Ann. Inst. Statist. Math.
34 (1982), Part A, 259-280

ASYMPTOTIC THEORY FOR ESTIMATING THE PARAMETERS OF A LÉVY PROCESS

MICHAEL G. AKRITAS

(Received Nov. 13, 1979; revised Sept. 30, 1981)

Summary

We consider consistency and asymptotic normality of maximum likelihood estimators (MLE) for parameters of a Lévy process of the discontinuous type. The MLE are based on a single realization of the process on a given interval $[0, t]$. Depending on properties of the Lévy measure we either consider the MLE corresponding to jumps of size greater than ε and, keeping t fixed, we let ε tend to 0, or we consider the MLE corresponding to the complete information of the realization of the process on $[0, t]$ and let t tend to ∞. The results of this paper improve in both generality and rigor previous asymptotic estimation results for such processes.

1. Introduction

Let X denote a Lévy process, i.e., a stochastic process with independent increments for which the distribution of $X(t+h)-X(t)$ depends only on h for all $t \in [0, \infty)$. Such a process is the continuous time analogue of a sequence of partial sums of independent, identically distributed random variables and thus may be used to model inputs to a dam or demands for a particular commodity, or certain other economic phenomena. See for instance Moran [16], Fama [9], Brockwell and Chung [7].

In this paper we establish consistency and asymptotic normality of maximum likelihood estimators using the approach of Huber [14]. Huber's results were extended to the Markovian case by Rao [18] and to the independent not identically distributed case by Inagaki [15]. Inagaki also proved a number of interesting asymptotic results which may be shown to carry over to the case of Lévy processes. Depending on properties of the Lévy measure, it is natural to consider two cases. Namely, the case in which consistent estimators are obtained from complete information over any finite interval $[0, t]$, and the case in which consistent estimators are obtained only by letting t tend to ∞. Roughly speaking, the distinction lies in whether, for any $a>0$, Fisher's information number $\int_{\{0\}^c} I_{(-a,a)}(x)\left[\frac{\partial/\partial\theta(d\mu_\theta(x))}{d\mu_\theta(x)}\right]^2 d\mu_\theta(x)$ corresponding to the jumps of the process $X(t)$, $0\leqq t\leqq 1$, of size in $(-a, a)$ is infinite or finite. Indeed, condition (BC 3) with $p(x, \theta)=-\log\frac{d\mu_\theta(x)}{d\mu_{\theta_0}(x)}$ and $a(x, \theta)=\left(1-\frac{d\mu_\theta(x)}{d\mu_{\theta_0}(x)}\right)$ cannot hold unless $\int_{\{0\}^c}\left[\frac{\partial/\partial\theta(d\mu_{\theta_0}(x))}{d\mu_{\theta_0}(x)}\right]^2 d\mu_{\theta_0}(x)<\infty$. Similarly for condition (BN 2). Finally note that when the last integral above is infinite the measures $P_{T,\theta}$ and P_{T,θ^*} cannot be mutually absolutely continuous because condition (L 2) of Proposition 2.4 is violated. The above arguments justify the subsequent labeling of the two cases. The important question regarding the rate of convergence of the estimators is investigated in Akritas [2].

Due to the fact that until recently (see Akritas [1], Chapter 4, or Akritas and Johnson [3]) there did not exist a convenient expression for the Radon-Nikodym derivative for Lévy processes, there has been no previous general treatment of maximum likelihood estimation. Rubin and Tucker [20] consider nonparametric estimation of quantities appearing in the characteristic function of independent increment processes, whereas Frost [12] was the first to study signal detection and estimation problems for such processes (see also Segall and Kailath [21]). Recently, Basawa and Brockwell [4] considered maximum likelihood estimation for gamma and stable processes.

Technically, let Ω be the space $D([0, \infty))$ of all real valued functions $X(t)$, $t\in[0, \infty)$ that are right continuous and have finite left-hand limits, and let $\mathcal{A}$ be the σ-field of cylinder sets in Ω. For each $\theta\in\Theta\subseteq R^k$, let P_θ be a probability measure on $(\Omega, \mathcal{A})$ and assume that, under P_θ, the coordinate process $\{X(t),\ t\in[0, \infty)\}$ has stationary, independent increments and characteristic function $f_t(u)=\exp[t\Psi_\theta(u)]$. It then follows that the process $X(t)$ is continuous in probability and $X(0)=0$ a.s. $[P_\theta]$. Here,

$$\Psi_\theta(u)=iu\beta(\theta)+\int_{\{0\}^c}\left(e^{iux}-1-\frac{iux}{1+x^2}\right)d\mu_\theta(x) \tag{1.1}$$

where, for each $\theta\in\Theta$, $\mu_\theta((-\infty, -a]\cup[a, \infty))<\infty$, for all $a>0$, and $\int_{-1}^{1} x^2 d\mu_\theta(x)<\infty$. The function $\Psi_\theta(u)$ is called the *exponent function* and μ_θ the *Lévy measure* of the process. If μ_θ is finite, $X(t)$ is a *jump process* or a *compound Poisson* process; if μ_θ is infinite, $X(t)$ is a *limit of jump processes*. In the next section we present a number of preliminary results that are needed in Sections 3 and 4 for the proofs of consistency and asymptotic normality respectively. In Section 5 we present a number of examples.

2. Some preliminary results

In this section we state some results which are needed in the rest of the paper. Unless otherwise stated, their proofs may be found in Akritas [1], Chapter 4, or in Akritas and Johnson [3].

The first proposition serves to delineate the relationship between Lévy measure and the first two moments of the process.

PROPOSITION 2.1. Let $X(t)$ be a process with stationary, independent increments and exponent function given by (1.1). Then if $\int_{\{0\}^c} x^2 d\mu_\theta(x) = \sigma^2 < \infty$ the first two moments exist and

$$\mathrm{E}\, X(t) = t\left(\beta + \int_{\{0\}^c} \frac{x^3}{1+x^2} d\mu_\theta(x)\right), \qquad \mathrm{Var}\, X(t) = t\sigma^2 .$$

Moreover, the additional assumption $\int_{\{0\}^c} |x| d\mu_\theta(x) < \infty$ implies that $\mathrm{E}\, X(t) = t\int_{\{0\}^c} x\, d\mu_\theta(x)$.

The next proposition generalizes a known result (see e.g., Breiman [5], Proposition 14.25) and is useful in determining distribution properties of certain random functions.

PROPOSITION 2.2. Let $X(t)$ be defined on $(\Omega, \mathcal{A}, P_\theta)$ with Lévy measure μ_θ and exponent function given by (1.1), and let g be a Borel measurable function such that $\int |g(x)| d\mu_\theta < \infty$. Then $Y(t) = \sum_i^{(t)} g(Z_i)$ is well defined, finite and $\{Y(t); t>0\}$ is a Lévy process with Lévy measure $\mu_\theta \circ g^{-1}$.

The symbol $\int$ used above, denotes integration over $\{0\}^c$. The symbol $\sum_i^{(t)}$ denotes summation over all jumps Z_i of $X(s)$, $0 \leqq s \leqq t$. Also $N_t(B)$ will denote the number of jumps of $X(s)$, $0 \leqq s \leqq t$ that are of size B, where B is any Borel set bounded away from zero.

PROPOSITION 2.3. Let $X(t)$ be as in Proposition 2.2 and let $a(x)$ be a Borel measurable function such that $\int \frac{|a(x)|^3}{1+[a(x)]^2} d\mu_\theta(x) < \infty$. For any partition B_m, $m \geqq 1$, of $R - \{0\}$, let

$$N_{mt} = N_t(B_m) \text{ and define } W(t) = \sum_{m=1}^{\infty} \left[\sum_{j=1}^{N_{mt}} a(Z_{mj}) - t\int_{B_m} a(x) d\mu_\theta(x) \right]$$

where Z_{mj} are the jumps of $X(s)$, $0 \leqq s \leqq t$, of size B_m. Then: (i) $W(t)$ is a well defined r.v. having an infinitely divisible distribution with Lévy measure $\mu_\theta \circ a^{-1}$; (ii) under the additional assumption that $\int [a(x)]^2 \cdot d\mu_\theta(x) < \infty$, $\mathcal{E}_\theta W(t) = 0$.

PROOF. (i) The r.v. $W(t)$ is the infinite convolution of the independent r.v.'s $W_m(t) = \sum_{j=1}^{N_{mt}} a(Z_{mj}) - t\int_{B_m} a(x)\, d\mu_\theta(x)$ where, according to Proposition 2.2, $W_m(t)$ has exponent function

$$iu\int_{B_m}\left[\frac{a(x)}{1+[a(x)]^2} - a(x)\right] d\mu_\theta + \int_{B_m}\left(e^{iux} - 1 - \frac{iux}{1+x^2}\right) d\mu_\theta \circ a^{-1}(x).$$

Summing these exponent functions over m we obtain

$$iu\int \frac{-[a(x)]^3}{1+[a(x)]^2}\, d\mu_\theta + \int\left(e^{iux} - 1 - \frac{iux}{1+x^2}\right) d\mu_\theta \circ a^{-1}(x).$$

Since the limiting exponent function is by assumption finite, the sequence of characteristic functions of $\sum_{m=1}^{n} W_m(t)$ converges and hence $W(t)$ is the limit in distribution of that series. But then (cf. Chung [8], p. 347) $W(t)$ is also the a.s. limit of the series.

