Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

K.O. Dzhaparidze

On asymptotic inference about intensity parameters of a counting process

Department of Mathematical Statistics Report MS-R8505 July

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

On asymptotic inference about intensity parameters of a counting process

Kacha Dzhaparidze
Centre for Mathematics and Computer Science, Amsterdam

The Cox regression model may be viewed as a special case (see (1.3)) of the general model described in this paper via the pair $(\bar{A}_t, \Psi_t)$ of predictable characteristics of an r-variate counting process $\mathbf{N}_t = (N_t^1, \cdots, N_t^r)$, associated with its compensator $\mathbf{A}_t = (A_t^1, \cdots, A_t^r)$ as follows: $\bar{A}_t = A_t^1 + \cdots + A_t^r$ and $\Psi_t = d\mathbf{A}/d\bar{A}_t$. It is supposed that the latter characteristic involves the real valued parameter β, i.e. $\Psi_t = \Psi_t(\beta)$, to be estimated by means of a given sample path of $\{\mathbf{N}_t, 0 \leqslant t \leqslant 1\}$. Treating this problem in its asymptotic setting, we consider our experiment (2.1) as n-th in a sequence of experiments, and let $\bar{A}_t$ meet Condition I of asymptotic stability. Under this and certain additional conditions introduced on demand, we study asymptotic properties of the estimator $\hat{\beta}$ for β defined by (1.4), which is in fact the Cox estimator extended to our situation. In particular, we characterize the consistency and asymptotic normality of $\hat{\beta}$ by estimating the probability of large deviations, and then showing the convergence in all moments of the distribution of $\hat{\beta}$ to a normal law. Finally, it is shown that $\hat{\beta}$ is the best within a class of (regular) estimators in the sense that neither of them can have an asymptotic distribution that is less spread out than that of $\hat{\beta}$.

1980 Mathematics Subject Classification : 62F12, 62G05, 62M99

Keywords & phrases : Cox's Regression Model, Multivariate counting process, compensator, parameter estimation.

Note: This is an invited paper at the Centenary Session of the International Statistical Institute, Amsterdam, 12-22 August 1985, and will appear in *Bull. Int. Stat. Inst. 51.*

Report MS-R8505
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. INTRODUCTION

1. Statistical inference on counting processes attracts considerable attention in the literature of recent years; see Bibliography where a number of related references is enclosed which may serve as a source for many further references. Typically, the approach taken in these works is inspired by the developments of the theory of stochastic processes related to the notion of martingales, see, e.g. Shiryayev (1981), as well as by the developments of the asymptotic theory of statistical decisions, see, e.g. Le Cam (1969) or Ibragimov/Has'minskii (1981); also Greenwood/Shiryayev (1985).

Within the framework of the theory of stochastic processes, these processes are defined on a complete probability space $(\Omega,\mathcal{F},P)$ equipped with a nondecreasing family $\{\mathcal{F}_t,t\geqslant 0\}$ of right-continuous sub-σ-algebras of $\mathcal{F}$ augmented by sets from $\mathcal{F}$ of zero probability. For the sake of simplicity, we discuss only the case in which $t\in[0,1]$.

Let $\mathbb{N} = (\mathbb{N}_t,\mathcal{F}_t,P)$ be r-variate counting process which by definition consists of components $N^i, i=1,...,r$ having stepwise sample paths: $N_0^i = 0$, $N_t^i-N_{t-}^i = \Delta N_t^i = 0$ or 1, $\Delta N_t^i\Delta N_t^j = 0$ if $i\neq j$ (no two component processes jumping at the same time), and $N_1^i<\infty$ P - a.s. With $\mathbb{N}$ one may associate an r-variate predictable increasing process $\mathbb{A} = \{\mathbb{A}_t,\mathcal{F}_t,P\}$ such that $\mathbb{N}-\mathbb{A}\equiv\mathbb{M} = \{\mathbb{M}_t,\mathcal{F}_t, P\}$ is an r-variate local square integrable martingale with the quadratic characteristic $<\mathbb{M}>_t = diag\mathbb{A}_t-[\mathbb{A}]_t$ (see Lemma 3.1).

If, in addition, the filtration is of special form $\mathcal{F}_t=\sigma\{\omega:\mathbb{N}_s,s\leqslant t\}$ then the probability measure P is completely defined by the compensator $\mathbb{A}$ (in the sense of Liptser/Shiryayev (1978), Section 18.3). Hence in this case the statistical model for the observed phenomena may be completely specified in terms of the compensator $\mathbb{A}_t$ or, for convenience, in terms of the so-called $(P,\mathcal{F}_t)$- predictable characteristics $(\overline{A}_t,\Psi_t)$ of $\mathbb{N}_t$, associated with $\mathbb{A}_t$ by the following relations $\Psi_t = d\mathbb{A}_t/d\overline{A}_t$ and $\overline{A}_t = \mathbb{1}_r^T\mathbb{A}_t$ (here $\mathbb{1}_r = col(1,...,1)$, and T indicates the transposition). Obviously, the first of these characteristics is the compensator of $\overline{N}_t=N_t^1+...+N_t^r$, while the r-variate nonnegative predictable process $\Psi=\{\Psi_t,\mathcal{F}_t,P\}$ consists of components $\Psi_t^i, i=1,...,r, \Psi_t^i$ being, roughly speaking, the probability of having a jump of N_t^i at time t, given $\mathcal{F}_{t-}$ and given that $\overline{N}_t$ jumps at time t; Brémaud (1981), pp. 34 and 236.

2. In applications the latter characteristic is usually parametrized: it is restricted to a certain parametric family $\Psi\in\{\Psi(\beta),\beta\in\mathcal{B}\}$ of nonnegative $\mathcal{F}_t$-predictable processes for each admisible value of the parameter $\beta\in\mathcal{B}$.

In such a case β is "the parameter of interest" - inference about β is required by means of a given sample path of $\{\mathbb{N}_t,0\leqslant t\leqslant 1\}$ drawn according to the pair $(\overline{A}_t,\Psi_t(\beta))$ of the characteristics of $\mathbb{N}$ for an unknown β and, typically, for the characteristic $\overline{A}_t$ specified only up to the restrictions of a general nature (to be introduced below). Actually, $\overline{A}_t$ itself may depend on the parameter of interest β, as well as on certain nuisance quantities, as it is illustrated by the following

Example 1.1. Let $\{P_{\alpha,\beta},\alpha\in\mathcal{Q}, \beta\in\mathcal{B}\}$ be a family of the probability measures, where $\mathcal{Q}$ is a set of deterministic nonnegative and nondecreasing functions $\alpha = \alpha_t$, $0\leqslant t\leqslant 1$, and $\mathcal{B}$ an open set of R^1. For each $\alpha\in\mathcal{Q}$ and $\beta\in\mathcal{B}$ let $\mathbb{N} = (\mathbb{N}_t,\mathcal{F}_t,P_{\alpha,\beta})$ be an r-variate counting process of the Poisson type (Liptser/Shiryayev (1978), p. 249) defined on the stochastic basis $(\Omega,\mathcal{F},\{\mathcal{F}_t, 0\leqslant t\leqslant 1\},P_{\alpha,\beta})$, with the compensator of form

$$\mathbb{A}_t = \mathbb{A}_t(\alpha,\beta) = \int_0^t\Phi_s(\beta)d\alpha_s,\ 0\leqslant t\leqslant 1 \tag{1.1}$$

where $\Phi(\beta)$ is an r-variate nonnegative $\mathcal{F}_t$-predictable process for each $\beta\in\mathcal{B}$. Obviously, the pair of the $(P_{\alpha,\beta},\mathcal{F}_t)$- predictable characteristics of the process $\mathbb{N}$ is given by the following relations

$$\Psi_t(\beta) = \Phi_t(\beta)\,/\,\overline{\Phi}_t(\beta)\quad\text{with}\quad\overline{\Phi}_t(\beta) = \mathbb{1}_r^T\Phi_t(\beta),\text{ and }\overline{A}_t(\alpha,\beta) = \int_0^t\overline{\Phi}_s(\beta)d\alpha_s \tag{1.2}$$

The most popular special case of the Cox regression model for censored survival data specifies these characteristics as follows:

$$\Psi_t^i(\beta) = Y_t^i e^{\beta Z_t^i} / \sum_{i=1}^{r} Y_t^i e^{\beta Z_t^i},\ \bar{A}_t(\alpha,\beta) = \int_0^t \sum_{i=1}^{r} Y_s^i e^{\beta Z_s^i} d\alpha_s \tag{1.3}$$

with certain $\mathcal{F}_t$-predictable processes Y_t^i and Z_t^i, free from β; see, e.g., Andersen/Gill (1982) (or, for a bit more general model, Prentice/Self (1983)). These authors and later Begun et.al. (1983) have shown that under wide conditions the particular estimator $\hat{\beta}$ for β, defined by the relation

$$\sup_{\beta\in\mathfrak{B}} \int_0^1 \ln^T \Psi_s(\beta) d\mathbb{N}_s = \int_0^1 \ln^T \Psi_s(\hat{\beta}) d\mathbb{N}_s,\ \ \ln\Psi = col\{\ln\Psi^i, i=1,...,r\} \tag{1.4}$$

possesses the desired asymptotic properties (to be specified in the next section)

Obviously, if $\Psi_s(\beta)$ is a sufficiently smooth function of β, then the estimator $\hat{\beta}$ of β is well defined by condition (1.4) also for the general set up discussed at the beginning of this subsection (and not only for the special Cox model; see (1.3)). Naturally, one can expect that under circumstances similar to those of the papers mentioned above, the estimator $\hat{\beta}$ preserves its desired properties. In the present paper this conjecture is confirmed, furthermore, a refined characterization of these properties is given (cp. Efron (1977)).

