11. ______, *Generalized curves*, Duke Math. J. **6** (1940), 513–536.

12. ______, *Existence theorems for Bolza problems in the calculus of variations*, Duke Math. J. **7** (1940), 28–61.

13. L. W. Neustadt, *A general theory of minimum-fuel space trajectories*, J. SIAM Control Ser. A **3** (1965), 317–356.

14. ______, *An abstract variational theory with applications to a broad class of optimization problems*.I, II, SIAM J. Control Optim. **4** (1966), 503–527; ibid. **5** (1967), 90–137.

15. ______, *Optimization—A theory of necessary conditions*, Princeton Univ. Press, Princeton, N. J., 1976.

16. L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko, *The mathematical theory of optimal processes*, Fitzmatgiz, Moscow, 1961; English transl., Wiley, New York, 1962.

17. R. W. Rishel, *An extended Pontryagin principle for control systems whose control laws contain measures*, SIAM J. Control Optim. **3** (1965), 191–205.

18. E. Roxin, *The existence of optimal controls*, Michigan Math. J. **9** (1962), 109–119.

19. W. W. Schmaedecke, *Optimal control theory for nonlinear vector differential equations containing measures*, SIAM J. Control Optim. **3** (1965), 231–280.

20. L. Tonelli, *Sugli integrali del calcolo delle variazioni in forma ordinaria*, Ann. Scuola Norm. Sup. Pisa (2) **3** (1934), 401–405.

21. F. A. Valentine, *The problem of Lagrange with differential inequalities as added side conditions*, Contributions to the Calculus of Variations, 1933–1937, Univ. of Chicago Press, Chicago, 1937, pp. 407–448.

22. J. Warga, *Relaxed variational problems*, J. Math. Anal. Appl. **4** (1962), 111–128.

23. ______, *Minimizing variational curves restricted to a preassigned set*, Trans. Amer. Math. Soc. **112** (1964), 432–455.

24. ______, *Variational problems with unbounded controls*, J. SIAM Control Ser. A **3** (1965), 424–438.

25. ______, *Optimal control of differential and functional equations*, Academic Press, New York, 1972.

26. ______, *Necessary conditions without differentiablility assumptions in optimal control*, J. Differential Equations **18** (1975), 41–62.

27. T. Ważewski, *Sur la généralisation de la notion des solutions d'une équation au contingent*, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. **10** (1962), 11–15.

28. ______, *Sur les systèmes de commande non linéaires dont le contredomaine de commande n'est pas forcément convexe*, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. **10** (1962), 17–21.

29. L. C. Young, *Generalized curves and the existence of an attained absolute minimum in the calculus of variations*, C. R. Sci. Lettres Varsovie, C III **30** (1937), 212–234.

30. ______, *Lectures on the calculus of variations and optimal control theory*, Saunders, Philadelphia, 1969.

J. WARGA

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 9, Number 3, November 1983

0273-0979/83 $1.00 + $.25 per page

Statistical analysis of counting processes, by Martin Jacobsen, Lecture Notes in Statistics, Vol. 12, Springer-Verlag, New York, 1982, vii + 226 pp., $14.80. ISBN 0-3879-0769-6

As the implications of a mathematical structure become more deeply understood, the number of applied problems that may be solved by that structure increases rapidly, often in some surprising directions. In *The statistical analysis of counting processes*, Martin Jacobson has given us an excellent account of just such a phenomenon: the application of ideas and results from the general theory of stochastic processes, as developed primarily by the French school of probability, to the compelling and delicate applied statistical problem of the analysis of the lifetime data arising for the most part in clinical trials, epidemiological studies, and engineering reliability experiments.

Jacobsen's monograph is far from being the first published account of how the abstract theory of stochastic processes, and the consequent results on martingales and stochastic integrals, can be used in analyzing lifetime data, or more generally, data that accrues over time in the form of observed counting processes. In fact, the literature on this topic is now quite extensive. He has, though, provided the first thorough "unified and essentially self-contained exposition" in this area. The need for such an account is clear, and a short account of the development of statistical methods in this fascinating area will help explain why.

