

ISBN 82-553-0561-0

No 9

October 29

1984

COUNTING PROCESS MODELS FOR
LIFE HISTORY DATA: A REVIEW*

by

Per Kragh Andersen
Statistical Research Unit, Copenhagen

Ørnulf Borgan
University of Oslo

* Presented at the 10th Nordic Conference on Mathematical Statistics
at Bolkesjø, Norway, 17th - 21st June 1984.

STATISTICAL RESEARCH REPORT - Institute of Mathematics,
University of Oslo



ABSTRACT

A survey is given of the development of statistical models for life history data based on counting processes. This development was initiated by Aalen's 1975 thesis from Berkeley. We review nonparametric estimation and testing procedures for counting process intensities, kernel function smoothing, parametric inference and various regression techniques, including a generalization of the Cox regression model for censored survival data.

Key words: Censoring, intensity, Markov chain, martingale, multiplicative intensity model, multivariate counting process, nonparametric inference, parametric inference, regression models, stochastic integral, survival analysis.

1. INTRODUCTION

Life history analysis (or event-history analysis) finds applications in actuarial science, demography, epidemiology, medical research, reliability analysis, micro-sociology, and possibly other fields. In this theory, individual life histories are seen as independent sample paths of stochastic processes moving between states in a discrete state-space. The states of the processes correspond to various statuses for an individual, an insurance policy, a technical component, or whatever we are studying while transitions between the states correspond to occurrence of the events of interest. Most often the object of study is the rate or intensity at which an event occurs. Thus, typically a statistical model for life history data includes a specification of how the various intensities depend on time and on individual characteristics and outside events that are being observed. The study of the simplest situation, in which there are only the two states "alive" and "dead" (or "functioning" and "not functioning"), is often called life-table analysis, survival analysis, or failure time analysis. In this case the intensity of the event "death" is simply the hazard rate function for the survival time distribution.

A special feature of this field of statistics is that one is rarely able to observe complete life-histories. This phenomenon, called censoring, may for instance, be due to the planned termination of a clinical trial, or due to the planned removal of certain test objects in a study of the life distribution of a technical component.

Starting with the work of J. Graunt in 1662 (cf. Glass, 1950; Benjamin, 1978), life-table analysis has been studied for centuries by actuaries and demographers. Other important elements in the theory of life history analysis, like the three state illness-death model (or disability model) and the product-limit estimator frequently named after Kaplan and Meier (1958), also have a history of more than 70 years, with roots back to Karup (1893) and Böhmer (1912). However, in spite of this long history, it seems appropriate to date a modern statistical approach to life history analysis to the beginning of the 1950's. Important contributions from this period are the stochastic illness-death model of Fix and Neyman (1951), and Halperin's (1952) and Epstein and Sobel's (1953) study of maximum likelihood estimation for parametric life time models under certain types of censorship.

In the years that have followed, most of the research effort has gone into the study of survival analysis, or failure time analysis, which has indeed been established as a field of its own. Some important contributions have also appeared on more general life history models, usually in a Markov chain setting. The works of Freund (1961), Sverdrup (1965), Chiang (1968), and Hoem (1972, 1976) are well worth mentioning. Only quite recently, however, has a theory been presented that allows for a unified treatment of the statistical methods of survival analysis and the more general life history models. To give a review of this theory and its applications is, in fact, the purpose of this paper. Before we turn to that, however, we will give a brief outline of the developments in survival analysis.

Following the papers by Halperin (1952) and Epstein and Sobel (1953), much work was done in the 1950's, and especially in the 1960's and 1970's, on developing parametric statistical models for

censored failure time data. Lawless (1983) reviews the work in this area. The parametric methods have found widespread use in the analysis of censored failure time data arising in engineering settings.

In biostatistical applications it was often found impossible to justify a particular parametric life-time model. Initiated by the paper by Kaplan and Meier (1958), which discussed the product-limit estimator for the survival distribution function, much effort has gone into the development of nonparametric methods for censored survival data. Some important contributions are the generalizations of the Wilcoxon, Kruskal-Wallis and Savage (or "log-rank") tests to censored data (Gehan, 1965; Breslow, 1970; Peto and Peto, 1972).

During the 1960's several papers appeared on parametric regression models for censored survival data, making it possible to include explanatory variables (or covariates) in the analysis. Kalbfleisch and Prentice (1980, p. 68) give references to such papers. In 1972, Cox proposed a semiparametric regression model for censored survival data, modelling the hazard rate function of the lifetime distribution as a product of one parametric term and one which was left completely arbitrary. Cox's regression model quickly became very popular, and it has had an enormous influence on applied as well as theoretical research in biostatistics. Like many of the recently proposed methods, Cox's regression model requires modern computing equipment to be applicable, so the concurrent development of modern computers has been one of the prerequisites for this methodological work.

The life history model that has been discussed most frequently in the literature, apart from the simple survival data model,

is the competing risks or multiple decrement model, where more than one cause of death (failure) is considered. But usually this model, as well as the survival data model, have been formulated by means of random variables, and the statistical methods have typically been derived and studied by means of results for i.i.d. random variables. In life history analysis, time and random phenomena occurring in time play an essential role, and it seems therefore more natural to study life history analysis in terms of the theory of stochastic processes. Thus, the formulation in terms of random variables may have contributed to hampering the researchers working in the field of survival analysis, or failure time analysis, from extending their otherwise fine methodology to more general life history models.

Such an extension was facilitated by the fundamental work of Aalen (1975, 1978), which was a decisive breakthrough for the use of modern theory of stochastic processes in life history analysis. Aalen showed how the theory of multivariate counting processes provides a general framework in which both censored failure time data and censored observations from inhomogeneous Markov chains may be analysed, and he studied the empirical cumulative intensity estimator (Nelson, 1969; Altshuler, 1970) and nonparametric two-sample tests. Later, this approach, which relies heavily on modern theory of time-continuous martingales and stochastic integrals, has been used to extend other well-known methods from the survival analysis literature, such as nonparametric k-sample and one-sample tests (Andersen et al., 1982), Cox's regression model (Andersen and Gill, 1982), kernel function smoothing of cumulative intensities (Ramlau-Hansen, 1983a,b), and maximum likelihood estimation in parametric settings (Borgan, 1984), to the more general models of life history analysis. This counting process approach also has

the important advantage of providing straightforward, but rigorous, proofs for the distributional properties of the various estimators and test statistics under very general censoring patterns (Aalen, 1978; Aalen and Johansen, 1978; Gill, 1980a).

The purpose of the present paper is to give an extensive review of the above mentioned works by Aalen and others. We will aim at interpreting the statistical models, discussing the theoretical results, and give illustrative applications. We will not go deeply into the probabilistic background for the methods we discuss, only in Subsection 3.3 the basic definitions are given, and some references for further reading are provided. Rather in this paper, emphasis will be put on a heuristic introduction to the mathematical framework following the lines of Gill (1984). It should therefore be possible to benefit from the reading of this paper without any prior exposure to the subject.

The plan of the paper is as follows. In Section 2 we present some introductory examples of life history models. A heuristic introduction to the notions of a multivariate counting process, an intensity process, a martingale, and a stochastic integral is given in Section 3, where we also present the fundamental multiplicative intensity model of Aalen (1978) with illustrative examples. The empirical cumulative intensity estimator (or Nelson-Aalen estimator) is introduced in Section 4. There we also show how this estimator may be smoothed by kernel function methods. In Section 5 we present results for nonparametric tests. Parametric alternatives to the nonparametric methods are given in Section 6, while Section 7 contains a discussion of regression models. The main message of this paper is that the theory of counting processes has been extremely useful in the study of statistical methods

for analysing life histories. No tree grows into heaven, however, and even the counting process approach has its limitations. These limitations are discussed in our final Section 8. In an appendix we illustrate the use of the background theory in the derivation of the properties of the Nelson-Aalen estimator and of the maximum likelihood estimator in a simple parametric model.

Throughout this paper we shall concentrate on statistical models for the intensities or rates at which the various events occur. It is, however, worth pointing out that counting processes also have been very useful for the study of product-limit type estimators for the survival distribution, or more generally for the transition probabilities of Markov chain models (Aalen & Johansen, 1978; Gill, 1980a, 1983a).

2. INTRODUCTORY EXAMPLES

To give a more specific introduction to the kind of models and data one encounters in life history analysis, let us in this section consider a few examples more in detail. As mentioned above, we model individual life histories by a stochastic process with finite state space. It is convenient to illustrate such a process by a diagram, where the states are represented as boxes, and where arrows between the boxes indicate the possible direct transitions. The time parameter of the process may be e.g. an individual's age or the time elapsed since the diagnosis of a certain disease. Only rarely will the time parameter correspond to calendar time. This should be kept in mind when we talk about "time" below.

2.1 Survival data

The simplest possible model for life history data is the model illustrated in Fig. 1, where one only has the two states 0 and 1, with state 1 absorbing. We will denote the states "alive" and "dead", respectively, although other names may be more appropriate in some applications. This model is the one underlying most work in survival or failure time analysis.

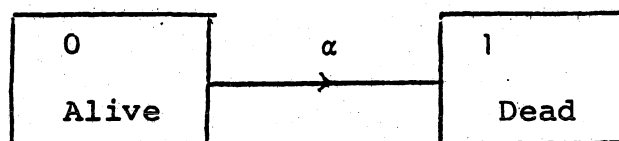


Fig 1. A simple survival data model.

In statistical analysis of survival data from a homogeneous population, one is interested in estimating and testing hypotheses concerning the death intensity (or force of mortality, or hazard function) α . This quantity is defined as follows. Let the random variable T represent the survival time for an individual from the population. Then

$$\alpha(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t < T < t + \Delta t | T > t), \quad (2.1)$$

i.e. $\alpha(t)dt$ is the probability that an individual dies in the small time interval from t to $t + dt$, given that the individual is alive at time t . In this respect α measures the instantaneous death risk.

In many applications, there are explanatory variables (or co-variates) upon which the survival times may depend. These may either be qualitative variables, as indicators for sex, treatment group and geographical region, or quantitative variables like age

when a certain disease was diagnosed and blood pressure. Generally, one has a vector of, possibly time-dependent, covariates $\underline{z}(t) = (z_1(t), \dots, z_p(t))'$ for each individual under study. For such situations one is interested in studying the effect of the covariates on the risk of dying. This is often accomplished by a regression type model, where the death intensity for an individual with covariate vector $\underline{z}(t)$ is assumed to have the multiplicative form

$$\alpha_0(t) e^{\underline{\beta}' \underline{z}(t)} \quad (2.2)$$

Here $\underline{\beta} = (\beta_1, \dots, \beta_p)'$ is a vector of regression parameters, and the underlying death intensity α_0 is the force of mortality for an individual with covariate vector $\underline{z} \equiv \underline{0}$. Within this framework, the effect of the covariates on the risk of dying may be measured by $\underline{\beta}$, while α_0 is a measure of the level of mortality.

The individuals under study may consist of patients at a given hospital suffering from some lethal disease (possibly randomized to one out of a given set of treatments), or they may consist of a cohort or a cross-sectional sample of individuals from some well defined population. The group under study is followed continuously in time, and the occurrences and times of deaths are recorded. Time will often be measured from the date of the entry into the study (the time of randomization). This kind of data collection has an inevitable consequence in the form of right censored data, since in practice one cannot continue the data collection until all individuals are observed to die. Some individuals will still be alive at the end of the study, and for these individuals it will only be known that their survival times exceed certain lower limits. Censoring may also occur because some individuals are lost from follow-up. Thus, statistical methods for

analysing survival data (and other kinds of life history data) must be able to deal with censored observations. A review of the application of counting process methods in the survival analysis set-up is given by Andersen (1982).

The following concrete example of survival data will be used below for illustrative purposes. We consider those among the population of insulin dependent diabetics alive in the county of Funen in Denmark at 1 July 1973 (Green et al., 1981) who had an age at onset of the disease not exceeding 29 years. This group consists of 413 males and 314 females. These individuals were followed until death or emigration or until 1 January 1982. We will show below how these data may be used to estimate the age specific force of mortality among diabetics, and how these estimators may be compared with the death intensity for the general population. Our analysis will also include a comparison of male and female diabetics and a discussion of the influence on survival of the age at onset of the disease (using a model of the form (2.2)).

2.2 Competing risks

When in a survival time study also the cause of death is of interest, the state "dead" in Fig. 1 can be split into, say k states "dead of cause 1", ..., "dead of cause k ", cf. Fig. 2.

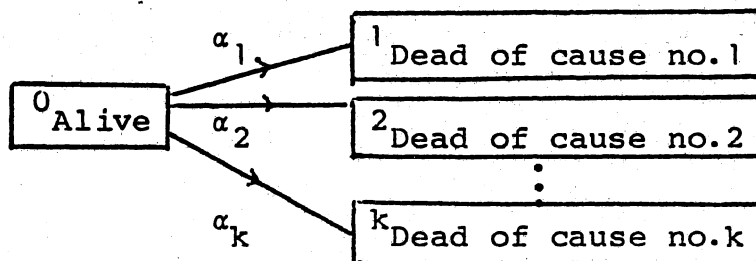


Fig.2. A model for competing risks

In an analysis of data on competing risks from a homogeneous population, the parameters of prime interest are the cause specific death intensities (or hazard functions) $\alpha_1, \alpha_2, \dots, \alpha_k$. These are defined in a similar manner as (2.1), such that $\alpha_h(t)dt$ is the probability that an individual will die of cause no. h in the small time interval from t to $t + dt$, given that the individual is alive at time t . The possible effect of certain covariates on the cause specific death intensities may be studied by a regression type model similar to (2.2).

When studying mortality among diabetics, it is sometimes of interest to analyse deaths caused from direct complications to the disease, deaths due to cardiovascular diseases and deaths due to other causes separately. In the set of data mentioned in Subsection 2.1, however, no reliable information on causes of death is available, and in this paper no real example of a competing risks model will be analysed. Such examples can be found in Prentice et al. (1978) and Aalen (1982a).

2.3. An illness-death model

In a study of life history data, one may be interested in the occurrence of a disease (or some other event) and how this affects the force of mortality. A model for such a situation is displayed in Fig.3. In biostatistics the model is usually called an illness-death model, while actuaries will recognize it as a disability model.

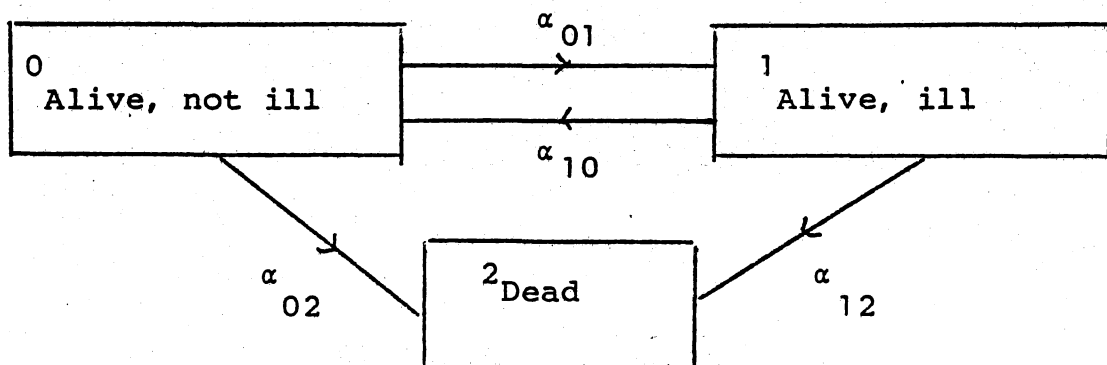


Fig.3. An illness-death or disability model.

For studies of data from a homogeneous population, the stochastic process is often assumed to be Markovian. Then the rates at which the various events occur are measured by the transition intensities, defined as

$$\alpha_{hj}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P_{hj}(t, t+\Delta t). \quad (2.3)$$

Here $P_{hj}(s, t)$ is the probability that an individual in state h at time s will be in state j at time $t \geq s$. Thus α_{01} is the force of morbidity, α_{10} is the cure rate, while α_{12} and α_{02} are death intensities for diseased and disease free individuals, respectively.

More generally, one may assume a semi-Markov structure where e.g. α_{10} and α_{12} depend on time as well as on the duration of the disease. It is also possible to incorporate the duration of the disease as a (time-dependent) covariate, along with other explanatory variables, in a regression model similar to (2.2). A semi-Markov specification is often appropriate in a model for cancer progression, where the state 1 of Fig. 3 corresponds to relapse of the disease (Voelkel and Crowley, 1984).

In this paper we will apply an illness-death model in connection with a study of survival with liver cirrhosis, CSL-I, a randomized clinical trial conducted by the Copenhagen Study Group for Liver Diseases. The purpose of the study was to compare the effect on survival of prednisone treatment versus placebo. (The Copenhagen Study Group for Liver Diseases, 1974). In the period 1962-69, 532 patients with histologically verified liver cirrhosis were included in the study and followed until death or censoring, the closing date of the study being 1 September 1974. The effect on survival of clinical, serological, biochemical and histological

variables measured at the time of entry into the trial were analysed by Schlichting et al. (1983). We shall be concerned also with the effect of "follow-up variables" on survival. These were recorded 3, 6, 9 and 12 months after start of treatment and thereafter once a year. In particular we shall study the effect of the biochemical variable prothrombine on survival during prednisone and placebo treatment, and also how changes in the level of prothrombine may themselves depend on treatment. Restricting attention to either low or normal level we obtain an "illness-death" model for each treatment, time t being measured from the date of randomization.

Another example of a model of the type shown on Fig. 3 was analysed by Andersen and Rasmussen (1982). They studied admissions to and discharges from psychiatric hospitals among women giving birth and women having induced abortion. Here the state 1 corresponded to a woman being resident in a psychiatric hospital, and the state 0 to a woman not being resident in such a hospital. Time t was measured relative to the date of birth /abortion.

2.4. Interaction between life history events

For studying the interaction between two separate events A and B in the life history of an individual, a model of the form displayed in Fig. 4 can sometimes be applied.

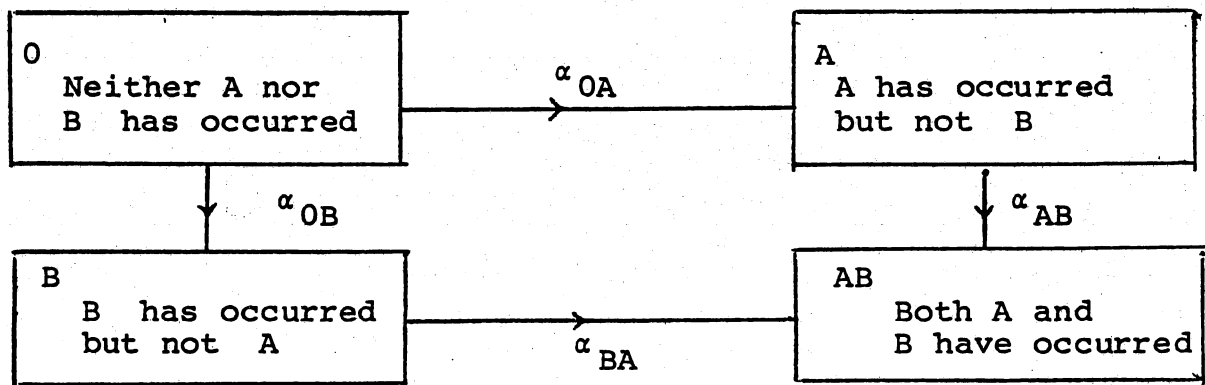


Fig. 4. A model for analysing the interaction between two separate life history events.

Aalen et al.(1980) used a Markov model of this type to study the possible effect of menopausal hormonal changes on the intensity of the outbreak of the chronical skin disease pustulosis palmo-plantaris. Similarly, Borgan et al. (1982) analysed a set of data concerning the interaction between nickel allergy and hand eczema among Danish women by the model of Fig.4.

2.5. Labour Market Dynamics

In longitudinal studies on labour market dynamics, the three states "unemployed", "employed" and "out of labour force" are considered, cf. Fig.5.

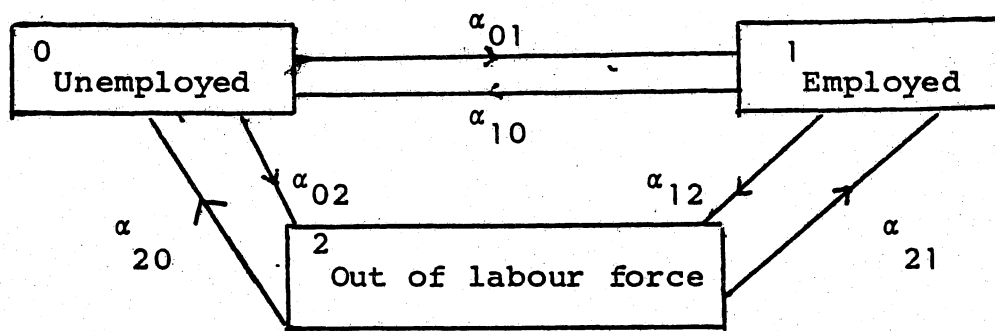


Fig.5. A model for labour force dynamics

Andersen (1985) discussed statistical models for this situation assuming continuous observation of a random sample of individuals from the potential labour force over a fixed calendar time period. Hoem (1977) used an illness-death type model to study the accession to and separation from the Danish labour force for the period 1972-74.

3. MULTIVARIATE COUNTING PROCESSES

In this section we introduce the important concept of a multivariate counting process and the corresponding intensity process, and show how this gives a general framework for analysing the type of situations discussed in Section 2. In Subsection 3.1 we give a somewhat informal introduction to these notions following Gill (1984), and provide illustrative examples. We also com-

ment upon the so-called multiplicative intensity model of Aalen (1978). The informal introduction to the mathematical framework is continued in Subsection 3.2 by a discussion of martingales and stochastic integrals. In the final Subsection 3.3 the precise mathematical results with references are given.

3.1 Multivariate counting processes. The multiplicative intensity model

A multivariate counting process

$\tilde{N} = \{(N_1(t), N_2(t), \dots, N_k(t)), t \in [0, 1]\}$ is a stochastic process with (say) k components, which can be thought of as counting the occurrences, as time t proceeds, of k different types of events, $N_h(t)$ being the number of type h events in the time interval $[0, t]$. In this paper the time parameter t is assumed to vary in a finite interval, which we for convenience in the general discussion will take to be $[0, 1]$. It is assumed that each component process N_h has jumps of size $+1$, and that no two component processes can jump simultaneously. Thus multiple events cannot occur. The events will typically correspond to the transitions, for an individual or a group of individuals, between the various states of a stochastic process as exemplified in Section 2.

The development in time of a multivariate counting process \tilde{N} is governed by its (random) intensity process $\tilde{\lambda} = \{(\lambda_1(t), \dots, \lambda_k(t)), t \in [0, 1]\}$, which is given as follows. Let I_{dt} be a small time interval of length dt around time t , then $\lambda_h(t)dt$ is the conditional probability that N_h jumps in I_{dt} given all that has happened till just before time t . If we let $dN_h(t)$ denote the increment of N_h over I_{dt} , and let \mathcal{F}_{t-} denote everything that has happened up to, but not including t , then we can write

$$\lambda_h(t)dt = \Pr[dN_h(t) = 1 | \mathcal{F}_{t-}]. \quad (3.1)$$

Here the "history" \mathcal{F}_{t-} includes a complete specification of the path of $N(u)$ on the interval $[0, t)$ as well as all other events implicitly or explicitly included in the model which have happened before (but not at) time t . As a consequence we have that $\mathcal{F}_s \subseteq \mathcal{F}_t$, whenever $s < t$, reflecting the fact that as time proceeds more and more is learnt about the process.

Let us see how the examples of Section 2 fit into this framework

Example 1. Survival data

We consider the situation of Subsection 2.1. To be concrete, let us suppose that a group of n patients indexed by $i, i=1, \dots, n$, suffering from a given (lethal) disease is followed at some hospital from the time of diagnosis of the disease to the time of death or to some fixed closing date of the study. Thus for each patient i we observe a disease duration \tilde{T}_i which is either his true survival time T_i , i.e. the length of time from diagnosis to death, or a censoring time; i.e. the length of time from diagnosis to the closing date. Let $D_i = 1$ if \tilde{T}_i is a true survival time; $D_i = 0$ otherwise. Moreover, we assume that the pairs $(T_i, D_i); i = 1, 2, \dots, n$; are independent.

