Lecture 1

A Review of Basic Probability

It is frequently the case when an experiment is performed that we are mainly interested in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 7 and not be concerned over whether the actual outcome was (1, 6), or (2, 5), or (3, 4). Also, in coin flipping we may be interested in the total number of heads that occur and not care at all about the actual head-tail sequence that results. These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as *random variables*.

Because the real value of a RV is determined by the outcome of the experiment, we may assign probabilities to these possible values of the RV. We call these outcomes *realizations* of the RV.

Example:
Suppose that our experiment consists of tossing 3 fair coins. If we let Y denote the number of heads appearing, then Y is a RV taking on one of the values 0, 1, 2, or 3 with respective probabilities. Furthermore, y is a realization of the RV, which in this case may be either 0, 1, 2, or 3.

$$P\{Y = 0\} = P\{(T, T, T)\} = 1/8$$

$$P\{Y = 1\} = P\{(T, T, H),\ (T, H, T),\ (H, T, T)\} = 3/8$$

$$P\{Y = 2\} = P\{(T, H, H),\ (H, H, T),\ (H, T, H)\} = 3/8$$

$$P\{Y = 0\} = P\{(H, H, H)\} = 1/8$$

Definition: Support
Let X be a RV, and let x be a realization of the RV. Let $\text{Supp}[X]$ denote the *support* of the RV. The support of a RV is the set of all possible values that could be realized in the RV, and is a subset of the real numbers.

Example:
Suppose that our experiment consists of rolling one balanced six-sided die. Let X be the random variable of this experiment, where the quantity of interest is the value on the top face of the rolled die. The support of this RV is as follows,

$$Supp[X] = \{1, 2, 3, 4, 5, 6\}$$

Definition: Discrete Random Variable
A random variable that can take on at most a countable number of possible values is said to be a *discrete random variable* (DRV). In other words, the cardinality of the support is at most countable such that the support is either finite or if infinite it is countably infinite.

$$|Supp[X]| \ \leq \ \mathbb{N}$$

Definition: Probability Mass Function
The *probability mass function* (PMF) is a function that gives the probability that a DRV is exactly equal to some value. For a DRV X, we define the probability mass function $p(a)$ of X by

$$p(a) = P\{X \ = \ a\}$$

$$p : Supp[X] \rightarrow (0, 1]$$

If we were to take the PMF of the support, we should always expect to receive a value between (0, 1], exclusive of 0. Why exclusive of 0 you ask? Because we mention that there are values in the support, that is a non-zero number of elements in the set. One property of the PMF is that the sum of the PMFs of all values in the support is 1, as so

$$\Sigma_{x \in Supp[X]} \ p(x) = 1$$

The RV has a number of realizations, or states, where each state is possible even if it's probability is 1 in a million. The sum of the probabilities of each individual state occurring must be 1. If the sum of the PMF's of all realizations of the RV is greater or less than 1, check your work.

Definition: Cumulative Distribution Function
The cumulative distribution function (CDF) of a real-valued RV X, or just distribution function of X, evaluated at x, is the probability that X will take a value less than or equal to x. The CDF may be an value between 0 and 1 inclusive. This is because the CDF is explicitly defined as a probability, and may take any real number as input. The CDF is denoted as so

$$F(x) = P(X = x)$$

$$F : R \rightarrow [0, 1]$$

One may write the CDF in terms of the PMF, given only realizations of the RV X are inputted and not just any real number and that the inputted value is less than x. In the equation below you may wonder why I use the variables x and y to describe the CDF. Both variables are used because x represents a constant value that exists in the support, whereas y is any arbitrary real number.

$$F(x) = P(X = x) = \Sigma_{\{y:y \in Supp[X] y \leq x\}}\ p(y)$$

Definition: Continuous Random Variables
Continuous random variables describe outcomes in probabilistic situations where the possible values some quantity can take form a continuum, which is often (but not always) the entire set of real numbers R. They are the generalization of discrete random variables to uncountably infinite sets of possible outcomes. such as the continuous set [0, 1].

