STATISTICAL TRAINING AT THE UNIVERSITIES—SOME VIEWS

S. RAO JAMMALAMADAKA
Department of Statistics and Applied Probability
University of California, Santa Barbara (USA)

1. INTRODUCTION

Statistics, as a discipline which incorporates inductive reasoning, is one of the maj building blocks on which new scientific advances are based. Statisticians help scientists a others make a guarded leap of faith from the (observed) sample to the (incompletely observe entirety and thus play a pivotal role in the progress of science.

In dealing with uncertainties around us in everyday life, from weather prediction forecasting the performance of our investments, we all "play statistician". So it is ve appropriate that everyone of us acquire some "statistical literacy"—just to be reasonab informed citizens. This exposure to basic ideas of probability and statistics should take place high schools, and it is gratifying that this is indeed happening in many countries. This topic h been adequately dealt with elsewhere (see e.g. Råde (1981)) and I shall not dwell on it here.

My objective is to share with you some ideas regarding statistical training at the universi level, i.e., in training Bachelors, Masters and Ph.D. degree students. These views evolved ov a period of twenty-five years during which I have been involved in statistical education at t university level—mostly in the USA—and from my own statistical training which has been the Indian Statistical Institute, Calcutta, under such people as C.R. Rao, P.C. Mahalanobis a J.B.S. Haldane—to name just a few.

I believe that, irrespective of whether one is training a student for a Bachelor's degree o Ph.D. degree, the training in statistics should always involve an appropriate mixture of bo theory and applications. Statistics derives its strength, usefulness as well as research motivati from applications and from the practical problems arising in other fields. It is the most inte disciplinary or multi-disciplinary subject. Given this, our main challenge is to find this "prope mix or balance between theory and applications. This balance, in my view, depends very mu on what we are training the students for—training people to fight the fires or training them invent and build new fire-fighting tools. We do, of course, need both kinds of training. majority of those we train at the universities at the Bachelors and Masters level work for indust and government and may be classified as the "firefighters". They should know how to use t given tools effectively and know their limitations. These students should concentrate more acquiring a broad knowledge of the statistical techniques along with the applications (includi computing, consulting and communicating) and less on the theory. On the other hand, one needs to understand a large body of theory and possess good mathematical skills to develop new statistical tools. However, even these "theoretical-types" who are working say for the Ph.D. degree, should constantly bear in mind the immediate or at least potential applications of these new statistical tools. These new tools are meant not so much to enrich and embellish the subject of statistics but to solve practical problems dealing with the real world.

Although this viewpoint is not new, I believe it bears repeating and will be addressed in a bit more detail in the next section. In Section 3, these ideas are given a more concrete shape in the form of an outline of courses etc. and related to our experience at the University of California, Santa Barbara (UCSB), in starting the B.A./B.S. degrees in Statistical Science, M.A. (Statistics) and Ph.D. (Statistics).

2. THEORY AND APPLICATIONS

Mahalanobis (1965), echoing a theme of Fisher, describes statistics as a "key technology"—much like medicine or engineering. He distinguishes science as "the effort to know nature more adequately" and technology as "the effort to use scientific knowledge for the fulfillment of specific purposes" and says further that "A scientist may have a wide or narrow range of interest or specialisation. (But) a technologist must have knowledge and experience of a wide range of scientific subjects". He quotes Fisher who pointed out that, "a professional statistician, as a technologist, must talk the language of both theoreticians and practitioners". I fully agree with this viewpoint and believe we should implement this philosophy in training statisticians, to the maximum extent possible. However from my own personal experiences as a student, I should caution against the possibility of getting carried away too far, in providing this broad-based education to statisticians. In 1960, I was among the first batch of students selected for a Bachelor of Statistics (B.Stat.) degree at the Indian Statistical Institute—a degree for training "professional statisticians with a broad base", formulated by Fisher, Mahalanobis, Haldane and C.R. Rao among others. A somewhat detailed description of this degree program is contained in the article "A multidisciplinary approach for teaching statistics and probability" by C.R. Rao (1969). While this degree program has been highly successful (and continues on a more modest level, even today), it might have been overly ambitious in trying to expose students to the multitude of subjects where statistics finds applications. This meant the students, besides learning mathematics and statistics, had to learn the quantitative techniques and the methodology used in a wide range of subjects including economics, biology, geology, genetics, engineering, chemistry, physics, psychology, sociology, demography etc. etc. To top it off, Professor J.B.S. Haldane taught us a course which he simply termed "General Science"—but to us, it meant an endless hopping from one subject to another—from discussions on space ships to

3/7

mining, from gerontology to philology, organic chemistry to microbiology and on and c
was certainly a most laudable attempt at training statisticians with a broad base—a
ambitious one.

