This paper gives a brief survey of the uses of stochastic calculus in survival analysis. The role played by martingale central limit theory in deriving asymptotic distributions of estimators and test statistics is described. The Nelson-Aalen estimator, Kaplan-Meier estimator, Cox’s proportional hazards model, Aalen’s additive risk model and a goodness-of-fit test for Cox’s model are discussed. Sketches of the proofs of the main results are included.
STOCHASTIC CALCULUS AND SURVIVAL ANALYSIS

by

Ian W. McKeague¹ and Klaus J. Utikal

FSU Technical Report No. M-790
USARO Technical Report No. D-102

June, 1988

Department of Statistics
The Florida State University
Tallahassee, FL 32306-3033

This is an invited paper for a special issue of Applied Mathematics and Computation devoted to "Theory and Applications of Stochastic Differential Equations (Ordinary and Partial)".

¹ Research supported by the Army Research Office under Grant DAAL03-86-K-0094.


Key words and phrases. Censored survival data, counting processes, martingale methods.
STOCHASTIC CALCULUS AND SURVIVAL ANALYSIS

by

Ian W. McKeague
and
Klaus J. Utikal

ABSTRACT

This paper gives a brief survey of the uses of stochastic calculus in survival analysis. The role played by martingale central limit theory in deriving asymptotic distributions of estimators and test statistics is described. The Nelson-Aalen estimator, Kaplan-Meier estimator, Cox's proportional hazards model, Aalen's additive risk model and a goodness-of-fit test for Cox's model are discussed. Sketches of the proofs of the main results are included.
1. Introduction

Beginning with Aalen's 1975 thesis from Berkeley, there has been a rapid increase in the use of stochastic calculus as a tool in the study of survival analysis. Aalen realized that multivariate counting processes provide a natural framework for the study of censored survival data and that a central role is played by martingales, predictable processes and stochastic integrals. The counting process approach has been successfully applied in the study of Nelson's cumulative hazard estimator (Aalen, 1978), the Kaplan-Meier product-limit estimator (Aalen and Johansen, 1978; Gill, 1980, 1983), nonparametric k-sample tests (Andersen et al., 1982) and Cox's proportional hazards regression model (Andersen and Gill, 1982), to name just a few examples.

The list of such applications is now quite long, and as new estimators, models and data structures are introduced, they too can often be treated under the umbrella of the counting process approach. Some of the benefits of this are: (1) we can bring to bear powerful results from the theory of stochastic processes; (2) more general censoring patterns can be allowed; (3) i.i.d. assumptions no longer play a central role; (4) straightforward, yet rigorous, proofs can be given. Although some background in stochastic processes is needed to appreciate these advantages, it turns out (as remarked by Arjas, 1985) that the σ-fields and martingales involved are far more concrete and practical than the traditional approach via elementary mathematics.

The purpose of the present paper is to outline these developments with special emphasis on some of the recent applications of Rebolledo's martingale central limit theorem to the study of asymptotic distributions of estimators and test statistics. We shall make no attempt to review the stochastic calculus relevant to survival analysis, but rather refer to the books of Liptser and Shiryayev (1978), Elliott (1982) and Kopp (1984) as necessary. Earlier survey articles and books containing material on the use of counting process theory in survival analysis have been written by Gill (1980), Jacobsen (1982), Davis (1983), Andersen and Borgan (1985), Shorack and Wellner (1986), Karr (1986) and Prakasa Rao (1987). We also mention the forthcoming book of Andersen, Borgan, Gill and Keiding (1988).

2. The Nelson-Aalen and Kaplan-Meier estimators

2.1. Cumulative hazard and survival functions

Let \( T \), representing the survival time of an individual, be a positive random variable with distribution function \( F \) having a density \( f \). The hazard function (or failure rate function) of \( T \) is defined by \( \lambda(t) = f(t)/S(t) \) for \( t \) such that \( S(t) = 1 - F(t) > 0 \). Here \( S \) is called the survival function. Since

\[
\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}
\]

we may interpret \( \lambda(t) \Delta t \) as the probability of failure in the time interval \( t \) to \( t + \Delta t \) given survival
up to time $t$. It is easily seen that the hazard function determines the distribution. For

$$\lambda(t) = -\frac{d}{dt} \log S(t),$$

so that, in terms of the cumulative hazard function $A(t) = \int_0^t \lambda(s) \, ds$, we have $S(t) = e^{-A(t)}$.

When $F$ does not have a density, the hazard function cannot be defined, but the cumulative hazard function is defined by

$$A(t) = \int_{(0,t]} \frac{dF(u)}{S(u-)}$$

for $t$ such that $S(t-):= \lim_{u\downarrow t} S(u) > 0$. The survival function can be represented in terms of the cumulative hazard function by $S = \mathcal{E}(-A)$, where $\mathcal{E}$ is the Doléans-Dade exponential defined by the following result.

**Theorem 2.1.** (A special case of Doléans-Dade (1970)) Let $X$ be a right-continuous function of bounded variation with $X_0 = 0$. Then the equation

$$Z_t = 1 + \int_{(0,t]} Z_{s-} \, dX_s$$

has a unique solution which is bounded on finite intervals. The solution, denoted $\mathcal{E}(X)_t$, is given by $Z_t$ equal to

$$\mathcal{E}(X)_t = e^{X_t} \prod_{s \leq t} (1 + \Delta X_s),$$

where $X_t = X_t - \sum_{s \leq t} \Delta X_s; \Delta X_s = X_s - X_{s-}$.

A proof of this result can be found in Liptser and Shiryayev (1978, pp.255-256). As an immediate consequence (see Wellner, 1985) we obtain the formula

$$S = \mathcal{E}(-A),$$

since $S$ satisfies the equation

$$S(t) = 1 + \int_{(0,t]} S(u-) \, d(-A(u)).$$

An alternative way of representing $S$ in terms of $A$ is to use product integration. Let $X$ be a finite Borel measure on the positive real line and denote its distribution function by the same symbol: $X(t) = X((0,t])$. The product integral of $X$ is defined by

$$\prod_{(0,t]} (1 + dX) = \lim_{\max|t_i - t_{i-1}| \to 0} \prod_i (1 + X([t_{i-1}, t_i])),$$

where $0 = t_0 < t_1 < ... < t_n = t$ is a partition of $(0,t]$. It can be shown (see Gill and Johansen, 1988) that

$$S(t) = \prod_{(0,t]} (1 - dA).$$
2.2. Random censorship models and the counting process formulation

Censoring is a general phenomenon which affects many types of data. In survival data it typically takes the form of random right or left censorship. For example, consider a study for determining the age \( T \) at which a certain chronic disease or other permanent condition appears in an individual. Right censorship occurs if the individual dies or leaves the study before the disease appears. However, the disease may have already appeared before the individual entered the study, and this results in left censoring. In each case, the exact value of \( T \) cannot be determined, but some useful information is still available. One of the most elegant features of the counting process formulation is that it unifies these and more general censoring schemes under the notion of "predictable censoring". Estimators of the survival function based on right-censored data (Kaplan and Meier, 1958), left-censored data (Woodroofe, 1985) or doubly-censored data (Chang and Yang, 1987) can then be studied in a unified way.

