Applied Stochastic Analyis
Homework 01

Author:	Jaime Cascante-Vega
Email:	jc12343@nyu.edu
University ID:	N19711762

Problem 1

Show that if U is uniformly distributed on the interval $(0,1)$ then $X = -\log(U)$ is an exponential random variable with mean 1. This means that the probability density of X is $f(x) = \exp(-x)$ for $x \geq 0$. But it will be easier to show that X has the right cumulative distribution function, i.e. that $P(X \geq x)$ equals the integral of $\exp(-y)$ from 0 to x.

If $X = -\log(U)$ then

$$
\begin{aligned}
P(X < x) =& P(-\log(U) < x) \\
P(X < x) =& P(U > \exp(-x)) \\
P(X \geq x) =& 1 - P(U > \exp(-x)) \\
& \text{as U} \sim \text{Unif}[0,1] \\
P(X \geq x) =& 1 - \int_0^x \exp(-x) dx \\
P(X \geq x) =& 1 - (-\exp(-x)|_0^x \\
P(X \geq x) =& 1 - (-\exp(-x) + 1) \\
P(X \geq x) =& \exp(-x) \\
P(X < x) =& 1 - \exp(-x)
\end{aligned}
$$

And $P(X < x)$ is the cumulative distribution thus the probability density function (pdf) can be obtained with differentiation, and the pdf corresponds to the distribution of a exponential random variable with mean 1.

$$p(x) = \exp(-x)$$

Write a simple Monte Carlo code to compute the sample average of N i.i.d. exponential random variables with mean 1.
This tutorial will get you started
https://towardsdatascience.com/monte-carlo-simulation-a-practical-guide-85da45597f0e
with random number generation and plotting in Python.
Finally, numerically verify the Weak Law of Large Numbers, the Central Limit Theorem, and the concentration inequality that we proved at the end of class on 1/22 (see notes from

that lecture). You have some freedom to interpret ”verify.” Use your imagination to choose experiments and plots that convey your message.

Weak Law of Large Numbers

The Monte Carlo approximation of the sample average if N i.i.d exponential random variables $X \sim \text{Poisson}(\lambda = 1)$ or $\bar{f}_N$, with variance $\sigma^2 = 1/\lambda^2 = 1$ was computed from a sample of $N = 1000$. Using the Chebyshev inequality the probability that it deviates from it’s true value by ε will be bounded above as shown in the equation below. In Figure 1 I present the upper bounds and the probability of the deviation of $\bar{f}_N$, and as expected for different values of ε Chebyshev’s inequality is satisfied.

$$P\left(\left|\bar{f}_N - \pi(f)\right| > \varepsilon\right) = \frac{\text{Var}(\bar{f}_N)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 \cdot N}$$

