ST521: Chapter 1

1 Overview Of Probability

Probability theory is a branch of mathematics that deals with uncertainty.

- A *random experiment* is an experiment for which the outcome can not be predicted with certainty.
- Probability = chance or the likelihood that some outcomes will happen. Common phrases used to describe the probability are "very likely", "almost impossible", "it is to be expected", etc
- The probability of A (a collection of possible outcomes) is the fraction of times that the outcome of the random experiment results in A in a large number of trials of the random experiment.
- Key people: Fermat, Pascal, Bernoulli, Laplace, Kolmogorov ...

Probability plays an essential role in decision making in day-to-day activities

- The insurance companies determine the premiums for customers; the manufacturing companies determine the warranty period of its product; the supermarkets decide on the number of counters to open ...
- DNA matching is important in a judicial process, especially for crime related activities
- Statistics builds upon the foundation of probability theory.

Question: How do we define probability?
(1) *Classical* approach
(2) *Frequency* approach
(3) *Subjective* approach
(4) Formal definition: axiomatic approach (principles accepted without proof).

(1) *Classical* approach supposes an experiment has a finite number of "equally likely" outcomes.

Let's take the simplest possible model. Coin — toss. Head or tail. Then what is the probability of head?

- It is intuitively clear that, if there is no essential difference between the two sides (fair coin), then the chance of head should be 1 in 2, that is 0.5.

- Similarly, if you roll a fair die, the chance that 5 or 6 comes up is 2/6.

The probability is calculated as

$$\frac{\text{the number of outcomes in the event}}{\text{total number of outcomes}}.$$

(2) *Probability is a long run relative frequency.*
Classical definition is not appropriate for every situation. For instance, if we want to know the chance of air crash, then 1/2 is clearly a false value of the real chance.

- P(air crash): empirical data. Proportion. Relative frequency.

One proposes probability based on experience, using the frequency interpretation of probability.

(3) Also, *chance could be subjective.*

- P(tomorrow stock index will rise): a subjective opinion.
- An experience physician may say “This patient has a 30% chance of recovery with a certain treatment”

This is a belief in the chance of an event occurring.

(4) To avoid conceptual difficulties, like in many branches of mathematics, probability is defined axiomatically (some principles accepted without proof). The formal definition of probability is based on tools from *set theory.*

2 Set Theory

Definitions:

- A *set* is a collection of well defined objects, called *elements*.
- If A is a set and x is an element of A, we say x *belongs to* A, and write $x \in A$.
- In a random experiment, one of several possible results will occur each time. These possible results are called *outcomes*.
- The collection of all possible outcomes is called the *sample space*. Denote it by $\mathcal{S}$.
- A set of some well defined outcomes is called an *event*.
- Any event is a subset of the sample space. We talk about the probabilities of the events.

EXAMPLES OF SAMPLE SPACE:

EX.: Toss a coin: $\mathcal{S} = \{H, T\}$

EX., Roll a die: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$

EX., Draw a card: 52 possibilities corresponding to every combination of one of the four suits (spade, heart, diamond, club) and one of the 13 denomination (2–10, J, Q, K, A).

EX., Observe how many calls received by a telephone:

EX., Toss a coin until you get a head:

EXAMPLES OF EVENTS:

EX., Head in a single throw of coin: {H}.

EX., In the die-rolling example, the possible events “an odd number”, “an even number”, “A number less than 3”.

Important: {H} and H are not the same. Why?

Two special events: $\emptyset = \{\}$ and $\mathcal{S}$.

Definition:

countable set: A set containing either finite elements or the same number of elements as natural numbers.

EXAMPLES: All odd numbers; The number of phone calls expected.

Basic operations on sets:

Subset/superset: $A \subset B$ if $x \in A \implies x \in B$
Equality: $A = B$ if they have exactly the same elements.

$$A = B \text{ iff both } A \subset B \text{ and } B \subset A.$$

Union: Forming a larger collection. $A \cup B$ consists of all elements which are in A or in B or in both A and B. We say "*either A or B*".

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

Intersection: Collecting the common elements. $A \cap B$ consists of all elements which are common to both A and B. We say "*both A and B*".

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

Complement: Opposite event. A^c consists of those elements which are not members of A: We say "*not A*".

$$A^c = \{x : x \notin A\}$$

Difference: $A \backslash B$, or $A - B$, consists of those elements which are in A but not in B. We say "A but not B".

$$A \backslash B = \{x : x \in A, x \notin B\}$$

Empty set: Contains no element! Denoted by $\emptyset$.

Disjoint or (*Mutually exclusive*): We say A and B *disjoint* if $A \cap B = \phi$.

