A Central Limit Theorem for Local Martingales with Applications to the Analysis of Longitudinal Data

S. A. MURPHY

September 20, 1996

Department of Statistics

Pennsylvania State University

Summary

A functional central limit theorem for a local square integrable martingale with persistent discontinuities is given. By persistent discontinuities, it is meant that the martingale has jumps which do not vanish asymptotically. This central limit theorem is motivated by problems in the analysis of longitudinal and life history data.

Running Headline: A Central Limit Theorem for Martingales

Key words: Longitudinal Data, Event History Analysis, Non-Classical Central Limit Theorem, Martingale

Research supported by NSF grant DMS-9307255 and partially carried out during the author's visit with the Econometrics Dept., Free University, Amsterdam.

1. Introduction

Very little recent work has been done on Non-Classical Central Limit Theorems in addition to the work by Gill (1982) and the paper by Liptser and Shiryaev (1983) which is later reformulated in the book by Jacod and Shiryaev (1987). The central limit theorem given here is based on the later two works, but the conditions given are amenable to applications in life history/ longitudinal data analysis. Life history/longitudinal data typically involves observation of entities or individuals over a period of time. Even though this type of data may be thought of as the observation of stochastic processes, the statistical analysis is quite different. The analysis of life history data is based on the observation of several or many stochastic processes each over a short time period instead of observation of a very few (or one) stochastic processes over a long period of time. Therefore the asymptotics given here will be for the number of individuals/processes increasing without bound.

Both longitudinal and life history data can expressed as observations of marked point processes (Arjas and Haara, 1992). The event times of the point process are the times at which one collects information on the individuals and the marks are the information collected. Asymptotic results for estimators and test statistics are based on a central limit theorem for an estimating equation. The estimating equation may be based on the derivative of the log of the full or partial likelihood. This derivative forms a local square integrable martingale under integrability conditions(Andersen et.al., 1993). More generally, estimating equations can be constructed by parametrizing aspects of the conditional distribution of the information collected at a time point given the past. These estimating equations are integrals with respect to the marked point process and under integrability conditions form locally square integrable martingales (Murphy and Li, 1993). Central limit theorems (Rebolledo, see Andersen et.al., 1993) for continuous time martingales assume that the intensity of the jumps of the martingale is (asymptotically) continuous. However it is easy to envision the situation in which one plans to make measurements on each individual at regular intervals (eg. every 3 months) but some individuals appear earlier or later for measurements. The time at which the individual appears could depend on past history, i.e. an appointment for a sicker patient may be scheduled earlier due to health concerns (doctor's care or patient self-selection, Grüger, Kay and Schumacher, 1991). The asymptotic analysis should allow for both measurements taken at random times and clumping of measurements (at the 3 month intervals). Additionally individuals may be lost to follow up or censored. The theorem presented in the next section will also allow for dependence between individuals which is due to the censoring mechanism and time dependent covariates.

The first theorem is given for a local square integrable martingale. Next this theorem is specialized to a central limit theorem for integrals with respect to a marked point process. Lastly motivating applications are discussed. All of the proofs are in the appendix.

2. A central limit theorem

The first theorem is the most general given here and is for a d-dimensional local square integrable martingale, M_n with $M_n(0) = 0$, defined on a stochastic basis $\left(\Omega^n, \mathbb{F}^n, \{\mathbb{F}_t^n\}_{t \in \Re^+}, P^n\right)$. Associated with M_n is a marked point process which counts the jumps of M_n, ΔM_n, and records the sizes of the jumps as follows, $\mu_n(d\boldsymbol{x}, dt) = \sum_{s \ni \Delta M_n(s) \neq 0} \epsilon_{\Delta M_n(s), s}(d\boldsymbol{x}, dt)$ $(\boldsymbol{x} \in \Re^d)$ where ϵ_u is a probability measure giving mass 1 to the point u. The marked point process has a predictable compensator given by, $\nu_n(d\boldsymbol{x}, dt)$. For the precise definition of the predictable sigma field, $\mathbb{F}_p^n$, and other terminology see Jacod and Shiryaev (Chapter 2, 1987). Using this marked point process one can decompose M_n into a continuous local square integrable martingale, M_n^c, plus the compensated jumps, $M_n(\cdot) = M_n^c(\cdot) + \int_0^{\cdot} \int \boldsymbol{x}(\mu_n(d\boldsymbol{x}, dt) - \nu_n(d\boldsymbol{x}, dt))$. It is always possible to write ν_n as, $\nu_n(d\boldsymbol{x}, dt) = K_n(d\boldsymbol{x}; t)\Lambda_n(dt)$ where $K_n(d\boldsymbol{x}; s)$ is a transition function from $\left(\Re^+ \times \Omega^n, \mathbb{F}_p^n\right)$ to $\left(\Re^d, \mathcal{B}(\Re^d)\right)$ and Λ_n is a predictable nondecreasing process. Let J_n be a subset of discontinuities of Λ_n. These will be the persistent jumps which will contribute to the fixed jumps of the limiting Gaussian process. The jumps of Λ_n which are not contained in J_n will be assumed to be asymptotically negligible.

The accumulation of information necessary for asymptotics on the continuous part (and the asymptotically negligible jumps) of M_n is formed by summing over ever smaller intervals in time, uncorrelated increments of M_n. Lipster and Shiryaev avoid the details of how one might accumulate information on the persistent jumps by requiring that the conditional distribution (given the past) of the jump sizes approach a normal distribution sufficiently fast (condition $\mathbf{R}^*$ in Liptser and Shiryaev, 1983).

In applications, it is necessary to give some thought to how one would accumulate information at the persistent jumps. In the discussion following Corollary 1, the jump sizes correspond to the derivative of a log density. The density is parametrized by a regression of the response observed at the jump time on covariates. In regression one often assumes that, given the covariates, the responses are independent. Denoting the covariates by the variable, θ, a general assumption, which mimics the above regression assumption, is that the conditional distribution of the size of a jump given that the jump occurred and the past is the mixture of a convolution, i.e., for $t \in J_n$, assume that

$$K_n(d\boldsymbol{x};t) = \int I\{\boldsymbol{x} \neq 0\}(\times_i \mathcal{F}_{n,i})(d\boldsymbol{x};\theta,t)\mathcal{G}_n(d\theta;t). \tag{1}$$

The above product $(\times_i \mathcal{F}_{n,i})$ denotes the convolution of probability transition functions, $\mathcal{F}_{n,i}$'s, each on $\Re^d$. For θ, of dimension $n \times p$, $\mathcal{G}_n$ is a probability transition function from $(\Re^+ \times \Omega^n, \mathbb{F}_p^n)$ to $(\Re^{np}, \mathcal{B}(\Re^{np}))$ and $\mathcal{F}_{n,i}$ is a probability transition function from $(\Re^+ \times \Omega^n \times \Re^{np}, \mathbb{F}_p^n \vee \mathcal{B}(\Re^{np}))$ to $(\Re^d, \mathcal{B}(\Re^d))$.

Note that for t a discontinuity point of Λ_n, the assumption of $M_n(0) = 0$, implies that $\int \boldsymbol{x} K_n(d\boldsymbol{x};t) = 0$ $a.e.P^n$. Make the further assumption that

$$\int \boldsymbol{x}(\times_i \mathcal{F}_{n,i})(d\boldsymbol{x};\theta,t) = 0 \tag{2}$$

$a.e.$ $(\mathcal{G}_n(d\theta;t)dP^n)$. Define $\sigma_n^2(t) = \int \boldsymbol{x}\boldsymbol{x}^T \nu_n(d\boldsymbol{x},\{t\})$ and $\langle M_n^c \rangle$ to be the predictable variation matrix of M_n^c. Without loss of generality, each component of the vector, M_n, is assumed to belong to the space of right continuous, left hand-limited functions on $[0,\infty)$, denoted by $D[0,\infty)$. Let $\mathcal{M}$ be a Gaussian martingale on $D[0,\infty)^d$ with $E(\mathcal{M}\mathcal{M}^T) = \Sigma$.

Theorem 1

For each $t \in D$, D a subset of $\Re^+$, consider the following assumptions.

