Stochastic Processes and their Applications 9 (1979) 291–305

ASYMPTOTIC TESTS OF COMPOSITE HYPOTHESES FOR NON-ERGODIC TYPE STOCHASTIC PROCESSES

I.V. BASAWA

LaTrobe University and Cornell University, Ithaca, NY 14850, U.S.A.

H.L. KOUL*

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, U.S.A.

Received 7 September 1979
Revised 22 October 1979

Limiting distributions of a score statistic and the likelihood ratio statistic for testing a composite hypothesis involving several parameters in non-ergodic type stochastic processes are obtained. It is shown that, unlike in the usual theory (ergodic type processes), the limiting distributions of these statistics are different both under the null and a contiguous sequence of alternative hypotheses. The results are applied to a regression model with explosive autoregressive Gaussian errors. In the discussion of this example a modified score statistic is suggested where the limiting null and non-null distributions are the same as those of the likelihood ratio statistic.

AMS (1980) Subj. CLass.: Primary 62 M07, 63 M10
Score test explosive autoregression
likelihood-ratio test

1. Introduction

This paper is concerned with the limiting distribution of the *score* and *likelihood-ratio* (L.R.) test-statistics for testing composite hypotheses involving several parameters in the non-ergodic type (see Section 2 for definitions) stochastic processes. Under a suitable sequence of alternatives the limit distributions of these statistics will be shown to be mixtures of non-central chi square distributions. We defer discussion of optimality of these tests to a forthcoming paper.

Non-ergodic processes (in the sense of this paper) arise in several applications such as supercritical branching processes, explosive autoregressive processes, classical mixture experiments leading to exchangeable processes, etc. The problem of testing a simple hypothesis about a single parameter against one-sided alternatives for such processes was broached by Basawa and Scott [3, 4] who discussed some simple tests

* This research was supported by the NIH Grant No. 1 RO1 GM 25449-01.

and noted certain difficulties regarding the efficiency question. Feigin [13] and Sweeting [22] investigated some further properties of the tests discussed by Basawa and Scott.

Our results in this paper extend the work of Weiss [24] and Dzhaparidge [11] concerning tests for ergodic models to the non-ergodic situation. Section 2 is concerned with the specification of a non-ergodic model. The limit distributions of the score and L.R. statistics are derived in Section 3 using ideas analogous to those used by Weiss [24] and Dzhaparidge [11]. In Section 4 we discuss an application of the results to a simple regression model with explosive autoregressive errors.

2. A regular non-ergodic model

Let $\boldsymbol{X}_n=(X_1,\ldots,X_n)$ be a realization from a stochastic process with a joint density $p_n(\boldsymbol{x}_n;\theta)$ with respect to some σ-finite measure, where θ is a $(k\times 1)$ vector of unknown real parameters taking values in a subset Ω of the k-dimensional euclidean space. Consider the random variable (for each fixed $\theta\in\Omega$),

$$\Lambda_n(\theta_n,\theta)=\log\{p_n(\boldsymbol{X}_n;\theta_n)/p_n(\boldsymbol{X}_n;\theta)\},\qquad \theta_n=\theta+I_n^{-1/2}(\theta)h, \tag{2.1}$$

where h is a $(k\times 1)$ vector of finite real numbers and $I_n(\theta)$ is a $(k\times k)$ non-random diagonal matrix, to be further specified later, with positive diagonal entries $(k_{ni}(\theta), i=1,2,\ldots,k)$ such that $k_{ni}(\theta)$ are continuous in θ and $\uparrow\infty$ as $n\to\infty$. The asymptotic behavior of $\Lambda_n(\theta_n,\theta)$ will determine in some sense the large sample performance of the tests to be discussed later on. Assume now that the following conditions are fulfilled:

(A.1) There exists a $(k\times 1)$ random vector $\Delta_n(\theta)$ and a $(k\times k)$ random matrix $B_n(\theta)$ such that as $n\to\infty$,

$$|\Lambda_n(\theta_n,\theta)-[h^{\mathrm{T}}\Delta_n(\theta)-\tfrac{1}{2}h^{\mathrm{T}}\{I_n^{-1/2}(\theta)B_n(\theta)I_n^{-1/2}(\theta)\}h]|\xrightarrow{\mathrm{p}}0$$

both under $P_n(\theta)$ and $P_n(\theta_n)$ probabilities.

(A.2) Let $G_n(\theta)=I_n^{-1/2}(\theta)B_n(\theta)I_n^{-1/2}(\theta)$. Then $G_n(\theta)\xrightarrow{\mathrm{p}}G(\theta)$, both under $P_n(\theta)$ and $P_n(\theta_n)$ probabilities, where $G(\theta)$ is a possibly random $(k\times k)$ non-negative definite matrix with rank $l\leq k$.

(A.3) $(\Delta_n(\theta), G_n(\theta))\xrightarrow{\mathrm{d}}(\Delta(\theta), G(\theta))$, under $P_n(\theta)$, where $\Delta(\theta)$ is some random vector.

Definition 2.1. The density $p_n(\boldsymbol{x}_n;\theta)$ is said to belong to an *ergodic family* if it satisfies (A.1)–(A.3) with $G(\theta)$ being non-random. If, on the other hand, at least one element of $G(\theta)$ is a non-degenerate random variable we say that $p_n(\boldsymbol{x}_n;\theta)$ belongs to a *non-ergodic* family.

The above definition is a generalization of the one given by Basawa [2]. Also, see [9]. The models discussed in [6–8, 11, 17, 20, 23, 24], belong to the ergodic family. Examples and subclasses of the non-ergodic family have been previously studied, in a different context in [1, 3–5, 10, 12–16, 19].

In the present paper we will be concerned with a special version of the non-ergodic family specified by conditions (B.1) to (B.3) below:

(B.1) Condition (A.1) holds under $P_n(\theta)$ with the choice

$$\Delta_n(\theta) = I_n^{-1/2}(\theta) S_n(\theta) \quad \text{and} \quad B_n(\theta) = \left(\left(\frac{\partial^2 \log p_n(X_n;\theta)}{\partial\theta_i\,\partial\theta_j}\right)\right)$$

$(i, j, = 1, 2, \ldots, k)$, where $S_n(\theta)$ is a $(k \times 1)$ vector of *scores*

$$\left(\frac{\partial \log p_n}{\partial \theta_i}, \quad (i = 1, 2, \ldots, k)\right).$$

(B.2) Condition (A.2) holds under $P_n(\theta)$ only with at least one element of $G(\theta)$ being a non-degenerate random variable.

