The Annals of Statistics
1994, Vol. 22, No. 2, 565–579

APPROXIMATION OF THE BASIC MARTINGALE

BY ALEX J. KONING

Erasmus University Rotterdam

Probability inequalities governing the approximation of the basic martingale, the rescaled difference between a counting process and its compensator, are derived. Applications to the random censoring model and to the field of goodness-of-fit tests are given. The inequalities are compared to inequalities derived earlier for the random censoring model using an empirical process approach.

1. Introduction. Let $(\Omega, \mathcal{A}, P)$ be a probability space. For each $n = 1, 2, \ldots$, let $N_n = \{N_n(t)\}_{t \in [0, \infty)}$ be a counting process adapted to some filtration $\{\mathcal{A}_{nt}\}_{t \in [0, \infty)}$ of σ-algebras contained in $\mathcal{A}$. Define A_n as the compensator of $n^{-1}N_n$ with respect to $\{\mathcal{A}_{nt}\}_{t \in [0, \infty)}$.

In Section 2 we construct a time-transformed Wiener process, which strongly approximates

$$M_n = n^{1/2}\{n^{-1}N_n - A_n\}. \tag{1}$$

Moreover, we give an exponential inequality for the distance (in the supremum metric) between M_n and the approximating process. This inequality is of a type similar to those given in Komlós, Major and Tusnády (1975) for the partial sum and the empirical process.

In the random censoring model (see Section 3) it is more or less customary to refer to M_n as the basic martingale. Although the random censoring model has provided the main motivation for this study, there are also useful applications in other fields. An example is given in Section 4.

Section 6 contains the proofs. The tools involved are presented in Section 5.

2. Main results. Throughout it is assumed that A_n is continuous. Define the inverse of A_n by

$$A_n^{-1}(t) = \inf\{s\colon A_n(s) \geq t\}. \tag{2}$$

Due to the continuity of A_n, it follows by Aalen and Hoem (1978) [see also Theorem 18.10 in Liptser and Shiryayev (1978), page 280, or Theorem II.16 in Brémaud (1981), page 41] that $N_n \circ A_n^{-1}$ is a Poisson process with intensity n, starting at zero and randomly stopped at $A_n(\infty)$. Observe that the independent standard exponential random variables of Barlow and Proschan [(1969), Lemma 1] basically are the interarrival times of this stopped Poisson process, multiplied by n.

Received October 1991; revised June 1993.

AMS 1991 *subject classifications.* Primary 60G55; secondary 62G10, 60F15.

Key words and phrases. Counting process, random time change, approximation, random censoring, goodness-of-fit.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

At time $A_n(\infty)$, start up an independent Poisson process with intensity n and attach it to $N_n \circ A_n^{-1}$, as in Gill (1986). In this way we have constructed a Poisson process on $[0, \infty]$, say π_n, which satisfies $\pi_n(\cdot \wedge A_n(\infty)) = N_n \circ A_n^{-1}$. Thus, for the centered Poisson process $\pi_n^c = \pi_n - nI$, we have

$$n^{-1/2}\pi_n^c(t) = M_n \circ A_n^{-1}(t) \quad \text{for } t \le A_n(\infty). \tag{3}$$

The process $n^{-1/2}\pi_n^c$ allows approximation by means of a standard Wiener process. This leads to the following theorem.

THEOREM 1. *Let $x > 0$, $\beta \le \frac{1}{2}$ and $n^{-2\beta}(21\log n + x) \le d_{n,x} \le n^{1/2-\beta}$. If the probability space $(\Omega, \mathcal{A}, P)$ is sufficiently rich, then there exists a standard Wiener process W_n such that for every $c_4, c_5 > 0$ there exist constants c_1, c_2 and c_3, depending only on c_4 and c_5, such that*

$$P\Bigl(\sup_{t\in[0,\infty)} \bigl|M_n \circ A_n^{-1} \circ \xi_n(t) - W_n \circ \xi(t)\bigr| > c_1 n^{\beta} d_{n,x}\Bigr) \le c_2 \exp\{-c_3 x\}, \tag{4}$$

for every random element $\xi_n \in D[0,\infty)$ satisfying

$$P\bigl(0 \le \xi_n(t) \le A_n(\infty) \text{ for every } t > 0\bigr) = 1, \tag{5}$$

$$P\Bigl(\sup_{t\in[0,\infty)} \bigl|\xi_n(t) - \xi(t)\bigr| > d_{n,x}\Bigr) \le c_4 \exp\{-c_5 x\}, \tag{6}$$

where $\xi \in D[0,\infty)$ is deterministic and bounded by 1. Moreover, if $\xi_n(t) = t \wedge \tau_n$, where τ_n is a random variable, then (4) also holds for $n^{-2\beta}x \le d_{n,x} < n^{-2\beta}(21\log n + x)$.

It is essential that ξ remains bounded. The requirement that ξ is in fact bounded by 1 is quite natural, as the applications in Section 3 and 4 will show.

In Theorem 1 the restriction $d_{n,x} < n^{1/2-\beta}$ occurs. Without this restriction (which does not exclude the more interesting values of x) the right-hand side of (4) should be replaced by $c_2\exp\{-c_3 x\psi(n^{1/2-\beta}d_{n,x})\}$, where the function ψ is described in Section 5. In fact, the restriction is used to bound $\psi(n^{1/2-\beta}d_{n,x})$ from below.

For given c_4 and c_5, explicit values of c_1, c_2 and c_3 can be obtained by reworking the proof of Theorem 1 in Bretagnolle and Massart (1989). In this proof a centered Poisson process and an approximating Wiener process are constructed on the same probability space.

