RANDOM VARIABLES

Definition: Recall that a **probability space** $(S, \mathcal{E}, P)$ is composed of a sample space S, the algebra $\mathcal{E}$ (see notes), and a probability function $P : \mathcal{E} \to [0,\ 1]$ that satisfies Kolmogorov's axioms.

In practice, we think of **random variables** (r.v.) in two ways.

1. We commonly think of a random variable as a "place holder" for the observed outcome of an experiment. Ex: *Let X be the number of heads in 10 coin tosses.*

2. Formally, if X is a random variable, it is a real-valued *measurable function* that maps one probability space into another (real-valued) probability space. That is, $X : (S, \mathcal{E}, P) \to (\Omega, \mathcal{F}, P_X)$ where $\Omega \subseteq \mathbb{R}$ and we define

   $$P_X(A) = P(s \in \Omega : X(s) \in A) \text{ for all events } A \in \mathcal{F}$$

   Q: How are these consistent?

   A: We tend to only be explicit about the real-valued representation of the outcome, and focus on X and P_X instead of explicitly defining all of the other details.

Definition: We refer to P as the **distribution of the random variable** and this often is sufficient to imply the structure of the associated probability space and experiment.

Example 1: Stating that "X is a Bernoulli r.v. with probability p of success" implies that $S = \{0, 1\}$ and $P(X = k) = p^k(1-p)^{1-k}$. That is, $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

Definition: A **Bernoulli process** $X = (X_1, ..., X_n)$ is a series of n *independent and identically distributed* (*iid*) Bernoulli trials (X_i) each with probability p of success.

Example 2: Stating that "Y is a binomial r.v. with parameters n and p" implies that $Y = \sum_{i=0}^{n} X_i$ is the number of successes in a Bernoulli process of length n, and therefore that $S = \{0, 1, ..., n\}$ and $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in S$ (zero otherwise).

Example 3: If X is a hypergeometric r.v., it represents the number of successes in n draws from a population of size N with K successes. Thus $S = \{0, ..., \min(K, n)\}$ and

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$