The Annals of Statistics
1982, Vol. 10, No. 4, 1163–1171

THE EFFECT OF DEPENDENCE ON CHI SQUARED TESTS OF FIT[1]

BY DAVID S. MOORE

Purdue University

Tests of fit that assume i.i.d. observations may be affected by dependence among the observations. This effect is studied for the Pearson chi squared test when the data form a stationary stochastic process. The general form of asymptotic distribution theory under the null hypothesis is outlined and examples are given. For testing fit to a specified normal law, it is shown that when observations come from a quite general class of Gaussian stationary processes, positive correlation among the observations is confounded with lack of normality.

1. Introduction. In testing the fit of a sequence of observations to a distribution or to a parametric family of distributions, it is commonly assumed that the observations are independent and identically distributed (i.i.d.). In practice, however, the observations may have substantial dependence, as when the data were collected as a time series. Suppose, then, that $X_1, \cdots, X_n$ are observations on a stationary stochastic process (SSP). A test of fit for the common distribution of the X_t is applied that is designed for the i.i.d. case. What is the effect of dependence on such a test when the null hypothesis concerning the univariate marginal distribution is correct?

Despite the statistical interest of this question, there is little literature on it. Gasser (1975) conducted a small simulation study of the effect of dependence on the validity and power of a chi squared test for normality. For the SSP's he studied, which were primarily Gaussian autoregressive processes, the Pearson test using (asymptotic) i.i.d. critical points rejected normality too often. Moreover, the test had low power against non-Gaussian autoregressive processes relative to its power against i.i.d. alternatives. Both effects of dependence were more marked for SSP's with positive autocorrelations. The purpose of the present paper is to initiate the theoretical study of the effects of dependence on the Pearson chi squared test.

Section 2 outlines the basic large-sample theory, following the pattern laid down in the i.i.d. case by Moore and Spruill (1975), hereafter referred to as MS. Since this development follows MS closely, details are largely omitted. It is shown that the vector of standardized cell counts is generally asymptotically multivariate normally distributed, and the form of the asymptotic covariance matrix is obtained for several methods of estimating unknown parameters. The distribution of the Pearson statistic, and of other quadratic forms in the cell counts, can then be expressed in terms of the characteristic roots of this covariance matrix. Because interest centers on the effect of dependence on the validity of chi squared tests of fit, only large sample behavior under the hypothesized univariate distribution of the X_t is discussed. Behavior under contiguous alternatives can be obtained by combining the results of Section 2 and the contiguity results of Roussas (1979).

Section 3 examines in more detail the case of testing fit to a specified normal distribution. It is shown that for a general class of Gaussian SSP's with positive autocorrelations, the asymptotic distribution of the Pearson statistic is stochastically larger than the i.i.d. case chi squared asymptotic distribution. That is, positive correlation is confounded with lack of normality. Moreover, the correct asymptotic critical points can be arbitrarily large for sufficiently strong autocorrelation. The arguments of Section 3 make essential use of only

Received August 1980; revised May 1982.

[1] Research supported by the National Science Foundation under Grant MCS 79-05815.

AMS 1980 *subject classifications.* Primary 62G10; secondary 62M99.

Key words and phrases. Chi squared tests, stationary stochastic processes, goodness of fit.

one property of the normal laws, and the results therefore extend to other null hypotheses. This section also points out how knowledge of the SSP structure can sometimes be used to construct tests of fit that, unlike the Pearson test, have chi squared limiting null distributions. Gasser's simulations suggest that the results of Section 3 may extend to cases in which unknown parameters are estimated, such as testing fit to the normal family. This is among the open questions remaining within the framework of the general theory of Section 2.

2. General large sample theory. Identically distributed observations $X_1, \cdots, X_n$ are to be tested for fit to a parametric family of distribution functions $\{F(x, \theta) : \theta \text{ in } \Omega\}$, where Ω is an open set in Euclidean m-space R^m. Choose M cells $A_i = (a_{i-1}, a_i]$, $i = 1, \cdots, M$ with boundaries $-\infty = a_0 < a_1 < \cdots < a_{M-1} < a_M = \infty$. Let $\delta_i(x)$ be the indicator function of A_i, so that the ith cell frequency is $N_i = \sum_{t=1}^{n} \delta_i(X_t)$. The corresponding cell probability is $p_i(\theta) = F(a_i, \theta) - F(a_{i-1}, \theta)$. A basic condition for applicability of chi squared tests of fit is that under $F(x, \theta_0)$, $N_i/n \to p_i(\theta_0)$ in probability as $n \to \infty$. The mean-square ergodic theorem (Karlin and Taylor, 1975, page 476) asserts that this is true for a SSP $\{X_t\}$ if and only if

$$n^{-1} \sum_{t=0}^{n-1} \{P[X_1 \text{ in } A_i, X_{1+t} \text{ in } A_i] - p_i^2\} \to 0. \tag{2.1}$$

Mixing conditions on $\{X_t\}$ give sufficient conditions for (2.1). For example, if $\{X_t\}$ is a strongly mixing SSP with mixing coefficient $\alpha(k)$, then (2.1) holds if $\sum_{k=1}^{\infty} \alpha(k) < \infty$. *All processes considered will be assumed to satisfy* (2.1). The examples appearing below do satisfy this condition.

