Basic Probability

Lecture 1

Aug. 19, 21 and 26, 2025

Outline

Course Overview

- Basic Probability (Gubner Chapter 1):
- Random Variables (Gubner Chapters 2, 3, 4)
- Multiple Random Variables/ Random Vectors (Gubner Chapters 6, 8)
- Random Processes (Gubner (parts of) Chapters 7, 9-12)

Classical Definition of Probability

This was the prevailing definition for many centuries.

- Define the probability of an event A as $P(A) = \frac{N_A}{N}$, where N is the number of possible outcomes of the random experiment and N_A is the number of outcomes favorable to the event A.
- For example, for a 6-sided die there are 6 outcomes and 3 of them are even, thus $P(even) = 3/6$.
- Problems with this classical definition:
  - Here the assumption is that all outcomes are equally likely (probable). Thus, the concept of probability is used to define probability itself! Cannot be used as basis for a mathematical theory.
  - In many random experiments, the outcomes are not equally likely.
  - The definition doesn't work when the number of possible outcomes is infinite.

Axiomatic Definition of Probability

- The axiomatic definition of probability was introduced by A. Kolmogorov in 1933. It provides rules for assigning probabilities to events in a mathematically consistent way and for deducing probabilities of events from probabilities of other events.
- Elements of axiomatic definition:
  - Set of all possible outcomes of the random experiment Ω (sample space)
  - Set of events, which are subsets of Ω
  - A probability law (measure or function) that assigns probabilities to events such that
    1. $P(A) \geq 0$
    2. $P(\Omega) = 1$
    3. $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint events (i.e., $A \cap B = \emptyset$).
- These rules are consistent with relative frequency interpretation.

Random Variables

- A **random variable** is a function defined on the outcomes in the sample space.
- Consider roll of a (6-sided) die with $\Omega = \{1, 2, 3, 4, 5, 6\}$. Suppose you bet \$10. that the outcome is 3. So you win \$10. if outcome $\omega = 3$, and lose \$10. if $\omega =$1, 2, 4, 5, or 6. Thus, the following function X defined on ω models the result of your bet:

$$X(\omega) = \begin{cases} 10 & \text{if } \omega = 3 \\ -10 & \text{otherwise} \end{cases}$$

  Assuming a fair die, $P(X = 10) = 1/6$ and $P(X = -10) = 5/6$.
- Randomness in X arises from randomness of outcomes in Ω, not from the definition of the function.

Random Processes

- A **random process** is a function of two variables: outcomes in a probability sample space, and time (or some other variable). It is typically denoted as $X(t)$ even though in fact, it is $X(\omega, t)$.
- Consider roll of a (6-sided) die with $\Omega = \{1, 2, 3, 4, 5, 6\}$. Suppose it represents one of 6 symbols in digital communications. Each symbol is coded into a voltage waveform $p_\omega(t)$, $0 \leq t \leq T$. Then $X(t) = p_\omega(t)$ is a random process.
- In the study of random variables and random processes, our main objective is to characterize their probabilistic properties which are derived from the basic probability concepts.

Set Theory

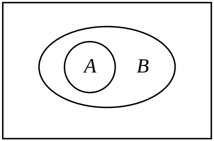
- A **set** is a collection of some "objects" (items, things, elements, ...). We assume a universal set (largest set) Ω. For instance, roll of a (6-sided) die yields $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Notation
  - If ω is a member of Ω, we write $\omega \in \Omega$.
  - If any element of A is also in B, then A is a subset of B, denoted as $A \subseteq B$.
  - But if there is at least one element in B that is not in A, then A is a proper subset of B, denote as $A \subset B$.
- Set Operations
  - **Union/OR**: $A \cup B$= set of elements that are either in A, or in B, or in both A and B.
  - **Intersection/AND**: $A \cap B$= set of elements that are common to A and B.
  - **Complement/NOT**: A^c or $\bar{A}$= set of elements in Ω that are not in A. We denote Ω^c as $\emptyset$, the null or empty set.