On a much more modest scale, I advocate that all undergraduate training in s
should incorporate at least one (preferably two or more) application area(s). If one acc
idea that statistics is a "technology" like medicine or engineering, one should develop a
speciality in one or two application areas, like an electrical engineer or a pediatrici
Statistical ideas could be learnt in the context of this speciality and this provides both mo
and a concrete setting for developing the statistical theory and concepts. For instanc
U.S. universities, this can be achieved at the Bachelors level through what is called a "
major"—majoring in statistics alongside say economics or biology. If it takes a bit l
obtain a double-major, so be it. One does not get a degree in medicine or engineering
putting in some extra time and effort.

In the U.S. where students typically spend four years to obtain a Bachelors d
believe the students should spend approximately equal amounts of time and effort
during the final three years—the first year is mostly lost to general education requirem
(a) mathematics, (b) computing, (c) statistics and (d) application area(s). In mathema
should have a year of calculus and a year of linear/matrix algebra. If time permits ad
courses on advanced calculus, real and complex analysis, numerical analysis,
mathematics and operations research should be recommended. Students should have
mathematical ability to work with e.g. Jacobians of transformations, linear models i
notations etc. Under (b), they should be proficient in computers and be able to write p
in one or two languages. Also to be included here is a course on "statistical computing
teaches students at least two software packages for statistical analysis. Given the curren
place, the more the students pack in here, the more employable they are. In statistics,
courses should include (i) a one-year course on probability and mathematical statistic
typically includes derivations of sampling distributions and parametric and some nonpa
inference, (ii) a course (of at least half-year duration) on linear models and generaliz
models with discussions on regression and analysis of variance and (iii) at least a ba
course on sample surveys and on design of experiments. Other optional courses offer
include time-series analysis, nonparametric methods, quality control and reliability, si
methods. Finally, under (d), application areas, students should take some "serious co
the subject area(s), specifically those that focus on quantitative methodologies use
area(s).

At the Masters degree level, it should be possible to separate students into at l
main streams depending on their goals and interests. For instance, an Applied Statistic

and a Mathematical Statistics stream. While the Applied stream trains students for jobs in government and industry (the "firefighters"), the other stream emphasizes more on theory and mathematics and prepares them for a Ph.D. Students doing Masters in applied statistics should possess a broad knowledge of a wide range of statistical techniques and methods—applied multivariate analysis, categorical data analysis, some biostatistics including survival analysis, sampling methods, time-series etc. They should also be able to effectively utilize statistical packages for data analysis, be able to solve general statistical problems from industry or government as consultants, and communicate with other scientists. At UCSB, we found that students who obtain their B.A./B.S. degrees in one of the substantive areas (e.g. biology or psychology) are equally successful at pursuing this stream, if they are willing to make up the deficiencies. There is an excellent report by the American Statistical Association Section on Statistical Education on this type of training (ASA (1980)) that I would recommend to your attention.

The other (Mathematical Statistics) stream consists of potential researchers—those who intend to obtain a Ph.D. in statistics. Their coursework should emphasize more statistical theory (optimality of various procedures, nonparametric and robust methods, decision theory etc.) and additional courses in probability theory, stochastic processes as well as other relevant mathematics courses.

Students who wish to pursue a Ph.D. degree should develop considerable breadth in statistics and familiarity with applications, before they are allowed to narrow down to their chosen research topic. This might be achieved, for example, through various course requirements and/or "qualifying" examinations. It is unfortunately still possible to get a Ph.D. in statistics in most U.S. universities and not ever be exposed to basic ideas of survey sampling, design of experiments, time-series, computing or even real applications. I quote here again from Mahalanobis (1965) about a Ph.D. from a well-known U.S. university who was seeking a job in India: "the results of his investigation were given in an abstract mathematical form and their implications were not immediately clear to me. As the post was for work mostly at an operating level, I asked him whether he had any occasion to use his results in practice. He was surprised at this question and said 'no'. I asked him whether he knew or could think of any situation in which his results could be used in practice; he thought over my question for a while and then again said 'no'." This state of affairs has changed very little. To quote from Box (1980), "the student may receive verbal and nonverbal messages that tell him that the kind of Ph.D. thesis that is motivated by a scientific problem with real data is not likely to be so acceptable as a pretty narrowly introspective mathematical exercise. Again the junior faculty member knows that he has only a few short years to make it to tenure. In many cases he will be judged by faculty who have no understanding of, or sympathy for, application. He may suspect that when his publications are assessed, those with any taint of application will be down-weight entirely discounted." Students often get a Ph.D. degree in statistics for abstract exe "statistical mathematics" which have no connection to real data analysis. The thesis co and supervisors should make certain that the research topic the student chooses to work potential use if not of immediate applicability. This test of "applicability of results" al "knowledge of computing", should be made requirements for all students doing th degree in statistics.