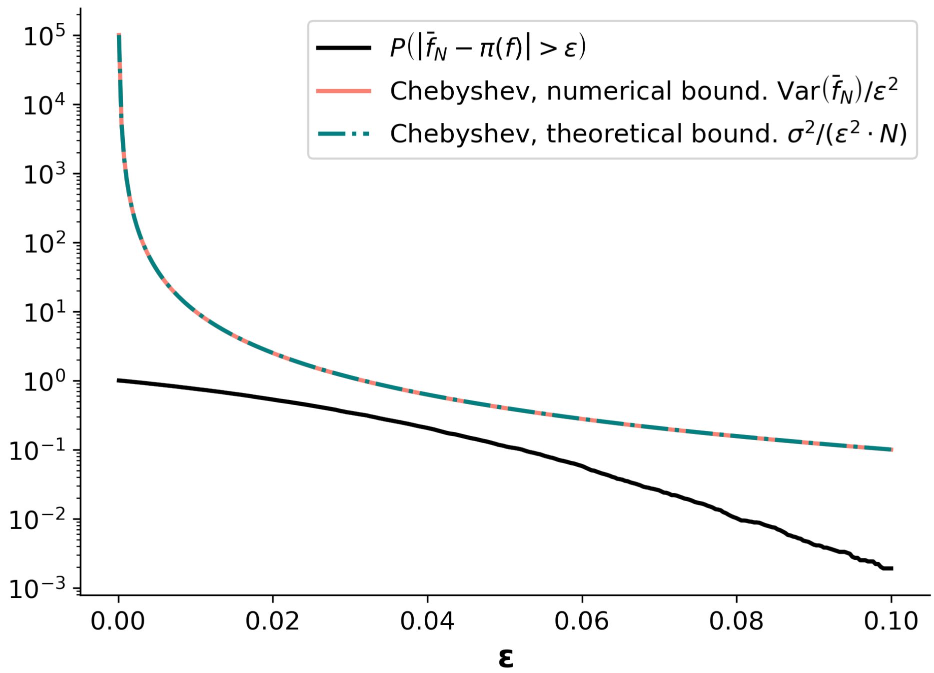

Figure 1: **Weak Lay of Large Numbers.** Black line shows Probability that the Monte Carlo approximation of the sample average $\bar{f}_N$ deviates from the true value more than ε. Salmon solid line and dark dashed green line shows the upper bound using Chebyshev’s inequality. The numerical variance $\bar{\mathrm{f}}_{\mathrm{N}}$ was computed from 1000 Monte Carlo approximation of 1000 samples and the theoretical variance σ^2 computed as $\sigma^2 = 1/\lambda^2 = 1$, the variance of the exponential distribution. The x-axis varies ε from 1e-3 to 0.1, y-axis is presented in log scale.

Central Limit Theorem (CLT)

The CLT for the Monte Carlo approximation of the sample average $\bar{f}_N$ is presented below.

$$Z_n = \sqrt{N}\left(\bar{f}_N - \pi\left[f\right]\right) \to \mathcal{N}\left(0, \sigma^2\right)$$

And using $\pi(f) = \lambda = 1$, the mean of the exponential distribution, and $\sigma^2 = 1/\lambda^2$, the variance of the exponential distribution would state that $\bar{f}_N$ converge in distribution to the normal distribution with parameters as shown below.

$$\bar{f}_N \to \mathcal{N}\left(\lambda, \frac{1}{\lambda^2\sqrt{N}}\right)$$

In Figure 2 I present both the empirical distribution of $\bar{f}_N$ and the theoretical prediction from the CLT as the sample size N increases. In Figure 3 I investigate: i) how does the empirical and theoretical uncertainty decrease as the sample size increase (left subplot) and ii) how well does the Monte Carlo estimate of the sample average $\bar{f}_N$ and the CLT normal distribution converge to $\pi(f)$ in distribution as the sample size increase (right subplot). As expected the MC uncertainty both measured empirically for different MC estimates and theoretically decreases as the sample size increase (see Figure 3 left plot). Similarly, the error was computed as the Continuous Ranked Probability Score (CRPS) between N samples from a distribution of $\bar{f}_N$ obtained from either MC or the CLT. The CRPS measures the distance between a probability distribution and a point. The error decreases as the sample size increase, and both errors are almost equal for samples sizes greater than 1000, suggesting that the MC estimate tends to a normal distribution as predicted by the CLT (see Figure 3 right plot).

