Higher Theory of Statistics
Math2901 UNSW

Hussain Nawaz
hussain.nwz000@gmail.com

2022T2

Contents

1 Introduction

1.1 Experiments, Sample Space and Events

Experiments An experiment is any process that leads to a recorded observation.

Outcome and Sample Space An outcome is possible result of the experiment. The set of all possible outcomes is called the sample space. The sample space is often denoted by Ω.

Observe that not all sample spaces are countable. An uncountable example would be the set of all real number between 0 and 1.

Events An event is a set of outcomes that is, a subset of the sample space Ω.

Mutual Exclusion Events A, B are mutually exclusive (disjoint) if they have no outcomes in common. That is, $A \cap B = \emptyset$.

Set Operation Revision If you have trouble recalling the following laws, for associativity and distributivity, you may replace $\cap$ with $\times$ and $\cup$ with $+$.

TODO: Associative and Distributive Law

1.2 Sigma Algebra

The σ algebra must be defined for rigorously working with probability. The formalization of this, is beyond the scope of this course.

The σ-algebra can be thought of as the family of all possible subsets or events in a sample space. Analogously, this may be conceptualised as the power-set of the sample space.

Probability The probability is a set function, often denoted by $\mathcal{P}$ that maps events from the σ-algebra to $[0, 1]$ and satisfies certain properties.

Probability Space The triplet $\Omega, \mathcal{A}, \mathbb{P}$ is the probability space where

- Ω is the sample space,
- $\mathcal{A}$ is the σ-algebra,
- $\mathbb{P}$ is the probability function.

Properties of Probability Given the probability/sample space $\Omega, \mathcal{A}, \mathbb{P}$, the probability function $\mathbb{P}$ must satisfy

- For all set $A \in \mathcal{A}$, $\mathbb{P}(A) \geq 0$
- $\mathbb{P}(\Omega) = 1$
- Countable additive. Suppose that the family of set A_i

Theorem: Continuity from below Given an increasing sequence of events $A_1 \subset A_2 \subset \cdots \subset A_n$ then,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} = \mathbb{P}(A_n)$$

Theorem: Continuity from above Given a decreasing sequence of events $A_1 \supset A_2 \supset \ldots A_n$ then,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} = \mathbb{P}(A_n)$$

More Probability Lemmas

- $\mathbb{P}(\emptyset) = 0$,
- For any $A \in \mathcal{A}$, $\mathbb{P}(A) \leq 1$ and $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- Suppose $A, B \in \mathcal{A}$ and $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

1.3 Conditional Probability and Independence

Conditional Probability The conditional probability that an event A occurs given that the event B has already occurred is denoted by

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

Independence The events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

A lemma on independence Given two events A, B, then

$$\mathbb{P}(A|B) = A \quad \text{if and only if} \quad \mathbb{P}(B|A) = B$$

Pairwise Independence of Sequences A countable sequence of events $A_{i\in\mathbb{N}}$ is pairwise independent if

$$\mathbb{P}(A_i0 \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \forall i \neq j.$$

Independence of Sequences A countable sequence of events $A_{i\in\mathbb{N}}$ is independent is independent if for any sub-collection $A_{i_1}, \ldots, A_{i_n}$ we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) = \prod_{j=1}^{n} \mathbb{P}(A_{i_j}).$$

Multiplicative Law Given A, B are events, then,

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

This is equivalent to the multiplication down a decision tree.

Additive Law Let A, B be events. Then,

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

This is analogous to the inclusion-exclusion principle from set theory.

Law of Total Probability Suppose that $(A_i)_{i=1,\ldots,k}$ are mutually exclusive and exhaustive of Ω. That is,

$$\bigcup_{i=1}^{k} A_i = \Omega.$$

Then for any event B, we have

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

1.4 Descriptive Statistics and R

Sample Variance and Mean Suppose that we are given observations x such that $x = (x_1, x_2, \ldots, x_n)$.

Then, the **sample mean** is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The **sample variance** is given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

2 Random Variables

2.1 Random Variables

Definition: Random Variables Suppose that we work on a probability space $(\Omega, \mathcal{A}, \mathbb{P}$ And the outcomes in Ω are denoted by ω.

Then, a random variable (r.v) X is a function from Ω to $\mathbb{R}$ such that $\forall x \in \mathbb{R}$, the set $A_x = \{\omega \in \Omega, X(\omega) \leq x\}$. That is, a random variable is a function that maps *Omega* to some space.

Convention on Random Variables Random variables are often denoted by capital letters while, the outcomes are denoted by the lower-case equivalent of the random variable.

