IS 606: Statistics and Probability for Data Analytics
Hands-On Laboratory Series
Continuous Probability Distributions: Fundamentals II

Overview

This exercise is designed to give you practice in working with a continuous probability distribution in order to provide key summary properties of the distribution. In particular, we will focus on the expected value, variance, and standard deviation of a distribution.

Prerequisites

You should have a good understanding of the first fundamentals lab material for continuous distribution. We will again be making use of the basic concepts of calculus (in particular, integration).

Materials

This lab exercise is self-contained.

Instructions

This lab exercise is to be completed step by step according to the instructions given. If you are struggling with a particular step, then our recommendation is that you look to the solution ***for only that step*** for help. Once you have sorted out the details of the step in question, proceed to the next task.

Interspersed within the instructions are some short, basic tutorials. Further reading is available in chapter 6 of the Dobrow textbook.

Expected Value

Recall that the expected value of a distribution is one measure of the typical value that will occur in a random sample. In particular, expected value is equivalent to what we would expect, on average, for the mean of a random sample of a distribution. (Another measure of typical value is the median, which was briefly covered in the last lab.)

We define expected value in the continuous case by replacing the summations used in the discrete case with integrals. Thus, the expected value of a random variable X is given by:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Again, we often will adjust the bounds for piecewise-defined functions. Let's consider an example. Suppose we have a probability distribution given by the following PDF:

$$f(x) = \frac{3}{x^4} \qquad\qquad x > 1$$

What is the expected value of this distribution?

$$E[X] = \int_{1}^{\infty} x \cdot \frac{3}{x^4}\, dx = \int_{1}^{\infty} \frac{3}{x^3}\, dx = \left[\frac{-3}{2x^2}\right]_1^{\infty} = 0 - \left(-\frac{3}{2}\right) = \frac{3}{2}$$

1. Consider the following probability density function:

   $$f(x) = \frac{3}{2}x^2 \qquad\qquad -1 < x < 1$$

   Determine the expected value of this probability distribution.

2. For the previous exercise, verify that your answer makes sense by plotting the PDF.

3. Consider the following probability density function:

   $$f(x) = xe^{-x} \qquad\qquad x \geq 0$$

   Determine the expected value of this probability distribution. (Hint: Integration by parts is useful here.)

4. Consider the following probability density function:

   $$f(x) = \frac{10}{x^2} \qquad\qquad x > 10$$

   Argue that there is no reasonable choice for an expected value of this distribution.

Variance and Standard Deviation

The variance of a random variable X is defined similarly to the variance of the discrete case:

$$Var(X) = E[(X - E(X))^2]$$

Notice that what we are asking is what we will see as a typical squared distance of an observation from the mean of the distribution. Thus, variance measures the typical spread of a distribution.

Just as in the case of a discrete random variable, it is often more convenient to work with an alternative formulation:

$$Var(X) = E[X^2] - (E[X])^2$$

Note that in the Dobrow textbook, variance is represented by $V(X)$ instead of $Var(X)$. I prefer the latter, as it is both more common and more natural to me. However, they can be used interchangeably.

As always, the standard deviation is just the square root of the variance:

$$SD(X) = \sqrt{Var(X)}$$

5. Consider the following probability density function:

   $$f(x) = \frac{3}{2}x^2 \qquad -1 < x < 1$$

   Determine the variance and standard deviation of this probability distribution.

6. Consider the following probability density function:

   $$f(x) = xe^{-x} \qquad x \geq 0$$

   Determine the variance and standard deviation of this probability distribution. (You will again need integration by parts.)

7. A statistics professor once asked for the variance of the following distribution on a test:

   $$f(x) = \frac{4}{x^3} \qquad x \geq \sqrt{2}$$

   What answer would you give in this case? (Hint: Start with $E[X^2]$ and see what that gives you. The professor had intended the problem to be solvable. What had she overlooked?)

Typical Ranges

When we ask what we expect to occur for a particular probability distribution, we often like to give an answer in the form of a range. For instance, one commonly used rule is that anything within two standard deviations of the expected value is to be considered typical. However, this is not always an effective rule to use. Consider the following problem:

8. Consider the following probability density function:

$$f(x) = \frac{3}{x^4} \qquad x > 1$$

   Using the expected value and standard deviation computed previously, what range of values would you consider typical? Do you notice any problems with your answer? A plot of the PDF may help.

For this reason, the rule about two standard deviations is often best applied only in the case of symmetric distributions.

9. Does the $E(x) \pm 2\,SD(X)$ rule make sense for the following distribution? Explain, using both calculated results and a plot.

$$f(x) = xe^{-x} \qquad x \geq 0$$

10. Does the $E(x) \pm 2\,SD(X)$ rule make sense for the following distribution? Explain, using both calculated results and a plot.

$$f(x) = \frac{3}{2}x^2 \qquad -1 < x < 1$$

11. Does the $E(x) \pm 2\,SD(X)$ rule make sense for the following distribution? Explain, using both calculated results and a plot.

$$f(x) = \frac{3}{4}(-x^2 + 1) \qquad -1 < x < 1$$

Summary

The probability density function can be used to determine measures of typical behavior. In particular, we are often interested in both the expected value and the variance of the distribution.

In each of the next several labs, we will work with a specific family of distributions.