



















## *Foreword*

The proofs in this set of notes will be more rigourous compared to the expectations of the course. If you are not the author and is interested in reading the notes, you may skip the proofs should you have little interest in them. The rigour is required almost exclusively for the author himself, for his own practice, and because he transferred his STAT230 course from a class that is clean of proofs.

Also, many of the common mathematical notations will be heavily used both in the author's notes and proofs.



# 1 Lecture 1 May 1st 2018

## 1.1 Introduction

### Definition 1.1.1 (Sample Space)

A **sample space**,  $S$  of a random experiment is the set of all possible outcomes of the experiment.

### Example 1.1.1

The following are some random experiments and their sample space.

- Flipping a coin  
 $S = \{H, T\}$  where  $H$  denotes head and  $T$  tail.
- Rolling a 6-faced dice twice  
 $S = \{(x, y) : x, y \in \mathbb{N}, 1 \leq x, y \leq 6\}$
- Measuring a patient's height  
 $S = \mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$

### Definition 1.1.2 ( $\sigma$ -field)

Let  $S$  be a sample space. The collection of sets  $\mathcal{B} \subseteq \mathbb{P}(S)$ <sup>1</sup>, is called a  $\sigma$ -field (or  **$\sigma$ -algebra**) on  $S$  if:

1.  $\emptyset \in \mathcal{B}$  and  $S \in \mathcal{B}$ ;
2.  $\forall A \in \mathcal{B} \quad A^C \in \mathcal{B}$ ; <sup>2</sup> and
3.  $\forall n \in \mathbb{N} \quad \forall \{A_j\}_{j=1}^n \subseteq \mathcal{B} \quad \cup_{j=1}^n A_j \in \mathcal{B}$ .

### Definition 1.1.3 (Measurable Space)

Given that  $S$  is a non-empty set, and  $\mathcal{B}$  is a  $\sigma$ -field,  $(S, \mathcal{B})$  is a **measurable space**.<sup>3</sup>

### Example 1.1.2

Consider  $S = \{1, 2, 3, 4\}$ . Check if  $\mathcal{B} = \{\emptyset, \{1, 2, 3, 4\}, \{1, 2\}, \{3, 4\}\}$  is a  $\sigma$ -field on  $S$ .

<sup>1</sup> The **power set** of  $S$ ,  $\mathbb{P}(S)$ , is defined as the set that contains all subsets of  $S$ .

<sup>2</sup> We shall denote the compliment of a set by a superscript  $C$  in this set of notes. The supplemental notes provided in the class uses an overhead bar, e.g.  $\overline{A}$ , while lecture notes will use  $A^C$  and  $A'$  interchangeably.

<sup>3</sup> A measurable space is a basic object in **measure theory**.

1. It is clear that  $\emptyset, S \in \mathcal{B}$ .
2. Note that  $S^C = \emptyset$  and  $\{1, 2\}^C = \{3, 4\}$ .
3. Note that the largest possible result of any countable union of the elements of  $\mathcal{B}$  is  $\{1, 2, 3, 4\}$ , which is an element of  $\mathcal{B}$ .

BECAUSE  $(S, \mathcal{B})$  is a measurable space, we can define a measure on it.

#### Definition 1.1.4 (Probability Measure)

Suppose  $S$  is a sample space of a random experiment. Let  $\mathcal{B} = \{A_1, A_2, \dots\} \subseteq \mathbb{P}(S)$  be the  $\sigma$ -field on  $S$ . The **probability set function** (or **probability measure**),  $P : \mathcal{B} \rightarrow [0, 1]$ , is a function that satisfies the following:<sup>4</sup>

- $\forall A \in \mathcal{B} \ P(A) \geq 0$ ;
- $P(S) = 1$ ;
- $\forall \{A_j\}_{j=1}^{\infty} \subseteq \mathcal{B} \ \forall i \neq j \in \mathbb{N} \ A_i \cap A_j = \emptyset \implies$

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \quad (1.1)$$

$(S, \mathcal{B}, P)$  is called a **probability space**.

