ON THE APPLICATION OF THE THEORY OF COUNTING PROCESSES IN THE STATISTICAL ANALYSIS OF CENSORED SURVIVAL DATA

Per Kragh Andersen

Statistical Research Unit, Danish Medical and Social Science Research Councils

0. SUMMARY

It was demonstrated by Aalen (1978) how the theory of multivariate counting processes gives a general framework in which both censored survival data and inhomogeneous Markov processes may be analyzed, and how by means of martingale central limit theory the asymptotic distribution for all the classical linear nonparametric two-sample tests and their generalizations to censored data may be derived. In this paper these results will be surveyed and further developed to both the case of the comparison of $k(\geq 2)$ distributions (see Andersen, Borgan, Gill & Keiding, 1981) and to the case of regression models for survival data (Cox, 1972; Andersen & Gill, 1981).

1. Introduction

In survival analysis one is interested in the distribution of the time T to some event, usually denoted death, and very often the object of a study is to relate this distribution to individual characteristics which in the simplest form are group indicators. Frequently statistical models for survival data are specified via the intensity or hazard function $\alpha(t)$ for T. The hazard function denotes the infinitesimal probability of dying at time t given survival up to time t, and hence $\alpha(t)$ may be interpreted as the rate at which the event in question occurs at time t.

A survival model is the simplest example of a Markov process model in that there are only two states, "alive" and "dead", with an intensity equal to $\alpha(t)$ of a transition from the former state to the latter. In more general Markov processes the basic parameters are the forces of transition between the states.

From these facts it seems obvious that the natural framework in which to analyze such phenomena is one where various types of events may happen during time and where the rate at which the events occur can be specified. One such framework can be introduced by the notation of a multivariate counting process.

A univariate point process is a countable random set of points on the real line, and a multivariate point process is a collection of, say k univariate processes. If $N_i(t)$ is defined as the number of points in $[0,1]$ from the i^{th} process, then N_i can be thought of as counting the events of type i before t, and N_i is called the counting process corresponding to the i^{th} point process. Let $(F_t)_{t\geq 0}$ be an increasing family of σ-algebras. One possibility would be to let F_t be the σ-algebra generated by the multivariate counting process $((N_1(s),\ldots,N_k(s)),\ s\in[0,t])$, but a larger family can also be considered. Note that the fact that (F_t) is increasing reflects that time moves in a certain direction. We shall assume that the limit

$$\text{(1)} \qquad \Lambda_i(t+) = \lim_{h\downarrow 0} \frac{1}{h} E(N_i(t+h) - N_i(t) | F_t), \quad t \geq 0, \ i=1,\ldots,k ,$$

exists and we shall call the random process $\Lambda_i(t)$ the intensity process of N_i; this concept generalizes the notion of a hazard rate.

The idea of using counting process theory in the analysis of survival data and other Markov processes is due to Aalen (1975, 1978). There the so-called multiplicative intensity model was introduced, this statistical method being specified by assuming that the intensity process has the form

$$\text{(2)} \qquad \Lambda(t) = \alpha(t)\ Y(t), \quad t \geq 0 .$$

Here $\alpha(t)$ is an unknown function and $Y(t)$ is an observable stochastic process adapted to F_{t-}. In a survival study $\alpha(t)$ will be the hazard function and $Y(t)$ the number of individuals at risk just before time t, while in a more general Markov chain $\alpha(t)$ is a force of transition and $Y(t)$ is the number at risk just before time t for the transition in question.

The intensity property (1) of $\Lambda_i(t)$ is (up to regularity conditions) equivalent to the fact that the processes

(3) $$M_i(t) = N_i(t) - \int_0^t \Lambda_i(u)du \ , \quad t \geq 0 \ , \quad i=1,\ldots,k \quad ,$$

are martingales, i.e., $E(M_i(t)|F_u) = M_i(u)$, $t > u$. This observation is the basis for making a unified approach to the proofs of asymptotic properties of many estimators and test statistics known from the survival data literature since these statistics can often be expressed as stochastic integrals with respect to martingales, and since furthermore central limit theorems and other properties of martingales are very well studied.

The typical feature of survival data is that one is not always able to observe all the lifetimes; rather, for some individuals it is only known that the true lifetime T_i exceeds some quantity t_i. We denote the corresponding observation a (right) censored observation. Another advantage of using the counting process description of survival data is that it accommodates fairly general censoring patterns (see e.g., Gill, 1980, Section 3.1, or Andersen, Borgan, Gill & Keiding, 1981, Sections 2D and 3D).

