Probability and Statistics Courses in the Universities in China

Yan Shi-Jian - Beijing, China

1. Introduction

Before the fifties, courses in statistics were offered only in a few specialities such as medicine and textiles in several universities in China. Courses in probability theory were even less common. Since the fifties, statistics courses have been offered in some social sciences specialities, but they emphasised the contents of socioeconomy and the intuitive analysis of comprehensive forms for reporting data, and did not go beyond uses of the mean value and standard deviation.

In the early 1960s, undergraduate courses in probability theory and mathematical statistics became more common. And in the eighties, any speciality requiring the study of differential and integral calculus also required these courses. Such specialities encompass almost all science and engineering specialities, certain specialities in agriculture and medicine, and a considerable number in sociology and economy.

So far, there has been a large difference between these two kinds of courses, not only in the teachers but also the contents. In Section 2 of this article, we give a brief introduction to the contents and general structure of the undergraduate requirement in probability theory and mathematical statistics, and in Section 3, the proposals from some experts in China will be presented. The efforts made by some teachers and some changes and opinions in some recent textbooks will be introduced in Section 4.

2. The contents and structure of courses in probability theory and mathematical statistics

At present, in most specialities, only one undergraduate course in probability theory and mathematical statistics is offered in Chinese universities. The contents and structure of the teaching materials are all quite similar, influenced by Gnedenko (1952) and Cramer (1946), and roughly as follows (cf. Fudan University, 1979; Zhongshan University, 1980; East China Normal University, 1983; Zhejiang University, 1979; Shen Heng-fan, 1982; Chen Jia-ding, 1982; Yan Shi-jian, 1982). The first part is probability theory and the second mathematical statistics. The probability part starts with the concept of probability, followed by elementary properties of probability, conditional probability, and several consequences. The next section is devoted to random variables, the main topics being discrete and continuous random variables; distribution functions and their properties; some important distributions; random vectors, independence of random variables, distribution of functions of random variables; and the deduction of some distributions which are commonly used in statistics. Of course, different specialities, texts, and teaching approaches may have the materials arranged in different order and depth. Following the distribution section, the concepts and properties of mathematical expectation, variance, and covariance are introduced. The last section on probability is on the concepts, properties, and elementary applications of the characteristic function, and includes the weak law of large numbers and the central limit theorem etc. Stochastic processes are generally excluded except for some very simple examples.

The part on mathematical statistics mainly illustrates statistical inference based on the contents of the probability part. Generally, the practical background and fundamental ideas and basic concepts of statistics come first, for example population, sample, histogram, and statistics. Very commonly, sampling methods are hardly touched upon. The second section deals with estimation: method of moments and method of maximum likelihood for point estimation; the unbiasedness, consistency, and efficiency of estimators; and interval estimation. Then come tests of statistical hypotheses, including z-test, t-test, χ^2-test and F-test. Finally, some comparatively practical topics such as quality control, regression, analysis of variance, and experimental design are introduced; in most cases, methods are stressed while the theories are barely elaborated. There is a great diversity in the arrangement of material on non-normal and nonparametric statistics, sufficient statistics, Cramer inequality, and asymptotic normality of MLE etc. in different textbooks and courses. In some textbooks, they are not even included at all. The total teaching periods are also different. The shortest is only 30-40 class hours (50 minutes each) and the longest may last 80-90 class hours.

The mathematical contents are emphasised quite heavily in these courses, even though measure theory is not used as the basis. Probably, the reasons are as follows: as the branch of mathematics dealing with stochastic phenomena, of course, probability theory should be treated in a way which stresses its mathematical contents. And for those students in specialities requiring statistics, probability theory, as the mathematical basis of mathematical statistics, should also be seriously studied. But I think there is still another factor which accounts for the wide acceptance of this arrangement in China, which is that most of the Chinese teachers giving such courses at university level have graduated from departments of mathematics. An advantage is that the concepts of probability and statistics can be understood more precisely by students. But in such a development, most of the class hours are used to introduce the mathematical concepts and proofs, and there is not enough time either to teach statistics or to give the students fundamental training in statistical thoughts and methods. Nor is there enough practice on computers. I think this situation mitigates against statistics playing an important role in the modernisation of China.

Stochastic processes are taught for mathematical undergraduate electives and statistics majors (cf. Wang Zi-kun, 1965; Wu Rong and Li Zhang-nan, 1987; He Sheng-wu, 1989). These courses may consist of introductions to Markov processes (Markov chains and Brownian motion as examples), martingales with discrete parameter, stationary processes, etc. Recently, some teachers taught a course similar to Ross (1983). Stochastic processes are also briefly introduced as required in some specialities, such as communications or automatic control theory. As for measure theory, it may be taught to statistics majors, mathematical electives, or as a part of real analysis to mathematics majors (cf. Yan Shi-jian et al., 1982; Xia Dao-xing et al., 1978).