Because values of β between $\frac{1}{4}$ and $\frac{1}{2}$ are in practice seldom encountered, the mathematical constraint $\beta \le \frac{1}{2}$ is outdone by the statistical constraint $\beta \le \frac{1}{4}$. In the important cases $\xi_n(t) = A_n(t)$ and $\xi_n(t) = t \wedge A_n(\infty)$, we often have $\beta = \frac{1}{4}$, as the following lemma shows.

LEMMA 1. *Suppose that there exist nonnegative nondecreasing stochastic* *processes* $A_{n1}, \ldots, A_{nn}$ *such that*

$$A_n = n^{-1} \sum_{i=1}^{n} A_{ni} \tag{7}$$

and

$$P\big(A_{ni}(\infty) > x\big) \le e^{-x}. \tag{8}$$

If the random variables $A_{n1}(\infty), \ldots, A_{nn}(\infty)$ *are independent, then*

$$P\Big(\big|A_n(\infty) - \mathcal{E}A_n(\infty)\big| > 4n^{-1/2}x\Big) \le \exp\bigg\{ -\frac{x^2}{1 + n^{-1/2}x} \bigg\}, \tag{9}$$

for every $x > 0$. *Moreover, if the processes* $A_{n1}, \ldots, A_{nn}$ *are independent, then*

$$P\bigg(\sup_{t \in [0,\infty)} \big|A_n(t) - \mathcal{E}A_n(t)\big| > 4n^{-1/2}(\log n + x)^{3/2} \bigg) \le 28e^{-x}. \tag{10}$$

Lemma 1 provides universal constants c_4 and c_5 for (6), and hence universal constants c_1, c_2 and c_3 for (4).

An example of a full application of Lemma 1 is given in the next section.

3. The random censoring model. Let the failure times $X_1, \ldots, X_n$ be nonnegative independent random variables defined on $(\Omega, \mathcal{A}, P)$, each having a continuous distribution function F_i. These failure times are not observed directly, and the only information concerning them is contained in the random variables

$$Z_i = X_i \wedge Y_i, \qquad \delta_i = 1_{\{X_i \le Y_i\}}, \qquad i = 1, \ldots, n, \tag{11}$$

where $Y_1, \ldots, Y_n$ are random variables also defined on $(\Omega, \mathcal{A}, P)$ and which are independent of $X_1, \ldots, X_n$. Depending on the value of the censoring indicator δ_i, Z_i is either called an observed ($\delta_i = 1$) or a censored ($\delta_i = 0$) failure time.

In many applications of this so-called random censoring model, the simultaneous distribution of $Y_1, \ldots, Y_n$ is an unknown nuisance parameter. Hence, assumptions concerning this distribution should be kept to a minimum.

In theoretical work on the model, one often encounters the stochastic process

$$n^{1/2}\{H_n^1 - A_n\} \tag{12}$$

[see Shorack and Wellner (1986), page 296], where

$$H_n^1 = n^{-1} \sum_{i=1}^{n} \delta_i 1_{\{Z_i \le \cdot\}} \quad \text{and} \quad A_n = n^{-1} \sum_{i=1}^{n} \int_0^{\cdot} 1_{\{Z_i \ge s\}}\, d\Lambda_i(s), \tag{13}$$

and Λ_i denotes the cumulative hazard function belonging to X_i, which is equal to $-\log(1-F_i)$ due to the continuity of F_i. Since nH_n^1 is a counting process with compensator nA_n with respect to the filtration

$$\mathcal{A}_{nt} = \sigma\big(1_{\{Z_i \le s\}}, \delta_i 1_{\{Z_i \le s\}}: i = 1, \dots, n, s \in [0, t]\big) \tag{14}$$

[see Shorack and Wellner (1986), page 310], it is easily seen that the process (12) is in fact the basic martingale M_n. Observe that A_n is continuous.

As the sample size n increases, the basic martingale converges to a time-transformed Wiener process. Under the additional assumption that $X_1, \dots, X_n$ have a common continuous distribution function F and that $Y_1, \dots, Y_n$ are independent and have a common continuous distribution function G, the refinement of Theorem 1 in Koning (1992) implies that for every $n > 1$ there exists a standard Wiener process $\widetilde{W}_n$ such that

$$\begin{aligned} &P\Big(\sup_{t \in [0,\infty)} \big|M_n(t) - \widetilde{W}_n \circ H^1(t)\big| > n^{-\gamma}(\widetilde{c}_1 \log n + x)^{\kappa}\Big) \\ &\quad \le \widetilde{c}_2 \exp\{-\widetilde{c}_3 x\}, \end{aligned} \tag{15}$$

for every $x > 0$. Here $H^1 = \int_0^{\cdot}(1-G(s))\,dF(s)$; $\widetilde{c}_1, \widetilde{c}_2$ and $\widetilde{c}_3$ are universal constants; and γ and κ take the values $\frac{1}{2}$ and 2, respectively. This result was obtained by using an empirical process approach.

The probability inequality (15) can be used to study the behavior of statistics of the form $T(M_n)$, which for instance occur in the field of goodness-of-fit tests. For functionals T fulfilling the Lipschitz condition

$$\big|T(\xi) - T(\zeta)\big| \le c_T \sup_{t \in [0,\infty)} \big|\xi(t) - \zeta(t)\big| \quad \text{for every } \xi, \zeta \in D[0,\infty), \tag{16}$$

it directly follows from (15) by setting x equal to $\log n$ that

$$n^{\gamma}(\log n)^{-\kappa}\big|T(M_n) - T(\widetilde{W}_n \circ H^1)\big| = O_P(1). \tag{17}$$

For Lipschitz functionals, (15) also implies a deviation result as the following lemma shows.

