Probability and Statistics

Editors
S K Basu
B K Sinha

Narosa

Statistical Inference from Discrete Life History Data

Debasis Sengupta and S. Rao Jammalamadaka

1. Introduction

The counting process approach to the analysis of life history data has been explored extensively in the last one and a half decades. This approach demands accurate observation of time, as pointed out by Arjas (1985). Inaccuracy of measurements often lead to tied data, which must be treated in an ad hoc manner. Therefore, in spite of the natural appeal of time as a continuous parameter, Arjas advocated a discrete time formulation of the counting process approach. Arjas and Haara (1987) proposed a discrete time logistic regression model and analysed it using a discrete time 'counting' process approach. The asymptotic normality of the regression estimators was established via a martingale convergence theorem, as the observation time goes to infinity. The authors used this approach also for a generalised Cox regression model; see Arjas and Haara (1988).

A discrete time formulation was also proposed by Hjort (1985). He defined the deterministic intensity in discrete time in the special case of survival data and presented its maximum likelihood estimator (mle) based on the likelihood function for a partially censored data set. In order to derive the asymptotic properties of his estimator, he assumed the discrete measurements to be samples of an originally continuous phenomenon and let the sampling interval go to zero at a desired rate, as the number of subjects in the study go to infinity.

In the earlier literature on Survival Analysis (see Kalbfleisch and Prentice, 1980 and Lawless, 1982) parametric and nonparametric estimation of survival function from censored survival data has been discussed. Testing and regression problems have also been considered. Usually the number of groups is finite in these formulations.

In addition to the arguments of measurement inaccuracy and discretised or grouped data, we would like to point out that discrete data may arise *naturally* in machines such as computers and computer-controlled devices which operate with a digital clock. We believe that the discrete time stochastic process approach deserves more attention than what it has received so far.

In this paper, inference on a discrete time counting process is viewed from a different angle. No assumption is made about the origin of the discrete nature of the data. The asymptotic arguments apply *as the number of individuals, rather than the length of the observation period, goes to infinity.* Even if the data originates from sampling a continuous time process, the time between the samples is *not* assumed to decrease at a specific rate.

The model is formulated in Section 2. The issues related to estimation and testing are discussed in Sections 3 and 4, respectively.

2. The Multiplicative Intensity Model in Discrete Time

Let $(\Omega, \mathcal{F}, P)$ be a probability space and for $i = 1, 2, \ldots, n$ and $h = 1, 2, \ldots, H$, $\{\Delta L_{ih}(k)\}_{k\geq 1}$ be a family of stochastic processes having discrete parameter k, state space $\{0,1\}$ and defined on $(\Omega, \mathcal{F})$. Let $\mathcal{F}_{n,0} = \{\phi, \Omega\}$ and $\mathcal{F}_{n,k}$ be the smallest sub-σ-algebra of $\mathcal{F}$ with respect to which $\{\Delta L_{ih}(l)\}_{1\leq l\leq k}$ are measurable for each i and h. We further define $\Delta N_{nh}(k) = \sum_{i=1}^{n} \Delta L_{ih}(k)$ and $N_{nh}(k) = \sum_{l=1}^{k} \Delta N_{nh}(k)$. The process $\{N_{nh}(k)\}_{k\geq 1}$ can be thought of as a 'counting' process with the discrete time parameter k. [Henceforth we shall drop the index k to indicate a sequence over $k \geq 1$.] The index n indicates the aggregate over n individuals, while h represents the type label. We impose the following restriction on $\Delta_{ig}(k)$:

A1. $\sum_{h=1}^{H} \Delta L_{ih}(k) \leq 1$.

This assumption says that the same individual can not have two different types of jump at the same time. We can define the predictable process λ_{nh} as

$$\lambda_{nh}(k) = E[\Delta N_{nh}(k)|\mathcal{F}_{n,k-1}].$$

This process is analogous to the stochastic intensity of a usual counting process. In the same manner as Aalen (1978) we postulate that

$$E[\Delta L_{ih}(k)|\mathcal{F}_{n,k-1}] = \alpha_h(k) X_{ih}(k) \quad a.s. \tag{1}$$

where for each h, $X_{1h}, \ldots, X_{nh}$ are predictable binary processes which are independent and identically distributed. One the other hand the deterministic sequence α_h is defined as

$$\alpha_h(k) = \begin{cases} P[\Delta L_{ih}(k) = 1|X_{ih}(k) = 1] & \text{if } P[X_{ih}(k) = 1] > 0 \\ 0 & \text{otherwise} \end{cases}$$

for any i. We put the following extra condition on $X_{ih}(k)$:

