The Annals of Statistics
1990, Vol. 18, No. 3, 1221–1258

GOODNESS OF FIT TESTS IN MODELS FOR LIFE HISTORY DATA BASED ON CUMULATIVE HAZARD RATES

BY NILS LID HJORT

Norwegian Computing Centre and University of Oslo

To check the validity of an assumed parametric model for survival data, one may compare $\hat{A}(t)$, the nonparametric Nelson–Aalen plot of the cumulative hazard rate, with $A(t, \hat{\theta})$, the estimated parametric cumulative hazard rate, $\hat{\theta}$ being for example the maximum likelihood estimator. Convergence in distribution of $\sqrt{n}\,(\hat{A}(t) - A(t, \hat{\theta}))$ and more general processes is studied in the present paper, employing the general framework of counting processes, which allows for quite general models for life history data and for quite general censoring schemes. The results are applied to the construction of χ^2-type statistics for goodness of fit. Cramér–von Mises and Kolmogorov–Smirnov type tests are presented in the case where the unknown parameter is one-dimensional. Power considerations are also included, and some optimality results are reached. Finally tests are constructed for the hypothesis that the unspecified hazard rate part in Cox's regression model follows a parametric form.

1. Introduction and summary. Statisticians often have to perform several difficult and highly interrelated tasks concurrently, including translating other scientists' problems into statistical terms, analyzing data and establishing models. The yes/no answer provided by a traditional goodness of fit test, for some parametric model, is usually only a small part, and dependent upon other parts, of the final statistical analysis. That many journal articles are devoted exclusively to goodness of fit testing, and usually without giving clues as to what to do should the model be rejected, for example, is perhaps better explained by the way in which we publish our papers, than by statistical practice. Having said this, however, this is an article developing general goodness of fit methods for parametric models for time-continuous survival data. It should find applications in the areas of actuarial statistics, biostatistics, demography, engineering, life history data, reliability and sociology. The methods work for much more general models than life length distribution models and should be applicable to a wide range of models for counts of transitions from one state to another, also in the presence of censoring.

Although workers in reliability theory and survival analysis and other fields often have been concerned about their parametric assumptions, relatively few general goodness of fit procedures seem to have been available for time-continuous data when censoring is present. An exception is the problem of testing for exponentiality, where numerous methods have been proposed; cf. Doksum

Received September 1985; revised September 1989.
AMS 1980 *subject classifications*. Primary 62G10; secondary 60B10.
Key words and phrases. Censoring, chi squared statistics, counting processes, goodness of fit, hazard rates, local power, weak convergence.

and Yandell (1984). Aalen and Hoem (1978) outlined a general method based on a technique using random time changes. Aalen (1982) elaborated a bit further on another method involving so-called Cox residuals; cf. Cox (1979). Independently of each other Burke (1981), Hjort [(1984), Section 3] and Habib and Thomas (1986) developed asymptotic theory for Kaplan–Meier curves when parameters are estimated and proposed ways of using this to test parametric models. Burke (1981), Csörgő and Horváth (1982b) and Horváth (1982) used empirical kernel transforms, but this results in somewhat awkward testing procedures. Habib and Thomas (1986) arrived at chi squared type tests. Gray and Pierce (1985) tested parametric models within larger parametric models using score functions. Akritas (1988) developed a chi squared test in the random censorship framework which in fact is identical to the test listed as Special Case 2 in Section 3.

Our approach is based upon *hazard rates* (or intensities, or forces of transition) and their cumulatives. These are perhaps more natural and fundamental quantities in survival data and counting process models than probability densities, cumulative distributions, quantile functions and transition probabilities. This gives rise to large classes of tests, most of which seem to be new even when specialized to the classical situation of random variables on $[0, \infty)$ with no censoring. We shall be concerned with quite general models for life history data. In its abstract form, Aalen's (1975, 1978a, 1982) general multiplicative model for counting process data is defined as follows: Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and let $\{\mathscr{F}_t;\ t \geq 0\}$ be an increasing, right-continuous family of sub-σ-algebras of $\mathscr{F}$. Furthermore, let $N = \{N(t);\ t \geq 0\}$ be a counting process defined on $(\Omega, \mathscr{F}, P)$, adapted to $\{\mathscr{F}_t;\ t \geq 0\}$, with hazard process $Y(s)\alpha(s)$ for $s \geq 0$, where Y is adapted, nonnegative and left-continuous. $N(\cdot)$ and $Y(\cdot)$ are observed, while $\alpha(\cdot)$, the parameter of the model, is (partly) unknown. See Andersen and Borgan (1985) for a review.

It is well demonstrated by now that the rather abstractly defined model above encompasses a fair range of important statistical models. The prime example is the homogeneous survival data model which involves data of the form $T_i = \min(X_i, c_i)$, $\delta_i = I\{T_i \leq c_i\}$ on n individuals having a common hazard rate $\alpha(\cdot)$; here X_i are lifetimes and c_i censoring times. In this situation $N(t) = \sum_{i=1}^n I\{T_i \leq t,\ \delta_i = 1\}$ counts the number of observed deaths in $[0, t]$ and $Y(t) = \sum_{i=1}^n I\{T_i \geq t\}$ is the number at risk just before time t. Other examples also typically let N count transitions of some kind, while Y has a number at risk interpretation and α is the hazard rate for one individual. There are important bonuses of the counting process formulation. Much more general models than the simple alive $\to$ dead model above can be treated, and in a unified way. It also encompasses quite general censoring schemes; see Aalen (1978a), Gill (1980) and Andersen and Borgan (1985).

The counting process framework above was originally created for the nonparametric case, where $\alpha(\cdot)$ is only assumed to be continuous (Aalen 1975, 1978a). However, recently several authors have studied *parametric* multiplicative models, thus generalizing and unifying (at least asymptotic aspects of) decades of work on parametric lifetime distributions. These models have

$\alpha(s) = \alpha(s, \theta)$ for some specified function of a p-dimensional parameter θ. Borgan (1984), Andersen and Borgan (1985), Hjort (1986) and Karr (1986) studied properties of maximum likelihood inference about θ; Hjort (1985) developed more general M-type estimators, while Hjort (1986) and Aven (1986) also considered Bayesian procedures.

Suppose one wants to check whether one's $\alpha(s, \theta)$ model is adequate. Introduce the *cumulative hazard rate* $A(t) = \int_0^t \alpha(s)\,ds$ for $t \geq 0$, which under the model assumption is equal to $A(t, \theta) = \int_0^t \alpha(s, \theta)\,ds$ for a certain value of the parameter. A sound procedure is to draw both

$$\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)} \quad \text{and} \quad A(t, \hat{\theta}) = \int_0^t \alpha(s, \hat{\theta})\,ds$$

in a diagram and compare them. Here $\hat{A}$ is the nonparametric Nelson–Aalen estimator (cf. Andersen and Borgan (1985)) and $A(t, \hat{\theta})$ the corresponding parametric one, with θ estimated from the data. If the assumed model is correct, or at least adequate, then these two functions should agree reasonably well. This method of diagnostic checking was introduced in Nelson (1969, 1972) and versions of it are widely used in engineering statistics. Some examples from biostatistics can be found in Aalen (1982).

To construct a proper test, hoping to make rigorous the "reasonably well" statement above, we choose to work with the maximum likelihood estimator $\hat{\theta}$, which has several asymptotic optimality properties according to Hjort (1986) and Dzhaparidze (1986), and study asymptotic properties of

$$Z_n(t) = \sqrt{n}\,\{\hat{A}(t) - A(t, \hat{\theta})\}$$

and the more general process

$$H_n(t) = \int_0^t K_n(s)\,dZ_n(s) = \sqrt{n}\int_0^t K_n(s)\{dN(s)/Y(s) - \alpha(s, \hat{\theta})\,ds\}, \qquad t \geq 0$$

The weight process K_n is assumed to be predictable, or to approximate one that is predictable, and is scaled in such a way that $K_n(s)$ converges to some $k(s, \theta)$ in probability. We show in Section 2 that $H_n \to_d H$, a certain zero-mean Gaussian process with covariance function given in Theorem 2.1. This generalizes the by now classic result about the limit in distribution of $\sqrt{n}\,\{\hat{A}(\cdot) - A(\cdot)\}$ available in Aalen (1978a), Gill (1980) and Andersen and Borgan (1985). It also generalizes results independently obtained by Csörgő and Horváth (1982b) for Z_n. Also included in Section 2 are results about the asymptotic behavior of the general goodness of fit process H_n outside model conditions. We establish results both for local alternatives and for a fixed alternative.

The weak convergence results are applied in Section 3 to the construction of several types of test statistics for model assumptions of the general form $\alpha(s) = \alpha(s, \theta)$. In particular, each choice of K_n above leads to a χ^2-type statistic. Tests of the Kolmogorov–Smirnov and Cramér–von Mises type can also be put up and in the one-parameter case explicit and readily usable limit distribution results are reached.

The general theory is applied to some special cases in Section 4, while power function performance of the tests is investigated in Section 5. In addition to providing approximations to power functions the results also have bearings on the problem of choosing the weight function K_n. Some optimality results are reached for local power against contiguous alternatives.

In Section 6 a parametric Cox model is studied. The model postulates that $\alpha_i(s) = \alpha(s, \theta)\exp(\beta' z_i)$ is the hazard rate for an individual with covariate vector z_i. Processes similar to Z_n and H_n above are defined and we show how the adequacy of the model can be tested. The tests are shown to be consistent against large classes of alternatives.

The theory and results of the present paper provide rigorous methods by which it is possible to judge whether the nonparametric $\hat{A}(t)$ fits the parametric $A(t, \hat{\theta})$. Formulae are also given sufficient for the drawing of $\hat{A}(t) - A(t, \hat{\theta})$ divided by an estimate of its standard deviation, as a function of time t, suggesting another graphical procedure which could be used to detect possible departures from the assumed model. Also informative in this respect is the comparison of a nonparametric estimator $\hat{\alpha}(s)$ of the hazard rate itself with the parametric $\alpha(s, \hat{\theta})$. Methods producing such $\hat{\alpha}$'s have been proposed by Ramlau-Hansen (1983), Hjort (1985) and others. Yet another proposal is to plot a simultaneous confidence band for the unknown cumulative hazard A, for which methods have been proposed by Csörgő and Horváth (1982a), Hjort [(1985), Section 1], Bie, Borgan and Liestøl (1987) and others.

A point worth making is that the proposed goodness of fit tests work perfectly well even when no censoring is present. Most of them seem to be new even in this classical framework. In particular, one of the χ^2-type statistics constructed in Section 3, referred to there as Special Case 2, can be used as an alternative to the classical Pearson χ^2. The new test still compares the number of observations in an interval with an estimate based on the model, but the latter is now evaluated in a more dynamic way, using the hazard function and the number at risk function instead of the overall probability of falling in the interval. This test was proposed in Hjort (1984, 1985), but was independently developed by Akritas (1988). A certain optimality property for this test is established in Section 5.

The article is concluded with a number of remarks placed in Section 7. Points taken up there include extensions to several counting processes, alternative estimators to be used in the various tests, a general search procedure for declaring certain departures from the model to be present and other goodness of fit tests that can be constructed using the same machinery.

2. Weak convergence of the general goodness of fit process. This section provides limit distribution results for the general goodness of fit process H_n described in Section 1. Section 2.1 briefly discusses various prerequisites that are needed and provides the basic Theorem 2.1 about the limit distribution of H_n under model conditions and from which goodness of fit tests are later derived in Section 3. Section 2.2 gives some results about H_n outside the model conditions, first w.r.t. a sequence of local alternatives and then in the context of a fixed alternative. These results have bearings on power functions for the various tests, a topic pursued in Section 5.

I learned after an earlier version of this paper had been written that a version of the special case of Theorem 2.1 that corresponds to weight function $K_n(s) \equiv 1$, i.e., $H_n = Z_n$, had been proved earlier in Csörgő and Horváth [(1982b), Section 12] and Horváth [(1982), Section 2]. These authors employ strong Hungarian approximation methods, which work under the random censorship model for i.i.d. random variables. Our approach, using counting processes and martingales, yields a more generally valid result. However, we include in our study only the maximum likelihood estimator for the unknown parameter, or asymptotically equivalent variants (cf. Remark 7A), whereas Csörgő and Horváth allow more general estimators.

2.1. *Weak convergence under model conditions.* Since we aim at asymptotic properties, we think of our model as being the n'th in a sequence of models having an increasing number (usually n, in fact) of individuals under study, each with the same hazard rate $\alpha(\cdot, \theta)$. We shall partly suppress n in the notation in what follows; this concerns in particular the counting process N, the number at risk process Y and the associate martingale M defined in (2.1). We employ $\hat{\theta} = \hat{\theta}_n$, the maximum likelihood estimator for θ in the n'th model; see (2.3). It will be convenient to keep θ as a free parameter, so single out the true value and denote it θ_0. Calculations and probability statements in the present Section 2.1 are w.r.t. the probability mechanism governed by this θ_0.

For simplicity, assume from now on that the stochastic processes involved are observed over a finite time interval only, say $[0, T]$. This is the customary framework, as, e.g., in most of Andersen and Borgan (1985). Generalizations are possible, for example to the case of an arbitrary stopping time T, or even to $[0, \infty)$ with some extra conditions; see Remark 7C.

Borgan [(1984), Section 4] and Hjort [(1986), Sections 2 and 3] studied the properties of $\hat{\theta}$ in this framework. We shall assume that Borgan's regularity conditions (A), (B), (C) and (D) are fulfilled. Although not strictly necessary, we shall also throw in the following condition:

(E) $Y(t) \le n$ and $N(t) \le n$ always, and $Y(t)/n$ tends to some positive, deterministic function $y(t)$, in probability

for good measure. In the random censorship model, for example, where $T_i = \min(X_i, c_i)$ and X_i's come from $F(\cdot)$ while censoring times c_i come from $G(\cdot)$, then $y(t) = F[t, \infty)G[t, \infty)$. A convenient set of *sufficient* conditions for (A)–(E) to hold is as follows:

(F) The convergence in (E) is uniform, i.e., $\max_{0 \le t \le T}|Y(t)/n - y(t)| \to_p 0$.

(G) There is a neighbourhood $N(\theta_0)$ of θ_0 in which the first, second and third order derivatives of $\alpha(s, \theta)$ w.r.t. θ exist and are continuous, for almost all s, and they are bounded in $[0, T] \times N(\theta_0)$.

(H) $\alpha(s,\theta)$ is bounded away from zero in $[0,T]\times N(\theta_0)$.
(I) The Σ-matrix defined in (2.5) is positive definite.

These conditions (F)–(I) are usually satisfied in practice.

More notation is needed. Let

$$M(t)=N(t)-\int_0^t Y(s)\alpha(s,\theta_0)\,ds,\qquad t\geq 0. \tag{2.1}$$

Then M is a local, quadratic integrable martingale w.r.t. $\{\mathscr{F}_t;\ t\geq 0\}$, and $\mathrm{Var}\{dM(s)|\mathscr{F}_{s-}\}=Y(s)\alpha(s,\theta_0)\,ds$. The regularity conditions ensure that $M(\cdot)/\sqrt{n}\to_d V(\cdot)$ in the function space $D[0,T]$ equipped with the Skorohod topology, where V is a zero mean Gaussian martingale with independent increments and

$$\mathrm{Var}\{dV(s)\}=y(s)\alpha(s,\theta_0)\,ds. \tag{2.2}$$

The integral of a function $g(s)$ w.r.t. $dV(s)$ is well defined whenever $\int_0^T g(s)^2 y(s)\alpha(s,\theta_0)\,ds$ is finite. Andersen, Borgan, Gill and Keiding (1982) and Andersen and Borgan (1985) are good sources to consult about the general martingale machinery that is going to be employed in this paper, and for other applications.