We can define a multivariate counting process \underline{N} by

$$N_i(t) = I(\tilde{T}_i < t, D_i = 1), \quad i = 1, \dots, n, \quad (3.2)$$

where $I(\cdot)$ is the indicator function. Thus N_i is zero before \tilde{T}_i and jumps to one at \tilde{T}_i if \tilde{T}_i is a true survival time; otherwise N_i does not jump at all. To find the corresponding intensity process we argue as follows. At any time t we know that either has the i th patient been observed to die, or he has

been censored, or he is still alive and uncensored. For the first two cases the conditional probability of observing N_i to jump in the interval I_{dt} is zero. For the latter case this conditional probability is $\alpha_i(t)dt$, where $\alpha_i(t)$ is the hazard function, or death intensity, for the true survival time T_i for this patient, cf. (2.1). Thus if we define

$$Y_i(t) = I(\tilde{T}_i > t), \quad (3.3)$$

then we have that

$$\Pr[dN_i(t)=1 | \mathcal{F}_{t-}] = \alpha_i(t) Y_i(t)dt, \quad (3.4)$$

where \mathcal{F}_{t-} represents all the information available on the course of the disease just before time t . (The independence assumption ensures that $\alpha_i(t)$ can be interpreted as the ordinary hazard rate function for individual no. i .) By (3.1) and (3.4), we see that the multivariate counting process $\tilde{N} = (N_1, \dots, N_n)$, given by (3.2), has an intensity process $\tilde{\lambda}$ with components λ_i given by

$$\lambda_i(t) = \alpha_i(t) Y_i(t) \quad ; \quad i = 1, 2, \dots, n. \quad (3.5)$$

In the example concerning survival with insulin dependent diabetes mellitus, the situation is more complicated. Studying mortality as a function of age, rather than as a function of disease duration, the individuals are not followed from age zero, but from their age at 1 July 1973. Denote this age for the i th individual by a_{i0} , and let $N_i(t)$ be 1 if this individual is observed to die in the age span from 0 to t years. Then $\tilde{N} = (N_1, \dots, N_n)$ is a multivariate counting process with intensity process of the form (3.5) with

$$Y_i(t) = I(\tilde{T}_i > t > a_{i0}) \quad , \quad i = 1, \dots, n.$$

In some situations it is reasonable to assume that the death intensities α_i are the same for all individuals, so that we have a homogeneous population. Denote the common value of the α_i s by α . Then we get a univariate counting process N by aggregating the "individual" counting processes (3.2), i.e.

$$N(t) = \sum_{i=1}^n N_i(t). \quad (3.6)$$

This process counts the total number of observed deaths in $[0, t]$. By (3.1), (3.5), and the fact that no two individuals die at the same time, it follows that (3.6) has intensity process λ given by

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t) = \alpha(t) Y(t) \quad (3.7)$$

where

$$Y(t) = \sum_{i=1}^n Y_i(t). \quad (3.8)$$

Note that the right hand side of (3.7) is just a product of the death intensity for a single individual and the "number at risk" just prior to t . \square

Example 2. Observations from a finite state Markov chain

Following the lines from the previous example, we consider a homogeneous group of individuals indexed by $i = 1, \dots, n$. For these individuals we observe, continuously in time, the events of interest, modelled as transitions between the states of a stochastic process with finite state space Γ (compare Subsections 2.2 - 2.5). We define the counting process $N_{hji}(t)$ to be the number of direct transitions from h to j ($h, j \in \Gamma$, $h \neq j$) observed for individual no. i in the time interval $[0, t]$, and assume that the individuals behave independently of each other. Let $Y_{hi}(t) = 1$ if the i th individual is observed to be in state h just prior to

time t , i.e "at risk" for a $h \rightarrow j$ transition; $Y_{hi}(t) = 0$ otherwise. Then, provided that the stochastic process is Markovian (and that the censoring mechanisms satisfy the general conditions discussed at the end of this section), arguments similar to those in Example 1 show that $N_{hji}(t)$ has intensity process $\alpha_{hj}(t) Y_{hi}(t)$, where α_{hj} is the intensity for a $h \rightarrow j$ transition, cf. (2.3). (Also for an inhomogeneous group of individuals or for a non-Markovian process the intensity process for N_{hji} will be a product of an "individual intensity" and the indicator Y_{hi} .) Thus $\tilde{N} = \{ (N_{hji}(t); h, j \in \Gamma, h \neq j, i=1, \dots, n, t \in [0, 1] \}$ is an nk -variate counting process, k being the number of possible types of direct transitions. Analogously to (3.6) and (3.8), we let

$$N_{hj}(t) = \sum_{i=1}^n N_{hji}(t)$$

be the total number of $h \rightarrow j$ transitions observed in $[0, t]$, and

$$Y_h(t) = \sum_{i=1}^n Y_{hi}(t)$$

be the total number of individuals observed to be in state h at t -. Then, since no two transitions occur simultaneously, and the transition intensities for different individuals are assumed to be identical, $\{ (N_{hj}(t), h, j \in \Gamma, h \neq j, t \in [0, 1] \}$ is a k -variate counting process with $N_{hj}(t)$ having intensity process $\alpha_{hj}(t) Y_h(t)$, cf (3.7). \square

Motivated by examples like the ones above, Aalen (1978) introduced the multiplicative intensity model for counting processes where it is assumed that the intensity process (3.1) may be given as

$$\lambda_h(t) = \alpha_h(t) Y_h(t), \quad h = 1, \dots, k, \quad t \in [0, 1]. \quad (3.9)$$

Here $\alpha_h(t)$ is a nonnegative deterministic function, while $Y_h(t)$ is a nonnegative observable stochastic process whose value at any time t is known just before t . We say that a process with these properties is predictable. In the examples, $\alpha_h(t)$ could be interpreted as an individual intensity for making the transition in question (i.e. of type h), and $Y_h(t)$ as the number "at risk" at t^- for making a transition of this type. As a consequence of this, we shall assume, in the general set-up as well, that $Y_h(t) = 0$ whenever $Y_h(t) < 1$. Other examples of the multiplicative intensity model are given by Aalen (1978, Sections 4 and 8; 1982a, Section 4).

Since the development in time of a multivariate counting process \tilde{N} is governed by its intensity process (3.1), we can specify a counting process model for life history data by giving a specification of the intensity process. Let $N_{hi}(t)$ be the number of type h transitions for individual i in $[0, t]$. Then all the statistical models studied in this paper (except for Subsection 7.3) have a common structure, namely that the intensity process of N_{hi} is given by

$$\lambda_{hi}(t) = \alpha_h(t, \tilde{Z}_i(t)) Y_{hi}(t) \quad (3.10)$$

$h = 1, \dots, k; i = 1, \dots, n; t \in [0, 1]$.

The individual intensities α_h may depend on the type h either nonparametrically (Sections 4, 5 and 7) or via a finite number of parameters (Sections 6 and 7), and they may depend on the individual i via a vector of predictable covariate processes $\tilde{Z}_i(t)$, either by a stratification according to the values of \tilde{Z}_i (Section 4, 5, 6) or by a regression model specification, cf. (2.2) (Section 7).

The Y_{hi} are predictable indicator processes, $Y_{hi}(t)$

indicating by the value 1 whether individual i is observed to be at risk just before t for making a type h transition. Let $\tilde{N}_{hi}(t)$ be the number of type h transitions in $[0, t]$ we had observed for individual i if there had been no censoring. Then the value of the indicator processes Y_{hi} is a result of both the development of $\tilde{N}_i = (\tilde{N}_{1i}, \dots, \tilde{N}_{ki})$ up to (but not including) t and of possible censoring. Suppose that the uncensored process \tilde{N}_i satisfies a model of the form (3.10), i.e. that it has intensity process

$$\tilde{\lambda}_{hi}(t) = \alpha_h(t, Z_i(t)) \tilde{Y}_{hi}(t),$$

where \tilde{Y}_{hi} is determined by \tilde{N}_i alone. Moreover, assume that censoring of individual i is determined by a predictable indicator process $C_i(t)$, indicating by the value 1 when this individual is under observation. Then the censored counting process $N_{hi}(t)$ is given by

$$N_{hi}(t) = \int_0^t C_i(s) d\tilde{N}_{hi}(s); \quad h = 1, \dots, k,$$

and it has intensity process given by (3.10) with

$$Y_{hi}(t) = \tilde{Y}_{hi}(t) C_i(t)$$

(Andersen et al., 1982).

Thus censoring by a predictable process C_i preserves the structure of the model, and inference on the $\alpha_h(\dots)$ may be drawn from observing the censored process $\tilde{N} = \{(N_{1i}, \dots, N_{ki}); i=1, \dots, n\}$. This means that the censoring mechanisms may be quite arbitrary, as long as they only depend on the past and outside random variation. It was discussed by Aalen (1978), Gill (1980a) and Andersen et al. (1982) how, in the case of survival data (Example 1, above), the most frequently used models for right

censoring (e.g. type I censorship and random censorship) can in fact be described in this way. Examples of censoring mechanisms which do not satisfy these conditions are given in Section 8.

As shown in the examples above, the model (3.10) may sometimes be reduced to the multiplicative intensity model (3.9) by aggregating the individual processes. Aalen (1982 b) discusses conditions under which such a reduction of the data is sufficient. Especially, he points out that for left censoring and "censoring on intervals" some information may be lost by the aggregation. The statistical methods derived from (3.9) will still be valid, however, as long as the censoring mechanisms satisfy the general conditions discussed above.

3.2. Martingales and stochastic integrals

The study of life history data by means of multivariate counting processes is intimately connected with the use of martingale methods for deriving the properties of the statistical estimation and testing procedures. In this subsection the link between counting processes and martingales is outlined, and stochastic integrals are introduced. Our informal presentation is modelled after Gill (1984).

The increment $dN_h(t)$ of N_h over the small interval I_{dt} of length dt around time t is a 0 - 1 variable. Therefore, by (3.1)

$$E(dN_h(t) | \mathcal{F}_{t-}) = \lambda_h(t)dt. \quad (3.11)$$

This implies that if we define stochastic processes M_h ; $h = 1, \dots, k$; by having increments

$$dM_h(t) = dN_h(t) - \lambda_h(t)dt \quad (3.12)$$

over I_{dt} (and satisfying $M_h(0) = 0$), then

$$E(dM_h(t) | \mathcal{F}_{t-}) = 0. \quad (3.13)$$

i.e. the processes

$$M_h(t) = N_h(t) - \int_0^t \lambda_h(u) du ; \quad (3.14)$$

$h = 1, \dots, k; t \in [0, 1]$; are martingales. In particular $EM_h(t) = 0$ for all $t \in [0, 1]$.

The relation (3.14) is the key to the "counting process approach" to life history analysis. As we will see in Sections 4-7 below, many estimators and test statistics may be expressed as, or approximated by, stochastic integrals with respect to the martingales (3.14). Moreover, central limit theorems and other properties for martingales, and therefore also for stochastic integrals (cf. below), are very well studied, and may be used to investigate the properties of the statistical procedures.

Martingale central limit theorems state conditions under which a sequence $(M^{(n)}(t); t \in [0, 1])$, $n = 1, 2, \dots$ of martingales (not necessarily of the form (3.14)) behaves as a continuous Gaussian martingale when n grows large. A continuous Gaussian martingale $(X(t), t \in [0, 1])$ is a (possibly) time-transformed Wiener process, and as such it has independent normally distributed increments with expectation zero. In particular the conditional variance of $dX(t)$, given all that has happened up to time t , i.e. \mathcal{F}_{t-} , equals $\text{Var}(dX(t))$, and hence it is deterministic. Furthermore, X has continuous sample paths. So if a sequence of martingales asymptotically should look like a continuous Gaussian martingale X , then firstly the jumps of $M^{(n)}$ should become negligible when n gets large, and secondly the conditional variances $\text{Var}(dM^{(n)}(t) | \mathcal{F}_{t-})$ should become deterministic in the limit. The conditional variance of a martingale M is given by the so-called

predictable variation process (or variance process) $\langle M \rangle$, defined by having the increments

$$d\langle M \rangle(t) = \text{Var}(dM(t) | \mathcal{F}_{t-})$$

over I_{dt} . So in conclusion, the second condition for the convergence of a sequence of martingales to a continuous Gaussian martingale, is that the predictable variation processes $\langle M^{(n)} \rangle$ converge to a deterministic function.

For the martingales M_h defined by (3.14), we find

$$\begin{aligned} d\langle M_h \rangle(t) &= \text{Var}(dN_h(t) - \lambda_h(t)dt | \mathcal{F}_{t-}) \\ &= \text{Var}(dN_h(t) | \mathcal{F}_{t-}), \end{aligned}$$

since $\lambda_h(t)$ is predictable, i.e. fixed given \mathcal{F}_{t-} . Because $dN_h(t)$ is a 0-1 variable, (3.1) yields

$$\begin{aligned} d\langle M_h \rangle(t) &= \lambda_h(t)dt (1 - \lambda_h(t)dt) \\ &\approx \lambda_h(t)dt, \end{aligned}$$

and therefore

$$\langle M_h \rangle(t) = \int_0^t \lambda_h(s)ds. \quad (3.15)$$

To study the transformation of a martingale M by stochastic integration, let H be a predictable stochastic process, and define a new process M' by the stochastic integral

$$M'(t) = \int_0^t H(s) dM(s). \quad (3.16)$$

Then M' is a martingale itself, because the increment $dM'(t) = H(t) dM(t)$ over I_{dt} has zero conditional expectation (cf. (3.13)):

$$E(H(t)dM(t) | \mathcal{F}_{t-}) = H(t) E(dM(t) | \mathcal{F}_{t-}) = 0.$$

Here the first equality is due to the predictability of H , while the second follows by (3.13). The predictable variation process of (3.16) is easily found: Since

$$\text{Var} (H(t) dM(t) | \mathcal{F}_{t-}) = H^2(t) d\langle M \rangle(t),$$

we get

$$\langle M \rangle(t) = \int_0^t H^2(s) d\langle M \rangle(s). \quad (3.17)$$

One final concept in the following is the orthogonality of two martingales M_1 and M_2 . To this end we introduce the predictable covariation process (or covariance process) $\langle M_1, M_2 \rangle$, defined by having the increments

$$d\langle M_1, M_2 \rangle(s) = \text{Cov}(dM_1(t), dM_2(t) | \mathcal{F}_{t-})$$

over I_{dt} , and we say that M_1 and M_2 are orthogonal if

$$\langle M_1, M_2 \rangle \equiv 0.$$

For any two martingales M_h and M_j , $h \neq j$, derived from a multivariate counting process by (3.14), we have

$$\begin{aligned} d\langle M_h, M_j \rangle(t) &= \text{Cov}(dN_h(t) - \lambda_h(t)dt, dN_j(t) - \lambda_j(t)dt | \mathcal{F}_{t-}) \\ &= E(dN_h(t) dN_j(t) | \mathcal{F}_{t-}) = 0. \end{aligned}$$

This follows since λ_h and λ_j are predictable, and by the fact that N_h and N_j do not jump simultaneously. Thus, the martingales defined by (3.14) are orthogonal.

3.3 Mathematical framework

After the informal introduction to the various mathematical concepts given in the preceding two subsections, we shall now turn to a precise mathematical formulation. Unlike the rest of the paper, this subsection (and the Appendix) requires knowledge of some basic concepts in measure and probability theory. It may safely be omitted by those of the readers who only want to get a brief review of the ideas and results in the "counting process approach" to life history analysis. Some important references to the theory of counting processes and martingales are Dolivo (1974), Meyer (1976), Bremaud and Jacod (1977), Jacod (1979), Gill (1980a), Bremaud (1981), Shiriyayev (1981) and Jacobsen (1982).

Definitions

Let (Ω, \mathcal{F}, P) be a complete probability space and $(\mathcal{F}_t)_{t \in [0,1]}$ a filtration on (Ω, \mathcal{F}) , i.e. an increasing, right-continuous family of sub- σ -algebras of \mathcal{F} . We also assume that \mathcal{F}_0 contains all P -null sets of \mathcal{F} . A multivariate counting process $\tilde{N} = ((N_1(t), \dots, N_k(t)), t \in [0,1])$ is a k -dimensional stochastic process adapted to the filtration (i.e. $\tilde{N}(t)$ is \mathcal{F}_t -measurable for each $t \in [0,1]$) with components N_h which have sample functions which are nondecreasing, right-continuous step functions, zero at time zero, and with jumps of unit size. Moreover it is assumed that, with probability one, no two components jump simultaneously, and that each $N_h(1)$ is almost surely finite.

An adapted stochastic process M , satisfying $M(0) = 0$, $E|M|(t) < \infty$; $t \in [0,1]$; and having right-continuous sample functions with left hand limits, is called a martingale if

$E(M(t) | \mathcal{F}_s) = M(s)$ a.s. (cf. (3.13)) and a submartingale if $E(M(t) | \mathcal{F}_s) > M(s)$ a.s. for $0 < s < t < 1$. A (sub-)martingale is square integrable if $\sup_{t \in [0,1]} EM^2(t) < \infty$.

A stopping time is a random variable T satisfying $\{T < t\} \in \mathcal{F}_t$ for all t . For a stochastic process X , we define the stopped process X^T by $X^T(\cdot) = X(\cdot \wedge T)$, $s \wedge t$ denoting the minimum of s and t . A stochastic process X is said to have a property locally if there exists an almost surely nondecreasing sequence (T_n) of stopping times with $P(T_n > t) \uparrow 1$, as $n \rightarrow \infty$ for all $t \in [0,1]$, such that for every n , the process X^{T_n} has the actual property. Thus a local martingale, a local square integrable martingale, a locally bounded process, etc. can be defined.

The precise definition of a predictable stochastic process can be found e.g. in Gill (1980a, p. 8-9). For our purpose it is sufficient to note that if a process is adapted and has left-continuous sample paths, then it is predictable and locally bounded. Moreover, any Borel measurable deterministic function is predictable.

A process X has a compensator Λ if $X - \Lambda$ is a local martingale, and Λ is predictable and has paths of locally bounded variation.

Results

Each component N_h of a multivariate counting process has a unique compensator Λ_h . Thus there exist local martingales M_h defined by (cf. (3.14))

$$M_h(t) = N_h(t) - \Lambda_h(t), \quad h = 1, \dots, k.$$

In fact the M_h are local square integrable martingales. This is the Doob-Meyer decomposition of the local submartingale N_h . Under

regularity conditions (e.g. Aalen, 1978, Section 3.2) Λ_h is absolutely continuous, so that there exist predictable processes λ_h such that (cf. (3.14))

$$\Lambda_h(t) = \int_0^t \lambda_h(s) ds; h = 1, \dots, k.$$

Furthermore,

$$\lambda_h(t+) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(N_h(t+\Delta t) - N_h(t) = 1 | \mathcal{F}_t),$$

and hence λ_h is denoted the intensity process for N_h (cf. (3.1)). Conversely, given a process λ_h with the above mentioned properties, then, subject to regularity conditions (e.g. Aalen, 1975, Sections 2C and 2D), a unique counting process can be determined which has λ_h as its intensity process. Throughout this paper we will assume the existence of an intensity process.

If M_1 and M_2 are local square integrable martingales, then $M_1 M_2$ has a unique compensator which we call the predictable covariation process of M_1 and M_2 and denote by $\langle M_1, M_2 \rangle$. We say that M_1 and M_2 are orthogonal when $\langle M_1, M_2 \rangle \equiv 0$. The counting process martingales given by (3.14) are orthogonal. The unique compensator for the local submartingale M^2 is called the predictable variation process of M and is denoted by $\langle M \rangle$. Thus $\langle M \rangle = \langle M, M \rangle$. For the counting process martingales (3.14) the predictable variation processes are given by (3.15).

If H_1 and H_2 are predictable and locally bounded processes, and M_1 and M_2 are local square integrable martingales with paths of locally bounded variation, then the stochastic integrals

$$\int H_h dM_h = \left(\int_0^t H_h(s) dM_h(s), t \in [0,1] \right);$$

$h = 1,2$; can be defined (Gill, 1980a). The stochastic integrals are local square integrable martingales themselves, with predictable covariation process

$$\langle \int H_1 dM_1, \int H_2 dM_2 \rangle(t) = \int_0^t H_1(s) H_2(s) d\langle M_1, M_2 \rangle(s). \quad (3.18)$$

In this paper we are mainly working with local square integrable martingales. This means that we cannot in general be sure that expectations, variances, covariances and correlations do exist. Therefore, when we in Sections 4 and 5 and Appendix A below talk about expectations etc., we do tacitly assume that they exist.

We finally state two theorems which are of fundamental importance when deriving the properties of the estimators and test statistics in the statistical models discussed in Sections 4-7 below.

Theorem 3.1

Let N be a univariate counting process with intensity process λ and let M be a local square integrable martingale (not necessarily given by (3.14)). Then for all $\delta, \eta > 0$

$$P(N(1) > \eta) < \frac{\delta}{\eta} + P\left(\int_0^1 \lambda(t) dt > \delta\right),$$

and

$$P\left(\sup_{t \in [0,1]} |M(t)| > \eta\right) < \frac{\delta}{\eta^2} + P(\langle M \rangle(1) > \delta).$$

This theorem is a consequence of Lenglart's (1977) inequality, cf. Andersen and Gill (1982, Appendix I).

Theorem 3.2 (A central limit theorem for local square integrable martingales)

Let $p > 1$ be fixed, and consider a sequence $\tilde{N}^{(n)}$ of k_n -variate counting processes with intensity process $\tilde{\lambda}^{(n)}$, and a sequence $\tilde{H}^{(n)}$ of $p \times k_n$ -matrices of predictable processes, such that the stochastic integrals

$$X_j^{(n)}(t) = \int_0^t \sum_{h=1}^{k_n} H_{jh}^{(n)}(s) (dN_h^{(n)}(s) - \lambda_h^{(n)}(s)ds);$$

$j = 1, \dots, p$; are well defined. If, as $n \rightarrow \infty$,

$$\langle X_j^{(n)}, X_l^{(n)} \rangle(t) \xrightarrow{P} G_{jl}(t);$$

$j, l = 1, \dots, p$, $t \in [0, 1]$, where G is a $p \times p$ matrix of continuous functions on $[0, 1]$ forming the covariance function of a p -variate Gaussian martingale $\tilde{X}^{(\infty)}$ with $\tilde{X}^{(\infty)}(0) = 0$, and if for all $\epsilon > 0$

$$\int_0^1 \sum_{h=1}^{k_n} [H_{jh}^{(n)}(t)]^2 \lambda_h^{(n)}(t) I(|H_{jh}^{(n)}(t)| > \epsilon) dt \xrightarrow{P} 0;$$

$j = 1, \dots, p$; then $\tilde{X}^{(n)} \xrightarrow{D} \tilde{X}^{(\infty)}$ as $n \rightarrow \infty$.

The weak convergence \xrightarrow{D} takes place in the space $D[0, 1]^p$ equipped with the Skorohod product topology (cf. Billingsley, 1968). Versions of this theorem were proved independently by Aalen (1977) and Rebolledo (1978). Our formulation of Theorem 3.2 is a consequence of the results of Rebolledo (1980), and can be found in Andersen and Gill (1982, Appendix I). Other important papers on martingale central limit theorems are Liptser and Shiriyayev (1980) and Helland (1982).

In the Appendix examples are given of the applicability of these two theorems.

4. NONPARAMETRIC ESTIMATION

Consider a multivariate counting process $N = \{(N_1(t), \dots, N_k(t)), t \in [0, 1]\}$ with intensity process satisfying the multiplicative intensity model (3.9). In this section we consider nonparametric estimation of the integrated α_h s, and how these estimators may be smoothed to obtain estimators for the α_h s themselves. Both methods are illustrated by means of the diabetes data discussed in Subsection 2.1.

4.1 Nonparametric estimation of the integrated α_h s

To derive an estimator for

$$A_h(t) = \int_0^t \alpha_h(s) ds, \quad (4.1)$$

we use (3.9) and (3.12) to write symbolically

" $dN_h(t) = \alpha_h(t)Y_h(t)dt + \text{noise}$ ". By this, a natural estimator of (4.1) would be $\int_0^t Y_h(s)^{-1} dN_h(s)$. However, one may have $Y_h = 0$, and in order to deal systematically with this possibility, we introduce the indicator $J_h(t) = I(Y_h(t) > 0)$, and define the estimator for (4.1) formally by

$$\hat{A}_h(t) = \int_0^t [J_h(s)/Y_h(s)] dN_h(s), \quad (4.2)$$

where $J_h(t)/Y_h(t)$ is interpreted as 0 whenever $Y_h(t) = 0$. It should be recalled that we assume that $Y_h(t) < 1$ implies $Y_h(t) = 0$.

The estimator (4.2) was introduced by Aalen (1975, 1978), and it generalizes the empirical cumulative intensity estimator, proposed independently by Nelson (1969, 1972) and Altshuler (1970), for the set-up with censored failure time data. We will denote (4.2) the Nelson-Aalen estimator.

It should be realized that the integral in (4.2) is just a simple sum. To see this, let $T_{h1} < T_{h2} < \dots$ be the successive

jump times for N_h . Then $dN_h(t) = 1$ when t equals one of these jump times, $dN_h(t) = 0$ otherwise. It follows that (4.2) may be written alternatively as

$$\hat{A}_h(t) = \sum_{\{j: T_{hj} < t\}} [Y_h(T_{hj})]^{-1}. \quad (4.3)$$

Thus, \hat{A}_h is an increasing, right-continuous step-function with increment $1/Y_h(T_{hj})$ at the observed jump time T_{hj} of N_h .