$$|Supp[X]| \ \leq \ \mathbb{R}$$

Definition: Probability Density Function
When working with continuous random variables the definition of the CDF remains the same, but the PMF does not exist. We instead define $f(x) = F'(x)$, the *probability density function*. Given,

$$P(x \in [a, b]) = F(b) - F(a) = \int_a^b f(x)dx$$

Lets pick apart the equation above. The probability that x is between a and b may be written as the CDF of b minus the CDF of a. The fundamental theorem of calculus allows you to redefine this as an integral. Hence, the probability that x is a value between two realizations of the RV is the integral from a to b of f(x). This allows us to define the following

$$Supp[X] = \{x : f(x) > 0\}$$

Note that the PDF is not a probability. It is a derivative, and is not limited by 1. Furthermore, the PDF may not be negative either. By construction, the PDF may only increase because as x gets larger you are collecting 'more' probability and probability is always non-negative. At the least the probability as x gets larger can stay constant, but it may never decrease. This means that when the support of X goes into the PDF, only a value between 0 and infinity may return.

$$f : Supp[X] \rightarrow (0, \infty)$$

Distributions

RV's are identified by their CDF/PMF if they're discrete, or their CDF/PDF if continuous. When working with random variables, $Bern_(p)$ acts as a place holder for either the PMF or the PDF depending on if the RV is a discrete or continuous variable. The notation $\sim$ means distributed as. The following four formulas define the Bernoulli RV, the binomial RV, the exponential RV, and the normal RV respectively.

$$X \sim Bern(p) = p^x(1-p)^{1-x} \rightarrow PMF\ (discrete)$$

$$X \sim Binom(n,p) = \binom{n}{x} p^x(1-p)^{n-x} \rightarrow PMF\ (discrete)$$

$$X \sim Exp(\lambda) = \lambda e^{-\lambda x} \rightarrow PDF\ (continuous)$$

$$X \sim \mathbb{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2} \rightarrow PDF\ (continuous)$$

That brings us back to the equation, $P(x \in [a,b]) = F(b) - F(a) = \int_a^b f(x)dx$. Hence, the probability of scoring a value between the minimum and maximum realizations of the support must be 1. This is a consequence of integrating the entire support. This is the analogue of the Humpty Dumpty principle.

Now lets take a closer look at the Bernoulli RV. If you plug in p(1), or 1 for x, you get p. If you plug in p(0), or 1 for x, you get 1-p. If you plug in p(1$\frac{}{7),or \frac{1}{7} for}$x, $yougetthevalueinthethirdequationbelow$

$$X \sim Bern(p) = p(1) = p^1(1-p)^{1-1} = p$$

$$X \sim Bern(p) = p(0) = p^0(1-p)^{1-0} = 1-p$$

$$X \sim Bern(p) = p(1) = p^{\frac{1}{7}}(1-p)^{\frac{6}{7}}$$

This specification is problematic. The problem is we don't know the support. If one defines the PMF as a function of the support, then we must also know the support. Every random variable has a support that must be defined. Otherwise, we run into this sticky situation where we are plugging in arbitrary values into inappropriate RVs such as negative numbers. This has to be noted so that we know what values are legal to put into the functions.

Distribution	Support
X $\sim Bern(p)$	Supp[X] = {0, 1}
X $\sim Binom(n,p)$	Supp[X] = {0, 1, ... , n}
X $\sim Exp(\lambda)$	Supp[X] = (0, ∞)
X $\sim \mathbb{N}(\mu, \sigma^2)$	Supp[X] = R

Tuning Knobs

What precisely is the constant p in the PMF? What about n, or λ, or μ, or σ? Let us look at an example.

