Three contributions to the Encyclopedia of Biostatistics: The Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen estimators

Ørnulf Borgan
Institute of Mathematics, University of Oslo,
P.O. Box 1053 Blindern, N-0316 Oslo, Norway

May, 1997

Abstract

This report contains three contributions to the Encyclopedia of Biostatistics. The Encyclopedia will be published by John Wiley and Sons Ltd in May 1998. The contributions in this report are

- The Nelson-Aalen estimator, page 2
- The Kaplan-Meier estimator, page 9
- The Aalen-Johansen estimator, page 19

NELSON-AALEN ESTIMATOR

Contribution to the Encyclopedia of Biostatistics

Ørnulf Borgan, University of Oslo

The Nelson-Aalen estimator is a nonparametric estimator which may be used to estimate the cumulative hazard rate function from censored survival data. Since no distributional assumptions are needed, one important use of the estimator is to check graphically the fit of parametric models, and this is the reason why it was originally introduced by Nelson (1969, 1972). Independently of Nelson, Altshuler (1970) derived the same estimator in the context of competing risks animal experiments. Later, by adopting a counting process formulation, Aalen (1978) extended its use beyond the survival data and competing risks set-ups and studied its small and large sample properties using martingale methods. The estimator is nowadays denoted the Nelson-Aalen estimator, although other names (the Nelson estimator, the Altshuler estimator, the Aalen-Nelson estimator, the empirical cumulative hazard estimator) are sometimes used as well. Below we present a number of situations where the Nelson-Aalen estimator may be applied and exemplify its use in one particular case. Further we indicate how counting processes provide a framework which allows for a unified treatment of all these diverse situations, and we summarize the most important properties of the Nelson-Aalen estimator. A detailed account is given in the monograph by Andersen et al. (1993, Section IV.1).

Survival data

Consider first the survival data situation where we want to study the time to death (or some other event) for a homogeneous population with hazard rate function $\alpha(t)$ and cumulative hazard rate function $A(t) = \int_0^t \alpha(s)\mathrm{d}s$. Assume that we have a sample of n individuals from this population. Our observation of the survival times for these individuals will typically be subject to right censoring, meaning that for some individuals we only know that their true survival times exceed certain censoring times. The censoring is assumed to be independent in the sense that the additional knowledge of censorings before any time t does not alter the risk of failure at t. We denote by $t_1 < t_2 < \cdots$ the times when deaths are observed and let d_j be the number of individuals who die at t_j.

The Nelson-Aalen estimator for the cumulative hazard rate function then takes the form

$$\widehat{A}(t) = \sum_{t_j \le t} d_j / r_j, \tag{1}$$

where r_j is the number of individuals at risk (i.e. alive and not censored) just prior to time t_j. Thus the Nelson-Aalen estimator is an increasing right-continuous step-function with increments d_j/r_j at the observed failure times. The variance of the Nelson-Aalen estimator may be estimated by

$$\widehat{\sigma}^2(t) = \sum_{t_j \le t} \frac{(r_j - d_j)d_j}{(r_j - 1)r_j^2}. \tag{2}$$

It may be shown (cf. below) that the Nelson-Aalen estimator (1) as well as the variance estimator (2) are almost unbiased. In large samples the Nelson-Aalen estimator, evaluated

at a given time t, is approximately normally distributed so that a standard $100(1-\alpha)\%$ confidence interval for $A(t)$ takes the form

$$\widehat{A}(t) \pm z_{1-\alpha/2}\widehat{\sigma}(t), \tag{3}$$

with $z_{1-\alpha/2}$ the $1-\alpha/2$ fractile of the standard normal distribution. The approximation to the normal distribution is improved by using a log-transform giving the confidence interval

$$\widehat{A}(t)\exp\left[\pm z_{1-\alpha/2}\widehat{\sigma}(t)/\widehat{A}(t)\right]. \tag{4}$$

This interval is satisfactory for quite small sample sizes (Bie et al., 1987).

Right censoring is not the only kind of data-incompleteness in survival analysis. Often, e.g. in epidemiological applications, individuals are not followed from time 0 (in the relevant time scale, typically age), but only from a later entry time (conditional on survival until this entry time). Thus, in addition to right censoring, the survival data are subject to left truncation. For such data we may still use the Nelson-Aalen estimator (1) and estimate its variance by (2). The number at risk r_j now is the number of individuals who have entered the study before time t_j and are still in the study just prior to t_j. For left truncated data the numbers at risk r_j may be low for small values of t_j. This will result in estimates $\widehat{A}(t)$ which have large sampling errors. But as the increments of the Nelson-Aalen estimator are uncorrelated (cf. below), the uncertainty induced for small time values have no influence when considering the increment $\widehat{A}(t)-\widehat{A}(s)$ of the Nelson-Aalen estimator over a later time interval $(s,t]$. An estimator for the variance of this increment is $\widehat{\sigma}^2(t)-\widehat{\sigma}^2(s)$.

Quite often we want to estimate the survival distribution function $S(t)=\exp[-A(t)]$ representing the probability that an individual will be alive at time t. This may be done from right censored and/or left-truncated survival data by the Kaplan-Meier estimator. The relation $A(t)=-\ln S(t)$ suggests that the cumulative hazard rate function alternatively may be estimated as minus the logarithm of the Kaplan-Meier estimator. Even though this estimator numerically will be close to the Nelson-Aalen estimator, the latter is the canonical one from a theoretical point of view (see entry on KAPLAN-MEIER ESTIMATOR). Further the Nelson-Aalen estimator may be used in a number of different situations (cf. below) while the alternative estimator only applies to the survival data situation.

An illustration

To give an illustration of the Nelson-Aalen estimator we use data from a randomized clinical trial for patients with histologically verified liver cirrhosis. Patients were recruited from several hospitals in Copenhagen between 1962 and 1969 and were followed until death, lost to follow-up or until the closing date of the study 1 October 1974. The time variable of interest is time since entry into the study. Patients are right censored if alive on 1 October 1974 or if lost to follow-up before that date.

We shall only consider the 138 placebo-treated male patients. Their median age at entry was 57 years, while the lower and upper quartiles were 51 and 66 years, respectively. Of the 138 patients 88 died during the study. The Nelson-Aalen estimate for these patients is shown in Figure 1 with 95% confidence intervals computed according to (4). Even though

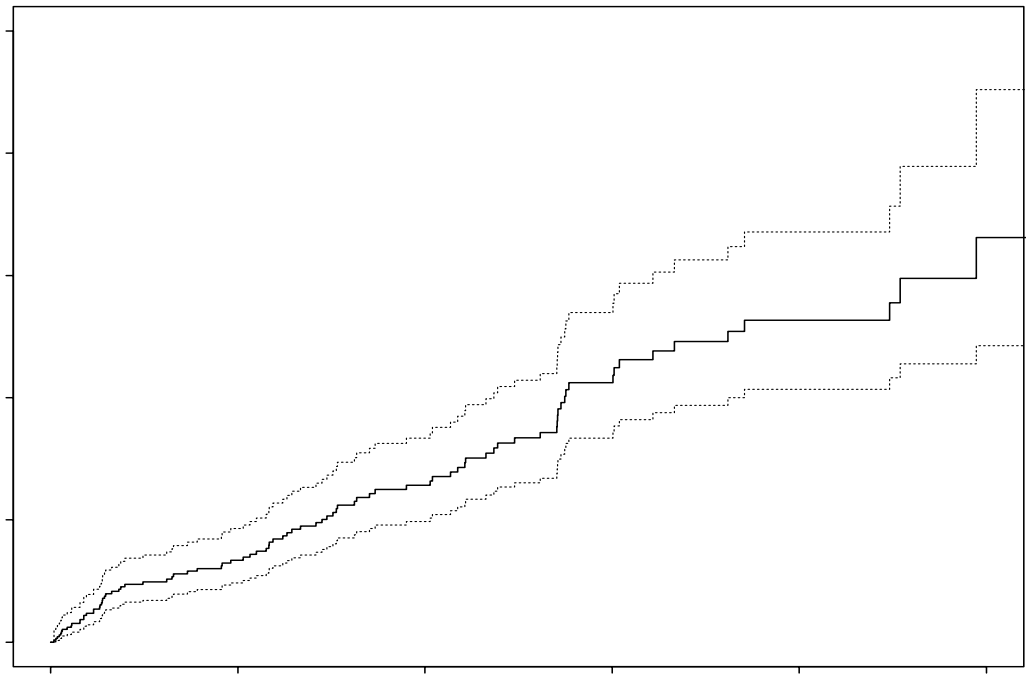

Figure 1: Nelson-Aalen estimate of the cumulative hazard rate function for death for 138 placebo-treated male patients with liver cirrhosis, with 95% log-transformed confidence intervals.

the cumulative hazard rate function provides a useful summary measure (e.g. Breslow and Day, 1980, Section 2.3), it is usually the hazard rate function itself which is the entity of real interest. So when interpreting the estimate in Figure 1, we mainly focus on the "slope" of the curve. The estimate of the cumulative hazard rate function is steeper the first 9–10 months after randomization than at later times. Therefore we have evidence that the risk of dying for these patients is highest just after randomization. (This may, at least in part, be due to heterogeneity which is not accounted for in our simple analysis.) The hazard rate function is approximately 0.3 per year the first 9–10 months and slightly below 0.2 per year thereafter when estimated as the average slope of the curve over the relevant time periods. More formal procedures for smoothing the Nelson-Aalen estimate in order to get an estimate for the hazard rate function itself are available but will not be considered here. A further discussion and analysis of the cirrhosis data are given by Schlichting et al. (1983). The data were also used for illustrative purposes by Andersen et al. (1993).

Multi-state models and recurrent events

The survival analysis set-up considered above may be generalized in two directions. More than one type of events may be considered for each individual under study, and/or the event in question may happen more than once for each individual. Examples of the first type are competing risks with two or more causes of death and the Markov illness-death model with the states "healthy," "diseased" and "dead." More generally we may consider any Markov process with a finite number of states which may be used to model the life-history of an individual. An example of the second type is an inhomogeneous Poisson process with intensity $\alpha(t)$ modeling the occurrence of some recurrent event like episodes of angina pectoris in patients with coronary heart disease or infections in AIDS patients. For both of these two types of situations we observe the times when events occur for a number of individuals (modeled as i.i.d. copies of the relevant process) who do not all need to be observed over the same interval of time. The Nelson-Aalen estimator may then be applied to estimate cumulative intensities.

