INADEQUACY OF ANSCOMBE-CHOW-ROBBINS TYPE PROCEDURES FOR SUPERCRITICAL BRANCHING PROCESSES

T. N. Sriram and A. N. Vidyashankar
Department of Statistics, 204 Statistics Building, University of Georgia
Athens, Georgia 30602, U.S.A.

ABSTRACT

This paper raises the question of adequacy of Anscombe-Chow-Robbins (ACR) type stopping rules in the inference for regular non-ergodic processes. It is shown that the usual recipe for constructing an ACR type stopping rule *fails* to provide an asymptotically risk efficient procedure for estimating the mean of a single-type supercritical branching process. However, the method does yield an asymptotically efficient sampling rule in the sense of Chow and Robbins (1965). This is established using tools from large deviation theory for branching processes. These results illustrate that while the ACR rules may perform well for stationary and ergodic processes, they could fail in non-ergodic models.

AMS 1991 *subject classification.* Primary 60J80, 62L12, 60G40.

Key words and Phrases. Branching processes, non-ergodic family, stopping rule, sequential estimation, asymptotic risk efficiency, asymptotic efficiency, iteration of generating functions, large deviations.

Short running title: Inadequacy of ACR rules.

1 Introduction

The purpose of this paper is to investigate the performance of a Anscombe-Chow-Robbins (ACR) type stopping rule and the associated sequential procedure for inference in supercritical branching processes. Sequential point estimation of parameters associated with independent and identically distributed (i.i.d.) random variables and even stationary and ergodic random variables have been extensively discussed in the literature. For these situations, it is rather folklore that the ACR type stopping rules can be used to produce asymptotically risk efficient sequential procedures. See, for instance, Woodroofe (1982), Martinsek(1983), and Ghosh, Mukhopadhyay and Sen (1997) for the i.i.d. case and Sriram (1987, 1988), Fakhre-Zakeri and Lee (1992), Lee (1994) and Sriram and Vidyashankar (1997) for the stationary and ergodic case.

A question of importance is, do the ACR type stopping rules continue to provide asymptotically risk efficient procedures for sequential estimation problems associated with processes belonging to, say, regular non-ergodic family? In this paper we show that the answer to the above question is in the *negative* when one considers sequential point estimation of the offspring mean in single-type supercritical branching processes. However, we establish that such a stopping rule is asymptotically efficient in the sense of Chow and Robbins (1965). It is interesting to note that such an anomaly can be explained using the large deviation theory for branching processes developed in Athreya (1994) and Vidyashankar (1994); see section 6 below.

2 Statement of the problem

Let $\{Z_n : n \geq 0\}$ be a Galton-Watson process with offspring probability distribution $p_j \equiv P[Z_1 = j]$ for $j \geq 0$ and $Z_0 \equiv 1$. Let $p_0 = 0$ and

$$m = \sum_{j \geq 1} j p_j \tag{2.1}$$

which will assumed to be greater than 1. Let σ^2 be the offspring variance which will be assumed to be positive and finite. Let $\{\xi_{i,j}\}$ be the offspring sequence which is i.i.d. with distribution $\{p_j : j \geq 0\}$. For basic aspects of branching processes see Athreya and Ney (1972).

Consider the problem of bounded risk point estimation of m by a simple estimator $\hat{m}_n$ defined by

$$\hat{m}_n = Z_n / Z_{n-1}, \quad n \geq 1 \tag{2.2}$$

subject to the squared error loss

$$L_n = (\hat{m}_n - m)^2. \tag{2.3}$$

The aim here is to find a minimum sample size for which the risk $R_n \equiv EL_n$ is no larger than b, i.e.,

$$R_n \leq b \tag{2.4}$$

for a prespecified value $b > 0$.

We will show that there exists a nonrandom sample size n_b^* satisfying (2.4) for all small values of b. However, n_b^* will depend on unknown parameters. To overcome this, it is customary to replace the unknown parameters in n_b^* by their consistent estimators. This leads to a Anscombe-Chow-Robbins type stopping rule, T_b, and an associated sequential procedure. In the stationary ergodic case such a sequential scheme, under regularity, leads simultaneously to asymptotic efficiency and risk-efficiency. However, we show that, in the above non-ergodic case such a recipe for constructing a sequential scheme *fails to provide* an asymptotically risk efficient procedure. Despite this, the sequential scheme *does yield* an asymptotically efficient sampling rule in the sense of Chow and Robbins (1965).

We begin by calculating the risk R_n. One can show, using a conditioning argument, that the risk in (2.4) is

$$\begin{aligned} R_n &= E(Z_n - mZ_{n-1})^2/Z_{n-1}^2 \\ &= \sigma^2 E(1/Z_{n-1}). \end{aligned} \tag{2.5}$$

At this point we would like to obtain an explicit expression for R_n and use it to find an n which minimizes the risk. In general, this is not possible. However, one can use the first order approximation for R_n to achieve this goal. To this end, we begin by recalling a recent result from Athreya (1994, see pages 784-786) [also see Athreya and Vidyashankar (1995)].

Theorem 2.1. Assume that $p_1 m > 1$. Then,

$$\begin{aligned} \lim_{n\to\infty} p_1^{-n} E(1/Z_n) &= \int_0^\infty Q(e^{-\theta})d\theta \\ &= \beta, \quad \text{say}, \end{aligned} \tag{2.6}$$

where

$$Q(s) = s + p_1^{-1} \sum_{r=0}^{\infty} g(f_r(s)p_1^{-r}, \tag{2.7}$$

$$f_r(s) = Es^{Z_r} = f(f(\dots f(s))) \tag{2.8}$$

is the r-th iterate of the probability generating function $f(s) = Es^{Z_1}$ and

$$g(s) = f(s) - p_1 s \tag{2.9}$$

for $0 \le s \le 1$. Furthermore, β is positive and finite. □
Now,

$$\begin{aligned}
\lim_{n\to\infty} p_1^{-n} R_n &= \lim_{n\to\infty} \sigma^2 p_1^{-n} E(1/Z_{n-1}) \\
&= \lim_{n\to\infty} (\sigma^2/p_1) p_1^{-(n-1)} E(1/Z_{n-1}) \\
&= p_1^{-1}\sigma^2\beta
\end{aligned} \tag{2.10}$$

and hence,

$$R_n = (p_1^{-1}\sigma^2\beta)p_1^n + o(p_1^n) \quad \text{as} \quad n \to \infty. \tag{2.11}$$