Note that unlike Andersen/Gill (1982) here only the case of the real valued parameter β is duscussed, and the abstract parameter α in (1.3) (or (1.2)) is considered as the nuisance quantity.

2. ASYMTOTIC INFERENCE

1. Following the usual device of the asymtotic theory (Le Cam (1969), Ibragimov/Has'minskii (1981)), we suppose that observed is an outcome of the experiment

$$\mathfrak{E}_n = (\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n, 0\leqslant t\leqslant 1\}, \{P^n\}) \tag{2.1}$$

(with a certain family of probability measures $\{P^n\}$), which is actually n-th in the sequence of experiments $\mathfrak{E}_1, \mathfrak{E}_2, \ldots$. Fix P^n at the right-hand side of (2.1), and define on that stochastic basis an r_n-variate counting process $\mathbb{N}^n = \{\mathbb{N}_t^n, \mathcal{F}_t^n, P^n\}$ where $r_n, n = 1,2,\ldots$ is a nondecreasing sequence of integers. As above, with the compensator $\mathbb{A}^n = \{\mathbb{A}_t^n, \mathcal{F}_t^n, P^n\}$ of $\mathbb{N}^n$ relate the pair $(\bar{A}_t^n, \Psi_t^n)$ of the $(P^n, \mathcal{F}_t^n)$-predictable characteristics, and let Ψ_t^n depend on $\beta\in\mathfrak{B}$.

The class of all admissible pairs $(\bar{A}_t^n, \Psi_t^n(\beta))$ of the $(P^n, \mathcal{F}_t^n)$-predictable characteristics of $\mathbb{N}^n$ determines the family of the probability measures $\{P^n\}$ in (2.1). The following basical condition restricts this class up to an asymptotic stability requirement on the sequence $F_t^n \equiv \bar{A}_t^n / k_n = \mathbb{1}_{r_n}^T \mathbb{A}_t^n / k_n$ with an unboundedly increasing sequence of numbers k_n, $n = 1,2,\ldots$.

Condition I. For each admissible pair $(\bar{A}_t^n, \Psi_t^n(\beta))$ of the $(P^n, \mathcal{F}_t^n)$-predictable characteristics there exists a continuous deterministic function of bounded variation F_t such that $F_t^n \to F_t$ in P^n probability as $n\to\infty$, each $t, 0\leqslant t\leqslant 1$.

Remark 2.1. In fact, by lemma 1 of McLeish (1978), p. 146 the continuity of F_t implies $\sup_{0\leqslant t\leqslant 1} |F_t^n - F_t| \to 0$ in P^n probability as $n\to\infty$

Define now the estimator $\hat{\beta}_n$ for β by condition (1.4) with $\mathbb{N} = \mathbb{N}^n$ and $\Psi = \Psi^n$. On deriving asymptotic (as $n\to\infty$) properties of $\hat{\beta}_n$, we require some regularity conditions on $\Psi^n(\beta)$; see Conditions II-IV in Section 4.

Condition II requires differentiability (in a certain sense) of $\sqrt{\Psi^n(\beta)} = col\{\sqrt{\Psi^{in}}, i=1,\ldots,r_n\}$ and existance of a positive number ν - the limit of $\int_0^1 |(\partial/\partial\beta)\sqrt{\Psi^n(\beta)}|^2 dF^n$ in P^n probability. Condition III (of the Lindeberg type), together with Condition II, leads to the conclusion of Corollary 4.1 needed for deriving asymptotic normality $N(0,1/4\nu)$ of the estimator $\hat{\beta}_n$.

As for Condition IV, it permits us (via Lemma 4.1) to apply a generalized version of Theorem 5.1 (Ibragimov/Has'minskii (1981), Section 1.5: Inequalities for Probabilities of Large Deviations) due to Sieders (1985), the conclusion of which can be informally described as follows: Let an estimator $\hat{\beta}_n$ for β be defined by maximizing with respect to β a certain functional of observations (e.g., the likelihood function). If this functional satisfies certain conditions, similar to the conditions imposed on the likelihood function in the above mentioned Theorem 5.1, then the estimator $\hat{\beta}_n$ is not only consistent (in P^n probability), but also the following holds: for sufficiently large values of n the P^n-probability that $2\sqrt{k_n\nu}(\hat{\beta}_n-\beta)$ exceeds in absolute value a (sufficiently large) number H is less then $C_0\exp-c_0H^2$, with some positive constants c_0 and C_0.

Hence, this way we get the first main result of Section 4 - the refinement of consistency of the estimator $\hat{\beta}_n$ (Proposition 4.1).

The second main result concerns the refinement of asymptotic normality of $\hat{\beta}_n$ based on a generalization of Theorem 10.1 (Ibragimov/Has'minskii (1981), p. 103): if the generalized version of Theorem 5.1 holds (Sieders (1985)), as well as Corollary 4.1 and Lemma 4.2, then all moments of $2\sqrt{k_n\nu}(\hat{\beta}_n-\beta)$ converge to the corresponding moments of the standard normal distribution (Propostion 4.2).

2. On discussing optimality properties of the estimator $\hat{\beta}_n$ in Section 5, we restrict our considerations to the processes of the Poisson type; see Example 1.1 in which all the introduced quantities are indexed now by n, except the parameters α and β, of course.

In the first place we show the LAN property of the family $\{P^n_{\alpha,\beta}, \alpha\in\mathcal{Q}, \beta\in\mathcal{B}\}$ of the probability measures defined on $\{\Omega^n,\mathcal{F}^n\}$; see Definition 4.1. Along with $P^n = P^n_{\alpha,\beta}$, let the probability measure $\underline{P}^n = P^n_{\underline{\alpha}^n,\underline{\beta}^n}$ be defined on $(\Omega^n,\mathcal{F}^n)$, where $\underline{\beta}^n = \beta+b/\sqrt{k_n}\in\mathcal{B}$, $b\in R^1$ and $\underline{\alpha}^n\in\mathcal{Q}$ is defined by the relation $\sqrt{d\underline{\alpha}^n_t/d\alpha_t} = 1 + a_t/\sqrt{k_n}$, $a_t\in L^2(dF)$ with $F_t=F_t(\alpha,\beta)$ of Condition VI.2. Then $\underline{P}^n \ll P^n$, and $d\underline{P}^n/dP^n$ is given by (5.1). The above mentioned LAN property is stated in Proposition 5.1 which tells us that under the Conditions V-VII the logarithm of $d\underline{P}^n/dP^n$ is in fact asymptotically quadratic with the asymptotically normal linear term $\delta^n_{\alpha,\beta}(a,b)$, and the quadratic term $-\frac{1}{2}g_{\alpha,\beta}(a,b)$ where $g_{\alpha,\beta}(a,b)$ is the limit in P^n probability of the quadratic characteristic of $\delta^n_{\alpha,\beta}(a,b)$.

Condition V.I requires continuous differentiability of $\sqrt{\Phi^n(\beta)}$ (in certain sense), and Condition V.2, together with Condition VI, determines the form of $g_{\alpha,\beta}(a,b)$. Condition VII (of the Lindeberg type) ensures the asymptotic normality of the linear term $\delta^n_{\alpha,\beta}(a,b)$.

Having the LAN, one can take advantage of its fairly general implications due to Le Cam and Hájek (see, e.g., Ibragimov/Has'minskii (1981), Ch. II and III, or Millar (1983)). Specifically, our conclusions about asymptotic optimality properties of the estimator $\hat{\beta}_n$ are based on the application of Hájek's convolution theorem to the situation under consideration (see Theorem 5.1).

For these purposes, define first the class of regular estimators $\{\beta^n_R\}$ for β. Under the conditions ensuring the LAN property of the family $\{P^n_{\alpha,\beta},\alpha\in\mathcal{Q},\beta\in\mathcal{B}\}$, at "point" $\alpha\in\mathcal{Q},\beta\in\mathcal{B}$ (Proposition 5.1), *the estimator* β^n_R *is called regular* (at the point $\alpha\in\mathcal{Q},\beta\in\mathcal{B}$) if for some nondegenerate distribution function G^0_R the following weak convergence takes place:

$$\mathcal{L}\{\sqrt{k_n}(\beta^n_R-\underline{\beta}^n)|\underline{P}^n\} \Rightarrow G^0_R \tag{2.2}$$

uniformly for each $|b|<c$ whatever $c>0$, and each bounded $a\in L^2(dF)(\underline{\alpha}^n$, $\underline{\beta}^n$ and $\underline{P}^n$ being defined as above).