LEMMA 2. *Let $\{T_n\}_{n=1}^{\infty}$ be a sequence of random variables. Suppose there exist positive constants a, γ, κ, $\widetilde{c}_1$, $\widetilde{c}_2$ and $\widetilde{c}_3$ and a sequence $\{\widetilde{T}_n\}_{n=1}^{\infty}$ of identically distributed random variables such that*

$$\lim_{t \to \infty} t^{-2} \log P(\widetilde{T}_n > t) = -a/2,$$

and, for every $0 < x < n^{2\gamma/(2\kappa-1)}$,

$$P\big(|T_n - \widetilde{T}_n| > n^{-\gamma}(\widetilde{c}_1 \log n + x)^{\kappa}\big) \le \widetilde{c}_2 \exp\{-\widetilde{c}_3 x\}.$$

Then

$$\lim_{n \to \infty} (t_n)^{-2} \log P(T_n > t_n) = -a/2,$$

for any sequence $\{t_n\}_{n=1}^{\infty}$ *such that* $t_n \to \infty$ *and* $t_n = o\big(n^{\gamma/(2\kappa-1)}\big)$ *as* $n \to \infty$.

Thus, if for any given nonnegative bounded nondecreasing function K the Lipschitz functional T satisfies

$$\lim_{t \to \infty} t^{-2} \log P\big(T(\widetilde{W}_n \circ K) > t\big) = -a_K/2, \tag{18}$$

for some $a_K > 0$, then (15) implies the deviation result

$$\lim_{n \to \infty} (t_n)^{-2} \log P\big(T(M_n) > t_n\big) = -a_{H^1}/2, \tag{19}$$

for any sequence $\{t_n\}_{n=1}^{\infty}$ such that $t_n \to \infty$ and $t_n = o(n^{1/6})$.

In particular, (18) holds for sublinear functionals, as follows from Borell (1975) [see also Adler (1990)]. A functional T: $D[0,\infty) \to R$ is said to be sublinear if $T(\xi + \zeta) \leq T(\xi) + T(\zeta)$ and $T(c\xi) = cT(\xi)$ for every $c \geq 0$ and $\xi, \zeta \in D[0,\infty)$. The functionals defined by

$$T_R(\xi) = \xi(\infty) \quad \text{and} \quad T_S(\xi) = \sup_{t \in [0,\infty)} \xi(t), \tag{20}$$

for every $\xi \in D[0,\infty)$, are both sublinear and Lipschitz.

The empirical process approach in Koning (1992) necessitated the aforementioned distributional assumptions on the failure times $X_1, \dots, X_n$ and the censoring times $Y_1, \dots, Y_n$. Lemma 3 enables us to apply the theory of Section 2 to the random censoring model described in the first paragraph of this section. This leads to the results given in Corollary 1, which remain true even if the empirical process approach assumptions do not hold.

LEMMA 3. *Suppose that* $Y_1, \dots, Y_n$ *are independent. Then there exist nonnegative nondecreasing and independent stochastic processes* $A_{n1}, \dots, A_{nn}$, *namely,* $A_{ni} = \Lambda_i(Z_i \wedge \cdot)$, *such that* (7) *and* (8) *hold.*

Because $\mathcal{E}A_n(t) = \mathcal{E}H_n^1(t) - n^{-1/2}\mathcal{E}M_n(t) = \mathcal{E}H_n^1(t)$, taking the expectation of A_n yields the (possibly defective) distribution function $\overline{H}^1$, defined by

$$\overline{H}^1(t) = n^{-1} \sum_{i=1}^{n} P(Z_i \leq t,\ \delta_i = 1). \tag{21}$$

Note that $\overline{H}^1$ and H^1 coincide if both $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ are identically distributed random variables.

COROLLARY 1. *There exists a standard Wiener process* W_n *such that*

$$P\Big(\sup_{t \in [0,\infty} \big|M_n \circ A_n^{-1}\big(t \wedge A_n(\infty)\big) - W_n\big(t \wedge \overline{H}^1(\infty)\big)\big| > n^{-1/4}x\Big) \leq c_2 \exp\{-c_3 x\},$$

$$P\Big(\sup_{t \in [0,\infty)} \big|M_n(t) - W_n \circ \overline{H}^1(t)\big| > n^{-1/4}(21 \log n + x)^{3/2}\Big) \leq c_2 \exp\{-c_3 x\},$$

where c_2 and c_3 are universal constants.

Define I^1 by $I^1(t) = 0 \vee t \wedge \overline{H}^1(\infty)$. By setting x equal to $\log n$, it follows by Corollary 1 that

$$n^{1/4}(\log n)^{-1}\left|T\left(M_n \circ A_n^{-1}\right) - T(W_n \circ I^1)\right| = O_P(1), \tag{22}$$

$$n^{1/4}(\log n)^{-3/2}\left|T(M_n) - T\left(W_n \circ \overline{H}^1\right)\right| = O_P(1), \tag{23}$$

for any Lipschitz functional T. Moreover, if T satisfies (18), then

$$\lim_{n\to\infty}(t_n)^{-2}\log P\left(T\left(M_n \circ A_n^{-1}\right) > t_n\right) = -a_{I^1}/2, \tag{24}$$

$$\lim_{n\to\infty}(t_n')^{-2}\log P\left(T(M_n) > t'_n\right) = -a_{\overline{H}^1}/2, \tag{25}$$

for all sequences $\{t_n\}_{n=1}^{\infty}$ and $\{t_n'\}_{n=1}^{\infty}$ such that $t_n, t_n' \to \infty$, $t_n = o(n^{1/4})$ and $t_n' = o(n^{1/8})$.