Cumulative Distributive The cumulative distribution of a r.v X is defined by

$$F_X(x) = \mathbb{P}(\{\omega : X(\omega \leq x)\}) = \mathbb{P}(X \leq x).$$

Cumulative Distribution Theorems Suppose that F_X is cumulative distribution function of X. Then,

- It is bounded between zero and one and

$$\lim_{x \downarrow -\infty} = 0 \quad \text{and} \lim_{x \uparrow \infty} = 1.$$

- It is non-decreasing. That is, if $x \leq y$ then, $F_X(x) \leq F_X(y)$.
- For any $x \leq y$,

$$\mathbb{P}(x < X < y) = \mathbb{P}(X \leq y) - \mathbb{P}(X \leq x) = F_X(y) - F_X(x).$$

- It is right continuous. That is,

$$\lim x \uparrow \infty F_X \left(x + \frac{1}{n} \right) = F_X(x).$$

- it has finite left-hand limit and

$$\mathbb{P}(X < x) = \lim_{n \to \infty} F_x \left(x - \frac{1}{n} \right),$$

  denoted by $F_X(x-)$. It is useful to observe that,

$$\mathbb{P}(X = x) = F_X(X) - F_X(x-) \coloneqq F_X(x).$$

Discrete Random Variables A r.v. is said to be discrete if the image of X consists of countable many values x where $\mathbb{P}(X = x) > 0$. The probability function is $\Delta F_X(x) = \mathbb{P}(X = x)$ and satisfies

$$\sum_{\text{all} x} \mathbb{P}(X = x) = 1.$$

Continuous Random Variables and Probability Density Functions A r.v is continuous if the image of X takes a continuum of values.

The probability density function of a r.v is a real-valued function f_x on $\mathbb{R} with$ the property that

$$\mathbb{P}(X \in A) = \int_A f_x(y) dy,$$

for any *Borel* subset of $\mathbb{R}$

Required Properties of a Density Function Valid density functions $f : \mathbb{R} \to \mathbb{R}$ must satisfy the following properties:

- $f(x) \geq 0, \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)dx = 1$.

Useful Properties of a Continuous Random Variable For all continuous random variables X, with density f_x,

1. If $A = (-\infty, x]$ and creating a cumulative distribution function F_x such that $F_X(x) = \mathbb{P}(X \in A) = \mathbb{P}(X \leq x)$ then,
$$F_X(x) = \int_{-infty}^{x} f_x(y)dy.$$
2. For all $a < b$,
$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_(a)^b f_X(x)dx.$$
3. By the fundamental theorem of calculus and property 1,
$$F_X'(x) = \frac{d}{dx}\int_{-\infty}^{x} f_x(y)dy = f_X(x).$$

2.2 Expectation and Variance

Expectation The expectation of a r.v X, denoted by $\mathbb{E}(X)$ may be computed depending on when X is discrete or continuous.

Expectation of Discrete Random Variables If X is a discrete random variable then,
$$\mathbb{E}(X) \coloneqq \sum_{\text{all } x} x\mathbb{P}(X = x) = \sum_{\text{all } x} x\Delta F_x(x).$$

Expectation of continuous Random Variables If X is a continuous random variable then,
$$\mathbb{E}(X) \coloneqq \int_{-infty}^{infty} xf_X(x)dx$$

Interpreting the Expectation Often $\mathbb{E}(x)$ is called the *mean* of X. Observe that mean and average are not neccesarily the same. $\mathbb{E}(X)$ may be thought as the long-run average of the outcomes of X. That is, the average observation of X converges to $\mathbb{E}(X)$.

Where our density function represents a physical model, $\mathbb{E}(X)$ is equivalent to the center of mass.

Linearity of the Expectation We note that the expectation is linear. That is, for all constants $a, b \in \mathbb{R}$,

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

Variance Let X be a r.v and set $\mu = \mathbb{E}(x)$. Then,

$$\mathrm{Var}(X) \coloneqq \mathbb{E}\left((X - \mu)^2\right).$$

The standard deviation is the square root of variance.

Properties of Variance Given a random variance X then, for any constants $a, b \in \mathbb{R}$,

1. $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.,
2. $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$,
3. $\mathrm{Var}(X + b) = \mathrm{Var}(X)$,
4. $\mathrm{Var}(b) = 0$.

2.3 Moment Generating Functions

Momnets A moment of the random variable is denoted by

$$\mathbb{E}[X^r], \quad r = 1, 2, \ldots$$

Moments measure mean, variance, skewness, and kurtosis, all ways of looking at the shape of the distribition.

Suppose that $f(x)$ is a probability density function. Then,

$$\mathbb{E}[X^r] = \int_{-\infty}^{\infty} x^r f(x) dx$$

Kurtosis The kurtosis is the standards 4th moment. It measures how *fat* the tail is. A positive kurtosis implies a thinner tail than negative kurtosis.

Moment Generating Function A moment generating function (MGF) is denoted as

$$M_x(u) = \mathbb{E}(e^{uX}) = \int_{\text{all } x} e^{uX} f_X(x) dx.$$

We say that the MGF of X exists if $M_X(u)$ is finite in some interval containing zero.

Using Moment Generating Function to Find Moments Suppose that the moment genrating funciton exists. Then,

$$\mathbb{E}(X^r) = \lim_{u \to 0} M_X^(r)(u) =: \lim_{x \to 0} \frac{d^r}{du} M_x(u).$$