Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

On efficiency bounds for estimating the offspring mean in a branching process

Alexander A. Gushchin

submitted: 19th September 1995

Steklov Mathematical Institute
Vavilova 42
117966 Moscow GSP–1
Russia

Preprint No. 175
Berlin 1995

1991 Mathematics Subject Classification. 62M05, 62G05.
Key words and phrases. Branching process, efficient estimation, local asymptotic quadraticity, Fisher information, Hellinger process.

Edited by
Weierstraß–Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
D — 10117 Berlin
Germany

Fax: + 49 30 2044975
e–mail (X.400): c=de;a=d400-gw;p=WIAS-BERLIN;s=preprint
e–mail (Internet): preprint@wias-berlin.de

ON EFFICIENCY BOUNDS FOR ESTIMATING THE OFFSPRING MEAN IN A BRANCHING PROCESS

Alexander A. Gushchin

Steklov Mathematical Institute, Moscow

ABSTRACT

Suppose that we observe the branching process with nonrandom immigration

$$x_k = \sum_{i=1}^{x_{k-1}} y_{k,i} + 1, \quad k = 1, 2, \dots, \quad x_0 = 1,$$

$y_{k,i}$ are i.i.d. random variables with an unknown distribution p on $\{0, 1, 2, \dots\}$ with finite second moment. We study the problem of efficient estimation of the offspring mean $\vartheta(p)$ on observing a single realization $\{x_1, \dots, x_n\}$. For a sequence of estimators $\widetilde{\vartheta}_n = \widetilde{\vartheta}_n(x_1, \dots, x_n)$ satisfying some "regularity" properties with respect to "small perturbations" near a point p, we prove an asymptotic lower bound on the deviation $\widetilde{\vartheta}_n - \vartheta(p)$. This lower bound has the same form in the subcritical, the supercritical and the critical cases. The estimator

$$\widehat{\vartheta}_n = \sum_{k=1}^{n} (x_k - 1) \left(\sum_{k=1}^{n} x_{k-1} \right)^{-1}$$

is asymptotically efficient at every point p in our approach.

1. INTRODUCTION

Suppose that we observe a stochastic process with an unknown distribution P which is assumed to belong to some large nonparametric family $\mathcal{P}$, and we wish to estimate a one-dimensional functional $\vartheta\colon \mathcal{P} \to \mathbf{R}^1$ of this family.

One can suggest the following scheme for the problem of asymptotically efficient estimation of ϑ.

1. For a fixed $P \in \mathcal{P}$, we introduce a class $\mathcal{P}(P)$ of one-parameter local submodels around the "point" P. Usually every submodel in $\mathcal{P}(P)$ must be approximated by a limit statistical experiment from a certain class.

2. Given an arbitrary submodel in $\mathcal{P}(P)$, we find a lower bound on the asymptotic performance of estimators in the parametric problem of estimating ϑ in this submodel.

3. We choose a "least favorable submodel" in $\mathcal{P}(P)$, i.e., a submodel for which the lower bound (Step 2) is maximized among the class $\mathcal{P}(P)$.

4. An estimator is said to be asymptotically efficient at the point P if it attains the lower bound corresponding to a "least favorable submodel" in $\mathcal{P}(P)$.

5. An estimator is said to be asymptotically efficient in $\mathcal{P}$ if it is asymptotically efficient at every point $P \in \mathcal{P}$.

To apply this scheme to a particular statistical model, we must specify the class $\mathcal{P}(P)$ (Step 1) and the asymptotic lower bound theorem (Step 2). It is important to take the class $\mathcal{P}(P)$ large enough in order not to miss a "least favorable submodel", so the appropriate choice of $\mathcal{P}(P)$ depends on the model under consideration. If we observe a stochastic process of ergodic type (in particular, if the observations are independent and identically distributed), the widely used approach consists of considering locally asymptotically normal submodels, i.e., submodels which can be approximated by Gaussian shift experiments. Then it is natural to take the asymptotic lower bounds given by the convolution theorem or by the asymptotic minimax theorem at Step 2. The lower bounds given by these theorems are inversely proportional to the value of the Fisher information (about ϑ) in the limit experiment. So both these theorems lead to the same description of the "least favorable submodels" and, moreover, they lead to the same description of asymptotically efficient estimators. This approach is originally due to Levit (Levit, 1973, 1975), see also the exposition in (Ibragimov and Has'minskii, 1981, Chapter 4.1). The problem of constructing estimators which are asymptotically efficient in $\mathcal{P}$ under this approach has been successfully solved for a wealth of semiparametric models, see e.g. (Bickel et al., 1993).

In this paper we consider a specially chosen example of estimating the offspring mean $\vartheta(P)$ in a branching process with nonrandom immigration. This process exhibits qualitatively different behavior for different parameter values. It is ergodic only if $\vartheta(P) < 1$ (the subcritical case). If $\vartheta(P) > 1$ (the supercritical case) then it grows exponentially fast, and it is unstable if $\vartheta(P) = 1$ (the critical case).

In this model the class of locally asymptotically normal submodels is too poor for "points" P with $\vartheta(P) \geq 1$. We suggest considering a larger class of one-parameter local submodels which can be approximated by a limit experiment of a rather general form, not necessarily by Gaussian shift experiments. We do not use the convolution theorem or the asymptotic minimax theorem as the asymptotic lower bound theorems at Step 2, though there are different versions of these theorems which are valid for much more general families of experiments than locally asymptotically normal families, see e.g. (Le Cam, 1986; Greenwood and Wefelmeyer, 1993; Shiryaev and Spokoiny, 1994). The reasons are that, firstly, it is not always possible to find a "least favorable submodel" (Step 3) or, secondly, the corresponding bounds are not sharp and asymptotically efficient estimators do not exist. Instead of this, we use the asymptotic lower bound given by an asymptotic version of the Cramér–Rao inequality, see (Gushchin, 1995b). The advantage of this approach is that the corresponding lower bound is inversely proportional to the value of the Fisher information (about ϑ) in the limit experiment again, which makes it possible to find a "least favorable submodel" easily. This lower bound is not always sharp, but it is sharp for locally asymptotically quadratic families, cf. (Gushchin, 1995a). Since in our model it is possible to find a "least favorable submodel" which is locally asymptotically quadratic, we can describe the estimators which are asymptotically efficient at a fixed P. Finally, we find an estimator which is asymptotically efficient in $\mathcal{P}$ in our model.

The following notation and definitions are used throughout the paper. $\mathbf{N} = \{1, 2, \ldots\}$ is the set of natural numbers, $\mathbf{N}_0 = \mathbf{N} \cup \{0\}$. Let P and $\widetilde{P}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$. We denote by $d\widetilde{P}/dP$ the density of $\widetilde{P}$ with respect to P defined as $d\widetilde{P}/dP = \widetilde{\mathfrak{z}}/\mathfrak{z}$, where $\widetilde{\mathfrak{z}}$ and $\mathfrak{z}$ are the Radon–Nikodym densities of $\widetilde{P}$ and P respectively with respect to $Q = (P + \widetilde{P})/2$. The distance in variation $\|P - \widetilde{P}\|$ between P and $\widetilde{P}$ is defined by

$$\|P - \widetilde{P}\| = 2 \sup_{B \in \mathcal{F}} |P(B) - \widetilde{P}(B)| = \int |\widetilde{\mathfrak{z}} - \mathfrak{z}| \, dQ = \int \left| \frac{d\widetilde{P}}{dP} - 1 \right| dP + \widetilde{P}\left\{ \frac{d\widetilde{P}}{dP} = \infty \right\}.$$

The Hellinger distance $\rho(P, \widetilde{P})$ between P and $\widetilde{P}$ is defined by

$$\begin{aligned}\rho^2(P, \widetilde{P}) &= \frac{1}{2} \int (\sqrt{\widetilde{\mathfrak{z}}} - \sqrt{\mathfrak{z}})^2 \, dQ = \frac{1}{2} \int (\sqrt{d\widetilde{P}/dP} - 1)^2 \, dP + \frac{1}{2} \widetilde{P}\left\{ \frac{d\widetilde{P}}{dP} = \infty \right\} \\ &= 1 - H(P, \widetilde{P}),\end{aligned}$$

where $H(P, \widetilde{P}) = \int \sqrt{\mathfrak{z}\widetilde{\mathfrak{z}}} \, dQ = \int \sqrt{d\widetilde{P}/dP} \, dP$ is the Hellinger integral (of order $\frac{1}{2}$) between P and $\widetilde{P}$.

Let $(P^\alpha, \alpha \in A \subseteq \mathbf{R}^1)$ be a family of probability measures on $(\Omega, \mathcal{F})$ and $\alpha_0 \in A$ a cluster point of A. Denote $Z^\alpha = dP^\alpha / dP^{\alpha_0}$. We shall say that (P^α) is L^2-*differentiable* at α_0 if there exists a random variable $V \in L^2(P^{\alpha_0})$ such that

$$\frac{\sqrt{Z^\alpha} - 1}{\alpha - \alpha_0} \longrightarrow \frac{V}{2} \quad \text{in} \quad L^2(P^{\alpha_0})$$

and

$$P^\alpha\{Z^\alpha = \infty\} = o(|\alpha - \alpha_0|^2)$$

as $\alpha \to \alpha_0$. In this case V is called the *score function* of (P^α) at α_0, we have $E^{\alpha_0}[V] = 0$ and the variance

$$I = \mathrm{Var}^{\alpha_0}[V] = E^{\alpha_0}[V^2] = 8 \lim_{\alpha \to \alpha_0} \frac{\rho^2(P^{\alpha_0}, P^\alpha)}{(\alpha - \alpha_0)^2} \tag{1.1}$$

is called the *Fisher information* about α in (P^α) at α_0 (where E^α and Var^α are the expectation and the variance respectively under P^α).

The next proposition is a slightly modified version of the result on differentiating under the expectation sign which is used to prove the Cramér–Rao inequality, cf. (Witting, 1985, Satz 2.136) and (Gushchin, 1995b).

Proposition 1. *Let (P^α) be L^2-differentiable at α_0 with the score function V and T a random variable such that* $\mathrm{Var}^\alpha[T] \le C < \infty$ *in a neighborhood of α_0. Then the function $m(\alpha) = E^\alpha[T]$ is differentiable at α_0 and*

$$m'(\alpha_0) = E^{\alpha_0}[TV].$$

If ξ^n are random vectors given on probability spaces $(\Omega^n, \mathcal{F}^n, P^n)$ with values in $\mathbf{R}^d$, then $\mathcal{L}(\xi^n \mid P^n) \Rightarrow \mathcal{L}(\xi)$ will mean that one can construct a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ and a random vector ξ defined thereon with values in $\mathbf{R}^d$ such that the distributions of ξ^n under P^n weakly converge to the distribution of ξ under P^* as $n \to \infty$. The notation E^* will indicate the expectation with respect to P^*.