Evidently it is difficult to date the first systematic collection and analysis of lifetime data. Life tables summarizing mortality rates for populations have existed since at least the third century a.d., when a table attributed to Ulpian was constructed to assist in the distribution of annuities in Rome. Inference problems for life table models have been present in the mathematical literature for some time as well: In 1760 Daniel Bernoulli **[4]** used Halley's **[12]** life table for the city of Breslau (now Wroclaw, Poland) to illustrate a method for estimating the effect on mortality rates of the elimination of smallpox. Both Cardano **[6]** and Euler **[9]** also worked on what we might now call statistical problems associated with lifetables. Generally, though, sample sizes used in classical lifetables are large enough to prevent sampling variation, the key issue in any real inference problem, from presenting many serious difficulties.

In the 1950s, however, life data began to accumulate in comparative experiments in fields such as engineering reliability studies, epidemiological and medical follow-up studies, and laboratory pharmacological experiments. The most important measurement on each experimental unit in such situations was the time to a specified event, e.g., failure time of a machine component, the time to death for subjects undergoing an arteriosclerosis treatment regimen, or the time at which a treated tumor begins to shrink. Sample sizes in these experiments were necessarily small, and the inference questions surrounding these data sets had to be handled with care. In the statistical literature focused research began in earnest about this time into the "best" statistical methods for these problems.

The common structure of the data in such problems can be described as follows. There is associated with each experimental unit a lifetime random variable T, a censoring time variable, Y, to be explained below, and a vector $\vec{Z} = (Z_1, \ldots, Z_p)$ of (possibly time dependent) covariates. The censoring time is that time measured from an appropriate origin, after which observation of the experimental unit is no longer possible. In applications, censoring times arise as, for instance, the planned termination time of a study, the loss of a subject during follow-up because of death from another disease, or the planned sacrifice of a laboratory animal to examine possible side effects of treatment. The data available for the ith subject in a lifetime experiment are thus

$X_i = \min(T_i, Y_i)$ and $\vec{Z}_i$; the crucial inference problems, however, remain those associated with the distribution of T_i or the conditional distribution of T_i given $\vec{Z}_i$. Since the presence of censoring variables distinguishes statistical problems in this area from more traditional regression or analysis of covariance problems, data of this sort are also commonly referred to as censored data.

The first attempts to develop methods for censored data and to understand the operating characteristics for those methods were for the most part in parametric settings of homogeneous samples (i.e. no covariates). Researchers assumed that the distribution of T was specified up to an unknown parameter θ, and $F_\theta(t) = P_\theta(T \leqslant t)$ was usually estimated by likelihood methods. Two early important papers using this approach were those by Halperin [**13**] and Epstein and Sobel [**8**]. These parametric methods were quickly extended to situations in which data arose in comparative experiments involving two or more samples and, eventually, to some situations in which the hazard function,

$$\lambda_\theta(t) = \frac{\frac{d}{dt}F_\theta(t)}{\{1 - F_\theta(t)\}},$$

depended on a covariate vector. Even in this parametric setting, however, the censoring variable made the subject unexpectedly messy. It was not always easy to defend proposed likelihood functions rigorously, nor was it clear that the asymptotic theory of likelihood based procedures was applicable. For lifetime distributions other than the exponential, even formulas based on heuristics were invariably cumbersome, and numerical methods were often needed to compute estimators.

The parametric methods were usually applied to data arising in engineering settings where, for instance, extreme value theory might justify a particular probability model. Biostatisticians, however, were skeptical about using such methods for, say, clinical trial data. It was often impossible to justify particular parametric models, and there was accumulating evidence that most of the parametric procedures lacked robustness. Statisticians began to examine nonparametric methods for censored data; some early seminal ideas were contained in the papers by Kaplan and Meier [**15**] and Mantel [**17**]. Kaplan and Meier proposed the first nonparametric estimator of $P(T \leqslant t)$ based on censored data, while Mantel proposed a nonparametric method for assessing the significance of observed differences in two samples of censored data. Cox's 1972 paper [**7**] was the first successful attempt to incorporate regressor variables into nonparametric methods for censored data.

