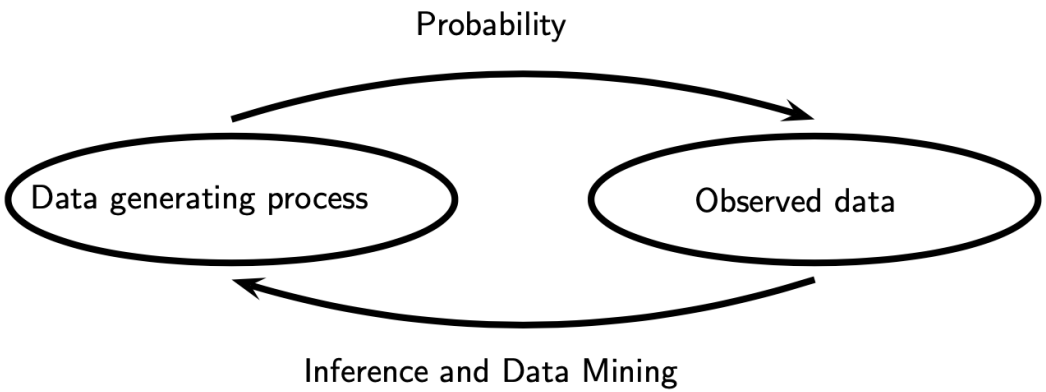
Lecture 01: Basics of Probability

Mathematical Statistics I, MATH 60061/70061

Tuesday August 31, 2021

Introduction

A phenomena (or experiment, in a general sense) is called *random* if the exact outcome is uncertain. The mathematical study of randomness is called the **theory of probability**.



To study statistical inference, we must have a good knowledge in probability theory.

Sample space

An essential piece of a **probability space** is the **sample space**.

The **sample space** S is the set of all possible outcomes of the experiment.

- An **event** A is a collection of outcomes. We can define an event by explicitly giving its outcomes,

  $$A = \{s_1, s_2, \ldots, s_n\}$$

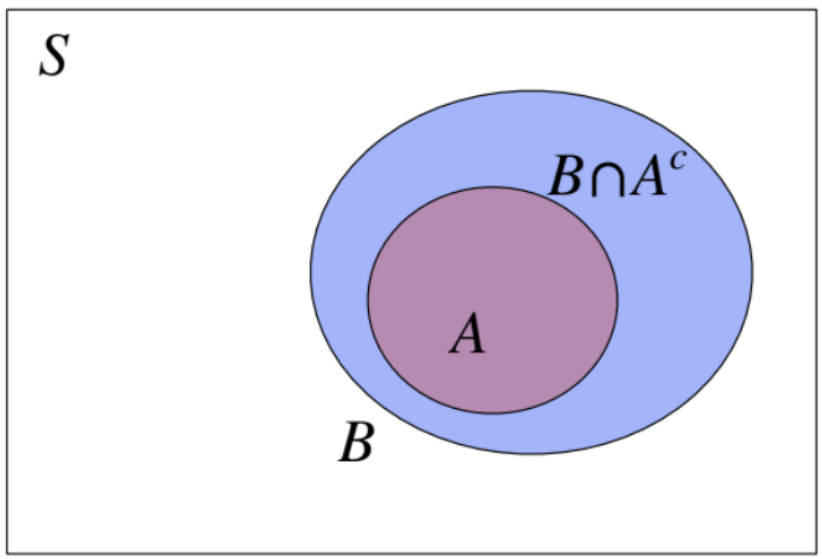
  or with a description

  $$A = \{x : x \text{ has property } \mathcal{P}\}$$

- A is a subset of the sample space S.