(B.3) Condition (A.3) is satisfied with the choice $\Delta(\theta) \overset{\mathrm{d}}{=} G^{1/2}(\theta)Z$, where Z is a $(k \times 1)$ vector of independent identically distributed $N(0, 1)$ variates and Z is independent of $G(\theta)$.

Remarks on conditions (B). (i) With the above choice of $\Delta_n(\theta)$ and $B_n(\theta)$, and under fairly mild regularity conditions, a large class of $p_n(x_n;\theta)$ will satisfy (B.1) via the Taylor expansion. The assumption that $I_n(\theta)$ is diagonal was also made by Weiss and Wolfowitz [25, Ch. 7], Weiss [24], and Dzhaparidge [11]. Note that this assumption does not necessarily imply that the Fisher information matrix $E(B_n(\theta))$ is diagonal. Also, the matrix $G(\theta)$ is not necessarily diagonal and it is allowed to be singular. The case when the norming matrix $I_n(\theta)$ is non-diagonal involves certain complexities in the derivation and will not be considered in the present paper.

(ii) Under a mild regularity conditions, $\{S_n(\vartheta), n = 1, 2, \ldots\}$ is known to be a zero-mean martingale. Under (B.2), and some further conditions such as those in [4, 5], a multivariate version of the central limit theorem for martingales ensures that (B.3) holds.

(iii) Under (B.1) to (B.3),

$$\Lambda_n(\theta_n, \theta) \overset{\mathrm{d}}{\to} \Lambda \overset{\mathrm{d}}{=} h^{\mathrm{T}} G^{1/2}(\theta) Z - \tfrac{1}{2} h^{\mathrm{T}} G(\theta) h$$

under $P_n(\theta)$ probability. Using the independence of Z and $G(\theta)$ one sees $\mathrm{E}(\mathrm{e}^{\Lambda}) = 1$. Thus, condition (S_3) of Roussas [20, Ch. 1], is satisfied and consequently, the sequences of probability measures $\{P_n(\theta)\}$ and $\{P_n(\theta_n)\}$ are *contiguous*. This particular property will enable us to derive the limit distributions of various statistics under the alternatives $\{\theta_n\}$ from the limit distributions under the null θ in an entirely standard way (see [20]).

3. Limit distributions of the score and likelihood-ratio tests

We assume throughout that $p_n(x_n;\theta)$ satisfies (B.1) to (B.3). Suppose we partition $\theta^{\mathrm{T}} = (\alpha^{\mathrm{T}}, \beta^{\mathrm{T}})$, where α is an $(s \times 1)$ vector and β, a $(k - s) \times 1$ vector. Consider the problem of testing $H: \beta = \beta_0$ against $K: \beta \neq \beta_0$, where α is treated as a nuisance parameter. We shall study the following test-statistics for this problem:

Score statistic. $T_{n1}=\Delta_n^{\mathrm{T}}(\tilde{\alpha}_{n0},\beta_0)\Delta_n(\tilde{\alpha}_{n0},\beta_0)$, where $\tilde{\alpha}_{n0}$ is an estimator of α to be specified later.

Likelihood-ratio (L.R.) statistic. The L.R. statistic is defined as

$$T_{n2}=2\Lambda_n((\tilde{\alpha}_n,\tilde{\beta}_n),(\tilde{\alpha}_{n0},\beta_0)),$$

where Λ_n is defined in (2.1) and $\tilde{\alpha}_{n0}$, $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ are the estimators of α and β to be specified later.

In order to derive the limit distributions of $\{T_{n1}\}$ and $\{T_{n2}\}$ we first introduce some conventions and notation.

Let $h^{\mathrm{T}}=(h_1^{\mathrm{T}},h_2^{\mathrm{T}})$ be a $1\times k$ vector of real numbers with h_1 being $s\times 1$ and h_2 being $(k-s)\times 1$. Write

$$I_n(\theta)=\begin{pmatrix} I_{n11}(\theta) & 0 \\ 0 & I_{n22}(\theta)\end{pmatrix},\qquad \theta^{\mathrm{T}}=(\alpha^{\mathrm{T}},\beta^{\mathrm{T}}),$$

where I_{n11} and I_{n22} are, respectively, $s\times s$ and $(k-s)\times(k-s)$ diagonal matrices. We now specify our alternatives

$$K_n:\qquad \beta=\beta_n=\beta_0+I_{n22}^{-1/2}(\alpha,\beta_0)h_2,$$

where β_0 is the H value of β and α is the nuisance parameter vector. Throughout the rest of the paper $\alpha_n=\alpha+I_{n11}^{-1/2}(\alpha,\beta_0)h_1$.

In what follows all limits are to be understood as $n\to\infty$; all $0_p(1)$ statements are under $\{P_n(\alpha,\beta_0)\}$. By remark (iii) following assumptions (B), $\{P_n((\alpha_n,\beta_n)\}$ and $\{P_n(\alpha,\beta_0)\}$ are mutually contiguous for any real k-vector h. Therefore taking $h_1=0$ in h implies that $\{P_n(\alpha,\beta_n)\}$ and $\{P_n(\alpha,\beta_0)\}$ are mutually contiguous and hence all $o_p(1)$ statements can be made under the alternatives $\{P_n(\alpha,\beta_n)\}$ also, whatever may be the parameter α. Finally for a sequence of r.vs. Y_n, $\mathscr{L}(Y_n|P_n)$ denotes the distribution of Y_n under a sequence of probability distributions P_n.

Limiting distribution of the score statistic. Partition the score matrix S_n as $S_n^{\mathrm{T}}(\alpha,\beta)=(S_{n1}^{\mathrm{T}}(\alpha,\beta),S_{n2}^{\mathrm{T}}(\alpha,\beta))$ corresponding to the first s parameters α and the last $(k-s)$ parameters β respectively. Similarly partition the matrix

$$G(\alpha,\beta)=\begin{pmatrix} G_{11}(\alpha,\beta) & G_{12}^{\mathrm{T}}(\alpha,\beta) \\ G_{12}(\alpha,\beta) & G_{22}(\alpha,\beta)\end{pmatrix}, \tag{3.1}$$

where G_{11} is $s\times s$, G_{22} is $(k-s)\times(k-s)$ and G_{12} is $s\times(k-s)$. We assume that the rank of the matrix G_{11} is s, $s<l=\mathrm{rank}(G)$ for all α,β.