To be specific, consider a finite-state Markov process with transition intensities $\alpha_{gh}(t)$ for $g \neq h$. Focusing on fixed g and h in the following, we drop the subscripts and write just $\alpha(t)$ for the $g \to h$ transition intensity. Further denote by $t_1 < t_2 < \cdots$ the times when transitions from g to h are observed, let d_j be the number of individuals who experience a $g \to h$ transition at t_j, and write r_j for the number of individuals in state g (i.e. at risk for a $g \to h$ transition) just prior to time t_j. Then the cumulative $g \to h$ transition intensity $A(t) = \int_0^t \alpha(s)\mathrm{d}s$ may be estimated by (1) and its variance by (2). Similarly the integrated intensity of an inhomogeneous Poisson process may be estimated with the t_js denoting the times of observed events, and the d_js and r_js being the corresponding numbers of events and numbers at risk, respectively. An illustration of the use of the Nelson-Aalen estimator to estimate integrated Markov transition intensities is given by Keiding and Andersen (1989).

Two other applications

For the situations considered so far, (1) and (2) apply with r_j the number at risk at t_j for the event in question. The use of the Nelson-Aalen estimator is, however, not restricted to such situations. We here mention two other applications and return to a general discussion below.

Relative mortality. Our first example considers right censored and/or left truncated survival data, but they no longer come from a homogeneous population. Rather we assume that the hazard rate function of the ith individual may be written as the product $\alpha(t)\mu_i(t)$, where $\alpha(t)$ is a relative mortality common to all individuals and $\mu_i(t)$ is the hazard rate function at time t for a person from an external standard population corresponding to the ith individual (e.g. of the same sex and age as individual i). Typically the $\mu_i(t)$ will be known from published life tables for the general population. In this situation the Nelson-Aalen estimator may be used to estimate the cumulative relative mortality $A(t) = \int_0^t \alpha(s)\mathrm{d}s$. All that is required is that r_j in (1) is taken to denote the sum of the external rates $\mu_i(t_j)$ for all individuals at risk just prior to t_j. An illustration of this use of the Nelson-Aalen estimator is provided by Breslow and Day (1987, Chapter 5).

An epidemic model. A simple model for the spread of an infectious disease in a community is the following. At the start of the epidemic, i.e. at time $t = 0$, some individuals make

contact with individuals from elsewhere and are thereby infected with the disease. There are no further infections from outside the community during the course of the epidemic. Let $S(t)$ and $I(t)$ denote the number of susceptibles and infectives, respectively, just prior to time t. Assuming random mixing, the infection intensity in the community at time t becomes $\alpha(t)S(t)I(t)$, where $\alpha(t)$ is the infection rate per possible contact. We denote by $0 < t_1 < t_2 < \cdots$ the times when individuals are infected and let d_j denote the number infected at t_j. Then the cumulative infection rate $A(t) = \int_0^t \alpha(s)\mathrm{d}s$ may be estimated by the Nelson-Aalen estimator (1) where now $r_j = S(t_j)I(t_j)$; see Becker (1989, Section 7.6) for an illustration.

Counting process formulation and small sample properties

In general we consider the occurrences of some events of interest (e.g. deaths, occurrences of a disease, infections), and denote by $0 < t_1 < t_2 < \cdots$ the times when an event is observed. We assume that two or more events cannot occur at the same time, so that there are no tied observations. (The handling of ties is discussed briefly below.) Then the process $N(t)$ counting the number of observed events in the time-interval $[0, t]$ is a (univariate) counting process. The behaviour of $N(t)$ is governed by its intensity process $\lambda(t)$ given heuristically by

$$\lambda(t)\mathrm{d}t = \Pr(\text{event occurs in } [t, t+\mathrm{d}t) \mid \mathcal{F}_{t-}).$$

Here $\mathcal{F}_{t-}$ represents all the information available to the researcher just before time t. The counting process satisfies Aalen's multiplicative intensity model if we may write its intensity process as

$$\lambda(t) = \alpha(t)Y(t), \tag{5}$$

for some unknown function $\alpha(t)$ and some observable process $Y(t)$ whose value at time t is known just prior to t. All the situations considered above give counting processes which fulfill (5). Survival data from a homogeneous population, finite-state Markov processes and the inhomogeneous Poisson process, all give an $Y(t)$ process which is the number at risk just prior to time t. For the model for relative mortality $Y(t)$ is the sum of the $\mu_i(t)$ for those at risk just before t, while for the epidemic model $Y(t) = S(t)I(t)$. The common structure of all these models when formulated as counting processes is the reason why the Nelson-Aalen estimator may be applied to all these diverse problems.

In fact the counting process formulation provides a framework which makes it simple to study the statistical properties of the Nelson-Aalen estimator. We will briefly indicate a few main steps and refer to Andersen et al. (1993, Section IV.1.1) for a thorough treatment. First we note that, with $r_j = Y(t_j)$, we may write the Nelson-Aalen estimator (1) as

$$\widehat{A}(t) = \int_0^t \frac{J(s)}{Y(s)}\mathrm{d}N(s), \tag{6}$$

where $J(s) = I(Y(s) > 0)$ and $0/0$ is interpreted as 0. Then using (5), (6) and the decomposition $N(t) = \int_0^t \lambda(s)\mathrm{d}s + M(t)$ of a counting process into a sum of its integrated intensity process and a local square integrable martingale $M(t)$, we get

$$\widehat{A}(t) = A^*(t) + M^*(t). \tag{7}$$

Here $A^*(t) = \int_0^t J(s)\alpha(s)\mathrm{d}s$ is almost the same as $A(t)$ when there is only a small probability that $Y(s) = 0$ for some $s \leq t$, while $M^*(t) = \int_0^t [J(s)/Y(s)]\mathrm{d}M(s)$ is a stochastic integral and as such a local square integrable martingale. The relation (7) is key to the study of the statistical properties of the Nelson-Aalen estimator. Since $M^*(t)$ has expected value zero for any given t, we have $\mathrm{E}\widehat{A}(t) = \mathrm{E}A^*(t)$ so the Nelson-Aalen estimator is almost unbiased. Further an unbiased estimator for the variance of $M^*(t)$ is its optional variation process $\int_0^t [J(s)/Y(s)^2]\mathrm{d}N(s)$. Thus the variance estimator (2) is almost unbiased when there are no ties. Finally a martingale has uncorrelated increments, and by (7) this is (almost) the case for the Nelson-Aalen estimator as well.

In the presence of ties, i.e. when the number of events d_j at t_j may exceed one, the process $N(t)$ counting occurrences of events in $[0, t]$ may have jumps of size two or larger and is therefore no longer a counting process. Often, however, we may write $N(t) = \sum_{i=1}^n N_i(t)$, where $N_i(t)$ is a counting process registering the events for individual i. If we consider a homogeneous population where the rates of occurrence of the events are the same for all individuals, we may adopt the discrete extension of the model described by Andersen et al. (1993, pp. 180-181). For this extended model, the arguments of Fleming and Harrington (1991, pp. 94-96) apply to show that the variance estimator (2) is almost unbiased also in the presence of ties. This justifies the use of the tie-corrected estimator (2) for all situations considered above except for the model with relative mortality and the epidemic model. Within the framework of the extended model the Nelson-Aalen estimator is a nonparametric maximum likelihood estimator; see Andersen et al. (1993, Section IV.1.5) for details and a further discussion.

Weak convergence and confidence bands

By (7) the martingale central limit theorem may be used to prove that, considered as a stochastic process, the Nelson-Aalen estimator (properly normalized) converges weakly to a mean zero Gaussian martingale. In particular for a fixed t it is asymptotically normally distributed, a fact which was used in connection with the confidence intervals (3) and (4). The weak convergence result also makes it possible to derive confidence bands for A, i.e. limits which contain $A(t)$ for all t in an interval $[\tau_1, \tau_2]$ with a prespecified probability.

One important class of such confidence bands are the equal precision bands. The standard and log-transformed equal precision bands are obtained by replacing $z_{1-\alpha/2}$ in (3) and (4) by $d_{1-\alpha}$, the $1-\alpha$ fractile in the distribution of the supremum of the absolute value of a standardized Brownian bridge (over a certain time-interval). This fractile may be found (approximately) by solving (w.r.t. d) the non-linear equation

$$4\phi(d)/d + 2\phi(d)(d - 1/d)\ln[\widehat{\sigma}(\tau_2)/\widehat{\sigma}(\tau_1)] = \alpha,$$

where $\phi(d)$ is the standard normal density function. The equal precision bands require $\widehat{\sigma}(\tau_1) > 0$, so they cannot be extended all the way down to $t = 0$. Typically one will also omit the largest values of t. The standard equal precision band has quite bad small sample properties, so even with sample sizes in the hundreds the use of the log-transformed confidence band is recommended (Bie et al., 1987). As an illustration we use once more the liver cirrhosis example. Considering the interval from 4 months (1/3 year) to 8 years, we have

$\widehat{\sigma}(1/3) = 0.027$ and $\widehat{\sigma}(8) = 0.163$ so that $d_{0.95} = 2.99$. Therefore the 95% log-transformed equal precision band for the cumulative hazard rate function between 4 months and 8 years may be obtained from (4) by using the fractile 2.99 instedad of the value 1.96 used for the pointwise confidence intervals in Figure 1. A detailed study of the weak convergence of the Nelson-Aalen estimator and the derivation of confidence bands are provided by Andersen et al. (1993, Section IV.1.2-3). Here another class of confidence bands, the Hall-Wellner bands, is also discussed.

We finally note that semi-Markov processes (or Markov renewal processes), where the transition intensities (only) depend on the sojourn times in the states, do not give rise to counting processes which fulfill the multiplicative intensity model (5). Thus the results outlined above do not immediately extend to such models. However, it turns out that enough of the above structure is preserved to be able to define Nelson-Aalen estimators also for such semi-Markov processes and to derive identical asymptotic results for these as for the case of Markov processes; see Andersen et al. (1993, Section X.1) for a discussion and further references.

References

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

Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. *Mathematical Biosciences* **6**, 1–11.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer Verlag, New York.

Bie, O., Borgan, Ø., and Liestøl, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scandinavian Journal of Statistics* **14**, 221–233.

Becker, N.G (1993). *Analysis of Infectious Disease Data.* Chapman and Hall, London.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research.* Volume 1 – *The Analysis of Case-Control Studies*, *IARC Scientific Publications*, Vol. 32. International Agency for Research on Cancer, Lyon.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research.* Volume 2 – *The Design and Analysis of Cohort Studies*, *IARC Scientific Publications*, Vol. 82. International Agency for Research on Cancer, Lyon.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Keiding, N. and Andersen, P. K. (1989). Nonparametric estimation of transition intensities and transition probabilities: a case study of a two-state Markov process. *Applied Statistics* **38**, 319–329.

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

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945–965.

