4.2 Discrete vs. Continuous

Thus far we have only studied discrete random variables, i.e. random variables that take on only up to *countably* many values. The word "countably" refers to a property of a set. We say a set is *countable* if we can describe a method to list out all the elements in the set such that for any particular element in the set, if we wait long enough in our listing process, we will eventually get to that element. In contrast, a set is called *uncountable* if we cannot provide such a method.

4.2.1 Countable vs. Uncountable

Let's first look at some examples.

Example 4.3. The set of all natural numbers

$$\mathbb{N} \doteq \{1, 2, 3, \dots\}$$

is countable. Our method of enumeration could simply be to start at 1 and add 1 every iteration. Then for any fixed element $n \in \mathbb{N}$, this process would eventually reach and list out n.

Example 4.4. The integers,

$$\mathbb{Z} \doteq \{0, 1, -1, 2, -2, 3, -3, \dots\}$$

is countable. Our method of enumeration as displayed above is to start with 0 for the first element, add 1 to get the next element, multiply by -1 to get the third element, and so on. Any integer $k \in \mathbb{Z}$, if we continue this process long enough, will be reached.

Example 4.5. The set of real numbers in the interval $[0, 1]$ is uncountable. To see this, suppose for the sake of contradiction that this set were countable. Then there would exist some enumeration of the numbers in decimal form. It might look like

$$\begin{array}{l}
0\ .\ 1\ 3\ 5\ 4\ 2\ 9\ 5\dots \\
0\ .\ 4\ 2\ 9\ 4\ 7\ 2\ 6\dots \\
0\ .\ 3\ 9\ 1\ 6\ 8\ 3\ 1\dots \\
0\ .\ 9\ 8\ 7\ 3\ 4\ 3\ 5\dots \\
0\ .\ 2\ 9\ 1\ 8\ 1\ 3\ 6\dots \\
0\ .\ 3\ 7\ 1\ 6\ 1\ 8\ 2\dots \\
\vdots
\end{array}$$

Consider the element along the diagonal of such an enumeration. In this case the number is

$$a \doteq 0.121318\ldots$$

Now consider the number obtained by adding 1 to each of the decimal places, i.e.

$$a' \doteq 0.232429\ldots$$

This number is still contained in the interval $[0, 1]$, but does not show up in the enumeration. To see this, observe that a' is not equal to the first element, since it differs in the first decimal place by 1. Similarly, it is not equal to the second element, as a' differs from this number by 1 in the second decimal place. Continuing this reasoning, we conclude that a' differs from the n^{th} element in this enumeration in the n^{th} decimal place by 1. It follows that if we continue listing out numbers this way, we will *never* reach the number a'. This is a contradiction since we initially assumed that our enumeration would *eventually* get to every number in $[0, 1]$. Hence the set of numbers in $[0, 1]$ is uncountable.

If you're left feeling confused after these examples, the important take away is that an uncountable set is *much* bigger than a countable set. Although both are infinite sets of elements, uncountable infinity refers to a "bigger" notion of infinity, one which has no gaps and can be visualized as a continuum.

4.2.2 Discrete Distributions

Definition 4.6. *A random variable X is called* ***discrete*** *if X can only take on finitely many or countably many values.*

For example, our coin flip example yielded a random variable X which could only take values in the set $\{0, 1\}$. Hence, X was a discrete random variable. However, discrete random variables can still take on infinitely many values, so long as this infinity is *countable*. An example of this is given below.

Example 4.7 (Poisson Distribution)**.** A useful distribution for modeling many real world problems is the *Poisson Distribution*. Suppose $\lambda > 0$ is a positive real number. Let X be distributed according to a Poisson distribution with parameter λ, i.e.

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

where $k \in \mathbb{N}$. The shorthand for stating such a distribution is $X \sim \text{Poi}(\lambda)$. Since k can be any number in $\mathbb{N}$, our random variable X has a positive probability on infinitely many numbers. However, since $\mathbb{N}$ is countable, X is still considered a discrete random variable.

On the website there is an option to select the "Poisson" distribution in order to visualize its probability mass function. Changing the value of λ changes the probability mass function, since λ shows up in the probability expression above. Drag the value of λ from 0.01 up to 10 to see how varying λ changes the probabilities.

Example 4.8 (Binomial Distribution)**.** Another useful distribution is called the *Binomial Distribution*. Consider n coin flips, i.e. n random variables $X_1, \ldots, X_n$ each of the form

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

Now consider the random variable defined by summing all of these coin flips, i.e.

$$S \doteq \sum_{i=1}^{n} X_i$$

The physical interpretation of the random variable S is the number of heads we flipped (since tails get mapped to 0, the sum of the X_i's is the total number of heads). We might then ask, "What is the probability distribution of S?" Based on the definition of S, it can take on values from 0 to n, however it can only take on the value 0 if all the coins end up tails. Similarly, it can only take on the value n if all the coins end up heads. But to take on the value 1, we only need one of the coins to end up heads and the rest to end up tails. This can be achieved in many ways. In fact, there are $\binom{n}{1}$ ways to pick which coin gets to be heads up. Similarly, for $S = 2$, there are $\binom{n}{2}$ ways to pick which two coins get to be heads up. It follows that for $S = k$, there are $\binom{n}{k}$ ways to pick which k coins get to be heads up. This leads to the following form,

$$P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The p^k comes from the k coins having to end up heads, and the $(1-p)^{n-k}$ comes from the remaining $n - k$ coins having to end up tails. The binomial coefficient $\binom{n}{k}$ comes from the number of ways in which we can flip k heads and $n - k$ tails. Here it is clear that k ranges from 0 to n, since the smallest value is achieved when no coins land heads up, and the largest number is achieved when all coins land heads up. Any value between 0 and n can be achieved by picking a subset of the n coins to be heads up.

