Probability Spaces and Random Variables

Gene Quinn

Probability Spaces

We defined the **probability space** associated with an experiment as having three components:

- The set of possible outcomes of the experiment Ω, called the **sample space**

Probability Spaces

We defined the **probability space** associated with an experiment as having three components:

- The set of possible outcomes of the experiment Ω, called the **sample space**
- A collection $\mathcal{F}$ of subsets of Ω called the **events** or **event space**

Probability Spaces

We defined the **probability space** associated with an experiment as having three components:

- The set of possible outcomes of the experiment Ω, called the **sample space**
- A collection $\mathcal{F}$ of subsets of Ω called the **events** or **event space**
- A probability measure $\rho : \mathcal{F} \to [0, 1]$, that is, a function that associates a number in the interval $[0, 1]$ (i.e., a probability) with each element of the event space $\mathcal{F}$.

Probability Spaces

We defined the **probability space** associated with an experiment as having three components:

- The set of possible outcomes of the experiment Ω, called the **sample space**
- A collection $\mathcal{F}$ of subsets of Ω called the **events** or **event space**
- A probability measure $\rho : \mathcal{F} \to [0, 1]$, that is, a function that associates a number in the interval $[0, 1]$ (i.e., a probability) with each element of the event space $\mathcal{F}$.

A **random variable** is a real-valued function defined on the sample space Ω

Bernoulli random variables

A **Bernoulli trial** is an experiment with two outcomes, **success** and **failure**: $\Omega = \{S, F\}$

Bernoulli random variables

A **Bernoulli trial** is an experiment with two outcomes, **success** and **failure**: $\Omega = \{S, F\}$

It is common to define a random variable Y on this sample space as:

$$Y = \begin{cases} 1 & \text{if} \quad \text{the outcome is } \textbf{success} \\ 0 & \text{if} \quad \text{the outcome is } \textbf{failure} \end{cases}$$

Bernoulli random variables

A **Bernoulli trial** is an experiment with two outcomes, **success** and **failure**: $\Omega = \{S, F\}$

It is common to define a random variable Y on this sample space as:

$$Y = \begin{cases} 1 & \text{if} \quad \text{the outcome is } \textbf{success} \\ 0 & \text{if} \quad \text{the outcome is } \textbf{failure} \end{cases}$$

But a random variable could just a well be defined so as to assign 1 to failure and 0 (or any other number) to success.

Random Variables

The experiment of tossing a coin until the first heads has the (infinite) sample space:

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$$

Random Variables

The experiment of tossing a coin until the first heads has the (infinite) sample space:

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$$

One way to define a random variable on this sample space would be to let Y be the number of tails. Then

$$Y = \begin{cases} 0 & \text{if the outcome is} \quad \text{H} \\ 1 & \text{if the outcome is} \quad \text{TH} \\ 2 & \text{if the outcome is} \quad \text{TTH} \\ 3 & \text{if the outcome is} \quad \text{TTTH} \\ \vdots & \qquad\qquad\qquad\qquad \vdots \end{cases}$$

Random Variables

More generally, we could conduct Bernoulli trials until the first success,

$$\Omega = \{S, FS, FFS, FFFS, FFFFS, \ldots\}$$

Random Variables

More generally, we could conduct Bernoulli trials until the first success,

$$\Omega = \{S, FS, FFS, FFFS, FFFFS, \ldots\}$$

One way to define a random variable on this sample space would be to let Y be the number of failures. Then

$$Y = \begin{cases} 0 & \text{if the outcome is} \quad \text{S} \\ 1 & \text{if the outcome is} \quad \text{FS} \\ 2 & \text{if the outcome is} \quad \text{FFS} \\ 3 & \text{if the outcome is} \quad \text{FFFS} \\ \vdots & \qquad\qquad\qquad\quad \vdots \end{cases}$$

Random Variables

If we conduct exactly 5 independent Bernoulli trials, we could represent the sample space Ω as the set of evey possible sequence of five letters, each one being either S or F (32 sequences in all).

Random Variables

If we conduct exactly 5 independent Bernoulli trials, we could represent the sample space Ω as the set of evey possible sequence of five letters, each one being either S or F (32 sequences in all).

One way to define a random variable on this sample space would be the number of successes in 5 trials. Then

$$Y = \begin{cases} 0 & \text{if there were no successes} \\ 1 & \text{if there was one success} \\ 2 & \text{if there were two successes} \\ 3 & \text{if there were three successes} \\ 4 & \text{if there were four successes} \\ 5 & \text{if there were five successes} \end{cases}$$

Random Variables

If we conduct Bernoulli trials until the second success, the sample space is:

$$\Omega = \{SS, FSS, SFS, FFSS, FSFS, SFFS, FFFSS, \ldots\}$$

Random Variables

If we conduct Bernoulli trials until the second success, the sample space is:

$$\Omega = \{SS, FSS, SFS, FFSS, FSFS, SFFS, FFFSS, \ldots\}$$

One way to define a random variable on this sample space would be to let Y be the number of failures. Then

$$Y = \begin{cases} 0 & \text{if the outcome is} \quad \text{SS} \\ 1 & \text{if the outcome is} \quad \text{FSS or SFS} \\ 2 & \text{if the outcome is} \quad \text{FFSS,FSFS, or SFFS} \\ \vdots & \qquad\qquad\qquad\qquad \vdots \end{cases}$$

Random Variables

It is customary to divide random variables into two classes, **discrete** and **continuous**

Random Variables

It is customary to divide random variables into two classes, **discrete** and **continuous**

A random variable whose underlying sample space contains a non-degenerate (i.e., not a single point) interval on the real line is **continuous**

Random Variables

It is customary to divide random variables into two classes, **discrete** and **continuous**

A random variable whose underlying sample space contains a non-degenerate (i.e., not a single point) interval on the real line is **continuous**

A random variable whose underlying sample space is finite or *countably* infinite is said to be **discrete**.

Generally, a random variable is discrete unless the underlying sample space contains an interval on the real line, including intervals of the form $(-\infty, a)$, (a, ∞) and $(-\infty, \infty)$.

Random Variables

The experiment associated with a continuous random variable can usually be described as "pick a number from the interval (a, b)" with some rule for specifying how likely a number in any given subinterval is to be chosen.

Random Variables

The experiment associated with a continuous random variable can usually be described as "pick a number from the interval (a, b)" with some rule for specifying how likely a number in any given subinterval is to be chosen.

Some examples and their associated sample spaces are:

A Normal or Bell Curve random variable	$(-\infty, \infty)$
A uniform random variable	$(0, 1)$
A chi-square random variable	$(0, \infty)$

Random Variables

The experiment associated with a continuous random variable can usually be described as "pick a number from the interval (a, b)" with some rule for specifying how likely a number in any given subinterval is to be chosen.

Some examples and their associated sample spaces are:

A Normal or Bell Curve random variable	$(-\infty, \infty)$
A uniform random variable	$(0, 1)$
A chi-square random variable	$(0, \infty)$

Generally for continuous random variables, we can think of the function on the sample space that defines them as the identity function, so the number selected in the experiment is identical to the value of the random variable.