Schlichting, P., Christensen, E., Andersen, P. K., Fauerholdt, L., Juhl, E., Poulsen, H., and Tygstrup, N., for The Copenhagen Study Group for Liver Diseases. (1983). Prognostic factors in cirrhosis identified by Cox's regression model. *Hepatology* **3**, 889–895.

KAPLAN-MEIER ESTIMATOR

Contribution to the Encyclopedia of Biostatistics

Ørnulf Borgan, University of Oslo

The Kaplan-Meier estimator is a nonparametric estimator which may be used to estimate the survival distribution function from censored data. The estimator may be obtained as the limiting case of the classical actuarial estimator, and it seems to have been first proposed by Böhmer (1912). It was, however, lost sight of by later researchers and not further investigated until the important paper by Kaplan and Meier (1958) appeared. The estimator is today usually named after these two authors, although it is sometimes denoted the product-limit estimator. Below we describe the Kaplan-Meier estimator, illustrate its use in one particular case, and discuss estimation of median and mean survival time. Further we show how the Kaplan-Meier estimator can be given as the product-integral of the Nelson-Aalen estimator, and indicate how this may be used to study its statistical properties. The Kaplan-Meier estimator has for almost four decades been one of the key statistical methods for analyzing censored survival data, and it is discussed in most textbooks on survival analysis. Rigorous derivations of the statistical properties of the estimator are provided in the books by Fleming and Harrington (1991) and Andersen et al. (1993). In particular the latter presents formal proofs of almost all the results reviewed below as well as an extensive bibliography.

The estimator and confidence intervals

Consider the survival data situation where we want to study the time to death (or some other event) for a homogeneous population with survival distribution function $S(t)$ representing the probability that an individual will be alive at time t. Assume that we have a sample of n individuals from this population. Our observation of the survival times for these individuals will typically be subject to right censoring, meaning that for some individuals we only know that their true survival times exceed certain censoring times. The censoring is assumed to be independent in the sense that the additional knowledge of censorings before any time t does not alter the risk of failure at t. We denote by $t_1 < t_2 < \cdots$ the times when deaths are observed and let d_j be the number of individuals who die at t_j.

The Kaplan-Meier estimator for the survival distribution function then takes the form

$$\widehat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{r_j}\right), \tag{8}$$

where r_j is the number of individuals at risk (i.e. alive and not censored) just prior to time t_j. If there are no censored observations, (8) reduces to one minus the empirical distribution function. The variance of the Kaplan-Meier estimator is estimated by Greenwood's formula

$$\widehat{\sigma}^2(t) = \widehat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{r_j(r_j - d_j)}. \tag{9}$$

In the case of no censoring (9) reduces to $\widehat{S}(t)[1 - \widehat{S}(t)]/n$, the standard binomial variance estimator.

In large samples the Kaplan-Meier estimator, evaluated at a given time t, is approximately normally distributed so that a standard $100(1-\alpha)\%$ confidence interval for $S(t)$ takes the form

$$\widehat{S}(t) \pm z_{1-\alpha/2}\widehat{\sigma}(t), \tag{10}$$

with $z_{1-\alpha/2}$ the $1-\alpha/2$ fractile of the standard normal distribution. The approximation to the normal distribution is improved by using the log-minus-log transformation giving the confidence interval

$$\widehat{S}(t)^{\exp\left\{\pm z_{1-\alpha/2}\widehat{\sigma}(t)/[\widehat{S}(t)\ln\widehat{S}(t)]\right\}}. \tag{11}$$

This interval is satisfactory for quite small sample sizes (Borgan and Liestøl, 1990). Confidence intervals with small sample properties which are comparable to (11), or even slightly better, may be obtained by using the arcsine-square-root transformation (Borgan and Liestøl, op. cit.) or by basing the confidence interval on the likelihood ratio (Thomas and Grunkemeier, 1975; Cox and Oakes, 1984, Section 4.3). Note that all these confidence intervals should be given a pointwise interpretation. Simultaneous confidence bands for the survival distribution function are considered below.

Right censoring is not the only kind of data-incompleteness in survival analysis. Often, e.g. in epidemiological applications, individuals are not followed from time 0 (in the relevant time scale, typically age), but only from a later entry time (conditional on survival until this entry time). Thus, in addition to right censoring, the survival data are subject to left truncation. For such data we may, in principle at least, still use the Kaplan-Meier estimator (8) and estimate its variance by (9). The number at risk r_j is now the number of individuals who have entered the study before time t_j and are still in the study just prior to t_j. However, for left truncated data the numbers at risk r_j will often be low for small values of t_j. This will result in estimates $\widehat{S}(t)$ which have large sampling errors and which therefore may be of little practical use. What can be usefully estimated in such situations is the conditional survival distribution function $S(t|t_0) = S(t)/S(t_0)$ representing the probability of survival to time t given that an individual is alive at time $t_0 < t$. It may be useful to estimate such conditional distribution functions for several values of t_0 (at which there are reasonable numbers at risk), there being nothing canonical about any particular value. The estimation is performed as described earlier, the only modification being that the product in (8) and the sum in (9) are restricted to those t_j for which $t_0 < t_j \leq t$.

An illustration

As an illustration we use data from a randomized clinical trial for patients with histologically verified liver cirrhosis. Patients were recruited from several hospitals in Copenhagen between 1962 and 1969 and were followed until death, lost to follow-up or until the closing date of the study 1 October 1974. The time variable of interest is time since entry into the study. Patients are right censored if alive on 1 October 1974 or if lost to follow-up before that date.

We shall only consider the 138 placebo-treated male patients. Their median age at entry was 57 years, while the lower and upper quartiles were 51 and 66 years, respectively. Of the 138 patients 88 died during the study. The Kaplan-Meier estimate of the survival distribution

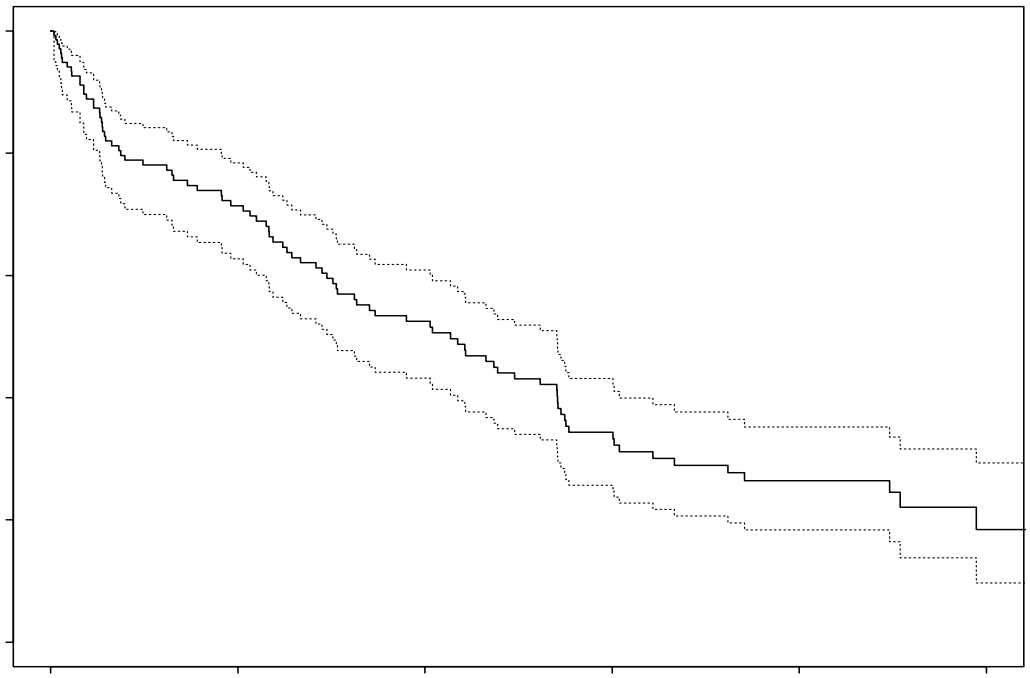

Figure 1: Kaplan-Meier estimate of the survival distribution function for 138 placebo-treated male patients with liver cirrhosis with 95% log-minus-log-transformed confidence intervals.

function for these patients is shown in Figure 1 with 95% confidence intervals computed according to (11). From the figure we see, e.g., that the five years survival probability is estimated to 43.0% with a 95% confidence interval from 34.0% to 51.9%, while the estimated ten years survival probability is 18.4% with confidence interval from 9.7% to 29.3%. We return to the liver cirrhosis example below in connection with median and mean survival times and simultaneous confidence bands. A further discussion and analysis of the data are given by Schlichting et al. (1983). The data were also used for illustrative purposes by Andersen et al. (1993).

Median survival time and related quantities

The use of the Kaplan-Meier estimator is not restricted to estimating survival probabilities for given times t. It may also be used to estimate fractiles such as the median survival time and related quantities like the interquartile range.

Consider the pth fractile ξ_p of the cumulative distribution function $F(t) = 1 - S(t)$, and

assume that $F(t)$ has positive density function $f(t) = F'(t) = -S'(t)$ in a neighborhood of ξ_p. Then ξ_p is uniquely determined by the relation $F(\xi_p) = p$, or equivalently, $S(\xi_p) = 1-p$. The Kaplan-Meier estimator is a step-function and hence does not necessarily attain the value $1-p$. Therefore a similar relation cannot be used to define the estimator $\widehat{\xi}_p$ of pth fractile. Rather we define $\widehat{\xi}_p$ to be the smallest value of t for which $\widehat{S}(t) \leq 1-p$, i.e. the time t where $\widehat{S}(t)$ jumps from a value greater than $1-p$ to a value less than or equal to $1-p$. In large samples $\widehat{\xi}_p$ is approximately normally distributed with a variance that may be estimated by

$$\widehat{\mathrm{var}}(\widehat{\xi}_p) = \frac{(1-p)^2 \widehat{\sigma}^2(\widehat{\xi}_p)}{[\widehat{f}(\widehat{\xi}_p)\widehat{S}(\widehat{\xi}_p)]^2}. \tag{12}$$

Here $\widehat{f}(t)$ is an estimator for the density function $f(t) = -S'(t)$. One may, e.g., use

$$\widehat{f}(t) = \frac{1}{2b}\left[\widehat{S}(t-b) - \widehat{S}(t+b)\right] \tag{13}$$

for a suitable bandwidth b (corresponding to a kernel function estimator with uniform kernel). Further, for $p < q$, $\widehat{\xi}_p$ and $\widehat{\xi}_q$ are approximately binormally distributed, and their correlation may be estimated by

$$\widehat{\mathrm{corr}}(\widehat{\xi}_p, \widehat{\xi}_q) = \frac{\widehat{\sigma}(\widehat{\xi}_p)\widehat{S}(\widehat{\xi}_q)}{\widehat{\sigma}(\widehat{\xi}_q)\widehat{S}(\widehat{\xi}_p)}. \tag{14}$$

Note that $\widehat{S}(\widehat{\xi}_p)$ in (12) and (14) is equal to or only slightly less than $1-p$, and that (12) could have been simplified if we had used this approximate equality. We have chosen not to do so since then $\widehat{S}(\widehat{\xi}_p)$ in (12) and (14) cancels with the same factor in $\widehat{\sigma}(\widehat{\xi}_p)$, cf. (9).

