Lecture 3 - Axioms of Probability

Sta102 / BME102

January 25, 2016

Colin Rundel

Axioms of Probability

What does it mean to say that:

The probability of flipping a coin and getting heads is 1/2?

What does it mean to say that:

The probability of flipping a coin and getting heads is 1/2?

The probability of rolling snake eyes is 1/36?

What does it mean to say that:

The probability of flipping a coin and getting heads is 1/2?
The probability of rolling snake eyes is 1/36?
The probability Apple's stock price goes up today is 3/4?

What does it mean to say that:

The probability of flipping a coin and getting heads is $1/2$?
The probability of rolling snake eyes is $1/36$?
The probability Apple's stock price goes up today is $3/4$?

Interpretations:

What does it mean to say that:

The probability of flipping a coin and getting heads is 1/2?
The probability of rolling snake eyes is 1/36?
The probability Apple's stock price goes up today is 3/4?

Interpretations:

- Symmetry: If there are n equally-likely outcomes, each has probability
$$P(E) = 1/n$$

What does it mean to say that:

The probability of flipping a coin and getting heads is 1/2?
The probability of rolling snake eyes is 1/36?
The probability Apple's stock price goes up today is 3/4?

Interpretations:

- Symmetry: If there are n equally-likely outcomes, each has probability
$$P(E) = 1/n$$
- Frequency: If you can repeat an experiment indefinitely,
$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}$$

What does it mean to say that:

The probability of flipping a coin and getting heads is 1/2?
The probability of rolling snake eyes is 1/36?
The probability Apple's stock price goes up today is 3/4?

Interpretations:

- Symmetry: If there are n equally-likely outcomes, each has probability
$$P(E) = 1/n$$
- Frequency: If you can repeat an experiment indefinitely,
$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}$$
- Belief: If you are indifferent between winning \$1 if E occurs or winning \$1 if you draw a blue chip from a box with $1,000,000 \times p$ blue chips, rest red,
$$P(E) = p$$

Terminology

Outcome space (Ω) - set of all possible outcomes (ω).

Examples:	3 coin tosses	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
	One die roll	{1,2,3,4,5,6}
	Sum of two rolls	{2,3,. . . ,11,12}
	Time waiting for bus	$[0, \infty)$

Terminology

Outcome space (Ω) - set of all possible outcomes (ω).

Examples:	3 coin tosses	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
	One die roll	{1,2,3,4,5,6}
	Sum of two rolls	{2,3,...,11,12}
	Time waiting for bus	$[0, \infty)$

Event (E) - subset of Ω ($E \subseteq \Omega$) that might or might not happen

Examples:	2 heads	{HHT, HTH, THH}
	Roll an even number	{2,4,6}
	Wait $<$ 2 minutes	$[0, 120)$

Terminology

Outcome space (Ω) - set of all possible outcomes (ω).

Examples:	3 coin tosses	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
	One die roll	{1,2,3,4,5,6}
	Sum of two rolls	{2,3,...,11,12}
	Time waiting for bus	$[0, \infty)$

Event (E) - subset of Ω ($E \subseteq \Omega$) that might or might not happen

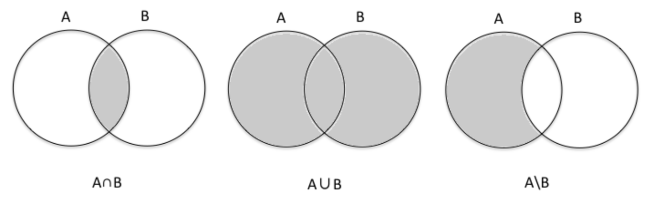
Examples:	2 heads	{HHT, HTH, THH}
	Roll an even number	{2,4,6}
	Wait $<$ 2 minutes	$[0, 120)$

Random Variable (X) - a value that depends somehow on chance

Examples:	# of heads	{3, 2, 2, 1, 2, 1, 1, 0}
	# flips until heads	{3, 2, 1, 1, 0, 0, 0, 0}
	$2^{\text{die value}}$	{2, 4, 8, 16, 32, 64}

Set Operations

Intersection	E and F, EF, $E \cap F$
Union	E or F, $E \cup F$
Complement	not E, E^c, E'
Difference	$E \backslash F = E$ and F^c



Axioms of Probability

Kolmogorov Axioms:

1. $P(E) \geq 0$ for all E

2. $P(\Omega) = P(\omega_1 \text{ or } \omega_2 \text{ or } \cdots \text{ or } \omega_n) = 1$

3. For mututally exclusive (disjoint) events E and F
   $P(E \text{ or } F) = P(E) + P(F)$

Axioms of Probability

Kolmogorov Axioms:

1. $P(E) \geq 0$ for all E

2. $P(\Omega) = P(\omega_1 \text{ or } \omega_2 \text{ or } \cdots \text{ or } \omega_n) = 1$

3. For mututally exclusive (disjoint) events E and F
   $P(E \text{ or } F) = P(E) + P(F)$

From these axioms we can also deduce:

- $P(\emptyset) = 0$
- If $A \subseteq B$ then $P(A) \leq P(B)$
- $0 \leq P(E) \leq 1$ for all $E \subseteq \Omega$

Equally Likely Outcomes

One of the proposed definitions of probably is the equally likely outcomes where $P(\omega_i) = 1/k$

Lets explore what happens with a simple case - rolling a six sided die. Let R be the value showing on the die after rolling.