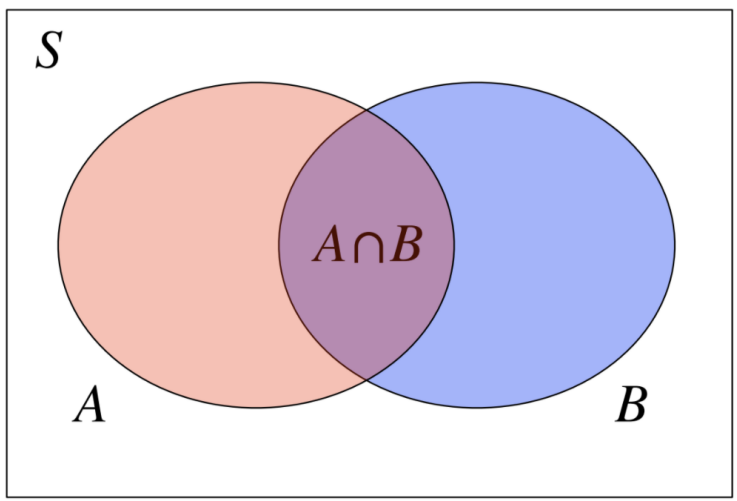
Sets and subsets

- A **set** is a collection of objects.
- A **subset** is defined as a 'set within a set'; set A is a subset of set B if and only if every element of A is also an element of B.
- The **empty set**, $\emptyset$, is the set that contains no elements.



Set operations

- The **union** of two sets A and B, $A \cup B$, is the set of all objects that are in A or B (or both).
- The **intersection** of two sets A and B, $A \cap B$, is the set of all objects that are in A and B.
- The **complement** of a set A, A^c, is the set of all objects in S that are *not* in A.
- **Disjoint sets** are sets that do not overlap.



Set identities

For any three events A, B, and C, defined on a sample space S, we have the following:

1. Commutivity. $A \cup B = B \cup A$, $A \cap B = B \cap A$.
2. Associativity. $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$.
3. Distributive 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. DeMorgan's Laws. $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$

Verify these identities for yourself.

Sigma algebras

A collection of subsets of S is called a **sigma algebra** (or **Borel field**) $\mathcal{B}$, if it satisfies the following properties:

1. $\emptyset \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$).
2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$.
3. If $A_1, A_2, \ldots \subset \mathcal{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Whenever S is finite (or countable), then we take $\mathcal{B}$ to be all subsets of S.

Axioms of Probability

Given a sample space S and an associated sigma algebra $\mathcal{B}$, a **probability function** is a function P with domain $\mathcal{B}$ that satisfies

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.
2. $P(S) = 1$.
3. If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The probability space consists of a sample space, an associated sigma algebra, and a probability function.

Consequences of the Axioms

If P is a probability function and A is any set in $\mathcal{B}$, then

- $P(\emptyset) = 0$, where $\emptyset$ is the empty set.
- $P(A) \leq 1$.
- $P(A^c) = 1 - P(A)$.

Consequences of the Axioms

If P is a probability function and A is any set in $\mathcal{B}$, then

- $P(\emptyset) = 0$, where $\emptyset$ is the empty set.
- $P(A) \leq 1$.
- $P(A^c) = 1 - P(A)$.

The sets A and A^c form a partition of the sample space, $S = A \cup A^c$. Therefore

$$P(A \cup A^c) = P(S) = 1$$

by the Axiom 2. Also, A and A^c are disjoint, so by the Axiom 3,

$$P(A \cup A^c) = P(A) + P(A^c).$$

So, $P(A^c) = 1 - P(A)$.

Consequences of the Axioms

If P is a probability function and A and B are any sets in $\mathcal{B}$, then

- $P(B \cap A^c) = P(B) - P(A \cap B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- If $A \subset B$, then $P(A) \leq P(B)$.
- $P(A \cap B) \geq P(A) + P(B) - 1$. (Bonferroni's Inequality)

If P is a probability function, then

- $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \ldots$
- $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets $A_1, A_2, \ldots$. (Boole's Inequality)

Counting

Calculating the probability of an event A involves counting the number of elements in A and the number of elements in S.

