EE 131A
Probability and Statistics
Instructor: Lara Dolecek

Homework 5 Solutions
Due: Tuesday, February 21, 2023
TA: Jayanth Shreekumar, Rushi Bhatt

Reading: Chapter 4 of *Probability, Statistics, and Random Processes* by A. Leon-Garcia

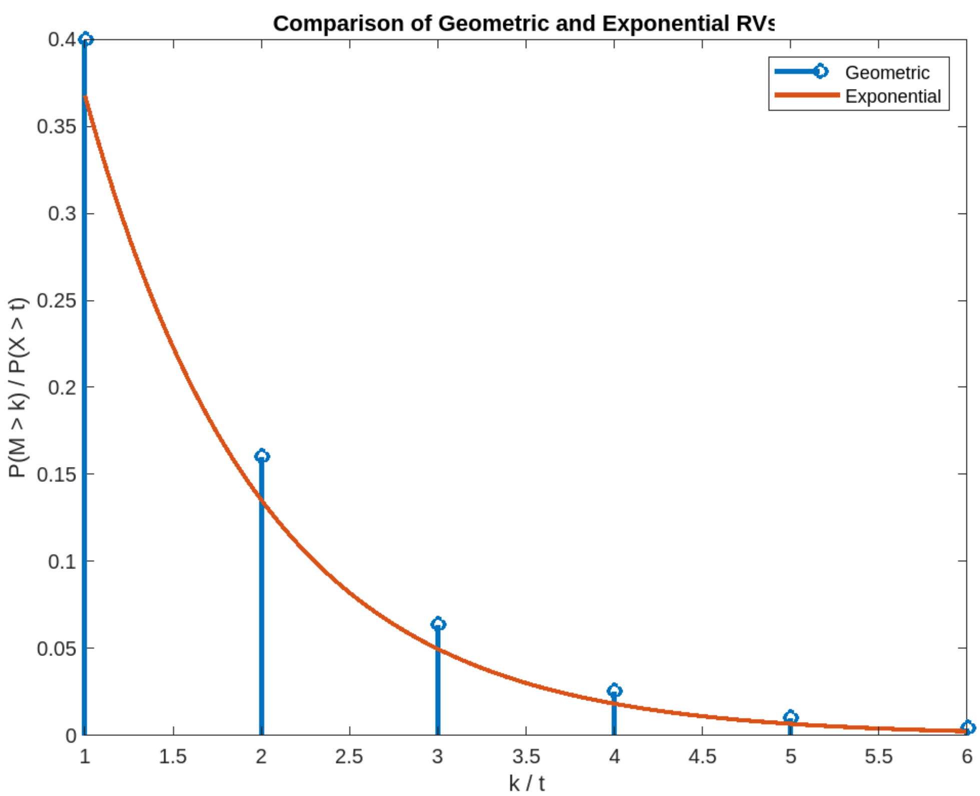
1. Let M be a geometric random variable with parameter p and let X be an exponential random variable with parameter λ.

   (a) Compute the tail probabilities $P(M > k)$ and $P(X > t)$ for the geometric and the exponential random variable where k is a positive integer and t is a non-negative real number.
   Solution: The tail probabilities are given by

$$P(M > k) = 1 - P(M \leq k) = 1 - \sum_{j=1}^{k} (1-p)^{j-1} p = (1-p)^k$$

$$P(X > t) = 1 - P(X \leq t) = 1 - F_X(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}$$

   (b) Use Matlab to plot $P(M > k)$ as a function of k and $P(X > t)$ as a function of t. Use $p = 0.6$ and $\lambda = 1$. Compare the two plots.
   Solution: $P(M > k)$ and $P(X > t)$ are shown below:



From the above plot, we can see that both the tail probabilities exponentially decay down to zero.

```

clc; clear; close;

p = 0.6;
k = 1:1:6;
t = 1:0.01:6;
lambda = 1;

geometric = (1 - p) .^(k);
exponential = exp(-lambda * t);

figure;
stem(k,geometric, LineWidth=2);
hold on;

plot(t, exponential, LineWidth=2);
xlabel("k / t");
ylabel("P(M > k) / P(X > t)");
legend("Geometric", "Exponential");
title("Comparison of Geometric and Exponential RVs");

```

(c) A continuous random variable X is said to satisfy the memoryless property if for all $t, h \geq 0$, $P(X > t+h | X > t) = P(X > h)$. Prove that the exponential random variable satisfies the memoryless property.
Recall that in Discussion 3, you proved that the geometric random variable satisfies the memoryless property.
Solution:

$$\begin{aligned} P(X > t+h | X > t) &= \frac{P(\{X > t+h\} \cap \{X > t\})}{P(X > t)} \\ &= \frac{P(X > t+h)}{P(X > t)} \quad \text{for } t > 0 \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} = P(X > h). \end{aligned}$$

Thus the exponential random variable satisfies the memoryless property.