2. A STATISTICAL MODEL

We consider a statistical experiment where the observation is a Galton–Watson branching process with (nonrandom) immigration defined by

$$x_k = \sum_{i=1}^{x_{k-1}} y_{k,i} + 1, \quad k = 1, 2, \ldots, \quad x_0 = 1, \tag{2.1}$$

$y_{k,i}$ are i.i.d. random variables with an unknown distribution p on $\mathbf{N}_0$. Let $p_j = p(\{j\})$. It will be assumed that $\sum_{j=0}^{\infty} j^2 p_j < \infty$. For the mean and the variance of the offspring distribution p write

$$\vartheta(p) = \sum_{j=0}^{\infty} j p_j, \qquad \sigma^2(p) = \sum_{j=0}^{\infty} (j - \vartheta(p))^2 p_j.$$

The estimation problem is to determine $\vartheta(p)$ on observing a single realization of the process $\{x_1, \ldots, x_n\}$. We do not consider estimation problems when all offspring sizes are observed.

There is a huge literature on inference for Galton–Watson branching processes with random immigration, see for instance (Winnicki, 1988; Wei and Winnicki, 1990) and references therein. Our assumption that the immigration is nonrandom is done to simplify the problem. On the other hand, if the observation is a Galton–Watson branching process without immigration:

$$x_k = \sum_{i=1}^{x_{k-1}} y_{k,i}, \quad k = 1, 2, \ldots, \quad x_0 = 1, \tag{2.1$'$}$$

$y_{k,i}$ being the same as above, the consistent estimation of $\vartheta(p)$ based on a single long realization is possible on the nonextinction set, which has a positive probability only in the supercritical case, $\vartheta(p) > 1$, whereas we are interested to study the subcritical, $\vartheta(p) < 1$, the critical, $\vartheta(p) = 1$, and the supercritical cases in a unified manner.

In our model a natural estimator for $\vartheta(p)$ is

$$\widehat{\vartheta}_n = \frac{\sum_{k=1}^n (x_k - 1)}{\sum_{k=1}^n x_{k-1}}. \tag{2.2}$$

(Note that for the model (2.1′) in the supercritical case a similar estimator $\widehat{\vartheta}_n = \sum_{k=1}^n x_k \Big/ \sum_{k=1}^n x_{k-1}$ was introduced by Harris (Harris,1948) and has been studied in many papers.) It can be obtained by minimizing

$$\sum_{k=1}^n \left(\frac{x_k - E(x_k \mid x_1, \ldots, x_{k-1})}{\sqrt{\operatorname{Var}(x_k \mid x_1, \ldots, x_{k-1})}} \right)^2,$$

i.e., $\widehat{\vartheta}_n$ is the weighted conditional least squares estimator in the sense of Wei and Winnicki (Wei and Winnicki, 1990). In parametric case, $\widehat{\vartheta}_n$ is the maximum likelihood estimator for a power series family of offspring distribution, cf. (Heyde, 1975) or (Basawa and Prakasa Rao, 1980, p. 22). $\widehat{\vartheta}_n$ can also be obtained as a nonparametric maximum likelihood estimator, cf. (Feigin, 1977).

Let $\mathbb{P}$ be the class of all probability measures p on $\mathbf{N}_0$ such that $0 < \sigma^2(p) < \Sigma$, where Σ is an arbitrary (large enough) fixed positive number. To every offspring distribution p there corresponds the distribution P of the entire process $\{x_1, x_2, \ldots, x_n, \ldots\}$ defined by (2.1). Let $\mathcal{P}$ be the image of $\mathbb{P}$ under this mapping. This mapping is one-to-one, so we can permit ourselves to use the notation $\vartheta(P)$ and $\sigma^2(P)$.

Distributions from $\mathcal{P}$ are defined on the coordinate space $\mathbf{N}^\infty = \{(x_1, \ldots, x_n, \ldots), x_n \in \mathbf{N}\}$ with the product σ-field $\mathcal{F}$. Let $\mathcal{F}_n$ be the sub-σ-field of $\mathcal{F}$ generated by the first n coordinates $x_1, \ldots, x_n$. The sub-index n will indicate the restriction of a measure to the σ-field $\mathcal{F}_n$: $P_n = P\big|_{\mathcal{F}_n}$.

Let P and Q be measures in $\mathcal{P}$ with the corresponding offspring distributions p and q. We shall need few times an explicit expression for the Hellinger process h of order $\frac{1}{2}$ between P and Q with respect to the filtration $(\mathcal{F}_n)$. Using Proposition IV.1.63 in (Jacod and Shiryaev, 1987) and the Markov property of the process $\{x_n\}$, we obtain the following version:

$$h_n = \sum_{k=1}^n [1 - H(x_{k-1})], \quad n = 1, 2, \ldots, \tag{2.3}$$

where $H(j)$, $j = 1, 2, \ldots$, is the Hellinger integral of order $\frac{1}{2}$ between the j-fold convolutions of p and q, i.e.,

$$1 - H(j) = \rho^2(\underbrace{p * \cdots * p}_{j \text{ times}}, \underbrace{q * \cdots * q}_{j \text{ times}}),$$

where $*$ means the convolution.

In the rest of this section let us fix a distribution $P \in \mathcal{P}$, E is the expectation with respect to P. Put $\vartheta = \vartheta(P)$, $\sigma^2 = \sigma^2(P)$. Let us define the following quantities:

$$\varphi_n = \begin{cases} \sqrt{1-\vartheta}\, n^{-1/2} & \text{if } \theta < 1, \\ \sqrt{2}\, n^{-1} & \text{if } \theta = 1, \\ (\vartheta - 1)\theta^{-(n+1)/2} & \text{if } \theta > 1, \end{cases} \tag{2.4}$$

$$\varepsilon_k = \sum_{i=1}^{x_{k-1}} (y_{k,i} - \vartheta) = x_k - 1 - \vartheta x_{k-1}, \tag{2.5}$$

$$S_n = \varphi_n \sum_{k=1}^{n} \varepsilon_k, \tag{2.6}$$

$$G_n = \varphi_n^2 \sum_{k=1}^{n} x_{k-1}. \tag{2.7}$$

Let $M_n = \sum_{k=1}^{n} \varepsilon_k$. It is easy to see that $(M_n, \mathcal{F}_n)$ is a P-square integrable martingale with the quadratic characteristic $\langle M \rangle_n = \sigma^2 \sum_{k=1}^{n} x_{k-1}$. In particular, $E\varepsilon_k = 0$ and

$$EM_n^2 = E\langle M \rangle_n = \sigma^2 E \sum_{k=1}^{n} x_{k-1} = \sigma^2 B_n(\vartheta),$$

where

$$B_n(t) = \sum_{k=1}^{n} \sum_{i=0}^{k-1} t^i. \tag{2.8}$$

It follows from direct calculations that $\lim_{n\to\infty} \varphi_n^2 B_n(\vartheta) = 1$; moreover,

$$\lim_{n\to\infty} \varphi_n^2 B_n(\vartheta + \varphi_n s) = \begin{cases} \frac{e^{\sqrt{2}\, s} - 1 - \sqrt{2}\, s}{s^2} & \text{if } \theta = 1, s \neq 0, \\ 1 & \text{otherwise.} \end{cases}$$

From monotonicity of $B_n(t)$ we now obtain

Lemma 1. *Let t_n be a sequence of positive numbers such that*

$$\limsup_{n\to\infty} \varphi_n^{-1} |t_n - \vartheta| \le C_1 < \infty.$$

Then

$$\limsup_{n\to\infty} \varphi_n^2 B_n(t_n) \le C_2 < \infty,$$

where C_2 depends only on C_1.

Now we have $ES_n = 0$, $ES_n^2 = \sigma^2 EG_n = \sigma^2 \varphi_n^2 B_n(\vartheta) \longrightarrow \sigma^2$ as $n \to \infty$, in particular, the sequence (S_n, G_n) is tight under P. In fact, this sequence weakly converges as it follows from the next proposition.

Proposition 2. *We have*

$$\mathcal{L}(S_n, G_n \mid P) \Rightarrow \mathcal{L}(S, G), \tag{2.9}$$

where $G > 0$ (P^*-a.s.),

$$E^*G = 1, \quad E^*S = 0, \quad E^*S^2 = \sigma^2 \tag{2.10}$$

and

$$E^* \exp\big(aS - \frac{a^2\sigma^2}{2}G\big) = 1 \tag{2.11}$$

for any real a. *More precisely, if* $\vartheta < 1$, $G = 1$ (P^*-a.s.) *and* S *has the normal* $(0, \sigma^2)$ *distribution. If* $\vartheta > 1$, *the distribution of* (S, G) *coincides with the distribution of* $(G^{1/2}N, G)$, *where a random variable* N *is independent of* G *and has the normal* $(0, \sigma^2)$ *distribution. If* $\vartheta = 1$, *the distribution of* (S, G) *coincides with the distribution of* $(\sqrt{2}(Y_1 - 1), 2\int_0^1 Y_t\,dt)$, *where* $(Y_t, 0 \le t \le 1)$ *is a non-negative diffusion process satisfying the stochastic differential equation*

$$dY_t = dt + \sigma\sqrt{Y_t}\,dW_t, \quad Y_0 = 0,$$

where W *is a standard Wiener process.*

Proof. If $\vartheta < 1$, the assertion of the proposition follows from ergodic properties of $\{x_n\}$, see e.g. the proof of Theorem 2.1 in (Wei and Winnicki, 1990). If $\vartheta > 1$, it is known from (Seneta, 1970) that $\vartheta^{-n}x_n \longrightarrow X$ (P-a.s.) as $n \to \infty$, where X is a random variable with a distribution on $(0, \infty)$. By Toeplitz's lemma, $G_n \longrightarrow \vartheta X$ (P-a.s.). The convergence result now follows from, e.g., Theorem 3 in (Touati, 1993). If $\vartheta = 1$, the convergence (2.9) to the described limit follows immediately from Remark 2.4 in (Wei and Winnicki, 1989). The equality (2.11) is trivial if $\vartheta \ne 1$. Let $\vartheta = 1$. It is known that $E^*\exp(-uY_t) = (1 + ut/2)^{-2}$, $u > 0$, see e.g. (Ikeda and Watanabe, 1981, p. 222), so it is easy to deduce that $\sup_{t \le 1} E^* e^{\delta Y_t} < \infty$ if $\delta < 2$. According to (Liptser and Shiryaev, 1977, p. 220, Example 3), this condition guarantees the last equality in the following relation:

$$\begin{aligned} E^* \exp\big(aS - \frac{a^2\sigma^2}{2}G\big) &= E^* \exp\left(\sqrt{2}a(Y_1 - 1) - a^2\sigma^2 \int_0^1 Y_t\,dt\right) \\ &= E^* \exp\left(\sqrt{2}a\sigma \int_0^1 \sqrt{Y_t}\,dW_t - a^2\sigma^2 \int_0^1 Y_t\,dt\right) = 1. \end{aligned}$$

Relations (2.10) follow easily from the facts which have just been proved, except the first equality in (2.10) in the case $\vartheta > 1$. But it follows from straightforward but tedious calculations that, for every $P \in \mathcal{P}$, $EG_n^2 = O(1)$ as $n \to \infty$, so the sequence G_n is P-uniformly integrable and, hence, $E^*G = \lim_{n\to\infty} EG_n = 1$.