More Set Operations

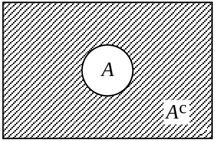
- Notation
  - $\cup_{i=1}^{n} A_i = A_1 \cup A_2 \cdots \cup A_n$
  - $\cap_{i=1}^{n} A_i = A_1 \cap A_2 \cdots \cap A_n$
- Definitions
  - A collection of sets $A_1, A_2, \cdots, A_n$ are **disjoint** or **mutually exclusive** if $A_i \cap A_j = \emptyset$ for all $i \neq j$, i.e., no two of them have a common element.
  - A collection of sets $A_1, A_2, \cdots, A_n$ **partition** Ω if they are disjoint and $\cup_{i=1}^{n} A_i = \Omega$.
- **DeMorgan's Laws**
  - $(\cup_{i=1}^{n} A_i)^c = \cap_{i=1}^{n} A_i^c$
  - $(\cap_{i=1}^{n} A_i)^c = \cup_{i=1}^{n} A_i^c$

Basic Relations

- $A \cap \Omega = A$
- $(A^c)^c = A$
- $A \cap A^c = \emptyset$
- Commutative law: $A \cup B = B \cup A$
- Associative law: $A \cup (B \cup C) = (A \cup B) \cup C$
- Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- These can all be proven using the definition of set operations or visualized using Venn Diagrams



(a)



(b)

Sample Spaces: Examples

- Sample space is called **discrete** if it contains a finite or a countable number of sample points
- Examples:
  - Flip a coin once: $\Omega = \{H, T\}$.
  - Flip a coin three times: $\Omega = \{HHH, HHT, HTH, ...\} = \{H, T\}^3 = \{H, T\} \times \{H, T\} \times \{H, T\}$.
  - Number of packets arriving in time interval $(0, T] = 0 < t \leq T$ at a node in a communication network : $\Omega = \{0, 1, 2, 3, \cdots\}$

  Note that the first two examples have **finite** Ω, whereas the last has countably infinite Ω. Both types are called discrete.
- Packet arrival time: $t \in (0, \infty)$, thus $\Omega = (0, \infty)$
- Arrival times for n packets: $t_i \in (0, \infty)$, for $i = 1, 2, \cdots, n$, thus $\Omega = (0, \infty)^n$
- Sample space is called **mixed** if it is neither discrete nor continuous, e.g., $\Omega = [0, 1] \cap \{3\}$

Axioms of Probability

Experiment: Some action that results in an outcome. **Random Experiment**: An experiment in which the outcomes are uncertain before the experiment is performed. **Sample Space** Ω (of a random experiment) is the set of all possible outcomes. In probability a subset of Ω is called an **event**.

Probability Space is the triple $(\Omega, \mathcal{F}, P)$ where Ω is the sample space, $\mathcal{F}$ is the sigma-field (σ-field, also σ-algebra) of (some) subsets of Ω, and P is a probability measure (set function) defined on sets in $\mathcal{F}$.

Probability P satisfies the following axioms:

1) $P(A) \geq 0$ for every $A \in \mathcal{F}$.
2) $P(\Omega) = 1$
3a) $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.
3b) $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ if all A_is are disjoint, i.e., $A_i \cap A_j = \emptyset$ for any $i \neq j$.

Discrete Probability Spaces

- For discrete sample spaces, the set of events $\mathcal{F}$ can be taken to be the set of all subsets of Ω, sometimes called the **power set** of Ω. For example, for the coin flipping experiment, $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$.
- The probability measure P can be defined by assigning probabilities to individual outcomes – single outcome events $\{\omega\}$ ("atoms") – so that

$$P(\{\omega\}) \geq 0 \text{ for every } \omega \in \Omega\,, \quad \Sigma_{\omega\in\Omega} P(\{\omega\}) = 1\,.$$

- The probability of any other event A is simply

$$P(A) = \Sigma_{\omega\in A} P(\{\omega\})$$

- Example: For the die rolling experiment, assign $P(i) = \frac{1}{6}$ for $i = 1, 2, \cdots, 6$. The probability of the event "the outcome is even," $A = \{2, 4, 6\}$, is

$$P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6} = \frac{1}{2}$$

Continuous Probability Spaces

- A **continuous sample space** has an uncountable number of elements. Example: $\Omega = [0, 1]$.
- For continuous Ω, we cannot in general define the probability measure P by first assigning probabilities to outcomes.
- To see why, consider assigning a uniform probability measure over $[0, 1]$.
  - In this case the probability of each single outcome event is zero
  - How do we find the probability of an event such as $A = [0.25, 0.75]$?
- Another difference for continuous Ω: we cannot take the set of events $\mathcal{F}$ as the power set of Ω. (To learn why, you need to study measure theory, which is beyond the scope of this course.)
- The set of events $\mathcal{F}$ cannot be an arbitrary collection of subsets of Ω. It must make sense, e.g., if A is an event, then its complement A^c must also be an event, the union of two events must be an event, and so on.