Note that β in (2.6) depends on $\{p_j : j \ge 1\}$. If p_j's are known, then we can choose a sample size n_b^* defined by

$$n_b^* = \text{first} \quad n \ge \log(\sigma^2\beta p_1^{-1}/b)/\log(p_1^{-1}). \tag{2.12}$$

One can show that $n_b^* \to \infty$ as $b \to 0$ and

$$\lim_{b\to 0} \left[\log(p_1^{-1}) n_b^* / \log(\sigma^2\beta p_1^{-1}/b)\right] = 1. \tag{2.13}$$

Moreover, from (2.5)

$$\begin{aligned}
R_{n_b^*}/b &= b^{-1}\sigma^2 E(1/Z_{n_b^*-1}) \\
&= b^{-1}\beta\sigma^2 p_1^{(n_b^*-1)} \left[\beta^{-1} p_1^{-(n_b^*-1)} E(1/Z_{n_b^*-1})\right] \\
&\le \beta^{-1} p_1^{-(n_b^*-1)} E(1/Z_{n_b^*-1}),
\end{aligned} \tag{2.14}$$

where the last inequality follows from the definition of n_b^* in (2.12). Therefore, from (2.6)

$$\limsup_{b\to 0} R_{n_b^*}/b \le 1. \tag{2.15}$$

The result in (2.15) says that if all p_j's are known then, for small values of $b > 0$, n_b^* provides a non-random sample size for which the risk $R_{n_b^*} \le b$. In fact, from (2.10) and the fact that $n_b^* \to \infty$ as $b \to 0$ we also have that

$$\lim_{b\to 0} p_1^{-n_b^*} R_{n_b^*} = p_1^{-1}\sigma^2\beta. \tag{2.16}$$

However, when p_j's are unknown, n_b^* cannot be used. For this case, we assume that the $\{\xi_{i,j}\}$ sequence is observable so that we can estimate the p_j's and define the following Anscombe-Chow-Robbins type stopping rule:

$$T_b = \inf\{n \geq n_{1b} : \quad n \geq \log(\hat{\sigma}_n^2 \hat{\beta}_n \hat{p}_{1,n}^{-1}/b)/\log(\hat{p}_{1,n}^{-1})\}, \tag{2.17}$$

where $n_{1b}(\geq 2)$ is the initial sample size depending on b and the estimators $\hat{\sigma}_n^2$, $\hat{\beta}_n$ and $\hat{p}_{1,n}$ are defined as follows:

$$\hat{p}_{k,n} = \sum_{j=1}^{Z_{n-1}} I_{\{\xi_{n-1,j}=k\}}/Z_{n-1}, k \geq 1, \quad \hat{\sigma}_n^2 = Z_{n-1}^{-1} \sum_{j=1}^{Z_{n-1}} (\xi_{n-1,j} - \hat{m}_n)^2 \tag{2.18}$$

and

$$\hat{\beta}_n = \int_0^\infty \hat{Q}(e^{-\theta})d\theta \tag{2.19}$$

where $\hat{Q}(s)$ is an estimator of $Q(s)$ defined in (2.7):

$$\hat{Q}(s) = s + \hat{p}_{1,n}^{-1} \sum_{r=0}^{\infty} \hat{g}(\hat{f}_r(s))\hat{p}_{1,n}^{-r} \tag{2.20}$$

with $\hat{g}(s)$ as the estimator of $g(s)$ defined in (2.9), i.e.,

$$\hat{g}(s) = \sum_{k \geq 2} \hat{p}_{k,n} s^k; \quad \hat{f}(s) = \hat{g}(s) + \hat{p}_{1,n} s \tag{2.21}$$

and

$$\hat{f}_r(s) = \hat{f}(\hat{f}(\ldots \hat{f}(s))) \text{ is the } r\text{-th iterate of } \hat{f}. \tag{2.22}$$

The dependence of the initial sample size on b in (2.17) ensures that the stopping rule T_b does not stop too soon (see Chow and Yu (1981)).

Analysis of the properties of the stopping rule T_b necessitates the study of rates of convergence of the estimators involved in it. These rates of convergence results are also of independent interest. The rest of the paper is organized as follows. In section 3 we will show that $R_b/R_{n_b^*}$ *does not converge* to 1 as $b \to 0$ for any choice of initial sample size n_{1b} in (2.17) where $R_b = E(\hat{m}_{T_b} - m)^2$ is the risk associated with the sequential estimator $\hat{m}_{T_b}$. In section 5, we establish the asymptotic efficiency of the stopping rule T_b which requires results on the rates of convergence of the estimators defined in (2.18) to (2.22). These and related rate results are given in section 4. Section 6 contains concluding remarks.

3 Inadequacy of Stopping Rule

In this section we show that sequential procedure $\hat{m}_{T_b}$ defined in section 2 is *not* asymptotically risk efficient.

Theorem 3.1. Under the assumptions of Theorem 2.1,

$$\liminf_{b\to 0} R_b/R_{n_b^*} = \begin{cases} 0 & \text{if } \lim_{b\to 0}(n_{1b}-n_b^*)=\infty \\ L & \text{if } \lim_{b\to 0}(n_{1b}-n_b^*)=c \\ \infty & \text{if } \lim_{b\to 0}(n_{1b}-n_b^*)=-\infty, \end{cases}$$

where c is some constant and $L = p^{(1+c)}(1-m)^2/(\sigma^2\beta)$.