Pairwise Disjoint or (or *Mutually exclusive*):
The events $A_1, A_2, \cdots, A_n$ are *pairwise disjoint* (or *mutually exclusive*) if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Some Facts: Assume $A, B \subset \mathcal{S}$.

- $(A \cap B) \subset A, (A \cap B) \subset B, A \subset (A \cup B), B \subset (A \cup B)$
- If $A \subset B$, then $A \cap B = A$ and $A \cup B = B$.
- $(A \backslash B) \subset A$ and $A - B = A - (A \cap B)$.

EXAMPLE. Select a card at random from a standard deck. The sample space $\mathcal{S}$={clubs (C), diamonds (D), hearts (H), spades (S)}. Some of the possible events are $A = \{C, D\}$ and $B = \{D, H, S\}$. Compute $A \cap B$, $A \cup B$ and A^c.

Theorem. For any 3 sets A, B, C, we have —

1. *Commutativity*: $A \cup B = B \cup A$, $A \cap B = B \cap A$.
2. *Associativity*:
$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$$
3. *Distributivity Laws*:
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
4. *De Morgan's laws*:
$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

We can define union and intersection of an infinite collection.
Let A_i, $i = 1, 2, \ldots$ be a collection of sets, all defined on a sample space $\mathcal{S}$,

- $\cup_{i=1}^{\infty} A_i$ is the grand collection of elements belonging to at least one A_i.
$$\cup_{i=1}^{\infty} A_i = \{x \in \mathcal{S} : x \in A_i, \text{ for some } i\}$$
Can be interpreted as "at least one of A's occurs".
- $\cap_{i=1}^{\infty} A_i$ is the collection of all elements common to all the A_i's.
$$\cap_{i=1}^{\infty} A_i = \{x \in \mathcal{S} : x \in A_i, \text{ for all } i\}$$
Can be interpreted as "all of A's must occur at the same time".

Axiomatic Foundation of Probability Theory.

For each event A in the sample space $\mathcal{S}$, the *probability* is a function which associates with A a number between zero and one. i.e. $P(A) \in [0,1]$.

Basic Principles of Probability:

(1) Probability is a measure of strength, like length, area, volume, taking the value between 0 and 1. The impossible event should have probability 0. The certain event (sample space) should have probability 1.

(2) Probability should be *additive* whenever two events are disjoint. This additivity should not only be true for two (or finitely many) events, but also for a whole sequence (countably many) of disjoint events.

(3) It is technically hard to define probability for each single event. In general, consider a good collection $\mathcal{B}$ of events which is large enough to contain all the useful events including $\emptyset$ and $\mathcal{S}$, and is closed under all possible countable set operations. This collection is called a *sigma algebra.* Probability is a set function defined only on this collection, i.e,

$$P : \mathcal{B} \longrightarrow [0,1]$$

Definition:

Let $\mathcal{B}$ be a collection of subsets of a sample space $\mathcal{S}$. $\mathcal{B}$ is called a σ-algebra or σ-field if and only if it satisfies the following three properties:

(i). $\emptyset \in \mathcal{B}$.

(ii). If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$. ($\mathcal{B}$ is closed under complementation)

(iii). If $A_1, A_2, \ldots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ (closed under countable Unions)

Note: (ii) and (iii) imply: if $A_1, A_2, \ldots \in \mathcal{B}$, then $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$

Ex. Tossing a fair die. The sample space $\mathcal{S} = \{1,2,3,4,5,6\}$.

- One possible σ-algebra is $\mathcal{B}_1 = \{\emptyset, \mathcal{S}\}$.
- Another σ-algebra is *power set*, the collection of all possible sets of $\mathcal{S}$.
- Another possible σ-algebra is $\mathcal{B} = \{\emptyset, \mathcal{S}, \{1\}, \{2,3,4,5,6\}\}$.

Ex. Let $\mathcal{S} = (-\infty, \infty)$ (real line). One special sigma-algebra $\mathcal{B}$ contains all sets of the form $(a,b), [a,b), (a,b], [a,b]$ and their countably unions and intersections. This σ-algebra is called *Borel σ-field.*

Definition: (Kolmogorov Axioms)

Given a sample space $\mathcal{S}$ and an associated σ-algebra $\mathcal{B}$, a set function P defined on $\mathcal{B}$ is called a *probability* if and only if it satisfies:

(i) $P(A) \geq 0$ for all $A \in \mathcal{B}$. (nonnegativity)

(ii) Total probability: $P(S) = 1$, where S is the sample space.

