Q_N)}{p_N} = \frac{1}{p_N}\sqrt{\frac{p_N(1-p_N)}{M}} \approx \frac{1}{p_N}\sqrt{\frac{p_N}{M}} = \frac{1}{\sqrt{M \cdot p_N}}$$

From the numerical simulation of p_N and Q_N, we plot the Log-log plot of them versus N. For $M = 10^2, 10^4$, and 10^6, the plots are shown in figure [7], [8], and [9] respectively.

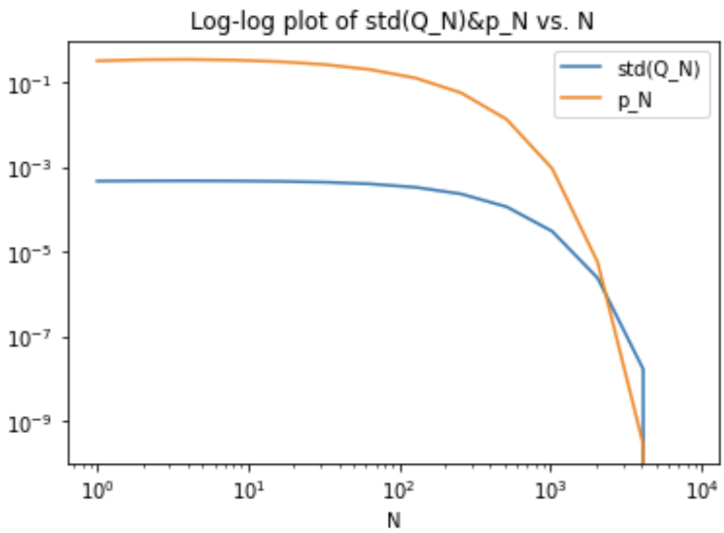
For fixed sample size M, the decay rate of $\sigma(Q_N)$ is slower than p_N. To render small error when estimating p_N by $\sigma(Q_N)$, we can scale $M \sim \frac{1}{p_N}$. We can also tell from the plots [7], [8], and [9] that with relatively smaller sample size M, the decay rates of $\sigma(Q_N)$ and p_N are closer, meaning the error of estimation is smaller.



Figur 7: $M = 10^2$



Figur 8: $M = 10^4$



Figur 9: $M = 10^6$