Now, Hájek's convolution theorem (Begun et al. (1983)) tells us that G^0_R at the right-hand side of (2.2) can be represented as the convolution of a certain normal law with another distribution law, G^1_R say. By Proposition 5.2, in our special case $G^0_R = N(0,1/4\nu)*G^1_R$ where $N(0,1/4\nu)$ coincides with the asymptotic distribution of $\sqrt{k_n}(\hat{\beta}_n-\beta)$; see the previous subsection.

Since convolution "spreads out mass", no regular estimator β_R^n can have an asymptotic distribution that is less spread out then $N(0,1/4v)$. Thus, in this sense the estimator $\hat{\beta}_n$ (which is regular under the conditions of the previous subsection; see Theorem 5.1) is best within the class $\{\beta_R^n\}$.

The proof of the results just mentioned uses the fact that the neighborhood about α that shrinks at rate $k_n^{-\frac{1}{2}}$ in the directions $\{a\}$, defined above, is "sufficiently fat" to include the function $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)}$ where $\phi_t(\beta)$ is the bounded limit in $P_{\alpha,\beta}^n$ probability of $\overline{\Phi}_t^n(\beta)/k_n$ (Condition VI). Simply, the variety $\{a\}$ includes $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)}$; see Proposition 5.2.

3. This inclusion typically fails in situations in which $\{a\}$ is a low dimentional subspace of $L^2(dF)$, namely, in the frequently encountered situations in which "the cumulative hazard function" α_t is also parametrized up to a certain number of nuisance parameters, and hence $\{a\}$ is taken as a linear subspace, $\mathrm{A}=\mathrm{A}(\alpha)$ say, spanned by the logarithmic derivatives of the density of α_t with respect to the nuisance parameters; see, e.g., Efron (1977), Jarupskin (1983), Borgan (1984), Hjort (1984). According to these works the following conclusions can be drawn about the maximum likelihood estimator β_{ML}^n for β, defined by maximizing the likelihood function (see (5.1)) simultaneously with respect to the parameter of interest β and the nuisance parameters.

Under certain regularity conditions $\mathcal{L}\{\sqrt{k_n}(\beta_{ML}^n-\beta^n)|\underline{P}^n\}\Rightarrow N\ (0,\mathcal{I}_{\alpha,\beta}^{-1})$ with $\mathcal{I}_{\alpha,\beta}$ defined as in (5.3), and this means that no $R^1\times\mathrm{A}$-*regular* estimator β_{RA}^n can have an asymptotic distribution less spread then that of β_{ML}^n. In fact, *the estimator* β_{RA}^n *is called* $R^1\times\mathrm{A}$-*regular* if for some nondegenerate distribution function G_{RA}^0 $\mathcal{L}\{\sqrt{k_n}(\beta_{RA}^n-\beta^n)|\underline{P}^n\}\Rightarrow G_{RA}^0$ for each $b\in R^1, a\in\mathrm{A}$, whereas Proposition 5.2 tells us that G_{RA}^0 may be represented as the convolution (5.3). In particular, β_{ML}^n is less dispersed then the estimator $\hat{\beta}_n$, for comparing their variances we have $\mathcal{I}^{-1}\leqslant(4v)^{-1}$ with equality iff $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)}\in\mathrm{A}$ (see Remark 5.3).

It is important, however, that there is a subclass of estimators for β whithin which no estimator has a less spread asymptotic distribution then $\hat{\beta}_n$ defined by (1.4). This is the subclass $\{\beta_R^n\}\subset\{\beta_{RA}^n\}$ of regular estimators defined as in the previous subsection by the condition: whatever the (bounded) direction $a\in L^2(dF)$ of approach to α there is some nondegenerate distribution function G_R^0 such that (2.2) takes place. Of course, β_{ML}^n is not regular in this sense, as for $a\notin A$ a bias appears in its limiting distribution. While the estimator $\hat{\beta}_n$ is regular, and it is the best among $\{\beta_R^n\}$ since by Proposition 5.3 G_R^0 may be always represented as the convolution $G_R^0=N(0,1/4v)*G_R^1$ (Theorem 5.1).

3. THE LIKELIHOOD RATIO

1. Let $(\Omega,\mathcal{F}_t,P)$ be a complete probability space with a filtration $\{\mathcal{F}_t,0\leqslant t\leqslant 1\}$ satisfying the usual conditions. Let $\mathbb{N}=\{\mathbb{N}_t,\mathcal{F}_t,P;\ 0\leqslant t\leqslant 1\}$ be a multivariate (r-variate) counting process: $\mathbb{N}=col\{N^1,...,N^r\}$. Consider its Doob-Meyer decomposition $\mathbb{N}=\mathbb{M}+\mathbb{A}$ where $\mathbb{M}=\{\mathbb{M}_t,\mathcal{F}_t,P;\ 0\leqslant t\leqslant 1\}$ is a local square integrable martingale, and $\mathbb{A}=\{\mathbb{A}_t,\mathcal{F}_t,P;\ 0\leqslant t\leqslant 1\}$ a predictable compensator.

Lemma 3.1. The quadratic variation and quadratic characteristic of $\mathbb{M}$ are given by the following relations:

1) $[\mathbb{M}]=diag\mathbb{N}-[\mathbb{A}]-[\mathbb{M},\mathbb{A}]-[\mathbb{A},\mathbb{M}]$

2) $<\mathbb{M}>=diag\mathbb{A}-[\mathbb{A}]$

Proof. By definition $[\mathbb{N}]=diag\mathbb{N}$, and this gives 1). To get 2) take the compensator of both sides of 1).

Remark 3.1. Denote $\overline{N}=N^1+...+N^r, \overline{N}=\overline{M}+\overline{A}$. From 2) follows

$$<\overline{M}>_t=\overline{A}_t-[\overline{A}]_t=\int_0^t(1-\Delta\overline{A})d\overline{A},\quad \Delta<\overline{M}>=(1-\Delta\overline{A})\Delta\overline{A},$$

hence $0\leqslant\Delta\overline{A}\leqslant 1$. For simplicity assume $\Delta\overline{A}<1$ (in fact one can easily dispence with this restriction; see e.g. Kabanov et al. (1975) or (1980)).

Remark 3.2. Consider $V_t = I_r-\Delta\mathbb{A}_t\otimes\mathbb{1}_r$ and $V_t^{-1} = I_r + (1-\Delta\overline{A}_t)^{-1}\Delta\mathbb{A}_t\otimes\mathbb{1}_r$ with $\mathbb{1}_r = col\{1,...,1\}$ and $I_r = diag\mathbb{1}_r$. Then $<\mathbb{M}>_t = \int_0^t V diag d\mathbb{A}$ and $diag\mathbb{A}_t = \int_0^t V^{-1}d<\mathbb{M}>$.

Lemma 3.2. Let $\mathcal{Q}_t = \int_0^t V^{-1}dA = \int_0^t(1-\Delta\overline{A})^{-1}d\mathbb{A}$ and $\mathfrak{M}_t = \int_0^t V^{-1}d\mathbb{M} = \mathbb{M}_t + [\mathcal{Q},\overline{M}]_t$. Then

1) $[\mathfrak{M}]_t = diag\mathbb{N}_t + \int_0^t(1-\Delta\overline{N})d[\mathcal{Q}]$,

2) $<\mathfrak{M}> = diag\mathbb{A} + [\mathcal{Q},\mathbb{A}]$.

Poof. As $\Delta\mathbb{N}^{\otimes 2} = diag\Delta\mathbb{N}$, $(1-\Delta\overline{N})^2 = (1-\Delta\overline{N})$ and $\Delta\mathbb{N}(1-\Delta\overline{N}) = 0$, 1) follows from

$$\Delta\mathfrak{M} = \Delta\mathbb{M} + \Delta\mathcal{Q}\Delta\overline{M} = \Delta\mathbb{N} - \Delta\mathcal{Q}(1-\Delta\overline{N}). \tag{3.1}$$

To get 2) take the compensators of both sides of 1).

2. Suppose that a probability measure $\underline{P}$ in addition to the probability measure P is given on a measurable space $(\Omega,\mathfrak{F})$ with a filtration of special form $\mathfrak{F}_t = \sigma\{\mathbb{N}_s:s\leqslant t\},0\leqslant t\leqslant 1$. Along with $\mathbb{N} = (\mathbb{N}_t,\mathfrak{F}_t,P)$, consider the counting process $\underline{\mathbb{N}} = (\mathbb{N}_t,\mathfrak{F}_t,\underline{P})$ with compensator $\underline{\mathbb{A}} = (\underline{\mathbb{A}}_t,\mathfrak{F}_t,\underline{P})$.

Theorem 3.1. (Kabanov et al. (1980)). 1) For absolute continuity of $\underline{P}$ with respect to $P(\underline{P}<<P)$ the following conditions are necessary and sufficient: $\underline{P}$ -a.s.