To give some examples of interesting test statistics for which these results hold, let L be a bounded nonrandom function of bounded variation, and define the functionals T_R^L and T_S^L by

$$T_R^L(\xi) = T_R\left(\xi L - \int_0^{\cdot} \xi(s)\,dL(s)\right), \tag{26}$$

$$T_S^L(\xi) = T_S\left(\xi L - \int_0^{\cdot} \xi(s)\,dL(s)\right), \tag{27}$$

for every $\xi \in D[0,\infty)$. Note that if we set L identical to 1, then $T_R^L = T_R$ and $T_S^L = T_S$. Both T_R^L and T_S^L are Lipschitz, which gives us (22) and (23). Moreover, as a consequence of sublinearity both T_R^L and T_S^L satisfy (18) with

$$a_K = \int_0^{\infty} \left(L(s)\right)^2 dK(s), \tag{28}$$

and hence we have (24) and (25) with

$$a_{I^1} = \int_0^{\overline{H}^1(\infty)} \left(L(s)\right)^2 ds, \qquad a_{\overline{H}^1} = \int_0^{\infty} \left(L(s)\right)^2 d\overline{H}^1(s). \tag{29}$$

Test statistics $T_R^L(M_n)$ and $T_S^L(M_n)$ belong to a larger class investigated by means of the empirical process approach in Koning (1992). Both $T_R(M_n)$ [see Breslow (1975)] and $T_S(M_n)$ [see Aki (1986)] are in a certain sense optimal for testing against proportional hazard alternatives. Recall that they coincide with $T_R(M_n \circ A_n^{-1})$ and $T_S(M_n \circ A_n^{-1})$, and note that for these two specific statistics (24) improves (19), the result obtained in Koning (1992). Statistics $T_R^{(1-F)}(M_n)$ and $T_S^{(1-F)}(M_n)$ [see Harrington and Fleming (1982)] are optimal with respect to logistic shift alternatives.

The rate of convergence results (22) and (23) are somewhat disappointing as compared with (17), derived in Koning (1992) by means of the empirical process approach. This is due to the random time change, employed in the present approach to transform the basic martingale into a rescaled centered Poisson process, which introduces randomness in the "horizontal" direction. In the empirical process approach we only encounter randomness in the "vertical" direction. The local characteristics of the centered Poisson process, as reflected by its modulus of continuity, make randomness in the horizontal direction essentially harder to deal with. However, the applicability of the empirical process approach is limited by the fact that in order to invoke the empirical process approximation in Komlós, Major and Tusnády (1975), we need $Z_1, \dots, Z_n$ to have a common continuous distribution function.

Another drawback of the empirical process approach is that the Wiener process approximating the basic martingale is constructed out of a Brownian bridge. Hence, the error at point t arising from the empirical process approximation is blown up as $F(t)$ tends to 1. This becomes especially noticeable when the approximation is used to bound tail probabilities of the basic martingale; it also explains why the deviation results (25) and (19) do not differ as dramatically as do (23) and (17): the increased "tail accurateness" of the present approach is partially making up for the negative effects of the random time change. This is even stronger for the process $M_n \circ A_n^{-1}$, as can be seen from (24).

4. Parametric compensators. We now apply the theory developed in Section 2 to processes which occur in the field of goodness-of-fit tests, in particular when the null hypothesis is composite [see Khmaladze (1981), (1982)]. The following lemma is the key result.

LEMMA 4. *Let $\widehat{\theta}_n$ be a random variable; let $\tau_1, \dots, \tau_n$ be a sequence of random variables satisfying $0 < \tau_1 < \cdots < \tau_n < \infty$ almost surely; and let the σ-algebras contained in the filtration $\{\mathcal{A}_{nt}\}_{t \in [0,\infty)}$ be of the form*

$$\mathcal{A}_{nt} = \sigma\Big(\widehat{\theta}_n, \big\{1_{\{\tau_1 \le s\}}\big\}_{s \in [0,t]}, \dots, \big\{1_{\{\tau_n \le s\}}\big\}_{s \in [0,t]}\Big). \tag{30}$$

Suppose $\Psi_1(t) = P(\tau_1 \le t \mid \widehat{\theta}_n)$ and $\Psi_i(t) = P(\tau_i \le t \mid \tau_{i-1}, \dots, \tau_1, \widehat{\theta}_n)$ are regular continuous conditional distribution functions.

Then the compensator A_n of the process $n^{-1}\sum_{i=1}^n 1_{\{\tau_i \le t\}}$ with respect to $\{\mathcal{A}_{nt}\}_{t \in [0,\infty)}$ is continuous and satisfies (7), where $A_{n1}, \dots, A_{nn}$ are nonnegative nondecreasing stochastic processes defined by $A_{ni} = -\log(1 - \Psi_i(\cdot \wedge \tau_i))$. Moreover, $A_{n1}(\infty), \dots, A_{nn}(\infty)$ are independent standard exponential random variables.

