Stat 381: Applied Statistical Methods I **Spring 2015**

February 2,4,6

Instructor: Brian Powers

9.1 Random Variables

Definition 9.1. A **random variable** is a function that maps each outcome in the sample space to a real number.

Example 9.2. Here are some examples of random variables.

- Suppose you roll a die. Let X be the number of the die.
- Flip a coin 10 times. Let Y be the number of heads.
- Sample a random frog from the lake. Let W be its weight.
- Pick a person at random. Let $G = 1$ if the person wears glasses, $G = 0$ otherwise.

In cases where a random variable is categorical as in the last example, we may call this a **dummy variable**. And for review, we repeat the following definitions:

Definition 9.3. If a sample space contains a finite number of possibilities or countably infinite, then it is called a **discrete sample space**. If the sample space is uncountably infinite, it is called **continuous**.

Definition 9.4. A random variable is a **discrete random variable** if its set of possible outcomes is countable. If a random variable can take any value on a continuous interval, it is a **continuous random variable**.

Typically a discrete random variable is something that is counted, while a continuous random variable is measured.

9.2 Discrete Probability Distributions

Definition 9.5. A function $f(x)$ is a **probability mass function (pmf)** of discrete random variable X which takes values $x_1, x_2, \ldots$, provided for each value x_i,:

- $f(x_i) \geq 0$
- $\sum_i f(x_i) = 1$
- $P(X = x_i) = f(x_i)$

A probability mass function can often be well defined by a table or a formula.

Example 9.6. An urn has 7 red balls and 4 blue balls. You reach in and grab 3. Let X be the number of red balls in the selection. Find its pmf.

Example 9.7. You take a multiple choice quiz: 5 questions with 4 choices each. Let X be the number of questions you get right. find its pmf.

It is often convenient rather thank looking at $P(X = x)$ to look at $P(X \leq x)$.

Definition 9.8. The **cumulative distribution function** of discrete random variable X is denoted $F(x)$, where

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{for } x \in \mathbb{R}.$$

The pmf may be well represented by a **probability histogram**, and the discrete cumulative distribution function may be graphically represented as well.

9.3 Continuous Probability Distributions

A continuous random variable has probability mass of 0 for any particular value the random variable can take. What is meaningful is the *probability density* for any particular value.

Definition 9.9. A function $f(x)$ is a **probability density function (pdf)** for a continuous random variable X taking values over $\mathbb{R}$ provided:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a < X < b) = \int_a^b f(x)dx$

Notice that

$$P(a \leq X \leq b) = P(a < X < b) + P(X = a) + P(X = b) = P(a < X < b) + 0 + 0.$$

Example 9.10. Show that the following is a valid pdf:

$$f(x) = \begin{cases} \frac{x^2}{3} & -1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

And find $P(0 < X \leq 1)$.

Definition 9.11. The **cumulative distribution function (cdf)** of a continuous random variable X is denoted $F(x)$ where

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \text{ for } x \in \mathbb{R}.$$

By the Fundamental Theorem of Calculus, we may write

$$P(a < X < b) = F(b) - F(a) \text{ and } P(X > a) = 1 - F(a).$$

Example 9.12. Find the cdf of the previous example, and use it to evaluate $P(.4 < X < .9)$.

Example 9.13. Find the constant c which makes the following a valid pdf, then find the cdf and evaluate $P(X > 4)$.

$$f(x) = \begin{cases} ce^{-2x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$