Classification Algorithms.

Mike James. New York: John Wiley, 1985. xii + 211 pp. $34.95.

This book is a clearly written exposition of basic classification methods. The emphasis is on ideas, not derivations, so very little math beyond some probability and matrix algebra is required. The general format is that a topic is introduced and discussed mainly in words, a computer program listing written in BASIC is given and discussed, and the computer output from an example (usually Fisher's iris data) is presented. There are no exercises. The audience that would find this book useful would be (a) owners of SAM (Statistical Analysis for Microcomputers), which contains all of the programs given here, who are looking for an extended explanation of the classification program or (b) practitioners in application areas who are looking for a good overview of the classification area. One of the book's few shortcomings is the limited number of references. Suggestions for further reading as given by Hand (1981) would be extremely useful in a book of this type.

The topics covered include linear and quadratic discriminants, nonparametric methods, and methods for use with discrete data. The main emphasis is on minimizing the total error of classification, but other criteria are briefly discussed. There are chapters on canonical analysis and stepwise variable selection. The discussion of estimating error rates is quite good and includes the naive estimate, the leave-one-out method, and splitting the data into a training set and a test set. The bootstrap method is not discussed. The final chapter discusses changes in the view of the classification problem necessitated by pattern recognition and artificial-intelligence problems.

A program to generate multivariate normal vectors is given in an appendix, but should be avoided because it uses the infamous sum of uniforms to generate normals. For superior methods, see Knuth (1981, pp. 117–127) or Box and Muller (1958).

The book covers all of the important beginning topics in identifying classes to which objects belong. The explanations are extremely clear and make good use of geometric ideas, they can quite easily be read without the computer sections to give a good overview of the subject. The computer sections are also clearly explained so that someone wanting to understand the methods involved can do so. The amount of material presented in the noncomputer section is extensive considering the book's brevity.

RONALD C. PRUITT
University of California, Davis

REFERENCES

Box, G. E. P., and Muller, M. E. (1958), "A Note on the Generation of Random Normal Deviates," *The Annals of Mathematical Statistics*, 29, 610–611.

Hand, D. J. (1981), *Discrimination and Classification*, New York: John Wiley.

Knuth, D. (1981), *The Art of Computer Programming* (Vol. 2, *Seminumerical Algorithms*), Reading, MA: Addison-Wesley.

Stochastic Differential Systems (Lecture Notes in Control and Information Sciences, Vol. 78).

N. Christopeit, K. Helmes, and M. Kolmann (eds.). Berlin: Springer-Verlag, 1986. vi + 372 pp. $28.20.

This volume contains the proceedings of the Third Conference on Stochastic Systems Control and Filtering organized by the University of Bonn. It covers a great variety of topics within the field; only a few can be listed here: nonlinear filtering for diffusion processes on manifolds, smoothing problems for point-process observations, optimal control of diffusions and reflected diffusions with state-space constraints, optimal control for nondegenerate diffusion models under the average-cost-per-unit time criterion, approximations and limit theorems for stochastic flows and for parabolic equations with wide-band noise driving terms, approximation methods for optimal controls under partial observations, existence results for optimal feedback controls for degenerate diffusion models, solutions of specific optimal control problems such as those arising in investment/consumption theory or in overload control of telephone networks. There are several papers (including a survey) in one of the main subareas of the conference—the points of intersection between stochastic analysis and quantum theory (along the lines of Nelson's stochastic mechanics). Generally, the papers represented the current state of stochastic systems theory.

HAROLD KUSHNER
Brown University

Point Processes and Their Statistical Inference (Probability: Pure and Applied, Vol. 2).

Alan F. Karr. New York: Marcel Dekker, 1986. xi + 490 pp. $89.75 (U.S.A. and Canada); $107.50 (all other countries).