Let $V_n(\theta)$ be the M-vector of standardized cell frequencies, having ith component $\{N_i - np_i(\theta)\}/\{np_i(\theta)\}^{1/2}$. Except in the simple null hypothesis case $\Omega = \{\theta_0\}$, the unknown parameter θ is estimated by $\theta_n = \theta_n(X_1, \cdots, X_n)$. General chi squared statistics are then non-negative definite quadratic forms in the components of $V_n(\theta_n)$. In particular, the Pearson chi squared statistic is the sum of squares $V_n(\theta_n)'V_n(\theta_n)$. In general, $V_n(\theta_n)$ is asymptotically normally distributed,

$$\mathscr{L}\{V_n(\theta_n)\} \to N_M(0, C(\theta_0)) \quad \text{under} \quad F(x, \theta_0), \tag{2.2}$$

so that the Pearson statistic has as its limiting null distribution that of $\sum_1^M \lambda_i Z_i^2$, where λ_i are the characteristic roots of C and $Z_1, \cdots, Z_M$ are i.i.d. $N(0, 1)$ random variables. We now develop these results in more detail.

We are concerned with the behavior of $V_n(\theta_n)$ under $F(x, \theta_0)$ where θ_0 is an arbitrary point in Ω. For convenience we often omit the argument θ when $\theta = \theta_0$; for example, $p_i = p_i(\theta_0)$, and derivatives and expected values not otherwise specified are evaluated at $\theta = \theta_0$. The following assumptions are required

ASSUMPTIONS A1. Under $F(x, \theta_0)$, $\theta_n - \theta_0 = O_p(n^{-1/2})$.

A2. For all i, $p_i > 0$, $p_i(\theta)$ is continuously differentiable at θ_0, and $\sum_1^M p_i = 1$.

A3. $F(x, \theta_0)$ is continuous at each cell boundary a_i.

Note that A1 requires that θ_n, which may be an estimator of θ predicated on i.i.d. observations, remain consistent when $\{X_t\}$ is in fact a SSP. We shall see that this often holds (e.g. when θ_n is the i.i.d.-case maximum likelihood estimator of θ). The results below can be generalized to include the case in which θ_n converges to some θ other than θ_0. In this case the limiting normal law of $V_n(\theta_n)$ does not have zero mean.

Assume now that A1–A3 are met, and let B be the $M \times m$ matrix with (i, j)th component

$$b_{ij} = p_i^{-1/2}(\partial p_i/\partial \theta_j).$$

As in MS, the following result is immediate.

LEMMA 2.1. *If* A1, A2 *and* A3 *hold, then under* $F(x, \theta_0)$

$$V_n(\theta_n) = V_n - Bn^{1/2}(\theta_n - \theta_0) + o_p(1). \tag{2.3}$$

Most estimators of interest, computed under the i.i.d. assumption, have the aysmptotic form specified by the following assumption:

ASSUMPTION A4. Under $F(x, \theta_0)$

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{t=1}^{n} h(X_t, \theta_0) + o_p(1),$$

where h has zero mean and finite covariance under $F(x, \theta_0)$.

If $\Delta(x)$ is the M-vector with components $\{\delta_\iota(x) - p_\iota\}/p_\iota^{1/2}$ and if we abbreviate $\Delta(X_t) = \Delta_t$ and $h(X_t, \theta_0) = h_t$, then from (2.3) and (A4)

$$V_n(\theta_n) = n^{1/2} \sum_{t=1}^{n} (\Delta_t - Bh_t) + o_p(1), \tag{2.4}$$

and asymptotic normality of $V_n(\theta_n)$ will follow from an appropriate central limit theorem for dependent variables. When θ_n is computed using the SSP structure of the data, a central limit theorem for triangular arrays will usually be required. Because of the great variety of applicable central limit theorems, we give only a single paradigm result for Gaussian processes, chosen for its usefulness in Section 3. We stress that the conclusion of Theorem 2.1 below, including the form of the limiting covariance matrix, holds whenever a central limit theorem applies to (2.4).