A2. There is a partition $\{H_1, H_2, \ldots, H_m\}$ of the set $\{1, 2, \ldots, H\}$ such that the following holds:

$$X_{ih}(k) = 1 \text{ and } h \in H_r \text{ implies } X_{ig}(k) = \begin{cases} 1 & \text{if } g \in H_r, \\ 0 & \text{if } g \in H_s,\ s \neq r. \end{cases}$$

This means the types of jumps can be grouped in such a manner that an individual can be ready for only one group of jump types at a time. This is a very reasonable assumption that helps simplify the likelihood expression in the next section. From (1) it follows that

$$\lambda_{nh}(k) = \alpha_h(k)Y_{nh}(k) \quad a.s.$$

where $Y_{nh}(k) = \sum_{i=1}^{n} X_{ih}(k)$, the number (out of n) of individuals ready for a type h jump at time k. We shall call α_h the deterministic intensity of type h transition. The above is similar to the celebrated multiplicative intensity model (Aalen, 1978). The crucial difference of the above model with Aalen's model is that while the time parameter in the latter is allowed to take values in an interval of the form $[0,T]$, no restriction of finiteness is necessary on the discrete time here. Further, multiple jumps of the process N_{nh} are allowed. In this respect it is comparable to Johansen's (1983) extension of the Aalen model. On the other hand, a specific decomposition of the processes N_{nh} and Y_{nh} are necessary here, unlike the corresponding processes in the continuous time case. The advantage of this assumption will be apparent in the next section. It may not be very restrictive, since we do not know of any real application of the continuous time model where this decomposition does not hold.

We conclude this section with discrete time versions of two examples from Andersen and Borgan (1985).

EXAMPLE 1: *Survival data with random right-censoring.* Suppose the independent and identically distributed (i.i.d.) censored lifetimes of n individuals are denoted by $T_1, T_2, \ldots, T_n$. These assumed to be positive and integer-valued, while any unit of time can be used. Let $\delta_1, \delta_2, \ldots, \delta_n$ be the corresponding censoring indicators. For $1 \le i \le n$ and each $k \in I\!N$ we define $\Delta L_i(k)$ to be the indicator of $\{T_i = k, \delta_i = 1\}$, dropping the subscript h for simplicity. Then (1) holds with $X_i(k)$ as the indicator of $\{T_i \ge k\}$. One can interpret $Y_n(k)$ as the number of individuals at risk at time k. If the censoring time and the notional lifetimes are independent, $\alpha(k)$ becomes the discrete hazard rate.

EXAMPLE 2: *Finite state Markov chain.* Suppose we have n samples from a discrete parameter Markov chain with a finite number of states and having H possible types of transition. For $h = 1, 2, \ldots, H$, $i = 1, 2, \ldots, n$ and $k \in I\!N$ we define $\Delta L_{ih}(k)$ as the indicator of the event that a 'type h' transition occurs to individual i at time k. Thus $\alpha_h(k)$ becomes a transition probability. Note that the assumption A2 holds here with m equal to the number of non-absorbing states in the Markov chain and the set H_r represents the indices of types of transitions that are possible from the rth state. In the special case of competing risk data, $\Delta L_{ih}(k)$, $X_{ih}(k)$ and $\alpha_h(k)$ are the indicator of death from cause h, the indicator of being alive (irrespective of i) and the discrete cause-specific hazard rate, respectively. Similar interpretations can be given in the special case of the illness-death model.

3. NONPARAMETRIC ESTIMATION OF DETERMINISTIC INTENSITY

3.1. *The maximum likelihood estimator.* The likelihood function is

$$\prod_k \left[\left\{ \prod_{h=1}^{H} \alpha_h(k)^{\Delta N_{nh}(k)} \right\} \prod_{i=1}^{n} \left(1 - \sum_{h=1}^{H} \alpha_h(k) X_{ih}(k) \right)^{1-\sum_{h=1}^{H} \Delta L_{ih}(k)} \right].$$