Fundamental methods for counting

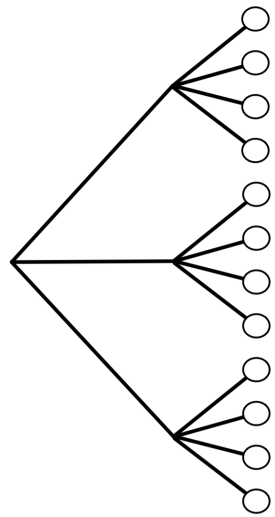
- Multiplication rule
  - Sampling with replacement
  - Sampling without replacement
  - Permutation
- Adjustment for overcounting
  - Binomial coefficient

Multiplication rule

Consider a compound experiment consisting of two sub-experiments, Experiment A and Experiment B:

- Experiment A has a possible outcomes.
- Experiment B has b possible outcomes for each of those outcomes in Experiment A.

The compound experiment has ab possible outcomes.



The same rule applies to cases with > 2 sub-experiments.

Ordered sampling

Possibilities of choosing k objects out of n, one at a time, where order matters (e.g., $\boxed{2}\boxed{3}$ and $\boxed{3}\boxed{2}$ represent different outcomes)

- **Sampling with replacement** (i.e., choosing a certain object does not preclude it from being chosen again):

- **Sampling without replacement** (i.e., choosing a certain object precludes it from being chosen again):

Ordered sampling

Possibilities of choosing k objects out of n, one at a time, where order matters (e.g., $\boxed{2}\boxed{3}$ and $\boxed{3}\boxed{2}$ represent different outcomes)

- **Sampling with replacement** (i.e., choosing a certain object does not preclude it from being chosen again):
  - k sub-experiments; each has n possible outcomes

$$\underbrace{n \cdot n \cdots n}_{k \text{ times}}$$

- **Sampling without replacement** (i.e., choosing a certain object precludes it from being chosen again):

Ordered sampling

Possibilities of choosing k objects out of n, one at a time, where order matters (e.g., $\boxed{2}\boxed{3}$ and $\boxed{3}\boxed{2}$ represent different outcomes)

- **Sampling with replacement** (i.e., choosing a certain object does not preclude it from being chosen again):
  - k sub-experiments; each has n possible outcomes

$$\underbrace{n \cdot n \cdots n}_{k \text{ times}}$$

- **Sampling without replacement** (i.e., choosing a certain object precludes it from being chosen again):
  - k sub-experiments; the # of possible outcomes decreases by 1 each time (0 possibilities for $k > n$)

$$n(n-1)\cdots(n-k+1), \text{ for } 1 \leq k \leq n$$

Permutations and factorials

A **permutation** of elements $1, 2, \ldots, n$ is an arrangement of them in some order. For example, $3, 5, 1, 2, 4$ is a permutation of $1, 2, 3, 4, 5$

Permutations can be viewed as a special case of ordered sampling without replacement, where $k = n$. There are

$$n(n-1) \cdots 1 = n!$$

permutations of $1, 2, \ldots, n$.

This can be used for adjusting unordered sampling.

Binomial coefficient

A **binomial coefficient** $\binom{n}{k}$ counts the number of subsets of a certain size, e.g., to choose a group of size k from a set of n people.

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

- Set and subsets are *unordered*, e.g., $\{2,3\} = \{3,2\}$.
- There are $n(n-1)\cdots(n-k+1)$ ways to make *ordered* choice of k elements without replacement.
- Each subset is overcounted by a factor of $k!$.

Sampling table

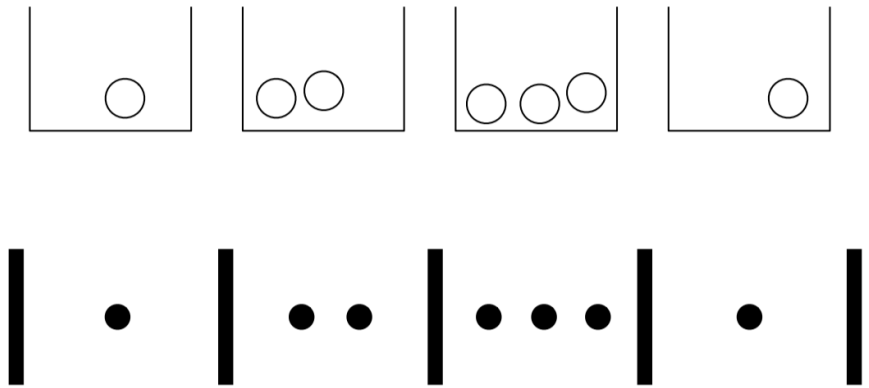
Choosing k times from a set of n objects