To further motivate the Nelson-Aalen estimator, let us see how it may be derived heuristically from the classical occurrence/exposure rates (cf. Hoem, 1976). To this end, we split the time interval $[0, t]$ by a partitioning $0=t_0 < t_1 < \dots < t_m = t$ which is so fine that in each subinterval at most one jump occurs, and such that $\alpha_h(\cdot)$ is (approximately) constant on each of the subintervals. Denote this constant value on $(t_j, t_{j+1}]$ by α_{hj} , and let Δ be the length of the subintervals (all assumed to be of equal size). Then the occurrence/exposure rate $\hat{\alpha}_{hj}$ for α_{hj} is given (almost) by $[Y_h(t_j)\Delta]^{-1}$ if N_h jumps in the actual subinterval, and it is 0 otherwise. Consequently a natural estimator for $A_h(t) \approx \sum_j \alpha_{hj}\Delta$ is $\sum_j \hat{\alpha}_{hj}\Delta$ which equals (4.3) approximately.

Still another motivation for working with the Nelson-Aalen estimators was given by Johansen (1983) who derived the estimators as maximum likelihood estimators in an extended model where the compensators (cf. Subsection 3.3) for the N_h s are not assumed to be absolutely continuous. Another extension of the model allowing maximum likelihood estimation, but giving rise to different estimators, was discussed by Jacobsen (1982, 1984). He derived the

asymptotic properties of his maximum likelihood estimators by proving that they are asymptotically equivalent to the Nelson-Aalen estimators. We shall not pursue these approaches any further, but motivate the use of the Nelson-Aalen estimators mainly by their nice and easily verifiable properties.

Breslow and Crowley (1974) studied the large sample properties of the Nelson-Aalen estimator for the special case of randomly censored survival data using results for i.i.d. random variables. Aalen (1978) studied these properties in general using the theory for multivariate counting processes, martingales and stochastic integrals (cf. Section 3). Let us recapitulate his line of reasoning. (We have slightly improved some of Aalen's arguments, in particular to make use of later developments in the theory of stochastic integrals.) We introduce

$$A_h^*(t) = \int_0^t \alpha_h(s) J_h(s) ds, \quad (4.4)$$

which is almost the same as (4.1) when there is only a small probability that $Y_h(s) = 0$ for some $s < t$. By (3.9) and (3.14), we then get

$$\hat{A}_h(t) - A_h^*(t) = \int_0^t \frac{J_h(s)}{Y_h(s)} dM_h(s). \quad (4.5)$$

Since J_h/Y_h is a bounded predictable process for each h , the right hand side of (4.5) is a stochastic integral w.r.t. a local square integrable martingale, and hence itself a mean-zero local square integrable martingale.

This fact is the key to the study of the properties of the Nelson-Aalen estimator. To illustrate the use of the theory of martingales and stochastic integrals in the study of statistical methods for life history data, we give in the Appendix a detailed study of the Nelson-Aalen estimator. Let us here just briefly state its properties. (The exact conditions under which the results hold true are given in the Appendix.)

By (4.5), we have for all $t \in [0, 1]$ (assuming that the

expectations exist, cf. remark just above Theorem 3.1)

$$E\hat{A}_h(t) = EA_h^*(t),$$

so that $\hat{A}_h(t)$ is an approximately unbiased estimator for (4.1).

Its variance may be estimated (almost unbiasedly) by

$$\hat{\tau}_h(t) = \int_0^t J_h(s) [Y_h(s)]^{-2} dN_h(s), \quad (4.6)$$

where the integral may be written as a simple sum in a similar fashion as (4.3). Furthermore, viewed as a process of t , $\hat{A}_h(t)$ has (approximately) uncorrelated increments, and $\hat{A}_h(t)$ is (approximately) uncorrelated with $\hat{A}_j(s)$ for any s, t and $h \neq j$. This latter fact is of great practical importance, since it implies that plots of the Nelson-Aalen estimators for $h = 1, 2, \dots, k$ may be judged independently of each other.

If each Y_h increases uniformly over $[0, 1]$, then an application of Theorem 3.1 shows that \hat{A}_h is a uniformly consistent estimator for A_h . Moreover, using Theorem 3.2 it can be shown that (suitably normalized) the \hat{A}_h s will be asymptotically distributed as independent Gaussian martingales. In particular $\hat{A}_h(t)$ will be asymptotically normally distributed with mean $A_h(t)$ and a variance which may be estimated by (4.6).

As in Aalen (1976) (for the special case of a multiple decrement model, cf. Subsection 2.2) one may develop $100(1-\alpha)$ per cent confidence bands for A_h of the form

$$A_h(t) \in [\hat{A}_h(t) - b_\alpha \hat{\tau}_h(t)^{1/2}, \hat{A}_h(t) + b_\alpha \hat{\tau}_h(t)^{1/2}]. \quad (4.7)$$

Here b_α is the upper α -fractile in the well-known distribution of $\sup_{t \in [0, 1]} |W(t)|$, where W is a standard Wiener process. (For a table see Walsh, 1962, p.334.) However, these confidence bands have constant width (determined by $\hat{\tau}_h(1)$) and may therefore be of little practical interest. Instead a transformation to a Brownian bridge can be applied. Let us sketch how this approach can be

used to obtain confidence bands for $A_h(t)$ with a width that increases as the estimated variance $\hat{\tau}_h(t)$ increases.

The idea is that if $X(t)$ is a mean zero Gaussian martingale on $[0,1]$ with $\text{Cov}(X(s),X(t)) = \int_0^{t \wedge s} g^2(u)du = G(t \wedge s)$, then $X(t)G(1)^{\frac{1}{2}}(G(1)+G(t))^{-1}$ is distributed as $W^0\left(\frac{G(t)}{G(1)+G(t)}\right)$, where W^0 is the standard Brownian bridge on $[0,1]$ (see Billingsley, 1968). Now we let X be the limiting process of \hat{A}_h (properly normalized), cf. the Appendix, and use the fact that, for this situation, $\hat{\tau}_h(t)$ (properly normalized) converges in probability to $G(t)$ uniformly in $t \in [0,1]$. Then it follows that $100(1-\alpha)$ per cent confidence bands for $A_h(t)$ are given by

$$A_h(t) \in \left[\hat{A}_h(t) - c_\alpha \hat{\tau}_h(1)^{\frac{1}{2}} \left(1 + \frac{\hat{\tau}_h(t)}{\hat{\tau}_h(1)}\right), \hat{A}_h(t) + c_\alpha \hat{\tau}_h(1)^{\frac{1}{2}} \left(1 + \frac{\hat{\tau}_h(t)}{\hat{\tau}_h(1)}\right) \right], \quad (4.8)$$

where c_α is the upper α -fractile in the distribution of $\sup_{t \in [0, \frac{1}{2}]} |W^0(t)|$, see Hall and Wellner (1980, p. 141).

Example. Survival among insulin dependent diabetics

As described in Example 1 in Subsection 3.1 each patient i in this example is followed from age a_{i0} at 1 July 1973 to age \tilde{T}_i at the exit from the study. In the first place we shall consider a two-dimensional counting process (N_1, N_2) , $N_1(t)$ counting the number of observed deaths among females, and $N_2(t)$ counting those among males. If $\alpha_1(t)$ and $\alpha_2(t)$ denote the age specific death intensities for females and males, respectively, then N_1 has intensity process $\alpha_1(t)Y_1(t)$, where

$$Y_1(t) = \sum_{i \text{ for females}} I(a_{i0} < t < \tilde{T}_i),$$

and similarly for N_2 (Example 1 of Subsection 3.1). Thus the integrated death intensities $A_1(t)$ and $A_2(t)$, see (4.1), can be estimated by the Nelson-Aalen estimators (4.2). Fig. 6 shows the estimates $\hat{A}_1(t)$ and $\hat{A}_2(t)$ for the age interval from 0 to 70 years. (For ages less than the lowest ages at which deaths are observed, 23 and 19 years, respectively, $\hat{A}_1(t)$ and $\hat{A}_2(t)$ are zero.) Also shown are approximate 95% pointwise confidence limits and 95% confidence bands computed from (4.7) (with $b_{0.05} = 2.25$). It is seen that both for females and for males the death intensity (the slope of the plots) seems to increase with age, and comparing $\hat{A}_1(t)$ and $\hat{A}_2(t)$, males seem to have a slightly higher mortality than females. Furthermore, it is seen that the confidence bands based on (4.7) are very wide, and that the pointwise confidence limits do not seem to reflect very well the uncertainty of the entire curve. Fig. 7 shows $\hat{A}_1(t)$ and $\hat{A}_2(t)$ with approximate 95% confidence bands computed from (4.8) (with $c_{0.05} = 1.27$). As expected, these confidence bands are not so wide for low ages as are those based on (4.7). In the next subsection we shall see how the influence of age on the mortality is much more clearly revealed by estimating the death intensities α_1 and α_2 directly. \square

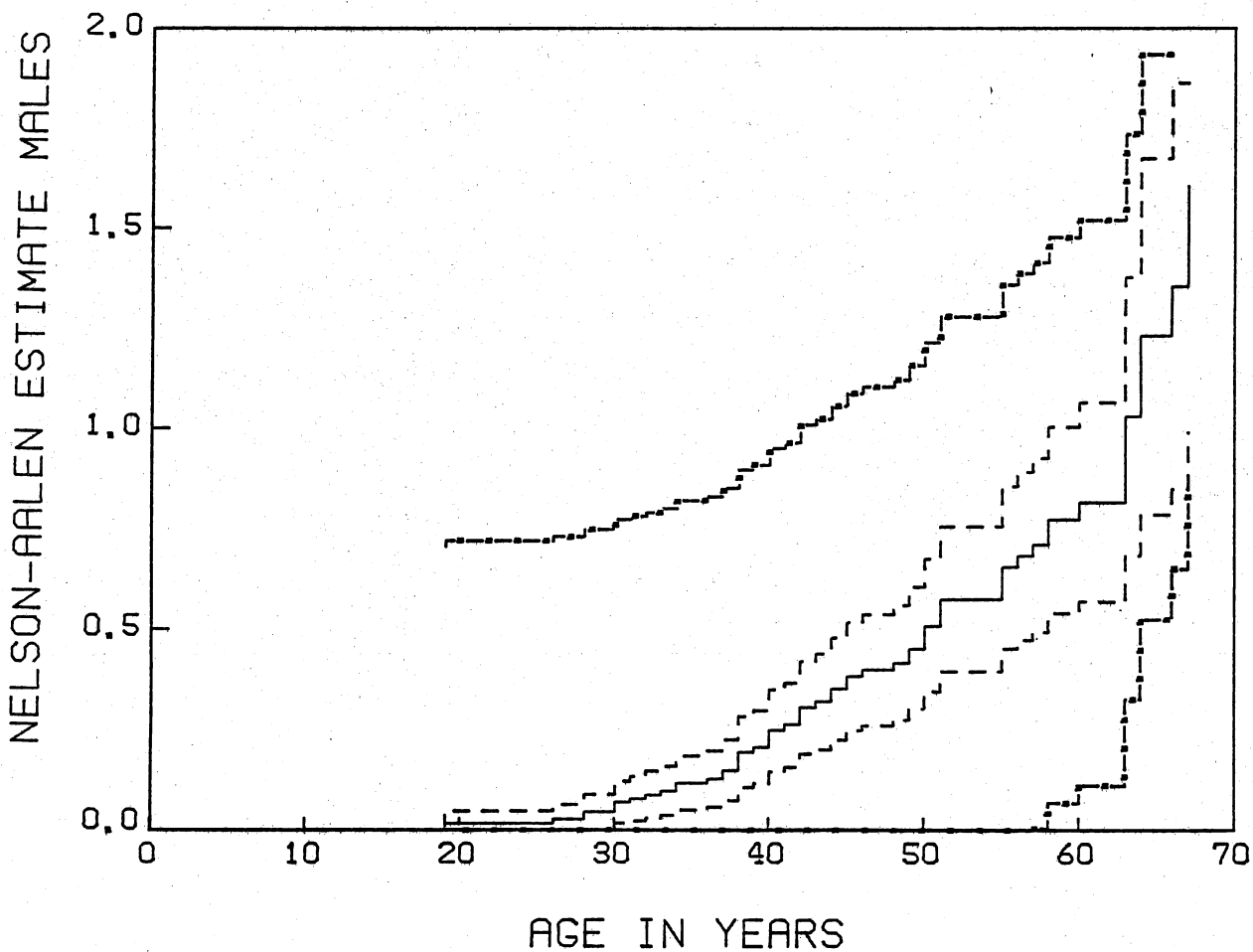
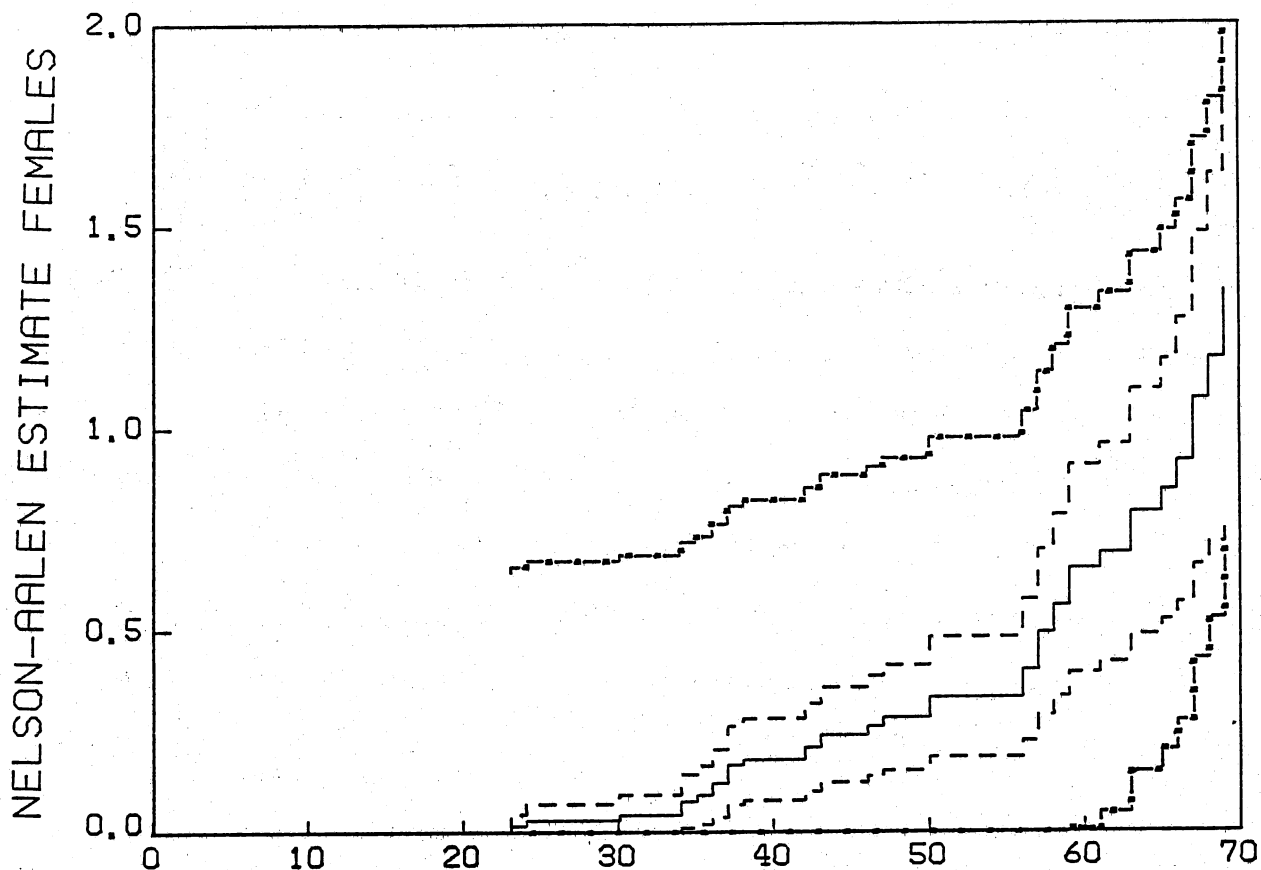


Fig. 6. Nelson-Aalen estimate for the integrated death intensity for female diabetics (upper figure) and for male diabetics (lower figure); 95% pointwise confidence limits (---) and 95% confidence bands (-·-·-) computed according to (4.7) are indicated.

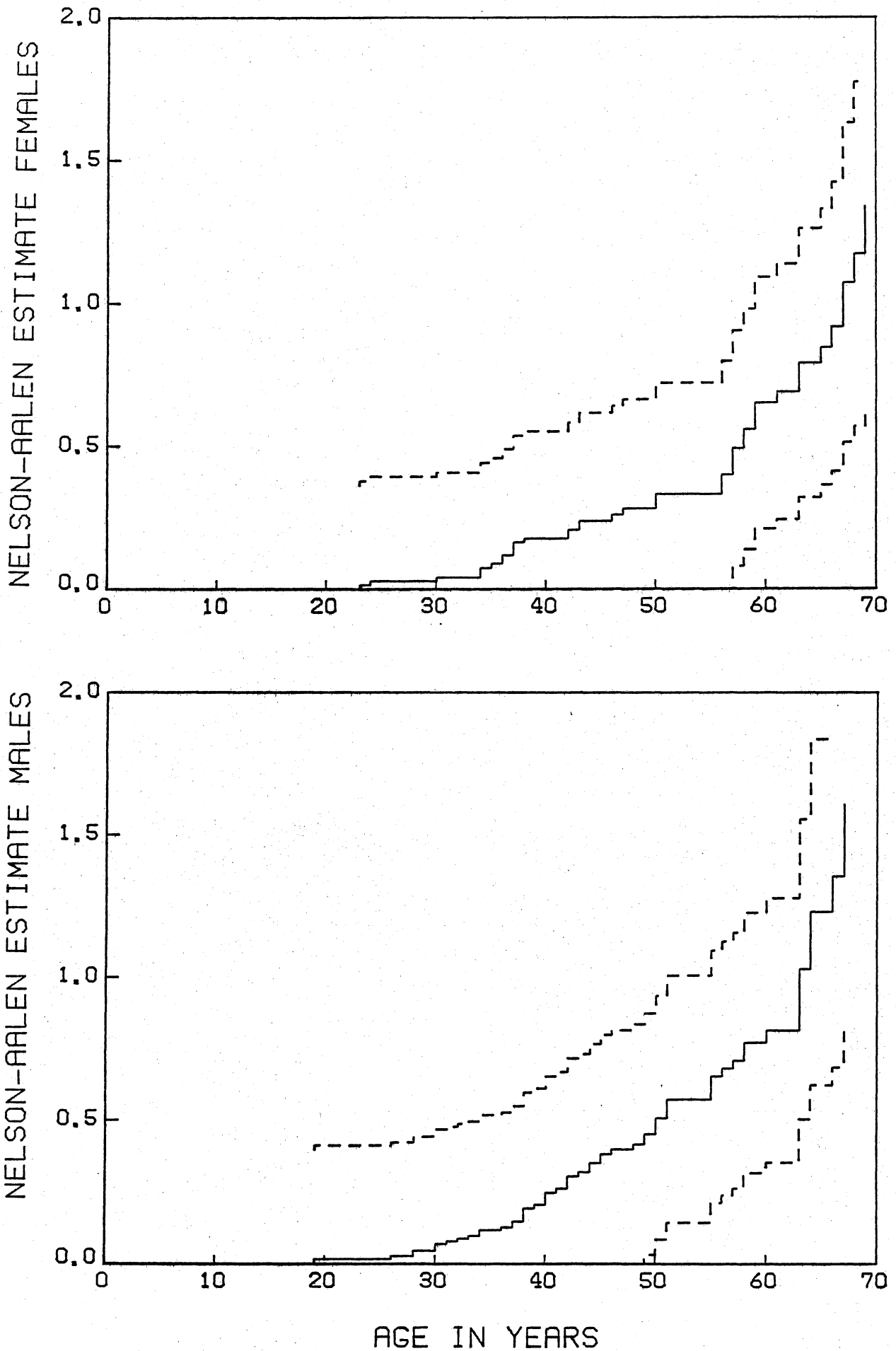


Fig. 7. Nelson-Aalen estimate for the integrated death intensity for female diabetics (upper figure) and for male diabetics (lower figure); 95% confidence bands computed according to (4.8) are indicated.

4.2 Kernel function smoothing

The Nelson-Aalen estimators (4.2) are estimators for the integrated α_h s given by (4.1). However, as seen in Subsection 3.1, it is the α_h s themselves which are the entities of real interest. Therefore, when studying plots of the Nelson-Aalen estimator, one mainly focuses on the slope of the curves. Hence it is useful to directly estimate the α_h s. Inspired by works on kernel function estimation of density functions (for a review, see Bean and Tsokos, 1980), Ramlau-Hansen (1983 a,b) proposed and studied nonparametric estimators for the α_h s in the multiplicative intensity model. Basically, the estimators are derived by smoothing the increments of the Nelson-Aalen estimators. Let us review the main results derived by Ramlau-Hansen. Since the Nelson-Aalen estimators are uncorrelated, we will in this subsection restrict our attention to one component of the multivariate counting process. We will therefore omit the subscript h in the notation.

As an estimator for $\alpha(t)$, Ramlau-Hansen (1983a) proposed

$$\hat{\alpha}(t) = \frac{1}{b} \int_0^1 K\left(\frac{t-s}{b}\right) d\hat{A}(s). \quad (4.9)$$

Here the kernel function K is a bounded function which is zero outside $[-1,1]$ and has integral 1. The window b is a positive parameter. The kernel function and the window have to be chosen in concrete applications. One frequently used kernel function is the Epanechnikov's kernel function $K(x) = 0.75(1-x^2)$, $|x| < 1$.

If we let $T_1 < T_2 < \dots$ denote the successive jump times of N , then \hat{A} may be given as in (4.3), and it follows that (4.9) may be written equivalently as

$$\hat{\alpha}(t) = \frac{1}{b} \sum_{T_j} K\left(\frac{t-T_j}{b}\right) \frac{1}{Y(T_j)}. \quad (4.10)$$

It should be realized that, since K vanishes outside $[-1,1]$, only values of j for which $t-b < T_j < t+b$ contribute to this sum. Given a window b , we will only discuss estimation of $\alpha(t)$ for $t \in [b, 1-b]$, since it is only for such values of t that (4.10) is a real average of the increments $1/Y(T_j)$ of \hat{A} . The remaining tail problem can be attacked in a similar manner (see Ramlau-Hansen, 1981).

Let us comment briefly upon the connection between (4.10) and the smoothing of occurrence/exposure rates by moving averages (cf. Borgan, 1979). We consider the set-up discussed just below (4.3), with $0 = t_0 < t_1 < \dots < t_m = 1$ being a fine partitioning of $[0,1]$, α_j the (almost) constant value of $\alpha(\cdot)$ on $(t_j, t_{j+1}]$, and $\hat{\alpha}_j$ the occurrence/exposure rate for α_j . Then, by a moving average, one would estimate α_j by $\sum_{v=-a}^a r_v \hat{\alpha}_{j-v}$ for some weights $\{r_v\}$. This equals (4.10) approximately, if we let $b = a\Delta$ and $r_v = \frac{\Delta}{b} K(\frac{-v\Delta}{b})$. The close connection between the theory of moving averages and kernel function smoothing by means of (4.9) is discussed by Ramlau-Hansen (1983b)

To study the properties of $\hat{\alpha}(t)$ we introduce the quantity (cf. (4.4))

$$\alpha^*(t) = \frac{1}{b} \int_0^1 K\left(\frac{t-s}{b}\right) dA^*(s) \quad (4.11)$$

and note that by (4.5)

$$\hat{\alpha}(t) - \alpha^*(t) = \frac{1}{b} \int_0^1 K\left(\frac{t-s}{b}\right) \frac{J(s)}{Y(s)} dM(s). \quad (4.12)$$

Thus, $\alpha(t) - \alpha^*(t)$ is a stochastic integral w.r.t. the local square integrable martingale M , a fact which provides the basis for studying the statistical properties of (4.9). We immediately get

for $t \in [b, 1-b]$ that $E\hat{\alpha}(t) = E\alpha^*(t)$, which means that the expected value of $\hat{\alpha}(t)$ equals

$$\frac{1}{b} \int_0^1 K\left(\frac{t-s}{b}\right) \alpha(s) ds \quad (4.13)$$

approximately when there is only a small probability that $Y(t)=0$ for some $t \in [0,1]$. Thus, in general, the kernel estimator is not even approximately unbiased for $\alpha(t)$ (however, cf. below). The variance of (4.9) may be estimated (almost unbiasedly) by

$$\hat{\sigma}^2(t) = \frac{1}{b^2} \int_0^1 K^2\left(\frac{t-s}{b}\right) \frac{J(s)}{Y^2(s)} dN(s) \quad (4.14)$$

(Ramlau-Hansen, 1983a, Proposition 3.2.1).