Students of statistics, no matter what the level, should get this sense that statistical well as the need for new theory is driven by applications. The best place to tea applications is alongside the relevant theory regardless of whether it is an eleme advanced course. Teachers do not always find this easy or convenient. Most of us tossing coins and dice, for 'concrete' examples. It seems that the training (or retraining actually begin at the faculty level, with the faculty developing interests in some applicati This can be achieved through joint appointments or by faculty from statistics department on temporary assignments to government or private agencies, to acquaint themselves real world problems. Participation in a statistical consulting laboratory at the university can be both an educational tool for the students and an extremely useful resource for th the university) and/or even private consulting are other ways for statisticians to keep with reality.

3. SOME EXPERIENCES AT UCSB

Courses in statistics have been offered at the University of California, Santa (UCSB) for over 15 years, mostly as part of the Mathematics curriculum. In the Fall of separate "Department of Statistics and Applied Probability" was created at UCSI Department now has close to 14 faculty members and offers:

(a) B.A. and B.S. in Statistical Science;

(b) M.A. (Statistics) with three streams leading up to Applied Statistics, Mathe Statistics and Operations Research specializations; and

(c) Ph.D. (Statistics).

The two Bachelors degrees require a substantial amount of coursework in mathe computing and statistics, as described in the earlier section. However the degrees do not time, require coursework in application areas that I discussed before. We hope we can in several double-majors, as for instance, statistics and economics, statistics and computer s statistics and biology. The B.A. degree requires 10 one-quarter (a quarter consists of 10 or 30 fifty-minute sessions of instruction) courses in statistics at the upper division (3rd year) level, while the B.S. degree requires 13 such one-quarter courses. Following is ar abbreviated list of courses (all are one-quarter, unless mentioned otherwise) categorized as:

Requirements: Probability and Mathematical Statistics (3 quarters)
Design and Analysis of Experiments
Regression Analysis
Statistical Computing

Options: Sampling Techniques
Statistics in Industry
Sequential Methods
Nonparametric Methods
Operations Research
Ranking and Selection Methods
Applied Stochastic Processes (2 quarters, includes some time-series)
Actuarial Statistics and Risk Theory (3 quarters)

Others: Internship in Statistics
Independent Studies in Statistics

While the "independent studies" course is a vehicle for learning any other topic that is nc taught as a regular course, the internship encourages students to do faculty-supervised academi internship in industrial or research firms in the area. This is an excellent way for students t relate the academic coursework to real problems that exist out there.

An M.A. degree with any of the three specializations requires about 11 one-quarte courses (mostly graduate with a few approved upper division courses allowed). An abbreviate list of graduate courses (all are 3-quarters unless mentioned otherwise), currently offere include:

Statistical Theory
Multivariate Analysis
Statistical Decision Theory
Linear Models
Life Testing and Reliability (one quarter)
Operations Research
Case Studies in Operations Research (one quarter)
Advanced Statistical Methods
Seminars and Projects in Statistical Consulting (one or two quarters)
Probability Theory
Advanced Probability Theory
Stochastic Processes
Advanced Stochastic Processes

Then we have a catch-all course called "Seminars in Probability and Statistics", where tl

3/7

topics (generally advanced) vary with the instructor's interests.

Students in M.A. (Applied Statistics) take the Advanced Statistical Methods cou covers a wide range of topics (not always with formal proofs) and have to do an inte the statistical consulting lab for a quarter or two. Here they participate in actual consult supervision, and write a project report on their statistical analysis and conclusions where the students are exposed to the 'inconvenient' real-world problems which ma traditional assumptions, have missing data and may indeed call for the developmen theory. The operations research stream requires a similar case-study project after th year-long graduate course in operations research. Students in the Mathematical Statisti take the more theory-oriented courses (statistical theory, probability theory, etc.). Stu encouraged to keep up their interest in an application area by providing academic credi courses taken outside the department.