Let $\tilde{\alpha}_{n0}$ be an estimator of α such that

$$G_{11}(\alpha,\beta_0)I_{n11}^{1/2}(\alpha,\beta_0)(\tilde{\alpha}_{n0}-\alpha)=I_{n11}^{-1/2}(\alpha,\beta_0)S_{n1}(\alpha,\beta_0)+o_p(1). \tag{3.2}$$

The restricted (under H) MLE of α typically satisfies (3.2). By assumption (B.1), $S_n(\theta)$ has first derivative in θ and hence the Taylor expansion, (3.2) and the continuity of $I_n(\theta)$ in θ imply

$$\begin{aligned}\Delta_{n1}(\tilde{\alpha}_{n0},\beta_0) &:= I_{n11}^{-1/2}(\tilde{\alpha}_{n0},\beta_0)S_{n1}(\tilde{\alpha}_{n0},\beta_0)\\ &= I_{n11}^{-1/2}(\alpha,\beta_0)S_{n1}(\alpha,\beta_0)\\ &\quad -\{I_{n11}^{-1/2}(\alpha,\beta_0)B_{n11}(\alpha,\beta_0)I_{n11}^{-1/2}(\alpha,\beta_0)\}\\ &\quad \times I_{n11}^{-1/2}(\alpha,\beta_0)(\tilde{\alpha}_{n0}-\alpha)+o_p(1).\\ &= o_p(1). \end{aligned}\tag{3.3}$$

By the same reasons we have

$$\begin{aligned}\Delta_{n2}(\tilde{\alpha}_{n0},\beta_0) &:= I_{n22}^{-1/2}(\tilde{\alpha}_{n0},\beta_0)S_{n2}(\tilde{\alpha}_{n0},\beta_0)\\ &= I_{n22}^{-1/2}(\alpha,\beta_0)S_{n2}(\alpha,\beta_0)\\ &\quad -\{I_{n22}^{-1/2}(\alpha,\beta_0)B_{n12}(\alpha,\beta_0)I_{n11}^{-1/2}(\alpha,\beta_0)\}\\ &\quad \times I_{n11}^{1/2}(\alpha,\beta_0)(\tilde{\alpha}_{n0}-\alpha)+o_p(1)\\ &= \Delta_{n2}(\alpha,\beta_0)-G_{12}(\alpha,\beta_0)G_{11}^{-1}(\alpha,\beta_0)\Delta_{n1}(\alpha,\beta_0)+o_p(1). \end{aligned}\tag{3.4}$$

From (3.3), (3.4) and assumption (B.3)' one has

$$\mathscr{L}(\Delta_{n2}(\tilde{\alpha}_{n0},\beta_0)\,|\,P_n(\alpha,\beta_0))\to\mathscr{L}(C^{1/2}Z_2),\tag{3.5}$$

where $C=G_{22}(\alpha,\beta_0)-G_{12}(\alpha,\beta_0)G_{11}^{-1}(\alpha,\beta_0)G_{12}^{\mathrm{T}}(\alpha,\beta_0)$ and where Z_2 is a $(k-s)\times 1$ vector of i.i.d. N(0, 1) r.vs. Note that the limit distribution in (3.5) is non-normal since C is a random variable.

A consequence of (3.5) is that

$$\mathscr{L}(T_{n1}\,|\,P_n(\alpha,\beta_0))\to\mathscr{L}(Z_2^{\mathrm{T}}CZ_2).\tag{3.6}$$

The distribution of $Z_2^{\mathrm{T}}CZ_2$ is not a chi-square but can be expressed as a mixture of a linear combination of independent chi-square variables. Next we note that the mutual contiguity of $\{P_n(\alpha,\beta_n)\}$ and $\{P_n(\alpha,\beta_0)\}$ imply

$$\mathscr{L}(T_{n1}\,|\,P_n(\alpha,\beta_n))\to\mathscr{L}((Z_2+C^{1/2}h_2)^{\mathrm{T}}C(Z_2+C^{1/2}h_2)).\tag{3.7}$$

This limiting distribution is somewhat involved, but can be expressed as a mixture of a linear combination of independent non-central chi-square variables.

Limit distribution of the L.R. statistic. Let $\tilde{\alpha}_n$, $\tilde{\beta}_n$ be estimators of α and β such that under $\{P_n(\alpha,\beta)\}$

$$G(\alpha,\beta)I_n^{1/2}(\alpha,\beta)\begin{pmatrix}\tilde{\alpha}_n-\alpha\\ \tilde{\beta}_n-\beta\end{pmatrix}=I_n^{-1/2}(\alpha,\beta)S_n(\alpha,\beta)+o_p(1).\tag{3.8}$$

The maximum likelihood estimates of α and β typically satisfy (3.8). The likelihood ratio (L.R.) statistic T_{n2} can be written as

$$T_{n2}=2\Lambda_n((\tilde{\alpha}_n,\tilde{\beta}_n),(\alpha,\beta_0))-2\Lambda_n((\tilde{\alpha}_{n0},\beta_0),(\alpha,\beta_0)),$$

where $\tilde{\alpha}_n$, $\tilde{\beta}_n$ satisfy (3.8) and $\tilde{\alpha}_{n0}$ satisfies (3.2).

Recall that the rank$(G) = l \leq k$. Rewrite, after interchanging rows and columns if necessary, the matrix G as

$$G = \begin{pmatrix} G^* & \square \\ \square & \square \end{pmatrix}_{k\times k}, \qquad G^* = \begin{pmatrix} G_{11} & G_{12}^{*\mathrm{T}} \\ G_{12}^* & G_{22}^* \end{pmatrix}. \tag{3.9}$$

The rank$(G^*) = l$, rank$(G_{11}) = s$ and the remaining rows and columns of G are partitioned appropriately in the submatrices indicate by $\square$. It then follows that a generalized inverse (see [18]) of G is

$$G^- = \begin{pmatrix} A & B^\mathrm{T} & 0 \\ B & C_*^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; C_* = G_{22}^* - G_{12}^* G_{11}^{-1} G_{12}^{*\mathrm{T}}, \tag{3.10}$$

where A and B are some matrices.