In view of assumption A2, the likelihood function reduces to

$$\prod_k \prod_{r=1}^{m} \left[\left\{ \prod_{h \in H_r} \alpha_h(k)^{\Delta N_{nh}(k)} \right\} \left(1 - \sum_{h \in H_r} \alpha_h(k) \right)^{Y_{nh}(k) - \sum_{h \in H_r} \Delta N_{nh}(k)} \right]$$

which is evidently maximised by the estimator

$$\widehat{\alpha}_{nh}(k) = J_{nh}(k) Y_{nh}^{-1}(k) \Delta N_{nh}(k) \tag{2}$$

where $J_{nh}(k)$ is the indicator of the event $Y_{nh}(k) > 0$. It is interesting to note that the maximisation of the likelihood function is much more difficult in the continuous time case (see Karr, 1987) unless the model is modified in some manner (Johansen, 1983; Jacobsen, 1984). We emphasize the importance of assumption A2 here. Since it holds in most practical situations, unrestricted maximisation of the likelihood may be misleading.

From (2) one can also find the mle of the discrete cumulative intensity (given by $\sum_{l=1}^{k} \alpha_h(l)$). Its computational form coincides with that of the Nelson-Aalen estimator (Aalen, 1978). In the case of survival data with censoring time independent of notional lifetime, the nonparametric ml or Kaplan Meier estimator of the survival function is given by $\prod_{i:k_i \le k}(1 - \widehat{\alpha}_n(k_i))$ where $k_1, k_2, \ldots$ are the times of observed death. When transformed to get an estimator of the hazard rate, it produces (2), as expected. It is also not surprising that Hjort (1985) found (2) as the mle of α in this special case.

3.2. *Bias of the mle.* Note that

$$\widehat{\alpha}_{nh}(k) - \alpha_h(k) = J_{nh}(k) Y_{nh}^{-1}(k) \Delta M_{nh}(k) - (1 - J_{nh}(k)) \alpha_h(k) \tag{3}$$

where $\Delta M_{nh}(k) = \Delta N_{nh}(k) - \lambda_{nh}(k)$ is a martingale difference with respect to the filtration $\{\mathcal{F}_{n,k}\}_{k \ge 1}$. It follows that

$$E[\widehat{\alpha}_{nh}(k) - \alpha_h(k)] = -\alpha_h(k) P^n[X_{1h}(k) = 0].$$

Thus the estimator has a finite negative bias (unless $X_{1h}(k) > 0$ *a.s.*) which goes to 0 at an exponential rate as $n \longrightarrow \infty$.

3.3. *Consistency.* Consider the space S of all real sequences endowed with the Fréchet metric defined by

$$\rho(x, y) = \sum_{k=1}^{\infty} 2^{-k} \frac{|x(k) - y(k)|}{1 + |x(k) - y(k)|}, \qquad x, y \in S,$$

where $x(k)$ is the kth coordinate of the sequence x. Suppose $(\Omega, \mathcal{S}, P)$ is a probability space and $Z_1, Z_2, \ldots$ is a sequence of functions mapping from Ω into S, while Z is another such function.

THEOREM 3.1. *The sequence $\{\rho(Z_n, Z)\}_{n \ge 1}$ converges to 0 in probability if and only if $\{Z_n(k)\}_{n \ge 1}$ converges to $Z(k)$ in probability for each $k \in I\!N$.*

PROOF. The 'only if' part is easy to prove. To prove the 'if' part let $0 < \epsilon < 1$ and k be an integer satisfying $2^{-k} < \epsilon$. Then

$$\begin{aligned} P[\rho(Z_n, Z) > \epsilon] &\leq P\left[\sum_{l=1}^{k} 2^{-l} \frac{|Z_n(l) - Z(l)|}{1 + |Z_n(l) - Z(l)|} > \epsilon - 2^{-k}\right] \\ &\leq P\left[\bigcup_{l=1}^{k} \left\{2^{-l} \frac{|Z_n(l) - Z(l)|}{1 + |Z_n(l) - Z(l)|} > \left(\frac{\epsilon - 2^{-k}}{1 - 2^{-k}}\right) 2^{-l}\right\}\right] \\ &\leq \sum_{l=1}^{k} P[|Z_n(l) - Z(l)| > (\epsilon - 2^{-k})(1 - \epsilon)^{-1}]. \end{aligned}$$

Since the limit of the last expression as $n \longrightarrow \infty$ is 0, the theorem is proved. □

