# Probability Theory and Statistics Notebook I

Zhang Ruiyang

# Contents

# Preface

This notebook includes concepts for the topics of Probability Theory and Statistics at an early undergraduate level. The note is primarily written to help with my own understanding and studying of these topics, but it should be able to help others too. I will include my own interpretation in the notebook, and along the way, I will be addressing the historical backgrounds of some key ideas to make the Maths even more entertaining.

I referred to many sources and I do not take much credit from this work. They will not be cited separately through the course of the notebook, but I will list the main ones below.

1. *Introduction to Probability Theory* by *Dimitri Bertsekas & John Tsitsiklis*

2. *Introduction to Probability Models* by *Sheldon Ross*

3. *Markov Chain* by *J. R. Norris*

4. *Statistical Inference* by *George Casella & Roger L. Berger*

5. *The Concise Encyclopedia of Statistics* by *Yadolah Dodge*

Mathematics textbooks are normally written in a the more rigorous 'Definition-Theorem-Corollary-Lemma' structure. It is easy to follow that same structure here but I did not. Although there are a lot of benefits and elegance with that format, I do not think it will be a good fit for the purpose of this notebook, as I aim to focus more on the explanations and connections between concepts rather than solely stating and proving them. The drawback will be that concepts are harder to spot if one is not reading from start to end, but there is the Index page at the end to help. It should not be too bad.

In terms of the content of the notebook, it mainly aims to cover the material of the first two years for a degree course of Mathematics and Statistics. In the US system, this should be enough for all the undergraduate level Mathematics and Statistics, since Measure Theory is usually a graduate course and that is essential to study the Measure-Theoretic Probability Theory.

For Probability Theory, it normally has three components - probability theory (dealing with distributions, expectations, etc.), stochastic process (dealing with stochastic models)

and stochastic calculus. The first two components will be covered here, where the stochastic process component is covered to a lesser degree. Normally, stochastic process at this level can be a course on its own, normally called 'applied probability' or just 'stochastic process'. There are a lot of real-life applications of stochastic processes, like Renewal Theory and Branching Processes. These topics, although very interesting, may not appeal to a wider audience and they can be picked up fairly easily for interested readers after knowing the more fundamental materials in this notebook. Also, I find it very hard, and almost impossible, to have a thorough discussion of stochastic calculus without more advanced Mathematical tools, which is why I omit any discussion on that here.

For the Statistics component, it should be enough Statistics material for anyone using Statistical tools but not involved in Statistics research. Just like how Probability Theory has many other applications, Statistics, too, have those applications, like Social Statistics and Biostatistics. These topics will not be covered, to make the notebook more concentrated. Another reason is that I am no expert in these topics, and I am not confident enough to write about them yet.

Normally, programming languages like R and Python will be studied to help with some computations in Statistics and Probability Theory. Although they are extremely useful and are the trend in this era of booming computing powers, we will omit these topics. I feel these topics should be treated separately, for a programming notebook, instead of being discussed in a Maths and Statistics notebook.

For readers of this notebook, I do recommend you to use this as a complementary material to a standard textbooks. This is due to the fact that this notebook does not include any exercises and problem sets to the topics, which are essential for a thorough understanding of any concepts. I would even recommend the readers to be inspired by this notebook and write their own version, since different people have different preference and style of learning.

Since I am no expert in these topics and am learning them myself not too long ago, I am sure there are things that I write wrongly or explain poorly. Typos and grammar errors are bound to happen too, such things will exist no matter how often one reviews the work. In such cases, I hope the readers can contact me to point out these issues or discussion alternative ways to discuss the concepts. My email address is: ruiyang.zhang.20@ucl.ac.uk.

Lastly, I would like to express my most sincere appreciation to those Mathematicians and Statisticians who have made contributions to the field. All of us are standing on their shoulders, and hopefully in the future, we can join them and offer our shoulders for the future generations.

Zhang Ruiyang