The Annals of Statistics
1992, Vol. 20, No. 1, 428–454

APPROXIMATION OF STOCHASTIC INTEGRALS WITH APPLICATIONS TO GOODNESS-OF-FIT TESTS[1]

BY ALEX J. KONING

Erasmus University

In this paper stochastic integrals with respect to the basic martingale are approximated by Gaussian processes. The probability inequalities governing this approximation are used to study goodness-of-fit tests based on sublinear functionals of weighted versions of these stochastic integrals. As special cases of these tests, generalized rank and supremum-type tests are considered.

1. Introduction. Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a probability space and $\mathscr{F}_n$ a subset of $\mathscr{B}[0, \infty)$. At stage n, each of the independent random variables $X_1, \dots, X_n$, $Y_1, \dots, Y_n$ maps $(\Omega, \mathscr{A})$ into $([0, \infty), \mathscr{F}_n)$. The probability measure induced by these random variables is denoted by P_n. Each pair (X_i, Y_i) is assumed to have the same distribution.

The distribution of Y_i, referred to as the censoring distribution, does not depend on n. Hence, there exists a cumulative distribution G such that $G(t) = P_n(Y_i \le t)$ for each n. Defectiveness of G is allowed.

The distribution of X_i, called the failure time distribution, is more complicated since it depends on n. This dependence is given structure in the following way: There exists a cumulative distribution function F, indexed by θ belonging to some set Θ and a sequence of points $\{\theta_n\}_{n=1}^{\infty}$ such that $F(t; \theta_n) = P_n(X_i \le t)$ for every $n \in \mathbb{N}$ (i.e., θ_n is the actual value of θ at stage n).

Now suppose θ_0 is an element of Θ which is of special interest to us, say because we want to know whether θ_n could possibly be equal to θ_0. Usually, this question is investigated by using techniques which are based on the empirical distribution function of the sample $X_1, \dots, X_n$. In some situations encountered in cancer-research or life-testing, however, the phenomenon of censoring occurs: The failure time X_i cannot be observed if it exceeds the censoring time Y_i. All that is observed in these situations are the random variable $Z_1, \dots, Z_n$ and $\delta_1, \dots, \delta_n$ defined by

$$Z_i = X_i \wedge Y_i, \tag{1}$$

$$\delta_i = 1_{\{X_i \le Y_i\}}. \tag{2}$$

The random variable δ_i is called the censoring indicator. Depending on the value of the censoring indicator, the random variable Z_i is called the observed failure time ($\delta_i = 1$) or the censored failure time ($\delta_i = 0$).

Received March 1990; revised May 1991.

[1]Part of this research was performed at the University of Limburg, Maastricht.

AMS 1980 *subject classifications*. Primary 62G10; secondary 60F15, 60H05.

Key words and phrases. Random censoring, stochastic integrals, approximation, goodness-of-fit tests.

The pair of random variables (Z_i, δ_i) follows the random censoring model. The sample $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$ can be represented without loss of information by means of the empirical distribution functions

$$H_n^1(t) = n^{-1} \sum_{i=1}^{n} 1_{\{Z_i \le t, \delta_i = 1\}}, \tag{3}$$

$$H_{n-}(t) = n^{-1} \sum_{i=1}^{n} 1_{\{Z_i < t\}}. \tag{4}$$

Observe that $H_n^1(t)$ may well be defective and that $H_{n-}(t)$ is left-continuous.

The understanding of the random censoring model has benefitted much from approaching $nH_n^1(t)$, the number of failures occurring in the time interval $[0, t]$ as a counting process [see Aalen (1976), Aalen and Johansen (1978), Gill (1980, 1983)]. The compensator of this counting process is $n\int_0^t (1 - H_{n-}(s))\, d\Lambda(s; \theta_n)$, where $\Lambda(t; \theta) = -\log(1 - F(t; \theta))$ is the cumulative hazard function belonging to $F(t; \theta)$. The difference between $nH_n^1(t)$ and its compensator is the martingale part of the Doob–Meyer decomposition of $nH_n^1(t)$ and provides the entry-point for invoking central limit theorems. Since this process arises in a natural way, it is sometimes believed that it is the most basic representation of the randomness in the sample $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$. Hence, an appropriately rescaled version of this process has become known under the name basic martingale [see Shorack and Wellner (1986), page 296].

If θ_n equals θ_0 (denote the probability measure corresponding to this situation by P_0 and the situation itself by under P_0), then the basic martingale takes the form

$$M_n(t; \theta_0) = n^{1/2} \left\{ H_n^1(t) - \int_0^t (1 - H_{n-}(s))\, d\Lambda(s; \theta_0) \right\}. \tag{5}$$

Note that we only know that $M_n(t; \theta_0)$ is a martingale under P_0. If θ_n is arbitrary (refer to this situation as under P_n), then the process $M_n(t; \theta_0)$ is in general not a martingale. Nevertheless, we shall also call $M_n(t; \theta_0)$ a basic martingale. Although misleading, this does not lead to difficulties since we do not encounter the original basic martingale $M_n(t; \theta_n)$ anymore.

The view that $M_n(t; \theta_0)$ is under P_0 the most basic representation of randomness in the sample gives rise to the conviction that $M_n(t; \theta_0)$ should give us an excellent answer to the question whether θ_n equals θ_0. Hence, tests for the simple null hypothesis that θ_n equals θ_0, on which we focus in this paper, should preferably be based on statistics constructed from the basic martingale, a view first expressed in Khmaladze (1981, 1982) and shared in Hjort (1990). In Section 3, we study tests based on a special type of functional of a weighted version of a process $Q_n(t)$, which is a stochastic integral with respect to the basic martingale, that is,

$$Q_n(t) = \int_0^t L_n(s)\, dM_n(s; \theta_0). \tag{6}$$

Here the weight process $L_n(t)$ is a stochastic process satisfying certain conditions.

The probability theory underlying the results in Section 3 is presented in Section 2, where the process $Q_n(t)$ is approximated. The common way to approach a stochastic integral with respect to the basic martingale is by martingale methods. In this paper an empirical process approach is followed, based on an alternative representation of $M_n(t;\theta_0)$ in terms of empirical processes. A slight drawback of the empirical process approach is that we must assume that both F and G are continuous on the complete real line. However, this assumption has its rewards: the knowledge obtained by the empirical process approach is far more precise than can be obtained by using standard martingale methods such as Rebolledo's central limit theorem.

Define the cumulative distribution functions $H^1(t;\theta_n)$ and $H(t;\theta_n)$ by

$$H^1(t;\theta_n) = P_n(Z_1 \le t, \delta_1 = 1) = \int_0^t (1 - G(s))\, dF(s;\theta_n), \tag{7}$$

$$H(t;\theta_n) = P_n(Z_1 \le t) = 1 - (1 - G(t))(1 - F(t;\theta_n)) \tag{8}$$

and the empirical processes $U_n^1(t;\theta_n)$ and $U_{n-}(t;\theta_n)$ by

$$U_n^1(t;\theta_n) = n^{1/2}\{H_n^1(t) - H^1(t;\theta_n)\}, \tag{9}$$

$$U_{n-}(t;\theta_n) = n^{1/2}\{H_{n-}(t) - H(t;\theta_n)\}. \tag{10}$$

Then we may decompose $M_n(t;\theta_0)$ conveniently into three parts:

$$M_n(t;\theta_0) = U_n^1(t;\theta_n) + \int_0^t U_{n-}(s;\theta_n)\, d\Lambda(s;\theta_0) + n^{1/2} D(t;\theta_0,\theta_n) \tag{11}$$

[compare to the decomposition given in equation (7.12) in Shorack and Wellner (1986)]. The first two parts involve empirical processes, and can be handled by empirical process theory. The third part is nonrandom and involves the function

$$D(t;\theta_0,\theta) = \int_0^t (1 - H(s;\theta))\{d\Lambda(s;\theta) - d\Lambda(s;\theta_0)\}. \tag{12}$$

Observe that if θ coincides with θ_0, then $D(t;\theta_0,\theta)$ is identical to zero. As can be expected, the function $D(t;\theta_0,\theta)$ will show up frequently in our results. Loosely speaking, it reflects the distance between the distribution functions $F(t;\theta)$ and $F(t;\theta_0)$.

Although for the empirical process approach it is needed that G is continuous on the complete real line, we may show that our approximation results remain true on $[0, t^*)$, where $t^* = \sup\{t: G(t) < 1\}$ is finite, if G is continuous only on $(-\infty, t^*)$, by appropriately modifying G on the interval $[t^*, \infty)$. Thus, our results also have implications for Type I censoring.

Finally, we point out to the reader that Sections 2 and 3 contain results only, and that proofs are gathered in Section 4.

2. Probability inequalities. In this section, we present probability inequalities which concern the approximation on the halfline $[0,\infty)$ of (a centered version of) the process $Q_n(t)$ by a one-parameter Gaussian process, both under

P_n as under P_0. For treatment of the former situation, the following condition is needed.

CONDITION 1. There exist constants $0 < \alpha < 1/2$ and $c_\alpha < \infty$ such that

$$\int_0^\infty (1 - F(s;\theta))^\alpha \, d\Lambda(s;\theta_0) < c_\alpha$$

for every $\theta \in \Theta$.

Essentially, Condition 1 relates the right tail behavior of $F(t;\theta)$ to the right tail behavior of $F(t;\theta_0)$. Note that if $(1 - F(t;\theta))/(1 - F(t;\theta_0))$ remains uniformly bounded in θ, then Condition 1 is satisfied for any $\alpha > 0$. Observe that Condition 1 implies

$$(13) \qquad |D(t_1;\theta_0,\theta) - D(t_2;\theta_0,\theta)| \le c_1(1 - H(t_1 \wedge t_2;\theta))^{1-\alpha},$$

where $c_1 = c_\alpha + 1/\alpha$. From (13) it immediately follows that $D(t;\theta_0,\theta)$ remains bounded by c_1.

The weight process $L_n(t)$ is assumed to satisfy Condition 2.

CONDITION 2. There exists a positive nondecreasing function $q(t)$ such that:

(a) $L_n(\cdot)/q(\cdot)$ is a random element in (the left-continuous right limits version of) $D[0,\infty)$ endowed with the $\mathscr{T}_1$ metric.

(b) There exists a constant c_2 not depending on θ_n such that

$$\sup_{t\in[0,\infty)} |L_n(t)|/q(t) < c_2, \qquad V(L_n(\cdot)/q(\cdot)) < c_2,$$

with P_n-probability 1 [here $V(f)$ denotes the total variation of f].

(c) There exists a deterministic function $L(t;\theta_n)$ such that for every $\beta < 1/2$,

$$P_n\Big(\sup_{t\in[0,\infty)} |L_n(t) - L(t;\theta_n)|/q(t) > c_3 n^{-\beta}\Big) \le c_4 n^{-c_5},$$

where c_3–c_5 are positive constants not depending on θ_n.