In view of the above theorem, the sequence $\widehat{\alpha}_{nh}$ is consistent as long as $E[\widehat{\alpha}_{nh}(k) - \alpha_h(k)]^2$ goes to 0 for each k as $n \longrightarrow \infty$. Indeed, some algebraic manipulations show that the mean squared error is

$$\alpha_h^2(k) P^n[X_{1h}(k) = 0] + \alpha_h(k)(1 - \alpha_h(k)) \sum_{i=1}^{n} \frac{1}{i} \binom{n}{i} P^i[X_{1h}(k) = 1] P^{n-i}[X_{1h}(k) = 0].$$

The above is identically 0 when $P[X_{1h}(k) = 1] = 0$. Otherwise the second term is upper-bounded by $2n^{-1}\alpha_h(k)(1 - \alpha_h(k))/P[X_{1h}(k) = 1]$. This establishes the required mean square convergence.

3.4. *Asymptotic normality.* The topology induced by ρ on S is the product topology. Indeed, the convergence of a sequence in S is equivalent to the convergence of each coordinate in $I\!R$ with respect to the Euclidian norm. It can be shown that the space S is complete and separable with respect to the metric ρ. Suppose $\mathcal{S}$ is the Borel σ-algebra on S and $\{P_n\}_{n \geq 1}$ is a sequence of probability measures on $(S, \mathcal{S})$. Then the tightness of this sequence is equivalent, by Prohorov's theorem, to its relative compactness. In such a case the convergence of all finite-dimensional marginal distributions of $\{P_n\}_{n \geq 1}$ implies weak convergence of the entire family (see Billingsley, 1968, page 35). Therefore one only needs a suitable compact subset of S in order to relate weak convergence of the coordinates to that of the family. It can be shown that a subset T of S has a compact closure if and only if the subset $\{x_k : x \in T\}$ of $I\!R$ is bounded for every k (see Billingsley, 1968, page 219). Thus we have the following theorem on the sequence $\{P_n\}_{n \geq 1}$.

THEOREM 3.2. *Suppose for each $\eta \in (0, 1)$ there is a positive sequence $\{B_k\}_{k \geq 1}$ such that $\inf_{n \in N} P_n[T] > 1 - \eta$ holds with $T = \bigcap_{k \geq 1}\{x : |x(k)| \leq B_k\}$. Then the weak convergence of all finite-dimensional marginal distributions of $\{P_n\}_{n \geq 1}$ implies the weak convergence of the entire family.*

Let Z and $Z_1, Z_2, \ldots$ be mappings from Ω to S. Then the above theorem can be restated as follows.

COROLLARY 3.3. *Suppose for each $\eta \in (0, 1)$ there is a positive sequence $\{B_k\}_{k \geq 1}$ such that*

$$\inf_{n \in N} P[|Z_n(k)| \leq B_k \text{ for each } k \in I\!N] > 1 - \eta.$$

Then $\{Z_n\}_{n\geq 1}$ *converges weakly to* Z *if and only if for each* $l \in I\!N$ *and every subset* $\{i_1, i_2, \ldots, i_l\}$ *of* $I\!N$ *the sequence of random vectors* $\{Z_n(i_1), Z_n(i_2), \ldots, Z_n(i_l)\}$ *converges weakly to the random vector* $\{Z(i_1), Z(i_2), \ldots, Z(i_l)\}$ *as* $n \longrightarrow \infty$.

The sufficient condition stated above is easily verified for $\sqrt{n}(\widehat{\alpha}_{nh}(k) - \alpha_h(k))$. Indeed

$$
\begin{aligned}
&P[\sqrt{n}|\widehat{\alpha}_{nh}(k) - \alpha_h(k)| < B_{hk} \text{ for each } h, k] \\
&= 1 - P\left[\bigcup_{k\geq 1}\bigcup_{h=1}^{H}\left\{|\widehat{\alpha}_{nh}(k) - \alpha_h(k)| \geq n^{-1/2}B_{hk}\right\}\right] \\
&\geq 1 - \sum_{k\geq 1}\sum_{h=1}^{H} n^{-1}B_{hk}^{-2}E[\widehat{\alpha}_{nh}(k) - \alpha_h(k)]^2.
\end{aligned}
$$

One can choose $B_{hk} = 1$ whenever $P[X_{1h}(k) = 1] = 0$. Otherwise it can be chosen in such a way that

$$C_{hk} + 2\alpha_h(k)(1 - \alpha_h(k))/P[X_{1h}(k) = 1] < B_{hk}^2\eta H^{-1}2^{-k}$$

where C_{hk} is an upper bound of the convergent sequence $\{n\alpha_h^2(k)P^n[X_{1h}(k) = 0]\}_{n\geq 1}$. The sufficient condition of Corollary 3.3 is then satisfied.

