BS1820: Maths and Statistics Foundations for Analytics

Probability 1

Zhe Liu

Imperial College Business School

Email: zhe.liu@imperial.ac.uk

Outline

1.1 Introduction to Probability

We start with an experiment where the possible **outcomes** are $\omega_1, \omega_2, \ldots$.

E.g. When we roll a die the possible outcomes are:

$$\omega_1 = 1,\ \omega_2 = 2,\ \omega_3 = 3,\ \omega_4 = 4,\ \omega_5 = 5,\ \omega_6 = 6.$$

Terminology:

- The set of all possible outcomes is the **sample space**, denoted by Ω.

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}.$$

- An **event** is a subset of the sample space Ω, e.g.,

$$A = \{\omega_2, \omega_4, \omega_6\} \subset \Omega, \quad \text{“getting an even number”}$$

- $\emptyset$ denotes the empty set which is the event consisting of no outcomes.

1.1 Introduction to Probability

Definition. A **probability** is a function defined on events that satisfies the following axioms:

1. $0 \leq \mathsf{P}(A) \leq 1$ for any event A.
2. $\mathsf{P}(\Omega) = 1$.
3. If A and B are disjoint, i.e. $A \cap B = \emptyset$, then

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B).$$

4. If $A_1, A_2, \ldots$ is an infinite sequence of pairwise disjoint events, i.e. $A_i \cap A_j = \emptyset$ whenever $i \neq j$, then

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

1.1 Introduction to Probability

Properties:

1. $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$ where A^c denotes the complement of A, i.e. the set of outcomes not in A.

2. For any events A and B we have

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

3. If $A \subset B$ then

$$\mathsf{P}(A) \leq \mathsf{P}(B).$$

1.2 Conditional Probability

Definition. Let the event A be such that $\mathsf{P}(A) > 0$. Then the **conditional probability** of an event B given A is

$$\mathsf{P}(B \mid A) := \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)}.$$

An implication is

$$\mathsf{P}(A \cap B) = \mathsf{P}(B \mid A)\mathsf{P}(A) = \mathsf{P}(A \mid B)\mathsf{P}(B).$$

Definition. Events A and B are **independent** if and only if

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B).$$

Equivalently and more intuitively, $\mathsf{P}(B \mid A) = \mathsf{P}(B)$ and $\mathsf{P}(A \mid B) = \mathsf{P}(A)$.

1.2 Conditional Probability

Let events $A_1, A_2, \ldots, A_n$ be a partition of Ω, i.e.

1. $A_i \cap A_j = \emptyset$ for $i \neq j$ (mutually exclusive)
2. $\cup_{i=1}^{n} A_i = \Omega$ (collectively exhaustive)

Let B be an event in Ω. We have

$$\mathsf{P}(B) = \sum_{j=1}^{n} \mathsf{P}(B \cap A_j). \tag{1}$$

By definition of conditional probability, we can write

$$\mathsf{P}(B \cap A_j) = \mathsf{P}(B \mid A_j)\mathsf{P}(A_j) \text{ for any } j = 1, \ldots, n. \tag{2}$$

Inserting (2) into (1) we get the **Law of Total Probability**:

$$\mathsf{P}(B) = \sum_{j=1}^{n} \mathsf{P}(B \cap A_j) = \sum_{j} \mathsf{P}(B \mid A_j)\mathsf{P}(A_j). \tag{3}$$

1.3 Bayes' Theorem

Bayes' Theorem:
Let A and B be two events for which $\mathsf{P}(B) \neq 0$. Then

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B \mid A)\mathsf{P}(A)}{\mathsf{P}(B)}.$$

Remark:
The denominator, $\mathsf{P}(B)$, can be substituted by the formula of total probability:

$$\mathsf{P}(B) = \sum_j \mathsf{P}(B \mid C_j)\mathsf{P}(C_j),$$

where C_j's are a partition of the sample space Ω.

E.g. Since A and A^c form a partition of Ω, we have

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(B \mid A)\mathsf{P}(A)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B \mid A)\mathsf{P}(A)}{\mathsf{P}(B \mid A)\mathsf{P}(A) + \mathsf{P}(B \mid A^c)\mathsf{P}(A^c)}.$$

1.3 Bayes' Theorem

Example: Mammogram Posterior Probabilities

- Approx. 1% of women aged 0 – 50 years have breast cancer.
- Woman with breast cancer has $\approx 90\%$ chance of positive mammogram test.
- Woman without breast cancer has $\approx 10\%$ chance of false-positive test result.

Question: What is the probability that a woman has breast cancer **given** that she just had a positive test result?

1.3 Bayes' Theorem

Example: Mammogram Posterior Probabilities

- Approx. 1% of women aged 0 – 50 years have breast cancer.
- Woman with breast cancer has $\approx 90\%$ chance of positive mammogram test.
- Woman without breast cancer has $\approx 10\%$ chance of false-positive test result.

Question: What is the probability that a woman has breast cancer **given** that she just had a positive test result?