σ-field

- σ-field $\mathcal{F}$ is a collection of sets that satisfies the following axioms:
  1) $\emptyset \in \mathcal{F}$
  2) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
  3) If all A_is, $i = 1, 2, \cdots$, are in $\mathcal{F}$, then $\cup_{i=1}^{\infty} A_i = B \in \mathcal{F}$.
- Of course, the power set is a σ-field. But we can define smaller σ-fields. For example, for rolling a die, we could define the set of events as $\mathcal{F} = \{\emptyset, \text{ odd}, \text{ even}, \Omega\}$.
- $\Rightarrow$ We will *ignore* $\mathcal{F}$ in the rest of the course. Just think of probability P as a set function defined on sets and subsets of Ω (and on sets obtained as a result of set operations on subsets of Ω).

- For $\Omega = (-\infty, \infty)$ (or, $(0, \infty)$, $(0, 1)$, etc.), $\mathcal{F}$ is typically defined as the family of sets obtained by starting from the intervals and taking countable unions, intersections, and complements.
- The resulting $\mathcal{F}$ is called the **Borel field**.
- Note: Amazingly there are subsets in R (real line) that cannot be generated in this way!
- To define a probability measure over a Borel field, we first assign probabilities to the intervals in a consistent way, i.e., in a way that satisfies the axioms of probability. For example, to define uniform probability measure over $(0, 1)$, we first assign $P((a, b)) = b - a$ to all intervals (a, b), $0 < a \leq b < 1$

Some Consequences of Axioms

- $P(\emptyset) = 0$.
  ($\Omega = \Omega \cup \emptyset$ where $\Omega \cap \emptyset = \emptyset$. Apply Axioms 2 and 3.)
- $P(A^c) = 1 - P(A)$.
  ($\Omega = A \cup A^c$ where $A \cap A^c = \emptyset$. Apply Axioms 2 and 3.)
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  (The set $A \cap B$ is counted twice, as a subset of A as well as of B, in $P(A)$ and in $P(B)$.
  Alternatively, note that $A \cup B = A \cup (A^c \cap B)$ and $B = (A^c \cap B) \cup (A \cap B)$. Then $P(A \cup B) = P(A) + P(A^c \cap B)$ and $P(B) = P(A^c \cap B) + P(A \cap B) \Rightarrow$ desired result.)
- $P(A) \leq P(B)$ if $A \subset B$.
  ($B = A \cup C$ where $C = B \cap A^c$. Since $A \cap C = \emptyset$, $P(B) = P(A) + P(C) \geq P(A)$ as $P(C) \geq 0$.)

Birthday Paradox

The"birthday paradox" examines the chances that two people in a group have the same birthday. It is a "paradox" not because of a logical contradiction, but because it goes against intuition. Take the number of days in a year to be 365. Suppose there are n people in a room. Let X_i be the birthday of the ith person. The sample space consists of all the n-tuples of birthdays: $|\Omega| = 365^n$. Let $A =$ "At least two people have the same birthday," and therefore, $A^c =$ "No two people have the same birthday." We have $P(A) = 1 - P(A^c)$. We will calculate $P(A^c)$, since it is easier, and then find $P(A)$. How many ways are there for no two people to have the same birthday? Well, there are 365 choices for the first person, 364 for the second, . . . , $(365 - n + 1)$ choices for the nth person, for a total of $365 \times 364 \times \cdots \times (365 - n + 1)$. Thus we have

$$P(A^c) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$$

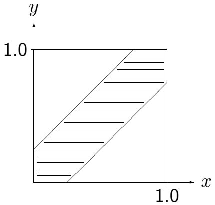
This allows us to compute $P(A) = 1 - P(A^c)$ as a function of the number of people, n. Denote it by $P_n(A)$.

$$P_2(A) = 0.0027\,, \quad P_4(A) = 0.0164\,, \quad P_{23}(A) = 0.5073\,, \quad P_{60}(A) = 0.9941$$

Another Example

Romeo and Juliet have a date. Each arrives late with a random delay of up to 1 hour. Each will wait only 1/4 of an hour before leaving. What is the probability that Romeo and Juliet will meet?

