STA286 Lecture 03

Neil Montgomery

Last edited: 2017-01-12 12:04

probability

the goal

Consider a variable in a dataset.

We need a mathematical model for the nature of the variation in that variable.

We need to start with a little set theory and to define what is meant by "probability."

We will not talk philosophy (i.e. "what is the meaning of probability?")—we will take an axiomatic approach without worrying too much about very deep meanings.

sample space

A random process will have a set of possible *outcomes* which together is called a *sample space*, often denoted by S.

sample space

A random process will have a set of possible *outcomes* which together is called a *sample space*, often denoted by S.

e.g. toss a coin

$$S = \{H, T\}$$

sample space

A random process will have a set of possible *outcomes* which together is called a *sample space*, often denoted by S.

e.g. toss a coin

$$S = \{H, T\}$$

e.g. roll a six-sided die

$$S = \{1, 2, 3, 4, 5, 6\}$$

sample space

A random process will have a set of possible *outcomes* which together is called a *sample space*, often denoted by S.

e.g. toss a coin

$$S = \{H, T\}$$

e.g. roll a six-sided die

$$S = \{1, 2, 3, 4, 5, 6\}$$

e.g. toss a coin repeatedly until the first time 'H' appears

$$S = \{H, TH, TTH, TTTH, \ldots\}$$

sample space

A random process will have a set of possible *outcomes* which together is called a *sample space*, often denoted by S.

e.g. toss a coin

$$S = \{H, T\}$$

e.g. roll a six-sided die

$$S = \{1, 2, 3, 4, 5, 6\}$$

e.g. toss a coin repeatedly until the first time 'H' appears

$$S = \{H, TH, TTH, TTTH, \ldots\}$$

e.g. run a backhoe, measuring the amount of time until it fails

$$S = (0, \infty)$$

events

An event is a subset of a sample space.

events

An event is a subset of a sample space.

e.g. If $S = \{H, T\}$, the events are:

$$\emptyset, \{H\}, \{T\}, S$$

events

An event is a subset of a sample space.

e.g. If $S = \{H, T\}$, the events are:

$$\emptyset, \{H\}, \{T\}, S$$

Note 1: H by itself is an *outcome*. The set containing $\{H\}$ is an *event*. But:

$$H \neq \{H\}$$

events

An event is a subset of a sample space.

e.g. If $S = \{H, T\}$, the events are:

$$\emptyset, \{H\}, \{T\}, S$$

Note 1: H by itself is an *outcome*. The set containing $\{H\}$ is an *event*. But:

$$H \neq \{H\}$$

Note 2: the empty event is needed for technical reasons. The book tries to make a practical example which is misleading (p.39 "detecting a microscopic…")

events

An event is a subset of a sample space.

e.g. If $S = \{H, T\}$, the events are:

$$\emptyset, \{H\}, \{T\}, S$$

Note 1: H by itself is an *outcome*. The set containing $\{H\}$ is an *event*. But:

$$H \neq \{H\}$$

Note 2: the empty event is needed for technical reasons. The book tries to make a practical example which is misleading (p.39 "detecting a microscopic…")

Another example: is $S = \{H, TH, TTH, TTTH, \ldots\}$ then

$$\{H, TTH, TTTTH, \ldots\}$$

is the event "odd number of tosses."

fun technicality

When the sample space is finite, or an infinite *list* of outcomes, we say the sample space is *countable*. All subsets of a countable sample space can be considered to be events.

All teaching of probability to undergraduate students is limited by some deep techicalities when the sample space is a real interval, such as with:

$$S = (0, \infty)$$

The sample space is *uncountable*, and not *all* subsets are allowed to be events.

This has to do with the hierarchy of sizes of infinite sets. If you're interested, there is a document with the lecture materials you can read.

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, ...

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, ...
- The outcomes in S but *not* in A is an event called A', the *complement* of A. (aka A^c aka $\overline{A}$)

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, ...
- The outcomes in S but *not* in A is an event called A', the *complement* of A. (aka A^c aka $\overline{A}$)
- The outcomes in both A **and** B is an event called $A \cap B$, the *intersection*.