The above results may be used in the usual way to determine approximate confidence intervals, e.g., for the median survival time $\xi_{0.50}$ and the interquartile range $\xi_{0.75} - \xi_{0.25}$ as illustrated below. For the purpose of determining a confidence interval for a fractile like the median it is, however, better to apply the approach of Brookmeyer and Crowley (1982). For the pth fractile one then use as a confidence interval all hypothesized values ξ_p^0 of ξ_p which are not rejected when testing the null hypothesis $\xi_p = \xi_p^0$ against the alternative hypothesis $\xi_p \neq \xi_p^0$ at the α levels. Such test-based confidence intervals can be read directly from the lower and upper confidence limits for the survival distribution function in exactly the same manner as $\widehat{\xi}_p$ can be read from the Kaplan-Meier curve itself (see MEDIAN SURVIVAL TIME: CONFIDENCE INTERVALS AND TESTS).

For the liver cirrhosis data an estimate of the median survival time is 4.27 years (standard error 0.66 years), while the lower and upper quartiles are estimated to 1.46 years (0.35 years) and 8.97 years (1.13 years), respectively, with an estimated correlation of 0.28. In these computations the bandwidth $b = 1$ year was used in (13). An estimate of the interquartile range of the survival distribution function is $8.97 - 1.46 = 7.51$ years with standard error $(0.35^2 + 1.13^2 - 2 \cdot 0.35 \cdot 1.13 \cdot 0.28)^{1/2} = 1.09$ years. From this an approximate 95% confidence interval for the median survival time is $4.27 \pm 1.96 \cdot 0.66$, i.e. from 2.98 to 5.56 years, while 95% confidence limits for the interquartile range are from 5.37 to 9.65 years. For the median survival time it is, as mentioned earlier, better to read the confidence limits directly from the pointwise confidence intervals for the survival distribution function given in Figure 1. This

gives 95% confidence limits for the median survival time from 3.02 years to 5.41 years. Note that no estimate of the density function is needed here.

Mean survival time

Due to right censoring it will in most survival studies not be possible to get reliable estimates for the mean survival time $\mu = \int_0^\infty tf(t)\mathrm{d}t = \int_0^\infty S(t)\mathrm{d}t$. This is one important reason why in survival analysis the median is a more useful measure of location than the mean. What may be usefully estimated from right censored survival data is the expected time lived in a given interval $[0,t]$, i.e. $\mu_t = \int_0^t S(u)\mathrm{d}u$. This is estimated by

$$\widehat{\mu}_t = \int_0^t \widehat{S}(u)\mathrm{d}u,$$

the area under the Kaplan-Meier curve between 0 and t. Such an estimate may be of interest in its own right, or it may be compared with a similar population based estimate to asses the expected number of years lost up to time t for a group of patients. In large samples $\widehat{\mu}_t$ is approximately normally distributed with a variance that may be estimated by

$$\widehat{\mathrm{var}}(\widehat{\mu}_t) = \sum_{t_j \le t} \frac{(\widehat{\mu}_t - \widehat{\mu}_{t_j})^2 d_j}{r_j(r_j - d_j)},$$

a result which may be used to give approximate confidence limits for μ_t. By letting t tend to infinity, the above results may be extended to the estimation of the mean μ itself (Gill, 1983). However, the conditions (mainly on the censoring) needed for such an extension to be valid are usually not met in practice.

In the liver cirrhosis study no patient was followed for more than 13 years, making the estimation of the mean survival time impossible. We may, however, estimate the expected number of years lived up to a given time t. In particular estimates for the expected number of years lived up to 5 years and 10 years after the start of the study are 3.29 years (standard error 0.17 years) and 4.73 years (0.33 years), respectively.

Redistribute-to-the-right algorithm and self-consistency

We mentioned earlier the relation between the Kaplan-Meier estimator and the empirical distribution function in the case of no censoring. The redistribute-to-the-right algorithm and the consept of self-consistency, both due to Efron (1967), further illustrate this relation.

For notational convenience we assume that there are no ties, and we denote by $t_1^0 < t_2^0 < \cdots t_n^0$ the ordered times of deaths and censorings combined. The redistribute-to-the-right algorithm is as follows. First we constuct the ordinary empirical (survival) distribution function which places probability mass $1/n$ at each of the observed times t_j^0. If $t_{j_1}^0$ is the smallest t_j^0 that corresponds to a censored observation, we remove its mass and redistribute it equally among the $n - j_1$ time-points to the right of it. Then, if $t_{j_2}^0$ is the second smallest censored observation, we remove its mass, which will be $1/n + 1/[n(n - j_1)]$, and redistribute it equally among the $n - j_2$ time-points to its right, etc. This algorithm will converge in a

finite number of steps to the Kaplan-Meier estimator (8) (with the modification that it is set equal to zero after t_n^0 also when this last time-point corresponds to a censored observation).

A self-consistent estimator $\widetilde{S}(t)$ for the survival distribution function equals $1/n$ times an estimate for the number of individuals who survive time t. More precisely

$$\widetilde{S}(t) = \frac{1}{n}\left[\#(t_j^0 > t) + \sum_{t_j^0 \le t} a_j(t)\right], \tag{15}$$

where $a_j(t) = \widetilde{S}(t)/\widetilde{S}(t_j^0)$ if the observation at t_j^0 corresponds to a censored observation, and $a_j(t) = 0$ if it corresponds to an observed death. It turns out that the Kaplan-Meier estimator (modified as just indicated) is the unique self-consistent estimator. Turnbull (1976) used the idea of self-consistency to derive an iterative procedure (a version of the EM-algorithm) for estimating the survival distribution function nonparametrically from arbitrarily grouped, censored and truncated data, while Gill (1989) showed that the self-consistency equation (15) may be interpreted as a generalized score-equation.

Product-integral representation and relation to Nelson-Aalen estimator

Usually one assumes that the survival distribution function $S(t)$ is absolute continuous with density function $f(t) = -S'(t)$, hazard rate function $\alpha(t) = f(t)/S(t)$ and cumulative hazard rate function $A(t) = \int_0^t \alpha(u)\mathrm{d}u$. On the other hand the Kaplan-Meier estimator is discrete in nature, and the same applies to the Nelson-Aalen estimator for the cumulative hazard rate function. This makes it useful to be able to handle both discrete and continuous distributions within a unified framework. Let us therefore review how the survival distribution function $S(t)$ and the cumulative hazard rate function $A(t)$ are related for distributions which neither need to be continuous nor discrete. For such distributions

$$A(t) = -\int_0^t \frac{\mathrm{d}S(u)}{S(u-)}, \tag{16}$$

where $S(t-)$ denotes the left-hand limit of the survival distribution function at t. For an absolute continuous distribution (16) specializes to $A(t) = -\ln S(t) = \int_0^t \alpha(u)\mathrm{d}u$. For a discrete distribution it gives $A(t) = \sum_{u \le t} \alpha_u$, where the discrete hazard α_t is the conditional probability of death exactly at time t given that death has not occurred earlier. To express the survival distribution function by the cumulative hazard rate function it is convenient to use the product-integral $\prod$ defined as the limit of approximating finite products in a similar manner as the ordinary integral $\int$ is defined as the limit of approximating finite sums. With the use of the product-integral we may write

$$S(t) = \prod_{u \le t} [1 - \mathrm{d}A(u)]. \tag{17}$$

For a continuous distribution (17) specializes to the well-known relation $S(t) = \exp[-A(t)]$, while for a discrete distribution it takes the form $S(t) = \prod_{u \le t}(1 - \alpha_u)$.

The Nelson-Aalen estimator for the cumulative hazard rate function is $\widehat{A}(t) = \sum_{t_j \le t} d_j/r_j$. This corresponds to a distribution with all probability mass concentrated at the observed

failure times and with discrete hazard $\widehat{\alpha}_j = d_j/r_j$ at t_j. Using (17) the corresponding survival distribution function takes the form

$$\widehat{S}(t) = \prod_{u \le t} \left[1 - \mathrm{d}\widehat{A}(u)\right] = \prod_{t_j \le t} (1 - \widehat{\alpha}_j), \tag{18}$$

i.e., it is the Kaplan-Meier estimator (8). Thus the Kaplan-Meier and Nelson-Aalen estimators are related in exactly the same way as are the survival distribution function and the cumulative hazard rate functions themselves. This fact is lost sight of when one considers the relations $A(t) = -\ln S(t)$ and $S(t) = \exp[-A(t)]$ only valid for the continuous case. In fact, the latter relations have lead researchers to suggest the estimators $-\ln \widehat{S}(t)$ and $\exp[-\widehat{A}(t)]$ for the cumulative hazard rate function and the survival distribution function, respectively. The numerical differences between these two estimators and the Nelson-Aalen and Kaplan-Meier estimators will be of little importance in most cases. But the fact that the Nelson-Aalen and Kaplan-Meier estimators are related through (16) and (17) indicate that they are the canonical nonparametric estimators for the cumulative hazard rate function and the survival distribution function. This statement is supported by the fact that they may both be given a nonparametric maximum likelihood interpretation (Johansen, 1978).

Martingale representation and statistical properties

The product-integral formulation (18) of the Kaplan-Meier estimator shows its close relation to the Nelson-Aalen estimator, and it is the key to the study of its statistical properties. In fact, these are closely related to those of the Nelson-Aalen estimator. We will here indicate a few main steps and refer to Andersen et al. (1993, Section IV.3) for a detailed account.