Proof. Write using the definition of $\hat{m}_n$ and T_b in (2.2) and (2.17) respectively that the risk

$$\begin{aligned} R_b &= E[(Z_{T_b}/Z_{T_b-1})-m]^2 \\ &\geq \int_{\{Z_1=1,\ldots,Z_{n_{1b}}=1\}} [(Z_{T_b}/Z_{T_b-1})-m]^2 dP. \end{aligned} \tag{3.1}$$

Since $n_{1b}\geq 2$, on the set $D_{n_{1b}} = \{Z_1=1,\ldots,Z_{n_{1b}}=1\}$ we have from (2.18) that

$$\hat{p}_{1,n_{1b}} = 1, \quad \hat{\sigma}^2_{n_{1b}} = (\xi_{n_{1b}-1,1}-1)^2 = 0, \tag{3.2}$$

since $\hat{m}_{n_{1b}} = 1$ and $\xi_{n_{1b}-1,1}=1$. Furthermore, on $D_{n_{1b}}$ we have from (2.18) to (2.22) that

$$\hat{p}_{k,n_{1b}} = 0 \text{ for } k\geq 2, \quad \hat{g}(s)=0 \text{ for all } s, \tag{3.3}$$

$$\hat{f}(s) = \hat{p}_{1,n_{1b}}s = s, \quad \hat{f}_r(s) = s \text{ for all } r\geq 1, \tag{3.4}$$

and hence

$$\hat{Q}(s) = s \tag{3.5}$$

which implies that $\hat{\beta}_{n_{1b}} = \int_0^\infty e^{-\theta}d\theta = 1$. Therefore, on the set $D_{n_{1b}}$ we have from the above calculations and the definition of T_b in (2.17) that $T_b = n_{1b}$. Hence, from (3.1)

$$\begin{aligned} R_b &\geq \int_{D_{n_{1b}}} (Z_{n_{1b}}/Z_{n_{1b}-1}-m)^2 dP \\ &= (1-m)^2 P\{D_{n_{1b}}\} \\ &= (1-m)^2 p_1^{n_{1b}}. \end{aligned} \tag{3.6}$$

From this and (2.16) we have that

$$\begin{aligned} \liminf_{b\to 0} R_b/R_{n_b^*} &= \liminf_{b\to 0} p_1^{-n_b^*} R_b/(p_1^{-n_b^*}R_{n_b^*}) \\ &\geq [p_1/(\sigma^2\beta)](1-m)^2 \liminf_{b\to 0} p_1^{(n_{1b}-n_b^*)}. \end{aligned}$$

The desired conclusion now follows. □

Remark. It is known (see Chow and Yu, 1981) that in the discrete case if the initial sample size is fixed (i.e. does not depend on b), then one will not obtain asymptotic risk efficiency. A usual remedy for this, in the stationary ergodic setup, is to choose the initial sample size $n_{1b} \to \infty$ appropriately so that $\frac{n_{1b}}{n_b^*} \to 0$. This is the only case of practical interest and it implies that $n_{1b} - n_b^* \to -\infty$ as $b \to 0$. Whence in the above theorem, the ratio of risks converges to ∞.

The above theorem shows that for the branching process case, there does not exist any choice of the initial sample size for which one obtains risk efficiency. See also the concluding remarks.

4 Rates of Convergence of Estimators

In the previous section we showed that the sequential procedure $\hat{m}_{T_b}$ is not asymptotically risk efficient. In the next section, however, we will show that the stopping rule T_b is asymptotically efficient in the sense of Chow and Robbins (1965); i.e., ET_b/n_b^* converges to 1 as $b \to 0$. The proof of asymptotic efficiency of T_b depends on the rate of convergence of the estimators $\hat{p}_{1,n}, \hat{m}_n$, $\hat{\beta}_n$, and $\hat{\sigma}_n^2$ defined in section 2, viz., (2.2) and (2.18) to (2.22). We establish these results in Lemma 4.2 and Theorem 4.3. We begin with a lemma whose proof is straightforward and is hence is omitted.

Lemma 4.1. Let $\{X_i : i \geq 1\}$ be any sequence of random variables, $\{d_i : i \geq 1\}$ be a nonnegative sequence of constants such that for $i \geq i_0$, $|X_i| \leq d_i$ for some $i_0 \geq 1$ and

$$\sum_{i \geq 1} d_i < \infty.$$

Then for any $M > 0$, there exists an r_0 and an $\tilde{M} > 0$ such that

$$P(|\sum_{i\geq 1} X_i| > M) \leq P(|\sum_{i=1}^{r_0} X_i| > \tilde{M}).$$

Lemma 4.2. Let $\epsilon > 0$ and $\hat{f}'$ denote the derivative of $\hat{f}$ defined in (2.21). If $p_1 m > 1$, then the following results hold as $n \to \infty$:

$$\text{(i)} \quad P\{|\hat{p}_{k,n} - p_k| > \epsilon\} = O(p_1^n), \ k \geq 1 \tag{4.1}$$

For any $0 < s^* < 1$,

$$\text{(ii)}\quad P\{\sup_{0\le s\le s^*}|\hat f(s) - f(s)| > \epsilon\} = O(p_1^n) \tag{4.2}$$

$$\text{(iii)}\quad P\{\sup_{0\le s\le s^*}|\hat f'(s) - f'(s)| > \epsilon\} = O(p_1^n) \tag{4.3}$$

$$\text{(iv)}\quad P\{\sup_{0\le s\le s^*}|\hat f_r(s) - f_r(s)| > \epsilon\} = O(p_1^n) \tag{4.4}$$

for each $r \ge 1$. If in addition $EZ_1^{2+\delta} < \infty$, for some $\delta > 0$ then

$$\text{(v)}\quad P\{|\hat m_n - m| > \epsilon\} = O(p_1^n) \tag{4.5}$$

furthermore, if $EZ_1^4 < \infty$ then

$$\text{(vi)}\quad P\{|\hat\sigma_n^2 - \sigma^2| > \epsilon\} = O(p_1^n). \tag{4.6}$$