The likelihood of what is observed, expressed as a Radon–Nikodym derivative w.r.t. a unit rate Poisson process, can be written

$$L(\theta)=\text{const.}\exp\left[\int_0^T\{\log\alpha(s,\theta)\,dN(s)-Y(s)\alpha(s,\theta)\,ds\}\right]; \tag{2.3}$$

see, for example, Borgan (1984). Let $\psi(s,\theta)$ be the vector of partial derivatives of $\log\alpha(s,\theta)$ w.r.t. θ; in particular $\hat\theta$ is the solution to $\int_0^T\psi(s,\hat\theta)\{dN(s)-Y(s)\alpha(s,\hat\theta)\,ds\}=0$. Then

$$\begin{aligned}U_n&=\frac{1}{\sqrt{n}}\frac{\partial\log L(\theta_0)}{\partial\theta}=\int_0^T\psi(s,\theta_0)\frac{dM(s)}{\sqrt{n}}\\&\to_d U=\int_0^T\psi(s,\theta_0)\,dV(s).\end{aligned} \tag{2.4}$$

Furthermore,

$$\begin{aligned}\partial^2\log L(\theta_0)/\partial\theta_j\,\partial\theta_l&=\int_0^T\partial\psi_j(s,\theta_0)/\partial\theta_l\,dM(s)\\&\quad-\int_0^T Y(s)\psi_j(s,\theta_0)\psi_l(s,\theta_0)\alpha(s,\theta_0)\,ds.\end{aligned}$$

The vector U has covariance matrix Σ with elements given by

$$-\frac{1}{n}\frac{\partial^2\log L(\theta_0)}{\partial\theta_j\,\partial\theta_l}\to_p\sigma_{j,l}=\int_0^T y(s)\psi_j(s,\theta_0)\psi_l(s,\theta_0)\alpha(s,\theta_0)\,ds. \tag{2.5}$$

Borgan (1984) showed in effect that the maximum likelihood estimator satisfies

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \Sigma^{-1}U_n + \varepsilon_n, \tag{2.6}$$

where ε_n tends to zero in probability; in particular $\sqrt{n}(\hat{\theta} - \theta_0)$ tends to $N_p(0, \Sigma^{-1})$ in distribution.

Now we possess the requisites and can proceed to processes relevant to goodness of fit testing. Let

$$Z_n(t) = \sqrt{n}\left\{\hat{A}(t) - A_J(t, \hat{\theta})\right\} = \sqrt{n}\int_0^T J(s)\{dN(s)/Y(s) - \alpha(s, \hat{\theta})\, ds\},$$

in which $J(s) = I\{Y(s) >\}$ and $A_J(t, \theta) = \int_0^t J(s)\alpha(s, \theta_0)\, ds$. We have $J(s) \to 1$ uniformly in probability from assumption (E), so that $A_J = A$ with high probability. Working with A_J instead of A makes the martingale apparatus work more smoothly and has a statistical reason, also: We cannot estimate $\alpha(s)$ in regions where $J(s) = 0$, i.e., where nothing is observed. More generally, define

$$\begin{aligned} H_n(t) &= \int_0^t K_n(s)\, dZ_n(s) \\ &= \sqrt{n}\int_0^t K_n(s)J(s)\{dN(s)/Y(s) - \alpha(s, \hat{\theta})\, ds\}, \qquad t \geq 0, \end{aligned} \tag{2.7}$$

where K_n is an a.s. bounded process scaled in such a way that $K_n(s) \to_p k(s, \theta_0)$, say. Of course $K_n \equiv 1$ gives us Z_n again.

We shall study weak convergence of H_n under suitable requirements on the weight function K_n. It is easiest to reach a result for cases where K_n is a *predictable* or *previsible* process. Sufficient conditions for K_n to be predictable are that $K_n(t)$ is $\mathscr{F}_t$-measurable for each t and that its sample paths are left continuous, i.e., $K_n(t)$ is known just prior to time t. It will, however, be worth the trouble to allow K_n to depend upon $\hat{\theta}$ as well, in which case it is not predictable, since $K_n(t)$ depends upon data recorded after time t. The following technical regularity condition is not the weakest possible, but will suffice for our purpose. Under this condition K_n at least approximates a predictable process and this will often suffice for the asymptotic martingale calculus to work.

(K) $K_n(s) = G_n(s, \hat{\theta})$, where the process $G_n(s, \theta_0)$ is predictable, converges uniformly to $k(s, \theta_0)$ in probability and is twice continuously differentiable w.r.t. θ. Furthermore, the partial derivative $G'_{n,j}(s, \theta_0)$ is predictable and converges uniformly to $g_j(s, \theta_0)$ in probability, and $g_j(s, \theta_0)^2 y(s)^{-1}\alpha(s, \theta_0)\, ds$ is integrable over $[0, T]$, for each j. Finally there is a neighbourhood $N(\theta_0)$ of θ_0 such that the second order derivatives satisfy

$$\max_{0 \leq t \leq T}\ \max_{\theta \in N(\theta_0)} \left|G''_{n,jl}(s, \theta)\right|/\sqrt{n} \to_p 0.$$

THEOREM 2.1. *Let the counting process model satisfy regularity conditions* (A)–(E) *or* (F)–(I) *and suppose that K_n satisfies the requirements of condition* (K), *in particular, $K_n(s) \to_p k(s, \theta_0)$. Let $B(t)$ be the vector $\int_0^t k(s, \theta_0)\psi(s, \theta_0)\alpha(s, \theta_0)\, ds$. Then $H_n \to_d H$ in $D[0, T]$, where*

$$H(t) = \int_0^t \{k(s, \theta_0)/y(s)\}\, dV(s) - B(t)'\Sigma^{-1} \int_0^T \psi(s, \theta_0)\, dV(s), \qquad t \geq 0.$$

H has covariance function

$$\operatorname{Cov}\{H(t_1), H(t_2)\} = \int_0^{t_1 \wedge t_2} \{k(s, \theta_0)^2/y(s)\}\alpha(s, \theta_0)\, ds - B(t_1)'\Sigma^{-1}B(t_2). \tag{2.8}$$

PROOF. Start out subtracting and adding $\sqrt{n}\, J(s)\alpha(s, \theta_0)$ in the defining expression for $Z_n(t)$, to obtain

$$dZ_n(s) = J(s)\left[\sqrt{n}\, dM(s)/Y(s) - \psi(s, \tilde{\theta})'\alpha(s, \tilde{\theta})\sqrt{n}\,(\hat{\theta} - \theta_0)\right],$$

in which $\tilde{\theta}$ is on the line segment between θ_0 and $\hat{\theta}$. Let $B_n(t) = \int_0^t K_n(s)\psi(s, \tilde{\theta})\alpha(s, \tilde{\theta})\, ds$. Then the expression above, the regularity assumption about K_n and (2.4) permit us to write

$$\begin{aligned} H_n(t) &= \int_0^t K_n(s) \frac{nJ(s)}{Y(s)} \frac{dM(s)}{\sqrt{n}} - B_n(t)'\sqrt{n}\,(\hat{\theta} - \theta_0) \\ &= \int_0^t G_n(s, \theta_0) \frac{nJ(s)}{Y(s)} \frac{dM(s)}{\sqrt{n}} - B_n(t)'(\Sigma^{-1}U_n + \varepsilon_n) \\ &\quad + \int_0^t \{K_n(s) - G_n(s, \theta_0)\} \frac{nJ(s)}{Y(s)} \frac{dM(s)}{\sqrt{n}}. \end{aligned}$$

This surely already suggests the result, in view of the following lemma:

LEMMA 1. *Suppose $(H_n^0, U_n) \to_d (H^0, U)$ in $D[0, T] \times \mathscr{R}^p$, which is the same as requiring that $H_n^0 \to_d H^0$ in $D[0, T]$ and that every set of $(H_n^0(t_1), \ldots, H_n^0(t_i), U_n)$ converge properly. In this case, if $C_{n,j}(s) \to c_j(s)$ uniformly in probability and the c_j's are continuous, then $H_n = H_n^0 - \sum_{j=1}^p C_{n,j} U_{n,j}$ converges in distribution to $H = H^0 - \sum_{j=1}^p c_j U_j$ in $D[0, T]$.*

PROOF. This lemma is proved upon noting two things, utilizing, respectively, Theorem 4.4 and Theorem 5.1 in Billingsley (1968). First, $(H_n^0, U_n, C_n) \to_d (H^0, U, c)$ in $D[0, T] \times \mathscr{R} \times D[0, T]^p$. Second, the mapping which takes (H_n^0, U_n, C_n) to $H_n^0 - (C_n)'U_n$ is measurable and continuous on $D[0, T] \times \mathscr{R}^p \times C[0, T]^p$. □

It remains to demonstrate that the regularity assumed really ensures $H_n \to_d H$. In view of Lemma 1, it suffices to show that:

(i) $$\left(\int_0^{\cdot} G_n(s,\theta_0)\frac{nJ(s)}{Y(s)}\frac{dM(s)}{\sqrt{n}}, U_n\right) \to_d \left(\int_0^{\cdot}\frac{k(s,\theta_0)}{y(s)}dV(s), U\right),$$

in $D[0,T] \times \mathscr{R}^p$.

(ii) $B_n(t) \to B(t)$ uniformly in probability, from which it then also follows that $B_n(t)'\varepsilon_n$ tends to zero uniformly in probability.

(iii) $$\max_{0 \le t \le T}\left|\int_0^T \{K_n(s) - G_n(s,\theta_0)\}\frac{nJ(s)}{Y(s)}\frac{dM(s)}{\sqrt{n}}\right| \to_p 0.$$

(i) is implied by the stronger statement

$$\left(\int_0^{\cdot} G_n(s,\theta_0)\frac{nJ}{Y}\frac{dM}{\sqrt{n}}, \int_0^{\cdot}\psi(s,\theta_0)\frac{dM}{\sqrt{n}}\right)$$
$$\to_d \left(\int_0^{\cdot}\frac{k(s,\theta_0)}{y(s)}dV(s), \int_0^{\cdot}\psi(s,\theta_0)\,dV(s)\right)$$

in $D[0,T]^{p+1}$. This follows, however, by an application of the general martingale convergence Theorem 2.1 of Andersen, Borgan, Gill and Keiding (1982). Details concerning the verification of the conditions of that theorem in the present situation are available in Hjort (1984). They involve the use of Lenglart's inequality as well as other martingale convergence arguments.

(ii) is true since consistency of $\hat{\theta}$, $\max_{0 \le s \le T}|K_n(s) - k(s,\theta_0)| \to_p 0$ and Borgan's regularity conditions together imply

$$\int_0^T \left|K_n(s)\psi_j(s,\tilde{\theta})\alpha(s,\theta_0) - k(s,\theta_0)\psi_j(s,\theta_0)\alpha(s,\theta_0)\right| ds \to_p 0.$$

To show (iii), write $K_n(s) = G_n(s,\hat{\theta}) + G_n'(s,\theta_0)(\hat{\theta} - \theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)' G_n''(s,\tilde{\theta})(\hat{\theta} - \theta_0)$, where $\tilde{\theta}$ again is between θ_0 and $\hat{\theta}$ and dependent upon both n and s. That

$$\max_{0 \le t \le T}\left|\int_0^t G_n'(s,\theta_0)(\hat{\theta} - \theta_0)\frac{nJ(s)}{Y(s)}\frac{dM(s)}{\sqrt{n}}\right|$$
$$\le \sum_{j=1}^p \max_{0 \le t \le T}\left|\int_0^t G_{n,j}'(s,\theta_0)\frac{nJ(s)}{Y(s)}\frac{dM(s)}{\sqrt{n}}\right|\left|\hat{\theta}_j - \theta_{0,j}\right| \to_p 0$$

holds, is seen via $\int_0^{\cdot} G_{n,j}'(nJ/Y)(dM/\sqrt{n}) \to_d \int_0^{\cdot} g_j/y\,dV$, which is another consequence of Andersen, Borgan, Gill and Keiding's Theorem 2.1. Finally,

$$\max_{0 \le t \le T}\left|\int_0^t (\hat{\theta}_j - \theta_{0,j})(\hat{\theta}_l - \theta_{0,l})G_{n,jl}''(s,\tilde{\theta})\frac{nJ(s)}{Y(s)}\frac{dM(s)}{\sqrt{n}}\right|$$
$$\le n\left|\hat{\theta}_j - \theta_{0,j}\right|\left|\hat{\theta}_l - \theta_{0,l}\right|\max_{0 \le s \le T}\left|G_{n,jl}''(s,\tilde{\theta})\frac{\sqrt{n}J(s)}{Y(s)}\right|\{1 + A(T)\}$$

and this converges to zero in probability by the assumption on $G''_{n,jl}$ stated in regularity condition (K). We have used here that $|\int_0^t C_n\, dM| \le n \max_s |C_n(s)| \times \{1 + A(T)\}$ holds, for every process C_n, by regularity condition (E).

The expression (2.8) for the covariance function of H is not difficult to verify, recalling that the Gaussian martingale V has independent increments with variances given by (2.2) and that $U = \int_0^T \psi(s, \theta_0)\, dV(s)$ has covariance matrix Σ. □

REMARK. The case where no unknown parameters are present is also covered by the efforts above. In that case $H_n(t) = \sqrt{n} \int_0^t K_n(s) J(s)\{dN(s)/Y(s) - \alpha(s)\, ds\}$ converges in distribution to $H(t) = \int_0^t \{k(s)/y(s)\}\, dV(s)$, provided only that K_n is a predictable process with a limit in probability function k. The limit process H becomes itself a Gaussian martingale with independent increments and $\mathrm{Var}\{dH(s)\} = \{k(s)^2/y(s)\}\alpha(s)\, ds$. This can be used to make inference about the hazard rate $\alpha(\cdot)$ in the nonparametric case; see Aalen (1978a, b) and Section 3.3. Versions of this result are in Aalen (1978a) and Andersen and Borgan [(1985), Appendix].

2.2. *Results outside model conditions.* Theorem 2.1 will be used in Section 3 to find the limiting null hypothesis distribution of various goodness of test statistics. The present subsection briefly discusses results about the distribution of H_n outside model conditions, which are relevant for power function considerations. We give two results, one pertaining to a sequence of local alternatives and one valid under a fixed alternative.

Our first result requires the Pitman-like framework of local alternatives. The null hypothesis is that $\alpha(\cdot, \theta)$ indeed is the true hazard rate for some appropriate θ_0. Consider some wider $(p + 1)$-dimensional parametric family $\alpha(\cdot, \theta, \eta)$, where $\eta = \eta_0$ corresponds to the simpler family $\alpha(\cdot, \theta)$. Assume that the true hazard rate for model n is

$$(2.9) \quad \alpha_n(s) = \alpha\big(s, \theta_0, \eta_0 + \delta/\sqrt{n}\big) \doteq \alpha(s, \theta_0)\{1 + \phi(s, \theta_0)\delta/\sqrt{n}\},$$

in which $\phi(s, \theta)$ is the derivative of $\log \alpha(s, \theta, \eta)$ w.r.t. η and evaluated at η_0.

THEOREM 2.2. *Under the sequence of local alternatives just described and under the conditions of Theorem* 2.1, H_n *tends in distribution to*

$$H(\cdot) + \delta\bigg[\int_0^{\cdot} k(s, \theta_0)\phi(s, \theta_0)\alpha(s, \theta_0)\, ds$$
$$-B(\cdot)'\Sigma^{-1} \int_0^T y(s)\phi(s, \theta_0)\psi(s, \theta_0)\alpha(s, \theta_0)\, ds\bigg]$$

in the function space $D[0, T]$.