Let $J(t) = 1$ if there is at least one individual at risk just before time t; $J(t) = 0$ otherwise. Further introduce $A^*(t) = \int_0^t J(u)\mathrm{d}A(u)$, and let

$$S^*(t) = \prod_{u \le t} \left[1 - \mathrm{d}A^*(u)\right]. \tag{19}$$

We note that (19) is almost the same as $S(t)$ (cf. (17)) when there is only a small probability that there is no one at risk at times $u \le t$. By a general result for product-integrals (Duhamel's equation), we may write

$$\frac{\widehat{S}(t)}{S^*(t)} - 1 = -\int_0^t \frac{\widehat{S}(u-)}{S^*(u)} \mathrm{d}(\widehat{A} - A^*)(u). \tag{20}$$

Here $\widehat{A} - A^*$ is a square integrable martingale (see NELSON-AALEN ESTIMATOR). It follows that the right hand side of (20) is a stochastic integral and hence itself a mean zero square integrable martingale. As a consequence of this $\mathrm{E}[\widehat{S}(t)/S^*(t)] = 1$ for any given t, so the Kaplan-Meier estimator is almost unbiased. Further the predictable variation process of the martingale on the right hand side of (20) may be used to arrive at an estimator for the variance of $\widehat{S}(t)/S^*(t)$. From this Greenwood's formula (9) follows provided one adopts a general model, not necessarily continuous. Greenwood's formula may also be derived through a standard information calculation starting with a binomial type likelihood for such a general model.

A further consequence of (20) is that $\sqrt{n}(\widehat{S} - S)/S$ is asymptotically equivalent to $-\sqrt{n}(\widehat{A} - A)$ and therefore converges weakly to a mean zero Gaussian martingale. In particular, for a fixed t, the Kaplan-Meier estimator (8) is asymptotically normally distributed, a fact which was used in connection with the confidence intervals (10) and (11). Also the asymptotic distributional results of the estimators for the median and mean survival times reviewed earlier are consequences of this weak convergence result.

Confidence bands

The weak convergence of $\sqrt{n}(\widehat{S} - S)/S$ to a mean zero Gaussian martingale also makes it possible to derive confidence bands for the survival distribution function, i.e. limits which contain $S(t)$ for all t in an interval $[\tau_1, \tau_2]$ with a prespecified probability. Two important types of such confidence bands are the equal precision bands (Nair, 1984) and the Hall-Wellner bands (Hall and Wellner, 1980). Borgan and Liestøl (1990) derived transformed versions of these confidence bands and compared them to the non-transformed ones.

The standard and log-minus-log transformed equal precision bands are obtained by replacing $z_{1-\alpha/2}$ in (10) and (11) by $d_{1-\alpha}(\widehat{c}_1, \widehat{c}_2)$, the $1-\alpha$ fractile in the distribution of the supremum of the absolute value of a standardized Brownian bridge over the interval from $\widehat{c}_1$ to $\widehat{c}_2$. Here

$$\widehat{c}_i = \frac{n[\widehat{\sigma}(\tau_i)/\widehat{S}(\tau_i)]^2}{1 + n[\widehat{\sigma}(\tau_i)/\widehat{S}(\tau_i)]^2}, \quad i = 1, 2. \tag{21}$$

The fractile $d_{1-\alpha}(\widehat{c}_1, \widehat{c}_2)$ may be found (approximately) by solving (w.r.t. d) the non-linear equation

$$\frac{4\phi(d)}{d} + \phi(d)\left(d - \frac{1}{d}\right) \ln\left[\frac{\widehat{c}_2(1-\widehat{c}_1)}{\widehat{c}_1(1-\widehat{c}_2)}\right] = \alpha$$

with $\phi(d)$ the standard normal density. The equal precision bands require $0 < \widehat{c}_1 < \widehat{c}_2 < 1$, so they cannot be extended all the way down to $t = 0$. Typically one will also omit the largest values of t.

The non-transformed Hall-Wellner band takes the form

$$\widehat{S}(t) \pm n^{-1/2} e_{1-\alpha}(\widehat{c}_1, \widehat{c}_2)\left(1 + n[\widehat{\sigma}(t)/\widehat{S}(t)]^2\right)\widehat{S}(t). \tag{22}$$

Here $e_{1-\alpha}(\widehat{c}_1, \widehat{c}_2)$ is the $1-\alpha$ fractile in the distribution of the supremum of the absolute value of a Brownian bridge over the interval from $\widehat{c}_1$ to $\widehat{c}_2$, cf. (21). For completely observed survival data the Hall-Wellner band reduces to the well-known Kolmogorov band $\widehat{S}(t) \pm n^{-1/2} e_{1-\alpha}(\widehat{c}_1, \widehat{c}_2)$. For the band (22) one will often let $\tau_1 = 0$ in which case tables of $e_{1-\alpha}(\widehat{c}_1, \widehat{c}_2) = e_{1-\alpha}(0, \widehat{c}_2)$ are given, e.g., by Koziol and Byar (1975) and Hall and Wellner (1980) for selected values of α and $\widehat{c}_2$. We note that (22) is obtained from (10) by substituting $n^{-1/2} e_{1-\alpha}(\widehat{c}_1, \widehat{c}_2)\left(1 + n[\widehat{\sigma}(t)/\widehat{S}(t)]^2\right)\widehat{S}(t)$ for $z_{1-\alpha/2}\widehat{\sigma}(t)$. The same substitution in (11) gives the log-minus-log transformed Hall-Wellner band. This transformed band require $\widehat{c}_1 > 0$ so it cannot be extended all the way down to $t = 0$. Due to the approximation $e_{1-\alpha}(\widehat{c}_1, \widehat{c}_2) \approx e_{1-\alpha}(0, \widehat{c}_2)$ the above mentioned tables may be used also for the transformed bands when $\widehat{c}_1$ is close to zero.

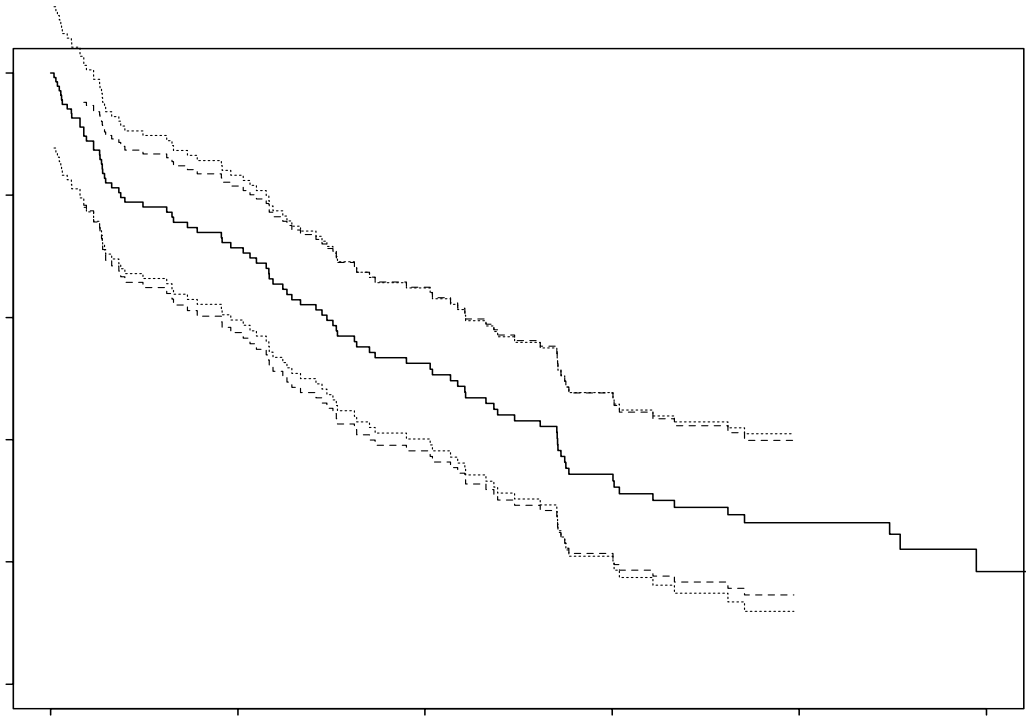

Figure 2: Kaplan-Meier estimate of the survival distribution function for 138 placebo-treated male patients with liver cirrhosis with 95% confidence bands: Log-minus-log transformed equal precision band over the interval from 4 months to 8 years (- - - - - - - -); Hall-Wellner band over the interval $[0, 8]$ years ($\cdots\cdots\cdots$).

The non-transformed equal precision band tends to achieve too high error rates when the number of observations is low, and the use of transformed bands is recommended even for samples of a hundred or more. The achieved error rates of the non-transformed Hall-Wellner band are fairly close to the nominal ones even in small samples, and the improvement obtained by using transformed bands are of less importance.

Figure 2 gives the Kaplan-Meier estimate for the liver cirrhosis data with 95% confidence bands. The bands shown are the log-minus-log transformed equal precision band over the interval from 4 months to 8 years and the non-transformed Hall-Wellner band valid from time zero to 8 years. Since $\tau_1 = 1/3$ year and $\tau_2 = 8$ years correspond to $\widehat{c}_1 = 0.090$ and $\widehat{c}_2 = 0.789$, the fractiles $d_{0.95}(\widehat{c}_1, \widehat{c}_2) = 2.99$ and $e_{1-\alpha}(0, \widehat{c}_2) = 1.36$ were used. It is seen that the equal precision band is more narrow than the Hall-Wellner band both for low and high values of t, while the Hall-Wellner band is slightly narrower than the equal precision band for intermediate values.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Verlag, New York.

Böhmer, P. E. (1912). Theorie der unabhängigen Warscheinlichkeiten. *Reports, Memoirs and Proceedings, Seventh International Congress of Actuaries, Amsterdam* **2**, 327–343.

Borgan, Ø. and Liestøl, K. (1990). A note on confidence intervals and bands for the survival curve based on transformations. *Scandinavian Journal of Statistics* **17**, 35–41.

Brookmeyer, R. and Crowley, J. J. (1982). A confidence interval for the median survival time. *Biometrics*, **38** 29–41.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* **4**, 831–853.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Gill, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Annals of Statistics* **11**, 49–58.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimation and the von Mises method (Part 1) *Scandinavian Journal of Statistics* **16**, 97–128.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–143.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. *Scandinavian Journal of Statistics* **5**, 195–199.

Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481, 562-563.

Koziol, J. A. and Byar, D. P. (1975). Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data. *Technometrics* **17**, 507–510.

Nair, V. N. (1984). Confidence bands for survival functions with censored data: A comparative study. *Technometrics* **26**, 265–275.

Schlichting, P., Christensen, E., Andersen, P. K., Fauerholdt, L., Juhl, E., Poulsen, H., and Tygstrup, N., for The Copenhagen Study Group for Liver Diseases. (1983). Prognostic factors in cirrhosis identified by Cox's regression model. *Hepatology* **3**, 889–895.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association* **70**, 865–871.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society B* **38**, 290–295.

