

# Honors Statistical Inference I - APMA 1655

Milan Capoor

Spring 2023

## 1 Lecture 1, Jan 25: Random outcomes & sample spaces

Features of random events:

- There is more than one possible outcome.
- Before doing or observing the experiment of interest, you do not know which outcome you will see.
- Some possible outcomes are likely, and some other outcomes are quite unlikely e.g., winning one billion dollars by buying a lottery ticket.

### Sample Spaces

*Sample space:* the set of all possible outcomes or results of that experiment (usually denoted  $\Omega$ ) Examples:

- Coin toss ( $\Omega = \{H, T\}$ )
- Schrodinger's cat ( $\Omega = \{\text{alive}, \text{dead}\}$ )
- The lifespan of a tree ( $\Omega = \{1, 2, \dots, 100, \dots\} = \mathbb{Z}_+$ )

Also note that not all elements/subsets of  $\Omega$  are equally likely - hence, "probability"

## 2 Lecture 2, Jan 27: Events, event operations, and infinite operations

Suppose  $\Omega$  is a sample space. Then:

- *Event*: each subset  $E$  of  $\Omega$
- *Impossible event*: the empty set  $\emptyset$

Example: Final exam scores

1. The sample space:  $\Omega = \{0, 1, 2, \dots, 100\}$
2. The event "score is higher than 50":  $E = \{51, 52, \dots, 100\} \subset \Omega$
3. The event "score is a negative number":  $\emptyset$

## Set/Event Operations

Suppose  $\Omega$  is a sample space and  $A$  and  $B$  are events  $\{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$ :

- *Intersection*: both  $A$  and  $B$  occur; the collection of elements that are in sets  $A$  AND  $B$

$$A \cap B$$

- *Union*: either  $A$  or  $B$  occurs; the collection of elements in  $A$  or  $B$

$$A \cup B$$

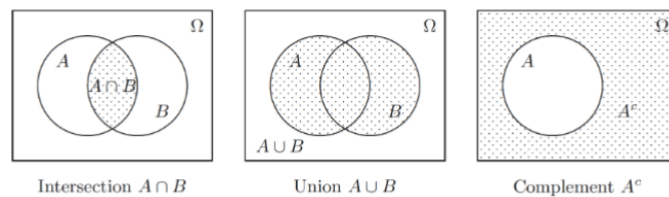
- *Complement*: the collection of elements that are not in  $A$ ; the opposite event of  $A$

$$A^c$$

Note that

$$\Omega^c = \emptyset$$

$$\emptyset^c = \Omega$$



*De Morgan's Laws*: for any two events  $A$  and  $B$  we have the following

$$(A \cup B)^c = A^c \cap B^c \quad (1)$$

$$(A \cap B)^c = A^c \cup B^c \quad (2)$$

## Infinite Sets

Suppose  $A_1, A_2, A_3, \dots, A_n, A_{n+1}, \dots$  are events. Some of them may be identical and some of them may be empty

Infinite Operations:

- *Infinite intersection*: the collection of events that are in ALL the sets  $A_1, \dots, A_n$ ; i.e. "all the events  $A_n$  for  $n = 1, 2, \dots$  happen"

$$\bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \forall n = 1, 2, 3, \dots\}$$

- *Infinite union*: the collection of elements in at least one of the sets; "at least one of these events happen"

$$\bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega : \exists n' | \omega \in A_{n'}\}$$

("there exists at least one  $n'$  such that omega is in the set")

## 3 Lecture 3, Jan 30: Probability space & properties of probability

*Disjoint*: for two events A and B, they are disjoint if  $A \cap B = \emptyset$

*Mutually disjoint*: if all pairwise intersections of  $A_1, A_2, \dots, A_n$  are empty ( $A_n \cap A_m = \emptyset$  if  $n \neq m$ )

*Definition of Probability*: from the following definition we can derive everything in probability theory.

Let  $\Omega$  be a sample space. Suppose  $\mathbb{P}$  is a real-valued function of subsets of  $\Omega$

$$\mathbb{P} : \{\text{subsets of } \Omega\} \rightarrow \mathbb{R}, \quad A \mapsto \mathbb{P}\{A\}$$

where A is an input and  $\mathbb{P}A$  is the corresponding output. If  $\mathbb{P}$  satisfies the following three axioms, the pair  $(\Omega, \mathbb{P})$  is a *probability space*

1.  $\mathbb{P}(A) \geq 0$  for any subset  $A \subset \Omega$  (the probability of an event must be non-negative)

2.  $\mathbb{P}(\Omega) = 1$

3. For any sequence of disjoint subsets  $\{A_i\}_{i=1}^{\infty}$  (i.e.  $A_i \cap A_j = \emptyset$ ) we have

$$\mathbb{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

The map  $\mathbb{P}$  is called a *probability*

We can define this as a specific function

$$\mathbb{P}(A) := \frac{\#A}{n} \quad A \subset \Omega$$

where  $\#A$  is the number of elements in  $A$  and  $\Omega = \{1, 2, \dots, n\}$  with a large  $n$ .

## Lecture 4, Feb 1: Properties of Probability

Let  $(\Omega, \mathbb{P})$  be a probability space. Then,

1.  $\mathbb{P}(\emptyset) = 0$

Note: while this implies that the probability of an impossible event is 0, there can be zero-probability events which are not themselves impossible

2. if two events  $E_1$  and  $E_2$  satisfy  $E_1 \cap E_2 = \emptyset$ , then

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$$

3. if  $A, B \subset \Omega$  and  $A \subset B$ , then  $\mathbb{P}(A) \leq \mathbb{P}(B)$

(Intuitively, if  $A$  happens,  $B$  must also happen so  $B$  is more likely)

4.  $0 \leq \mathbb{P}(A) \leq 1$  for  $A \subset \Omega$

5.  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

6. for any  $A, B \subset \Omega$

$$\mathbb{P}\{A \cup B\} = \mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \cap B\}$$

7. for any countable collection of subsets

$$\mathbb{P}\left\{\bigcup_{n=1}^{\infty} A_n\right\} \leq \sum_{n=1}^{\infty} \mathbb{P}\{A_n\}$$

Note: the equality is obvious from axiom three in the case where all events are mutually disjoint. in the case of intersections, though, rule 6 must be generalized to account for overlap, hence the less than or equal to

## 4 Lecture 5, Feb 3: Conditional Probability

### Part I - Motivating Problem

We know that a family has two children.

$$\begin{aligned}\Omega &= \{(g, g), (b, b), (g, b), (b, g)\} \\ \mathbb{P}(A) &= \frac{\#A}{\#\Omega} = \frac{\#A}{4} \quad A \subset \Omega\end{aligned}$$

Event 1:  $A = \{(g, g), (b, g), (g, b)\}$  ("at least one is girl")

$$\mathbb{P}(A) = \frac{3}{4}$$

Now suppose we get further information that the family has at least one boy:

$B = \{(b, g), (b, b), (g, b)\}$ :

$$\mathbb{P}(A|B) = \frac{\#(A \cap B)}{\#B} = \frac{\#\{(b, g), (g, b)\}}{\#\{(b, g), (b, b), (g, b)\}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{2}$$

("knowing that event B occurs, what is the updated likelihood of A?")