Proof. For the proof of (i) note by Chebychev's inequality that for any fixed $k \ge 1$

$$\begin{aligned}
P\{|\hat p_{k,n} - p_k| > \epsilon\} &\le (1/\epsilon)^2 E\{|\sum_{j=1}^{Z_{n-1}}(I_{\{\xi_{n-1,j}=k\}} - p_k)|^2/Z_{n-1}^2\} \\
&= (1/\epsilon)^2 EZ_{n-1}^{-2}E\{|\sum_{j=1}^{Z_{n-1}}(I_{\{\xi_{n-1,j}=k\}} - p_k)|^2|Z_{n-1}\} \\
&\le K_0E(1/Z_{n-1}), \\
&= O(p_1^n) \qquad (4.7)
\end{aligned}$$

and the last step follows from (2.6). The proofs of (ii), (iii), and (iv) follow from (i), Lemma 4.1, and the observations that for any integer $j \ge 1$, $|\hat p_{j,n} - p_j|s^j \le 2(s^*)^j$, and $|j\hat p_{j,n} - jp_j|s^j \le 2j(s^*)^j$. Proof of (v) is given in Athreya (1994); see Corollary 1. As for (vi), rewrite $\hat\sigma_n^2$ in (2.18) as

$$\hat\sigma_n^2 = Z_{n-1}^{-1}\sum_{j=1}^{Z_{n-1}}(\xi_{n-1,j} - m)^2 - (\hat m_n - m)^2. \tag{4.8}$$

Therefore,

$$P\{|\hat\sigma_n^2 - \sigma^2| > \epsilon\} \le P\{|\sum_{j=1}^{Z_{n-1}}[(\xi_{n-1,j} - m)^2 - \sigma^2]/Z_{n-1}| > \epsilon/2\} + P\{|\hat m_n - m| > \sqrt{\epsilon/2}\}. \tag{4.9}$$

Once again use Chebyschev's inequality, conditioning w.r.t. Z_{n-1} and (2.6) to show that the first term on the right hand side of (4.9) is also $O(p_1^n)$. The required result in (4.6) now follows from (4.5).

The next theorem establishes, using large deviation theory, that $\hat{\beta}_n$ converges to β geometrically fast. More precisely, the result established is an upper bound on the rate of convergence of $\hat{\beta}_n$ to β. In fact, one can strengthen this to an exact limiting result. This and related large deviation issues are being considered by Vidyashankar.

Theorem 4.3. Under the assumptions that $p_1 m > 1$ and $EZ_1^{2+\delta} < \infty$ for some $\delta > 0$,

$$P(|\hat{\beta}_n - \beta| > \epsilon) = O(p_1^n) \quad \text{as } n \to \infty. \tag{4.10}$$

Proof: Note that $(\hat{\beta}_n - \beta)$ can be expressed as a sum of three terms where

$$Term_1 = \sum_{r \geq 0} (\hat{p}_{1,n}^{-(r+1)} - p_1^{-(r+1)}) \int_0^1 \frac{g(f_r(s))}{s} ds, \tag{4.11}$$

$$Term_2 = \sum_{r \geq 0} \hat{p}_{1,n}^{-(r+1)} \int_0^1 \frac{[g(\hat{f}_r(s)) - g(f_r(s))]}{s} ds, \tag{4.12}$$

and

$$Term_3 = \sum_{r \geq 0} \hat{p}_{1,n}^{-(r+1)} \int_0^1 \frac{(\hat{g}(\hat{f}_r(s)) - g(\hat{f}_r(s)))}{s} ds. \tag{4.13}$$

Thus,

$$P(|\hat{\beta}_n - \beta| > \epsilon) \leq P(|Term_1| > \epsilon/3) + P(|Term_2| > \epsilon/3) + P(|Term_3| > \epsilon/3). \tag{4.14}$$

We will show that for each $i = 1, 2, 3$ and $\delta > 0$

$$P(|Term_i| > \delta) = O(p_1^n) \quad \text{as } n \to \infty.$$

The proof will use Lemmas 4.1 and 4.2 crucially.

We begin with the proof of $P(|Term_1| > \delta) = O(p_1^n)$ as $n \to \infty$. Note that

$$P(|Term_1| > \delta) \leq P(|Term_1| > \delta, |\hat{p}_{1,n} - p_1| \leq \eta_1) + P(|\hat{p}_{1,n} - p_1| > \eta_1),$$

for any $\eta_1 > 0$. Thus, by (4.1) it is enough to establish that the first term of the above expression is $O(p_1^n)$ as $n \to \infty$. Note that, since β in (2.6) is finite we have that

$$\sum_{r\geq 0} p_1^{-(r+1)} \int_0^1 \frac{g(f_r(s))}{s} ds < \infty \tag{4.15}$$

and on the set $|\hat{p}_{1,n} - p_1| \leq \eta_1$

$$\begin{aligned}\sum_{r\geq 0} \hat{p}_{1,n}^{-(r+1)} \int_0^1 \frac{g(f_r(s))}{s} ds &\leq \sum_{r\geq 0} (p_1 - \eta_1)^{-(r+1)} \int_0^1 \frac{g(f_r(s))}{s} ds \\ &< \infty, \end{aligned} \tag{4.16}$$

where the finiteness follows from arguments similar to those used to establish the result in (4.15). Hence, by Lemma 4.1 there exists $r_0 \geq 1$ and $\tilde{\delta}$ such that

$$\begin{aligned} &P(|Term_1| > \delta, |\hat{p}_{1,n} - p_1| \leq \eta_1) \\ &\leq P\left(\left|\sum_{r=0}^{r_0} (\hat{p}_{1,n}^{-(r+1)} - p_1^{-(r+1)}) \int_0^1 \frac{g(f_r(s))}{s} ds\right| > \tilde{\delta}\right) \end{aligned}$$

which is $O(p_1^n)$ as $n \to \infty$ using calculations and (4.1) of Lemma 4.2.

We now proceed to establish that $P(|Term_2| > \delta)$ is $O(p_1^n)$ as $n \to \infty$. Again, observe that

$$P(|Term_2| > \delta) \leq P(|Term_2| > \delta, A_n) + P(A_n^c), \tag{4.17}$$

where $A_n = B_n \cap C_n$, with

$$B_n = \{|\hat{p}_{1,n} - p_1| \leq \eta_1\} \cap \{|\hat{m}_n - m| \leq \eta_2\}$$

and

$$C_n = \{sup_{0\leq s\leq s^*} |\hat{f}(s) - f(s)| \leq \eta_3\} \cap \{sup_{0\leq s\leq s^*} |\hat{f}'(s) - f'(s)| \leq \eta_4\}$$

for $s^* \in (0, 1)$ and $\eta_i > 0$ for i=1,2,3,4 are chosen appropriately in the proof. From Lemma 4.2, it follows that $P(A_n^c)$ is $O(p_1^n)$ as $n \to \infty$. Thus we only have to establish that the first term on the right side of (4.17) is of order $O(p_1^n)$ as $n \to \infty$. For this, from Lemma 4.1 it is enough to show that on A_n

$$\hat{p}_{1,n}^{-(r+1)} \int_0^1 \frac{g(\hat{f}_r(s))}{s} ds \tag{4.18}$$

can be dominated by a deterministic sequence, say d_r, which is summable. We will first show that on A_n, for any $0 < u^* < 1$

$$\hat{p}_{1,n}^{-(r+1)} \int_0^{u^*} \frac{g(\hat{f}_r(s))}{s} \; ds \leq d_r \tag{4.19}$$

where $\sum_{r\geq 0} d_r < \infty$.

