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 $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ and $(S, \mathcal{B}(S))$. 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.

Definition 1.7 (Continuous RV)**.** A random variable X is continuous if there exists a function f_X such that $f_X(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and for every $a \leq b$,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx.$$

The function f_X is called the probability density function (PDF). We have that

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

and $f_X(x) = F'_X(x)$ at all points x at which F_X is differentiable.

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

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

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

Exercise 1.2. 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:
$$F(x) = F(x+) = \lim_{y \searrow x} F(y)\,.$$

1.2 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