STAT330S18 - Mathematical Statistics

Johnson Ng

May 2, 2018

Table of Contents

List of Definitions

List of Theorems

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*
  $S = R^+ = \{x \in \mathbb{R} : x \geq 0\}$

Definition 1.1.2 (σ-field)
Let S be a sample space. The collection of sets $\mathscr{B} \subseteq \mathbb{P}(S)$[1], is called a σ-field (or ***σ-algebra****) on S if:*

1. $\emptyset \in \mathscr{B}$ *and* $S \in \mathscr{B}$;
2. $\forall A \in \mathscr{B} \quad A^C \in \mathscr{B}$; [2] *and*
3. $\forall n \in \mathbb{N} \quad \forall \{A_j\}_{j=1}^n \subseteq \mathscr{B} \quad \cup_{j=1}^n A_j \in \mathscr{B}$.

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

[2] 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' interchangably.

Definition 1.1.3 (Measurable Space)
Given that S is a non-empty set, and $\mathscr{B}$ is a σ-field, $(S, \mathscr{B})$ is a ***measurable space****.*[3]

[3] A measurable space is a basic object in **measure theory**.

Example 1.1.2
Consider $S = \{1, 2, 3, 4\}$. Check if $\mathscr{B} = \{\emptyset, \{1, 2, 3, 4\}, \{1, 2\}, \{3, 4\}\}$ is a σ-field on S.

1. *It is clear that* $\emptyset, S \in \mathscr{B}$.
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* $\mathscr{B}$ *is* $\{1,2,3,4\}$, *which is an element of* $\mathscr{B}$.

Because $(S, \mathscr{B})$ 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* $\mathscr{B} = \{A_1, A_2, ...\} \subseteq \mathbb{P}(S)$ *be the* σ*-field on* S. *The* ***probability set function*** *(or* ***probability measure****),* $P : \mathscr{B} \to [0,1]$, *is a function that satisfies the following:*[4]

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

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

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

[4] These conditions are also known as **Kolmogorov Axioms**, or **probability axioms**.

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

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

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* $\mathscr{B}$. *Prove the following:*[5]

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

[5] Many among these properties illustrate that the probability is indeed a **measure**.

4. $P(A \cap B^C) = P(A) - P(A \cap B)$

5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6. $A \subseteq B \implies P(A) \leq P(B)$

Proof

Let S be the sample space for P.

Exercise 1.1.1
Prove that $A \subseteq B \iff B^C \subseteq A^C$.

1. *Note that*
   $A \in \mathscr{B} \implies A \in \mathbb{P}(S) \iff A \subseteq S$
   $A \in \mathscr{B} \iff A^C \in \mathscr{B} \implies A^C \subseteq S$. *Also, since A^C is the complement of A, it is clear that $S = A \cup A^C$.*

   $$\therefore P(S) = 1 \iff P(A \cup A^C) = 1 \overset{1}{\iff} P(A) + P(A^C) = 1$$

   where 1 is by condition 3 in Definition 1.1.4 since $A \cap A^C = \emptyset$ by definition of a complement of a set.

2. *Note that $S \cup \emptyset = S$ and $S \cap \emptyset = \emptyset$. Using a similar argument as above,*

   $$1 = P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset) \implies P(\emptyset) = 0$$

3. *By 1 from above, $P(A) = 1 - P(A^C)$. Since $0 \leq P(A^C) \leq 1$, we have that $P(A)$ is at most 1, as required.*

4. *Note that $A = (A \cap B) \cup (A \cap B^C)$. Clearly, $(A \cap B) \cap (A \cap B^C) = \emptyset$.*[6] *Hence by condition 3 in Definition 1.1.4,*

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

5. *Consider $P(A \cup B) + P(A \cap B)$. By definition,*

   $$A \cup B = (A \cap B^C) \cup (A \cap B) \cup (A^C \cap B)$$

   where each of the sets in brackets are disjoint from each other[7]*. By condition 3 of Definition 1.1.4, we would then have*

   $$\begin{aligned} &P(A \cup B) + P(A \cap B) \\ &= P(A \cap B^C) + P(A \cap B) + P(A^C \cap B) + P(A \cap B) \\ &= 2P(A \cap B) + P(A) - P(A \cap B) + P(B) - P(A \cap B) \quad \text{by } 4 \\ &= P(A) + P(B) \end{aligned}$$

[6] This is an easy proof using the basic way of proving membership.

[7] Again, this is not hard to show

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

$$P(B) = P(B \cap A^C) + P(A). \tag{†}$$

Suppose $A \subsetneq B$. *Then* $B \cap A^C \neq \emptyset$. *I shall make the claim that* $B \cap A^C \in \mathscr{B}$. *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*[8]. *Therefore,* $(B \cap A^C) \subseteq B \in \mathscr{B}$ *as required.*

[8] 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 (†), 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*

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

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 $A = \cup_{i=1}^n A_i$, then

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

Proof
Proof shall be provided later

1.2 *Random Variable*

Definition 1.2.1 (Random Variable)
In a given probability space $(S, \mathscr{B}, P)$, the function $X : S \to \mathbb{R}$ is called a ***random variable***[9] *if*

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

is defined for all $x \in \mathbb{R}$[10].

[9] We shall use rv as shorthand for random variable in this set of notes.

[10] $X \leq x$ is an abbreviation for $\{\omega \in S : X(\omega) \leq x\} \in \mathscr{B}$.

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

$$\begin{aligned}
x < 0 &\implies P(X \leq x) = P(\{\omega \in S : X(\omega) < 0\}) = P(\emptyset) = 0 \\
x \geq 1 &\implies P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\}) = P(\{H, T\}) \\
&\qquad = P(\{H\}) + P(\{T\}) = 1 \text{ by Independence} \\
0 \leq x < 1 &\implies P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\}) = P(T) \geq 0
\end{aligned}$$

which shows that P is deiined 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. $\lim_{x \to -\infty} = 0 \land \lim_{x \to \infty} = 1$
3. $\lim_{x \to a^+} F(x) = F(a)$ [11]
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. $P(X = b) = F(b) - \lim_{a \to b^-} F(a)$ [12]

[11] F is a **right-continuous** function.

[12] 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. $X \in \{x_1, x_2, ...\}$. The function*

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

*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*

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

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

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

Example 1.3.1
Consider the function

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

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

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 Index