Define the basic counting process \( N_t^* = I(T < t) \). In the absence of censoring, \( N_t^* \) is the only counting process we need to consider. To introduce censoring into the picture, let \((C_t, t > 0)\) be a \((0,1)\)-valued stochastic process, called the censoring process, which is understood to indicate censorship at time \( t \) if \( C_t = 0 \). The observed counting process is given by

\[
N_t = \int_0^t C_s dN_s^* = \begin{cases} 1 & \text{if } T \leq t \text{ and } C_T = 1 \\ 0 & \text{otherwise} \end{cases}
\]

Next, define \( \mathcal{F}_t^T = \sigma(N_s^*, s \leq t) \), the \( \sigma \)-field generated by \( N^* \) up to time \( t \). Let \( (\mathcal{G}_t) \) be a right-continuous filtration such that \( \mathcal{G}_t \) is independent of \( T \) for all \( t \) and write \( \mathcal{F}_t = \mathcal{F}_t^T \vee \mathcal{G}_t \), the \( \sigma \)-field generated by \( \mathcal{F}_t^T \) and \( \mathcal{G}_t \). Also define the processes \( R_t = I(T \geq t) \) and \( Y_t = C_t R_t \), so that \( Y_t \) is the indicator that the individual is observed to be at risk at time \( t \). The following result is crucial.

**Theorem 2.2.** Suppose that the censoring process \((C_t)\) is predictable with respect to the filtration \( (\mathcal{F}_t) \). Then the process

\[
M_t = N_t - \int_0^t Y_s \, d\Lambda_s
\]

is an \( \mathcal{F}_t \)-martingale.

In the language of counting processes this result is saying that the counting process \( N_t \) has compensator \( \Lambda_t = \int_0^t Y_s \, d\Lambda_s \) with respect to the filtration \( (\mathcal{F}_t) \), see Liptser and Shiryaev (1978, p. 239). If \( T \) has a hazard function \( \lambda(t) \), then the result implies that \( N_t \) has intensity process \( Y_t \lambda(t) \) and we may write (2.3) in the form of a stochastic differential equation:

\[
dN_t = Y_t \lambda(t) \, dt + dM_t.
\]

The result is intuitively reasonable in view of the interpretation of \( Y_t \lambda(t) \, dt \) as the probability of observing a failure in the time interval \( t \) to \( t + dt \) given survival up to time \( t \).

The censoring process \((C_t)\) is left-continuous in most applications, so in order to show that it is predictable, it suffices to show that it is \((\mathcal{F}_t)\)-adapted. In particular, the usual random right or
left censorship schemes are obtained by taking \( C_t = I(L < t \leq C) \), where \( L \) and \( C \) are the left and right “censoring times” respectively, assuming that \((L, C)\) is independent of \( T \), and setting \( \mathcal{G}_t = \mathcal{G}_0 = \sigma(L, C) \). Note, however, that the definition of predictable censoring is much more general than this.

**Proof of Theorem 2.2.** The main step in the proof is to show that \( M_t^* = N_t^* - \int_0^t R_u dA_u \) is an \( \mathcal{F}_t^T \)-martingale. Then, since \( \mathcal{G}_t \) is independent of \( \mathcal{F}_t^T \), it follows from a standard result on conditional expectation (see Chung, 1974, p.308) that \( M_t^* \) is an \( \mathcal{F}_t \)-martingale. Consequently, since \( M_t = \int_0^t C_u dM_t^* \) can be interpreted as a stochastic integral (see Kopp, 1984, p.149), \( M_t \) is an \( \mathcal{F}_t \)-martingale. Now to show that \( M_t^* \) is a martingale, let \( u < t \) and note that

\[
E(M_t^* - M_u^* | \mathcal{F}_u^T) = E(I(T > u)(M_t^* - M_u^*) | \mathcal{F}_u^T) = I(T > u) E(M_t^* - M_u^* | T > u)
\]

since \( A \cap \{T > u\} \) is either the empty set or \( \{T > u\} \) for any \( A \in \mathcal{F}_u^T \). Also

\[
E(N_t^* - N_u^* | T > u) = \frac{F(t) - F(u)}{S(u)}
\]

and with \( A_t = \int_0^t R_u dA_u \),

\[
E(A_t - A_u | T > u) = E(A_t - A_u | T > t) P(T > t | T > u)
\]

\[
+ E(A_t - A_u | u < T \leq t) P(T \leq t | T > u)
\]

\[
= (\Lambda(t) - \Lambda(u)) \frac{S(t)}{S(u)} + \int_{[u,t]} \frac{\Lambda(u) - \Lambda(v)}{F(t) - F(u)} dF(v) \frac{F(t) - F(u)}{S(u)}
\]

\[
= \frac{1}{S(u)} \int_{[u,t]} S(u-) d\Lambda(v)
\]

\[
= \frac{F(t) - F(u)}{S(u)} = E(N_t^* - N_u^* | T > u),
\]

where we have used the integration by parts formula for Stieltjes integrals (see Liptser and Shiryayev, 1978, p.253). This completes the proof.

The basic idea in the above proof is that the martingale property of \((M_t^*, \mathcal{F}_t^T)\) is preserved when independent events are added to \( \mathcal{F}_t^T \). It is natural to ask “How much can we enlarge \( \mathcal{F}_t^T \) while preserving the martingale property of \( M_t^* \)?” We know of no general answer to this question, although some recent work of Jacobsen (1986, 1988) dealing with right-censoring is of interest in this regard.