(d) There exist constants c_6–$c_7 > 0$ such that for every $x > 0$,

$$P_0\Big(\sup_{t\in[0,\infty)} |L_n(t) - L(t;\theta_0)|/q(t) > n^{-1/2}(c_6 \log n + x)\Big) \le c_7 \exp\{-c_8 x\}.$$

Condition 2 implies that both $\sup_{t\in[0,\infty)}|L(t;\theta_n)|/q(t)$ and $V(L(\cdot;\theta_n)/q(\cdot))$ do not exceed c_2.

Subsequent results may be viewed as bearing upon $Q_n(t)/q(t)$ rather than upon $Q_n(t)$ itself. Hence, the choice of $q(t)$ will often be inflicted by the projected application of these results. For example, to study processes of the type $(1 - F(t;\theta_0))^\rho \int_0^t (1 - F(s;\theta_0))^{-\rho} \, dM_n(s;\theta_0)$ for some $\rho \ge 0$, choosing $q(t) = (1 - F(t;\theta_0))^{-\rho}$ would be appropriate.

Under Conditions 1 and 2, we have the following result. The expectation operator under P_n is denoted by $\mathscr{E}_n$.

THEOREM 1. *There exists a sequence* $\{W_n(t)\}_{n=1}^{\infty}$ *of mean zero Gaussian processes which have covariance function*

$$\begin{aligned}\mathscr{E}_n W_n(t_1) W_n(t_2) &= H^1(t_1 \wedge t_2; \theta_n) - D(t_1; \theta_0, \theta_n) D(t_2; \theta_0, \theta_n) \\ &\quad + \int_0^{t_1 \wedge t_2} \{2D(s; \theta_0, \theta_n) \\ &\qquad - D(t_1; \theta_0, \theta_n) - D(t_2; \theta_0, \theta_n)\} \, d\Lambda(s; \theta_0) \end{aligned} \tag{14}$$

such that for every $\beta < (1/2 - \alpha) \wedge 1/6$, *there exist positive constants* c_9–c_{11} *not depending on* θ_n *such that*

$$\begin{aligned} P_n\Bigg(\sup_{t \in [0,\infty)} \Bigg| \Bigg\{ Q_n(t) - n^{1/2} \int_0^t L_n(s) \, dD(s; \theta_0, \theta_n) \Bigg\} \\ - \int_0^t L(s; \theta_n) \, dW_n(s) \Bigg| \Bigg/ q(t) > c_9 n^{-\beta} \Bigg) \le c_{10} n^{-c_{11}}. \end{aligned} \tag{15}$$

Moreover, there exist positive constants c_{12}–c_{14} *such that for every* $x > 0$,

$$\begin{aligned} P_0\Bigg(\sup_{t \in [0,\infty)} \Bigg| Q_n(t) - \int_0^t L(s; \theta_0) \, dW_n(s) \Bigg| \Bigg/ q(t) > n^{-1/6} (c_{12} \log n + x)^2 \Bigg) \\ \le c_{13} \exp\{-c_{14} x\}. \end{aligned} \tag{16}$$

COROLLARY 1. *Under* P_0, *the sequence* $\{Q_n(t)/q(t)\}_{n=1}^{\infty}$ *converges in distribution to* $X(t; \theta_0)/q(t)$, *where* $X(t; \theta_0)$ *is a time-transformed Wiener process with variance function* $\int_0^t (L(s; \theta_0))^2 \, dH^1(s; \theta_0)$.

In Theorem 1, a stochastic integral with respect to a Gaussian process appears. This integral is defined in the usual way, that is, $\int_0^t L(s; \theta_n) \, dW_n(s)$ denotes $L(t; \theta_n) W_n(t) - \int_0^t W_n(s) \, dL(s; \theta_n)$.

There is a refinement of Theorem 1 worth mentioning. If the stochastic process $L_n(t)$ coincides with the function $L(t; \theta_n)$ with P_n-probability 1, then (15) holds for every $\beta < (1/2 - \alpha)$. Moreover, the term $n^{-1/6}$ in (16) may be replaced by $n^{-1/2}$.

An approximation of $Q_n(t)$ by a two-parameter Gaussian process, only valid under P_0 and on some fixed closed interval, can be found in Einmahl and Koning (1992), where it is used to derive complete analogues of the Chibisov–O'Reilly theorem, the Lai–Wellner Glivenko–Cantelli theorem and the James' law of the iterated logarithm.

Convergence in distribution under P_0 of the sequence $\{Q_n(t)/q(t)\}_{n=1}^{\infty}$ may also be obtained by using Rebolledo's martingale central limit theorem [Shorack and Wellner (1986), page 895], provided that the weight process $L_n(t)$ is predictable. However, such an approach does not lead to probability inequalities of the same type as (16), which will prove to be essential for deriving moderate deviation results for the test statistics considered in the next section.

3. Sublinear goodness-of-fit tests. In this section we study sublinear tests for the simple null hypothesis that θ_n equals θ_0. These tests are based on statistics of the form $T(Q_n(\cdot)/q(\cdot))$, where T is a special type function mapping $D[0,\infty)$ into $\mathbb{R}$. We consider the behavior of these test statistics under the null hypothesis and under fixed and local alternatives, as well as efficiencies of the corresponding tests. Generalized rank and supremum-type tests receive special attention and are shown to be in some sense optimal for specific choices of the weight process.

CONDITION 3. The function T satisfies:
(a) There exists a constant c_T such that

$$|T(\xi) - T(\eta)| \le c_T \sup_{t\in[0,\infty)} |\xi(t) - \eta(t)| \quad \text{for every } \xi, \eta \in D[0,\infty).$$

(b) $T(\xi + \eta) \le T(\xi) + T(\eta)$ for every $\xi, \eta \in D[0,\infty)$.
(c) $T(c\xi) = cT(\xi)$ for every $c \ge 0$, $\xi \in D[0,\infty)$.

By definition, Conditions 3(b) and 3(c) imply that T is a sublinear function. In Borell (1975) [see also Adler (1990)] results are derived for the tail behavior of a random variable $T(\xi)$, where T is sublinear and ξ is a Gaussian process. Condition 3(a), which appears in Komlós, Major and Tusnády (1975) and in Inglot and Ledwina (1990), enables us to carry these results to some extent over to non-Gaussian processes close to ξ.

Now, let $\mathscr{S}$ be the set of functions

$$\mathscr{S} = \left\{ f \in C[0,1]\colon f \text{ absolutely continuous and } \int_0^\infty (f'(s))^2\, ds \le 1 \right\}$$

[compare with Strassen (1964)] and define

$$a = \left\{ \sup_{f\in\mathscr{S}} T\left(\int_0^{\cdot} L(s;\theta_0) f'(H^1(s;\theta_0))\, dH^1(s;\theta_0)/q(\cdot) \right) \right\}^{-2}. \tag{17}$$

One may interpret $a^{-1/2}$ as the norm of the function T induced by the reproducing kernel Hilbert space [see Aronszajn (1950)] of the Gaussian process $X(t;\theta_0)/q(t)$, the limit in distribution of the sequence $\{Q_n(t)/q(t)\}_{n=1}^{\infty}$ under P_0. Suppose that:

CONDITION 4. a is positive.

Theorem 2 describes the tail behavior of $T(X(\cdot,\theta_0)/q(\cdot))$ [and thus the tail behavior of the asymptotic distribution of $T(Q_n(\cdot)/q(\cdot))$] and presents a moderate deviation result for $T(Q_n(\cdot)/q(\cdot))$. It holds if equation (16) and Conditions 3 and 4 hold.

THEOREM 2. *We have*

$$\lim_{t\to\infty} t^{-2}\log P_0\big(T(X(\cdot\,;\theta_0)/q(\cdot)) > t\big) = -a/2. \tag{18}$$

Furthermore,

$$\lim_{n\to\infty} (s_n)^{-2}\log P_0\big(T(Q_n(\cdot)/q(\cdot)) > s_n\big) = -a/2 \tag{19}$$

for any sequence $\{s_n\}_{n=1}^{\infty}$ *such that* $s_n \to \infty$ *and* $s_n = o(n^{1/18})$ *as* $n \to \infty$.

Since $s_n = \mathscr{O}((\log n)^{1/2})$ is a special case of $s_n = o(n^{1/18})$, Theorem 2 implies that a moderate deviation result holds for $T(Q_n(\cdot)/q(\cdot))$. Moderate deviation results are important from a statistical perspective, because they play a role in evaluating the performance of a test.

As with Theorem 1, Theorem 2 can be refined in the special case where $L_n(t)$ coincides with $L(t;\theta_0)$ with P_0-probability 1. Now $s_n = o(n^{1/18})$ may be replaced by $s_n = o(n^{1/6})$, and thus we have obtained a Cramér-type large deviation result. A Cramér-type large deviation result should be distinguished from a Chernoff-type large deviation result which allows $s_n = \mathscr{O}(n^{1/2})$.

To transform (18) into a moderate or large deviation result, probability inequalities of the type (16) are needed. Hence, Theorem 2 does not follow from Rebolledo's central limit theorem.

It is possible to generalize Theorem 2 by relaxing Condition 3(a). For example, in Inglot and Ledwina (1989), it is assumed that there exist a constant $c_T > 0$ and a weight function $\tilde{q}(t)$ belonging to some special class such that

$$|T(\xi) - T(\eta)| \le c_T \sup_{t\in[0,\infty)} |\xi(t) - \eta(t)|/\tilde{q}(t) \quad \text{for all } \xi,\eta \in D[0,\infty).$$

However, since the functions of our primary interest, those leading to generalized rank and supremum-type tests, already satisfy Condition 3(a), we have preferred to present Theorem 2 in the simple version.

Since in general the distribution function $H^1(t;\theta_0)$ is involved in a, practical problems arise when the censoring distribution G is unknown. In this case it seems best to multiply the original weight process by the square root of an estimator for a. Typically, estimators for a are obtained by replacing $H^1(t;\theta_0)$ by $\int_0^t(1 - H_{n-}(s))\,d\Lambda(s;\theta_0)$. Of course, it should be verified whether the newly constructed weight process meets all requirements.

Next, we consider the behavior of $T(Q_n(\cdot)/q(\cdot))$ under the alternative hypothesis. Before turning to local alternatives, we first briefly discuss the behavior under a fixed alternative. Suppose that θ is an element of Θ, not necessarily equal to θ_0, and that we have $\theta_n = \theta$ for all $n \in \mathbb{N}$ (we shall refer to this situation as under P_θ). Combining equation (15) with Conditions 2(c), 3(a) and 3(c) yields for every $\beta > 0$,

$$n^{-\beta}\left|T(Q_n(\cdot)/q(\cdot)) - n^{1/2}T\left(\int_0^{\cdot} L(s;\theta)\,dD(s;\theta_0,\theta)/q(\cdot)\right)\right| \to_{P_\theta} 0. \tag{20}$$

To make treatment of the behavior under local alternatives possible, we need additional notation and conditions. First we assume:

CONDITION 5. Θ is a convex subset of $\mathbb{R}^p$.

