estimated, with applications to goodness of fit tests in models with censoring. *Research report.* Norwegian Computing Centre, Oslo.
Hjort, N. L. (1984e). Non-parametric Bayes estimators of cumulative intensities in models with censoring. *Research report.* Norwegian Computing Centre, Oslo.
Kronmal, R. & Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Am. Statist. Ass.* **63,** 925–952.
Lehmann, E. L. (1983). *Theory of point estimation.* Wiley, New York.
Millar, P. W. (1979). Asymptotic minimax theorems for the sample distributions function. *Z. Wahr. verw. Geb.* **48,** 233–252.
Padgett, W. J. & Wei, L. J. (1981). A Bayesian non-parametric estimator of survival probability assuming increasing failure rate. *Commun. Statist.* **A10,** 49–63.
Roussas, G. R. (1972). *Contiguity of probability measures: some applications in statistics.* Cambridge University Press.
Susarla, V. & Van Ryzin, J. (1976). Non-parametric Bayesian estimation of survival curves from incomplete observations. *J. Am. Statist. Ass.* **71,** 897–902.
Wellner, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10,** 595–602.

Elja Arjas (University of Oulu)

I would like to congratulate Andersen and Borgan on their excellent paper. I think that they were able to achieve a very fine blending of theoretical and practical aspects. It is perhaps not always recognized that advanced mathematics, such as σ-fields and martingales in this context, can sometimes be far more concrete and practical than the more traditional elementary mathematics.

I was particularly delighted by the authors' remark in the introduction concerning the use of i.i.d. random variables versus stochastic processes as tools for describing "random phenomena occurring in time". As there are so many things of interest in this world that belong under this heading, I believe that this remark applies far more generally than to the modelling of failure data.

Since I would like to emphasize the role of stochastic processes as a basic notion in modelling, and since a stochastic process is a time-indexed family of random variables, there is a good reason to go somewhat deeper into the analysis of the *time parameter*. In particular, I shall elaborate on the following remark the authors make in the beginning of section 2: "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."

To understand the meaning and the implications of this comment, let us first consider briefly how failure data usually comes about: suppose one starts collecting information at calendar time T_0. After T_0, events of interest are then registered in the order of their occurrence. Restricting ourselves for simplicity to the findings concerning a single individual in the follow-up, say the ith, we would typically register $T_{\mathrm{ENTRY}(i)}$, $T_{\mathrm{EXIT}(i)}$ and the status D_i at $T_{\mathrm{EXIT}(i)}$, i.e. $D_i=1$ if individual i died at $T_{\mathrm{EXIT}(i)}$ and $D_i=0$ if he was censored. The original data could also contain treatment dates together with treatment specifications. Supposing that there was only a single treatment type and at most one treatment per individual, we would register for every treated individual $T_{\mathrm{TREAT}(i)}$, the time of treatment. Other important dates could be $T_{\mathrm{BIRTH}(i)}$ and $T_{\mathrm{DIAG}(i)}$, the latter being the date at which the disease in question was first diagnosed in individual i. Here we shall assume for simplicity that $T_{\mathrm{DIAG}(i)}=T_{\mathrm{ENTRY}(i)}$.

Considering then the risk for the ith individual at (calendar) time t, it is natural to relate the risk to measurements such as calendar time itself, $(t-T_{\mathrm{BIRTH}(i)})^+=\text{age}$, $(t-T_{\mathrm{DIAG}(i)})^+=\text{time from diagnosis}$, $(t-T_{\mathrm{TREAT}(i)})^+=\text{time from treatment}$, etc. In other words, such measurements, or functions of such measurements, could be used as covariates in modelling hazard.

To make all these "times" more concrete, we could imagine that every individual in the study carries a collection of stopwatches. A first watch would simply show the calendar time. A second watch was set to go at $T_{\mathrm{BIRTH}(i)}$ and would then show the individual's age. A third watch was set to go at $T_{\mathrm{DIAG}(i)}$, etc. By "times" we would simply mean the collection of all clock times.

Let us now return to the remark about "time" usually being the time from diagnosis. In such a case, individually for each i, $T_{\mathrm{DIAG}(i)}$ becomes a new time origin. This leads to a reordering of events which were originally registered according to calendar time; we call this reordering operation *reshuffling*.

Note how reshuffling changes the meaning of the word "simultaneous": after reshuffling, events occurring to different individuals are considered to be simultaneous if the corresponding *clock times from diagnosis* are equal. This must be remembered when interpreting for survival data the assumption "no two component processes can jump simultaneously" about multivariate counting processes (see section 3.1.)