### Part II - A more rigorous definition

Let  $A, B \subset \Omega$  such that

1. if  $\mathbb{P}(B) > 0$  we call the following "the *conditional probability of A given B*"

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

2. otherwise, if  $\mathbb{P}(B) = 0$  then  $\mathbb{P}(A|B)$  is not well defined in the scope of this course (see real analysis)

**Theorem:** Let  $(\Omega, \mathbb{P})$  be a probability space. Suppose  $B \subset \Omega$  and  $\mathbb{P}(B) > 0$ . Then let

$$\tilde{\mathbb{P}}(A) := \mathbb{P}(A|B) \quad \forall A \subset \Omega$$

Then,  $\tilde{\mathbb{P}}$  is another probability defined on  $\Omega$  such that  $\tilde{\mathbb{P}}$  also satisfies the 3 axioms

### Part III - Properties of Conditional Probabilities

Assuming  $\mathbb{P}(B) > 0$ , and  $A, B \subset \Omega$ :

1. Multiplication Law

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$$

2. Let  $B_1, B_2, \dots, B_n \subset \Omega$ . We say the events provide a *partition* of  $\Omega$  if they are mutually disjoint and they satisfy<sup>1</sup>

$$\bigcup_{i=1}^n B_i = \Omega$$

3. *The law of total probability*

Let  $B_1, B_2, \dots, B_n$  be events and they provide a partition of  $\Omega$  and  $\mathbb{P}(B_i) > 0$ . Then,

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

**Proof:**

$$\Omega = \bigcup_{i=1}^n B_i \implies A = A \cap \Omega$$

Then, by the laws above,  $A \cap B_1, A \cap B_2, \dots, A \cap B_n$  is mutually disjoint. So

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^n A \cap B_i\right) = \sum_{i=1}^n \mathbb{P}(A \cap B_i)$$

---

<sup>1</sup>Note that for  $B_i = \{B_1, \dots, B_n\}$ ,

$$A \cap \left(\bigcup_{i=1}^n B_i\right) = A \cap (B_1 \cup B_2 \cup \dots \cup B_n) = \bigcup_{i=1}^n A \cap B_i$$

Then from the multiplication law,  $\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$  and the desired result follows

4. Corollary: Let  $B$  be an event with  $0 < \mathbb{P}(B) < 1$ . Then we have

$$\mathbb{P}(A) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c)$$

**Proof:** Let  $B_1 = B$ ,  $B_2 = B^c$ ,  $n = 2$ .  $B_1$  and  $B_2$  obviously partition  $\Omega$  and

$$\mathbb{P}(B_1) = \mathbb{P}(B) > 0, \quad \mathbb{P}(B_2) = \mathbb{P}(B^c) = 1 - \mathbb{P}(B) > 0$$

Then the corollary follows from the law of total probability

## 5 Lecture 6, Feb 6: Bayes' formula

### Part I - Bayes' Rule

Suppose  $B_1, \dots, B_n$  provide a partition of  $\Omega$ . In addition,  $\mathbb{P}(B_i) > 0 \quad \forall i \in [1, n]$ . Let  $A$  be any event such that  $\mathbb{P}(A) > 0$ . Then,

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)} \quad i = 1, 2, \dots, n$$

**Proof:** The definition of conditional probability implies

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)}$$

The multiplication law implies

$$\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

The law of total probability implies

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

Then the desired result follows.

Though this proof is trivial, the formula is quite meaningful in that it allows an exchange between the conditions and results. Note that this is not the general Bayes formula.

## 6 Lecture 7, Feb 8: Independence

Independent events do not affect each other's outcomes (e.g. flipping two coins). That is

$$\begin{cases} \mathbb{P}(A|B) = \mathbb{P}(A) \\ \mathbb{P}(B|A) = \mathbb{P}(B) \end{cases}$$

In other words, knowing A does not help predict B.

Together with those equations and the multiplication law, we have

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

(for independent events)

**Definition:** Let  $(\Omega, \mathbb{P})$  be a probability space.

1. Suppose A and B are two events. They are *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

2. Suppose  $A_1, A_2, \dots, A_n$  are a sequence of events. The sequence is *mutually independent* if

$$\mathbb{P}(A_m \cap A_n) = \mathbb{P}(A_m) \cdot \mathbb{P}(A_n) \quad m \neq n$$

**Theorem:** Let  $(\Omega, \mathbb{P})$  be a probability space with  $A, B \subset \Omega$  such that  $\mathbb{P}(A) > 0$  and  $\mathbb{P}(B) > 0$  (this is necessary to have the conditional probabilities well-defined). Then the following three equations are equivalent:

1.  $\mathbb{P}(A|B) = \mathbb{P}(A)$
2.  $\mathbb{P}(B|A) = \mathbb{P}(B)$
3.  $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Note that we choose this third equation to be the definition of independence stated above specifically because it does not depend on the positive probability condition of the other two

**Example:** Suppose a family has 3 children of unknown gender

$$\Omega = \{(g, g, g), (g, g, b), (g, b, g), (b, g, g), (g, b, b), (b, g, b), (b, b, g), (b, b, b)\}$$



$$\#\Omega = 2^3 = 8 \implies \mathbb{P}(A) := \frac{\#A}{8}$$

Consider the events A ("the family has boys and girls") and B ("the family has at most one girl")

Question: Are A and B independent?

$$\begin{aligned}\mathbb{P}(A) &= \frac{6}{8} \\ \mathbb{P}(B) &= \frac{4}{8} \\ \mathbb{P}(A) \cdot \mathbb{P}(B) &= \frac{3}{8} \\ \mathbb{P}(A \cap B) &= \frac{3}{8} = \mathbb{P}(A) \cdot \mathbb{P}(B)\end{aligned}$$

So the events are independent

## 7 Lecture 8, Feb 10: Random Variable

Probability theory has 3 building blocks:

1. Sample space ( $\Omega$ )
2. Probability ( $\mathbb{P}$ ) Together, these two give us probability space  $(\Omega, \mathbb{P})$
3. Random Variable

**Motivating Example:** Let  $\Omega$  be the collection of all undergraduate students at Brown. Then  $\mathbb{P}(A) := \frac{\#A}{\#\Omega}$ . And, for each student  $\omega \in \Omega$ ,

$$X(\omega) = \text{the SAT score of the given student} \quad \{X : \Omega \rightarrow \mathbb{R}, \omega \mapsto X(\omega)\}$$

but where  $\omega$  is unknown (say to protect anonymity) so it can be any value in  $\Omega$ . Note, however, that if  $\omega$  is unknown, then  $X$  is also uncertain.

**Definition:** Let  $(\Omega, \mathbb{P})$  be a probability space. Suppose  $X$  is a real-valued function defined on  $\Omega$ ,

$$\begin{aligned}X &: \Omega \rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega)\end{aligned}$$

We thus call  $X$  a *random variable*

## 8 Cumulative Distribution Functions

**Motivating Example:**  $\Omega$  is all the undergrads at Brown,  $\omega \in \Omega$  is a student at Brown,  $X(\omega)$  is a random variable denoting the SAT score of a Brown student

Let  $A_{100}$  be the event "the SAT score of a Brown student is  $\leq 100$ ". Then,

$$A_{100} = \{\omega \in \Omega : X(\omega) \leq 100\}$$

**Definition:** Let  $(\Omega, \mathbb{P})$  be a probability space and  $X$  be a random event. Then for any real number  $x \in \mathbb{R}$ , we define the event  $A_x$  by

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\}$$

We define a real-valued function  $F$  on  $\mathbb{R}$  by

$$F(x) := \mathbb{P}(A_x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

This function  $F$  is the *cumulative distribution function (CDF)* of the random variable  $X$ , usually written  $F_X$

## 9 Lecture 9, Feb 13: Cumulative distribution function

### Part I - Review

A random variable  $X$  is a function defined on a sample space  $\Omega$

For each real number  $x$ , we define an event

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\}$$

We then define a function by

$$F_X : \begin{array}{l} \mathbb{R} \rightarrow [0, 1] \\ x \mapsto \mathbb{P}(A_x) \end{array}$$