If Y increases uniformly in a neighbourhood of t , and at the same time the window tends to zero, then, subject to some regularity conditions, $\hat{\alpha}(t)$ is a consistent estimator for $\alpha(t)$. Moreover, it is asymptotically normally distributed with mean $\alpha(t)$ and a variance which may be estimated by (4.14) (Ramlau-Hansen, 1983a, Proposition 4.1.1 and Theorem 4.2.2). Finally, $\hat{\alpha}(s)$ and $\hat{\alpha}(t)$ are asymptotically independent when $s \neq t$.

To apply the kernel function estimator (4.9) in practice, one has to decide upon a choice for the kernel function and the window. Some guidelines to the choice of K are given by the results derived by Ramlau-Hansen (1983b). He argues, much the way one reasons in moving average theory (cf. Borgan, 1979), that one should choose a kernel function such that $\hat{\alpha}(t)$ is almost unbiased, i.e. such that (4.13) is approximately equal to $\alpha(t)$. This is possible if $\alpha(s)$ may be approximated by a polynomial of a certain degree over each interval of the form $[t-b, t+b]$. Subject to such an unbiasedness condition, one then chooses the kernel function which minimizes a specified risk function. Kernel

functions which are optimal in this sense are given by Ramlau-Hansen (1983b). Among other things, he shows that Epanechnikov's kernel function minimizes the variance of the first derivative of (4.9) when $\alpha(s)$ may be approximated by a linear function over each interval of the form $[t-b, t+b]$. The choice of the window b seems to be much a question of trial and error. Some guidelines are, however, given by Rudemo (1982).

Example (continued)

The kernel function smoothing method outlined above was used to obtain estimates for the age specific forces of mortality $\alpha_1(t)$ for female diabetics and $\alpha_2(t)$ for male diabetics in the example of Subsection 4.1. Fig. 8 shows the estimates $\hat{\alpha}_1(t)$ and $\hat{\alpha}_2(t)$ together with approximate 95% pointwise confidence intervals for $\alpha(t)$ (using the approximate normality of $\hat{\alpha}(t)$). In the estimation Epanechnikov's kernel function was used, and a window $b = 5$ years was chosen for the age interval $[24 \text{ years}, 66 \text{ years}]$. For ages t outside this interval, the largest window $b = b(t)$ such that $[t-b(t), t+b(t)] \subseteq [20 \text{ years}, 70 \text{ years}]$ was chosen.

From Figure 8 the level of the mortality is clearly seen. For ages less than about 55 years the mortality is close to 2% per year for both sexes, with a tendency to a lower mortality for females. From age 55 the death intensity increases for both men and women, and at age 65 the level of the mortality is about 10% per year. The fluctuations for ages above 65 years are due to less smoothing because of the narrow window used for these ages.

In the next section we shall see how a comparison of the mortality for men and women can be carried out, and also how the mortality among diabetics can be compared with that of the general Danish population. \square

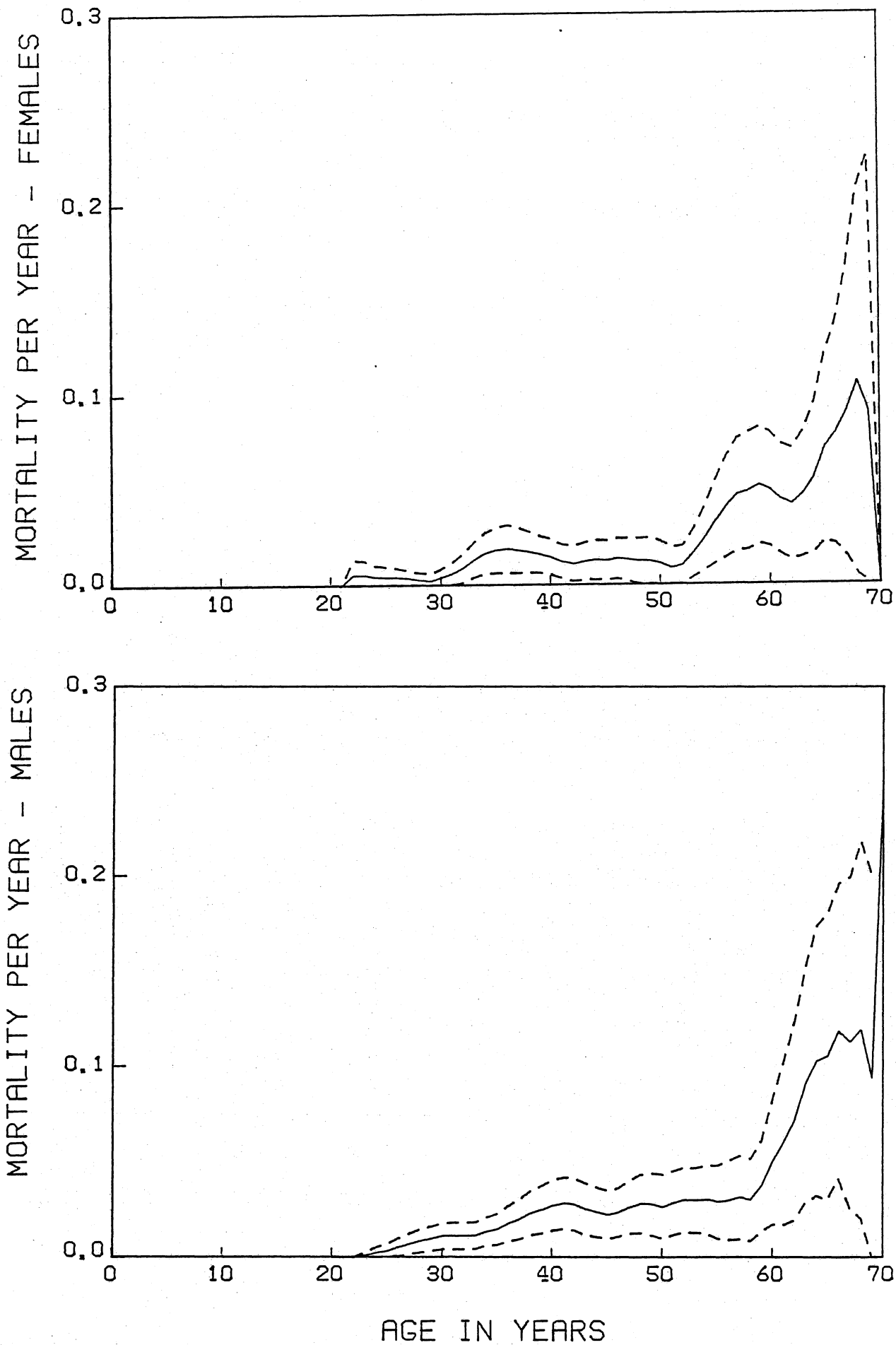


Fig. 8. Estimated absolute mortality per year for female diabetics (upper figure) and for male diabetics (lower figure); 95% pointwise confidence limits are indicated.

5. NONPARAMETRIC TESTING

This section is concerned with nonparametric testing for the multiplicative intensity model. In Subsection 5.1 we discuss how one may test whether one or more of the α_h s in (3.9) equal certain known functions, while we in Subsection 5.2 show how testing of the hypothesis that all α_h s are identical may be carried out. The procedures are illustrated by examining whether the mortality among the diabetes patients coincide with that of the general Danish population, and by a test for equality of the survival of male and female diabetics.

5.1 Tests of completely specified hypotheses

Let $\tilde{N} = (N_1, N_2, \dots, N_k)$ be a multivariate counting process satisfying the multiplicative intensity model (3.9). We want to derive tests for the hypotheses

$$H_0: \alpha_h = \alpha_h^0 \quad (5.1)$$

and

$$H_0: \alpha_1 = \alpha_1^0, \alpha_2 = \alpha_2^0, \dots, \alpha_k = \alpha_k^0 \quad (5.2)$$

where the α_h^0 s are known functions.

Let us first consider testing of the hypothesis (5.1). Andersen et al. (1982) studied a class of test statistics for this problem. Their approach was as follows. Introduce

$$A_h^0(t) = \int_0^t J_h(s) \alpha_h^0(s) ds, \quad (5.3)$$

and note that, under the hypothesis, (5.3) equals $A_h^*(t)$ defined by (4.4).

Therefore, when (5.1) holds true, we have

$$\hat{A}_h(t) - \hat{A}_h^0(t) = \int_0^t \frac{J_h(s)}{Y_h(s)} dM_h(s), \quad (5.4)$$

so that, except for random variations, \hat{A}_h and \hat{A}_h^0 are equal under the hypothesis. It is therefore natural to base a test for (5.1) on a comparison of these quantities. To do this, introduce a locally bounded predictable "weight process" K_h , and define the stochastic process

$$R_h(t) = \int_0^t K_h(s) d(\hat{A}_h - \hat{A}_h^0)(s). \quad (5.5)$$

When (5.1) holds true this is a stochastic integral (cf. (5.4)), and hence a mean zero local square integrable martingale.

By an application of the martingale central limit theorem (Theorem 3.2) it follows that R_h (properly normalized) converges weakly as Y_h increases to a Gaussian martingale when the hypothesis is true (Andersen et al., 1982, Theorem 4.1). In particular $R_h(1)$ is asymptotically normally distributed with mean zero and a variance that may be estimated by

$$\langle R_h \rangle(1) = \int_0^1 K_h^2(s) \frac{J_h(s)}{Y_h(s)} \alpha_h^0(s) ds. \quad (5.6)$$

Thus

$$U_h = R_h(1) / [\langle R_h \rangle(1)]^{1/2} \quad (5.7)$$

is an asymptotically standard normally distributed statistic for testing the hypothesis (5.1). Alternatively one may instead of (5.6) use the variance estimate

$$[R_h](1) = \int_0^1 K_h^2(s) \frac{J_h(s)}{Y_h^2(s)} dN_h(s), \quad (5.8)$$

obtained by replacing $\alpha_h^0(s)$ by the Nelson-Aalen estimate

$dN_h(s)/Y_h(s)$.

We note that (5.6) is an estimator of the variance of $R_h(1)$ when the hypothesis (5.1) is true, while (5.8) is valid in general. Thus if the alternative to the hypothesis (5.1) is $\alpha_h(t) > \alpha_h^0(t)$, with strict inequality for some t , then (5.8) will tend to be greater than (5.6). This implies that (5.6) is the best variance estimator to use for testing purposes in such situations.

By choosing different weight processes K_h , we get a number of possible test statistics of the form (5.7). It was shown by Andersen et al. (1982) how several one sample tests suggested for the survival data situation (cf. Subsection 2.1) are special cases of (5.7).

For the choice of weight process $K_h = Y_h$, the test statistic (5.7) reduces to

$$U_h = \frac{O_h - E_h}{\sqrt{E_h}}, \quad (5.9)$$

where

$$O_h = \int_0^1 J_h(s) dN_h(s) = N_h(1)$$

is the total observed number of type h events, while

$$E_h = \int_0^1 J_h(s) Y_h(s) \alpha_h^0(s) ds = \int_0^1 \alpha_h^0(s) Y_h(s) ds,$$

by (3.9) and (3.14), is the "expected" number of this type of events under the hypothesis. The choice $K_h = Y_h$ corresponds in the survival data situation to the one sample logrank test (Breslow, 1975; Hyde, 1977), which in this case is known to be optimal against proportional hazards alternatives

$$\alpha_h(s) = \theta_h \alpha_h^0(s). \quad (5.10)$$

Also in our more general set-up the model (5.10) is often approp-

riate, and when H_0 is rejected it is of interest to estimate the parameter θ_h . Using the results of our Section 6, the maximum likelihood estimator of θ_h is found to be

$$\hat{\theta}_h = \frac{O_h}{E_h} . \quad (5.11)$$

Moreover, $\hat{\theta}_h$ is asymptotically normally distributed with mean θ_h and a variance that can be estimated by $\hat{\theta}_h/E_h = V_h$, say. In particular the test statistic (5.9) equals $(\hat{\theta}_h - 1)V_{0h}^{-1/2}$, where $V_{0h} = 1/E_h$ is the estimated variance of $\hat{\theta}_h$ under the hypothesis $\theta_h = 1$. Note that in the survival data situation (5.11) is simply the well known standardized mortality ratio.

It is possible to use the convergence of the process R_h (properly normalized) to a Gaussian martingale to derive Kolmogorov-Smirnov type tests for (5.1), cf. Aalen (1976), Fleming et al. (1980), and our Subsection 4.1. We will not discuss the details here, however.

To test the hypothesis (5.2), we use the fact that the local martingales given by (5.4) for $h = 1, 2, \dots, k$ are orthogonal (cf. Subsections 3.2 and 3.3). This implies that the U_h given by (5.7) are asymptotically independent. Thus we may use the statistic

$$\chi^2 = \sum_{h=1}^k U_h^2 \quad (5.12)$$

for testing the hypothesis (5.2). Under the hypothesis, (5.12) is asymptotically chi-squared distributed with k degrees of freedom. For $K_h = Y_h$; $h = 1, \dots, k$; (5.12) gives a statistic of the well known form

$$\chi^2 = \sum_{h=1}^k \frac{(O_h - E_h)^2}{E_h} .$$

Example. Survival among diabetics

Among the 314 female diabetics, $O_1 = 39$ were observed to die in the period in question, and among the 413 males we found $O_2 = 65$. To investigate whether these figures are larger than expected, judged from published life-tables for the general Danish population, the test statistic (5.9) was calculated for females and males separately. The "expected" number of females dying was found to be $E_1 = 6.34$, while for males we found $E_2 = 13.86$. For both sexes the test statistic (5.9) is highly significant in that $U_1 = 12.97$ and $U_2 = 13.74$. Assuming a proportional hazards model (5.10) (in Section 7 we shall return to a discussion of this assumption), we find $\hat{\theta}_1 = O_1/E_1 = 6.15$ and $\hat{\theta}_2 = O_2/E_2 = 4.69$ indicating higher excess mortality among females than among males. Using the variance estimates $V_h = \hat{\theta}_h/E_h$, confidence intervals for θ_h can be constructed, e.g. by transforming a symmetric confidence interval for $\log \theta_h$. Thus we find approximate 95% confidence intervals $[2.82, 13.39]$ and $[2.77, 7.93]$ for θ_1 and θ_2 , respectively. \square

5.2 The k-sample problem

As in the preceding subsection, we consider a k-variate counting process \tilde{N} satisfying the multiplicative intensity model (3.9). We want to derive a test for the hypothesis

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k. \quad (5.13)$$

The common value of the α_h s will be denoted α .

Following Aalen (1978), who considered the two-sample problem, Andersen et al. (1982) introduced a class of statistics for testing the hypothesis (5.13). Their idea was to construct a test statistic by comparing the Nelson-Aalen estimators $\hat{A}_h(t)$, cf.

(4.2), with an estimator of the hypothesized common value

$A(t) = \int_0^t \alpha(s) ds$. This latter quantity can be estimated by

$$\hat{A}(t) = \int_0^t \frac{J(s)}{Y.(s)} dN.(s),$$

where $N. = \sum_{h=1}^k N_h$, $Y. = \sum_{h=1}^k Y_h$ and $J(t) = I(Y.(t) > 0)$. This follows

as in Subsection 4.1, since under the hypothesis, $N.(t)$ is a (univariate) counting process with intensity process $\alpha(t)Y.(t)$.

We introduce

$$\tilde{A}_h(t) = \int_0^t J_h(s) d\hat{A}(s) = \int_0^t \frac{J_h(s)}{Y.(s)} dN.(s),$$

and note that, when (5.13) holds true, we have

$$\hat{A}_h(t) - \tilde{A}_h(t) = \int_0^t \frac{J_h(s)}{Y.(s)} dM_h(s) - \int_0^t \frac{J_h(s)}{Y.(s)} dM.(s), \quad (5.14)$$

where $M. = \sum_{h=1}^k M_h$. Thus, except for random variations, \hat{A}_h and

\tilde{A}_h are equal under the hypothesis. In a similar manner as in Subsection 5.1, we introduce locally bounded predictable weight processes K_h , and define stochastic processes

$$Z_h(t) = \int_0^t K_h(s) d(\hat{A}_h - \tilde{A}_h)(s); \quad h = 1, \dots, k. \quad (5.15)$$

When (5.13) holds true, (5.14) yields that the Z_h s are linear combinations of stochastic integrals. Especially $EZ_h(t) = 0$ for all h and $t \in [0, 1]$.

It turns out that the special choice of weight processes

$$K_h(t) = Y_h(t) L(t); \quad h = 1, \dots, k; \quad (5.16)$$

where L is a locally bounded predictable process that only depends on $(N., Y.)$, covers most examples of interest. Then (5.15) may be written

$$Z_h(t) = \int_0^t L(s) dN_h(s) - \int_0^t L(s) \frac{Y_h(s)}{Y.(s)} dN.(s); \quad h=1, \dots, k. \quad (5.17)$$

We note that $\sum_{h=1}^k Z_h = 0$. By an application of the martingale central limit theorem (Theorem 3.2), it follows that the Z_h s given by (5.17) (properly normalized) converge weakly to a k -variate Gaussian martingale under the hypothesis, as the Y_h s increase (Andersen et al. 1982, Theorem 3.1, see also Helland, 1983, Section 4). Especially $\underline{z}(1) = (Z_1(1), \dots, Z_k(1))'$ is asymptotically multivariate normally distributed with mean zero and a (singular) covariance matrix that can be estimated by $\underline{v}(1) = \{v_{hj}(1)\}$, where

$$v_{hj}(1) = \int_0^1 L^2(s) \frac{Y_h(s)}{Y.(s)} \left(\delta_{hj} - \frac{Y_j(s)}{Y.(s)} \right) dN.(s), \quad (5.18)$$

and δ_{hj} is a Kronecker delta.

Thus, under the hypothesis (5.13), the statistic

$$\chi^2 = \underline{z}(1)' \underline{v}(1)^{-} \underline{z}(1), \quad (5.19)$$

where $\underline{v}(1)^{-}$ is a generalized inverse, is asymptotically chi-squared distributed with $k-1$ degrees of freedom (Andersen et al., 1982, Section 3.1). Note that (5.19) may be computed by deleting the last component of $\underline{z}(1)$ and the last row and column of $\underline{v}(1)$, to give $\underline{z}_0(1)$ and $\underline{v}_0(1)$ say, and then using the relation $\underline{z}(1)' \underline{v}(1)^{-} \underline{z}(1) = \underline{z}_0(1)' \underline{v}_0(1)^{-1} \underline{z}_0(1)$. Also, as in Section 4 the integrals (5.17) and (5.18) are finite sums.

Andersen et al. (1982, Section 3.2-3.4) studied various conservative approximations to the test statistic (5.19), and discussed how it generalizes the classical nonparametric tests as well as their generalizations to censored data. For example the choice $L(t) = I(Y.(t) > 0)$ corresponds to the logrank (or Savage) test (Peto and Peto, 1972), while $L(t) = Y.(t)$ gives a generalization of the Kruskal-Wallis test (Breslow, 1970). Also the tests suggested by Tarone and Ware (1977), Prentice (1978) and Harrington

and Fleming (1982) are special cases of (5.19).

For the two-sample problem, i.e. $k=2$, Aalen (1975, Theorem 9.1) showed that the logrank test is the asymptotically optimal similar test against local (or more precisely, contiguous) alternatives where α_1 and α_2 are proportional. Gill (1980a) studied asymptotic relative efficiencies between various tests for the two-sample problem, and showed how one can derive tests with optimality properties against specified alternatives.

Following the lines of the preceding subsection, let us now study the two-sample model with proportional intensities

$$\alpha_2(t) = \theta \alpha_1(t) \quad (5.20)$$

in more detail. When $k=2$ the Aalen (1978) test statistic based on the process

$$Z(t) = \int_0^t K(s) (d\hat{A}_2(s) - d\hat{A}_1(s)) \quad (5.21)$$

is equivalent to the general k -sample test statistic (5.17). (With $K(t) = Y_1(t)Y_2(t)L(t)/Y.(t)$, (5.21) is seen to equal $Z_2(t)$ given by (5.17).) This two-sample test takes the form $Z(1)V(1)^{-\frac{1}{2}}$ with V given by

$$V(t) = \int_0^t K^2(s) \frac{dN.(s)}{Y_1(s)Y_2(s)}$$

(i.e. $V(t)$ equals $V_{22}(t)$ given by (5.18) with the above mentioned choice of $K(t)$). The optimal test under the model (5.20), the logrank test, corresponds to the weight process

$$K_L(t) = Y_1(t) Y_2(t) / Y.(t).$$

Under the proportional intensity model it is of interest to estimate the intensity ratio θ , in particular if H_0 is rejected. A class of consistent and asymptotically normally distributed estimators of θ is given by

$$\hat{\theta}_K = \frac{\int_0^1 K(t) d\hat{A}_2(t)}{\int_0^1 K(t) d\hat{A}_1(t)} \quad (5.22)$$

(Andersen, 1983a), and in this class, the "log-rank estimator" $\hat{\theta}_{K_L}$ has smallest variance under the null hypothesis $\theta=1$. The variance of $\hat{\theta}_K$ can be estimated by

$$\hat{\sigma}_K^2 = \hat{\theta}_K^2 \frac{\int_0^1 K^2(t) \frac{dN_{\cdot}(t)}{Y_1(t) \hat{\theta}_K Y_2(t)}}{\left(\int_0^1 K(t) \frac{dN_1(t)}{Y_1(t)} \right)^2} \quad (5.23)$$

It is seen that the Aalen test statistic defined below (5.21) equals $(\hat{\theta}_K - 1) [\hat{\sigma}_{OK}^2]^{-1/2}$, where $\hat{\sigma}_{OK}^2$ is the estimated variance of $\hat{\theta}_K$ under the hypothesis $\theta = 1$, obtained by substituting 1 for $\hat{\theta}_K$ in (5.23).

The appropriateness of the model (5.20) can be checked graphically by plotting $\hat{A}_2(t)$ against $\hat{A}_1(t)$, or by plotting $\log \hat{A}_2(t)$ and $\log \hat{A}_1(t)$ against t (or $\log t$). Under the model (5.20) the former plot should approximate a straight line through the origin (with slope $\approx \theta$) and the latter should yield approximately parallel curves (with vertical distance $\approx \log \theta$). Alternatively, test statistics for proportionality can be constructed of the form

$$U(t) = \int_0^t \tilde{K}(t) (d\tilde{A}_2(t) - \tilde{\theta} d\tilde{A}_1(t)), \quad (5.24)$$

where \tilde{A}_h is some estimate for A_h , $h=1,2$, and $\tilde{\theta}$ is an estimate for θ . Gill and Schumacher (1984) studied the case $\tilde{A}_h = \hat{A}_h$ and $\tilde{\theta} = \hat{\theta}_K$ (cf. (5.22)) and obtained when $\tilde{K}(t)/K(t)$ is increasing a test $U(1)$ consistent against the alternative that $\alpha_2(t)/\alpha_1(t)$ is monotone. Wei (1983) showed that with $\tilde{K}(t) = Y_2(t)$, $\tilde{A}_2(t) = \hat{A}_2(t)$ and $\tilde{\theta}$ and $\tilde{A}_1(t)$ based on the Cox regression model (cf.

Section 7 below), the process $U(t)$ is distributed asymptotically as a time-transformed Brownian bridge. Finally Andersen (1983 b) showed that for $\tilde{A}_h = \hat{A}_h$, $h = 1, 2$, and $\tilde{\theta} = \hat{\theta}_{K_T}$ with $K_T(t) = Y_1(t)Y_2(t)/(Y_1(t) + \tilde{\theta}Y_2(t))$ a test equivalent to that of Wei (1983) is obtained, where $\tilde{\theta}$ is any consistent estimator for θ . In the case of survival data the estimator $\hat{\theta}_{K_T}$ is the two step estimator of Begun and Reid (1983).

The test statistic (5.19) is based only on the weak convergence of (5.17) for $t=1$. However, we have weak convergence of the entire processes given by (5.17). For the two-sample problem this may be used to derive Kolmogorov-Smirnov type statistics suitable for testing against "crossing intensities alternatives" (e.g. Fleming et al., 1980; Gill, 1980a; Fleming and Harrington, 1981; cf. also Subsection 4.1 above).

Example

As an example of the applicability of the testing procedures discussed in this subsection, let us examine whether male and female diabetics have identical mortalities. The figures in Section 4 indicate that males have a slightly higher mortality than females. The two-sample logrank test for the hypothesis $\alpha_2(t) = \alpha_1(t)$ takes the value 1.62 corresponding to a two-sided p-value of 0.11. The visual impression from the figures is confirmed by the logrank estimator $\hat{\theta}_{K_L} = 1.39$, assuming proportionality $\alpha_2(t) = \theta\alpha_1(t)$. The estimated standard error of $\hat{\theta}_{K_L}$ is 0.24 yielding an approximate 95% confidence interval [0.87, 2.22] for θ (by transforming a symmetric confidence interval for $\log \theta$). Thus the analyses in this section indicate that male diabetics have a higher absolute mortality than female diabetics, but compared to the mortality in the general Danish population female diabetics seem to have a higher excess mortality than males. \square

6. PARAMETRIC MODELS

In the two preceding sections we have considered nonparametric estimation and testing procedures for the multiplicative intensity model. Alternatively the α_h s in (3.9) can be given via a parametric specification and maximum likelihood methods can be applied. In this section we shall review the results of Borgan (1984) for this situation.

