# Chapter 2. Basics of Probability

## Basic Terminology

- Sample space: Usually denoted by $\Omega$ or $S$ (textbook). Collection of all possible outcomes, and each outcome corresponds to one and only one element in the sample space (i.e., sample point).
- Event: Any subset of $\Omega$. What's the meaning of an event *happens*?

Examples:

1. Toss coin twice. Event: One Heads and one Tails.

2. Draw two cards from a well-shuffled deck of 52 cards. Event: Black Jack.

3. Tossing a fair coin until a Heads appears. Sample points have different probabilities. Event: (a) first Heads on the 3rd toss; (b) no heads in the first two tosses.

4. Choose a point from interval $[0, 1]$. Event: the point chosen is smaller than 0.5.

5. Stock price after one month from today. (Two different ways, discrete or continuous). Event: the stock price is at least 20 dollars.

- Operations on events: $A \subseteq B$, $A \cap B$, $A \cup B$, and $\bar{A}$ (complement). Two events $A$ and $B$ are mutually exclusive (or disjoint) if $A \cap B = \emptyset$.

EXAMPLES: Express the following events in terms of $A$, $B$, and $C$.

1. $A$ happens but not $B$.
2. None of $A$ or $B$ happens.
3. Exactly two of $A$, $B$, and $C$ happen.
4. At most two of $A$, $B$, and $C$ happen.

- Distributive laws:

  $$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

- DeMorgan's laws:

$$\overline{A \cap B} = \bar{A} \cup \bar{B}, \quad \overline{A \cup B} = \bar{A} \cap \bar{B}.$$

# Probability Axioms

Let $P(A)$ denote the probability of event $A$. Then

1. $0 \leq P(A) \leq 1$ for every event $A \subset \Omega$.
2. $P(\Omega) = 1$.
3. If $A_1$, $A_2$,..., is a sequence of mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$

## A few elementary consequences from axioms:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(\bar{A}) = 1 - P(A)$. In particular, $P(\emptyset) = 0$.

## Examples

1. Consider a finite sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. A uniform distribution $\Omega$ is such that
$$P(\omega_1) = P(\omega_2) = \cdots = P(\omega_n).$$
What is $P(A)$ for any $A \subset \Omega$?

2. Toss a coin with $P(H) = p$ until first heads. What is the probability that the first head appears within three tosses? Within $n$ tosses?

3. The freshmen class has 400 students, 250 are women, 60 are majoring in art, and 40 art majors are women. If a student is randomly selected, what is the probability that the student will be either an art major or a woman?

4. A continuous sample space example. Picking a point randomly from interval $[0, 1]$. Identify the sample space and the probability of any event.

Remark: It is usually the case that in the calculation, the sample space will NOT be explicitly identified. But one should keep in mind that all the calculations are based on the underlying sample space.

# Counting method

Requirement: The sample space $\Omega$ has finitely many sample points, and each sample point is equally likely.

$$P(A) = \frac{\text{number of sample points in } A}{\text{total number of sample points in } \Omega}$$

- *Factorial:* $n! \doteq n \times (n-1) \times \cdots \times 1$, and $0! \doteq 1$.
- *Binomial coefficients:* For $0 \leq k \leq n$,

$$\binom{n}{k} \doteq \frac{n!}{k!(n-k)!}.$$

- *Multinomial coefficients:* For $n_1 + n_2 + \cdots + n_k = n$ with $n_i \geq 0$,

$$\binom{n}{n_1\ n_2\ \cdots\ n_k} \doteq \frac{n!}{n_1!n_2!\cdots n_k!}.$$

# Examples

1. Suppose a license plate consists of two letters followed by four integers from 0 to 9. What is the total number of possible license plates? What is the probability that a license plate starts with the letter “R”?

2. Five balls are to be put into 6 boxes. What is the probability that none of the boxes contain two or more balls?

3. A class of $k$ students. Probability that at least two students share the same birthday.

4. For a randomly dealt five-card poker hand, what is the probability that the hand will be four of a kind? a flush (5 cards of same suit)?

5. 30 students are to be randomly divided into 3 groups, each group consisting of 10 students. What is the probability that three good friends, Joe, Jane, and Jack, will each be assigned to a different group?