The central limit theorem leading to our paradigm result is given by Gastwirth and Rubin (1975). Combining the theorem of Sun (1963) with an argument on L_2 approximation of functions f with finite variance by polynomials, they show (page 816) that *if* $\{X_t\}$ *is any Gaussian* SSP *with* $\sum |\rho_k| < \infty$, *then for any function f such that* $f(X_t)$ *has finite variance,*

$$\mathscr{L}\{n^{-1/2} \sum_{t=1}^{n} [f(X_t) - EF(X_t)]\} \rightarrow N(0, \sigma^2), \qquad \sigma^2 = \lim n^{-1} \operatorname{Var}\{\sum_{t=1}^{n} f(X_t)\} < \infty. \tag{2.5}$$

Here ρ_k is the correlation between X_t and X_{t+k}. The condition $\sum |\rho_k| < \infty$ implies Gastwirth and Rubin's "Δ-mixing," and hence strong mixing, if $\{X_t\}$ is a Gaussian Markov process, but in general implies no mixing condition.

The statement of Theorem 2.1 requires some additional notation. Denote by I the $M \times M$ identity matrix, and by Q_{ts} the $M \times M$ matrix with (i, j)th component

$$\{P[X_t \text{ in } A_\iota, X_s \text{ in } A_j] - p_\iota p_j\}/(p_\iota p_j)^{1/2} \tag{2.6}$$

and let

$$Q = \lim n^{-1} \sum_{t,s=1,t\neq s}^{n} Q_{ts}$$

and $q = (p_1^{1/2}, \cdots, p_M^{1/2})'$. Clearly $Q = 0$ in the i.i.d. case. The following result is now immediate.

THEOREM 2.1. *Suppose that* $\{X_t\}$ *is a Guassian* SSP *with* $\sum |\rho_k| < \infty$ *and that Assumption* A4 *holds. Then* (2.2) *holds with*

$$C = I - qq' + Q + \lim n^{-1} \sum_{t,s=1}^{n} \{BE(h_t h_s')B' + BE(h_t\Delta_s') + E(\Delta_t h_s')B'\}.$$

REMARK. It is often desirable in practice to utilize data-dependent cells in a chi squared test of fit. If cells $A_{\iota n} = (a_{\iota-1,n}, a_{\iota n}]$ are employed, where $a_{\iota n} = a_{\iota n}(X_1, \cdots, X_n)$ and $a_{\iota n} \rightarrow a_\iota(P)$ under $F(x, \theta_0)$, then the argument in MS shows that the limiting law of $V_n(\theta_n)$ is the same as if the limiting fixed cells were used whenever the empiric c.d.f. process $n^{1/2}\{F_n(x) - F(x, \theta_0)\}$ converges weakly to a process with a.s. continuous paths. Such weak convergence results appear in, e.g. Gastwirth and Rubin (1975) and Withers (1975). In particular Gastwirth and Rubin obtain the required convergence for any Gaussian SSP with $\sum |\rho_k| < \infty$. *Thus Theorem* 2.1 *continues to hold when converging random cells are employed.*

We now record the form of the limiting covariance matrix C in several cases of general interest. The matrices appear intractible, but the example of the first order Gaussian autoregressive process that concludes this section shows that considerable simplification can occur. All of these results presuppose that a suitable expansion of θ_n and central limit theorem have been applied in (2.3). The expansions are discussed in each case.

CASE 1. *No estimation.* When testing fit to a specified distribution $F(x, \theta_0)$, then

$$C = I - qq' + Q.$$

In this case results of form (2.4) for uniformly bounded f are adequate to conclude (2.2). For example, (2.2) holds with C as above for a wide class of mixing processes by Theorem 2.1 of Gastwirth and Rubin (1975) and Corollary 1 of Withers (1975).

CASE 2. *Minimum chi squared estimation.* Suppose that a statistician, believing $\{X_t\}$ to be i.i.d., employs the minimum chi squared estimator $\bar{\theta}_n$ and the classical Pearson-Fisher statistic $V_n(\bar{\theta}_n)'V_n(\bar{\theta}_n)$. The condition $N_i/n \to p_i(\theta_0)$ in probability, assured by (2.1), is sufficient for consistency of $\bar{\theta}_n$ and for the usual expansion

$$n^{1/2}(\bar{\theta}_n - \theta_0) = (B'B)^{-1}B'V_n + o_p(1). \tag{2.7}$$

See Lemma 1 and Theorem 3 of Moore (1978) for an exact statement and proof in the i.i.d. case, which remain unchanged for SSP's satisfying (2.1). Though the *form* of (2.7) and hence of (2.3) is the same as in the i.i.d. case, the *distributions* differ due to dependence among the $\{X_t\}$. The result is now

$$C = I - qq' - P_B + (I - P_B)\, Q\, (I - P_B),$$

where $P^B = B(B'B)^{-1}\, B'$ is the orthogonal projection onto the range of B. This expression rarely simplifies.