The hazard function $\lambda(t;\theta)$ is defined as the derivative with respect to t of $\Lambda(t;\theta)$. Let $\psi_i(t;\theta)$ denote the first order partial derivative of $\log \lambda(t;\theta)$ with respect to the ith component of θ and $\psi_{ij}^{(1)}(t;\theta)$ the second order partial derivative of $\log \lambda(t;\theta)$ with respect to the ith and jth component of θ.

CONDITION 6. For every $\theta \in \Theta$ and $i, j = 1, \dots, p$ the functions $\lambda(t;\theta)$, $\psi_i(t;\theta)$ and $\psi_{ij}^{(1)}(t;\theta)$ exist. For some $\beta > 0$, there exists a constant c_{15} such that for every $\theta \in \Theta$,

$$\int_0^\infty (\psi_i(s;\theta))^2(1 - H(s;\theta))^{(1-2\alpha)-\beta}\, d\Lambda(s;\theta) < c_{15} \quad \text{for } i = 1, \dots, p,$$

$$\int_0^\infty \left(\psi_{ij}^{(1)}(s;\theta)\right)^2(1 - H(s;\theta))^{2(1-2\alpha)-\beta}\, d\Lambda(s;\theta) < c_{15} \quad \text{for } i, j = 1, \dots, p.$$

In the description of the behavior of the test statistics under local alternatives, the p-dimensional vector function $K_a(t)$ is involved. The ith component of $K_a(t)$ is defined by

$$K_{ai}(t) = \int_0^t L(s;\theta_0)\psi_i(s;\theta_0)\, dH^1(s;\theta_0). \tag{21}$$

It is easily proved that $K_{ai}(t)/q(t)$ remains uniformly bounded in t and θ_0 under Conditions 1 and 6.

Recall that $L(t;\theta)$ is the limiting function of the process $L_n(t)$ under P_θ. Let $L_i^{(1)}(t;\theta)$ denote the first order partial derivative of $L(t;\theta)$ with respect to the ith component of θ and $L_{ij}^{(2)}(t;\theta)$ the second order partial derivative of $L(t;\theta)$ with respect to the ith and jth component of θ.

CONDITION 7. For every $\theta \in \Theta$ and $i, j = 1, \dots, p$, the functions $L_i^{(1)}(t;\theta)$ and $L_{ij}^{(2)}(t;\theta)$ exist. There exists a constant c_{16} such that for every $\theta \in \Theta$,

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{1/2} \left| L_i^{(1)}(t;\theta) \right| q(t) < c_{16} \quad \text{for } i = 1, \dots, p,$$

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{1-\alpha} \left| L_{ij}^{(2)}(t;\theta) \right| / q(t) < c_{16} \quad \text{for } i, j = 1, \dots, p.$$

THEOREM 3. *Suppose the sequence* $\{\theta_n\}_{n=1}^\infty$ *converges to the point* θ_0. *Let* h *be the p-dimensional unit vector defined by* $h = \lim_{n\to\infty}(\theta_n - \theta_0)/|\theta_n - \theta_0|$ *and let* σ *denote* $\lim_{n\to\infty} n^{1/2}|\theta_n - \theta_0|$.

(a) *If* $\sigma = \infty$, *then* $|(n^{1/2}|\theta_n - \theta_0|)^{-1}T(Q_n(\cdot)/q(\cdot)) - T(h^T K_a(\cdot)/q(\cdot))|$ *converges to zero in* P_n-*probability*.

(b) *If* $\sigma < \infty$, *then* $\{T(Q_n(\cdot)/q(\cdot))\}_{n=1}^\infty$ *converges in* P_n-*distribution to the random variable* $T(\{X(\cdot;\theta_0) + \sigma h^T K_a(\cdot)\}/q(\cdot))$.

Theorem 3 reveals three types of behavior of the sequence of test statistics $\{T(Q_n(\cdot)/q(\cdot))\}_{n=1}^{\infty}$, depending on the rate at which the alternatives converge to the null hypothesis. If the rate is faster than $n^{-1/2}$, then we have convergence in distribution to the same limit as under the null hypothesis. If the rate is of the order $n^{-1/2}$, then we also have convergence in distribution, but to a limit different from the one under the null hypothesis. If the rate is slower than $n^{-1/2}$, then the convergence in distribution is lost, since the sequence of test statistics blows up as n tends to infinity.

Now that we have investigated the behavior of the test statistics, it is time to evaluate the corresponding tests. For assessing the performance of a test, a multitude of efficiency concepts are available. A few of them are discussed. We start with approximate Bahadur efficiency.

DEFINITION 1. A sequence of test statistics $\{T_{in}\}_{n=1}^{\infty}$ is said to be a standard sequence if the following three conditions are satisfied.

(a) The sequence $\{T_{in}\}_{n=1}^{\infty}$ converges in P_0-distribution to a random variable T_i.

(b) There exists a positive constant a_i such that

$$\lim_{t\to\infty} t^{-2}\log P_0(T_i > t) = -a_i/2.$$

(c) For every fixed $\theta \in \Theta - \{\theta_0\}$, there exists a constant $b_i(\theta) > 0$ such that $|n^{-1/2}T_{in} - b_i(\theta)|$ converges to zero in P_θ-probability.

The approximate Bahadur slope of a standard sequence $\{T_{in}\}_{n=1}^{\infty}$ is defined as $a_i(b_i(\theta))^2$. The approximate Bahadur efficiency of a standard sequence $\{T_{1n}\}_{n=1}^{\infty}$ with respect to another standard sequence $\{T_{2n}\}_{n=1}^{\infty}$ is defined as the ratio of their respective Bahadur slopes $a_1(b_1(\theta))^2/a_2(b_2(\theta))^2$.

By Corollary 1 and equations (18) and (20), it immediately follows that the approximate Bahadur slope of the sequence $\{T(Q_n(\cdot)/q(\cdot))\}_{n=1}^{\infty}$ is given by $a\{T(\int_0^{\cdot} L(s;\theta)\,dD(s;\theta_0,\theta)/q(\cdot))\}^2$.

Approximate Bahadur efficiency has been subject to some criticism. Already in Bahadur (1960) it is advocated that conclusions should not be entirely based on approximate Bahadur slopes. In Wieand (1976) a condition is given under which the existence of the limiting (as the alternative approaches the null hypothesis) approximate Bahadur efficiency implies the existence of the limiting (as the size of the test approaches zero) asymptotic Pitman efficiency and the equality of the two limits. This condition obviates most of the difficulties involved in the interpretation of approximate Bahadur efficiencies, at least for θ in the vicinity of θ_0.

DEFINITION 2. A standard sequence $\{T_{in}\}_{n=1}^{\infty}$ is said to be a Wieand sequence if there is a constant $\varepsilon^* > 0$ such that for every $\varepsilon > 0$ and $\delta \in (0,1)$, there exists an integer N such that

$$P_\theta\big(\big|n^{-1/2}T_{in} - b_i(\theta)\big| > \varepsilon b_i(\theta)\big) < \delta$$

for all θ satisfying $|\theta - \theta_0| < \varepsilon^*$ and $n > N/(b_i(\theta))^2$.

THEOREM 4 [Wieand (1976)]. *Let $\{T_{1n}\}_{n=1}^{\infty}$ and $\{T_{2n}\}_{n=1}^{\infty}$ be two Wieand sequences. Suppose $\lim_{\theta \to \theta_0} b_i(\theta) = 0$ for $i = 1, 2$ and suppose that for every p-dimensional unit vector, the limit*

$$\lim_{\theta \to \theta_0} \left\{ a_1(b_1(\theta))^2 \right\} \Big/ \left\{ a_2(b_2(\theta))^2 \right\} \tag{22}$$

exists if $(\theta - \theta_0)/|\theta - \theta_0|$ tends to h. Then the limiting (as the size of the tests approaches zero) asymptotic Pitman efficiency exists and is equal to the limit given in (22).

THEOREM 5. *Let h be a p-dimensional unit vector and let θ approach θ_0 from the direction h (i.e., $\lim_{\theta \to \theta_0}(\theta - \theta_0)/|\theta - \theta_0| = h$). If*

$$e(h) = a\left\{ T\left(h^T K_a(\cdot)/q(\cdot)\right)\right\}^2 \tag{23}$$

is not equal to zero, then $\{T(Q_n(\cdot)/q(\cdot))\}_{n=1}^{\infty}$ is a Wieand sequence with approximate Bahadur slope of the form

$$|\theta - \theta_0|^2 \{e(h) + o(1)\}. \tag{24}$$

The Wieand approach to efficiency is based on letting both the size of the test tend to zero and the alternative tend to the null hypothesis. However, both operations are done separately. In Kallenberg (1983), a concept of efficiency is proposed based on performing both operations simultaneously. It can be considered as intermediate between the asymptotic Pitman and the exact Bahadur approach.

DEFINITION 3. A sequence of test statistics $\{T_{in}\}_{n=1}^{\infty}$ is said to be a Kallenberg sequence if the following conditions are satisfied.

(a) There exists a positive constant a_i such that

$$\lim_{n \to \infty} (s_n)^{-2} \log P_0(T_{in} > s_n) = -a_i/2$$

for all sequences $\{s_n\}_{n=1}^{\infty}$ such that $s_n \to \infty$ and $s_n = o(n^{1/6})$ as $n \to \infty$.

(b) There exists a positive function $b_i(\theta)$ such that $n^{-1/2}T_{in}/b_i(\theta_n) \to 1$ in P_n-probability for all sequences $\{\theta_n\}_{n=1}^{\infty}$ such that $\theta_n \to \theta_0$ and $n^{1/2}|\theta_n - \theta_0| \to \infty$ as n tends to infinity.

If the sequences $\{T_{1n}\}_{n=1}^{\infty}$ and $\{T_{2n}\}_{n=1}^{\infty}$ both are Kallenberg and if the limit $\lim_{n \to \infty} a_1(b_1(\theta_n))^2/a_2(b_2(\theta_n))^2$ exists, then the asymptotic intermediate efficiency of $\{T_{1n}\}_{n=1}^{\infty}$ with respect to $\{T_{2n}\}_{n=1}^{\infty}$ is defined as this limit.

Typically, $b_i(\theta)$ behaves near θ_0 as a linear function of $|\theta - \theta_0|$, which justifies introducing the intermediate slope $\lim_{n \to \infty} a_i(b_i(\theta_n)/|\theta_n - \theta_0|)^2$. If the weight process $L_n(t)$ coincides with $L(t; \theta_0)$ with P_n-probability 1, then $T(Q_n(\cdot)/q(\cdot))\}_{n=1}^{\infty}$ is a Kallenberg sequence with intermediate slope equal to the quantity $e(h)$ defined by (23), as follows from refinements of Theorem 2 and Theorem 3.