PROOF. Expression (2.7) for H_n can be rewritten by replacing $dN(s)/Y(s) - \alpha(s,\hat{\theta})\,ds$ with

$$dM_n(s)/Y(s) + \{\alpha_n(s) - \alpha(s,\theta_0)\}\,ds - \{\alpha(s,\hat{\theta}) - \alpha(s,\theta_0)\}\,ds,$$

in which $M_n(t) = N(t) - \int_0^t Y(s)\alpha_n(s)\,ds$ is the accompanying true martingale for the n'th model. We can then proceed as in the proof of Theorem 2.1, with a couple of manageable extra complications. We omit details here, but mention that the limiting distribution of $\sqrt{n}\,(\hat{\theta} - \theta_0)$ is needed in this framework of local alternatives. Going carefully through arguments of Borgan [(1984), Section 4] valid for the nonlocal case, which corresponds to $\delta = 0$ in (2.9), one ends up with

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \Sigma^{-1}(U_n + \delta v) + \varepsilon_n \to_d \Sigma^{-1}U + \delta\Sigma^{-1}v \sim N_p(\delta\Sigma^{-1}v, \Sigma^{-1})$$

as the appropriate generalization of (2.6), where v is the vector $\int_0^T y(s)\phi(s,\theta_0)\psi(s,\theta_0)\alpha(s,\theta_0)\,ds$. The technical arguments used also resemble those used by Hjort [(1986), Section 3] in connection with the local asymptotic normality property of the counting process model. □

Assume next that the counting process model has a true, underlying hazard rate $\alpha(\cdot)$ which is not among the parametric candidates $\alpha(\cdot\ \theta)$. Thus, replacing (2.1), there is a true martingale $M(t) = N(t) - \int_0^t Y(s)\alpha(s)\,ds$ with $\mathrm{Var}\{dM(s)|\mathscr{F}_{s-}\} = Y(s)\alpha(s)\,ds$ and the limit in distribution $V(\cdot)$ of $M(\cdot)/\sqrt{n}$ has $\mathrm{Var}\{dV(s)\} = y(s)\alpha(s)\,ds$, instead of (2.2). Using Lenglart's martingale inequality one can show from (2.3) that $(1/n)\log L(\theta)$ converges in probability to the function $\int_0^T y(s)\{\alpha(s)\log\alpha(s,\theta) - \alpha(s,\theta)\}\,ds$. The maximum likelihood estimator $\hat{\theta}$ is consistent for the parameter value θ_0 that maximizes this function, as shown by Hjort [(1986), Section 2.2]. This is a statistically meaningful parameter in that it is *least false*, or most fitting, according to the distance measure

$$\Delta\{\alpha(\cdot),\alpha(\cdot,\theta)\} = \int_0^T y(s)\left[\alpha(s)\{\log\alpha(s) - \log\alpha(s,\theta)\} - \{\alpha(s) - \alpha(s,\theta)\}\right]ds \tag{2.10}$$

between hazard rates. This is a generalization of the Kullback–Leibler information distance to models with censoring; see Hjort (1986).

From (2.7) and with the extra notation introduced above one finds

$$H_n(t)/\sqrt{n} = \int_0^t K_n(s)J(s)\,dM(s)/Y(s) + \int_0^t K_n(s)J(s)\{\alpha(s) - \alpha(s,\hat{\theta})\}\,ds,$$

from which it is not difficult to prove

THEOREM 2.3. *Suppose that the true underlying hazard rate is $\alpha(\cdot)$ and let the conditions of Theorem 2.1 be in force. If there is a unique least false* *parameter value* θ_0 *minimizing* (2.10), *then*

$$\hat{\pi}(t) = H_n(t)/\sqrt{n} \to_p \int_0^t k(s,\theta_0)\{\alpha(s) - \alpha(s,\theta_0)\}\,ds = \pi(t).$$

The notation used here emphasizes that $H_n(\cdot)/\sqrt{n} = \hat{\pi}(\cdot)$ usefully can be thought of as an estimator of the population parameter function $\pi(\cdot)$, where the null hypothesis model amounts to this function being equal to zero. One can for completeness formulate a version of the theorem for the very rare cases of a nonunique θ_0 and can also prove that $H_n(t) - \sqrt{n}\,\pi(t) = \sqrt{n}\{\hat{\pi}(t) - \pi(t)\}$ has a limiting normal distribution, under natural conditions. Both the limit distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ outside model conditions, which can be worked out, and the limit process of $\sqrt{n}\{K_n(s) - k(s,\theta_0)\}$ play important roles here. See also Remark 7B.

Among the consequences of the above is that goodness of fit tests based on H_n generally will be *consistent*, in that any such test will detect any piecewise continuous alternative hazard rate with probability tending to 1 as n increases. The theorem also suggests how to design K_n so as to make possibly interesting departures from the model more easily detectable. If it is not more important to detect $\alpha \neq \alpha(\cdot,\theta)$ in some intervals than in others, then one might use $K_n(s) \equiv 1$. See Section 5 for more precise results.

3. Goodness of fit tests. Theorem 2.1 gave the limiting distribution of the general goodness of fit process $H_n = \int K_n\,dZ_n$ under model conditions. The model should be rejected if H_n is significantly different from zero, as measured by some suitable functional. Natural test statistics could, for example, be $\max_{0 \le t \le T}|H_n(t)|$ or $\int_0^T H_n(t)^2\,dt$. We know that these have limits in distribution $\max_{0 \le t \le T}|H(t)|$ and $\int_0^T H(t)^2\,dt$, respectively, but this can only rarely be utilized. The limit distributions would be very intractable and would also typically depend upon the particular $y(s)$ function. One could estimate $y(s)$ using $Y(s)/n$, but this function varies with the censoring mechanism and from experiment to experiment, making it impossible to construct general tables.

However, perhaps unexpectedly, some interesting explicit results for such Kolmogorov–Smirnov and Cramér–von Mises type statistics are available in the case where the unknown parameter θ is one-dimensional. This one-dimensional case is studied in Section 3.2. Section 3.1, on the other hand, considers χ^2-type statistics based on H_n and a division of the observation period into cells, in the general p-dimensional case. [Creating χ^2 statistics based on such cells and weak convergence is of course a well-known device, which at least goes back to Schoenfeld (1980).] Several interesting special cases come forward. Finally the fully specified case, which is much easier and corresponds to $p = 0$ unknown parameters in the hazard rate of the null hypothesis, is briefly discussed in Section 3.3.

Most of the tests proposed here seem to be new, even for specialized $\alpha(s,\theta)$ models like exponentiality and even in the case of no censoring. A referee has pointed out that the Kolmogorov–Smirnov type test for exponentiality given in Section 4 is essentially identical to one of the total-time-on-test related statistics proposed by Barlow and Campo (1975), which were later justified for use with censored data by Aalen and Hoem (1978). The chi squared tests of Section 3.1 were originally put forward in Hjort (1984, 1985). Akritas (1988) has independently worked out the theory for the important case that corresponds to weight function $K_n(s) = Y(s)/n$, listed as Special Case 2. Akritas used the i.i.d.-like framework of the random censorship model, so our results are in one way more general than his. On the other hand, Akritas also studied the benefits of using a particular minimum chi square estimator for θ and included a simulation study of the test's performance in an example with a fully specified null hypothesis.

Some specializations of the general methods of Section 3 are listed in Section 4. Section 5 provides results about the power functions of the tests, some of which pertain to the problem of choosing an appropriate weight function $K_n(\cdot)$.

3.1. *Chi squared tests.* Let $0 = a_0 < \cdots < a_m = T$ be a division of the observation interval $[0, T]$ into m cells $I_i = (a_{i-1}, a_i]$. Let

$$Q_{n,i} = H_n(a_{i-1}, a_i) = \sqrt{n}\int_{I_i} K_n(s)J(s)\{dN(s)/Y(s) - \alpha(s,\hat{\theta})\,ds\}$$

and $Q_i = H(a_{i-1}, a_i]$, where H is given in Theorem 2.1. Then Q_n with elements $Q_{n,i}$ converges in distribution to Q with elements Q_i, since H has no fixed discontinuities. The covariance matrix of Q is readily found from Theorem 2.1 to be of the form

$$R = \operatorname{Var} Q = D - S'\Sigma^{-1}S. \tag{3.1}$$

Here D is diagonal with elements $d_i = \int_{I_i}\{k(s,\theta_0)^2/y(s)\}\alpha(s,\theta_0)\,ds$, S is the $p \times m$ matrix $(b_1, \dots, b_m)$, in which

$$b_i = B(a_{i-1}, a_i] = \int_{I_i} k(s,\theta_0)\psi(s,\theta_0)\alpha(s,\theta_0)\,ds$$

(cf. Theorem 2.1), and Σ is given in (2.5).

Let now $\hat{R}$ be any consistent estimator of the covariance matrix of Q, say of the form $\hat{D} - \hat{S}'\hat{\Sigma}^{-1}\hat{S}$ with elements $\hat{r}_{i,j} = \hat{d}_i\delta_{i,j} - \hat{b}_i'\hat{\Sigma}^{-1}\hat{b}_j$. Natural choices are

$$\hat{d}_i = \int_{I_i} K_n(s)^2\frac{nJ(s)}{Y(s)}\frac{dN(s)}{Y(s)}, \qquad \hat{b}_i = \int_{I_i} K_n(s)\psi(s,\hat{\theta})\frac{dN(s)}{Y(s)}, \tag{3.2a}$$

$$\hat{\Sigma} = \int_0^T \frac{Y(s)}{n}\psi(s,\hat{\theta})\psi(s,\hat{\theta})'\frac{dN(s)}{Y(s)}. \tag{3.2b}$$

Some options are available here, since what matters is only that $\hat{d}_i$, $\hat{b}_i$ and $\hat{\Sigma}$

are consistent for d_i, b_i and Σ under the null hypothesis that the parametric model is correct. First, $dN(s)/Y(s)$ may everywhere be replaced by $J(s)\alpha(s,\hat{\theta})\,ds$ in the expressions above. Second, under some censoring schemes the limit in probability $y(t)$ of $Y(t)/n$ may be explicitly expressible as a function of θ, say $y(t,\theta)$. In such cases $y(t,\hat{\theta})$ may replace $Y(s)/n$ above. Third, it may be advantageous to use an alternative estimator $(\Sigma^*)^{-1}$ for the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}-\theta_0)$ that is consistent even outside the null hypothesis, i.e., when the parametric model fails. See Remark 7B, where also the benefits of using a model-robust estimator R^* for the full covariance matrix of Q_n are discussed.

Our general test statistic will be

$$X_n^2 = Q_n'\hat{R}^-Q_n = (Q_{n,1},\dots,Q_{n,m})\hat{R}^-(Q_{n,1},\dots,Q_{n,m})'. \tag{3.3}$$

It is clear that $X_n^2 \to_d X^2 = Q'R^-Q$, which is χ^2 distributed with df $=$ Rank(R) degrees of freedom. Here $\hat{R}^-$ and R^- are, if necessary, generalized inverses. In most cases R has full rank m, whereas df $= m-1$ in a particular class of cases, where $H_n(T) = \sum_{i=1}^m Q_{n,i} = 0$. To test the model assumption, choose m intervals, choose a weight function K_n among those allowed by condition (K) preceding Theorem 2.1, and compare X_n^2 above to its approximate χ^2_{df} distribution.

There is a convenient matrix identity that both simplifies the computation and aids the understanding of X_n^2. It is not very difficult to verify that

$$R = D - S'\Sigma^{-1}S \quad \text{implies} \quad R^- = D^{-1} + D^{-1}S'G^-SD^{-1}, \tag{3.4}$$

in which G^- is the $p\times p$ (generalized) inverse of $G = \Sigma - SD^{-1}S'$. This matrix equation is well known in the $p=1$ case, when Σ is simply a scalar, and is then sometimes called Bartlett's identity. The generalization invented for the present occasion leads to

$$\begin{aligned} X_n^2 &= Q_n'\hat{D}^{-1}Q_n + Q_n'\hat{D}^{-1}\hat{S}'\hat{G}^-\hat{S}\hat{D}^{-1}Q_n \\ &= \sum_{i=1}^m \frac{Q_{n,i}^2}{\hat{d}_i} + \left\{\sum_{i=1}^m \frac{Q_{n,i}}{\hat{d}_i}\hat{b}_i\right\}'\hat{G}^-\left\{\sum_{i=1}^m \frac{Q_{n,i}}{\hat{d}_i}\hat{b}_i\right\}, \end{aligned} \tag{3.5}$$

where

$$\hat{G}^- = (\hat{\Sigma} - \hat{S}\hat{D}^{-1}\hat{S}')^- = \left(\hat{\Sigma} - \sum_{i=1}^m \hat{b}_i\hat{b}_i'/\hat{d}_i\right)^-,$$

and there is a similar uncareted expression for X^2. In particular, only a $p\times p$ matrix needs to be inverted.

REMARK. The (generalized) inverse of $G = \Sigma - SD^{-1}S'$ plays a role in the limit distribution X^2 for X_n^2. Let us write $k(s,\theta_0) = y(s)h(s)$ and $d\nu(s) =$

$y(s)\alpha(s, \theta_0)\,ds$. Then G is the sum of m matrices

$$G_i = \int_{I_i} \psi\psi'\,d\nu - \left(\int_{I_i} h\psi\,d\nu\right)\left(\int_{I_i} h\psi\,d\nu\right)' \Big/ \int_{I_i} h^2\,d\nu.$$

From the Cauchy–Schwarz inequality it is seen that $u'G_i u$ always is nonnegative and is zero only if $h(s)$ is proportional to $u'\psi(s, \theta_0)$ on I_i, so G has full rank p unless there are u_i's for which $h(s) = u_i'\psi(s, \theta_0)$ on each I_i. In particular, df $=$ Rank(R) $= m$ in all other cases. In the special lower rank case $h = c'\psi$, which essentially corresponds to $K_n(s) = \{Y(s)/n\}c'\psi(s, \hat{\theta})$, it holds that $u'Gu$ is zero only if $u'\psi(s, \theta_0)$ is proportional to $c'\psi(s, \theta_0)$ on each cell. So when $h = c'\psi$, G will in all but very rare cases have rank $p - 1$ and df $= m - 1$. The model with constant hazards on different cells can have even lower rank.

SPECIAL CASE 1. If $K_n(s) = 1$, then $H_n = Z_n$ and $Q_{n,i} = \sqrt{n}\{\hat{A}(a_{i-1}, a_i] - \int_{I_i} J(s)\alpha(s, \hat{\theta})\,ds\}$. In this case, X_n^2 compares observed and expected cumulative hazards over intervals.

SPECIAL CASE 2. If $K_n(s) = Y(s)/n$, then $Q_{n,i} = (N_i - E_i)/\sqrt{n}$, where $N_i = N(a_{i-1}, a_i]$ counts the number of observed transitions in interval I_i and $E_i = \int_{I_i} Y(s)\alpha(s, \hat{\theta})\,ds$ is a model-based estimate of N_i. So X_n^2 now compares observed and expected number of transitions over intervals. It is interesting to compare the classical Pearson type χ^2 procedure to the present one, in the case of no censoring. Whereas the classical χ^2 uses $E_i = n\int_{I_i} f(x, \tilde{\theta})\,dx$, for a minimum χ^2 estimator $\tilde{\theta}$, our version is *dynamic*, ignores the overall probability of falling in I_i and prefers predicting N_i using the number of items under risk and their combined hazard rate $Y(s)\alpha(s, \hat{\theta})\,ds$; our χ^2 test also allows censoring.

Let us elaborate further. We have $k(s, \theta_0) = y(s)$, and $d_i = \int_{I_i} y(s)\alpha(s, \theta_0)\,ds$ is naturally estimated either by N_i/n or E_i/n, see (3.2) and its following lines. Accordingly, expression (3.5) simplifies to

$$X_n^2 = \sum_{i=1}^{m} \frac{(N_i - E_i)^2}{E_i} + W_n'\hat{G}^- W_n,$$

say, in which $W_n = \sqrt{n}\sum_{i=1}^m \{(N_i - E_i)/E_i\}\hat{b}_i$ and where N_i is equally allowable in the denominator. The limit distribution is χ_m^2 except in cases where $\hat{G}$ has lower rank than p and then $X_n^2 \to_d \chi_{m-1}^2$, with a further exception; see the preceding remark and Special Case 3. Let us also point out that a very simple but slightly conservative test procedure is to reject the model if $X_{0,n}^2 = \sum_{i=1}^m (N_i - E_i)^2/E_i$ exceeds the upper ε point of the χ_{df}^2. Finally, in the fully specified case, which corresponds to $p = 0$ unknown parameters and where $E_i = \int_{I_i} Y(s)\alpha_0(s)\,ds$, there is no second term and $X_n^2 = X_{0,n}^2$.