Note that on A_n, establishing the above amounts to showing that

$$(p_1 - \eta_1)^{-(r+1)} \int_0^{u^*} \frac{g(\hat{f}_r(s))}{s} \; ds \leq d_r. \tag{4.20}$$

The idea of proof is as follows. Let us pretend for the moment that we have f_r instead of $\hat{f}_r$ in the above expression. Then note that, since $\frac{g(f_r(s))}{s} \leq K(f_r(s))^2 \leq Kp_1^{2r}$ for $s \leq u^*$ we have that (4.20) is bounded above by $K(p_1 - \eta_1)^{-(r+1)} p_1^{2r}$ and hence

$$\sum_{r\geq 1} \hat{p}_{1,n}^{-(r+1)} \int_0^{u^*} \frac{g(\hat{f}_r(s))}{s} ds < \infty.$$

Our method of proof will make precise the idea that, for large r, we are interested in the behavior of $\frac{g(\hat{f}_r(s))}{s}$ near 0. We begin by setting, for $0 \leq s \leq 1$, $l(s) = f^{-1}(s)$. Consider $\int_0^{u^*} \frac{g(\hat{f}_r(s))}{s} \; ds$. Set $s = l_r(t)$, where $l_r(\cdot)$ is the rth iterate of l. Then

$$\begin{aligned}
\int_o^{u^*} \frac{g(\hat{f}_r(s))}{s} ds &= \int_0^{f_r(u^*)} \frac{g(\hat{f}_r(l_r(t)))}{l_r(t)} \; l_r'(t) dt \\
&= \int_0^{f_r(u^*)} \left[\frac{g(\hat{f}_r(l_r(t)))}{t}\right] \left[\frac{t l_r'(t)}{l_r(t)}\right] dt \\
&\leq \left\{\sup_{0\leq t\leq f_r(u^*)} \left[\frac{t l_r'(t)}{l_r(t)}\right]\right\} \left\{\int_0^{f_r(u^*)} \frac{(\hat{f}_r(l_r(t)))^2}{t^2} \; t \; dt\right\} \\
&\leq \left\{\sup_{0\leq t\leq f_r(u^*)} \left[\frac{t l_r'(t)}{l_r(t)}\right]\right\} \left\{\sup_{0\leq t\leq f_r(u^*)} \frac{(\hat{f}_r(l_r(t)))^2}{t^2}\right\} \frac{(f_r(u^*))^2}{2}.
\end{aligned}$$

Note that,

$$\begin{aligned}
\sup_{0\leq t\leq f_r(u^*)} \left[\frac{t}{l_r(t)} l_r'(t)\right] &\leq \sup_{0\leq t\leq f_r(u^*)} \left[\frac{t}{l_r(t)} (l'(l(t)))^r\right] \\
&\leq \left\{\sup_{0\leq t\leq f_r(u^*)} \left[\frac{t}{l_r(t)}\right]\right\} (l'(0))^r
\end{aligned}$$

since $(l'(t))^r$ is decreasing in t for all $r \geq 1$. Now,

$$\begin{aligned}
\sup_{0\leq t\leq f_r(u^*)} \left[\frac{t}{l_r(t)}\right] &\leq \sup_{0\leq t\leq f_r(u^*)} \left|\frac{t}{l_r(t)} - \frac{1}{l_r'(t)}\right| + \sup_{0\leq t\leq f_r(u^*)} \left(\frac{1}{l_r'(t)}\right) \\
&\leq C(l'(0))^{-r}
\end{aligned}$$

for some constant C. Hence

$$\sup_{0\le t\le f_r(u^*)}\left[\frac{t}{l_r(t)}l_r'(t)\right]\le C. \tag{4.21}$$

Next consider $\sup_{0\le t\le f_r(u^*)}\left[\frac{\hat{f}_r(l_r(t))}{t}\right]^2$. Note that for $0\le s<1$ and all integers $k\ge 2$, $s^k\le s^2$ and hence we have that

$$\hat{f}(s)\le \max(\hat{p}_{1,n},1-\hat{p}_{1,n})(s+s^2).$$

Furthermore, on A_n, $p_1-\eta_1\le\hat{p}_{1,n}\le p_1+\eta_1$ and hence,

$$\hat{f}(s)\le \max(p_1+\eta_1,(1-p_1)+\eta_1)(s+s^2).$$

Observe that there exists η_1^* such that for $\eta_1<\eta_1^*$, $\gamma=\max(p_1+\eta_1,(1-p_1)+\eta_1)<1$, and hence $\hat{f}(s)\le\gamma(s+s^2)$. Now, setting $v(s)=s+s^2$ we have that on A_n,

$$\hat{f}(s)\le\gamma v(s).$$

Using the above inequality one can show that for $0\le s\le 1$, on A_n

$$\hat{f}_r(s)\le Cr\gamma^r v(s), \tag{4.22}$$

for some constant C. From this (since for $t\le f_r(u^*)$ $l_r(t)<1$)

$$(\frac{\hat{f}_r(l_r(t))}{t})^2\le r^2\gamma^{2r}[\frac{l_r(t)}{t}+\frac{(l_r(t))^2}{t}]^2. \tag{4.23}$$