A variant of asymptotic intermediate efficiency, also proposed in Kallenberg (1983), is weak asymptotic intermediate efficiency. Here only sequences $\{s_n\}_{n=1}^{\infty}$ such that $s_n \to \infty$ and $s_n = \mathscr{O}((\log n)^{1/2})$ as $n \to \infty$ are considered. Observe that the sequence $\{T(Q_n(\cdot)/q(\cdot))\}_{n=1}^{\infty}$ has weak intermediate slope $e(h)$, even if the weight process does not coincide with its limiting weight function. Hence, the weak intermediate approach yields the same picture as the Wieand approach.

The concepts in Kallenberg (1983) were proposed so as to correspond with several types of moderate and large deviation results. In light of Theorem 2, it is tempting to propose a variant of asymptotic intermediate efficiency, which considers sequences $\{s_n\}_{n=1}^{\infty}$ such that $s_n \to \infty$ and $s_n = o(n^{1/18})$ as n tends to infinity.

In the beginning of this section, we assumed that the functional T is sublinear. A close look reveals that Condition 3(b) is used in the derivation of (18) only. Thus, we may set up an equivalent theory for functionals other than sublinear, provided a result similar to (18) holds. As examples we mention the functionals occurring in Cramér–von Mises and chi-square tests [see Durbin (1973)].

However, we have preferred to restrict our attention to the class of sublinear tests, since it comprises two particularly appealing subclasses, the class of generalized rank tests and the class of supremum-type tests. The remainder of this section is devoted to these two subclasses.

Generalized rank tests are based on tests statistics of the form $T_R(Q_n)$, where the functional T_R is defined by

$$T_R(\xi) = \xi(\infty) \quad \text{for } \xi \in D[0,\infty). \tag{25}$$

Examples of generalized rank tests can be found in Breslow (1975) [see also Woolson (1981)] and Harrington and Fleming (1982). Supremum-type tests are based on test statistics of the form $T_S(Q_n)$, where T_S is defined by

$$T_S(\xi) = \sup_{t\,[0,\infty)} \xi(t) \quad \text{for } \xi \in D[0,\infty). \tag{26}$$

For examples of tests of supremum type, refer to Aki (1986) and Harrington and Fleming (1982).

It is easily seen that both T_R and T_S satisfy Condition 3. Hence, our theory applies [set $q(t)$ equal to 1 for all $t \in [0,\infty)$]. Defining a_R and a_S according to (17), with T replaced by T_R and T_S, respectively, it follows that

$$a_R = a_S = \left\{\int_0^{\infty} (L(s;\theta_0))^2\, dH^1(s;\theta_0)\right\}^{-1}. \tag{27}$$

Similarly defining $e_R(h)$ and $e_S(h)$ according to (23), we obtain

$$e_R(h) = a_R\{h^T K_a(\infty)\}^2, \tag{28}$$

$$e_S(h) = a_S\left\{\sup_{t\in[0,\infty)} h^T K_a(t)\right\}^2. \tag{29}$$

It should be noted that if the weight process coincides with its limiting function with P_n-probability 1, then $T_R(Q_n)$ can be written as the sum of i.i.d. random variables [see Proposition 3.1 in Aki (1986)] and hence results for this special type generalized rank test may be proven in a simpler manner. For example, Theorem 2 is now an immediate consequence of Theorem 1 on page 549 of Feller (1971). Observe that this alternative proof also leads to a Cramér-type large deviation result.

As opposed to general sublinear tests, general rank tests do allow us to compute asymptotic relative Pitman efficiencies. By Corollary 1 and Theorem 3(b), it follows that the asymptotic power against local alternatives $\theta_n = \theta_0 + n^{-1/2}h$ of the test based on $T_R(Q_n)$ of size $\tilde{\alpha}$ equals

$$P_0\big(X(\infty;\theta_0) > z_{\tilde{\alpha}}(a_R)^{-1/2} - h^T K_a(\infty)\big),$$

where $z_{\tilde{\alpha}}$ is the $(1-\tilde{\alpha})$ quantile of the standard normal distribution. This implies that the efficacy of the sequence of test statistics $\{T_R(Q_n)\}_{n=1}^{\infty}$ is equal to $e_R(h)$.

Since applying the Cauchy–Schwarz inequality to $h^T K_a(\infty)$ yields

$$e_R(h) \le \int_0^{\infty} \big(h^T \psi(s;\theta_0)\big)^2 \, dH^1(s;\theta_0), \tag{30}$$

where $\psi(t;\theta_0)$ is the p-dimensional vector with elements $\psi_i(t;\theta_0)$, it follows that $e_R(h)$ (and hence asymptotic relative Pitman efficiency, limiting approximate Bahadur efficiency and weak intermediate efficiency of generalized rank tests) is maximized by those tests based on weight processes with limiting weight function satisfying

$$L(t;\theta_0) \propto h^T \psi(t;\theta_0) \tag{31}$$

provided that the weight process conforms to Condition 2. Furthermore, it can be shown [see Lemma 1 in Koning (1991)] that the Fisher information matrix $I(\theta_0)$ equals the $p \times p$ matrix with elements

$$I_{ij}(\theta_0) = \int_0^{\infty} \psi_i(s;\theta_0)\psi_j(s;\theta_0) \, dH^1(s;\theta_0), \tag{32}$$

from which it follows that the upper bound for the efficacy derived in Rao (1963) coincides with the right-hand side of inequality (30) and thus generalized rank tests based on weight processes with limiting weight function satisfying (31) are asymptotically most powerful.

Clearly, the use of a generalized rank test instead of a classical test (the likelihood ratio test, say) does not have to result in loss of asymptotic relative efficiency. This raises the question whether the same conclusion holds for a supremum-type test. After all, in contrast to the generalized rank test which is not consistent against alternatives approaching the null hypothesis in a direction perpendicular to $K_a(\infty)$, the supremum-type test has the character of an omnibus test. Needless to say, it is rather attractive to have an efficient omnibus test at our disposal.

Unfortunately, the asymptotic distribution of the supremum-type test statistic is not normal. Although the concept of asymptotic Pitman efficacy can be extended to test statistics which have nonnormal asymptotic distributions [e.g., refer to Wieand (1976)], it turns out that in the case of the supremum-type test, this efficacy depends on the size of the test in a rather complicated way.

In virtually the same way as with $e_R(h)$, we obtain that $e_S(h)$ is bounded by the right-hand side of (30) and that this upper bound is attained by supremum-type tests based on weight processes with limiting weight function satisfying (31). It follows that these tests have efficiency 1 (in the sense of limit asymptotic Pitman efficiency, limiting approximate Bahadur efficiency and weak intermediate efficiency) with respect to the locally most powerful generalized rank test based on the same weight process.

This last result may impel to question the usefulness of concepts which are not able to distinguish between generalized rank tests and supremum-type tests. However, as recent results on the tail behavior of the supremum of a Gaussian process show, this inability is basically a consequence of letting the size of the test tend to zero, which is the sensible thing to do if we are committed to avoiding making errors, Type I as well as Type II. From Rubin and Sethuraman (1965) it follows that minimizing the Bayes risk leads to letting the size of the test tend to zero at a rate n^{-1}. Observe that this is exactly the situation to which weak intermediate efficiency refers.

For any nonnegative value of ρ, the foregoing theory is applicable to the one-parameter family of alternatives,

$$F(t;\theta) = \begin{cases} 1 - \left[1 + \left((1 - F(t;\theta_0))^{-\rho} - 1\right)\exp\{\theta - \theta_0\}\right]^{-1/\rho}, & \text{if } \rho > 0, \\ 1 - (1 - F(t;\theta_0))^{\exp\{\theta - \theta_0\}}, & \text{if } \rho = 0, \end{cases}$$

which was introduced in Harrington and Fleming (1982). A particularly attractive property of this family is the simple form of $h^T\psi(t;\theta)$ (here h is the one-dimensional unit vector):

$$h^T\psi(t;\theta) = (1 - F(t;\theta))^{\rho}.$$

By making use of the fact that the first order partial derivative of $h^T\psi(t;\theta)$ with respect to θ is equal to $h^T\psi(t;\theta)(h^T\psi(t;\theta) - 1)$, we may verify our conditions and reach the conclusion that the generalized rank and supremum-type tests proposed in Harrington and Fleming (1982), based on weight processes $(1 - F(t;\theta_0))^{\rho}$, are optimal in the sense of limiting asymptotic Pitman efficiency, limiting approximate Bahadur efficiency and weak intermediate efficiency. Moreover, the generalized rank tests are optimal in the sense of asymptotic Pitman efficiency.

The family of Harrington and Fleming alternatives couples computational simplicity with practical relevance. It includes proportional hazards alternatives ($\rho = 0$) and logistic shift alternatives ($\rho = 1$) as important special cases. Observe that in the theory of rank tests the score functions for these alternatives lead to Savage and Wilcoxon tests.

We end this section by stressing that generalized rank and supremum-type tests are not the only sublinear tests worth noticing. For example, refer to Aki and Kashiwagi (1989) for a sublinear test based on a functional other than T_R and T_S.

4. Proofs. In this section we prove the theorems presented in Sections 2 and 3. We shall make repeated use of the following inequality, an adapted version of the inequality of Fernique (1970, 1971) [see also Adler (1990)].

INEQUALITY 1. There exist positive constants c_{17} and c_{18} such that for every separable mean zero Gaussian process $Z(t)$ satisfying

$$P\Big(\sup_{t\in[0,\infty)} |Z(t)| < \infty\Big) = 1,$$

and for every $x > 0$,

$$P\bigg(\sup_{t\in[0,\infty)} |Z(t)| > x\Big\{\sup_{t\in[0,\infty)} E\{Z(t)\}^2\Big\}^{1/2}\bigg) \le c_{17}\exp\{-c_{18}x^2\}.$$

PROOF OF THEOREM 1. As in Einmahl and Koning (1992), proof of Proposition 1 [see also Theorem 3.1 in Burke, Csörgő and Horváth (1981)], let $\tilde{U}_n$ denote the empirical process based on the uniform (0, 1) random variables

$$\tilde{Z}_i = \delta_i H^1(Z_i;\theta_n) + (1-\delta_i)\{H^1(\infty;\theta_n) + H^0(Z_i;\theta_n)\},$$

where

$$H^0(t;\theta_n) = H(t;\theta_n) - H^1(t;\theta_n)$$

is the cumulative distribution function of the censored failure times under P_n. Note that

$$U_n^1(t;\theta_n) = \tilde{U}_n(H^1(t;\theta_n)),$$

$$U_{n-}(t;\theta_n) = \tilde{U}_{n-}(H^1(t;\theta_n)) + \tilde{U}_{n-}(H^1(\infty;\theta_n) + H^0(t;\theta_n)) - \tilde{U}_{n-}(H^1(\infty;\theta_n)),$$

where $\tilde{U}_{n-}(t)$ denotes the left-continuous version of $\tilde{U}_n(t)$. The approximation theorem of Komlós, Major and Tusnády (1975) yields the existence of a sequence $\{\tilde{B}_n(t)\}_{n=1}^{\infty}$ of Brownian bridges with continuous sample paths such that for all $x > 0$,

$$P_n\Big(\sup_{t\in[0,\infty)} |\tilde{U}_n(t) - \tilde{B}_n(t)| > n^{-1/2}(c_{19}\log n + x)\Big) \le c_{20}\exp\{-c_{21}x\}, \tag{33}$$

where c_{19}–c_{21} are universal constants.