To extend the counting process formulation to \( n \) individuals with corresponding survival times \( T_1, ..., T_n \), introduce processes \( N_i, Y_i, M_i \), \( i = 1, ..., n \) and filtrations \((\mathcal{F}_t)\), \( i = 1, ..., n \) having the same structure as \( N, Y, M \) and \((\mathcal{F})\). It is convenient to view the martingales \( M_i \), \( i = 1, ..., n \) with respect to the same filtration. If \( \mathcal{F}_{1t}, ..., \mathcal{F}_{nt} \) are independent \( \sigma \)-fields for all \( t \), then \( M_{1t}, ..., M_{nt} \) are martingales (in fact orthogonal martingales) with respect to the filtration \( \mathcal{F}_t^{(n)} = \mathcal{F}_{1t} \vee \cdots \vee \mathcal{F}_{nt} \). For some applications, however, it may be too restrictive to assume that the \( \mathcal{F}_{it} \) are independent.
The general counting process formulation does not require any such assumption; it only requires that \(M_1, ..., M_n\) are martingales with respect to some filtration \(\mathcal{F}_t^{(n)}\).

Aalen's (1978) multiplicative intensity model can now be formulated as follows. Let \(N(t) = (N_1(t), ..., N_n(t))^t\), \(t \in [0, 1]\) be a multivariate counting process with respect to a right-continuous filtration \(\mathcal{F}_t^{(n)}\), i.e. \(N\) is adapted to the filtration and has components \(N_i\) which are right-continuous step functions, zero at time zero, with jumps of size +1 such that no two components jump simultaneously. Suppose that the counting process \(N_i\) has intensity

\[
\lambda_i(t) = Y_i(t) \lambda(t),
\]

where \(Y_i(t)\) is a predictable \(\{0, 1\}\)-valued process and \(\lambda\) is a fixed hazard function. By Davis (1983, Proposition 4.2)

\[
M_i(t) = N_i(t) - \int_0^t \lambda_i(s) \, ds, \quad i = 1, ..., n
\]

are orthogonal local square integrable martingales on \([0, 1]\) with predictable variation process

\[
\langle M_i, M_i \rangle_t = \int_0^t \lambda_i(s) \, ds.
\]

If \(E N_i(1) < \infty\) then \(M_i\) is in fact a square integrable martingale on \([0, 1]\). As Aalen does, we shall assume throughout that \(M_i, i = 1, ..., n\) are square integrable martingales.

2.9. The Nelson-Aalen estimator

An estimator for the cumulative hazard function \(\Lambda\) was introduced by Nelson (1969) in the case of right-censored survival data, and extended by Aalen (1975, 1978) to his multiplicative intensity model. The so called Nelson-Aalen estimator is defined by

\[
\hat{\Lambda}(t) = \int_0^t \frac{dN^{(n)}(s)}{Y^{(n)}(s)},
\]

where \(N^{(n)} = \sum_{i=1}^n N_i\), \(Y^{(n)} = \sum_{i=1}^n Y_i\) and \(1/0 \equiv 0\). A motivation for this estimator is provided by formally solving the stochastic differential equation

\[
dN^{(n)}_t = Y^{(n)}_t \, d\Lambda_t + dM^{(n)}_t
\]

(where \(M^{(n)} = \sum_{i=1}^n M_i\)) for \(\Lambda_t\) and ignoring the "noise" term \(\int_0^t \frac{dM^{(n)}(s)}{Y^{(n)}(s)}\) (which is a martingale).

The asymptotic distribution of \(\hat{\Lambda}\) can be derived under the following asymptotic stability condition:

(AS) There exists a function \(\rho\), bounded away from zero on \([0, 1]\), such that \(\bar{Y}^{(n)}(t) \equiv \frac{1}{n} Y^{(n)}(t)\) satisfies

\[
\sup_{t \in [0,1]} |\bar{Y}^{(n)}(t) - \rho(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.
\]
Note that in the i.i.d. case, in which \( Y_i, i = 1, \ldots, n \) are i.i.d. replicates of one another and \( Y_i \) has left-continuous paths with right hand limits, the asymptotic stability condition (AS) can be checked using Ranga Rao's (1963) law of large numbers. Let \( D[0,1] \) denote the space of functions on \([0,1]\) which are right-continuous on \([0,1]\) with left limits on \((0,1]\), and equip it with the Skorohod topology (see Billingsley, 1968, p.111). Convergence in distribution will be denoted \( \mathcal{D} \).

**Theorem 2.3.** (Aalen, 1978) Suppose that the asymptotic stability condition (AS) holds. Then, under Aalen's multiplicative intensity model, \( \sqrt{n}(\hat{\Lambda} - \Lambda) \xrightarrow{\mathcal{D}} m \) in \( D[0,1] \) as \( n \to \infty \), where \( m \) is a continuous Gaussian martingale with covariance function

\[
\text{Cov}(m_s, m_t) = \int_0^{t \wedge t} \frac{\lambda(u)}{\rho(u)} \, du.
\]

**Proof.** Using (2.7) we have

\[
\sqrt{n}(\hat{\Lambda}(t) - \Lambda(t)) = \hat{M}_t^{(n)} - R_t^{(n)},
\]

where

\[
\hat{M}_t^{(n)} = \frac{1}{\sqrt{n}} \int_0^t \frac{dM_t^{(n)}(s)}{Y_t^{(n)}(s)}, \quad R_t^{(n)} = \sqrt{n} \int_0^t I(Y_t^{(n)}(s) = 0) \, dA_t.
\]

Since the condition (AS) implies that

\[
\sup_{t \in [0,1]} |R_t^{(n)}| \xrightarrow{\mathcal{L}} 0,
\]

(2.9) to complete the proof it suffices to show that \( \hat{M}_t^{(n)} \xrightarrow{\mathcal{D}} m \) in \( D[0,1] \). Note that \( \hat{M}_t^{(n)} \) is a square integrable \( \mathcal{F}_t^{(n)} \)-martingale. Now apply the version of Rebolledo's (1980) martingale central limit theorem stated in Andersen and Gill (1982) with \( p = 1 \) and \( H_t^{(n)}(t) = n^{-\frac{1}{2}} (Y_t^{(n)}(t))^{-1} \). By (2.5), (2.6) and (AS) we have

\[
\langle \hat{M}^{(n)}, \hat{M}^{(n)} \rangle_t = \int_0^t \frac{\lambda(s)}{Y_t^{(n)}(s)} \, ds \xrightarrow{\mathcal{D}} \int_0^t \frac{\lambda(s)}{\rho(s)} \, ds.
\]

(2.10) The Lindeberg condition, here given by

\[
\frac{1}{n} \sum_{i=1}^n \int_0^1 \frac{1}{(Y_t^{(n)}(s))^2} \lambda_i(s) I\left(\left|\frac{1}{Y_t^{(n)}(s)}\right| > c \sqrt{n}\right) \, ds \xrightarrow{\mathcal{L}} 0
\]

for all \( c > 0 \), follows from (AS). This completes the proof.

