I Basic Definitions

I.1 Bertrand's paradox

Circle of radius 2

Choose a chord of this circle at random (uniformly)

What is the probability $P$ that this chord intersects the concentric circle of radius 1?

Sol.1 Chord is uniquely determined by the position of its midpoint

$$P = \frac{\text{area of inner circle}}{\text{area of larger circle}} = \frac{1}{4}$$

Sol.2 Symmetry under rotation → assume that the chord is vertical

$$P = \frac{\text{diameter of inner circle}}{\text{diameter of larger circle}} = \frac{2}{4} = \frac{1}{2}$$
Sol. 3 Symmetry - One end of the chord is at the far left point of the larger circle.

\[-\frac{\pi}{2} < \theta < \frac{\pi}{2}\]

chord hits the inner circle \(\iff\) \(-\frac{\pi}{6} < \theta < \frac{\pi}{6}\)

\[P = \frac{\frac{2\pi}{6}}{2\pi/2} = \frac{1}{3}\]

\[\Rightarrow \text{We must carefully define what we mean by the term "random"}\]
1.2 Probability spaces and $\mathcal{B}$-algebras

**Sample Space**

$\Omega$: a non-empty set

- set of possible outcomes of a random experiment
- called: sample space, universal sample space, universe, outcome set

6-faced die $\Omega = \{1, 2, 3, 4, 5, 6\}$

**$\mathcal{B}$-algebra**

**Def 1.2.1**

A $\mathcal{B}$-algebra is a collection $\mathcal{U}$ of subsets of $\Omega$ with these properties:

(i) $\emptyset, \Omega \in \mathcal{U}$

(ii) If $A \in \mathcal{U}$ then $A^c \in \mathcal{U}$

(iii) If $A_1, A_2, \ldots \in \mathcal{U}$ then

\[ \bigcup_{k=1}^{\infty} A_k \in \mathcal{U} \text{ and } \bigcap_{k=1}^{\infty} A_k \in \mathcal{U} \]

$\mathcal{U}$ is also called a $\sigma$-field
- sometimes not written $\mathcal{F}$

elements of $\mathcal{U}$ are called "events"
6-faced die

\[ U = \{ \Omega, \emptyset \} \quad \text{(trivial \( \sigma \)-algebra)} \]

or \[ U = \{ \Omega, \emptyset, \{1, 3, 5\}, \{2, 4, 6\} \} \]

or \[ U = \{ \Omega, \emptyset, \{1, 2, 3\}, \{4, 5, 6\} \} \]

Let \( U' = \{ \Omega, \emptyset, \{1, 2\}, \{3, 4\}, \{5, 6\} \} \)

is \( U' \) a \( \sigma \)-algebra?

The outcome of the die is an element of \( \Omega \)

the "event": "the result of the die is odd"

is not an element of \( \Omega \)

Ex. Let \( A \) and \( B \) belong to some \( \sigma \)-field \( U \)

Does \( U \) contain:

- \( A \cap B \)
- \( A \setminus B \)
- \( A \Delta B \)

Ex. 6-faced die. Assume that \( U \) contains the events:

- the result is odd,
- the result is between 1 and 2
- the result is 1 or 2 or 3... or 6?

Does \( U \) contain the events:

- The result is 1 or 2 or 3... or 6?
- The result is 2 or 3 or 5?
Probability measure

Def I.2.2
Let \( \mathcal{U} \) be a \( \sigma \)-algebra of subsets of \( \Omega \). We call \( P : \mathcal{U} \rightarrow [0,1] \) a probability measure provided

(i) \( P(\emptyset) = 0 \), \( P(\Omega) = 1 \)
(ii) If \( A_1, A_2, \ldots \in \mathcal{U} \) then
\[
P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)
\]
(iii) If \( A_1, A_2, \ldots \) are disjoint sets in \( \mathcal{U} \) then
\[
P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)
\]

Probability space

Def I.2.3
A triple \((\Omega, \mathcal{U}, P)\) is called a probability space provided that \( \Omega \) is any set, \( \mathcal{U} \) is a \( \sigma \)-algebra of subsets of \( \Omega \), and \( P \) is a probability measure on \( \mathcal{U} \).