Note that  $F_X = \mathbb{P}(X \leq x)$  and this function is called *the cumulative distribution function*

## Part II - The CDF

**Example 1:** Flipping a coin

- $\Omega = \{H, T\}$
- $\mathbb{P}(A) := \frac{\#A}{\#\Omega} = \frac{\#A}{2} \quad (A \subset \Omega)$
- 

$$\begin{cases} X(H) = 1 \\ X(T) = 0 \end{cases}$$

*Claim:* the CDF of X is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

*Proof:*

1. When  $x < 0$ , then

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\} = \emptyset$$

(this is impossible because X is only 0 or 1 so not negative). So  $F_X(x) = \mathbb{P}(A_x) = 0$

2. When  $0 \leq x < 1$ ,

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\} = \{\omega \in \Omega : X(\omega) = 0\} = \{T\}$$

$$\text{so } F_X(x) = \mathbb{P}(\{T\}) = \frac{1}{2}$$

3. When  $x \geq 1$

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\} = \{\omega \in \Omega : X(\omega) \leq 1\} = \{T, H\}$$

$$\text{so } P(A_x) = 1$$

**Example 2:** Flipping a biased coin

- $\Omega = \{H, T\}$
- 

$$\mathbb{P}(A) := \begin{cases} p & A = \{H\} \\ 1 - p & A = \{T\} \end{cases}$$

for some  $p \in [0, 1]$

- $X(H) = 1, X(T) = 0$

*Claim:* the CDF here is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

*Proof:* same as above

Note: this example actually refers to the Bernoulli Distribution **Definition:** Let  $X$  be a random variable on  $(\Omega, \mathbb{P})$  If the CDF of  $X$  is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

for some  $p \in [0, 1]$ , we say " $X$  follows the Bernoulli Distribution with success probability  $p$ " and denote  $X \sim \text{Bernoulli}(p)$

**Theorem:** Let  $(\Omega, \mathbb{P})$  be a probability space. Suppose  $X$  is a random variable defined on  $\Omega$  and its CDF is  $F_X$

$F_X$  is non-decreasing ( $F_X(x_1) \leq F_X(x_2)$   $x_1 \leq x_2$ ) for any CDF.

**Proof:**

$$\begin{aligned} A_{x_1} &= \{\omega \in \Omega : X(\omega) \leq x_1\} \\ A_{x_2} &= \{\omega \in \Omega : X(\omega) \leq x_2\} \\ x_1 &\leq x_2 \implies A_{x_1} \subset A_{x_2} \\ F_X(x_1) &= \mathbb{P}(A_{x_1}) \leq \mathbb{P}(A_{x_2}) \leq F_X(x_2) \end{aligned}$$

## Lecture 10, Feb 15:

### Part I - Review

Let  $X$  be a random variable defined on a probability space  $(\Omega, \mathbb{P})$ . For any real number  $x$ ,

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\}$$

Then the cumulative distribution function of  $X$  is

$$F_X(x) = \mathbb{P}(A_x)$$

## Part II - Properties of the CDF

Let  $X$  be any random variable on any probability space  $(\Omega, \mathbb{P})$  with  $F_X(x)$  as the corresponding CDF

1. Any CDF is non-decreasing ( $F_X(x_1) \leq F_X(x_2) \quad x_1 \leq x_2$ )

2.

$$\lim_{x \rightarrow -\infty} F_X(x) = 0$$

3.

$$\lim_{x \rightarrow \infty} F_X(x) = 1$$

(Note the rigorous proof needs real analysis)

4.  $F_X$  is right-continuous, i.e.

$$\forall x_0 \in \mathbb{R}, \quad F_X(x_0) = \lim_{x \rightarrow x_0^+} F_X(x)$$

Note that the "right-continuous" property is implied from the  $\leq$  sign. With a strict less-than inequality, the CDF becomes left-continuous

5. For any  $x_0 \in \mathbb{R}$ ,

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = x_0\}) = F_X(x_0) - \lim_{x \rightarrow x_0^-} F_X(x)$$

Note that this is zero if the CDF is continuous

## Lecture 11, Feb 17

### Part I - Review

The building blocks of probability together define the CDF:

$$\begin{cases} \Omega \\ \mathbb{P} \\ X \end{cases} \implies F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

## Part II - Moving backwards

**Question:** Given a function  $F$  (satisfying some conditions), do there exist a sample space, a probability, and a random variable corresponding to the given  $F$ ?

**Theorem:** Suppose we have a  $F : \mathbb{R} \rightarrow [0, 1]$  satisfying

- $F$  is non-decreasing
- With end behavior

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1$$

- $F$  is right-continuous

Then, there exist a sample space, a probability, and a random variable such that

$$F(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

**Proof:** far beyond the scope of this course

## Part III - Classification of Random Variables

Types of random variables:

- Continuous
- Discrete
- Neither continuous nor discrete

**Definition:** a function  $F$  is continuous if

$$\lim_{x \rightarrow x_0^-} F(x) = \lim_{x \rightarrow x_0^+} F(x) = F(x_0) \quad \forall x_0$$

**Definition:** a random variable  $X$  is a continuous random variable if the CDF  $F_X : \mathbb{R} \rightarrow [0, 1]$  is a continuous function

**Theorem:** Let  $X$  be a random variable on  $(\Omega, \mathbb{P})$ . If  $X$  is a continuous random variable,

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = x_0\}) = 0 \quad \forall x_0 \in \mathbb{R}$$

**Proof:**

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = x_0\}) = F_X(x_0) - \lim_{x \rightarrow x_0^-} F_X(x) = F_X(x_0) - F_X(x_0) = 0$$

## Lecture 12, Feb 22: Continuous Random Variables

### Part I - A “theorem”

**“Theorem”:** Let  $F_X$  be the CDF of a continuous random variable  $X$ . Then,  $F_X$  is differentiable.

**Remark:** the true and rigorous version of this “theorem” requires lots of pure math so this version does have some edge cases such that “differentiable” really means “piecewise differentiable”

*Example:*

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

This is not technically differentiable because of the sharp points at  $x = \{0, 1\}$  but we can use a generalized derivative to cheat:

$$F'(x) = \begin{cases} 0 & x < 0 \text{ or } x > 1 \\ 1 & 0 < x < 1 \\ k & x = 0 \text{ or } x = 1 \end{cases}$$

where  $k$  is any value whatsoever.

**Definition:** Let  $F_X$  be the CDF of a continuous random variable  $X$ .

$$p_X(x) = F'_X(x)$$

We call  $p_X(x)$  the *probability density function (PDF)* of the random variable  $X$

### Part II - The Rigorous Treatment (optional)

**Definition:**  $X$  is a continuous random variable if  $F_X$  is absolutely continuous

**Theorem:**  $X$  is a continuous random variable. Then its CDF  $F_X$  is differentiable “almost everywhere with respect to the Lebesgue measure”

**Definition:**  $p_X(x) = F'_X(x)$  is the probability density function for a continuous random variable  $X$

## Part III - The Probability Density Function

Less rigorously,

$$p_X(x) = F'_X(x)$$

is the probability density function and is the piecewise derivative of  $F_X$  for the continuous random variable  $X$

**Theorem:** Let  $X$  be a continuous RV with CDF  $F_X(x)$  and PDF  $p_X(x)$ . Then,

$$F_X(x) = \int_{-\infty}^x p_X(t) dt$$

**Remarks:**

- $p_X(x) := F'_X(x)$  so CDF determines PDF
- $F_X(x) = \int_{-\infty}^x p_X(t) dt$  so PDF determines CDF

## Part IV - Examples of continuous random variables

**Definition:** Let  $X$  be a RV. If the CDF of  $X$  is

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

(for  $a < b$ ), then  $X$  follows the uniform distribution between  $a$  and  $b$ .