1) Asymptotic Negligibility.

a) For all $\epsilon > 0$,

$$\int_0^t I\{s \notin J_n\} \int \boldsymbol{x}\boldsymbol{x}^T I\{||\boldsymbol{x}|| > \epsilon\}\nu_n(d\boldsymbol{x},ds) \xrightarrow{\mathcal{P}} 0.$$

b) For all $\epsilon > 0$,

$$\sum_{s \in J_n, s \leq t} \int \sum_i \int \boldsymbol{x}\boldsymbol{x}^T I\{||\boldsymbol{x}|| > \epsilon\}\mathcal{F}_{n,i}(d\boldsymbol{x};\theta,s)\mathcal{G}_n(d\theta;s)\Lambda_n(\{s\}) \xrightarrow{\mathcal{P}} 0.$$

2) Convergence of the variance.

a)

$$\langle M_n^c\rangle(t) + \int_0^t \int \boldsymbol{x}\boldsymbol{x}^T \nu_n(d\boldsymbol{x}, ds) \xrightarrow{\mathcal{P}} \Sigma(t).$$

b) For all l, j,

$$\sum_{s\in J_n,\ s\le t} \int \left|\sigma_n^2(s)_{lj} - \sum_i \int x_l x_j \mathcal{F}_{n,i}(d\boldsymbol{x};\theta,s)\right| \mathcal{G}_n(d\theta;s)\Lambda_n(\{s\}) \xrightarrow{\mathcal{P}} 0,$$

and

$$\sum_{s\in J_n, s\le t} |\Lambda_n(\{s\}) - 1| \xrightarrow{\mathcal{P}} 0. \tag{$*$}$$

3) Tightness in $D[0,\infty)$.

$$\sum_{s\le t}\Big(\sum_{j=1}^d \sigma_n^2(s)_{jj}\Big)^2 \xrightarrow{\mathcal{P}} \sum_{s\le t}\Big(\sum_{j=1}^d \Delta\Sigma(s)_{jj}\Big)^2.$$

Under conditions 1) and 2), the finite dimensional distributions of M_n in D converge to the finite dimension distributions of $\mathcal{M}$. If D is dense on the real line then the additional assumption of 3) implies that M_n converges weakly with respect to the Skorohod metric on $D[0,\infty)$ to $\mathcal{M}$.

Remarks:

(1) Intuitively, the joint conditional distribution of a jump at a time $t \in J_n$ and the jump size is,

$$\int_\theta (\times_i \mathcal{F}_{n,i})(d\boldsymbol{x};\theta,t)\mathcal{G}_n(d\theta,t)\Lambda_n(\{t\}) + \epsilon_0(d\boldsymbol{x})(1-\Lambda_n(\{t\}). \tag{3}$$

The assumption, 2b), is present because the distribution (3) is for fixed time, t, a mixture of a convolution. If $\int_\theta (\times_i \mathcal{F}_{n,i})(d\boldsymbol{x};\theta,t)\mathcal{G}_n(d\theta,t)$ is a convolution then the first part of 2b) is implied by the second part of 2b), assumption (*) on the persistent jumps. Furthermore if for fixed t, (3) can be written as a convolution then assumption 2b) need not be made.

(2) In applications the above result may be more useful when the weak convergence is in supremum norm on the space of bounded functions on a compact interval, ($l^\infty([0,\tau])$). The space, $l^\infty([0,\tau])$ is discussed by van der Vaart and Wellner (1993). This can be achieved by employing a different version of 3) above. The more restrictive version essentially requires the a priori knowledge of the locations of the persistent jumps (points of discontinuity of Σ). Replace 3) by:

$3'$) Tightness in $l^\infty([0,\tau])$. Suppose $D = [0,\tau]$ and in addition, for each s, a discontinuity point of Σ and each j, we have

$$\sigma_n^2(s)_{jj} \xrightarrow{\mathcal{P}} \Delta\Sigma(s)_{jj}.$$

Then there exists a version of $\mathcal{M}$ which is a tight Borel measurable, Gaussian process in $l^\infty([0,\tau])$ and M converges weakly to $\mathcal{M}$.

(3) The Gaussian martingale, $\mathcal{M}$, possess the following properties. For a fixed countable set of times, t, $\Sigma(t) - \Sigma(t-) = \Delta\Sigma(t)$ will be nonzero, symmetric and nonnegative definite. These are the fixed times of discontinuities of $\mathcal{M}$. Outside of these fixed times of discontinuity, almost all paths of $\mathcal{M}$ are continuous. Additionally $\Delta\mathcal{M}(t) = \mathcal{M}(t) - \mathcal{M}(t-)$ is multivariate normal with mean zero and variance-covariance matrix, $\Delta\Sigma(t)$. The covariance of $\mathcal{M}_k(t)$ and $\mathcal{M}_l(s)$, for t less than s, is the entry in the kth row, lth column of $\Sigma(t)$. For more properties of a Gaussian martingale see Jacod and Shiryaev, (pg. 111, 1987).

Now specialize to local square integrable martingales which can be expressed as integrals with respect to a nonexplosive marked point process. The marked point process, $\mathcal{N}_n$, is defined on the stochastic basis $\left(\Omega^n, \mathbb{F}^n, \{\mathbb{F}_t^n\}_{t\in\Re^+}, P^n\right)$ and is a composite of n marked point processes so that the event times of $\mathcal{N}_n$, say $\{T_j\}_{j\geq 1}$, are the ordered event times of the individual marked point processes. Part of the mark at each T_j is an indicator of which of the individual marked point processes jumped at that time. Denote the mark at time T_j be $(\boldsymbol{X}_j, \boldsymbol{Z}_j, \boldsymbol{\delta}_j)$. In applications, $\boldsymbol{X}_j$ is the matrix of reponses and $\boldsymbol{Z}_j$ is the matrix of covariates collected at time T_j. So the mark is a matrix of real components and it has row dimension n; if the ith marked point process has a jump at T_j, δ_{ji} is set to one and the ith row of $(\boldsymbol{X}_j, \boldsymbol{Z}_j)$ corresponds to the mark of the i process. Otherwise, the ith row of $(\boldsymbol{X}_j, \boldsymbol{Z}_j)$ is set to the empty set and δ_{ji} is set to zero. Denote the compensator of $\mathcal{N}_n$ by $\boldsymbol{\Lambda}_n$. If the filtration, $\mathbb{F}$ is the internal filtration, then $\boldsymbol{\Lambda}_n$, written as a measure, is given by, $\boldsymbol{\Lambda}_n(d\boldsymbol{x}, d\boldsymbol{z}, d\boldsymbol{\delta}, ds) = P[\boldsymbol{X}_j \in d\boldsymbol{x}, \boldsymbol{Z}_j \in d\boldsymbol{z}, \boldsymbol{\delta}_j \in d\boldsymbol{\delta}, T_j \in ds \big| T_j \geq s,\ \mathcal{F}_{T_{j-1}}]$ on $T_{j-1} < s \leq T_j,\ j \geq 1$, where ds denotes the interval $[s, s+ds)$. For $\{H_i\}_{i\geq 1}$, each a d dimensional deterministic function and $\{\psi_s^i\}_{i\geq 1}$ predictable, consider,

$$M_n(\cdot) = \int_0^{\cdot} \int n^{-1/2} \sum_{i=1}^n H_i(x_i; z_i, \psi_s^i, s)\delta_i \mathcal{N}_n(d\boldsymbol{x}, d\boldsymbol{z}, d\boldsymbol{\delta}, ds).$$

To ensure that M_n is a local square integrable martingale, assume that the compensator of M_n is zero and that $\int_0^t \int \left(\sum_{i=1}^n H_i(x_i; z_i, \psi_s^i, s)\delta_i \right)^2 \boldsymbol{\Lambda}_n(d\boldsymbol{x}, d\boldsymbol{z}, d\boldsymbol{\delta}, ds)$ is locally integrable (Jacod and Shiryaev, pg. 73, 1987). After the corollary is a discussion of how equations of the form of M_n arise as estimating equations in survival analysis and in the analysis of life history data.