The nonparametric methods quickly became popular in biometric settings and the model proposed by Cox grew into widespread use in analyzing data from cancer studies. Cox's model incorporated covariates into the distribution for T by assuming that the conditional hazard rate for T, given the covariates, was

$$\lambda(t \mid \vec{Z}) = \lambda_0(t) \exp\left(\sum_{j=1}^{p} Z_j \beta_j \right),$$

where λ_0 was an unspecified baseline hazard function corresponding to $Z_j = 0$, $1 \leqslant j \leqslant p$, and $\vec{\beta} = (\beta_1, \ldots, \beta_p)$ was a vector of unknown regression coefficients. Inference about $\vec{\beta}$ was based on a likelihood-type function, which Cox called a partial likelihood, and asymptotic distribution theory for statistics derived from the partial likelihood. Some of the distribution results were obtained through long, difficult proofs (e.g. Tsiatis [**19**]), some were "established" by analogy with classical likelihood theory, and some were simply hoped for. As in the parametric setting, usable results established with current standards of mathematical rigor were difficult to come by.

In 1975 Odd Aalen wrote a Ph.D. dissertation under Lucien LeCam at Berkeley which suggested an entirely new way of modeling lifetime data with what Aalen called multiplicative intensity models. The data for a homogeneous subsample (possibly of size 1) was viewed as a counting process $\{N(t), t \geqslant 0\}$ on the real line; this counting processs was trivially a submartingale and, under mild regularity conditions, its Doob-Meyer decomposition was of the form

$$N(t) = M(t) + \int_0^t Y(s)\lambda(s)\,ds,$$

where $\{M(t), t \geqslant 0\}$ was a locally square integrable martingale with respect to an appropriately defined filtration, $\{Y(s), s \geqslant 0\}$ was a predictable stochastic process easily determined from the problem at hand, and $\lambda(s)$ was the hazard function and object of inference for the units in the subsample. It turned out that almost without exception, statistical procedures being used for lifetime data were based on processes of the form $\int h\,dM$, or linear combinations of such processes. Since at most a minor modification was needed to insure that h was a predictable process, the theory of stochastic integrals implied that such statistics were themselves martingales. Fundamental results on central limit theorems and weak convergence to Wiener processes for sequences of martingales now reduced difficult proofs of asymptotic distribution theory to a much more standard exercise in checking Lindeberg type conditions. In his dissertation and a subsequent paper [**1**], Aalen provided the first weak convergence theorems specifically tailored to sequences of stochastic integrals with respect to counting process martingales.

When the significance of Aalen's ideas was understood, the direction of mathematical research into censored data methods changed quickly. M. Rebolledo [**18**] simplified considerably the conditions in Aalen's weak convergence theorem. Richard Gill [**11**] used the approach in the first comprehensive rigorous study of the operating characteristics of two sample censored data linear rank statistics, both under null and alternative hypotheses. Gill also showed how martingale inequalities and invariance principles could provide direct and simple proofs of known and new asymptotic properties of Kaplan's and Meier's estimator; in conjunction with Per Andersen [**3**] and later with Per Andersen, Ornulf Borgan and Niels Keiding [**2**], he applied these ideas to the Cox model and to multisample nonparametric statistics. Several authors found the stochastic integral formulation both powerful and convenient for studying entirely new approaches to censored data, both in the form of new statistics and new group sequential experimental designs (e.g. [**10**] and [**14**]).

The martingale based proofs all have some common characteristics: they are structurally simple, invariably short, and are based on results that applied statisticians have never seen. As universal as the Doob-Meyer decomposition, stochastic integrals with respect to L^2 martingales, and progressively measurable and predictable processes with respect to right continuous, complete filtrations all seem to probabilists, these topics have not filtered into the classical statistical literature at all. Both Aalen and Gill provided excellent summaries of these notions, but for detailed explanations readers were referred to a large body of literature, most of it in French and essentially inaccessible to those who needed it most in this context. Research statisticians cannot thoroughly understand a tool they have only seen summarized. Now Martin Jacobsen has taken that tool apart in this special setting and shown in detail how it works.