#### Example 1.1.3

Consider flipping a coin where  $S = \{H, T\}$ . Let  $P$  be defined as follows

$$P(\{H\}) = \frac{1}{3} \quad P(\{T\}) = \frac{2}{3} \quad P(\emptyset) = 0 \quad P(S) = 1$$

Conditions 1 and 2 of Definition 1.1.4 are met. Notice that

$$P(\{H\} \cup \{T\}) = P(S) = 1 \quad \text{and} \quad P(\{H\}) + P(\{T\}) = \frac{1}{3} + \frac{2}{3} = 1.$$

Hence condition 3 is also fulfilled.

#### Proposition 1.1.1 (Properties of Probability Set Functions)

Let  $P$  be a probability set function and  $A, B$  be any set in  $\mathcal{B}$ . Prove the following:<sup>5</sup>

1.  $P(A^C) = 1 - P(A)$
2.  $P(\emptyset) = 0$
3.  $P(A) \leq 1$

<sup>4</sup> These conditions are also known as **Kolmogorov Axioms**, or **probability axioms**.

<sup>5</sup> Many among these properties illustrate that the probability is indeed a **measure**.



6. Note that  $B = B \cap S = B \cap (A^C \cup A) = (B \cap A^C) \cup A$ . Clearly,  $A \cap (B \cap A^C) = \emptyset$ . By condition 3 in Definition 1.1.4, we thus have that

$$P(B) = P(B \cap A^C) + P(A). \quad (\dagger)$$

Suppose  $A \subsetneq B$ . Then  $B \cap A^C \neq \emptyset$ . I shall make the claim that  $B \cap A^C \in \mathcal{B}$ . Since  $A \subseteq B$  we have that

$$\begin{aligned} a \in (B \cap A^C) &\iff a \in B \wedge a \in A^C \\ &\iff a \in B \wedge a \notin A \\ &\iff a \in (B \setminus A). \end{aligned}$$

But  $B \setminus A$  is a subset of  $B$  from the above steps<sup>8</sup>. Therefore,  $(B \cap A^C) \subseteq B \in \mathcal{B}$  as required.

<sup>8</sup> This is rather obvious from the steps, since  $\forall a \in (B \cap A^C), a \in B$ .

With that done, by condition 1 in Definition 1.1.4,  $P(B \cap A^C) \geq 0$ . Hence from Equation  $(\dagger)$ , we have that

$$\begin{aligned} P(B) &= P(B \cap A^C) + P(A) \\ &\geq P(A) \end{aligned}$$

as required. □

### Definition 1.1.5 (Conditional Probability)

Suppose  $S$  is a sample space of a random experiment, and  $A, B \subseteq S$ . The **conditional probability of  $A$  given  $B$**  is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0. \quad (1.2)$$

### Definition 1.1.6 (Independent Events)

Suppose  $S$  is a sample space of a random experiment, and  $A, B \subseteq S$ .  $A$  and  $B$  are said to be **independent of each other** if

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

### Proposition 1.1.2 (Boole's Inequality)

If  $\{A_j\}_{j=1}^{\infty}$  is a sequence of events, then

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} P(A_j)$$

**Proof**

*Proof shall be provided later*

**Proposition 1.1.3 (Bonferroni's Inequality)**

If  $\{A_j\}_{j=1}^k$  is a set of events where  $k \in \mathbb{N}$ , then

$$P\left(\bigcap_{j=1}^k A_j\right) \geq 1 - \sum_{j=1}^k P(A_j^C)$$

**Proof**

*Proof shall be provided later*

**Proposition 1.1.4 (Continuity Property)**

If  $A_1 \subset A_2 \subset \dots$  is a sequence where  $A = \bigcup_{i=1}^{\infty} A_i$ , then

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) = P(A)$$

**Proof**

*Proof shall be provided later*

## 1.2 Random Variable

**Definition 1.2.1 (Random Variable)**

In a given probability space  $(S, \mathcal{B}, P)$ , the function  $X : S \rightarrow \mathbb{R}$  is called a **random variable**<sup>9</sup> if

$$P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\}) \quad (1.3)$$

is defined for all  $x \in \mathbb{R}$ <sup>10</sup>.

<sup>9</sup> We shall use rv as shorthand for random variable in this set of notes.

<sup>10</sup>  $X \leq x$  is an abbreviation for  $\{\omega \in S : X(\omega) \leq x\} \in \mathcal{B}$ .