Hence,

$$\begin{aligned}\sup_{0\le t\le f_r(u^*)}(\frac{\hat{f}_r(l_r(t))}{t})^2 &\le r^2\gamma^{2r}\sup_{0\le t\le f_r(u^*)}[\frac{l_r(t)}{t}+\frac{(l_r(t))^2}{t}]^2\\ &\le C_1r^2\gamma^{2r}(\frac{1}{p_1^{2r}}+C_2)\\ &\le Cr^2(\frac{\gamma}{p_1})^{2r}\end{aligned}$$

for some constants C, C_1 and C_2, where we have use the calculation similar to (4.21) in the above inequalities. Thus,

$$\begin{aligned}\int_0^{u^*}\frac{g(\hat{f}_r(s))}{s}\;ds &\le C[r(\frac{\gamma}{p_1})^r]^2\hat{f}_r^2(u^*)\\ &\le C[r^2\gamma^{2r}].\end{aligned}$$

Hence,

$$\sum_{r\geq 0}(p_1-\eta_1)^{-r}\int_0^{u^*}\frac{g(\hat{f}_r(s))}{s}\ ds \leq C\sum_{r\geq 0}(p_1-\eta_1)^{-r}r[\max(p_1+\eta_1,(1-p_1)+\eta_1)]^{2r}.$$

Also, there exists an η_1^{**} so that for $\eta_1 \leq \eta_1^{**}$

$$(p_1-\eta_1)^{-1}[\max(p_1+\eta_1,(1-p_1)+\eta_1)]^2 \leq \gamma^* < 1.$$

Now choosing $\eta_1 \leq \min(\eta_1^*, \eta_1^{**})$ we have that

$$\sum_{r\geq 0}(p_1-\eta_1)^{-r}\int_0^{u^*}\frac{g(\hat{f}_r(s))}{s}\ ds \leq c\sum_{r\geq 0}r(\gamma^{**})^r < \infty,$$

thus completing the proof of the first part.

To complete the proof, we will establish that

$$(p_1-\eta_1)^{-(r-1)}\int_{u^*}^1\frac{g(\hat{f}_r(s))}{s}ds \leq \tilde{d}_r \tag{4.24}$$

where $\sum \tilde{d}_r < \infty$. We begin by considering

$$I_r = \int_{u^*}^1\frac{g(\hat{f}_r(s))}{s}ds. \tag{4.25}$$

Set $s=\hat{f}(u)$. Then, since $\hat{f}^{-1}(u)\geq u$,

$$\begin{aligned} I_r &= \int_{\hat{f}^{-1}(u^*)}^1\frac{g(\hat{f}_{r+1}(u))}{\hat{f}(u)}\hat{f}'(u)du \\ &= \int_{u^*}^1 g(\hat{f}_{r+1}(u))[\frac{\hat{f}'(u)}{\hat{f}(u)}]du - \int_{u^*}^{\hat{f}^{-1}(u^*)} g(\hat{f}_{r+1}(u))[\frac{\hat{f}'(u)}{\hat{f}(u)}]du. \end{aligned} \tag{4.26}$$

Thus,

$$\begin{aligned} I_r &= \int_{u^*}^1\frac{g(\hat{f}_{r+1}(u))}{\hat{f}(u)}\hat{f}'(u)du - \int_{u^*}^{\hat{f}^{-1}(u^*)}\frac{g(\hat{f}_{r+1}(u))}{\hat{f}(u)}\hat{f}'(u)du \\ &\geq \left[\int_{u^*}^1\frac{g(\hat{f}_{r+1}(u))}{u}du\right](\hat{f}'(u^*)) - \int_{u^*}^{\hat{f}^{-1}(u^*)} g(\hat{f}_{r+1}(u))\frac{\hat{f}'(u)}{\hat{f}(u)}du \end{aligned}$$

where the last inequality follows from $\hat{f}(u)\leq u$ for $0\leq u\leq 1$ and $\hat{f}'(u)$ is increasing in u. Let $\eta_5>0$, $\eta_3>0$ be such that $m-(\eta_3+\eta_5)>1$. Let $u^*<1$ (which depends on η_5) be such that $f'(u^*)>m-\eta_5$. Note that on A_n,

$$\hat{f}'(u^*) \geq f'(u^*)-\eta_3 \geq m-(\eta_5+\eta_3) = \tilde{m} > 1. \tag{4.27}$$

Thus,

$$I_r \geq \tilde{m}\left[\int_{u^*}^{1} \frac{g(\hat{f}_{r+1}(u))}{u}\, du\right] - \int_{u^*}^{\hat{f}^{-1}(u^*)} g(\hat{f}_{r+1}(u))\frac{\hat{f}'(u)}{\hat{f}(u)}du.$$

Note that,

$$\begin{aligned}
\int_{u^*}^{\hat{f}^{-1}(u^*)} g(\hat{f}_{r+1}(u))\left\{\frac{\hat{f}'(u)}{\hat{f}(u)}\right\} du &\leq g(\hat{f}_r(u^*))\left[\sup_{u^*\leq u\leq \hat{f}^{-1}(u^*)} \frac{\hat{f}'(u)}{\hat{f}(u)}\right][\hat{f}^{-1}(u^*) - u^*] \\
&\leq (\hat{f}_r(u^*))^2\left[\sup_{u^*\leq u\leq 1} \frac{\hat{f}'(u)}{\hat{f}(u)}\right] \\
&\leq (\hat{f}_r(u^*))^2\left[\frac{\hat{f}'(1)}{\hat{f}(u^*)}\right] \\
&\leq (\hat{f}_r(u^*))^2(\frac{m+\eta_2}{f(u^*)-\eta_3})
\end{aligned}$$

where the last inequality follows from the fact that the calculations are being performed on A_n. Thus,

$$I_r \geq \tilde{m}I_{r+1} - (\hat{f}_r(u^*))^2(\frac{m+\eta_2}{f(u^*)+\eta_3})$$

and hence

$$I_{r+1} \leq \frac{1}{\tilde{m}}I_r + K(\hat{f}_r(u^*))^2 \tag{4.28}$$

where $K = \left[\frac{m+\eta_2}{f(u^*)+\eta_3}\right]\left[\frac{1}{\tilde{m}}\right]$. Now, iterating the above inequality one gets,

$$\begin{aligned}
I_{r+1} &\leq \frac{1}{\tilde{m}}I_r + K(\hat{f}_r(u^*))^2 \\
&\leq \frac{1}{\tilde{m}}[\frac{1}{\tilde{m}}I_{r-1} + K(\hat{f}_{r-1}(u^*))^2] + K(\hat{f}_r(u^*))^2 \\
&= \frac{1}{\tilde{m}^2}I_{r-1} + \frac{K}{\tilde{m}}(\hat{f}_{r-1}(u^*))^2 + K(\hat{f}_r(u^*))^2.
\end{aligned}$$