Now define mean zero Gaussian processes $B_n^1(t)$, $B_n(t)$ and $W_n(t)$ by

$$B_n^1(t) = \tilde{B}_n\big(H^1(t;\theta_n)\big),$$
$$B_n(t) = \tilde{B}_n\big(H^1(t;\theta_n)\big) + \tilde{B}_n\big(H^1(\infty;\theta_n) + H^0(t;\theta_n)\big) - \tilde{B}_n\big(H^1(\infty;\theta_n)\big),$$
$$W_n(t) = B_n^1(t) + \int_0^t B_n(s)\, d\Lambda(s;\theta_0).$$

The processes $B_n^1(t)$ and $B_n(t)$ are used to approximate $U_n^1(t;\theta_n)$ and $U_{n-}(t;\theta_n)$, respectively. Thus, it follows by (11) that $W_n(t)$ approximates $M_n(t;\theta_0) - n^{1/2}D(t;\theta_0,\theta_n)$. Before studying the implications of this approximation, we first pay attention to the covariance structure of $W_n(t)$. Covariance calculations yield

$$\mathscr{E}_n B_n^1(t_1)B_n^1(t_2) = H^1(t_1 \wedge t_2;\theta_n) - H^1(t_1;\theta_n)H^1(t_2;\theta_n),$$
$$\mathscr{E}_n B_n^1(t_1)B_n(t_2) = H^1(t_1 \wedge t_2;\theta_n) - H^1(t_1;\theta_n)H(t_2;\theta_n),$$
$$\mathscr{E}_n B_n(t_1)B_n(t_2) = H(t_1 \wedge t_2;\theta_n) - H(t_1;\theta_n)H(t_2;\theta_n),$$

and hence (14). Observe that (14) implies

$$\begin{aligned}\mathscr{E}_n\{W_n(t_1) - W_n(t_2)\}^2 &\\ = H^1(t_1 \vee t_2;\theta_n) - H^1(t_1 \wedge t_2;\theta_n) - \big(D(t_1;\theta_0,\theta_n) - D(t_2;\theta_0,\theta_n)\big)^2 &\\ + 2\int_{t_1\wedge t_2}^{t_1\vee t_2}\{D(s;\theta_0,\theta_n) - D(t_1 \vee t_2;\theta_0,\theta_n)\}\, d\Lambda(s;\theta_0), &\end{aligned} \tag{34}$$

from which we may infer

$$\begin{aligned}\mathscr{E}_n\{W_n(t_1) - W_n(t_2)\}^2 &\le H^1(t_1 \vee t_2;\theta_n) - H^1(t_1 \wedge t_2;\theta_n)\\ &\quad + 2c_1\int_{t_1\wedge t_2}^{t_1\vee t_2}(1 - H(s;\theta_n))^{1-\alpha}\, d\Lambda(s;\theta_0)\\ &\le c_{22}\int_{t_1\wedge t_2}^{t_1\vee t_2}(1 - H(s;\theta_n))^{\alpha}\{d\Lambda(s;\theta_n) + d\Lambda(s;\theta_0)\},\end{aligned} \tag{35}$$

where $c_{22} = 1 + 2c_1$. To gain some probabilistic insight in the process $W_n(t)$, remark that

$$W_n(t) - \int_0^t \frac{B_n(s)}{1 - H(s;\theta_n)}\, dD(s;\theta_0,\theta_n)$$

is a time-transformed Wiener process with variance function $H^1(t;\theta_n)$. Moreover, the process $B_n(t)/1 - H(t;\theta_n)$ is a time-transformed Wiener process with variance function $H(t;\theta_n)/(1 - H(t;\theta_n))$.

Recall that $W_n(t)$ approximates $M_n(t;\theta_0) - n^{1/2}D(t;\theta_0,\theta_n)$. As a direct consequence we have that $\int_0^t L_n(s)\{dM_n(s;\theta_0) - n^{1/2}\, dD(s;\theta_0,\theta_n)\}$ is approximated by $\int_0^t L_n(s)\, dW_n(s)$. Unfortunately, the latter process is not easy to work with (it may not even be Gaussian), so we prefer replacing $\int_0^t L_n(s)\, dW_n(s)$ by $\int_0^t L(s;\theta_n)\, dW_n(s)$. To evaluate the effects of this replacement, we make use of a pure jump process $J_n(t)$, which is defined by

$$J_n(t) = W_n(x_{i,n}) \quad \text{for } t \in I_{i,n}, \tag{36}$$

where $I_{i,n} = [x_{i,n}, x_{i+1,n})$ and $0 = x_{0,n} < x_{1,n} < \cdots < x_{m(n),n} = \infty$ is a grid chosen so as to satisfy

$$\begin{aligned} &\int_{I_{i,n}} (1 - H(s;\theta_n))^{\alpha}\{d\Lambda(s;\theta_n) + d\Lambda(s;\theta_0)\} \\ &\qquad \le n^{-1/3} \int_0^{\infty} (1 - H(s;\theta_n))^{\alpha}\{d\Lambda(s;\theta_n) + d\Lambda(s;\theta_0)\} \end{aligned} \tag{37}$$

for $i = 0, \dots, m(n) - 1$. If the grid is chosen carefully, then there is no need for $m(n)$ to exceed $n^{1/3} + 1$. We shall assume that this is indeed the case.

It follows by (35) that the variance of the mean zero Gaussian process $W_n(t)$ is bounded by $c_{22}c_1$, while the variance of the mean zero Gaussian process $J_n(t) - W_n(t)$ is bounded by $c_{22}c_1 n^{-1/3}$. Hence, by applying Inequality 1 we obtain

$$P_n\Big(\sup_{t\in[0,\infty)} |J_n(t)| > x\Big) \le P_n\Big(\sup_{t\in[0,\infty)} |W_n(t)| > x\Big) \le c_{17}\exp\{-c_{23}x^2\}, \tag{38}$$

$$P_n\Big(\sup_{t\in[0,\infty)} |J_n(t) - W_n(t)| > xn^{-1/6}\Big) \le c_{17}\exp\{-c_{23}x^2\}, \tag{39}$$

where $c_{23} = c_{18}/(c_{22}c_1)$. Note that the application of Inequality 1 to $J_n(t) - W_n(t)$ is justified since this process is separable.

For any sequence $\{d_n\}_{n=1}^{\infty}$ of points in $(0,\infty)$, we may now write

$$\begin{aligned} &\sup_{t\in[0,\infty)} \bigg| \bigg\{ Q_n(t) - n^{1/2}\int_0^t L_n(s)\,dD(s;\theta_0,\theta_n) \bigg\} \\ &\qquad\qquad - \int_0^t L(s;\theta_n)\,dW_n(s) \bigg| \bigg/ q(t) \le \sum_{i=1}^{6} \Delta_{ni}, \end{aligned} \tag{40}$$

where

$$\Delta_{n1} = \sup_{t\in[0,d_n]} \bigg| Q_n(t) - \int_0^t L_n(s)\{dW_n(s) + n^{1/2}\,dD(s;\theta_0,\theta_n)\} \bigg| \bigg/ q(t),$$

$$\Delta_{n2} = \sup_{t\in[d_n,\infty)} |Q_n(t) - Q_n(d_n)|/q(t),$$

$$\Delta_{n3} = \sup_{t\in[d_n,\infty)} \bigg| \int_{d_n}^t L_n(s)\,dW_n(s) \bigg| \bigg/ q(t),$$

$$\Delta_{n4} = n^{1/2} \sup_{t\in[d_n,\infty)} \bigg| \int_{d_n}^t L_n(s)\,dD(s;\theta_0,\theta_n) \bigg| \bigg/ q(t),$$

$$\Delta_{n5} = \sup_{t\in[0,\infty)} \bigg| \int_0^t L_n(s)\{dW_n(s) - dJ_n(s)\} \bigg| \bigg/ q(t),$$

$$\Delta_{n6} = \sup_{t\in[0,\infty)} \bigg| \int_0^t \{L_n(s) - L(s;\theta_n)\}\,dJ_n(s) \bigg| \bigg/ q(t).$$

Observe that Δ_{n5} and Δ_{n6} relate to the replacement of $\int_0^t L_n(s)\,dW_n(s)$ by $\int_0^t L(s;\theta_n)\,dW_n(s)$.

Later in this proof we will meet two specific choices of $\{d_n\}_{n=1}^{\infty}$. A first choice leads to (15) and a second to (16). Both choices have in common that $d_n \to \infty$ as $n \to \infty$. But before these choices are made, we explore the behavior of Δ_{ni}, $i = 1, \dots, 6$, for general sequences. Integration by parts yields, with P_n-probability 1,

$$\begin{aligned}\Delta_{n1} &\le \Big\{ \sup_{t\in[0,\infty)} |L_n(t)/q(t)| + V(L_n(\cdot)/q(\cdot))\Big\}\\ &\quad\times\Big\{ \sup_{t\in[0,d_n]} \big|M_n(t;\theta_0) - n^{1/2}D(t;\theta_0,\theta_n) - W_n(t)\big|\Big\}\\ &\le 2c_2\Big\{ \sup_{t\in[0,d_n]} |U_n^1(t;\theta_n) - B_n^1(t)|\\ &\qquad + \sup_{t\in[0,d_n]}\Big|\int_0^t (U_{n-}(s;\theta_n) - B_n(s))\,d\Lambda(s;\theta_0)\Big|\Big\}\\ &\le 2c_2(1 + 3\Lambda(d_n;\theta_0)) \sup_{t\in[0,1]} |\tilde{U}_n(t) - \tilde{B}_n(t)|,\end{aligned}$$

and therefore it follows by (33) that

$$\begin{aligned}&P_n\big(\Delta_{n1} > 2c_2(1 + 3\Lambda(d_n;\theta_0))n^{-1/2}(c_{19}\log n + x)\big)\\ &\qquad \le c_{20}\exp\{-c_{21}x\}.\end{aligned} \tag{41}$$