(ii) Let $D_n = \bigcup_{m=1}^{n} B_m$. Then

$$\mathcal{E}_\theta\left[\sum_{m=1}^{n} W_m(t)\right]^2 = t\int_{D_n} [a(x)]^2 d\mu_\theta(x) \leqq t\int [a(x)]^2 d\mu_\theta(x) < \infty.$$

Therefore, the r.v.'s $\left[\sum_{m=1}^{n} W_m(t)\right]^r$, $n \geqq 1$, are uniformly integrable for all $r < 2$, so that $\mathcal{E}_\theta W(t) = \lim_n \mathcal{E}_\theta\left[\sum_{m=1}^{n} W_m(t)\right] = 0$.

Let $P_{T,\theta}$ denote the restriction of P_θ on $\mathcal{A}_T = \sigma\{X(s), 0 \leqq s \leqq T\}$.

PROPOSITION 2.4. For $\theta, \theta^* \in \Theta$, $P_{T,\theta}$ and P_{T,θ^*} are mutually absolutely continuous ($\approx$) if and only if

(L 1) $\mu_\theta \approx \mu_{\theta^*}$,

(L 2) $\int\left[1 - \left(\frac{d\mu_{\theta^*}}{d\mu_\theta}\right)^{1/2}\right]^2 d\mu_\theta < \infty$, and

(L 3) $\beta(\theta^*) - \beta(\theta) - \int \frac{x}{1+x^2}\, d(\mu_{\theta^*} - \mu_\theta)(x) = 0$.

The following useful remark is due to Newman [17].

Remark. Condition (L 2) implies $\int \frac{|x|}{1+x^2}\, d|\mu_{\theta^*} - \mu_\theta|(x) < \infty$. This, in particular, implies that the integral appearing in (L 3) exists and is finite.

PROPOSITION 2.5. Let $\theta, \theta^* \in \Theta$ and assume that $P_{T,\theta} \approx P_{T,\theta^*}$. Let $\{D_n\}$ be any sequence of neighborhoods of the origin such that $D_n \downarrow \{0\}$. Let $B_1 = D_1^c$, $B_m = D_{m-1} - D_m$, $m \geqq 2$, and set $N_{mT} = N_T(B_m)$. Then

$$\frac{dP_{T,\theta}}{dP_{T,\theta^*}}(X(t),\, t \in [0, T]) = \prod_{m=1}^{\infty}\left[e^{-T(\mu_\theta - \mu_{\theta^*})(B_m)} \prod_{j=1}^{N_{mT}} \frac{d\mu_\theta(Z_{mj})}{d\mu_{\theta^*}(Z_{mj})}\right].$$

It follows that

$$\begin{aligned}\Lambda_T(\theta) &= \log \frac{dP_{T,\theta}}{dP_{T,\theta^*}}(X(t),\, t \in [0, T]) \\ &= \sum_{m=1}^{\infty}\left[T \int_{B_m}\left(1 - \frac{d\mu_\theta(x)}{d\mu_{\theta^*}(x)}\right) d\mu_{\theta^*}(x) + \sum_{j=1}^{N_{mT}} \log \frac{d\mu_\theta(Z_{mj})}{d\mu_{\theta^*}(Z_{mj})}\right].\end{aligned}$$

Note. The proof of Proposition 2.5 was originally carried out in Akritas [1] using a method based on martingales. The same result was established independently and using a different method by Brockett, Hudson and Tucker [6].

3. Consistency

In this section we establish consistency of the MLE using the approach of Huber [14] which is relevant for robustness considerations. Methodologically, this approach is a generalization of the classical paper of Wald [22]. In the first case, the estimators correspond to jumps of $X(s)$, $0 \leqq s \leqq t$, of size greater than ε and, keeping t fixed, we let ε tend to 0. In the second case, the estimators correspond to the complete information of the realization of the process on $[0, t]$ and we let t tend to ∞.

Case A: Infinite Fisher information. For a Borel set B, bounded away from zero, define the process $X(B, t) = \sum_{s \leq t} [X(s) - X(s-)] I_B(X(s) - X(s-))$ whose jumps are of size B, and set $X_\varepsilon^+(t) = X([\varepsilon, \infty), t)$, $\varepsilon > 0$. Let Z_j, $j = 1, \cdots, N_\varepsilon$ denote the jumps of $X_\varepsilon^+(s)$, $0 \leqq s \leqq t$, and assume without loss of generality that $t = 1$. Define

$$Y_\varepsilon(\theta) = \sum_{i=1}^{N_\varepsilon} p(Z_i, \theta) - \int_{(\varepsilon, \infty)} a(x, \theta)\, d\mu_{\theta_0}(x) \tag{3.1}$$

for some functions $p(x, \theta)$, $a(x, \theta)$. (Note that for

$$p(x, \theta) = -\log d\mu_\theta(x), \qquad a(x, \theta) = 1 - d\mu_\theta(x)/d\mu_{\theta_0}(x) \tag{3.2}$$

$Y_\varepsilon(\theta)$ is minus the log-likelihood based on jumps of size greater than ε).

Let $\hat{\theta}_\epsilon=\hat{\theta}_\epsilon(X_\epsilon^+(s), 0\leqq s\leqq 1): \Omega\to\Theta$ be a sequence of estimators such that

$$\frac{1}{n}Y_{\epsilon_n}(\hat{\theta}_{\epsilon_n})-\frac{1}{n}\inf_{\theta\in\Theta}Y_{\epsilon_n}(\theta)\to 0 \text{ a.s. } [P_{\theta_0}] \tag{3.3}$$

for any sequence of positive numbers ε_n satisfying assumption (AC 3). We want to give sufficient conditions that any sequence of estimators satisfying (3.3) converges in P_{θ_0}-probability to θ_0.

ASSUMPTIONS (AC). (AC 1) Θ is locally compact.

(AC 2) For each fixed $\theta\in\Theta$, $p(x, \theta)$ is Borel measurable and $p(x, \theta)$ is separable as a stochastic process in θ.

(AC 3) Let ε_n be a sequence of positive numbers such that $\varepsilon_n\downarrow 0$. The quantities

$$\gamma_i(\theta)=\int_{B_i}(p(x, \theta)-a(x, \theta))d\mu_{\theta_0}(x),$$

where $B_1=(\varepsilon_1, \infty)$, $B_i=(\varepsilon_{i+1}, \varepsilon_i]$ are well defined, finite, and $\gamma_i(\theta)$ is lower semicontinuous in θ uniformly in i, that is

$$\inf\{\gamma_i(\theta'); \theta'\in U\}\to\gamma_i(\theta) \text{ uniformly in } i, \tag{3.4}$$

as the neighborhood U of θ shrinks to $\{\theta\}$. Moreover,

$$\bar{\gamma}_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\gamma_i(\theta)\to\gamma(\theta), \qquad \text{for all } \theta\in\Theta. \tag{3.5}$$

(AC 4) When $\theta_0\in\Theta$ obtains, $\gamma(\theta)>\gamma(\theta_0)$ for all $\theta\neq\theta_0$.

(AC 5) (i) For each compact subset K of Θ with $\theta_0\in K$ and for each $m\geqq 1$, there exists a function $g_m(x)$ such that

$$\int_{B_m}g_m(x)d\mu_{\theta_0}(x)<\infty \text{ and } |p(x, \theta)|\leqq g_m(x), \qquad \forall x\in B_m \text{ and } \forall\theta\in K.$$

(ii) There exists a sequence S_i of increasing measurable sets such that $\mu_{\theta_0}\left(R-\bigcup_{i=1}^{\infty}S_i\right)<\infty$ and that for each i, $\mu_{\theta_0}(B_m)p(x, \theta)$ is equicontinuous in θ for $x\in B_m\cap S_i$ uniformly in m.

(AC 6) For $\theta, \theta'\in\Theta$, $\sup_i\int_{B_i}|p(x, \theta')|d\mu_\theta(x)<\infty$ and $\sup_i\int_{B_i}|a(x, \theta')|\cdot d\mu_{\theta_0}(x)<\infty$. If Θ is not compact, let ∞ denote the point at infinity in its one point compactification.

(AC 7) (i) The quantity $\int_{B_i}(p(x, \theta)-a(x, \theta))d\mu_{\theta_0}(x)$ is differentiable with respect to each component θ_l of θ and is increasing as $\theta_l\to\infty$, $l=1, \cdots, k$. Moreover, the slope is bounded away from zero uniformly in $i\geqq 1$.