The marked point process of the jumps of M_n is,

$$\mu_n(du, dt) = \int \epsilon_{\{n^{-1/2}\sum_i H_i(x_i, z_i, \psi_s^i, s)\delta_i\}}(du) I\{n^{-1/2} \sum_i H_i(x_i, z_i, \psi_s^i, s)\delta_i \neq 0\} \mathcal{N}_n(d\boldsymbol{x}, d\boldsymbol{z}, d\boldsymbol{\delta}, ds).$$

Of course ν_n follows the same formula as the above but with $\boldsymbol{\Lambda}_n$ in place of $\mathcal{N}_n$. Assume (1) with $\theta = (\boldsymbol{z}, \boldsymbol{\delta})$ and $\mathcal{F}_{n,i}(dv; \theta, s) = \int I\{n^{-1/2} H_i(x; z_i, \psi_s^i, s)\delta_i \in dv\} F_{n,i}(dx; z_i, s)$. The distribution function $F_{n,i}$ is allowed to be a function of the past, that is, $F_{n,i}$ is a probability transition function from $(\Re^+ \times \Omega^n \times \Re^p, \mathbb{F}_p^n \vee \mathcal{B}(\Re^p))$ to $(\Re^k, \mathcal{B}(\Re^k))$ where p is the column dimension of $\boldsymbol{Z}$ and k is the column dimension of $\boldsymbol{X}$. Assume that $\int H_i(x; z_i, \psi_s^i, s) F_{n,i}(dx; z_i, s) = 0$ which implies (2). Define $\boldsymbol{\Lambda}_n^{\{i\}}$, $({\mathcal{G}_n}^{\{i\}})$ to be $\delta_i \boldsymbol{\Lambda}_n$, $(\delta_i \mathcal{G}_n)$ integrated over all $\boldsymbol{\delta}$, $(x_l, z_l,\ l \neq i)$. Let J_n be the set of all discontinuities of $\boldsymbol{\Lambda}_n$. This set certainly includes the discontinuities of ν_n. If $t \in J_n$ then $\boldsymbol{\Lambda}_n^{\{i\}}(dx, dz, \{t\}) = F_{n,i}(dx; z, t){\mathcal{G}_n}^{\{i\}}(dz; t)\Lambda_n(\{t\})$ and $\sigma_n^2(t)_{lj} = n^{-1} \sum_{i=1}^n \int_{x,z} H_i(x; z, \psi_t^i, t)_l H_i(x; z, \psi_t^i, t)_j F_{n,i}(dx; z, t)$ ${\mathcal{G}_n}^{\{i\}}(dz; t)\Lambda_n(\{t\})$. Assume that at the continuity points of $\boldsymbol{\Lambda}_n$, at most one of the individual marked point processes can jump, that is,

$$\int_{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\delta}, s} I\{s \notin J_n\} \delta_i \delta_j \boldsymbol{\Lambda}_n(d\boldsymbol{x}, d\boldsymbol{z}, d\boldsymbol{\delta}, ds) = 0. \tag{4}$$

The assumptions of theorem 1 simplify nicely.

Corollary 1

For each $t \in D$ consider the following assumptions.

1) Asymptotic negligibility.

For all $\epsilon > 0$,

$$n^{-1} \sum_{i=1}^n \int_0^t \int H_i(x; z, \psi_s^i, s) H_i(x; z, \psi_s^i, s)^T I\left\{ \left|\left| H_i(x; z, \psi_s^i, s) \right|\right| > \epsilon\sqrt{n} \right\} \boldsymbol{\Lambda}_n^{\{i\}}(dx, dz, ds) \xrightarrow{\mathcal{P}} 0.$$

2) Convergence of the variance.

a)

$$n^{-1}\sum_{i=1}^{n}\int_0^t\int H_i(x;z,\psi_s^i,s)H_i(x;z,\psi_s^i,s)^T\boldsymbol{\Lambda}_n^{\{i\}}(dx,dz,ds)\xrightarrow{\mathcal{P}}\Sigma_t.$$

b) For all l, j,

$$\sum_{s\le t}\int\left|\sigma_n^2(s)_{lj}-n^{-1}\sum_{i=1}^{n}\int H_i(x;z,\psi_s^i,s)_l H_i(x;z,\psi_s^i,s)_j F_{n,i}(dx;z_i,s)\delta_i\right|\mathcal{G}_n(d\boldsymbol{z},d\boldsymbol{\delta};s)\Lambda_n(\{s\})\xrightarrow{\mathcal{P}}0$$

and (*)

3) Tightness.

$$\sum_{s\le t}\Big(\sum_{j=1}^{d}\sigma_n^2(s)_{jj}\Big)^2\xrightarrow{\mathcal{P}}\sum_{s\le t}\Big(\sum_{j=1}^{d}\Delta\Sigma(s)_{jj}\Big)^2.$$

3′) Tightness in $l^\infty([0,\tau])$. Suppose that for each s, a discontinuity point of Σ and each j, we have

$$\sigma_n^2(s)_{jj}\xrightarrow{\mathcal{P}}\Delta\Sigma(s)_{jj}.$$

Under conditions 1) and 2), the finite dimensional distributions of M_n in D converge to the finite dimensional distributions of $\mathcal{M}$. If D is dense on the real line then the additional assumption of 3) implies that M_n converges weakly with respect to the Skorohod metric on $D[0,\infty)$ to $\mathcal{M}$. If $D=[0,\tau]$ and conditions 1), 2) and 3′) hold, there exists a version of $\mathcal{M}$ which is a tight Borel measurable process in $l^\infty([0,\tau])$ and M converges weakly to $\mathcal{M}$.

As before, the assumption of 2b) is only necessary because the conditional distribution of X given the time of the jump and the past is a mixture of a convolution of distributions. If the distributions of the rows of $(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\delta})$ are independent given the time of the jump and the past then the first part of 2b) is implied by the assumption (*) on the persistent jumps. Furthermore if the joint conditional distribution of the rows of $(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\delta})$ and jump time given the past can be expressed as a product of distributions for fixed s ((3) is a convolution), then assumption 2b) need not be assumed.

A marked point process can be used to model life history data (Arjas and Haara, 1992), in which case the event times of the process, indicate the ordering in which observations are collected. In many applications $T_j\equiv j$. The mark at time T_j is the information collected. A simplified version is as follows. In contrast to survival analysis there are two indicators of observability. First an indicator of censoring, $\boldsymbol{\Delta}_j$, an $n \times 1$ vector with ith entry equal to one if the ith subject leaves the study at "time" T_j or is presently absent from the study and zero otherwise. Second an indicator of measurement time, $\boldsymbol{\delta}_j$, an $n \times 1$ vector with ith entry equal to one if the ith subject contributes both a covariate Z and a response, X, at "time", T_j and zero otherwise. Subject i can only contribute a response at time T_j, if $\Delta_{j-1,i} = 0$. Then $(\boldsymbol{X}_j, \boldsymbol{Z}_j, \boldsymbol{\delta}_j, \boldsymbol{\Delta}_j)$ is the mark at time T_j.

The marked point process approach is best illustrated using the terminology of a longitudinal study. This illustration is a generalization of the work by Scheike, (1994). In an longitudinal study of n individuals, T_j would be time of the jth event (either an appointment or a censoring time); if individual i contributes a measurement of X at time T_j then $\delta_{ji} = 1$ otherwise $\delta_{ji} = 0$. Interest lies in regressing X on Z and possibly other variables measured prior to the present event time. These other variables can include the present time, and measurements of X and Z made on previous appointments. Since some appointments may be scheduled a priori at regular intervals (e.g. every three months), several individuals can contribute measurements at any one time. The times of the other appointments or censorings may depend on previous measurements of X and Z. Let $U^{(j-1)} = ((T_l, \boldsymbol{X}_l, \boldsymbol{Z}_l, \boldsymbol{\delta}_l, \boldsymbol{\Delta}_l), l < j, T_j, \boldsymbol{Z}_j, \boldsymbol{\delta}_j)$ and define, $\mathbb{F}_{T_j} = \sigma\{(T_l, \boldsymbol{X}_l, \boldsymbol{Z}_l, \boldsymbol{\delta}_l, \boldsymbol{\Delta}_l), l \leq j\}$. Let $\mathcal{I}$ be an arbitrary subset of the integers, 1 through n. Assume that responses from different individuals present at an appointment time are conditionally independent. That is, the conditional independence assumption is : on $\prod_{i \in \mathcal{I}} \delta_{ji} \prod_{i \notin \mathcal{I}} (1 - \delta_{ji}) = 1$,

$$P[X_{ji} \leq x_i, i \in \mathcal{I} | U^{(j-1)}] = \prod_{i \in \mathcal{I}} F_{n,i}(x_i; Z_{ji}, \psi_t^i, t)|_{t=T_j},$$

where for each l and $t \in (T_{l-1}, T_l]$, ψ_t^i is a function of the variables in $\mathcal{F}_{T_{l-1}}$ (i.e. ψ^i is predictable). This assumption means that on $\prod_{i \in \mathcal{I}} \delta_{ji} \prod_{i \notin \mathcal{I}} (1 - \delta_{ji}) = 1$, the $X_{ji}, i \in \mathcal{I}$ are conditionally independent with distribution functions $F_{n,i}, i \in \mathcal{I}$. The conditional distribution, $F_{n,i}$, can be parametrized to reflect the effect of the covariates on X_j. A partial likelihood for the response X is

$$\prod_{j \geq 1} \prod_{i=1}^{n} \left(F_{n,i}(dX_{ji}; Z_{ji}, \psi_{T_j}^i, T_j) \right)^{\delta_{ji}}.$$

By modeling only $F_{n,i}$, we model only the probabilistic evolution of X through time and not the evolution of the other variables, $\boldsymbol{Z}$ and T.