Convergence of any finite-dimensional marginal distribution to a multivariate normal one essentially rests on the fact that the second term in (3) decreases faster than the first and that $\Delta M_{nh}(k)$ can be written as the sum of iid variables $\Delta L_{ih}(k) - \alpha_h(k)X_{ih}(k)$. A standard multivariate central limit theorem (e.g., Theorem 29.5 of Billingsley, 1985) can be used. The limiting covariance of $\sqrt{n}(\widehat{\alpha}_{nh}(k) - \alpha_h(k))$ with $\sqrt{n}(\widehat{\alpha}_{ng}(l) - \alpha_g(l))$ can be shown to be zero unless $l = k$, in which case it is given by

$$
\tau_{hg}(k) = \begin{cases}
\alpha_h(k)(1 - \alpha_h(k))/E[X_{1h}(k)] & \text{if } g = h, \\
-\alpha_h(k)\alpha_g(k)/E[X_{1h}(k)] & \text{if } g \neq h, \quad h, g \in H_r \text{ for some } r, \\
0 & \text{if } g \neq h, \quad h \in H_r, g \in H_s,\ s \neq r.
\end{cases}
$$

The fact that $X_{1g}(k)$ is either equal to $X_{1h}(k)$ or $1 - X_{1h}(k)$ (depending on whether h and g belong to the same H_r) has been used in deriving the above. That $\tau_{hg}(k)$ is not necessarily 0 for $g \neq h$ underscores the complications arising from different kinds jumps being allowed to occur (to different individuals) at the same time. The covariance can be consistently estimated by replacing $\alpha_h(k)$ and $\alpha_g(k)$ by their respective mle's and $E[X_{1h}(k)]$ by $n^{-1}Y_{nh}(k)$ in the above expression. We call this estimator $\widehat{\tau}_{ngh}(k)$ for future reference.

4. Tests of Hypotheses

4.1. *The one-sample problem.* Suppose $\alpha_1^0, \ldots, \alpha_H^0$ are specified sequences and we want to test

$$\mathcal{H}_0 : \alpha_h = \alpha_h^0, \quad h = 1, 2, \ldots, H \tag{4}$$

One can use $T_{nh} = \sqrt{n}\sum_{k\geq 1} W_{nh}(k)(\widehat{\alpha}_{nh}(k) - J_{nh}(k)\alpha_h^0(k))$ for this purpose, where W_{nh} are predictable sequences converging to (possibly unknown) non-random real sequences for each h. The covariance of T_{nh} and T_{ng} can be consistently estimated by

$\sum_{k\geq 1} W_{ng}(k)W_{nh}(k)\hat{\tau}_{ngh}(k)$. Thus one can form an asymptotically normal statistic for $H = 1$ and asymptotically χ^2_H-distributed statistic for $H > 1$.

4.2. *The K-sample problem.* Suppose $N^{(1)}_{n_1}, \ldots, N^{(K)}_{n_K}$ are independent H-variate counting processes and we want to test

$$\mathcal{H}_0 : \alpha^{(1)}_h = \alpha^{(2)}_h = \cdots = \alpha^{(K)}_h, \quad h = 1, 2, \ldots, H \tag{5}$$

Following Aalen (1978) one can combine the samples and compare the pooled estimator of α_h with the estimator from individual samples. The reader is referred to Sengupta and Jammalamadaka(1990) for a detailed discussion of the resulting test statistic.

In the special case of $K = 2$, the statistic $U_n = T'_n \widehat{C}_n^{-1} T_n$ can be used, where the components of T_n and $\widehat{C}_n$ are given by

$$\begin{aligned} T_{nh}(k) &= \sqrt{n} \sum_{k\geq 1} W_{nh}(k)(\widehat{\alpha}^{(1)}_{n_1 h}(k) - \widehat{\alpha}^{(2)}_{n_2 h}(k)) \\ \widehat{C}_{nhg} &= \sum_{k\geq 1} W_{nh}(k)W_{ng}(k)(\widehat{\tau}^{(1)}_{n_1 hg}(k) + (\widehat{\tau}^{(2)}_{n_2 hg}(k)) \end{aligned}$$

where $W_{nh}(k)$ is a weight function as before. The asymptotic distribution of U_n is χ^2_H.