This is continuous so we can take the piecewise derivative as follows:

$$p_X(x) = \begin{cases} 0 & x < a \text{ or } x > b \\ \frac{1}{b-a} & a \leq x \leq b \end{cases}$$

Notice that the graph of this function is a rectangle with base  $b - a$  and height  $1/(b - a)$ .

**Experiment:** Randomly select a number between 0 and 1

- $\Omega = (0, 1)$
- $X : \Omega \rightarrow \mathbb{R}, \quad X(\omega) = \omega, \quad \omega \in \Omega = (0, 1)$
- Because we randomly select numbers, we assume  $X \sim \text{Unif}(0, 1)$



- Let  $E = \{0.5\} = \{\omega \in \Omega : X(\omega) = 0.5\} = \emptyset$  be the event “the selected number is exactly 0.5”

So,

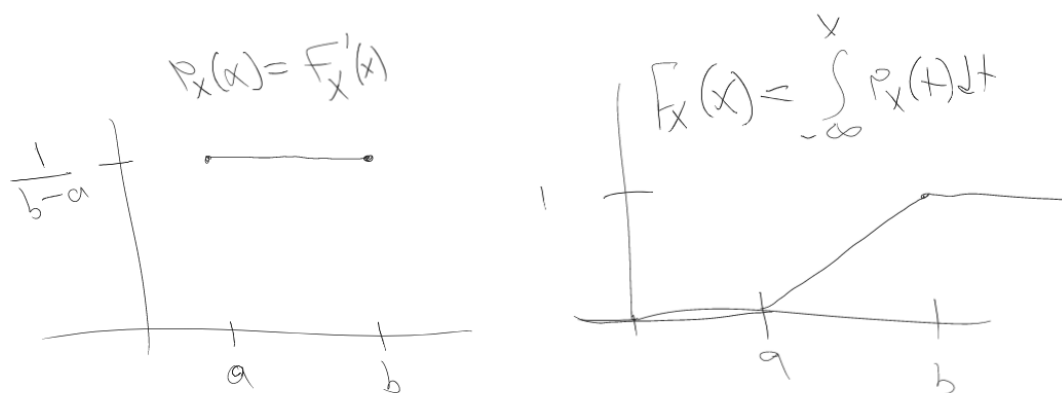
$$\mathbb{P}(E) = \mathbb{P}(X = 0.5) = F_X(0.5) - \lim_{x \rightarrow 0.5^-} F_X(x) = 0$$

because  $F_X$  is continuous so the limit is equal to the value

## Lecture 13, Feb 24:

### Part I - Review

- $X$  is a continuous random variable iff  $F_X$  is a continuous function
- $p_X(x) = F'_X(x)$  is the “probability density function” of  $X$
- The following are the graphs of the PDF and CDF of the random variable  $X \sim \text{Unif}(a, b)$   $a < b$



- “Impossible” implies “zero probability” but zero probability does not imply impossible

### Part II - The Probability Density Function

**Theorem:** Let  $X$  be a continuous random variable and its PDF is  $p_X(x)$ . Then

$$\int_{-\infty}^{\infty} p_X(t) dt = 1$$

**Proof:**

$$\begin{aligned}\int_{-\infty}^{\infty} p_X(t) dt &= \lim_{x \rightarrow +\infty} \int_{-\infty}^x p_X(t) dt \\ &= \lim_{x \rightarrow +\infty} F_X(x) = 1 \quad \blacksquare\end{aligned}$$

**Interlude:** Applications of the above law

- Bayesian statistics

$$f(\theta|x) = \frac{f(x, \theta)}{\int_{-\infty}^{\infty} f(x, \theta) d\theta}$$

- Quantum mechanics For each fixed  $t$ ,  $|\psi(x, t)|^2$  is the PDF of  $x$  so

$$\int_{-\infty}^{\infty} |\psi(x, t)|^2 dx = 1$$

which is the “normalization condition of the Schrodinger equation”

## Part III - Normal Distributions

**Definition:** Let  $X$  be a RV. We say “ $X$  follows the normal distribution with mean  $\mu$  and variance  $\sigma^2$ ” (denoted  $X \sim N(\mu, \sigma^2)$ ), if

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

additionally,

$$p_X(x) = F'_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

## Lecture 14, Feb 27: Discrete Random Variables

### Part I - Preparation

**Definition:** Let  $A$  be a subset of  $\mathbb{R}$ . Then the *indicator function* of  $A$  is

$$\mathbf{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

**Example 1:** If  $A = [0, +\infty)$ , then

$$\mathbf{1}_A(x) = \mathbf{1}_{[0, +\infty)}(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

**Example 2:** If  $A = [\sqrt{2}, +\infty)$ , then

$$\mathbf{1}_A(x) = \mathbf{1}_{[\sqrt{2}, +\infty)}(x) = \mathbb{1}_{[0, +\infty)}(x - \sqrt{2}) = \begin{cases} 1 & x \geq \sqrt{2} \\ 0 & x < \sqrt{2} \end{cases}$$

Also note that for  $X \sim \text{Bernoulli}(\frac{1}{2})$ ,

$$\begin{aligned} F_X(x) &= \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases} \\ &= \frac{1}{2} \mathbf{1}_{[0, +\infty)}(x) + \mathbf{1}_{[1, +\infty)}(x) \\ &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) \end{aligned}$$

Remember that the random variable  $X$  that conforms to the bernoulli distribution can only take 2 values which thus map to the two left endpoints 0 and 1 of the indicator functions.

More generally, say  $X$  takes values in  $\{x_k\}_{k=1}^K = \{x_1, x_2, \dots, x_K\}$  so the CDF can also be written

$$F_X(x) = \sum_{k=1}^K \mathbb{P}(X = x_k) \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

Heuristically, this is the sum of probabilities at each point  $x_k$

## Lecture 15, March 1: Discrete Random Variable Definition

**Definition:** Let  $X$  be a RV.  $X$  is a discrete random variable if its CDF can be written in the form

$$F_X(x) = \sum_{k=1}^K p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

where  $p_k \geq 0$  and  $\sum_{k=1}^K p_k = 1$  as  $p_k$  represents a probability. Also,  $x_k$  is a real number, and  $K$  is allowed to be  $+\infty$ .

Finally, the ordered sequence  $\{p_k\}_{k=1}^K$  is the *probability mass function* of  $X$  and represents the probability of each endpoint event

Remark: that the sequence  $\{p_k\}_{k=1}^K$  is referred to as a “function” is simply a convention as the PMF is the counterpart to the CDF. It may also refer to any function

$$k \mapsto p_k, \quad k \in \mathbb{Z}$$

**Example:**  $X \sim \text{Bernoulli}(p)$

$$F_X(x) = (1 - p) \cdot \mathbf{1}_{[0, +\infty)}(x) + p \cdot \mathbf{1}_{[1, +\infty)}(x)$$

**Remark:** From the CDF

$$F_X(x) = \sum_{k=1}^K p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

we can derive the PMF  $\{p_k\}_{k=1}^K$ . And if we know  $\{x_k\}_{k=1}^K$  we can go from PMF to CDF

**Theorem:** Let  $X$  be a RV defined on  $(\Omega, \mathbb{P})$  and its CDF is

$$F_X(x) = \sum_{k=1}^K p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

then we have

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = x_k\}) = p_k$$

**Proof:** Omitted. The theorem is cheating. To make it true, make some topological assumptions such that  $\{x_k\}_{k=1}^K$  as a subset of the real line is locally finite (or its derived set is  $\emptyset$ )