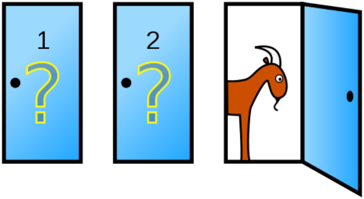
Solution: Let $A =$ "woman has breast cancer$''$ and $B =$ "positive test result$''$. We want $\mathsf{P}(A \mid B)$. From Bayes' Theorem we have

$$
\begin{aligned}
\mathsf{P}(A \mid B) &= \frac{\mathsf{P}(B \mid A)\mathsf{P}(A)}{\mathsf{P}(B \mid A)\mathsf{P}(A) + \mathsf{P}(B \mid A^c)\mathsf{P}(A^c)} \\
&= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.1 \times .99} \\
&= 8.33\%.
\end{aligned}
$$

Source: Resnick's *Elementary Probability for Applications*

1.3 Bayes' Theorem

Example: The Monty Hall Problem



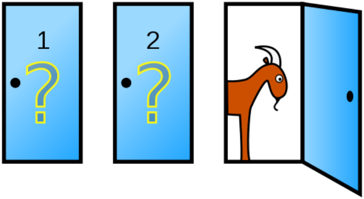
Source: Wikipedia

1. There are three closed doors:
   - behind one door lies a car; behind each of the other two doors lies a goat.
2. You don't know which door has the car, so you randomly choose one.
3. Before your chosen door is opened, Monty Hall opens a different door
   - this door always has a goat behind it, i.e. no prize
4. Monty now gives you the option to switch to another unopened door.

Question: Should you switch?

1.3 Bayes' Theorem

Example: The Monty Hall Problem



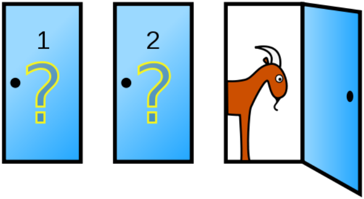
Source: Wikipedia

Approaches:

1. Discuss what's behind your chosen door.
   - Revealing a goat in an unchosen door doesn't change the initial probability.
2. Choosing one door vs choosing two doors together.
   - The $\frac{2}{3}$ chance of finding a car hasn't been changed by opening one of them.
3. Bayes' Theorem.
   - Hypothesis (H): your chosen door has a car behind it.
   - Evidence (E): Monty reveals a door with a goat behind it.

1.3 Bayes' Theorem

Example: The Monty Hall Problem



Source: Wikipedia

Approaches:

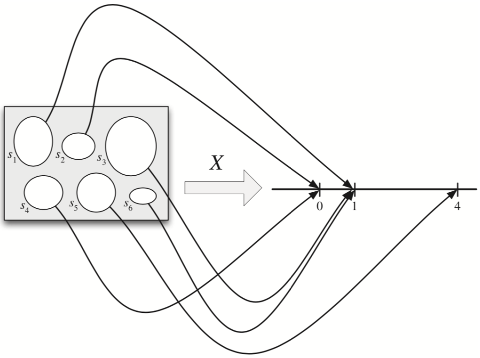
1. Discuss what's behind your chosen door.
   - Revealing a goat in an unchosen door doesn't change the initial probability.
2. Choosing one door vs choosing two doors together.
   - The $\frac{2}{3}$ chance of finding a car hasn't been changed by opening one of them.
3. Bayes' Theorem.
   - Hypothesis (H): your chosen door has a car behind it.
   - Evidence (E): Monty reveals a door with a goat behind it.

Follow-up question:

What if Monty **randomly** opens another door and it has a goat behind it?

1.4 Random Variables

Definition. A **random variable** (R.V.) maps the outcome of an experiment from the sample space to a numerical quantity, i.e., $\omega \in \Omega \mapsto X(\omega) \in \mathbb{R}$.



Source: Joseph K. Blitzstein and Jessica Hwang (2014)

Note: The randomness comes from the event while the R.V. as a function that assigns a real number to each outcome is itself deterministic.

1.4 Random Variables

E.g. $X = \#$ heads in 10 coin tosses.

E.g. Consider experiment where a die is rolled with possible outcomes

$$\omega_1 = 1,\ \omega_2 = 2,\ \omega_3 = 3,\ \omega_4 = 4,\ \omega_5 = 5,\ \omega_6 = 6.$$

Let ω denote the outcome of rolling the die, and $\Omega = \{\omega_1, \ldots, \omega_6\}$.

We can then define a random variable X by setting $X(\omega) := \omega$.

We can also define another random variable Y according to

$$Y(\omega) \quad := \quad \begin{cases} 1, & \text{if } \omega \leq 4 \\ 0, & \text{otherwise.} \end{cases}$$

Definition. The **probability** that X takes on a value in a set S is given by

$$\mathsf{P}(X \in S) \ = \ \mathsf{P}(\omega \in \Omega \mid X(\omega) \in S).$$

Exercises

Question: Consider the experiment of rolling a fair die. Let X be the random variable which assigns 1 if the number appears to be even and 0 if it is odd.