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, ...
- The outcomes in S but *not* in A is an event called A', the *complement* of A. (aka A^c aka $\overline{A}$)
- The outcomes in both A **and** B is an event called $A \cap B$, the *intersection*.
- The outcomes in either A **or** B is an event called $A \cup B$, the *union*.

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, ...
- The outcomes in S but *not* in A is an event called A', the *complement* of A. (aka A^c aka $\overline{A}$)
- The outcomes in both A **and** B is an event called $A \cap B$, the *intersection*.
- The outcomes in either A **or** B is an event called $A \cup B$, the *union*.
- If $A \cap B = \emptyset$ we say A and B are *disjoint* (aka *mutually exclusive*.)

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, $\ldots$
- The outcomes in S but *not* in A is an event called A', the *complement* of A. (aka A^c aka $\overline{A}$)
- The outcomes in both A **and** B is an event called $A \cap B$, the *intersection*.
- The outcomes in either A **or** B is an event called $A \cup B$, the *union*.
- If $A \cap B = \emptyset$ we say A and B are *disjoint* (aka *mutually exclusive*.)
- If $A_1, A_2, A_3, \ldots$ are disjoint and

$$\bigcup_{i=1}^{\infty} A_i = S$$

we say it is a *partition* of S.

reminders (?!) of basic set theory

- Events are typically given names from the beginning of the capital Roman alphabet: A, B, C, A_1, A_2, A_3, ...
- The outcomes in S but *not* in A is an event called A', the *complement* of A. (aka A^c aka $\overline{A}$)
- The outcomes in both A **and** B is an event called $A \cap B$, the *intersection*.
- The outcomes in either A **or** B is an event called $A \cup B$, the *union*.
- If $A \cap B = \emptyset$ we say A and B are *disjoint* (aka *mutually exclusive*.)
- If $A_1, A_2, A_3, \ldots$ are disjoint and

  $$\bigcup_{i=1}^{\infty} A_i = S$$

  we say it is a *partition* of S.
- All of the above work for finite and infinite collections

formal definition of probability

A probability is a *function*.

formal definition of probability

A probability is a *function*.

Given a sample space S and a collection of events $\mathcal{A}$, a *probability* is a function $P : \mathcal{A} \rightarrow \mathbb{R}$ which satisfies:

1. $P(S) = 1$

formal definition of probability

A probability is a *function*.

Given a sample space S and a collection of events $\mathcal{A}$, a *probability* is a function $P : \mathcal{A} \to \mathbb{R}$ which satisfies:

1. $P(S) = 1$
2. $P(A) \geq 0$

formal definition of probability

A probability is a *function*.

Given a sample space S and a collection of events $\mathcal{A}$, a *probability* is a function $P : \mathcal{A} \to \mathbb{R}$ which satisfies:

1. $P(S) = 1$
2. $P(A) \geq 0$
3. If $A_1, A_2, A_3, \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

formal definition of probability

A probability is a *function*.

Given a sample space S and a collection of events $\mathcal{A}$, a *probability* is a function $P : \mathcal{A} \to \mathbb{R}$ which satisfies:

1. $P(S) = 1$
2. $P(A) \geq 0$
3. If $A_1, A_2, A_3, \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

formal definition of probability

A probability is a *function*.

Given a sample space S and a collection of events $\mathcal{A}$, a *probability* is a function $P : \mathcal{A} \to \mathbb{R}$ which satisfies:

1. $P(S) = 1$
2. $P(A) \geq 0$
3. If $A_1, A_2, A_3, \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

"The three axioms"

formal definition of probability

A probability is a *function*.

Given a sample space S and a collection of events $\mathcal{A}$, a *probability* is a function $P : \mathcal{A} \to \mathbb{R}$ which satisfies:

1. $P(S) = 1$
2. $P(A) \geq 0$
3. If $A_1, A_2, A_3, \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

"The three axioms"

One can prove that 3. also works for finite collections of disjoint events.

example of a probability function

Coin toss.

example of a probability function

Coin toss.