SPECIAL CASE 3. Now try $K_n(s) = \{Y(s)/n\}c'\psi(s, \hat{\theta})$ for some coefficients $c_1, \ldots, c_p$. This choice is feasible according to Theorem 2.1. Then $\sum_{i=1}^m Q_{n,i} = 0$

by the definition of the maximum likelihood estimator and R has rank $m-1$, as explained in the preceding remark. If $c=(1,0,\dots,0)'$, for example, then $Q_{n,i}=(1/\sqrt{n})\int_{I_i}\psi_1(s,\hat\theta)\{dN(s)-Y(s)\alpha(s,\hat\theta)\,ds\}$ and one can derive a simplified expression for (3.5), involving the elements of $\hat\Sigma_i=\int_{I_i}\psi(s,\hat\theta)\psi(s,\hat\theta)'\,dN(s)/n$ written in block notation:

$$X_n^2=\sum_{i=1}^m\frac{Q_{n,i}^2}{\hat\sigma_{i,11}}+\left(\sum_{i=1}^m\frac{Q_{n,i}}{\hat\sigma_{i,11}}\hat\Sigma_{i,12}\right)\left\{\sum_{i=1}^m\left(\hat\Sigma_{i,22}-\hat\Sigma_{i,21}\hat\sigma_{i,11}^{-1}\hat\Sigma_{i,12}\right)\right\}^{-1}\times\left(\sum_{i=1}^m\frac{Q_{n,i}}{\hat\sigma_{i,11}}\hat\Sigma_{i,21}\right).$$

In particular only a $(p-1)\times(p-1)$ matrix needs to be inverted, and in the $p=1$ case the second term vanishes; see (3.8).

Unlike the classical χ^2 tests, our test statistics (3.5) accept censored data and do not require special estimators for their use, but simply the maximum likelihood estimator, computed from the original, ungrouped data.

The problem of choosing the m cells in (3.5) is of course present. An old and very conservative rule of thumb from the traditional χ^2 tests, stating that each cell should contain at least five observations, seems reasonable here too; see also Remark 7H. Akritas (1988), who independently of the present author developed the tests given as Special Case 2, investigated the actually achieved level in a simulation study, but only in a situation with a fully specified null hypothesis hazard rate.

The versatility of the class of tests is illustrated by Special Cases 1–3. The choice of weight function K_n is up to the user, for whom the power considerations in Section 5 should be pertinent.

3.2. *The one-dimensional case.* The case where the parameter θ of the model is one-dimensional deserves special attention. Before turning to continuous type test statistics, let us record some explicit simplifications of the X_n^2 statistics considered above. We have

$$X_n^2=\sum_{i=1}^m\frac{Q_{n,i}^2}{\hat d_i}+\left\{\sum_{i=1}^m\frac{Q_{n,i}\hat b_i}{\hat d_i}\right\}^2\Bigg/\left\{\hat\sigma^2-\sum_{i=1}^m\frac{\hat b_i^2}{\hat d_i}\right\},\tag{3.6}$$

where $\hat\sigma^2=\int_0^T\psi(s,\hat\theta)^2\,dN(s)/n$ or some other estimator of $\sigma^2=\int_0^Ty(s)\psi(s,\theta_0)^2\alpha(s,\theta_0)\,ds$ that is consistent under the model.

For the rest of this subsection we will stick to the important special case already touched upon in the preceding remark, where $K_n(s)$ is chosen to be $\{Y(s)/n\}\psi(s,\hat\theta)$, so that

$$H_n(t)=\left(1/\sqrt{n}\right)\int_0^t\psi(s,\hat\theta)\{dN(s)-Y(s)\alpha(s,\hat\theta)\,ds\}.\tag{3.7}$$

In this case, $Q_{n,i} = (1/\sqrt{n})\int_{I_i}\psi(s,\hat\theta)\{dN(s) - Y(s)\alpha(s,\hat\theta)\,ds\}$, $k(s,\theta_0) = y(s)\psi(s,\theta_0)$ and the elements of $R = \operatorname{Var} Q$ can be written $r_{i,j} = \sigma^2(p_i\delta_{i,j} - p_ip_j)$, in which $p_i = b_i/\sigma^2 = B(I_i)/B([0,T])$; cf. the notation in Theorem 2.1. So in this case R has rank $m-1$ and

$$X_n^2 = Q_n'\hat R^- Q_n = \sum_{i=1}^m Q_{n,i}^2/\hat d_i \to_d \chi_{m-1}^2. \tag{3.8}$$

In fact, fuller information awaits us in this case, as H_n behaves asymptotically as a time-transformation of a Brownian bridge W^0. For

$$\operatorname{Cov}\{H(t_1), H(t_2)\} = \sigma^2\{p(t_1\wedge t_2) - p(t_1)p(t_2)\},$$

where $p(t) = B(t)/\sigma^2 = B(t)/B(T)$, which means that

$$H_n \to_d H = \sigma W^0(p(\cdot)) \quad \text{in } D[0,T]. \tag{3.9}$$

Accordingly, both Kolmogorov–Smirnov and Cramér–von Mises type test statistics can be constructed in the one-parameter case.

For example, $\max_{0\le t\le T}|H_n(t)|/\hat\sigma \to_d \max_{0\le s\le 1}|W^0(s)|$, so rejecting the parametric model if

$$\max_{0\le t\le T}\left|\int_0^t \psi(s,\hat\theta)\{dN(s) - Y(s)\alpha(s,\hat\theta)\,ds\}\right| > 1.36\hat\sigma\sqrt{n}$$

constitutes a test with asymptotic level 5%. This test is universally consistent by an application of Theorem 2.3. One may also construct weighted versions of this test criterion, for example, dividing by an estimate of the limiting standard deviation.

Another test that is consistent against each piecewise continuous alternative hazard rate α can be constructed, utilizing the fact that

$$\int_0^T H_n(t)^2\,d\hat B(t)/\hat\sigma^2 \to_d \int_0^T \sigma^2 W^0(p(t))^2\,dB(t)/\sigma^2 =_d \sigma^2\int_0^1 W^0(s)^2\,ds,$$

where $d\hat B(t)$ could be either $\psi(t,\hat\theta)^2\,dN(t)/n$ or $\{Y(t)/n\}\psi(t,\hat\theta)^2\alpha(t,\hat\theta)\,dt$. So, for example, one can reject the model whenever $\int_0^T H_n(t)^2\psi(t,\hat\theta)^2 dN(t) > n\lambda_\varepsilon\hat\sigma^4$, where λ_ε is the upper ε point of the distribution of $\int_0^1 W^0(s)^2\,ds$.

We record two variations on this theme, both leading to goodness of fit tests with good overall properties. Let for convenience $\hat p(t) = \hat B(t)/\hat\sigma^2$ and $d\hat p(t) = d\hat B(t)/\hat\sigma^2$. First, the null hypothesis distribution of $\int_0^T|H_n(t)|\,d\hat p(t)/\hat\sigma$ tends to that of $\int_0^T\sigma|W^0(p(t))|\,dp(t)/\sigma = \int_0^1|W^0(s)|\,ds$. The latter has been found by Shepp (1982), and a table can be found in Johnson and Killeen (1983). Next, the limiting variance for $H_n(t)$ is $\sigma^2 p(t)(1-p(t))$ and it is natural to try to weight H_n in an Anderson–Darling manner. One can show that the limiting null distribution of

$$A_n^2 = \int_0^T \frac{H_n(t)^2}{\hat p(t)(1-\hat p(t))}\frac{d\hat p(t)}{\hat\sigma^2}$$

is that of

$$\int_0^1 \sigma^2 W^0(p(t))^2/\{p(t)(1-p(t))\}\, dp(t)/\sigma^2 = \int_0^1 W^0(s)^2/\{s(1-s)\}\, ds.$$

Rejecting the model if $A_n^2 > 2.50$ gives, for example, an asymptotic level 5% test; tables are available in Lewis (1961).

3.3. *The fully specified case.* It is sometimes of interest to test the hypothesis $\alpha(s) = \alpha_0(s)$ for s in $[0, T]$, or perhaps only on a subinterval, where α_0 is fully specified. For example, the group of individuals under study might be compared to some established average hazard curve for a larger population.

This corresponds to the case of $p = 0$ unknown parameters in the model and is covered by our earlier efforts; see the remark ending Section 2.1. The process $H_n(t) = \sqrt{n}\int_0^t K_n J(dN/Y - \alpha_0\, ds)$ converges in distribution to a time-transformed Brownian motion process $H(t) = W(\tau^2(t))$, where $\tau^2(t) = \operatorname{Var} H(t) = \int_0^t \{k(s)^2/y(s)\}\alpha_0(s)\, ds$. A consistent estimator of the latter is

$$\hat{\tau}^2(t) = \int_0^t K_n(s)^2 \frac{nJ(s)}{Y(s)} \frac{dN(s)}{Y(s)}, \tag{3.10}$$

where one alternatively could use $\alpha_0(s)\, ds$ instead of $dN(s)/Y(s)$. It is therefore easy to construct a variety of large-sample tests, for example, of the Kolmogorov–Smirnov and Cramér–von Mises types, as in the previous subsection. We mention here but two examples of this sort. First,

$$\frac{\max_{0\le t\le T} |H_n(t)|}{\hat{\tau}(T)} \to_d \frac{\max_{0\le t\le T} |W(\tau^2(t))|}{\tau(T)} =_d \max_{0\le s\le 1} |W(s)|, \tag{3.11}$$

under the null hypothesis. A derivation of the distribution of the limit variable can be found in Billingsley [(1968), Section 11] and a table in Walsh [(1962), page 334]. Second,

$$\max_{a\le t\le c} \frac{|H_n(t)|}{\hat{\tau}(t)} \to_d \max_{a\le t\le c} \frac{|W(\tau^2(t))|}{\tau(t)} =_d \max_{\tau^2(a)/\tau^2(c)\le s\le 1} \frac{|W(s)|}{\sqrt{s}}, \tag{3.12}$$

providing a naturally weighted version of the first test. One must use a positive a here. Upper quantiles for the last limit distribution have been obtained via simulation in a study by Gringorten (1968). Easy to use approximations to such quantiles can be constructed from Miller and Siegmund (1982) and an explicit expression for the exact limit distribution has been obtained in an unpublished report of Hjort.

Let us finally provide a χ^2-type test for $\alpha = \alpha_0$. Divide the time observation period into m intervals $(a_{i-1}, a_i]$ once more. Then

$$X_n^2 = \sum_{i=1}^m H_n(a_{i-1}, a_i]^2/\hat{d}_i \to_d \chi_m^2, \tag{3.13}$$

under null hypothesis conditions, where $\hat{d}_i$ is a consistent estimator for $d_i = \tau^2(a_{i-1}, a_i] = \int_{I_i}(k^2/y)\alpha_0\, ds$; cf. the two possibilities for $\hat{\tau}^2(\cdot)$ noted above.

The classes of tests derived in this way are very rich, since we are free to choose $K_n(\cdot)$; see Special Cases 1–3 in Section 3.1. As explained in Special Case 2, the particular choice $K_n(s) = Y(s)/n$ leads to the familiar expression $X_n^2 = \sum_{i=1}^m (N_i - E_i)^2/E_i$. This test was independently developed by Akritas (1988). Some guidelines for choosing K_n based on power considerations are offered in Section 5.

4. Some applications.

4A. The most important parametric model for a hazard rate is of course $\alpha(s,\theta) = \theta$, a constant hazard, corresponding to an exponential lifetime distribution in that framework. It is easy to write down a variety of tests for this model assumption or for the slightly more general $\alpha(s,\theta) = \theta\alpha_0(s)$, where α_0 is specified. The maximum likelihood estimator is $\hat{\theta} = N(T)/R(T)$, where $R(t) = \int_0^t Y(s)\alpha_0(s)\,ds$ and σ^2 of (2.5) becomes $r(T)/\theta$, where $r(t) = \int_0^t y(s)\alpha_0(s)\,ds$ is the limit in probability of $R(t)/n$. The χ^2 tests Special Cases 2 and 3 both simplify to $X_n^2 = \sum_{i=1}^m (N_i - E_i)^2/E_i$, where $E_i = \int_{I_i} Y\hat{\theta}\alpha_0\,ds$ is the dynamic model-based estimate of N_i. X_n^2 tends to χ^2_{m-1} under the model.

Let us also present a Kolmogorov–Smirnov type test. Following the procedure of Section 3.2, let $H_n(t) = \sqrt{n}\,\hat{\pi}(t)$, where $\hat{\pi}(t) = \{N(t) - \hat{\theta}R(t)\}/n\hat{\theta}$ consistently estimates $\pi(t) = \int_0^t y(s)\{\alpha(s) - \theta_0\}\,ds/\theta_0$ under a priori circumstances and $\theta_0 = \int_0^T y\alpha\,ds / \int_0^T y\,ds$ is least false. Let $D_n = \max_{0\le t\le T}|\hat{\pi}(t)|$. The test rejects the model if $\sqrt{n}\,D_n/\hat{\sigma}$ exceeds the upper ε point of the distribution of $\max_{0\le s\le 1}|W^0(s)|$; see also Section 5.3.

4B. Assume that each individual of a population has a constant hazard rate, but that these individual rates vary according to a gamma distribution with parameters $(\theta/\eta, 1/\eta)$. Then the life length of a randomly chosen individual can be seen to have a distribution with hazard rate $\alpha(s) = \theta/(1+\eta s)$. This is an important model which seems to explain many phenomena in biostatistics; see Aalen (1982). To test this model, or more generally $\alpha(s) = \theta g(s,\eta)$, where g is specified, one might use the χ^2 test given as Special Case 3 of Section 3.1. In the notation of that section, $\psi_1(s,\theta,\eta) = 1/\theta$ and $\psi_2(s,\theta,\eta) = \partial \log g(s,\eta)/\partial\eta = \psi_2(s,\eta)$ and one arrives at

$$X_n^2 = \frac{\sum_{i=1}^m (N_i - E_i)^2}{E_i} + \left(W_n^0\right)^2 F_n, \tag{4.1}$$

where $E_i = \int_{I_i} Y(s)\hat{\alpha}\,ds$, $W_n^0 = \sqrt{n}\sum_{i=1}^m \{(N_i - E_i)/E_i\}\int_{I_i}\hat{y}\hat{\psi}_2\hat{\alpha}\,ds$ and $F_n = \int_0^T \hat{y}(\hat{\psi}_2)^2\hat{\alpha}\,ds - \sum_{i=1}^m (\int_{I_i}\hat{y}\hat{\psi}_2\hat{\alpha}\,ds)^2/\{E_i/n\}$, using $\hat{y}$ for $Y(s)/n$, $\hat{\alpha}$ for $\alpha(s,\hat{\theta},\hat{\eta})$, etc. Under the model X_n^2 goes to a χ^2_m.

4C. The Weibull model $\alpha(s) = \theta^\eta \eta s^{\eta-1}$ can be tested similarly, with an X_n^2 as in (4.1) and the very same expressions for E_i and W_n^0, but for the fact that $\hat{\psi}_2$ now means $\psi_2(s,\hat{\theta},\hat{\eta}) = 1/\hat{\eta} + \log(s\hat{\theta})$.

4D. The gamma distribution and the log-normal distribution for life lengths can also be tested similarly, most conveniently using Special Case 3 of Section 3, to avoid matrix inversion. It would however require numerical integration over cells in each case.

4E. Suppose that α_0 is normal but that one suspects that a tiny proportion of the population may have some substantially higher hazard rate $\alpha_1(\cdot)$. The population probability density is $f = (1-\varepsilon)f_0 + \varepsilon f_1$, say, and the hazard rate for a randomly chosen individual is $\alpha(s,\varepsilon) = \pi_0(s)\alpha_0(s) + \pi_1(s)\alpha_1(s)$, where $\pi_0(s) = (1-\varepsilon)F_0[s,\infty)/\{(1-\varepsilon)F_0[s,\infty) + \varepsilon F_1[s,\infty)\}$ and $\pi_0(s) + \pi_1(s) = 1$. This is but one example of interesting mixture models for lifetimes and its adequacy can be tested using the methods of earlier sections. The structurally simplest case has three parameters, namely constant hazards θ_0 and θ_1 in addition to the mixture proportion ε.