This book is a research monograph on inference problems for point processes that focuses on limit theorems for estimators. Most of the literature on point processes concerns point processes on the real line that model times of occurrences of a certain event (e.g., times at which data packets arrive to a computer, or times at which a stochastic process enters a special state). Point processes on Euclidean or other topological spaces that model such phenomena as locations of galaxies or locations or random elements (functions, sets, lines, etc.) are generally more difficult to analyze because the state space is not totally ordered. Point processes are sometimes viewed as special random measures, random sets, or random fields. The major types of point processes and the decades of their primary development are Poisson processes (last 150 years), renewal processes (1950s, 1960s), stationary point processes on groups such as R^n (1960s, 1970s), point processes on locally compact Hausdorff spaces with countable bases (1970s), and point processes on R_+ with attendant marks and martingale dynamics (1970s, 1980s). This book addresses inference problems for each of these.

Chapters 1 and 2 are devoted to defining these processes, and Chapter 3 is an introductory discussion of the inference problems that lie ahead. The main results begin in Chapter 4, "Empirical Processes Associated With Point Processes." Suppose $N_1, N_2, \ldots$ are iid versions of a point process N on a compact space E (N takes values in the space $\mathfrak{N}$ of all counting measures on E and $N(B)$ is the number of points in $B \subset E$). The *empirical random measure* $\hat{\mathfrak{L}} = n^{-1} \sum_{i=1}^{n} \varepsilon_{N_i}$ is the maximum likelihood estimator of the probability law $\mathfrak{L}$ of N [$\varepsilon_N(\Gamma) = 0$ or 1 according to whether N is or is not in $\Gamma \subset \mathfrak{N}$]. In addition, if $H(\mathfrak{L})$ is a functional of N (or $\mathfrak{L}$), then $H(\hat{\mathfrak{L}})$ is a natural choice for an estimator of it. For example, a natural estimator of the Laplace functional $E_{\mathfrak{L}}(e^{-Nf}) = \int_{\mathfrak{N}} e^{-\mu f}\, d\mathfrak{L}(\mu)$, where $\mu f = \int_E f(x)\, d\mu(x)$, is the *empirical Laplace functional* $n^{-1} \sum_{i=1}^{n} e^{-N_i f} = \int e^{-\mu f}\, d\hat{\mathfrak{L}}(\mu)$. Other functionals of interest are the zero-probability functional $P_{\mathfrak{L}}(N(A) = 0)$ (which characterizes $\mathfrak{L}$ for simple point processes), reduced palm distributions, and thinned point processes. The focus of this chapter is the application of the theory of empirical distributions on general measure spaces to describe the limiting behavior as $n \to \infty$ of the difference between $\hat{\mathfrak{L}}$ and $\mathfrak{L}$ and between $\mathcal{H}(\hat{\mathfrak{L}})$ and $\mathcal{H}(\mathfrak{L})$ for the preceding functionals.

Chapter 5 deals with the general theory of intensity-based inference. Suppose $N_t(k) = \sum_n 1(T_n \le t, Z_n = k)$ is a *marked point process* (adapted to an increasing family of σ fields) with the times $0 < T_1 < T_2 < \cdots$ having attendant discrete marks $Z_1, Z_2, \ldots,$ and with *compensator* $A_t(k) = \int_0^t \lambda_s(k)\, ds$, where $\lambda_t(k)$ is the *random predictable intensity*. The process $N_t(k) - A_t(k)$, $t \ge 0$, is a zero-mean martingale for each k. Suppose the intensity takes the form $\lambda_t(\alpha, k) = H_t(\alpha, k)\lambda_t(k)$, where α is an unknown parameter and H_t is another predictable process. A special case is $\lambda_t(\alpha, k) = \alpha_t(k)\lambda_t(k)$. The focus here is on martingale estimators of the unknown α or law of N based on observations of the process on a finite time interval. This includes a filtering problem, hypothesis testing, and a Cox regression model with covariates.

Chapters 6–9 apply the basic results of Chapters 4 and 5 to analyze a variety of estimators for the most tractable point processes: Poisson, Cox (doubly stochastic or conditional Poisson), renewal, and stationary processes. Chapter 10 ends the book with a study of inference of Markov processes, stationary processes on R, and stationary random fields based on observations resulting from a point-process sampling mechanism. Each chapter contains several interesting exercises that supplement the main results.

This book would be ideal for an advanced graduate statistics course or seminar on inference for point processes. It could also serve as a supplement for a graduate probability course on point processes. The writing style is rigorous, clean, and balanced; the focus is on concrete, detailed inference models as well as on general principles.