(iii) If $A_1, A_2, \ldots$ are mutually exclusive events, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. (countable additivity)

Note: If $\mathcal{S}$ is finite, then (iii) only requires the union of finite events. In measure theory, the pair $(\mathcal{S}, \mathcal{B})$ is called a *measurable space*, and the elements of $\mathcal{B}$ are called *measurable sets*. The triple $(\mathcal{S}, \mathcal{B}, P)$ is called a *probability space.*

Ex 1. If we toss a coin, $\mathcal{S} = \{H, T\}$. $\mathcal{B} = \{\emptyset, \{H\}, \{T\}, \mathcal{S}\}$.

Ex 2. *Classical definition.* The finite sample space $S = \{1, \ldots, N\}$. $\mathcal{B}$ is the power set. Show $P(A) = \#A/N$ is a valid probability function.

Ex 3. (Theorem 1.2.6 in textbook) More generally, let $\mathcal{S} = (1, \ldots, N)$ and $\mathcal{B}$ be any sigma algebra of subsets of $\mathcal{S}$. Let $p_1, \ldots, p_N$ be nonnegative numbers which add up to 1. For any $A \in \mathcal{B}$, define $P(A)$ by

$$P(A) = \sum_{i \in A} p_i.$$

Then P is a probability function on $\mathcal{B}$. This remains true if $N = \infty$.

Ex 4 *Geometrical probability.* Events are described by area on a plane. $P(A)$ = area of A/the total area. For instance, this could be the model when hitting a dart board. As area is additive, so will be P.

(In three dimension, area will be replaced by volume. Gives a sense of uniform distribution of probability).

3 Calculus of Probabilities

Theorem 1.2.8

(i) $P(\emptyset) = 0$.
(ii) $P(A) \leq 1$.
(iii) $P(A^c) = 1 - P(A)$.

Proof:

(i) $1 = P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$.

(ii) and (iii). $\{A, A^c\}$ is a disjoint collection. So $1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$. This gives (iii), and since $P(A^c) \geq 0$, also (ii).

Theorem 1.2.9. For any A, B,

(i) $P(B \cap A^c) = P(B) - P(A \cap B)$
(ii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(iii) $A \subset B$ implies $P(A) \leq P(B)$.

Proof:

(i) $B = (A \cap B) \cup (A^c \cap B)$ and these two sets are disjoint. Thus by additivity of P, (i) follows.

(ii) $A \cup B = A \cup (A^c \cap B)$. Apply additivity, $P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B)$ by (i).

(iii) If $A \subset B$, then $A \cap B = A$. So $0 \leq P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)$.

Probability of Event Unions and Intersections

Inequalities: (find bounds)

1.*Boole's inequality*: $P(A \cup B) \leq P(A) + P(B)$.
2.*Boole's inequality for n events*: $P(A_1 \cup A_2 \cdots \cup A_n) \leq \sum_{i=1}^n P(A_i)$.
3.*Bonferroni's inequality*: $P(A \cap B) \geq P(A) + P(B) - 1$.
4. *Bonferroni's inequality for n events*:

$$P(A_1 \cap \cdots \cap A_n) \geq P(A_1) + \cdots + P(A_n) - (n - 1).$$

Equations:

5. *Union of three events*:

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

6. *Union of n events: inclusion-exclusion principle.* (*Jordan Formula*)

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P_1 - P_2 + P_3 - \cdots + (-1)^{n-1} P_n,$$

where

$$P_1 = \sum_{1 \le i \le n} P(A_i), P_2 = \sum_{1 \le i < j \le n} P(A_i \cap A_j),$$

$$P_3 = \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k), \cdots, P_n = P(A_1 \cap A_2 \cap \cdots \cap A_n).$$

PROOF:

1. Boole follows from (ii) of the last theorem as $P(A \cap B) \ge 0$.

2. Proved either by disjointification or by induction. Define the disjoint sets $A_1^* = A_1, A_i^* = A_i \setminus (\cup_{i=1}^{i-1} A_i)$ for $i = 2, \cdots, n$. Then $\cup_{i=1}^n A_i^* = \cup_{i=1}^n A_i$ and $A_i^* \subset A_i$ for all i. Then $P(\cup_{i=1}^n A_i) = P(\cup_{i=1}^n A_i^*) = \sum_{i=1}^n P(A_i^*) \le \sum_{i=1}^n P(A_i)$.

3. Bonferroni's follows from (ii) of the last theorem as $P(A \cup B) \le 1$.

4. Boole's inequality easily carries over to any finite number of events by successive applications. By de Morgan's and Boole's inequality,

$$\begin{aligned} & P(A_1 \cap \cdots \cap A_n) = P((A_1^c \cup \cdots \cup A_n^c)^c) \\ = \; & 1 - P(A_1^c \cup \cdots \cup A_n^c) \ge 1 - \sum_{i=1}^n P(A_i^c) = 1 - \sum_{i=1}^n (1 - P(A_i)) = \sum_{i=1}^n P(A_i) - n + 1. \end{aligned}$$

5. Notice that

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

Here

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

and

$$\begin{aligned} P(A \cap (B \cup C)) &= P((A \cap B) \cup (A \cup C)) = P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C). \end{aligned}$$

Thus,

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

6. This can be proved by induction, using the same idea as above. (We shall use this in complicated probability calculations such as the envelope matching problem discussed shortly).

