Probability Theory Review

This handout includes a review of concepts from probability theory that we will make use of this semester. Along with the distribution handout, these should cover (almost) everything you need to know for DSST330. Note that only the information in the random variables sections below will be actively used in the first unit of the course, so focus on reviewing that section for now.

PROBABILITY SPACES

A **probability function** $\mathbb{P}$ defined over a set S called the **sample space** associates a number between 0 and 1 to subsets $A \subset S$, known as **events**, with the following properties:

1. $\mathbb{P}[S] = 1$.
2. For every pair of events A and B such that $A \cap B = \emptyset$, called **mutually exclusive** events, we have:

$$\mathbb{P}(A \cup B) = \mathbb{P}A + \mathbb{P}B.$$

The pair $(S, \mathbb{P})$ is called a probability space.

Let A and B be events from a sample space S such that $\mathbb{P}(B) > 0$. The **conditional probability** of A given B is written $\mathbb{P}(A|B)$ and defined as:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

A set of events are called **(mutually) independent** if the probability of their intersection is equal to the product of their individual probabilities. In particular, two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

RANDOM VARIABLES

A **random variable** X is a mapping from a sample space into the real numbers.[1] We can describe a random variable through the **cumulative distribution function (cdf)**, given by:

$$F_X(x) = \mathbb{P}[X \leq x], \quad x \in \mathbb{R}.$$

If the cdf is a step function, we say that X is a discrete random variable. We say that X is a continuous random variable if the cdf is continuous. For a discrete random variable we can define the **probability mass function (pmf)** $p_X(x)$ by:

$$p_X(x) = \mathbb{P}[X = x].$$

[1] We could write a random variable as $f(s)$ or $X(s)$, but almost always avoid the function notation in favor of capital letters.

We can drop the subscript when it is clear what random variable we are writing the mass function for. For a continuous random variable, we instead define the **probability density function (pdf)** $f_X(x)$ as the derivative of the cdf. Through the fundamental theorem of calculus we have:

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x)dx.$$

Usually, we define random variables by providing their pmf or pdf.[2]

The **expected value** $\mathbb{E}[X]$ of a discrete random variable X is defined by the pmf as:

$$\mathbb{E}[X] = \sum_x x \cdot p_X(x) = \sum_x x \cdot \mathbb{P}[x = X],$$

For a continuous random variable, we define the expected value as:

$$\mathbb{E}X = \int_{-\infty}^{\infty} x \cdot f_X(x)\ dx.$$

The **variance** $\text{Var}[X]$ of a random variable X is given by the expected squared distance away from the expected value: $\mathbb{E}[(X - \mathbb{E}X)^2]$.

For any random variable X and constants a and b, we have that:

$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}[X] + b, \quad \text{Var}[aX + b] = a^2 \cdot \text{Var}[X].$$

For two random variables X and Y, we also have that $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. The variance of the sum is equal to the sum of the variances if the random variables are independent, a concept defined in the following section.

[2] A table of common families of distributions is linked to at the top of the course website. Typically, we denote the pdf/cdf of a random variable by indicating which family and with which parameters a random variable comes from. For example, $X \sim N(0, 1)$ would indicate that $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

JOINT DISTRIBUTIONS

Let $X_1, \ldots, X_n$ be a sequence of n random variables defined over the same sample space S. We can define the **joint probability density function** as a function f such that:

$$\mathbb{P}[(a_1 \leq X_1 \leq b_1) \cap \cdots \cap (a_n \leq X_n \leq b_n)] = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\ dx_1 \cdots dx_n.$$

We say that the sequence of random variables is **independent** if we have the following factorization for all values of x_i:

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

Similarly, we have the following definition of a **conditional probability density function** for two random variables X and Y:

$$f_{X|Y}(x, y) = \frac{f(x, y)}{f_Y(y)}$$

While we will work with sets of independent random variables starting in the first week of the class, we will not make deep use of these joint distributions until the second unit.