CASE 3. i.i.d. *maximum likelihood estimation.* Suppose that a statistician, believing $\{X_t\}$ to be i.i.d., employs the MLE $\hat{\theta}_n$ of θ in $F(x, \theta)$ from i.i.d. $X_1, \cdots, X_n$. Then $V_n(\hat{\theta}_n)'V_n(\hat{\theta}_n)$ is the Chernoff-Lehmann (1954) statistic. It is again true that $\hat{\theta}_n$ remains consistent and has its usual expansion

$$n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2}J^{-1} \textstyle\sum_{t=1}^{n} \dfrac{\partial \log f(X_t \mid \theta)}{\partial \theta} + o_p(1)$$

for quite general SSP's. Here J is the Fisher information matrix of $F(x, \theta_0)$ and $\partial \log f/\partial\theta$ is the m-vector of partial derivatives evaluated at $\theta = \theta_0$. Such results are given in detail in work of B. Ranneby, described by Basawa and Prakasa Rao (1980). In this case,

$$C = I - qq' + Q - BJ^{-1}B' + \lim n^{-1} \textstyle\sum_{t,s=1,t\neq s}^{n} \{BJ^{-1}J_{ts}J^{-1}B' - BJ^{-1}J_{ts}^{*} - J_{ts}^{*\prime}J^{-1}B'\},$$

where

$$J_{ts} = E\left\{\frac{\partial \log f(X_t \mid \theta)}{\partial \theta}\, \frac{\partial \log f(X_s \mid \theta)'}{\partial \theta}\right\} (m \times m)$$

and

$$J_{ts}^{*} = E\left\{\frac{\partial \log f(X_t \mid \theta)}{\partial \theta} \Delta(X_s)'\right\} \ (m \times M).$$

In the i.i.d. case, $Q = J_{ts} = J_{ts}^{*} = 0$. The example below demonstrates that in Case 3, unlike Case 2, the matrix C can simplify considerably.

CASE 4. *SSP maximum likelihood estimation.* If the dependence structure of $\{X_t\}$ is known, the statistician may employ the correct MLE $\tilde{\theta}_n$ for θ in $F(x, \theta)$ based on $X_1, \cdots, X_n$. In regular cases (see Roussas, 1979, and references given there) $\tilde{\theta}_n$ has the following properties. Let f_t be the conditional density function of X_t given $X_1, \cdots, X_{t-1}$, evaluated at

$X_1, \cdots, X_t$, and set

$$\Gamma_n = n^{-1} \sum_{t=1}^{n} E\left\{\frac{\partial \log f_t}{\partial \theta} \frac{\partial \log f_t'}{\partial \theta}\right\}.$$

Then $\Gamma_n \to \Gamma = \Gamma(\theta_0)$ under $F(x, \theta_0)$, where Γ is positive definite, and

$$n^{1/2}(\tilde{\theta}_n - \theta_0) = n^{1/2}\Gamma^{-1} \sum_{t=1}^{n} \frac{\partial \log f_t}{\partial \theta} + o_p(1). \tag{2.8}$$

Note that (2.8) does not have the form stated in Assumption A4, so that a central limit theorem for triangular arrays must now be applied to (2.4), where the summands h_t now have the form $h_t(X_1, \cdots, X_t)$ with the function h_t varying with t. Potentially applicable theorems are given in Section 4 of Philipp (1969) and Section 2 of Withers (1975). In cases where such a result implies (2.2), considerable calculation shows that

$$C = I - qq' + Q - B\Gamma^{-1}B'.$$

Since $\tilde{\theta}_n$ is an asymptotically efficient estimator of θ from the X_t, this form of C also follows from the result of Pierce (1982). In the i.i.d. case, $Q = 0$ and $\Gamma = J$, producing the classical Chernoff-Lehmann result.