Ex.: If A and B are two events and each has probability 0.95. Then the probability of both occurring is at least 0.90.

Definition: We say $A_1, \cdots, A_n$ are *mutually exclusive* (or *pairwise disjoint*) if any pair of them are disjoint, i.e., if there is no outcome in common between any pair.

Ex: The collection $A_i = [i, i+1), \quad i = 0, 1, \ldots$ are pairwise disjoint.

Definition: If $C_1, C_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} C_i = \mathcal{S}$, then the collection $C_1, C_2, \ldots$ forms a *partition* of $\mathcal{S}$.

Ex. Let $S = [1, \infty]$. Define $C_i = [i, i+1)$ for $i = 1, 2, \cdots$. They form a partition.

Remark: Any collection of events $A_1, A_2, \ldots$ can be disjointified as

$A_1^* = A_1, \quad A_2^* = A_2 \cap A_1^c = A_2 \setminus A_1, \quad A_i^* = A_i \setminus (\cup_{j=1}^{i-1} A_j), i = 3, \cdots, n.$
Then

(i) A_i^*'s are disjoint

(ii) $A_i^* \subset A_i$ for each i.

(iii) $\cup_{i=1}^{m} A_i = \cup_{i=1}^{m} A_i^*$ for all $m = 1, 2, \cdots, n$,

All these are also true if $n = \infty$.

Theorem 1.2.11. (Continuity of Probability)

(i) $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \ldots$.

(ii) If $A_1 \subset A_2 \subset \cdots$, then $P(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} P(A_n)$. (upper continuity). Sometimes we write $\cup_{i=1}^{\infty} A_n = \lim_{n\to\infty} A_n$).

(iii) If $A_1 \supset A_2 \supset \cdots$, then $P(\cap_{i=1}^{\infty} A_i) = \lim_{n\to\infty} P(A_n)$. (lower continuity). Sometime we write $\cap_{i=1}^{\infty} A_n = \lim_{n\to\infty} A_n$).

(iv) $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$. (Generalized Boole's inequality).

Proof.

(i) $A = A \cap \mathcal{S} = A \cap (\cup_{i=1}^{\infty} C_i) = (A \cap C_1) \cup (A \cap C_2) \cap \cdots$, disjoint break up. Apply countable additivity of P.

(ii) Consider $B_1 = A_1$, $B_2 = A_2 \cap A_1^c$, $B_3 = A_3 \cap A_2^c$, ... This technique is called disjointification. Then clearly $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i$ and B_i's are disjoint, so $P(\cup_{i=1}^{\infty} A_i) = P(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) =$

$\lim_{n\to\infty} P(\cup_{i=1}^{n} B_i)$. Now for any n, $\cup_{i=1}^{n} B_i = \cup_{i=1}^{n} A_i = A_n$, so the result follows.

(iii). Apply (ii) with A_i replaced by A_i^c and apply de Morgan's rule.

(iv) Consider $B_n = \cup_{i=1}^{n} A_i$. Then $B_n \uparrow \cup_{i=1}^{\infty} A_i$ so that by (ii), $P(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i)$.

4 Counting: Combinatorics

Counting problems can be complicated. The trick is to break them into a series of simple tasks that are easy to count.

(1) **Multiplication Rule**

To count the number of possible pairs, if there are n_1 choices of the first component and for every choice of the first component, there are n_2 choices of the second, then there are $n_1 \times n_2$ possible pairs altogether.

This also generalizes to k-tuples. The number will be $n_1 \times \cdots \times n_k$.

Ex. Suppose there are n items on the menu of a restaurant and k people go there. Each person orders exactly one item. What is the total number of possible orders.

Each person has n choices, irrespective of others. So the total number is k fold product n^k.

Each person is like a ball, which can go to any of the n cells (choose an item). The number of possibilities is cell to the power ball.

(2) **Addition Rule (Sum Rule)**

Assume there are K tasks. There are n_1 ways to finish Task 1 and n_2 ways to finish Task 2. If these tasks can be done at the same time, then there are $n = n_1 + n_2$ ways to finish one of these tasks.

Ex. A person sends a package from city A to city B. There are three ways of delivery via ground and two ways of delivery by air. The totally number of ways for the package delivery is $3 + 2 = 5$.