Remark. Let $(Y_t^a, 0 \le t \le 1)$ be a non-negative diffusion process satisfying the stochastic differential equation

$$dY_t^a = (\sqrt{2}\,a\sigma^2 Y_t^a + 1)\,dt + \sigma\sqrt{Y_t^a}\,dW_t, \quad 0 \le t \le 1, \quad Y_0^a = 0.$$

If Q^a is the distribution of Y_t^a in $\mathbf{C}[0,1]$, then (Q^0-a.s.)

$$\frac{dQ^a}{dQ^0}(Y) = \exp\left(\sqrt{2}a(Y_1 - 1) - a^2\sigma^2 \int_0^1 Y_t\,dt\right).$$

Since $\widehat{\vartheta}_n - \vartheta = \varphi_n G_n^{-1} S_n$ by (2.2) and (2.5)–(2.7), it follows from Proposition 2 that $\widehat{\vartheta}_n$ is a consistent estimator of $\vartheta(P)$ in $\mathcal{P}$. Moreover,

$$\mathcal{L}(\varphi_n^{-1}(\widehat{\vartheta}_n - \vartheta) \mid P) \Rightarrow \mathcal{L}(S/G)$$

and

$$\mathcal{L}(\varphi_n^{-1} G_n(\widehat{\vartheta}_n - \vartheta) \mid P) \Rightarrow \mathcal{L}(S).$$

It is also important to study the asymptotic behavior of $\widehat{\vartheta}_n$ under some alternatives. Let $P^n \in \mathcal{P}$, $n \ge 1$, be a sequence of distributions. Put $\vartheta_n = \vartheta(P^n)$, $\sigma_n^2 = \sigma^2(P^n)$.

Lemma 2. *Assume that*

$$\limsup_{n\to\infty} \varphi_n^{-1}|\vartheta_n - \vartheta| \le C_1 < \infty.$$

Then the sequence of distributions $\mathcal{L}(\varphi_n^{-1} G_n(\widehat{\vartheta}_n - \vartheta_n) \mid P^n)$ is tight and, if its subsequence weakly converges to a distribution L on $\mathbf{R}$, then

$$\int x L(dx) = 0 \qquad \text{and} \qquad \int x^2 L(dx) \le C_2 < \infty,$$

where C_2 depends only on C_1 and Σ.

Proof. Denote $\widetilde{S}_n = \varphi_n^{-1} G_n(\widehat{\vartheta}_n - \vartheta_n) = \varphi_n \sum_{k=1}^n (x_k - 1 - \vartheta_n x_{k-1})$. Similarly as above, $E^n \widetilde{S}_n = 0$, $E^n \widetilde{S}_n^2 = \varphi_n^2 \sigma_n^2 B_n(\vartheta_n)$ (E^n is the expectation with respect to P^n). By Lemma 1, $\limsup_{n\to\infty} E^n \widetilde{S}_n^2 \le C_2 < \infty$, where C_2 depends only on C_1 and Σ, and the result follows from the uniform integrability of $(\widetilde{S}_n, P^n)$ and from Fatou's lemma.

3. MAIN RESULTS

In this section we follow the set-up of the previous section.

Let $\widetilde{\vartheta}_n = \widetilde{\vartheta}_n(x_1, \dots, x_n)$ be a sequence of estimators of $\vartheta(P)$. Our aim is to give an asymptotic lower bound on the performance of $(\widetilde{\vartheta}_n)$, that is, given an arbitrary point $P \in \mathcal{P}$, we shall obtain a lower bound for a quantity characterizing the asymptotic behavior of the deviation $\widetilde{\vartheta}_n - \vartheta(P)$. This will be done for estimators $(\widetilde{\vartheta}_n)$ satisfying some "regularity" properties with respect to "small perturbations" of P. Our lower bound is based on an asymptotic version of the Cramér–Rao inequality, and in this sense our approach is close to that of Dzhaparidze and Spreij (Dzhaparidze and Spreij, 1993). Let us also mention that there is a connection between our approach and quasi-likelihood methods, see Section 5.

To give a meaning to the words "small perturbations", we introduce a notion of a one-dimensional local submodel passing through $P \in \mathcal{P}$. Roughly speaking, a one-dimensional local submodel passing through P is a family $(P^{\alpha,n})$, $P^{\alpha,n} \in \mathcal{P}$, of measures, where α is a real parameter and $n \in \mathbf{N}$, $P^{0,n} \equiv P$, satisfying the following property: the sequence of statistical experiments $(\mathbf{N}^\infty, \mathcal{F}_n, P_n^{\alpha,n})$ (recall that $P_n^{\alpha,n} = P^{\alpha,n}|_{\mathcal{F}_n}$) weakly converges (in the sense of Le Cam, see (Strasser, 1985; Le Cam, 1986; Le Cam and Yang, 1990) to an appropriate limit experiment $\mathcal{E}^*$. If we supposed that $\mathcal{E}^*$ is a Gaussian shift (as in the LAN theory), we would obtain a very poor class of one-dimensional local submodels in the case $\vartheta(P) \geq 1$. Moreover, since the next step will be to find a submodel which is the most difficult (in a sense) for estimating $\vartheta(P)$, so at this stage we are interested to embrace so many submodels as possible. In particular, we do not include the assumption of the contiguity $P_n^{\alpha,n} \lhd P_n$ for all α into the next definition since we do not need it, though this assumption seems to be very natural and simplifies the proofs.

In the sequel P is an arbitrary but fixed point in $\mathcal{P}$. The sequence φ_n is defined according to (2.4) with $\vartheta = \vartheta(P)$.

Let A be a subset of $\mathbf{R}$ such that $0 \in A$ and 0 is a cluster point of A. A family $(P^{\alpha,n}, \alpha \in A, n \in \mathbf{N})$ is called a *one-dimensional local submodel passing through* $P \in \mathcal{P}$ if

(i) $P^{\alpha,n} \in \mathcal{P} \quad \forall \alpha \in A \quad \forall n \in \mathbf{N}$;

(ii) $P^{0,n} = P \quad \forall n \in \mathbf{N}$;

(iii) the statistical experiments $(\mathbf{N}^\infty, \mathcal{F}_n, P_n^{\alpha,n}, \alpha \in A)$ weakly converge to a statistical experiment $\mathcal{E}^* = (\Omega^*, \mathcal{F}^*, Q^\alpha, \alpha \in A)$;

(iv) the family (Q^α) is L^2-differentiable at $\alpha = 0$ and the Fisher information $I^{(\alpha)}$ about α in (Q^α) at $\alpha = 0$ is strictly positive; the measures Q^α and Q^0 are not singular for every $\alpha \in A$.

The class of one-dimensional local submodels passing through P will be denoted by $\mathcal{P}(P)$.

An interesting (unsolved) question is formulated as follows. Let p be the offspring distribution corresponding to $P \in \mathcal{P}$. Let $v = v(j)$, $j \in \mathbf{N}_0$, be a function on $\mathbf{N}_0$ such that $\int v\, dp = 0$ and $0 < \int v^2\, dp < \infty$. Let (p^t), $t \in (-\varepsilon, \varepsilon)$ be a "path" in $\mathbb{P}$ such that $p^0 = p$ and (p^t) is L^2-differentiable at $t = 0$ with the score function v (such a path always exists by Lemma 4 in Appendix). Let $(P^{\alpha,n}, \alpha \in \mathbf{R})$ be a family in $\mathcal{P}$ such that $P^{\alpha,n}$ is the distribution of the branching process (2.1) under $p^{\varphi_n \alpha}$ if $|\varphi_n \alpha| < \varepsilon$. Is it true that $(P^{\alpha,n}) \in \mathcal{P}(P)$? By Lemma 5 in Appendix, the answer depends only on p and v. We know that the answer is positive if $\vartheta(P) < 1$, the limit experiment $\mathcal{E}^*$ being a Gaussian shift (the proof can be based on Theorem 1.24 in (Höpfner et al., 1990)), or if $v(j)$ is proportional to $j - \vartheta(P)$ (see Theorem 2 below). It is easy to check with the use of Hellinger processes that $(P_n^{\alpha,n}) \lhd\rhd (P_n)$ for all $\alpha \in \mathbf{R}$, so the hypothesis seems to be true. But even if the answer is positive in the general case, there are no reasons to restrict ourselves a priori to considering only such submodels: in the supercritical case it is possible to construct a family $P^\alpha \in \mathcal{P}$ such that $P^{\alpha,n} \equiv P^\alpha$ is a one-dimensional local submodel in our sense, cf. Lemma 3 in (Le Cam and Yang, 1988) and Section 3 in (Wei and Winnicki, 1990).

It follows from Lemma 6 in Appendix that if $(P^{\alpha,n}) \in \mathcal{P}(P)$, the sequence

$$\frac{\vartheta(P^{\alpha,n}) - \vartheta(P)}{\varphi_n} \tag{3.1}$$

is bounded for all $\alpha \in A$. The functional ϑ will be called *differentiable in the limit* along $(P^{\alpha,n}) \in \mathcal{P}(P)$ if the sequence (3.1) has a limit $k(\alpha)$ for every $\alpha \in A$ and there exists a limit

$$\lim_{\alpha \to 0} \frac{k(\alpha)}{\alpha} = \varkappa$$

(it follows from Theorem 1 that this limit is necessarily finite). The class of submodels $(P^{\alpha,n}) \in \mathcal{P}(P)$ such that ϑ is differentiable in the limit along $(P^{\alpha,n})$, will be denoted by $\mathcal{P}_d(P)$.

