Assignment 7

CSCI2244-Randomness and Computation
due Tuesday, April 10 at 11:59 PM

This assignment consists of a single problem (well, okay, maybe 8 problems, but they're all about the same distribution) that is a sort of 'everything you need to know about the fundamentals of continuous densities, their expectation and variance'. To set the stage, consider the function

$$p(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{x+1}, & x \geq 0. \end{cases}$$

Although its graph superficially resembles that of the exponential density, this function cannot be a probability density function, nor can any constant multiple of it be a density function. This is because

$$\int_0^x p(t)dt = \ln(x+1),$$

which approaches infinity as $x \to \infty$. Thus the area under the graph of p is infinite.

One way to produce a practical version of this distribution is to have a cutoff value $M > 0$ such that $P(X > M) = 0$. This assignment is devoted to exploring the resulting random variable.

You will have to dust off some of your calculus to do this! If you evaluate all of the integrals by hand, the only special calculus fact you need is the one given above, that the derivative of $\ln(x+1)$ is $\frac{1}{x+1}$; you don't have to use integration by parts, or partial fractions, or fancy substitutions. On the other hand, it's easy to get bogged down and make a computational error, especially when computing the variance. Feel free to use a tool like Wolfram Alpha, which performs both symbolic and numerical integration, to evaluate the integrals in question or to check your work. I am more interested in your ability to write down the appropriate integrals than your skills at evaluating them by hand. Just be sure to show all your work.

1. Let $M > 0$. Define

$$q_M(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{x+1}, & 0 \leq x \leq M \\ 0, & x > M. \end{cases}$$

Determine a constant $C_M > 0$ such that $C_M \cdot q_M(x)$ is the probability density function of a random variable.

2. For this and the remaining parts of the problem, we let $M = 20$ and let X be the random variable with density

$$P_X(x) = C_{20} \cdot q_{20}(x).$$

Find an expression for the cumulative distribution $F_X(x)$. (Note that the 'expression', like our expression for $q_M(x)$, consist of a collection of different formulas for different intervals.)

3. Find $E(X)$.

4. Find the *median* of X—that is, find the value a such that $P(X < a) = \frac{1}{2}$.

5. Write a Python function that generates a value having the same distribution as X. (Recall that you do this by using the inverse function of the cumulative distribution $F_X(x)$: First generate a random uniformly distributed r with $0 < r < 1$, and return the unique x such that $F_X(x) = r$.)

   Write a second function that calls your generator N times and returns the average value. For large N, you should find agreement with the value you found for $E(X)$.

6. Compute $Var(X)$ and $\sigma(X)$.

7. Chebyshev's inequality guarantees that the probability of X differing by more than k standard deviations from the mean is no more than $\frac{1}{k^2}$. If $k = 2$, this gives

$$P(|X - E(X)| \geq 2\sigma(X)) \leq \frac{1}{4}.$$

Determine the left-hand side exactly, and verify that the inequality holds. (Usually Chebyshev's inequality gives a very crude bound, so the actual probability could turn out to be much smaller.) What about *three* standard deviations?

8. Find the standard deviation of the average of N independent samples each with the same distribution as X. In principle, this should provide a rough estimate of how far the experimental result in problem 5 differs from $E(X)$. How large does N need to be for this difference to be less that 10^{-3}? Is this borne out by the experimental results?