We partition Δ_n and h in conformity with (3.9) and write

$$\Delta_n = \begin{pmatrix} \Delta_{n1} \\ \Delta_{n2}^* \\ \square \end{pmatrix}, \qquad h = \begin{pmatrix} h_1 \\ h_2^* \\ \square \end{pmatrix}, \tag{3.11}$$

where Δ_{n1} and h_1 are $s \times 1$ vectors as before, Δ_{n2}^* is a $(l-s)\times 1$ vector and h_2^* is a $(l-s)\times 1$ vector.

After some lengthy but routine calculations similar to those that occur in the ergodic case (e.g. see [11]) one can verify, using the above notation, that

$$\lim_{n\to\infty} \mathscr{L}(T_{n2}|P_n(\theta_0)) = \lim_{n\to\infty} \mathscr{L}(\Delta_{n2}^{*\mathrm{T}}(\theta_0)C_*^{-1}(\theta_0)\Delta_{n2}^*(\theta_0)|P_n(\vartheta_0)) = \mathscr{L}(Z_2^{*\mathrm{T}}Z_2^*), \tag{3.12}$$

where $\theta_0 = (\alpha, \beta_0)$ and Z_2^* is a vector of $(l-s)\times 1$ i.i.d. N(0, 1) r.vs. Using contiguity one also can deduce that

$$\mathscr{L}(T_{n2}|P_n(\alpha, \beta_n)) \to \mathscr{L}((Z_2^* + C_*^{1/2}h_2^*)^\mathrm{T}(Z_2^* + C_*^{1/2}h_2^*)). \tag{3.13}$$

The limit distribution in (3.13) is a mixture of non-central chi-square r.vs. with mixing variable being the noncentrality parameter $\lambda = h_2^* C_* h_2^*$, where h_2^* is as in (3.11).

Thus, the limit distributions of T_{n1} and T_{n2} are different both under H and the sequence of alternatives K_n. In the ergodic models it is well known (see e.g. [11, 24]) that these statistics have asymptotically the same limit distributions both under H and K_n, except for a scale change. If G were known and *non-random* (i.e. in the ergodic model) one may consider a modified score-statistic $T_{n1}^* = \Delta_n^\mathrm{T}(\tilde{\alpha}_{n0}, \beta_0)G^-\Delta_n(\tilde{\alpha}_{n0}, \beta_0)$ whose limit distribution would be the same as that of T_{n2} (cf. [11]). However, in the present case G is an *unobserved* random matrix. In the following application we briefly consider T_{n1}^* where G^- is replaced by its consistent estimator. In the latter case results similar to the ergodic case are obtained (insofar as the asymptotic equivalence of T^*_{n1} and T_{n2} is concerned); however, the limiting non-null distributions are still non-standard (i.e. mixtures of non-central chi-squares rather than non-central chi-squares).

4. An application

Let the observable r.vs. $\{X_i, 1 \leqslant i \leqslant N\}$ satisfy

$$X_i = \alpha C_i + Y_i, \tag{4.1}$$

where the errors $\{Y_i\}$ satisfy

$$Y_i = \beta_1 Y_{i-1} + \beta_2 Y_{i-2} + Z_i, \quad 1 \leqslant i \leqslant n \tag{4.2}$$

with $\{Z_i, 1 \leqslant i \leqslant n\}$ i.i.d. N(0, 1) r.vs. and $Z_i = 0 = C_i$ for $i \leqslant 0$. We assume that the roots m_1 and m_2 of the equation

$$m^2 - \beta_2 m - \beta_1 = 0 \tag{4.3}$$

are distinct and there exists a unique root, say m_1 and call it ρ, such that

$$|\rho| > \max(|m_2|, 1). \tag{4.4}$$

The process $\{X_i\}$ satisfying (4.1) through (4.4) is known as the regression process with explosive autocorrelated Gaussian errors. The numbers $\{C_i\}$ are known and the parameter vector is $\theta^{\mathrm{T}} = (\alpha, \beta_1, \beta_2)$.

Before discussing the inferential problems about θ, we first study the likelihood function and obtain the limiting matrix G and the limiting distribution of various underlying entities. To that effect observe that the log likelihood function in this case is

$$\log p_n(X_n, \theta) = \text{const} - \tfrac{1}{2} \sum_{i=1}^{n} \{(X_i - \alpha C_i) - \beta_1(X_{i-1} - \alpha C_{i-1}) - \beta_2(X_{i-2} - \alpha C_{i-2})\}^2.$$

Therefore when θ is the true parameter vector

$$\begin{aligned} A_{n1}(\theta) &= \frac{\partial}{\partial \alpha} \log P_n \\ &= \sum_{i=1}^{n} [(X_i - \alpha C_i) - \beta_1(X_{i-1} - \alpha C_{i-1}) - \beta_2(X_{i-2} - \alpha C_{i-2})] \\ &\quad \times (C_i - \beta_1 C_{i-1} - \beta_2 C_{i-2}) \\ &= \sum_{i=1}^{n} (C_i - \beta_1 C_{i-1} - \beta_2 C_{i-2}) Z_i, \end{aligned}$$

$$A_{nj}(\theta)=\frac{\partial}{\partial\beta_j}\log p_n$$
$$=\sum_{i=1}^{n}(X_{i-j}-\alpha C_{i-j})$$
$$\times[(X_i-\alpha C_i)-\beta_1(X_{i-1}-\alpha C_{i-1})-\beta_2(X_{i-2}-\alpha C_{i-2})]$$
$$=\sum_{i=j}^{n}Y_{i-j+1}Z_i,\quad j=2,3.$$

The $B_n(\theta)$ matrix of the assumption (B.1) will have the entries $((b_{nij}))$ given by

$$b_{n11}(\theta)=\sum_{i=1}^{n}(C_i-\beta_1C_{i-1}-\beta_2C_{i-2})^2,$$

$$b_{n1j}(\theta)=\sum_{i=j}^{n}C_{i-j+1}Z_i+\sum_{i=j}^{n}(C_i-\beta_iC_{i-1}-\beta_2C_{i-2})Y_{i-j+1},\quad j=2,3,$$

$$b_{njj}(\theta)=\sum_{i=j}^{n}Y_{i-j+1}^2,\quad j=2,3$$

and

$$b_{n23}(\theta)=\sum_{i=3}^{n}Y_{i-1}Y_{i-2}.$$

Now let us consider some preliminary facts. In what follows all calculations are carried out under the assumption that (α,β_1,β_2) are the true parameters and the above model holds.