EXAMPLE. Let $\{X_t\}$ be a first-order Gaussian autoregressive process, so that each X_t is $N(\mu, \sigma^2)$ and $\rho_k = \rho^k$ for some ρ satisfying $-1 < \rho < 1$. For Cases 1-3, (2.2) follows from Theorem 2.1. For Case 4, (2.8) holds by Example 7.1 of Roussas (1979), and then (2.2) follows by Theorem 5 of Philipp (1969). Calculation shows that the covariance matrix in Case 3 ((μ, σ) estimated by the i.i.d.-case MLE $(\bar{X}, s)$) simplifies to

$$C = I - qq' + Q - \sigma^2 B \begin{pmatrix} \dfrac{1+\rho}{1-\rho} & 0 \\ 0 & \dfrac{1}{2}\dfrac{1+\rho^2}{1-\rho^2} \end{pmatrix} B'.$$

In Case 4 (MLE for the first order autoregressive model),

$$C = I - qq' + Q - \sigma^2 B \begin{pmatrix} \dfrac{1+\rho}{1-\rho} & 0 \\ 0 & \dfrac{1}{2} \end{pmatrix} B'.$$

Note the effect on the last term of the fact that the MLE's of μ but not of σ are (asymptotically) the same in the two cases. Both results should be compared with that for $\{X_t\}$ i.i.d. with (μ, σ) estimated by the MLE $(\bar{X}, s)$, the Chernoff-Lehmann case:

$$C = I - qq' - \sigma^2 B \begin{pmatrix} 1 & 0 \\ 0 & ½ \end{pmatrix} B'.$$

Most of the simplification in this example occurs for general Gaussian SSP's. For example, if $\{X_t\}$ is any SSP satisfying the conditions of Theorem 2.1, and (μ, σ) are estimated by $(\bar{X}, s)$,

$$C = I - qq' + Q - \sigma^2 B \begin{pmatrix} 1 + 2\sum_1^\infty \rho_k & 0 \\ 0 & ½ + \sum_1^\infty \rho_k^2 \end{pmatrix} B'.$$

3. The Gaussian no-estimation case. Let now $\{X_t\}$ be a Guassian SSP, so that the X_t are each $N(\mu, \sigma^2)$ but are not independent. If $\theta_0 = (\mu, \sigma^2)$, the vector of standardized cell frequencies $V_n(\theta_0)$ employed in the Pearson test for fit to the specific distribution $N(\mu, \sigma^2)$ has limiting null distribution $N_M(0, C)$ with $C = I - qq' + Q$ whenever $\sum |\rho_k| < \infty$, by Theorem 2.1. The matrix Q represents the effect of dependence. Note that qq' is an orthogonal projection of rank 1, orthogonal to $Q(q'C = q'Q = 0)$. Concentrate first on a single "incidence matrix" P_{ts} with components $P[X_t \text{ in } A_i, X_s \text{ in } A_j]$ for $i, j = 1, \cdots, M$.

LEMMA 3.1. *If (X_t, X_s) has a symmetric bivariate normal distribution with correlation $\rho(X_t, X_s) = \rho > 0$, then P_{ts} is positive definite.*

PROOF. Represent (X_s, X_t) in the form $X_s = Y + Z_s$, $X_t = Y + Z_t$, where Y, Z_s, Z_t are independent, Y is $N(\mu, \rho\sigma^2)$ and Z_t, Z_s are identically distributed as $N(0, (1-\rho)\sigma^2)$. Then conditional on $Y = y$, X_t and X_s are independent $N(y, (1-\rho)\sigma^2)$. Therefore if $p(y)$ is the M-vector of probabilities $p_i(y)$ of A_i under $N(y, (1-\rho)\sigma^2)$,

$$P_{ts} = \int p(y)p'(y)\, dF_Y(y).$$

For any M-vector x,

$$x'P_{ts}x = \int \{x'p(y)\}^2\, dF_Y(y)$$

and it follows that P_{ts} is positive definite.

For Q_{ts} as in (2.6), let

$$Q_n = n^{-1} \textstyle\sum_{t,s=1,t\neq s}^{n} Q_{ts} \tag{3.1}$$

so that $Q = \lim Q_n$. Denote by D the diagonal matrix with the cell probabilities p_i as diagonal entries. Then, if all $\rho_{ts} = \rho(X_t, X_s) > 0$,

$$Q_n = n^{-1}D^{-1/2}(\textstyle\sum P_{ts})D^{-1/2} - (n-1)qq', \tag{3.2}$$

where $\sum P_{ts}$ is over the same range of summation as in (3.1). Finally, set $C_n = I - qq' + Q_n$ so that $C_n \to C$.

THEOREM 3.1. *Suppose that $\{X_t\}$ is any Gaussian* SSP *with $\rho_k = \rho(X_t, X_{t+k}) \geq 0$ for all k and $\sum\rho_k < \infty$. Then* (2.2) *holds, and C has rank $M-1$ with all nonzero roots satisfying $\lambda \geq 1$.*

PROOF. Since $I + Q_n = C_n + qq'$ and $q'C_n = 0$, C_n has rank $r(C_n) = M - 1$ if and only if $I + Q_n$ is nonsingular. Suppose then that x is an M-vector such that $(I + Q_n)x = 0$. Since $q'Q_n = 0$, $q'x = 0$ also, and from (3.2) x satisfies

$$\{I + n^{-1}D^{-1/2}(\textstyle\sum P_{ts})D^{-1/2}\}x = 0$$

and $-n$ is therefore a characteristic root of $D^{-1/2}(\sum P_{ts})D^{-1/2}$ if $x \neq 0$. This contradicts Lemma 3.1, and therefore $I + Q_n$ is nonsingular.