I $\Delta\underline{\overline{A}} = 1$ implies $\Delta\overline{A} = 1$.

II. The components $\underline{A}^i$ and $A^i, i=1,...,r$ of $\underline{\mathbb{A}}$ and $\mathbb{A}$ are related as $\underline{A}_t^i = \int_0^t\lambda^i dA^i$ where $col\{\lambda^1,\ldots,\lambda^r\} = \Lambda = \{\Lambda_t,\mathfrak{F}_t\}$ is a nonnegative predictable process such that the associate Hellinger process is bounded: $\mathcal{H}_t = \int_0^t\sum_{i=1}^r(\sqrt{d\underline{A}^i}-\sqrt{dA^i})^2 + \sum_{\substack{s\leqslant t\\ 0<\Delta A_s<1}}(\sqrt{1-\Delta\underline{\overline{A}}_s}-\sqrt{1-\Delta\overline{A}_s})^2<\infty$.

2) Assume $\underline{P}<<P$, and denote z_t a right-continuous modification of the martingale $E(d\underline{P}/dP|\mathfrak{F}_t)$ $0\leqslant t\leqslant 1$. Then $z_t = \exp\{m_t + \sum_{s\leqslant t}\Phi_1(\Delta m_s)\}$ where $m_t = \int_0^t(\Lambda-\mathbb{1}_r)^T d\mathfrak{M}$ and $\Phi_1(x) = \ln(1+x)-x$.

Remark 3.3. By (3.1) and $(\Lambda-\mathbb{1}_r)^T\Delta\mathcal{Q} = 1-\dfrac{1-\Delta\underline{\overline{A}}}{1-\Delta\overline{A}}$, $\Phi_1(\Delta m) = \Phi_1^T(\Lambda-\mathbb{1}_r)\Delta\mathbb{N} + (1-\Delta\overline{N})\Phi_1(\dfrac{1-\Delta\underline{\overline{A}}}{1-\Delta\overline{A}}-1)$. Hence $z_t = \exp\{\int_0^t\ln^T\Lambda d\mathbb{N}-\underline{\overline{A}}_t^c + \overline{A}_t^c + \sum_{s\leqslant t}(1-\Delta\overline{N}_s)\ln\dfrac{1-\Delta\underline{\overline{A}}_s}{1-\Delta\overline{A}_s}\}$ (cp. Liptser/Shiryayev (1978), p. 312).

Remark 3.4. Process $z = (z_t,\mathfrak{F}_t,P)$, being a nonnegative supermartingale with $E(z|P) = 1$ as well as a local martingale, is a solution of the Doleans-Dade equation $z_t = 1 + \int_0^t z_{s-}dm_s, 0\leqslant t\leqslant 1$ (Liptser/Shiryayev (1978), p. 288).

3. Let $\{\Omega^n,\mathfrak{F}^n,(\mathfrak{F}_t^n,0\leqslant t\leqslant 1),P^n\}$, $n=1,2,...$ be a sequence of stochastic basises of the same type as above. Let $\mathbb{N}^n = (\mathbb{N}_t^n,\mathfrak{F}_t^n,P^n)$ be an r_n-variate counting process with the Doob-Meyer decomposition $\mathbb{N}^n = \mathbb{M}^n + \mathbb{A}^n$, where r_n, $n=1,2,...$ is a nondecreasing sequence of integers.

Define also $\mathfrak{M}_t^n = \int_0^t(V^n)^{-1}d\mathbb{M}^n$ where $V^n = I_{r_n}-\Delta\mathbb{A}^n\otimes\mathbb{1}_{r_n}$. Assume that the compensator $\mathbb{A}^n$ satisfies

Condition I of Section 2.

Lemma 3.3. Under Condition I for each $t, 0\leq t\leq 1$ $k_n^{-1}<\overline{\mathfrak{M}}^n>_t\to F_t$ in P^n probability as $n\to\infty$.

Proof. By 2) of Lemma 3.2 $k_n^{-1}<\overline{\mathfrak{M}}^n>_t = F_t^n + k_n^{-1}\sum_{s\leq t}(1-\Delta\overline{A}_s^n)^{-1}(\Delta\overline{A}_s^n)^2$, where the second term satisfies the inequality

$$\frac{1}{k_n}\sum_{\substack{s\leq t\\ 0<\Delta A_s^n<\frac{1}{2}}}(1-\Delta\overline{A}_s^n)^{-1}(\Delta\overline{A}_s^n)^2+\frac{1}{k_n}\sum_{\substack{s\leq t\\ \frac{1}{2}<\Delta A_s^n<1}}(1-\Delta\overline{A}_s^n)^{-1}(\Delta\overline{A}_s^n)^2$$

$$\leq\frac{2}{k_n}[\overline{A}^n]_t+\frac{1}{k_n}\sum_{\substack{s\leq t\\ \frac{1}{2}<\Delta A_s^n<1}}(1-\Delta\overline{A}_s^n)^{-1}$$

and therefore tends to 0 in P^n probability, as $k_n^{-1}[\overline{A}^n]_t\leq k_n^{-1}\overline{A}_t^{nd}\to 0$ and the number of jumps of $\overline{A}_s^n$, $s\leq t$, exceeding $\frac{1}{2}$ is finite.

The last argument is used also in (3.2) and (3.4) below.

Theorem 3.2. Let $W^1 = (W_t^1,\mathfrak{F}_t)$ be a continuous Gaussian martingale with the quadratic variation $<W^1>_t = F_t$. Then, under Condition I, $k_n^{-\frac{1}{2}}\overline{\mathfrak{M}}^n\xrightarrow{D}W^1$ as $n\to\infty$ in $D[0,1]$.

Proof follows from Liptser/Shiryayev (1980), Corollary 2. In fact, the condition (12) of this corollary is met (Lemma 3.3). As for the Lindeberg condition (L_2) of the corollary, for $k_n^{-\frac{1}{2}}\overline{\mathfrak{M}}_t^n = k_n^{-\frac{1}{2}}\int_0^t(1-\Delta\overline{A}_s^n)^{-1}d\overline{M}_s^n$ it is satisfied as for each ϵ, $0<\epsilon<\frac{1}{2}$

$$k_n^{-1}\int_0^t I(\Delta\overline{A}_s^n>1-\epsilon)(1-\Delta\overline{A}_s^n)^{-2}d<\overline{M}>_s = k_n^{-1}\int_0^t I(\Delta\overline{A}_s^n>1-\epsilon)(1-\Delta\overline{A}_s^n)^{-1}d\overline{A}_s^n\leq k_n^{-1}\sum_{\substack{s\leq t\\ 1-\epsilon<\Delta A_s^n<1}}(1-\Delta\overline{A}_s^n)^{-1}\to 0 \tag{3.2}$$

in P^n probability as $n\to\infty$.

Lemma 3.4. Let $\mathbb{H}^n = \{\mathbb{H}_t^n,\mathfrak{F}_t^n,P^n\}$, $n=1,2,...$ be a sequence of r_n-vector valued predictable processes such that there is a function σ_t^2 satisfying $\int_0^t\sigma^2dF>0$ for which (each t, $0\leq t\leq 1$)

$$\int_0^t\mathbb{H}^{nT}diagd\mathbb{A}^n\,\mathbb{H}^n\to\int_0^t\sigma^2dF \tag{3.3}$$

in P^n probability as $n\to\infty$. Then $<\int_0^t\mathbb{H}^{nT}d\mathfrak{M}^n>\to\int_0^t\sigma^2dF$ in P^n probability as $n\to\infty$.

Proof. In view of the continuity of F

$$\int_0^t\mathbb{H}^{nT}diagd\mathbb{A}^{nd}\mathbb{H}^n = \sum_{s\leq t}\mathbb{H}^{nT}diag\Delta\mathbb{A}^n\mathbb{H}^n\to 0$$

in P^n probability as $n\to\infty$, and this implies

$$\int_0^t\mathbb{H}^{nT}d[\mathfrak{C}^n,\mathbb{A}^n]\mathbb{H}^n = \sum_{s\leq t}(1-\Delta\overline{A}_s^n)^{-1}(\mathbb{H}_s^{nT}\Delta\mathbb{A}_s^n)^2\leq\sum_{s\leq t}\frac{\Delta\overline{\mathbb{A}}_s^n}{1-\Delta\overline{\mathbb{A}}_s^n}\mathbb{H}_s^{nT}diag\Delta\mathbb{A}_s^n\mathbb{H}_s^n$$

$$\leq\sum_{\substack{s\leq t\\ 0<\Delta A_s^n<\frac{1}{2}}}\mathbb{H}_s^{nT}diag\Delta\mathbb{A}_s^n\mathbb{H}_s^n+\sum_{\substack{s\leq t\\ \frac{1}{2}<\Delta A_s^n<1}}(1-\Delta\overline{A}_s^n)^{-1}\mathbb{H}_s^{nT}diag\Delta\mathbb{A}_s^n\mathbb{H}_s^n\to 0 \tag{3.4}$$

in P^n probability as $n\to\infty$, by $\mathcal{C}_t^n = \int_0^t (1-\Delta\bar{A}^n)^{-1}d\mathbb{A}^n$ and the Schwartz inequality. Now, the desired result follows from (3.3), (3.4) and 2) of Lemma 3.2.