Permutation and Combination

Depending on whether "order" is taken into account or not in counting, we have the notation of *Permutation* and *Combination.*

Permutation: Permutation means arrangement of things when the order matters. Formally, a permutation is an ordered sequence of r elements taken from a set of n distinct objects, without repetitions. We require $r \le n$.

- For example, given the set of letters C, E, G, I, N, R. Some permutations of four letters are RING, RICE, RCIE, RECI.
- Another example. Suppose there are r numbered seats and $n(\ge r)$ people to occupy these seats. In how many ways these seats can be filled up? (each seat only allows for one person to sit on)
  First seat: n ways, second: $n-1$ ways, ..., rth seat: $n-r+1$ ways. The total number $n(n-1)\cdots(n-r+1)$.
- In general, the number of permutations is denoted by $(n)_r$, defined as

  $$(n)_r = n(n-1)\cdots(n-r+1),$$

  where n is the number of elements available for selection, and r is the number of elements to be selected ($0 \le r \le n$). This read as n permutation r.

Example of Permutations:

(1) In the seating example above, what is the answer if there are exactly n seats?

Notation: n factorial — $n! = 1 \times 2 \times \cdots \times n$. Convention: $0! = 1$.

(2) Suppose that there are 10 eligible faculty members in a department and we need to choose a head and an assistant head.

The number of ways this can be done is $(10)_2 = 10 \times 9 = 90$.

(3) Three problems, to be solved by 35 students on the chalk board. The total number of ways the teacher can ask this is $35 \times 35 \times 35$, if a student can be asked to do more than once. If that is not practiced, the number is $(35)_3 = 35 \times 34 \times 33$.

Combination: Combination means selection of things when the order of things has no importance. Formally, a combination is an un-ordered collection of r elements selected from a set of n distinct objects (without repetition).

- For example, given the set of letters C, E, G, I, N, R, the RICE, RCIE, RECI refer to one same combination.
- Another example. There are 20 people and a committee of 4 is to be formed. How many was can this be done?

  If we keep the ordering, the number is 20!/16!. Now among the selected 4, the 4! permutations among themselves do not change anything. That is, all these 4! possibilities lead to the same committee. Thus the actual number of different possible committees is 20!/16! divided by 4!, that is 20!/(4!16!).
- More generally, if r people have to be chosen from n disregarding the order, then total number of different possibilities is

  $$\binom{n}{r} = n!/(r!(n-r)!), \qquad \text{(read as } n \text{ choose } r).$$
- Note: $\binom{n}{r} = \binom{n}{n-r}$, $\binom{n}{n} = 1$.
- Prove that $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$. In particular, $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.
- For $n \geq 2$, the following holds

  $$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0, \quad \sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}, \quad \sum_{k=1}^{n} (-1)^{k+1} k \binom{n}{k} = 0.$$

EXAMPLE OF COMBINATIONS:

(1) There are 10 men and 12 women, and we want to form a committee of 5, so that there should be exactly 2 men and 3 women in the committee. What is the number of ways this can be formed?

Choose 2 men from 10 in $\binom{10}{2}$ ways and 3 women from 12 in $\binom{12}{3}$ ways, so the total number is $\binom{10}{2} \times \binom{12}{3}$.

(2) New York state lottery scheme. From the numbers 1,2,...,44, a person may pick any six for her ticket. The win ning number is decided by randomly selecting six numbers from the forty-four. What is the probability of winning?

The total number is $\binom{44}{6} = 7,059,052$, so the winning probability is $1/7,059,052 = 1.4 \times 10^{-7}$.

(3) Card games. 52 cards. Poker hand consists of 5 cards. ($\binom{52}{5} \approx 2.6$ million.)

More examples: a mixture of ordered and unordered —

(1) 35 people, committee of 5 with a chair. (Answer: $35 \times \binom{34}{4}$.)

(2) 5 people go to a theater show. If there is a couple who want to sit together, then the number of ways this can be done is $4! \times 2!$.

A couple behaves like a single person (inseparable) so $1 + 3 = 4$ people now, but 2! internal rearrangements possible with every sitting allocation.

Ex. Die throwing.

What is the probability of getting an even number? 3/6=1/2.

In two throws, what is the probability of getting 8 as the sum? 5/36.

Envelope matching problem.

n envelopes, n letters, randomly assigned. P(at least one match)=?

Let $A_i = i$th letter goes to ith envelope.

