MATH 136-02, 136-03 Calculus 2, Fall 2018

Section 7.8: Probability and Integration

This section focuses on a key type of function in the theory of probability, namely the **probability density function** (PDF). Probability is a subject that focuses on the likelihood or chance that a particular event will occur. Events are described by **random variables**, such as the income of someone in the United States, or the height of a random Holy Cross student.

The notation used for probabilities is fairly straight-forward. $P(a \leq x \leq b)$ means the probability that the variable x (measuring income, height, GPA, etc.) lies between the values a and b. For instance, if x represents the yearly income of a typical US citizen, then

$$P(20{,}000 \leq x \leq 30{,}000) \;=\; 0.24$$

means the probability that a typical US citizen makes between \$20,000 and \$30,000 in one year is 24%. The value of a probability is always a percent, that is, a number between 0 and 1. A probability of 0 means the event has no chance of occurring while a probability of 1 means the event will absolutely take place. The statement

$$P(x \geq 300{,}000) \;=\; 0.01$$

means that 1% of the US population has an income greater than \$300,000.

Probability Density Functions

The primary purpose of a probability density function is to compute probabilities. This is done by evaluating a definite integral.

Definition 0.1. *A probability density function $p(x)$ satisfies the following:*

(i) $p(x) \geq 0$ *for all* x,

(ii) $\displaystyle\int_{-\infty}^{\infty} p(x)\,dx \;=\; 1$

(iii) $\displaystyle P(a \leq x \leq b) \;=\; \int_a^b p(x)\,dx.$

Notes about PDF's:

- The third item is the real point of the definition. We compute the probability that x lies between a and b by evaluating the integral of p from a to b. In other words, the probability that x lies between a and b is equal to the area under the PDF from a to b.

- The first item in the definition states that the graph of p cannot lie below the x-axis. This means that $\int_a^b p(x)\,dx \geq 0$, so that $P(a \leq x \leq b) \geq 0$. This is to be expected because the value of a probability should always be positive or 0.

- The second item in the definition states that the total area under the graph of p is equal to 1. Taken with the third item in the definition, this means that $P(-\infty < x < \infty) = 1$, which makes logical sense; the probability that a real random variable lies somewhere on the real line is 100%. Moreover, since $p(x) \geq 0$, and the total area under the graph of p is 1, $\int_a^b p(x)\,dx \leq 1$ always. It follows that

$$0 \;\leq\; \int_a^b p(x)\,dx \;\leq\; 1 \qquad \text{or} \qquad 0 \;\leq\; P(a \leq x \leq b) \;\leq\; 1,$$

  which agrees with the fact that probabilities are always percentages between 0% and 100%.

Exercises

1. Find the value of C that makes $p(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{C}{(x+2)^2} & \text{if } x \geq 0 \end{cases}$ a probability density function. Then compute $P(0 \leq x \leq 1)$ and $P(x \geq 1)$.

2. Show that $f(x) = \begin{cases} 0 & \text{if } x < 0 \\ ke^{-kx} & \text{if } x \geq 0 \end{cases}$ is a probability density function for any constant $k > 0$.

   This PDF is known as the **exponential density function**.

3. Suppose that the probability a telephone call made in the US lasts between a and b minutes is modeled by the exponential density function with $k = 1/4$.

   a) What is the probability that a call lasts between 2 and 3 minutes?

   b) What is the probability that a call lasts over an hour?

   c) Find the value of T so that the probability of a call selected at random lasting longer than T minutes is 50%. T is known as the **median**.

Mean or Average Value

One important quantity associated to any probability density function is the **mean**. Intuitively, the mean measures the average value of x over the long run.

Definition 0.2. *The* <u>*mean*</u> *of a PDF $p(x)$, denoted as μ (pronounced "mu"), is*

$$\mu = \int_{-\infty}^{\infty} x\, p(x)\, dx.$$

It is the average value of the random variable x over the long run.

4. Show that the mean of the exponential density function is $1/k$.

5. Show that $f(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & \text{if } -2 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$ is a probability density function, and then calculate its mean. *Hint:* Draw a graph of f and interpret the integrals in terms of area.