

1. Introduction

1-1. Preface

I have no experience to learn mathematics or statistics in professional course in university, though I need to analyze obtained data for making report of my researches. On such occasion, I have to read various text books and, I feel an unpleasantness to the text books commonly used. Most of them started from the explanation of probability distribution of mutual exclusive event. That is binomial distribution. Readers should have concept of probability distribution, and introduction of binomial distribution is necessary. However, normal distribution is introduced suddenly in next chapter as extreme form of binomial distribution without prior knowledge. This makes confusion of readers, because the formulas are completely different and binomial distribution is a discontinuous distribution of discrete phenomena, though normal distribution is continuous distribution. Following the introduction of the concept of probability distribution, t-test is introduced as an example of application of probability distribution for the confirmation of statistically significant difference of mean value between two data groups. For this, we need a method to estimate population variance (variance of parent population) from variance of samples. Value of the population variance is an estimated value and fluctuate stochastically. Because of the fluctuation, the estimated average does not fluctuate in normal distribution, even the original parent population distribute in normal distribution. We learn that estimated average from sample population distribute in t distribution without any explanation of t distribution. There is no explanation why extreme form of binomial distribution is normal distribution or how we can obtain t distribution from normal distribution. We are requested to accept the dogma having no doubt. I was dissatisfied such lectures and text books.

I worked in a university as a professor and taught many students. I have compassion to lecturers and authors of the text books. I understand that there was a compelling reason. Normal distribution is obtained by polar coordinate transformation after superposition of two binomial distributions. Chi distribution is obtainable folding of normal distribution by squaring. T distribution is the result of multiplying of Chi square distribution and normal distribution, and F distribution is probability distribution of ratio of two chi square distribution. In order to make new probability distribution, we have to select combination of probability distributions and overlay the two integrations sterically, and then the volume of the distribution should be calculated by multiple integration. For this, we need knowledge of calculus which is lectured in lecture of mathematics in general education curriculum in universities and the process

is very long. It is difficult to finish the work without any mistakes. In addition to this, chi square analysis is not used in the statistical analysis of continuous variables and explanation of chi square distribution is not necessary at this stage. So, the explanation of chi square distribution is generally skipped at this stage. However, chi square distribution is a base of t distribution and F distribution.

Time is limited and there are mountains of topics to learn. The skips are unavoidable. In addition, I was lazy student who escape from the lecture of calculus. Probably, I could not understand the contents of the lecture, even if I had lectured the process of deriving normal distribution, t distribution and F distribution.

Statistics is only a skill for writing report, and some say that it is enough to understand how to use the skill. I agree that those opinion is a wisdom and it was enough old scientist as me. However, I think, such smattering understanding is not enough for future scientists. Technologies are rapidly advancing and speed of calculation by computer is reached unbelievable level. Now a day, even a laptop computer can finish the complicated calculation, which needed long time to be finished by hand calculation, within a second. Depending on the machine power, we can implement calculation which was impossible in old days by time limitation and many new concept and methods of statistical analyses are proposed and being popularized, such as most likelihood method in 1980s, Bayes statistics in 1990s, more recently MCMC so on. Soft wares for such analyses are obtainable through internet by free. This is not only the trend in natural science but also in social science. Many scientists in various fields of science are using computer for statistical analysis. There are many scientists who are familiar with recent statistical methods in social science fields. We can and should do various statistical analysis using computer, and should learn and understand various new methods in computerized statistical analyses, which are developed day by day. We have to manage flooding of the information. Learning each method as skill for solution of each specific topic is rather ineffective because of time limitation. It is better to learn from the basic for various applications.

You are not living in same age as me, who learned statistical skill without understanding of basic logics. I will not ignore basic explanation of each small step as basic policy. However, I was grown up in old days and is not good in mathematics. I may make mistakes in explanation. I ask readers who are good in mathematics to point out my mistakes for improvement of this text book. My purpose of writing this pdf text book is to make excellent text book in future.