The courses in probability and mathematical statistics for graduate students are very different, the former including courses in analytic probability, stochastic processes, stochastic analysis, mathematical statistics based on measure theory etc., the latter courses in statistical inference, linear models, multivariate analysis, time series, etc.

3. Proposals from some experts

In 1986, when a plan for the teaching programme was being discussed in the Division of Probability, Statistics and Operational Research of the appropriate committee in the Chinese State Education Commission, some of the experts suggested compiling new teaching materials. Some suggestions were: (1) "An Elementary Course in Mathematical Statistics", in which statistical methods and applications are emphasised paticularly, also the basic ideas of statistics, suitable for the specialities of economics, administration, and some related areas in liberal arts (such as education, etc.); (2) a new kind of textbook on probability and statistics, suitable for teaching mathematics in teacher's colleges; (3) one or two kinds of textbook on applied mathematical statistics, in which the fundamental concepts of statistics are stressed and statistical methods and applications are emphasised particularly, perhaps through case studies so the texts could cater for undergraduate students who had completed an elementary course in probability and statistics and were going to use statistics; (4) various kinds of textbooks for undergraduate electives and statistics majors on applied stochastic processes, in some of which measure theory would not be used as the basis, but many models and examples with practical backgrounds given, and some heuristic ideas and methods used in explanation, while others, which would be based on measure theory, should be written for undergraduate electives and statistics majors.

I think that these proposals may have originated from the following situation and viewpoints. In the view of Chinese visiting scholars coming back from abroad, and from the China-Japan Symposium on Statistics and Sino-American Statistical Meeting, it seems that Chinese scholars in probability theory and statistics are quite strong at mathematics, but weaker in statistical ideas and their applications. And we also lack some original ideas. Students educated in China are in the same situation. So real backgrounds and fundamental concepts should be emphasised in probability and statistics courses. Otherwise, students will feel alienated and fail to grasp the fundamental ideas and characteristics. And in mathematical statistics, the fundamental concepts should be expounded in detail (for example, to infer population values from observations, from which the problem of sampling arises, ...), and the students should be assisted to thoroughly understand the heuristic and original ideas underlying probability-statistics and to appreciate the practical significance of the conclusions.

For students to grasp the fundamental thoughts, they must not be distracted at first by formal mathematical theories; so measure theory need not be the basis in some elementary courses. For students who are going to apply statistics in other areas, we should develop their ability to solve practical problems through the use of probability and statistics. Here it is not enough for students to grasp only the mathematical theories and understand the practical backgrounds; they should also master a certain range of statistical methods. For this reason, we need to lecture on statistical methods and demonstrate the techniques and effects of applying these methods through examples (even case studies). In addition, for students in the areas of economics, administration, and liberal arts, who are relatively weak in mathematics, it may be necessary to compile teaching materials in which the statistical methods and their applications are emphasised particularly and the mathematical proofs are omitted.

As regards probability theory and stochastic processes, although they embrace extensive theoretical research and should be actively developed within mathematics, attention should also be paid to practical backgrounds. The reason is that probability theory and stochastic processes, as branches of mathematics, were set up comparatively late, but have rich and extensive practical backgrounds. Many research problems concerning them have been posed only very recently. Hence, a course in applied stochastic processes should be offered in universities, and many models and examples which have actual backgrounds should be discussed in lectures. This may not only arouse students' interest, but also help open up new domains of research in probability theory and stochastic processes.

4. Some characteristics of the teaching materials compiled recently

In the last three or four years, some teachers have made efforts to reform the textbooks on probability and statistics. For example, some teachers started compiling the teaching materials of applied statistics courses based on case studies (Xie Zhong-jie, 1988). Some teachers combined probability with statistics even from the beginning of their teaching material on probability theory and mathematical statistics (Zhang Yao-ting, 1984). In 1989, I compiled one such text for the mathematics departments of teachers' colleges, in collaboration with Professor Liu Xiu-Fang and Professor Xu Cheng-Yi (Yan Shi-jian et al., 1989). In this book we begin with the classical model of probability, immediately followed by the hypergeometric distribution, testing hypotheses for the numbers of failures, and estimating the population size using the maximum likelihood principle. After introducing multiple Bernoulli trials, we immediately develop the estimation of probabilities using the maximum likelihood principle, and introduce testing hypotheses of probability.