(ii) for $\theta, \theta' \in \Theta$, $\sup_i \int_{B_i} \left| \frac{\partial}{\partial \theta_l} p(x, \theta') \right| d\mu_\theta(x) < \infty$, $l=1, \cdots, k$.

Note. The sets B_i in (AC 7) are specified through a sequence $\varepsilon_n \downarrow 0$ as in (AC 3) but we don't require the sequence $\{\varepsilon_n\}$ to be the same. Actually the sequence can be different for each component θ_l, $l=1, \cdots, k$.

Remark. If Θ is compact (AC 7) is redundant. Condition (AC 7) is only used in the proof of Lemma 3.1 and the monotonicity it requires is, admittedly, strong. If in some particular case it is not satisfied, one may try to verify the conclusion of Lemma 3.1 by other methods.

LEMMA 3.1. *Under assumptions* (AC 1), (AC 2), (AC 3), (AC 6) *and* (AC 7), *there exists a compact set* $K \subseteq \Theta$ *such that any sequence* $\hat{\theta}_\varepsilon$ *satisfying condition* (3.3) *ultimately stays in* K *with probability tending to one.*

PROOF. First note that

$$\frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} (p(Z_{ij}, \theta_0) - a(Z_{ij}, \theta_0)) \right] \to \gamma(\theta_0) \tag{3.6}$$

in P_{θ_0}-probability, where $N_i = N(B_i)$ and Z_{ij}, $j=1, \cdots, N_i$ are the jumps of $X(t)$, $0 \leqq t \leqq 1$ of size B_i. Indeed from Breiman [5] pp. 310–312, the r.v.'s $W_i = \sum_{j=1}^{N_i} (p(Z_{ij}, \theta_0) - a(Z_{ij}, \theta_0))$ are independent. Moreover, assumption (AC 6) implies that, except for finite many i's, $\sup_i \operatorname{Var}(W_i) < \infty$ (see also Proposition 2.1) so that the law of large numbers for independent variables (cf. Gnedenko [13], p. 226 or 232) applies. Next, by Proposition 2.2, W_i has mean value $\gamma_i(\theta_0)$ so that (AC 3) implies (3.6). Similarly, assumption (AC 6) implies

$$\frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} a(Z_{ij}, \theta_0) - \int_{B_i} a(x, \theta_0) d\mu_{\theta_0}(x) \right] \to 0 \tag{3.7}$$

in P_{θ_0}-probability. From (3.6) and (3.7) it follows that for n sufficiently large and with probability greater than $1-\delta$,

$$\begin{aligned} \inf_{\theta \in \Theta} \frac{1}{n} Y_{\varepsilon_n}(\theta) &\leqq \frac{1}{n} Y_{\varepsilon_n}(\theta_0) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} (p(Z_{ij}, \theta_0) - a(Z_{ij}, \theta_0)) \right] \\ &\quad + \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} a(Z_{ij}, \theta_0) - \int_{B_i} a(x, \theta_0) d\mu_{\theta_0}(x) \right] \\ &\leqq \gamma(\theta_0) + \delta . \end{aligned} \tag{3.8}$$

Next we are going to show that there exists a compact set $K \subseteq \Theta$ such that for all n large enough

$$\inf_{\theta \notin K} \frac{1}{n} Y_{\epsilon_n}(\theta) \geqq \gamma(\theta_0) + \delta \tag{3.9}$$

with probability greater than $1-\delta$. Relations (3.8), (3.9) and condition (3.3) imply the result of the lemma.

In order to show (3.9) note that with probability tending to one as $n \to \infty$, the quantity $\frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} p(Z_{ij}, \theta) - \int_{B_i} a(x, \theta) d\mu_{\theta_0}(x) \right]$ is increasing as each coordinate θ_l of θ tends to ∞. Indeed, differentiating with respect to θ_l we obtain

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} \frac{\partial}{\partial \theta_l} p(Z_{ij}, \theta) - \frac{\partial}{\partial \theta_l} \int_{B_i} a(x, \theta) d\mu_{\theta_0}(x) \right] \\ &\quad = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} \frac{\partial}{\partial \theta_l} p(Z_{ij}, \theta) - \int_{B_i} \frac{\partial}{\partial \theta_l} p(x, \theta) d\mu_{\theta_0}(x) \right] \\ &\quad\quad + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_l} \int_{B_i} (p(x, \theta) - a(x, \theta)) d\mu_{\theta_0}(x) . \end{aligned}$$

By the law of large numbers and assumption (AC 7)-(ii) the first part of the above expression tends to zero while by (AC 7)-(i) the second part is either positive or negative as θ_l is on the right or left of θ_{0l}, respectively. Therefore, if K is a large enough compact set it follows that, with probability tending to one as $n \to \infty$, $\inf \left\{ \frac{1}{n} Y_{\epsilon_n}(\theta);\ \theta \notin K \right\} = \frac{1}{n} Y_{\epsilon_n}(\theta_1)$ where θ_1 belongs in the boundary of K. Thus, with probability tending to one as $n \to \infty$,

$$\begin{aligned} \inf_{\theta \notin K} \frac{1}{n} Y_{\epsilon_n}(\theta) &= \frac{1}{n} Y_{\epsilon_n}(\theta_1) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} (p(Z_{ij}, \theta_1) - a(Z_{ij}, \theta_1)) \right] \\ &\quad + \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} a(Z_{ij}, \theta_1) - \int_{B_i} a(x, \theta_1) d\mu_{\theta_0}(x) \right] \\ &\geqq \gamma(\theta_1) - \delta > \gamma(\theta_0) + \delta , \end{aligned}$$

where $\delta > 0$, for all n large enough, by (AC 3), (AC 4) and the law of large numbers.

The next lemma provides an extension of Theorem 1 of Rubin [19] in the context of Lévy processes.

LEMMA 3.2. *Let K be a compact subset of Θ containing θ_0 and let assumptions* (AC 2), (AC 5) *and* (AC 6) *hold. Then*

$$\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj},\theta)-\int_{B_m} p(x,\theta)d\mu_{\theta_0}(x)\right]\to 0$$

as $n\to\infty$ *a.s.* $[P_{\theta_0}]$ *uniformly in* $\theta\in K$.

PROOF. From the assumptions made we may select I large enough so that, for all $i\geqq I$

$$\int_{S^c_{m,i}} g_m(x)d\mu_{\theta_0}(x)<\frac{\varepsilon}{8}\,,\qquad \text{where } S_{m,i}=B_m\cap S_i\,, \tag{3.10}$$

holds except for finite many m. Since K is compact we may by (AC 5)-(ii), select $\theta_l\in K$, $l=1,\cdots,q$ and open neighborhoods K_l of θ_l with $K\subseteq\bigcup_{l=1}^{q} K_l$ and

$$\mu_{\theta_0}(B_m)|p(x,\theta)-p(x,\theta_l)|<\frac{\varepsilon}{4} \tag{3.11}$$

for all m, $\theta\in K_l$ and $x\in S_{m,I}$.

As usual, let Z_{mj} denote a jump of size B_m and define

$$U_{mj}=g_m(Z_{mj}) \text{ if } Z_{mj}\notin S_{m,I}\,,\qquad 0 \text{ otherwise}\,. \tag{3.12}$$

By definition of a.s. convergence, $\forall\,\varepsilon>0$, $\delta>0$ there exists $N>0$ with

$$\begin{aligned}\mathrm{P}_{\theta_0}\Bigg[\text{some } n>N,\ \Bigg|\frac{1}{n}\sum_{m=1}^{n}\Bigg[\sum_{j=1}^{N_m} p(Z_{mj},\theta_l)\\ -\int_{B_m} p(x,\theta_l)d\mu_{\theta_0}(x)\Bigg]\Bigg|\geqq\frac{\varepsilon}{4}\Bigg]<\frac{\delta}{2q}\,,\qquad l=1,\cdots,q\end{aligned} \tag{3.13}$$

and

$$\mathrm{P}_{\theta_0}\left[\text{some } n>N,\ \left|\frac{1}{n}\sum_{m=1}^{n}\left(\sum_{j=1}^{N_m} U_{mj}\right)\right|\geqq\frac{\varepsilon}{8}\right]<\frac{\delta}{2} \tag{3.14}$$

where for (3.14) we also used (3.10) and (3.12). Next, for each $\theta\in K_l$, $l=1,\cdots,q$

$$\begin{aligned}&\left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj},\theta)-\int_{B_m} p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|\\ &\leqq\left|\frac{1}{n}\sum_{m=1}^{n}\sum_{j=1}^{N_m}(p(Z_{mj},\theta)-p(Z_{mj},\theta_l))\right|\\ &\quad+\left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj},\theta_l)-\int_{B_m} p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|.\end{aligned} \tag{3.15}$$