But reshuffling has consequences that are far more important: The σ-fields generated by reshuffled events do not coincide with the "natural" histories based on calendar time ordering. This is important since, when verifying the martingale property, it is always necessary that the histories be specified. Thus a martingale with respect to the natural history in calendar time might not be a martingale with respect to the reshuffled history, and vice versa.

The example in section 8 called *testing with replacement* teaches us to be cautious: there life lengths and censoring time have an extremely simple structure, the former being i.i.d. random variables in a renewal process and the latter being constant. Yet, if reshuffling is done and the history is changed accordingly, the natural martingale structure with $\alpha(\cdot)$ as the hazard rate breaks down. If there were not reshuffling, such problems would not arise; problems arise when one wants to find non-parametric or semi-parametric estimates. This is illustrated in the two examples below.

Note that there is a minor difference in how "the breakdown of the counting process approach" in testing with replacement is diagnosed: Andersen and Borgan point at the non-predictability of the censoring processes $C_i(\cdot)$ whereas I would like to put the blame on reshuffling and the consequent change of histories. (Predictability is not a problem if the processes $C_i(\cdot)$ are adapted since we are dealing with the absolutely continuous case and can always choose $C_i(\cdot)$ to be left continuous. I think that the processes $C_i(\cdot)$ should be always included in the list of generators of the history, even for the reshuffled one, in which case being adapted becomes automatic.)

The first example is an extension of testing with replacement.

Example 1. Consider the case of "staggered entry" where the times $T_{\mathrm{DIAG}(i)}$ can be different for different individuals (cf. Chapter 6 in Gill (1980a) and Sellke & Siegmund (1983)). We can then define the counting processes N_i and the risk indicators Y_i almost exactly as in (3.2) and (3.3), by setting

$$N_i(t)=I(T_{\mathrm{EXIT}(i)}\leqslant t,\ D_i=1)$$

and

$$Y_i(t)=I(T_{\mathrm{DIAG}(i)}<t\leqslant T_{\mathrm{EXIT}(i)}).$$

Writing $(\mathscr{F}_t)$ for the natural history according to calendar time t, we can postulate a statistical model by assuming that

$$P\{dN_i(t)=1\,|\,\mathscr{F}_{t-}\}=Y_i(t)\cdot\alpha(t-T_{\mathrm{DIAG}(i)})\,dt,$$

where $\alpha=\alpha(\cdot)$ is some (unknown) function. More precisely then, for each i,

$$M_i(t)=N_i(t)-\int_0^t Y_i(s)\,\alpha(s-T_{\mathrm{DIAG}(i)})\,ds, \qquad t\geq 0. \tag{1}$$

is an $(\mathscr{F}_t)$-martingale. This is almost exactly as (3.4) and (3.12), except that t is explicitly the calendar time and $(\mathscr{F}_t)$ the corresponding history.

But how can we estimate the function $\alpha(\cdot)$ in this case? A moment's reflection shows that reshuffling would really be necessary for the non-parametric Nelson–Aalen method to work because $\alpha(\cdot)$ is considered with different time delays for different individuals (recall the "clock times"!). In doing so, the martingale property of (1) may be lost. However, parametric estimation is more successful: Sticking to the natural history $(\mathscr{F}_t)$, assume that (1) holds with $\alpha(\cdot)$ belonging to a parametric family $\{\alpha(\cdot, \theta);\ \theta\in\mathscr{R}^1\}$ of hazard rates. Also assume that censoring is non-informative in the sense of Kalbfleish & Prentice (1980, section 5.2). Then, up to factors not depending on θ, the logarithmic likelihood function for data collected before calendar time t has the expression (cf. (6.1))

$$\log L_t(\theta)=\sum_i\int_0^t Y_i(s)\,\{\log \alpha(s-T_{\mathrm{DIAG}(i)}, \theta)dN_i(s)-\alpha(s-T_{\mathrm{DIAG}(i)}, \theta)\,ds\}.$$

Differentiating with respect to θ we obtain (cf. (6.4))

$$\frac{d}{d\theta}\log L_t(\theta)=\sum_i\int_0^t Y_i(s)\,\frac{d\log\alpha(s-T_{\mathrm{DIAG}(i)}, \theta)}{d\theta}\,dM_i(s),$$

where M_i is as in (1) with $\alpha(\cdot)=\alpha(\cdot, \theta)$. Thus

$$\left\{\frac{d}{d\theta}\log L_t(\theta)\right\}_{t\geq 0}$$

is an $(\mathscr{F}_t)$-martingale for the "correct value" of θ. If $\alpha(\cdot, \theta)$ is estimated by the maximum likelihood method, martingale-based asymptotic results are at hand in very much the same way as in the paper. It remains a question of terminology whether one should then say that one is still using the counting process approach (cf. section 8). □

The next example aims at clarifying the role of "times" in the *semi-parametric* regression models.

