MATH 310: Mathematical Statistics (brief notes)

Truong-Son Van

Contents

Disclaimer

This is class notes for Mathematical Statistics at Fublbright University Vietnam. I claim no originality in this work as it is mostly taken from the reference books. However, all errors and typos are solely mine.

PART 1: Background

Chapter 1

Probability

"If we have an atom that is in an excited state and so is going to emit a photon, we cannot say when it will emit the photon. It has a certain amplitude to emit the photon at any time, and we can predict only a probability for emission; we cannot predict the future exactly."

— Richard Feynman

1.1 Review

1.1.1 Probability Space

Definition 1.1 (Sigma-algebra)**.** Let Ω be a set. A set $\Sigma \subseteq \mathcal{P}(\Omega)$ of subsets of Ω is called a σ-algebra of Ω if

1. $\Omega \in \Sigma$
2. $F \in \Sigma \implies F^C \in \Sigma$
3. If $F_n \in \Sigma$ for all $n \in \mathbb{N}$, then
$$\bigcup_n F_n \in \Sigma.$$

A Borel σ-algebra is the smallest σ-algebra that contains all the open sets.

Definition 1.2 (Probability Space)**.** A *Probability Space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set called *sample space*, $\mathcal{F}$ is a σ-algebra on Ω, $\mathbb{P} : \mathcal{F} \to [0, 1]$, called a *Probability Measure*, is a function that satisfies the following:

1. $P(\Omega) = 1$,
2. If F is a disjoint union of $\{F_n\}_{n=1}^{\infty}$, then
$$\mathbb{P}(F) = \sum_{n=1}^{\infty} \mathbb{P}(F_n).$$

Each element $\omega \in \Omega$ is called an *outcome* and each subset $A \in \mathcal{F}$ is called an *event*.

Philosophically, the σ-algebra represents the details of information we could have access to. There are certain events that are building blocks of knowledge and that we don't have access to finer details.

Definition 1.3 (Independent Events)**.** Let $A, B \in \mathcal{F}$ be events. We say that A and B are independent if

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

Definition 1.4 (Conditional Probability)**.** Let $A, B \in \mathcal{F}$ be events such that $\mathbb{P}(B) > 0$. Then, the conditional probability of A given B is

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

Theorem 1.1 (Bayes's Theorem)**.** *Let $A, B \in \mathcal{F}$ be events such that $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Then,*

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

In modern statistics, there are names for the above terms:

1. $\mathbb{P}(A|B)$ is called *Posterior Probability*,
2. $\mathbb{P}(B|A)$ is called *Likelihood*,
3. $\mathbb{P}(A)$ is called *Prior Probability*,
4. $\mathbb{P}(B)$ is called *Evidence*.

The theorem is often expressed in words as:

$$\text{Posterior Probability} = \frac{\text{Likelihood} \times \text{Prior Probability}}{\text{Evidence}}$$

It is a good idea to ponder why those mathematical terms have those names.

1.1.2 Random Variables

The notion of probability alone isn't sufficient for us to describe ideas about the world. We need to have a notion of objects that associated with probabilities. This brings about the idea of *random variable*.

Definition 1.5 (Random Variable)**.** Let $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ be a probability space and $(S, \mathcal{B}(S))$ a σ-algebra. A random variable is a (Borel measurable) function from $\Omega \to S$.

- S is called the *state space* of X.

In this course, we will restrict our attentions to two types of random variables: discrete and continuous.

Definition 1.6 (Discrete RV)**.** $X : \Omega \to S$ is called a discrete RV if S is a countable set.

- A *probability function* or *probability mass function* for X is a function $f_X : S \to [0, 1]$ defined by

$$f_X(x) = \mathbb{P}(X = x) .$$

In contrast to the simplicity of discrete RV. Continuous RVs are a little bit messier to describe. This is because of the lack of background in measure theory so we can talk about this concept in a more precise way.

Definition 1.7 (Continuous RV)**.** A *continuous random variable* is a measurable function $X : \Omega \to S$ is continuous if it satisfies the following conditions:

1. $S = \mathbb{R}^n$ for some $n \in \mathbb{N}$.
2. There exists an (integrable) function f_X such that $f_X(x) \geq 0$ for all x, $\int_{\mathbb{R}^n} f_X(x)dx = 1$ and for every open cube $C \subseteq \mathbb{R}^n$,

$$\mathbb{P}(X \in C) = \int_C f_X(x)dV .$$

The function f_X is called the *probability density function (PDF)*.

There are more general concepts of continuous RV where we don't need to require S to be a Euclidean space as in the above definition. However, such concepts require the readers to be familiar with advanced subjects like Topology and Measure Theory. It is particularly important to know these two subjects in order to thoroughly understand Stochastic Processes.

Exercise 1.1. Create a random variable that represents the results of n coin flips.

For real-valued RV $X : \Omega \to \mathbb{R}$ we have the concept of cumulative distribution function.