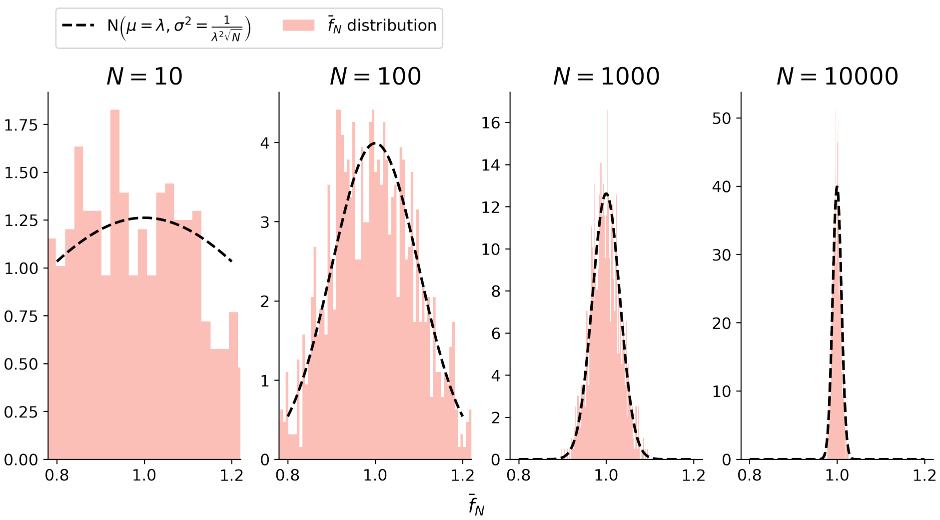

Figure 2: **Central Limit Theorem approximation.** In all plots the salmon histogram represent the empirical distribution of a 1000 different Monte Carlo estimates of the sample average $\bar{f}_N$. The dashed black line indicate the prediction from the CLT with mean μ and variance σ^2 as indicated in the legend. The parameter $\lambda = 1$ is the rate of the exponential distribution and N is the sample size, that increase from left to right.

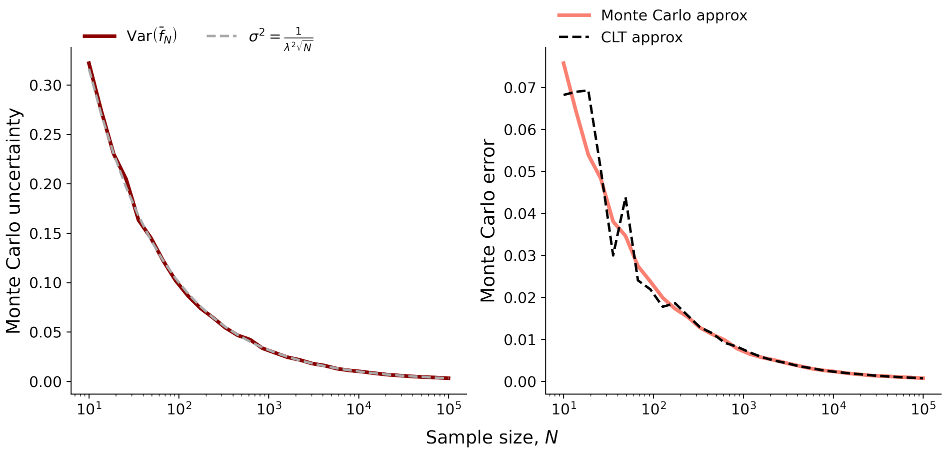

Figure 3: **Monte Carlo and CLT errors.** Left plot presents the uncertainty in the Monte Carlo estimate computed from the empirical variance of different estimates in dark red, and theoretically as the variance of the exponential distribution is $1/\lambda^2$. Right plot presents the error from a distribution of MC estimates in salmon and in black from estimates draw from a normal distribution as predicted by the Central Limit Theorem.

Concentration inequality

Using Chernoff's bound the probability that the error of the Monte Carlo estimate of $\pi[x]$, $\bar{f}_N$, is bounded above in probability is shown in the expression below.

$$
\begin{aligned}
P\left(\bar{f}_N - \pi[f] \geq a\right) \leq & \exp\left(-N \cdot I(a)\right) \\
& \text{where} \\
& I(a) = \sup_{\lambda > 0} \{a \cdot \lambda - \log\left(\mathrm{E}\left[\exp\left(\lambda \cdot (f(X) - \pi[f])\right]\right)\right\}
\end{aligned}
$$

In Figure 4 I present both the probability of the error of the Monte Carlo estimate varying the error allowed a, and varying λ from the moment generating function of $f(X)$. As expected the upper bound is always above however as λ increases the gap between both gets tighter.