Let us assume for a moment that the limit experiment $\mathcal{E}^*$ for a submodel $(P^{\alpha,n})$ is a Gaussian shift, i.e., $A = \mathbf{R}$, $\mathcal{E}^* = (\mathbf{R}, \mathcal{B}, \mathcal{N}(I^{(\alpha)}\alpha, I^{(\alpha)}))$; moreover, assume that ϑ is differentiable in the limit along $(P^{\alpha,n})$ and $k(\alpha)$ is a linear function: $k(\alpha) = \varkappa\alpha$ and $\varkappa \neq 0$. In the limit the problem of estimating ϑ in the parametric submodel $(P^{\alpha,n})$ is no easier than estimating $k(\alpha)$ in $\mathcal{E}^*$. The latter is equivalent to the problem of estimating the mean in a Gaussian shift experiment again but with the Fisher information $I^{(\alpha)}/\varkappa^2$.

Similarly, if $\mathcal{E}^*$ is the limit experiment for $(P^{\alpha,n}) \in \mathcal{P}_d(P)$, the amount $I^{(\alpha)}/\varkappa^2$ is the Fisher information about $k(\alpha)$ in $\mathcal{E}^*$ at $\alpha = 0$.

Generalizing this concept, for $(P^{\alpha,n}) \in \mathcal{P}(P)$, define

$$I^{(\vartheta)} = \left[\sup_{\{n(\nu)\}} \limsup_{\alpha \to 0} \frac{1}{\alpha^2} \lim_\nu \left(\frac{\vartheta(P^{\alpha,n(\nu)}) - \vartheta(P)}{\varphi_{n(\nu)}} \right)^2 \right]^{-1} I^{(\alpha)},$$

where the supremum is taken over the set of all subnets $\{n(\nu)\}$ in $\mathbf{N}$ such that there exists a limit

$$\lim_\nu \frac{\vartheta(P^{\alpha,n(\nu)}) - \vartheta(P)}{\varphi_{n(\nu)}}$$

for all $\alpha \in A$ (this set is nonempty by Tychonov's theorem). By the definition, $0 \leq I^{(\vartheta)} \leq \infty$. We call $I^{(\vartheta)}$ the *asymptotic Fisher information* about ϑ in $(P^{\alpha,n})$.

Remark. By (1.1),

$$I^{(\alpha)} = 8 \lim_{\alpha \to 0} \frac{\rho^2(Q^\alpha, Q^0)}{\alpha^2} = 8 \lim_{\alpha \to 0} \lim_{n \to \infty} \frac{\rho^2(P_n^{\alpha,n}, P_n)}{\alpha^2}.$$

So the definition of the asymptotic Fisher information can be given without referring to the limit experiment and thus it can be extended to arbitrary families $(P^{\alpha,n})$ satisfying (i) and (ii) above. But it seems that the result of Theorem 1 below is not true for such arbitrary families: L^2-differentiability of the limit experiment is essentially used in the proof.

As the lower bound on the asymptotic performance of estimators, we intend to use the asymptotic information-type inequality given by Theorem 2 in (Gushchin, 1995b). For convenience, we now reformulate it conformably to our model.

Proposition 3. *Let $(P^{\alpha,n}) \in \mathcal{P}_d(P)$ and $\widetilde{\vartheta}_n = \widetilde{\vartheta}_n(x_1, \ldots, x_n)$ be a sequence of estimators. Assume that there exist a subnet $\{n(\nu)\}$ in $\mathbf{N}$ and a net of measurable mappings $H_\nu \colon (\mathbf{N}^\infty, \mathcal{F}_{n(\nu)}) \to [0, +\infty)$ such that*

$$\mathcal{L}(\varphi_{n(\nu)}^{-1} H_\nu(\widetilde{\vartheta}_{n(\nu)} - \vartheta(P^{\alpha,n(\nu)})) \mid P^{\alpha,n(\nu)}) \Rightarrow L_\alpha, \quad \alpha \in A,$$

where L_α is a probability measure on $\mathbf{R}$,

$$\int x L_\alpha(dx) = 0 \quad \text{and} \quad \int x^2 L_\alpha(dx) \leq C < \infty$$

for all $\alpha \in A$. Moreover, assume that $\mathcal{L}(H_\nu \mid P) \Rightarrow H$, where H is a variable with values in $[0, +\infty]$ and $P^*(H = 0) < 1$. Then

$$\int x^2 L_0(dx) \geq (E^* H)^2 / I^{(\vartheta)} \tag{3.2}$$

(here $\infty/\infty = 0$ and $\infty/a = \infty$ if $0 \leq a < \infty$). In particular, $E^* H < \infty$ if $I^{(\vartheta)} < \infty$.

Remark. After replacing $I^{(\vartheta)}$ by

$$\left[\limsup_{\alpha \to 0} \frac{1}{\alpha^2} \lim_\nu \left(\frac{\vartheta(P^{\alpha,n(\nu)}) - \vartheta(P)}{\varphi_{n(\nu)}} \right)^2 \right]^{-1} I^{(\alpha)}$$

in (3.2), the statement of Proposition 3 remains true not only for $(P^{\alpha,n}) \in \mathcal{P}_d(P)$ but also for $(P^{\alpha,n}) \in \mathcal{P}(P)$ such that there exists a limit

$$\lim_\nu \frac{\vartheta(P^{\alpha,n(\nu)}) - \vartheta(P)}{\varphi_{n(\nu)}}$$

for all $\alpha \in A$.

Our next goal is to find "least favorable submodels". According to Proposition 3, the asymptotic lower bound is inversely proportional to $I^{(\vartheta)}$ (if $I^{(\vartheta)} < \infty$, we can normalize H_ν so that $E^*H = 1$). Hence, we must find submodels with the minimal asymptotic Fisher information. This will be done in Theorems 1 and 2.

Theorem 1. *For any one-parameter local submodel passing through P, its asymptotic Fisher information $I^{(\vartheta)}$ about ϑ satisfies the inequality*

$$I^{(\vartheta)} \geq 1/\sigma^2(P).$$

The class of submodels $(P^{\alpha,n}) \in \mathcal{P}_d(P)$ such that $I^{(\vartheta)} = 1/\sigma^2(P)$ will be denoted by $\mathcal{P}_m(P)$. This class is nonempty for any $P \in \mathcal{P}$ as it follows from Theorem 2.

Lemma 3. *Let $P \in \mathcal{P}$ and p be the offspring distribution corresponding to P. There exists a mapping $t \rightsquigarrow p^t$ from an interval $(-\varepsilon, \varepsilon)$ to $\mathbb{P}$ such that $p^0 = p$ and (p^t) is L^2-differentiable at $t = 0$ with the score function $v(j) = j - \vartheta$.*

A particular case of the next theorem was considered in Example 3 in (Gushchin, 1995a). The variables S_n and G_n are defined in (2.6) and (2.7) with $\vartheta = \vartheta(P)$; $\sigma^2 = \sigma^2(P)$.

Theorem 2. *Let a mapping $t \rightsquigarrow p^t$ from $(-\varepsilon, \varepsilon)$ to $\mathbb{P}$ satisfy the statement of Lemma 3. Let P^t be the distribution of the branching process (2.1) under p^t. Define $P^{\alpha,n}$, $\alpha \in \mathbf{R}$, as follows: $P^{\alpha,n} = P^{\varphi_n \alpha}$ if $|\varphi_n \alpha| < \varepsilon$ and $P^{\alpha,n} = P$ otherwise.*
1) *For any bounded sequence $\{\alpha_n\}$, the sequences $P_n^{\varphi_n \alpha_n}$ and P_n are mutually contiguous and*

$$\log \frac{dP_n^{\varphi_n \alpha_n}}{dP_n} - (\alpha_n S_n - \frac{\alpha_n^2}{2}\sigma^2 G_n) \xrightarrow{P} 0 \quad \textit{as} \quad n \to \infty. \tag{3.3}$$

2) $(P^{\alpha,n}) \in \mathcal{P}_m(P)$.

Remark. The first statement of the theorem means that the sequence $(\mathbf{N}^\infty, \mathcal{F}_n, P_n^t, t \in (-\varepsilon, \varepsilon))$ of statistical experiments is *locally asymptotically quadratic* at $t = 0$ (see (Le Cam and Yang, 1990) for the definition of locally asymptotically quadratic families); moreover, it follows from Proposition 2 that this sequence is *locally asymptotically mixed normal* if $\vartheta(P) \neq 1$ and *locally asymptotically normal* if $\vartheta(P) < 1$.

Let us now proceed to study properties of estimators of $\vartheta(P)$. As a simple corollary of Proposition 3, we prove (in fact, under redundant assumptions) that a "good" estimator cannot converge to $\vartheta(P)$ faster than $\widehat{\vartheta}_n$.

Theorem 3. *Let* $(P^{\alpha,n}) \in \mathcal{P}_d(P)$ *with* $I^{(\vartheta)} < \infty$ *and* $\widetilde{\vartheta}_n = \widetilde{\vartheta}_n(x_1, \dots, x_n)$ *be a sequence of estimators. Assume that the sequence of distributions* $\mathcal{L}(\psi_n^{-1}(\widetilde{\vartheta}_n - \vartheta(P^{\alpha,n})) \mid P^{\alpha,n})$ *is tight for every* $\alpha \in A$ *for some sequence* $\psi_n > 0$. *Then*

$$\limsup_{n\to\infty} \psi_n^{-1}\varphi_n < \infty.$$

The first part of the next theorem is a slight reformulation of Proposition 3 in the case $(P^{\alpha,n}) \in \mathcal{P}_m(P)$. As it was mentioned in the introduction, the lower bound (3.2) is not sharp in general but it is sharp for locally asymptotically quadratic submodels, cf. Theorems 3 and 4 in (Gushchin, 1995a). Though not every submodel from $\mathcal{P}_m(P)$ can be obtained from a locally asymptotically quadratic sequence as the submodel considered in Theorem 2, the structure of all submodels from $\mathcal{P}_m(P)$, roughly, is the same as $\alpha \to 0$. This fact explains the second part of the next theorem, which means that the estimator $\widehat{\vartheta}_n$ is asymptotically efficient at all $P \in \mathcal{P}$ in our sense.