COROLLARY 2. *Under the conditions of Lemma 4 there exists a standard Wiener process W_n such that*

$$P\Big(\sup_{t \in [0,\infty)} \big|M_n \circ A_n^{-1}(t \wedge A_n(\infty)) - W_n(t \wedge \mathcal{E}A_n(\infty))\big| > n^{-1/4}x\Big) \\ \le c_2 \exp\{-c_3 x\},$$

where c_2 and c_3 are universal constants.

Lemma 4 is directly applicable in the situation where $\{\mathcal{A}_{nt}\}_{t \in [0, \infty)}$ is the minimal filtration corresponding to the counting process N_n [see Section 18.2 in Liptser and Shiryayev (1978), pages 244–252]. In this case the random variable $\widehat{\theta}_n$ is degenerate, but there are also useful applications which involve a less trivial random variable $\widehat{\theta}_n$.

For instance, it may be the case that the compensator with respect to the minimal filtration involves an unknown parameter (say, θ) which for operational purposes has to be estimated. Khmaladze (1981, 1982) advocates using a different compensator, computed with respect to the filtration where $\mathcal{A}_{nt}$ equals the σ-algebra generated not only by the random variables $\{1_{\{\tau_1 \leq s\}}\}_{s \in [0,t]}, \dots, \{1_{\{\tau_n \leq s\}}\}_{s \in [0,t]}$, but also by the estimator of θ.

As an example, let $\tau_1, \dots, \tau_n$ be the first n order statistics of a random sample of size $n + 1$ from an exponential distribution with unknown mean θ, and let N_n be the counting process corresponding to these order statistics. We set $\widehat{\theta}_n$ equal to the sample mean, the maximum likelihood estimator of θ. Conditional on $\tau_{i-1}, \dots, \tau_1$ and $\widehat{\theta}_n$, the random variable τ_i has a Beta$(1, n - i + 1)$ distribution on the interval (τ_{i-1}, τ_i^u), where

$$\tau_i^u = \frac{(n+1)\widehat{\theta}_n - \sum_{j=1}^{i-1} \tau_j}{n - i + 2}. \tag{31}$$

With respect to the filtration defined by (30) we have (7) with

$$A_{ni} = \int_{\cdot \wedge \tau_{i-1}}^{\cdot \wedge \tau_i} \frac{n - i + 1}{\tau_i^u - s}\, ds, \tag{32}$$

where τ_0 is equal to zero. We have deliberately deleted the largest order statistic from the sample, since its distribution given $\tau_1, \dots, \tau_n$ and $\widehat{\theta}_n$ is degenerate.

By setting x equal to $\log n$, it follows from Corollary 2 that,

$$n^{1/4}(\log n)^{-1}\left|T(M_n \circ A_n^{-1}) - T(W_n)\right| = O_P(1), \tag{33}$$

for any Lipschitz functional T. Moreover, if T satisfies (18), then

$$\lim_{n \to \infty} (t_n)^{-2} \log P\big(T(M_n \circ A_n^{-1}) > t_n\big) = -a_I/2, \tag{34}$$

for all sequences $\{t_n\}_{n=1}^{\infty}$ such that $t_n \to \infty$, and $t_n = o(n^{1/4})$.

Because the stochastic processes $A_{n1}, \dots, A_{nn}$ as constructed in the proof of Lemma 4 are certainly not independent [$A_{ni}(t) = 0$ implies $A_{nj}(t) = 0$ for $i < j \leq n$], we cannot make use of the second part of Lemma 1 to obtain results for $T(M_n)$.

However, for some choices of the functional T, the statistics $T(M_n)$ and $T(M_n \circ A_n^{-1})$ coincide. In particular, this is true for the sublinear Lipschitz functionals T_R and T_S defined by (20). Thus, (34) and (33) remain valid if $T(M_n \circ A_n^{-1})$ is replaced by either $T_R(M_n)$ or $T_S(M_n)$.

In the situation considered here, it is possible to compute the exact compensator of the counting process N_n with respect to the filtration (30). However, in most other situations the exact compensator is not tractable. In Khmaladze (1981, 1982) a method is given to find an approximation K_n to the exact compensator A_n. In light of Corollary 2, the identity

$$n^{-1}N_n - K_n = (n^{-1}N_n - A_n) + (A_n - K_n) \tag{35}$$

shows that it is important to know the behavior of $A_n - K_n$ in order to obtain results for the process $n^{-1}N_n - K_n$. We shall not pursue this point further here.

5. Tools. In this section three exponential inequalities are given which are instrumental in proving the results of the previous sections. The first inequality [see Shorack and Wellner (1986), page 855] concerns the sum of independent random variables. Note that the random variables need neither to be bounded nor to have identical distributions.

INEQUALITY 1 (Bernstein). *Let $V_1, \dots, V_n$ be independent random variables having mean zero. Assume there exist positive constants $a_1, \dots, a_n$ and c such that*

$$\mathcal{E}|V_i|^r \le a_i r! c^{r-2}/2, \qquad i = 1, \dots, n,$$

for all $r \ge 2$. Then, for all $x > 0$,

$$P\left(\sum_{i=1}^n V_i > x\sqrt{n}\right) \le \exp\left\{-\frac{x^2/2}{n^{-1}\sum_{i=1}^n a_i + cx/\sqrt{n}}\right\}.$$

The next two inequalities are versions of Inequalities 14.5.3 and 14.5.5 of Shorack and Wellner [(1986), page 571] and concern the centered Poisson process π_n^c, introduced in Section 2. They involve a function ψ defined by

$$\psi(t) = \frac{2h(1+t)}{t^2},$$

where

$$h(x) = x(\log x - 1) + 1.$$

INEQUALITY 2. *For every $y > 0$ and $b > 0$,*

$$P\left(\sup_{0 \le t \le b} |\pi_n^c(t)| > y\sqrt{bn}\right) \le 2\exp\left\{-\frac{y^2}{2}\psi\left(\frac{y}{\sqrt{bn}}\right)\right\}.$$