Example 2. Recall from (7.1) the basic factorization of the intensity

$$\alpha_h\{t, \mathbf{Z}_i(t)\}=\alpha_{0h}(t)\,g\{\boldsymbol{\beta}'\mathbf{Z}_{hi}(t)\}, \tag{2}$$

where $\alpha_{0h}(\cdot)$ is an unknown *baseline hazard* and $g(\cdot)$ a given *relative risk function*. In the *partial likelihood* approach $\alpha_{0h}(\cdot)$ is suppressed and $\boldsymbol{\beta}$ is estimated by maximizing an expression which does not involve $\alpha_{0h}(\cdot)$.

Questions of interest to us here are: which one of the "times" should t be in (2)? How should one distinguish the baseline hazard from the relative risk function in a case where also the latter can depend on time through time-dependent covariates?

A possible guideline is the fact that $\alpha_{0h}(\cdot)$ does not depend on the individual. One could take t as the calendar time, in which case $\alpha_{0h}(\cdot)$ would represent a calendar time trend common to all individuals. As noted in the paper, however, this choice is made only rarely: it is more likely that time from diagnosis or time from treatment has a stronger effect on the hazard than calendar time. Therefore, it is tempting to reshuffle the events and to replace calendar time t individually by $t-T_{\mathrm{DIAG}(i)}$ or $t-T_{\mathrm{TREAT}(i)}$.

But then there is a similar danger as in example 1: it can be that the difference

$$N_i(t)-\int_0^t Y_i(s)\,\alpha_{0h}(s-T_{\mathrm{DIAG}(i)})\,g\{\boldsymbol{\beta}'\mathbf{Z}_{hi}(s)\}\,ds \tag{3}$$

is a martingale with respect to the natural history (t being the calendar time), but that reshuffling and the consequent change of history destroys the martingale property. The danger is here made only bigger by the fact that the covariates can be random and they can depend on time.

There is also another point to make: it can be that, for a realistic statistical model, one should consider simultaneously several clock times, say t, $(t-T_{\mathrm{DIAG}(i)})^+$ and $(t-T_{\mathrm{TREAT}(i)})^+$. Clearly, the covariate vector $\mathbf{Z}_{hi}(t)$ can accommodate all such times in its coordinates so that this is well within the framework of (2). But it is important to note that the baseline hazard can be synchronized with only one clock time. If the model uses more than a single clock time, the remaining ones must be included in the relative risk function and become then a part of the parametric model. I think that statisticians who are using the partial likelihood should be cautioned against thinking that the baseline hazard is a "pool which contains all unspecified (and maybe not clearly understood) time dependencies". When a method becomes standard and available in statistical computer program packages, there is always a great danger of uncritical use. □

Finally, I would like to comment briefly on the use of *continuous* versus *discreter* time parameter. In the paper nothing is said about discrete time models, and so, implicitly, "the counting process approach" means automatically continuous time parameter. Some authors state explicitly their preference of continuous time models (see e.g. Gill, 1984).

While physical time is most naturally described as a continuous variate, it is less clear to me that common types of failure data should be modelled by continuous time models. It has been pointed out by many authors (see e.g. Thompson, 1977) that some rounding of the time measurements is always present. The ties in failure data are mostly a result from such rounding. It is somewhat unfortunate if a statistical model, through the assumption "no two component processes can jump simultaneously", puts a zero probability density on the observed data whenever there is at least one tied failure time. More or less *ad hoc* methods are then needed (in the non-parametric and semi-parametric models) to overcome this.

One possibility to finding a resolution in this question would be to say that continuous time models give a more natural physical description about evolution in time, but that discrete time models can be more consistent with data arising from rounded time measurements. Discrete time models have also two pleasant properties: first, orthogonality of the component martingales M_i is implied by a simple conditional independence assumption between the individuals, without a need to rule out simultaneous deaths. Secondly, the required mathematics is much more elementary. For example, all struggling about the exact definition and the correct interpretation of the term "predictable" becomes unnecessary. This may be a valuable asset when trying to make the counting process approach popular among statisticians.

Additional reference

Thompson, W. A., Jr. (1977). On the treatment of grouped observations in life studies. *Biometrics* **33**, 463–470.

Jon Stene (University of Copenhagen)

It should be stressed, what the authors only briefly indicate in section 8, that the practical applicability of the methods presented in this paper relies heavily on the error one makes by using as approximations limit results derived by means of the martingale central limit theorem.