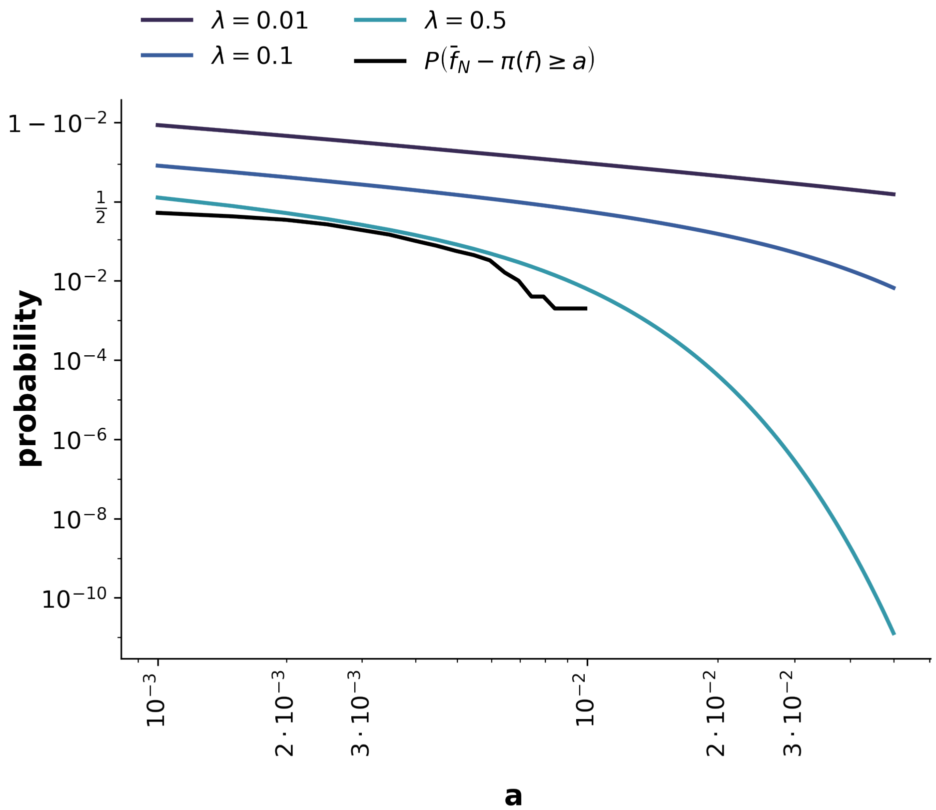

Figure 4: **Concentration inequality using Chernoff's bound.**

Problem 2

Problem 1.5 from Applied Stochastic Analysis (ASA)

Suppose $X \sim P(\lambda), Y \sim P(\mu)$ are two independent Poisson random variables.

(a) Prove that $Z = X + Y \sim P(\lambda + \mu)$.

Using the characteristic function of the two random variables:

$$\begin{aligned} f_x(\zeta;\lambda) &= e^{\lambda(\exp(i\zeta)-1)} \\ f_y(\zeta;\mu) &= e^{\mu(\exp(i\zeta)-1)} \end{aligned}$$

As they are independent $f_{z=x+y}(\zeta) = f_x(\zeta) f_y(\zeta)$:

$$\begin{aligned} f_z(\zeta;\lambda,\mu) &= e^{\lambda(\exp(i\zeta)-1)} e^{\mu(\exp(i\zeta)-1)} \\ f_z(\zeta;\lambda,\mu) &= e^{\lambda(\exp(i\zeta)-1)+\mu(\exp(i\zeta)-1)} \\ f_z(\zeta;\lambda,\mu) &= e^{(\lambda+\mu)(\exp(i\zeta)-1)} \end{aligned}$$

$f_z(\zeta)$ is the characteristic function of a Poisson distribution with parameter $\lambda + \mu$, thus $Z \sim P(\lambda + \mu)$.

(b) Prove that the conditional distribution of X (or Y), conditioning on $X + Y$ being fixed, i.e., $X + Y = N$, is a binomial distribution with parameter $n = N$ and $p = \lambda/(\lambda + \mu)$ (or $p = \mu/(\lambda + \mu)$).

The probability mass functions to compute $X = m$ and $Y = n$ are given by

$$\begin{aligned} P(x = m; \lambda) &= \frac{\lambda^m}{m!} e^{-\lambda} \\ P(y = n; \mu) &= \frac{\mu^n}{n!} e^{-\mu} \end{aligned}$$

We want to compute $P(y = k | x + y = N)$, using Bayes one can write this as

$$P(y = k | x + y = N) = \frac{P(x + y = N | y = k) P(y = k)}{P(x + y = N)}$$

The probability mass function for $P(y = k)$ and $P(x + y = N)$ are known as $Y \sim P(\mu)$ and $X + Y \sim P(\lambda + \mu)$ as shown in the previous numeral. The expression for this known probabilities are shown below.