5. Power considerations. This section reports on a brief investigation into the power properties of the various goodness of fit tests. Sections 5.1 and 5.2 study local asymptotic power, first in the case of a fully specified null hypothesis and then in the parametric case, whereas Section 5.3 considers fixed alternatives to the model hypothesis. In addition to actually giving approximations to the power of the tests, the results below are also relevant for the problem of specifying the weight function K_n and in fact some optimality results are established for local power.

5.1. *Local asymptotic power in the fully specified case.* Let us start with the simplest case, which is the $p = 0$ case of a completely specified null hypothesis $\alpha = \alpha_0$; see Section 3.3. Consider a contiguous sequence of alternative hazard rates, of the form

$$\alpha_n(s) = \alpha(s, \eta_0 + \delta/\sqrt{n}) \doteq \alpha_0(s)\{1 + \phi(s)\delta/\sqrt{n}\}, \tag{5.1}$$

where $\alpha_0(s) = \alpha(s,\eta_0)$ and $\phi(s) = \partial \log \alpha(s,\eta_0)/\partial\eta$. This is as in (2.9), but actually simpler. Let H_n and H be as in Section 3.3; H_n depends upon a predictable weight function K_n and this function's limit in probability k features in H. The proof of Theorem 2.2 can be used to show that H_n tends in distribution to $H(\cdot) + \delta\int_0^{\cdot} k(s)\phi(s)\alpha_0(s)\,ds$.

How should $K_n(\cdot)$ be chosen in order to achieve high local power for a test based upon H_n? The simplest such test is based on $H_n(a,b] = H_n(b) - H_n(a)$, i.e., the increment in H_n over a single cell, and rejects the model if $|H_n(a,b]|/\{\hat\tau^2(b) - \hat\tau^2(a)\}^{1/2}$ exceeds $z_{\varepsilon/2}$, the upper $\varepsilon/2$ point of the standard normal, where $\hat\tau(t)$ is given in (3.10).

PROPOSITION 5.1. *The local asymptotic power is*

$$\Pr\{\text{reject}|\alpha_n(\cdot)\} \to \Pr\{\chi_1^2(\delta^2\lambda(a,b]) > z_{\varepsilon/2}^2\}, \tag{5.2}$$

where the eccentricity parameter of the χ_1^2 *distribution involves*

$$(5.3)\quad \lambda(a,b] = \left[\int_{(a,b]} k(s)\phi(s)\alpha_0(s)\,ds\right]^2 \Big/ \int_{(a,b]} \{k(s)^2/y(s)\}\alpha_0(s)\,ds.$$

The choice of $K_n(s)$ *that brings optimal asymptotic power against the alternatives* (5.1) *is* $K_n(s) = \{Y(s)/n\}\phi(s)$.

PROOF. Expressions (5.2) and (5.3) follow readily from $H_n \to_d H + \delta\int_0^{\cdot} k\phi\alpha_0\,ds$, upon using the fact that $\hat{\tau}^2(t)$ converges in probability to $\tau^2(t) = \int_0^t (k^2/y)\alpha_0\,ds$ not only under the null hypothesis model but also along the sequence of models (5.1). Maximizing the local asymptotic power amounts to maximizing $\lambda = \lambda(a,b]$. But this can easily be done using the Cauchy–Schwarz inequality. Let us for clarity write $k(s) = y(s)h(s)$ and $d\nu(s) = y(s)\alpha_0(s)\,ds$, which leads to the expression $\lambda = (\int_{(a,b]} h\phi\,d\nu)^2/\int_{(a,b]} h^2\,d\nu$. The optimizing choice is h proportional to ϕ, i.e., $k(s) = y(s)\phi(s)$, and we are led to $K_n(s) = \phi(s)Y(s)/n$. □

Next, consider the more serious χ^2 test X_n^2 of (3.13). It is not difficult to see that X_n^2 converges to a noncentral $\chi_m^2(\delta^2\lambda)$ distribution, along the alternatives (5.1), where in fact $\lambda = \sum_{i=1}^m \lambda(a_{i-1}, a_i]$. Since the best choice is $k(s) = y(s)\phi(s)$ on each interval, this is also true for the sum. Accordingly, one should use $X_n^2 = \sum_{i=1}^m Q_{n,i}^2/\hat{d}_i$, with $Q_{n,i} = \int_{I_i}\phi(s)\{dN(s) - Y(s)\alpha_0(s)\,ds\}$ and $\hat{d}_i = \int_{I_i} Y(s)\phi(s)^2\alpha_0(s)\,ds/n$ to detect alternatives (5.1), and the optimal local power is $\Pr\{\chi_m^2(\delta^2\int_0^T y\phi^2\alpha_0\,ds) > \gamma_{m,\varepsilon}\}$, where $\gamma_{m,\varepsilon}$ is the upper ε point of the χ_m^2.

EXAMPLE 1. Suppose that the null hypothesis specifies a constant hazard rate θ_0. Other constant hazards θ as alternatives correspond to $\phi(s)$ also being constant and $K_n(s) = Y(s)/n$ is optimal. This is Special Case 2 of Sections 3.1 and 3.3 and accordingly the χ_m^2 statistic $X_n^2 = \sum_{i=1}^m (N_i - E_i)^2/E_i$ is optimal for detecting other constant hazards. This is also the test independently proposed by Akritas (1988). The optimality property just derived explains theoretically the simulation results of Akritas' Table 1.

EXAMPLE 2. Weibull alternatives correspond to cumulative hazard rate $(\theta_0 t)^\eta$ or hazard rate $\theta_0^\eta \eta s^{\eta-1}$. The optimal weight function is $K_n(s) = \{Y(s)/n\}\{1 + \log(\theta_0 s)\}$.

EXAMPLE 3. Next consider $\alpha(s) = \theta_0/(1 + \eta s)$, as in 4B. To detect such alternatives to a constant hazard rate, which is the degenerate case $\eta = 0$, one should use $K_n(s) = sY(s)/n$.

EXAMPLE 4. Consider the situation of 4E. How should K_n be chosen in order to detect such alternatives to the null hypothesis α_0? Following the general procedure one finds that $\phi(s)$, the derivative of $\log\alpha(s,\varepsilon)$ w.r.t. ε and evaluated at zero, is equal to $\exp[-\{A_1(s) - A_0(s)\}]\{\alpha_1(s) - \alpha_0(s)\}/\alpha_0(s)$. If

$\alpha_0(s) = \theta_0$ is the null hypothesis hazard and $\alpha_1(s) = \theta_1$, for example, then the optimal choice is $K_n(s)$ equal to (or proportional to) $\exp\{-(\theta_1 - \theta_0)s\}Y(s)/n$. This means giving much more weight to the shorter life lengths than the longer ones, which is natural, in that individuals with hazard rate α_1 are those that die first.

EXAMPLE 5. Consider the gamma distribution density $f(t, \theta, \eta) = \{\theta^n/\Gamma(\eta)\}t^{\eta-1}\exp(-\theta t)$. The best choice for K_n, in order to find gamma alternatives to the exponential with parameter θ_0, which is the special case $\eta = 1$, can be shown to be $K_n(s) = q(s\theta_0)Y(s)/n$, in which $q(t) = E\{\log(X/t)|X \geq t\} = \int_t^\infty \log(x/t)e^{-x}\,dx/\int_t^\infty e^{-x}\,dx$ for X a unit exponential.

One can also derive expressions for the local asymptotic power of the Kolmogorov–Smirnov and Cramér–von Mises type statistics mentioned in Section 3.3. In the first case these expressions involve crossing probabilities for Brownian motion with nonlinear boundaries and in the second case distributions of integrals of squares of shifted Brownian motion. Accordingly the local asymptotic power would in general have to be computed by means of nontrivial numerical devices, for each $k(\cdot)$ and each δ, making general results very difficult to obtain.

In a couple of instances explicit optimality results are obtainable, however. Consider the test statistic of (3.12), defined as the maximum over an interval $[a, c]$ of a weighted version of H_n. One can show that it converges in distribution along the path of (5.1) to

$$\max_{\tau^2(a)\leq s\leq\tau^2(c)}\left|W(s)/\sqrt{s} + \delta\int_0^{(\tau^2)^{-1}(s)} k\phi\alpha_0\,du/\sqrt{s}\right|,$$

and use this to demonstrate that the best choice of $k(s)$, also for this test statistic, is $y(s)\phi(s)$. The optimal local power is

$$\Pr\left\{\max_{\tau^2(a)/\tau^2(c)\leq t\leq 1}\left|W(t)/\sqrt{t} + \delta\tau(c)\sqrt{t}\right| > c_\varepsilon\right\}, \tag{5.4}$$

where c_ε is the appropriate quantile of the distribution on the right-hand side of (3.12).

We should stress that not every test based on H_n will achieve its optimal local power against (5.1) alternatives by using $\phi(s)Y(s)/n$ for $K_n(s)$. The $\{\max_{0\leq t\leq T}|H_n(t)|\}/\hat{\tau}(T)$ test of (3.12) is a case in point.

REMARK. The important remaining problem is to compare the power performance of the best tests proposed here with that of other classical tests. One might, for example, compare the local power given in (5.4), which can be computed numerically using methods of Durbin (1971) and Folkeson [(1984), Chapter 2], with similar expressions for the local power of Kolmogorov–Smirnov or Cramér–von Mises tests. This is not pursued here. We point out, however, that our new χ^2 tests beat the classical ones in many cases. Let us illustrate this in a simple situation. Assume the problem is to test $\alpha(s) = \theta_0$

on $[0, T]$ for a large T based on survival times with no censoring. Then the classical Pearson χ^2_{m-1} is $Z_n = \sum_{i=1}^m (N_i - np_{i,0})^2/np_{i,0}$, where $p_{i,0} = \int_{I_i} f_0(x)\,dx$ and f_0 is the exponential density $\theta_0 \exp(-\theta_0 x)$ and $N_i = N(a_{i-1}, a_i]$ is the number of observed failures in cell number i. One can now work out an expression for the local asymptotic power of the Z_n test, along the Pitman path (5.1) of alternatives. In general terms it becomes $\Pr\{\chi^2(\delta^2\lambda_{\mathrm{KP}}) > \gamma_{m-1,\varepsilon}\}$, where $\lambda_{\mathrm{KP}} = \sum_{i=1}^m c_i^2/p_{i,0}$ and $c_i = d(a_{i-1})\exp\{-A_0(a_{i-1})\} - d(a_i)\exp\{-A_0(a_i)\}$, writing $d(t)$ for $\int_0^t \phi(s)\alpha_0(s)\,ds$. If the alternatives are other constant hazards $\theta_0(1 + \delta/\sqrt{n})$, then $c_i = a_{i-1}\theta_0 \exp(-a_{i-1}\theta_0) - a_i\theta_0 \exp(-a_i\theta_0)$ and one can show that λ_{KP} is always less than $\lambda_{\mathrm{new}} = \int_0^T y(s)\,ds = 1 - \exp(-\theta_0 T)$. Accordingly the new test $X_{0,n}^2 = \sum_{i=1}^m (N_i - E_i)^2/E_i$, which uses $E_i = \int_{I_i} Y(s)\alpha_0(s)\,ds$ and even has an additional degree of freedom, is stronger than the classical Z_n, which uses $E_i = np_{i,0}$. This again provides a theoretical explanation for the simulation results of Akritas [(1988), Table 1]. Akritas' Table 2 indicates that $X_{0,n}^2$ also is better than Z_n against Weibull alternatives. Example 2 shows that X_n^2 with the K_n given there performs even better. Further research could compare the new tests with also the more general χ^2 tests of Cressie and Read (1984).

5.2. *Local asymptotic power in the parametric case.* Let us next turn to the case of a parametric null hypothesis. It is reasonably straightforward to obtain expressions for the asymptotic local power of the various tests, employing Theorem 2.2, but it becomes harder to find explicit optimality results than in the fully specified case considered above.

Encapsule the null model $\alpha(s, \theta)$ in a larger family $\alpha(s, \theta, \eta)$, where $\eta = \eta_0$ gives the original model. We are interested in local power for tests, along the Pitman path of alternatives (2.9). The simplest test based on H_n of (2.7) is the ratio $|H_n(I)|/\{\hat{d}(I) - \hat{B}(I)\hat{\Sigma}^{-1}\hat{B}(I)\}^{1/2}$ based on a single cell $I = (a, b]$ in the notation of Section 3. From Theorem 2.2 the asymptotic local power is found to be as in (5.2), but with a more complicated expression than (5.3) for λ, namely,

$$\lambda = \frac{\left[\int_I k\phi\alpha\,ds - \left(\int_I k\psi\alpha\,ds\right)'\Sigma^{-1}\left(\int_0^T y\phi\psi\alpha\,ds\right)\right]^2}{\int_I (k^2/y)\alpha\,ds - \left(\int_I k\psi\alpha\,ds\right)'\Sigma^{-1}\left(\int_I k\psi\alpha\,ds\right)}, \tag{5.5}$$

in which α is shorthand for $\alpha(s, \theta_0)$, etc.

PROPOSITION 5.2. *The function k that maximizes the noncentrality parameter λ of* (5.5) *is $k(s) = y(s)h_0(s) = y(s)\{\phi(s, \theta_0) - c_0'\psi(s, \theta_0)\}$, where $c_0 = (\int_{I^c} y\psi\psi'\alpha\,ds)^{-1}\int_{I^c} y\phi\psi\alpha\,ds$ and I^c denotes the complement $[0, T] - I$. The choice of K_n that yields optimal local power against alternatives* (2.9) *for the single-cell test described above is*

$$K_n(s) = \{Y(s)/n\}\{\phi(s, \hat{\theta}) - (\hat{c}_0)'\psi(s, \hat{\theta})\},$$

where $\hat{c}_0$ is the plug-in estimator for c_0.

PROOF. It is again helpful to write $k = yh$ and $d\nu(s) = y(s)\alpha(s, \theta_0)\,ds$. Then λ can be written $[o]^2/[u]$, where

$$[o] = \int_I h\phi\,d\nu - \left(\int_I h\psi\,d\nu\right)'\Sigma^{-1}\left(\int_0^T \phi\psi\,d\nu\right)$$

and

$$[u] = \int_I h^2\,d\nu - \left(\int_I h\psi\,d\nu\right)'\Sigma^{-1}\left(\int_I h\psi\,d\nu\right).$$

Let us define $(g, h) = \int_I gh\,d\nu - (\int_I g\psi\,d\nu)'\Sigma^{-1}(\int_I h\psi\,d\nu)$, for any pair of functions g and h for which the integrals exist. This actually defines a bona fide inner product, satisfying the Cauchy–Schwarz inequality. Furthermore, $[u] = (h, h)$ and inspection shows that $[o]$ can be written $(h, \phi - c'\psi)$, irrespective of h, when c is chosen as $c_0 = (\int_{I^c}\psi\psi'\,d\nu)^{-1}\int_{I^c}\phi\psi\,d\nu$. Accordingly λ is maximal when h is proportional to $h_0 = \phi - c_0'\psi$, as claimed. The weight process K_n given in the proposition is consistent for $k = yh_0$ and obeys regularity condition (K) of Theorem 2.1. □

It is slightly awkward that the best choice K_n depends upon the cell I. If I is a small cell, an approximation to the best k is

$$k^*(s) = y(s)h^*(s) = y(s)\{\phi(s, \theta_0) - (c^*)'\psi(s, \theta_0)\}, \tag{5.6}$$

where $c^* = \Sigma^{-1}\int_0^T y\phi\psi\alpha\,ds$, with an accompanying proposal for $K_n(s)$; see (5.7).