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

$$P(\omega_i) = 1/6$$

Equally Likely Outcomes

One of the proposed definitions of probably is the equally likely outcomes where $P(\omega_i) = 1/k$

Lets explore what happens with a simple case - rolling a six sided die. Let R be the value showing on the die after rolling.

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

$$P(\omega_i) = 1/6$$

Lets see if Kolmogorov Axioms hold:

Equally Likely Outcomes

$$P(E) = \frac{\#(E)}{\#(\Omega)} = \frac{1}{\#(\Omega)} \sum_i 1_{\omega_i \in E}$$

Notation:

Cardinality - $\#(S) =$ number of elements in set S

Indicator function - $1_{x \in S} = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$

Probability of rolling an even number with a six sided die?

Equally Likely Outcomes

$$P(E) = \frac{\#(E)}{\#(\Omega)} = \frac{1}{\#(\Omega)} \sum_i 1_{\omega_i \in E}$$

Notation:

Cardinality - $\#(S)$ = number of elements in set S

Indicator function - $1_{x \in S} = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$

Probability of rolling an even number with a six sided die?

$$E = \{2, 4, 6\} \text{ } and \text{ } \Omega = \{1, 2, 3, 4, 5, 6\}$$
$$P(E) = 3/6 = 1/2$$

Useful Identities

Complement Rule:

$$P(\text{not } A) = P(A^c) = 1 - P(A)$$

Difference Rule:

$$\text{If } A \subseteq B \text{ then } P(B \text{ and } A^c) = P(B) - P(A)$$

Inclusion-Exclusion:

$$P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$$

Useful Identities (cont)

Commutativity & Associativity:

$$A \text{ or } B = B \text{ or } A$$

$$(A \text{ or } B) \text{ or } C = A \text{ or } (B \text{ or } C)$$

$$A \text{ and } B = B \text{ and } A$$

$$(A \text{ and } B) \text{ and } C = A \text{ and } (B \text{ and } C)$$

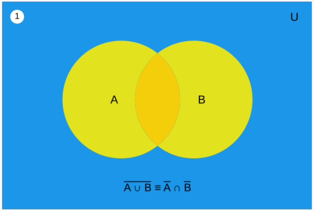
Distributivity:

$$(A \text{ or } B) \text{ and } C = (A \text{ and } C) \text{ or } (B \text{ and } C)$$

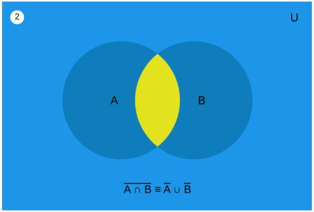
Think of union (or) as addition and intersection (and) as multiplication $(A + B) \times C = AC + BC$

DeMorgan's Rules

not (A or B) = (not A) and (not B)



not (A and B) = (not A) or (not B)



Generalized Inclusion-Exclusion

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i \leq n} P(E_i) - \sum_{i<j \leq n} P(E_i E_j) + \sum_{i<j<k \leq n} P(E_i E_j E_k) - \ldots + (-1)^{n+1} P(E_1 \ldots E_n)$$

Generalized Inclusion-Exclusion

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i \leq n} P(E_i) - \sum_{i<j \leq n} P(E_i E_j) + \sum_{i<j<k \leq n} P(E_i E_j E_k) - \ldots + (-1)^{n+1} P(E_1 \ldots E_n)$$

For the case of $n = 3$:

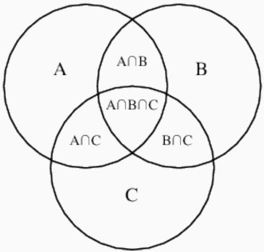
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Generalized Inclusion-Exclusion

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i \le n} P(E_i) - \sum_{i<j\le n} P(E_i E_j) + \sum_{i<j<k\le n} P(E_i E_j E_k) - \ldots + (-1)^{n+1} P(E_1 \ldots E_n)$$

For the case of $n = 3$:

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



Conditional Probability

Conditional Probability

This is the probability an event will occur when another event is known to have already occurred.

With equally likely outcomes we define the probability of A *given* B as

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)}$$

(the proportion of outcomes in B that are also in A)

Conditional Probability, cont.