The pair of delays is equivalent to that achievable by picking two random numbers between 0 and 1. Define probability of an event as its area. The event of interest is represented by the cross hatched region:
$|x - y| \leq 0.25$



Probability of the event is given by the area of crosshatched region

$$1 - 2 \times \frac{1}{2}(0.75)^2 = 0.4375$$

Conditional Probability

- Conditional probability allows us to compute probabilities of events based on partial knowledge of the outcome of a random experiment
- Examples
  - We are told that the sum of the outcomes from rolling a die twice is 9. What is the probability the outcome of the first die was a 6?
  - A spot shows up on a radar screen. What is the probability that there is an aircraft?
  - You receive a 0 at the output of a digital communication system. What is the probability that a 0 was sent?

Conditional Probability

- Let B be an event such that $P(B) \neq 0$. The conditional probability of event A given B is defined to be

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

- The function $P(\cdot|B)$ is a probability measure over $\mathcal{F}$, i.e., it satisfies the axioms of probability.
- $P(A, B) = P(A)P(B|A) = P(B)P(A|B)$
- Law of Total Probability: Let $A_1, A_2, \cdots, A_n$ partition Ω, i.e., they are disjoint and $\cup_{i=1}^{n} A_i = \Omega$. Then for any event B,

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

  Using conditional probability, we have

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

Bayes Rule

- Let B be an event such that $P(B) \neq 0$ and let Let $A_1, A_2, \cdots, A_n$ partition Ω. Then

$$P(A_j|B) = \frac{P(A_j, B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{P(B)}, \quad j = 1, 2, \cdots n$$

- By law of total probability,

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

- This yields the Bayes rule

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Example (from Pishro-Nik)

A box contains three coins: two regular coins and one fake two-headed coin ($P(H) = 1$).

- You pick a coin at random and toss it. What is the probability that it lands heads up?
- You pick a coin at random and toss it, and get heads. What is the probability that it is the two-headed coin?

Let C_1 be the event that you choose a regular coin, and let C_2 be the event that you choose the two-headed coin. C_1 and C_1 form a partition of the sample space. Given $P(H|C_1) = 0.5$ and $P(H|C_2) = 1$.

- By total probability,

$$P(H) = P(H|C_1)P(C_1) + P(H|C_2)P(C_2) = \frac{1}{2} \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{2}{3}$$

- Use Bayes rule:

$$P(C_2|H) = \frac{P(H|C_2)P(C_2)}{P(H)} = \frac{1 \times \frac{1}{3}}{\frac{2}{3}} = 0.5$$

(Statistical) Independence

- Events A and B are said to be (statistically) independent if

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

- If $P(B) \neq 0$, then the above statement is equivalent to $P(A|B) = P(A)$, i.e., knowing whether B occurs does not change the probability of A.
- Events A_i, $i = 1, 2, \cdots, n$, are said to be independent if

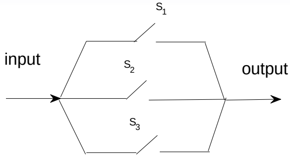
$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) \text{ for } i \neq j \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) \text{ for } i \neq j \neq k \\ \vdots \quad &\vdots \end{aligned}$$

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

- $P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$ alone is not sufficient for independence

Each switch in the figure shown operates independently and it remains closed with probability p..

- What is the probability that there is a closed path from input to output.
- Suppose there exists a closed path. What is the probability that switch S_1 is open?



Let S_i denote the event switch i is closed, and C_P denote the event that there is a closed path. Then

$$P(C_P) = 1 - P(C_P^c) = 1 - P(\cap_{i=1}^3 S_i^c) = 1 - \prod_{i=1}^{3} P(S_i^c) = 1 - (1-p)^3$$

where we used independence to deduce $P(\cap_{i=1}^3 S_i^c) = \prod_{i=1}^3 P(S_i^c)$. The conditional probability of S_1^c is

$$P(S_1^c | C_P) = \frac{P(C_P|S_1^c)P(S_1^c)}{P(C_P)} = \frac{(1-(1-p)^2)(1-p)}{1-(1-p)^3}$$

Roll two fair dice independently, and define the following events: A : first die is 1,2,3; B : first die is 2,3,6; C : sum of outcomes is 9. Are the events A, B and C independent?