Ex: Can we choose \( P \) as a mapping \( \Omega \rightarrow [0,1] \) (discrete prob.)? Is it to take \( w \) or pick \( w \) uniformly in \([0,1]\)?
Terminology

(i) \( A \in \Omega \): an event

(ii) \( w \in \Omega \): sample point

(iii) \( P(A) \): probability of the event \( A \)

(iv) A property which is true except for an event of probability zero is said to hold almost surely (abbreviated "a.s.")

**Example**

\( \Omega = \{w_1, ..., w_N\} \) finite set (6-faced die \( \Omega = \{1, 2, 3, 4, 5, 6\} \))

\( p_1, ..., p_N \): numbers such that

\[ 0 \leq p_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{N} p_i = 1 \]

\( \mathcal{U} \): set of all subsets of \( \Omega \)

For each \( A = \{w_{j_1}, ..., w_{j_m}\} \) with \( 1 \leq j_1 < j_2 < ... < j_m \leq N \)

\[ P(A) := p_{j_1} + p_{j_2} + ... + p_{j_m} \]
Ex 2  The smallest σ-algebra containing all the open subsets of IR^n is called the Borel σ-algebra, denoted B (or B(IR^n))

\[ \int_{IR^n} f(x) \, dx = 1 \]

Define
\[ IP[B] := \int_B f(x) \, dx \quad \forall B \in B \]

Then (IR^n, B, IP) is a probability space

\[ p: \text{density of the probability measure } IP \]

Ex 3  Fix \( z \in IR^n \)

\[ IP(B) = 1 \quad \text{if } z \in B \quad \forall B \in B \]

(\( IR^n, B, IP \)): probability space

\[ IP: \text{Dirac mass centered at } z \]

\[ IP = \delta_z \]
Ex: Buffon's needle problem

The plane is ruled by parallel lines 2 inches apart and a 1-inch long needle is dropped at random on the plane.

What is the probability that it hits one of the parallel lines?

\[ h = \text{distance from the center of needle to nearest line} \]
\[ \theta = \text{angle (} \leq \frac{\pi}{2} \text{) that the needle makes with the horizontal} \]

\[ \Omega = \left[ 0, \frac{\pi}{2} \right) \times \left[ 0, h \right] \]

values of \( \theta \) values of \( h \)

\( \mathcal{U} \): Borel subset of \( \Omega \)

\[ P(B) = \frac{2 \cdot \text{area of } B}{\pi} \quad \forall B \in \mathcal{B} \]

\( A := \text{needle hits a horizontal line} \)

\[ (\theta, h) \in A \iff h \leq \sin \theta \cdot \frac{1}{2} \]

\[ A = \{ (\theta, h) \in \Omega \mid h \leq \frac{\sin \theta}{2} \} \]

\[ P(A) = \frac{2}{\pi} \int_0^{\pi/2} \frac{1}{2} \sin \theta \, d\theta = \frac{1}{\pi} \]
I.3 Random Variables

I.3.1 Definition

Example: Suppose that our experiment consists of seeing how long a battery can operate before wearing down. Suppose also that we are not primarily interested in the actual lifetime of the battery but are concerned only about whether or not the battery lasts at least 2 years.

Ω: lifetime of the battery

or chemical composition, room temperature, pressure, moisture, in position

prob. space: most an essential mathematical construct not directly observable

introduce mappings from Ω to \( \mathbb{R}^n \) the values of which we can observe

battery: random variable \( X \)

\[
X = \begin{cases} 
1 & \text{if lifetime } \geq 2 \text{ years} \\
0 & \text{otherwise} 
\end{cases}
\]
Def. I.3.1.1

Let \((\Omega, \mathcal{U}, \mathbb{P})\) be a probability space. A mapping \(X: \Omega \to \mathbb{R}\) is called an \(n\)-dimensional random variable if for each \(B \in \mathcal{B}\) we have \(X^{-1}(B) \in \mathcal{U}\).