We can rewrite the counting definition of conditional probability as

$$\begin{aligned} P(A|B) &= \frac{\#(A \cap B)}{\#(B)} \\ &= \frac{\#(A \cap B)/\#(\Omega)}{\#(B)/\#(\Omega)} \\ &= \frac{P(A \cap B)}{P(B)} \end{aligned}$$

which is the general definition of conditional probability.

Note that $P(A|B)$ is undefined if $P(B) = 0$.

Multiplication rule

Very often we may know the probability of events and their conditional probabilities but not probabilities of the events together, in which case we can use

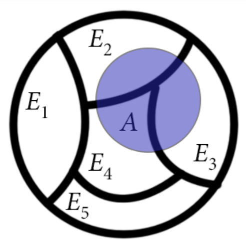
$$P(A \cap B) = P(A|B)P(B)$$

Law of total probability

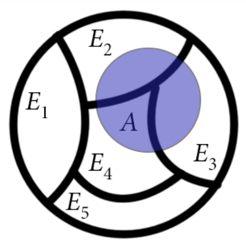
Other cases where we do not have the probability of one of the events, we can use
For a partition $B_1, \dots, B_n$ of Ω, with $B_i \cap B_j = \emptyset$ for all $i \neq j$. (In other words we need to use all possible mutually exclusive outcomes for B)

$$
\begin{aligned}
P(A) &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) \\
&= P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)
\end{aligned}
$$

Law of total probability (cont.)

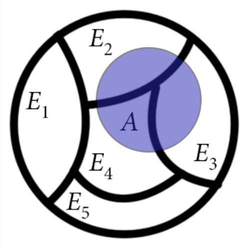


Law of total probability (cont.)



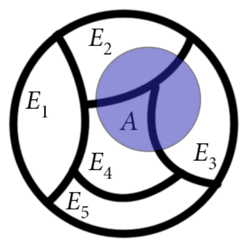
$$P(A) = P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A) + P(E_4 \cap A) + P(E_5 \cap A)$$

Law of total probability (cont.)



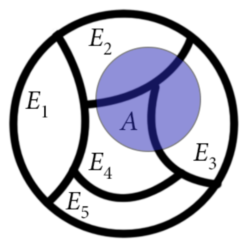
$$
\begin{aligned}
P(A) &= P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A) + P(E_4 \cap A) + P(E_5 \cap A) \\
&= \sum_{i=1}^{5} P(E_i \cap A)
\end{aligned}
$$

Law of total probability (cont.)



$$
\begin{aligned}
P(A) = & P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A) + P(E_4 \cap A) + P(E_5 \cap A) \\
= & \sum_{i=1}^{5} P(E_i \cap A) \\
= & P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + P(A|E_3)P(E_3) \\
& + P(A|E_4)P(E_4) + P(A|E_5)P(E_5)
\end{aligned}
$$

Law of total probability (cont.)



$$
\begin{aligned}
P(A) = & P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A) + P(E_4 \cap A) + P(E_5 \cap A) \\
= & \sum_{i=1}^{5} P(E_i \cap A) \\
= & P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + P(A|E_3)P(E_3) \\
& + P(A|E_4)P(E_4) + P(A|E_5)P(E_5) \\
= & \sum_{i=1}^{5} P(A|E_i)P(E_i)
\end{aligned}
$$

Example - Hiking

A quick example of the application of the rule of total probability:

Whether or not I go hiking depends on the weather, if it is sunny there is a 60% chance I will go for a hike, while there is only a 10% chance if it is raining and a 30% chance if it is snowing.

The weather forecast for tomorrow calls for 50% chance of sunshine, 40% chance of rain, and a 10% chance of snow.

What is the probability I go for a hike tomorrow?

Independence

We defined events A and B to be independent when

$$P(A \cap B) = P(A)P(B)$$

which also implies that

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

This should *not* to be confused with disjoint (mutually exclusive) events where

$$P(A \cap B) = P(\emptyset) = 0$$

Example - Eye and hair color

Table 3.3.1 Hair color and eye color

		Hair color			
		Brown	Black	Red	Total
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770

1. Are brown and black hair disjoint?
2. Are brown and black hair independent?
3. Are brown eyes and red hair disjoint?
4. Are brown eyes and red hair independent?

Bayes' Rule

Expands on the definition of conditional probability to give a relationship between $P(B|A)$ and $P(A|B)$

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

In the case where $P(A)$ is not known we can extend this using the law of total probability

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Example - House

If you've ever watched the TV show *House*, you know that Dr. House regularly states, "It's never lupus."

Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease.

The test for lupus is very accurate if the person actually has lupus, however is very inaccurate if the person does not. More specifically, the test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease.

Is Dr. House correct even if someone tests positive for Lupus?