One purpose of this arrangement is to help students comprehend the fundamental ideas of statistics, and the logical relation between statistical inference and probability, through simple examples. Another is to make the students at teachers' colleges realise that in senior high school teaching, it is not only possible to explain some basic statistical methods using only sample mean and sample variance, but also to teach some statistical inference. When the concept of independence is introduced, we should relate it to statistical sampling. When the normal distribution is lectured on, we should take examples of histograms drawn from the sample data, to illustrate the extensive uses of the normal distribution and to show how to recognise heuristically whether a population has a normal distribution or not. And when we come to the distribution of functions of random variables, we should demonstrate the statistical backgrounds of the χ^2-distribution, t-distribution and so on. We hope such an arrangement may help students comprehend the concepts of statistics more precisely and better understand the applications of probability to statistics.

Since students find it hard to initially comprehend the general concepts of probability and statistics, in this book we do not introduce the concepts through general descriptive explanations; instead, we insist on beginning with the most concrete situations and then expanding and deepening progressively the concepts so discovered. For instance, in dealing with the concept of probability, we do not describe stochastic phenomena, events, and so on, generally at first. We try to induce the concepts and the calculating methods directly from examples of classical models (such as coin tossing). We then cite some examples of geometrical probability and explain them statistically. Finally, we make a summary and put forward the general concept of probability, which is an axiomatic definition. As for the concept of random variable, we point out certain quantities which can be quantified by numbers as early as when calculating probabilities in some classical models; for example, the quantity representing the number of occurrences of a head in coin tossing, etc.

The ideas in this textbook have been put into effect in our teaching practice, but they have not yet been consistently applied throughout the whole course.

5. Concluding remarks

As mentioned above, it is only recently that courses in probability and statistics have been comprehensively offered in the undergraduate programme in China. Although some experience has accumulated from teaching practice, the teaching methods still need further improvement. Many requirements are still far from being fulfilled, especially as regards the diversity of textbooks, the needs of different specialities, the treatment of applied topics, and training in consulting. A more essential issue is to build up a first-rate contingent of teachers in probability and statistics. Since 1979, nearly one hundred summer seminars have been held throughout the country by the probability and statistics circles for teachers in these fields in each summer vacation. And recently, a large number of students and young teachers have been trained and received Masters and Doctoral degrees in the two areas of probability and statistics. But raising the levels of academic research and teaching of this contingent of teachers is still a long-term and heavy task.

References

Chen Jia-ding and Liu Wan-ru (1982) *Lectures on Probability and Statistics* (1st ed) (in Chinese). Higher Education Press, Beijing.

Cramer, H (1946) *Mathematical Methods of Statistics.* Princeton University Press, Princeton, New Jersey.

East China Normal University (1983) *Course in Probability Theory and Mathematical Statistics* (in Chinese). Higher Education Press, Beijing.

Fudan University (1979) *Probability Theory* (Vol 1) (in Chinese). Higher Education Press, Beijing.

Gnedenko, B V *The Theory of Probability.* Chinese 1st ed, People's Education Press, Beijing, 1952; 2nd ed, Higher Education Press, Beijing, 1956; English edition, Chelsea, New York, 1967.

He Sheng-wu (1989) *Introduction to Stochastic Processes* (in Chinese). East China Normal University Press, Shanghai.

Ross, S M (1983) *Stochastic Processes.* John Wiley and Sons, New York.

Shen Heng-fan (1982) *Lectures on Probability Theory* (2nd ed) (in Chinese). Higher Education Press, Beijing.

Wang Zi-kun (1965) *Theory of Stochastic Processes* (in Chinese). Science Press, Beijing.

Wu Rong and Li Zhang-nan (1987) *Course in Stochastic Processes* (in Chinese). Higher Education Press, Beijing.

Xia Dao-xing, Yan Shao-zong, Wu Zhuo-ren and Shu Wu-chang (1978) *Functions of Real Variables and Functional Analysis* (in Chinese). Higher Education Press, Beijing.

Xie Zhong-jie (1988) *Case Studies in Applied Time Series Analysis* (in Chinese). Lecture Notes of the Department of Probability and Statistics, Peking University, Beijing.

Yan Shi-jian, Liu Xiu-fang and Xu Cheng-yi (1989) *Probability Theory and Mathematical Statistics* (in Chinese). Higher Education Press, Beijing.

Yan Shi-jian, Wang Jun-xiang and Liu Xiu-fang (1982) *Foundations of Probability Theory* (in Chinese). Science Press, Beijing.

Yan Shi-jian, Wang Jun-xiang, and Xu Cheng-yi (1982) *Fundamentals of Probability Theory and Mathematical Statistics* (in Chinese). Shanghai Science and Technology Press, Shanghai.

Zhang Yao-ting (1984) *Probability and Statistics* (1st ed). Central R and TV University Press, Beijing. (2nd ed, 1986)

Zhejiang University (1979) *Mathematics for Engineers - Probability Theory and Mathematical Statistics* (in Chinese). Higher Education Press, Beijing.

Zhongshan University (1980) *Probability Theory and Mathematical Statistics* (in Chinese). Higher Education Press, Beijing.