Theorem 4. *Let* $(P^{\alpha,n}) \in \mathcal{P}_m(P)$ *and* $\widetilde{\vartheta}_n = \widetilde{\vartheta}_n(x_1, \dots, x_n)$ *be a sequence of estimators.*
1) *Assume that there exists a sequence of measurable mappings* $H_n\colon (\mathbf{N}^\infty, \mathcal{F}_n) \to [0, +\infty)$ *such that the sequence of distributions*

$$\mathcal{L}(\varphi_n^{-1} H_n(\widetilde{\vartheta}_n - \vartheta(P^{\alpha,n})) \mid P^{\alpha,n}) \tag{3.4}$$

is tight in $\mathbf{R}$ *for every* $\alpha \in B$, *where* B *is a subset of* A *such that* $0 \in B$ *and* 0 *is a cluster point of* B. *Denote* $\mathcal{L}_\alpha$ *the set of cluster points of the sequence* (3.4). *Assume that, for some* $C < \infty$, $\int x L_\alpha(dx) = 0$ *and* $\int x^2 L_\alpha(dx) \le C$ *for all* $\alpha \in B$ *and* $L_\alpha \in \mathcal{L}_\alpha$, *and* $\mathcal{L}(H_n \mid P) \Rightarrow H$, $E^* H = 1$. *Then*

$$\int x^2 L_0(dx) \ge \sigma^2(P) \tag{3.5}$$

for every $L_0 \in \mathcal{L}_0$.
2) *The sequence* $(\widehat{\vartheta}_n)$ *satisfies the assumptions of the first part of the theorem with* $H_n = G_n$, $n \in \mathbf{N}$, *and attains the lower bound in* (3.5).
3) *If* $(\widetilde{\vartheta}_n)$ *satisfies the assumptions of the first part of the theorem with* $H_n = G_n$, $n \in \mathbf{N}$, *and we have equality in* (3.5) *for every* $L_0 \in \mathcal{L}_0$, *then* $\varphi_n^{-1}(\widetilde{\vartheta}_n - \widehat{\vartheta}_n) \xrightarrow{P} 0$ *as* $n \to \infty$.

Remarks. 1. It follows easily from the second part of the theorem that the converse to the third part of the theorem is true if $P_n^{\alpha,n} \lhd P_n$ for all $\alpha \in A$.

2. If we have equality in (3.5) for a sequence $(\widetilde{\vartheta}_n)$ of estimators satisfying the assumptions of the first part of the theorem with an arbitrary sequence $\{H_n\}$, then we cannot assert that $\varphi_n^{-1}(\widetilde{\vartheta}_n - \widehat{\vartheta}_n) \xrightarrow{P} 0$ as $n \to \infty$.

4. PROOFS

Proof of Theorem 1. Since the limit experiment $\mathcal{E}^* = (\Omega^*, \mathcal{F}^*, Q^\alpha, \alpha \in A)$ is L^2-differentiable, $\lim_{\alpha\to 0} \|Q^\alpha - Q^0\| = 0$ and we can assume without loss of generality that $\|Q^\alpha - Q^0\| < 2(1-\varepsilon)$ for some $\varepsilon > 0$ for all $\alpha \in A$.

It follows from Lemmas 2 and 6 that the sequence of distributions

$$\mathcal{L}(\varphi_n^{-1} G_n(\widehat{\vartheta}_n - \vartheta(P^{\alpha,n})) \mid P^{\alpha,n}) \tag{4.1}$$

is tight in $\mathbf{R}$ for every $\alpha \in A$; moreover, if $\mathcal{L}_\alpha$ denotes the set of cluster points of the sequence (4.1),

$$\int x L_\alpha(dx) = 0 \qquad \text{and} \qquad \int x^2 L_\alpha(dx) \le C < \infty \tag{4.2}$$

for every $\alpha \in A$ and $L_\alpha \in \mathcal{L}_\alpha$, where C does not depend on α.

Let $\{n(\nu)\}$ be a subnet such that

$$\lim_\nu \frac{\vartheta(P^{\alpha,n(\nu)}) - \vartheta(P)}{\varphi_{n(\nu)}}$$

exists (and finite by Lemma 6) for all $\alpha \in A$. We can find a further subnet, abusing notation denoted $\{n(\nu)\}$ again, such that the net of distributions

$$\mathcal{L}(\varphi_{n(\nu)}^{-1} G_{n(\nu)}(\widehat{\vartheta}_{n(\nu)} - \vartheta(P^{\alpha,n(\nu)})) \mid P^{\alpha,n(\nu)})$$

weakly converges to a distribution $L_\alpha \in \mathcal{L}_\alpha$ for all $\alpha \in A$. Note that $L_0 = \mathcal{L}(S)$ by Proposition 2. In view of (4.2), we can apply Theorem 2 in (Gushchin, 1995b), see Remark after Proposition 3, which gives

$$\frac{E^* S^2}{(E^* G)^2} \le \left[\limsup_{\alpha\to 0} \frac{1}{\alpha^2} \lim_\nu \left(\frac{\vartheta(P^{\alpha,n(\nu)}) - \vartheta(P)}{\varphi_{n(\nu)}} \right)^2 \right] \Big/ I^{(\alpha)}. \tag{4.3}$$

The left-hand side of (4.3) is equal to $\sigma^2(P)$ by Proposition 2, which completes the proof.

Proof of Lemma 3. Let $a(t) = p(|tv| \le 1)$ and $f(t) = \int \exp(tv)\mathbf{I}(|tv| \le 1)\, dp$, $t \in \mathbf{R}$. Then $a(t) > 0$ if t is small enough, so we can define a probability measure p_t on $\mathbf{N}_0$ by

$$dp^t = \frac{\exp(tv)\mathbf{I}(|tv| \le 1)}{f(t)}\, dp. \tag{4.4}$$

The function $\sigma^2(p^t)$ is continuous by Lemma 4 in Appendix, hence $p^t \in \mathbb{P}$ if $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. The L^2-differentiability of (p^t) at $t = 0$ follows also from Lemma 4.

Proof of Theorem 2. 1) Let p^t be defined by (4.4). It is enough to prove the statement only for this family. Indeed, let $\{\widetilde{p}^t\}$, $\widetilde{p}^0 = p$, be another family which is L^2-differentiable at $t = 0$ with the same score function. Put $\mathfrak{z}_t = dp^t/dp$ and $\widetilde{\mathfrak{z}}_t = d\widetilde{p}^t/dp$. Then

$$\rho^2(p^t, \widetilde{p}^t) \le \frac{1}{2}\int (\sqrt{\widetilde{\mathfrak{z}}_t} - \sqrt{\mathfrak{z}_t})^2\, dp + \frac{1}{2}p^t(\mathfrak{z}_t + \widetilde{\mathfrak{z}}_t = \infty) + \frac{1}{2}\widetilde{p}^t(\mathfrak{z}_t + \widetilde{\mathfrak{z}}_t = \infty) = o(t^2)$$

owing to the definition of L^2-differentiability and since $p^t(\mathfrak{z}_t < \infty, \widetilde{\mathfrak{z}}_t = \infty) = \widetilde{p}^t(\mathfrak{z}_t = \infty, \widetilde{\mathfrak{z}}_t < \infty) = 0$. Moreover, $\lim_{t\to 0} t^{-1}(\vartheta(p^t) - \vartheta(p)) = \sigma^2(p)$ by Proposition 1. Therefore, $\|\widetilde{P}_n^{\varphi_n\alpha_n} - P_n^{\varphi_n\alpha_n}\| \longrightarrow 0$ (where $\widetilde{P}^t$ is the distribution of the branching process (2.1) under $\widetilde{p}^t$) as $n \to \infty$ for any bounded sequence $\{\alpha_n\}$ by Lemma 5 in Appendix.

Let us also define a probability measure q^t on $\mathbf{N}_0$ by

$$dq^t = \frac{\mathbf{I}(|tv| \le 1)}{a(t)}\, dp.$$

Denote by Q^t the distribution of the branching process (2.1) under q^t. By Lemma 4, $\rho^2(p, q^{\varphi_n\alpha_n}) = o(\varphi_n^2)$ for any bounded sequence $\{\alpha_n\}$, so

$$\|P_n - Q_n^{\varphi_n\alpha_n}\| = o(1) \tag{4.5}$$

by Lemma 5.

Since $p^t \sim q^t$ and

$$\frac{dp^t}{dq^t} = \frac{a(t)}{f(t)}\exp(tv) \qquad q^t\text{-a.s.},$$

it is easy to deduce, using the Markov property of $\{x_1, \ldots, x_n, \ldots\}$, that, Q^t-a.s.,

$$\begin{aligned}\log\frac{dP_n^t}{dQ_n^t} &= t\sum_{k=1}^{n}(x_k - 1 - \vartheta x_{k-1}) + \Big(\log\frac{a(t)}{f(t)}\Big)\sum_{k=1}^{n} x_{k-1}\\ &= \varphi_n^{-1} t S_n + \varphi_n^{-2}\Big(\log\frac{a(t)}{f(t)}\Big)G_n.\end{aligned}$$

By Lemma 4,

$$\log\frac{a(t)}{f(t)} = -\frac{\sigma^2 t^2}{2}(1 + o(1)) \qquad \text{as} \quad t \to 0,$$

hence

$$\log\frac{dP_n^{\varphi_n\alpha_n}}{dQ_n^{\varphi_n\alpha_n}} = \alpha_n S_n - \frac{\alpha_n^2\sigma^2}{2}(1 + o(1))G_n$$

for sufficiently large n for any bounded sequence $\{\alpha_n\}$. Taking into account (4.5) and Proposition 2, we obtain (3.3).

To prove the mutual contiguity of $\{P_n^{\varphi_n \alpha_n}\}$ and $\{P_n\}$ for any bounded sequence $\{\alpha_n\}$ it is enough to consider the case $\alpha_n \to \alpha$, $n \to \infty$. Then, by Proposition 2 and (3.3),

$$\mathcal{L}(dP_n^{\varphi_n \alpha_n}/dP_n \mid P) \Rightarrow \exp(\alpha S - \frac{\alpha^2\sigma^2}{2}G) > 0 \qquad (P^*\text{-a.s.}),$$

hence $P_n \lhd P_n^{\varphi_n \alpha_n}$. In view of (2.11), $P_n^{\varphi_n \alpha_n} \lhd P_n$ by the Le Cam first lemma.