AALEN-JOHANSEN ESTIMATOR

Contribution to the Encyclopedia of Biostatistics

Ørnulf Borgan, University of Oslo

The survival data situation may be described by the Markov process with the two states "alive" and "dead." Splitting the state "dead" into two or more states, corresponding to different causes of death, a Markov model for competing risks is obtained. Another Markov model of importance for biostatistical research is the illness-death model with states "healthy," "diseased" and "dead." For survival data the probability of a transition from state "alive" to state "dead" may be estimated as one minus the Kaplan-Meier estimator. The Kaplan-Meier estimator may be generalized to non-homogeneous Markov processes with a finite number of states. Such a generalization was considered by Aalen (1978) for the competing risks model and independently by Aalen and Johansen (1978) and Fleming (1978a,b) for the general case. In particular the product-integral formulation of Aalen and Johansen (1978) shows how the estimator, usually denoted the Aalen-Johansen estimator, can be seen as a matrix version of the Kaplan-Meier estimator.

Below we first consider the competing risks model and the Markov illness-death model for a chronic disease. This gives illustrations of the Aalen-Johansen estimator in two simple situations where its elements take an explicit form. Then we present the Aalen-Johansen estimator in general, and show how it is obtained as the product-integral of the Nelson-Aalen estimators for the cumulative transition intensities. We also briefly indicate how this may be used to study its statistical properties. A detailed account is given in the monograph by Andersen et al. (1993, Section IV.4).

Competing risks

Assume that we want to study the time to death and cause of death in a homogeneous population. This situation with competing causes of death may be modeled by a Markov process with one transient state 0, corresponding to "alive," and k absorbing states corresponding to "dead by cause h," $h = 1, 2, \ldots, k$. The transition intensity from state 0 to state h is denoted $\alpha_{0h}(t)$ and describes the instantaneous risk of dying from cause h, i.e. $\alpha_{0h}(t)\mathrm{d}t$ is the probability that an individual will die of cause h in the small time interval $[t, t + \mathrm{d}t)$ given that it is still alive just prior to t. The $\alpha_{0h}(t)$ are also termed cause specific hazard rate functions. For $h = 1, 2, \ldots, k$, we write $P_{0h}(s, t)$ for the probability that an individual in state 0 (i.e. alive) at time s will be in state h (i.e. dead from cause h) at a later time t. These transition probabilities are often termed cumulative incidence functions. Finally let $P_{00}(s, t)$ denote the probability that an individual who is alive (i.e. in state 0) at time s will still be alive at a later time t. Then

$$P_{00}(s,t) = \exp\left[-\int_s^t \sum_{h=1}^{k} \alpha_{0h}(u)\mathrm{d}u\right], \tag{23}$$

and

$$P_{0h}(s,t) = \int_s^t P_{00}(s,u)\alpha_{0h}(u)\mathrm{d}u \tag{24}$$

for $h = 1, 2, \ldots, k$.

Assume that we have a sample of n individuals from the population under study. Each individual is followed from an entry time to death or censoring, i.e. our observations may be subject to right censoring and/or left truncation. We denote by $t_1 < t_2 < \cdots$ the times when deaths (of any cause) are observed, and let d_{0hj} be the number of individuals who die from cause h (i.e. make a transition from state 0 to state h) at t_j. We also introduce $d_{0j} = \sum_{h=1}^{k} d_{0hj}$ for the number of deaths at t_j due to any cause, and let r_{0j} be the number of individuals at risk (i.e. in state 0) just prior to time t_j.

Then the survival probability (23) may be estimated by the Kaplan-Meier estimator

$$\widehat{P}_{00}(s,t) = \prod_{s<t_j\le t} (1 - d_{0j}/r_{0j}) \,, \tag{25}$$

while the cumulative incidence function (24) may be estimated by

$$\widehat{P}_{0h}(s,t) = \sum_{s<t_j\le t} \widehat{P}_{00}(s,t_{j-1})(d_{0hj}/r_{0j}) \tag{26}$$

for $h = 1, 2, \ldots, k$. Note that (26) is obtained from (24) by replacing $P_{00}(s,u) = P_{00}(s,u-)$ by $\widehat{P}_{00}(s,u-)$ and $\alpha_{0h}(u)\mathrm{d}u$ by $\mathrm{d}\widehat{A}_{0h}(u)$, the increment of the Nelson-Aalen estimator $\widehat{A}_{0h}(t) = \sum_{t_j\le t} d_{0hj}/r_{0j}$ for the cumulative cause specific hazard rate function $A_{0h}(t) = \int_0^t \alpha_{0h}(u)\mathrm{d}u$.

The variance of the Kaplan-Meier estimator (25) may in the usual way be estimated by Greenwood's formula, while when there are no ties in the data,

$$\begin{aligned}\widehat{\mathrm{var}}\,\widehat{P}_{0h}(s,t) = &\sum_{s<t_j\le t} [\widehat{P}_{00}(s,t_{j-1})\widehat{P}_{0h}(t_j,t)]^2 (r_{0j}-1) r_{0j}^{-3} d_{0j} \\ &+ \sum_{s<t_j\le t} \widehat{P}_{00}(s,t_{j-1})^2 [1 - 2\widehat{P}_{0h}(t_j,t)](r_{0j}-1) r_{0j}^{-3} d_{0hj}.\end{aligned} \tag{27}$$

By breaking the ties at random, this variance estimator may also be used when there is a small amount of tied observations. A more systematic treatment of variance estimation in the presence of ties is discussed below.

To illustrate the above results, we consider data on a cohort of uranium miners from the Colorado Plateau (e.g., Hornung and Meinhardt, 1987). The cohort consists of 3,347 Caucasian male miners recruited between 1950 and 1960 and was traced for mortality outcomes through December 31, 1982, by which time there were 258 lung cancer deaths and 2087 deaths from other causes. Of these deaths 145 and 1348 occurred between 40 and 60 years of age. The data were collected to study the effects of radon exposure and smoking on mortality, but for our illustrative purposes we will study the (marginal) risk of death from lung cancer disregarding the information on these exposures.

We use the competing risks model with two competing causes of death, corresponding to "dead from lung cancer" (state 1) and "dead from other causes" (state 2), and with age as time-scale. Figure 1 shows $\widehat{P}_{01}(40,t)$ for $40 < t \le 60$, i.e. the estimated risk that a 40 years old miner will die from lung cancer between 40 and t years of age taking into account the risk of death from other causes. Pointwise 95 % (log-transformed) confidence intervals based on the approximate normality of the Aalen-Johansen estimator (cf. below) are also shown. For comparison, Figure 1 also shows the estimated risk of lung cancer death disregarding the competing causes of death (computed as one minus the Kaplan-Meier estimator treating deaths from other causes as censorings). This estimate is sometimes interpreted as estimating the probability of death due to lung cancer assuming this to be the only possible cause of death. Such an interpretation may be quite speculative, however, see the discussion in Kalbfleisch and Prentice (1980, Chapter 7). The estimate disregarding competing risks is of course larger than the estimate which take the competing causes of death into account, the difference between them increases with age as the risk of dying from other causes increases.

An illness-death model

To study the occurrence of a chronic disease as well as death in a homogeneous population, we may adopt the Markov illness-death model with states 0, 1 and 2 corresponding to "healthy," "diseased" and "dead," respectively, and where no recovery (i.e. transition from state 1 to state 0) is possible. The

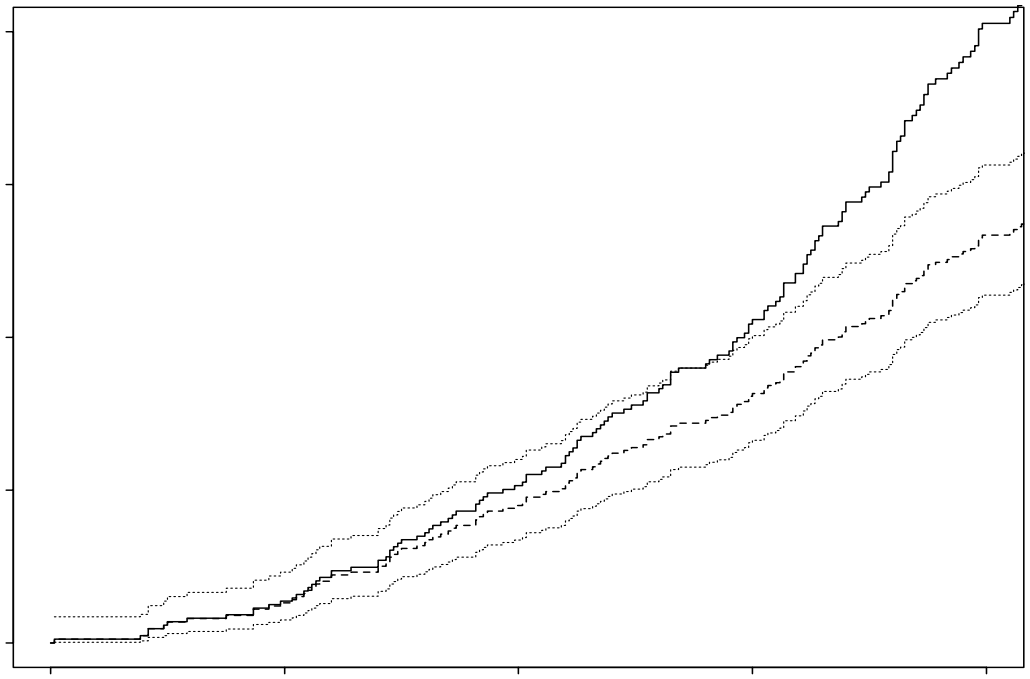

Figure 1: Aalen-Johansen estimate for the risk of dying from lung cancer taking into account the risk of death from other causes (– – – – – –) with 95 % log-transformed confidence intervals (- - - - - - -). Risk estimate disregarding other causes of death is also given (———).

transition intensities of the model are denoted $\alpha_{01}(t)$, $\alpha_{02}(t)$ and $\alpha_{12}(t)$ and describe the instantaneous risks of transitions between the states, i.e., $\alpha_{01}(t)\mathrm{d}t$ is the probability that an individual who is healthy just prior to time t will get diseased in the small time interval $[t, t+\mathrm{d}t)$, while $\alpha_{02}(t)\mathrm{d}t$ and $\alpha_{12}(t)\mathrm{d}t$ are the probabilities that an individual who is disease-free, respectively diseased, just before time t will die in the small time interval $[t, t+\mathrm{d}t)$. For an individual who is healthy (i.e. in state 0) at time s, we write $P_{01}(s,t)$ for the probability that he is diseased (i.e. in state 1) at a later time t, while $P_{00}(s,t)$ is the probability that he is still healthy (i.e. in state 0) at that time. Similarly, for an individual who is diseased (i.e. in state 1) at time s, we let $P_{11}(s,t)$ denote the probability that he is still alive (i.e. in state 1) at time t. Then we have

$$P_{00}(s,t) = \exp\left\{-\int_s^t [\alpha_{01}(u) + \alpha_{02}(u)]\,\mathrm{d}u\right\}, \tag{28}$$

$$P_{11}(s,t) = \exp\left[-\int_s^t \alpha_{12}(u)\mathrm{d}u\right], \tag{29}$$

$$P_{01}(s,t) = \int_s^t P_{00}(s,u)\alpha_{01}(u)P_{11}(u,t)\mathrm{d}u. \tag{30}$$

It is seen that (28) and (29) are of the same form as the survival probability in the survival data situation.

