Brief Review of Probability
Çinlar, *Introduction to Stochastic Processes*, *pp.* 1–20.

EVENTS, RANDOM VARIABLES, AND SAMPLE SPACES

Colloquially a *random variable* is some quantity that depends on the outcome of some *random experiment* (toss of a die, draw from an urn, public opinion poll, laboratory measurement, *etc.*); we'll need to be a little more precise. A *Sample Space* is just the collection of possible outcomes of such an experiment; it's often denoted by the upper-case Greek letter Ω, and the possible outcomes themselves by lower-case ω's. The Random Variable, a "random quantity depending on the outcome...", is now seen to be a *function* $X : \Omega \to \mathbb{R}$, for a Real-valued random variable, or more generally $X : \Omega \to \mathcal{S}$, for a random variable taking values in some "state space" $\mathcal{S}$.

An *event* in common speach is something that might happen, and then again might not; in probability theory we represent such a thing by the **set** of possible outcomes $E \subset \Omega$ for which the event occurs. This allows us to use set notation to describe most of the usual combinations of events ("A and B" becomes the intersection $A \cap B$, "A or B" becomes the union $A \cup B$, "Not A" becomes the complement $A^c = \{\omega \in \Omega : \omega \notin A\}$, "At least one of $\{A_i\}$" becomes the countable union $\cup_{i=1}^{\infty} A_i$, "A implies B" becomes $A \subset B$, *etc.*). Be careful not to confuse OUTCOME (one of the ω's) with EVENT (subset of Ω, possibly containing infinitely-many outcomes ω). The collection of events we will consider is often denoted by $\mathcal{F}$ or a nearby letter ($\mathcal{E}$, $\mathcal{G}$...); the biggest possible such collection would be the "power set" or set of all possible subsets of the sample space, sometimes denoted $\wp(\Omega)$ or 2^{Ω}; in small problems (dice, coins, *etc.*) this works fine, but it turns out that it's "too big" for most problems with continuous distributions... so $\mathcal{F} \subset 2^{\Omega}$ but $\mathcal{F} \neq \Omega$. For technical reasons $\mathcal{F}$ is required to have certain closure properties (for example, it must contain A^c for each event $A \in \mathcal{F}$) and so is what mathematicians call a "Sigma algebra" or "Borel Field"; this is often shortened simply to "Field" of events ("set of sets" would be too confusing!)

- **Sample Space:** $\Omega = \{\omega\}$, the set of *possible outcomes* of some random experiment;
- **Outcome:** $\omega \in \Omega$, a single element of the Sample Space;
- **Event:** $E \subset \Omega$, a subset of the Sample Space;
- **Field:** $\mathcal{F} = \{E : E \subset \Omega\}$, the collection of Events we'll consider;
- **Random Variable:** $X : \Omega \to \mathcal{S}$, a function from the Sample Space Ω to a State Space $\mathcal{S}$;
- **State Space:** $\mathcal{S}$, a space containing the possible values of a random variable— common choices are the integers $\mathbb{N}$, reals $\mathbb{R}$, k-vectors $\mathbb{R}^k$, complex numbers $\mathbb{C}$, positive reals $\mathbb{R}_+$, *etc*;
- **Probability:** $\mathsf{P} : \mathcal{F} \to [0, 1]$, obeying rules 1–3 below;
- **Distribution:** $\mu : \mathcal{B} \to [0, 1]$, where $\mathcal{B} \subset \{A : A \subset \mathbb{R}\}$ is the Borel sets (intervals, *etc.*).

Probabilities and Distributions

What is missing so-far is an assignment of Probability to those random events that "might happen and, then again, might not"... an assignment of a number $0 \le \mathsf{P}[A] \le 1$ to each event $A \in \mathcal{F}$ (NOT to each outcome $\omega \in \Omega$!). The assignment must satisfy some rules consistent with our notions of probability... for example, we must have $\mathsf{P}[A] \le \mathsf{P}[B]$ if $A \subset B$ (for then B certainly occurs when A does, so its probability must be no smaller); the usual definition of a Probability Assignment (or Measure) requires:

1. For any $A \in \mathcal{F}$, $0 \le P[A] \le 1$;
2. $\mathsf{P}[\Omega] = 1$;
3. For any sequence $A_1, A_2, ...$ of disjoint events, $\mathsf{P}\big(\cup A_i\big) = \sum \mathsf{P}[A_i]$.

The usual properties of probabilities follow from these rules and some work... see me for outside readings if this is unfamiliar or confusing.

The Distribution of a random variable X is just the assignment of probabilities to the events $X \in A$ for each set $A \subset \mathcal{S}$, *i.e.*, the specification of probabilities to sets of outcomes of X. For real-valued random variables it turns out that it's enough to just specify the probabilities for events of the form $[X \le t]$ (*i.e.*, $X \in A_t$ for sets $A_t = (-\infty, t]$); the function these define is called the cumulative **Distribution Function** (or CDF) for X, usually denoted by upper-case $F(t) \equiv \mathsf{P}[X \le t] = \mathsf{P}\big[X \in (-\infty, t]\big]$. If this is a continuous and differentiable function, its derivative $f(t) \equiv F'(t)$ is called the probability **density function** (or pdf) for X, and X is said to have a **continuous** distribution; if instead it's piecewise constant, changing only by jumps of size p_k at the possible values x_k of X, then the function assigning $p_k = f(x_k)$ is called the **probability function** for X and X is called **Discrete**.

By definition, the probability assignment $A \mapsto \mathsf{P}[X \in A]$ *is* the **Probability Distribution** of X; it's easy to show that this assignment satisfies rules 1–3 above, so it's a probability measure (often denoted μ_X) on $\mathcal{B}$, *the subsets of* $\mathbb{R}$, telling what is the probability $\mu_X[A] = \mathsf{P}[X \in A]$ that X will lie in each subset $A \subset \mathcal{S}$ of possible values of X. For continuous and discrete random variables, respectively, the distributions are given by

$$\mu_X[A] = \int_A f(x)\,dx \qquad \text{or} \qquad \mu_X[A] = \sum_{x_i \in A} p_i.$$

The "same" random variable can be constructed in lots of different ways, with lots of different sample-spaces (and even different state spaces)— but the distribution μ_X is unique, whatever Ω we use. In fact, for real-valued random variables we can *always* use the **canonical probability space** $(\mathbb{R}, \mathcal{B}, \mu_X)$, thinking of the "random experiment" as simply observing X, and the "possible outcomes" as all real numbers $x \in \mathbb{R}$; now the probability assignment is just the distribution, $\mu_X[A] = \mathsf{P}[X \in A]$. Similarly for $\mathcal{S}$-valued random variables we can always use the space $\Omega = \mathcal{S}$ with its Borel sets $\mathcal{B}(\mathcal{S})$ and distribution μ_X; in particular, we'll see later that a Stochastic Process can be thought of simply as an $\mathcal{S}^*$-valued random variable for the path-space $\mathcal{S}^* = \{x_t : \mathcal{T} \to \mathcal{S}\}$ of $\mathcal{S}$-valued trajectories on $\mathcal{T}$. Even though we won't have density functions on infinite-dimensional spaces like $\mathcal{S}^*$, this will allow us to construct stochastic processes with any desired distribution simply by building probability measures on the path space $\mathcal{S}^*$.