The characteristic roots of C_n are one zero and $M - 1$ $\lambda_i = 1 + \delta_i$ where δ_i are the nonzero roots of Q_n, corresponding to characteristic vectors in $\mathscr{R}_Q = \mathscr{R}_{qq'}^{\perp}$. Thus $\delta_i > 0$ since they are roots of $P_n = n^{-1}D^{-1/2}(\sum P_{ts})D^{-1/2}$. (Because $q'P_n = (n-1)q'$, P_n has root $n-1$ in the qq' direction. The δ_i are the remaining roots of P_n.) The theorem now follows from Theorem 2.1 (which ensures the existence of $C = \lim C_n$) and continuity of characteristic roots as C_n converges to C.

The Pearson statistic $V_n'V_n$ has as its limiting null distribution the law of $\sum_{i=1}^{M-1}\lambda_i Z_i^2$ where λ_i are the nonzero roots of C and Z_i are i.i.d. $N(0, 1)$. If $\{X_t\}$ were i.i.d., this distribution would be $\chi^2(M-1)$, with all $\lambda_i = 1$. Note that $Q \neq 0$ if any $\rho_k > 0$.

COROLLARY 3.1. *Suppose that $\{X_t\}$ satisfies the conditions of Theorem* 3.1 *and that some $\rho_k > 0$. Then the limiting null distribution of the Pearson chi squared statistic for testing fit to a single normal law is stochastically larger than in the* i.i.d. *(all $\rho_k = 0$) case.*

Aside from guaranteeing Assumptions A2, A3 and (2.2), normality has been employed only in the proof that P_{ts} is positive definite. Corollary 3.1 therefore holds for suitably regular non-Gaussian processes for which P_{ts} is positive definite, and in particular whenever (X_t, X_s) are conditionally i.i.d. It is not true that every pair of random variables having a symmetric distribution with $\rho > 0$ are conditionally i.i.d.; see Dykstra, Hewett and Thompson (1973, page 676) for a counterexample. DeFinetti's theorem (see Kingman, 1978) asserts that the conditional i.i.d. property holds if (X_t, X_s) can be extended to a countable exchangeable sequence. Though $\rho > 0$ is a necessary condition for such an extension, I know of no sufficient condition.

It remains to investigate the extent to which the asymptotic critical points of the Pearson statistic in the case of positively correlated observations can exceed the chi squared critical points. This depends on the extent of the dependence among the observations, but we shall see that the characteristic roots of C, and hence the critical points, can be arbitrarily large. However, for C_n and hence for m-dependent processes, upper bounds are obtainable.

LEMMA 3.2. *Suppose* $\{X_t\}$ *is any Gaussian* SSP *with* $\rho_k > 0$ *for all* k. *Then the nonzero characteristic roots of* C_n *satisfy* $\lambda < n$. *If* $\rho_k = 0$ *for* $k > m$, *then* $\lambda < 2m + 1$ *for all* $n \geq m$.

PROOF. Returning to the proof of Theorem 3.1, we need only show that the roots δ_i of P_n satisfy $\delta_i < n - 1$. This follows if

$$G_n = (n-1)I - P_n = n^{-1}D^{-1/2}\{\sum(D - P_{ts})\}D^{-1/2}$$

is non-negative definite of rank $M - 1$. Represent (X_t, X_s) as in the proof of Lemma 3.1 and condition on $Y = y$ to obtain

$$D - P_{ts} = \int A(y)\, dF_Y(y),$$

where by conditional independence $A(y)$ has entries

$$A_{ij}(y) = \begin{cases} p_i(y)\{1 - p_i(y)\}, & i = j, \\ -p_i(y)p_j(y), & i \neq j, \end{cases}$$

with $p_i(y)$ as in the proof of Lemma 3.1. If $\varepsilon = (1, 1, \cdots, 1)'$, then $A(y)\varepsilon = 0$ and if $x \neq 0$, $x \perp \varepsilon$, then $x'A(y)x > 0$ for all y. Hence $x'(D - P_{ts})x > 0$ for all t, s and $x'\{\sum(D - P_{ts})\}x > 0$ for such x. Thus G_n is n.n.d. of rank $M - 1$ as claimed. If $\rho_k = 0$ for $k > m$, the $n - 1$ in the last term of (3.2) is replaced by $2m$, and the argument above can be repeated to yield the better bound.