To make sure that the Ph.D. students have the appropriate breadth, they "qualifying examinations" in mathematical analysis (real and complex variables), compulsory and in any three of the following five areas: Mathematical Statistics Statistics, Operations Research, Stochastic Processes and Probability Theory. We still foreign-language requirements and one of them can be fulfilled by demonstrating profi a computer language—almost all the students opt for this and do a rather substantial e computing. Even those who opt for a very theoretical piece of work for their Ph.D. th to answer my colleague Joseph Gani's familiar question during their thesis defence. apply your results to any data sets, real or simulated?" So all in all, I believe we have training program that needs only minor fine-tuning and is mostly in line with m recommendations.

4. CONCLUDING REMARKS

Although there is no unique formula which works for training all categ statisticians, there is, I believe, general agreement that statistical training as well as research, should be driven by applications. We are in the fortunate position of bein motivate all the theory we teach, through real-world applications and we should co synergy between theory and applications to our students. A good understanding of t along with its limitations, an appreciation for applications and familiarity with " computing" should form the basic themes for any university degree in statistics.

Statistics is a discipline which is growing both in scope and depth, especially d last few decades. This makes it imperative that we re-examine the courses and curri periodic basis and update them, as necessary. For instance, the tremendous pow computer which is now at our disposal, has revolutionary implications for our discipline and we should take advantage of this. Computer intensive methods like resampling techniques including bootstrapping, iteratively reweighted least squares, density estimation, simulation, dynamic graphics, etc. have all become topics of great interest. See for instance Diaconis and Efron (1983), Speed (1984). Also it is possible to do a large part of data-scrutiny and exploratory data analysis through interactive statistical graphics packages, i.e., plot, look for outliers and try various transformations etc. See for instance Tierney (1990). We should incorporate these topics as well as a host of new and emerging theoretical fields like inference for stochastic processes, robustness and influence functions, directional data analysis, spatial statistics, image reconstruction, interacting particle systems, etc. into our curricula alongside the more traditional courses that we already offer.

BIBLIOGRAPHY

1. American Statistical Association (1980): Preparing Statisticians for Careers in Industry: Report of the ASA Section on Statistical Education—Committee on Training of Statisticians for Industry, *The American Statistician*, 34, 65-75.

2. Box, G.E.P. (1980): Comment on the above article, *The American Statistician*, 34, 76-77.

3. Diaconis, P. and Efron, B. (1983): Computer intensive methods in statistics, *Scientific American*, 248, 96-108.

4. Mahalanobis, P. (1965): Statistics as a key technology, *The American Statistician*, 19, 43-46.

5. Råde, L. (editor) (1981): *Proceedings of the Gothenburg Symposium on Teaching Statistics*.

6. Rao, C.R. (1969): A multidisciplinary approach to teaching statistics and probability, *Sankhyā, Ser. B*, 31, 321-340.

7. Speed, T.P. (1984): Teaching statistics at university level: How computers can help us find realistic models for real data and reasonably assess their reliability, in *Teaching of Statistics in the Computer Age* (Ed. L. Råde and T. Speed), Chartwell-Bratt Ltd.

8. Tierney, L. (1990): LISP-STAT: *An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*, Wiley, New York.

3/7

SUMMARY

This paper presents some views regarding the training and academic prepara statisticians at the University level. Statistics is a "key technology" which helps solve pr arising in a wide variety of disciplines. Therefore our goal should be to train students only understand the mathematical methods and the theory but also have a clear appreci the practical problems and applications in at least one or two subject areas. The proper between theory and applications will, of course, vary depending on the students' go needs. The argument is made that there should be no dichotomy splitting the profess theoretical statisticians and applied statisticians. There is no room in the profession for theoretical statistician whose work has no relevance to the real world or for one who grasp the basic theory, assumptions and limitations behind the techniques. Some gene regarding the coursework and a comparison with the degree programs at the Unive California, Santa Barbara, are provided.

RESUME

Cet article traite des idées concernant la formation des statisticiens. Notre but devrait etre d'enseigner aux étudiants à comprendre la théorie mathematiques et à apprecier les problèmes pratiques. On affirme que la profession ne support pas une dichotomie entre ceux qui s'interessent à la théorie et ceux que s'interessent à la statistique appliquée. Finalement, a décrit quelques idées concernant les cours à l'Université de Californie, Sa Barbara.