To give a motivation for the general set-up, we first consider the situation with censored failure time data. Let T_1, \dots, T_n be independent and identically distributed (true) survival times with hazard rate function $\alpha(t; \theta_0)$. Some commonly used forms for $\alpha(t; \theta_0)$ are reviewed by Kalbfleisch and Prentice (1980) and Miller (1981). We do not observe the T_i s, but only censored survival times \tilde{T}_i and indicators $D_i = I(\tilde{T}_i = T_i)$; $i = 1, \dots, n$ (cf. Example 1, Subsection 3.1). Then for a very broad class of censoring mechanisms (the important part of) the likelihood is

$$L(\theta) = \prod_{i=1}^n \left\{ \alpha(\tilde{T}_i; \theta)^{D_i} \exp\left(-\int_0^{\tilde{T}_i} \alpha(s; \theta) ds\right) \right\}$$

(e.g. Kalbfleisch and Prentice, 1980, Section 5.2).

With $N(t)$ defined as in (3.2) and (3.6), and $Y(t)$ defined as in (3.3) and (3.8), we may write

$$\log L(\theta) = \int_0^1 \log(\alpha(s; \theta)) dN(s) - \int_0^1 \alpha(s; \theta) Y(s) ds, \quad (6.1)$$

and the maximum likelihood estimator $\hat{\theta}$ is defined as a solution to the set of equations $\partial \log L(\theta) / \partial \theta = 0$.

Turning to the general formulation, we let $\tilde{N} = (N_1, \dots, N_k)$ be a multivariate counting process satisfying the multiplicative intensity model (3.9) with parametric α_h s, i.e. the intensity process is given by

$$\lambda_h(t) = \alpha_h(t; \underline{\theta}_0) Y_h(t); \quad h = 1, \dots, k; \quad (6.2)$$

where $\underline{\theta}_0 = (\theta_{10}, \dots, \theta_{q0})'$ belongs to some open subset θ of \mathbb{R}^q . Under some regularity conditions (e.g. Aalen, 1978, Section 3.3), the log-likelihood function now takes the form (cf. (6.1))

$$\log L(\underline{\theta}) = \sum_{h=1}^k \int_0^1 \log(\alpha_h(s; \underline{\theta})) dN_h(s) - \sum_{h=1}^k \int_0^1 \alpha_h(s; \underline{\theta}) Y_h(s) ds, \quad (6.3)$$

and the maximum likelihood estimator $\hat{\underline{\theta}}$ is defined as a solution to the set of equations

$$\sum_{h=1}^k \int_0^1 \frac{\frac{\partial}{\partial \theta_j} \alpha_h(s; \underline{\theta})}{\alpha_h(s; \underline{\theta})} dN_h(s) - \sum_{h=1}^k \int_0^1 \frac{\partial}{\partial \theta_j} \alpha_h(s; \underline{\theta}) Y_h(s) ds = 0; \quad (6.4)$$

$j = 1, \dots, q.$

Borgan (1984, Theorems 1 and 2) shows that, under certain regularity conditions on the α_h s, the likelihood equations (6.4) have, with a probability tending to one, exactly one consistent solution $\hat{\underline{\theta}}$ as the Y_h s increase. Moreover, $\hat{\underline{\theta}}$ is asymptotically multnormally distributed with mean $\underline{\theta}_0$ and a covariance matrix that may be estimated by $-\underline{I}(\hat{\underline{\theta}})^{-1}$, where $\underline{I}(\underline{\theta}) = \partial^2 \log L(\underline{\theta}) / \partial \theta^2$. Thus, the usual results for maximum likelihood estimation in the i.i.d. case continue to hold under our more general model (6.2).

The methods of proofs used in that paper are similar to the classical i.i.d. case (Cramér, 1945). But in the present context Lenglar's inequality (Theorem 3.1) is used to establish the consistency results derived by the law of large numbers in the classical set-up, while asymptotic normality is derived by the martingale central limit theorem (Theorem 3.2). These results may be used, since by (3.14) and (6.2), the left hand sides of (6.4), evaluated at the true parameter value $\underline{\theta}_0$, equal the stochastic integrals

$$\sum_{h=1}^k \int_0^1 \frac{\frac{\partial}{\partial \theta_j} \alpha_h(s; \underline{\theta}_0)}{\alpha_h(s; \theta_0)} dM_h(s); h = 1, \dots, k. \quad (6.5)$$

(We have used the notation $\frac{\partial}{\partial \theta_j} \alpha_h(s; \underline{\theta}_0)$ for $\frac{\partial}{\partial \theta_j} \alpha(s; \underline{\theta})|_{\underline{\theta}=\underline{\theta}_0}$.)
 To illustrate how powerful the martingale techniques are in deriving such general results, we give in the Appendix a detailed derivation of the properties of the maximum likelihood estimator for a univariate counting process when $q = 1$.

By combining the results of Borgan (1984) with the argument used to derive the properties of the likelihood ratio test for the i.i.d. case (e.g. Serfling, 1980, Section 4.4), we have that minus two times the logarithm of the likelihood ratio test statistic is asymptotically chi-squared distributed, also in our more general setting.

Example. Survival with liver cirrhosis

In Subsection 2.3 a brief introduction to the CSL-I study was given. Here some analyses of the data from that trial are reviewed.

We shall be concerned with the interaction between the treatment (prednisone or placebo) and the biochemical variable prothrombine, the value of which will be considered as either "low" or "normal". The variable was recorded according to the follow-up scheme described in Subsection 2.3. In the following analyses the assumption is made that for each patient and for each time t , the prothrombine value for this patient at t is the one that was recorded at the last follow-up preceding t .

For each of the two treatments an illness-death model as shown in Figure 3 is considered. Figure 9 shows the Nelson-Aalen plots for the integrated death intensities. It is seen that for

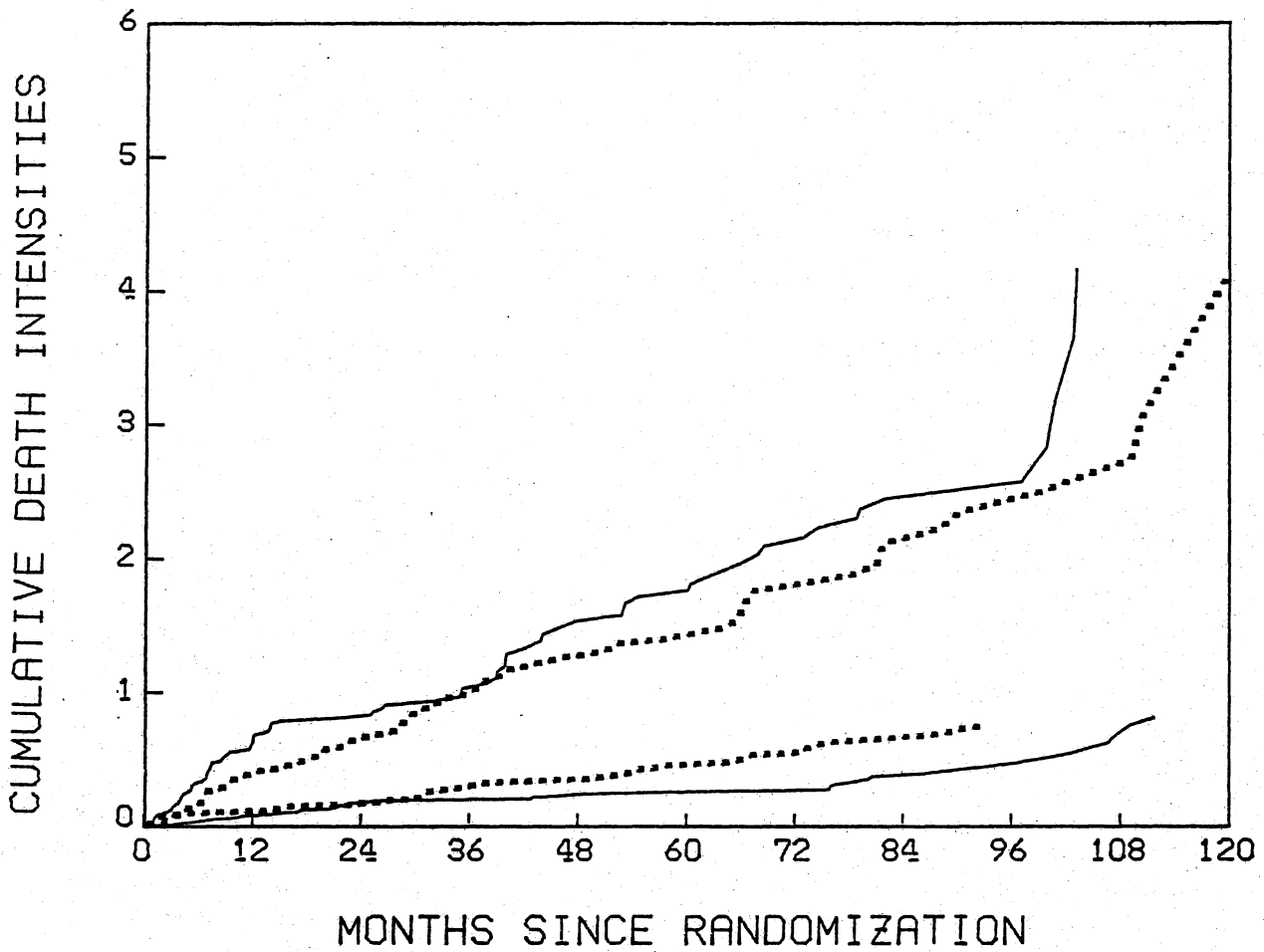


Fig. 9. Nelson-Aalen estimates for the integrated death intensities for patients with liver cirrhosis: upper pair of curves ~ low prothrombine, lower pair of curves ~ normal prothrombine.

Prednisone treatment: ———
Placebo treatment: ·····

both treatments the death intensities are higher when the prothrombine value is low. Moreover, the "difference" between the "low" and "normal" curve is smaller for the placebo treated patients than for those treated with prednisone.

Assuming proportionality between the death intensities for patients with normal and low value within each treatment group we can estimate the hazard ratio by the logrank version of (5.22). For prednisone treatment we find $\hat{\theta}_{\text{pred}} = 0.159$ and for placebo treatment $\hat{\theta}_{\text{plac}} = 0.294$. A comparison between these hazard ratios is most easily carried out by noticing that it follows from (5.23) that $\hat{\theta}_K^{1/2}$ has a variance that can be estimated by $\hat{\sigma}_{OK}^2/4$ where $\hat{\sigma}_{OK}^2$ is defined just below (5.23). Thus we find that when $\theta_{\text{plac}} = \theta_{\text{pred}}$ the statistic

$$U = 2 \frac{\hat{\theta}_{\text{plac}}^{1/2} - \hat{\theta}_{\text{pred}}^{1/2}}{(\hat{\sigma}_{0,\text{plac}}^2 + \hat{\sigma}_{0,\text{pred}}^2)^{1/2}}$$

has an approximate standard normal distribution. Inserting the estimates $\hat{\theta}_{\text{pred}}$, $\hat{\theta}_{\text{plac}}$ and $\hat{\sigma}_{0,\text{plac}} = 0.095$ and $\hat{\sigma}_{0,\text{pred}} = 0.078$ we get the significant value $U = 2.29$. Thus the indication, looking only at the death intensities, is that prednisone treatment should only be given to patients with normal prothrombine value.

Figure 10 shows the Nelson-Aalen plots for the integrated intensities for transitions from low to normal prothrombine. We see that this intensity is highest for prednisone treated patients. Figure 10 also shows the corresponding curves for transitions from normal to low prothrombine, and in this case the prednisone intensity seems to be lowest.

A comparison between the transition intensities for the two treatment groups can be performed e.g. by using the two-sample logrank test. For the transitions from normal to low value the test statistic takes the value - 3.16 corresponding to the significance probability 0.002, and for transitions from low to normal

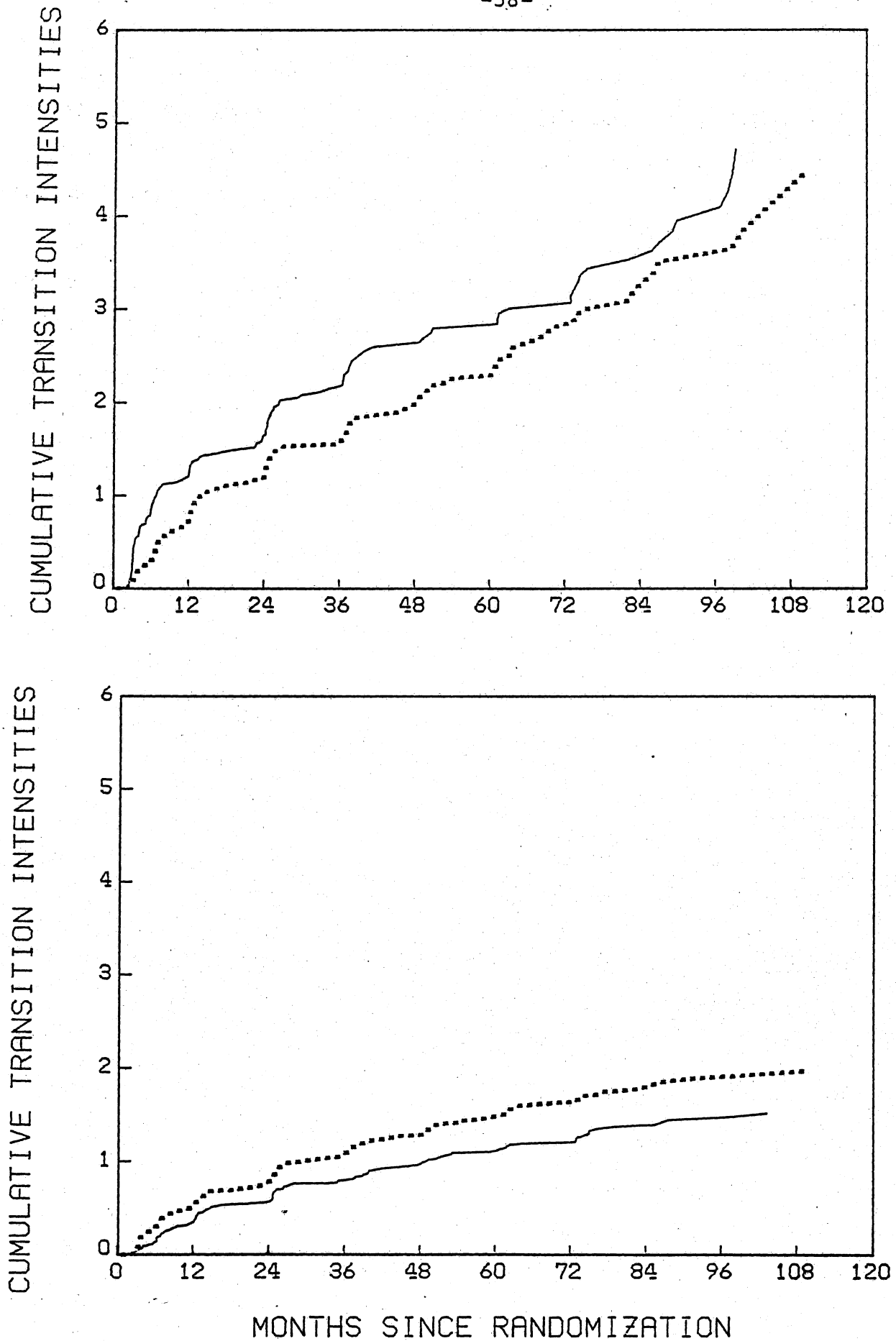


Fig. 10. Nelson-Aalen estimate for the integrated transition intensity from low to normal prothrombine (upper figure) and from normal to low prothrombine (lower figure) for patients with liver cirrhosis. Prednisone (—). Placebo (···).

the value is 2.09 ($P=0.04$). Thus both of the tendencies seen from the figures are significant. Equivalent to these values of the test statistic are the logrank estimates $\hat{\theta} = 0.72$ for the ratio between the transition intensities from normal to low value for prednisone treated and for placebo treated patients with an approximate 95% confidence interval (0.57, 0.89). For transitions from low to normal value we find $\hat{\theta} = 1.29$ with an approximate 95% confidence interval (1.01, 1.60).

So for patients with normal prothrombine value prednisone treatment seems to be beneficial in that both the death intensity and the tendency for getting a low prothrombine value is smaller for prednisone treated patients. For patients with low prothrombine value the situation is more complicated because both the death intensity and the tendency for getting a normal prothrombine value is higher during prednisone treatment. So in this case the decision whether or not to treat a patient with prednisone cannot be based solely on the estimated intensities. What is needed is an estimate for each treatment of the probability of being alive at any time t given the initial state (low or normal). These probabilities can be estimated nonparametrically following the lines of Aalen & Johansen (1978). We shall estimate these probabilities under the parametric assumption of constant transition intensities. Judged from the Figures 9 and 10 this assumption (which is equivalent to linear cumulative intensities) is not too unreasonable even though there is a tendency towards higher transition intensities shortly after start of treatment.

It is easily seen from the likelihood equations (6.4) that the estimators in this simple model are occurrence/exposure rates (see also Borgan, 1984, Section 5.1). That is, in order to estimate the

intensities, we only need to divide the total number of transitions of the various kinds by the total amount of time spent in the relevant states. The estimates are shown in Table 1.

From these estimates the same hypotheses as discussed above can be tested and the same hazard ratios can be estimated. In no case any major discrepancies from the earlier results are found. From the estimated transition intensities we can estimate the probabilities of being alive at any time t using the relations between the transition intensities and the transition probabilities in a Markov chain with constant intensities (see e.g. Tuma et al., 1979). The estimated probabilities are shown in Figure 11, and for patients having normal prothrombine value at start of treatment the survival probability (as expected) is larger during prednisone treatment. For patients having low value at start of treatment no clear picture is seen and the conclusion is that for these patients there seems to be no treatment effect. □

Table 1. Model with constant transition intensities for patients with liver cirrhosis.

Transition	Prednisone treatment		Placebo treatment	
	No. transitions	Intensity per year	No. transitions	Intensity per year
From low to normal	164	0.69	159	0.53
From normal to low	137	0.20	150	0.26
From low to death	95	0.40	99	0.33
From normal to death	47	0.070	51	0.089

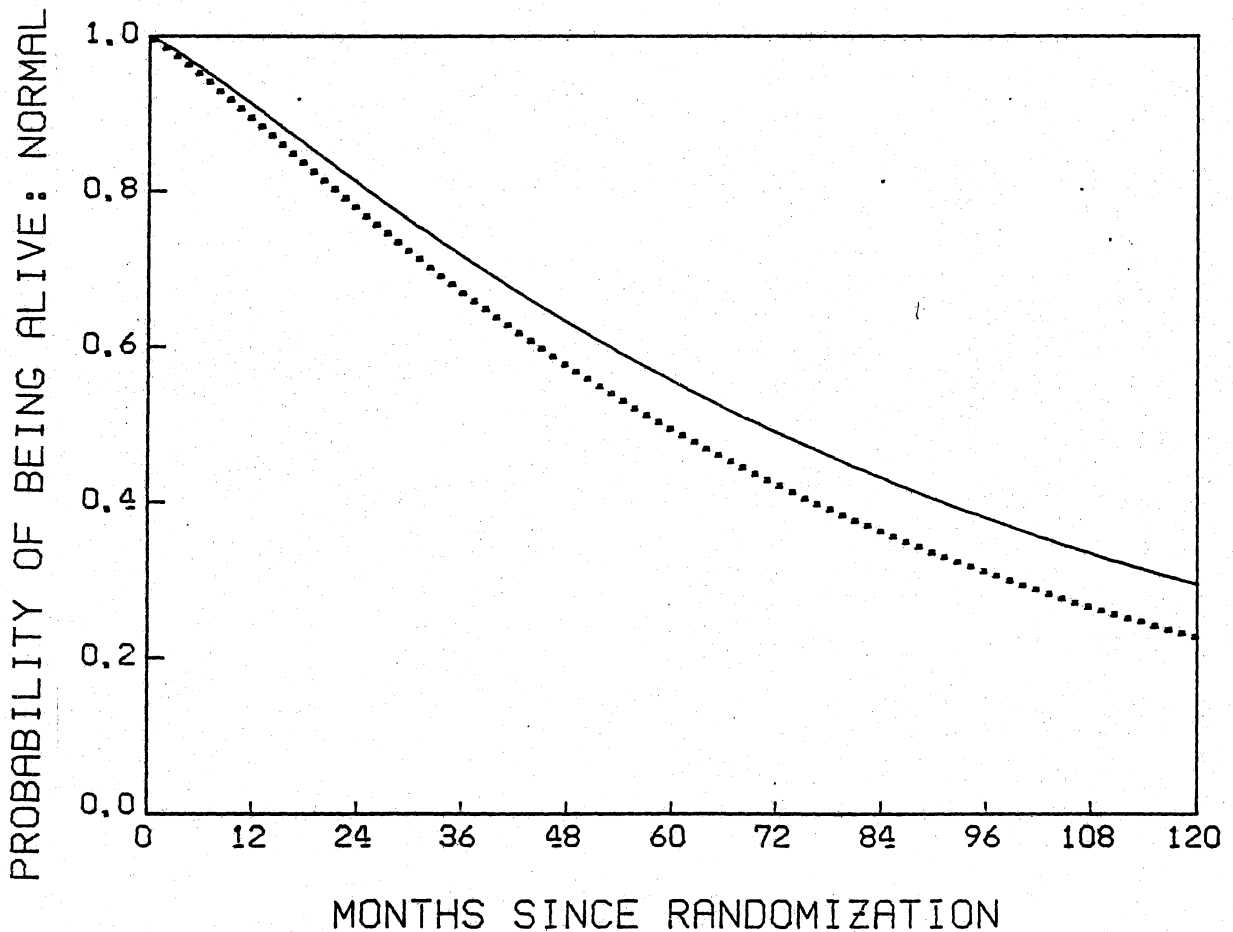
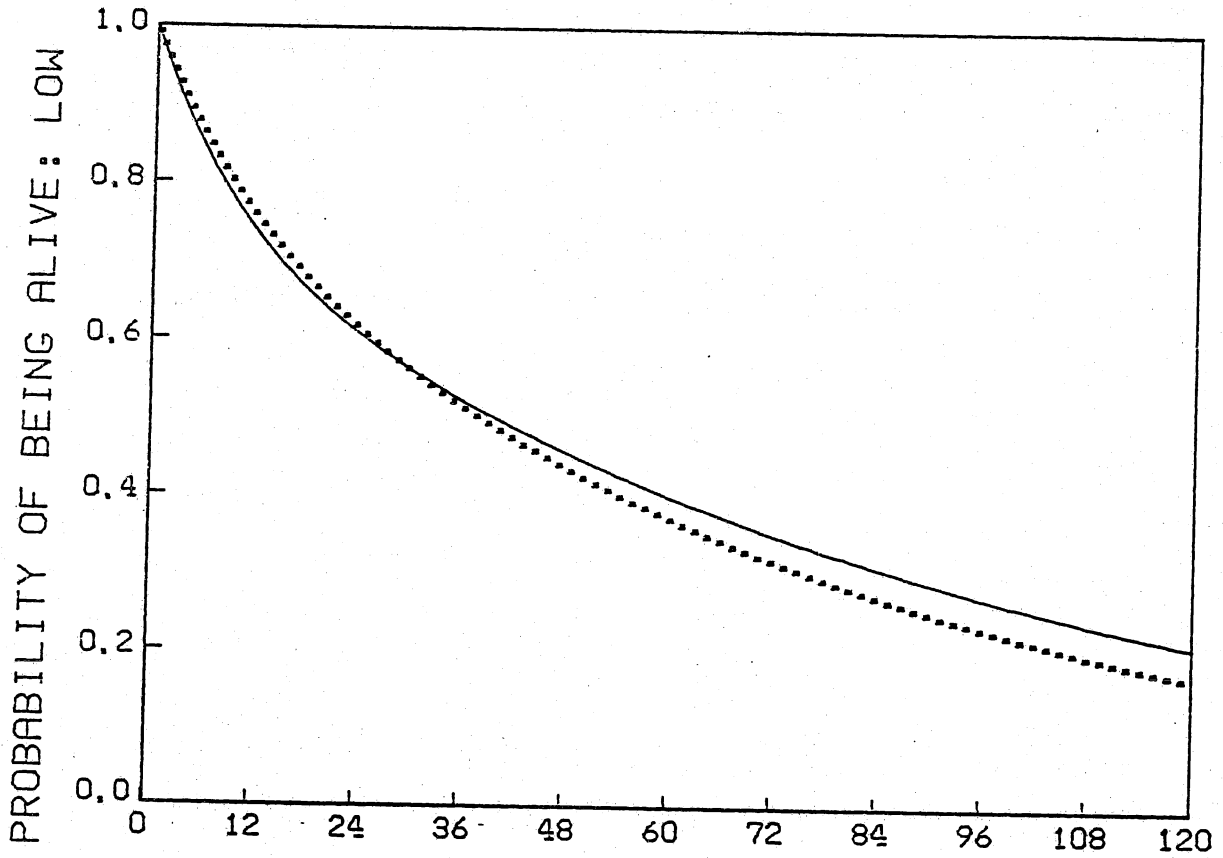


Fig. 11. Estimated probability of being alive at time t after randomization for patients with liver cirrhosis; upper figure ~ low prothrombine at time 0, lower figure ~ normal prothrombine at time 0. Prednisone (—). Placebo (···).