	Order matters	Order doesn't matter
With replacement	n^k	$\binom{n+k-1}{k}$
Without replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

The top-right case is much harder to solve than the others.

Sampling with replacement, order doesn't matter

Consider an equivalent problem to count the number of ways to put k identical balls into n boxes, e.g., putting 7 balls into 4 boxes



Same as repeating the procedure for k times: "sample one box (out of n), and add a ball to it"

- Sampling boxes *with replacement*: we can put as many balls in any box as we want
- *Order doesn't matter*: balls are identical

Sampling with replacement, order doesn't matter

Consider an equivalent problem to count the number of ways to put k identical balls into n boxes, e.g., putting 7 balls into 4 boxes

Encoding balls as $\bullet$ and walls of boxes as $|$, any sequence of k $\bullet$'s and $n-1$ $|$'s in between a starting $|$ and an ending $|$ is a valid sample. The number of possible such sequence is

$$\binom{n+k-1}{k}$$

Definition of conditional probability

If A and B are events with $P(B) > 0$, then the **conditional probability** of A given B, denoted by $P(A \mid B)$, is defined as

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

- A is the event whose *uncertainty* we want to update.
- B is the *evidence* we observe (or want to treat as given).

Definition of conditional probability

If A and B are events with $P(B) > 0$, then the **conditional probability** of A given B, denoted by $P(A \mid B)$, is defined as

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

- $P(A)$ is the **prior** probability of A.
- $P(A \mid B)$ is the **posterior** probability of A.

Probability of the intersection of multiple events

For any two events A and B with positive probabilities,

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

This is immediate from the definition of conditional probability.

Probability of the intersection of multiple events

For any two events A and B with positive probabilities,

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

This is immediate from the definition of conditional probability.

For any events $A_1, \ldots, A_n$ with $P(A_1, \ldots, A_{n-1}) > 0$,

$$\begin{aligned} P(A_1, A_2, \ldots, A_n) &= P(A_1, \ldots, A_{n-1})P(A_n \mid A_1, \ldots, A_{n-1}) \\ &= P(A_1)P(A_2 \mid A_1) \cdots P(A_n \mid A_1, \ldots, A_{n-1}), \end{aligned}$$

where the commas denote intersections.

- There are in fact $n!$ expressions for the RHS.
- Often the RHS will be easier to compute with some orderings.