2) By the first part of the theorem, the sequence $(\mathbf{N}^\infty, \mathcal{F}_n, P_n^{\alpha,n}, \alpha \in \mathbf{R})$ weakly converges to $(\Omega^*, \mathcal{F}^*, Q^\alpha, \alpha \in \mathbf{R})$, where $dQ^\alpha = Z_\alpha\, dP^*$, $Z_\alpha = \exp(\alpha S - \alpha^2\sigma^2 G/2)$, $\alpha \in \mathbf{R}$. The function Z_α is differentiable in α with the derivative $(S - \alpha\sigma^2 G)Z_\alpha$. Using the explicit representation for the distribution of (S, G) given by Proposition 2 and Remark after it, it is easy to obtain that

$$\int (S - \alpha\sigma^2 G)^2\, dQ^\alpha = E^* S^2 = \sigma^2 \qquad \text{if} \qquad \vartheta \neq 1$$

and

$$\int (S - \alpha\sigma^2 G)^2\, dQ^\alpha = \begin{cases} \frac{e^{\sqrt{2}\,\alpha\sigma^2} - 1 - \sqrt{2}\,\alpha\sigma^2}{\alpha^2\sigma^2} & \text{if } \vartheta = 1, \alpha \neq 0, \\ \sigma^2 & \text{if } \vartheta = 1, \alpha = 0. \end{cases}$$

In the both cases the function $\int (S - \alpha\sigma^2 G)^2\, dQ^\alpha$ is continuous in α. Therefore, the family (Q^α) is L^2-differentiable at every α, see (Strasser, 1985, p. 393, Discussion 77.5) and $I^{(\alpha)} = \sigma^2(P)$. Furthermore,

$$\lim_{n\to\infty} \frac{\vartheta(P^{\alpha,n}) - \vartheta(P)}{\varphi_n} = \alpha\sigma^2(P)$$

by Proposition 1, so ϑ is differentiable in the limit along $(P^{\alpha,n})$. Finally, we deduce $I^{(\vartheta)} = 1/\sigma^2(P)$.

Proof of Theorem 3. Assume that $\lim_\nu \psi_{n(\nu)}^{-1}\varphi_{n(\nu)} = \infty$ for some subsequence $\{n(\nu)\}$. Then

$$\varphi_{n(\nu)}^{-1}(\widetilde{\vartheta}_{n(\nu)} - \vartheta(P^{\alpha,n(\nu)})) \xrightarrow{P^{\alpha,n(\nu)}} 0$$

for all $\alpha \in A$. Applying Proposition 3 (with $H_\nu = 1$), we arrive to a contradiction.

Proof of Theorem 4. 1) Let us fix $L_0 \in \mathcal{L}_0$. There is a subsequence $\{n(m)\}$ such that

$$\mathcal{L}(\varphi_{n(m)}^{-1} H_{n(m)}(\widetilde{\vartheta}_{n(m)} - \vartheta(P)) \mid P) \Rightarrow L_0.$$

By hypotheses of the theorem, there exists a subnet $\{n(\nu)\}$ of $\{n(m)\}$ such that

$$\mathcal{L}(\varphi_{n(\nu)}^{-1} H_{n(\nu)}(\widetilde{\vartheta}_{n(\nu)} - \vartheta(P^{\alpha,n(\nu)})) \mid P^{\alpha,n(\nu)})$$

weakly converges to a distribution $L_\alpha \in \mathcal{L}_\alpha$ for all $\alpha \in B \setminus \{0\}$. Applying Proposition 3, we obtain

$$\int x^2 L_0(dx) \geq (E^* H)^2 / I^{(\vartheta)} = \sigma^2(P).$$

2) It was implicitly shown in the proof of Theorem 1 that $\widehat{\vartheta}_n$ satisfies the assumptions of the first part of the theorem with $H_n = G_n$, $n \in \mathbf{N}$, and

$$B = \{\alpha \in A : \limsup_{n\to\infty} \|P_n^{\alpha,n} - P_n\| < 2(1-\varepsilon)\}, \quad 0 < \varepsilon < 1.$$

Evidently, $\mathcal{L}_0 = \{\mathcal{L}(S \mid P^*)\}$ and we have equality in (3.5) by Proposition 2.

3) Assume that the joint distributions

$$\mathcal{L}(\varphi_n^{-1} G_n(\widetilde{\vartheta}_n - \vartheta(P)), \varphi_n^{-1} G_n(\widehat{\vartheta}_n - \vartheta(P)) \mid P)$$

weakly converge along a subsequence $\{n(m)\}$ to a distribution $\mathcal{L}(\widetilde{S}, S \mid P^*)$. Since the sequence of estimators $\beta\widetilde{\vartheta}_n + (1-\beta)\widehat{\vartheta}_n$ satisfies the assumptions of the first part of the theorem with $H_n = G_n$ for every $\beta \in [0,1]$, we have

$$E^* \widetilde{S}^2 = \sigma^2, \quad E^* S^2 = \sigma^2, \quad E^*[\beta\widetilde{S} + (1-\beta)S]^2 \geq \sigma^2.$$

This is possible only if $\widetilde{S} = S$ (P^*-a.s.). This means that $\varphi_n^{-1} G_n(\widetilde{\vartheta}_n - \widehat{\vartheta}_n) \xrightarrow{P} 0$, and the result follows since the limit distribution of $\mathcal{L}(G_n \mid P)$ is concentrated on $(0, \infty)$.

5. CONCLUDING REMARKS

We follow the set-up of Section 2.

Let $\vartheta > 0$ be a real number. Put $\mathcal{P}_\vartheta = \{P \in \mathcal{P} : \vartheta(P) = \vartheta\}$, $M_n(\vartheta) = \sum_{k=1}^n (x_k - 1 - \vartheta x_{k-1})$. We implicitly exploit in the proofs of Theorems 1 and 4 the fact that

$$(M_n(\vartheta), \mathcal{F}_n) \quad \text{is a } P\text{-square integrable martingale for every } P \in \mathcal{P}_\vartheta. \qquad (5.1)$$

To illustrate this we shall deduce some simple consequences of (5.1) in this section.

First, we shall obtain counterparts of Theorems 1 and 2 for finite samples. Let (p^t), $t \in (-\varepsilon, \varepsilon)$, be a path in $\mathbb{P}$ which is L^2-differentiable at $t = 0$ with a score function v. Denote by P^t the distribution of the branching process (2.1) under p^t; $p = p^0$, $P = P^0$, $\vartheta = \vartheta(p)$, $\sigma^2 = \sigma^2(p)$, $\varepsilon_n = x_n - 1 - \vartheta x_{n-1}$. It is easy to check that, for every $n \in \mathbf{N}$, the family (P_t^n) is L^2-differentiable at $t = 0$ with the score function

$$V_n = \sum_{k=1}^{n} \sum_{i=1}^{x_{k-1}} E[v(y_{k,i}) \mid \mathcal{F}_k]. \qquad (5.2)$$

The process $(V_n, \mathcal{F}_n)$ is a P-square integrable martingale. Put $\Delta V_n = V_n - V_{n-1}$, $V_0 = 0$.

By Proposition 1, $\vartheta(p^t)$ is differentiable at $t = 0$ and

$$\dot{\vartheta} = \lim_{t \to 0} \frac{\vartheta(P^t) - \vartheta}{t} = \sum_{j=0}^{\infty} j v(j) p_j. \tag{5.3}$$

Hence,

$$\begin{aligned} E(\Delta V_n \varepsilon_n \mid \mathcal{F}_{n-1}) &= E\left[\sum_{i=1}^{x_{n-1}} v(y_{n,i}) \sum_{i=1}^{x_{n-1}} (y_{n,i} - \vartheta(p)) \middle| \mathcal{F}_{n-1} \right] \\ &= \sum_{i=1}^{x_{n-1}} \sum_{j=1}^{x_{n-1}} E[v(y_{n,i})(y_{n,j} - \vartheta(p)) \mid \mathcal{F}_{n-1}] = \dot{\vartheta} x_{n-1}. \end{aligned} \tag{5.4}$$

(In fact, the relation (5.4) can be directly deduced from (5.1). For a general result of this kind see (Gushchin, 1994, Theorem 8.1).) Put

$$V_n = \frac{\dot{\vartheta}}{\sigma^2} \sum_{k=1}^{n} \varepsilon_k + V_n^{\perp}. \tag{5.5}$$

Then $E(\Delta V_n^{\perp} \varepsilon_n \mid \mathcal{F}_{n-1}) = 0$, so the terms in the right-hand side of (5.5) are orthogonal in L^2. In particular, we have the following lower bound for the Fisher information $I_n^{(t)}$ about t in (P_n^t) at $t = 0$:

$$I_n^{(t)} \geq \sigma^{-2} \dot{\vartheta}^2 B_n(\vartheta).$$

Let us suppose that $\dot{\vartheta} \neq 0$ and $\vartheta(p^t)$ is a monotone function in a neighborhood of 0. Then we can reparametrize $\{P_n^t\}$ by the offspring mean, and the Fisher information $I_n^{(\vartheta)}$ about ϑ at $\vartheta(p)$ satisfies the inequality

$$I_n^{(\vartheta)} \geq \sigma^{-2} B_n(\vartheta). \tag{5.6}$$

It follows from (5.5) that if we have equality in (5.6) for some n, then we have equality in (5.6) for $n = 1$, i.e., $\int v^2 \, dp = \sigma^{-2} \dot{\vartheta}^2$. Comparing with (5.3), we obtain

$$v(j) = \sigma^{-2} \dot{\vartheta}(j - \vartheta), \quad j \in \mathbf{N}_0, \quad p\text{-a.s.} \tag{5.7}$$

Conversely, if (5.7) holds then (5.2) implies

$$V_n = \frac{\dot{\vartheta}}{\sigma^2} \sum_{k=1}^{n} \varepsilon_k,$$

so we have equality in (5.6) for all n.

The existence of a family (p^t) passing through an arbitrary $p \in \mathbb{P}$ with the score function (5.7) is proved in Lemma 4 in Appendix. If $g(t) = \sum_j p_j(1+t)^j < \infty$ (in particular, if $-1 < t \le 0$), we can define p^t by

$$\frac{dp^t}{dp}(j) = \frac{\dot{\vartheta}}{\sigma^2}\frac{p_j(1+t)^j}{g(t)}, \quad j \in \mathbf{N}_0.$$

This means that the power series family of offspring distributions has the minimal Fisher information about ϑ among other parametric families passing through a fixed point p. This also explains why we prove that the estimator $\widehat{\vartheta}_n$ which is the maximum likelihood estimator for power series families of offspring distributions, is asymptotically efficient at every point p.

Let us now draw a parallel between Theorem 4 and quasi-likelihood estimation. We refer to (Godambe and Heyde, 1987) for a discussion of the general quasi-likelihood framework. Quasi-likelihood estimation for branching processes with immigration has been considered in (Heyde and Lin, 1992).