RICHARD F. SERFOZO
University of North Carolina and
Georgia Institute of Technology

Empirical Processes With Applications to Statistics.

Galen R. Shorack and Jon A. Wellner. New York: John Wiley, 1986. xxxvii + 938 pp. $59.95.

For a sample of real-valued random variables $X_1, \ldots, X_n$, the empirical distribution function $F_n(x) = (\#\text{ of } X_i\text{'s} \le x)/n$ often summarizes the information in the sample. After centering by $F(x) = \Pr(X_i \leq x)$ and multiplying by $n^{1/2}$, we get the stochastic process (in x)

$$L_n(x) = n^{1/2}[F_n(x) - F(x)],$$

which is called the *empirical process*. This process has attracted much attention because interesting statistics are related to L_n in a natural way.

For example, the sample mean after standardizing may be written as

$$n^{1/2}(\overline{X} - \mu) = n^{1/2} \int_{-\infty}^{\infty} x d[F_n(x) - F(x)] = \int_{-\infty}^{\infty} x dL_n(x),$$

and the Cramér–von Mises goodness-of-fit statistic is

$$n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, dF(x) = \int_{-\infty}^{\infty} L_n^2(x) \, dF(x).$$

Now it is clear how to get the limiting distribution of $n^{1/2}(\overline{X} - \mu)$ directly from the central limit theorem without consideration of L_n. But the convergence of $\int L_n^2(x)\, dF(x)$ requires first an enlarged understanding of convergence in distribution of L_n—that is, "weak convergence" of L_n to a stochastic process L on $(-\infty, \infty)$. A continuity argument then gives

$$\int L_n^2(x)\, dF(x) \xrightarrow{d} \int L^2(x)\, dF(x).$$

Fortunately, $L_n(x)$ has the same distribution as $U_n(F(x))$, where $U_n(t)$ is the empirical process for a sample $U_1, \ldots, U_n$ of uniform [0, 1] random variables. Thus we can usually focus attention on this simpler *uniform empirical process* $U_n(t)$ and derive results from it.

This book then has two main focuses. First, the authors have attempted to give many of the important probabilistic results concerning $U_n(t)$ and its cousin $V_n(t)$, the uniform quantile process. These results include process versions of laws of large numbers, central limit theorems, laws of the iterated logarithm, exponential bounds, and so forth. Also included are the various Skorohod and Hungarian constructions, a survey of classical probability inequalities, and an introduction to the recent counting process methods used in censored-data situations. Various extensions involving general sample spaces are introduced in Chapter 26, but the details are left for the complementary books by Gaenssler (1983) and Pollard (1984).

Second, the authors show how to use weak convergence results for $U_n(t)$ to get important limit theorems for large classes of statistics such as those based on ranks, order statistics, and spacings. These are often elegant and among the best possible. Moreover, such applications do not require a deep understanding of weak convergence in order to be applied. In the authors' own words we find, "However, our interest is in learning to use the tool of weak convergence, not to rederive the basic theorems" (p. 44). This practical approach is possible because they emphasize the Skorohod construction that essentially converts weak convergence to almost sure convergence. Coupled with the authors' own pioneering work with q norms $\|U_n/q\| = \sup_x|U_n(x)/q(x)|$, many "weak convergence" proofs are reduced to straightforward calculus and Slutsky's theorem. For example, the proof of convergence of the Anderson–Darling statistic takes only several trivial steps to get a fairly nontrivial result (p. 148).

The book includes an incredible number of results and is therefore quite long, more than 900 pages. This is both a strength and a weakness. Often one feels that explanations are too terse, that applications are not well motivated. The too-simple labeling of theorems and sections makes life very difficult for the reader. There are six Theorem 1s in Chapter 3. To find Theorem 3.3.1, one must first find Chapter 3, then find Section 3, then realize that Theorem 1 is actually Theorem 3.3.1. This may be a standard approach, but I found it very hard to find referenced results. On the other hand, the short and expanded tables of contents are quite helpful.