If it is possible to completely model the density of $F_{n,i}$, say $f_{n,i}$, as a function of a parameter, β, then a partial likelihood analysis would be based on the partial likelihood score. That is the local martingale, M_n is given by putting $H_i = \frac{\partial}{\partial\beta}\ln\left(f_{n,i}\right)$. On the other hand, if one is willing to parametrize at most the conditional mean and conditional variance of $\boldsymbol{X}_j$, then an analysis based on the projected partial score is still possible. Suppose that the conditional mean and variance of $F_{n,i}(\cdot; Z_{ji}, \psi^i_{T_{ji}}, T_{ji})$ is given by $\mu_{ji} = \mu_i(Z_{ji}, \psi^i_{T_{ji}}, T_{ji}; \beta)$, and $V_{ji} = V_i(Z_{ji}, \psi^i_{T_{ji}}, T_{ji}; \beta)$, respectively. The projected partial score is given by, $\sum_{j\geq 1}\sum_{i=1}^n \frac{\partial}{\partial\beta}\mu_{ji}V_{ji}^{-1}(X_{ji} - \mu_{ji})\delta_{ji}$ (Murphy and Li, 1993). Here, put $H_i(x; z, \psi^i_t, t) = \frac{\partial}{\partial\beta}\mu_i(z, \psi^i_t, t; \beta)V_i^{-1}(z, \psi^i_t, t; \beta)(x - \mu_i(z, \psi^i_t, t; \beta))$ to form the local martingale, M_n. Scheike (1994) considers examples in which there are no Z's so that the mean and variance of $\boldsymbol{X}_j$ are functions of $\psi^i_{T_j}$ only.

3. Examples

1. *Conditionally Independent Marked Point Processes*

A marked point process which satisfies, (1) and (4) can be formed by combining conditionally independent nonexplosive point processes. Conditionally independent point processes share with the multivariate point process, the property, that at time points which are continuity points of the compensator only one of the component processes is allowed to jump. Suppose that on a given stochastic basis, there exists n marked point processes, $N_i,\ i \geq 1$ each with marks $(X_j, Z_j) \in E$, the mark space, $j \geq 1$ and locally integrable compensators, $\Lambda_i,\ i \geq 1$. Define the composite marked point process, $\mathcal{N}_n$ to have event times, T_j, at the ordered event times of the N_i's and the ith row of the mark at T_j to be the empty set if N_i does not jump at T_j and to be the mark from N_i if N_i does jump at T_j. In addition add to the mark at time T_j, the n by one vector $\boldsymbol{\delta}_j$, with i row equal to one if N_i jumps at T_j and zero otherwise. So the mark at time, T_j, is $(\boldsymbol{X}_j, \boldsymbol{Z}_j, \boldsymbol{\delta}_j)$ which has n rows, one for each of the N_i's. The conditional independence assumption is;

$$\sum_{s\leq \cdot}\prod_{i\in S} N_i(A_i, \{s\}) - \int_0^{\cdot}\prod_{i\in S}\Lambda_i(A_i, ds)$$

is a local martingale for all S, subsets of $\{1,2,\ldots,n\}$ and A_i, measurable subsets of the mark space, E. This means (use the monotone class theorem) that the compensator of $\mathcal{N}_n$ is $\boldsymbol{\Lambda}_n(d\boldsymbol{x},d\boldsymbol{z},d\boldsymbol{\delta},ds) = I\{\delta. \geq 1\}\prod_{i=1}^n \Lambda_i(dx_i,dz_i,ds)^{\delta_i}(1-\Lambda_i(E,ds))^{1-\delta_i}\epsilon_\emptyset(dx_i,dz_i)^{1-\delta_i}$. The dot in the place of the index for δ indicates that δ is summed over the index. Note that at the continuity points (in s) of $\boldsymbol{\Lambda}_n$ only one of the N_i can jump, that is, (4) holds. So conditionally independent marked point processes behave like a multivariate point process at the continuity points of $\boldsymbol{\Lambda}_n$. Write $\Lambda_i(dx_i,dz_i,ds)$ as $F_i(dx_i;z_i,s)G_i(dz_i;s)\lambda_i(ds)$. Then, $\boldsymbol{\Lambda}_n(d\boldsymbol{x},d\boldsymbol{z},d\boldsymbol{\delta},ds) = I\{\delta. \geq 1\}\prod_{i=1}^n \big(F_i(dx_i;z_i,s)G_i(dz_i;s)\lambda_i(ds)\big)^{\delta_i}$ $\big((1-\lambda_i(ds))\epsilon_\emptyset(dx_i,dz_i)\big)^{1-\delta_i}$ and $\boldsymbol{\Lambda}_n^{\{i\}}(dx,dz,ds) = F_i(dx;z,s)G_i(dz;s)\lambda_i(ds)$.

To write $\boldsymbol{\Lambda}_n$ in the form specified by (1), put $\mathcal{F}_{n,i}(dv;\theta,s) = \int I\{(n^{-1/2}H_i(x;z_i,\psi_s^i,s)\delta_i \in dv\}F_{n,i}(dx;z_i,s)G_i(dz_i;s)$, put $\mathcal{G}_n(d\boldsymbol{\delta};s) = I\{\delta. \geq 1\}\frac{\prod_{i=1}^n \big(\lambda_i(ds)\big)^{\delta_i}\big(1-\lambda_i(ds)\big)^{1-\delta_i}}{1-\prod_{i=1}^n(1-\lambda_i(ds))}$ and put $\Lambda_n(ds) = 1-\prod_{i=1}^n(1-\lambda_i(ds))$.

If $\int_0^t\int\big(\sum_{i=1}^n H_i(x_i;z_i,\psi_s^i,s)\delta_i\big)^2\boldsymbol{\Lambda}_n(d\boldsymbol{x},d\boldsymbol{z},d\boldsymbol{\delta},ds) < \infty$ holds for all t then M_n (defined in the previous section) is a locally square integrable martingale, and since (3) can be written as a convolution for fixed s, only assumptions 1), 2a) and 3) of the corollary need be verified in order to prove weak convergence for M_n. Assumptions 1) and 2a) are identical to the assumptions (2.5.1) and (2.5.3) of Rebolledo's theorem as stated by Andersen et. al. (1993). The only additional assumption here is 3) concerning the persistent jumps.

2.*Nelson-Aalen Estimator for Cumulative Hazards with Jumps*

Suppose that for each n, there exists mutually independent $X_{1n},X_{2n},\ldots,X_{nn}$ and $U_{1n},U_{2n},\ldots,U_{nn}$ where each X_{in} has distribution function F and integrated hazard rate, $A(\cdot) = \int_0^\cdot(1-F(u-))^{-1}dF(u)$ and each U_{in} has distribution function G. In the following we consider estimation of A. Let $N_i^X(t) = I\{X_{in} \leq t\}$, $N_i^U(t) = I\{U_{in} \leq t\}$ and $Y_i(t) = I\{X_{in}\wedge U_{in} \geq t\}$ for all $t \in \Re^+$. The filtration, $\mathbb{F}$, is the internal filtration of the $(N_i^X,\ N_i^U)$'s. These processes can be represented by a marked point process with event times at each of the U_{in} and X_{in} and marks indicating which of the U_{in} and X_{in} occur at each event time. Then $\mathbb{F}$ is the internal filtration for this process and Jacod's formula for the compensator (Andersen et. al. pg. 96, 1993) yields, $\int_0^\cdot\prod_{i\in S}\big(I\{X_{in} \geq u\}dA(u)\big)$