Relation (3.11) implies that for $l=1,\cdots,q$,

$$\frac{1}{n}\sum_{m=1}^{n}|N_m(p(x_m,\theta)-p(x_m,\theta_l))|<\frac{\varepsilon}{4} \tag{3.16}$$

with probability tending to one as $n \to \infty$ uniformly in $\theta \in K_l$ and $x_m \in S_{m,I}$. Also for $\theta \in K_l$ there exists $x_m \in S_{m,I}$ with

$$\left|\sum_{j=1}^{N_m} (p(Z_{mj}, \theta) - p(Z_{mj}, \theta_l))\right| \leqq |N_m(p(x_m, \theta) - p(x_m, \theta_l))| + 2\sum_{j=1}^{N_m} U_{mj} . \tag{3.17}$$

Thus, by (3.16), (3.17) and (3.14),

$$P_{\theta_0}\left[\text{some } n > N,\ \left|\frac{1}{n}\sum_{m=1}^{n}\sum_{j=1}^{N_m}(p(Z_{mj}, \theta) - p(Z_{mj}, \theta_l))\right| \geqq \frac{\varepsilon}{2}\right] < \frac{\delta}{2} . \tag{3.18}$$

Relations (3.15), (3.18) and (3.13) imply

$$P_{\theta_0}\left[\text{some } n > N,\ \text{some } \theta \in K_l,\ \left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj}, \theta) - \int_{B_m} p(x, \theta_l)d\mu_{\theta_0}(x)\right]\right| \geqq \varepsilon\right] < \frac{\delta}{2} + \frac{\delta}{2q}\cdot q = \delta . \tag{3.19}$$

Finally, from (3.10) and (3.11), it follows that

$$\left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj}, \theta) - \int_{B_m} p(x, \theta)d\mu_{\theta_0}(x)\right]\right| \leqq \left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj}, \theta) - \int_{B_m} p(x, \theta_l)d\mu_{\theta_0}(x)\right]\right| + \frac{\varepsilon}{2}$$

so that, by (3.19)

$$\begin{aligned} &P_{\theta_0}\left[\text{some } n > N,\ \text{some } \theta \in K_l,\ \left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj}, \theta) - \int_{B_m} p(x, \theta)d\mu_{\theta_0}(x)\right]\right| \geqq 2\varepsilon\right] \\ &\leqq P_{\theta_0}\left[\text{some } n > N,\ \text{some } \theta \in K_l,\ \left|\frac{1}{n}\sum_{m=1}^{n}\left[\sum_{j=1}^{N_m} p(Z_{mj}, \theta) - \int_{B_m} p(x, \theta_l)d\mu_{\theta_0}(x)\right]\right| > \varepsilon\right] < \delta . \end{aligned}$$

THEOREM 3.1. *Let assumptions* (AC 1)–(AC 6) *hold. Then every sequence* $\hat{\theta}_t$ *satisfying* (3.3) *and the conclusion of Lemma* 3.1 *converges to* θ_0 *in* P_{θ_0}*-probability.*

PROOF. We restrict attention to the compact set K and let U be an open neighborhood of θ_0. By (AC 3) γ is lower semicontinuous. Indeed, if U is a neighborhood of θ, $\inf\{\gamma(\theta');\ \theta' \in U\} \geqq \lim_{n\to\infty} \sum_{i=1}^{n} \inf\{\gamma_i(\theta');\ \theta' \in U\}$ so that if $U \downarrow \{\theta\}$ it follows, via (3.4), that $\liminf_{U\downarrow\{\theta\}} \{\gamma(\theta');\ \theta' \in U\} \geqq \gamma(\theta)$. Thus its infimum on $K \backslash U$ is attained and by (AC 4) this is greater than $\gamma(\theta_0)$, say $\geqq \gamma(\theta_0) + 4\delta$, some $\delta > 0$. Next by (AC 3) again, each $\theta \in K \backslash U$ admits a neighborhood U_θ such that for all n large enough,

$$\frac{1}{n} \sum_{i=1}^{n} \inf \{\gamma_i(\theta');\ \theta' \in U_\theta\} \geq \gamma(\theta_0) + 3\delta \,. \tag{3.20}$$

Select a finite number of points θ_r, $r=1, \cdots, N$ such that $U_r = U_{\theta_r}$ cover $K \backslash U$ and (3.20) holds with $\theta=\theta_r$. For each $r=1, \cdots, N$,

$$\begin{aligned} \inf_{\theta \in U_r} n^{-1} Y_{\varepsilon_n}(\theta) &= \inf_{\theta \in U_r} n^{-1} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} p(Z_{ij}, \theta) - \int_{B_i} p(x, \theta) d\mu_{\theta_0}(x) + \gamma_i(\theta) \right] \\ &\geq \inf_{\theta \in U_r} n^{-1} \sum_{i=1}^{n} \left[\sum_{j=1}^{N_i} \left(p(Z_{ij}, \theta) - \int_{B_i} p(x, \theta) d\mu_{\theta_0}(x) \right) \right] \\ &\quad + \inf_{\theta \in U_r} n^{-1} \sum_{i=1}^{n} \gamma_i(\theta) \geq \gamma(\theta_0) + 2\delta \,, \end{aligned}$$

for all n large enough, by virtue of Lemma 3.2 and (3.20). Therefore, if (3.3) holds, $\hat{\theta}_\varepsilon$ should belong to U by virtue of (3.8).

The proof for $X_\varepsilon^-(t) = X((-\infty, -\varepsilon], t)$ and $X_\varepsilon(t) = X((-\varepsilon, \varepsilon)^c, t)$ carries over with identical arguments.

Case B: Finite Fisher information. Let $p(x, \theta)$, $a(x, \theta)$ be two measurable functions and assume that for each $\theta \in \Theta$, $t>0$, for any realization $X(s)$, $0 \leq s \leq t$, and for any partition B_m, $m \geq 1$, of $R - \{0\}$

$$Y_t(\theta) = \sum_{m=1}^{\infty} \left[\sum_{i=1}^{N_{mt}} p(Z_{mi}, \theta) - t \int_{B_m} a(x, \theta) d\mu_{\theta_0}(x) \right] \tag{3.21}$$

is well defined and finite, where $N_{mt} = N_t(B_m)$ and Z_{mi}, $i=1, \cdots, N_{mt}$ are the jumps of $X(s)$, $0 \leq s \leq t$, of size B_m. Note that for

$$p(x, \theta) = -\log \frac{d\mu_\theta(x)}{d\mu_{\theta_0}(x)} \,, \qquad a(x, \theta) = 1 - \frac{d\mu_\theta(x)}{d\mu_{\theta_0}(x)} \tag{3.22}$$

$Y_t(\theta)$ is $-\log \dfrac{dP_{T,\theta}}{dP_{T,\theta_0}}$ as defined in Proposition 2.5. Next, let $\hat{\theta}_T = \hat{\theta}_T(X(s),$ $s \in [0, T])$: $\Omega \to \Theta$ be a sequence of estimators such that

$$T^{-1} Y_T(\hat{\theta}_T) - \inf_{\theta \in \Theta} T^{-1} Y_T(\theta) \to 0 \text{ a.s. } [P_{\theta_0}] \,. \tag{3.23}$$

It will be shown that any sequence of estimators $\hat{\theta}_T$ satisfying (3.23) converges in P_{θ_0}-probability to θ_0 provided the following assumptions are met.

ASSUMPTIONS BC. (BC 1) Θ is locally compact.

(BC 2) For each $\theta \in \Theta$, $p(x, \theta)$, $a(x, \theta)$ and Borel measurable and separable as stochastic processes in θ.

(BC 3) (i) The quantity

$$\gamma(\theta)=\int_{\{0\}^c}(p(x,\theta)-a(x,\theta))d\mu_{\theta_0}(x)$$

is well defined, finite and

(ii) it is lower semicontinuous, that is

$$\inf\{\gamma(\theta');\ \theta'\in U\}\to\gamma(\theta)$$

as the neighborhood U of θ shrinks to $\{\theta\}$.

(BC 4) When $\theta_0\in\Theta$ obtains, $\gamma(\theta)>\gamma(\theta_0)$ for all $\theta\neq\theta_0$.

(BC 5) (i) The derivatives of $p(x,\theta)$, $a(x,\theta)$ with respect to θ_l, $l=1,\cdots,k$ exist and for $\theta\in K$, any compact set, $\left|\frac{\partial}{\partial\theta_l}p(x,\theta)\right|\leqq g_1(x)$, $\left|\frac{\partial}{\partial\theta_l}a(x,\theta)\right|\leqq g_2(x)$ with $\int_{\{0\}^c}g_i(x)d\mu_{\theta_0}(x)<\infty$, $i=1,2$.

(ii) The quantity $\gamma(\theta)$ is increasing as each component θ_l, $l=1,\cdots,k$ tends to infinity.

(iii) $\int_{\{0\}^c}(a(x,\theta))^2d\mu_{\theta_0}(x)<\infty$.