INEQUALITY 3. *For every $y > 0$ and $0 < b < \frac{1}{2}$,*

$$P\left(\sup_{\substack{0 \le t_1 \le t_2 \le 1 \\ 0 \le t_2 - t_1 \le b}} |\pi_n^c(t_2) - \pi_n^c(t_1)| > y\sqrt{bn}\right) \le \frac{160}{b}\exp\left\{-\frac{y^2}{16}\psi\left(\frac{y}{\sqrt{bn}}\right)\right\}.$$

According to Proposition 11.1.1 of Shorack and Wellner [(1986), page 441] the function ψ possesses the following properties:

$$\psi(t) \text{ is decreasing for } t \geq -1 \text{ with } \psi(1) = 0.7726; \tag{36}$$

$$\psi(t) \geq \frac{1}{1+t/3} \quad \text{for } t \geq -1. \tag{37}$$

6. Proofs. In this section the proofs of Theorem 1 and Lemmas 1–4 are given.

PROOF OF THEOREM 1. By property (36) we have $\psi(n^{1/2-\beta}d_{n,x}) \geq \psi(1)$. Hence, applying Inequality 2 with $b = d_{n,x}$ and $y = n^{\beta}(d_{n,x})^{1/2}$ yields

$$P\left(\sup_{0 \leq t \leq d_{n,x}} \left|\pi_n^c(t)\right| > n^{\beta+1/2}d_{n,x}\right) \leq 2\exp\left\{-x\frac{\psi(1)}{2}\right\}. \tag{38}$$

A similar application of Inequality 3 yields, for $d_{n,x} < \frac{1}{2}$,

$$\begin{aligned} &P\left(\sup_{\substack{0 \leq t_1 \leq t_2 \leq 1 \\ 0 \leq t_2 - t_1 \leq d_{n,x}}} \left|\pi_n^c(t_2) - \pi_n^c(t_1)\right| > n^{\beta+1/2}d_{n,x}\right) \\ &\quad \leq \frac{160}{d_{n,x}} \exp\left\{-(21\log n + x)\frac{\psi(1)}{16}\right\} \\ &\quad \leq \frac{160n^{2\beta}}{\log n + x} \exp\left\{-\left(\log n + \frac{x}{21}\right)\right\} \\ &\quad \leq 160\exp\left\{-\frac{x}{21}\right\}. \end{aligned} \tag{39}$$

Moreover, for $\frac{1}{2} < d_{n,x} \leq n^{1/2-\beta}$, we have

$$\begin{aligned} &P\left(\sup_{\substack{0 \leq t_1 \leq t_2 \leq 1 \\ 0 \leq t_2 - t_1 \leq d_{n,x}}} \left|\pi_n^c(t_2) - \pi_n^c(t_1)\right| > n^{\beta+1/2}d_{n,x}\right) \\ &\quad \leq P\left(2\sup_{0 \leq t \leq 1} \left|\pi_n^c(t)\right| > n^{\beta+1/2}d_{n,x}\right) \\ &\quad \leq P\left(4\sup_{0 \leq t \leq d_{n,x}} \left|\pi_n^c(t)\right| > n^{\beta+1/2}d_{n,x}\right) \\ &\quad \leq 2\exp\left\{-x\frac{\psi(1)}{32}\right\} \end{aligned}$$

[use Inequality 2 with $b = d_{n,x}$ and $y = n^{\beta}(d_{n,x})^{1/2}/4$]. Since

$$
\begin{aligned}
&P\left(\sup_{t\in[0,\infty)}\left|M_n\circ A_n^{-1}\circ\xi_n(t)-n^{-1/2}\pi_n^c\circ\xi(t)\right|>2n^{\beta}d_{n,x}\right)\\
&\quad\le P\left(\sup_{\substack{0\le t_1\le t_2\le 1\\ 0\le t_2-t_1\le d_{n,x}}}\left|\pi_n^c(t_2)-\pi_n^c(t_1)\right|>n^{\beta+1/2}d_{n,x}\right)\\
&\qquad+P\left(\sup_{0\le t\le d_{n,x}}\left|\pi_n^c(t)\right|>n^{\beta+1/2}d_{n,x}\right)\\
&\qquad+P\left(\sup_{t\in[0,\infty)}\left|\xi_n(t)-\xi(t)\right|>d_{n,x}\right)
\end{aligned}
\tag{40}
$$

(the second term of the right-hand side of the latter inequality deals with possible excursions of ξ_n into the interval $[1, 1+d_{n,x}]$), it suffices to prove the existence of a Wiener process W_n such that

$$
P\left(\sup_{t\in[0,1]}\left|n^{-1/2}\pi_n^c(t)-W_n(t)\right|>c_1'n^{\beta}d_{n,x}\right)\le c_2'\exp\{-c_3'x\}.
\tag{41}
$$

