# MATH 134-02, Spring 2014

## Probability Density Functions (Section 6.8)

This worksheet 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 **continuous random variables**, such as the income of someone in the United States, or the GPA 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 random 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. 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* <u>*probability density function*</u> $f(x)$ *satisfies the following:*

**(i)** $f(x) \geq 0$

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

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

**Notes about PDF's:**

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

- The first item in the definition states that the graph of $f$ cannot lie below the $x$-axis. This means that $\int_a^b f(x)\,dx \geq 0$, 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 $f$ 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 $f(x) \geq 0$, and the total area under the graph of $f$ is 1, $\int_a^b f(x)\,dx \leq 1$ always. It follows that

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

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

**Exercises**

1. 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**.

2. Suppose that the probability a telephone call made in the US lasts between $a$ and $b$ minutes is modeled by the exponential PDF 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 and Median**

There are two important quantities associated to any probability density function, called the **mean** and the **median**. Intuitively, the mean measures the average value of $x$ over the long run. It is the average value of our random variable $x$. The median is the value that divides the probability in half. In other words, the probability of $x$ being less than the median equals the probability that $x$ is greater than the median equals 1/2. Geometrically, the median is the value on the $x$-axis that divides the area under $f$ in half. Here are the integral definitions.

**Definition 0.2.** **(i)** *The* <u>*mean*</u> *of $f$, denoted as $\mu$ (pronounced "mu"), is*

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

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

**(ii)** *The* <u>*median*</u> *of $f$, denoted $m$, is the $x$-value that divides the area under $f$ in half. It is found by solving*

$$\int_{m}^{\infty} f(x)\,dx = \frac{1}{2} \quad \text{or} \quad \int_{-\infty}^{m} f(x)\,dx = \frac{1}{2}$$

*for $m$.*

3. 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 and median. *Hint:* Draw a graph of $f$ and interpret the integrals in terms of area.

4. Find the mean and median of the PDF $g(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$

5. Show that the mean of the exponential density function is $1/k$, and that its median is $\ln(2)/k$.