Selecting the "Binomial" distribution on the website will allow you to visualize the probability mass function of S. Play around with n and p to see how this affects the probability distribution.

4.2.3 Continuous Distributions

Definition 4.9. *We say that X is a* ***continuous*** *random variable if X can take on uncountably many values.*

If X is a continuous random variable, then the probability that X takes on any particular value is 0.

Example 4.10. An example of a continuous random variable is a Uniform$[0,1]$ random variable. If $X \sim$ Uniform$[0,1]$, then X can take on any value in the interval $[0,1]$ (which we showed was uncountable in Example 4.5), where each value is equally likely. The probability that X takes on any particular value in $[0,1]$, say $\frac{1}{2}$ for example, is 0. However, we can still take probabilities of subsets in a way that is intuitive. The probability that x falls in some interval (a,b) where $0 \leq a < b \leq 1$ is

$$P(X \in (a,b)) = b - a.$$

The probability of this event is simply the length of the interval (a,b).

A continuous random variable is distributed according to a *probability density function*, usually denoted f, defined on the domain of X. The probability that X lies in some set A is defined as

$$P(X \in A) = \int_A f$$

This is informal notation but the right hand side of the above just means to integrate the density function f over the region A.

Definition 4.11. *A* ***probability density function*** *f (abbreviated* ***pdf****) is valid if it satisfies the following two properties.*

1. *$f(x) \geq 0$ for all $x \in \mathbb{R}$*
2. *$\int_{-\infty}^{\infty} f(x)dx = 1$*

Example 4.12 (Exponential Distribution)**.** Let $\lambda > 0$ be a positive real number. Suppose X is a continuous random variable distributed according to the density

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

In this case, we would write $X \sim \exp(\lambda)$.

Let's check that f defines a valid probability density function. Since $\lambda > 0$ and e^y is positive for any $y \in \mathbb{R}$, we have $f(x) \geq 0$ for all $x \in \mathbb{R}$. Additionally, we have

$$\begin{aligned}\int_0^\infty f(x)dx &= \int_0^\infty \lambda e^{-\lambda x} \\ &= \left[\lambda \frac{-1}{\lambda} e^{-\lambda x}\right]_0^\infty \\ &= 0 - (-1) \\ &= 1\end{aligned}$$

Since f is nonnegative and integrates to 1, it is a valid pdf.

Example 4.13. Suppose $X \sim \exp(\lambda)$. We want to find the probability that X is greater than 10. To do this, we integrate over the region of interest A, is $A \doteq \{\omega : X(\omega) > 10\}$. Hence,

$$\begin{aligned}P(X > 10 &= \int_A f \\ &= \int_{\{\omega : X(\omega) > 10\}} \lambda e^{-\lambda x} dx \\ &= \int_{10}^\infty \lambda e^{-\lambda x} dx \\ &= \left[- e^{-\lambda x}\right]_{10}^\infty \\ &= 0 - (-e^{-10\lambda}) \\ &= e^{-10\lambda}\end{aligned}$$

Example 4.14 (Normal Distribution)**.** We arrive at perhaps the most known and used continuous distributions in all of statistics. The Normal distribution is specified by two parameters, the mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. To say X is a random variable distributed according to a Normal distribution with mean μ and variance σ^2, we would write $X \sim N(\mu, \sigma^2)$. The corresponding pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Some useful properties of normally distributed random variables are given below.

Proposition 4.15. *If $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ are independent random variables, then*

(a) The sum is normally distributed, i.e.

$$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

(b) Scaling by a factor $a \in \mathbb{R}$ results in another normal distribution, i.e. we have

$$aX \sim N(a\mu_x, a^2\sigma_x^2)$$

(c) Adding a constant $a \in \mathbb{R}$ results in another normal distribution, i.e.

$$X + a \sim N(\mu_x + a, \sigma_x^2)$$

Heuristic. In order to rigorously prove this proposition, we need to use moment generating functions, which aren't covered in these notes.

However, if we believe that $X+Y$, aX, and $X+a$ are all still normally distributed, it follows that the specifying parameters (μ and σ^2) for the random variables in (a), (b), and (c) respectively are

$$\begin{aligned} E(X+Y) &= EX + EY = \mu_x + \mu_y \\ \mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) = \sigma_x^2 + \sigma_y^2 \end{aligned}$$

and

$$\begin{aligned} E(aX) &= aEX = a\mu_x \\ \mathrm{Var}(aX) &= a^2\mathrm{Var}(X) = a^2\sigma_x^2 \end{aligned}$$

and

$$\begin{aligned} E(X+a) &= EX + a = \mu_x + a \\ \mathrm{Var}(X+a) &= \mathrm{Var}(X) + \mathrm{Var}(a) = \mathrm{Var}(X) = \sigma_x^2 \end{aligned}$$

□