Example: Let the experiment be flipping a single coin, and the value on the top face of the coin acts as the Bernoulli random variable (RV) X. Let Tails = 0, and Heads = 1, such that $Supp[X] = 0, 1$. Lets define success as flipping Heads. The constant p defines the probability of success, or in this case, the probability of flipping Heads. In the real world the value of the constant p must surely be 0.5. But what if, in the mathematical sense, we let p be 0.8. We find now that we will be much more likely to flips a value of Heads, and if one flips the coin enough times they may find that they flip Heads eight out of ten times. This example serves to make sense of how these constants act as a tuning knob.

Parameters

Definition: Parameters
These constants p, n, λ, μ and σ act as tuning knobs, but they are formally defined as *parameters*. These parameters are something you need to know in order to compute the PMF or PDF, but they are something that can be changed, resulting in an entirely new model. Looking back at the previous example, the parameter p controls how often 0s and 1s occur in the Bernoulli random variable.

What are the legal values for the parameter p for the Bernoulli RV you ask. We must consider what values respect its support. $p \in (0, 1)$. This is the parameter space for the parameter p. The parameter p in the Bernoulli Random Variable is a probability and therefore must be between 0 and 1. You can make the parameter p to 1, we find the following

$$X \sim Bern(1) = p(x) = 1^x(1-1)^{1-x}$$

$$p(1) = 1^1(1-1)^{1-1} = 1$$
$$p(0) = 1^0(1-1)^{1-0} = 0$$

This should make sense intuitively. If you declare the probability of success of a Bernoulli RV to be 1, then you can say confidently that the RV will be successful every time. Conversely, If you declare the probability of success of a Bernoulli RV to be 1, then you can say confidently that the RV will never fail. We call this *degenerate*, and we call the RV a *degenerate random variable* denoted $Deg(1)$. The support of the degenerate random variable is defined as so, $Supp[Deg(1)] = \{1\}$. Although this is possible, it is not interesting because it is not random. As a standard convention, degenerate parameter values are not considered part of the parameter space. For these reasons, we cannot say that the degenerate random variable is a Bernoulli random variable. Deg(0) is also possible, implying p = 0 and that failure is certain. The support of this degenerate random variable is denoted, $Supp[Deg(1)] = \{1\}$.

Let θ denote unknown parameters, let $\vec{\theta}$ denote multiple unknown parameters, and let Θ denote the parameter space.

$$\theta \in \Theta$$

We can now rewrite the definitions of the specified random variables in terms of unknown parameters, $\vec{\theta}$, instead of known parameters, constants. We will take a look at $Bern(\theta)$ and $Binom(\theta_2, \theta_1)$.

$$X \sim Bern(\theta) = \theta^x (1-\theta)^{1-x} \rightarrow PMF \ (discrete)$$

$$X \sim Binom(\theta_2, \theta_1) = \binom{\theta_2}{x} \theta_1^x (1-\theta_1)^{\theta_2 - x} \rightarrow PMF \ (discrete)$$

The parameter space of these random variables are defined as such

Distribution	Support
X $\sim Bern(\theta)$	$\Theta = \{0,\ 1\}$
X $\sim Binom(\theta_2, \theta_1)$	$\Theta = (0,\ 1) \times \mathbb{N}$

The set of all possible Bernoulli parameters is denoted by $_{Bern}$ and is defined as such

$$F_{Bern} = \{\theta^x (1-\theta)^{1-x} : \theta \in (0,\ 1)\}$$

The general parametric model represents all possible parametric models. The general parametric model for discrete random variables (PMF) may be defined as so

$$F = \{\ p(x) : \vec{\theta} \in \Theta\ \}$$

For continuous random variables with unknown parameters (PDF) the notation of the general parametric model is very similar

$$F = \{\ f(x) : \vec{\theta} \in \Theta\ \}$$

The parametric model for Bernouli random variables is defined below. Note a semi-colon is used instead of a comma

$$F = \{\ p(x;\vec{\theta}) : \vec{\theta} \in \Theta\ \}$$