**Poisson Distribution** ( $X \sim \text{Pois}(\lambda)$ )

“ $X$  follows the Poisson distribution with rate parameter  $\lambda$ ” if  $K = \infty$ ,  $x_k = k \quad \forall k \in [0, \infty)$ ,

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}$$

where that quotient represents the probability that “k events occur in a fixed time interval”

Then,

$$\begin{aligned}\sum_{k=0}^{\infty} p_k &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} e^{\lambda} \quad (\text{taylor series for } e^x) \\ &= 1\end{aligned}$$

Note: the starting index must be 0 for the taylor series to work

## Random Variables that are neither continuous nor discrete

### Part I - Preparation

Let Y and Z be random variables. They are independent if for *any* subsets  $A, B \subset \mathbb{R}$ ,

$$\mathbb{P}(\{\omega \in \Omega : (Y(\omega) \in A) \cap (Z(\omega) \in B)\}) = \mathbb{P}(\{\omega \in \Omega : Y(\omega) \in A\}) \cdot \mathbb{P}(\{\omega \in \Omega : Z(\omega) \in B\})$$

### Part II - Definition

**Example:** Let Y, Z be independent random variables such that

$$\begin{cases} Y \sim \text{Bernoulli}(\frac{1}{2}) \\ Z \sim N(0, 1) \end{cases}$$

Then let

$$X(\omega) = Y(\omega) + (1 - Y(\omega)) \cdot Z(\omega)$$

Claim:

$$F_X(x) = \frac{1}{2} \mathbf{1}_{[0, +\infty)}(x) + \frac{1}{2} F_N(x) = \frac{1}{2} \mathbf{1}_{[0, +\infty)}(x) + \frac{1}{2} \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Then the graph of  $F_X$  is continuous for all x except at  $x = 0$  where there is a jump discontinuity.

Proof:

$$\begin{cases} A_X = \{\omega \in \Omega : X(\omega) \leq x\} \\ B = \{\omega \in \Omega : Y(\omega) = 1\} \end{cases}$$

Then by definition of CDF,

$$F_X(x) = \mathbb{P}(A_x) = \mathbb{P}(A_X|B) \cdot \mathbb{P}(B) + \mathbb{P}(A_X|B^c) \cdot \mathbb{P}(B^c)$$

(by law of total probability)

## Lecture March 3: Independence

### Part I - Independence between two events (Review)

**Definition:** Let  $(\Omega, \mathbb{P})$  be a probability space.  $\tilde{A}, \tilde{B} \subset \Omega$  are two events that are independent if

$$\mathbb{P}(\tilde{A} \cap \tilde{B}) = \mathbb{P}(\tilde{A}) \cdot \mathbb{P}(\tilde{B})$$

### Part II - Independence between two random variables

**Definition Version 1:** Let  $Y$  and  $Z$  be RVs defined on  $(\Omega, \mathbb{P})$ .  $Y$  and  $Z$  are independent if *for any* subsets  $A, B \subset \mathbb{R}$  we define events via  $Y$  and  $Z$  by

$$\begin{aligned} \tilde{A} &= \{\omega \in \Omega : Y(\omega) \in A\} \\ \tilde{B} &= \{\omega \in \Omega : Z(\omega) \in B\} \end{aligned}$$

where  $A$  and  $B$  are independent

**Definition Version 2:** Let  $Y$  and  $Z$  be random variables defined on  $(\Omega, \mathbb{P})$ .  $Y$  and  $Z$  are independent if

$$\mathbb{P}((Y \in A) \cap (Z \in B)) = \mathbb{P}(Y \in A) \cdot \mathbb{P}(Z \in B)$$

*for any*  $A, B \in \mathcal{R}$

Note that

$$\mathbb{P}((Y \in A) \cap (Z \in B)) = \mathbb{P}(\tilde{A} \cap \tilde{B})$$

showing that the two definitions are equivalent

## Part III - Examples

### Example 1 (Intuition):

$Y$  = outcome of Michael flipping a coin

$Z$  = outcome of Taylor Swift flipping a coin

Obviously, these are independent.

### Example 2 (with math):

$Y$  and  $Z$  are two independent RVs defined on  $(\Omega, \mathbb{P})$ . Let

$$Y \sim \text{Bernoulli}\left(\frac{1}{2}\right)$$

$$Z \sim N(0, 1)$$

$$X(\omega) := Y(\omega) + (1 - Y(\omega)) \cdot Z(\omega)$$

Then for a fixed real number  $x$ ,

$$\mathbb{P}((X \leq x) \cap (Y = 0)) = \mathbb{P}((Z \leq x) \cap (Y = 0))$$

by definition of  $X$  and the fixed value of  $Y$ . The above is then equal to

$$\mathbb{P}((Y \in \{0\}) \cap (Z \in (-\infty, x]))$$

where if  $A = \{0\}$  and  $B = (-\infty, x]$ ,

$$= \mathbb{P}(Y \in \{0\}) \cdot \mathbb{P}(Z \in (-\infty, x])$$

$$= \mathbb{P}(Y = 0) \cdot \mathbb{P}(Z \leq x)$$

$$= \frac{1}{2} F_Z(x)$$

## 10 Lecture March 6: Mean/Expected Value

### Part I - Motivation

**Example 1:** Flip a fair coin 1000 times. For the  $i$ -th flip

$$X_i = \begin{cases} 1 & \text{if Heads} \\ 0 & \text{if Tails} \end{cases}$$

Then intuitively, the average value (when the number of flips is large) will be

$$\bar{X}_{1000} = \frac{\sum_{i=0}^{1000} X_i}{1000} \approx \frac{1}{2}$$

or

$$\lim_{n \rightarrow \infty} \bar{X}_n = \frac{1}{2}$$

Then, in general,

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \approx \sum_{k=0}^K x_k \cdot p_k$$

**Example 2:** Office hours are 2-4pm on Mondays. On the i-th Monday,  $X_i$  students come. This can be modelled as

$$X_i \sim \text{Pois}(\lambda)$$

so

$$\bar{X}_n \approx \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-k}}{k!} = \lambda$$

## Part II - Conjecture

**The law of large numbers:** for almost all discrete CDFs and n sufficiently large

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \approx \sum_{k=0}^K x_k \cdot p_k$$

## Part III - Definition

**Discrete Definition:** Let X be a discrete random variable with CDF

$$F_X(x) = \sum_{k=0}^K p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

If

$$\sum_{k=0}^K |x_k| \cdot p_k < +\infty$$



then we call the following sum *the expected value/mean of X*:

$$\mathbb{E}(X) = \sum_{k=0}^K x_k \cdot p_k$$

**Continuous Definition:** Let X be a continuous random variable with PDF  $p_X(x)$ .  
If

$$\int_{-\infty}^{\infty} |x| \cdot p_X(x) dx < +\infty$$

then we call the following the *expected value/mean of X*:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot p_X(x) dx$$

## 11 Lecture March 8

### Part I - Review

**Definition:** Let X be a discrete random variable with CDF

$$F_X(x) = \sum_{k=0}^K p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

1. If  $\sum_{k=0}^K |x_k| \cdot p_k < \infty$ , then we call  $\sum_{k=0}^K x_k \cdot p_k$  the expected value of X denoted  $\mathbb{E}(X)$
2. If  $\sum_{k=0}^K |x_k| \cdot p_k = \infty$  we say  $\mathbb{E}(X)$  does not exist