The rest of this paper contains examples from survival analysis of the use of the theory of counting processes, martingales and stochastic integrals. For a more detailed survey of the probabilistic background the reader is referred to Aalen (1978), Gill (1980) and Andersen, Borgan, Gill & Keiding (1981) and the references therein.

2. The One-Sample Situation

The simplest situation with censored data is the one-sample set-up, where out of n independent identically distributed lifetimes T_i, some are observed, but for the rest it is only known that they are larger than some times t_i. Let $X_i = \min(T_i, t_i)$ and $\delta_i = I(T_i \le t_i)$. From these data we want to estimate the distribution F of the T_i's. The product-limit estimator (Kaplan & Meier, 1958) of the survivorship functions $S = 1-F$ is given by

(4) $$\hat{S}(t) = \prod_{i:X_i \le t} \left(1 - \frac{\delta_i}{Y(X_i)}\right), t \ge 0 \quad ,$$

where $Y(t) = \#\{i:X_i \ge t\}$. Let $N(t) = \#\{i:X_i \le t,\ \delta_i = 1\}$. Then N is a counting process with intensity process $\alpha(t)\ Y(t)$ (cf. (2)), where $\alpha(t) = -\frac{d}{dt} \log S(t)$ is the hazard function (see Aalen (1978, Example 1)). It follows that $\hat{S}(t) = \prod_{u \le t} (1 - \frac{dN(u)}{Y(u)})$ and from this fact and from (3) it was noted by Aalen and Johansen (1978) (see also Gill, 1980, Lemma 3.2.1) that $\hat{S}(t)/\tilde{S}(t)$ is a martingale, where $\tilde{S}(t)$ converges to S(t) at an exponential rate; hence the asymptotic properties when $n \to \infty$ of $\hat{S}$ also proved by Breslow and Crowley (1974) can be derived very simply (see Gill, 1980, 1981).

The Nelson estimator (Nelson, 1969, 1972) for the cumulative hazard function $\beta(t) = \int_0^t \alpha(u)du$ is given by

(5) $$\hat{\beta}(t) = \int_0^t \frac{dN(u)}{Y(u)} , \ t \ge 0 \quad ,$$

(Aalen, 1978, Section 6.1), and using (3) we see that $\hat{\beta}(t) - \beta(t)$ is a martingale (aside from a term that converges to zero at an exponential rate), being a stochastic integral of the process Y^{-1} with respect to a martingale; hence the asymptotic properties (see also Breslow & Crowley, 1974) can be found directly. Ramlau-Hansen (1981) used counting process and martingale techniques to study kernel function estimation

$$\hat{\alpha}(t) = (1/b) \int_0^\infty K((t-s)/b)\ d\hat{\beta}(s) \quad ,$$

of the hazard function $\alpha(t)$ itself rather than the integrated hazard $\beta(t)$. Here the kernel function K is non-negative with integral 1 and the window b is a positive parameter. Thus simple proofs of consistency and asymptotic normality of $\hat{\alpha}(t)$ were obtained.

We shall conclude this section by noting that the tests studied by Breslow (1975), Hyde (1977), Hollander and Proschan (1979) and Harrington and Fleming (1981) for comparing the distribution of the T_i's with a known distribution F_0 (with hazard function α_0, say) can be shown to have essentially the form

$$(6) \qquad Z(t) = \int_0^t L(u)\ (d\hat{\beta}(u) - \alpha_0(u)du)\ ,\ t \geq 0 \quad ,$$

for various choices of the process L(u) (see Andersen, Borgan, Gill & Keiding, 1981, Section 4). It follows that under $H_0: F = F_0$, Z is a martingale, and from this fact the asymptotic distribution of the test statistics can be derived.

3. The k-Sample ($k \geq 2$) Situation

In this situation the problem is one of comparing the survival of k distinct groups. In each group i we have the Nelson estimator $\hat{\beta}_i(t)$ given by (5) for the cumulative hazard function $\beta_i(t) = \int_0^t \alpha_i(u)du$. Under the null hypothesis H_0: $\alpha_1 = \ldots = \alpha_k$ $(=\alpha, \text{say})$ we can estimate $\beta(t) = \int_0^t \alpha(u)du$ by

$$(7) \qquad \hat{\beta}(t) = \int_0^t \frac{d\bar{N}(u)}{\bar{Y}(u)}\ ,\ t \geq 0 \quad ,$$