Consider next the full χ^2 test (3.3). Using Theorem 2.2 again it follows that $X_n^2 \to_d \chi^2_{\text{df}}(\delta^2\lambda)$, where $\lambda = g'R^-g$ and g is the vector with components

$$\begin{aligned} g_i &= \int_{I_i} k\phi\alpha\,ds - b_i'\Sigma^{-1}\int_0^T y\phi\psi\alpha\,ds \\ &= \int_{I_i} h\phi\,d\nu - \left(\int_{I_i} h\psi\,d\nu\right)'\Sigma^{-1}\left(\int_0^T \phi\psi\,d\nu\right) = \int_{I_i} hh^*\,d\nu, \end{aligned}$$

again writing $k = yh$, and using h^* of (5.6). Upon using (3.4), it follows that

$$\begin{aligned} \lambda &= \sum_{i=1}^m \frac{\left(\int_{I_i} hh^*\,d\nu\right)^2}{\int_{I_i} h^2\,d\nu} \\ &\quad + \left\{\sum_{i=1}^m \frac{\int_{I_i} hh^*\,d\nu}{\int_{I_i} h^2\,d\nu}\int_{I_i} h\psi\,d\nu\right\}' G(h)^- \left\{\sum_{i=1}^m \frac{\int_{I_i} hh^*\,d\nu}{\int_{I_i} h^2\,d\nu}\int_{I_i} h\psi\,d\nu\right\}, \end{aligned}$$

where $G(h)^-$ is the (generalized) inverse of $\Sigma - SD^{-1}S' = \sum_{i=1}^m[\int_{I_i}\psi\psi'\,d\nu - (\int_{I_i} h\psi\,d\nu)(\int_{I_i} h\psi\,d\nu)'/\int_{I_i} h^2\,d\nu]$. Choosing $h = h^*$ maximizes the first term and makes the second term vanish. I have not been able to prove that h^* really maximizes λ, but it is at least possible to show that λ, as a function of the possible choice $h = h^* + (\varepsilon_i)'\psi$ on I_i, has a local maximum for $\varepsilon_1 = \cdots = \varepsilon_m = 0$. This indicates that h^* is a very good choice for h and we propose

$$K_n(s) = \{Y(s)/n\}\{\phi(s, \hat\theta) - \hat c'\psi(s, \hat\theta)\}, \tag{5.7}$$

where $\hat c = \hat\Sigma^{-1}\int_0^T \hat y\hat\phi\hat\psi\hat\alpha\,ds$ estimates c^*.

EXAMPLE 1. An omnibus test for exponentiality, which at the same time is clever at detecting alternatives of the form $\alpha(s) = \theta\{1 + \phi(s, \theta)\delta\}$ for small δ, uses an appropriate estimated version of $k(s) = y(s)\{\phi(s, \theta) - \bar{\phi}(\cdot, \theta)\}$, where $\bar{\phi}(\cdot, \theta) = \int_0^T y(s)\phi(s, \theta)\,ds / \int_0^T y(s)\,ds$, for $K_n(s)$. This is seen following the general program above. Weibull alternatives correspond to $\phi(s, \theta) = 1 + \log(\theta s)$ and gamma distribution alternatives correspond to $\phi(s, \theta) = -q(s\theta)$, where q is the function given in Example 5 of the previous subsection. Several interesting models have $\alpha(s) = \theta g(s, \eta)$ for some specified g, where some η_0 gives the constant hazard rate case. An example is the gamma frailty mixture case $\theta/(1 + \eta s)$ considered in Example 3 of the previous subsection. Then $\phi(s,\theta) = \partial \log g(s, \eta_0)/\partial\eta = \phi(s)$ is independent of θ; in the example it is equal to $-s$. In this case, an empirical version of $y(s)\{\phi(s) - \bar{\phi}\}$ is used for $K_n(s)$.

EXAMPLE 2. One can use the general method to establish a test for Weibull cumulative hazards $(\theta t)^\beta$, which is strong against three-parameter alternatives $(\theta t + \eta)^\beta$, and one may similarly devise a K_n weight function which makes a test for exponentiality strong against a ray through or a mixture of this three-parameter alternative model, etc.

5.3. *Power against fixed alternatives.* Here we shall only outline results that can be obtained. Let us first consider the χ^2 test X_n^2 of (3.3). We are interested in the distribution of X_n^2 outside the null model, say under a general hazard rate $\alpha(\cdot)$ not belonging to the parametric family. It can be written $X_n^2 = Q_n' \hat{R}^- \hat{Q}_n = n\hat{\pi}'\hat{R}^-\hat{\pi}$, where $\hat{\pi}$ is the vector with components $\hat{\pi}_i = Q_{n,i}/\sqrt{n}$. From Theorem 2.3 it is seen that $\hat{\pi}$ really estimates the vector π with elements $\pi_i = \int_{I_i} k(s, \theta_0)\{\alpha(s) - \alpha(s, \theta_0)\}\,ds$ under a priori circumstances and X_n^2 in effect tests whether the m π_i's are equal to zero; see also Remark 7B.

Under natural conditions $\sqrt{n}\,(\hat{\pi} - \pi)$ will have some Gaussian zero mean limit, say Z, with a covariance matrix R_{gen} more general and more cumbersome than the null model R previously considered. Simultaneously, $\sqrt{n}\,(\hat{R}^- - R_0^-)$ will have some zero mean Gaussian matrix limit, say M, where R_0 is the limit of $\hat{R}$ under α. One can now show that

$$\sqrt{n}\left(X_n^2/n - \pi' R_0^- \pi\right) = \sqrt{n}\left(\hat{\pi}'\hat{R}^-\hat{\pi} - \pi' R_0^- \pi\right)$$

converges in distribution to the variable $\pi' M \pi + 2\pi' R_0^- Z$, which is normal with zero mean and a complicated variance depending upon the specific alternative α. This variance is zero under the null model.

Let us also consider another type of example, namely the Kolmogorov–Smirnov type test for $\alpha(s) = \theta\alpha_0(s)$ described in 4A. The variable $D_n = \max_{0 \le t \le T} |\hat{\pi}(t)|$ converges in probability to $\lambda_\alpha = \max_{0 \le t \le T} |\pi(t)|$. Under model conditions λ_α is zero and $\sqrt{n}\, D_n$ has a limit distribution. Under regular conditions $\sqrt{n}\,\{\hat{\pi}(\cdot) - \pi(\cdot)\}$ tends to a certain zero mean normal process, say $U(\cdot)$; cf. comments made after Theorem 2.3. Assume for concreteness that $\lambda_\alpha = |\pi(t_0)|$ for a unique t_0. Then careful arguments similar to those used in

Raghavachari (1973) can be used to show that $\sqrt{n}(D_n - \lambda_\alpha)$ is asymptotically equivalent to $\sqrt{n}\{|\hat{\pi}(t_0)| - |\pi(t_0)|\}$ and this variable tends to $U(t_0)$. Additional analysis brings an end result for $\sqrt{n}(D_n/\hat{\sigma} - \lambda_\alpha/\sigma_\alpha)$, where σ_α is the limit in probability of $\hat{\sigma}$ under α conditions.

These results can be used to explore problems like choosing a test with high power for log-normal alternatives under gamma distribution conditions and vice versa. See also Remark 7B.

6. Testing the parametric Cox model. So far we have considered lifetimes to have been drawn from a homogeneous population of individuals with a common hazard rate $\alpha(s)$. Assume now that certain covariate measurements are available for each individual, say $z_i = (z_{i1}, \dots, z_{iq})'$ for number i, and that these are thought to influence this individual's hazard rate. Among several possible models for this kind of situation, by far the most popular is Cox's regression model for proportional hazards; see for example Cox (1972), Kalbfeisch and Prentice (1980) or Andersen and Borgan (1985). This is a semiparametric model which postulates that number i has hazard rate

$$\alpha_i(s) = \alpha(s)\exp(\beta' z_i), \tag{6.1}$$

where β is an unknown q-dimensional parameter and $\alpha(\cdot)$, the hazard rate for an individual with covariate vector zero, is left unspecified. Gill (1984) gives a good account of the martingale approach to the understanding and analysis of the Cox model.

The success of the Cox model has perhaps led to the unintended side effect that practitioners too seldomly invest efforts in studying the baseline hazard $\alpha(\cdot)$. A parametric version, say

$$\alpha_i(s) = \alpha(s, \theta)\exp(\beta' z_i), \tag{6.2}$$

for some p-dimensional θ, if found to be adequate, would lead to more efficient estimation of β and related quantities, like survival probabilities, and concurrently contribute to a better understanding of the survival phenomenon under study. References to work where models of type (6.2) have been used, with $\alpha(s,\theta)$ corresponding to the exponential, Weibull, log-normal distribution, or to piecewise constant hazards, can be found in Kalbfleisch and Prentice [(1980), Chapter 3], Friedman (1982) and Borgan (1984). Efron (1977) also studied this model. Borgan (1984) studied the asymptotic properties of the maximum likelihood estimators $(\hat{\theta}, \hat{\beta})$ in the general setting of (6.2). Observe that $\hat{\beta}$ is now more efficient than the usual Cox estimator $\hat{\beta}_{\text{cox}}$. This can be shown using arguments as in Altham (1984). [The Cox estimator is usually almost as good as $\hat{\beta}$, as shown by Efron (1977) and others.]

In Section 6.1 we motivate goodness of fit processes similar to those studied in Section 2 and derive their limit distributions under model conditions. The results are used in Section 6.2 to construct explicit goodness of fit tests for the parametric model (6.2). The general line of reasoning is analogous to that of Sections 2 and 3, but necessarily becomes more cumbersome, both w.r.t. notation and proof technicalities. We leave a fair amount of details behind in Hjort (1984).

REMARK. We operate here with covariates z_i that are constant over time, but only for the sake of concreteness and notational convenience. Our methods and results go through for time-dependent and even random processes $z_i(s)$ as long as these are uniformly bounded and predictable; see Andersen and Gill (1982). Similarly results below can be generalized to different forms of the relative risk function, as in Prentice and Self (1983).

Techniques for checking the adequacy of the tranditional *semiparametric* Cox model have been proposed by Schoenfeld (1980), Andersen (1982), Moreau, O'Quigley and Mesbah (1985), Gill and Schumacher (1987) and Arjas (1988), among others. Wei (1984) provided a consistent goodness of fit test for proportional hazards in the two-sample case. I am not aware of any earlier tests for the *parametric* Cox model.

6.1. *Weak convergence of goodness of fit processes.* In the following we shall employ the same framework, and partly the same techniques, as laid out for us in Andersen and Gill (1982), Gill (1984) and Borgan (1984). Matters are more complicated than in the usual semiparametric Cox model that Andersen and Gill studied, because of the presence of θ and $\hat{q}$ and because we use $\hat{\beta}$ instead of $\hat{\beta}_{\text{cox}}$.

We need to introduce some notation. Results will be derived under the conditions of the model (6.2). Let θ_0 and β_0 denote true parameter values. Probability statements below are w.r.t. the underlying true model, except for Section 6.3 and the paragraph where $\hat{A}$ of (6.7) is motivated. Let $[0, T]$ be the observation period and let N_i, Y_i and M_i be the counting process, the at-risk indicator and the corresponding martingale for individual number i. The σ-algebras in question are $\mathscr{F}_t = \bigvee_{i=1}^n \sigma\{N_i(s), Y_i(s);\ s \le t\}$, while

$$(6.3)\quad dM_i(s) = dN_i(s) - Y_i(s)\exp(\beta_0' z_i)\alpha(s, \theta_0)\, ds, \quad \text{for } i = 1, \dots, n.$$

Note that both N_i and Y_i are zero–one processes.

The logarithm of the observed likelihood can be written

$$\log L(\theta, \beta) = \text{const.} + \sum_{i=1}^n \int_0^T [\log\{\alpha(s,\theta)\exp(\beta' z_i)\}\, dN_i(s) - Y_i(s)\alpha(s,\theta)\exp(\beta' z_i)\, ds];$$

cf. (2.3). Write again $\psi(s,\theta) = \partial \log \alpha(s,\theta)/\partial\theta$. The maximum likelihood estimators $(\hat{\beta}, \hat{\theta})$ are consistent solutions to the equations

$$\frac{\partial \log L(\theta,\beta)}{\partial\theta} = \sum_{i=1}^n \int_0^T \psi(s,\theta)\{dN_i(s) - Y_i(s)\exp(\beta z_i)\alpha(s,\theta)\, ds\} = 0,$$

$$\frac{\partial \log L(\theta,\beta)}{\partial\beta} = \sum_{i=1}^n \int_0^T z_i\{dN_i(s) - Y_i(s)\exp(\beta z_i)\alpha(s,\theta)\, ds\} = 0.$$

Next, let

$$(6.4)\qquad \begin{aligned} U_n &= \frac{1}{\sqrt{n}} \frac{\partial \log L(\theta_0, \beta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^T \psi(s, \theta_0)\, dM_i(s), \\ V_n &= \frac{1}{\sqrt{n}} \frac{\partial \log L(\theta_0, \beta_0)}{\partial \beta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^T z_i\, dM_i(s). \end{aligned}$$

We shall assume that the regularity conditions of Borgan [(1984), Section 6] are in force. They ensure that

$$\begin{aligned} R(s, \beta) &= \frac{1}{n} \sum_{i=1}^{n} Y_i(s) \exp(\beta' z_i) \to_p r(s, \beta), \\ R_{(1)}(s, \beta) &= \frac{1}{n} \sum_{i=1}^{n} Y_i(s) z_i \exp(\beta' z_i) \to_p r_{(1)}(s, \beta), \\ R_{(2)}(s, \beta) &= \frac{1}{n} \sum_{i=1}^{n} Y_i(s) z_i z_i' \exp(\beta' z_i) \to_p r_{(2)}(s, \beta), \end{aligned}$$

at least for β values in a neighbourhood of the true one. Note that $R_{(1)}(s, \beta)$ is a q vector and $R_{(2)}(s, \beta)$ a $q \times q$ matrix. It is furthermore the case that $(U_n', V_n')'$ converges in distribution to $(U', V')'$, a zero mean Gaussian distribution with covariance matrix Σ, defined blockwise as

$$(6.5)\qquad \begin{aligned} \Sigma_{11} &= \int_0^T \psi(s, \theta_0) \psi(s, \theta_0)' r(s, \beta_0) \alpha(s, \theta_0)\, ds, \\ \Sigma_{12} &= \int_0^T \psi(s, \theta_0) r_{(1)}(s, \beta_0)' \alpha(s, \theta_0)\, ds, \\ \Sigma_{22} &= \int_0^T r_{(2)}(s, \beta_0) \alpha(s, \theta_0)\, ds. \end{aligned}$$

The technical arguments presented in Borgan [(1984), Section 6] can be used to demonstrate that

$$(6.6)\qquad \begin{bmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ \sqrt{n}(\hat{\beta} - \beta_0) \end{bmatrix} = \Sigma^{-1} \begin{bmatrix} U_n + \delta_n \\ V_n + \varepsilon_n \end{bmatrix},$$

in which δ_n and ε_n both go to zero in probability. In particular the limiting covariance matrix for $(\hat{\theta}, \hat{\beta})$ is Σ^{-1}/n.

Depart for a moment from the parametric model assumption (6.2) and assume only that (6.1) does hold for some β_0, in order to motivate a semiparametric estimator, as it were, for the cumulative baseline hazard $A(t) = \int_0^t \alpha(s)\, ds$. Let $N = \sum_{i=1}^n N_i$ and $M = \sum_{i=1}^n M_i$ be the accumulated counting process and martingale, respectively, where M_i for the moment has $\alpha(s)\, ds$ instead of $\alpha(s, \theta_0)\, ds$ in its definition; cf. (6.3). Then $dM(s) = dN(s) - \sum_{i=1}^n Y_i(s) \exp(\beta_0' z_i) \alpha(s)\, ds$, so that $J(s)\{dN(s)/n\}/R(s, \beta_0)$ can be written as

$J(s)\alpha(s)\,ds$ plus martingale noise, where $J(s) = \mathrm{I}\{\Sigma_{i=1}^n Y_i(s) > 0\}$. This motivates

$$\hat{A}(t) = \int_0^t \frac{J(s)\,dN(s)/n}{R(s,\hat{\beta})} = \int_0^t \frac{J(s)\,dN(s)}{\Sigma_{i=1}^n Y_i(s)\exp(\hat{\beta}'z_i)} \tag{6.7}$$

as an estimator of $A(t)$. The traditional estimator for A in the Cox model is identical to (6.7) except from using $\hat{\beta}_{\text{cox}}$ instead of the present $\hat{\beta}$; see Johansen (1983) or Gill (1984).