We equivalently say \(X\) is \(\mathcal{U}\)-measurable.

Notation: We write \(X\) for \(X(\omega)\) and \(\mathbb{P}(X \in B)\) for \(\mathbb{P}(X^{-1}(B))\).

Ex1: Let \(A \in \mathcal{U}\). Let \(X_A(\omega) = 1\) if \(\omega \in A\) and \(0\) if \(\omega \notin A\). The indicator function of \(A\) it is a random variable.
Ex 2: If \( A_1, A_2, \ldots, A_m \in \mathcal{U} \) with \( \Omega = \bigcup_{i=1}^{m} A_i \) and \( a_1, a_2, \ldots, a_m \in IR \), then \( X = \sum_{i=1}^{m} a_i X_{A_i} \) is a random variable called a "simple" function.

Ex: Let \( X \) denote the random variable that is defined as the sum of two fair dice; then

\[
\begin{align*}
\mathbb{P}(X = 2) &= \mathbb{P}[(1, 1)] = \frac{1}{36} \\
\mathbb{P}(X = 3) &= \mathbb{P}[(1, 2), (2, 1)] = \frac{2}{36} \\
\mathbb{P}(X = 4) &= \mathbb{P}[(1, 3), (2, 2), (3, 1)] = \frac{3}{36} \\
\mathbb{P}(X = 5) &= \mathbb{P}[(1, 4), (2, 3), (3, 2), (4, 1)] = \frac{4}{36} \\
\vdots
\end{align*}
\]
Example: The experiment consists of tossing two fair coins.

- \( Y \): number of heads appearing
- \( Y \): a random var. taking one of the values 0, 1, 2

\[
\begin{align*}
P(Y = 0) &= P((T, T)) = \frac{1}{4} \\
P(Y = 1) &= P((T, H), (H, T)) = \frac{1}{2} \\
P(Y = 2) &= P((H, H)) = \frac{1}{4}
\end{align*}
\]

\[
P(Y = 0) + P(Y = 1) + P(Y = 2) = 1
\]

Lemma I.31.2

Let \( X : \Omega \to \mathbb{R}^n \) be a random variable.

Then \( \mathcal{U}(X) := \{ X^{-1}(B) \mid B \in \mathcal{B} \} \)

is a \( \sigma \)-algebra, called the \( \sigma \)-algebra generated by \( X \).

This is the smallest sub-\( \sigma \)-algebra of \( \mathcal{U} \) with respect to which \( X \) is measurable.

Notation: \( \mathcal{U}(X) \leftrightarrow \mathcal{O}(X) \)
Ex 6: 6-faced die
\[ \Omega = \{1, 2, 3, 4, 5, 6\} \]
\[ \mathcal{U}: \text{set of all subsets of } \Omega \]
\[ X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is odd} \\ 0 & \text{otherwise} \end{cases} \]
\[ \mathcal{U}(X) = \{ \emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\} \} \]

Remark
\[ \mathcal{U}(X) \text{ can be interpreted as } \text{"(containing) all relevant information" about the random variable } X \]

In particular
\[ Y \text{ is a function of } X \iff Y \text{ is a } \mathcal{U}(X)-\text{measurable} \]

i.e., \[ Y = \phi(X) \]
\( (\phi \text{ "reasonable"} ) \)
1.3.2 Discrete Random Variables

A random variable that can take at most a countable number of possible values is said to be discrete.

\[ X: \Omega \rightarrow \{x_1, \ldots, x_n, \ldots\} \]

\[ p: \{x_1, \ldots, x_n, \ldots\} \rightarrow [0, 1] \]

probability mass function of \( X \)

\[ \sum_{i=1}^{\infty} p(x_i) = 1 \]