Assume then that we have a sample of n individuals from the population under study, and that each individual is followed from an entry time to death or censoring. Exact times of disease occurrences and deaths are recorded, and we denote by $t_1 < t_2 < \cdots$ the times of any observed event (disease occurrence or death). Further we let d_{01j} be the number of individuals who get diseased (i.e. make a transition from state 0 to state 1) at t_j, while d_{02j} and d_{12j} denote the numbers of disease-free, respectively diseased, individuals who die at that time. Finally we introduce $d_{0j} = d_{01j} + d_{02j}$ for the total number of transitions out of state 0, and let r_{0j} and r_{1j} be the number of healthy (i.e. in state 0) and diseased (i.e. in state 1) individuals, respectively, just prior to time t_j. Then (28) and (29) may be estimated by the Kaplan-Meier estimators

$$\widehat{P}_{00}(s,t) = \prod_{s<t_j\le t} (1 - d_{0j}/r_{0j})\,, \tag{31}$$

$$\widehat{P}_{11}(s,t) = \prod_{s<t_j\le t} (1 - d_{12j}/r_{1j}), \tag{32}$$

while an estimator for (30) is

$$\widehat{P}_{01}(s,t) = \sum_{s<t_j\le t} \widehat{P}_{00}(s,t_{j-1})(d_{01j}/r_{0j})\widehat{P}_{11}(t_j,t). \tag{33}$$

Note that (33) is obtained from (30) by replacing $P_{00}(s,u) = P_{00}(s,u-)$ by $\widehat{P}_{00}(s,u-)$, $P_{11}(u,t)$ by $\widehat{P}_{11}(u,t)$ and $\alpha_{01}(u)\mathrm{d}u$ by $\mathrm{d}\widehat{A}_{01}(u)$, the increment of the Nelson-Aalen estimator $\widehat{A}_{01}(t) = \sum_{t_j\le t} d_{01j}/r_{0j}$ for the cumulative disease intensity $A_{01}(t) = \int_0^t \alpha_{01}(u)\mathrm{d}u$. The variance of the Kaplan-Meier estimators (31) and (32) may be estimated by Greenwood's formula, while

$$\begin{aligned}
\widehat{\mathrm{var}}\,\widehat{P}_{01}(s,t) \;=\; & \sum_{s<t_j\le t} \widehat{P}_{00}(s,t_{j-1})^2[\widehat{P}_{11}(t_j,t) - \widehat{P}_{01}(t_j,t)]^2(r_{0j}-1)r_{0j}^{-3}d_{01j} \\
& + \sum_{s<t_j\le t} [\widehat{P}_{00}(s,t_{j-1})\widehat{P}_{01}(t_j,t)]^2(r_{0j}-1)r_{0j}^{-3}d_{02j} \\
& + \sum_{s<t_j\le t} [\widehat{P}_{01}(s,t_{j-1})\widehat{P}_{11}(t_j,t)]^2(r_{1j}-1)r_{1j}^{-3}d_{12j}
\end{aligned} \tag{34}$$

when there are no ties in the data, or when a few ties have been broken at random.

Before we illustrate these results, let us mention that other interpretations of the states are possible than the ones given above. In particular in a study involving treatment of cancer, state 0 could correspond to "no response to treatment," state 1 to "response to treatment" and state 2 to "relapse." The probability $P_{01}(s,t)$ is then the probability of being in response function suggested by Temkin (1978) and sometimes used as an outcome measure when studying the efficacy of cancer chemotherapy. Another interpretation arise in the study of complications to a disease. Here state 0 could correspond to "diseased with no complications," state 1 to "diseased with complications" and state 2 to "dead." This interpretation of the states is the one relevant for the following illustration.

The Steno Memorial Hospital in Greater Copenhagen has since 1933 served as a diabetes specialist hospital for patients from the whole of Denmark. From the medical records at Steno we use for illustration data on the 374 female patients referred between 1933 and 1981 and in whom the diagnosis insulin-dependent diabetes mellitus was established (usually by a general practitioner or another hospital) before the age of 10 years and between 1933 and 1972. The patients were followed from first contact with Steno to death, emigration, or 31 December 1984. One of the major complications of insulin-dependent diabetes is diabetic nephropathy which is a sign of kidney failure. Seventeen patients had diabetic nephropathy at first admission to Steno, while 76 developed this complication during the observation period. The seriousness of diabetic nephropathy is reflected by the fact that among these 93 patients 54 were observed to die, whereas only 30 of the 281 patients who did not develop diabetic nephropathy died during the observation period.

We model the disease-histories of the patients by the Markov illness-death model with the states 0 and 1 corresponding to "alive without diabetic nephropathy" and "alive with diabetic nephropathy," respectively, and with diabetes duration as time-scale. Figure 2 shows $\widehat{P}_{01}(5,t)$, i.e. the estimated probability of being alive with diabetic nephropathy for patients without this complication five years after the onset of the disease. Pointwise 95 % (log-transformed) confidence intervals based on the approximate normality of the Aalen-Johansen estimator (cf. below) are also shown. It is seen that the probability of being alive with diabetic nephropathy (among the group of patients we consider) first increases up to an estimated value of 17 % after 23 years of diabetes duration, and then declines due to the high mortality among these patients.

It should be realized that Figure 2 is based on two crude assumptions. Firstly, calendar time trends in mortality and incidence of diabetic nephropathy are not taken into account. Secondly, by using a Markov process to model the disease-histories, the effect on mortality of duration of diabetic nephropathy has been neglected. A point of less importance is that the exact times of onset of diabetic nephropathy were not known for 9 of the 93 patients with this complication. For these nine patients predicted times for the occurrence of diabetic nephropathy were used. A further discussion and analysis of the data are given, e.g., by Borch-Johnsen et al. (1985). The data were used for illustrative purposes by Andersen et al. (1993) who also describe how the nine predicted times have been calculated.

The general case

We then consider a general Markov process with a finite number of states which may be used to model the life-histories of individuals from a homogeneous population. Let $\mathcal{I} = \{0, 1, \ldots, k\}$ be the state space of the Markov process, and denote by $\alpha_{gh}(t)$ the transition intensity from state $g \in \mathcal{I}$ to state $h \in \mathcal{I}$, $g \neq h$. The transition intensities describe the instantaneous risks of transitions between the states, so $\alpha_{gh}(t)\mathrm{d}t$ is the probability that an individual who is in state g just before time t will make a transition to state h in the small time interval $[t, t + \mathrm{d}t)$. Further for all $g, h \in \mathcal{I}$, we let $P_{gh}(s,t)$ denote the probability that an individual who is in state g at time s will be in state h at a later time t, and we write $\mathbf{P}(s,t)$ for the $(k+1) \times (k+1)$ matrix of these transition probabilities. Only for simple Markov processes, like the competing risks and illness-death models considered earlier, is it

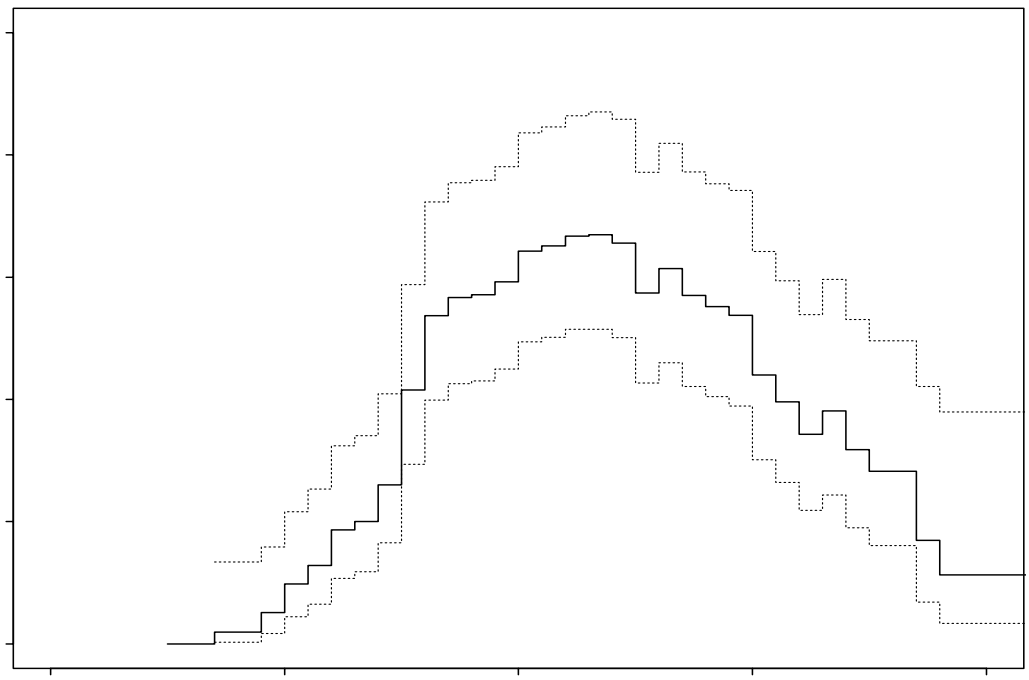

Figure 2: Aalen-Johansen estimate of the probability of being alive with diabetic nephropathy for female patients with diabetes onset before 10 years of age and with no sign of diabetic nephropathy five years after the onset of the disease (————). Pointwise 95 % log-transformed confidence intervals are also shown (- - - - - - -).

possible to give explicit expressions for the $P_{gh}(s,t)$ in terms of the transition intensities, cf. (23), (24) and (28)–(30). We will see later, however, that the transition probability matrix $\mathbf{P}(s,t)$ itself can be expressed in terms of the $(k+1)\times(k+1)$ matrix $\boldsymbol{\alpha}(t)$ of the transition intensities. First we review the Aalen-Johansen estimator for $\mathbf{P}(s,t)$ and discuss estimation of (co)variances.