where $\bar{N} = N_1 + \cdots + N_k$, $\bar{Y} = Y_1 + \cdots + Y_k$, $N_i(t)$ is the stochastic process counting the number of failures in group i in [0,1], and $Y_i(t)$ is the number at risk in group i at time t-. A general test statistic for H_0 based on the processes

$$(8) \qquad Z_i(t) = \int_0^t L(u)\ Y_i(u)\ (d\hat{\beta}_i(u) - d\hat{\beta}(u)),\ t \geq 0\ ,\ i=1,\ldots,k \quad ,$$

comparing the individual estimates $\hat{\beta}_i(t)$ with the common value $\hat{\beta}(t)$ was introduced by Andersen, Borgan, Gill & Keiding (1981, Section 3A). These authors proved that special choices of the stochastic process $L(t)$ correspond to various previously suggested test statistics from the literature. Thus, $L(t)=1$ yields the logrank test (Peto & Peto; 1972), $L(t)=\bar{Y}(t)$ corresponds to the generalized Kruskal-Wallis test of Breslow (1970); $L(t)=(\bar{Y}(t))^{\rho}$ for ρ in [0,1] corresponds to the family of statistics suggested by Tarone & Ware (1977); and $L(t)=[\hat{S}(t-)]^{\rho}$ (cf. (4)) gives the class of tests considered by Harrington & Fleming (1981) generalizing the Kruskal-Wallis type test of Prentice (1978), which is obtained for $\rho=1$. The counting process formulation reveals that under H_0 we may write

$$(9) \qquad Z_i(t) = \int_0^t L(u) \left(dM_i(u) - \frac{Y_i(u)}{\bar{Y}(u)} d\bar{M}(u)\right), \quad t \geq 0, \quad i=1,\ldots,k,$$

where $M_i(t) = N_i(t) - \int_0^t \alpha(u)\, Y_i(u)du$, and $\bar{M} = M_1+\cdots+M_k$; hence Z_i is a martingale. This gives a general way of finding the asymptotic distribution of the test statistics. (See Crowley & Thomas (1975) for a derivation of the asymptotic distribution of the logrank test using a different approach).

In the case k=2, we can equivalently test for H_0 using the process

$$(10) \qquad Z(t) = \int_0^t K(u) \left(d\hat{\beta}_2(u) - d\hat{\beta}_1(u)\right), \quad t \geq 0 \quad ;$$

see Aalen (1978, Section 7). As special cases of (10) we get the logrank test ($K = Y_1Y_2/(Y_1+Y_2)$), the Wilcoxon test of Gehan (1965) ($K=Y_1Y_2$), and the test of Efron (1967) ($K=\hat{S}_1\hat{S}_2$). These tests were studied carefully by Gill (1980), who verified the conditions for normality for special censoring schemes and gave a discussion of efficiency properties of the tests.

The problem of estimating hazard ratios using counting process techniques in the two-sample model was discussed by Andersen (1981) following up the results of Crowley (1975). See also the paper by Crowley, Liu & Voelkel (this volume).

The use of the tests (8) and (10) in more general Markov processes was discussed by Aalen (1978), Aalen, Borgan, Keiding & Thormann (1980) and

Andersen & Rasmussen (1982). Examples of the applicability in analyses of Markov processes is found in Borgan (1980).

4. The Cox Regression Model

The semiparametric regression model of Cox (1972) specifies the hazard function of an individual i with (possibly time-dependent) covariates $z_i(t)$ to have the form

$$\alpha_i(t) = \lambda_0(t)\, e^{\beta_0' z_i(t)} , \quad t \geq 0 \quad . \tag{11}$$

Here β_0 is a vector of unknown regression coefficients and λ_0 is an unknown and unspecified underlying hazard function. The problem of estimating β_0 and λ_0 was discussed by Cox (1972,1975), Breslow (1972,1974), Kalbfleisch & Prentice (1973) and Tsiatis (1981a), and reviewed by Kalbfleisch & Prentice (1980).

The counting process formulation of (11) (cf. Andersen & Gill, 1981) specifies the intensity process Λ_i for the counting process N_i corresponding to the i^{th} individual to have the form

$$\Lambda_i(t) = \lambda_0(t)\, e^{\beta_0' z_i(t)}\, Y_i(t) , \quad t \geq 0 \quad , \tag{12}$$

where $Y_i(t) = 1$ if i is under observation at time t- and 0 otherwise. It was proven by Johansen (1981) that in an extended model, replacing the absolutely continuous measure $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ by an arbitrary measure Λ on $[0,\infty)$ and allowing jumps with a Poisson-distributed size, the joint likelihood $L(\beta,\Lambda)$ for β_0 and $\Lambda_0(t)$ based on independent processes $N_1,\ldots,N_n$ is maximized for fixed β by

$$\hat{\Lambda}_0(t) = \int_0^t \frac{d\bar{N}(u)}{\sum_{j=1}^{n} Y_j(u)\, e^{\beta' z_j(u)}} , \quad t \geq 0 \quad . \tag{13}$$