It is possible to use Theorem 2.3 to construct confidence bands for \( \Lambda \), see Andersen and Borgan (1985, p.114). In many applications it is of interest to estimate the hazard function \( \lambda \) itself. It is possible to develop an asymptotic distribution theory for pointwise estimators of \( \lambda(t) \), \( \hat{\lambda}(t) \) say, using Rebolledo's martingale central limit theorem, much as in the proof of Theorem 2.3. This has been done for kernel estimators (Ramlau-Hansen, 1983), spline sieve estimators (Karr,
grouped data based estimators (Borgan and Ramlau-Hansen, 1985; McKeague, 1988b) and penalized maximum likelihood estimators (Antoniadis, 1987). Integrating any one of these estimators provides another estimator \( \int_0^t \hat{\Lambda}(s) \, ds \) of \( \Lambda \), which (not surprisingly) turns out to have the same asymptotic distribution as the Nelson-Aalen estimator.

2.4. The Kaplan-Meier estimator

In view of (2.1) and (2.2) it is reasonable to estimate the survival function \( S \) by the Doléans-Dade exponential or product integral of \(-\hat{\Lambda}\), where \( \hat{\Lambda} \) is the Nelson-Aalen estimator. Define

\[
\hat{S}(t) = \mathcal{E}(-\hat{\Lambda})_t = \prod_{(0,t]} (1 - d\hat{\Lambda}).
\]

Since the continuous part of \( \hat{\Lambda} \) is zero, \( \hat{S} \) reduces to the so called "product-limit" estimator

\[
\hat{S}(t) = \prod_{s \leq t} \left(1 - \frac{\Delta N_j^{(n)}(s)}{Y_j^{(n)}}\right),
\]

which was originally introduced by Kaplan and Meier (1958) in the case of right-censored survival data.

Breslow and Crowley (1974) gave the first proof of weak convergence of the Kaplan-Meier estimator; Gill (1980, 1983) gave a proof based on martingale methods. We shall present Gill's proof in the context of the multiplicative intensity model. In the classical i.i.d. random censorship model other proofs are possible. Gill and Johansen (1988) recently gave a proof using Hadamard differentiability and a functional version of the delta method. That approach works for general (not necessarily continuous) survival functions and can be used to study the asymptotic behavior of the bootstrapped Kaplan-Meier estimator; see Gill (1987). David Pollard (1988) has informed us that a proof via the theory of empirical processes is also possible. We refer the reader to Akritas (1986), Horváth and Yandell (1987) and Lo and Singh (1986) for results on the bootstrapped Kaplan-Meier estimator.

**Theorem 2.4.** Suppose that the asymptotic stability condition \((AS)\) holds. Then, under Aalen's multiplicative intensity model, \( \sqrt{n}(\hat{S} - S) \rightarrow S(\cdot) \, m(\cdot) \) in \( D[0,1] \) as \( n \to \infty \), where \( m \) is the Gaussian martingale of Theorem 2.3.

**Proof.** First note that \( \mathcal{E}(\Lambda - \hat{\Lambda})_t / \mathcal{E}(\Lambda)_t = \hat{S}_t / S_t \), so, by Theorem 2.1,

\[
\frac{\hat{S}_t}{S_t} = 1 + \int_0^t \frac{\hat{S}_u - S_u}{S_u} d(\Lambda - \hat{\Lambda})(u).
\]

Thus, using (2.8), we obtain

\[
\sqrt{n}(\hat{S}_t - S_t) = -S_t \int_0^t \frac{\hat{S}_u - S_u}{S_u} d\hat{M}_u^{(n)} + R_t^{(n)}, \tag{2.11}
\]
where $R_4(n)$ as a remainder term (different from the original $R_4(n)$) satisfying (2.9). To complete the proof, it suffices to show that $\hat{m}(n) \rightarrow m$ in $D[0,1]$, where

$$\hat{m}_t(n) = \int_0^t \frac{\hat{S}_u}{S_u} d\hat{M}_u(n).$$

Now $\hat{m}(n)$ is a square integrable martingale with predictable variation process

$$\langle \hat{m}(n), \hat{m}(n) \rangle_t = \int_0^t \left[ \frac{\hat{S}_u}{S_u} \right]^2 \lambda(u) \frac{\varphi(n)(u)}{\bar{F}(n)(u)} du.$$

Thus, by (AS), we have $\langle \hat{m}(n), \hat{m}(n) \rangle_1 = O_P(1)$. Using Lenglart's (1977) inequality and (2.11) it follows that $\sup_{t \in [0,1]} |\hat{S}_t - S_t|^P \rightarrow 0$. Hence

$$\langle \hat{m}(n), \hat{m}(n) \rangle_t \rightarrow \int_0^t \frac{\lambda(u)}{\rho(u)} du = (m, m)_t.$$}

The Lindeberg condition for $\hat{m}(n)$ is checked in the same way it was checked for $\hat{M}(n)$. The result follows by Rebolledo’s martingale central limit theorem.

It is natural to ask whether the above results have any extension to two-dimensional survival times, $T = (T_1, T_2)$ say. Data of that kind can arise, for example, in a study of the ages $T_1, T_2$ at which two different chronic diseases appear in an individual. Unfortunately, many of the techniques that are useful in the univariate case are no longer applicable in the bivariate case. In particular, a two-parameter martingale central limit theorem is not available. The best results that are currently available are all for the i.i.d. case with right censoring and rely on classical methods; see Tsai, Leurgans and Crowley (1986) for instance.

3. Regression models for survival data

In most applications of survival analysis it is important to consider the effects that covariates may have upon the survival times of individuals in the study. This can be done by using a regression model for the conditional hazard function $\lambda(t, z) = \lambda(t|x)$ of the survival time of an individual who has a covariate vector $z = (z_1, \ldots, z_p)'$, say, at time $t$. The well known proportional hazards model of Cox (1972) has been the most popular model, but in recent years other models have begun to be considered. We list Cox’s model and various other alternative nonparametric and semiparametric models with which we are familiar as follows.