7. REGRESSION MODELS

In the preceding sections attention has been focused on statistical models applicable when analysing homogeneous groups of individuals. The methods of inference discussed in the Sections 4, 5 and 6 were based on processes $N_h(t)$ and $Y_h(t)$ obtained by aggregating individual processes corresponding to the type h event, cf. Subsection 3.1. In this section we shall study regression models for the individual intensity processes (3.10). In order to include individual covariates or covariate processes in the models, some assumptions have to be made about the way in which these enter the individual intensity processes. We shall mainly concentrate on extensions of the Cox regression model, that is models where the factor $\alpha_h(t, Z_i(t))$ in (3.10) factorises as

$$\alpha_h(t, Z_i(t)) = \alpha_{0h}(t)g(\beta' \tilde{z}_{hi}(t)). \quad (7.1)$$

It is convenient to let the vector of regression parameters $\beta = (\beta_1, \dots, \beta_p)'$ be the same for all types h . This can always be obtained, if necessary by introducing extra type specific covariates. Therefore, we have introduced type specific covariate vectors $\tilde{z}_{hi}(t) = (z_{hi1}(t), \dots, z_{hip}(t))'$ on the right-hand side of (7.1).

In the now classical regression model for survival data of Cox (1972) (see also Kalbfleisch and Prentice, 1980), the relative risk function $g(\cdot)$ was chosen to be the exponential function. In this presentation we shall also discuss this case in greatest detail, and only briefly mention models with a general form of the relative risk function $g(\cdot)$.

In Subsection 7.1 we discuss the counting process formulation of the Cox regression model, and study the statistical properties of the estimators (Andersen & Gill, 1982). A brief discussion of the modifications needed to allow a general relative risk function $g(\cdot)$ is also included (Prentice & Self, 1983). Parametric regression models (Borgan, 1984) are treated in Subsection 7.2, while in Subsection 7.3 an alternative, linear regression model for the intensity process (Aalen, 1980) is introduced. As an example of the application of a Cox regression model to life history data, some results from an analysis of the diabetes survival data are given.

7.1. The Cox regression model

We consider an nk -dimensional counting process $(N_{hi}(t), h = 1, \dots, k; i = 1, \dots, n), t \in [0, 1]$ where $N_{hi}(t)$ counts the number of type h events in $[0, t]$ for individual i , see Subsection 3.1. We assume that N_{hi} has intensity process of the form

$$\lambda_{hi}(t) = \alpha_{0h}(t) \exp(\beta_0' Z_{hi}(t)) Y_{hi}(t), \quad (7.2)$$

cf. (2.2), (3.10) and (7.1). Here $\alpha_{0h}, h = 1, \dots, k$ are unspecified type specific underlying intensities whose integrals

$$A_{0h}(t) = \int_0^t \alpha_{0h}(s) ds \text{ are assumed to satisfy } A_{0h}(1) < \infty, h = 1, \dots, k.$$

Furthermore $\beta_0 = (\beta_{01}, \dots, \beta_{0p})'$ is a vector of unknown regression coefficients and $Z_{hi}(t) = (Z_{hi1}(t), \dots, Z_{hip}(t))'$ a vector of predictable and locally bounded (type specific) covariate processes. Some further boundedness conditions on the covariate processes are also needed, see Andersen & Gill (1982, p. 1105 and 1110, item C).

Finally $Y_{hi}(t)$ is a predictable indicator process taking the value 1 if the i th individual is at risk at time t for experiencing a type h event; otherwise $Y_{hi}(t) = 0$.

The basic assumption in the extended Cox model (7.2) is that each covariate $Z_{hij}(t)$ has a multiplicative effect on the intensity; in particular for time-independent covariates we have a model with proportional intensities. The asymptotic properties of the Cox model for survival data have been studied by Tsiatis (1981) and Næs (1982), while the properties of the model (7.2) were studied by Andersen & Gill (1982) in the case $k = 1$, i.e. of individual univariate counting processes. In the following we shall formulate the corresponding results valid also when $k > 1$. The proofs of Andersen and Gill (1982) go through almost unchanged for our more general model. The estimator $\hat{\beta}$ for β_0 is defined as the solution to the equations $(\partial/\partial\beta_j)C(\beta, l) = 0$, $j = 1, \dots, p$, where

$$C(\beta, t) = \sum_{h=1}^k \left[\sum_{i=1}^n \int_0^t \beta' Z_{hi}(s) dN_{hi}(s) - \int_0^t \log \left\{ \sum_{i=1}^n Y_{hi}(s) \exp(\beta' Z_{hi}(s)) \right\} dN_h(s) \right], \quad (7.3)$$

and $N_h = N_{h1} + \dots + N_{hn}$. The process (7.3) can be regarded as the logarithm of a generalized Cox's partial likelihood function (Cox, 1972, 1975; Johansen, 1983). We shall not, however, use the interpretation of $\hat{\beta}$ as a maximum partial likelihood estimator in the following; only the fact that it is a solution to $(\partial/\partial\beta)C(\beta, l) = 0$. Hence we shall neither go into a discussion of the extended model of Johansen (1983), in which $\hat{\beta}$ (and the estimate $\hat{A}_{0h}(t)$ for $A_{0h}(t)$ given below, see (7.8)) is the maximum likelihood estimator, nor of that of Jacobsen (1984) in which a maximum likelihood estimator differing from $\hat{\beta}$ is obtained.

The key step in the derivation of the statistical properties of $\hat{\beta}$ is to notice the fact that, evaluated at the true value β_0 of the parameter vector, the derivatives $U(\beta, t) = (\partial/\partial\beta)C(\beta, t)$ of (7.3) are local square integrable martingales. To see this we introduce the predictable processes

$$\begin{aligned} S_h^{(0)}(\beta, t) &= \frac{1}{n} \sum_{i=1}^n y_{hi}(t) \exp(\beta' z_{hi}(t)), \\ S_{hj}^{(1)}(\beta, t) &= \frac{1}{n} \sum_{i=1}^n y_{hi}(t) z_{hij}(t) \exp(\beta' z_{hi}(t)), \\ S_{hj\lambda}^{(2)}(\beta, t) &= \frac{1}{n} \sum_{i=1}^n y_{hi}(t) z_{hij}(t) z_{hi\lambda}(t) \exp(\beta' z_{hi}(t)), \end{aligned} \quad (7.4)$$

$h = 1, \dots, k$; $j, \lambda = 1, \dots, p$; and we define

$$E_{hj}(\beta, t) = \frac{S_{hj}^{(1)}(\beta, t)}{S_h^{(0)}(\beta, t)}.$$

Then the j th component of $U(\beta, t)$ is given as

$$U_j(\beta, t) = \sum_{h=1}^k \left[\int_0^t \sum_{i=1}^n z_{hij}(s) dN_{hi}(s) - \int_0^t E_{hj}(\beta, s) dN_h(s) \right],$$

and using (3.14) and (7.2) we see that

$$U_j(\beta_0, t) = \sum_{h=1}^k \sum_{i=1}^n \int_0^t [z_{hij}(s) - E_{hj}(\beta_0, s)] dM_{hi}(s), \quad j=1, \dots, p \quad (7.5)$$

are linear combinations of stochastic integrals. Thus the martingale central limit theorem (Theorem 3.2) can be applied to prove that the process $n^{-1/2} U(\beta_0, \cdot)$ asymptotically, as $n \rightarrow \infty$, is distributed as a mean zero Gaussian martingale. To transform this result into a theorem concerning the asymptotic distribution of $\hat{\beta}$ we Taylor expand $U_j(\beta, t)$ around β_0 :

$$U_j(\beta, 1) = U_j(\beta_0, 1) + \sum_{\lambda=1}^P I_{j\lambda}(\beta^*)(\beta_\lambda - \beta_{0\lambda}), \quad (7.6)$$

where β^* is between β and β_0 and $I_{j\lambda}(\beta)$ is the second order partial derivative of $C(\beta, 1)$ w.r.t. β_j and β_λ . Thus

$$I_{j\lambda}(\beta) = - \sum_{h=1}^k \int_0^1 \left[\frac{s_{hj\lambda}^{(2)}(\beta, s)}{s_h^{(0)}(\beta, s)} - E_{hj}(\beta, s) E_{h\lambda}(\beta, s) \right] dN_h(s). \quad (7.7)$$

We write $\underline{I}(\beta)$ for the $p \times p$ matrix with components $I_{j\lambda}(\beta)$. Inserting $\hat{\beta}$ in (7.6) we get (since $U_j(\hat{\beta}, 1) = 0$ by definition of $\hat{\beta}$)

$$n^{-\frac{1}{2}} U_j(\beta_0, 1) = \sum_{\lambda=1}^P \left(-\frac{1}{n} I_{j\lambda}(\beta^*) \right) n^{\frac{1}{2}} (\hat{\beta}_\lambda - \beta_{0\lambda}),$$

with β^* between $\hat{\beta}$ and β_0 . It now remains to be proved that $\hat{\beta}$ is consistent as $n \rightarrow \infty$, and that for each β^* with $\beta^* \xrightarrow{P} \beta_0$

$$-\frac{1}{n} I_{j\lambda}(\beta^*) \xrightarrow{P} \sigma_{j\lambda}$$

where $\Sigma = \{\sigma_{j\lambda}\}$ is positive definite.

Sufficient conditions for these results to hold were given by Andersen & Gill (1982) in the case $k = 1$. For $k > 1$ the conditions include an assumption of the sums (7.4) converging uniformly for $t \in [0, 1]$ and β in some neighbourhood of β_0 to functions $s_h^{(0)}(\beta, t)$, $s_{hj}^{(1)}(\beta, t)$ and $s_{hj\lambda}^{(2)}(\beta, t)$, respectively, in probability. Furthermore some regularity assumptions on the limiting s-functions are needed (Andersen & Gill, 1982, p. 1105, item D). Under such conditions $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ is asymptotically multnormally distributed $N_p(0, \Sigma^{-1})$ and $\sigma_{j\lambda}$ can be estimated consistently by $-\frac{1}{n} I_{j\lambda}(\hat{\beta})$, see (7.7).

Under the same set of conditions the estimates for $A_{0h}(t)$:

$$\hat{A}_{0h}(t) = \int_0^t [nS_h^{(0)}(\hat{\beta}, s)]^{-1} dN_h(s), \quad h = 1, \dots, k, \quad (7.8)$$

(with $S_h^{(0)}$ defined in (7.4)) will be distributed asymptotically as Gaussian processes. The proof for this goes as in Andersen and Gill (1982, Theorem 3.4), the main step being to notice that

$$\int_0^t \{ [nS_h^{(0)}(\hat{\beta}_0, s)]^{-1} dN_h(s) - I(Y_h(s) > 0) \alpha_{0h}(s) ds \}$$

with $Y_h = Y_{h1} + \dots + Y_{hn}$, $h = 1, \dots, k$ are local square integrable martingales which are orthogonal to (7.5).

One should notice that for a homogeneous group of individuals, i.e. when all $Z_i \equiv 0$ the estimator (7.8) reduces to the Nelson-Aalen estimator (4.2).

The results mentioned so far make it possible to draw asymptotic inference on the regression parameters β in the presence of the "nuisance" functions $\alpha_{01}, \dots, \alpha_{0k}$. This property of the model was in fact Cox's original motivation for introducing the semi-parametric specification (7.1). In some cases, however, the underlying intensities are also of interest, and we have seen how their integrals $A_{0h}(t)$ can be estimated. The underlying intensities themselves can be estimated by smoothing $\hat{A}_{0h}(t)$ by means of a kernel function similarly to the approach in Subsection 4.2:

$$\hat{\alpha}_{0h}(t) = \frac{1}{b} \int_0^1 K\left(\frac{t-s}{b}\right) d\hat{A}_{0h}(s), \quad t \in [b, 1-b]. \quad (7.9)$$

Combining the asymptotic results of Ramlau-Hansen (1983a) with those concerning $\hat{A}_{0h}(t)$ mentioned above, it can be seen that the asymptotic distribution of (7.9) when $n \rightarrow \infty$ (and $b_n \rightarrow 0$ in such a way that $nb_n \rightarrow \infty$) is normal with mean $\alpha_{0h}(t)$ and variance $\sigma_h^2(t)/n$, where $\sigma_h^2(t)$ is given by

$$\sigma_h^2(t) = \frac{\alpha_{0h}(t)}{s_h^{(0)}(\hat{\beta}_0, t)} \int_0^1 K^2(s) ds.$$

The latter quantity can be estimated consistently by

$$\hat{\sigma}_h^2(t) = \frac{n}{b} \int_0^1 K^2\left(\frac{t-s}{b}\right) [nS_h^{(0)}(\hat{\beta}, s)]^{-2} dN_h(s).$$

Furthermore $\hat{\alpha}_{0h_1}(t_1)$ and $\hat{\alpha}_{0h_2}(t_2)$ are independent for $h_1 \neq h_2$ or $t_1 \neq t_2$. (For these asymptotic results to hold it is crucial that $b_n \rightarrow 0$. For fixed b the asymptotic variance will be larger than $\sigma_h^2(t)$ and the independence result will not be true. The relevant variance formula in this case was given by Andersen & Rasmussen (1982).)

As noted above, the estimator $\hat{\alpha}_{0h}(t)$ reduces to the ordinary Nelson-Aalen estimator when all $Z_{hi} = 0$. Another link between the Cox regression model and the models described in the preceding sections is the fact that certain score test statistics based on the extended Cox model (7.2) coincide with the nonparametric tests discussed in Section 5. For example, consider the k-sample problem with proportional intensities, i.e. introduce covariates $Z_{hi}(t) = (Z_{hi1}(t), \dots, Z_{hip}(t))'$, where $p = k-1$, and $Z_{hij}(t) = \delta_{hj} Y_{hi}(t)$, $j = 1, \dots, p$, and δ_{hj} as usual is a Kroneches delta. Then the score test statistic for the hypothesis that each of the corresponding regression coefficients is zero is

$$S = \underline{U}(\underline{0}, 1)' (-\underline{I}(\underline{0}))^{-1} \underline{U}(\underline{0}, 1)$$

with \underline{U} defined just above (7.5) and \underline{I} in (7.7). It is easily seen that S is equal to the k-sample logrank test statistic ((5.16), (5.17) and (5.18) with $L(t) = I(Y.(t) > 0)$). Also the other k-sample test statistics in Subsection 5.2 can be obtained

by appropriate choices of time-dependent covariates, see Lustbader (1980), Oakes (1981), Andersen & Gill (1982, p. 1110) and Andersen et al. (1982, p. 236).

Example

In the example in Section 5 the influence of sex on the mortality among insulin dependent diabetics was examined using a nonparametric approach. We saw how it was possible to give an "after the fact estimate" of the hazard ratio θ by first estimating the integrated intensities $A_1(t)$ and $A_2(t)$, without assuming proportionality, and afterwards estimating θ from $\hat{A}_1(t)$ and $\hat{A}_2(t)$, see (5.22). Using the Cox model as described above, the hazard ratio $\theta = \exp(\beta)$ and the underlying hazard function $\alpha_0(t)$ can be estimated simultaneously. So, assume now that each individual counting process $N_i(t)$ corresponding to a female diabetic i has intensity process $\alpha_0(t)Y_i(t)$, with $Y_i(t)$ being defined in the usual fashion as the predictable indicator for individual i being at risk at $t-$. For a male diabetic $N_i(t)$ has intensity process $\alpha_0(t)\exp(\beta)Y_i(t)$. Thus if we define the covariate Z_{i1} by

$$Z_{i1} = \begin{cases} 1 & \text{if } i \text{ is a man,} \\ 0 & \text{if } i \text{ is a woman,} \end{cases}$$

then any individual counting process has an intensity process which can be written as

$$\lambda_i(t) = \alpha_0(t)\exp(\beta_1 Z_{i1})Y_i(t),$$

cf. (7.2). In this example the estimated sex effect is $\hat{\beta}_1 = 0.33$ with an estimated standard error 0.21. These results are in close agreement with those of the example in Subsection 5.2 in that

$\exp(\hat{\beta}_1) = 1.39$ with an approximate 95% confidence interval $[0.93, 2.08]$. One should notice that the confidence interval based on the Cox regression model is more narrow than the one based on the estimate $\hat{\theta}_{K_L}$. This is a consequence of the general result that any estimator $\hat{\theta}_K$ of the form (5.22) has larger asymptotic variance than $\exp(\hat{\beta})$ (Andersen, 1983a).

Figure 12 shows the estimated integrated underlying intensity $\hat{A}_0(t)$, in this case estimating the mortality among female diabetics, and also the smoothed estimate $\hat{a}_0(t)$ using the same smoothing procedure as in the example in Subsection 4.2. Comparing these two figures with the Figures 6 to 8, a fairly close agreement is seen, indicating that the hypothesis of proportional hazards seems reasonable.

Another covariate of interest is the age at onset of the disease. Thus we introduce the covariate Z_{i2} = age (in years) at onset for individual no. i and consider the model

$$\lambda_i(t) = \alpha_0(t) \exp(\beta_1 Z_{i1} + \beta_2 Z_{i2}) Y_i(t).$$

Notice that the interpretation of β_1 is now different. In the first model $\exp(\beta_1)$ was the ratio between the hazard functions for any male and any female diabetic, whereas now $\exp(\beta_1)$ is the ratio between the hazard functions for a male and a female with the same age at onset. The estimates in the new model become $\hat{\beta}_1 = 0.36$ (0.21) and $\hat{\beta}_2 = -0.015$ (0.014) with the estimated standard errors given in brackets. The estimated correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$ is -0.13 . Due to this negative correlation, indicating that in this data set female diabetics tend to have a slightly earlier disease onset than males, the estimated sex effect increased slightly from 0.33 to 0.36. The negative value

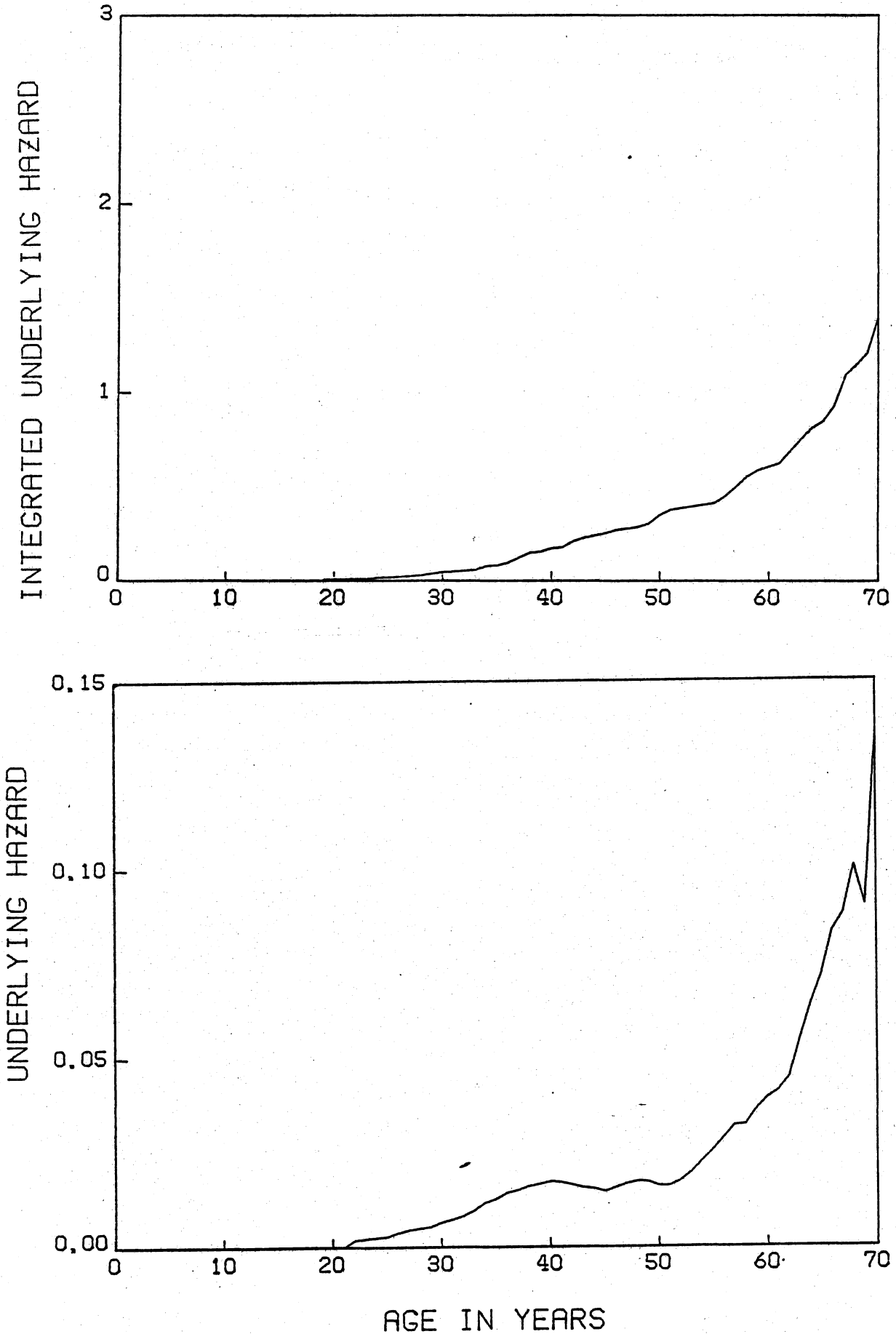


Fig. 12. Estimated integrated underlying death intensity (upper figure) and estimated underlying death intensity (lower figure) for diabetics in Cox regression model with sex included as covariate.

$\hat{\beta}_2 = -0.015$ indicates a worse prognosis for diabetics with early disease onset.

In Subsection 5.1 we studied an alternative model for the mortality among insulin dependent diabetics, where the mortality was assumed to be proportional to the mortality in the general Danish population. Within the framework of the Cox regression model it is possible to examine the proportionality assumption in the model more closely. Andersen (1984) studied a model of the form

$$\lambda_i(t) = \alpha_0(t) \exp(\beta' Z_i(t)) \mu_i(t) Y_i(t), \quad (7.10)$$

where $\mu_i(t)$ is the known age-, sex- and calendar time specific population based mortality for an individual similar to no.i. In this model $\alpha_0(t)$ is an underlying excess mortality and the covariates $Z_i(t)$ have a multiplicative effect on the excess mortality. The properties of the model (7.10) can be studied in a way similar to the Cox model because $\mu_i(t) Y_i(t)$ is predictable.

For the diabetes data a model of the form (7.10) was analysed. No covariates $Z_i(t)$ were included in the exponent part of the model, but possibly different excess mortalities $\alpha_{01}(t)$ and $\alpha_{02}(t)$ among females and males, respectively, were allowed. Figure 13 shows the estimates $\hat{\alpha}_{01}(t)$ and $\hat{\alpha}_{02}(t)$ obtained by smoothing estimates $\hat{A}_{01}(t)$ and $\hat{A}_{02}(t)$ by means of a kernel function. For high and low values of age t the standard error of the estimates are large due to the narrow window. It is seen that both females and males tend to have a high excess mortality around age 35-40 years indicating that the assumption of the intensities being proportional to the population mortalities is probably unrealistic.