The estimate (13) was also considered by Tsiatis (1981a), and by interpolating (13) between failure times we get the estimate of Breslow (1972, 1974). Inserting (13) in $L(\beta,\Lambda)$ yields a partially maximized likelihood $L_c(\beta) = \max_{\Lambda} L(\beta,\Lambda)$ proportional to the well known Cox's likelihood (cf. Cox 1972, 1975). Hence the Cox likelihood is an also reasonable basis for the estimation of β_0 in the counting process model (12). It was noted by Andersen & Gill (1981) that evaluated at the true value β_0 the score statistic $U(\beta) = \frac{d}{d\beta} \log L_c(\beta)$ has the form

$$(14) \qquad U(\beta_0) = \sum_{i=1}^{n} \int_0^\infty z_i(u)\, dM_i(u) - \int_0^\infty \frac{\sum_{i=1}^{n} Y_i(u)\, z_i(u)\, e^{\beta_0' z_i(u)}}{\sum_{i=1}^{n} Y_i(u)\, e^{\beta_0' z_i(u)}}\, d\overline{M}(u) \quad ,$$

where M_i is the martingale $N_i(t) - \int_0^t \Lambda_i(u)du$ and $\overline{M} = M_1 + \cdots + M_n$. Hence, the process $U(\beta_0,t)$, obtained by replacing ∞ by t in (14), is a martingale, and this fact gives a simple way of proving asymptotic normality of the score statistic (see also Tsiatis, 1981b and Sen, 1981). In the usual way this result extends to a proof of asymptotic normality of the solution $\hat{\beta}$ to the likelihood equation $U(\beta) = 0$. (See Andersen & Gill, 1981, for details). Næs (1982) obtained the same results under stronger conditions using discrete time martingale results.

The weak convergence of $\hat{\Lambda}_0(\cdot) - \Lambda_0(\cdot)$ to a Gaussian process on a compact interval $[0,\tau]$ (Tsiatis, 1981a) can be obtained by first rewriting this difference as

$$(15) \qquad \begin{aligned} \hat{\Lambda}_0(t) - \Lambda_0(t) = & \int_0^t \left(\frac{1}{\sum_{j=1}^{n} Y_j(u)\, e^{\hat{\beta}' z_j(u)}} - \frac{1}{\sum_{j=1}^{n} Y_j(u)\, e^{\beta_0' z_j(u)}} \right) d\overline{N}(u) \\ & + \left(\int_0^t \frac{d\overline{N}(u)}{\sum_{j=1}^{n} Y_j(u)\, e^{\beta_0' z_j(u)}} - \Lambda_0(t) \right) \quad . \end{aligned}$$

The second term in (15) is (asymptotically equivalent to) a martingale, and by a Taylor expansion of the first term around β_0 this fact can be combined with the asymptotic normality of $\hat{\beta}$ to prove the weak convergence of $\hat{\Lambda}_0(\cdot) - \Lambda_0(\cdot)$.

An example of using the model (12) in a Markov process situation is found in Andersen & Rasmussen (1982) (see also Andersen & Gill, 1981).

Lustbader (1980) and Oakes (1981) showed how several well-known two-sample test statistics could be obtained as score tests from (11) by appropriate choices of time-dependent covariates. In fact, every k-sample test statistic of the form (8) can be obtained from (12) as a score test by letting $z_{ij}(t) = L(t)$ if individual i belongs to group j at time t and 0 otherwise.

ACKNOWLEDGEMENT

The author's participation in the IMS special topics meeting on survival analysis in Columbus, Ohio, 26-28 October 1981 was supported by the Danish Medical Research Council (Grant No. 12-2590) and the Danish Social Science Research Council (Grant No. 14-1997).

REFERENCES

Aalen, O.O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation. Department of Statistics, University of California, Berkeley.

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. Annals of Statistics 6, 701-726.

Aalen, O.O., Borgan, Ø., Keiding, N. and Thormann, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring. Scandinavian Journal of Statistics 7, 161-171.

Aalen, O.O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics 5, 141-150.

Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1981). Linear non-parametric tests for comparison of counting processes, with applications to censored survival data. Research Report 81/4, Statistical Research Unit, Danish Medical and Social Science Research Councils. To appear in the International Statistical Review, 1982.