1. What are the domain and range of X?

2. Find $\mathsf{P}(X = 1)$ and $\mathsf{P}(X = 0)$.

Based on Hsu (1997).

Exercises

Question: Consider the experiment of rolling a fair die. Let X be the random variable which assigns 1 if the number appears to be even and 0 if it is odd.

1. What are the domain and range of X?

   Answer: The domain is the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the range is $\{0, 1\}$. Furthermore,

   $$X(\omega) = \begin{cases} 1 & \text{if } \omega \in \{2, 4, 6\} \\ 0 & \text{if } \omega \in \{1, 3, 5\} \end{cases}.$$

2. Find $\mathsf{P}(X = 1)$ and $\mathsf{P}(X = 0)$.

   Answer: $\mathsf{P}(X = 1) = \mathsf{P}(\omega \in \{2, 4, 6\}) = 1/2$,
   $\mathsf{P}(X = 0) = \mathsf{P}(\omega \in \{1, 3, 5\}) = 1/2$.

Based on Hsu (1997).

1.4 Random Variables

There are two major types of random variables:

1. **Disrete Random Variables:** the range contains a finite or countbly infinite sequence of values, usually representing a "counting".

   E.g. # of heads in 10 coin tosses; # of coin tosses until a head appears

2. **Continuous Random Variables:** the range is uncountably infinite, usually represents a "measurement".

   E.g. the time that passes until a head appears in repetitive coin tossing

1.5 Distribution Functions

Definition. The cumulative distribution function (CDF), $F(\cdot)$, of a random variable X, is defined by

$$F(x) := \mathsf{P}(X \leq x).$$

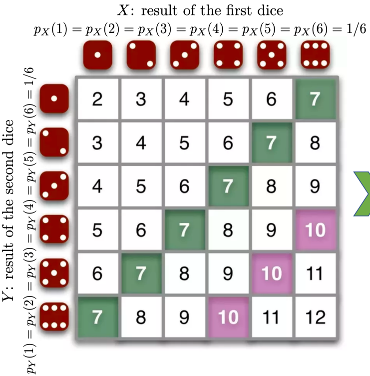
Definition. A **discrete** R.V. X has probability mass function (PMF) $p(\cdot)$ if $p(x) \geq 0$ and for all events A we have

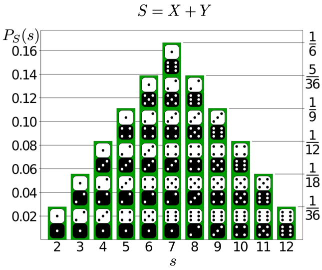
$$\mathsf{P}(X \in A) = \sum_{x \in A} p(x).$$

1.5 Distribution Functions

Example:

Consider the R.V. S that counts the sum of two (six-sided) die rolls. The PMF of S is illustrated below:





1.6 The Mean of a Random Variable

Definition. The expected value or mean of a discrete random variable, X, is given by

$$\mathsf{E}[X] := \sum_i x_i \, p(x_i).$$

E.g. Let X be the result of tossing a fair 6-sided die. Then

$$\begin{aligned}\mathsf{E}[X] &= 1 \times \mathsf{P}(X=1) + 2 \times \mathsf{P}(X=2) + \cdots + 6 \times \mathsf{P}(X=6) \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} \\ &= 3.5\end{aligned}$$

1.7 The Variance of a Random Variable

Definition. The variance of any random variable, X, is defined as

$$\begin{aligned} \mathsf{Var}(X) &:= \mathsf{E}\left[(X - \mathsf{E}[X])^2\right] \\ &= \mathsf{E}[X^2] \ - \ \mathsf{E}[X]^2. \end{aligned}$$

The variance of X is a measure of how dispersed the possible values of X are around its mean.

E.g. Returning to our die example we see

$$\begin{aligned} \mathsf{Var}(X) &= (1-3.5)^2 \times \mathsf{P}(X=1) \ + \ \cdots \ + \ (6-3.5)^2 \times \mathsf{P}(X=6) \\ &= (1-3.5)^2 \times \frac{1}{6} \ + \ \cdots \ + \ (6-3.5)^2 \times \frac{1}{6} \\ &\approx 2.92 \end{aligned}$$

The standard deviation of X is then defined as

$$\mathsf{SD}(X) \ = \ \sqrt{\mathsf{Var}(X)}.$$

1.8 SAXBY Formula

Let X, Y be any two R.V.s and a, b, c be any constant real numbers, we have

$$\mathsf{E}(aX + bY + c) = a\mathsf{E}(X) + b\mathsf{E}(Y) + c$$

$$\mathsf{Var}(aX + bY + c) = a^2\mathsf{Var}(X) + b^2\mathsf{Var}(Y) + 2ab\mathsf{Cov}(X, Y)$$

More generally, if Y_i's are independent for $i = 1, \ldots, n$ then

$$\mathsf{Var}\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_i^2 \, \mathsf{Var}(Y_i).$$