Define independent standard Poisson random variables $U_1,\dots,U_n$ by $U_i = \pi_n(t_i)-\pi_n(t_{i-1})$, where $t_i = i/n$. Since

$$
\pi_n^c(t_i)=\sum_{j=1}^{i}(U_j-1),
\tag{42}
$$

the approximation theorem of Komlós, Major and Tusnády (1975) for partial sums yields the existence of a Wiener process W_n such that

$$
\begin{aligned}
&P\left(\max_{i=1,\dots,n}\left|n^{-1/2}\pi_n^c(t_i)-W_n(t_i)\right|>n^{-1/2}\left(c_1''\log n+x\right)\right)\\
&\quad\le c_2''\exp\{-c_3''x\}.
\end{aligned}
\tag{43}
$$

Now (41) follows from (43) since for each $i = 1,\dots,n$ we have, for every $x > 0$,

$$
\begin{aligned}
&P\left(\sup_{t_{i-1}\le t\le t_i}\left|\pi_n^c(t)-\pi_n^c(t_{i-1})\right|>x\right)\\
&\quad\le 2\exp\left\{-\frac{x^2/2}{1+x/3}\right\}\\
&\quad\le 2\exp\left\{-\frac{x}{3}\right\}
\end{aligned}
\tag{44}
$$

and

$$
\begin{aligned}
&P\left(\sup_{t_{i-1}\le t\le t_i}\left|W_n(t)-W_n(t_{i-1})\right|>n^{-1/2}x\right)\\
&\quad\le 4P(N(0,1)>x)\\
&\quad\le 2\exp\left\{-\frac{x^2}{2}\right\}.
\end{aligned}
\tag{45}
$$

The second line of (44) is again a consequence of Inequality 2. Now use $b = n^{-1}$ and (37).

Actually, we have just shown that (41) holds with $c_1' n^{\beta} d_{n,x}$ replaced by $n^{-1/2}$ $(c_1' \log n + x)$, which is of the same order as $n^{-\beta}(21 \log n + x)$, the lower bound to $n^{\beta} d_{n,x}$. Since (41) is combined with (40), there is no gain in using the sharper version. Hence, the critical part of this proof deals with the effects of random time change on the centered Poisson process.

To obtain the special result for $\xi_n(t) = t \wedge \tau_n$, we note that $\xi(t)$ should be of the form $t \wedge \tau$, where τ is a constant. Now we may write

$$
\begin{aligned}
&P\Big(\sup_{t \in [0,\infty)} \big| M_n \circ A_n^{-1} \circ \xi_n(t) - n^{-1/2} \pi_n^c \circ \xi(t) \big| > n^{\beta} d_{n,x} \Big) \\
&\qquad \le P\Big(\sup_{0 \le t \le d_{n,x}} \big| \pi_n^c(t) \big| > n^{\beta+1/2} d_{n,x} \Big) + P\big(|\tau_n - \tau| > d_{n,x} \big).
\end{aligned}
\tag{46}
$$

This completes the proof of Theorem 1. □

Proof of Lemma 1. Since $\mathcal{E}(A_{ni}(\infty))^r = r \int_0^{\infty} x^{r-1} P(A_{ni}(\infty) > x)\, dx \le r!$, the C_r-inequality bounds the rth central moment of $A_{ni}(\infty)$ by $8r!2^{r-2}/2$. Inequality 1 now yields

$$
P\Big(\big| A_n(\infty) - \mathcal{E} A_n(\infty) \big| > n^{-1/2} x \Big) \le \exp\Big\{ - \frac{x^2/2}{8 + 2n^{-1/2} x} \Big\}, \tag{47}
$$

by which (9) follows.

It remains to show (10). Fix $t \in [0,\infty)$ for the moment. Due to the monotonicity of A_{ni} we have $P(A_{ni}(t) > x) \le \exp\{-x\}$, and hence the rth central moment of $A_{ni}(t)$ is also bounded by $8r!2^{r-2}/2$. It follows from the independence of $A_{n1}, \dots, A_{nn}$ and Inequality 1 that

$$
P\Big(\big| A_n(t) - \mathcal{E} A_n(t) \big| > n^{-1/2} (\log n)^{1/2} \Big) < \tfrac{26}{27} \tag{48}
$$

for every $n > 1$.

This inequality for fixed t allows us to perform similar symmetrizations as in the first two steps of the proof given in Pollard [(1984), pages 14–15] of the Glivenko–Cantelli theorem, leading to

$$
\begin{aligned}
&P\Big(\sup_{t \in [0,\infty)} \big| A_n(t) - \mathcal{E} A_n(t) \big| > 4n^{-1/2} (\log n + x)^{3/2} \Big) \\
&\qquad \le 27 P\left(\sup_{t \in [0,\infty)} \left| \sum_{i=1}^{n} \sigma_i A_{ni}(t) \right| > (2n)^{1/2} (\log n + x)^{3/2} \right).
\end{aligned}
\tag{49}
$$

Here $\sigma_1, \dots, \sigma_n$ are independent Rademacher random variables [see Shorack and Wellner (1986), page 879], independent of $A_{n1}, \dots, A_{nn}$. Observe that, for every $\omega \in \Omega$,

$$\sup_{t \in [0,\infty)} \left| \sum_{i=1}^{n} \sigma_i A_{ni}(t) \right| = \max_{j=1,\dots,n} \left| \sum_{i=1}^{n} \sigma_i A_{ni}(\tau_j) \right|, \tag{50}$$

where $\tau_j = \inf\{t: A_{nj}(t) = A_{nj}(\infty)\}$.