### Part II - Example

$$K := \infty$$

$$x_k := (-1)^{k+1} \cdot k \quad \forall k = 0, 1, 2, \dots$$

$$p_0 := 0$$

$$p_k := \frac{6}{\pi^2} \frac{1}{k^2} \quad \forall k = 1, 2, \dots$$

Then

$$\sum_{k=0}^K x_k \cdot p_k = \sum_{k=1}^{\infty} (-1)^{k+1} \cdot k \cdot \frac{6}{\pi^2} \cdot \frac{1}{k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$$

from calculus, that series is equal to  $\log 2$  so

$$= \frac{6}{\pi^2} \log 2$$

**Question:** is  $\mathbb{E}(X) = \frac{6}{\pi^2} \log 2$ ? **Answer:** No. The expected value does not exist.

$$\sum_{k=0}^K |x_k| \cdot p_k = \sum_{k=1}^{\infty} k \cdot \frac{6}{\pi^2} \frac{1}{k} = \infty$$

because the sum diverges

### Part III - Theory of Series

**Definition:** Let  $\{a_k\}_{k=0}^{\infty}$  be an ordered series

1. If  $\lim_{n \rightarrow \infty} \sum_{k=0}^n a_k$  exists, we say  $\sum_{k=0}^{\infty} a_k$  is convergent

**Theorem:** If  $\sum_{k=0}^{\infty} |a_k|$  is convergent, then so is

$$\sum_{k=0}^{\infty} a_k$$

**Proof:** Follows from Cauchy's convergence test

**Definition:**

1. If  $\sum_{k=0}^{\infty} |a_k|$  is convergent,  $\sum_{k=0}^{\infty} a_k$  is *absolutely convergent*
2. If  $\sum_{k=0}^{\infty} a_k$  is convergent but  $\sum_{k=0}^{\infty} |a_k|$  is divergent, then  $\sum_{k=0}^{\infty} a_k$  is *conditionally convergent*

**Remark:** This reduces the condition of the expected value definition to "The expected value exists if the sum is absolutely convergent"

**Definition:** Suppose  $\mathbb{N} = \{0, 1, 2, \dots\}$ . Any bijective map (i.e. the map is one-to-one and onto;  $\forall i \in \mathbb{N} \exists$  a unique  $j : \sigma(j) = i$ )  $\sigma$  from  $\mathbb{N} \mapsto \mathbb{N}$  is called a *permutation*

**Theorem:**

1. If  $\sum_{k=0}^{\infty} a_k$  is absolutely convergent,

$$\sum_{k=0}^{\infty} a_{\sigma(k)} = \sum_{k=0}^{\infty} a_k$$

for all permutations  $\sigma$

2. (Riemann series theorem) If  $\sum_{k=0}^{\infty} a_k$  is conditionally convergent, for any  $A \in \mathbb{R} \cup \{-\infty, +\infty\}$ , there exists a permutation  $\sigma$  such that

$$\sum_{k=0}^{\infty} a_{\sigma(k)} = A$$

**Interpretation:** For a conditionally convergent sum, you can select a permutation such that the sum equals any value you want. This also offers a way to define the reals via results of sums of permutations.

**Application to Expected Value:**

$$\sum_{k=0}^{\infty} p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

where “all  $x_k$ ’s are created equal” ( $x_k$  is unordered). Then for an absolutely convergent sum

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} x_k \cdot p_k = \sum_{k=0}^{\infty} x_{\sigma(k)} \cdot p_{\sigma(k)}$$

## Lecture 19, March 10:

### Part I - Review

**Definition:** Let  $X$  be a discrete RV with CDF

$$\sum_{k=0}^K p_k \cdot \mathbf{1}_{[x_k, +\infty)}(x)$$

1. If  $\sum_{k=0}^K |x_k| \cdot p_k < \infty$  then  $\mathbb{E}(X) = \sum_{k=0}^K x_k \cdot p_k$
2. If  $\sum_{k=0}^K |x_k| \cdot p_k = \infty$  then  $\mathbb{E}(X)$  does not exist

**Question:** In the summation  $\sum_{k=0}^{\infty} x_k \cdot p_k = x_0 p_0 + x_1 p_1 + \dots$ , does the order matter?

**Answer:** No! So long as

$$\sum_{k=0}^{\infty} x_k \cdot p_k = \sum_{k=0}^{\infty} x_{\sigma(k)} \cdot p_{\sigma(k)}$$

for all permutations  $\sigma$

1. If  $\sum_{k=0}^{\infty} |x_k| \cdot p_k < \infty$  then the permutation invariance is true and  $\sum_{k=0}^K x_k \cdot p_k < \infty$  is *absolutely convergent*
2. If  $\sum_{k=0}^{\infty} |x_k| \cdot p_k = \infty$  then the permutation invariance is false

**Definition:** Let  $X$  be a continuous RV with PDF  $p_X(x)$

1. If  $\int_{-\infty}^{\infty} |x| \cdot p_X(x) dx < \infty$ , then  $\mathbb{E}X = \int_{-\infty}^{\infty} x \cdot p_X(x) dx$
2. If  $\int_{-\infty}^{\infty} |x| \cdot p_X(x) dx < \infty$ , then  $\mathbb{E}X$  does not exist

**Remark:** the condition “ $\int_{-\infty}^{\infty} |x| \cdot p_X(x) dx < \infty$ ” is more subtle and requires Lebesgue integrals

**Example:** Let  $X$  be a continuous RV with PDF of the Cauchy-Lorentz distribution

$$p_X(x) = \frac{1}{\pi(1+x^2)}$$

Claim:  $\mathbb{E}X$  does not exist Proof: HW 5

## Transformations of RV

Motivation:

Let  $\Omega$  be the students taking APMA 1655.  $X(\omega)$  is the score student  $\omega$  gets in the final exam. The professor wants to curve the scores such that

$$Y(\omega) = \min\{X(\omega)^2, 100\}$$

$$g(x) = \min\{x^2, 100\} \implies Y(\omega) = g(X(\omega))$$

**More rigorously:** Suppose we are given

- RV  $X: \Omega \rightarrow \mathbb{R}$
- a function  $g: \mathbb{R} \rightarrow \mathbb{R}$
- $\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$
- $\omega \mapsto X(\omega) \mapsto g(X(\omega))$
- $g(X)$  is a random variable  $\omega \mapsto g(X(\omega))$

What is  $\mathbb{E}[g(X)]$ ?

**Definition:** Suppose  $X$  and  $g$  are given.

1. Suppose  $X$  is discrete and has CDF  $\sum_{k=0}^K p_k \cdot \mathbf{1}_{[x_k, \infty)}$ . If  $\sum_{k=0}^K |g(x_k)| \cdot p_k < \infty$  then

$$\mathbb{E}[g(X)] = \sum_{k=0}^K g(x_k) \cdot p_k$$

Otherwise,  $\mathbb{E}[g(X)]$  does not exist.

2. Suppose  $X$  is continuous and has PDF  $p_X(x)$ . If  $\int_{-\infty}^{\infty} |g(x)| \cdot p_X(x) dx < \infty$  then  $\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) \cdot p_X(x) dx$  Otherwise,  $\mathbb{E}[g(x)]$  does not exist

**Definition:** Let  $X$  be a RV with CDF

$$F_X(x) = p \cdot F_Z(x) + (1 - p) \cdot F_W(x)$$

where

1.  $0 \leq p \leq 1$
2.  $Z$  is a discrete RV with CDF  $F_Z$
3.  $W$  is a continuous RV with CDF  $F_W$

Then

$$\mathbb{E}X = p \cdot \mathbb{E}Z + (1 - p)\mathbb{E}W$$

(assuming  $\mathbb{E}Z$  and  $\mathbb{E}W$  exist)

**Example:**

$$X = YZ + (1 - Y)W$$

where

$$X \sim \text{Bernoulli}\left(\frac{1}{3}\right)$$

$$Y \sim \text{Pois}(\lambda)$$

$$Z \sim N(1000, 1)$$

and all three are independent.