Theorem 3.3. Along with (3.3) let the following Lindeberg condition be satisfied: for each t, $0\leqslant t\leqslant 1$ and $\epsilon>0$ (here $\mathbb{H}^n = col\{H^{in}, i=1,...,r_n\}$)

$$\int_0^t \sum_{i=1}^{r_n} I(|H^{in}|>\epsilon)(H^{in})^2 dA^{in}\to 0 \tag{3.5}$$

in P^n probability as $n\to\infty$. Let $W = (W_t,\mathcal{F}_t)$ be a continuous Gaussian martingale with the quadratic variation $<W>_t = \int_0^t \sigma^2 dF$. Then $\int_0^t \mathbb{H}^{nT} d\mathfrak{M}^n \xrightarrow{D} W_t$ as $n\to\infty$ in $D([0,1])$.

Proof is immediate consequence of Liptser/Shiryayev (1980), Corollary 2, for its conditions (12) and (L_2) are varified by Lemma 3.4 and, respectively, (3.4) and (3.5).

4. Suppose that a probability measure $\underline{P}^n$ in addition to P^n is given on a measurable space $\{\Omega^n,\mathcal{F}^n\}$ in the sequence of stochastic basises of the preceding subsection. Suppose in addition that the filtration $\{\mathcal{F}_t^n, 0\leqslant t\leqslant 1\}$ is minimal: $\mathcal{F}_t^n = \sigma\{\mathbb{N}_s^n : s\leqslant t\}$ where $\mathbb{N}^n = (\mathbb{N}_t^n,\mathcal{F}_t^n,P^n)$ is an r_n-variate counting process with the compensator $\mathbb{A}^n = (\mathbb{A}_t^n,\mathcal{F}_t^n,P^n)$. Let $\underline{\mathbb{N}}^n = (\mathbb{N}_t^n,\mathcal{F}_t^n,\underline{P}^n)$ be another counting process with the compensator $\underline{\mathbb{A}}^n = (\underline{\mathbb{A}}_t^n,\mathcal{F}_t^n,\underline{P}^n)$.

For each, n assume $\underline{P}^n << P^n$ and, in accordance with II of Theorem 3.1, define the Hellinger process

$$\mathcal{H}_t^n = \int_0^t \mathbb{U}^{nT} diag d\mathbb{A}^n \mathbb{U}^n + \sum_{\substack{s\leqslant t\\ 0<\Delta A_s^n<1}} (\sqrt{1-\Delta \underline{\bar{A}}_s^n} - \sqrt{1-\Delta \bar{A}_s^n})^2$$

where $\mathbb{U}^n = col\{U^{in} = \sqrt{d\underline{A}^{in}/dA^{in}} - 1, i=1,...,r_n\}$. Obviously, $\Lambda^n = col\{\lambda^{in} = (U^{in}+1)^2, i=1,...,r_n\}$.

Theorem 3.4. Let there be a function σ_t^2 satisfying $\int_0^t \sigma^2 dF>0$ such that for each $t, 0\leqslant t\leqslant 1$ $\mathcal{H}_t^n\to\int_0^t \sigma^2 dF$ in P^n probability as $n\to\infty$. Let $\mathbb{U}^n, n=1,2,...$ satisfy the Lindeberg condition (3.5). Then $\ln z_1^n = m_1^n + \sum_{s\leqslant 1}\Phi_1(\Delta m_s^n)$ (with $m_t^n = \int_0^t (\Lambda^n - \mathbb{I}_{r_n})^T d\mathfrak{M}^n$) is asymptotically normal:

$$\mathcal{L}\{\ln z_1^n | P^n\}\Rightarrow N(-2\int_0^1 \sigma^2 dF,\ 4\int_0^1 \sigma^2 dF). \tag{3.6}$$

Proof of this theorem follows the same line as that of Gill (1979) Proposition 5.3.1, and therefore it will be shortly sketched in the following two remarks.

Remark 3.5. Since $\mathbb{U}^n$ satisfies the conditions of Theorem 3.3, $m_1^n = 2\int_0^1 \mathbb{U}^{nT} d\mathfrak{M}^n + \int_0^1 \mathbb{U}^{nT} diag d\mathfrak{M}^n \mathbb{U}^n$ is asymptotically normal with zero mean and variance $4\int_0^1 \sigma^2 dF$, that is the limit in P^n probability of $4<\int_0^1 \mathbb{U}^{nT} d\mathfrak{M}^n>$.

Remark 3.6. As

$$\frac{1}{2}\sum_{s\leqslant 1}\Phi_1(\Delta m_s^n) = \int_0^1 \Phi_2^T(\mathbb{U}^n) d\mathbb{N}^n + \sum_{s\leqslant 1}(1-\Delta\bar{N}^n)\Phi_2(1-\sqrt{\frac{1-\Delta\underline{\bar{A}}^n}{1-\Delta\bar{A}^n}}) - \int_0^1 \mathbb{U}^{nT} diag\ d\mathbb{A}^n \mathbb{U}^n$$

where the first two terms tend to zero, while the third term tends to $-\int_0^1 \sigma^2 dF$ in P^n probability, the desired limiting value is obtained for the mean in (3.6). Here $\Phi_2(x)$ for $x = col(x^1,...,x^r)$ denotes $col\{\ln(1+x^i) - x^i + \frac{1}{2}x^{i2},\ i=1,...,r\}$.

In the sequel we deal with a situation in which the following condition is satisfied: There is a sequence of bounded r_n-variate predictable processes $\mathbb{S}^n = \{\mathbb{S}_t^n, \mathcal{F}_t^n, P^n\}, n=1,2,...$ such that

$$\int_0^1 (\mathbb{U}^n - k_n^{-\frac{1}{2}}\mathbb{S}^n)^T diag\ d\mathbb{A}^n(\mathbb{U}^n - k_n^{-\frac{1}{2}}\mathbb{S}^n) \to 0 \tag{3.7}$$

in P^n probability as $n\to\infty$.

Corollary 3.1. If $k_n^{-\frac{1}{2}}\mathbb{S}^n, n=1,2,...$ satisfies (3.3), (3.5) and (3.7) (with $\mathbb{H}^n = k_n^{-\frac{1}{2}}\mathbb{S}^n$), then $z_1^n = \exp\{\frac{2}{\sqrt{k_n}}\int_0^1 \mathbb{S}^{nT} d\mathfrak{M}^n - 2\int_0^1 \sigma^2 dF + \eta_n\}$ where $\eta_n \to 0$ in P^n probability as $n\to\infty$ and $\mathfrak{L}\{k_n^{-\frac{1}{2}}\int_0^1 \mathbb{S}^{nT} d\mathfrak{M}^n\} \Rightarrow N(0, \int_0^1 \sigma^2 dF)$.

4. CONSISTENCY AND ASYMPTOTIC NORMALITY

1. Consider the situation described in Subsection 2.1, and suppose that the r_n-vector valued process $\sqrt{\Psi^n(\beta)}$ is continuously differentiable (in β) in the following sense:

Condition II. There is a sequence of continuous in β r_n-vector valued predictable processes

$$\mathbb{L}^n(\beta)(\equiv \frac{\partial}{\partial\beta}\sqrt{\Psi^n(\beta)}) = \{\mathbb{L}_t^n(\beta), \mathcal{F}_t^n, P^n\}\ n=1,2,...$$

such that if $(\overline{A}_t^n, \Psi_t^n(\beta))$ is the pair of the $(P^n, \mathcal{F}_t^n)$-predictable characteristics, then

II 1. For each real valued b such that $\underline{\beta}^n = \beta + b/\sqrt{k_n} \in \mathfrak{B}$, eventually,

$$\int_0^1 |\sqrt{\Psi^n(\underline{\beta}^n)} - \sqrt{\Psi^n(\beta)} - \frac{b}{\sqrt{k_n}}\mathbb{L}^n(\beta)|^2 d\overline{A}^n \to 0$$

in P^n probability as $n\to\infty$.

II 2. For some deterministic function $\sigma_t^2(\beta)$ such that $v_t(\beta) = \int_0^t \sigma^2(\beta) dF > 0,\ 0 \leqslant t \leqslant 1$

$$\mathcal{V}_t^n(\beta) = \int_0^t |\mathbb{L}^n(\beta)|^2 dF^n \to v_t(\beta)$$

in P^n probability as $n\to\infty$.

We shall show that the estimator $\hat{\beta}_n$, defined by (1.4) with $\Psi = \Psi^n$ and $\mathbb{N} = \mathbb{N}^n$ is consistant and asymptotically normal $N(0, 1/4v)$, $v = v_1(\beta)$. For this we need some additional conditions stipulated in the next subsection.