Suppose that we have a sample of n individuals from the population under study. The individuals may be followed over different periods of time, so our observations of their life-histories may be subject to left-truncation and/or right censoring. A crucial assumption, however, is that truncation and censoring is independent so that the entry and censoring times do not carry any information on the risks of transitions between the states; cf. Andersen et al. (1993, Sections III.2-3) for a general discussion. We assume that exact times for transitions between the states are recorded, and denote by $t_1 < t_2 < \cdots$ the times when transitions between any two states are observed. Further for $g, h \in \mathcal{I}$, $g \neq h$, we let d_{ghj} be the number of individuals who experience a transition from state g to state h at t_j, and introduce $d_{gj} = \sum_{h\neq g} d_{ghj}$ for the number of transitions out of state g at that time. Finally we let r_{gj} be the number of individuals in state g just prior to time t_j. Then the Aalen-Johansen estimator takes the form

$$\widehat{\mathbf{P}}(s,t) = \prod_{s<t_j\leq t} (\mathbf{I} + \widehat{\boldsymbol{\alpha}}_j)\,. \tag{35}$$

Here $\mathbf{I}$ is the $(k+1)\times(k+1)$ identity matrix, $\widehat{\boldsymbol{\alpha}}_j$ is the $(k+1)\times(k+1)$ matrix with entry (g,h) equal to $\widehat{\alpha}_{ghj} = d_{ghj}/r_{gj}$ for $g \neq h$ and entry (g,g) equal to $\widehat{\alpha}_{ggj} = -d_{gj}/r_{gj}$, and the matrix product is taken in the order of increasing t_js. For simple models like the competing risks model and the illness-death model considered earlier, we are able to give explicit expressions for the elements of (35), cf. (25), (26) and (31)–(33). In general, however, this is not possible. But in any case a direct implementation of (35) is simple using software which can handle matrix multiplications.

For any $g, h, m, r \in \mathcal{I}$, the covariance between $\widehat{P}_{gh}(s,t)$ and $\widehat{P}_{mr}(s,t)$ may be estimated by

$$\widehat{\mathrm{cov}}(\widehat{P}_{gh}(s,t), \widehat{P}_{mr}(s,t)) = \tag{36}$$
$$\sum_{i=0}^{k}\sum_{l\neq i}\sum_{s<t_j\leq t} \widehat{P}_{gi}(s,t_{j-1})\widehat{P}_{mi}(s,t_{j-1})[\widehat{P}_{lh}(t_j,t) - \widehat{P}_{ih}(t_j,t)][\widehat{P}_{lr}(t_j,t) - \widehat{P}_{ir}(t_j,t)](r_{ij}-1)r_{ij}^{-3}d_{ilj}$$

provided that there are no ties in the data or that a small amount of tied observations have been broken at random. Formulas (27) and (34) given earlier are special cases of (36). As an alternative to (36), or to handle ties in a systematic manner, one may use the recursion formula

$$\begin{aligned}\widehat{\mathrm{cov}}\left(\widehat{P}_{gh}(s,t_j), \widehat{P}_{mr}(s,t_j)\right) &= \\ \sum_{i=0}^{k}\sum_{l=0}^{k} \widehat{\mathrm{cov}}\left(\widehat{P}_{gi}(s,t_{j-1}), \widehat{P}_{ml}(s,t_{j-1})\right)(\delta_{ih} + \widehat{\alpha}_{ihj})(\delta_{lr} + \widehat{\alpha}_{lrj}) &\qquad (37)\\ + \sum_{i=0}^{k} \widehat{P}_{gi}(s,t_{j-1})\widehat{P}_{mi}(s,t_{j-1})\widehat{\mathrm{cov}}(\widehat{\alpha}_{ihj}, \widehat{\alpha}_{irj}).\end{aligned}$$

which describes how the estimated (co)variances are updated at the times of the observed transitions. (The estimates are constant between the t_js.) Here δ_{ih} is a Kronecker delta, while $\widehat{\mathrm{cov}}(\widehat{\alpha}_{ihj}, \widehat{\alpha}_{irj})$ equals $(\delta_{hr}r_{ij} - d_{ihj})r_{ij}^{-3}d_{irj}$ when $h \neq i$, $r \neq i$; it equals $-(r_{ij} - d_{ij})r_{ij}^{-3}d_{irj}$ when $h = i \neq r$; and it equals $(r_{ij} - d_{ij})r_{ij}^{-3}d_{ij}$ when $h = r = i$. When there are no ties in the data (36) and (37) give identical results.

Product-integral representation and relation to Nelson-Aalen estimator

We then review how the transition probability matrix may be derived from the transition intensities $\alpha_{gh}(t)$ and describe how the Aalen-Johansen estimator is related to the Nelson-Aalen estimators for the cumulative transition intensities. To this end introduce $\alpha_{gg}(t) = -\sum_{h \neq g} \alpha_{gh}(t)$ and write $\boldsymbol{\alpha}(t)$ for the $(k+1) \times (k+1)$ matrix with element (g,h) equal to $\alpha_{gh}(t)$. Then the transition probability matrix $\mathbf{P}(s,t)$ is the unique solution to the Kolmogorov forward differential equation $(\partial/\partial t)\mathbf{P}(s,t) = \mathbf{P}(s,t)\boldsymbol{\alpha}(t)$ with initial condition $\mathbf{P}(s,s) = \mathbf{I}$. By a general result for product-integrals (Volterra's equation) this solution takes the form $\mathbf{P}(s,t) = \prod_{(s,t]}[\mathbf{I} + \boldsymbol{\alpha}(u)\mathrm{d}u]$. Alternatively, if we introduce the $(k+1) \times (k+1)$ matrix $\mathbf{A}(t)$ with elements $A_{gh}(t) = \int_0^t \alpha_{gh}(s)\mathrm{d}s$, we may write

$$\mathbf{P}(s,t) = \prod_{(s,t]} [\mathbf{I} + \mathrm{d}\mathbf{A}(u)]\,. \tag{38}$$

This product-integral representation of the transition probability matrix of a Markov process is not restricted to the situation where transition intensities exist. In fact (38) only assumes the existence of cumulative transition intensities $A_{gh}(t)$ which do not need to be absolutely continuous.

For $g \neq h$ we may estimate the cumulative transition intensity $A_{gh}(t)$ by the Nelson-Aalen estimator $\widehat{A}_{gh}(t) = \sum_{t_j \leq t} \widehat{\alpha}_{ghj}$, while $\widehat{A}_{gg}(t) = -\sum_{h \neq g} \widehat{A}_{gh}(t) = \sum_{t_j \leq t} \widehat{\alpha}_{ggj}$. Let $\widehat{\mathbf{A}}(t) = \sum_{t_j \leq t} \widehat{\boldsymbol{\alpha}}_j$ be the $(k+1) \times (k+1)$ matrix with these elements. By (38) it is reasonable to estimate the transition probability matrix by $\widehat{\mathbf{P}}(s,t) = \prod_{(s,t]}[\mathbf{I} + \mathrm{d}\widehat{\mathbf{A}}(u)]$. But since $\widehat{\mathbf{A}}(t)$ is a matrix of step-functions with a finite number of jumps on $(s,t]$, this is nothing but the Aalen-Johansen estimator (35). Thus the Aalen-Johansen and Nelson-Aalen estimators are related in exactly the same way as are the transition probability matrix and the cumulative transition intensities themselves. This suggests that the Aalen-Johansen estimator is the canonical nonparametric estimator for the matrix of transition probabilities in a Markov process with a finite number of states. This statement is supported by the fact that it may also be given a nonparametric maximum likelihood interpretation (Johansen, 1978).

Martingale representation and statistical properties

The product-integral formulation of the Aalen-Johansen estimator is useful for the study of its statistical properties. We will here indicate a few main steps and refer to Andersen et al. (1993, Section IV.4) for a detailed account. For each $g \in \mathcal{I}$ we introduce an indicator $J_g(t)$ which is one if there is at least one individual in state g just before time t, and which is zero otherwise. Further for all $g, h \in \mathcal{I}$ define $A^*_{gh}(t) = \int_0^t J_g(u)\mathrm{d}A_{gh}(u)$, and let $\mathbf{A}^*(t)$ be the $(k+1) \times (k+1)$ matrix with these elements. Finally we introduce $\mathbf{P}^*(s,t) = \prod_{(s,t]}[\mathbf{I} + \mathrm{d}\mathbf{A}^*(u)]$, and note that this is almost the same as $\mathbf{P}(s,t)$ (cf. (38)) when there is only a small probability that one or more states will be empty at times u between s and t. By a general result for product-integrals (Duhamel's equation), we may then write

$$\widehat{\mathbf{P}}(s,t)\mathbf{P}^*(s,t)^{-1} - \mathbf{I} = \int_{(s,t]} \widehat{\mathbf{P}}(s,u-)\mathrm{d}(\widehat{\mathbf{A}} - \mathbf{A}^*)(u)\mathbf{P}^*(s,u)^{-1}. \tag{39}$$

Here $\widehat{\mathbf{A}} - \mathbf{A}^*$ is a $(k+1) \times (k+1)$ matrix of square integrable martingales (see NELSON-AALEN ESTIMATOR). It follows that the right hand side of (39) is a matrix-valued stochastic integral, and therefore itself a $(k+1) \times (k+1)$ matrix of mean zero square integrable martingales. As a consequence of this

$$\mathrm{E}\left[\widehat{\mathbf{P}}(s,t)\mathbf{P}^*(s,t)^{-1}\right] = \mathbf{I},$$

so the Aalen-Johansen estimator is almost unbiased. Further the predictable variation process of the matrix-valued martingale (39) suggests an estimator for the covariance matrix of $\widehat{\mathbf{P}}(s,t)\mathbf{P}^*(s,t)^{-1}$, and based on this the (co)variance estimators (36) and (37) may be derived.

The martingale representation (39) is also key to the study of the large sample properties of the Aalen-Johansen estimator. For fixed s it may be shown that $\widehat{\mathbf{P}}(s, \cdot)$, properly normalized, converges weakly to a matrix-valued Gaussian process. In particular when also t is given, the Aalen-Johansen estimator (35) is asymptotically multinormally distributed, a fact which was used earlier in connection with the construction of confidence intervals.

References

Aalen, O. O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models *Annals of Statistics* **6**, 534–545.

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

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Verlag, New York.

Borch-Johnsen, K., Andersen, P. K., and Deckert, T. (1985). The effect of proteinuria on relative mortality in Type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* **28**, 590–596.

Fleming, T. R. (1978a). Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. *Annals of Statistics* **6**, 1057–1070.

Fleming, T. R. (1978b). Asymptotic distribution results in competing risks estimation. *Annals of Statistics* **6**, 1071–1079.

Hornung, R. and Meinhardt, T. (1987). Quantitative risk assessment of lung cancer in U. S. uranium miners. *Health Physics*, **52**, 417–30.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. *Scandinavian Journal of Statistics* **5**, 195–199.

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

Temkin, N R. (1978). An analysis for transient states with application to tumor shrinkage. *Biometrics* **34**, 571–580.