as the predictable compensator of $\sum_{s\leq \cdot}\prod_{i\in S} N_i^X(\{s\})$ for S any subset of $1,\ldots,n$. Furthermore, since $I\{U_{in}\geq u\}$ is predictable, $\sum_{s\leq \cdot}\prod_{i\in S} I\{U_{in}\geq u\}N_i^X(\{s\}) - \int_0^\cdot \prod_{i\in S}\big(Y_i(u)dA(u)\big)$ forms a zero mean martingale. Put $N_i(t)=\int_0^t I\{U_{in}\geq u\}dN_i^X(s)$ and $J(s)=I\{Y_\cdot(s)>0\}$ (the dot in place of the index denotes the sum over that index). Note that the above form of the compensator for $\sum_{s\leq \cdot}\prod_{i\in S} N_i(\{s\})$ implies the conditional independence of the N_i as defined in the last section. Define $\mathcal{N}_n$ to be the point process with event times, T_j at the ordered event times of the N_i's and at time T_j, the mark, $\boldsymbol{\delta}_j$, with ith row of $\boldsymbol{\delta}_j$ equal to 1 if N_i jumps at time T_j and zero otherwise. The conditional independence property implies that the compensator of $\mathcal{N}_n$ is given by $\boldsymbol{\Lambda}_n(d\boldsymbol{\delta},ds)=I\{\delta_\cdot\geq 1\}\prod_{i=1}^n\big(Y_i(s)A(ds)\big)^{\delta_i}\big(1-Y_i(s)A(ds)\big)^{1-\delta_i}I\{\delta_\cdot>0\}$.

We are unable to observe the $(N_i^X,\ N_i^U)$'s, rather we observe only the (N_i,Y_i)'s. The Nelson Aalen estimator is defined for t such that $A(t)$ is finite and is given by, $\hat{A}(t)=\int_0^t J(s)Y_\cdot(s)^{-1}dN_\cdot(s)$ (in the integrand, interpret 0/0 as 0). The form of the compensator given in the last paragraph implies that the compensator of $\hat{A}$ is $\int_0^t J(s)dA(s)$. An asymptotic analysis of the Nelson-Aalen estimator is based on $X_n(t)=\sqrt{n}\big(\hat{A}(t)-\int_0^t J(s)dA(s)\big)$ and on $\sqrt{n}\int_0^t J(s)-1\ dA(s)$. Suppose that for a specified T, $A(T)<\infty$ and $G(T-)<1$. The Glivenko-Cantellli theorem implies that $\sup_{0\leq t\leq T}|n^{-1}Y_\cdot(t)-(1-F(t-)(1-G(t-)|$ converges in probability to zero which further implies that $\sup_0|\sqrt{n}\int_0^t J(s)-1\ dA(s)|$ goes to zero in probability.

Since $J(s)Y_\cdot(s)^{-1}$ bounded above by 1, X_n is a square integrable martingale (see Andersen et. al., 1993, pg. 181) and Theorem 1 can be used to derive the asymptotic distribution for X_n on $[0,T]$. Put $M_n(t)=X_n(t\wedge T)$ and let J_n be the set of discontinuities of A. The marked point process, μ_n recording the jumps of M_n is given by

$$\mu_n(dx,dt)=I\{t\leq T\}\int_{\boldsymbol{\delta}} I\{x\neq 0\}\epsilon_{\sqrt{n}J(t)Y_\cdot^{-1}(t)(\delta_\cdot-Y_\cdot(t)\triangle A(t))}(dx)\mathcal{N}_n(d\boldsymbol{\delta},dt)+$$
$$I\{t\leq T\}I\{x\neq 0\}\epsilon_{(-\sqrt{n}J(t)\triangle A(t))}(dx)\left(1-\int_{\boldsymbol{\delta}}\mathcal{N}_n(d\boldsymbol{\delta},dt)\right).$$

The compensator of μ_n, ν_n, is given by same formula as above but with $\boldsymbol{\Lambda}_n$ in place of $\mathcal{N}_n$. For t a continuity point of A, $\nu_n(dx,dt)$ simplifies considerably to $\nu_n(dx,dt)=I\{x\neq 0\}Y_\cdot(t)\epsilon_{(\sqrt{n}J(t)/Y_\cdot(t))}(dx)A(dt)$. Otherwise note that $\mathcal{F}_{n,i}(\cdot;t)$ is the distribution of $\sqrt{n}(B-Y_i(t)\triangle A(t))J(t)/Y_\cdot(t)$ where B is a Bernoulli with success probability, $Y_i(t)\triangle A(t)$ and $\Lambda_n(dt)=\epsilon_{\{s:\triangle A(s)\neq 0\}}(dt)$. This means that (3) can be written as a convolution and assumption 2b) is not necessary. Assumptions 1a), 1b) and 2a) all follow by the Glivenko- Cantelli theorem applied to $n^{-1}Y.$ and the assumption that $(1-F(T-)(1-G(T-)>0$. The equation in assumption 1a) is $\int_0^{t\wedge T} nJ(s)/Y.(s)I\{J(s)/Y.(t)>\epsilon/\sqrt{n}\epsilon\}A^c(ds)$. And to verify assumption 1b, note that $\sum_{s\in J_n, s\leq t\wedge T}\sum_i\int x^2I\{|x|>\epsilon\}\mathcal{F}_{n,i}(dx;s)\Lambda_n(\{s\})=$

$$\sum_{s\in J_n, s\leq t\wedge T} nJ(s)/Y.(s)(1-A(\{s\}))^2A(\{s\})I\{J(s)/Y.(s)(1-A(\{s\}))>\epsilon/\sqrt{n}\}+$$

$$\sum_{s\in J_n, s\leq t\wedge T} nJ(s)/Y.(s)A(\{s\})^2(1-A(\{s\}))I\{J(s)/Y.(s)A(\{s\})>\epsilon/\sqrt{n}\}.$$

Since $\int_0^{t\wedge T} nJ(s)/Y.(s)(1-A(\{s\}))dA(s)$ converges in probability to $\int_0^{t\wedge T}\big((1-F(t-)(1-G(t-)\big)^{-1}(1-A(\{s\}))dA(s)$ assumption 2a) holds. This is sufficient for finite dimensional convergence to a Gaussian martingale with covariance function, $\Sigma(t)=\int_0^{t\wedge T}\big((1-F(s-)(1-G(s-)\big)^{-1}(1-A(\{s\}))dA(s)$. Functional weak convergence using the Skorohod metric follows also but this problem is nice enough so that it is possible to prove functional convergence on the space of bounded functions on $[0,T]$, called $l^\infty[0,T]$. All that is necessary is to prove asymptotic tightness. It is sufficient to prove that for each ϵ, $\eta>0$ there exists a finite partition of $[0,T]$, say $0=t_1<,\ldots,<t_k=T$, such that $\limsup_n P\big(\max_i\sup_{t\in[t_i,t_{i+1})}|M_n(t)-M_n(t_i)|>\epsilon\big)<\eta$ (van der Vaart and Wellner, 1993). This is easily enough done by choosing the partition to contain the larger jump points of A and using Lenglart's inequality (see Andersen et.al., 1993). Gill (1980) proved a similar result for the Kaplan-Meier Estimator of F by inserting an interval at the jump points of A. This theorem yields a quick proof in that setting also.

3. *The Proportional Hazards Model.*

The marked point process, $\mathcal{N}_n$ is the composite of an n-variate multivariate counting process. The event times, T_j, are the ordered event times of the component counting processes and the mark at time T_j is simply $\boldsymbol{\delta}_j$ where $\delta_{ji}=1$ if the ith counting process jumps at time T_j and is zero otherwise. The ith counting process has intensity, $Y_i(t)e^{\beta^TZ_i(t)}\lambda(t)dt$, where the covariate process, Z_i is locally bounded and predictable and the baseline intensity, λ is locally integrable. The process, Y_i is also predictable and is the censoring process, that is, Y_i is one as long as it is possible to observe a jump of the ith process and zero thereafter. Since the components of a multivariate counting process can not have common jumps, the component counting processes are conditionally independent. This implies that the marked point process has compensator,

$$\mathbf{\Lambda}_n(d\boldsymbol{\delta}, dt) = I\{\delta. \geq 1\} \prod_{i=1}^{n} \Big(Y_i(t) e^{\beta^T Z_i(t)}\Big)^{\delta_i} \lambda(t)dt.$$

Note that $\mathbf{\Lambda}_n^{\{i\}}$ is the intensity of the ith counting process. The derivative with respect to β of the natural log of Cox's partial likelihood (Andersen et.al., 1993 pg. 483) is given by,

$$\int \sum_{i=1}^{n} \delta_i \Big(Z_i(t) - \frac{\sum_l Y_l(t) e^{\beta^T Z_l(t)} Z_l(t)}{\sum_l Y_l(t) e^{\beta^T Z_l(t)}} \Big) \mathcal{N}_n(d\boldsymbol{\delta}, dt).$$

The above is a local square integrable martingale, since the Z_i's are locally bounded. Define M_n by setting $H_i = \Big(Z_i(t) - \frac{\sum_l Y_l(t) e^{\beta^T Z_l(t)}}{\sum_l Y_l(t) e^{\beta^T Z_l(t)}}\Big)$. Since $\mathbf{\Lambda}_n$ is continuous, only assumptions 1) and 2a) need to be proved. These are the conditions of Rebolledo's theorem as stated in Andersen et. al. (1993, pg. 83). This model can be developed as in the previous example above (see Andersen et. al., 1993).