$$\begin{aligned} P(Y = k) &= \frac{\mu^k}{k!} e^{-\mu} \\ P(X + Y = N) &= \frac{(\lambda + \mu)^N}{N!} e^{-(\lambda+\mu)} \end{aligned} \tag{1}$$

Thus only $P(x+y=N|y=k)$ remains unknown. However, re-writing as $P(x=N-k|y=k)$ and noting that X and Y are independent $P(x=N-k|y=k)=P(x=N-k)$ where $X \sim P(\lambda)$.

$$P(x=N-k|y=k)=\frac{\lambda^{N-k}e^{-\lambda}}{(N-k)!}$$

$$\begin{aligned}
P(y=k|x+y=N) &= \frac{\frac{\lambda^{N-k}}{(N-k)!}e^{-\lambda}\cdot\frac{\mu^k}{k!}e^{-\mu}}{\frac{(\lambda+\mu)^N}{N!}e^{-(\lambda+\mu)}} \\
P(y=k|x+y=N) &= \frac{\frac{\lambda^N\lambda^{-k}}{(N-k)!}\frac{\mu^k}{k!}}{\frac{(\lambda+\mu)^N}{N!}} \\
P(y=k|x+y=N) &= \left(\frac{\lambda}{\lambda+\mu}\right)^N\left(\frac{\mu}{\lambda}\right)^k\frac{N!}{k!\,(N-k)!} \\
P(y=k|x+y=N) &= \binom{N}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^N\left(\frac{\mu}{\lambda}\right)^k \\
&\qquad \text{multiplying by } (\lambda/(\lambda+\mu))^k/(\lambda/(\lambda+\mu))^k \\
P(y=k|x+y=N) &= \binom{N}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^{N-k}\left(\frac{\mu}{\lambda}\frac{\lambda}{\lambda+\mu}\right)^k \\
P(y=k|x+y=N) &= \binom{N}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^{N-k}\left(\frac{\mu}{\lambda+\mu}\right)^k
\end{aligned}$$

That is a Binomial distribution for Y conditioned on $X+Y=N$ with $p=\mu/(\lambda+\mu)$, $1-p=\lambda/(\lambda+\mu)$ and $n=N$. If we would prove the conditional for X, $p=\lambda/(\lambda+\mu)$ and $n=N$.

Problem 3

Problem 1.6 from ASA.

Prove the following statements:

(a) (Memoryless property of exponential distribution) Suppose $X \sim E(\lambda)$. Prove that

$$P(X > s+t|X > s) = P(X > t), \quad \forall s,t > 0$$

Recall that a random variable X that have a exponential distribution with rate λ

$$P(X \leq t) = 1 - e^{-\lambda t}, \quad \forall t \geq 0$$

As the time s is contained in $s+t$, or $s \subset s+t$, $P(X > s|X > x+t) = 1$. Using the Bayes theorem

$$\begin{aligned}
P(X > s+t|X > s) =& \frac{P(X > s|X > t+s) \cdot P(X > t+s)}{P(X > s)} \\
P(X > s+t|X > s) =& \frac{P(X > t+s)}{P(X > s)} \\
P(X > s+t|X > s) =& \frac{e^{-\lambda \cdot (t+s)}}{e^{-\lambda s}} \\
P(X > s+t|X > s) =& e^{-\lambda t}
\end{aligned}$$

And by definition $P(X > t) = e^{-\lambda t}$, thus $P(X > s+t|X > s) = P(X > t)$

(b) Let $X \in (0, \infty)$ to be a random variable such that

$$P(X > s+t) = P(X > s)P(X > t) \quad \forall s,t > 0$$

Prove that there exists $\lambda > 0$ such that $X \sim E(\lambda)$.

Let's assume that $X \sim E(\lambda), \quad \lambda > 0$, then $P(X > s) = e^{-\lambda t}$ and $P(X > s) = e^{-\lambda t}$ and thus $P(X > s)P(X > t) = e^{-\lambda t} \cdot e^{-\lambda t} = e^{-\lambda(s+t)}$. By definition $P(X > s+t) = e^{-\lambda(s+t)}$, and in consequence $P(X > s)P(X > t) = P(X > s+t)$.

