Introduction to probability theory and statistical inference

Jesús Urtasun Elizari

Imperial College London

December 25, 2024

ii

Contents

Index

Chapter 1

Introduction

In the following pages one will find an introductory text to one of the key subjects within mathematical sciences. The text is composed by four chapters, together with some appendix reviewing basic mathematical concepts, and a bibliographic note. The purpose of this lecture notes is to make both probability and statistical analysis an easy, interesting and engaging topic for anyone interested, without the need for prior experience with mathematical training.

First, we will introduce and explore the concept of probability itself, and we will discuss how to model information, surprise, and various random processes, also referred to as *stochastic*. Then we will introduce the idea of a function and how functions need to be adapted to implement uncertainty when discussing random events.

In the second part we will address the difference between prediction and inference, and discuss a set of subjects normally grouped under the name of hypothesis testing. Here we will introduce how to quantify certainty and bias, and how to model significance and false positives, that is, to compute p-values.

Finally, we will discuss bayesian statistics.

Chapter 2

Introduction to probability and random events

2.1 What is probability?

Probability theory is one of the oldest subjects within mathematical studies. Paradoxically, almost every topic explained nowadays in modern courses of statistics is extremely new, ranging broadly a century. (...)

So let's first ask ourselves the question. What *is* probability in the first place? What do we mean by it and what does it describe? Probability is nothing more, and nothing less, that a number we make up, a quantity we come up with, to quantify certainty. A number we will use to describe the amount of information we have about a random, or stochastic, event. For simplicity, we can make it range from 0 to 1, in the following way.

- If I'm sure A will never happen, P(A) = 0.

- If I'm sure A will always happen, P(A) = 1.

- Anything in between, if exists degree of surprise, P(A) in [0, 1]

$$\sum_{i=0}^{n} p(x_i) = 1 \tag{2.1}$$

Once we have a definition for probability in the abstract case, we should have a way to compute for particular cases. A way of doing that, referred to as frequentist approach, is by dividing the number of favorable outcomes by the total number of outcomes.

$$P(\text{A happening}) = \frac{\text{Number of times A happens}}{\text{Total number of trials}} \tag{2.2}$$

Probabilities must follow a property we call unitarity. Unitarity ensures that, if we consider and add up the probabilities for all possible events in a given experiment, we get the total. That means, at least one of the scenarios will happen.

Indeed, the literal meaning of probability comes from latin probabilis. American logician and philosopher Richard Jeffrey, "Before the middle of the seventeenth century, the term "probable" (Latin probabilis) meant just approvable, and was applied in that sense, univocally, to opinion and to action. A probable action or opinion was one such as sensible people would undertake or hold, in the circumstances."[12] However, in legal contexts especially, "probable" could also apply to propositions for which there was good evidence.

The sixteenth-century Italian polymath Girolamo Cardano demonstrated the efficacy of defining odds as the ratio of favourable to unfavourable outcomes (which implies that the probability of an event is given by the ratio of favourable outcomes to the total number of possible outcomes[14]). Aside from the elementary work by Cardano, the doctrine of probabilities dates to the correspondence of Pierre de Fermat and Blaise Pascal (1654). Christiaan Huygens (1657) gave the earliest known scientific treatment of the subject.[15] Jakob Bernoulli's Ars Conjectandi (posthumous, 1713) and Abraham de Moivre's Doctrine of Chances (1718) treated the subject as a branch of mathematics.[16] See Ian Hacking's The Emergence of Probability[10] and James Franklin's The Science of Conjecture[17] for histories of the early development of the very concept of mathematical probability.

Like other theories, the theory of probability is a representation of its concepts in formal terms – that is, in terms that can be considered separately from their meaning. These formal terms are manipulated by the rules of mathematics and logic, and any results are interpreted or translated back into the problem domain.

There have been at least two successful attempts to formalize probability, namely the Kolmogorov formulation and the Cox formulation. In Kolmogorov's formulation (see also probability space), sets are interpreted as events and probability as a measure on a class of sets. In Cox's theorem, probability is taken as a primitive (i.e., not further analyzed), and the emphasis is on constructing a consistent assignment of probability values to propositions. In both cases, the laws of probability are the same, except for technical details.

2.2   Discrete and continous

Discrete and continous

2.3   Probability distributions

Once we are comfortable with the idea of random events, and we have been introduced to probability as a number to quantify surprise, we can agree that not all random phenomena are equal. A basic way to classify and separate random events is according to how their probabilities are distributed.

Binomial distribution

What is the probability of obtaining x successes in N trials, if the probability of success each time is p

Poisson distribution

What is the probability of counting x events in a time interval, if the observed average is given by λ

Uniform distribution

Uniform distribution

Exponential distribution

Exponential distribution

Gaussian distribution

Gaussian distribution

Chapter 3

Parameter estimation

3.1 Prediction vs inference

Prediction vs inference

3.2 Parameter estimation

Parameter estimation

3.3 The law of large numbers

The law of large numbers

3.4 The central limit theorem

The central limit theorem

Chapter 4

Introduction to statistical inference

4.1 Prediction vs inference

4.2 Hypothesis testing

Probability theory is one of the oldest subjects within mathematical studies. Paradoxically, almost every topic explained nowadays in modern courses of statistics is extremely new, ranging broadly a century. (...)

So let's first ask ourselves the question. What is probability in the first place? What do we mean by it and what does it describe? Probability is nothing more, and nothing less, that a number we make up, a quantity we come up with, to quantify certainty. A number we will use to describe the amount of information we have about a random, or stochastic, event. For simplicity, we can make it range from 0 to 1, in the following way.

4.3 Significance and p-values

Significance and p-values

4.4 Statistic tests

Statistic tests

Compare sample mean with hypothesized value - One sample t-test

The one-sample t-test is used to determine whether the mean of a single sample differs significantly from a known or hypothesized population mean. It assumes that the sample data are drawn from a normally distributed population. The test statistic is given by:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

where $\bar{x}$ is the sample mean, μ is the population mean, s is the sample standard deviation, and n is the sample size.

Compare sample means of two groups - Two sample t-test

Compare sample variances of two groups - Fisher test

Compare more than two groups - ANOVA

Compare distributions - χ^2 test

4.5   Parametric and non-parametric

Chapter 5

Introduction to bayesian statistics

5.1 The Bayes' theorem

The Bayes' theorem.

5.2 Bayesian vs frequentist

Bayesian vs frequentist.

5.3 Bayesian statistics

Bayesian statistics.

Chapter 6

Stochasticity and Markov processes

6.1 Stochasticity and Markov processes

6.2 Markov chains and hidden Markov models

6.3 Modern applications

Bibliography