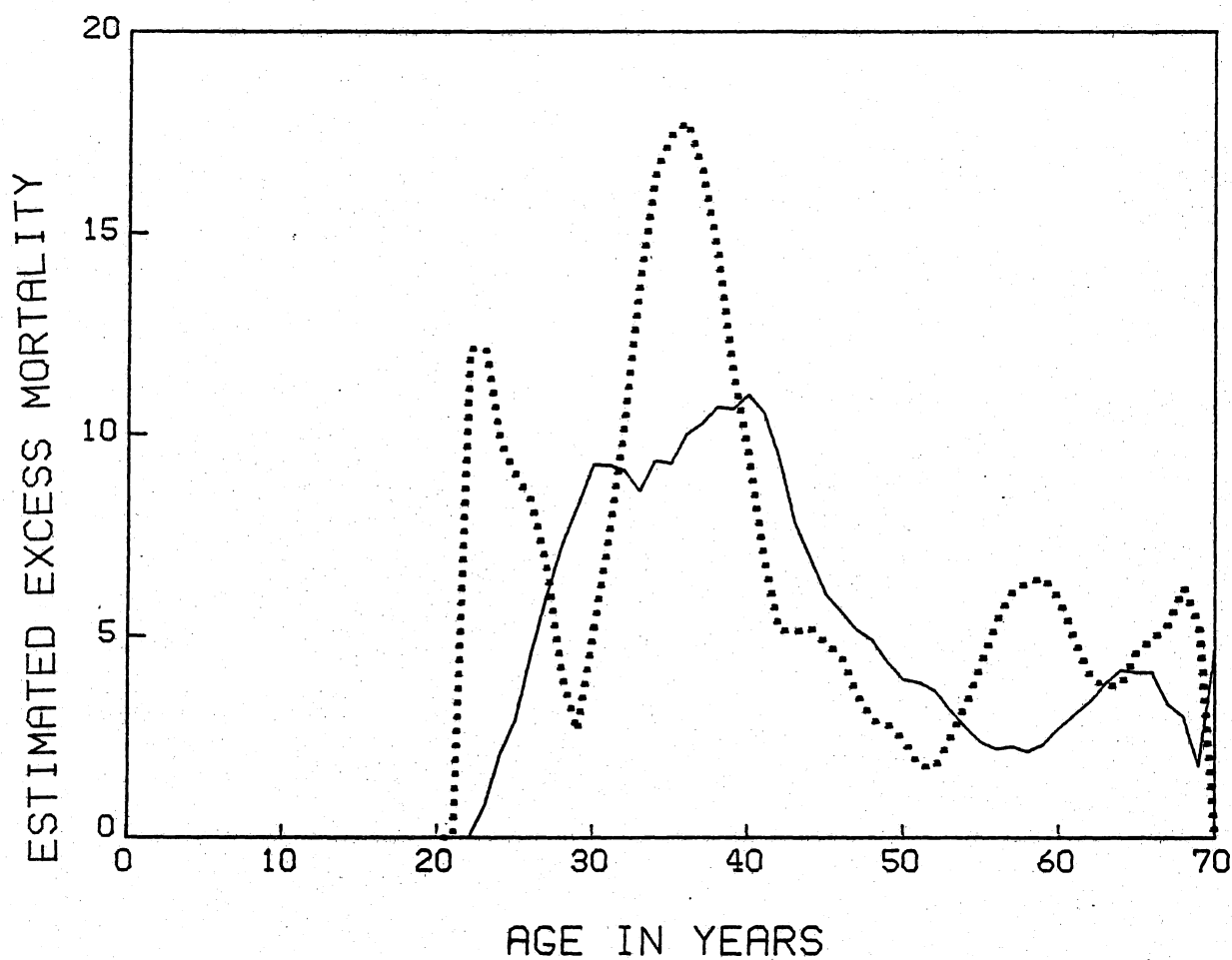


Fig. 13. Estimated excess mortality in relation to the general Danish population for female diabetics (···) and for male (—) diabetics.

Assuming that the excess mortality among males and females are proportional (Figure 13 indicates that they are possibly identical) the ratio between the excess mortalities can be estimated using the model (7.10) with the covariate Z_1 included. In this model we obtain the estimate $\hat{\beta}_1 = -0.19 (0.21)$ confirming the tendency seen in the example in Subsection 5.1 that females have a slightly higher excess mortality. There we found $\hat{\theta}_2/\hat{\theta}_1 = 0.76 = \exp(-0.27)$. \square

In this subsection we have so far only considered versions of the Cox regression model with an exponential form of the relative risk function $g(\cdot)$, see (7.1). Prentice and Self (1983) studied Cox-type models with a general $g(\cdot)$ (when $k=1$), and proved consistency and asymptotic normality of the estimator $\hat{\beta}$ which maximizes (7.3) with $t=1$ and $\beta'Z_{hi}(s)$ replaced by $\log g(\beta'Z_{hi}(s))$. The conditions under which these asymptotic results were proved include those of Andersen & Gill (1982) sketched above, but in addition some extra conditions are needed to ensure the positivity of $g(\beta'Z_{hi}(s))$ and positive definiteness of the estimator for ξ in a neighbourhood of β_0 , see Prentice and Self (1983) for details.

7.2 Parametric regression models.

In the models discussed in the previous subsection the factor $\alpha_{0h}(t)$ in (7.1) was left completely arbitrary in that no specific form of $\alpha_{0h}(t)$ was assumed. In this subsection we consider an nk -variate counting process as in Subsection 7.1, but instead of (7.2) we assume that N_{hi} has intensity process of the form

$$\lambda_{hi}(t) = \alpha_{hi}(t; \theta_0) \exp(\beta_0' Z_{hi}(t)) Y_{hi}(t), \quad (7.11)$$

with θ_0 belonging to an open subset of R^q . References to papers

where parametric models of the form (7.11) have been studied in the special case of survival data are given by Kalbfleisch and Prentice (1980, Section 3). Borgan (1984, Section 6) studied maximum likelihood estimation for the model (7.11) for the special case of $k = 1$, i.e. of individual univariate counting processes. We shall briefly present these results extended to our more general setting.

The log-likelihood function is analogous to (6.3), i.e.

$$\log L(\underline{\theta}, \underline{\beta}) = \sum_{h=1}^k \sum_{i=1}^n \left[\int_0^1 \{ \log \alpha_h(s; \underline{\theta}) + \underline{\beta}' \underline{z}_{hi}(s) \} dN_{hi}(s) - \int_0^1 \alpha_h(s, \underline{\theta}) \exp(\underline{\beta}' \underline{z}_{hi}(s)) Y_{hi}(s) ds \right],$$

and the maximum likelihood estimators $\hat{\underline{\theta}}$ and $\hat{\underline{\beta}}$ are defined as solutions to the set of equations

$$\sum_{h=1}^k \sum_{i=1}^n \left[\int_0^1 \frac{\frac{\partial}{\partial \theta_j} \alpha_h(s, \underline{\theta})}{\alpha_h(s, \underline{\theta})} dN_{hi}(s) - \int_0^1 \frac{\partial}{\partial \theta_j} \alpha_h(s, \underline{\theta}) \exp(\underline{\beta}' \underline{z}_{hi}(s)) Y_{hi}(s) ds \right] = 0$$

$j = 1, \dots, q$ and (7.12)

$$\sum_{h=1}^k \sum_{i=1}^n \left[\int_0^1 \underline{z}_{hil}(s) dN_{hi}(s) - \int_0^1 \alpha_h(s, \underline{\theta}) \underline{z}_{hil}(s) \exp(\underline{\beta}' \underline{z}_{hi}(s)) Y_{hi}(s) ds \right] = 0$$

$l = 1, \dots, p$. The proof for the asymptotic properties of the estimators $\hat{\underline{\theta}}$ and $\hat{\underline{\beta}}$ now proceeds exactly as in Borgan (1984) by first of all noticing that the left hand sides of the likelihood equations (7.12) evaluated at the true parameter values $\underline{\theta}_0$ and $\underline{\beta}_0$ are linear combinations of stochastic integrals of the predictable processes $\frac{\partial}{\partial \theta_j} \alpha_h(s, \underline{\theta}_0) / \alpha_h(s, \underline{\theta}_0)$ and $\underline{z}_{hil}(s)$, respectively, w.r.t. the local square integrable martingales $M_{hi}(t) =$

$N_{hi}(t) - \int_0^t \lambda_{hi}(s) ds$. Hence a central limit theorem for these mar-

tingales can be proved using Theorem 3.2. Subject to some regularity conditions this result can be used to prove that the maximum likelihood estimators are asymptotically multinormally distributed with the proper expectation and with a covariance matrix that may be estimated in the usual manner. The regularity conditions needed are slight extensions of those of Borgan (1984, p.14), in particular asymptotic stability conditions on the sums (7.4) (as well as a similar "third order" condition) have to be fulfilled.

The results of Borgan (1984, Section 6) may also be used to derive that minus two times the logarithm of the likelihood ratio test statistic is asymptotically chi-squared distributed in the usual manner (cf. our Section 6).

As it was the case for the Cox regression model (Subsection 7.1) other forms of the relative risk function than the exponential one can be considered, but the regularity conditions get more complicated.

7.3 A linear regression model for the intensity process

An alternative regression model for a multivariate counting process $\underline{N}(t) = (N_1(t), \dots, N_k(t))$ was introduced by Aalen (1980). He suggested a matrix version of the multiplicative intensity model, assuming that the intensity process satisfies

$$\underline{\lambda}(t) = \underline{Y}(t) \underline{\alpha}(t), \quad (7.13)$$

$\underline{\alpha} = (\alpha_1, \dots, \alpha_p)'$, $p < k$, being a vector of unknown functions and $\underline{Y}(t) = (Y_{hj}(t))$, $h = 1, \dots, k$, $j = 1, \dots, p$, a matrix of predictable processes. Thus the model is given by

$$\lambda_h(t) = Y_{h1}(t)\alpha_1(t) + \dots + Y_{hp}(t)\alpha_p(t), \quad h = 1, \dots, k.$$

As an example h could refer to the single individuals, $Y_{h1}(t)$ could be the "usual" indicator of individual no. h being at risk at time $t-$, and for $j = 2, \dots, p$ Y_{hj} could be defined as

$$Y_{hj}(t) = Z_{hj}(t) Y_{h1}(t),$$

for some predictable covariate processes $Z_{hj}(t)$.

Estimators for

$$A_j(t) = \int_0^t \alpha_j(s) ds, \quad j = 1, \dots, p$$

can be defined by

$$\hat{A}(t) = \int_0^t J(s) \underline{Y}^-(s) dN(s), \quad (7.14)$$

where $\underline{Y}^-(t)$ is a generalized inverse of $\underline{Y}(t)$ (i.e. \underline{Y}^- satisfies $\underline{Y}^-(t)\underline{Y}(t) =$ the $r \times r$ identity matrix) and J is defined by

$$J(t) = \lim_{\Delta t \rightarrow 0} I[\text{rank } \underline{Y}(t-\Delta t) = r]$$

(thus J is predictable). It follows now that with $\underline{A}^*(t) = \int_0^t J(s) \underline{\alpha}(s) ds$, $\hat{A} - \underline{A}^*$ is an r -variate local square integrable martingale with predictable covariation process (a $p \times p$ matrix)

$$\langle \hat{A} - \underline{A}^* \rangle(t) = \int_0^t J(s) \underline{Y}^-(s) \underline{\lambda}(s) \underline{Y}^-(s)' ds,$$

where $\underline{\lambda}$ is the matrix $\text{diag}(\lambda_1, \dots, \lambda_k)$. From this result some exact and asymptotic properties may be derived.

Except for the computation of \underline{Y}^- the estimators (7.14) are simple and the model (7.13) is a truly nonparametric alternative to the semi-parametric and parametric regression models discussed in the two previous subsections. Aalen (1980, section 5) presents one application of this regression model.

8. LIMITATIONS OF THE COUNTING PROCESS APPROACH

As seen in the preceding sections, the "counting process approach" to life history analysis is very useful in the study of a number of statistical estimation and testing procedures. However, in its simple form, as presented in this paper, it does not solve every problem, and in this final section we comment upon some points which illustrate the limitations. Some areas where further research is needed are also mentioned.

Only for the very simplest situations in life history analysis, like type II censored exponential life times (Epstein and Sobel, 1953), is it possible to derive useful expressions for the exact distribution of the estimators and test statistics. Thus most statistical procedures for analysing life history data have to rely on large sample results, and it is important to know "how large" a sample must be to make this appropriate. We have reviewed how the martingale central limit theorem may be used to study the asymptotic properties of many estimators and test statistics. However, little is known about the rate of convergence for the martingale central limit theorem, so the counting process approach cannot help us solve this problem.

For the special case of a competing risks model, Csörgö and Horvath (1982) have studied uniform rates of convergence for the Nelson-Aalen estimators and certain transforms of these. Their results are quite disappointing in general, in that they indicate that quite large sample sizes may be needed for the asymptotics to hold. However, we believe that their results are too general to give guidance about the sample sizes needed in concrete applications. Our guess is that simulation studies for "typical" situations encountered in practice will show that the large sample results are satisfactory for much smaller sample sizes than those indicated

by the results of Csörgö and Horvath (1982).

Some simulation studies have in fact been performed to study the small sample properties of some of the statistical methods discussed in this paper in the special case of censored survival data. We mention the studies of the performance of parametric and nonparametric two-sample tests of Gehan and Thomas (1969), Lee et al. (1975) and Latta (1981), the studies of parametric regression models and the Cox regression model of Peace and Flora (1978) and Lee et al. (1983), and Schou and Væth's (1980) study of failure time data from an exponential distribution under various types of censorship. However, much work remains to be done before a satisfactory knowledge of the small sample properties of the various estimators and test statistics has been established. Especially we will mention the possibility of using transformations to improve the approximations to the asymptotic distributions. Kalbfleisch and Prentice (1980, pp. 14-15) mention this possibility in connection with the Kaplan-Meier estimator for the survival function, and Schou and Væth (1980) show that the cubic-root of the occurrence/exposure rate may be considered to be approximately normally distributed for much smaller sample sizes than are needed for the occurrence/exposure rate itself.

In Subsection 3.1 we indicated how the results of this paper are valid under quite arbitrary censoring mechanisms, as long as censoring only depends on the "past" and outside random variation. Speaking in technical terms the censoring processes $C_i(\cdot)$ of Subsection 3.1 have to be predictable. There are some important situations in which this is not the case.

In testing with replacement, items are life tested one at a time. At each failure, the failed item is replaced by a new one. If observation stops after a fixed period of time, then the last item is censored. Moreover, its censoring time is determined by the possibly longer life times of the preceding items, i.e. the censoring does not depend solely on the "past". This means that the results presented above cannot be applied. It is, however, shown e.g. by Gill (1981) how more classical arguments for i.i.d. random variables can be used to study this situation.

For semi-Markov models, or Markov renewal processes, the intensity for a transition between two states depends on the time elapsed since the entry into the current state. Thus "time" starts anew at zero after each transition into a new state. For censored observations of such a process, we get the same type of problem as discussed above, and in general the counting process approach in its simple form cannot be used (Gill, 1980b). However, Voelkel and Crowley (1984) show how one via a random time change may apply the counting process methods of this paper for some hierarchical semi-Markov processes.

For sequential analysis with staggered entry one has to consider two time scales simultaneously, a fact which makes the theoretical problems much more complicated. Sellke and Siegmund (1983) discuss this situation for Cox's (1972) proportional hazards model with one regression parameter. Here the derivative of the log-partial-likelihood function, evaluated at the true parameter value, will no longer be a martingale as was the case in our Subsection 7.1. However, Sellke and Siegmund (1983) show that it may be approximated by a martingale, so that the martingale central limit theorem may still be used. Another paper on sequential

analysis with staggered entry is by Slud (1984), who discusses nonparametric two-sample tests for this situation.

Thus there are important situations with censored data that are not covered by the results reviewed in this paper. Another important limitation is that most statistical methods presented in the Sections 4-7 above make strict demands in terms of data accuracy; in particular dated events on the individual level are usually needed. (An important exception, where only aggregate level data are needed, is parametric regression models with piecewise constant underlying intensities and solely qualitative time-independent covariates, cf. Borgan (1984, Section 5.2) for a particular simple example of such a situation.) Real life data are often less comprehensive. For instance, it happens that information is only available on the exact time for some of the events of interest, or data may be missing for a systematic part of the study population. Development of statistical methods for situations with incomplete data is therefore of considerable interest. There seems, however, to be no general solution to the problem of estimating the intensities of partially observed Markov chains or more general stochastic process models. Also we believe that it will be more the exception than the rule that the counting process approach will help in solving such problems. A few examples will illustrate these points.

When the number of study subjects is large the computations needed to evaluate the maximum partial-likelihood estimator $\hat{\beta}$ of Subsection 7.1 may be very time-consuming, especially if some covariates are time-dependent. Considerable reduction in computing time may therefore be achieved by comparing each failure with a

random sample of the corresponding risk set (Thomas, 1977; see also Oakes, 1981, Section 3.4; Breslow et al., 1983, Section 6). The same idea may be applied to implement the Cox regression model for case-control studies (Prentice and Breslow, 1978). To our knowledge the distributional properties of the resulting estimators for the regression parameters have not been studied (see, however, Oakes, 1981; Breslow et al., 1983). Also it seems as if the counting process approach does not work for this situation, the reason being that a simple relation like (3.14) (combined with (7.2)) will no longer hold for the modified version of (7.3) valid when we sample from the risk set. Probably more classical results for i.i.d. random variables may still be used, however, see Borgan and Gill (1982) who use a Skorohod construction as in Breslow and Crowley (1974) to study Nelson-Aalen-type estimators and nonparametric tests for case-control studies in a Markov chain setting.

Similar problems arise in the study of demographic incidence rates by Borgan and Ramlau-Hansen (1983). They consider a special case of partially observed Markov chains, in which transitions within a subset of states are observed in detail, while counts of transitions out of this subset are only observed aggregated over the states. Such situations arise in demography, where one for instance in a study of marriage formation and dissolution in a female birth cohort may have detailed information about the marriages, but no information about the distribution of the women over the various marital statuses. Borgan and Ramlau-Hansen (1983) study estimators of the Nelson-Aalen- and occurrence/exposure-type for this situation. Since relations like (3.14) (combined with (3.9)) do not apply for the estimators they consider, the martingale central limit theorem cannot be used, however, and they

have to study the distributional properties of the estimators by more classical methods.

As a final example of a situation where incomplete data occur, and where the methods discussed in this paper do not work, let us mention experiments with laboratory animals. For such experiments one is seldom able to observe the exact time for the onset of a disease, and alternative incomplete observational plans have to be used. Such observational plans may include the killing of certain animals at prespecified times (serial sacrifice) or periodic diagnosis of live animals, see Borgan et al. (1984) who also provide further references to the literature. It is hard to see how the counting process approach can be of much use in studying statistical procedures for analysing data from such experiments.

One problem in which incomplete data occur and where martingale methods have proved useful is the epidemics model studied by Becker (1977, 1981) and Becker & Hopper (1983). They consider a closed population of "susceptibles" to which an "infected" individual arrives and study the intensity at which the infectious disease is spread. In this situation it is unreasonable to assume that one is able to observe both the number of susceptibles and the number of infected at any time t . However, under the assumption of a constant infection intensity α these authors derive an estimator $\tilde{\alpha}$ that can be expressed in terms of observable quantities, and they study the properties of $\tilde{\alpha}$ using martingale methods.

In conclusion, detailed life history data may be given a thorough analysis using the methods based on counting processes discussed in this paper. However, if less precise information is available, then alternative techniques are necessary. Some such techniques can, as we have seen in this final section, also be based on counting process ideas, but not in the simple form as presented in this paper.

ACKNOWLEDGEMENTS

We are grateful to Anders Green and Anne Katrin Sjølie for allowing us to use the diabetes data, and to Erik Christensen for allowing us to use the CSL-I data. We wish to thank Richard Gill for his comments on an earlier draft of the manuscript.

APPENDIX

A. Derivation of the statistical properties of the Nelson-Aalen estimators

As mentioned in Subsection 4.1, the statistical properties of the Nelson-Aalen estimators (4.2) are most conveniently derived by introducing the quantities $A_h^*(t)$, see (4.4). It should be realized, however, that provided that Y_h increases uniformly (in probability) over $[0,1]$, the difference between A_h and A_h^* will eventually vanish, and A_h^* may be replaced by A_h everywhere in the asymptotic results below.

The key formula for deriving the properties of $\hat{A}_h(t)$ is (4.5) which we repeat here:

$$\hat{A}_h(t) - A_h^*(t) = \int_0^t \frac{J_h(s)}{Y_h(s)} dM_h(s); \quad (A.1)$$

$h=1, \dots, k$. It follows that the $\hat{A}_h - A_h^*$ are mean-zero local square integrable martingales, and the orthogonality of the M_h s and (3.18) give that (A.1) have predictable covariation processes

$$\langle \hat{A}_h - A_h^*, \hat{A}_j - A_j^* \rangle(t) = \delta_{hj} \int_0^t \frac{J_h(s)}{Y_h(s)} \alpha_h(s) ds, \quad (A.2)$$

where δ_{hj} is a Kronecker delta, and we write $\langle M, M \rangle$ for $\langle M \rangle$. Thus, the local martingales (A.1) are also orthogonal. From these facts follow the "unbiasedness" property for all $t \in [0,1]$ (assuming that the expectations exist, cf. remark just above Theorem 3.1)

$$E\hat{A}_h(t) = EA_h^*(t); \quad h = 1, \dots, k,$$

and furthermore that the processes $\hat{A}_h - A_h^*$, $h = 1, 2, \dots, k$, have uncorrelated increments and that $\hat{A}_h(t) - A_h^*(t)$ is uncorrelated with $\hat{A}_j(s) - A_j^*(s)$ for any s, t and $h \neq j$.

The mean squared error function of (4.2) is given by

$$\begin{aligned} \tau_h(t) &= E[\hat{A}_h(t) - A_h^*(t)]^2. \\ &= E\langle \hat{A}_h - A_h^* \rangle(t), \end{aligned} \tag{A.3}$$

where the latter equality follows from the definition of $\langle M \rangle$ as the compensator for M^2 . As an estimator for $\tau_h(t)$, we use $\hat{\tau}_h(t)$ given by (4.6). The difference between (4.6) and (A.3) equals the stochastic integral $\int_0^t J_h(s) [Y_h(s)]^{-2} dM_h(s)$, so that (4.6) is unbiased.

To study the large sample properties of the Nelson-Aalen estimators, we consider a sequence of counting processes indexed by $n = 1, 2, \dots$, each satisfying the multiplicative intensity model with the α_h s being the same for all n . By a direct application of Lenglart's inequality (Theorem 3.1) we get, using (A.2), that

$$\sup_{t \in [0,1]} | \hat{A}_h^{(n)}(t) - A_h^{*(n)}(t) | \xrightarrow{P} 0 \tag{A.4}$$

as $n \rightarrow \infty$, if only

$$\int_0^1 \frac{J_h^{(n)}(s)}{Y_h^{(n)}(s)} \alpha_h(s) ds \xrightarrow{P} 0 \tag{A.5}$$

as $n \rightarrow \infty$. We note that (A.5) essentially requires that $Y_h^{(n)}(t)$ becomes large for all $t \in [0,1]$.

Let us then study the asymptotic distribution of the Nelson-Aalen estimators. We will do this by applying the martingale central limit theorem (Theorem 3.2). By (A.2), what essentially is needed for the applications considered in this paper is that $Y_h^{(n)}(t)/n$ converge to some deterministic function as $n \rightarrow \infty$. For other applications (e.g. Aalen, 1978, Section 8) we need to normalize the Y_h s by other constants than $\{n\}$. By Theorem 3.2 (with $p=k_n=k$), we may therefore state the following general result.

Theorem. Assume that there exist a sequence of positive constants $\{a_n\}$, increasing to infinity as $n \rightarrow \infty$, and non-negative square integrable functions g_h , $h = 1, 2, \dots, k$, defined on $[0, 1]$, such that

A) For each $t \in [0, 1]$ and $h = 1, 2, \dots, k$

$$a_n^2 \int_0^t \frac{J_h^{(n)}(s)}{Y_h^{(n)}(s)} \alpha_h(s) ds \xrightarrow{P} \int_0^t g_h^2(s) ds \quad \text{as } n \rightarrow \infty.$$

B) For all h and $\epsilon > 0$

$$a_n^2 \int_0^t \frac{J_h^{(n)}(s)}{Y_h^{(n)}(s)} \alpha_h(s) I(|a_n \frac{J_h^{(n)}(s)}{Y_h^{(n)}(s)}| > \epsilon) ds \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

Then

$$a_n \begin{pmatrix} \hat{A}_1^{(n)} & - & \hat{A}_1^{*(n)} \\ & \vdots & \\ \hat{A}_k^{(n)} & - & \hat{A}_k^{*(n)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix},$$

where X_1, X_2, \dots, X_k are independent Gaussian martingales with $X_h(0) = 0$ and $\text{Cov}(X_h(s), X_h(t)) = \int_0^{t \wedge s} g_h^2(u) du$.

The $(k \times k)$ -matrix of predictable processes H_{jh} in Theorem 3.2 is in this case diagonal with $H_{hh} = a_n^2 J_h Y_h^{-1}$. In all applications in this paper we will have $a_n = n^{1/2}$.

