Homework 3

Problem 1: CDFs and PDFs

Part A. Define a discrete random variable X as follows:

- $X = 2$ with probability $1/10$
- $X = 3$ with probability $1/10$
- $X = 5$ with probability $8/10$

Draw a picture of the CDF of X over the interval $x \in [0, 8]$. What are $P(2 < X \leq 4.5)$ and $P(2 \leq X < 4.5)$?

Part B. Let X be a (continuous) uniform random variable on $[0, 1]$.

i. Compute $P(X^2 \leq 0.25)$.
ii. For any number a, compute $P(X^2 \leq a)$.
iii. From (ii), find the PDF of the random variable $Y = X^2$.
iv. Compute $E(Y)$ and $\text{var}(Y)$ directly from the PDF.

Problem 2: practice with expected value

Part A. Suppose that we have d independent standard normal random variables $Z_1, \ldots, Z_d$, where each $Z_i \sim N(0, 1)$. We say that a continuous random variable X follows a chi-squared (χ^2) distribution with d degrees of freedom if

$$X \stackrel{D}{=} Z_1^2 + \cdots + Z_d^2 \,.$$

Remember that $\stackrel{D}{=}$ means "equal in distribution," i.e. the left and right-hand side of the equation have the same CDF. Our shorthand for the chi-squared distribution is $X \sim \chi^2_d$.

The chi-squared distribution plays an important role in statistical inference. For now, your job is simply to compute $E(X)$ using the relationship between the normal and the chi-squared. (That is, do not look up the PDF of the chi-squared distribution and try to compute the mean and variance directly, by integration.)

Note: an optional practice exercise is to compute $\text{var}(X)$ using the same basic strategy.

Part B. Markov lives in Austin, 2 miles from campus. During the winter, if the weather outside is cold, then Markov prefers to wear a fur cap with ear flaps and to walk to school. But if the weather is warm, like many winter days in Austin, Markov leaves the fur cap at home, wears a helmet instead (to protect the probabilistic inference engine in his skull), and rides to campus on one of those electric rental scooters that he typically finds strewn about the corners of central Austin. If the weather is cold (which happens with probability 0.4 in the winter in Austin), Markov walks the 2 miles to school at a brisk speed of V = 5 miles per hour. Otherwise he travels by scooter at a speed of V = 10 miles per hour.

Markov wants to calculate the expected time T that it takes him to get to class on a random winter day. He reasons as follows. His expected speed V is equal to

$$E(V) = 0.4 \cdot 5 + 0.6 \cdot 10 = 2 + 6 = 8 \,.$$

Therefore, since he must travel two miles to class, his expected time T to get to class is $E(2/V) = 2/E(V) = 2/8$ hours, or 15 minutes.

Do you agree with Markov's reasoning that $E(T) = 2/8$ hours? Explain why or why not, and if you don't agree, show how you would calculate $E(T)$ correctly.

Problem 3: inverse CDF

Suppose that U is a continuous random variable with a uniform distribution on $[0,1]$. Now suppose that f is the PDF of some continuous random variable of interest, that F is the corresponding CDF, and assume that F is invertible, so that the function $F^{-1}(u)$ exists and assigns a unique value to each $u \in (0,1)$.

Show that the random variable $X = F^{-1}(U)$ has PDF $f(x)$—that is, that X has the desired PDF. Hint: use results on transformations of random variables.

This cute result allows you to simulate random variables with any known (invertible) CDF, assuming that you have a source of uniform random variables.

Problem 4: simulation

This simulation exercise will serve as a bridge between probability and statistical inference. For now, you can take it as an illustration of the relationship between the expected value (i.e. the theoretical mean) of a random variable and the sample average constructed from a number of (independent) observations on that random variable.

Part A. Suppose that $X_N \sim \text{Binomial}(N, P)$ be the (random) number of successes in a sequence of N binary trials. Let $\hat{p}_N = X_N/N$ denote the proportion of observed successes. Calculate $E(\hat{p}_N)$ and $\text{sd}(\hat{p}_N)$.

Part B. Use R to simulate at least 1000 (or more) realizations of the random variable $\hat{p}_5$, assuming that the true $P = 0.5$. In other words, for each of the 1000 realizations, you will simulate five coin flips to get one value of X_5, and then set $\hat{p}_5 = X_5/5$. This process is repeated 1000 times (or more). Verify that the Monte Carlo mean and standard deviation of your simulated $\hat{p}_5$'s agree, at least approximately, with the theoretical mean and standard deviation computed from your result in (A).

Now repeat the process in part (B) for $\hat{p}_{10}$, $\hat{p}_{25}$, $\hat{p}_{50}$, and $\hat{p}_{100}$. Make a graph that overlays two sets of points: (i) the sample standard deviation of $\hat{p}_N$ versus N for the five different values of N that you used in your simulations; and (ii) the corresponding theoretical standard deviations versus N, calculated from your result in Part A. Comment on the patterns you see in the graph.

Problem 5: more PDF/CDF practice

A non-negative continuous random variable X is said to follow an exponential distribution with parameter $\lambda > 0$ if its PDF is given by

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$, and is zero otherwise. The parameter λ is usually referred to as the rate.

Suppose that $X_1, \ldots, X_N$ are a set of N independent samples from an exponential distribution with rate λ. Let $Y_N = \max\{X_1, \ldots, X_N\}$ be the maximum value in your sample. Derive the PDF of Y_N for fixed N.

Here are some hints:

- For any y, we know that $\max\{X_1, \ldots, X_N\} \leq y$ if and only if $X_i \leq y$ for all $i = 1, \ldots N$.
- The following integral will help:
$$\int_0^x \lambda e^{-\lambda t} \, dt = 1 - e^{-\lambda x} \, .$$