Sankhyā: The Indian Journal of Statistics
2000, Volume 62, Series A, Pt. 1, pp. 154–156

Mathematical Statistics:
A Unified Introduction by George R. Terrell.
Springer Texts in Statistics
Springer Verlag, New York
ISBN: 0-387-98621-9
Pages: (ix) + 453

This is an attempt at presenting, in a unified way, probability and statistics at an introductory level. The thirteen chapters are on Structural Models for Data, Least Squares Method, Combinatorial Probability, Other Probability Models, Discrete Random Variables I and II, Random Vectors and Random Samples, Maximum Likelihood Estimates for Discrete Models, Continuous Random Variables I and II, Continuous Random Vectors, Sampling Statistics for the Linear Model and, Representing Distributions. Each chapter has a reasonable number and variety of problems. Hints and solutions are given to most of them at the end of the book.

You may have noticed the unusual sequencing of chapters. This is an effect of the author's conviction that "Statistics precedes probability". He starts off with the discussion of models for one and two way layouts, regressions (including multiple and logistic) and independence models for contingency tables in Chapter 1. Chapter 2 introduces the principles of least squares which is then used to motivate sample variances, covariances and correlations. A natural follow up is decomposition of the sum of squares in ANOVA and some insight into the idea of degrees of freedom. Chapters 1 and 2 do not use any probability considerations. In particular, the discussions are free of any models for errors.

Chapter 3 is on combinatorial probability with emphasis on the useful principle that all probabilities are conditional.

Chapter 4 uses geometric examples to introduce continuous probability models, considers a set of adequate probability axioms (finitely additive), does Bayes' theorem, and has some discussion on Kolmogorov's axioms and Borel sigma algebra.

Chapter 5 introduces discrete random variables with finite population sampling, giving rise naturally to the negative hypergeometric family. The first example of a test, Fisher's exact test in 2×2 tables, arises as an application. Expectation is introduced and is immediately put to use to define a moment estimate. This in turn becomes a convenient vehicle to introduce the idea of confidence bound and confidence level.

Chapter 6 introduces the geometric, negative binomial, binomial and Poisson families. The connection, via asymptotic limits, between these and the negative hypergeometric are explored. General discrete expectation and variance definitions are introduced. Estimation of binomial p is considered and it leads to the general definition of confidence bounds and of two sided hypothesis tests.

Chapter 7 introduces random vectors, conditional expectations, covariances, correlations and the formula for expectations of linear combinations. Rudiments of Bayesian inference find a place in this chapter. There is also a small section on Markov's inequality and probability convergence.

Chapter 8 investigates estimation and goodness of fit in discrete models, log-likelihood ratio statistics and its relation to Pearson's chi-squared statistics, maximum likelihood in loglinear and logistic regression models.

Chapter 9 is on continuous random variables, starting with a Poisson process and building up the gamma and the beta family from that - spacings, transformations, ordered statistics arise here naturally. Inference about the shape parameter leads to likelihood ratio test, idea of a most powerful test and to families with monotone likelihood ratio.

Chapter 10 starts with the problem of generating variables with given distribution. Thus comes into picture the quantile function. Expectation for any random variable is now defined via the quantile function. The "usual" expectation definition is then shown to be equivalent. Approximation to gamma density is derived and this leads to the definition of a normal density. Using relation between Poisson and gamma, normal approximation to Poisson is derived and continuity correction is explained in a nontraditional way.

Chapter 11 is on multivariate distributions and does Dirichlet, chi-squared, and bivariate normal. Taking advantage of the multivariate nature of the chapter, it introduces the idea of conjugate priors.

Chapter 12 is on statistics that arise in linear models. It discusses spherical errors, maximum likelihood with normal errors, and likelihood ratio statistics in linear models. The general linear model is introduced with matrix notations and the Gauss-Markov theorem is proved. Fisher's information and MVUE are defined.

Chapter 13 starts with the idea of compounding distribution and uses it to get to probability generating functions. Given a random variable X with support $\{0, 1, 2, \ldots\}$, perform X independent trials each with success probability "t". Then probability that all trials are successes is $E(t^X)$. The moment generating function (for nonnegative X) is introduced as $P(Y \geq X)$ where Y is independent of X and is negative exponential. Limits of generating functions is briefly discussed. Let $\{X_i\}$ be iid with mean μ and suppose the moment generating function exists in a neighbourhood of zero. It is shown that $Ee^{t\bar{X}} \rightarrow e^{t\mu}$. This is labeled as a law of large numbers.

The central limit theorem is done via the m.g.f. The form of the log likelihood of the binomial is used as a motivation for defining a natural exponential family. Some ideas of sufficient statistics and Rao-Blackwell theorem are covered. The chapter ends with a brief coverage of exponential tilting and normal and Poisson tail approximation.

Some of the things I liked are: (i) the emphasis that all probabilities are conditional (ii) bounds on the accuracy of approximate probability calculations throughout the book, starting early with the birthday problem (iii) emphasis on and exploitation of symmetry and duality in probability experiments (iv) putting the contents of the first two chapters at the beginning, although, I suspect some of the material there may be hard to put across to the beginners.

I felt uneasy about no mention of probability trees. These could have been used for introducing conditional ideas, bivariate sample spaces/random variables and even to do Bayes theorem. If early on it can be established that $g \leq h \Longrightarrow Eg \leq Eh$ then it would give all the probability inequalities. The linearity formula (even in bivariate case) could also be established early on. On the binomial confidence interval problem (page 196) there is of course the standard observation that replacing $p(1-p)$ by $1/4$ would remain conservative. Probability space has been introduced as some algebra of events (pages 121, 124 and 139). The author's wish of avoiding a "triplet" definition of random variable is understandable but what appears as a definition (page 146) "A random variable is a probability space whose outcomes are real numbers" is likely to confuse students. Couple of other obvious oversights which I spotted and which can be rectified easily are:

(i) Page 147, "with this random variable the probabilities of different numerical outcomes are not all the same, so it is an example of a discrete random variable".

(ii) Page 415, after the "law of large number" proposition (see above) the author writes: "... convergence in m.g.f. means that we also have convergence in distributions. Therefore, the existence of a variance really will turn out to be unnecessary in laws of large numbers". The existence of m.g.f. (in a neighborhood of 0) already implies finiteness of the variance.

Moreover, if the m.g.f. exists, it is rather easy to prove that $\bar{X}$ converges to μ in distribution. Simply use Markov's inequality appropriately.

The book does not carry a bibliography. You will find somewhat different flavour and style in the way your personal favourites are introduced and discussed here. Most likely this will not change the fundamental ways you approach the subject. But it may bring some added insight and incline you towards some adjustments in your teaching.

Indian Statistical Institute and Purdue University Arup Bose