(BC 6) The function $p(x,\theta)$ is continuous in θ, $\forall x$, and for each compact subset K of Θ with $\theta_0\in K$ there exists a function $g_3(x)$ such that $|p(x,\theta)|\leqq g_3(x)$ for all x and for all $\theta\in K$, with $\int_{\{0\}^c}g_3(x)d\mu_{\theta_0}(x)<\infty$.

LEMMA 3.3. *Under assumptions* (BC 1), (BC 2), (BC 3)-(i) *and* (BC 5) *there exists a compact set* $K\subseteq\Theta$ *such that any sequence* $\hat{\theta}_T$ *satisfying condition* (3.23) *ultimately stays in* K *with probability tending to one.*

PROOF. First note that

$$T^{-1}S(T)=T^{-1}\sum_{m=1}^{\infty}\left[\sum_{i=1}^{N_{mT}}(p(Z_{mi},\theta_0)-a(Z_{mi},\theta_0))\right]\to\gamma(\theta_0) \tag{3.24}$$

in P_{θ_0}-probability, where $N_{mT}=N_T(B_m)$ is the number of jumps Z_{mi} of $X(t)$, $0\leqq t\leqq T$, of size B_m. Indeed if t_0 is an arbitrary but fixed positive number, define $S_j=S(jt_0)-S(jt_0-t_0)$, $j=1,\cdots,n$, where $n=[T/t_0]$ is the greatest integer smaller than T/t_0, so that $S(T)=\sum_{j=1}^{n}S_j+R_n$, where R_n is a remainder term. Thus, by (BC 3)-(i), the law of large numbers for i.i.d. r.v.'s applies (see also Propositions 2.1, 2.2). Similarly, by (BC 5)-(iii) and Propositions 2.1, 2.2,

$$T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}a(Z_{mj},\theta_0)-T\int_{B_m}a(x,\theta_0)d\mu_{\theta_0}(x)\right]\to 0 \tag{3.25}$$

in P_{θ_0}-probability. From (3.24) and (3.25) it follows that for T sufficiently large and with probability greater than $1-\delta$,

$$\inf_{\theta\in\Theta} T^{-1}Y_T(\theta)\leqq T^{-1}Y_T(\theta_0)\leqq\gamma(\theta_0)+\delta\,. \tag{3.26}$$

Again we prove the result of the lemma by showing that there exists a compact set $K\subseteq\Theta$ such that for all n large enough

$$\inf_{\theta\notin K}\frac{1}{T}Y_T(\theta)\geqq\gamma(\theta_0)+\delta \tag{3.27}$$

with probability greater than $1-\delta$. To do this note that

$$T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta)-T\int_{B_m}a(x,\theta)d\mu_{\theta_0}(x)\right]$$

is increasing as each coordinate θ_l of θ tends to ∞. Indeed, differentiating with respect to θ_l we obtain

$$T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}\frac{\partial}{\partial\theta_l}p(Z_{mj},\theta)-T\int_{B_m}\frac{\partial}{\partial\theta_l}a(x,\theta)d\mu_{\theta_0}(x)\right].$$

Note that by virtue of (BC 5)-(i) we can differentiate under the summation sign (see also Proposition 2.2) and further we can differentiate under the integral sign. We rewrite the above as

$$T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}\frac{\partial}{\partial\theta_l}p(Z_{mj},\theta)-T\int_{B_m}\frac{\partial}{\partial\theta_l}p(x,\theta)d\mu_{\theta_0}(x)\right]$$
$$+\int\frac{\partial}{\partial\theta_l}(p(x,\theta)-a(x,\theta))d\mu_{\theta_0}(x)\,.$$

By (BC 5)-(i) the first term tends to zero so that by (BC 5)-(ii) it is either positive or negative as θ_l is on the right or left of θ_{0l} respectively. The proof can now be completed by copying the concluding arguments of Lemma 3.1.

LEMMA 3.4. *Let K be a compact subset of Θ containing θ_0 and let assumptions* (BC 2), (BC 6) *hold. Then*

$$T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta)-T\int_{B_m}p(x,\theta)d\mu_{\theta_0}(x)\right]\to 0\,,\qquad as\ T\to\infty$$

a.s. $[P_{\theta_0}]$ *uniformly in* $\theta\in K$.

PROOF. Since K is compact we may, by (BC 6), select $\theta_l\in K$, $l=1,\cdots,q$ and open neighborhoods K_l of θ_l with $K\subseteq\bigcup_{l=1}^{q}K_l$ and

$$\left|\int(p(x,\theta)-p(x,\theta_l))d\mu_{\theta_0}(x)\right|\leqq\frac{\delta}{4},\quad\forall\,\theta\in K_l\,,\quad l=1,\cdots,q\,. \tag{3.28}$$

Thus

$$(3.29)\quad P_{\theta_0}\left[\text{some } n>N,\ \left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}(p(Z_{mj},\theta)-p(Z_{mj},\theta_l))\right]\right|>\frac{\varepsilon}{2}\right]<\frac{\delta}{2}.$$

Also, for $l=1,\cdots,q$

$$(3.30)\quad P_{\theta_0}\left[\text{some } n>N,\ \left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta_l)-\int_{B_m}p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|\geqq\frac{\varepsilon}{2}\right] \leqq\frac{\delta}{2q}.$$

Next, for $l=1,\cdots,q$,

$$\begin{aligned}&\left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta)-\int_{B_m}p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|\\&\leqq\left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}(p(Z_{mj},\theta)-p(Z_{mj},\theta_l))\right]\right|\\&\quad+\left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta_l)-\int_{B_m}p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|\end{aligned}$$

so that by (3.29), (3.30)

$$\begin{aligned}(3.31)\quad &P_{\theta_0}\Big[\text{some } n>N,\ \text{some } \theta\in K_l,\\&\left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta)-\int_{B_m}p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|\geqq\varepsilon\Big]\\&\leqq\frac{\delta}{2}+\frac{\delta}{2q}q=\delta.\end{aligned}$$

Finally, by (3.28), (3.31),

$$\begin{aligned}&P_{\theta_0}\Big[\text{some } n>N,\ \text{some } \theta\in K_l,\\&\left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta)-\int_{B_m}p(x,\theta)d\mu_{\theta_0}(x)\right]\right|\geqq\varepsilon+\frac{\delta}{4}\Big]\\&\leqq P_{\theta_0}\Big[\text{some } n>N,\ \text{some } \theta\in K_l,\\&\left|T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}}p(Z_{mj},\theta)-\int_{B_m}p(x,\theta_l)d\mu_{\theta_0}(x)\right]\right|\geqq\varepsilon\Big]\leqq\delta.\end{aligned}$$

THEOREM 3.2. *Let assumptions* (BC 1)-(BC 4) *and* (BC 6) *hold. Then every sequence* $\hat{\theta}_T$ *satisfying* (3.23) *and the conclusion of Lemma* 3.3 *converges to* θ_0 *in* P_{θ_0}-probability.

PROOF. We restrict attention to the compact set K and let U be an open neighborhood of θ_0. By (BC 3)-(ii) the infimum of $\gamma(\theta)$ on $K^*=K\backslash U$ is attained and, by (BC 4),

$$(3.32)\qquad \inf\{\gamma(\theta');\ \theta'\in K^*\}\geqq\gamma(\theta_0)+3\delta,\qquad \text{some } \delta>0.$$

Thus,

$$\begin{aligned}\inf_{\theta\in K^*} T^{-1}Y_T(\theta) &= \inf_{\theta\in K^*}\left\{T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}} p(Z_{mj},\theta)-T\int_{B_m} p(x,\theta)d\mu_{\theta_0}(x)\right]+\gamma(\theta)\right\}\\ &\geqq \inf_{\theta\in K^*} T^{-1}\sum_{m=1}^{\infty}\left[\sum_{j=1}^{N_{mT}} p(Z_{mj},\theta)-T\int_{B_m} p(x,\theta)d\mu_{\theta_0}(x)\right]+\inf_{\theta\in U_T}\gamma(\theta)\\ &\geqq \gamma(\theta_0)+2\delta\end{aligned}$$

for all n large enough, by virtue of Lemma 3.4 and (3.32). Therefore, if (3.23) holds, $\hat{\theta}_T$ should belong to U by virtue of (3.26).

4. Asymptotic normality

In this section we take Θ to be open, we assume that some $\theta_0\in\Theta$ obtains, i.e. P_{θ_0} is the true underlying probability measure, and we will impose assumptions that are local in nature. Detailed proofs will only be given for case B (finite Fisher information) while for case A we simply state the assumptions.