Thus,

$$I_{r+1} \leq K\sum_{j=1}^{r}\frac{\hat{f}_j(u^*))^2}{\tilde{m}^{r-j}} + \frac{1}{\tilde{m}^{r+1)}}I_0 \leq K\frac{1}{\tilde{m}^{(r+1)}}$$

where the constants K are different. Thus,

$$(p_1-\eta_1)^{-(r+1)}\int_{u^*}^{1}\frac{g(\hat{f}_r(s))}{s}ds \leq K(p_1-\eta_1)^{-(r+1)}\frac{1}{\tilde{m}^{(r+1)}}$$

and by the choice of η_1, η_5 and η_3 and the assumption that $p_1 m > 1$, we have the convergence of $\sum_r \frac{(p_1-\eta_1)^{-r+1}}{\tilde{m}^{r+1}}$. This completes the proof that $P(|Term_2| > \delta) = O(p_1^n)$ as $n \to \infty$.

Finally, $P(|Term_3| > \delta)$ can be shown to be $O(p_1^n)$ by establishing that

$$\int_0^1 \frac{\hat{g}(\hat{f}_r(s))}{s} ds \leq d_r$$

such that $\sum_r d_r < \infty$. This follows using the fact that $g(s) \leq s^2$ and the arguments used in establishing that the $P(|Term_2| > \delta) = O(p_1^n)$ as $n \to \infty$.

5 Expected Stopping Time

The following theorem shows that the stopping rule T_b defined in (2.17) is asymptotically efficient.

Theorem 5.1. Under the assumption that $p_1 m > 1$, $EZ_1^{2+\delta} < \infty$ for some $\delta > 0$ and that the initial sample size n_{1b} in (2.17) is such that $n_{1b} \to \infty$ and $n_{1b}/n_b^* \to 0$ as $b \to 0$, we have the following:

$$\text{(i) For each } b > 0,\ T_b < \infty \text{ almost surely (a.s.)},\ T_b \to \infty \text{ a.s. as } b \to 0 \tag{5.1}$$

$$\text{(ii) } T_b/n_b^* \to 1 \text{ a.s. as } b \to 0. \tag{5.2}$$

If in addition $EZ_1^4 < \infty$, then

$$\text{(iii) } E|T_b/n_b^* - 1| \to 0 \text{ as } b \to 0. \tag{5.3}$$

Proof. From strong law and Theorem 4.3 it follows immediately that, as $n \to \infty$,

$$\hat{m}_n \to m \text{ a.s., } \hat{p}_{k,n} \to p_k \text{ a.s. for each } k \geq 1, \tag{5.4}$$

$$\hat{\sigma}_n^2 \to \sigma^2 \text{ a.s. and } \hat{\beta}_n \to \beta \text{ a.s.,} \tag{5.5}$$

for the estimators defined in (2.2) and (2.18) to (2.19). From this it follows that $T_b < \infty$ a.s. for each fixed $b > 0$. Also, $T_b \to \infty$ a.s. as $b \to 0$ since $T_b \geq n_{1b}$ and $n_{1b} \to \infty$ as $b \to 0$. Furthermore, from (2.17) it follows that

$$T_b \geq \log(\hat{\sigma}_{T_b}^2 \hat{\beta}_{T_b} \hat{p}_{1,T_b}^{-1}/b)/\log(\hat{p}_{1,T_b}^{-1}) \tag{5.6}$$

and

$$T_b \leq n_{1b} + \log(\hat{\sigma}_{T_b-1}^2 \hat{\beta}_{T_b-1} \hat{p}_{1,T_b-1}^{-1}/b)/\log(\hat{p}_{1,T_b-1}^{-1}). \tag{5.7}$$

The assertion in (5.2) now follows from these inequalities, (2.13), (5.4), (5.5) and the fact that $T_b \to \infty$ as $b \to 0$.

Finally, for the assertion in (5.3) let $n_{2b} = [n_b^*(1+\epsilon)] + 1$ for n_b^* defined in (2.12) and $0 < \epsilon < 1$. Then, for T_b defined in (2.17) and $n \geq n_{2b}$

$$\begin{aligned} P[T_b > n] &\leq P[\{\log(\hat{\sigma}_n^2 \hat{\beta}_n \hat{p}_{1,n}^{-1}/b)/\log(\hat{p}_{1,n}^{-1})\} \geq n] \\ &\leq P[\{\log(\hat{\sigma}_n^2 \hat{\beta}_n \hat{p}_{1,n}^{-1}/b)/\log(\hat{p}_{1,n}^{-1})\} \geq n_{2b}] \\ &\leq P[(n_b^*)^{-1}\{\log(\hat{\sigma}_n^2 \hat{\beta}_n \hat{p}_{1,n}^{-1}/b)/\log(\hat{p}_{1,n}^{-1})\} \geq (1+\epsilon)] \\ &\leq P[\{(U_n - c_b)/(1-c_b)\}V_n \geq (1+\epsilon)], \end{aligned} \tag{5.8}$$

where we used (2.12) and let

$$U_n = \log(\hat{\sigma}_n^2 \hat{\beta}_n \hat{p}_{1,n}^{-1})/\log(\sigma^2 \beta p_1^{-1}) \tag{5.9}$$

$$V_n = \log(p_1^{-1})/\log(\hat{p}_{1,n}^{-1}) \quad \text{and} \quad c_b = \log(b)/\log(\sigma^2 \beta p_1^{-1}). \tag{5.10}$$

After some algebra the right hand side of (5.8) can be rewritten as

$$\begin{aligned} &= P[\{(U_n V_n - 1) - c_b(V_n - 1)\}(1-c_b)^{-1} \geq \epsilon] \\ &\leq P[|U_n V_n - 1| \geq (\epsilon/2)|1-c_b|] + P[|V_n - 1| \geq (\epsilon/2)|1 - 1/c_b|]. \end{aligned} \tag{5.11}$$