(1) Cox’s (1972) proportional hazards model:

$$\lambda(t, z) = \lambda_0(t) e^{\beta^T z},$$
where $\lambda_0$ is an unknown baseline hazard function and $\beta_0$ is a vector of $p$ unknown parameters.

(2) Aalen's (1980) additive risk model:

$$\lambda(t, z) = \sum_{j=1}^{p} \alpha_j(t) z_j,$$

where $\alpha_1, \ldots, \alpha_p$ are unknown functions.

(3) The general nonparametric model. Beran (1981) considered

$$\lambda(t, z)$$

is arbitrary.

(4) Variations on Cox's model. The general proportional hazards model

$$\lambda(t, z) = \lambda_0(t) r(z),$$

where $\lambda_0$ is an unknown baseline hazard function and $r$ is an unknown "relative risk" function was proposed by Thomas (1983). Hastie and Tibshirani (1987) suggested the generalized additive model $r(z) = \sum_{j=1}^{p} r_j(z_j)$, where $r_1, \ldots, r_p$ are unknown functions. Prentice and Self (1983) take $r(z) = r_0(\beta_0 z)$, where $r_0$ is known and $\beta_0$ is a vector of $p$ unknown parameters. Zucker (1986) and Zucker and Karr (1987) generalized Cox's model by allowing $\beta_0$ to be time-dependent.

Since the papers of Andersen and Gill (1982) and Aalen (1980), which developed asymptotic theory for the models (1) and (2), martingale methods have been used to obtain asymptotic theory for most of these models. In this section we review some of that work. Throughout we shall use the following counting process framework, extending the multiplicative intensity model to allow for covariates. Suppose that $N(t) = (N_1(t), \ldots, N_n(t))'$ is a multivariate counting process with respect to a right-continuous filtration ($\mathcal{F}_t^{(n)}$). The counting process $N_i$, which records events in the life of the $i$th individual, is assumed to have intensity process $\lambda_i(t) = Y_i(t) \lambda(t, Z_i(t))$, where $Y_i(t)$ is a predictable $\{0, 1\}$-valued process as before, and $Z_i(t) = (Z_{i1}(t), \ldots, Z_{ip}(t))'$ is a $p$-vector of predictable covariate processes. The martingales $M_1, \ldots, M_n$ are again defined by (2.5). For simplicity, we shall only consider the i.i.d. case in which $(N_i, Y_i, Z_i)$, $i = 1, \ldots, n$ are i.i.d. replicates of $(N, Y, Z)$. Also, assume that $Y$ and $Z$ are left-continuous with right hand limits and the covariate processes are bounded.

3.1. Cox's proportional hazards model

In this section we shall briefly sketch the main results of Andersen and Gill (1982). We refer to the review paper of Davis (1983) for an informal discussion of these results and the motivation behind the estimators. For simplicity of presentation, we shall assume that the covariates are scalar valued ($p = 1$).
Cox (1972, 1975) proposed that inference for $\beta_0$ in (1) be based on the partial likelihood function

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta Z_i(T_i)}}{\sum_{j \in R_i} e^{\beta Z_j(T_j)}} \right\}^{\delta_i},$$

where $\delta_i$ and $T_i$ are the indicator of noncensorship and the survival time for the $i$th individual respectively, and $R_i$ is the "risk set" consisting of all individuals which are observed to be at risk at time $T_i$. Let $\hat{\beta}$ be the value that maximizes $L(\beta)$. In terms of the underlying counting processes, the estimate $\hat{\beta}$ is the unique solution to $\frac{\partial}{\partial \beta} \log L(\beta) = U(\beta, 1) = 0$, where

$$U(\beta, t) = \sum_{i=1}^{n} \int_{0}^{t} \left\{ Z_i(u) - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right\} dN_i(u),$$

$$S^{(j)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Z_i(t)^j Y_i(t) e^{\beta Z_i(t)},$$

for $j = 0, 1, 2$, where $0^0 = 1$. The following theorem gives the asymptotic distribution of $\hat{\beta}$. Define $s^{(j)}(\beta, t) = E S^{(j)}(\beta, t), e = s^{(1)}/s^{(0)}, v = s^{(2)}/s^{(0)} - e^2$ and $\Sigma = \int_{0}^{1} v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt$.

**Theorem 3.1.** Suppose that $\lambda_0$ is integrable over $[0, 1]$, $s^{(0)}(\cdot, \cdot)$ is bounded away from 0 in a neighborhood of $\beta_0$, and $\Sigma > 0$. Then $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \Sigma^{-1})$.

**Proof.** (Sketch) By the mean value theorem

$$U(\hat{\beta}, 1) - U(\beta_0, 1) = - I(\beta^\ast, 1)(\hat{\beta} - \beta_0),$$

where $\beta^\ast$ lies between $\beta_0$ and $\hat{\beta}$, and

$$I(\beta, t) = - \frac{\partial}{\partial \beta} U(\beta, t) = \int_{0}^{t} \left\{ \frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - \left( \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right)^2 \right\} dN^{(n)}(u),$$

where $N^{(n)} = \sum_{i=1}^{n} N_i$. But $U(\hat{\beta}, 1) = 0$, so from (3.3) we obtain

$$\sqrt{n}(\hat{\beta} - \beta_0) = \frac{n^{-\frac{1}{2}} U(\beta_0, 1)}{n^{-1} I(\beta^\ast, 1)}. \quad (3.4)$$

The key step in the proof is to see that $U(\beta_0, \cdot)$ is a martingale:

$$U(\beta_0, t) = \sum_{i=1}^{n} \int_{0}^{t} \left\{ Z_i(u) - \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} \right\} dM_i(u).$$