# Conditional Probability and Independence

Conditional probability: Given that event $B$ happens, what is the probability that $A$ happens?

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

1. A graphical explanation of conditional probability.
2. Rewrite the definition:

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

3. Is conditional "probability" really a probability? (Verify the axioms)

Example: The distribution of 1237 college students by sex and major is given below.

| Sex | Arts and Sciences | Business | Music | Total |
|---|---|---|---|---|
| Male | 127 | 383 | 40 | 550 |
| Female | 380 | 242 | 65 | 687 |
| Total | 507 | 625 | 105 | 1237 |

Find (a) the probability a randomly selected student is Music major; (b) given that the randomly selected student is Music major, what is the probability that it is a female?

Example: Consider the Polya's urn model. An urn contain 2 red balls and 1 green balls. Every time one ball is randomly drawn from the urn and it is returned to the urn together with another ball with the same color. (a) The probability that the first draw is a red ball? (b) The probability that the second draw is a red ball?

Let $\{B_1, B_2, \ldots, B_n\}$ be a partition of the sample space $\Omega$.

1. Law of total probability:

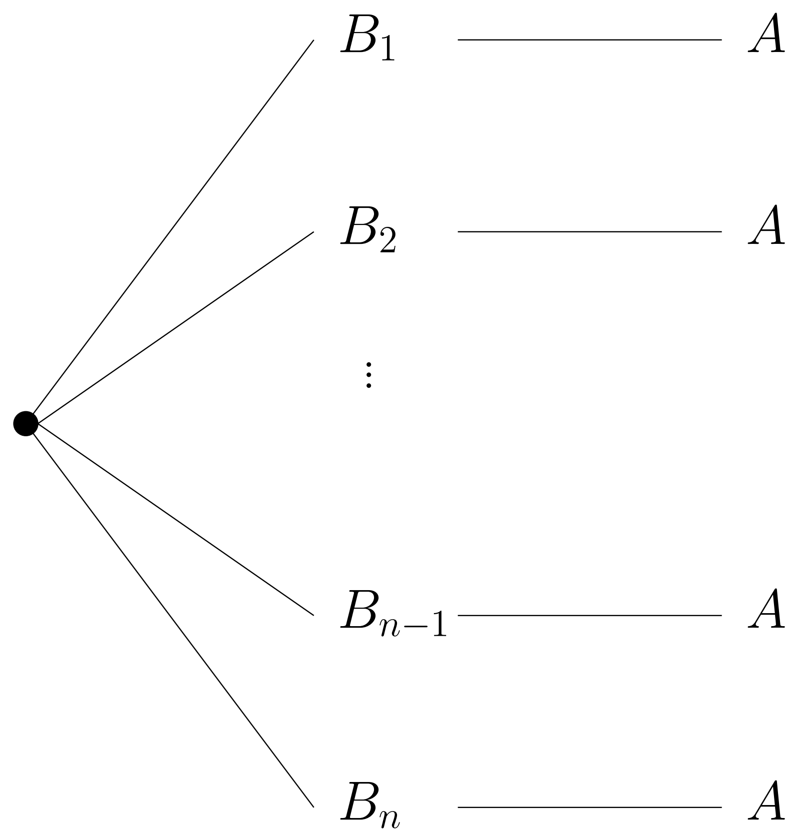
$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

2. Bayes' rule:

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

# Tree method

Law of total probability and Bayes' rule are easily manifested by tree method. Recall the lie-detector example.



# Example

1. [Redo the lie detector example]. A collection of suspects, 90% are honest men, 10% are thieves. An 80% accurate lie detector – an honest man will pass the test with probability 80%, a thief will fail with probability 80%. Randomly pick a suspect and put him through the test. What is the probability that the suspect is a thief if he fails the lie detector test?

2. Three identical cards, one red on both sides, one black on both sides, and the third red on one side and black on the flip side. A card is randomly drawn and tossed on the table.

   (a) Probability that the up face is red?

   (b) Given the up face is red, probability that the down face is also red?