**Example 1.2.1**

In a coin flip experiment, we have that  $S = \{H, T\}$  where  $\mathbb{P}(S) = \{\emptyset, S, \{H\}, \{T\}\}$ . Define  $X$  : the number of heads in a flip, i.e.

$$X(\{H\}) = 1 \text{ and } X(\{T\}) = 0$$

To prove why  $X$  is a random variable given this definition, notice that

$$x < 0 \implies P(X \leq x) = P(\{\omega \in S : X(\omega) < 0\}) = P(\emptyset) = 0$$

$$\begin{aligned} x \geq 1 &\implies P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\}) = P(\{H, T\}) \\ &= P(\{H\}) + P(\{T\}) = 1 \text{ by Independence} \end{aligned}$$

$$0 \leq x < 1 \implies P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\}) = P(\{T\}) \geq 0$$

which shows that  $P$  is defined for all  $x \in \mathbb{R}$ . Hence  $X$  is a random

variable.

### Definition 1.2.2 (Cumulative Distribution Function)

The **cumulative distribution function (c.d.f)** of a random variable  $X$  is defined as

$$\forall x \in \mathbb{R} \quad F(x) = P(X \leq x)$$

#### Note

NOTICE that  $F(x)$  is defined for **all** real numbers, and since it is a probability, we have  $0 \leq F(x) \leq 1$ .

### Proposition 1.2.1 (Properties of the cdf)

1.  $\forall x_1 < x_2 \in \mathbb{R} \quad F(x_1) \leq F(x_2)$
2.  $\lim_{x \rightarrow -\infty} F(x) = 0 \wedge \lim_{x \rightarrow \infty} F(x) = 1$
3.  $\lim_{x \rightarrow a^+} F(x) = F(a)$  <sup>11</sup>
4.  $\forall a < b \in \mathbb{R} \quad P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$
5.  $P(X = b) = F(b) - \lim_{a \rightarrow b^-} F(a)$  <sup>12</sup>

<sup>11</sup>  $F$  is a **right-continuous** function.

<sup>12</sup> This is also called **the magnitude of the jump**.

#### Proof

**Proof shall be provided later**

#### Note

The definition and properties of the cdf hold for the rv  $X$  regardless of whether  $S$  is discrete (finite or countable) or not.

## 1.3 Discrete Random Variable

### Definition 1.3.1 (Discrete Random Variable)

An rv  $X$  is a **discrete random variable** when its image is finite or countably infinite, i.e.  $X \in \{x_1, x_2, \dots\}$ . The function

$$\forall x \in \mathbb{R} \quad f(x) := P(X = x) = F(x) - \lim_{\varepsilon \rightarrow 0^+} F(x - \varepsilon)$$

is its probability function, commonly known as the **probability mass function (pmf)**. The set  $A := \{x : f(x) > 0\}$  is called the **support set** of  $X$ , and

$$\sum_{x \in A} f(x) = \sum_{i=1}^{\infty} f(x_i) = 1. \quad (1.4)$$



**Proposition 1.3.1 (Properties of pmf)***Prove that*

1.  $\forall x \in \mathbb{R} \quad f(x) \geq 0$
2.  $\sum_{x \in A} f(x) = 1$

**Proof***Proof shall be provided later***Exercise 1.3.1***Consider an urn containing  $r$  red marbles and  $b$  black marbles. Find the pmf of the rv for the following:*

1.  $X =$  number of red balls in  $n$  selections without replacement.
2.  $X =$  number of red balls in  $n$  selections with replacement.
3.  $X =$  number of black balls selected before obtaining the first red ball if sampling is done with replacement.
4.  $X =$  number of black balls selected before obtaining the  $k$ th red ball if sampling is done with replacement.