$$P(A \cap B \cap C) = P(\{(3,6)\}) = \frac{1}{36}$$
$$P(A) = P(B) = \frac{1}{2}$$
$$P(C) = P(\{(3,6),(6,3),(4,5),(5,4)\}) = \frac{4}{36} = \frac{1}{9}.$$

Hence, we have $P(A \cap B \cap C) = P(A)P(B)P(C)$, but events A and B are not (pairwise) independent, i.e.,

$$P(A \cap B) = \frac{1}{3} \neq \frac{1}{2} \times \frac{1}{2} = P(A)P(B).$$

Permutations

Permutations How many different ordered arrangements of the letters a, b, and c are possible? By direct enumeration we see that there are 6, namely, abc, acb, bac, bca, cab, and cba. Each arrangement is known as a permutation. Thus, there are 6 possible permutations of a set of 3 objects.

Suppose now that we have n objects. Reasoning similar to that we have just used for the 3 letters then shows that there are

$$n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n!$$

different permutations of the n objects.

The term $n!$ is read as "n factorial." By convention, $0! = 1$.

Combinations

Combinations We are interested in determining the number of different groups of r objects that could be formed from a total of n $(\geq r)$ objects, when the order in which the objects are selected is not relevant.
For instance, how many different groups of 3 could be selected from the 5 items A, B, C, D and E? Since there are 5 ways to select the initial item, 4 ways to then select the next item, and 3 ways to select the final item, there are thus $5 \times 4 \times 3$ ways of selecting the group of 3 when the order in which the items are selected is relevant. However, since every group of 3 – say, the group consisting of items A, B and C – will be counted 6 times (that is, all of the permutations ABC, ACB, BAC, BCA, CAB and CBA will be counted when the order of selection is relevant), it follows that the total number of groups that can be formed is

$$\frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10 \,.$$

In general, as $n(n-1)\cdots(n-r+1)$ represents the number of different ways that a group of r items could be selected from n items when the order of selection is relevant, and as each group of r items will be counted $r!$ times in this count, it follows that the number of different groups of r items that could be formed from a set of n items is

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!}\,.$$

Notation We define $\binom{n}{r}$, for $r \leq n$, by

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$$

and say that $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time. We take $\binom{n}{r}$ to be 0 if $r < 0$ or $r > n$. MATLAB function *nchoose(m,r)* implements $\binom{m}{r}$

A committee of 3 is to be formed from a group of 20 people. How many different committees are possible? There are

$$\binom{20}{3} = \frac{20 \times 19 \times 18}{3 \times 2 \times 1} = 1140 \text{ possible committees.}$$

Bernoulli Trials

A Bernoulli Trial is a random experiment that has two possible outcomes which we can label as "success" and "failure," or events A and A^c. A Binomial Experiment consists of n independent Bernoulli trials where count the total number of successes (or failures).

In a given trail, we have $\Omega = \{s, f\}$. Conduct n trials resulting in $\Omega^n = \Omega \times \Omega \cdots \times \Omega = \{ss \cdots s,\ fs \cdots s, \cdots, ff \cdots f\}$= set of 2^n possible sequences. Typically the probability of success is denoted by p and probability of failure by $q = 1 - p$.

Let $A_k = \{k \text{ successes in } n \text{ trials}\}$. What is $P(A_k)$? The probability of a particular sequence of s and fs that is in A_k is $p^k(1-p)^{n-k}$. All such sequences have the same probability. There are $\binom{n}{k}$ sequences in A_k. Hence

$$P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

A sequence of 10 bits (zeros or ones) is decoded at a receiver. The error probability (i.e., a zero is decoded as a one, or vice versa) is 0.001, and each bit is decoded independently of the other bits. Find the probability of at least one error in the sequence.

$$
\begin{aligned}
P(\text{At least one error}) &= 1 - P(\text{no error}) \\
&= 1 - \binom{10}{0}(1-0.001)^{10}(0.001)^0 \\
&= 1 - (0.999)^{10} \\
&= 1 - (1-0.001)^{10} \approx 1 - (1 - 10 \times 0.001) = 0.01
\end{aligned}
$$

where we use $(1-x)^n \approx 1 - nx$ (Taylor series expansion around $x = 0$) for $|x| \ll 1$.