Case B: Finite Fisher information. Let $\eta(x,\theta)$ be an R^k-valued function on $R\times\Theta$ and set

$$V_t(\theta)=\sum_{m=1}^{\infty}\left[\sum_{i=1}^{N_{mt}}\eta(Z_{mi},\theta)-t\int_{B_m}\eta(x,\theta)d\mu_\theta(x)\right] \tag{4.1}$$

where B_m, $m\geqq 1$, is a partition of $R-\{0\}$, $N_{mt}=N_t(B_m)$, and Z_{mj}, $j=1,\cdots,N_{mt}$, the jumps of $X(s)$, $0\leqq s\leqq t$, whose size belongs in B_m. It follows from Proposition 2.2 and assumption (BN 2) below that $V_t(\theta)$ is well defined and finite a.s. P_θ (and hence a.s. P_{θ_0} if $P_{\theta_0}\approx P_\theta$) in a neighborhood of θ_0. Next, let $\hat{\theta}=\hat{\theta}_t(X(t), t\in[0,T])$: $\Omega\to\theta$ be a sequence of estimators of θ such that

$$T^{-1/2}V_T(\hat{\theta}_T)\to 0\,,\qquad \text{in } P_{\theta_0}\text{-probability}\,. \tag{4.2}$$

We want to give sufficient conditions ensuring that any sequence of estimators satisfying (4.2) is asymptotically normal. In the following $|\theta|=\max\{|\theta_1|,\cdots,|\theta_k|\}$.

ASSUMPTIONS (BN). (BN 1) For each fixed $\theta\in\Theta$, $\eta(x,\theta)$ is Borel measurable and $\eta(x,\theta)$ is separable as a stochastic process in θ.

(BN 2) $\int|\eta(x,\theta)|^i d\mu_\theta(x)<\infty$, $i=1,2$, for $|\theta-\theta_0|\leqq d_0$, some $d_0>0$. Let $u(x,\theta,\delta)=\sup\{|\eta(x,\theta)-\eta(x,\theta')|;\ |\theta-\theta'|\leqq\delta\}$. Then,

(BN 3) There exist strictly positive numbers α, β, γ such that

(i) $\int u(x,\theta,\delta)d\mu_{\theta_0}(x)\leqq\beta$ for $|\theta-\theta_0|+\delta\leqq d_0$.

(ii) $\int u(x,\theta,\delta)^2 d\mu_{\theta_0}(x)\leqq\gamma$ for $|\theta-\theta_0|+\delta\leqq d_0$.

Define $\lambda(\theta)$ by $\mathrm{E}_{\theta_0} V_t(\theta)=t\lambda(\theta)$, so that $\lambda(\theta)=\int \eta(x, \theta)d\mu_{\theta_0}(x)-\int \eta(x, \theta)d\mu_\theta(x)$, and $\lambda(\theta_0)=0$.

(iii) $|\lambda(\theta)|\geqq\alpha|\theta-\theta_0|$ for $|\theta-\theta_0|\leqq d_0$.

(BN 4) λ has a nonsingular derivative S at θ_0.

LEMMA 4.1. *Let*

$$Z_T(\theta', \theta)=\frac{|V_T(\theta')-V_T(\theta)-T\lambda(\theta')+T\lambda(\theta)|}{\sqrt{T}+T|\lambda(\theta')|}.$$

Assumptions (BN 1), (BN 2), (BN 3), *imply that as* $T\to\infty$

$$\sup\{Z_T(\theta, \theta_0);\ |\theta-\theta_0|\leqq d_0\}\to 0 \text{ in } P_{\theta_0}\text{-probability}.$$

PROOF. Assume that $\theta_0=0$ and $d_0=1$. Let $M\geqq 2$, $q=1/M$, $p=1-q$ and consider the subdivision of the cube $|\theta|\leqq 1$ into concentric cubes $C_m=\left\{\theta \,\middle|\, |\theta|\leqq p^m\right\}$, $m=0, 1, \cdots, m_0$. Split $C_{m-1}\backslash C_m$ into smaller cubes with edges of length $2d=p^{m-1}q$. It follows that there are $N<m_0(2M)^k$ cubes in $C_0\backslash C_{m_0}$ which we number $C_{(1)}, \cdots, C_{(N)}$. Note that if ξ is the center of a cube in $C_{m-1}\backslash C_m$, $|\xi|=p^{m-1}(1-q/2)$.

Next, given $\varepsilon>0$ and $\frac{1}{2}<f<1$, let $M\geqq\frac{3\beta}{\varepsilon\alpha}$ and define $m_0=m_0(T)$ by $p^{m_0}\leqq T^{-f}<p^{m_0-1}$. If then follows

$$(4.3)\qquad m_0(T)-1<\frac{f\log T}{|\log(1-q)|}\leqq m_0(T), \qquad \text{so that } N=O(\log T).$$

We will show that the r.h.s. of (4.4) tends to 0 as $T\to\infty$, where,

$$(4.4)\qquad \begin{aligned}&\mathrm{P}(\sup_{|\theta|\leqq 1} Z_T(\theta, 0)\geqq 2\varepsilon)\\ &\leqq \mathrm{P}(\sup_{\theta\in C_{m_0}} Z_T(\theta, 0)\geqq 2\varepsilon)+\sum_{j=1}^{N}\mathrm{P}(\sup_{\theta\in C_{(j)}} Z_T(\theta, 0)\geqq 2\varepsilon).\end{aligned}$$

Let ξ be the center of $C_{(i)}$ and $\theta\in C_{(i)}$. Then by (BN 3)-(i),

$$(4.5)\qquad |\lambda(\theta)-\lambda(\xi)|\leqq\int u(x, \xi, d)d\mu_{\theta_0}(x)\leqq\beta d\leqq\beta p^{m_0}q,$$

so that

$$(4.6)\qquad \begin{aligned} Z_T(\theta, 0)&\leqq\frac{|V_T(\theta)-V_T(\xi)-T\lambda(\theta)+T\lambda(\xi)|}{\sqrt{T}+T|\lambda(\theta)|}+\frac{|V_T(\xi)-V_T(0)-T\lambda(\xi)|}{\sqrt{T}+T|\lambda(\theta)|}\\ &\leqq\frac{\sum_i^{(T)} u(Z_i, \xi, d)+T\int u(x, \xi, d)d\mu_{\theta_0}(x)}{T\alpha p^{m_0}}\end{aligned}$$

$$+\frac{|V_T(\xi)-V_T(0)-T\lambda(\xi)|}{T\alpha p^{m_0}}=U_T+W_T$$

where in the second inequality we made use of (4.5) and the fact that, by (BN 3)-(iii), $|\lambda(\theta)|\geqq\alpha|\theta|\geqq\alpha p^{m_0}$, for $\theta\in C_{(i)}$. Next, by (BN 3)-(i), $\int u(x, \xi, \delta)d\mu_{\theta_0}(x)\leqq\beta\delta\leqq\beta p^{m_0}q$, which, with the definitions of M and q imply

$$\varepsilon\alpha p^{m_0}-2\int u(x, \xi, d)d\mu_{\theta_0}(x)\geqq\varepsilon\alpha p^{m_0}-2\beta qp^{m_0}\geqq\beta qp^{m_0}\,. \tag{4.7}$$

Thus,

$$\begin{aligned}
&\mathrm{P}\,(U_T\geqq\varepsilon)\\
&=\mathrm{P}\left\{\sum_i^{(T)}u(Z_i, \xi, d)-T\int u(x, \xi, d)d\mu_{\theta_0}\geqq T\varepsilon\alpha p^{m_0}-2T\int u(x, \xi, d)d\mu_{\theta_0}\right\}\\
&\leqq\mathrm{P}\left\{\sum_i^{(T)}u(Z_i, \xi, d)-T\int u(x, \xi, d)d\mu_{\theta_0}\geqq T\beta qp^{m_0}\right\}\\
&\leqq\frac{\gamma}{\beta^2pq}\cdot\frac{1}{Tp^{m_0-1}}\,,
\end{aligned}$$

by (4.7), (BN 3)-(ii) and Chebyshev's inequality. Similarly, $\mathrm{P}\,(W_T\geqq\varepsilon)\leqq\frac{\gamma}{9\beta^2q^2p^2}\frac{1}{Tp^{m_0-1}}$. Hence,

$$\mathrm{P}\,(\sup_{\theta\in C_{(i)}}Z_T(\theta, 0)\geqq2\varepsilon)\leqq KT^{f-1}\,,\qquad \text{where } K=\frac{\gamma}{\beta^2pq}+\frac{\gamma}{9\beta^2q^2p^2}\,. \tag{4.8}$$

Moreover, with $d=(1-q)^{m_0}\leqq T^{-f}$, and for T large enough,

$$\begin{aligned}
&\mathrm{P}\,(\sup_{\theta\in C_{m_0}}Z_T(\theta, 0)\geqq2\varepsilon)\\
&\leqq\mathrm{P}\left(\sum_i^{(T)}u(Z_i, 0, d)-T\int u(x, 0, d)d\mu_{\theta_0}(x)\geqq\sqrt{T}\,\varepsilon-2T\int u(x, 0, d)d\mu_{\theta_0}\right)\\
&\leqq\mathrm{P}\left(\sum_i^{(T)}u(Z_i, 0, d)-T\int u(x, 0, d)d\mu_{\theta_0}\geqq\sqrt{T}\,\varepsilon\right),
\end{aligned}$$

since $\int u(x, 0, d)d\mu_{\theta_0}<\beta d\leqq\beta T^{-f}$, so that, for T large enough, $2\sqrt{T}\,\varepsilon-2T\int u(x, 0, d)d\mu_{\theta_0}\geqq\sqrt{T}\,\varepsilon$. Thus, by Chebyshev's inequality,

$$\mathrm{P}\,(\sup_{\theta\in C_{m_0}}Z_T(\theta, 0)\geqq2\varepsilon)\leqq\gamma\varepsilon^{-2}T^{-f}\,. \tag{4.9}$$

Combining (4.3), (4.4), (4.8), and (4.9) we have $\mathrm{P}\,(\sup_{\theta\in C_0}Z_T(\theta, 0)\geqq2\varepsilon)\leqq O(T^{-f})+O(T^{f-1}\log T)$ which proves the lemma.

