Chapter 2. Basics of Probability

Basic Terminology

- Sample space: Usually denoted by Ω 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 Ω.

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 coin until a Heads appears. Sample points have different probabilities.
4. Choose a point from interval or a square.
5. Stock price after one month from today.

- 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. $0 \le P(A) \le 1$ 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. Revisit some of the old examples.
2. Choose a point "uniformly" from an interval. Discuss the meaning of zero-probability events.
3. The base and the altitude of a right triangle are chosen uniformly from $[0, 1]$. What's the probability that the area of the triangle so formed will be less than $1/4$.

Counting method

Requirement: The sample space Ω 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. An elevator with $k = 5$ passengers and stops at $n = 6$ floors. The probability that no two passengers alight at the same floor.

2.(a) A class of k students. Probability that at least two students share the same birthday.

(b) A class of k students, including you. Probability that at least one student has the same birthday as yours.

3. Probability that four bridge players each holds an Ace (about 10%).

4. A professor sign n letters of recommendation for one of his student and put them randomly into n pre-addressed envelopes. Probability that none of the letters is put in its correct envelope.

Solution: Ω is the sample space consisting of all possible outcomes ($n!$ in total), each outcome with equal probability $1/n!$. Let A_i be the event that i-th letter is put into the correct envelope. The event of interest is

$$E = \text{the complement of } \ \cup_{i=1}^{n} A_i.$$

Note the formula

$$P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots \pm P(A_1 \cap A_2 \cap \cdots \cap A_n),$$

and

$$\sum_{i_1<i_2<\cdots<i_k} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \sum_{i_1<\cdots<i_k} \frac{(n-k)!}{n!} = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}$$

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. Recall the example of lie-detector ...
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: 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?

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

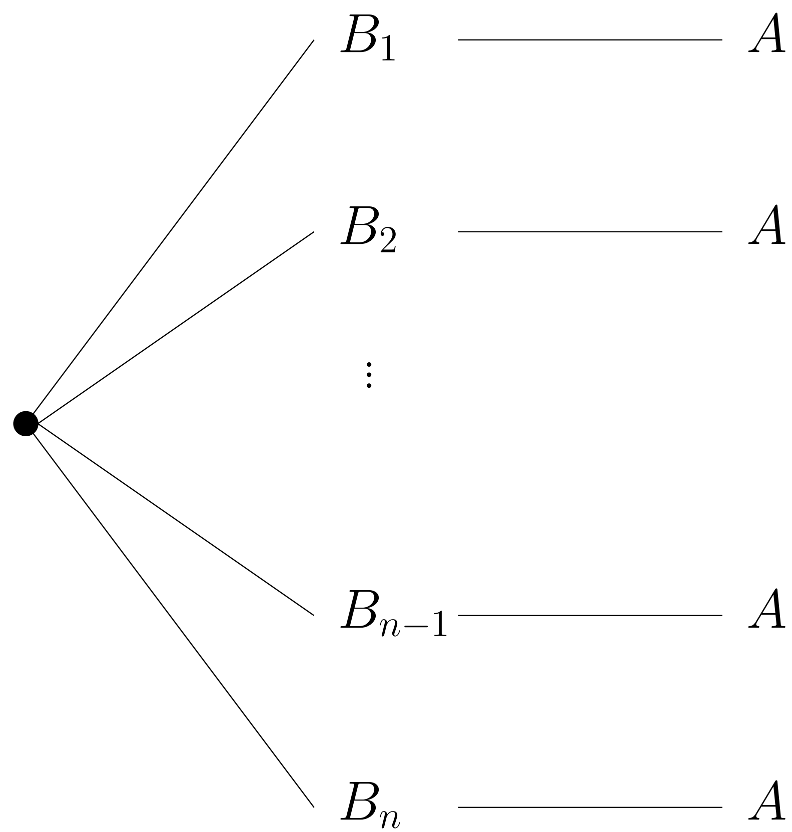
$$\begin{aligned} A &= \{\text{First letter on card is } a\} \\ B &= \{\text{Second letter on card is } a\} \\ C &= \{\text{Third letter on card is } a\}. \end{aligned}$$

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

Examples: We have used the idea of independence intuitively many times in coin tossing. Examples of serial system and parallel system.

Tree method

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



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

1. Law of total probability:

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

2. Bayes' rule:

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

Example

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.

1. Probability that the up face is red?
2. Given the up face is red, probability that the down face is also red?

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)!

Examples: Discrete random variable

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

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.

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.

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

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

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, $f(x) = F'(x)$.

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 Ω. 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).

A Challenging Example: Buffon's Needle

A table of infinite expanse has inscribed on it a set of parallel lines spaced one unit apart. A needle of length ℓ ($\ell < 1$) is dropped on the table at random. What is the chance that the needle crosses a line?

Solution: (1) Define some convenient random variables for this problem: $X =$ angle, $Y =$ position.

(2) Density for (X, Y). It is uniform.

(3) Draw out the region for which a crossing will happen.

(4) Compute the probability of (X, Y) falls into that region.

(5) Answer: $2\ell/\pi$.