The outline for Jacobsen's monograph is tightly organized and well thought out. Chapter 1 provides basic definitions associated with probability measures on $(0, \infty]$ and a careful construction of an appropriate probability space for one-dimensional counting processes. Appropriate filtrations are established constructively, and their necessary properties are established directly, rather than by reference to other more general results from measure theory. The simple structure of the path space here gives an algebraic flavor to many of the proofs, some of which are quite elegantly constructed. The chapter closes with a self-contained proof of the existence-half of the Doob-Meyer decomposition in this setting, a rigorous derivation of a likelihood function, and some well-chosen exercises.

With some exceptions, Chapter 2 repeats the program of Chapter 1 for multidimensional counting processes. Discrete time counting processes and processes defined on product probability spaces appear in this chapter, but not in Chapter 1.

The third chapter is the shortest, but much of the intended audience will find it the most useful. It is a self-contained treatment, for this special setting, of stochastic integrals with respect to martingales. The roles of predictable processses, increasing processes, the Doob-Meyer decomposition and quadratic variation processes are explained with careful definitions and detailed proofs. This chapter also closes with some useful exercises, although Exercise 3.E.2 is a result by Boel, Varaiya and Wong [5] and a reference to the literature should have been provided.

Chapters 4 and 5 show how this material is applied to counting process data. Aalen's multiplicative intensity model is developed in Chapter 4, and the model is illustrated for homogeneous samples of censored and uncensored lifetime data, Markov chains with time dependent transition intensities, and the Cox regression model. Chapter 5 explores in detail the application of invariance principles for sequences of martingales to counting process statistics. Specific results are illustrated for cumulative intensity estimators, the Kaplan-Meier estimator, and some two sample statistics.

The monograph closes with a short but lucid appendix on the principle of repeated conditioning and weak convergence of probability measures.

Some general comments about the book are perhaps in order at this point. It has obviously been prepared with great care. There are exceptionally few errors of substance. Example 5.2.6 does contain a subtle error in the use of the Lebesgue convergence theorem. Lapses like this are rare, however, and typographical errors are almost equally rare.

There are some drawbacks, which may seem more serious to some readers than to others. The title is a bit of a misnomer. Don't expect to see any statistical analyzing going on here. The models and their asymptotic properties are examined in detail, but not a single datum is hiding anywhere between the covers. The results in Chapters 4 and 5 stop a good deal short of what is currently in the literature, so mathematicians interested in research in this area will have to look elsewhere for the problems statisticians consider important. §5.4 on the comparison of two intensities appears almost as an afterthought. Gill's monograph and the work of others has shown, though, that this may be the most natural way to think about the two sample censored data problem. Statisticians have told me, and I tend to agree, that Martin Jacobsen has underestimated the prerequisites necessary for reading the text. Jacobsen assumes "some knowledge of probability... especially conditional probabilities, weak convergence and basic martingale theory." Some statisticians may have been introduced to basic martingale theory in a treatment like that in Karlin and Taylor **[16]**, but they will still struggle with the measure theoretic and path algebraic proofs Jacobsen has constructed in the Meyer-Dellacherie style.

We should not ask too much of this text, though. Martin Jacobsen chose a well-defined target, and hit it squarely.