Lemma 3.2 implies that the nonzero roots of C satisfy $1 \leq \lambda \leq 2m + 1$ in the m-dependent case. It is easy to see by arguments given in the proof of Theorem 3.2 below that all $\lambda \to 2m + 1$ as all $\rho_k \to 1$, and in this sense the upper bound of Lemma 3.2 cannot be improved. However, the correlations ρ_k of an m-dependent process are constrained by the requirement that the correlation matrix be positive definite, and cannot approach 1 arbitrarily closely. For example, the upper bounds on the largest correlation are ½ for $m = 1$, $\sqrt{2}/2 = 0.707$ for $m = 2$ and $(5 + 3\sqrt{5})/(10 + 2\sqrt{5}) = 0.809$ for $m = 3$. Thus the upper bound in the m-dependent case is conservative, especially for small m. One can nonetheless assert that, e.g., for a Gaussian 1-dependent process with positive correlation the critical points of the Pearson statistic fall between those of $\chi^2(M-1)$ and $3\chi^2(M-1)$.

Let $Q_k = Q_{ts}$ for $|t - s| = k$. Collecting terms in the definition of Q shows that

$$Q = 2\sum_{k=1}^{\infty} Q_k - 2\lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{n} Q_k$$

whenever the expressions on the right converge. In particular, if $\{X_t\}$ is m-dependent, then $Q = 2\sum_{k=1}^{m} Q_k$. These expressions are useful in the theorem and example that follow.

THEOREM 3.2. *Let $\{X_t\}$ be a first-order Gaussian autoregressive process with one-step correlation ρ. Then as $\rho \to 1$ all nonzero characteristic roots of C increase without bound.*

PROOF. The bound given by Gastwirth and Rubin (1975, page 811), shows that the entries of Q_k are $O(\rho^k)$. It follows easily that $Q = 2\sum_1^\infty Q_k$. For any integer $m > 1$, write $C = A + B$, where $A = I - qq' + 2\sum_1^m Q_k$ and $B = 2\sum_{m+1}^\infty Q_k$. Let $p_{ij} = P[X_1 \text{ in } A_i, X_{1+k} \text{ in } A_j]$. Using the representation of $P_{1,1+k}$ from the proof of Lemma 3.1 and the dominated convergence theorem, one sees that as $\rho^k = \rho(X_1, X_{1+k}) \to 1$, $p_{ij} \to 0$ for $i \neq j$ and $p_{ii} \to p_i = P[X_t \in A_i]$. Therefore

$$\lim_{\rho\to 1} A = (2m + 1)(I - qq').$$

Now B has rank $M - 1$ and range $\mathscr{R}_B = \mathscr{R}_Q$, the space onto which $I - qq'$ is the orthogonal projection. Further, $x'Bx > 0$ for any $x \neq 0$ in $\mathscr{R}_Q$, since for such x

$$x'Bx = 2x'D^{-1/2}(\textstyle\sum_{m+1}^\infty P_k)D^{-1/2}x$$

and $P_k = P_{ts}$ for $|t - s| = k$ is positive definite by Lemma 3.1. It follows easily that for any $\varepsilon > 0$, the $M - 1$ nonzero roots of C all exceed $2m + 1 - \varepsilon$ for ρ sufficiently near 1. But m was arbitrary, which establishes the theorem.

EXAMPLE. Computation of the roots of C is difficult in general because of the complexity of Q. Direct computation is possible using the bivariate normal quadrant probability in the case where $\mu = 0$ and $M = 2$ with cells $(-\infty, 0]$ and $(0, \infty)$.

Suppose first that $\{X_t\}$ is a Gaussian m-step moving average or other m-dependent process. The nonzero root of C is

$$\lambda = 1 + \frac{4}{\pi}\sum_{k=1}^m \sin^{-1}\rho_k.$$

As all ρ_k cover $0 < \rho_k < 1$, λ covers the range $1 < \lambda < 2m + 1$ of Lemma 3.2. In fact, the ρ_k cannot approach 1, and in the $m = 1$ case as ρ_1 covers $0 < \rho_1 < 1/2$, the nonzero root covers the range $1 < \lambda < 5/3$.

If $\{X_t\}$ is a first order Gaussian autoregressive process, on the other hand, the nonzero root of C_n ranges over all of $1 < \lambda < n$ as ρ ranges over $0 < \rho < 1$, and the nonzero root of C is

$$1 + \frac{4}{\pi}\sum_{k=1}^\infty \sin^{-1}\rho^k$$

which is finite but increases without bound as $\rho \to 1$, as announced in Theorem 3.2.