2. Define $\frac{\partial}{\partial\beta}\Psi^n(\beta) = 2\{diag\Psi^n(\beta)\}^{1/2}\frac{\partial}{\partial\beta}\sqrt{\Psi^n(\beta)}$ and

$$\frac{\partial}{\partial\beta}\ln\Psi^n(\beta) = 2\{diag\Psi^n(\beta)\}^{-1/2}\frac{\partial}{\partial\beta}\sqrt{\Psi^n(\beta)}.$$

Obviously,

$$\int_0^t \{\frac{\partial}{\partial \beta} \ln \Psi^n(\beta)\}^T d\mathbb{A}^n = \int_0^t \mathbb{I}_{r_n}^T \frac{\partial}{\partial \beta} \Psi^n(\beta) d\bar{A}^n = 0. \tag{4.1}$$

Hence,

$$\int_0^t \{\frac{\partial}{\partial \beta} \ln \Psi^n(\beta)\}^T d\mathfrak{M}^n = \int_0^t \{\frac{\partial}{\partial \beta} \ln \Psi^n(\beta)\}^T d\mathbb{N}^n \tag{4.2}$$

and

$$\frac{1}{k_n} \int_0^t \{\frac{\partial}{\partial \beta} \ln \Psi^n(\beta)\}^T diag\ d\mathbb{A}^n \{\frac{\partial}{\partial \beta} \ln \Psi^n(\beta)\} = 4 \int_0^t |\mathbb{L}^n(\beta)|^2 dF^n \rightarrow 4 v_t(\beta) \tag{4.3}$$

in P^n probability as $n \rightarrow \infty$ (see II.2)).

Now we apply Theorem 3.3. to derive asymptotic normality of the intergral in (4.2), taking account of (4.1) and (4.3).

Corollary 4.1. Along with Conditions II, let the following Lindeberg condition hold ($\mathbb{L}^n = col\{L^{in}, i=1,...r_n\}$):

Condition III. If $(\bar{A}_t^n, \Psi_t^n(\beta))$ is the pair of the $(P^n, \mathfrak{F}_t^n)$-predictable characteristics, then for each $\epsilon > 0$

$$\frac{1}{k_n} \int_0^1 \sum_{i=1}^{r_n} I(k_n^{-1/2} |L^{in}(\beta)| > \epsilon) |L^{in}(\beta)|^2 dA^{in} \rightarrow 0$$

in P^n probability as $n \rightarrow \infty$ Then

$$\mathfrak{L}(\frac{1}{k_n} \int_0^1 \{\frac{\partial}{\partial \beta} \ln \Psi^n(\beta)\}^T d\mathbb{N}^n | P^n) \Rightarrow N(0, 4v)$$

Condition IV. If $(\bar{A}_t^n, \Psi_t^n(\beta))$ is the pair of the $(P^n, \mathfrak{F}_t^n)$-predictable characteristics, then

$$\int_0^1 | \sqrt{\Psi^n(\beta_n^2)} - \sqrt{\Psi^n(\beta_n^1)} |^2 d\bar{A}^n,$$

where

$$\beta_n^i = \beta + b_i / 2\sqrt{k_n v}, \quad b_i \in B_n = \mathfrak{B} - \beta / 2\sqrt{k_n v}, \ i = 1,2,$$

obeys the following bounds: there are constants $C_1 > 0$ and $C_2 > C_1$ (independent of n) such that for sufficiently large values of n and each $b_i \in B_n$, $i = 1,2$

$$P^n \text{ a.s. } C_1 (b_2 - b_1)^2 \leqslant \int_0^1 | \sqrt{\Psi^n(\beta_n^2)} - \sqrt{\Psi^n(\beta_n^1)} |^2 d\bar{A}^n \leqslant C_2 (b_2 - b_1)^2$$

Lemma 4.1. Let Condition IV hold. Define $Y^n(b) = \exp \int_0^1 \{\ln \Psi^n(\beta_n) - \ln \Psi^n(\beta)\}^T d\mathbb{N}^n$, $b \in B_n$. Then there are constants $c_1 > 0$ and $c_2 > 0$ such that for each $b \in B_n$ and $b_i \in B_n$, $i = 1,2$

(i) $E\{ \sqrt{Y^n(b)} | P^n \} \leqslant e^{-c_1 b^2}$,

(ii) $E\{ | \sqrt{Y^n(b_2)} - \sqrt{Y^n(b_1)} |^2 | P^n \} \leqslant c_2 |b_2 - b_1|^2$

Proof.

(i). In accordance with the remarks 3.3 and 3.4

$$E\{\sqrt{Y^n(b)}\exp(\frac{1}{2}\int_0^1|\sqrt{\Psi^n(\beta_n)}-\sqrt{\Psi^n(\beta)}|^2 d\bar{A}^{nc}+\sum_{s\leqslant 1}(1-\Delta\bar{N}^n_s)\ln(1+\frac{1}{2}\frac{\Delta\bar{A}^n_s}{1-\Delta\bar{A}^n_s}|\sqrt{\Psi^n_s(\beta_n)}-\sqrt{\Psi^n_s(\beta)}|^2))|P^n\} = 1.$$

Thus

$$1\geqslant\exp\frac{1}{2}C_1 b^2 E\{\sqrt{Y^n(b)}\exp-\frac{1}{2}\sum_{s\leqslant 1}|\sqrt{\Psi^n_s(\beta_n)}-\sqrt{\Psi^n_s(\beta)}|^2\Delta A^n_s|P^n\}\geqslant E\{\sqrt{Y^n(b)}|P^n\}\exp c_1 b^2.$$

(ii) As

$$E\{Y^n(b)|P^n\} = E\{\exp\int_0^1(\ln\Psi^n(\beta_n)-\ln\Psi^n(\beta))^T d\mathbb{N}^n|P^n\} = 1, \tag{4.4}$$

and

$$E\{\sqrt{Y^n(b_1)Y^n(b_2)}\exp(\frac{1}{2}\int_0^1|\sqrt{\Psi^n(\beta^2_n)}-\sqrt{\Psi^n(\beta^1_n)}|^2 d\bar{A}^{nc}+$$

$$+\sum_{s\leqslant 1}(1-\Delta\bar{N}^n_s)\ln(1+\frac{1}{2}\frac{\Delta\bar{A}^n_s}{1-\Delta\bar{A}^n_s}|\sqrt{\Psi^n_s(\beta^2_n)}-\sqrt{\Psi^n_s(\beta^1_n)}|^2))|P^n\} = 1,$$

we have

$$1\leqslant 2\exp\frac{1}{2}C_2(b_2-b_1)^2 E\{\sqrt{Y^n(b_1)Y^n(b_2)}\exp\frac{1}{2}\sum_{s\leqslant 1}\frac{\Delta\bar{A}^n_s}{1-\Delta\bar{A}^n_s}|\sqrt{\Psi^n(\beta^2_n)}-\sqrt{\Psi^n(\beta^1_n)}|^2|P^n\}$$

$$\leqslant 2\exp\frac{1}{2}C_2(b_2-b_1)^2 E\{\sqrt{Y^n(b_1)Y^n(b_2)}\exp(\sum_{\substack{s\leqslant 1\\ 0<\Delta A^n_s<\frac{1}{2}}}|\sqrt{\Psi^n_s(\beta^2_n)}-\sqrt{\Psi^n_s(\beta^1_n)}|^2\Delta\bar{A}^n_s$$

$$+\frac{1}{2}\sum_{\substack{s\leqslant 1\\ \frac{1}{2}<\Delta A^n_s<1}}\frac{\Delta\bar{A}^n_s}{1-\Delta\bar{A}^n_s}|\sqrt{\Psi^n_s(\beta^2_n)}-\sqrt{\Psi^n_s(\beta^1_n)}|^2)|P^n\}\leqslant E\{\sqrt{Y^n(b_1)Y^n(b_2)}|P^n\}e^{\frac{1}{2}c_2(b_2-b_1)^2}$$

(cp. the proof of Lemma 3.3), and

$$E\{|\sqrt{Y^n(b_2)}-\sqrt{Y^n(b_1)}|^2|P^n\}\leqslant 2(1-E\{\sqrt{Y^n(b_2)Y^n(b_1)}|P^n\})\leqslant 2(1-e^{-\frac{1}{2}c_2(b_2-b_1)^2})\leqslant c_2(b_2-b_1)^2.$$

Lemma 4.1 and the equation (4.4) allow us to apply the result of Sieders (1985) mentioned in Subsection 2.1.

Proposition 4.1. Under the conditions stipulated above there are certain positive constants C_0 and c_0 such that the estimator $\hat{\beta}_n$ is consistent: $\hat{\beta}_n\to\beta$ in P^n probability, and for sufficiently large values of n and H

$$P^n\{|2\sqrt{k_n v}(\hat{\beta}_n-\beta)|>H\}\leqslant C_0 e^{-c_0 H^2}$$

Repeating the arguments leading to Corollary 3.5, and taking into account Corollary 4.1 we arrive at the assertion of the following lemma.

Lemma 4.2. Under the conditions stipulated above finite dimentional distributions of $Y^n(b)$ tend to finite dimentional distributions of $e^{b\zeta-\frac{1}{2}b^2}$, $\mathfrak{L}(\zeta) = N(0,1)$.