Let λ_1 and λ_2 be real numbers such that

$$\lambda_1+\lambda_2=1,\qquad \lambda_1\rho^{-1}+\lambda_2m_2^{-1}=0. \tag{4.5}$$

Then one can write [19]

$$\rho^{-j}Y_j=\rho^{-j}\sum_{r=1}^{j}(\lambda_1\rho^{j-r}+\lambda_2m_2^{j-r})Z_r,\quad j\geq 1$$
$$=\lambda_1W_j+\lambda_2R_j\quad\text{(say)}, \tag{4.6}$$
$$W_j=\sum_{r=1}^{j}\rho^{-r}Z_r,\qquad R_j=\rho^{-j}\sum_{r=1}^{j}m_2^{j-r}Z_r,\qquad j\geq 1.$$

Observe that

$$\sum_{j=1}^{\infty}\mathbf{E}|R_j|^2=\sum_{j=1}^{\infty}\rho^{-2j}\sum_{k=0}^{j-1}(m_2)^{2k}<\sum_{j=1}^{\infty}j(m_2/\rho)^{2j}<\infty \tag{4.7}$$

because $|m_2/\rho|<1$.

Therefore the Markov inequality and the Borel–Cantelli lemma imply

$$R_j\to 0\text{ a.s.}\quad\text{and}\quad R_j\to 0\text{ in }L_2,\quad\text{as }j\to\infty. \tag{4.8}$$

Next note that $\{W_j, j \geq 1\}$ form an L_2-martingale and $|\rho| > 1$ implies

$$\sup_{j\geq 1} \mathbf{E} W_j^2 = \sup_{j\geq 1} \sum_{r=1}^{j} \rho^{-2r} = (\rho^2 - 1)^{-1} < \infty. \tag{4.9}$$

Hence by the Martingale Convergence Theorem $\exists a$ r.v. W such that

$$W_j \to W \text{ a.s.} \quad \text{and} \quad W_j \to W \text{ in } L_2, \quad \text{as } j \to \infty. \tag{4.10}$$

The assumption of normality yields that for every real t

$$\mathbf{E}\, \mathrm{e}^{itW_j} = \mathrm{e}^{-(t^2/2)\sum_{r=1}^{j} \rho^{-2r}} \to \mathrm{e}^{-(t^2/2)(\rho^2-1)^{-1}} \qquad \text{as } j \to \infty. \tag{4.11}$$

Hence W of (4.10) is a

$$\mathrm{N}(0, (\rho^2 - 1)^{-1} \text{ r.v.} \tag{4.12}$$

Combining (4.10) and (4.8) with (4.6) one gets

$$\rho^{-j} Y_j \to \lambda_1 W \text{ a.s. and in } L_2, \quad \text{as } j \to \infty. \tag{4.13}$$

Next define

$$k_n^2(\theta) = \sum_{i=1}^{n} (C_i - \beta_1 C_{i-1} - \beta_2 C_{i-2})^2 \quad (C_j = 0, j \leq 0)$$

$$\sigma_n^2 = \sum_{i=1}^{n} C_i^2$$

and

$$I_n(\theta) = \begin{pmatrix} k_n^2(\theta) & 0 & 0 \\ 0 & s_n^2(\theta) & 0 \\ 0 & 0 & s_n^2(\theta) \end{pmatrix}, \quad \text{where } s_n = \frac{|\lambda_1| \rho^n}{(\rho^2 - 1)}.$$

Lemma. *Suppose the model* (4.1)–(4.4) *holds. Suppose, in addition*

$$\max_{1\leq i\leq n} C_i^2 \sigma_n^{-2} \to 0 \quad \text{as } n \to \infty \tag{4.14}$$

and

$$\sigma_n^{-2} \sum_{i=1}^{n} C_i C_{i-1} \to b \quad \text{as } n \to \infty, (b < \infty). \tag{4.15}$$

Then

(a) $\sigma_n^{-2} k_n^2(\theta) \to 1 - b^2 + (b - \beta_1 - \beta_2)^2 > 0$,

(b) $I_n^{-1/2}(\theta) B_n(\theta) I_n^{-1/2}(\theta) \xrightarrow{\text{a.s.}} G(\theta)$ *as* $n \to \infty$, *where*

$$G(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & V^2 & \rho^{-1} V^2 \\ 0 & \rho^{-1} V^2 & \rho^{-2} V^2 \end{pmatrix}, \qquad V = (\rho^2 - 1)^{1/2} W,$$

(c) $\{k_n^{-1}(\theta)A_{n1}(\theta), \quad s_n^{-1}(\theta)A_{n2}(\theta), \quad s_n^{-1}(\theta)A_{n3}(\theta)\} \xrightarrow{d} G^{1/2}Z$, *where* $Z^{\mathrm{T}} = (Z_1, Z_2, Z_3) \sim N_3(0, I_{3\times 3})$ *and is independent of G and where G is the same as in* (b).

Proof. (a) follows from (4.14), (4.15), the expansion of k_n^2 and the Cauchy–Schwarz inequality.

(b) A variation of this result with a.s. replaced by "in probability" has been proved by Rao [19]. His proof is direct but somewhat lengthy. A careful evaluation of his proof together with the Borel–Cantelli lemma gives the a.s. result. Or alternatively we can use (4.13), the Borel–Cantelli and the Toeplitz lemmas directly to conclude (b) above. We demonstrate this just for two terms. Suppress θ and let $D_n = I_n^{-1/2}B_nI_n^{-1/2} = ((d_{nij}))_{3\times 3}$. Clearly

$$d_{n11} = 1, \quad n \geqslant 1.$$

$$d_{nij} = k_n^{-1}s_n^{-1}b_{nij} = k_n^{-1}\sigma_n\sigma_n^{-1}s_n^{-1}b_{nij}, \quad i, j = 2, 3.$$

Now

$$\begin{aligned}\sigma_n^{-1}s_n^{-1}b_{n12} &= \sigma_n^{-1}s_n^{-1}\sum_{i=2}^{n} C_{i-1}Z_i + \sum_{i=2}^{n}(C_i - \beta_1C_{i-1} - \beta_2C_{i-2})Y_{i-1} \\ &= \mathbf{U}_n + \mathbf{U}_n' \quad \text{(say)}.\end{aligned} \tag{4.16}$$