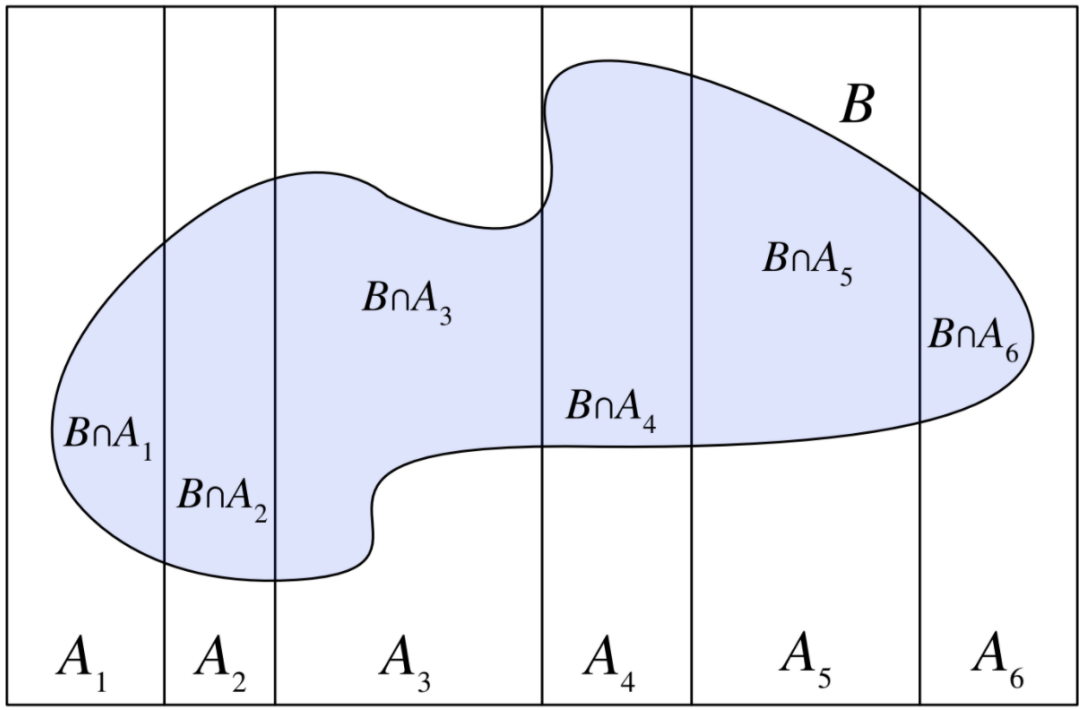
Bayes' rule

Bayes' rule is extremely useful, which relates $P(A \mid B)$ to $P(B \mid A)$:

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

- Often the $P(B \mid A)$ is easier to find than $P(A \mid B)$.
- Finding $P(B)$ can be non-trivial.

The law of total probability (LOTP)



Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i) > 0$ for all i,

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

Independence of two events

Definition: Events A and B are **independent** if

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

Independence of two events

Definition: Events A and B are **independent** if

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

Since $P(A \mid B) = P(A \cap B)/P(B)$, if $P(A) > 0$ and $P(B) > 0$, then this is equivalent to

$$P(A \mid B) = P(A), \text{ and } P(B \mid A) = P(B).$$

A and B are independent if learning that B occurred gives us no information about the probability of A occurring (and vice versa).

Independence of complements

If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.

Independence of complements

If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.

Assuming $P(A) \neq 0$ (since if $P(A) = 0$, A is independent of every event), if A and B are independent, then $P(B \mid A) = P(B)$, and

$$P(B^c \mid A) = 1 - P(B \mid A) = 1 - P(B) = P(B^c).$$

So A and B^c are independent.

Independence of complements

If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.

Assuming $P(A) \neq 0$ (since if $P(A) = 0$, A is independent of every event), if A and B are independent, then $P(B \mid A) = P(B)$, and

$$P(B^c \mid A) = 1 - P(B \mid A) = 1 - P(B) = P(B^c).$$

So A and B^c are independent.

Swapping A and B, we have that A^c and B are independent.

Independence of complements

If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.

Assuming $P(A) \neq 0$ (since if $P(A) = 0$, A is independent of every event), if A and B are independent, then $P(B \mid A) = P(B)$, and

$$P(B^c \mid A) = 1 - P(B \mid A) = 1 - P(B) = P(B^c).$$

So A and B^c are independent.

Swapping A and B, we have that A^c and B are independent.

Using the fact that A, B independent implies A, B^c independent, with A^c, B independent, we have that A^c and B^c are independent.

Independence of three events

Events A, B, and C are said to be independent if *all* of the following equations hold:

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

If the first three conditions hold, we say that A, B, and C are *pairwise independent*.

Independence of many events

For n events $A_1, A_2, \ldots, A_n$ to be independent, we require

- any pair to satisfy $P(A_i \cap A_j) = P(A_i)P(A_j)$ (for $i \neq j$),
- any triplet to satisfy $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ (for i, j, k distinct),
- similarly for all quadruplets, all quintuplets, ...