2. On a special sale day, the UCLA store provides discounts for students only if the total sum of all their purchases ever made is at least \$500. Let X (in 100s of \$) be the total discount availed by a student. Assume X has the PDF:

$$f(x;\alpha) = \begin{cases} \frac{k}{x^\alpha} & x \geq 5 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

   (a) Find the value of k. What restriction on α is necessary?

   (b) What is the CDF of X?

   (c) What is the expected discount availed by a randomly chosen student, if $\alpha > 2$?

   (d) Show that $\ln(X/5)$ has an exponential distribution with parameter $\alpha - 1$.

   Hint: k must be greater than 0.
   Solution:

   (a) We know that the integral over the real line for a PDF must equal 1. Therefore,

$$1 = \int_5^\infty \frac{k}{x^\alpha} dx \tag{2}$$

$$= k.\frac{5^{1-\alpha}}{\alpha - 1} \tag{3}$$

   Therefore, $k = (\alpha - 1)5^{\alpha-1}$. Clearly, α cannot be less than or equal to 1 for the PDF to be valid. Therefore, $\alpha > 1$.

   (b) For $x \geq 5$, the CDF is given by:

$$\mathrm{F}(x) = \int_5^x \frac{k}{y^\alpha} dy \tag{4}$$

$$= 1 - \left(\frac{5}{x}\right)^{\alpha-1}. \tag{5}$$

   (c) The expectation is given by:

$$\mathbb{E}(X) = \int_5^\infty x\frac{k}{x^\alpha} dx \tag{6}$$

$$= \frac{k}{5^{\alpha-2}.(\alpha-2)}. \tag{7}$$

   The first term vanishes because of the condition $\alpha > 2$.

   (d) Finding the CDF,

$$P\left(ln\left(\frac{X}{5}\right) \leq y\right) = P\left(\frac{X}{5} \leq e^y\right) \tag{8}$$

$$= P\left(X \leq 5e^y\right) = F\left(5e^y\right) \tag{9}$$

$$= 1 - \left(\frac{5}{5e^y}\right)^{\alpha-1} \tag{10}$$
$$= 1 - e^{-(\alpha-1)y} \tag{11}$$

which is the CDF of an exponential RV with parameter $\alpha - 1$.

3. Find the PDF of $X = -\ln(4-4U)$, where U is a continuous random variable, uniformly distributed on the $[0, 1]$ interval.
   Solution:
   First, we get the CDF. For $x < -\ln(4)$, $F_X(x) = 0$. For $x \geq -\ln(4)$,

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(-\ln(4-4U) \leq x) = P(4 - 4U \geq e^{-x}) \\
&= P(4U \leq 4 - e^{-x}) = P\left(U \leq 1 - \frac{e^{-x}}{4}\right) \\
&= 1 - \frac{e^{-x}}{4}.
\end{aligned}$$

   Now we get the PDF. For $x < -\ln(4)$, $f_X(x) = 0$, and for $x \geq -\ln(4)$,

$$\begin{aligned}
f_X(x) &= \frac{d}{dx} F_X(x) \\
&= \frac{e^{-x}}{4}, \; x \geq -\ln(4) = -1.3863.
\end{aligned}$$

4. In this problem, you will show that the Poisson random variable is a good approximation for the Binomial random variable in the limit using characteristic functions:

   (a) Find the characteristic function of the Binomial random variable.
   Hint: The Binomial theorem states the following:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

   (b) Find the characteristic function of the Poisson random variable.
   Hint: We know from Taylor series:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

   (c) As n approaches ∞, denote np as λ and use the fact that

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

   to prove the question.

(d) What can you say about the value of n, p, and λ at this point? When is the Poisson random variable a good approximation for the Binomial random variable?
Hint: λ is the mean of the Poisson distribution: $\mathbb{E}[X] = \lambda$ where $X \sim Poisson(\lambda)$.