5. Appendix

Proof of Theorem 1

For simplicity, the proof for convergence of the finite dimensional distributions is given for a one dimensional local square integrable martingale, M_n, with $M_n(0) = 0$. The Cramér-Wold device can then be used to extend the result to higher dimensions. The superscripts c, d will denote the continuous, respectively discrete, parts of the martingale and compensators.

Intuitively the conditional distribution of $M_n(dt)$ given the past is a convolution of a continuous gaussian increment $M_n^c(dt)$ with mean zero and variance, $\langle M_n^c \rangle(dt)$ and a random variable $M_n^d(dt)$ which assumes the value $x - \int_y y\nu_n(dy, dt)$ according to, $\nu_n(dx, dt)$ and the value, $-\int_y y\nu_n(dy, dt)$ with probability, $(1 - \nu_n(\Re, dt)$. Since $M_n(0) = 0$ implies that $\int_x x\nu_n(dx, \{t\})$ is zero $a.s.$, $M_n^d(\{t\})$ will have conditional distribution $\nu_n(dx, \{t\}) + \epsilon_0(dx)(1 - \nu_n(\Re, \{t\}))$ (recall that ϵ_0 is a probability measure giving mass 1 to the point 0). The characteristic function (conditional on the past) of the random variable, $(M_n^d(dt))$, is given by, $e^{-iu\int_x x\nu_n(dx,dt)}\big(1 + \int e^{iux} - 1 \ \nu_n(dx, dt)\big)$, it's conditional mean is

$\int_x x\nu_n(dx, dt)$ and it's conditional variance is $\int_x x^2\nu_n(dx, dt)$. This intuition helps one understand why the following proof works.

Theorem VIII.1.18 on page 418 of Jacod and Shiryaev (1987) gives a product integral which acts much as a characteristic function when the limiting process is a process of independent increments (as is the case here). This product integral is the product of the conditional characteristic functions of $M_n(dt)$. Define the process of locally bounded variation, $A_n(t;u) = -1/2u^2\langle M_n^c\rangle(t) + \int_0^t \int e^{iux} - 1 - iux\ \nu_n(dx, ds)$. Then the product integral is

$$G_n(t;u) = \prod_{s\leq t}(1 + A_n(ds;u)).$$

Since A_n is one dimensional and $\int x\nu_n(dx, \{t\})$ is zero, G_n simplifies to

$$G_n(t;u) = \exp\Big\{ -1/2u^2\langle M_n^c\rangle(t)\Big\}\prod_{s\leq t}\left(1 + \int_x e^{iux} - 1\ \nu_n(dx, ds)\right)e^{-iu\int_x x\nu_n^c(dx,ds)}$$

(Andersen et. al. pg. 90, 1993). Theorem VIII.1.18 states that if $G_n(t;u)$ converges in probability to $\exp\{ -1/2u^2\Sigma(t)\}$ for all $u \in \Re$ and for all $t \in D$, then the finite dimensional distributions in D of M_n converge to the finite dimensional distributions in D of a Gaussian martingale with covariance function, Σ.

Set $\alpha_n = \exp\{ -1/2u^2\langle M_n^c\rangle(t) - 1/2u^2\int_0^t\int x^2\nu_n(dx, ds)\}$ and $\beta_n = \exp\{ \int_0^t\int e^{iux} - 1 - iux + 1/2u^2x^2\nu_n^c(dx, ds)\}$. Of course, α_n converges in probability to $\exp\{ -1/2u^2\Sigma(t)\}$ by assumption 2a) and in the next paragraph it is proved that β_n converges to one in probability. Consider

$$G_n(t;u) - \alpha_n\beta_n = \exp\left\{ -1/2u^2\langle M_n^c\rangle(t) + \int_0^t\int e^{iux} - 1 - iux\ \nu_n^c(dx, ds)\right\}$$

$$\left[\prod_{s\leq t}(1 + \int e^{iux} - 1\ \nu_n(dx, \{s\})) - \exp\left\{ -1/2u^2\sum_{s\leq t}\int x^2\nu_n(dx, \{s\})\right\}\right].$$

The term before the square brackets can be shown to be bounded in probability by using the results on α_n and β_n. All that needs to be done is to show that β_n goes to one in probability and that the above term in square brackets goes to zero in probability.

Since there is a constant, C for which $|e^{iux} - 1 - iux + u^2x^2/2| \leq C(|x|^3 \wedge x^2)$, the exponent in β_n can be bounded in absolute value by $C\int_0^t\int |x|^3 \wedge x^2\nu_n^c(dx, ds)$. For ϵ small,

$$\int_0^t\int |x|^3 \wedge x^2\nu_n^c(dx, ds) \leq \int_0^t\int x^2I\{|x| > \epsilon\}\nu_n^c(dx, ds) + \epsilon\int_0^t\int x^2I\{|x| \leq \epsilon\}\nu_n^c(dx, ds)$$

But this goes to zero in probability because of assumption 1a), assumption 2a) and the fact that ϵ may be taken arbitrarily small.

All that is left is to deal with the jumps of M_n. Intuitively, the term,

$$\prod_{s\le t}\big(1+\int e^{iux}-1\;\nu_n(dx,\{s\})\big)-\exp\Big\{-1/2u^2\sum_{s\le t}\int x^2\nu_n(dx,\{s\})\Big\}$$

is just the difference between the conditional characteristic function of the jumps and the characteristic function of a normal random variable with the same variance as the jump.

Both terms above can be written as two products, the first over the persistent jumps (in J_n) and the second over the asymptotically negligible jumps (jumps not in J_n). Since $A_1A_2-B_1B_2=(A_1-B_1)A_2+B_1(A_2-B_2)$ all that is necessary is to prove that

$$\prod_{s\le t,s\notin J_n}\big(1+\int e^{iux}-1\;\nu_n(dx,\{s\})\big)-\exp\Big\{-1/2u^2\sum_{s\le t,s\notin J_n}\int x^2\nu_n(dx,\{s\})\Big\}\xrightarrow{\mathcal{P}}0 \qquad (5)$$

and

$$\prod_{s\le t,s\in J_n}\big(1+\int e^{iux}-1\;\nu_n(dx,\{s\})\big)-\exp\Big\{-1/2u^2\sum_{s\le t,s\in J_n}\int x^2\nu_n(dx,\{s\})\Big\}\xrightarrow{\mathcal{P}}0. \qquad (6)$$

The assumptions, (1), 1b) and 2b), are only used in proving (6) for the persistent jumps.

First (5) is proven as it is the easier. Using Taylor series expansions, it is possible to prove that for any η,

$$|e^{iux}-1-iux+u^2x^2/2I\{|x|\le\eta\}|\le|ux|^3/6I\{|x|\le\eta\}|+(tx)^2/2I\{|x|>\eta\}|.$$

The notation to follow is made simpler by denoting $E_{n,s}$ as expectation with respect to the distribution, $\nu_n(dx,\{s\})+\epsilon_0(dx)(1-\nu_n(\Re,\{s\})$. Note that this distribution has mean zero and variance $\sigma_n^2(s)$. Let $\{\eta_n\}_{\eta\ge1}$ be a sequence converging so slowly to zero so that 1a) holds for η_n in place of ϵ. Using the last inequality given above, one gets that

$$E_{n,s}(e^{iux})-1+u^2/2E_{n,s}(x^2I\{|x|\le\eta_n\})=\psi_{n,s}\Big(\frac{|u|^3}{6}\eta_nE_{n,s}(x^2I\{|x|\le\eta_n\})+u^2/2E_{n,s}(x^2I\{|x|>\eta_n\})\Big)$$

for $\psi_{n,s}$ which is bounded in absolute value by one. Denote

$$-u^2/2E_{n,s}(x^2I\{|x|\le\eta_n\})+\psi_{n,s}\Big(|u|^3/6\eta_nE_{n,s}(x^2I\{|x|\le\eta_n\})+u^2/2E_{n,s}(x^2I\{|x|>\eta_n\})\Big)$$

by $\gamma_{n,s}$. These terms, $\gamma_{n,s},\ s \notin J_n$ possess the following nice properties,

1) the $\sup_{s\notin J_n,\ s\leq t} |\gamma_{n,s}|$ converges to zero in probability and

2) the $\sum_{s\notin J_n,\ s\leq t} |\gamma_{n,s}|$ is bounded in probability.