Let $m_1$ be a continuous Gaussian martingale with variation process

$$\langle m_1 \rangle_t = \int_{0}^{t} v(\beta_0, u) s^{(0)}(\beta_0, u) \lambda_0(u) du. \quad (3.5)$$
Then
\[ \langle n^{-1/2} U(\beta_0, \cdot) \rangle_t = \int_0^t \left\{ S^{(2)}(\beta_0, u) - \frac{(S^{(1)}(\beta_0, u))^2}{S^{(0)}(\beta_0, u)} \right\} \lambda_0(u) \, du \mathcal{P}(m_1)_t, \]
so by Rebolledo's martingale central limit theorem \( n^{-1/2} U(\beta_0, \cdot) \overset{D}{\rightarrow} m_1 \) in \( D[0,1] \). Consequently, \( n^{-1/2} U(\beta_0, 1) \overset{D}{\rightarrow} N(0, \Sigma) \) and from (3.4), to complete the proof it suffices to show that \( n^{-1} I(\beta^*, 1) \overset{P}{\rightarrow} \Sigma \). This is done in Andersen and Gill (1982, p.1108), but it is to be expected because they show that \( \beta^* \beta_0 \), so \( \beta^* \mathcal{P} \beta_0 \), and we can write \( n^{-1} I(\beta_0, 1) \) in the form
\[ \int_0^1 \left\{ S^{(2)}(\beta_0, u) - \frac{(S^{(1)}(\beta_0, u))^2}{S^{(0)}(\beta_0, u)} \right\} \lambda_0(u) \, du + \frac{1}{n} \int_0^1 \left\{ \frac{S^{(2)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} - \left( \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} \right)^2 \right\} dM^{(n)}(u), \]
where \( M^{(n)} = \sum_{i=1}^n M_i \). The first term above tends in probability to \( \Sigma \) and the second term tends in probability to zero, by Lenglart's inequality.

Breslow (1972, 1974) suggested that the cumulative baseline hazard function \( \Lambda_0(t) = \int_0^t \lambda_0(s) \, ds \) could be estimated by a piecewise linear approximation to the Nelson-Aalen type estimator
\[ \hat{\Lambda}(t) = \int_0^t \frac{dN^{(n)}(u)}{nS^{(0)}(\beta, u)}. \]

**Theorem 3.2.** Under the conditions of Theorem 3.1, \( \sqrt{n}(\hat{\Lambda} - \Lambda_0) \overset{D}{\rightarrow} m_0(\cdot) + \psi(\cdot) m_1(1) \) in \( D[0,1] \), where \( m_0 \) and \( m_1 \) are independent zero mean Gaussian martingales, \( m_1(t) \) is defined by (3.5),
\[ \langle m_0 \rangle_t = \int_0^t \frac{\lambda_0(u) \, du}{S^{(0)}(\beta_0, u)} \quad \text{and} \quad \psi(t) = \Sigma^{-1} \int_0^t e(\beta_0, u) \lambda_0(u) \, du. \]

### 3.2. Aalen's additive risk model

In some applications, additive risk models are more appropriate than proportional hazards models. However, although parametric additive risk models have been used in survival analysis (especially in epidemiology) for many years (see the references in Breslow, 1986; Muirhead and Darby, 1987), the nonparametric additive risk model (2) has only been studied recently (Aalen, 1980; McKeague, 1986, 1988a, 1988b; Huffer and McKeague, 1988).

Let \( \alpha = (\alpha_1, \ldots, \alpha_p)' \) and denote \( Y_{ij}(t) = Y_i(t) Z_{ij}(t) \), \( A(t) = \int_{t_0}^t \alpha(s) \, ds \) for fixed \( t_0, 0 \leq t_0 \leq 1 \). Aalen (1980) proposed estimators \( \hat{A} \) of \( A \) of the form \( \hat{A}(t) = \int_{t_0}^t Y^-(s) \, dN(s) \), where \( Y^-(s) \) is a predictable generalized inverse of the \( n \times p \) matrix \( Y(s) = (Y_{ij}(s)) \). In the case \( p = 1 \), with \( (Y^-(s))_{1j} = (\sum_{k=1}^n Y_{k1}(s))^{-1} \), \( i = 1, \ldots, n \), \( \hat{A} \) is the Nelson-Aalen estimator. For \( p > 1 \), Aalen suggested using \( Y^-(s) = (Y'(s) Y(s))^{-1} Y'(s) \), where here and in the sequel, for any square matrix (or scalar) \( D \), \( D^{-1} \) denotes the inverse of \( D \) if \( D \) is invertible, the zero matrix otherwise. Aalen observed that this choice of \( Y^- \) can be motivated by a formal least squares principle and that
the resulting estimator $\hat{A}(t) = \int_{t_0}^t (Y'(s)Y(s))^{-1}Y'(s)\,dN(s)$, referred to as Aalen's least squares estimator, probably gives reasonable but not optimal estimates of $A$. Huffer and McKeague (1988) proposed using the following generalized inverse of $Y(s)$:

$$Y^{-}(s) = (Y'(s)\hat{W}(s)Y(s))^{-1}Y'(s)\hat{W}(s),$$

where $\hat{W}(t)$ is the $n \times n$ diagonal matrix with $i$th diagonal entry $\hat{W}_i(t)$, and

$$\hat{\lambda}_i(t) = \sum_{j=1}^p \hat{\alpha}_j(t) Y_{ij}(t),$$

where $\hat{\alpha}_j$ is a predictable estimator of $\alpha_j$. The estimator $\hat{\lambda}_i$ is taken to be the $i$th component of the smoothed least squares estimator

$$\hat{\alpha}(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) d\hat{A}(s),$$

where $K$ is a left-continuous bounded kernel function having integral 1, support $[0, 1]$ and $b_n > 0$ is a bandwidth parameter. The choice of generalized inverse (3.6) defines the so-called weighted least squares estimator

$$\hat{A}(t) = \int_{t_0}^t (Y'(s)\hat{W}(s)Y(s))^{-1}Y'(s)\hat{W}(s)\,dN(s).$$

In the case of a single covariate the weighted least squares estimator coincides with the Nelson-Aalen estimator.

A heuristic explanation for the choice of weight matrix $\hat{W}(t)$ is as follows. By conditioning on the past $\mathcal{F}^{(n)}(t)$ we may interpret the stochastic differential equation $dN(t) = Y(t)\alpha(t)\,dt + dM(t)$, increment by increment, as standard linear regression model with heteroscedastic errors $dM(t)$, where $M = (M_1, ..., M_n)'$. We should choose the weight matrix to be proportional to the inverse of the error covariance matrix $\text{Cov}(dM(t)|\mathcal{F}^{(n)}(t)) = \text{the } n \times n \text{ diagonal matrix with } i\text{th diagonal entry } \lambda_i(t)\,dt$. However, $\lambda_i(t)$ depends on the unknown $\alpha(t)$. Estimating $\lambda_i(t)$ by (3.7), where the estimator $\hat{\alpha}(t)$ only depends on the past (since the kernel $K$ has support $[0, 1]$), leads to $\hat{W}(t)$.