Table 1 contains values of the nonzero root λ and of the true asymptotic probability that the Pearson statistic exceeds chi squared critical points in this example for 1-dependent and first order autoregressive processes with various ρ. The choices of ρ are those of Gasser (1975) except for $\rho = 0.5$, which is the largest possible correlation in the 1-

TABLE 1
True significance levels of the Pearson statistic, two-cell example

Process	λ	Nominal $\alpha = 0.10$	Nominal $\alpha = 0.05$	Nominal $\alpha = 0.01$
1-dependent, $\rho = 0.3$	1.388	0.163	0.096	0.029
1-dependent, $\rho = 0.5$	1.667	0.203	0.129	0.046
autoregressive, $\rho = 0.3$	1.550	0.186	0.115	0.039
autoregressive, $\rho = 0.5$	2.275	0.276	0.194	0.088
autoregressive, $\rho = 0.6$	2.864	0.331	0.247	0.128
autoregressive, $\rho = 0.75$	4.356	0.431	0.348	0.217
autoregressive, $\rho = 0.9$	7.993	0.561	0.488	0.362

dependent case. The effect of dependence in this two cell example is somewhat stronger than for Gasser's simulated ten cell case, especially for small ρ.

A final remark: In some cases, knowledge of the SSP structure of the data and the theory given here allow construction of usable tests of fit for the univariate marginals. If $I + Q$ is nonsingular (which is usually the case), then a generalized inverse of C is $C^- = (I + Q)^{-1}$. Consequently $V_n'(I + Q)^{-1}V_n$ has the $\chi^2(M - 1)$ limiting null distribution, and the same will be true of $V_n'(I + \hat{Q}_n)^{-1}V_n$ for suitable estimators $\hat{Q}_n = \hat{Q}_n(X_1, \cdots, X_n)$ of Q. Suppose, for example, that $\{X_t\}$ is known to be a 1-dependent SSP, and that it is desired to test whether individual X_t are $N(\mu, \sigma^2)$. In this case Q has components $2(P_{ij} - p_i p_j)/(p_i p_j)^{1/2}$, where $p_i = P[X_t \text{ in } A_i]$ and $P_{ij} = P[X_t \text{ in } A_i, X_{t+1} \text{ in } A_j]$. The proof of Theorem 3.1 shows that $I + Q$ is nonsingular. Take $\hat{Q}_n$ to be the obvious count estimator of Q from $X_1, \cdots, X_n$. Then $\hat{Q}_n$ is a consistent estimator of Q and

$$\mathscr{L}\{V_n'(I + Q_n)^{-1}V_n\} \to \chi^2(M - 1)$$

under the null hypothesis. Similar statistics for m-dependent processes are immediate, but estimation of Q by counting cell frequencies rapidly becomes impractical as m increases.

Acknowledgment. I am grateful to Professors S. Geller and H. Rubin for helpful discussions.

REFERENCES

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980). *Statistical Inference for Stochastic Processes.* Academic, New York.

CHERNOFF, H. and LEHMANN, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25** 579–586.

DYKSTRA, R. L., HEWETT, J. E. and THOMPSON, W. A. (1973). Events which are almost independent. *Ann. Statist.* **1** 674–681.

GASSER, T. (1975). Goodness-of-fit tests for correlated data. *Biometrika* **62** 563–570.

GASTWIRTH, J. L. and RUBIN, H. (1975). The asymptotic distribution theory of the empiric c.d.f. for mixing processes. *Ann. Statist.* **3** 809–824.

IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theor. Probab. Appl.* **7** 349–382.

KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes,* Second Edition. Academic, New York.

KINGMAN, J. F. C. (1978). Uses of exchangeability. *Ann. Probab.* **6** 183–197.

MOORE, D. S. (1978). Chi-square tests. In. R. V. Hogg (Ed.) *Studies in Statistics.* Mathematical Association of America, Washington.

MOORE, D. S. and SPRUILL, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3** 599–616.

PIERCE, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10** 475–478.

PHILIPP, W. (1969). The central limit problem for mixing sequences of random variables. *Z. Wahrsch. verw. Gebiete* **12** 155–171.

ROUSSAS, G. G. (1979). Asymptotic distribution of the log-likelihood function for stochastic processes. *Z. Wahrsch. verw. Gebiete* **47** 31–46.

SUN, T. C. (1963). A central limit theorem for non-linear functions of a normal stationary process. *J. Math. Mech.* **12** 945–977.

WITHERS, C. S. (1975). Convergence of empirical processes of mixing rv's on [0, 1]. *Ann. Statist.* **3** 1101–1108.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907