COROLLARY 4.1. *Suppose that* $\hat{\theta}_T$ *is a consistent estimator satisfying* (4.2). *Then, under assumptions* (BN 1), (BN 2), (BN 3),

$$T^{-1/2}V_T(\theta_0)+T^{1/2}\lambda(\hat{\theta}_T)\to 0 \qquad \text{in } P_{\theta_0}\text{-}probability\,.$$

PROOF. The identity

$$V_T(\theta_0)+T\lambda(\hat{\theta}_T)=[V_T(\theta_0)-V_T(\hat{\theta}_T)+T\lambda(\hat{\theta}_T)]+V_T(\hat{\theta}_T)$$

and the fact that $P_{\theta_0}(|\hat{\theta}_T-\theta_0|\leqq d_0)\to 1$, as $T\to\infty$, imply that with probability tending to one,

$$\frac{V_T(\theta_0)+T\lambda(\hat{\theta}_T)}{\sqrt{T}+T|\lambda(\hat{\theta}_T)|}\leq \sup_{|\theta-\theta_0|\leq d_0} Z_T(\theta,\theta_0)+T^{-1/2}V_T(\hat{\theta}_T)\,. \tag{4.10}$$

Therefore, Lemma 4.1 and assumption (4.2) imply that the left-hand side of (4.10) tends to zero in P_{θ_0}-probability, so that, for sufficient large T,

$$P\left(|T^{-1/2}V_T(\theta_0)+T^{1/2}\lambda(\hat{\theta}_T)|>\varepsilon(1+T^{1/2}|\lambda(\hat{\theta}_T)|)\right)\leqq\frac{\varepsilon}{4}\,. \tag{4.11}$$

By Chebyshev's inequality,

$$P\left(T^{-1/2}V_T(\theta_0)\geqq D\right)\leqq\frac{\varepsilon}{4}\,, \qquad \text{where } D^2=2\int[n(x,\theta_0)]^2d\mu_{\theta_0}(x)\Big/\varepsilon\,. \tag{4.12}$$

From (4.11) and (4.12) we have

$$P\left(T^{1/2}|\lambda(\hat{\theta}_T)|\geqq\frac{D+\varepsilon}{1-\varepsilon}\right)\leqq\frac{\varepsilon}{2}$$

which, together with (4.11) implies

$$P\left(|T^{-1/2}V_T(\theta_0)+T^{1/2}\lambda(\hat{\theta}_T)|\geqq\frac{(D+1)\varepsilon}{1-\varepsilon}\right)\leqq\varepsilon\,.$$

A straightforward application of Corollary 4.1 and the multivariate central limit theorem yields

THEOREM 4.1. *Suppose that $\hat{\theta}_T$ is a consistent estimator such that it satisfies* (4.2). *Then, under assumptions* (BN 1)–(BN 5)

$$\mathcal{L}[T^{1/2}(\hat{\theta}_T-\theta_0)|P_{\theta_0}]\Longrightarrow N(0,S^{-1}\Gamma(S')^{-1})$$

where $\Gamma=\int[\eta(x,\theta_0)\eta'(x,\theta_0)]d\mu_{\theta_0}$.

Assume now that for $\theta,\theta^*\in\Theta$, $P_{T,\theta}$ and P_{T,θ^*} are mutually absolutely continuous (see Proposition 2.3 for conditions), fix $\theta^*\in\Theta$ and let the logarithm $\Lambda_T(\theta)$ of the Radon-Nikodym derivative of $P_{T,\theta}$ with respect to P_{T,θ^*} be given by the expression in Proposition 2.4. Let

$$\eta(x,\theta)=\frac{\partial}{\partial\theta}\log\phi(x,\theta)\,, \qquad \text{where } \phi(x,\theta)=\frac{d\mu_\theta(x)}{d\mu_{\theta^*}(x)} \tag{4.13}$$

and assume that assumptions (BN 1)-(BN 4) hold locally uniformly in θ_0. Assumptions (BN 2), (BN 3)-(ii) and Proposition 2.2 imply that $\Lambda_T(\theta)$ may be differentiated under the summation and integral signs so that $V_t(\theta)$ in (4.1) with $\eta(x, \theta)$ given by (4.13) is indeed the likelihood estimating function. Hence there exists a sequence of maximum likelihood estimators $\hat{\theta}_T$ such that (4.2) is satisfied. Next it is easy to show, perhaps under some additional assumptions, that the derivative of λ at θ_0 is $-\Gamma(\theta_0)=-\int[\eta(x, \theta_0)\eta'(x, \theta_0)]d\mu_{\theta_0}$ (see Huber [14], p. 231) and that $\Gamma(\theta_0)$ plays the role of Fisher's information matrix in the asymptotic theory of estimation. Thus the efficiency of the maximum likelihood estimator follows from Theorem 4.1.

Case A: Infinite Fisher information. Let $X_\varepsilon^+(s)$, $0\leqq s\leqq 1$, and Z_j, $j=1, \cdots, N$ be as in Section 3. Let $\eta(x, \theta)$ be an R^k-valued function on $R\times\Theta$ and set

$$V_\varepsilon(\theta)=\sum_{j=1}^{N}\eta(Z_j, \theta)-\int_{(\varepsilon,\infty)}\eta(x, \theta)d\mu_\theta(x)\,. \tag{4.14}$$

Let $\hat{\theta}_\varepsilon=\hat{\theta}_\varepsilon(X_\varepsilon^+(s), 0\leqq s\leqq 1): \Omega\rightarrow\Theta$ be a sequence of estimators such that

$$\frac{1}{n}V_{\varepsilon_n}(\hat{\theta}_{\varepsilon_n})\rightarrow 0\,, \qquad \text{in } P_{\theta_0}\text{-probability} \tag{4.15}$$

for any sequence of positive numbers ε_n satisfying assumption (AN 2). The assumptions under which a sequence of estimators $\hat{\theta}_\varepsilon$ satisfying (4.15) is asymptotically normal are given below.

ASSUMPTIONS (AN). (AN 1) Same as (BN 1).

(AN 2) Let ε_n be a sequence of positive numbers such that $\varepsilon_n\downarrow 0$, the quantities

$$\lambda_i(\theta)=\int_{B_i}\eta(x, \theta)d\mu_{\theta_0}(x)-\int_{B_i}\eta(x, \theta)d\mu_\theta(x) \tag{4.16}$$

where $B_1=(\varepsilon_1, \infty)$, $B_i=(\varepsilon_{i+1}, \varepsilon_i]$ are well defined, finite, and

$$\bar{\lambda}_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\lambda_i(\theta)\rightarrow\lambda(\theta)\,, \qquad \theta\in\Theta\,. \tag{4.17}$$

It follows that $\lambda(\theta_0)=0$.

(AN 3) $\sup_i\int_{B_i}|\eta(x, \theta)|d\mu_\theta(x)<\infty$ and $\sup_i\int_{B_i}|\eta(x, \theta)|^2d\mu_{\theta_0}(x)<\infty$ for $|\theta-\theta_0|\leqq d_0$, some $d_0>0$.

Let $u(x, \theta, \delta)=\sup\{|\eta(x, \theta)-\eta(x, \theta')|;\ |\theta-\theta'|\leqq\delta\}$. Then

(AN 4) There exist strictly positive numbers α, β, γ such that

(i) $\sup_i \int_{B_i} u(x, \theta, \delta) d\mu_{\theta_0}(x) \leqq \beta\delta$ for $|\theta-\theta_0|+\delta \leqq d_0$

(ii) $\sup_i \int_{B_i} [u(x, \theta, \delta)]^2 d\mu_{\theta_0}(x) \leqq \gamma\delta$ for $|\theta-\theta_0|+\delta \leqq d_0$

(iii) $|\lambda(\theta)| \geqq \alpha|\theta-\theta_0|$ for $|\theta-\theta_0| \leqq d_0$.