Observe that

$$\sum_{n=1}^{\infty} \mathbf{E}\mathbf{U}_n^2 = \sum_{n=1}^{\infty} \sigma_n^{-2}s_n^{-2}\sum_{i=1}^{n} C_i^2 < (\rho^2-1)^2\lambda_1^{-2}\sum_{n=1}^{\infty}\rho^{-2n} < \infty. \quad (\because |\rho| > 1).$$

The Borel–Cantelli lemma implies

$$\mathbf{U}_n \xrightarrow{\text{a.s.}} 0. \tag{4.17}$$

Next write

$$\mathbf{U}_n' = \sum_{i=2}^{n} a_{ni}(Y_{i-1}/\rho^{i-1}) = \sum_{i=2}^{n} a_{ni}[(Y_{i-1}/\rho^{i-1}) - \lambda_1 W] + \sum_{i=2}^{n} a_{ni}\lambda_1 W,$$

where for $2 \leqslant i \leqslant n$

$$a_{ni} = \sigma_n^{-1}s_n^{-1}\rho^{i-1}(C_i - \beta_1C_{i-1} - \beta_2C_{i-2}) = e\cdot\sigma_n^{-1}\rho^{-n+i-1}(C_i - \beta_1C_{i-1} - \beta_2C_{i-2})$$

with $e = |\lambda_1|^{-1}(\rho^2 - 1)$.

Using (4.14) and $|\rho| > 1$, one gets that $|a_{ni}| \to 0$ for every $2 \leqslant i \leqslant n$ as $n \to \infty$. Using the Cauchy–Schwarz inequality one has

$$\sum_{2}^{n} |a_{ni}| \leqslant e\{2(1 + \beta_1^2 + \beta_2^2)\}^{1/2}\rho^{-n}\left\{\sum_{i=2}^{n}\rho^{2(i-1)}\right\}^{1/2} < \infty, \quad n > 1.$$

Moreover

$$\sum_{i=2}^{n} |a_{ni}| \leqslant e\sigma_n^{-1}\max|C_i|(1 + |\beta_1| + |\beta_2|)\rho^{-n}\sum_{i=0}^{n}\rho^i \to 0 \quad \text{by (4.14)}.$$

Hence using (4.13) and the Toeplitz lemma one has

$$\mathbf{U}'_n \xrightarrow{\text{a.s.}} 0. \tag{4.18}$$

Therefore in view of the part (a) of the lemma, (4.18) and (4.17) one has

$$d_{n12} \xrightarrow{\text{a.s.}} 0.$$

Next consider

$$d_{n22} = s_n^{-2} b_{n22} = e^2 \rho^{-2n} \sum_{i=2}^{n} Y_{i-1}^2 = e^2 \sum_{i=2}^{n} \rho^{-2(n-1+1)} (Y_{i-1}/\rho^{i-1})^2$$

$$\xrightarrow{\text{a.s.}} e^2 \cdot \lambda_1^2 (\rho^2-1)^{-1} W^2 = (\rho^2-1) W^2 = V^2$$

by (4.13) and the Toeplitz lemma. One completes the rest of the proof by using similar arguments.

(c) Consider

$$s_n^{-1} A_{nj} = e\rho^{-n} \sum_{i=j}^{n} Y_{i-j+1} Z_i, \quad j=2,3$$

$$= e \sum_{n=j}^{n} \rho^{-(n-i+j-1)} Z_i (Y_{i-j+1}/\rho^{i-j+1}) \quad (e = |\lambda_1|^{-1}(\rho^2-1)).$$

Let

$$\xi_{nj} = \sum_{i=j}^{n} \rho^{-(n-i+1)} Z_i, \quad j=2,3.$$

Observe that with W defined in (4.13),

$$\mathbf{E}|s_n^{-1} A_{nj} - e\rho^{-(j-2)} \xi_{nj} \lambda_1 W| \geq$$

$$\leq e \sum_{i=j}^{n} |\rho|^{-(n-i+j-1)} \{\mathbf{E}((Y_{i-j+1}/\rho^{i-j+1}) - \lambda_1 W)^2\}^{1/2} \to 0 \tag{4.19}$$

by the Toeplitz lemma and the L_2 part of (4.13).

We next show that the vector $A_n^{\mathrm{T}} = (k_n^{-1} A_{n1}, \xi_{n2}, \xi_{n3})$ is asymptotically independent of W_n where W_n is defined in (4.5). We also show that the asymptotic distribution of A_n^{T} is $\mathrm{N}_3(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma^2 & \sigma^2 \\ 0 & \sigma^2 & \sigma^2 \end{pmatrix} \quad \text{with } \sigma^2 = (\rho^2-1)^{-1}. \tag{4.20}$$

To this effect let $a^{\mathrm{T}} = (a_1, a_2, a_3, a_4)$ be any real vector of dimension 4. Let $d_i = (C_i - \beta_1 C_{i-1} - \beta_2 C_{i-2}) k_n^{-1}$, $1 \leq i \leq n$. Then

$$L_n = a^{\mathrm{T}}(A_n, W_n)$$

$$= a_1 \sum_{i-1}^{n} d_i Z_i + a_2 \sum_{i=2}^{n} \rho^{-(n-i+1)} Z_i + a_3 \sum_{i=3}^{n} \rho^{-(n-i+1)} Z_i + a_4 \sum_{i=1}^{n} \rho^{-i} Z_i$$

$$= (a_1 d_1 + a_4 \rho^{-1}) Z_1 + (a_1 d_2 + a_2 \rho^{-(n-1)} + a_4 \rho^{-2}) Z^2$$

$$+ \sum_{i=3}^{n} \{a_1 d_1 + (a_2 + a_3) \rho^{-(n-i+1)} + a_4 \rho^{-i}\} Z_i.$$

Now $\{Z_i\}$ i.i.d. N(0, 1) imply

$$\mathbf{E}\, e^{iL_n} = \exp\{-\tfrac{1}{2}[(a_1d_1 + a_4\rho^{-1})^2 + (a_1d_2 + a_2\rho^{-(n-1)} + a_4\rho^{-2})^2 + \sum_{j=3}^{n} \{a_1d_j + (a_2+a_3)\rho^{-(n-j+1)} + a_4\rho^{-j}\}^2]\}.$$

Part (a) of the lemma and assumptions (4.14), (4.15) imply

$$\max\left\{ \sum_{i=1}^{n} |d_i\rho^{-(n-i+1)}|,\quad \sum_{i=1}^{n} |d_i\rho^{-i}|,\quad \sum_{i-1}^{n} |\rho^{-(n-i+1)}\rho^{-i}| \right\} \to 0 \quad \text{as } n\to\infty.$$

Hence one readily sees that

$$\mathbf{E}\, e^{iL_n} \to \exp\{-\tfrac{1}{2}[a_1^2 + (a_2+a_3)^2(\rho^2-1)^{-1} + a_4^2(\rho^2-1)^{-1}]\} \tag{4.21}$$

which establishes the claimed result.