We can finally define the general goodness of fit process. Let $A_J(t,\theta) = \int_0^t J(s)\alpha(s,\theta)\,ds$ and $Z_n(t) = \sqrt{n}\{\hat{A}(t) - A_J(t,\hat{\theta})\}$ and consider

$$\begin{aligned} H_n(t) &= \int_0^t K_n(s)\,dZ_n(s) \\ &= \sqrt{n}\int_0^t K_n(s)\left[\frac{dN(s)}{\Sigma_{i=1}^n Y_i(s)\exp(\hat{\beta}'z_i)} - J(s)\alpha(s,\hat{\theta})\,ds\right]. \end{aligned} \tag{6.8}$$

To ensure the desired convergence result we need to impose restrictions on the weight function $K_n(s)$, similar to condition (K) of Section 2.

(K*) $K_n(s) = G_n(s,\hat{\theta},\hat{\beta})$, where the process $G_n(s,\theta_0,\beta_0)$ is predictable, converges to a suitable $k(s,\theta_0,\beta_0)$ uniformly in probability and is twice continuously differentiable in (θ,β). The partial derivatives $G'_{n,j}(s,\theta_0,\beta_0)$ are predictable and converge uniformly in probability to functions $g_j(s,\theta_0,\beta_0)$ satisfying $\int_0^T g_j(s,\theta_0,\beta_0)^2 r(s,\beta_0)^{-1}\alpha(s,\theta_0)\,ds < \infty$. There is finally a neighbourhood N of (θ_0,β_0) for which the variable

$$\max_{0\le s\le T}\ \max_{(\theta,\beta)\in N} \frac{\left|G''_{n,jl}(s,\theta,\beta)\right|}{\sqrt{n}}$$

tends to zero in probability.

THEOREM 6.1. *Assume that regularity conditions* (1)–(5) *of Borgan* [(1984), *page* 14] *are satisfied, with the* $r(s,\beta_0)$ *defined above being positive on* $[0,T]$, *and assume that* $\Sigma_{i=1}^n \exp(\beta_0' z_i)/n$ *is bounded in probability. Let* $\{K_n\}$ *be a sequence of weight functions obeying condition* (K*) *and define*

$$B(t) = \begin{bmatrix} \int_0^t k(s,\theta_0,\beta_0)\psi(s,\theta_0)\alpha(s,\theta_0)\,ds \\ \int_0^t k(s,\theta_0,\beta_0)e(s,\beta_0)\alpha(s,\theta_0)\,ds \end{bmatrix},$$

in which $e(s,\beta) = r_{(1)}(s,\beta)/r(s,\beta)$. *Then* H_n *of* (6.8) *converges in* $D[0,T]$ *to a Gaussian zero-mean process* H *with covariance function*

$$\operatorname{Cov}\{H(t_1),H(t_2)\} = \int_0^{t_1\wedge t_2} \frac{k(s,\theta_0,\beta_0)^2}{r(s,\beta_0)}\alpha(s,\theta_0)\,ds - B(t_1)'\Sigma^{-1}B(t_2).$$

PROOF. We start out rewriting $dZ_n(s)$ in a manner similar to the manipulations of Andersen and Gill [(1982), page 1104], but taking also the variability of $\hat{\theta}$ around θ_0 into account. By also using arguments similar to those used in the proof of Theorem 2.1, one finds

$$\begin{aligned} dZ_n(s) &= \sqrt{n}\,J(s)\big[(1/n)\,dN(s)/R(s,\hat{\beta}) - \alpha(s,\hat{\theta})\,ds\big] \\ &= J(s)R(s,\beta_0)^{-1}\,dM(s)/\sqrt{n} - J(s)\psi(s,\tilde{\theta})'\alpha(s,\tilde{\theta})\,ds\,\sqrt{n}\,(\hat{\theta}-\theta_0) \\ &\quad - J(s)\{dM(s)/n + R(s,\beta_0)\alpha(s,\theta_0)\,ds\} \\ &\qquad \times\Big\{R_{(1)}(s,\tilde{\beta})/R(s,\tilde{\beta})^2\Big\}\sqrt{n}\,(\hat{\beta}-\beta_0), \end{aligned}$$

where $\tilde{\theta}$ is between θ_0 and $\hat{\theta}$ and, similarly, $\tilde{\beta}$ is between β_0 and $\hat{\beta}$. By (6.6), this leads to an expression of the type

$$H_n(t) = \int_0^t K_n(s)\,dW_n(s) - B_n(t)'\Sigma^{-1}\begin{bmatrix} U_n \\ V_n \end{bmatrix} + \text{remainder}_n,$$

in which $W_n(t) = \int_0^t J(s)R(s,\beta_0)^{-1}\,dM(s)/\sqrt{n}$ is a martingale, $B_n(t)$ is a random vector which converges to $B(t)$ uniformly in probability and remainder$_n$ goes to zero in probability, by several careful arguments parallelling those used in proof of Theorem 2.1. Details are available in Hjort (1984).

Rebolledo's central limit theorem for martingales, in the form of Theorem 1.2 in Andersen and Gill (1982), can be used to verify that $W_n \to_d W$ in $D[0,T]$, a Gaussian martingale with $\operatorname{Var}\{dW(s)\} = r(s,\beta_0)^{-1}\alpha(s,\theta_0)\,ds$. We have earlier mentioned that $(U_n', V_n')'$ of (6.4) tends to a zero-mean Gaussian $(U', V')'$ with covariance matrix Σ given in (6.5). What is needed to conclude the proof is the stronger simultaneous statement $(W_n, U_n, V_n) \to_d (W, U, V)$ in $D[0,T] \times \mathscr{R}^{p+q}$, where (W, U, V) has the appropriate covariance structure: $\operatorname{Cov}\{W(t), U\} = \int_0^t \psi(s,\theta_0)\alpha(s,\theta_0)\,ds$ and $\operatorname{Cov}\{W(t), V\} = \int_0^t e(s,\beta_0)\alpha(s,\theta_0)\,ds$. This can be shown using Rebolledo's theorem once more. In consequence, H_n converges in distribution to

$$H = \int_0^{\cdot} k(s,\theta_0,\beta_0)\,dW(s) - B(\cdot)'\Sigma^{-1}\begin{bmatrix} U \\ V \end{bmatrix}.$$

This is a Gaussian process with mean zero, and it is not difficult to verify that its covariance structure is as given in the theorem. □

6.2. *Goodness of fit tests.* To check the validity of one's parametric Cox model with $\alpha(s) = \alpha(s,\theta)$, one could draw the graph of $H_n(t)/\hat{\tau}(t)$, where

$$\hat{\tau}^2(t) = \int_0^t \frac{K_n(s)^2}{R(s,\hat{\beta})}\,d\hat{A}(s) - \hat{B}(t)'\hat{\Sigma}^{-1}\hat{B}(t)$$

is a consistent estimator for $\tau^2(t) = \operatorname{Var} H(t)$. Estimators $\hat{\Sigma}$ and $\hat{B}(t)$ are easy to construct by replacing $k(s,\theta_0,\beta_0)$ by $K_n(s)$, $\psi(s,\theta_0)$ by $\psi(s,\hat{\theta})$, etc., in the expressions defining Σ and $B(t)$. If the model is correct, then $H_n(t)/\hat{\tau}(t)$ is asymptotically standard normal for each t, but otherwise it would drift off from zero. Special Cases 1–3 below lead to three different possibilities for $H_n(t)/\hat{\tau}(t)$.

A class of rigorous χ^2 tests can also be derived. Let $0 = a_0 < \cdots < a_m = T$ divide $[0, T]$ into cells $I_i = (a_{i-1}, a_i]$, as in Section 3.1, and let

$$Q_{n,i} = H_n(a_{i-1}, a_i] = \sqrt{n} \int_{I_i} K_n(s) J(s) \left[\frac{dN(s)/n}{R(s, \hat{\beta})} - \alpha(s, \hat{\theta})\, ds \right].$$

Then the vector Q_n with elements $Q_{n,i}$ tends in distribution to the vector Q with elements $Q_i = H(a_{i-1}, a_i]$. As in Section 3.1 it is seen that Q has

$$\Omega = \operatorname{Var} Q = D - S'\Sigma^{-1}S,$$

in which D is diagonal with elements $d_i = \int_{I_i} \{k^2/r\}\alpha\, ds$ and $S = (b_1, \dots, b_m)$ is $(p + q) \times m$ with elements $b_i = (\int_{I_i} k\psi'\alpha\, ds, \int_{I_i} ke'\alpha\, ds)'$. We allow ourselves to write k for $k(s, \theta_0, \beta_0)$, ψ for $\psi(s, \theta_0)$, etc., here.

Let $\hat{\Omega}$ be any estimator of Ω which is consistent under model conditions, say with elements $\hat{\omega}_{i,j} = \hat{d}_i \delta_{i,j} - \hat{b}_i'\hat{\Sigma}^{-1}\hat{b}_j$. Our test statistic will be $X_n^2 = Q_n'\hat{\Omega}^- Q_n$, which upon reemployment of the matrix identity (3.4) can be written

$$(6.9) \qquad X_n^2 = Q_n'\hat{\Omega}^- Q_n = \sum_{i=1}^m \frac{Q_{n,i}^2}{\hat{d}_i} + \left\{ \sum_{i=1}^m \frac{Q_{n,i}}{\hat{d}_i} \hat{b}_i \right\}' \hat{G}^- \left\{ \sum_{i=1}^m \frac{Q_{n,i}}{\hat{d}_i} \hat{b}_i \right\},$$

involving the (generalized) inverse of the $(p + q) \times (p + q)$ matrix $\hat{G} = \hat{\Sigma} - \hat{S}\hat{D}^{-1}\hat{S}' = \hat{\Sigma} - \sum_{i=1}^m \hat{b}_i\hat{b}_i'/\hat{d}_i$. The limit distribution of X_n^2 under model conditions (6.2) is χ^2_{df}, where $\mathrm{df} = \operatorname{Rank}(\Omega)$. In most cases $\mathrm{df} = m$, whereas $\mathrm{df} = m - 1$ for choices of K_n that give $H_n(T) = \sum_{i=1}^m Q_{n,i} = 0$; see the Remark in Section 3 and Special Case 3 below.

Interesting choices for the K_n function include the following:

Special case 1. Let $K_n(s) \equiv 1$. Then $Q_{n,i} = \sqrt{n}\{\hat{A}(I_i) - \int_{I_i} J(s)\alpha(s, \hat{\theta})\, ds\}$ compares two estimates of the cumulative hazard rate over interval I_i, one based on the semiparametric Cox model and one based on the parametric Cox model.

Special case 2. Let $K_n(s) = \sum_{j=1}^n Y_j(s)\exp(\hat{\beta}'z_j)/n = R(s, \hat{\beta})$, which obeys condition (K*) of the theorem, as a consequence of the other regularity conditions. In this case $Q_{n,i} = (N_i - E_i)/\sqrt{n}$, where $N_i = N(I_i)$ counts the observed transitions in interval I_i and $E_i = \int_{I_i} \sum_{j=1}^n Y_j(s)\exp(\hat{\beta}'z_j)\alpha(s, \hat{\theta})\, ds$ is a dynamic, hazard rate based estimate of N_i based on model assumptions. One has $X_n^2 = \sum_{i=1}^m (N_i - E_i)^2/E_i + W_n'\hat{G}^- W_n$, in which $W_n = \sqrt{n}\sum_{i=1}^m \{(N_i - E_i)/E_i\}\hat{b}_i$.

Special case 3. Let this time $K_n(s) = R(s, \hat{\beta})c'\psi(s, \hat{\theta})$ for some coefficients $c_1, \dots, c_p$. Then $\sum_{i=1}^m Q_{n,i} = 0$, $\mathrm{df} = m - 1$ and (6.9) can be evaluated using the inverse of a $(p + q - 1) \times (p + q - 1)$ matrix, as in the corresponding situation of Section 3.1.

THE ONE-DIMENSIONAL CASE. Assume that θ is one-dimensional and choose $K_n(s) = R(s, \hat\beta)\psi(s, \hat\theta)$. Then

$$Q_{n,i} = \int_{I_i} \psi(s, \hat\theta)\left\{dN(s) - \sum_{j=1}^{n} Y_j(s)\exp(\hat\beta' z_j)\alpha(s, \hat\theta)\, ds\right\}/\sqrt{n}$$

and $\hat d_i = \int_{I_i} R(s, \hat\beta)\psi(s, \hat\theta)^2\alpha(s, \hat\theta)\, ds$. One can show that

$$X_n^2 = \sum_{i=1}^{m} Q_{n,i}^2/\hat d_i + (W_n^0)'(\hat G^0)^{-1}(W_n^0),$$

where $W_n^0 = \sum_{i=1}^m Q_{n,i}\int_{I_i} R_{(1)}(s, \hat\beta)\alpha(s, \hat\theta)\, ds/\hat d_i$ and $\hat G^0$ is the $q \times q$ matrix $\sum_{i=1}^m [\int_{I_i} \hat R_{(2)}\hat\alpha\, ds - \{\int_{I_i} \hat R_{(1)}\hat\psi\hat\alpha\, ds\}\{\int_{I_i} \hat R_{(1)}\hat\psi\hat\alpha\, ds\}'/\hat d_i]$. The test statistic X_n^2 has a limiting χ^2_{m-1} distribution if the model conditions (6.2) hold. In particular one can write down a test for exponential regression, that is, $\alpha_i(s) = \theta\exp(\beta' z_i)$, of the form $\sum_{i=1}^m (N_i - E_i)^2/E_i + (W_n^0)'(\hat G^0)^{-1}(W_n^0)$.

6.3. *Consistency of the tests.* What type of departures from the parametric Cox model will the tests just proposed be able to detect? Assume only that the true hazard rate structure is of the form $\alpha_i(s) = \alpha(s)g(z_i)$, say, where we might take $g(0) = 1$ to identify the baseline hazard rate $\alpha(\cdot)$. The parametric Cox model amounts to $\alpha(s) = \alpha(s, \theta)$ for some θ and $g(z) = \exp(\beta' z)$ for some β. But what happens to H_n of (6.8) in the wider model?

One has $dN(s) = \sum_{i=1}^n Y_i(s)g(z_i)\alpha(s)\, ds + dM(s)$, where M is a sum of n individual martingales. Assume that the average $\sum_{i=1}^n Y_i(s)g(z_i)/n$ tends in probability to some function $t(s)$. Under regularity conditions similar to those described in Hjort [(1986), Section 4] one can demonstrate that $\hat\theta$ and $\hat\beta$ converge in probability to certain appropriate least false parameter values θ_0 and β_0. Lenglart's inequality [see Andersen and Gill (1982)] can be used to show that $\hat A(t)$ of (6.7) tends to $\int_0^t \{t(s)/r(s, \beta_0)\}\alpha(s)\, ds$, and these results combine to give

$$\hat\pi(t) = \frac{H_n(t)}{\sqrt{n}} \to_p \int_0^t k(s, \theta_0, \beta_0)\left[\frac{t(s)}{r(s, \beta_0)}\alpha(s) - \alpha(s, \theta_0)\right] ds = \pi(t);$$

cf. Theorem 2.3 and its proof. It follows that appropriate tests based upon H_n will detect *any* departure from (6.2) of the type $\alpha(s)g(z_i)$, with probability tending to 1. One can also show that $H_n(t) - \sqrt{n}\,\pi(t) = \sqrt{n}\,\{\hat\pi(t) - \pi(t)\}$ has a limiting normal distribution under traditional regularity assumptions; cf. the closing paragraph of Section 2 and Remark 7B.

Two types of departure from (6.2) are of particular interest. If the $g(z) = \exp(\beta_0' z)$ part is correctly modelled, but $\alpha(\cdot)$ differs from $\alpha(\cdot, \theta_0)$, then $H_n(t)/\sqrt{n}$ tends to $\pi(t) = \int_0^t k(s, \theta_0, \beta_0)\{\alpha(s) - \alpha(s, \theta_0)\}\, ds$. Even in cases where the $\alpha(s) = \alpha(s, \theta_0)$ part of the model is correct, will departures of $g(z)$ from $\exp(\beta_0' z)$ be detected, in that

$$\pi(t) = \int_0^t k(s, \theta_0, \beta_0)\{t(s)/r(s, \beta_0) - 1\}\alpha(s, \theta_0)\, ds.$$

One may also derive results about local asymptotic power, as in Theorem 2.2 and Section 5, but this is not pursued here.