In view of the assertions of the Lemmas 4.1 and 4.2, we can make use of Theorem I.10.1 and III.1.2 of Ibragimov/Has'minskii (1981). The result can be formulated as

Proposition 4.2. Under the conditions stipulated above, for each $\delta>0$ as $n\to\infty$

$$P^n\{|2\sqrt{k_n\nu}(\hat{\beta}_n-\beta)-\frac{1}{2\sqrt{k_n\nu}}\int_0^1\{\frac{\partial}{\partial\beta}\ln\Psi^n(\beta)\}^T d\mathbb{N}^n|>\delta\}\to 0$$

$$\mathfrak{L}\{2\sqrt{k_n\nu}(\hat{\beta}_n-\beta)|P^n\}\Rightarrow N(0,1)$$

and

$$E\{|2\sqrt{k_n\nu}(\hat{\beta}_n-\beta)|^K|P^n\}\to\frac{2^{K/2}}{\sqrt{\pi}}\Gamma(\frac{K+1}{2}).$$

5. ASYMPTOTIC OPTIMALITY

1. As in Subsection 2.2, suppose that the process $\mathbb{N}^n$ is of the Poisson type with the compensator (1.1) where $\Phi = \Phi^n$ satisfies the conditions stipulated below.

Condition V. There is a sequence of r_n-vector valued $\mathcal{F}^n_t$-predictable processes, continuously dependent on β, say $(\partial/\partial\beta)\sqrt{\Phi^n(\beta)}, n=1,2,...,$ such that for each $\alpha\in\mathcal{Q}$ and $\beta\in\mathcal{B}$ the following holds:

V.1. For each b such that $\underline{\beta}^n=\beta+b/\sqrt{k_n}\in\mathcal{B}$ eventually,

$$\int_0^1|\sqrt{\Phi^n(\underline{\beta}^n)}-\sqrt{\Phi^n(\beta)}-\frac{b}{\sqrt{k_n}}\frac{\partial}{\partial\beta}\sqrt{\Phi^n(\beta)}|^2 d\alpha\to 0$$

in $P^n_{\alpha,\beta}$ probability as $n\to\infty$;

V.2. For some deterministic function $\rho^2_t(\beta)$ such that $w_t(\alpha,\beta)=\int_0^t\rho^2_s(\beta)d\alpha_s>0$, and for each t, $0\leqslant t\leqslant 1$, $\alpha\in\mathcal{Q}$, $\beta\in\mathcal{B}$

$$\mathcal{W}^n_t(\alpha,\beta) = \frac{1}{k_n}\int_0^1|\frac{\partial}{\partial\beta}\sqrt{\Phi^n_s(\beta)}|^2 d\alpha_s\to w_t(\alpha,\beta)$$

in $P^n_{\alpha,\beta}$ probability as $n\to\infty$.

Condition VI. There is a positive bounded deterministic function $\phi_t(\beta)$ (uniformly in t and β $m<\phi_t(\beta)<M$ where $0<m<M<\infty$) having continuous bounded derivative in β, such that for each $\alpha\in\mathcal{Q}$, $\beta\in\mathcal{B}$ and $0\leqslant t\leqslant 1$

VI. 1. $F^n_t(\alpha,\beta) = \int_0^t\frac{1}{k_n}\overline{\Phi}^n_s(\beta)d\alpha_s\to F_t(\alpha,\beta)$ in $P^n_{\alpha,\beta}$ probability as $n\to\infty$ where $F_t = F_t(\alpha,\beta) = \int_0^t\phi_s(\beta)d\alpha_s>0$ (cp. Condition I).

VI.2. For $(\partial/\partial\beta)\overline{\Phi}^n = \mathbb{I}^T_{r_n}(\partial/\partial\beta)\Phi^n = 2\mathbb{I}^T_{r_n}\{diag\Phi^n\}^{1/2}(\partial/\partial\beta)\sqrt{\Phi^n}$

$$\int_0^t\frac{\partial}{\partial\beta}\overline{\Phi}^n_s(\beta)d\alpha_s\to\int_0^t\frac{\partial}{\partial\beta}\phi_s(\beta)d\alpha_s<\infty$$

in $P^n_{\alpha,\beta}$ probability as $n\to\infty$.

Remark 5.1. By the Conditions II.2, V.2, VI.1, VI.2 and (4.3)

$$\mathcal{V}_t(\alpha,\beta) = \mathcal{W}_t(\alpha,\beta)-\int_0^t|\frac{\partial}{\partial\beta}\sqrt{\overline{\Phi}^n_s}|^2 d\alpha_s\to w_t(\alpha,\beta)-\int_0^t|\frac{\partial}{\partial\beta}\sqrt{\phi_s}|^2 d\alpha_s = v_t(\alpha,\beta)$$

in $P^n_{\alpha,\beta}$ probability as $\to\infty$.

2. Let the probability measures P^n and $\underline{P}^n$ be defined on $(\Omega^n,\mathcal{F}^n)$ as in Subsection 2.2. Then (see Remark 3.3)

$$\frac{d\underline{P}^n}{dP^n} = \exp\{\int_0^1(\ln\Phi_s^n(\underline{\beta}^n)-\ln\Phi_s^n(\beta))^T d\mathbb{N}_s^n + 2\int_0^1 \ln\frac{d\underline{\alpha}_s^n}{d\alpha_s^n}d\overline{N}_s^n - \int_0^1 \overline{\Phi}_s^n(\underline{\beta}^n)d\underline{\alpha}_s^n + \int_0^1 \overline{\Phi}_s^n(\beta)d\alpha_s\} \tag{5.1}$$

We shall now apply Corollary 3.1 to show the LAN of the $\{P^n_{\alpha,\beta},\alpha\in\mathcal{C}, \beta\in\mathcal{B}\}$ in the sense of

Definition 5.1. This family is called Locally Asymptotically Normal (LAN) at the "point" $\alpha\in\mathcal{C}$, $\beta\in\mathcal{B}$ if for each $b\in R^1$ and each bounded $a\in L^2(dF)$ such that $\underline{\alpha}^n\in\mathcal{C},\underline{\beta}^n\in\mathcal{B}$ eventually, there is a sequence of asymptotically normal variables $\delta^n_{\alpha,\beta}(a,b)$, $n=1,2,\ldots$:

$$\mathcal{L}\{\delta^n_{\alpha,\beta}(a,b)|P^n_{\alpha,\beta}\}\Rightarrow N(0,g_{\alpha,\beta}(a,b))$$

as $n\to\infty$ with $g_{\alpha,\beta}(a,b)>0$ for which $d\underline{P}^n/dP^n= \exp\{\delta^n_{\alpha,\beta}(a,b)-\frac{1}{2}g_{\alpha,\beta}(a,b) + \eta^n_{\alpha,\beta}(a,b)\}$ where $\eta^n_{\alpha,\beta}(a,b)\to 0$ in $P^n_{\alpha,\beta}$ probability as $n\to\infty$.

Note first that if

$$\mathbb{U}_t^n = col\{\sqrt{\Phi_t^{in}(\underline{\beta}^n)/\Phi_t^{in}(\beta)}\sqrt{d\underline{\alpha}_t^n/d\alpha_t}-1,\ i=1,\ldots,r_n\},$$

then (3.7) is satisfied by $\mathbb{S}_t^n = \mathbb{S}^n_{t,\beta}(a,b) = \frac{1}{2}b(\partial/\partial\beta)\ln\Phi_t^n(\beta) + a_t\mathbb{1}_{r_n}$, for which

$$\frac{1}{k_n}\int_0^t \mathbb{S}^{nT}diagd\mathbb{A}^n\mathbb{S}^n\to b^2w_t(\alpha,\beta) + 2b\int_0^t\frac{(\partial/\partial\beta)\sqrt{\phi_s(\beta)}}{\sqrt{\phi_s(\beta)}}a_s dF_s(\alpha,\beta) + \int_0^t a_s^2 dF_s(\alpha,\beta) \tag{5.2}$$

in P^n probability as $n\to\infty$, by the Conditions V and VI. Finally, let the following Lindeberg condition hold:

Condition VII. For each $\alpha\in\mathcal{C}$, $\beta\in\mathcal{B}$ and $\epsilon>0$

$$\frac{1}{k_n}\int_0^t\sum_{i=1}^{r_n} I(k_n^{-1/2}\frac{\partial}{\partial\beta}\sqrt{\Phi_s^{in}(\beta)}|>\epsilon)|\frac{\partial}{\partial\beta}\sqrt{\Phi_s^{in}(\beta)}|^2 d\alpha_s\to 0$$

in $P^n_{\alpha,\beta}$ as $n\to\infty$.

Proposition 5.1. Under the Conditions V-VII the family $\{P^n_{\alpha,\beta},\alpha\in\mathcal{C}, \beta\in\mathcal{B}\}$ is LAN at the "point" $\alpha\in\mathcal{C}$, $\beta\in\mathcal{B}$ for $\delta^n_{\alpha,\beta}(a,b) = \frac{1}{\sqrt{k_n}}\int_0^1 \mathbb{S}^n_\beta(a,b)^T d(\mathbb{N}^n-\mathbb{A}^n(\alpha,\beta))$ and $g_{\alpha,\beta}(a,b)$ that equals to 4 times the right-hand side of (5.2) evaluated at $t=1$.