Observing that $\Delta_{n2} = 0$ if $Z_{n:n} < d_n$, where $Z_{n:n}$ denotes the largest order statistic of the sample $Z_1, \dots, Z_n$, we obtain

$$\begin{aligned}P_n(\Delta_{n2} \ne 0) &\le P_n(Z_{n:n} \ge d_n)\\ &\le 1 - (H(d_n;\theta_n))^n \le n(1 - H(d_n;\theta_n)).\end{aligned} \tag{42}$$

Twice integrating by parts yields, with P_n-probability 1,

$$\begin{aligned}\Delta_{n3} &\le \Big\{V(L_n(\cdot)/q(\cdot)) + \sup_{t\in[0,\infty)} |L_n(t)|/q(t)\Big\}\\ &\quad\times\Big\{ \sup_{t\in[d_n,\infty)} \Big|\int_{d_n}^t q(s;\theta_0)\,dW_n(s)\Big|\Big/q(t)\Big\}\\ &\le 4c_2 \sup_{t\in[d_n,\infty)} |W_n(t) - W_n(d_n)|.\end{aligned}$$

Thus, applying Inequality 1 yields

$$\begin{aligned}&P_n\Bigg(\Delta_{n3} > 4c_2x\Big\{\sup_{t\in[d_n,\infty)} \mathscr{E}_n\{W_n(t) - W_n(d_n)\}^2\Big\}^{1/2}\Bigg)\\ &\qquad \le c_{17}\exp\{-c_{18}x^2\}.\end{aligned} \tag{43}$$

Similarly, combining integration by parts with inequality (39) produces

$$P_n\big(\Delta_{n5} > 4c_2 x n^{-1/6}\big) \le c_{17}\exp\{-c_{23}x^2\}. \tag{44}$$

Furthermore, we have with P_n-probability 1,

$$\begin{aligned}\Delta_{n6} \le &\left\{\sup_{t\in[0,\infty)}\left|\int_0^t q(s;\theta_0)\,dJ_n(s)\right|\Big/q(t)\right\}\\ &\times\Bigg\{\Bigg\{\sum_{i=0}^{m(n)-1}\big|\{L_n(x_{i+1,n}) - L(x_{i+1,n};\theta_n)\}/q(x_{i+1,n})\\ &\qquad -\{L_n(x_{i,n}) - L(x_{i,n};\theta_n)\}/q(x_{i,n})\big|\Bigg\}\\ &\qquad + \sup_{t\in[0,\infty)}|L_n(t) - L(t;\theta_n)|/q(t)\Bigg\}\\ \le &\,(4n^{1/3}+6)\Big\{\sup_{t\in[0,\infty)}|J_n(t)|\Big\}\Big\{\sup_{t\in[0,\infty)}|L_n(t) - L(t;\theta_n)|/q(t)\Big\}.\end{aligned} \tag{45}$$

A general investigation of the behavior of Δ_{n4} is rather useless, since it depends too much on the actual situation. Thus, we content ourselves with (41)–(45).

Next we turn to the aforementioned specific choices of $\{d_n\}_{n=1}^{\infty}$. For both choices, the inequalities just derived are sharpened, supplemented by an inequality for Δ_{n4} and combined.

To prove (15), choose $\beta < (1/2-\alpha)\wedge 1/6$ and d_n so as to satisfy $H(d_n;\theta_n) = 1 - n^{-(\gamma+1)}$, where $\gamma = (1/2-\alpha-\beta)/\alpha$. Note that $\gamma > 0$. Since

$$\begin{aligned}\Lambda(d_n;\theta_0) &< (1 - H(d_n;\theta_n))^{-\alpha}\int_0^{d_n}(1 - H(s;\theta_n))^{\alpha}\,d\Lambda(s;\theta_0)\\ &\le c_\alpha n^{1/2-\beta},\end{aligned} \tag{46}$$

it follows from (41) that

$$P_n\big(\Delta_{n1} > 2c_2(1+3c_\alpha)(1+c_{19})n^{-\beta}\log n\big) \le c_{20}n^{-c_{21}}. \tag{47}$$

By (42) we immediately have

$$P_n(\Delta_{n2} \ne 0) \le n^{-\gamma}. \tag{48}$$

From (34) we may infer for $t \ge d_n$,

$$\mathscr{E}_n\{W_n(t;\theta_0) - W_n(d_n;\theta_0)\}^2 \le c_{22}c_1(1 - H(d_n;\theta_n))^{1-2\alpha},$$

and hence (43) implies

$$P_n\Big(\Delta_{n3} > 4c_2\{c_{22}c_1 n^{-(\gamma+2\beta)}\log n\}^{1/2}\Big) \le c_{17}n^{-c_{18}}. \tag{49}$$

Integration by parts yields with P_n-probability 1

$$\Delta_{n4} \le 2c_2 n^{1/2} \sup_{t \in [d_n, \infty)} |D(t;\theta_0,\theta_n) - D(d_n;\theta_0,\theta_n)|,$$

which leads in combination with (13) to

$$P_n\big(\Delta_{n4} > 2c_2c_1 n^{-(\beta+\gamma)}\big) = 0. \tag{50}$$

From (44) we obtain

$$P_n\big(\Delta_{n5} > 4c_2 n^{-1/6}(\log n)^{1/2}\big) \le c_{17} n^{-c_{23}}. \tag{51}$$

Combining Condition 2(c) with equations (38) and (45), it follows that

$$P_n\big(\Delta_{n6} > 10c_3 n^{-\beta}(\log n)^{1/2}\big) \le (c_4 + c_{17}) n^{-(c_5 \wedge c_{23})} \tag{52}$$

(note that we have used $\beta + 1/3 < 1/2$). Now, (47)–(52) together with (40) yield (15).

If $\theta_n = \theta_0$, then we may obtain a sharper result by making a different choice of d_n. Let $x > 0$ and choose d_n so as to satisfy $H(d_n;\theta_0) = 1 - \exp\{-x\}/n$. By noting that $\Lambda(d_n;\theta_0) \le -\log(1 - H(d_n;\theta_0)) = \log n + x$, we obtain from (41) for $n > 1$,

$$P_0\big(\Delta_{n1} > 2c_2 n^{-1/2}(5\log n + 3x)(c_{19}\log n + x)\big) \le c_{20}\exp\{-c_{21}x\}. \tag{53}$$

Furthermore, we have by (42),

$$P_0(\Delta_{n2} \ne 0) \le \exp\{-x\}, \tag{54}$$

and since by (34),

$$\begin{aligned} x \sup_{t \in [d_n,\infty)} \mathscr{E}_n\{W_n(t) - W_n(d_n)\}^2 &\le x\big(H^1(\infty;\theta_0) - H^1(d_n;\theta_0)\big) \\ &\le x\big(1 - H(d_n;\theta_0)\big) \le n^{-1}, \end{aligned}$$

it follows by (43)

$$P_0\big(\Delta_{n3} > 4c_2 n^{-1/2}\big) \le c_{17}\exp\{-c_{18}x\}. \tag{55}$$

As a simple consequence of the fact that $D(t;\theta_0,\theta_0)$ is identical to zero, we have

$$P_0(\Delta_{n4} \ne 0) = 0. \tag{56}$$

Equation (44) implies

$$P_0\big(\Delta_{n5} > 4c_2 n^{-1/6}x\big) \le 2c_{17}\exp\{-c_{23}x\}, \tag{57}$$

while combining Condition 2(d) with equations (38) and (45) yields

$$P_0\big(\Delta_{n6} > 10 n^{-1/6}x^{1/2}(c_6\log n + x)\big) \le (c_7 + c_{17})\exp\{-(c_8 \wedge c_{23})x\}. \tag{58}$$

Hence, (53)–(58) together with (40) yield (16).

We conclude the proof of Theorem 1 with the remark that if $L_n(t)$ coincides with $L(t;\theta_n)$ with P_n-probability 1, it suffices to bound the left-hand side of (40) by $\sum_{i=1}^4 \Delta_{ni}$. This yields the refinement of Theorem 1 mentioned at the end of Section 2. □

PROOF OF THEOREM 2. Equation (18) directly follows from Theorem 5.2 in Borell (1975).

To prove (19), observe that as a consequence of Theorem 1 there exists a process $X_n(t;\theta_0)$, equal in P_0-distribution to $X(t;\theta_0)$, which satisfies

$$\begin{aligned}(59)\qquad & P_0\Big(\sup_{t\in[0,\infty)}|Q_n(t)-X_n(t;\theta_0)|/q(t)>n^{-\gamma}(c_{12}\log n+x)^{\tau}\Big)\\ &\quad\le c_{13}\exp\{-c_{14}x\},\end{aligned}$$

with $\gamma=1/6$ and $\tau=2$. Now, let $\rho=\gamma/(2\tau-1)$ and choose $\tau^{-1}<\beta<2$. Note that for $n\to\infty$,

$$n^{(2-\beta)\rho}(s_n)^{\beta}\gg n^{(2-\beta)\rho}\gg\log n,$$
$$n^{(2-\beta)\rho}(s_n)^{\beta}=(n^{-\rho}s_n)^{\beta-2}(s_n)^2\gg(s_n)^2,$$
$$(n^{-\rho}s_n)^{\beta\tau-1}\ll 1$$

(here $a_n\gg b_n$ denotes $b_n/a_n\to 0$) and hence (59) implies

$$\begin{aligned}& P_0\big(|T(Q_n(\cdot)/q(\cdot))-T(X_n(\cdot;\theta_0)/q(\cdot))|>c_T n^{(1-\beta\tau)\rho}(s_n)^{\beta\tau}\big)\\ &\quad\le P_0\Big(\sup_{t\in[0,\infty)}|Q_n(t)-X_n(t;\theta_0)|/q(t)>n^{(1-\beta\tau)\rho}(s_n)^{\beta\tau}\Big)\\ (60)\qquad &\quad\le c_{13}\exp\Big\{-c_{14}\Big(n^{(2-\beta)\rho}(s_n)^{\beta}-c_{12}\log n\Big)\Big\}\\ &\quad\ll\exp\big\{-a(s_n)^2/2\big\}.\end{aligned}$$

Since $T(X_n(\cdot;\theta_0)/q(\cdot))$ and $T(X(\cdot;\theta_0)/q(\cdot))$ are equal in distribution under P_0, we may bound $P_0(T(Q_n(\cdot)/q(\cdot))>s_n)$ from below by

$$\begin{aligned}& P_0\Big(T(X(\cdot;\theta_0)/q(\cdot))>s_n\Big(1+c_T(n^{-\rho}s_n)^{\beta\tau-1}\Big)\Big)\\ &\qquad -P_0\big(|T(X_n(\cdot;\theta_0)/q(\cdot))-T(Q_n(\cdot)/q(\cdot))|>c_T n^{(1-\beta\tau)\rho}(s_n)^{\beta\tau}\big)\end{aligned}$$

and from above by

$$\begin{aligned}& P_0\Big(T(X(\cdot;\theta_0)/q(\cdot))>s_n\Big(1-c_T(n^{-\rho}s_n)^{\beta\tau-1}\Big)\Big)\\ &\qquad +P_0\big(|T(X_n(\cdot;\theta_0)/q(\cdot))-T(Q_n(\cdot)/q(\cdot))|>c_T n^{(1-\beta\tau)\rho}(s_n)^{\beta\tau}\big)\end{aligned}$$

and thus it follows that (18) and (60) together yield (19). This concludes the proof of Theorem 2. □

The proofs of Theorem 3 and Theorem 5 make repeated use of the following lemma, which holds under Conditions 1 and 6.