Exercises are sprinkled throughout the book so that it can be used as a textbook. Since the level is quite high, it would be most appropriate for a seminar-type course for faculty and third- or fourth-year graduate students. The authors suggest that "the prerequisite is a standard graduate course in probability and some exposure to nonparametric statistics" (p. x). More specifically, I think that a good understanding of chapters 1 and 2 of Serfling (1980) and perhaps chapter 1 of Billingsley (1968) would also be helpful.

This book will be an important resource for mathematical statisticians and probabilists for years to come. The authors should be commended for bringing together such a large number of important results.

DENNIS D. BOOS
North Carolina State University

REFERENCES

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.
Gaenssler, P. (1983), *Empirical Processes* (Lecture Notes, Monograph Series), Hayward, CA: Institute of Mathematical Statistics.
Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer-Verlag.
Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Long-Range Forecasting: From Crystal Ball to Computer (2nd ed.).

J. Scott Armstrong. New York: John Wiley, 1985. xxiii + 689 pp. $22.95 (paperback).

To be useful, a textbook on forecasting has to recognize the unique nature of forecasting problems and the appropriate role of statistics in solving those problems. *Long-Range Forecasting* (LRF) does just that. It provides a practical, understandable approach to a discipline that has shown scant progress for much of the time since the days of the Greek oracles. The scope of topics covered in LRF includes economics, strategic planning, marketing, sales, communications, and—oh yes—statistics. Those students and teachers who feel uncomfortable venturing outside their chosen discipline will be annoyed and disturbed at this approach. This is hardly a problem, because those people probably should not get involved in forecasting in the first place.

In using a cross-discipline approach, Armstrong, on page 65, warns against becoming too much of a specialist ("one who knows more and more about less and less until eventually he knows everything about nothing"). He suggests on page 66 that, in forecasting, there are distinct advantages to becoming more of a generalist ("one who knows less and less about more and more until eventually he knows nothing about everything"). Armstrong believes that venturing beyond the immediate confines of one's own discipline is essential to successful forecasting.

The greatest contribution that LRF makes to would-be forecasters is its practical advice. My own experience after almost 20 years in the business confirms many of Armstrong's insights. These recommendations include recognizing the difference between a forecasting model and a measurement model, getting user inputs to the process, placing a premium on simplicity, and searching for disconfirming evidence. Forecasters could improve their product substantially by heeding such advice.

Armstrong advises statisticians who wish to play the forecasting game to recognize statistics as a bag of tools for accomplishing an objective. The greatest challenge for the statistician involves choosing the appropriate tool and recognizing that the tool is not the end but merely a means to an end. From a statistical standpoint, the most professionally satisfying approach to a forecasting problem is often the wrong approach. Hence, success in the statistical field may be incompatible with success in forecasting. One may have to choose between dazzling colleagues with examples of statistical brilliance or a useful forecast. Armstrong quotes Tukey's sage warning: "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise" (p. 334).

Communication is particularly important in the field of forecasting. The greatest strength of LRF is that it is understandable. Although it is often easy to explain math to mathematicians or statistics to statisticians, it is generally not easy to explain a forecasting methodology to clients. And yet no responsible client will act on a forecast without a good understanding of what it is based on. Explaining that you have minimized the root mean squared error simply will not do. Complexity may gain you respect, but it will not produce a usable forecast.

Armstrong tells of the Dr. Fox studies to illustrate this point:

> Dr. Fox was an actor who looked distinguished and sounded authoritative. Armed with a fictitious but impressive resume, he lectured on a subject about which he knew nothing. The lecture, "Mathematical Game Theory as Applied to Physical Education," was delivered on three occasions to a total of 55 psychologists, educators, psychiatrists, social workers, and administrators. The talk lasted an hour and was followed by 30 minutes of questions and answers. It consisted entirely of double talk, meaningless words, false logic, contradictory statements, irrelevant humor, and meaningless references to unrelated topics. According to a questionnaire administered at the end of the session, the audience found the lecture to be clear and stimulating. They had a high regard for Dr. Fox. None of the subjects realized that the lecture was pure nonsense. (pp. 45–46).

In reflecting on the importance of making things simple, Armstrong reveals what I believe is far more common than anyone might suspect: "I found that many of the fancy ways of saying things were not necessary. Occasionally I had said something in a complex way because I did not know what I was talking about." (p. xii). Through consistently clear