If \( \{x_1, \ldots, x_n, \ldots\} \) is an ordered set we may define

\[ F: \text{cumulative distribution function of } p \]

\[ F(a) = \sum_{\text{all } x_i \leq a} p(x_i) \]

\[ E_X X / p(1) = \frac{1}{2} \quad p(2) = \frac{1}{3} \quad p(3) = \frac{1}{6} \]

\[ F(a) = \begin{cases} 
0 & a < 1 \\
\frac{1}{2} & 1 \leq a < 2 \\
\frac{5}{6} & 2 \leq a < 3 \\
1 & 3 \leq a 
\end{cases} \]
Expectation of a Discrete Random Variable

Let \( X : \Omega \rightarrow \{ x_1, \ldots, x_n, \ldots \} \) be a discrete random var. \( \{ x_1, \ldots, x_n, \ldots \} \) are real numbers or reals.

Def: If \( \sum_{i=1}^{\infty} p(x_i) |x_i| < \infty \) or \( \sum_{i=1}^{\infty} p(x_i) (-x_i)_+ < \infty \) \((a)_+ = \max(a, 0)\)

then the expected value of \( X \) is defined by

\[
E[X] = \sum_{i=1}^{\infty} x_i p(x_i)
\]

\[
= \sum_{i=1}^{\infty} x_i p(x_i) - \sum_{i=1}^{\infty} \max(-x_i, 0) p(x_i)
\]

Rk: The expectation of a positive random var is always defined but it can be \( \infty \).

\( X_+ = \max(X, 0) \quad X_- = \max(-X, 0) \)

then \( X = X_+ - X_- \)

and \( E[X] = E[X_+] - E[X_-] \)
Variance of a discrete random variable

**Def** If \( \mathbb{E}[|X|] < \infty \) we call

\[
V(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]
\]

the variance of \( X \)

**Rk**

\[
V(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\]

\[
= \sum_{i=1}^{\infty} p(x_i) x_i^2 - \left( \sum_{i=1}^{\infty} p(x_i) x_i \right)^2
\]

**Moment Generating Function**

Assume \( x_i \in \mathbb{R} \)

**Def** We call moment generating

\[
\phi : \mathbb{R} \rightarrow \mathbb{R}^+
\]

\[
t \rightarrow \phi(t) := \mathbb{E}[e^{tX}]
\]

\[
= \sum_{i=1}^{\infty} e^{tx_i} p(x_i)
\]

is the moment generating function of \( X \)
Rk: All the moments of $X$ can be obtained by successively differentiating $\phi(t)$:

$$\phi(t) = \sum_{i=1}^{\infty} x_i e^{t \xi_i} = E[X e^{tx}]$$

1. $\phi'(0) = E[X]$
2. $\phi''(t) = \frac{d}{dt} E[X e^{tx}]$
   $$= E[X^2 e^{tx}]$$
   $$\phi''(0) = E[X^2]$$
3. $\phi'''(0) = E[X^3]$
The Bernoulli Random Variable

\[ X \in \{0, 1\} \]

\[ p(0) = \Pr[X = 0] = 1 - p \]

\[ p(1) = \Pr[X = 1] = p \]

\[ p: 0 \leq p \leq 1 \] is the probability that the/a trial whose outcome \( X \) can be classified as either a "success" or a "failure" is a success.

A trial, or an experiment, whose outcome can be classified as either a "success" or as a "failure" is performed.

\[ X = 1 \iff \text{outcome is a success} \]

\[ X = 0 \iff \text{outcome is a failure} \]

\( p: \) probability of a success

\[ 0 \leq p \leq 1 \]

\[ \mathbb{E}[X] = p \]

\[ \text{Var}(X) = p(1-p) \]

\[ \phi(t) = \mathbb{E}[e^{tX}] = 1 - p + pe^t \]
The Binomial Random Variable

\[ X \in \{0, 1, \ldots, n\} \]

for \( i \in \{0, 1, \ldots, n\} \)

\[ p(i) = \binom{n}{i} p^i (1-p)^{n-i} \]

\[ \binom{n}{i} = \frac{n!}{i! (n-i)!} \]