Clearly $c_b \to -\infty$ as $b \to 0$. Therefore, for any given $\Delta > 0$ there exists b_0 small such that for all $b \leq b_0$, $1 - c_b > \Delta$. Also, $1 - 1/c_b \to 1$ as $b \to 0$. Hence, for small $b > 0$ and some $\tilde{\epsilon} > 0$ the right hand side of (5.11) is

$$\begin{aligned} &\leq P\{|U_n V_n - 1| \geq (\epsilon/2)\Delta\} + P\{|V_n - 1| > \tilde{\epsilon}\} \\ &= O(p_1^n) \quad \text{as} \quad n \to \infty, \end{aligned} \tag{5.12}$$

where we used the fact that U_n and V_n are continuous functions of $\hat{m}_n$, $\hat{p}_{1,n}$, $\hat{\sigma}_n^2$ and $\hat{\beta}_n$, Lemma 4.2, Theorem 4.3 and a result similar to Lemma 1 of Sriram (1987). Therefore, from (5.8) to (5.12), for small b

$$\sum_{n=1}^{\infty} P(T_b > n) < \infty. \tag{5.13}$$

Furthermore, for $n_{2b} = [n_b^*(1+\epsilon)] + 1$

$$\begin{aligned} \int_{\{T_b > 2n_{2b}\}} T_b \, dP &\leq 2\int_{\{T_b > 2n_{2b}\}} (T_b - n_{2b}) dP \\ &\leq 2\int_{\{T_b > n_{2b}\}} (T_b - n_{2b}) dP \\ &= 2\sum_{n=n_{2b}}^{\infty} P(T_b > n) \\ &\to 0 \quad \text{as} \quad b \to 0, \end{aligned} \tag{5.14}$$

by (5.13). It follows from (5.14) that

$$\{T_b/n_b^*; 0 < b < \tilde{b}_0\} \text{ is uniformly integrable,} \tag{5.15}$$

for some $\tilde{b}_0$. Therefore, the assertion in (5.3) follows from (5.2). Hence the theorem. □

6 Concluding Remarks

Why did the ACR type stopping rule fail to produce a risk efficient procedure? To explain this, let $\hat{\theta}_n$ be an estimator of a parameter θ. Then in the stationary ergodic case, under regularity, the mean squared error is of order n^{-1} and hence $E(\hat{\theta}_{n_0} - \theta)^2 = O(n_0^{-1})$ where n_0 is the best fixed sample size. In this case, under regularity, the ACR method *does yield* an asymptotically efficient stopping rule ($ET = O(n_0)$) which in turn yields risk-efficient procedures.

In the branching process case, while the ACR recipe yields an asymptotically efficient stopping rule (i.e., $ET_b = O(n_b^*)$) it fails to produce an asymptotically risk-efficient procedure due to the fact that $R_{n_b}^* = O(p_1^{n_b^*})$. This difference in rates (algebraic growth of ET_b and geometric decay of $R_{n_b^*}$) is the reason for the failure of the ACR rule in our situation. Note that the geometric decay of the risk R_n is due to the large deviation behavior of $\hat{m}_n$.

One might superficially argue that the reason for the failure is due to the fact that the risk for estimating the mean in the branching process case is a function of $E(1/Z_n)$. It must be noted that even in the stationary ergodic case, if one considers different loss functions, e.g. relative squared error loss (see Etemadi, Sriram and Vidyashankar (1997) and Martinsek (1995)), one can show that the risk is a function of $E(1/S_n)$ where S_n is an appropriate partial sum. However, $E(1/S_n)$ still decays like n^{-1} under regularity; see Etemadi, Sriram and Vidyashankar (1997). In these cases, the usual method for constructing ACR type stopping rules *will continue to yield* risk-efficient procedures; once again see Etemadi, Sriram and Vidyashankar (1997). This explains the role played by the underlying processes in the risk of the sequential procedures.

All these suggest that, for non-ergodic models, one may need to modify the ACR recipe in order to obtain simultaneously asymptotic efficiency and risk efficiency. These and related issues are being pursued by the authors.

References

Athreya, K. B. (1994). Large deviation rates for branching processes -I, the single type case, *Ann. Appl. Probab.*, 4, 779-790.

Athreya, K.B. and Ney, P.E. (1972). *Branching Processes*, Springer, Berlin.

Athreya, K.B. and Vidyashankar, A.N. (1995): Large deviation rates for branching processes -II, the multi-type case, *Ann. Appl. Probab.*, 5, 566-576.

Chow, Y. S. and Robbins, H. E. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* 36, 457-462.

Chow, Y. S. and Yu, K. F. (1981). The performance of a sequential procedure for the estimation of the mean. *Ann. Statist.* 9, 184-189.

Etemadi, K., Sriram, T.N., and Vidyashankar, A.N. (1997). L_p convergence of reciprocal of sample means with applications to sequential estimation in Linear Regression, *Journ. Statist. Plan. Inf.* 65, 1-15.

Fakhre-Zakeri, I. and Lee, S. (1992). Sequential estimation of the mean of a linear process. *Sequen. Anal.* 11, 181-197.

Ghosh, M. Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation.* Wiley, New York.

Lee, S. (1994). Sequential estimation for the parameters of a stationary autoregressive model. *Sequen. Anal.* 13, 301-317.

Martinsek, A. T. (1983). Second order approximation to the risk of a sequential procedure. *Ann. Statist.* 11 827-836.

Martinsek, A.T. (1995). Estimation of a slope parameter in regression with prescribed proportional accuracy. *Statistics and Decisions*, 13, 363-377.

Sriram, T. N. (1987). Sequential estimation of the mean of a first-order stationary autoregressive process. *Ann. Statist.* 15, 1079-1090.

Sriram, T. N. and Vidyashankar, A. N. (1997). Sequential point estimation for branching processes-I, subcritical case. *Tech. Report* 97-16, Dept. Statistics, University of Georgia. Submitted.

Vidyashankar, A.N. (1994). Large deviation results for branching processes in fixed and random environments, *Unpublished Ph.d. Thesis*, Department of Statistics and Mathematics, Iowa State University, Ames, IA 50011.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* SIAM, Philadelphia.