Probability Functions 5M

There are two types of random variables:

- *Discrete*
  Takes values from a specified set of exact countable values
- *Continuous*
  Takes values from a specified interval or a collection of intervals

Probability Mass Function (PMF)

If a random variable X is discrete, then for any value x, the probability of $X = x$ can be calculated by evaluating its ***probability mass function (PMF)*** at x, which is commonly denoted by one of the following:

- $\Pr(X = x)$
- $p_X(x)$
- $p(x)$

Coach's Remarks

The subscript X in $p_X(x)$ refers to the random variable X and is usually included to distinguish between multiple random variables. However, if the random variable is implied, it is common to drop the subscript.

The PMF produces probabilities and all probabilities must be between 0 and 1. The sum of the probabilities of all possible values of a random variable must equal 1. Therefore, a valid PMF must satisfy the following properties:

- $0 \leq p(x) \leq 1$
- $\sum_{\text{all } x} p(x) = 1$

Probability Density Function (PDF)

A ***probability density function (PDF)***, or ***density function*** for short, is the equivalent of a PMF for continuous random variables. For a continuous random variable X, its PDF is denoted by:

- $f_X(x)$
- $f(x)$

Unlike a PMF, evaluating a PDF at a certain value does not produce the probability at that value. In other words,

$$f(x) \neq \Pr(X = x)$$

Instead, calculate probabilities for a continuous random variable by integrating the PDF. More generally,

$$\Pr(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

Using the same logic, we can conclude the probability of X being any particular value is zero:

$$\begin{aligned}\Pr(X = a) &= \Pr(a \leq X \leq a) \\ &= \int_a^a f(x)\,\mathrm{d}x \\ &= 0\end{aligned}$$

This is because, for continuous random variables, the possible values are uncountably infinite, making any exact value immensely improbable. Thus, non-zero probabilities exist over **intervals** of values.

Coach's Remarks

Although evaluating a PDF at a particular value does not yield a probability, it describes the relative likelihood for the continuous random variable to take on a given value. For example, if the PDF evaluated at a is greater than at b, then X is more likely to be in the neighborhood of a than in the neighborhood of b.

Recall that PMFs must satisfy two properties to be valid. There are two analogous properties for PDFs.

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)\, \mathrm{d}x = 1$

All PDFs must be non-negative. However, they are allowed to exceed 1 because PDFs do not represent actual probabilities. A PDF above 1 **does not** translate to a probability exceeding 1. The integral of the PDF over all values of the random variable must equal 1.