Definition 1.8 (Cumulative Distribution Function)**.** Given a RV $X : \Omega \to \mathbb{R}$. The *cumulative distribution function of* X or CDF, is a function $F_X : \mathbb{R} \to [0, 1]$ defined by

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

Notationally, we use the notation $X \sim F$ to mean RV X with distribution F.

Exercise 1.2. Given a real-valued continuous RV $X : \Omega \to \mathbb{R}$, prove that

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

is differentiable for every x and $f_X(x) = F'_X(x)$.

Exercise 1.3. Let X be an RV with CDF F and Y with CDF G. Suppose $F(x) = G(x)$ for all x. Show that

$$\mathbb{P}(X \in A) = \mathbb{P}(Y \in B)\,.$$

Exercise 1.4. Let $X : \Omega \to \mathbb{R}$ be an RV and F_X be its CDF. Prove the following:

1. F is non-decreasing: if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
2. F is normalized:
$$\lim_{x \to -\infty} F(x) = 0\,,$$
and
$$\lim_{x \to \infty} F(x) = 1\,.$$
3. F is right-continuous:
$$F(x) = F(x+) = \lim_{y \searrow x} F(y)\,.$$

1.1.3 Joint distribution of RVs

Let $X : \Omega \to S$ and $Y : \Omega \to S$ be RVs. We denote

$$\mathbb{P}(X \in A; Y \in B) = \mathbb{P}(\{X \in A\} \cap \{Y \in B\})\,.$$

For discrete RVs, the joint probability function of X and Y is

$$f_{X,Y}(x, y) = \mathbb{P}(X = x; Y = y)$$

For continuous RVs, the situations are more complicated.

1.1.4 Some important random variables

1. Point mass distribution (Dirac delta): Given a discrete probability $X : \Omega \to S$. X has a point mass distribution at $a \in S$ if
$$\mathbb{P}(X = a) = 1.$$
We call X a point mass RV and write $X \sim \delta_a$.

   Question. Suppose $S = \mathbb{N}$. Write down F_X for the point mass RV X.

2. Discrete uniform distribution: $f_X(k) = \frac{1}{n}$, $k \in \{1, \dots, n\}$.

3. Bernoulli distribution: let $X : \Omega \to \{0, 1\}$ be RV that represents a binary coin flip. Suppose $\mathbb{P}(X = 1) = p$ for some $p \in [0, 1]$. Then X has a Bernoulli distribution, written as $X \sim \text{Bernoulli}(p)$. The probability function is
$$f_X(x) = p^x(1-p)^{1-x}.$$
We write $X \sim \text{Bernoulli}(p)$.

4. Binomial distribution: let $X : \Omega \to \mathbb{N}$ be the RV that represents the number of heads out of n independent coin flips. Then
$$f_X = \begin{cases} \binom{n}{x} p^x(1-p)^{n-x}, & x \in \{0, 1, \dots, n\} \\ 0, & \text{otherwise} \end{cases}$$
We write $X \sim \text{Binomial}(n, p)$.

5. Poisson distribution: $X \sim \text{Poisson}(\lambda)$.
$$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}, x \geq 0.$$
X is a RV that describe a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event

Exercise 1.5. Let $X_{n,p} \sim \text{Binomial}(n, p)$. Suppose that as $n \to \infty$, $p \to 0$ in such a way that $np = \lambda$ always. Let $x \in \mathbb{N}$.

1. For n very very large, what is the behaviour of
$$\frac{n!}{(n-x)!}.$$
(You should just get some power of n)

2. Show that
$$\lim_{n \to \infty} f_{X_{n,p}}(x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

3. Interpret this result.

1.1.5 Independent random variables

Definition 1.9. Let $X : \Omega \to S$ and $Y : \Omega \to S$ be RVs. We say that X and Y are independent if, for every $A, B \in \mathcal{B}(S)$, we have
$$\mathbb{P}(X \in A; Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B),$$
and write $X \perp Y$.

Definition 1.10. Let $X : \Omega \to S$ and $Y : \Omega \to S$ be RVs. Suppose that $f_Y(y) > 0$. The conditional probability mass function of X given Y is
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

1.1.6 Transformations of RVs

1.2 Inequalities

1.3 Law of Large Numbers

1.4 Central Limit Theorem

PART 2: Inference

Chapter 2

Sampling, Estimating CDF and Statistical Functionals

2.1 Empirical Distribution

2.2 Statistical Functionals

2.3 Bootstrap

Chapter 3

Parametric Inference (Parameter Estimation)

3.1 Method of Moments

3.2 Method of Maximum Likelihood

3.3 Bayesian Approach

3.4 Expectation-Maximization Algorithm

3.5 Unbiased Estimators

3.6 Efficiency: Cramer-Rao Inequality

3.7 Sufficiency and Unbiasedness: Rao-Blackwell Theorem

Chapter 4

Hypothesis Testing

4.1 Neyman-Pearson Lemma

4.2 Wald Test

4.3 Likelihood Ratio Test

4.4 Comparing samples

PART 3: Models

Chapter 5

Linear Least Squares

5.1 Simple Linear Regression

5.2 Matrix Approach

5.3 Statistical Properties