Then $P(A_i) = 1/n = (n-1)!/n!$, $P(A_i \cap A_j) = 1/(n(n-1)) = (n-2)!/n!$ for $i < j$, $P(A_i \cap A_j \cap A_k) = 1/(n(n-1)(n-2)) = (n-3)!/n!$ for $i < j < k$, ..., $P(A_1 \cap \cdots \cap A_n) = 1/n!$. Thus

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= 1 - \binom{n}{1}\frac{(n-1)!}{n!} + \binom{n}{2}\frac{(n-2)!}{n!} - \cdots + (-1)^{n-1}\binom{n}{n}\frac{(n-n)!}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!} \end{aligned}$$

so that P(no match)=$P_n = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \to e^{-1}$ as $n \to \infty$.

Also P(exactly k matches) can be computed. $\binom{n}{k} \times 1 \times (n-k)! P_{n-k}/n! = (1/k!)P_{n-k} \to (1/k!)e^{-1}$ as $n \to \infty$ for any fixed k.

Birthday Problem.

365 days, 35 students. What is the probability that all birthdays are different? Should you bet for all different or at least one pair same?

Capture-recapture problem.

n fish in a pond. Catch m, mark them and return. After a while, catch again k fish. How many of these captured fish will be marked?

$$P(l \text{ marked fish}) = \frac{\binom{m}{l}\binom{n-m}{k-l}}{\binom{n}{k}}, \quad l = 0, 1, 2 \ldots, \min(m, k).$$

5 Conditional Probability and Independence

Probability of A given B is defined as

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

provided that $P(B) > 0$.

Several Facts:

- $P(B|B) = 1$
- $P(A|B) = P(A)/P(B)$ for $A \subset B$.
- $P(B|A) = 1$ if $A \subset B$.
- the multiplication formula is useful when it is easier to obtain conditional probability.

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

- Similarly,

  $$P(A) = P(A \cap B) + P(A \cap B^c) = P(B)P(A|B) + P(B^c)P(A|B^c).$$

- If A and B are mutually exclusive, then

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

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

Ex. Let box 1 contain 2 red balls and 3 green balls and box 2 contain 5 red balls and 4 green balls. One box is chosen at random and one ball is drawn randomly from the chosen box. What is the probability of getting a red ball?

$A =$ red ball, $B =$ first box. Then $P(B) = P(B^c) = \frac{1}{2}$, $P(A|B) = 2/5$, $P(A|B^c) = 5/9$. Thus $P(A) = \frac{1}{2} \times \frac{2}{5} + \frac{1}{2} \times \frac{5}{9} = 43/90$.

Bayes' rule. (Inversion of conditional probability)

Let $B_1, B_2, \ldots$ be a partition of the sample space, and let A be any event. Let the probabilities $P(B_1), P(B_2), \ldots$ be given. Let the conditional probabilities $P(A|B_i)$, $i = 1, 2, \ldots$. Then the posterior probabilities are given by

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}.$$

Proof. $P(B_i|A) = P(A \cap B_i)/P(A) = P(B_i)P(A|B_i)/P(A)$. Now as $B_1, B_2, \ldots$ form a partition, $P(A) = \sum_j P(A \cap B_j) = \sum_j P(B_j)P(A|B_j)$.

Ex. (last example continued.)

Given that we obtained a red ball, what is the probability that box 1 was selected?

Answer

$$\frac{\frac{1}{2} \times \frac{2}{5}}{\frac{1}{2} \times \frac{2}{5} + \frac{1}{2} \times \frac{5}{9}} = \frac{\frac{1}{5}}{\frac{43}{90}} = 18/43.$$

Application. Rare disease.

The probability of occurrence of a disease is 1/1000. A test accurately predicts the occurrence with 99% accuracy and negates the disease with 98% accuracy. Given that the test has shown positive, what is the probability of actually having the disease?

B_1 = disease, B_2=healthy, A=test positive.

We have $P(B_1) = 0.001$, $P(B_2) = 0.999$, $P(A|B_1) = 0.99$, $P(A|B_2) = 0.02$.

By the Bayes rule, $P(B_1|A) = 0.001 \times 0.99/(0.001 \times 0.99 + 0.999 \times 0.02) = 0.047210$.

6 Independence

Definition:

Two events are *independent* if the occurrence of one of the events gives us no information about whether or not the other event will occur; that is, the events have no influence on each other. In probability theory we say that two events, A and B, are *independent* if $P(A|B) = P(A)$. (applies even if $P(A)$ or $P(B)$ is 0).

- This is equivalent to $P(A \cap B) = P(A)P(B)$, i.e, the probability that they both occur is equal to the product of the probabilities of the two individual events, i.e.

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

- symmetric relationship: A is *independent* of B implies B is *independent* of A

- Clearly, unrelated events must be independent.