7. Concluding remarks.

REMARK 7A. Our results have been derived using maximum likelihood estimators for the unknown parameters. The results continue to hold for other estimators that are asymptotically equivalent. A Bayesian is allowed the privilege of substituting for $\hat{\theta}$ some appropriate Bayes estimator $\hat{\theta}_B$ in the test statistics of Sections 3, 4 and 6, for example, in that $\sqrt{n}(\hat{\theta}_B - \hat{\theta})$ vanishes in probability, even outside model conditions, according to Hjort (1986). One can also derive results parallelling Theorem 2.1 for other estimators, but the covariance structure will in general be more complicated, which in turn usually would lead to more complicated test statistics. (An exception is the minimum chi square estimator used by Akritas [(1988), Section 3].) This remark applies also to the possibility of using $\hat{\beta}_{\text{cox}}$ instead of $\hat{\beta}$ in Section 6. A statistical reason for sticking to $\hat{\theta}$ or some asymptotically equivalent relative is that these are asymptotically optimal estimators; see Hjort [(1986), Section 3].

REMARK 7B. To make our next point, let us write $\Pr_\alpha$ for probability statements w.r.t. the counting process model with hazard rate $\alpha(\cdot)$. The defining property of a bona fide test statistic with asymptotic level ε is that

$$\limsup_{n\to\infty} \Pr_\alpha\{\text{reject}\} \le \varepsilon \quad \text{for all } \alpha \in H_0, \tag{7.1}$$

where H_0 is the subset of α's that agree with the null hypothesis model, i.e., $\alpha(s) = \alpha(s, \theta)$ for some value of the parameter θ. The χ^2-type test statistics of Section 3 employ certain estimators $\hat{d}_i, \hat{b}_i, \hat{\Sigma}$ for quantities d_i, b_i, Σ present in the asymptotic covariance matrix $R = D - S'\Sigma^{-1}S$ for Q_n; see (3.1) and (3.2). As pointed out there, the single requirement is that these estimators are consistent *under H_0 conditions*, i.e., (7.1) holds, and indeed with $=$ replacing $\le$ and lim replacing lim sup, provided only that $\hat{R} \to_p R = R_\alpha$ for each $\alpha \in H_0$.

But what should a statistician do next, if the null hypothesis model is rejected? The natural desire is to search for specific *departures* from H_0 and declare certain such departures to be present. This calls for a *search rule*; see Hjort (1988) for a general discussion. In the present situation the χ^2 test $X_n^2 = Q_n'\hat{R}^-Q_n = n\hat{\pi}'\hat{R}^-\hat{\pi}$ of (3.3) in effect tests whether $\pi_1 = 0, \dots, \pi_m = 0$, in the notation introduced in Theorem 2.3 and Section 5.3. Call a smooth function $f(\pi) = f(\pi_1, \dots, \pi_m)$ a *contrast* if $f(0, \dots, 0) = 0$. A possible search rule for departures from the null state is to declare the contrast $f(\pi)$ to be positive if $\sqrt{n} f(\hat{\pi})/\hat{\sigma}(f) > c$, for appropriate level c, where $\hat{\sigma}^2(f) = D_f(\hat{\pi})'\hat{R}D_f(\hat{\pi})$ and $D_f(\pi)$ is the vector of partial derivatives. A good search rule should have the property

$$\limsup_{n\to\infty} \Pr_\alpha\{\text{there is some } f(\pi) \le 0, \text{but } f(\pi) \text{ is declared positive}\} \le \varepsilon, \tag{7.2}$$

for *all* a priori α hazards, not only those under H_0. The results of Hjort (1988) can be used to show that (7.2) indeed does hold, for the quoted search rule, with $c = (\gamma_{\mathrm{df}, \varepsilon})^{1/2}$, provided that $\hat{\sigma}(f)$ is computed with an $\hat{R}$ that is *a priori consistent*. This means that $\hat{R}$ is required to tend to the general limiting covariance matrix R_α of $\sqrt{n}(\hat{\pi} - \pi)$, also for α's outside H_0. There is a link to significance testing in that the event "at least one contrast is declared positive" is asymptotically equivalent to $X_n^2 = n\hat{\pi}'\hat{R}^-\hat{\pi} > \gamma_{\mathrm{df}, \varepsilon}$.

Seen in this light X_n^2 takes the perhaps more natural role of a *clearance test* and one avoids the somewhat artificial probability calculations under a model that one perhaps knows in advance cannot be exactly true. The *caveat* and the price to pay is that a model-robust $\hat{R}$ is called for; $\hat{R} = \hat{D} - \hat{S}'\hat{\Sigma}^{-1}\hat{S}$ is not good enough. Accordingly, if (7.2) is strived for [not only (7.1)], then X_n^2 cannot be used in the simple forms (3.5), (3.6) and (3.13), for example. One can derive general but rather cumbersome expressions for R_α and a priori consistent estimators can be constructed based on this; see the closing paragraph of Section 2. It might be simpler to use nonparametric bootstrapping.

Of course the points made here are of a general nature and apply to most of the classical χ^2-type tests as well.

REMARK 7C. We have employed a framework with a fixed, bounded time interval $[0, T]$, since the martingale and counting process apparatus works best there. We do not consider the boundedness a serious practical problem, but it is of course satisfying to have results applicable for the full half-line. Extensions of this article's results are possible, to T being a random stopping time and to $T = \infty$, under appropriate extra conditions; see Gill (1980), Andersen, Borgan, Gill and Keiding (1982), Andersen and Gill (1982), Helland (1982), Wei (1984) and Aven (1986) for similar comments and details. Basically one needs conditions that ensure contributions to estimators and test statistics from data on $[\tau, \infty)$ to be arbitrarily small, uniformly in n for large enough τ, and the matrix $\Sigma = \int_0^\infty y(s)\psi(s, \theta_0)\psi(s, \theta_0)'\alpha(s, \theta_0)\, ds$ must be finite. Explicit conditions seem easiest to put up following Aven [(1986), Section 5].

REMARK 7D. Often a framework is needed with more than one counting process. Assume, for example, that n individuals in a homogeneous population move among different states in a Markov chain manner, with hazard rate $\alpha_{i,j}(s)$ for transitions from state i to state j. Let $N_{i,j}(t)$ count the number of observed transitions from i to j in the time interval $[0, t]$. Then the collection of $N_{i,j}$'s form a multivariate counting process, with hazard processes $Y_i(s)\alpha_{i,j}(s)$, where $Y_i(s)$ is the number of individuals in state i just prior to time s; see Aalen (1978a, b). Some examples of parametric models for several hazard rates, used in actuarial sciences and demography, can be found in Borgan (1984) and his references.

The methods and results of Sections 2 and 3 first of all apply to any of the given counting processes, so that the hypothesis that the hazard rate from i to j is constant over a certain time interval, for example, can be tested with some appropriate test among those proposed in Section 3. Second, the methods and results can be extended to the multivariate framework, with surprisingly few extra difficulties, using asymptotic martingale techniques. This is due to the fact that the associate basic martingales become orthogonal. Details of this extension are not given here, but are available along with some examples of applications in Hjort (1984).

REMARK 7E. There are several relatives to the goodness of fit tests in Section 3. Several tests can be based on weak convergence of $\sqrt{n}\{\hat{F}(t) - F(t, \hat{\theta})\}$ or a weighted version, where $\hat{F}$ is the Kaplan–Meier estimator. Results for this process were independently derived, thereby simultaneously generalizing classic theorems of Durbin (1973) and Aalen and Johansen (1978), by Burke (1981), Hjort [(1984), Section 3] and Habib and Thomas (1986). The latter authors proposed a class of χ^2 tests. Other related tests that can be constructed using essentially the same machinery as in this paper's Sections 2 and 3 are based on parametric transformations of the cumulative hazard. To test whether $A(t)$ is of the Weibull type $(\theta t)^\beta$, for example, one might consider functionals of $\log \hat{A}(t) - \hat{\beta}\log(\hat{\theta} t)$, or perhaps inventing a test for linearity of $\log \hat{A}(t)$ in $\log t$.

REMARK 7F. The limit process H of Theorem 2.1 is a p times tied-down time-transformed Brownian motion process. Write $H(t) = H^0(t) - B(t)'\Sigma^{-1}U$, where $H^0(t)$ has independent increments and can be represented as $W\{\int_0^t (k^2/y)\alpha\, ds\}$ and $U = \int_0^T \psi(s, \theta_0)\, dV(s)$. One can now show that the distribution of H^0, conditioned on the vector U being zero, is exactly that of H.

REMARK 7G. It is worth pointing out that the null distribution of some of the tests proposed here can be obtained through simulation.

REMARK 7H. The χ^2 tests of Sections 3 and 6 can use certain types of random cell boundaries without affecting the limit distribution results. This can be established using random time change techniques of Billingsley [(1968), Section 17]; see Habib and Thomas [(1986), Section 3].

REMARK 7I. It should be possible to construct classes of goodness of fit tests for certain parametric *semi-Markov* processes, at least for the class studied in Voelkel and Crowley (1984), extending the techniques used in this paper.

Acknowledgments. Conversations with Ørnulf Borgan triggered some of the research that led to the present paper. I am grateful for the careful work of an associate editor and the encouraging persistence of editor van Zwet, which finally caused me to carry out the necessary revision.

REFERENCES

AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. thesis, Dept. Statist., Univ. California, Berkeley.

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

AALEN, O. O. (1978b). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6** 534–545.

AALEN, O. O. (1982). Practical applications of the nonparametric statistical theory for counting processes. Statistical Research Report, Inst. of Math., Univ. of Oslo.

AALEN, O. O. and HOEM, J. M. (1978). Random time changes for multivariate counting processes. *Seand. Actuar. J.* 81–101.

AALEN, O. O. and JOHANSEN, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5** 141–150.

AKRITAS, M. G. (1988). Pearson-type goodness-of-fit tests: the univariate case. *J. Amer. Statist. Assoc.* **83** 222–230.

ALTHAM, P. M. E. (1984). Improving the precision of estimation by fitting a model. *J. Roy. Statist. Soc. Ser. B* **46** 118–119.

ANDERSEN, P. K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics* **38** 67–78.

ANDERSEN, P. K. and BORGAN, Ø. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.* **12** 97–158.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Internat. Statist. Rev.* **50** 2199–258.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

ARJAS, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *J. Amer. Statist. Assoc.* **83** 204–212.

AVEN, T. (1986). Bayesian inference in a parametric counting process model. *Scand. J. Statist.* **13** 87–97.

BARLOW, R. and CAMPO, R. A. (1975). Total time on test processes and application to failure data analysis. In *Reliability and Fault Tree Analysis* (R. E. Barlow, J. B. Fussel and N. D. Singpurwalla, eds.) 451–481. SIAM, Philadelphia.

BIE, O., BORGAN, Ø. and LIESTØL, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scand. J. Statist.* **14** 221–233.

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

BORGAN, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* **11** 1–16.

BRESLOW, N. E. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimators under random censorship. *Ann. Statist.* **2** 437–453.

BURKE, M. D. (1981). Tests for exponentiality based on randomly censored data. In *Nonparametric Statistical Inference* (L. Rejto, ed.) *Çoll. Math. Soc. János Bolyai* **32**. North-Holland, Amsterdam.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

COX, D. R. (1979). A note on the graphical analysis of survival data. *Biometrika* **66** 188–190.

CRESSIE, N. and READ, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46** 440–464.

CSÖRGŐ, S. and HORVÁTH, L. (1982a). Confidence bands from censored samples. *Canad. J. Statist.* **14** 131–144.

CSÖRGŐ, S. and HORVÁTH, L. (1982b). Statisztikai következtetés cenzorált mintákból (in Hungärian). *Alkalmazott Matematikai Lapok* **8** 1–89.

DOKSUM, S. A. and YANDELL, B. (1984). Tests for exponentiality. In *Handbook of Statistics* (P. R. Krishnaiah, ed.) **4**. North-Holland, Amsterdam.

DURBIN, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing power of the Kolmogorov–Smirnov test. *J. Appl. Probab.* **8** 431–453.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

DZHAPARIDZE, K. (1986). On asymptotic inference about intensity parameters of a counting process. *Papers on Semiparametric Models at the* ISI *Centenary Session* CWI, Amsterdam. 31–54.

EFRON, B. (1977). Efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

FOLKESON, M. (1984). Goodness of fit tests based on quantile processes. Graduate thesis, Dept. Math. Statist., Univ. of Oslo. (In Norwegian.)

FRIEDMAN, M. (1982). Piecewise exponential models for survival data with covariates. *Ann. Statist.* **10** 101–113.

GILL, R. D. (1980). *Censoring and Stochastic Integrals*. Math. Centre Tracts **124**, Mathematical Centre, Amsterdam.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

GILL, R. D. (1984). Understanding Cox's regression model: A martingale approach. *J. Amer. Statist. Assoc.* **79** 441–447.

GILL, R. D. and SCHUMACHER, M. (1987). A simple test for the proportional hazard assumption. *Biometrika* **74** 289–300.

GRAY, R. J. and PIERCE, D. A. (1985). Goodness-of-fit tests for censored survival data. *Ann. Statist.* **13** 552–563.

GRINGORTEN, I. I. (1968). Estimating finite-time maxima and minima of a stationary Gaussian Ornstein–Uhlenbeck process by Monte Carlo simulation. *J. Amer. Statist. Assoc.* **63** 1517–1521.

HABIB, M. G. and THOMAS, D. R. (1986). Chi-square goodness-of-fit tests for randomly censored data. *Ann. Statist.* **14** 759–765.

HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–94.

HJORT, N. J. (1984). Weak convergence of cumulative intensity processes when parameters are estimated, with applications to goodness of fit tests in models with censoring. Research Report No. 763, Norwegian Computing Centre, Oslo.

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

HJORT, N. L. (1986). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scand. J. Statist.* **13** 63–85.

HJORT, N. L. (1988). On large sample multiple comparison methods. *Scand. J. Statist.* **15** 259–271.

HORVÁTH, L. (1982). Parameter estimation and testing composite hypotheses from censored data. *Coll. Math. Soc. János Bolayi* **36** 587–616.

JOHANSEN, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* **51** 258–262.

JOHNSON, B. MCK. and KILLEEN, T. (1983). An explicit formula for the cdf of the L_1 norm of the Brownian bridge. *Ann. Probab.* **11** 807–808.

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

KARR, A. F. (1986). *Point Processes and their Statistical Inference*. Dekker, New York.

LEWIS, P. A. W. (1961). Distribution of the Anderson–Darling statistic. *Ann. Math. Statist.* **32** 1118–1124.

MILLER, R. and SIEGMUND, D. (1982). Maximally selected chi square statistics. *Biometrics* **38** 1011–1016.

MOREAU, T., O'QUIGLEY, J. and MESBAH, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. *Appl. Statist.* **34** 212–218.

NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1** 27–52.

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

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

RAGHAVACHARI, M. (1973). Limiting distribution of Kolmogorov–Smirnov type statistics under the alternative. *Ann. Statist.* **1** 67–73.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

SCHOENFELD, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards model. *Biometrika* **67** 145–153.

SHEPP, L. A. (1982). On the integral of the absolute value of the pinned Wiener process. *Ann. Probab.* **10** 234–239.

VOELKEL, J. G. and CROWLEY, J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Ann. Statist.* **12** 142–160.

WALSH, J. E. (1962). *Handbook of Nonparametric Statistics: Investigations of Randomness, Moments, Percentiles and Distributions* **1**. Van Nostrand, Princeton, N.J.

WEI, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Amer. Statist. Assoc.* **79** 649–652.

NORWEGIAN COMPUTING CENTRE
GAUSTADALLÉEN 23
P.B. 114 BLINDERN
N-0314 OSLO 3
NORWAY