(AN 5) λ has a nonsingular derivative S at θ_0.

5. Some examples

In this section we present four examples that illustrate the generality of the theory.

Example 1. Consider the class of compound Poisson processes. Here the Lévy measure is of the form $\lambda \cdot F$ where, λ is the intensity and F denotes both the distribution function of the jumps and the corresponding measure. F may be any of the standard parametric families and λ may be an additional parameter. It is then easy to check that for many parametric families, minus the log-likelihood function [likelihood estimating function] satisfies Assumptions (BC) [Assumptions (BN)].

Example 2 (Parametric signal detection). Consider now the following general class of problems. Let μ be a Lévy measure, and let $X(t)$ be the corresponding process (e.g. μ could be the Lévy measure corresponding to a stable process). Further let ν_θ be a second Lévy measure, corresponding to the process $Y(t)$, that depends on a parameter θ (e.g. ν_θ could be θ_1 $(N(\theta_2, \theta_3))$, $\theta_1, \theta_3>0$, $\theta_2 \in R$, where $N(\theta_2, \theta_3)$ denotes the normal distribution with mean θ_2 and variance θ_3). It is assumed that we observe $Z(t)=X(t)+Y(t)$ and want to estimate θ. It is easily seen that the process $Z(t)$ has Lévy measure $\mu+\nu_\theta$.

Example 3. Let $\{X(t), t \in (0, \infty)\}$ be the gamma process. That is, the characteristic function of $X(t)$ under P_θ is $f_t(u)=\left(1-i\frac{u}{\theta}\right)^{-t}$, $\theta>0$. The exponent function is given by (1.1) with (cf. Feller [10], p. 567)

$$\beta(\theta)=\int_{(0,\infty)} \frac{e^{-\theta x}}{1+x^2} dx \quad \text{and} \quad \mu_\theta(A)=\int_{A\cap(0,\infty)} \frac{e^{-\theta x}}{x} dx\,, \quad A \in B\,. \tag{5.1}$$

In order to check whether Assumptions (BC) are satisfied for minus the log-likelihood, note that (5.1) implies

$$p(x, \theta)=(\theta-\theta_0)x\,, \qquad a(x, \theta)=1-e^{-(\theta-\theta_0)x}\,. \tag{5.2}$$

It follows that assumption (BC 2) holds since $p(x, \theta)$, $a(x, \theta)$ are continuous in x for each fixed θ and continuous in θ for each fixed x.

Next (BC 3) is easily seen to be satisfied while (BC 4) holds by the well-known properties of the Kullback-Leibler information number. For (BC 5)-(i) take $g_1(x)=x$, $g_2(x)=xe^{-(\varepsilon-\theta_0)x}$, where $0<\varepsilon\leqq\inf(\theta;\ \theta\in K)$. (BC 5)-(ii) is satisfied since the integrand of $\gamma(\theta)$ possesses the stated property, while (BC 5)-(iii) is easily seen to be satisfied. Finally, for (BC 6) take $g_3(x)=cx$, where $c=\sup(\theta-\theta_0;\ \theta\in K)$.

Next, in order to check that Assumptions (BN) are satisfied note that $\eta(x, \theta)=x$, so that (BN 1), (BN 2) and (BN 3)-(i), (ii) are clearly satisfied. Also since $\lambda(\theta)=\frac{\theta-\theta_0}{\theta\theta_0}$, (BN 4) is satisfied and (BN 3)-(iii) holds with $a=\theta_0(\theta_0+d_0)$.

Example 4. Let $X(t)$, $t\geqq 0$ be a stable process so that $d\mu_\theta(x)=\alpha\beta x^{-1-\beta}dx$, $\theta=(\alpha, \beta)'$, $\alpha>0$, $0<\beta<1$. Thus in order to check that Assumptions (AC) are satisfied for minus the log-likelihood, take

$$p(x, \theta)=-\log(\alpha\beta)+(1+\beta)\log x\,, \qquad a(x, \theta)=1-\frac{\alpha\beta}{\alpha_0\beta_0}x^{-(\beta-\beta_0)}\,. \tag{5.3}$$

It is easy to see that if we select the sequence ε_n, $n\geqq 1$ in such a way that $\sup_i[(\varepsilon_{i+1}^{-\beta_0}-\varepsilon_i^{-\beta_0})\cdot\log\varepsilon_{i+1}^{-1}]<\infty$ then (AC 3), (AC 6) will hold while (AC 4) holds by the well-known properties of the Kullback-Leibler information number. (AC 5)-(i) is clearly true while the condition $\sup_i[(\varepsilon_{i+1}^{-\beta_0}-\varepsilon_i^{-\beta_0})\log\varepsilon_{i+1}^{-1}]<\infty$ implies that $\mu_{\theta_0}(B_i)p(x, \theta)=\alpha_0[\varepsilon_{i+1}^{-\beta_0}-\varepsilon_i^{-\beta_0}]\cdot[(1+\beta)\log x-\log(\alpha\beta)]$ is equicontinuous in θ for $x\in B_i$ uniformly in i, so that (AC 5)-(ii) is also satisfied. Next to check (AC 7)-(i) for the parameter α take $\{\varepsilon_n\}$ so that $(\varepsilon_i^{-\beta_0}-\varepsilon_{i+1}^{-\beta_0})$ remains bounded away from zero and note that the derivative of $\int_{B_i}(p(x, \theta)-a(x, \theta))d\mu_{\theta_0}(x)$ with respect to α evaluated at (α_0, β_0) is $(\varepsilon_i^{-\beta_0}-\varepsilon_{i+1}^{-\beta_0})\left(\frac{\alpha_0}{\alpha}-1\right)$. (AC 7)-(ii) is easily seen to be satisfied. Similarly for the parameter β.

Acknowledgement

I am thankful for the careful remarks of a referee that helped improve the paper significantly.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

[1] Akritas, M. G. (1978). Contiguity of probability measures associated with continuous time stochastic processes, Ph.D. dissertation, University of Wisconsin-Madison.

[2] Akritas, M. G. (1981). Rates of convergence of estimates for Lévy processes, in preparation.

[3] Akritas, M. G. and Johnson, R. A. (1981). Asymptotic inference in Lévy processes of the discontinuous type, *Ann. Statist.*, **9**, 604-614.

[4] Basawa, I. V. and Brockwell, P. J. (1978). Inference for gamma and stable processes, *Biometrika*, **65**, 129-133.

[5] Breiman, L. (1968). *Probability*, Addison-Wesley.

[6] Brockett, P. L., Hudson, W. N. and Tucker, H. G. (1978). The distribution of the likelihood ratio for additive processes, *Notices Amer. Math. Soc.*, A-149 (to appear in *J. Multivariate Anal.*).

[7] Brockwell, P. J. and Chung, K. L. (1975). Emptiness times of a dam with stable input and general release function, *J. Appl. Prob.*, **12**, 212-217.

[8] Chung, K. L. (1974). *A Course in Probability Theory*, Academic Press.

[9] Fama, E. F. (1965). The behavior of stock market prices, *J. of Business of the University of Chicago*, **XXXVIII**, 38-105.

[10] Feller, W. (1971). *An Introduction to Probability Theory and its Applications* II, John Wiley & Sons.

[11] Fristedt, B. (1972). Sample functions of stochastic processes with stationary, independent increments, *Tech. Rep.* No. 1166, M.R.C., University of Wisconsin-Madison.

[12] Frost, P. (1972). Optimal estimation and detection for conditionally independent increment processes, *1972 IEEE International Symposium on Information Theory*, Asilomar, Pacific Grove.

[13] Gnedenko, B. V. (1963). *The Theory of Probability*, Chelsea.

[14] Huber, P. J. (1967). The behavior of maximum likelihood estimators under nonstandard conditions, *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, **1**, 221-233.

[15] Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator, *Ann. Inst. Statist. Math.*, **25**, 1-26.

[16] Moran, P. A. P. (1959). *The Theory of Storage*, Methuen, London.

[17] Newman, C. M. (1973). On the orthogonality of independent increment processes, *Topics of Probability Theory* (eds. D. W. Stroock and S. R. S. Varathan), Courant Institute of Math. Science, N.Y.

[18] Rao, B. L. S. Prakasa (1972). Maximum likelihood estimation for Markov processes, *Ann. Inst. Statist. Math.*, **24**, 333-345.

[19] Rubin, H. (1956). Uniform convergence of random functions with applications to statistics, *Ann. Math. Statist.*, **27**, 200-203.

[20] Rubin, H. and Tucker, H. G. (1959). Estimating the parameters of a differential process, *Ann. Math. Statist.*, **30**, 641-658.

[21] Segall, A. and Kailath, T. (1975). Radon-Nikodym derivatives with respect to measures induced by discontinuous independent increment processes, *Ann. Prob.*, **3**, 449-464.

[22] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, **20**, 595-601.