Remark: Independent and mutually exclusive are not the same. In fact, If $P(A) > 0$ and $P(B) > 0$, then

(a) If A and B are mutually exclusive, they can not be independent;

(b) If A and B are independent, they can not be mutually exclusive (disjoint).

Theorem. If A and B are independent, so are the pairs, (A, B^c), (A^c, B) and (A^c, B^c).

PROOF. It is enough to prove the first; the second follows by reversal of roles, and the third by applying the first assertion on (A^c, B).

Now $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(AP(B) = P(A)(1 - P(B)) = P(A)P(B^c)$.

EX: Suppose that a man and a woman each have a pack of 52 playing cards. Each draws a card from his/her pack. Find the probability that they each draw the ace of clubs.

$(1/52 \cdot 1/52 = 0.00037)$

Ex. Two throws of a die.

A=first throw even, B=second throw even, C=sum even.

Clearly A must be independent of B, as the pair is unrelated.

What about A and C?

Note $P(C) = P(A)P(B) + P(A^c)P(B^c) = \frac{1}{2}$.

$P(A \cap C) = P(A \cap B) = P(A)P(B) = \frac{1}{4} = P(A)P(C)$, so that A and C are also independent. Similarly, B and C are also independent.

Ex. (Chevalier de Mere) What is the chance of getting at least one six in 4 throws of a die? $1 - (5/6)^4 = 0.518 > 0.5$.

What is the probability of getting at least one double six in 24 throws of a pair of dice? $1 - (35/36)^{24} = 0.4914$.

So it is profitable to bet on the first rather than the second.

Independence of a collection of events.

The idea of independence can be extended to more than two events. For example, A, B and C are independent if:

(1). A and B are independent; A and C are independent and B and C are independent (pairwise independence); and

(2). $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Definition:

A collection is *mutually independent* if for *every* subcollection, the probability of the intersection is equal to the product of probabilities.

The collection is called *pairwise independent* if every pair is independent.

Ex. Two throws of a die example.

A, B, C are pairwise independent.

Now $P(A \cap B \cap C) = P(A \cap B) = \frac{1}{4}$ while $P(A)P(B)P(C) = \frac{1}{8}$, so the collection is not mutually independent.

Theorem. $A_1, \ldots, A_n$ are mutually independent iff

$$P(A_1^{\epsilon_1} \cap \cdots \cap A_n^{\epsilon_n}) = P(A_1^{\epsilon_1}) \cdots P(A_n^{\epsilon_n})$$

for every choice of $\epsilon_i = 0, 1$, $i = 1, \ldots, n$, where $A^1 = A$ and $A^0 = A^c$.

Proof. Only if: first prove $n = 2$ and then use induction. If part: If an event is not present in an intersection, combine intersections with it and its complement, and go down one by one.

7 Random variable

[For those who are interested: A random variable is actually required to satisfy the condition that the set of all outcomes leading values inside an interval (a, b) is a member of the sigma algebra, for every $a < b$. But this condition holds in all the examples we can think about.]

Definitions:
A *random variable* is a real-valued function defined on the sample space $\mathcal{S}$, that is a rule which assigns a number to each outcome.
- A random variable is a function $\mathcal{S} \to R$.

A random variable is called *discrete* if it takes only a finite or countably many values.

Probability mass function: (pmf) The probability law of a discrete random variable can be described by the function defined as

$$p(x) = P(X = x).$$

- Only at a countable number of values for p can be positive.
- Let $x_1, x_2, \ldots$ be the points where p gives positive masses $p_1, p_2, \ldots$ respectively. Thus $P(X = x_j) = p_j$, $j = 1, 2, \ldots$.
- Clearly, $p_j \geq 0$ and $\sum_{j=1}^{\infty} p_j = 1$. These two conditions determine whether a given function is a pmf or not.

Cumulative distribution function: (cdf) for any $x \in R$, define

$$F(x) = P(X \leq x).$$

- Note $F(b) - F(a) = P(X \in (a, b])$ for $a < b$.
- Properties: (i) $0 \leq F(x) \leq 1$; (ii) $F(x)$ non-decreasing, (iii) right continuous; (iv) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.

PROOF. First clear, second from the note, third by continuity of probability as $\{X \leq x_n\}$ decreases to the $\{X \leq x\}$ as $x_n \downarrow x$, fourth by continuity of probability as $\{X \leq x\}$ decreases to the empty set as $x \to -\infty$, and the fifth similarly as $\{X \leq x\}$ increases to the sample space as $x \to \infty$.

Remark: Any function satisfying the above properties is the cdf of some random variable, in the sense that we can construct a random variable whose cdf is the given function.

Ex. Let $P(X = 0) = \frac{1}{3}, P(X = 1) = \frac{1}{2}$, and $P(X = 2) = \frac{1}{6}$.