These two properties follow from assumptions 1a) and 2a). The proof of this follows the proof in Chung (1974, pg. 200).

Rewrite equation (5) as,

$$\left[\exp\left\{\sum_{s\notin J_n,\ s\leq t} \ln(E_{n,s}e^{iux}) + \frac{1}{2}u^2 \sum_{s\notin J_n,\ s\leq t} \sigma_n^2(s)\right\} - 1\right] e^{-1/2u^2 \sum_{s\notin J_n,\ s\leq t} \sigma_n^2(s)}.$$

Since $\sum_{s\notin J_n,\ s\leq t} \sigma_n^2(s)$ is bounded in probability, it is sufficient to show that,

$$\sum_{s\notin J_n,\ s\leq t} \ln(E_{n,s}e^{iux}) + \frac{1}{2}u^2 \sum_{s\notin J_n,\ s\leq t} \sigma_n^2(s) = \sum_{s\notin J_n,\ s\leq t} \ln(1+\gamma_{n,s}) + \frac{1}{2}u^2 \sum_{s\notin J_n,\ s\leq t} \sigma_n^2(s)$$

converges to zero in probability in order to finish the proof of (5). Since the $\sup_{s\notin J_n,\ s\leq t} |\gamma_{n,s}|$ converges to zero in probability, $|\ln(1+\gamma_{n,s}) - \gamma_{n,s}| = |\sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m!}\gamma_{n,s}^m|$ can be bounded above by $|\gamma_{n,s}|^2/2\sum_{m=2}^{\infty}(1/2)^{m-2} \leq |\gamma_{n,s}|^2$ on a set of probability going to one. Now,

$$\left|\sum_{s\notin J_n,\ s\leq t} \ln(1+\gamma_{n,s}) + \frac{1}{2}u^2 \sum_{s\notin J_n,\ s\leq t} \sigma_n^2(s)\right| \leq \left|\sum_{s\notin J_n,\ s\leq t} \ln(1+\gamma_{n,s}) - \gamma_{n,s}\right| + \left|\sum_{s\notin J_n,\ s\leq t} \gamma_{n,s} + \frac{1}{2}u^2\sigma_n^2(s)\right|.$$

But $|\sum_{s\notin J_n,\ s\leq t} \ln(1+\gamma_{n,s}) - \gamma_{n,s}| \leq \sum_{s\notin J_n,\ s\leq t} |\gamma_{n,s}|^2$ on a set of probability converging to 1 and therefore converges to zero in probability. Also $|\sum_{s\notin J_n,\ s\leq t} \gamma_{n,s} + \frac{1}{2}u^2\sigma_n^2(s)|$ can also be shown to converge to zero using assumptions 1a) and 2a).

Now condition (6) for the persistent jumps must be proved. Assumptions (1), 1b) and 2b) are sufficient to prove (6). As in the above paragraph one only need prove that $\sum_{s\in J_n,\ s\leq t} \ln[E_{n,s}(e^{iux})] + u^2/2\sigma_n^2(s)$ goes to zero in probability. Note that for $s \in J_n$, the expectation, $E_{n,s}$, is with respect to the distribution, $\int_\theta (\times_i \mathcal{F}_{n,i})(dx;\theta,s)\mathcal{G}_n(d\theta;s)\Lambda_n(\{s\}) + \epsilon_0(dx)(1-\Lambda_n(\{s\}))$ and $\sigma_n^2(s) = \int \sum_i \int x^2 \mathcal{F}_{n,i}(dx;\theta,s)\mathcal{G}_n(d\theta;s)\Lambda_n(\{s\})$. Further, recall that, $\sum_i \int x\mathcal{F}_{n,i}(dx;\theta,s) = 0$ for almost all θ. If necessary, redefine the $\mathcal{F}_{n,i}(\cdot;\theta,s)$ so that $\int x\mathcal{F}_{n,i}(dx;\theta,s) = 0$. Using the same Taylor series argument as in the proof of (5), one gets, $\int(e^{iux})\mathcal{F}_{n,i}(dx;\theta,s) - 1 = \gamma_{n,s}(\theta,i)$ for

$$\gamma_{n,s}(\theta,i) = -u^2/2\int(x^2 I\{|x|\leq\eta_n\})\mathcal{F}_{n,i}(dx;\theta,s) +$$

$$\psi_{n,s}^i\Big(\frac{|u|^3}{6}\eta_n \int x^2 I\{|x|\leq\eta_n\}\ \mathcal{F}_{n,i}(dx;\theta,s) + u^2/2\int(x^2 I\{|x|>\eta_n\}\ \mathcal{F}_{n,i}(dx;\theta,s)\Big)$$

where $\psi^i_{n,s}$ which is bounded in absolute value by one. If $\big(\prod_i \mathcal{F}_{n,i}\big)\mathcal{G}_n$ can also be written as a product distribution then the above integrals should be with respect to the ith distribution in the product. Or if the expectation, $E_{n,s}$ can be written as an expectation with respect to a product distribution, then the integrals in the Taylor series should be with respect to each distribution in the product. Then similar steps to the following can be taken to prove that $\sum_{s\in J_n,\ s\leq t} |\ln[e^{u^2/2\sigma_n^2(s)}E_{n,s}(e^{iux})]|$ goes to zero in probability. Indeed if $E_{n,s}$ can be written as an expectation with respect to a product distribution, then each of the integrals in the above expansion should be replaced by the each distribution in the product and then assumptions 2b) and (*) are no longer necessary.

All that is necessary to finish the proof is to show that $\sum_{s\in J_n,\ s\leq t} |e^{u^2/2\sigma_n^2(s)}E_{n,s}(e^{iux}) - 1|$ goes to zero in probability. This is the case, since on $\sum_{s\in J_n,\ s\leq t} |e^{u^2/2\sigma_n^2(s)}E_{n,s}(e^{iux}) - 1| < 1/2$, $\sum_{s\in J_n,\ s\leq t} |\ln[e^{u^2/2\sigma_n^2(s)}E_{n,s}(e^{-ux})] - 1|$ is bounded above by $\sum_{s\in J_n,\ s\leq t} |e^{u^2/2\sigma_n^2(s)}E_{n,s}(e^{iux}) - 1|$. This in turn is bounded above by,

$$\sum_{s\in J_n} \Big| \int \exp\Big\{ \sum_i \ln(1+\gamma_{n,s}(\theta,i)) + 1/2u^2\sigma_n^2(s)\Big\} - 1 \ \mathcal{G}_n(d\theta;s)\Lambda_n(\{s\})\Big| +$$

$$\sum_{s\in J_n,\ s\leq t} |\Lambda_n(\{s\}) - 1|. \tag{7}$$

The last term goes to zero in probability by assumption (*). If $|e^{g_1(v)+g_2(v)} - 1| \leq C$ for all v, A is a set and L is a positive measure, then the integral,

$$\Big| \int e^{g_1(v)+g_2(v)} - 1 \ L(dv)\Big| \leq C * L\Big(|g_1(v)|I\{v\in A\} > \epsilon \ \cup |g_2(v)| > \epsilon \ \cup A^c\Big) +$$

$$2\int |g_1(v)I\{v\in A\} + g_2(v)| I\Big\{|g_1(v)|I\{v\in A\} < \epsilon \ \cap \ |g_2(v)| < \epsilon \ \cap \ A\Big\} L(dv)$$

for ϵ sufficiently small. This in turn is bounded above by

$$C/\epsilon \int |g_1(v)I\{v\in A\}| L(dv) + C/\epsilon \int |g_2(v)| L(dv) +$$

$$CL(A^c)) + 2\int |g_1(v)I\{v\in A\}| L(dv) + 2\int |g_2(v)| L(dv).$$

To apply this inequality to the first term in (7), let ϵ' be arbitrarily greater than zero and set $g_1 = [\sum_i \ln(1+\gamma_{n,s}(\theta,i)) - \sum_i \gamma_{n,s}(\theta,i)]$, $A = \{\sup_i |\gamma_{n,s}(\theta,i))| \leq \epsilon'\}$, $L(d\theta) = \mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$, $C =$