Problem 4

Problem 1.17 from ASA
Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be $i.i.d.$ continuous random variables with common distribution F and density $\rho(x) = F'(x)$. $\left(X_{(1)}, X_{(2)}, \cdots, X_{(n)}\right)$ are called the *order statistics* of $\mathbf{X}$ if $X_{(k)}$ is the $k-th$ smallest of these random variables. Prove that the density of the order statistics is given by

$$p(x_1, x_2, \ldots, x_n) = n! \prod_{k=1}^{n} \rho(x_k), \quad x_1 < x_2, \cdots < x_n.$$

$$\begin{aligned}
S_1 &= \{(X_1, X_2, X_3) | x_1 < x_2 < x_3\} \\
S_2 &= \{(X_1, X_2, X_3) | x_1 < x_3 < x_2\} \\
S_3 &= \{(X_1, X_2, X_3) | x_2 < x_1 < x_3\} \\
S_4 &= \{(X_1, X_2, X_3) | x_2 < x_3 < x_1\} \\
S_5 &= \{(X_1, X_2, X_3) | x_3 < x_1 < x_2\} \\
S_6 &= \{(X_1, X_2, X_3) | x_3 < x_2 < x_1\}
\end{aligned} \tag{2}$$

As the random variables X_i are i.i.d the order of sampling each does not matter thus $P(S_1) = \rho(x_1) \cdot \rho(x_2) \cdot \rho(x_3)$, for the second set $P(S_2) = \rho(x_1) \cdot \rho(x_3) \cdot \rho(x_2)$, for the third $P(S_3) = \rho(x_2) \cdot \rho(x_1) \cdot \rho(x_3)$ and so on. Here I'm assuming that $P(S_i)$ implies that the order is the one indicated in the presentation of the sets in previous equation, and that the order does not matter as the random variables are i.i.d.

In consequence the joint distribution of the order statistics is presented below

$$p(x_1, x_2, x_3) = \sum_{i=1}^{6} P(S_i) = 3! \rho(x_1) \cdot \rho(x_2) \cdot \rho(x_3), \quad x_1 < x_2 < x_3$$

As for n random variables the sample space will in general be $n!$ by a similar argument as the one above the joint distribution will be given by

$$p(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} P(S_i) = n! \prod_{i=1}^{n} \rho(x_i), \quad x_1 < x_2 < \ldots < x_n$$

Problem 5

Problem 1.19 from ASA.
Prove that if the moment generating function $M(t)$ can be defined on an open set U, then $M \in C^\infty(U)$.

I randomly found this webpage https://math.stackexchange.com/questions/4018712/prove-that-mgf-defined-on-an-open-set-is-infinitely-differentiable when looking the definition of $C^\infty(U)$.

The generating function of a random variable X is given by, following the notation in the textbook.

$$M(t) = \mathrm{E}\left[e^{tX}\right]$$

Defining a point $t_0 \in [0, \varepsilon)$ and $\varepsilon > 0$ to linearize the MGF around, using the Taylor expansion:

$$\frac{1}{t}\mathrm{E}\left[e^{(t_0+t)X} - e^{t_0 X}\right] = \mathrm{E}\left[e^{t_0 X} X\right] + o(t)\mathrm{E}\left[X^2 e^{t_0 X}\right]$$

Using Holder's inequality.

$$\mathrm{E}\left[X^2 e^{t_0 X}\right] \leq ||e^{t_0 X}||_p \cdot ||X^2||_q = e^{t_0 p X} X^{2q}, \quad \text{where } p > 1, \frac{1}{p} + \frac{1}{q} = 1$$

Thus in the limit $t \to 0$ and evaluating at $t_0 = 0$ thus term becomes 0 and the derivative is.

$$\lim_{t \to 0} \frac{1}{t}\mathrm{E}\left[e^{(t_0+t)X} - e^{t_0 X}\right] = \frac{d}{dt}M(t)|_{t=t_0} = \mathrm{E}\left[e^{t_0 X} X\right]$$

One could repeatedly apply this argument such that higher order terms in the Taylor expansion, after applying Holder's inequality become 0. Thus the MGF is infinitely differentiable and it's expression is

$$\frac{d^n}{dt^n}M(t)|_{t=t_0} = \mathrm{E}\left[e^{t_0 X} X^n\right]$$