The following result, which gives the asymptotic distribution of $\hat{A}$, is a special case of Theorem 3.2 of McKeague (1988a). Let $L(t)$ and $V(t)$ denote the $p \times p$ matrices with entries $L_{jk}(t) = EY_{j1}(t)Y_{1k}(t)$ and $V_{jk}(t) = EY_{j1}(t)Y_{1k}(t)\lambda_1^{-1}(t)$, respectively, and let $D[t_0, 1]^p$ denote the product of $p$ copies of the Skorohod space $D[t_0, 1]$.

**Theorem 3.3.** Suppose that $\alpha_1, ..., \alpha_p, L(\cdot), V(\cdot)$ are continuous, $L(t)$ and $V(t)$ are nonsingular for all $t \in [0, 1]$, $\lambda(t, Z_t)$ is bounded away from zero, $b_n \to 0$, $nb_n^2 \to \infty$, and the kernel function $K$ has bounded variation. Let $0 < t_0 < 1$. Then, under Aalen's additive risk model, $\sqrt{n}(\hat{A} - A) \overset{D}{\to} m$ in $D[t_0, 1]^p$, where $m$ is a $p$-variate continuous Gaussian martingale with mean zero and covariance function

$$\text{Cov}(m_j(t), m_k(t)) = \int_{t_0}^t (V^{-1}(s))_{jk} \, ds.$$
3.9. The general nonparametric model

The fully nonparametric model (3) was first studied by Beran (1981). It can be applied successfully only when the sample size is very large and there are a small number of covariates. Inference for this model has been studied further by Doksum and Yandell (1982), Dabrowska (1987a, 1987b), McKeage and Utikal (1987, 1988a) and Cheng (1987). In this section we discuss the main result of McKeage and Utikal (1988a) who introduced an estimator for the "doubly" cumulative hazard function $A(t, x) = \int_0^t \int_0^x \lambda(s, x) ds dx$, $(t, x) \in [0, 1]^2$. This estimator turns out to be important in the development of goodness-of-fit tests for specific regression models. For simplicity we shall only consider the case of a single covariate $(p = 1)$.

Let $I_r$, $r = 1, ..., d_n$ be a partition of the unit interval, where $I_r = [x_{r-1}, x_r]$, $x_r = r/d_n$ and $d_n$ is an increasing sequence of positive integers. Let $N_i(t)$ be the counting process which registers the jumps of $N_i(t)$ when $Z_i(t) \in I_r$, so that $N_i(t) = \sum_{s \leq t} I\{Z_i(s) \in I_r\} dN_i(s)$. Beran (1981) suggested that the cumulative conditional hazard function $A(t, x) = \int_0^t \lambda(s, x) ds$ could be estimated by the Nelson-Aalen type estimator

$$\hat{A}(t, x) = \int_0^t \frac{dN_{r_{(n)}}(s)}{Y_{r_{(n)}}(s)} , \quad \text{for } x \in I_r,$$

and that the conditional survival function $S(t|x) = e^{-A(t,x)}$ could be estimated by the product-limit estimator

$$\hat{S}(t|x) = \prod_{s \leq t} (1 - \hat{A}(s, x)),$$

where $Y_{r_{(n)}}(s) = \sum_{i=1}^n I\{Z_i(s) \in I_r\} Y_i(s)$ and $N_{r_{(n)}} = \sum_{i=1}^n N_i$. Here $d_n$ should tend to infinity at a suitable rate as $n \to \infty$. Dabrowska (1987a, 1987b) obtained weak convergence results for such estimators in the case of right-censoring and non-time-dependent covariate, using the classical approach of Breslow and Crowley (1974). McKeage and Utikal (1987), using the martingale approach, obtained asymptotic results for $\hat{A}$ under general predictable censoring and time-dependent covariates.

McKeage and Utikal (1988a) proposed to estimate $A$ by

$$\tilde{A}(t, x) = \int_0^t \tilde{A}(t, x) dx , \quad (3.9)$$

and obtained the following weak convergence result for $\tilde{A}$. Let $\int_0^t \int_0^x \phi(s, x) dW(s, x)$ denote a continuous version of the Wiener integral of a function $\phi \in L^2([0,1]^2, ds dx)$ with respect to a Brownian sheet $W$; see Wong and Zakai (1974). Suppose that for each $t \in [0,1]$, the random vector $(Z_t, Y_t)$ is absolutely continuous with respect to the product of Lebesgue measure on $[0,1]$ and counting measure, and denote the corresponding density by $f_{Z(t) Y(t)}(z, y)$. Also, assume that $f_{Z(t) Y(t)}(x, 1)$ is a positive, continuous function of $(t, x) \in [0,1]^2$. Let $D_2$ denote the extension of Skorohod space $D[0,1]$ to functions on $[0,1]^2$, as defined in Neuhaus (1971).
Theorem 3.4. Suppose that $\lambda$ is Lipschitz, $d_{n}/n \to \infty$ and $d_{n} = o(n^{\delta})$ for some $\delta \in (\frac{1}{2}, 1)$. Then $\sqrt{n}(\tilde{A} - \lambda) \overset{D}{\to} m$ in $D_{2}$ as $n \to \infty$, where $m = (m(t, z), (t, z) \in [0, 1]^{2})$ is given by

$$m(t, z) = \int_{0}^{t} \int_{0}^{z} \sqrt{h(s, x)} \, dW(s, x),$$

$$h(s, z) = \frac{\lambda(s, z)}{I_{2}(s) Y(s)(z, 1)}.$$

Proof. (Sketch) It can be shown easily that $\sqrt{n}(\tilde{A} - \lambda)$ is asymptotically equivalent in distribution to $\tilde{M}(n)$, where

$$\tilde{M}(n)(t, z) = \frac{\sqrt{n}}{d_{n}} \sum_{r=1}^{[zd_{n}]} \int_{0}^{t} \frac{dM_{r}^{(n)}(s)}{Y_{r}(n)(s)},$$

$$M_{r}^{(n)}(t) = \sum_{i=1}^{n} \int_{0}^{t} I(Z_{i}(s) \in I_{r}) Y_{i}(s) \, dM_{i}(s), \quad r = 1, \ldots, d_{n}.$$  \hfill (3.10)

Since $\tilde{M}(n)(t, z)$ is a martingale for each fixed $z$, Rebolledo's martingale central limit theorem can be used to show that the finite dimensional distributions of $\tilde{M}(n)$ converge to those of $m$ (cf. the proof of Theorem 2.3). Finally, $\{\tilde{M}(n), n \geq 1\}$ is shown to be tight in $D_{2}$ by checking the moment conditions of Bickel and Wichura (1971).