$2e^{u^2/2\sigma_n^2(s)}$ and $g_2 = \sum_i \gamma_{n,s}(\theta, i) + u^2/2\sigma_n^2(s)$. To finish the proof the following must converge to zero in probability;

1) $\sum_{s\in J_n,\ s\leq t} \int [\sum_i |\ln(1+\gamma_{n,s}(\theta,i)) - \sum_i \gamma_{n,s}(\theta,i)] I\{\sup_i |\gamma_{n,s}(\theta,i))| \leq \epsilon'\} \mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$,

2) $\sum_{s\in J_n,\ s\leq t} \int [|\sum_i \gamma_{n,s}(\theta,i) + u^2/2\sigma_n^2(\{s\})|\ \mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$ and

3) $\sum_{s\in J_n,\ s\leq t} \mathcal{G}_n\big(\{\sup_i |\gamma_{n,s}(\theta,i)| > \epsilon'\}, s\big)\Lambda_n(\{s\})$.

First consider 1). This is bounded above by $\sum_{s\in J_n,\ s\leq t} \int [\sum_i |\gamma_{n,s}(\theta,i)|^2 I\{\sup_i |\gamma_{n,s}(\theta,i))| \leq \epsilon'\} \mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$ which is equal to $O_p(1)\epsilon' \sum_{s\in J_n,\ s\leq t} \sigma_n^2(s)$. Since $\sum_{s\in J_n,\ s\leq t} \sigma_n^2(s)$ is bounded in probability as n increases, 1) can be made as small as desired by reducing the value of ϵ'. Next using the definition of $\gamma_{n,s}(\theta,i)$ note that 2) is bounded above by the sum of $u^2/2 \sum_{s\in J_n,\ s\leq t} \int [|\sum_i \int x^2 \mathcal{F}_{n,i}(dx;\theta,s) - \sigma_n^2(s)| \mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$ and $u^2/2 \sum_{s\in J_n,\ s\leq t} \int [\sum_i \int x^2 I|\{|x| > \eta_n\} \mathcal{F}_{n,i}(dx;\theta,s)\mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$ and $O_p(1)\eta_n \sum_{s\in J_n,\ s\leq t} \sigma_n^2(s)$. All three terms converge to zero in probability. All that is left is 3). It is bounded above by a constant times $1/\epsilon' \sum_{s\in J_n,\ s\leq t} \int [\sum_i \int x^2 I|\{|x| > \eta_n\} \mathcal{F}_{n,i}(dx;\theta,s)\mathcal{G}_n(d\theta,s)\Lambda_n(\{s\})$ plus a constant times $(1/\epsilon')\eta_n^2 \sum_{s\in J_n,\ s\leq t} \sigma_n^2(s)$. These two terms also converge to zero in probability.

To prove tightness in the Skorohod metric, assume that D is dense on $\Re^d$. Theorem VI.5.17 of Jacod and Shiryaev, (1987) implies that it is sufficient to show that the predictable variation of M_n converges in the Skorohod metric. That is, there exists a nondecreasing deterministic function, Σ, for which,

1) $\langle M_n\rangle_{jj}(t) \xrightarrow{\mathcal{P}} \Sigma_{jj}(t)$ for $j = 1, \ldots, d$ and

2) $\sum_{j=1}^d \sum_{s\leq t} \big(\triangle\langle M_n\rangle_{jj}(s)\big)^2 \xrightarrow{\mathcal{P}} \sum_{j=1}^d \sum_{s\leq t} \big(\triangle\Sigma_{jj}(s)\big)^2$,

for all $t \in D$. The predictable variation of M_n is given by $\langle M_n\rangle(\cdot) = \langle M_n^c\rangle(\cdot) + \int_0^{\cdot} \int xx^T \nu_n(dx, ds)$. So 1) is assumption 2a) and 2) is assumption 3).

To prove tightness in $l^\infty([0,\tau])$, we must prove that for any $\epsilon, \eta > 0$ there exists a partition, $0 = t_1 < t_2 <, \ldots, < t_k = \tau$, for which

$$\limsup_n P(\sup_i \sup_{t\in[t_i,t_{i+1})} |M_n(t) - M_n(t_i)| > \epsilon) < \eta.$$

For a reference see van der Vaart and Wellner (1993, chapter 1, section 4). Choose the partition to contain all of the (finitely many) jumps of Σ which are of size greater than $\epsilon/4$. In addition choose the partition fine enough so that $\Sigma(t_{i+1}-) - \Sigma(t_i) < \epsilon/4$. To show that $\sum_i P(\sup_{t\in[t_i,t_{i+1})} |M_n(t) - M_n(t_i)| >$

$\epsilon) < \eta$, employ Lenglart's inequality as proved in Jacod and Shiryaev (1987, pg. 35). Assumption $3'$ is used in proving that $\langle M_n \rangle$ converges uniformly to Σ.

Proof of Corollary 1

Since (4) is assumed, 2) and 3) of theorem 1 translate directly into the assumptions 2) and 3) of this corollary. The equation in assumption 1b) of theorem 1 is

$$n^{-1} \sum_{i=1}^{n} \sum_{s \in J_n, s \leq t} \int H_i(x; z, \psi_s^i, s) H_i(x; z, \psi_s^i, s)^T I \left\{ \left| \left| H_i(x; z, \psi_s^i, s) \right| \right| > \epsilon\sqrt{n} \right\} F_{n,i}(dx; z, s) \mathcal{G}_n{}^{\{i\}}(dz; s) \Lambda_n(\{s\}).$$

Since there are no non-persistent jumps, the equation in assumption 1a) is

$$\int_0^t \int x x^T I\{||x|| > \epsilon\} \nu_n^c(dx, ds).$$

The above can be expressed in terms of $\mathbf{\Lambda}_n$;

$$\int_0^t \int n^{-1} \left(\sum_{i=1}^{n} H(x_i; z_i, \psi_s^i, s)\delta_i \right) \left(\sum_{i=1}^{n} H(x_i; z_i, \psi_s^i, s)\delta_i \right)^T I\{|| \sum_{i=1}^{n} H(x_i; z_i, \psi_s^i, s)\delta_i|| > \epsilon\sqrt{n}\} \mathbf{\Lambda}_n(d\boldsymbol{x}, d\boldsymbol{z}, d\boldsymbol{\delta}, ds).$$

Recall that no two of the δ_i's can be one at the continuity points of $\mathbf{\Lambda}_n$, so that the summation over i comes out of the square term and the indicator term, resulting in assumption 1) of the corollary.

5. References

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993). *Statistical models based on counting processes.* Springer Verlag, New York.

Arjas, E. and P. Haara (1992). Observation scheme and likelihood. *Scand. J. Statist.* **19**, 111-132.

Chung,K.L. (1974). *A course in probability theory.* Academic Press, Inc., San Diego.

Gill, R.D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tract 124. Mathematisch Centrum, Amsterdam.

Grüger, J., R. Kay, and M. Schumacher (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics* **47**, 595-605.

Jacod, J. and A.N. Shiryaev (1987). *Limit theorems for stochastic processes.* Springer Verlag, New York.

Liptser, R.S. and A.N. Shiryaev (1983). On the invariance principle for semi-martingales: the "nonclassical" case. *Theory Prob. Appl.* **28**, 1-34.

Murphy, S.A. and B. Li (1993). Projected partial likelihood and its application to the longitudinal data. to appear *in Biometrika.*

Scheike, T.H. (1994). Parametric regression for longitudinal data with counting process measurement times. *Scand. J. Statist.* **21**, 245-264.

van der Vaart, A.W. and J.A. Wellner (1993). *Weak convergence and empirical processes.* IMS Lecture Notes-Monograph Series (to appear).

Susan A. Murphy

Department of Statistics, Pennsylvania State University,

326 Classroom Building, University Park, PA 16802, U.S.A.