Taking (5.1) into account, we confine attention to martingale estimating functions belonging to the class

$$\mathcal{H} = \{H\colon H_n(\vartheta) = \sum_{k=1}^{n} c_k(\vartheta)(x_k - 1 - \vartheta x_{k-1}) \quad \text{for } \mathcal{F}_{k-1}\text{-measurable } c_k(\vartheta)\}.$$

It is assumed that $c_k(\vartheta)$, $k \in \mathbf{N}$, are differentiable with respect to ϑ. The estimator $\widehat{\vartheta}_{H,n}$ corresponding to $H \in \mathcal{H}$ is defined as the solution of the equation $H_n(\vartheta) = 0$. Usually it is implicitly assumed that comparisons are to be made between estimators which, with appropriate norming, are asymptotically normally distributed:

$$\langle H(\vartheta)\rangle_n^{-1/2}\overline{H}_n(\vartheta)(\widehat{\vartheta}_{H,n} - \vartheta) \Rightarrow \mathcal{N}(0,1) \tag{5.8}$$

under $P \in \mathcal{P}_\vartheta$ as $n \to \infty$, where

$$\overline{H}_n(\vartheta) = -\sum_{k=1}^{n} c_k(\vartheta)x_{k-1}$$

is the P-compensator of $\partial H_n(\vartheta)/\partial\vartheta$ and

$$\langle H(\vartheta)\rangle_n = \sigma^2(P)\sum_{k=1}^{n} c_k^2(\vartheta)x_{k-1}$$

is the P-quadratic characteristic of $H_n(\vartheta)$. The relation (5.8) leads to maximizing the expression

$$\frac{\overline{H}_n^2(\vartheta)}{\langle H(\vartheta)\rangle_n} = \sigma^{-2}(P)\Big(\sum_{k=1}^{n} c_k(\vartheta)x_{k-1}\Big)^2\Big(\sum_{k=1}^{n} c_k^2(\vartheta)x_{k-1}\Big)^{-1},$$

which is maximized if $c_k(\vartheta) = \text{const}$. Therefore, the optimal estimating function in the sense of the asymptotic optimality criterion of Godambe and Heyde (Godambe and Heyde, 1987), i.e., the quasi-score estimating function, is $H_n^*(\vartheta) = \text{const} \cdot \sum_{k=1}^{n}(x_k - 1 - \vartheta x_{k-1})$ (as it usually happens, the quasi-score estimating function is the true score function for some parametric submodel, a power series family in our case), and the quasi-likelihood estimator $\widehat{\vartheta}_{H^*,n}$ coincides with $\widehat{\vartheta}_n$. H^* also satisfies the fixed sample criterion in (Godambe and Heyde, 1987).

To compare the quasi-likelihood approach and our results let us first note that, according to Proposition 2, $\widehat{\vartheta}_n$ does not satisfy (5.8) if $\vartheta(P) = 1$, so the quasi-likelihood theory does not justify optimality properties of $\widehat{\vartheta}_n$ in the critical case. Next, Taylor's expansion gives

$$\frac{\partial H_n}{\partial \vartheta}(\vartheta'_n)(\widehat{\vartheta}_{H,n} - \vartheta) = -H_n(\vartheta),$$

where ϑ'_n lies between ϑ and $\widehat{\vartheta}_{H,n}$. It follows from the martingale property of $H(\vartheta)$ that under mild regularity conditions (which are of less restrictive type than those leading to (5.8)), with appropriate norming, the estimator $\widehat{\vartheta}_{H,n}$ is asymptotically unbiased, which corresponds to the assumptions of Theorem 4. Thus, the class of competing estimators in the quasi-likelihood approach is smaller than that in our approach. On the other hand, both approaches compare the quality of estimators in a similar manner, from the point of view of minimum dispersion distance.

6. APPENDIX

Lemma 4. *Let ξ be a variable on a probability space $(\Omega, \mathcal{F}, P)$ with $E\xi^2 < \infty$. Put $a(t) = P(|t\xi| \leq 1)$ and $f(t) = E[\exp(t\xi)\mathbf{I}(|t\xi| \leq 1)]$, where $\mathbf{I}(\cdot)$ is the indicator function. If $a(t) > 0$, define probability measures Q_t and P_t as follows:*

$$dQ_t = \frac{\mathbf{I}(|t\xi| \leq 1)}{a(t)}\, dP$$

and

$$dP_t = \frac{\exp(t\xi)\mathbf{I}(|t\xi| \leq 1)}{f(t)}\, dP.$$

Then

$$a(t) = 1 + o(t^2), \qquad f(t) = 1 + tE\xi + t^2 E\xi^2/2 + o(t^2) \tag{6.1}$$

and

$$\rho^2(P, Q_t) = o(t^2) \tag{6.2}$$

as $t \to 0$ and the family (P_t) is L^2-differentiable at $t = 0$ with the score function $\xi - E\xi$. The functions $\int \xi\, dP_t$ and $\xi^2\, dP_t$ are continuous at $t = 0$.

Proof. Since $E\xi^2 < \infty$, we have

$$1 - a(t) = P(|t\xi| > 1) \leq t^2 E[\xi^2 \mathbf{I}(|t\xi| > 1)] = o(t^2) \tag{6.3}$$

as $t \to 0$. It follows immediately from the definitions that $\rho^2(P, Q_t) = 1 - \sqrt{a(t)}$, hence (6.2) follows from (6.3).

Write

$$e^x = 1 + x + x^2/2 + x^2 R(x), \quad R(0) = 0,$$

then $R(x)$ is a continuous function. One has

$$t^{-2}[f(t) - 1 - tE\xi - t^2 E\xi^2/2] = E[\xi^2 R(t\xi)\mathbf{I}(|t\xi| \le 1)] - E[t^{-2}(1 + t\xi + t^2\xi^2/2)\mathbf{I}(|t\xi| > 1)]. \quad (6.4)$$

Evidently, the expressions in the square brackets in the right-hand side of (6.4) tend to 0 as $t \to 0$ and they are majorized by integrable variables $\xi^2 \sup_{|x|\le 1} |R(x)|$ and $5\xi^2/2$ respectively, so (6.1) follows.

Let $Z_t = \exp(t\xi)\mathbf{I}(|t\xi| \le 1)/f(t)$. It now follows from (6.1) that

$$\frac{\sqrt{Z_t} - 1}{t} \to \frac{\xi - E\xi}{2}$$

as $t \to 0$. To prove that this convergence holds also in $L^2(P)$, it is enough to check that

$$\frac{E(\sqrt{Z_t} - 1)^2}{t^2} \to \frac{1}{4}E(\xi - E\xi)^2$$

or, equivalently,

$$\frac{1 - E\sqrt{Z_t}}{t^2} \to \frac{E\xi^2 - (E\xi)^2}{8}. \quad (6.5)$$

But $E\sqrt{Z_t} = f_1(t/2)/\sqrt{f(t)}$, where $f_1(t) = E[\exp(t\xi)\mathbf{I}(|t\xi| \le 1/2)]$. The function $f_1(t)$ has the same expansion (6.1) as $f(t)$, and (6.5) follows after simple calculations.

The last assertion of the lemma follows from the dominated convergence theorem.

In the next lemma we follow the set-up of Section 2. The sequence φ_n is defined by (2.4) with some number $\vartheta > 0$. We assume that P^n and Q^n, $n \in \mathbf{N}$, are measures from $\mathcal{P}$ and p^n and q^n are the corresponding offspring distributions. The sub-index n means the restriction of a measure to the σ-field $\mathcal{F}_n$.

Lemma 5. *Let $\vartheta(p^n) - \vartheta = O(\varphi_n)$ and $\rho^2(p^n, q^n) = o(\varphi_n^2)$ as $n \to \infty$. Then $\|P_n^n - Q_n^n\| = o(1)$ as $n \to \infty$.*

Proof. The Hellinger process h^n of order $\frac{1}{2}$ between P^n and Q^n with respect to the filtration $(\mathcal{F}_n)$ is given, according to (2.3), by

$$h_t^n = \sum_{k=1}^{t} [1 - H^n(x_{k-1})], \quad t = 1, 2, \dots,$$

where $H^n(j)$, $j = 1, 2, \ldots$, is the Hellinger integral of order $\frac{1}{2}$ between the j-fold convolutions of p^n and q^n. It is clear that

$$H^n(j) \geq [H^n(1)]^j = [1 - \rho^2(p^n, q^n)]^j \geq 1 - j\rho^2(p^n, q^n).$$

Hence

$$h_t^n \leq \rho^2(p^n, q^n) \sum_{k=1}^{t} x_{k-1}$$

and

$$E^n h_n^n \leq B_n(\vartheta(p^n))\rho^2(p^n, q^n),$$

where E^n is the expectation with respect to P^n and $B_n(\cdot)$ is defined by (2.8). The hypotheses of the lemma and Lemma 1 imply now that $E^n h_n^n \to 0$, $n \to \infty$. Hence $h_n^n \xrightarrow{P^n} 0$, and the result follows from Theorem V.4.31 in (Jacod and Shiryaev, 1987).

In the next lemma P and P^n, $n = 1, 2, \ldots$, are measures in $\mathcal{P}$, p and p^n are corresponding offspring distributions, $\vartheta = \vartheta(P)$, $\vartheta_n = \vartheta(P^n)$, $\sigma^2 = \sigma^2(P)$ and $\sigma_n^2 = \sigma^2(P^n)$; φ_n and G_n are defined according to (2.4) and (2.7).

Lemma 6. *Let*

$$\limsup_{n\to\infty} \|P_n - P_n^n\| < 2(1 - \varepsilon), \qquad \varepsilon > 0. \tag{6.6}$$

Then

$$\limsup_{n\to\infty} \varphi_n^{-1}|\vartheta_n - \vartheta| \leq C_1 < \infty,$$

where C_1 depends only on P, Σ and ε.

Proof. Let $H^n(j)$, $j = 1, 2, \ldots$, be the Hellinger integral of order $\frac{1}{2}$ between the j-fold convolutions of p and p^n. Put $L_n(j) = -\log H^n(j)$ and

$$\zeta_n = \exp\big(-\sum_{k=1}^{n} L_n(x_{k-1})\big).$$

The proof consists of two steps. First we shall prove that condition (6.6) implies

$$\liminf_{n\to\infty} E\zeta_n > \varepsilon^2. \tag{6.7}$$

Then we shall show that

$$\varphi_n^{-2}(\vartheta_n - \vartheta)^2 G_n \zeta_n^2 \leq C_2 < \infty, \tag{6.8}$$

where C_2 depends only on Σ. The statement of the lemma follows from (6.7) and (6.8). Indeed, $\zeta_n \leq 1$, hence, by (6.7),

$$P(\zeta_n \geq \varepsilon^2/2) \geq \varepsilon^2/2 \tag{6.9}$$

for n large enough. On the other hand, $\mathcal{L}(G_n \mid P_n) \Rightarrow G$ and $G > 0$ (P^*-a.s.) by Proposition 2, therefore,

$$P(G_n \geq \delta) \geq 1 - \varepsilon^2/4, \qquad n = 1, 2, \ldots, \tag{6.10}$$

for some $\delta > 0$ depending only on P. From (6.9) and (6.10) we obtain $P(G_n \zeta_n^2 \geq \delta\varepsilon^4/4) \geq \varepsilon^2/4$, so (6.8) implies $\varphi_n^{-2}(\vartheta_n - \vartheta)^2 \leq 4C_2/\delta\varepsilon^4$.