The above test was applied to the data on mating rates of 'Ebony' and 'Oregon' flies, which was also examined by Aalen (1978). The data consists of time measured in seconds from introduction in control chamber to the initiation of mating. Since the measurements are discrete, the method discussed here is directly applicable. The statistic U_n produced two-sided p-values of 3.1×10^{-4} and 2.5×10^{-6} for $W_n(k) = n^{-1}Y^{(1)}_{n_1}(k)Y^{(2)}_{n_2}(k)/[Y^{(1)}_{n_1}(k) + Y^{(2)}_{n_2}(k)]$ and $W_n(k) = n^{-2}Y^{(1)}_{n_1}(k)Y^{(2)}_{n_2}(k)$, respectively. Thus the hypothesis of equal (discrete) mating rates is rejected.

4.3 *Other tests.* In the same manner as above, one can formulate a test for equality of components or groups of components of a multivariate intensity function. The asymptotic results of Section 3 will again be useful. An important potential application of this test is in comparing cause-specific hazards of competing risks.

Another hypothesis of interest in the context of two samples is $\alpha^{(2)}(k) = \theta\alpha^{(2)}(k)$ for some unknown θ. A starting point would be to define the estimator

$$\widehat{\theta}_{n,W} = \frac{\sum_{k\geq 1} W_n(k)\widehat{\alpha}^{(2)}_{n_2}(k)}{\sum_{k\geq 1} W_n(k)\widehat{\alpha}^{(1)}_{n_1}(k)}$$

where W_n is a suitable weight function. The asymptotic normality of $\sqrt{n}(\widehat{\theta}_{n,W} - \theta)$ is easy to establish, while a computational formula for the asymptotic variance can also be obtained routinely. This gives a test for $\alpha^{(2)} = \theta_0\alpha^{(1)}$ for fixed θ_0. A reasonable statistic for the original problem is $\sqrt{n}\sum_{k\geq 1} M_n(k)(\widehat{\alpha}^{(2)}_{n_2}(k) - \widehat{\theta}_{n,W}\widehat{\alpha}^{(1)}_{n_1}(k))$ where M_n is another weight function. Further details of these tests may be found in Sengupta and Jammalamadaka(1990).

Concluding Remarks

The asymptotic approach used here is not a competitor of that used by Arjas (1985). His approach of letting the observation time go to infinity is motivated by a desire to process the data in real time and is suitable for regression models, where information tends to accumulate with time. Our approach uses all the data at a time and is suitable for estimation and testing involving the deterministic Intensity. We believe that the two approaches will be useful in analysing real data in a somewhat complementary manner.

Reference

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.

Andersen, P.K. and Borgan, Ø (1985). Counting process models for life history data: a review. *Scand. J. Statist.* **12**, 97–140.

Arjas, E. (1985). Contribution to the discussion of Andersen and Borgan's paper "Counting process models for life history data: a review". *Scand. J. Statist.* **12**, 150–153.

Arjas, E. and Haara, P. (1987). A logistic regression model for hazard: asymptotic results. *Scand. J. Statist.* **14**, 1–18.

Arjas, E. and Haara, P. (1988). A note on the asymptotic normality in the Cox regression model. *Ann. Statist.* **16**, 1133–1140.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Billingsley, P. (1968). *Probability and Measure.* Wiley, New York.

Hjort, N.L. (1985). Contribution to the discussion of Andersen and Borgan's paper "Counting process models for life history data: a review". *Scand. J. Statist.* **12**, 141–150.

Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model: a survey. *Int. Statist. Rev.* **52**, 193–207.

Johansen, S. (1983). An extension of Cox's regression model. *Int. Statist. Rev.* **51**, 165–174.

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

Karr, A.F. (1987). Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann. Statist.* **15**, 473–490.

Lawless, J.F. (1982). *Statistical Models and Methods for Censored Survival Data.* Wiley, New York.

Sengupta, D. and Jammalamadaka, S.R. (1990). Inference on discrete life history data. *Tech. Report # 116*, Univ. of Calif. Santa Barbara.