References

1. O. O. Aalen, *Weak convergence of stochastic integrals related to counting processes*, Z. Wahrsch. Verw. Gebiete **38** (1977), 261–277.
2. P. K. Andersen, O. Borgan, R. D. Gill and N. Keiding, *Linear nonparametric tests for comparison of counting processes, with applications to censored survival data*, Internat. Statist. Rev. **50** (1982), 219–258.
3. P. K. Andersen and R. D. Gill, *Cox's regression model for counting processes: a large sample study*, Ann. Statist. **10** (1982), 1100–1120.
4. D. Bernoulli, *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et les avantages de l'inoculation pour la prévenir*, Histoire de l'Académie Royale des Sciences, 1960, pp. 1–45.
5. R. Boel, P. Varaiya and E. Wong, *Martingales on jump processes.* I: *Representation results*, SIAM J. Control Optim. **13** (1975), 999–1021.
6. G. Cardano, *Opus Novuum de Proportionibvs Nvmerorvm*, 1570, Basel.
7. D. R. Cox, *Regression models and lifetables*, J. Roy. Statist. Soc. Ser. B **34** (1972), 187–220. (with discussion)
8. B. Epstein and M. Sobel, *Life-testing*, J. Amer. Statist. Assoc. **48** (1953), 486–502.
9. L. Euler, *Recherches générales sur la mortalité et la multiplication du genre humaine*, Histoire de l'Académie Royale des Sciences et Belles Lettres **16** (1760), 144–164; English transl. by N. and B. Keyfitz in Theoret. Population Biol. **1** (1970), 307–314.
10. T. R. Fleming, J. R. O'Fallon, P. C. O'Brien and D. P. Harrington, *Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right censored data*, Biometrics **36** (1980), 607–625.
11. R. D. Gill, *Censoring and stochastic integrals*, Math. Centre Tracts 124, Math. Centrum, Amsterdam, 1980.

12. E. Halley, *An estimate of the degrees of mortality of mankind...*, Philos. Trans. Roy. Soc. London Ser. A **17** (1693), 596–610, 653–656.

13. M. Halperin, *Maximum likelihood estimation in truncated samples*, Ann. Math. Statist. **23** (1952), 226–238.

14. D. P. Harrington, T. R. Fleming and S. J. Green, *Procedures for serial testing in censored survival data*, Survival Analysis (J. Crowley and R. A. Johnson, eds.), IMS Lecture Notes-Monograph Ser., Vol. 2, 1982, pp. 269–286.

15. E. L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **53** (1958), 457–481.

16. S. Karlin and H. M. Taylor, *A first course in stochastic processes*, 2nd ed., Academic Press, New York, 1975.

17. N. Mantel, *Evaluations of survival data and two new rank order statistics arising in its consideration*, Cancer Chemother. Rep. **50** (1966), 163–170.

18. R. Rebolledo, *Sur les applications de la theorie des martingales à l'étude statistique d'une famille de processus ponctuels*, Journées de Statistique des Processus Stochastiques (Proc. Conf., Grenoble, 1977), Lecture Notes in Math., Vol. 636, Springer-Verlag, Berlin, 1978.

19. A. A. Tsiatis, *A large sample study of Cox's regression model*, Ann. Statist. **9** (1981), 93–108.

DAVID P. HARRINGTON

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 9, Number 3, November 1983

0273-0979/83 $1.00 + $.25 per page

Nonlinear analysis on manifolds. Monge-Ampère equations, by Thierry Aubin, Grundlehren der mathematischen Wissenschaften, vol. 252, Springer-Verlag, Berlin and New York, 1982, xii + 302 pp., $29.50. ISBN 0-3879-0704-1

No one is very surprised if an area of mathematics can solve its own problems. The surprise is when one area of mathematics can help solve those of another. In recent years it has been our good fortune to see problems from places like algebraic geometry and differential topology solved using nonlinear partial differential equations. Of course, an area should not be judged solely on how it helps other branches of mathematics—but the publicity sure helps convince the skeptical of its current relevance. With this in mind, it is important to note that these developments have taken place as part of a vigorous general advance in our understanding of nonlinear partial differential equations.

Linear problems dominated analysis in the first half of this century, which saw the emergence of the now classical linear functional analysis of Hilbert and Banach spaces. A major source of motivation for this work came from an attempt to understand the wave, Laplace, heat, and Schrödinger partial differential equations of mathematical physics. The development of Fourier analysis was one of the most fruitful discoveries, serving simultaneously as both a tool and as a subject in its own right. Now many of the linear differential equations are linear for a very simple reason: one takes a nonlinear equation and bluntly linearizes it (see any derivation of the wave equation