Andersen, P.K. and Gill, R.D. (1981). Cox's regression model for counting processes: a large sample study. Research Report 81/6, Statistical Research Unit, Danish Medical and Social Science Research Councils. To appear in the Annals of Statistics, 1982.

Andersen, P.K. and Rasmussen, N.K. (1982). Admissions to psychiatric hospitals among women giving birth and women having induced abortion. A statistical analysis of a counting process model. Research Report from Statistical Research Unit, Danish Medical and Social Science Research Councils, Copenhagen.

Andersen, P.K. (1981). Comparing survival distributions via hazard ratio estimates. Research Report 81/7, Statistical Research Unit, Danish Medical and Social Science Research Councils. To appear in the Scandinavian Journal of Statistics, 1983.

Borgan, Ø. (1980). Applications of non-homogeneous Markov chains to medical studies. Nonparametric analysis for prospective and retrospective data. In Explorative Datenanalyse. Frühjahrstagung, München, 1980. Proceedings (eds. N. Victor, W. Lehmacher and W. van Eimeren), pp. 102-115. Springer's series Medizinische Informatik und Statistik, Band 26.

Breslow, N.E. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 57, 579-594.

Breslow, N. (1972). Contribution to the discussion of the paper by D.R. Cox. Journal of the Royal Statistical Society B 34, 187-220.

Breslow, N. (1974). Covariance analysis of censored survival data. Biometrics **30**, 89-99.

Breslow, N.E. (1975). Analysis of survival data under the proportional hazards model. International Statistical Review **43**, 45-58.

Breslow, N.E. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Annals of Statistics **2**, 437-453.

Cox, D.R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society B **34**, 187-220.

Cox, D.R. (1975). Partial likelihood. Biometrika **62**, 269-276.

Crowley, J. (1975). Estimation of relative risk in survival studies. Technical Report No. 423. Department of Statistics, University of Wisconsin-Madison.

Crowley, J. and Thomas, D.R. (1975). Large sample theory for the logrank test. Technical Report No. 415. Department of Statistics, University of Wisconsin-Madison.

Efron, B. (1967). The two sample problem with censored data. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol, IV, University of California Press, Berkeley, California, 831-853.

Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika **52**, 203-223.

Gill, R.D. (1980). Censoring and stochastic integrals. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Gill, R.D. (1981). Large sample behaviour of the product-limit estimator on the whole line. Preprint 74/81, Mathematical Centre, Amsterdam. To appear in the Annals of Statistics, 1983.

Harrington, D.P. and Fleming, T.R. (1981). A class of rank test procedures for censored survival data. Technical Report Series No. 12, Section of Medical Research Statistics, Mayo Clinic. To appear in Biometrika, 1983.

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. Biometrics **35**, 393-401.

Hyde, J. (1977). Testing survival under right censoring and left truncation. Biometrika 64, 225-230.

Johansen, S. (1981). An extension of Cox's regression model. Preprint No. 11, Institute of Mathematical Statistics, University of Copenhagen. To appear in the International Statistical Review, 1983.

Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika 60, 267-278.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457-481.

Lustbader, E.D. (1980). Time-dependent covariates in survival analysis. Biometrika 67, 697-698.

Nelson, W. (1969). Hazard plotting for incomplete failure data. Journal of Quality Technology 1, 27-52.

Nelson, W. (1972). Theory and application of hazard plotting for censored failure data. Technometrics 14, 945-966.

Næs, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. Scandinavian Journal of Statistics 9, 107-116.

Oakes, D. (1981). Survival times: aspects of partial likelihood (with discussion). International Statistical Review 49, 235-264.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). Journal of the Royal Statistical Society A 135, 185-206.

Prentice, R.L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-179.

Ramlau-Hansen, H. (1981). Smoothing counting process intensities by means of kernel functions. Working Paper No. 43, Laboratory of Actuarial Mathematics, University of Copenhagen. To appear in the Annals of Statistics, 1983.

Sen, P.K. (1981). The Cox regression model, invariance principles for some induced quantive processes and some repeated significance tests. Annals of Statistics **9**, 109-121.

Tarone, R.E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. Biometrika **64**, 156-160.

Tsiatis, A.A. (1981a). A large sample study of Cox's regression model. Annals of Statistics **9**, 93-108.

Tsiatis, A.A. (1981b). The asymptotic distribution of the efficient scores test for the proportional hazards model caclulated over time. Biometrika **68**, 311-315.