Then

$$\mathbb{E}Z = \lambda \quad \mathbb{E}W = 1000$$

and

$$F_X(x) = \frac{1}{3}F_Z(x) + \frac{2}{3}F_W(x)$$

so

$$\mathbb{E}X = \frac{1}{3}\lambda + \frac{2}{3}(1000)$$

## Lecture March 13: Variance

### Part I - Motivation

- $X$  follows a distribution
- The distribution generates numbers  $X_1, \dots, X_n$
- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \approx \mathbb{E}X$
- $e_n = |\bar{X}_n - \mathbb{E}X|$  (that is, the error is the distance between the average of  $n$  numbers and the expected value)

### Part II - Exploring error

**Simulation 1:** Generate  $X_1, \dots, X_n$  from Bernoulli( $p$ ) *Claim:*

$$e_n = |\bar{X}_n - p| \stackrel{\text{likely}}{<} \sqrt{p(1-p) \frac{2 \log(\log n)}{n}}$$

*Proof:* Derive from Brownian Motion

**Simulation 2:** Generate  $X_1, \dots, X_n$  from Poisson( $\lambda$ ) *Claim:*

$$e_n = |\bar{X}_n - \lambda| \stackrel{\text{likely}}{<} \sqrt{\lambda \cdot \frac{2 \log(\log n)}{n}}$$

**Conjecture:** Generate  $X_1, \dots, X_n$  from a distribution. Then

$$e_n = |\bar{X}_n - \mathbb{E}X| \stackrel{\text{likely}}{<} \sqrt{V \cdot \frac{2 \log(\log n)}{n}}$$

where  $V$  is the variance and both the expected value and variance exist for the given distribution

**Remark:** this is known as the “law of the iterated logarithm”

## Part III - Variance

Let  $X$  be a random variable whose expected value  $\mathbb{E}X$  exists. We define a function

$$g(x) = (x - \mathbb{E}X)^2$$

Then the *variance of  $X$*  is

$$\text{Var}(X) = \mathbb{E}[(x - \mathbb{E}X)^2]$$

or the “expected squared deviation from  $\mathbb{E}X$ ”

## Lecture March 15

### Part I - Review of $\mathbb{E}[g(x)]$

If  $X$  is a discrete RV and has CDF  $\sum_{k=0}^K p_k \cdot \mathbf{1}_{[x_k, \infty)}(x)$  then (if it exists),

$$\mathbb{E}[g(x)] = \sum_{k=0}^K g(x_k) \cdot p_k$$

If  $X$  is a continuous RV and has PDF  $p_X(x)$  then (if it exists)

$$\int_{-\infty}^{\infty} g(x) \cdot p_X(x) \, dx$$

### Part II - Review of Variance

**Definition:** Let  $X$  be a RV whose expected value exists. Then the variance of  $X$  (if it exists) is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$$

#### Remark:

1. If  $\mathbb{E}[(X - \mathbb{E}X)^2] = \infty$ , we say the variance does not exist
2. IF  $X$  is discrete and has CDF  $\sum_{k=0}^K p_k \cdot \mathbf{1}_{[x_k, \infty)}(x)$ ,

$$\text{Var}(X) = \sum_{k=0}^K (x_k - \mathbb{E}X)^2 \cdot p_k$$

3. If  $X$  is continuous and has PDF  $p_X(x)$ ,

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 \cdot p_X(x) \, dx$$

### Part III - Properties of Expected Values

Let  $X$  be a continuous or discrete RV.

1. For a constant  $c$ ,  $\mathbb{E}c = c$
2. Let  $a$  and  $b$  be two constants, then

$$\mathbb{E}[aX + b] = a(\mathbb{E}X) + b$$

3. Let  $g_1(x), g_2(x), \dots, g_J(x)$  be functions. Suppose  $\mathbb{E}[g_j](x)$  exists for all  $k = 1, 2, \dots, J$ . Then

$$\mathbb{E}[g_1(X) + g_2(X) + \dots + g_J(X)] = \mathbb{E}\left[\sum_{k=1}^J g_k(X)\right] = \sum_{j=1}^J \mathbb{E}[g_j(X)]$$

exists and shows that the expected value is linear

**Proof:** all the properties derive from the properties of the summation and the integral. Details are homework.

### Part IV - Properties of Variance

1.  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}(X^2) - (\mathbb{E}X)^2$  **Proof:**

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}X)^2] &= \mathbb{E}[X^2 - 2(\mathbb{E}X) \cdot X + (\mathbb{E}X)^2] \\ &= \mathbb{E}X^2 + \mathbb{E}[-2(\mathbb{E}X) \cdot X] + (\mathbb{E}X)^2 \\ &= \mathbb{E}X^2 - 2(\mathbb{E}X)(\mathbb{E}X) + (\mathbb{E}X)^2 \\ &= \mathbb{E}(X^2) - (\mathbb{E}X)^2 \quad \blacksquare\end{aligned}$$

2.  $\text{Var}(aX + b) = a^2\text{Var}(X)$

**Proof:** HW

3. For any constant  $x$ ,  $\text{Var}(c) = 0$

**Proof:**  $\text{Var}(c) = \mathbb{E}(c^2) - (\mathbb{E}c)^2 = c^2 - c^2 = 0$

4. Let  $X$  be a RV. If  $\text{Var}(X) = 0$  then there exists a constant  $c$  such that  $\mathbb{P}(\{\omega \in \Omega : X(\omega) = c\}) = 1$

**Proof:** See real analysis.



## Lecture March 17:

### Part I - Review:

If  $X(\omega) = c$  for all  $\omega \in \Omega$  for some random variable  $X$  and constant  $c$ ,

$$\text{Var}(X) = 0$$

Additionally, if  $\text{Var}(X) = 0$  then  $\exists c$  such that

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = c\}) = 1$$

### Part II - Law of Large Numbers (LLN)

**A sloppy version:** Let  $X$  be a RV that follows a distribution which generates random numbers  $X_1, X_2, \dots, X_n$ . Then if  $\mathbb{E}X$  exists

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \approx \mathbb{E}X$$

e.g. if

- $X \sim \text{Bernoulli}(p) \implies \mathbb{E}X = p$
- $X \sim \text{Pois}(\lambda) \implies \mathbb{E}X = \lambda$

However, this raises some questions: What does  $\approx$  mean? In addition to the existence criteria, what other conditions are necessary?