3. Suppose that the underling model confines "the directions" a to a linear subspace $\mathrm{A}\in L^2(dF)$, and let $\{\beta^n_{RA}\}$ be a class of $R^1\times\mathrm{A}$-regular estimators for β, which includes a subclass of regular estimators $\{\beta^n_R\}\subset\{\beta^n_{RA}\}$ (see Subsection 2.3). Then, by Hájek's convolution theorem (Begun et al. (1983), Theorem 3.1), we have

Proposition 5.2. Let the Condition V-VII hold. Then

(i) $$\mathcal{L}\{\sqrt{k_n}(\beta^n_{RA}-\underline{\beta}^n)|\underline{P}^n\}\Rightarrow G^0_{RA} = N(0,\mathcal{I}^{-1}_{\alpha,\beta})*G^1_{RA} \tag{5.3}$$

as $n\to\infty$ with some distribution law G^1_{RA}, where $\mathcal{I}_{\alpha,\beta}=4\{w_1(\alpha,\beta) - \int_0^1\pi_s^2(\alpha,\beta)dF_s(\alpha,\beta)\}$, $\pi_s(\alpha,\beta)$ being the projection of $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)}$ into A, that is, it satisfies the equation

$$\int_0^1\{\frac{(\partial/\partial\beta)\sqrt{\phi_s(\beta)}}{\sqrt{\phi_s(\beta)}}-\pi_s(\alpha,\beta)\}a_s dF_s(\alpha,\beta) = 0$$

for each $a_s \in A$.

(ii) $$\mathfrak{L}\{\sqrt{k_n}(\beta_R^n - \beta^n)|\underline{P}^n\} \Rightarrow G_R^0 = N(0, 1/4v_1(\alpha,\beta)) * G_R^1 \tag{5.4}$$

as $n \to \infty$ with some distribution law G_R^1 (see Remark 5.1), uniformly for each $|b| < c$ whatever $c > 0$, and each bounded $a \in L^2(dF)$

Remark 5.2. For the "least favorable" direction $a_t = -b\pi_t(\beta)$ the quantity $g_{\alpha,\beta}(a,b)$ coincides with $b^2 \mathcal{I}_{\alpha,\beta}$, $\mathcal{I}_{\alpha,\beta}$ being Fisher's information for β (Begun et al. (1983), Section 3).

Remark 5.3. Evidently, $\mathcal{I}_{\alpha,\beta} \geq 4v_1(\alpha,\beta)$ with equality iff $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)} \in A$ (see Remark 5.1).

Having the limiting distribution of $\sqrt{k_n}(\hat{\beta}_n - \beta)$ under P^n (Proposition 4.2), one can apply the usual contiguity arguments (allowed by Proposition (5.2)) to arrive at the formula (5.4) for $\beta_R^n = \hat{\beta}_n$ with G_R^1 that degenerates at 0. These considerations can be summerized as the following statement on the optimality properties of the estimator $\hat{\beta}_n$.

Theorem 5.1. Under the conditions stipulated above, the estimator $\hat{\beta}_n$ is regular and $\mathfrak{L}\{\sqrt{k_n}(\hat{\beta}_n - \beta^n)|\underline{P}^n\} \Rightarrow N(0, 1/4v_1(\alpha,\beta))$ for each $b \in R^1$ and each bounded $a \in L^2(dF)$ determining $\underline{\alpha}^n, \underline{\beta}^n$ and $\underline{P}^n$ as in Subsection 2.2.

The estimator $\hat{\beta}_n$ is the best among $\{\beta_R^n\}$ in the sense that no regular estimator can have the limiting distribution less spread then $\hat{\beta}_n$. Besides, iff $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)} \in A$, then it is the best among $\{\beta_{RA}^n\} \supset \{\beta_R^n\}$ in the same sense. Finally, observe that Proposition 5.2 and Theorem 11.2 of Ibragimov/Has'minskii (1981), Chapter II, allow us to obtain the lower bounds for the risk of $R^1 \times A$-regular and regular estimators and, consequently, to give yet another characterization of the optimality of $\hat{\beta}_n$. Namely, the following corollary holds:

Corollary 5.1 Let the conditions stipulated above be satisfied. Let $w(x) \geq 0$ $x \in R^1$ be a continuous even loss function. Then for fixed $\alpha \in \mathcal{Q}, \beta \in \mathfrak{B}$

$$\liminf_{n\to\infty} E\{w(\zeta^n) \mid P_{\alpha,\beta}^n\} \geq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} w(x) e^{-\frac{1}{2}x^2} dx$$

where $\zeta^n = 2\sqrt{k_n v}(\beta_R^n - \beta)$. The same inequality holds also for $\zeta^n = \sqrt{k_n \mathcal{I}}(\beta_{RA}^n - \beta)$. In particular

$$\liminf_{n\to\infty} k_n var\{\beta_R^n \mid P_{\alpha,\beta}^n\} \geq (4v)^{-1} \geq \mathcal{I}^{-1}, \quad \liminf_{n\to\infty} k_n var\{\beta_{RA}^n \mid P_{\alpha,\beta}^n\} \geq \mathcal{I}^{-1}$$

If, in addition, w allows a polynominal majorant, then by the last assertion of Proposition 4.2

$$\lim_{n\to\infty} E\{w(2\sqrt{k_n v}(\hat{\beta}_n - \beta) \mid P_{\alpha,\beta}^n\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} w(x) e^{-\frac{1}{2}x^2} dx,$$

hence $\hat{\beta}_n$ attains the lower bound for the risks of regular estimators. Besides, iff $(\partial/\partial\beta)\sqrt{\phi_t(\beta)}/\sqrt{\phi_t(\beta)} \in A$, then $\mathcal{I} = 4v$ and $\hat{\beta}_n$ attains the lower bound also for the risks of $R^1 \times A$-regular estimators.

BIBLIOGRAPHY

Andersen, P. K. and Gill, R.D. (1982) Cox's regression model for counting processes: a large sample study, *Ann. Statist.* **10** 1100-1120.

Begun, J.M., Hall, W.J., Huang, W.M. and Wellner, J.A. (1983) Information and asymptotic efficiency in parametric and non-parametric models, *Ann. Statist.* **11** 432-452.

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

Brémaud, P. (1981) Point Processes and Queues. Martingale Dinamics. Springer-Verlag. New York.

Efron, B. (1977) The efficiency of Cox's likelihood function for censored data *J. Amer. Statist. Assoc.* **72** 557-565.

Gill, R.D. (1979) Censoring and Stochastic Integrals. Mathematisch Centrum Amsterdam.

Greenwood, P.E. and Shiryayev, A.N. (1985) Continuity and the Statistical Invariance Principle. Stochastic Monographs. Vol. 1, Gordon and Breach, New York.

Hjort, N.L. (1984)Bayes estimators and asymptotic efficiency in parametric counting process models. Research report. Norw. Comp. Centre, Oslo (submitted to *Scand. J. Statist).*

Ibragimov, I.A. and Has'minskii A.N. (1981) Statistical Estimation. Asymptotic Theory. Springer-Verlag. New York.

Jarupskin, B.D.S. (1983) Maximum likelihood and related estimation methods in point processes and point process systems. Ph. D. Dept. Statist. University of California. Berkeley.

Kabanov, Yu. M., Lipcer, R.S. and Shiryaev A.N. (1976) Criteria of absolute continuity of measures corresponding to multivariate point processes. Proc. Third Japan-USSR Sympos. Probability Theory (Tashkent, 1975), Lecture Notes in Math., vol. 550, Springer-Verlag, New York.

Kabanov, Yu. M., Lipcer, R. S. and Shiryaev A.N. (1980) Absolute continuity and singularity of locally absolute continuous probability distributions. II *Math. USSR Sbornik.* **36** 31-58.

Kabanov, Yu. M., Liptser R. Sh. and Shiryayev A.N. (1980) Some limit theorems for simple point processes (a martingale approach). *Stochastics.* **3** 203-216.

Le Cam, L. (1969) Theorie Asymptotic de la Decision Statistique, Les Presses de l'Universite de Montreal. Montreal.

Liptser, R.S. and Shiryayev A.N. (1978) Statistics of Random Processes II. Applications. Springer-Verlag. New York.

Liptser, R. Sh. and Shiryaev A.N. (1980) A functional central limit theorem for samimartingales. *Theory Probab. Appl.* **25** 667-688.

McLeish D.L. (1978) An extended martingale principle. *Ann. Probab.* **6** 144-150.

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

Shiryayev, A.N. (1981) Martingales: Recent Developments, Results and Applications. *Int. Statist. Revieuw.* **49** 199-233.

Sieders A. (1985) An extention of a Theorem by Ibragimov and Has'minskii. Research report, TH Delft.