LEMMA 1. *Let $g(t;\theta)$ be a real-valued function, $g_i^{(1)}(t;\theta)$ the first order partial derivative of $g(t;\theta)$ with respect to the ith component of θ. Suppose* *there exists a constant* c_{24} *such that*

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{1/2} |g(t;\theta)| < c_{24}, \tag{61}$$

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{1-\alpha} |g_i^{(1)}(t;\theta)| < c_{24} \tag{62}$$

for every $\theta \in \Theta$. *Then there exists a constant* c_{25} *such that*

$$\sup_{t \in [0,\infty)} \left| \int_0^t g(s;\theta)\, dD(s;\theta_0,\theta) \right| \le c_{25} |\theta - \theta_0|, \tag{63}$$

$$\sup_{t \in [0,\infty)} \left| \int_0^t g(s;\theta)\, dH^1(s;\theta) - \int_0^t g(s;\theta_0) dH^1(s;\theta_0) \right| \le c_{25} |\theta - \theta_0| \tag{64}$$

for every $\theta \in \Theta$. *Let* $g_{ij}^{(67)}(t;\theta)$ *be the second order partial derivative of* $g(t;\theta)$ *with respect to the ith and jth components of* θ. *If*

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{\alpha} |g(t;\theta)| < c_{24}, \tag{65}$$

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{1/2} |g_i^{(1)}(t;\theta)| < c_{24}, \tag{66}$$

$$\sup_{t \in [0,\infty)} (1 - H(t;\theta))^{1-\alpha} |g_{ij}^{(2)}(t;\theta)| < c_{24}, \tag{67}$$

then

$$\begin{aligned} \sup_{t \in [0,\infty)} \Bigg| \int_0^t g(s;\theta)\, dD(s;\theta_0,\theta) \\ - \int_0^t (\theta - \theta_0)^T \psi(s;\theta_0) g(s;\theta_0)\, dH^1(s;\theta_0) \Bigg| \\ \le c_{25} |\theta - \theta_0|^2 \end{aligned} \tag{68}$$

for every $\theta \in \Theta$.

COROLLARY 2. *There exists a constant* c_{25} *such that*

$$\sup_{t \in [0,\infty)} \left| \int_0^t (1 - H(s;\theta))^{-1/2}\, dD(s;\theta_0,\theta) \right| \le c_{25} |\theta - \theta_0|$$

for every $\theta \in \Theta$.

PROOF OF LEMMA 1. For convenience, we introduce the functions

$$\Lambda_0(t;\theta) = \Lambda(t;\theta) - \Lambda(t;\theta_0),$$
$$\lambda_0(t;\theta) = \lambda(t;\theta) - \lambda(t;\theta_0).$$

Let $\Lambda_i^{(1)}(t;\theta)$ and $\lambda_i^{(1)}(t;\theta)$, respectively, denote the first order partial derivatives of $\Lambda_0(t;\theta)$ and $\lambda_0(t;\theta)$ with respect to the ith component of θ. Furthermore, let $\Lambda_{ij}^{(2)}(t;\theta)$ and $\lambda_{ij}^{(2)}(t;\theta)$, respectively, denote the second order partial derivatives of $\Lambda_0(t;\theta)$ and $\lambda_0(t;\theta)$ with respect to the ith and jth components of θ. By Condition 6 we have for every $t \in [0,\infty)$,

$$\begin{aligned}
&\int_0^t (1 - H(s;\theta))^{1/2-\alpha} |\lambda_i^{(1)}(s;\theta)|\, ds \\
&\quad = \int_0^t (1 - H(s;\theta))^{1/2-\alpha} |\psi_i(s;\theta)|\, d\Lambda(s;\theta) \\
&\quad \le \left\{ \int_0^\infty (1 - H(s;\theta))^{\beta}\, d\Lambda(s;\theta) \right\}^{1/2} \\
&\qquad \times \left\{ \int_0^\infty (\psi_i(s;\theta))^2 (1 - H(s;\theta))^{1-2\alpha-\beta}\, d\Lambda(s;\theta) \right\}^{1/2} \\
&\quad \le (c_{15}/\beta)^{1/2},
\end{aligned} \tag{69}$$

and hence

$$\begin{aligned}
&(1 - H(t;\theta))^{1/2-\alpha} |\Lambda_i^{(1)}(t;\theta)| \\
&\quad = (1 - H(t;\theta))^{1/2-\alpha} \left| \int_0^t \lambda_i^{(1)}(s;\theta)\, ds \right| \\
&\quad \le \int_0^t (1 - H(s;\theta))^{1/2-\alpha} |\lambda_i^{(1)}(s;\theta)|\, ds \\
&\quad \le (c_{15}/\beta)^{1/2}.
\end{aligned} \tag{70}$$

Now, we find by using the identity

$$1 - H(t;\theta) = (1 - H(t;\theta_0))\exp\{-\Lambda_0(t;\theta)\}$$

that the first order partial derivative of $\int_0^t g(s;\theta)\lambda_0(s;\theta)(1 - H(s;\theta))\, ds$ with respect to the ith component of θ is given by

$$\int_0^t \{\{g_i^{(1)}(s;\theta) - g(s;\theta)\Lambda_i^{(1)}(s;\theta)\}\lambda_0(s;\theta) + g(s;\theta)\lambda_i^{(1)}(s;\theta)\}(1 - H(s;\theta))\, ds$$

and is bounded in t and θ, as follows from Condition 1, (61), (62), (69) and (70). By expressing $D(t;\theta_0,\theta)$ as $\int_0^t (1 - H(s;\theta))\lambda_0(s;\theta)\, ds$ we obtain (63).

As for (64), by writing $H^1(t;\theta) = \int_0^t (1 - H(s;\theta))\lambda(s;\theta)\, ds$, it follows that the first order partial derivative of $\int_0^t g(s;\theta)\, dH^1(s;\theta)$ with respect to the ith component of θ is given by

$$\int_0^t \{g_i^{(1)}(s;\theta) + g(s;\theta)\{\psi_i(s;\theta) - \Lambda_i^{(1)}(s;\theta)\}\}(1 - H(s;\theta))\lambda(s;\theta)\, ds$$

and is bounded in t and θ.

Finally (68). For every $t \in [0,\infty)$, we have

$$\int_0^t (1 - H(s;\theta))^{1-2\alpha} \left|\lambda_{ij}^{(2)}(s;\theta)\right| ds$$

$$= \int_0^t (1 - H(s;\theta))^{1-2\alpha} \left|\psi_{ij}^{(1)}(s;\theta) + \psi_i(s;\theta)\psi_j(s;\theta)\right| d\Lambda(s;\theta)$$

$$\le \left\{\int_0^\infty (1 - H(s;\theta))^{\beta}\, d\Lambda(s;\theta)\right\}^{1/2}$$

$$\times\left\{\int_0^\infty \left(\psi_{ij}^{(1)}(s;\theta)\right)^2 (1 - H(s;\theta))^{2(1-2\alpha)-\beta}\, d\Lambda(s;\theta)\right\}^{1/2} \tag{71}$$

$$+\left\{\int_0^\infty (\psi_i(s;\theta))^2 (1 - H(s;\theta))^{1-2\alpha-\beta}\, d\Lambda(s;\theta)\right\}^{1/2}$$

$$\times\left\{\int_0^\infty (\psi_j(s;\theta))^2 (1 - H(s;\theta))^{1-2\alpha-\beta}\, d\Lambda(s;\theta)\right\}^{1/2}$$

$$\le (c_{15}/\beta)^{1/2} + c_{15}$$

and

$$(1 - H(t;\theta))^{1-2\alpha} \left|\Lambda_{ij}^{(2)}(t;\theta)\right| \le (c_{15}/\beta)^{1/2} + c_{15}. \tag{72}$$

Note that the first order partial derivative of $\int_0^t g(s;\theta)\lambda_0(s;\theta)(1 - H(s;\theta))\,ds$ with respect to the ith component of θ equals $\int_0^t g(s;\theta_0)\psi_i(s;\theta_0)\,dH^1(s;\theta_0)$ when evaluated at $\theta = \theta_0$. Furthermore, the second order partial derivative with respect to the ith and jth component of θ is given by

$$\int_0^t \Big\{\big\{g_{ij}^{(2)}(s;\theta) - g_i^{(1)}(s;\theta)\Lambda_j^{(1)}(s;\theta) - g_j^{(1)}(s;\theta)\Lambda_i^{(1)}(s;\theta)$$

$$-g(s;\theta)\Lambda_{ij}^{(2)}(s;\theta) + g(s;\theta)\Lambda_i^{(1)}(s;\theta)\Lambda_j^{(1)}(s;\theta)\big\}\lambda_0(s;\theta)$$

$$+\big\{g_j^{(1)}(s;\theta) - g(s;\theta)\Lambda_j^{(1)}(s;\theta)\big\}\lambda_i^{(1)}(s;\theta)$$

$$+\big\{g_i^{(1)}(s;\theta) - g(s;\theta)\Lambda_i^{(1)}(s;\theta)\big\}\lambda_j^{(1)}(s;\theta)$$

$$+g(s;\theta)\lambda_{ij}^{(2)}(s;\theta)\Big\}(1 - H(s;\theta))\,ds.$$

Hence, (68) follows from Condition 1, Condition 6, (62), (65), (67) and (69)–(72). This concludes the proof of Lemma 1. □

PROOF OF THEOREM 3. Let $\{W_n(t)\}_{n=1}^\infty$ be the sequence of mean zero Gaussian processes given in Theorem 1. We may write

$$\sup_{t\in[0,\infty)} \left|Q_n(t) - n^{1/2}\left|\theta_n - \theta_0\right| h^T K_a(t)\right|/q(t) \le \sum_{i=1}^{4} \Delta_{ni}, \tag{73}$$

where

$$\Delta_{n1} = \sup_{t \in [0,\infty)} \left| \left\{ Q_n(t) - n^{1/2} \int_0^t L_n(s)\, dD(s;\theta_0,\theta_n) \right\} - \int_0^t L(s;\theta_n)\, dW_n(s) \right| \Big/ q(t),$$

$$\Delta_{n2} = \sup_{t \in [0,\infty)} \left| \int_0^t L(s;\theta_n)\, dW_n(s) \right| \Big/ q(t),$$

$$\Delta_{n3} = n^{1/2} \sup_{t \in [0,\infty)} \left| \int_0^t (L_n(s) - L(s;\theta_n))\, dD(s;\theta_0,\theta_n) \right| \Big/ q(t),$$

$$\Delta_{n4} = n^{1/2} \sup_{t \in [0,\infty)} \left| \int_0^t L(s;\theta_n)\, dD(s;\theta_0,\theta_n) - |\theta_n - \theta_0| h^T K_a(t) \right| \Big/ q(t).$$