### Part III - Preparations for the Formal

**Definition:** Let  $\{X_i\}_{i=1}^\infty$  be an infinitely long sequence of RVs defined on  $(\Omega, \mathbb{P})$ . We say  $X_1, X_2, \dots$  are independent if

$$\mathbb{P}(\{\omega \in \Omega : X_i \in A_i \quad \forall i = 1, 2, \dots, n\}) = \prod_{i=1}^n \mathbb{P}(\{\omega \in \Omega : X_i(\omega \in A_i)\})$$

for any positive integer  $n$ , and any subsets  $A_1, \dots, A_n \subset \mathbb{R}$

**Definition:** Let  $\{X_i\}_{i=1}^\infty$  be an infinitely long sequence of RVs defined on  $(\Omega, \mathbb{P})$ . We say  $X_1, X_2, \dots$  are independently and identically distributed if

1.  $X_1, X_2, \dots$  are independent
2.  $X_1, X_2, \dots$  share the same CDF, i.e.  $F_{X_1} = F_{X_2} = F_{X_3} = \dots$

## Part IV - The Law of Large Numbers

**Theorem:** Let  $\{X_i\}_{i=1}^{\infty}$  be an infinitely long sequence of RVs defined on  $(\Omega, \mathbb{P})$ . Suppose  $X_1, X_2, \dots$  are independently and identically distributed and

$$A\{\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{X_1(\omega) + X_2(\omega) + \dots + X_n(\omega)}{n} = \lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}X_1\}$$

where  $\mathbb{E}X_1 = \mathbb{E}X_2 = \dots$  because the CDFs are equal. Then

$$\mathbb{P}(A) = 1$$

**Example:** Flip a fair coin infinitely many times.

- An outcome  $\omega$  is then an infinitely long sequence of H and T.
- $\Omega = \{\omega = (\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \dots) : \text{each } \omega^{(i)} \text{ is either H or T}\}$
- 

$$X_i(\omega) = \begin{cases} 1 & \omega^{(i)} = H \\ 0 & \omega^{(i)} = T \end{cases}$$

- $\exists \mathbb{P}$  such that

$$\mathbb{P}(\{\omega \in \Omega : X_i(\omega) = 1\}) = \mathbb{P}(\{\omega \in \Omega : X_i(\omega) = 0\}) = \frac{1}{2}$$

and  $X_1, X_2, \dots$  are independent which implies  $X_1, X_2, \dots$  are identically and independently distributed.

Notice that

$$A = \{\omega \in \Omega : \lim_{n \rightarrow \infty} \bar{X}_n = \frac{1}{2}\}$$

And by LLN,  $\mathbb{P}(A) = 1$  but  $A \neq \Omega$ . Note: To see why, imagine a sequence of infinite tails such that

$$X(\omega^*) = 0 \implies \lim_{n \rightarrow \infty} \bar{X}_n(\omega^*) = 0 \implies \omega^* \notin A$$

## Lecture March 20

### Part I - Review of the Law of Large Numbers

**Theorem:** Let  $\{X_i\}_{i=1}^{\infty}$  be an infinitely long sequence of RVs defined on  $(\Omega, \mathbb{P})$ . Suppose the random variables are independently and identically distributed (they

share the same CDF) and  $\mathbb{E}X_1$  exists. Then

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}X_1\}\right) = 1$$

Note that because the CDFs are the same,

$$\mathbb{E}X_1 = \mathbb{E}X_2 = \mathbb{E}X_3 = \dots$$

Additionally, denote

$$\bar{X}_n = \frac{X_1(\omega) + \dots + X_n(\omega)}{n}$$

as the *sample average* which determines  $\mathbb{E}X_1$  (the *population average*).

Then, the law of large numbers can be expressed “the sample average converges to the population average with probability 1”

Note that the independence condition is crucial and without it, the LLN is not necessarily true.

**Example:**  $X$  is a RV on  $(\Omega, \mathbb{P})$  where  $X \sim \text{Bernoulli}(\frac{1}{2})$  For each  $i = 1, 2, \dots$

$$X_i(\omega) = X(\omega) \quad \forall \omega \in \Omega$$

$$X_1 = X_2 = X_3 = \dots = X \implies X_1, X_2, \dots \text{ are not independent}$$

so

$$A = \{\omega \in \Omega : \bar{X}_n = \frac{1}{2}\} = \{\omega \in \Omega : X(\omega) = \frac{1}{2}\}$$

The LLN anticipates  $\mathbb{P}(A) = 1$  however

$$\mathbb{P}(A) = \mathbb{P}(X = \frac{1}{2}) = 0$$

so the LLN does not hold

## Part II - Monte Carlo Integration

**Motivation:** Integrals are HARD.

$$I = \int_0^1 \cos^{-1}\left(\frac{\cos(\frac{\pi}{2}X)}{1 + 2\cos(\frac{\pi}{2}x)}\right) dx = \frac{5\pi}{12}$$

But we can approximate it with the LLN! First, let  $U \sim \text{Unif}(0, 1)$  whose PDF is  $\mathbf{1}_{[0,1]}(x)$  Denote the crazy integrand  $g(x)$  then

$$I = \mathbb{E}[g(U)]$$

**Theorem: Generalized Law of Large Numbers** Let  $\{X_i\}_{i=1}^{\infty}$  be an infinitely long sequence of RVs defined on  $(\Omega, \mathbb{P})$ . Let  $g(x)$  be a continuous function. Suppose the random variables are independently and identically distributed (they share the same CDF) and  $\mathbb{E}[g(X_1)]$  exists. Then

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{g(X_1(\omega)) + \dots + g(X_n(\omega))}{n} = \mathbb{E}[g(X_1)]\right\}\right) = 1$$

So applied to the example above using this generalized LLN, we can generate  $X_1(\omega), X_2(\omega), \dots \stackrel{iid}{\sim} \text{Unif}(0, 1)$ . Then the sample average of  $g(X_i)$  will approximate the value of the integral for enough random numbers.

## Lecture March 22: Monte Carlo Integration

### Part I - Introduction

Let  $X$  be a continuous RV with PDF  $p_X(x)$  and  $g(x)$  is a real-values function. Then

$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) \cdot p_X(x) dx$$

and

$$\text{Var}[g(x)] = \int_{-\infty}^{\infty} (g(x) - \mathbb{E}[g(x)])^2 \cdot p_X dx$$

(if the expected exists).

### Part II - An Example

Let  $X \sim \text{Unif}(0, 1)$  and

$$g(x) = \cos^{-1}\left(\frac{\cos(\frac{\pi}{2}X)}{1 + 2\cos(\frac{\pi}{2}x)}\right) \quad 0 < x < 1$$

and the PDF of  $X$  is

$$p_X(x) = \mathbf{1}_{(0,1)}(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

So

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot \mathbf{1}_{(0,1)}(x) dx = \int_0^1 g(x) dx = \int_0^1 \cos^{-1} \left( \frac{\cos(\frac{\pi}{2}x)}{1 + 2 \cos(\frac{\pi}{2}x)} \right) dx = \frac{5\pi}{12}$$

This is a VERY HARD integral to calculate but we can approximate it with Monte Carlo integration.

Using the Generalized Law of Large Numbers, we define an infinite sequence of independently and identically distributed RVs,

$$X_1, X_2, X_3, \dots \sim \text{Unif}(0, 1)$$

so

$$\frac{g(X_1) + g(X_2) + \dots + g(X_n)}{n} \approx \mathbb{E}[g(X_1)] = \int_0^1 \cos^{-1} \left( \frac{\cos(\frac{\pi}{2}x)}{1 + 2 \cos(\frac{\pi}{2}x)} \right) dx$$

This is especially useful for higher dimensional integrals as it allows us to take the running sum of only  $n$  numbers rather than say a 100-dimensional Riemann sum.

Note that for other integrals with bounds  $(a, b)$  rather than  $(0, 1)$  we can still use the same method by defining a new random variable from  $U \sim \text{Unif}(0, 1)$ , where:

$$X = a + (b - a) \cdot U \sim \text{Unif}(a, b)$$

and then calculating  $\bar{X}_n$

Also note that there will be some error between the approximated value and the true value

$$e_n(\omega) = \frac{g(X_1(\omega)) + \dots + g(X_n(\omega))}{n} - \mathbb{E}[g(X_1)]$$

and by *The Law of the Iterated Logarithm*  $|e_n(\omega)|$  is likely to be smaller than

$$\sqrt{\text{Var}[g(X_1)] \cdot \frac{2 \log(\log n)}{n}}$$