The pmf of X, $p(x)$, is given by

$$p(0) = \frac{1}{3}, \quad p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{6}, \quad p(x) = 0 \text{ for all other } x.$$

The cdf of X, $F(x)$, is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0; \\ \frac{1}{3} & \text{for } 0 \le x < 1; \\ \frac{5}{6} & \text{for } 1 \le x < 2; \\ 1 & \text{for } x \ge 2. \end{cases}$$

The graph of F is a step function, having jumps at the points where X give mass.

This is a common feature of the cdf's of all discrete random variables.

Ex. Possible values of X are $1, 2, \ldots$.

$$p_k = P(X = k) = q^{k-1}p, \quad \text{where } 0 < p < 1, q = 1 - p, k = 1, 2, \ldots.$$

This is a pmf because

$$\begin{aligned} p_k &\ge 0, \quad \text{for all } k = 1, 2, \ldots; \\ \sum_{k=0}^{\infty} p_k &= p(1 + q + q^2 + \cdots) = p/(1 - q) = 1. \end{aligned}$$

Cdf

$$F(x) = p(1+q+q^2+\cdots+q^{k-1}) = 1-q^k, \quad \text{for } k \le x < k+1, \ k = 1, 2, \ldots.$$

Ex. Three tosses of a fair coin. $X = \#$ heads =0, 1, 2, 3.

$p_0 = \frac{1}{8}$, $p_1 = \frac{3}{8}$, $p_2 = \frac{3}{8}$, $p_3 = \frac{1}{8}$. This gives the pmf.

Cdf: $F(1.5) = F(1) = p_0 + p_1 = \frac{1}{2}$.

Ex. Let X take values $1, 2, 3, 4$ with $p_1 = 2c$, $p_2 = 4c$, $p_3 = 3c$ and $p_4 = c$. What is the value of c?

Definition:
A random variable is called *continuous* if it can assume all the values in an interval, and the cdf $F(x)$ is a continuous function.

- Continuity of cdf is the feature of all continuous random variables.
- Note that for any x, $P(X = x) = P(X \leq x) - P(X < x) = F(x) - \lim_{y \uparrow x} F(y) = 0$ by the continuity of F. This is in sharp contrast with discrete variables.

Probability density function: (pdf) In most cases, we have

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

for some function f. Such f is the pdf.

- If $F(x)$ is differentiable at a given point, then $f(x) = F'(x)$.
- This f must satisfy

  $$f(t) \geq 0, \qquad \int_{-\infty}^{\infty} f(t)dt = 1.$$

- Pdf measures the concentration of X values at a given x.
- Probability of an interval can be easily found from the density function:

  $$P(a \leq X \leq b) = \int_{a}^{b} f(t)dt.$$

  End points a and b may or may not be included. They do not matter for continuous random variable X.

identically distributed:

Let X and Y be two random variables with $P(X \in A) = P(Y \in A)$ for all sets A. Then they are called *identically distributed*.

Equivalent condition: $F_X(x) = F_Y(x)$ for all x, i.e. the two cdfs are equal. The same can be said about pmf/pdf, whenever applicable.

Ex. Hit a dart board of radius R randomly. Let X be the distance of the chosen point from the center. Probabilities are proportional to area, so $F(x) = P(X \leq x) = (\pi x^2)/(\pi R^2) = (x/R)^2$, $0 \leq x \leq R$.

Here X can assume any value between 0 and R, so that X is continuous.

Note that the cdf is a smooth continuous function, no jumps.

Ex. Show that $F(x) = 1/(1 + e^{-x})$ is a cdf. What is the corresponding pdf? Also compute $P(1 < X < 2)$.

Ex. $f(x) = cx^2$, $0 < x < 1$. What is c so that $f(x)$ is a pdf? What is $P(X < 1/2)$?

Ex. $f(x) = \lambda e^{-\lambda x}$, $x > 0$. Compute $F(x)$.

Ex. $f(x) = 3x^2 e^{-x^3}$, $x > 0$. Compute $F(x)$.

Note: A variable may neither be discrete nor continuous. Could be a mixture such as $X = 1/2$ wp $1/3$ and X has density $f(x) = 1$ on $(0, 1)$ wp $2/3$.

The cdf has jump at only $1/2$ but increases continuously otherwise.

Ex. Five throws of a fair coin. X is the number of heads, Y is the number of tails. Clearly because of symmetry the two random variables must be identically distributed. They are not identical, though.

Ex. X has density $f(x) = 1$ for $0 < x < 1$. $Y = 1 - X$ has identical distribution.

$P(Y \leq x) = P(X \geq 1 - x) = \int_{1-x}^{1} dt = x = P(X \leq x)$.