(e) Haemophilia is a rare genetic bleeding disorder that affects 1 in every 10,000 people in the world, where the person's blood does not clot by itself. In a sample of 2,000 people, what is the probability that exactly 1 person has haemophilia? Solve this question using the Binomial RV as well as its Poisson approximation and compare.

Solution

(a) The PMF of the Binomial random variable is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The characteristic function of a discrete random variable is given by:

$$\phi_X(\omega) = \mathbb{E}[e^{j\omega X}] = \sum_k p_X(x_k) e^{j\omega x_k}$$

Substituting the PMF of the Binomial random variable, we get:

$$\begin{aligned}
\phi_X(\omega) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{j\omega k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k (1-p)^{n-k} \\
&= (1 - p + pe^{j\omega})^n \quad \text{(from hint)}
\end{aligned}$$

(b) The PMF of the Poisson random variable is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Substituting the PMF of the Poisson random variable in the equation for the characteristic function of a discrete random variable, we get

$$\begin{aligned}
\phi_X(\omega) &= \sum_{k=0}^{\infty} e^{j\omega k} \frac{\lambda^k e^{-\lambda}}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} \\
&= e^{-\lambda} e^{\lambda e^{j\omega}} \quad \text{(from hint)} \\
&= e^{\lambda(e^{j\omega} - 1)}
\end{aligned}$$

(c) Writing down the characteristic function of the Binomial random variable as follows so that we can exploit the hint:

$$(1-p+pe^{j\omega})^n = \left(1+\frac{np(e^{j\omega}-1)}{n}\right)^n = \left(1+\frac{\lambda(e^{j\omega}-1)}{n}\right)^n$$
$$= e^{\lambda(e^{j\omega}-1)} \quad \text{(from hint)}.$$

(d) As λ is the mean of the Poisson distribution, it is a constant. Therefore, as n is very large, then, p must be very small. In other words, the Poisson random variable is a good approximation to the Binomial random variable if the probability of success is very small and the number of independent trials is very large.

(e) From the question, we have $n = 2000, p = \frac{1}{10000}$, and $k = 1$. Therefore, we calculate λ as

$$\lambda = np = 0.2$$

Using the PMF of the Binomial random variable X, the probability that 1 person has haemophilia is given by:

$$P(X=1) = \binom{2000}{1}(\frac{1}{10000})^1(\frac{9999}{10000})^{1999} = 0.16376088.$$

Now, using the PMF of the Poission random variable Y, the probability that 1 person has haemophilia is given by:

$$P(Y=1) = \frac{\lambda^1 e^{-\lambda}}{1!} = 0.2 * e^{-0.2} = 0.1637461.$$

5. Find the characteristic function of a normal distribution with mean m and variance σ^2.
Hint:

- Completion of square. Let $k = m + j\omega\sigma^2$, then $j\omega x - \frac{(x-m)^2}{2\sigma^2} = \frac{-(x-k)^2+2mj\omega\sigma^2-\omega^2\sigma^4}{2\sigma^2}$.
- $\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}$.

Solution:
The characteristic function of a random variable X is defined as following:

$$\Phi_X(\omega) = \mathbb{E}[e^{j\omega X}] = \int_{-\infty}^{\infty} f_X(x)e^{j\omega x}dx$$

Thus, the characteristic function of a normal distribution with mean m and variance σ^2 is:

$$\begin{aligned}\Phi_X(\omega) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} e^{j\omega x}dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{j\omega x - \frac{(x-m)^2}{2\sigma^2}} dx\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-(x-k)^2 + 2mj\omega\sigma^2 - \omega^2\sigma^4}{2\sigma^2}} dx \quad \text{(from hint)} \\
&= e^{\frac{2mj\omega\sigma^2 - \omega^2\sigma^4}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-k)^2}{2\sigma^2}} dx \\
&= e^{mj\omega - \frac{1}{2}\omega^2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\left(\frac{x-k}{\sqrt{2\sigma^2}}\right)^2} dx \\
&= e^{mj\omega - \frac{1}{2}\omega^2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-z^2} \sqrt{2\sigma^2} dz \\
&= e^{mj\omega - \frac{1}{2}\omega^2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{2\sigma^2} \int_{-\infty}^{\infty} e^{-z^2} dz \\
&= e^{mj\omega - \frac{1}{2}\omega^2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{2\pi\sigma^2} \quad \text{(from hint)} \\
&= e^{mj\omega - \frac{1}{2}\omega^2\sigma^2}
\end{aligned}
$$