Finally in view of (4.21), (4.19) and the fact that $W_n \to W$ a.s., part (c) of the lemma is completely proved.

Remarks. (1) From the above calculations it is clear that the lemma holds even when the errors in (4.1) are obeying a pth order autoregressive equation where one root, say ρ, of the equation

$$m^p - \beta_p m^{p-1} - \cdots - \beta_1 = 0 \tag{4.22}$$

is larger than 1 and all others in absolute value.

Here the analogue of G would be

$$G = \begin{pmatrix} 1 & 0 \\ 0 & V^2\Gamma \end{pmatrix}_{p+1\times p+1}, \qquad \Gamma = \begin{pmatrix} 1 & \rho^{-1} & \cdots & \rho^{-(p-1)} \\ \rho^{-1} & \rho^{-2} & \cdots & \rho^{-(p-2)} \\ \rho^{-(p-1)} & & & \rho^{-2(p-1)} \end{pmatrix}_{p\times p} \tag{4.23}$$

and V^2 is a r.v.

(2) If one would be interested in looking at the explosive autoregressive errors, where some (more than 1) roots of (4.22) are larger than 1 and some are smaller then one can also get an analogue of the lemma using Stigum's [21] results. Perhaps it should be noted that Stigum also does not use the Martigale Convergence Theorem and Toeplotz's lemma.

(3) Finally we should point out, as is evident from either the above proof or otherwise, that part (b) of the lemma is also true when one has non-Gaussian errors but finite second moments. However, part (c) of the lemma need not hold for non-Gaussian errors.

Now turning to the testing problem about the parameters $\theta^{\mathrm{T}} = (\alpha, \beta_1, \beta_2)$ we first observe that in view of the definition of $\{A_{nj}, j = 1, 2, 3\}$ and part (c) of the lemma the likelihood function of our model (4.1)–(4.4) satisfies the assumptions (B.1)–(B.3) with $S_n^{\mathrm{T}} = (A_{n1}, A_{n2}, A_{n3})$, I_n defined prior to the above lemma and G as in (b).

Now suppose we wish to test H_0: $\beta_j = \beta_{j0}$, $j = 1, 2$ treating α as a nuisance parameter. Here $s = 1$, $k = 3$, $k - s = 2$. By part (b) of the lemma

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & V^2 & \rho^{-1}V^2 \\ 0 & \rho^{-1}V^2 & \rho^{-2}V^2 \end{pmatrix}, \qquad V \text{ a } \mathrm{N}_1(0, 1) \text{ r.v.}$$

The rank l of G is 2 and a generalized inverse is

$$G^- = \begin{pmatrix} 1 & 0 & 0 \\ 0 & V^{-2} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that

$$C = V^2\begin{pmatrix} 1 & \rho^{-1} \\ \rho^{-1} & \rho^{-2} \end{pmatrix}, \qquad C_* = V^2.$$

We use $\tilde{\alpha}_{n0}$ as the MLE of α under H and $\tilde{\alpha}_n, \tilde{\beta}_{1n}, \tilde{\beta}_{2n}$ as the MLE of α, β_1, β_2 under no restrictions. These estimators are readily seen to satisfy (3.1) through (3.3) with $S_{n1} = A_{n1}$, $S_{n2}^{\mathrm{T}} = (A_{n2}, A_{n3})$. With these choices of $\tilde{\alpha}_{n0}$ and $\tilde{\beta}_{nj}$, $j = 2, 3$ the asymptotic distributions of T_{n1} and T_{n2} are given by (3.6), (3.7) and (3.12), (3.13) respectively with the above C and C_*.

Note that the limit distribution of T_{n2} is χ_1^2 under H_0 and $\chi_1^2(\delta)$–a noncentral χ_1^2 with $\delta = (h_2^* V)^2$ (i.e. a mixture of noncentral χ_1^2 r.vs. mixed with another χ_1^2 r.v.) under the sequence of alternatives $K_{(n)} = \beta_j = \beta_{j0} + s_n^{-1}(\beta_{j0})h_j^*$, $j = 2, 3$. The limit distribution of T_{n1}, however, does not have a simple form.

For the first order autoregressive case (i.e. $p = 1$ in Remark (1)) we get

$$G = \begin{pmatrix} 1 & 0 \\ 0 & V^2 \end{pmatrix},$$

where V^2 now is a χ_1^2 random variable, and G is non-singular. For this special case (i.e. $p = 1$) note that, for testing $\beta_1 = \beta_0$, we have $C = V^2 = C_*$. Thus, $T_{n1} \xrightarrow{\mathrm{d}} U^2 V^2$ under the null hypothesis, where U is a N(0, 1) variate independent of V^2, i.e. the asymptotic null distribution of T_{n1} is that of a prodict of two independent χ_1^2 variates. The asymptotic distribution of T_{n1} under the sequence of alternatives does not have a simple form. It is interesting that the limiting null and non-null distributions of T_{n2} for the case $p = 1$ remain the same as in the case $p = 2$ with V^2 now having a χ_1^2 distribution. In the special case $p = 1$ we can consider a *modified* score statistic

$$T_{n1}^* = \Delta_n^{\mathrm{T}}(\tilde{\alpha}_{n0}, \beta_0)\begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_n^{-2} \end{pmatrix}\Delta_n(\tilde{\alpha}_{n0}, \beta_0),$$

where $\hat{V}_n^2 = \sum_1^n (X_j - \tilde{\alpha}_{n0}C_j)^2/\beta_0^{2n}$ is a consistent estimator of V^2. Here one may take

$$\tilde{\alpha}_{n0} = \left[\sum_{i=1}^n (X_i - \beta_0 X_{i-1})(C_i - \beta_0 C_{i-1})\right] \times \left[\sum_{i=1}^n (C_i - \beta_0 C_{i-1})^2\right]^{-1}.$$

This is the MLE of α under $H_0: \beta_1 = \beta_0$ and it has all desirable properties. It can be shown (we omit details) that T^*_{n1} has the same asymptotic distributions as T_{n2} both under the null and the alternative hypotheses.