Let h_t^n, $t = 1, 2, \ldots$, be the Hellinger process of order $\frac{1}{2}$ between P and P^n with respect to the filtration $(\mathcal{F}_n)$. According to (2.3), we have $h_t^n = \sum_{k=1}^t [1 - H^n(x_{k-1})]$. Hence,

$$\zeta_n = \prod_{k=1}^n H^n(x_{k-1}) = \prod_{k=1}^n (1 - h_k^n + h_{k-1}^n) = \mathcal{E}(-h^n)_n,$$

where $\mathcal{E}(\cdot)$ is the Doléans exponential. (6.7) follows now from (6.6) and from the following estimate for the variation $\|P_n - P_n^n\|$:

$$\|P_n - P_n^n\| \geq 2(1 - \sqrt{E\mathcal{E}(-h^n)_n}),$$

see (Kabanov et al., 1986, Theorem 2.1).

Since the j-fold convolution of p (respectively, p^n) has the mean $j\vartheta$ (respectively, $j\vartheta_n$) and the variance $j\sigma^2$ (respectively, $j\sigma_n^2$), we have the estimate

$$\frac{1 - H^n(j)}{[H^n(j)]^2} \geq \frac{(\vartheta_n - \vartheta)^2}{4(\sigma^2 + \sigma_n^2)} j,$$

which follows immediately from the distance inequality proved in (Kholevo, 1973) or (Le Cam and Yang, 1990, p. 128, Corollary 1). In other terms,

$$e^{2L_n(j)} - e^{L_n(j)} \geq C_2^{-1}(\vartheta_n - \vartheta)^2 j, \tag{6.11}$$

where $C_2 = 8\Sigma$.

The function $f(x) = e^{2x} - e^x$, $x \geq 0$, is convex and $f(0) = 0$, hence $f(x_1) + \cdots + f(x_n) \leq f(x_1 + \cdots + x_n)$ for arbitrary n and $x_i \geq 0$, $i = 1, \ldots, n$. Using this property and (6.11), we get

$$\begin{aligned} C_2^{-1}(\vartheta_n - \vartheta)^2 \sum_{k=1}^n x_{k-1} &\leq \sum_{k=1}^n f(L_n(x_{k-1})) \leq f\Big(\sum_{k=1}^n L_n(x_{k-1})\Big) \\ &= f(-\log \zeta_n) = \zeta_n^{-2} - \zeta_n^{-1} \leq \zeta_n^{-2}, \end{aligned}$$

which completes the proof of (6.8) and the lemma.

Acknowledgments. This work was mostly done while the author was a research visitor at the University of Bonn in the cooperation program supported by the Volkswagen-Stiftung. Research was partially supported by the Russian Foundation of Fundamental Researches, Grant 93-011-1440, and by the International Science Foundation Grant MMK000. The work was completed at the Weierstrass—Institut für Angewandte Analysis und Stochastik.

REFERENCES

Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and adaptive estimation in semiparametric models.* John Hopkins Univ. Press, Baltimore.

Dzhaparidze, K. and Spreij, P. (1993). On optimality of regular projective estimators in semimartingale models. *Stochastics and Stoch. Reports.* **43**, 161–178.

Feigin, P. D. (1977). A note on maximum likelihood estimation for simple branching processes. *Austral. J. Statist.* **19**, 152–154.

Godambe, P. E. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55**, 231–244.

Greenwood, P. E. and Wefelmeyer, W. (1993). Asymptotic minimax results for stochastic process families with critical points. *Stoch. Processes Appl.* **44**, 107–116.

Gushchin, A. A. (1994). On the convergence of sequences of semimartingales and their components. In: *Proc. of the Steklov Institute of Mathematics*, Issue 4, pp. 35–95.

Gushchin, A. A. (1995a). On asymptotic optimality of estimators of parameters under the LAQ condition. *Theory Probab. Appl.* **40**, 286–300 (in Russian).

Gushchin, A. A. (1995b). On some inequalities of Cramér–Rao type. In: *Frontiers in Pure and Applied Probability* II. Ed. by A. N. Shiryaev et al., TVP, Moscow (to appear).

Harris, T. E. (1948). Branching processes. *Ann. Math. Statist.* **19**, 474–494.

Heyde, C. C. (1975). Remarks on efficiency in estimation for branching processes. *Biometrika.* **62**, 49–55.

Heyde, C. C. and Lin, Y.-X. (1992). On quasi-likelihood methods and estimation for branching processes and heteroscedastic regression models. *Austral. J. Statist.* **34**, 199–206.

Höpfner, R., Jacod, J., and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Rel. Fields.* **86**, 105–129.

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

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin, New York.

Kabanov, Yu. M., Liptser, R. Sh. and Shiryaev, A. N. (1986). On the variation distance for probability measures defined on a filtered space. *Probab. Theory Rel. Fields.* **71**, 19–35.

Kholevo, A. S. (1973). A generalization of the Rao–Cramér inequality. *Theory Probab. Appl.* **18**, 359–362.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

Le Cam, L. and Yang, G. L. (1988). Distinguished statistics, loss of information and a theorem of Robert B. Davies. In: *Statistical Decision Theory and Related Topics* IV, Vol. 2. Ed. by S. S. Gupta and J. O. Berger. Springer-Verlag, New York.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer-Verlag, New York.

Levit, B. Ya. (1973). On optimality of some statistical estimates. In: *Proc. of the Prague Symposium on Asymptotic Statistics*, Vol. 2. Charles University, Prague, pp. 215–238.

Levit, B. Ya. (1975). On the efficiency of a class of nonparametric estimates. *Theory Probab. Appl.* **20**, 723–740.

Liptser, R. S. and Shiryaev, A. N. (1977). *Statistics of Random Processes.* Springer-Verlag, New York.

Seneta, E. (1970). A note on the supercritical Galton–Watson process with immigration. *Math. Biosc.* **6**, 305–311.

Shiryaev, A.N. and Spokoiny, V.G. (1994). *Statistical Experiments and Decisions: Asymptotic Theory.* Springer-Verlag (to appear).

Strasser, H. (1985). *Mathematical Theory of Statistics.* De Gruyter, Berlin.

Touati, A. (1993). Two theorems on convergence in distribution for stochastic integrals and statistical applications. *Theory Probab. Appl.* **38**, 95–117.

Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stoch. Processes Appl.* **31**, 261–282.

Wei, C. Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.* **18**, 1757–1773.

Winnicki, J. (1988) Estimation theory for the branching process with immigration. *Contemporary Math.* **80**, 301–322.

Witting, H. (1985). *Mathematische Statistik.* I. Teubner, Stuttgart.

Recent publications of the
Weierstraß–Institut für Angewandte Analysis und Stochastik

Preprints 1995

146. Anton Bovier, Véronique Gayrard: An almost sure large deviation principle for the Hopfield model.

147. Hans Babovsky: Limit theorems for deterministic Knudsen flows between two plates.

148. Björn Sandstede: Stability of multiple–pulse solutions.

149. Björn Sandstede: Constructing dynamical systems possessing homoclinic bifurcation points of codimension two.

150. Boris N. Khoromskij, Siegfried Prössdorf: Multilevel preconditioning on the refined interface and optimal boundary solvers for the Laplace equation.

151. Anton Bovier, Christof Külske: There are no nice interfaces in 2+1 dimensional SOS–models in random media.

152. Ilja Schmelzer: Covariant geometry description.

153. Alexander Korostelev, Michael Nussbaum: Density estimation in the uniform norm and white noise approximation.

154. Peter Hall, Michael Nussbaum, Steven E. Stern: On the estimation of a support curve of indeterminate sharpness.

155. Lev D. Pustyl'nikov: On the stability of solutions and absence of Arnol'd diffusion in a nonintegrable Hamiltonian system of a general form with three degrees of freedom.

156. Ralf Kornhuber: Adaptive monotone multigrid methods for some non–smooth optimization problems.

157. Sergej Rjasanow, Wolfgang Wagner: A generalized collision mechanism for stochastic particle schemes approximating Boltzmann type equations.

158. Nikolai Nefedov, Klaus Schneider: Singularly perturbed systems: Case of exchange of stability.

159. Rainer Dahlhaus, Michael H. Neumann, Rainer von Sachs: Nonlinear wavelet estimation of time–varying autoregressive processes.

160. Henri Schurz: Numerical regularization for SDEs: Construction of nonnegative solutions.

161. Anton Bovier, Véronique Gayrard: The retrieval phase of the Hopfield model: A rigorous analysis of the overlap distribution.

162. Boris N. Khoromskij, Gunther Schmidt: A fast interface solver for the biharmonic Dirichlet problem on polygonal domains.

163. Michael H. Neumann: Optimal change–point estimation in inverse problems.

164. Dmitry Ioffe: A note on the extremality of the disordered state for the Ising model on the Bethe lattice.

165. Donald A. Dawson, Klaus Fleischmann: A continuous super–Brownian motion in a super–Brownian medium.

166. Norbert Hofmann, Peter Mathé: On quasi–Monte Carlo simulation of stochastic differential equations.

167. Henri Schurz: Modelling, analysis and simulation of stochastic innovation diffusion.

168. Annegret Glitzky, Rolf Hünlich: Energetic estimates and asymptotics for electro–reaction–diffusion systems.

169. Pierluigi Colli, Jürgen Sprekels: Remarks on the existence for the one–dimensional Frémond model of shape memory alloys.

170. Klaus R. Schneider, Adelaida B. Vasil'eva: On the existence of transition layers of spike type in reaction–diffusion–convection equations.

171. Nikolaus Bubner: Landau–Ginzburg model for a deformation–driven experiment on shape memory alloys.

172. Reiner Lauterbach: Symmetry breaking in dynamical systems.

173. Reiner Lauterbach, Stanislaus Maier–Paape: Heteroclinic cycles for reaction diffusion systems by forced symmetry–breaking.

174. Michael Nussbaum: Asymptotic equivalence of density estimation and Gaussian white noise.