By (15) we have for $\beta < (1/2 - \alpha) \wedge 1/6$,

$$P_n(\Delta_{n1} > c_9 n^{-\beta}) \le c_{10} n^{-c_{11}}. \tag{74}$$

From integration by parts we obtain with P_n-probability 1

$$\Delta_{n2} \le 2c_2 \sup_{t \in [0,\infty)} \left| \int_0^t q(s)/q(t)\, dW_n(s) \right| \le 4c_2 \sup_{t \in [0,\infty)} |W_n(t)|.$$

Since by (35) and Condition 1 the variance function of the process $W_n(t)$ is bounded by $(c_{22})^2$, it follows by Inequality 1 that

$$P_n(\Delta_{n2} > 4c_2 c_{22} x) \le c_{17} \exp\{-c_{18} x^2\}. \tag{75}$$

Lemma 1 yields the existence of a constant c_{26} such that

$$\Delta_{n4} \le c_{26} n^{1/2} \{ |\theta_n - \theta_0|^2 + |(\theta_n - \theta_0) - |\theta_n - \theta_0| h| \}, \tag{76}$$

and for every $\theta \in \Theta$ and $\beta < 1/2$,

$$P_n\left(n^\beta \sup_{t \in [0,\infty)} \int_0^t (L_n(s) - L(s;\theta_n))\, dD(s;\theta_0,\theta)/q(t) > c_{26} |\theta - \theta_0| \right)$$

$$\le P_n\left(n^\beta \sup_{t \in [0,\infty)} \sup_{0 \le s \le t} |L_n(s) - L(s;\theta_n)|/q(t) > c_3 \right)$$

[observe that $L_n(t) - L(t;\theta_n)$ does not depend on θ and hence the first order partial derivatives with respect to θ are all equal to zero]. Note that by Condition 2(c) this last inequality implies

$$P_n\left((n^{1/2} |\theta_n - \theta_0|)^{-1} \Delta_{n3} > c_{26} n^{-\beta} \right) \le c_4 n^{-c_5}. \tag{77}$$

The first part of the theorem is now easily proved by means of the inequality

$$\begin{aligned} &\left| (n^{1/2} |\theta_n - \theta_0|)^{-1} T(Q_n(\cdot)/q(\cdot)) - T(h^T K_a(\cdot)/q(\cdot)) \right| \\ &\quad \le c_T \left(n^{1/2} |\theta_n - \theta_0| \right)^{-1} \sum_{i=1}^4 \Delta_{ni}. \end{aligned} \tag{78}$$

To prove the second part of the theorem, we first note that the supremum over t of the absolute difference between the processes $Q_n(t)/q(t)$ and $\{\int_0^t L(s;\theta_n)\,dW_n(s) + n^{1/2}|\theta_n - \theta_0| h^T K_a(t)\}/q(t)$ is bounded by $\Delta_{n1} + \Delta_{n3} + \Delta_{n4}$, and hence converges to zero in P_n-probability if $n^{1/2}|\theta_n - \theta_0|$ tends to a finite limit as $n \to \infty$. Thus, it remains to show that the latter process converges in P_n-distribution to $\{X(t;\theta_0) + \sigma h^T K_a(t)\}/q(t)$, which boils down to the convergence in P_n-distribution of $\int_0^t L(s;\theta_n)\,dW_n(s)/q(t)$ to $X(t;\theta_0)/q(t)$.

Let $B_n(t)$ be as defined in the proof of Theorem 1. Split the process $\int_0^t L(s;\theta_n)\,dW_n(s)/q(t)$ into the two parts

$$\int_0^t L(s;\theta_n)\left\{dW_n(s) - \frac{B_n(s)}{1 - H(s;\theta_n)}\,dD(s;\theta_0,\theta_n)\right\}\bigg/q(t)$$

and

$$\int_0^t L(s;\theta_n)\frac{B_n(s)}{1 - H(s;\theta_n)}\,dD(s;\theta_0,\theta_n)\bigg/q(t).$$

The first part may be interpreted as a time-transformed Wiener process divided by $q(t)$, the time-transformed Wiener process having variance function $\int_0^t (L(s;\theta_n))^2\,dH^1(s;\theta_n)$. It follows from Lemma 1 and Theorem VI.10 in Pollard (1984) that the first part converges in P_n-distribution to $X(t;\theta_0)/q(t)$. Use Lemma 1 to show that the supremum over t of the second part converges to zero in P_n-probability. This completes the proof of Theorem 3. □

Proof of Theorem 5. Let Δ_{n1}, Δ_{n2}, Δ_{n3} and Δ_{n4} be as in the proof of Theorem 3 and define $b(\theta_n)$ as $T(\int_0^\cdot L(s;\theta_n)\,dD(s;\theta_0,\theta_n)/q(\cdot))$. We may write

$$(79)\qquad \left|n^{-1/2}T(Q_n(\cdot)/q(\cdot)) - b(\theta_n)\right| \le \Delta_{n1} + \Delta_{n2} + \Delta_{n3}.$$

Furthermore, we have

$$(80)\qquad \left|b(\theta_n) - |\theta_n - \theta_0| T\big(h^T K_a(\cdot)/q(\cdot)\big)\right| \le c_T n^{-1/2}\Delta_{n4}.$$

Since $e(h)$ is not equal to zero, this yields the existence of positive constants ε^* and c_{27} such that for θ_n satisfying $|\theta_n - \theta_0| < \varepsilon^*$,

$$(81)\qquad c_{27}|\theta_n - \theta_0| < b(\theta_n) < 1.$$

Now suppose $\theta \in \Theta - \{\theta_0\}$ satisfies $|\theta - \theta_0| < \varepsilon^*$ and set θ_n equal to θ. Choose $\varepsilon > 0$ and $\delta \in (0,1)$. By (74) and (75), it follows that there exists an integer N_1 not depending on θ such that for $n > N_1$,

$$P_\theta\big(\Delta_{n1} > (N_1)^{1/2}\varepsilon/4c_T\big) < \delta/4,$$
$$P_\theta\big(\Delta_{n2} > (N_1)^{1/2}\varepsilon/4c_T\big) < \delta/4.$$

Hence, for $n > N_1/(b(\theta))^2$, we have

$$(82)\qquad P_\theta\big(n^{-1/2}\{\Delta_{n1} + \Delta_{n2}\} > \varepsilon b(\theta)/2c_T\big) < \delta/2,$$

since $n > N_1$ and $(N_1/n)^{1/2} < b(\theta)$. Moreover, (77) implies the existence of an integer $N > N_1$ depending on θ such that for $n > N$,

$$(83) \quad P_\theta(\Delta_{n3} > \varepsilon b(\theta)/2c_T) \le P_\theta\Big(\big(n^{1/2}|\theta - \theta_0|\big)^{-1}\Delta_{n3} > c_{27}\varepsilon\Big) < \delta/2.$$

Combining (79)–(83) now yields that $\{T(Q_n(\cdot)/q(\cdot))\}_{n=1}^{\infty}$ is indeed a Wieand sequence.

Finally, (24) immediately follows from (76) and (80). This concludes the proof of Theorem 5. □

Acknowledgments. The author is indebted to Wim Albers and John Einmahl for detailed comments on earlier drafts.

REFERENCES

AALEN, O. O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27.

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.

ADLER, R. J. (1990). *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*. IMS, Hayward, Calif.

AKI, S. (1986). Some test statistics based on the martingale term of the empirical distribution function. *Ann. Inst. Statist. Math.* **38** 1–21.

AKI, S. and KASHIWAGI, N. (1989). Asymptotic properties of some goodness-of-fit tests based on the L_1 norm. *Ann. Inst. Statist. Math.* **41** 753–764.

ARONSZAJN, N. (1950). The theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

BAHADUR, R. R. (1960). Stochastic comparisons of tests. *Ann. Math. Statist.* **31** 276–295.

BORELL, C. (1975). The Brunn–Minkowsky inequality in Gauss space. *Invent. Math.* **30** 207–216.

BRESLOW, N. E. (1975). Analysis of survival data under proportional hazards model. *Internat. Statist. Rev.* **43** 45–58.

BURKE, M. D., CSÖRGŐ, S. and HORVÁTH, L. (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrsch. Verw. Gebiete* **56** 87–112.

DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, Philadelphia.

EINMAHL, J. H. J. and KONING, A. J. (1992). Limit theorems for a general weighted process under random censoring. *Canad. J. Statist.* **20**.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

FERNIQUE, X. (1970). Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris Ser. A* **270** 1698–1699.

FERNIQUE, X. (1971). Régularité des processus gaussiens. *Invent. Math.* **12** 304–320.

FLEMING, T. R. and HARRINGTON, D. P. (1981). A class of hypothesis tests for one and two samples of censored data. *Comm. Statist. Theory Methods* **13** 2469–2486.

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

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

HARRINGTON, D. P. and FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69** 553–566.

HJORT, N. L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18** 1221–1258.

INGLOT, T. and LEDWINA, T. (1989). Large and moderate deviations of some functionals of weighted empirical process. Technical report, Inst. Mathematics, Technical Univ., Wroclaw.

INGLOT, T. and LEDWINA, T. (1990). On probabilities of excessive deviations for Kolmogorov–Smirnov, Cramér–von Mises and chi-square statistics. *Ann. Statist.* **18** 1491–1495.

KALLENBERG, W. C. M. (1983). Intermediate efficiency, theory and examples. *Ann. Statist.* **11** 170–182.

KHMALADZE, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.

KHMALADZE, E. V. (1982). Some applications of the theory of martingales to statistics. *Uspekhi Mat. Nauk* **37** 193–212. [Translation: *Russian Math. Surveys* **37** 215–237.]

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **34** 33–58.

KONING, A. J. (1991). Stochastic integrals and goodness-of-fit tests. Ph.D. dissertation, Univ. Twente.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

RAO, C. R. (1963). Criteria of estimation in large samples. *Sankhyā Ser. A* **25** 189–206.

RUBIN, H. and SETHURAMAN, J. (1965). Bayes risk efficiency. *Sankhyā Ser. A* **27** 347–356.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226.

WIEAND, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* **4** 1003–1011.

WOOLSON, R. F. (1981). Rank tests and a one-sample logrank test for comparing observed survival data to a standard population. *Biometrics* **37** 687–696.

ECONOMETRIC INSTITUTE
ERASMUS UNIVERSITY
P.O. BOX 1738
3000 DR ROTTERDAM
THE NETHERLANDS