In practice the verification of the conditions A and B is not always so direct, and it is useful to have alternative and more easily verifiable sets of conditions. A simple set of conditions, sufficient for A and B to hold true, and which is often fulfilled in practice, is

A') For $h = 1, 2, \dots, k$

$$\sup_{t \in [0, 1]} |a_n^2 J_h^{(n)}(t) [Y_h^{(n)}(t)]^{-1} \alpha_h(t) - g_h^2(t)| \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

B') For $h = 1, 2, \dots, k$

$$\sup_{t \in [0,1]} | a_n J_h^{(n)}(t) [Y_h^{(n)}(t)]^{-1} | \xrightarrow{P} 0$$

as $n \rightarrow \infty$.

Alternative sets of sufficient conditions are discussed by Gill (1980a, 1983b), Andersen et al. (1982) and Helland (1983). To apply the weak convergence result in practice, one must be able to estimate the covariance function of the limiting Gaussian martingale. By Rebolledo (1980) we have that, if Conditions A and B hold true, then for all h and $t \in [0,1]$

$$a_n^2 \hat{\tau}_h^{(n)}(t) \xrightarrow{P} \int_0^t g_h^2(u) du, \quad (\text{A.6})$$

where $\hat{\tau}_h^{(n)}(t)$ is defined as in (4.6).

An application of Lenglart's inequality (Theorem 3.1) shows that we have uniform convergence for $t \in [0,1]$ in probability in (A.6) provided that

$$a_n^4 \int_0^1 \frac{J_h^{(n)}(s)}{[Y_h^{(n)}(s)]^3} \alpha_h(s) ds \xrightarrow{P} 0$$

as $n \rightarrow \infty$, and that we have uniform convergence for $t \in [0,1]$ in probability in Condition A. It is straightforward to see that Conditions A' and B' are sufficient for this to hold true.

B. Asymptotic properties of the maximum likelihood estimator for a one-parameter univariate counting process model

We consider a sequence $\{N^{(n)}\}$ of univariate counting processes, where $N^{(n)}(t)$ has intensity process $\alpha(t; \theta_0)Y^{(n)}(t)$, with θ_0 belonging to an open interval θ of R . By (6.3) the log-likelihood function for the n -th model takes the form

$$\log L(\theta) = \int_0^1 \log(\alpha(s; \theta)) dN^{(n)}(s) - \int_0^1 \alpha(s; \theta) Y^{(n)}(s) ds, \quad (B.1)$$

and the maximum likelihood estimator $\hat{\theta}_n$ is defined as a solution to the equation

$$\frac{\partial \log L(\theta)}{\partial \theta} = \int_0^1 \frac{\frac{\partial}{\partial \theta} \alpha(s; \theta)}{\alpha(s; \theta)} dN^{(n)}(s) - \int_0^1 \frac{\partial}{\partial \theta} \alpha(s; \theta) Y^{(n)}(s) ds = 0. \quad (B.2)$$

We will here derive the properties of $\hat{\theta}_n$ under the following sufficient set of conditions. (For a general set of conditions, see Borgan (1984, Section 4).)

1) There exist a sequence of nonnegative constants a_n , increasing to infinity as $n \rightarrow \infty$, and a function y such that $Y^{(n)}/a_n^2$ converges uniformly on $[0, 1]$ to y in probability as $n \rightarrow \infty$.

2) There exists a neighbourhood θ_0 of θ_0 such that $\alpha(t; \theta)$ and its derivatives of first, second and third order w.r.t. θ exist and are continuous functions of $\theta \in \theta_0$. Moreover, they are bounded on $[0, 1] \times \theta_0$.

3) $\alpha(t; \theta)$ is bounded away from zero on $[0, 1] \times \theta_0$.

4) $\sigma^2(\theta_0) > 0$, where $\sigma^2(\theta) = \int_0^1 \frac{\frac{\partial^2}{\partial \theta^2} \alpha(s; \theta)}{\alpha(s; \theta)} y(s) ds$.

In all applications in this paper we will have $a_n = n^{\frac{1}{2}}$

We may prove:

Theorem With a probability tending to 1, the likelihood equation (B.2) has exactly one consistent solution $\hat{\theta}_n$ under Conditions 1-4. Moreover

$$a_n(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, [\sigma^2(\theta_0)]^{-1}),$$

where $\sigma^2(\theta_0)$ is defined in Condition 4.

Proof: By a Taylor series expansion we have for $\theta \in \theta_0$

$$\frac{1}{a_n^2} \frac{\partial \log L(\theta)}{\partial \theta} = A_n + B_n(\theta - \theta_0) + \frac{1}{2} C_n(\theta - \theta_0)^2. \quad (B.3)$$

Here

$$A_n = \frac{1}{a_n^2} \frac{\partial \log L(\theta_0)}{\partial \theta},$$

$$B_n = \frac{1}{a_n^2} \frac{\partial^2 \log L(\theta_0)}{\partial \theta^2},$$

and

$$C_n = \frac{1}{a_n^2} \frac{\partial^3 \log L(\theta_n^*)}{\partial \theta^3},$$

where θ_n^* is between θ and θ_0 . Let us study the behaviour of these three terms as $n \rightarrow \infty$.

By (6.5)

$$A_n = \frac{1}{a_n^2} \int_0^1 \frac{\frac{\partial}{\partial \theta} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} dM^{(n)}(s),$$

such that (3.17) and an application of Lenglart's inequality (Theorem 3.1) gives that for all $\delta, \eta > 0$ we have

$$P(|A_n| > \eta) < \frac{\delta}{\eta^2} + P\left(\left\langle \frac{1}{a_n^2} \int \frac{\frac{\partial}{\partial \theta} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} dM^{(n)}(s) \right\rangle (1) > \delta\right)$$

$$= \frac{\delta}{\eta^2} + P\left(\frac{1}{a_n^2} \int_0^1 \left[\frac{\frac{\partial}{\partial \theta} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} \right]^2 \alpha(s; \theta_0) \frac{Y^{(n)}(s)}{a_n^2} ds > \delta\right).$$

By Conditions 1-3 the last term on the right hand side converges to zero as $n \rightarrow \infty$, so we have that

$$A_n \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty. \quad (\text{B.4})$$

Using (3.14) and (B.2) we see that

$$B_n = - \int_0^1 \frac{\frac{\partial^2}{\partial \theta^2} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} \frac{Y^{(n)}(s)}{a_n^2} ds + \frac{1}{a_n^2} \int_0^1 \frac{\partial^2 \log(\alpha(s; \theta_0))}{\partial \theta^2} dM^{(n)}(s).$$

Here the second term converges in probability to zero by an argument similar to the one giving (B.4). Therefore Conditions 1-3 give that

$$B_n \xrightarrow{P} -\sigma^2(\theta_0) \quad \text{as } n \rightarrow \infty. \quad (\text{B.5})$$

Finally by Conditions 2 and 3

$$|C_n| < K_1 \frac{N^{(n)}(1)}{a_n^2} + K_2 \int_0^1 \frac{Y^{(n)}(s)}{a_n^2} ds$$

for some constants K_1 and K_2 not depending on θ . Another application of Lengart's inequality (Theorem 3.1) and Condition 1 therefore give that there exists a finite constant M not depending on θ such that

$$\lim_{n \rightarrow \infty} P(|C_n| < M) = 1. \quad (\text{B.6})$$

From (B.3) - (B.6) it follows as for the classical i.i.d. case that there exists a (weakly) consistent solution to the likelihood equation (B.2). (See e.g. Serfling (1980, pp. 147-148), with convergence almost surely replaced by convergence in probability.) It is shown in Billingsley (1961, pp.12-13) that if $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(2)}$ are two consistent solutions of (B.2), then the

probability that $\hat{\theta}_n^{(1)} = \hat{\theta}_n^{(2)}$ goes to one as $n \rightarrow \infty$, so that (B.2) has an essentially unique (weakly) consistent solution.

To prove the second assertion of the theorem, we use (B.2) and (B.3) to write

$$0 = \frac{1}{a_n^2} \frac{\partial \log L(\hat{\theta}_n)}{\partial \theta} = A_n + B_n(\hat{\theta}_n - \theta_0) + \frac{1}{2} C_n(\hat{\theta}_n - \theta_0)^2$$

when $\hat{\theta}_n \in \theta_0$. By this

$$a_n(\hat{\theta}_n - \theta_0) - \frac{-a_n A_n}{B_n + \frac{1}{2} C_n(\hat{\theta}_n - \theta_0)} \xrightarrow{P} 0$$

as $n \rightarrow \infty$. By (B.5), (B.6) and the consistency of

$\hat{\theta}_n$, $B_n + \frac{1}{2} C_n(\hat{\theta}_n - \theta_0) \xrightarrow{P} -\sigma^2(\theta_0)$ as $n \rightarrow \infty$, and it follows that $a_n(\hat{\theta}_n - \theta_0)$ has the same asymptotic distribution as $a_n A_n / \sigma^2(\theta_0)$.

Now

$$a_n A_n = \frac{1}{a_n} \int_0^1 \frac{\frac{\partial}{\partial \theta} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} dM^{(n)}(s)$$

is a stochastic integral with respect to a square integrable martingale. Since, by Conditions 1-3,

$$\begin{aligned} &< \frac{1}{a_n} \int \frac{\frac{\partial}{\partial \theta} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} dM^{(n)}(s) > \quad (1) \\ &= \int_0^1 \left[\frac{\frac{\partial}{\partial \theta} \alpha(s; \theta_0)}{\alpha(s; \theta_0)} \right]^2 \alpha(s; \theta_0) \frac{Y^{(n)}(s)}{a_n^2} ds \xrightarrow{P} \sigma^2(\theta_0), \end{aligned}$$

as $n \rightarrow \infty$, an application of the martingale central limit theorem (Theorem 3.2) gives that $a_n A_n \xrightarrow{D} N(0, \sigma^2(\theta_0))$, and the theorem is proved. Note that, except for the application of the martingale central limit theorem, the proof of the asymptotic normality is exactly as for the classical i.i.d. case (e.g. Serfling, 1980, p.148). \square

It is shown in Borgan (1984) that $\sigma^2(\theta_0)$ may be estimated consistently by $-a_n^{-2} \partial^2 \log L(\hat{\theta}_n) / \partial \theta^2$. The proof of this result uses (B.5) and the fact that $\hat{\theta}_n$ is consistent for θ_0 .

REFERENCES

- Aalen, O.O. (1975). Statistical inference for a family of counting processes. Ph.D. Dissertation, Dept. of Stat., Univ. of Calif. Berkeley.
- Aalen, O.O. (1976). Nonparametric inference in connection with multiple decrement models. Scand. J. Statist. 3, 15-27.
- Aalen, O.O. (1977). Weak convergence of stochastic integrals related to counting processes. Z. Wahr. verw. Geb. 38, 261-277. Correction (1979) 48, 347.
- Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. Ann. Statist. 6, 701-726.
- Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. Springer Lect. Notes in Statist. 2, 1-25.
- Aalen, O.O. (1982a). Practical applications of the nonparametric statistical theory for counting processes. Statistical Research Report 1982, No. 2. Institute of Mathematics, University of Oslo.
- Aalen, O.O. (1982b). Contribution to the discussion of a paper by Andersen et al. Int. Statist. Rev. 50, 219-258.
- Aalen, O.O., Borgan, Ø., Keiding, N. & Thormann, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring. Scand. J. Statist. 7, 161-171.
- Aalen, O.O. & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand. J. Statist. 5, 141-150.
- Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. Math. Biosc. 6, 1-11.
- Andersen, P.K. (1982). On the application of the theory of counting processes in the statistical analysis of censored survival data. Pp. 1-13 in J. Crowley and R.A. Johnson (eds.) Survival analysis. IMS Lecture Notes-Monographs 2.
- Andersen, P.K. (1983a). Comparing survival distributions via hazard ratio estimates. Scand. J. Statist. 10, 77-85.
- Andersen, P.K. (1983b). Testing for proportional hazards. Research Report 83/1. Statistical Research Unit, Copenhagen.
- Andersen, P.K. (1984). A Cox regression model for the excess mortality in long term follow-up studies. Proceedings form XII Int. Biometric Conf., Tokyo, Japan, pp. 157-162.
- Andersen, P.K. (1985). Statistical models for longitudinal labour market data based on counting processes. In: Longitudinal Analysis of Labour Market Data (Ed. J.J. Heckman & B. Singer), Cambridge University Press, New York.

- Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). Int. Statist. Rev. 50, 219-258.
- Andersen, P.K. & Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. Ann. Statist. 10, 1100-1120.
- Andersen, P.K. & 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 82/6. Statistical Research Unit, Copenhagen.
- Bean, S.J. & Tsokos, C.P. (1980). Developments in non-parametric density estimation. Int. Statist. Rev. 48, 267-287.
- Becker, N. (1977). Estimation for discrete time branching processes with application to epidemics. Biometrics 33, 515-522.
- Becker, N. (1981). The infectiousness of a disease within households. Biometrika 68, 133-141.
- Becker, N. & Hopper, J. (1983). The infectiousness of a disease in a community of households. Biometrika 70 29-39.
- Begun, J. & Reid, N. (1983). Estimating the relative risk with censored data. J. Amer. Statist. Assoc. 78, 337-341.
- Benjamin, B. (1978). Graunt, John. Pp. 435-437 in International Encyclopedia of Statistics, Vol. 1. Eds. W.H. Kruskal and J.M. Tanur. Free Press, Macmillan Publishing Co., New York.
- Billingsley, P. (1961). Statistical inference for Markov processes. The University of Chicago Press, Chicago.
- Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.
- Bøhmer, P.E. (1912). Theorie der unabhängigen Wahrscheinlichkeiten. Rapp., Mém. et Procès-Verbaux 7^e Congr. Internat. Act., Amsterdam 2, 327-346.
- Borgan, Ø. (1979). On the theory of moving average graduation. Scand. Actuarial J. 83-105.
- Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. Scand. J. Statist. 11, 1-16.

- Borgan, Ø. & Gill, R.D. (1982). Case-control studies in a Markov chain setting. Preprint SW 89/82, Mathematical Centre, Amsterdam.
- Borgan, Ø., Liestøl, K. & Ebbesen, P. (1984). Efficiencies of experimental designs for an illness-death model. Biometrics, to appear.
- Borgan, Ø., Menné, T. & Green, A. (1982). Interaction between nickel allergy and hand eczema in the Danish female population. Scand. J. Statist. 9, 183-185.
- Borgan, Ø. & Ramlau-Hansen, H. (1983). Demographic incidence rates and estimation of intensities with incomplete information. Stat. Res. Report. 2/83. Inst. of Math., Univ. of Oslo.
- Brémaud, P. (1981). Point processes and queues. Martingale dynamics. Springer-Verlag, New York, Heidelberg, Berlin.
- Brémaud, P. & Jacod, J. (1977). Processus ponctuels et martingales: Résultats récents sur la modélisation et filtrage. Adv. Appl. Prob. 9, 362-416.
- 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.E. (1975). Analysis of survival data under the proportional hazards model. Int. Statist. Rev. 43, 45-58.
- Breslow, N.E. & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2, 437-453.
- Breslow, N.E. Lubin, J.H., Marek, P. & Langholz, B. (1983). Multiplicative models and cohort analysis. J. Amer. Statist. Assoc. 78, 1-12.
- Chiang, C.L. (1968). Introduction to stochastic processes in biostatistics. Wiley, New York.
- Cox, D.R. (1972). Regression models and life tables (with discussion). J. Roy. Statist. Soc. B. 34, 187-220.
- Cox, D.R. (1975). Partial likelihood. Biometrika 62, 269-276.
- Cramér, H. (1945). Mathematical methods of statistics. Almqvist & Wiksell, Uppsala.
- The Copenhagen Study Group for Liver Diseases (1974). Sex, ascites and alcoholism in survival of patients with cirrhosis. New England J. of Medicine 291, 271-273.
- Csörgö, S. & Horvath, L. (1982). On cumulative hazard processes under random censorship. Scand. J. Statist. 9, 13-21.
- Dolivo, F.G. (1974). Counting processes and integrated conditional rates: a martingale approach with application to detection theory. Ph.D. thesis, University of Michigan.

- Epstein, B. & Sobel, M. (1953). Life testing. J. Amer. Statist. Assoc. **48**, 486-502.
- Fix, E. & Neyman, J. (1951). A simple stochastic model of recovery relapse, death and loss of patients. Human Biol. **23**, 205-241.
- Fleming, T.R. & Harrington, D.P. (1981). A class of hypothesis tests for one and two sample censored survival data. Commun. Statist.-Theor. Meth. **A10**, 763-794.
- Fleming, T.R., O'Fallon, J.R., O'Brien, P.C. & Harrington, D.P. (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right censored data; Biometrics **36**, 607-625.
- Freund, J.E. (1961). A bivariate extension of the exponential distribution. J. Amer. Statist. Assoc. **56**, 971-977.
- Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika **52**, 203-223.
- Gehan, E.A. & Thomas, D.G. (1969). The performance of some two-sample tests in small samples with and without censoring. Biometrika **56**, 127-132.
- Gill, R.D. (1980a). Censoring and stochastic integrals. Math. Centre Tracts **124**, Mathematical Centre, Amsterdam.
- Gill, R.D. (1980b). Nonparametric estimation based on censored observations of a Markov renewal process. Z. Wahr. verw. Geb. **53**, 97-116.
- Gill, R.D. (1981). Testing with replacement and the product limit estimator. Ann. Statist. **9**, 853-860.
- Gill, R.D. (1983a). Large sample behaviour of the product-limit estimator on the whole line. Ann. Statist. **11**, 49-58.
- Gill, R.D. (1983b). Contribution to the discussion of papers by Helland and Kurtz. 44th Session of the Int. Stat. Inst., Madrid, 1983.
- Gill, R.D. (1984). Understanding Cox's regression model: a martingale approach. J. Amer. Statist. Assoc. **79**, 441-447.
- Gill, R.D. & Schumacher, M. (1984). A simple test of the proportional hazards assumption. Preprint. Univ. Heidelberg.
- Glass, D.V. (1950). Graunt's life table. J. Inst. Actuaries **76**, 60-64.
- Green, A., Hauge, M., Holm, N.V. & Rasch, L.L. (1981). Epidemiological studies of Diabetes Mellitus in Denmark: II. A prevalence study based on insulin prescriptions. Diabetologia **20**, 468-470.

- Hall, W.J. & Wellner, J.A. (1980). Confidence bands for a survival curve with censored data. Biometrika 67, 133-143.
- Halperin, M. (1952). Maximum likelihood estimation in truncated samples. Ann. Math. Stat. 23, 226-238.
- Harrington, D.P. & Fleming, T.R. (1982). A class of rank test procedures for censored survival data. Biometrika 69, 533-546.
- Helland, I.S. (1982). Central limit theorems for martingales with discrete or continuous time. Scand. J. Statist. 9, 79-94.
- Helland, I.S. (1983). Applications of central limit theorems for martingales with continuous time. Proc. 44th session of the Int. Stat. Inst., Madrid, 1983 1, 346-360.
- Hoem, J.M. (1972). On the statistical theory of analytic graduation. Proc. Sixth Berkeley Symp. 1, 569-600.
- Hoem, J.M. (1976). The statistical theory of demographic rates (with discussion). Scand. J. Statist. 3, 169-185.
- Hoem, J.M. (1977). A Markov chain model for working life tables. Scand. Actuarial J., 1-20.
- Hyde, J. (1977). Testing survival under right censoring a left truncation. Biometrika 64, 225-230.
- Jacobsen, M. (1982). Statistical Analysis of Counting Processes. Lecture Notes in Statistics 12, Springer-Verlag, Berlin.
- Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model: a survey. Int. Statist. Rev. 52, 193-207.
- Jacod, J. (1979). Calcul stochastique et problèmes de martingales. Lecture Notes in Mathematics 714, Springer-Verlag, Berlin.
- Johansen, S. (1983). An extension of Cox's regression model. Int. Statist. Rev. 51, 258-262.
- Kalbfleisch, J.D. & Prentice, R.L. (1980). The statistical analysis of failure time data. Wiley, New York.
- Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.
- Karup, I. (1893). Die Finanzlage der Gothaischen Staatsdiener-Wittwen-Societät, Dresden.
- Latta, R.B. (1981). A Monte Carlo study of some two-sample rank tests with censored data. J. Amer. Statist. Assoc. 76, 713-719.

- Lawless, J.F. (1983). Statistical methods in reliability (with discussion). Technometrics 25, 305-335.
- Lee, E.T., Desu, M.M. & Gehan, E.A. (1975). A Monte Carlo study of the power of some two-sample tests. Biometrika 62, 425-432.
- Lee, K.L., Harrell, F.E., Tolley, H.D. & Rosati, R.A. (1983). A comparison of test statistics for assessing the effects of concomitant variables in survival analysis. Biometrics 39, 341-350.
- Lenglart, E. (1977). Relation de domination entre deux processus. Ann. Inst. Henri Poincaré 13, 171-179.
- Liptser, R. Sh. & Shiriyayev, A.N. (1980). A functional limit theorem for semimartingales. Theor. Probab. Appl. 25, 667-688.
- Lustbader, E.L. (1980). Time-dependent covariates in survival analysis. Biometrika 67, 697-698.
- Meyer, P.A. (1976). Un cours sur les intégrales stochastiques. In Lecture Notes in Mathematics 511, 245-400. Springer-Verlag, Berlin.
- Miller, R.G. Jr. (1981). Survival analysis. Wiley, New York.
- Nelson, W. (1969). Hazard plotting for incomplete failure data. J. Qual. Technol. 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. Scand. J. Statist. 9, 107-115.
- Oakes, D. (1981). Survival times: aspects of partial likelihood (with discussion). Int. Statist. Rev. 49, 235-264.
- Peace, K.E. & Flora, R.E. (1978). Size and power assessment of tests of hypotheses on survival parameters. J. Amer. Statist. Assoc. 73, 129-132.
- Peto, R. & Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). J. Roy. Statist. Soc. Ser. A 135, 185-206.
- Prentice, R. L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-179.
- Prentice, R.L. & Breslow, N.E. (1978). Retrospective studies and failure time models. Biometrika 65, 153-158.

- Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V. Jr., Flournoy, N.S. Farewell, V.T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. Biometrics 34, 541-554.
- Prentice, R.L. & Self, S.G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. Ann. Statist. 11, 804-813.
- Ramlau-Hansen, H. (1981). Udglatning med kernefunktioner i forbindelse med tælleprocesser (in Danish). Working Paper no. 41 Lab. Actuarial Math., Univ. Copenhagen.
- Ramlau-Hansen, H. (1983a). Smoothing counting process intensities by means of kernel functions. Ann.Statist. 11, 453-466.
- Ramlau-Hansen, H. (1983b). The choice of a kernel function in the graduation of counting process intensities. Scand. Actuarial J., 165-182.
- Rebolledo, R. (1978). Sur les applications de la théorie des martingales a l'étude statistique d'une famille de processus ponctuels. Lecture Notes in Mathematics 636, 27-70, Springer-Verlag, Berlin.
- Rebolledo, R. (1980). Central limit theorems for local martingales. Z. Wahrsch. verw. Gebiete 51, 269-286.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. Scand. J. Statist. 9, 65-78.
- Schlichting, P., Christensen, E., Andersen, P.K., Fauerholdt, L., Juul, E., Poulsen, H., Tygstrup, N. & The Copenhagen Study Group for Liver Diseases (1983). Identification of prognostic factors in cirrhosis using Cox's regression model. Hepatology 3, 889-895.
- Schou, G. & Vath, M. (1980). A small sample study of occurrence/exposure rates for rare events. Scand. Actuarial J., 209-225.
- Sellke, T. & Siegmund, D. (1983). Sequential analysis of the proportional hazards model. Biometrika 70, 315-326.
- Serfling, R.J. (1980). Approximation theorems of mathematical statistics. Wiley, New York.
- Shiryayev, A.N. (1981). Martingales: Recent developments, results and applications. Int. Statist. Rev. 49, 199-234.
- Slud, E.V. (1984). Sequential linear rank tests for two-sample censored survival data. Ann.Statist. 12, 551-571.
- Sverdrup, E. (1965). Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. Skand. Aktuar. Tidssk. 52, 184-211.

- Tarone, R.E. & Ware, J. (1977). On distribution-free tests for equality of survival distributions. Biometrika 64, 156-160.
- Thomas, D.C. (1977). Addendum to the paper "Methods of cohort analysis: Appraisal by application to asbestos mining" by F.D.K. Liddell, J.C. McDonald & D.C. Thomas. J. Roy. Statist. Soc. A 140, 469-491.
- Tsiatis, A.A. (1981). A large sample study of Cox's regression model. Ann. Statist. 9, 93-108.
- Tuma, N.B., Hannan, M.T. & Groeneveld, L.P. (1979). Dynamic analysis of event histories. Amer. J. Sociol. 84, 820-854.
- Voelkel, J.G. & Crowley, J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations Ann. Statist. 12, 142-160.
- Walsh, J.E. (1962). Handbook of nonparametric Statistics. Vol I: Investigation of randomness, moments, percentiles and distributions. Van Nostrand, Princeton.
- Wei, L.J. (1983). Testing goodness of fit for proportional hazards model with censored observations. Preprint, Dept. of Statistics, George Washington Univ.