In general, whenever a consistent estimator of $G^-(\alpha, \beta_0)$ can be found, say $\hat{G}^-_n(\tilde{\alpha}_{n0}, \beta_0)$, one could consider the modified score statistic

$$T^*_{n1} = \Delta^{\mathrm{T}}_n(\tilde{\alpha}_{n0}, \beta_0)\hat{G}^-_n(\tilde{\alpha}_{n0}, \beta_0)\Delta_n(\tilde{\alpha}_{n0}, \beta_0)$$

which will have the same limiting distributions as T_{n2}. In the present example with $p = 2$, we have

$$\hat{G}^-_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \hat{V}^{-2}_n & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Suppose now one has the regression model with errors following a pth order explosive autoregressive Gaussian process as described in Remark (1). Then the G matrix is given by (4.22) and its rank is still 2. For testing $H: \beta_j = \beta_{j0}$, $j = 1, \ldots p$ treating α as a nuisance parameter the limit distribution of T_{n1} and T_{n2} remain the same as for the case $p = 2$. This is so because of the high singularity of G. Part of the reason is that the scores A_{nj+1}'s corresponding to β_j's, $j = 1, \ldots, p$, are asymptotically linearly related among themselves as is evident from (4.19), which continues to hold for the present model. Furthermore β's enter G only through the single root ρ (the dominating root). These observations seem to lead to the conclusion that we should either treat all β's as nuisance parameters or all β's should be under test in order to avoid degeneracy in the limit distribution of T_{n1} and T_{n2}.

If we consider the dual problem of testing $\alpha = \alpha_0$ treating the β's as nuisance parameters in the present example we find that T_{n1} and T_{n2} for this problem are

$$T_{n1} = \Delta^{\mathrm{T}}_n(\alpha_0, \tilde{\beta}_{n0})\Delta_n(\alpha_0, \tilde{\beta}_{n0})$$

and

$$T_{n2} = -2 \log \frac{p_n(X^n; \alpha_0, \tilde{\beta}_{n0})}{p_n(X^n; \tilde{\alpha}_n, \tilde{\beta}_n)}.$$

They have asymptotically the same limiting distributions both under the null and non-null hypotheses. This is due to the special form of the G matrix which leads the problem of testing $\alpha = \alpha_0$ to a problem in an ergodic type model and it is not surprising that the ergodic type results obtain for this case.

References

[1] T.W. Anderson, On asymptotic distributions of estimates of parameters of stochastic difference equations, Ann. Math. Statist. 30 (1959) 676–687.

[2] I.V. Basawa, Asymptotic curvature for dependent observations. Bull. Internat. Statist. Inst. 41st session, to appear.

[3] I.V. Scott and D.J. Scott, Efficient tests for branching processes, Biometrika 63 (1976) 531–536.
[4] I.V. Basawa and D.J. Scott, Efficient tests for stochastic processes, Sankhyā Ser. A 39 (1977) 21–31.
[5] I.V. Basawa and D.J. Scott, Efficient estimation of stochastic processes, J. Indian Statist. Assoc. (1978) to appear.
[6] I.V. Basawa, P.D. Feigin and C.C. Heyde, Asymptotic properties of maximum likelihood estimators for stochastic processes, Sankhya Ser. A 38 (1976) 259–270.
[7] B.R. Bhat, On the method of maximum likelihood for dependent observations, J. Roy. Statist. Soc. Ser. B 36 (1974) 48–53.
[8] P. Billingsley, Statistical Inference for Markov Processes (Univ. Chicago Press, Chicago, 1961).
[9] R.B. Davies, Asymptotic inference when the amount of information is random, Tech. Report, Wellington Univ. (1978).
[10] J.P. Dion, Estimation of the mean and the initial probabilities of a branching process, J. Appl. Probability 11 (1974) 687–694.
[11] K.O. Dzhaparidge, Tests of composite hypotheses for random variables and stochastic processes, Theor. Probability Appl. XXII (1977) 104–118.
[12] P.D. Feigin, Maximum likelihood estimation for stochastic processes—A Martingale approach, Ph.D. Thesis, Austr. National Univ. (1975).
[13] P.D. Feigin, The efficiency criteria problem for stochastic processes, Stochastic Processes Appl. 6 (1978) 115–127.
[14] C.C. Heyde and P.D. Feigin, On efficiency and exponential families in stochastic process estimation, in: G.P. Patil, S. Kotz and J.K. Ord, Eds., Statistical Distributions in Scientific Work, Vol. 1 (Reidel, Dordrecht, 1975) 227–240.
[15] C.C. Heyde, Remarks on efficiency in estimation for branching processes, Biometrika 62 (1975) 49–55.
[16] C.C. Heyde, On an optimal property of the maximum likelihood estimator of a parameter from a stochastic process, Stochastic Processes Appl. 8 (1978) 1–9.
[17] L. LeCam, Locally asymptotically normal families of distributions, Univ. Calif. Publ. Statist. 3 (1960) 37–98.
[18] C.R. Rao, Linear Statistical Inference and its Applications (Wiley, New York, 1973).
[19] M.M. Rao, Consistency and limit distribution of estimators of parameters in explosive stochastic difference equations, Ann. Math. Statist. 32 (1961) 195–218.
[20] G.G. Roussas, Contiguity of probability mneasures (Cambridge Univ. Press, Cambridge, 1972).
[21] B.P. Stigum, Asymptotic properties of dynamic stochastic parameter estimates (III), J. Multivariate Anal. 4 (1974) 351–381.
[22] T.J. Sweeting, On efficient tests for branching processes, Biometrika, 65 (1978) 123–127.
[23] A. Wald, Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process, Ann. Math. Statist. 20 (1948) 595–601.
[24] L. Weiss, The asymptotic distribution of the likelihood ratio in some nonstandard cases, J. Amer. Statist. Assoc. 70 (1975) 204–208.
[25] L. Weiss and J. Wolfowitz, Maximum probability estimators and related topics, Lecture Notes in Mathematics 424 (Springer, Berlin, 1974).