Independence:

1. Two events $A$ and $B$ are independent if $P(A \cap B) = P(A)P(B)$.
2. $n$ events $A_1, \ldots, A_n$ are (mutually) independent if

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) && i < j \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) && i < j < k \\ &\vdots \\ P(A_1 \cap A_2 \cdots \cap A_n) &= P(A_1)P(A_2) \cdots P(A_n). \end{aligned}$$

Caution: A collection of events $A_1, \ldots, A_n$ can be pairwise independent yet fail to be (mutually) independent. Example: Four cards marked $aaa$, $abb$, $bab$, $bba$. Random draw a card, and let

$$\begin{aligned} A &= \{\text{First letter on card is } a\} \\ B &= \{\text{Second letter on card is } a\} \\ C &= \{\text{Third letter on card is } a\}. \end{aligned}$$

$A$, $B$, and $C$ are pairwise independent, but not (mutually) independent!

# Example

1. Serial system and parallel system.

2. The distinction between independence and mutually exclusiveness.

3. Family with two children and each child is 50-50 to be boy or girl. A randomly selected family. Events

$$\begin{aligned} A &= \{\text{the family has children of both sex}\} \\ B &= \{\text{the family has at most one girl}\} \end{aligned}$$

Are event $A$ and $B$ independent?

Repeat the same question for a family with 3 children.

## Random Variables

Random variable is a variable whose value is a numerical outcome of a random phenomenon.

Random Variable: A random variable $X$ is a function $X : \Omega \longrightarrow \mathbb{R}$. In other words, for every sample point (i.e., possible outcome) $\omega \in \Omega$, its associated numerical value is $X(\omega)$.

Notation: For any subset $I \subset \mathbb{R}$, the event $\{X \in I\} \doteq \{\omega : X(\omega) \in I\}$.

Do not be afraid to use more than one random variables (random vector)!

Why random variables?

## Examples: Discrete random variable

Discrete random variable: A random variable that can only take finitely many or countably many possible values.

How to describe a discrete random variable $X$: Let $\{x_1, x_2, \ldots\}$ be the possible values of $X$. Let

$$P(X = x_i) = p_i.$$

We should have $p_i \geq 0$ and

$$\sum_i p_i = 1.$$

$\{p_i\}$ is called a probability function.

1. Toss a fair coin. Win \$1 if heads, and lose \$1 if tails. $X$ is the total winning.
   (a) Toss coin once.
   (b) Toss coin twice.
   (c) Toss coin $n$ times.

2. Toss a fair coin. $X$ is the first time a heads appears.

3. Randomly select a college student. His/her IQ and SAT score.

## Examples: Continuous random variable

Continuous random variable: A random variable that can take any value on an interval on $\mathbb{R}$.

1. The decimal part of a number randomly chosen.
2. Stock price, waiting times, height, weight, etc.

How to describe a continuous random variable $X$: Use a non-negative density function $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ such that

$$P(X \in I) = \int_I f(x)dx$$

for every subset $I \subset \mathbb{R}$. It is necessary that $f \geq 0$ and

$$\int_{\mathbb{R}} f(x)\, dx = 1.$$

## A general description of a random variable

Cumulative distribution function (cdf): For a random variable $X$, its cdf $F$ is defined as

$$F(x) \doteq P(X \leq x)$$

for every $x \in \mathbb{R}$.

1. $F$ is non-decreasing. $F(-\infty) = 0$, $F(\infty) = 1$.
2. For a discrete random variable, $F$ is a step function, with jumps at $x_i$ and jump sizes $p_i$.
3. For a continuous random variable, $f(x) = F'(x)$.

Remark: From cdf one can calculate $P(X \in I)$ for all subsets $I$.

## A few random comments

- All the descriptions for discrete or continuous random variables transfer to random vectors in an obvious fashion.
- It is sometimes convenient to describe a discrete random variable in the continuous fashion. For example, IQ or SAT scores.
- The specification of the cdf, or the density function, or the probability function, completely determines the statistical behavior (distribution) of the random variable.
- The advantage of random variables over sample points. It is usually too big a task to explicitly identify the sample space $\Omega$. Fortunately, such an identification is not necessary, since one is often interested in some numerical characteristics of the system behavior instead of individual sample points. Example of Nasdaq Index and interest rate.
- There are other types of random variables (mixed distribution).