3.4. The general proportional hazards model

Tibshirani (1984) and Hastie and Tibshirani (1986) considered a local partial likelihood technique for estimating the log relative risk function $\eta(z) = \log r(z)$ in the general proportional hazards model (4) with $p = 1$. O'Sullivan (1986a, 1986b) studied a penalized partial likelihood estimator for $\eta$ and established consistency of that estimator.

McKeague and Utikal (1988a) considered estimating the cumulative relative risk function $R(z) = \int_{0}^{z} r(x) \, dx$ by $\hat{R}(x) = \tilde{A}(1, x)$, where $\tilde{A}$ is defined by (3.9). By Theorem 3.4 and the continuous mapping theorem (Billingsley, 1968) we obtain that $\sqrt{n}(\hat{R} - R) \overset{D}{\to} m^{R}$, where $m^{R}$ is a continuous Gaussian martingale with covariance function

$$\text{Cov}(m^{R}(x_{1}), m^{R}(x_{2})) = \int_{0}^{x_{1}} \int_{0}^{x_{2}} h(s, z) \, ds \, dx,$$

provided that $A_{0}$ is constrained to satisfy $A_{0}(1) = 1$ (to ensure identifiability). Similarly, the cumulative baseline hazard function $A_{0}$ can be estimated by $\hat{A}(t) = \tilde{A}(t, 1)$. If $R$ is constrained to satisfy $R(1) = 1$, then $\sqrt{n}(\hat{A} - A_{0}) \overset{D}{\to} m^{0}$, where $m^{0}$ is a continuous Gaussian martingale with covariance function

$$\text{Cov}(m^{0}(t_{1}), m^{0}(t_{2})) = \int_{0}^{t_{1}} \int_{0}^{t_{2}} h(s, z) \, dx \, ds.$$
3.5. Goodness-of-fit tests

There is an extensive literature on goodness-of-fit tests for Cox's proportional hazards regression model, see the references in Arias (1988). Recently, McKeague and Utikal (1988a, 1988b) have developed consistent goodness-of-fit tests for Cox's model, Aalen's additive risk model and the general proportional hazards model against the alternative of the general nonparametric model (3).

Consider testing the null hypothesis $H_0$: Cox's proportional hazards model (1) holds over the region $(t, z) \in [0, 1]^2$. Under $H_0$, the natural estimator of $\mathcal{A}$ is

$$\hat{\mathcal{A}}(t, z) = \hat{\lambda}(t) \int_0^z e^{\beta z} dx,$$

where $\hat{\beta}$ and $\hat{\lambda}$ are defined in Section 3.1 and, if $(T_i, Z_i(T_i))$ falls outside $[0, 1]^2$, the survival time $T_i$ is regarded as being censored (i.e. $\delta_i$ is set to 0). The Kolmogorov-Smirnov type test statistic $T^{(n)} = \sqrt{n} \sup_{(t, z) \in [0, 1]^2} |\hat{A}(t, z) - \hat{A}(t, z)|$ could be used for testing $H_0$. The following result provides a way of determining (based on simulation) an appropriate critical region for $T^{(n)}$. Define

$$S^{(j)}(\beta, t) = n^{-1} \sum_{i=1}^n Z_i(t)^j Y_i(t) I(0 \leq Z_i(t) \leq 1) e^{\beta Z_i(t)},$$

and define the quantities $s^{(j)}$, $\Sigma$ etc. of Section 3.1 in terms of this $S^{(j)}$.

Theorem 3.5. (McKeague and Utikal, 1988b) Suppose that the conditions of Theorems 3.1 and 3.4 hold. Then, under $H_0$, $\sqrt{n}(\hat{\mathcal{A}} - \mathcal{A}) \overset{D}{\rightarrow} m'$ in $D_2$, where

$$m'(t, z) = \int_0^t \int_0^z \sqrt{h(u, z)} dW(u, z) - b(z) \int_0^t \int_0^z \frac{g(u, z)}{s^{(0)}(\beta_0, u)} dW(u, z)$$

$$- c(t, z) \int_0^t \int_0^1 \left\{ x - s^{(1)}(\beta_0, u) \right\} \sqrt{g(u, z)} dW(u, z),$$

$$h(u, z) = \frac{\lambda_0(u) e^{\beta_0 z}}{f_Z(u) g(u, z)(z, 1)},$$

$$g(u, z) = \lambda_0(u) e^{\beta_0 z} f_Z(u) g(u, z)(z, 1),$$

$$b(z) = \int_0^z e^{\beta_0 z} dx,$$

$$c(t, z) = \Sigma^{-1}(\lambda_0(t)) \int_0^z xe^{\beta_0 z} dx - b(z) \int_0^t e(\beta_0, u) \lambda_0(u) du.$$
Using Rebolledo's martingale central limit theorem (see the proof of Theorem 3.4 of Andersen and Gill, 1982) it can be shown that \((\hat{M}_0, \hat{M}_1)^D(m_0, m_1)\) jointly in \(D[0,1]^2\), where \(m_0\) and \(m_1\) are the independent Gaussian martingales defined in Theorem 3.2. The key step in the proof is to see that \((m_0, m_1)\) can be represented in terms of a single Brownian sheet process \(W:\)

\[
m_0(t) = \int_0^t \int_0^1 \frac{\sqrt{g(u, x)}}{\beta_0, u} dW(u, x),
\]

\[
m_1(t) = \int_0^t \int_0^1 \left\{ x - \frac{s^{(1)}(\beta_0, u)}{\beta_0, u} \right\} \sqrt{g(u, x)} dW(u, x).
\]

Then, using Rebolledo's martingale central limit theorem again, and also using Theorem 3.4, it can be shown that \((\hat{M}, \hat{M}_0, \hat{M}_1)^D(m, m_0, m_1)\) jointly in \(D_2 \times D[0,1]^2\). Applying the continuous mapping theorem, we obtain \(\sqrt{n}(\hat{\lambda} - \lambda)^D m - bm_0 - cm_1(1) = n'\). This completes the proof.
References