Now introduce $\Omega_{n,x}$, the subset of Ω containing all ω for which $A_{ni}(\infty) \le \log n + x$ for all $i = 1,\dots,n$. Conditioning on $\Omega_{n,x}$ and the sample paths of $A_{n1},\dots,A_{nn}$, we obtain

$$\begin{aligned}
&P_{n,x}\left(\sup_{t \in [0,\infty)} \left|A_n(t) - \mathcal{E}A_n(t)\right| > 4n^{-1/2}(\log n + x)^{3/2} \right) \\
&\quad \le 27 P_{n,x}\left(\max_{j=1,\dots,n} \left| \sum_{i=1}^{n} \sigma_i A_{ni}(\tau_j) \right| > (2n)^{1/2}(\log n + x)^{3/2} \right) \\
&\quad \le 27 \sum_{j=1}^{n} P_{n,x}\left(\left| \sum_{i=1}^{n} \sigma_i \frac{A_{ni}(\tau_j)}{\log n + x} \right| > (2n)^{1/2}(\log n + x)^{1/2} \right) \\
&\quad \le 27n \exp\{-(\log n + x)\},
\end{aligned} \tag{51}$$

where $P_{n,x}$ denotes $P(\cdot \mid \Omega_{n,x}, A_{n1},\dots,A_{nn})$. The last line follows from Hoeffding's inequality [Hoeffding (1963), see also Shorack and Wellner (1986), page 855, or Pollard (1984), page 191] by using the fact that $A_{ni}(\tau_j)/(\log n + x)$ remains bounded by 1 on $\Omega_{n,x}$. Thus,

$$\begin{aligned}
&P\left(\sup_{t \in [0,\infty)} \left|A_n(t) - \mathcal{E}A_n(t)\right| > 4n^{-1/2}(\log n + x)^{3/2} \right) \\
&\quad \le 27 \exp\{-x\} + P(\Omega_{n,x}^C).
\end{aligned} \tag{52}$$

It is easily seen that $P(\Omega_{n,x}^C)$ does not exceed e^{-x}. □

PROOF OF LEMMA 2. Denote $\gamma/(2\kappa - 1)$ by ρ. Define the sequence $\{x_n\}_{n=1}^{\infty}$ by $x_n = n^{\rho} t_n$. Observe that $x_n = o(n^{2\rho})$ as n tends to ∞. Moreover, we have $\log n = o(x_n)$ and $(t_n)^2 = o(n^{\rho} t_n) = o(x_n)$, and thus

$$P\left(\left|T_n - \widetilde{T}_n\right| > \widetilde{c}_1 n^{-\gamma}(x_n)^{\kappa}\right) \le \widetilde{c}_2 \exp\{-\widetilde{c}_3(x_n - \log n)\} = o\left(\exp\left\{-\frac{a(t_n)^2}{2}\right\}\right).$$

The lemma follows by bounding $P(T_n > t_n)$ between

$$P(\widetilde{T}_n > t_n + \widetilde{c}_1 n^{-\gamma}(x_n)^{\kappa}) - P\left(\left|T_n - \widetilde{T}_n\right| > \widetilde{c}_1 n^{-\gamma}(x_n)^{\kappa}\right)$$

and

$$P(\widetilde{T}_n > t_n - \widetilde{c}_1 n^{-\gamma}(x_n)^{\kappa}) + P\left(\left|T_n - \widetilde{T}_n\right| > \widetilde{c}_1 n^{-\gamma}(x_n)^{\kappa}\right),$$

since $n^{-\gamma}(x_n)^{\kappa} = n^{\rho\kappa - \gamma}(t_n)^{\kappa - 1} \cdot t_n = o(n^{\rho(2\kappa - 1) - \gamma} \cdot t_n) = o(t_n)$. □

PROOF OF LEMMA 3. We have

$$P(A_{ni}(\infty) > x) = P(\Lambda_i(Z_i) > x) \le P(\Lambda_i(X_i) > x), \tag{53}$$

and $\Lambda_i(X_i)$ is a standard exponential random variable. □

PROOF OF LEMMA 4. It follows as in the proof of Theorem 18.2 in Liptser and Shiryayev [(1978), page 245] that (7) holds with A_{ni} as indicated. Now note that

$$\begin{aligned} P\big(A_{ni}(\infty) > x \mid \tau_{i-1}, \dots, \tau_1, \widehat{\theta}_n\big) &= P\Big(-\log\big(1 - \Psi_i(\tau_i)\big) > x \mid \tau_{i-1}, \dots, \tau_1, \widehat{\theta}_n\Big) \\ &= P\big(\Psi_i(\tau_i) > 1 - e^{-x} \mid \tau_{i-1}, \dots, \tau_1, \widehat{\theta}_n\big) \\ &= e^{-x}, \end{aligned}$$

which does not depend on $\tau_{i-1}, \dots, \tau_1$ or $\widehat{\theta}_n$. □

Acknowledgment. The author wishes to thank an Associate Editor for remarks that led to a considerable improvement in presentation.

REFERENCES

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

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

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

BARLOW, R. E. and PROSCHAN, F. (1969). A note on tests for monotone failure rate based on incomplete data. *Ann. Math. Statist.* **40** 595–600.

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

BRÉMAUD, P. (1981). *Point Processes and Queues, Martingale Dynamics.* Springer, New York.

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

BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256.

GILL, R. D. (1986). The total time on test plot and the cumulative time on test statistic for a counting process. *Ann. Statist.* **14** 1234–1239.

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

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

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. (In Russian. Translation in *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 F.I. *Z. Wahrsch. Verw. Gebiete* **34** 33–58.

KONING, A. J. (1992). Approximation of stochastic integrals with applications to goodness-of-fit tests. *Ann. Statist.* **20** 428–454.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes, II: Applications.* Springer, New York.

POLLAND, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

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