$S = \{H, T\}$

example of a probability function

Coin toss.

$S = \{H, T\}$

$\mathcal{A} = \Big\{\emptyset, \{H\}, \{T\}, S\Big\}$

example of a probability function

Coin toss.

$S = \{H, T\}$

$\mathcal{A} = \Big\{\emptyset, \{H\}, \{T\}, S\Big\}$

Define P as follows:

$$
\begin{aligned}
P(\emptyset) &= 0 \\
P(\{H\}) &= 0.5 \\
P(\{T\}) &= 0.5 \\
P(S) &= 1
\end{aligned}
$$

P satisfies the required properties.

a class of valid probability functions

If $S = \{\omega_1, \omega_2, \omega_3, \dots\}$ is *countable* then $\mathcal{A}$ can be the collection of all subsets of S and any function P with:

- $P(\{\omega_i\}) = p_i$

a class of valid probability functions

If $S = \{\omega_1, \omega_2, \omega_3, \ldots\}$ is *countable* then $\mathcal{A}$ can be the collection of all subsets of S and any function P with:

- $P(\{\omega_i\}) = p_i$
- $p_i \geq 0$

a class of valid probability functions

If $S = \{\omega_1, \omega_2, \omega_3, \ldots\}$ is *countable* then $\mathcal{A}$ can be the collection of all subsets of S and any function P with:

- $P(\{\omega_i\}) = p_i$
- $p_i \geq 0$
- $\sum_{i=1}^{\infty} p_i = 1$

a class of valid probability functions

If $S = \{\omega_1, \omega_2, \omega_3, \ldots\}$ is *countable* then $\mathcal{A}$ can be the collection of all subsets of S and any function P with:

- $P(\{\omega_i\}) = p_i$
- $p_i \geq 0$
- $\sum_{i=1}^{\infty} p_i = 1$

a class of valid probability functions

If $S = \{\omega_1, \omega_2, \omega_3, \ldots\}$ is *countable* then $\mathcal{A}$ can be the collection of all subsets of S and any function P with:

- $P(\{\omega_i\}) = p_i$
- $p_i \geq 0$
- $\sum_{i=1}^{\infty} p_i = 1$

is a probability function.

a class of valid probability functions

If $S = \{\omega_1, \omega_2, \omega_3, \ldots\}$ is *countable* then $\mathcal{A}$ can be the collection of all subsets of S and any function P with:

- $P(\{\omega_i\}) = p_i$
- $p_i \geq 0$
- $\sum_{i=1}^{\infty} p_i = 1$

is a probability function.

Example: toss a fair die; assign each side probability $1/6$.

sample spaces with equally likely outcomes

Denote the size of a set A by $|A|$.

sample spaces with equally likely outcomes

Denote the size of a set A by $|A|$.

A common special case of the previous slide is with $|S| = n$ finite and $p_i = 1/n$ for all i.

sample spaces with equally likely outcomes

Denote the size of a set A by $|A|$.

A common special case of the previous slide is with $|S| = n$ finite and $p_i = 1/n$ for all i.

In this case the probabilities will all be:

$$P(A) = \frac{|A|}{|S|}$$

sample spaces with equally likely outcomes

Denote the size of a set A by $|A|$.

A common special case of the previous slide is with $|S| = n$ finite and $p_i = 1/n$ for all i.

In this case the probabilities will all be:

$$P(A) = \frac{|A|}{|S|}$$

This fact explains the existence of textbook sections such as our 2.3 that go on (and on and on) about counting sizes of events. Not a focus of this course.

the axiomatic approach

Some of the basic rules can be formally proven, which is great fun!

Theorem 1 $P(\emptyset) = 0$

Theorem 2 $P(A') = 1 - P(A)$

Theorem 3 If $A \subset B$ then $P(A) \leq P(B)$

Theorem 4 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Corollaries to Theorem 4: $P(A \cup B) = P(A) + P(B)$ when A and B are disjoint, and $P(A \cup B) \leq P(A) + P(B)$ (always).