**Solution**

1. Let  $d = \max\{n, r + b\}$ . The desired pmf is therefore the pmf from the hypergeometric distribution

$$\forall x \in \mathbb{Z}_{\leq r}^+ \quad f(x) = \frac{\binom{r}{x} \binom{b}{d-x}}{\binom{r+b}{d}}.$$

2.  $\forall x \in \mathbb{Z}^+ \quad f(x) = \binom{n}{x} \left(\frac{r}{r+b}\right)^x \left(\frac{b}{r+b}\right)^{n-x}$ , which is the pmf of the binomial distribution.

$$3. \quad \forall x \in \mathbb{Z}^+ \quad f(x) = \left(\frac{b}{r+b}\right)^x \left(\frac{r}{r+b}\right)$$

$$4. \quad \forall x \in \mathbb{Z}^+ \quad f(x) = \binom{x+k-1}{k-1} \left(\frac{b}{r+b}\right)^x \left(\frac{r}{r+b}\right)^k$$

**Example 1.3.1***Consider the function*

$$f(x) = \begin{cases} \frac{C\mu^x}{x!} & x \in \mathbb{Z}^+, \mu > 0 \\ 0 & \text{otherwise} \end{cases}$$

*Find  $C$  such that  $f(x)$  is a pmf for the rv  $X$ .*

**Solution**

We have that

$$\begin{aligned} 1 &= \sum_{x \in \mathbb{Z}^+} \frac{C\mu^x}{x!} \\ &= C \sum_{x \in \mathbb{Z}^+} \frac{\mu^x}{x!} \\ &= Ce^\mu \end{aligned}$$

Thus  $C = e^{-\mu}$ .

**Exercise 1.3.2**

Prove that the pdf of  $X \sim \text{POI}(\mu)$  sums to 1 over all of its values.

**Solution**

$$\begin{aligned} \sum_{x \in \mathbb{N}} \frac{\mu^x e^{-\mu}}{x!} &= e^{-\mu} \sum_{x \in \mathbb{N}} \frac{\mu^x}{x!} \\ &= e^{-\mu} e^\mu \quad \because \sum_{x \in \mathbb{N}} \frac{k^x}{x!} = e^k \\ &= 1 \end{aligned}$$

**Exercise 1.3.3**

If  $X$  is a random variable with pmf

$$f(x) = \frac{-(1-p)^x}{x \log p}, \quad x = 1, 2, \dots; \quad 0 < p < 1,$$

show that

$$\sum_{x \in \mathbb{N}} f(x) = 1$$

**Solution**

$$\begin{aligned} \sum_{x \in \mathbb{N}} \frac{-(1-p)^x}{x \log p} &= -\frac{1}{\log p} \sum_{x \in \mathbb{N}} \frac{(-1)^x (p-1)^x}{x} \\ &= -\frac{1}{\log p} \underbrace{\left[ -(p-1) + \frac{(p-1)^2}{2} - \frac{(p-1)^3}{3} + \dots \right]}_{\text{Taylor expansion of } -\log p} \\ &= 1 \end{aligned}$$

This gives us that  $\forall x \in \mathbb{Z}^+$ ,  $f(x) = \frac{e^{-\mu} \mu^x}{x!}$ , and this is, of course, the pmf of the **Poisson distribution**.

## 2 Lecture 2 May 03rd 2018

### 2.1 Continuous Random Variable

#### Definition 2.1.1 (Continuous Random Variable)

Suppose  $X$  is an rv with cdf  $F$ . If  $F$  is a continuous function for all  $x \in \mathbb{R}$  and  $F$  is differentiable except possibly at countably many points, then  $X$  is a **continuous rv**. The probability function, or more commonly known as the **probability density function** (pdf), of  $X$  is  $f(x) = F'(x)$  wherever  $F$  is differentiable on  $x$  and 0 otherwise.

The set  $A = \{x : f(x) > 0\}$  is called the **support set** of  $X$  and

$$\int_{x \in A} f(x) dx = 1$$



## 3 Index

$\sigma$ -algebra, 11

$\sigma$ -field, 5, 11

Bonferroni's Inequality, 7, 15

Boole's Inequality, 7, 14

Conditional Probability, 5, 14

Continuity Property, 7, 15

Continuous Random Variable, 5, 19

Cumulative Distribution Function, 5,  
16

Discrete Random Variable, 5, 16

Independent Events, 5, 14

Kolmogorov Axioms, 12

Measurable Space, 5, 11

power set, 11

probability axioms, 12

probability density function, 19

probability mass function, 16

Probability Measure, 5, 12

probability set function, 12

probability space, 12

Properties of pmf, 7, 17

Properties of the cdf, 7, 16

Random Variable, 5, 15

right-continuous, 16

Sample Space, 5, 11

support set, 16