\( \binom{n}{i} \): number of different groups of \( i \) objects that can be chosen from a set of \( n \) objects

\[ \sum_{i=0}^{n} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1 \]

\( X = \) # of successes in \( n \) "independent" Bernoulli trials

Distribution law

\( Y_1, \ldots, Y_n : \) Bernoulli, indep., \( \text{IP}[Y_i = 1] = p \)

\( Y_1 + \ldots + Y_n : \) # of heads after tossing \( n \) times a loaded coin

\( \text{IP}[X = i] = \text{IP}[Y_1 + \ldots + Y_n = i] \)

We say that \( X \) has a binomial distribution with parameters \( (n, p) \)
$E[X] = np$

$\text{Var}[X] = np(1-p)$

$E[e^{tX}] = (1-p + pe^t)^n$

$\left( \sum_{i=0}^{n} \binom{n}{i} e^{ti} p^i (1-p)^{n-i} \right)$
Ex. It is known that any item produced by a certain machine will be defective with probability 0.1 independently of any other item. What is the probability that in a sample of 3 items at most 1 will be defective?

Sol: \( X: \# \text{ of defective items} \)

\[ X \text{ binomial r.v. r.v. with parameters } (3, 0.1) \]

\[ P(X = 0) + P(X = 1) = \binom{3}{0} (0.1)^0 (0.9)^3 + \binom{3}{1} (0.1)^1 (0.9)^2 = 0.972 \]

Ex. Suppose that an airplane engine will fail when in flight with prob. \( 1 - p \) independently from engine to engine; suppose that the airplane will make a successful flight if at least 50% of its engines remains operative. For what value of \( p \) is a four-engine plane preferable to a two-engine plane?
Sol: Each engine is assumed to fail or function independently of what happens with the other engines.

\[ X \rightarrow \text{# of engines remaining operative} \]

Binomial random var.

Prob. that a four-engine plane makes a successful flight:

\[
\binom{4}{2} p^2 (1-p)^2 + \binom{4}{3} p^3 (1-p) + \binom{4}{4} p^4 (1-p)^0
\]

\[
= 6p^2(1-p)^2 + 4p^3(1-p) + p^4
\]

Corresponding prob. for a two-engine plane:

\[
\binom{2}{1} p (1-p) + \binom{2}{2} p^2 = 2p(1-p) + p^2
\]

A four-engine plane is safer if

\[
6p^2(1-p)^2 + 4p^3(1-p) + p^4 \geq 2p(1-p) + p^2
\]

\[
\Rightarrow 6p(1-p)^2 + 4p^2(1-p) + p^3 \geq 2 - p
\]

\[
\Rightarrow 3p^3 - 8p^2 + 7p - 2 \geq 0
\]

\[
\Rightarrow (p-1)^2(3p-2) \geq 0
\]

\[
\Rightarrow 3p - 2 \geq 0
\]

\[
\Rightarrow p \geq \frac{2}{3}
\]
Suppose that a particular trait of a person (such as eye color or left handedness) is classified on the basis of one pair of genes and suppose that D represents a dominant gene and d a recessive gene.

Thus, a person with DD genes is pure dominance, one with dd is pure recessive, and one with Dd is hybrid.

The pure dominance and the hybrid are alike in appearance. Children receive one gene from each parent. If both parents are of two hybrid parents, have a total of 4 children, what is the probability that exactly 3 of the 4 children have the outward appearance of the dominant gene.

\[ p(DD) = \frac{1}{4} \]
\[ p(Dd) = \frac{1}{2} \]
\[ p(dd) = \frac{1}{4} \]

The number of children with the outward appearance of the dominant gene is binomial \((4, \frac{3}{4})\).

\[ p(X=3) = \binom{4}{3} \left( \frac{3}{4} \right)^3 \left( \frac{1}{4} \right) = \frac{27}{64} \]