The Annals of Statistics
1989, Vol. 17, No. 4, 1723–1734

STOCHASTIC INEQUALITIES RELATING A CLASS OF LOG-LIKELIHOOD RATIO STATISTICS TO THEIR ASYMPTOTIC χ^2 DISTRIBUTION

BY B. T. PORTEOUS

Durham University

For decomposable covariance selection models, stochastic inequalities which relate the null distribution of the log-likelihood ratio statistic to its asymptotic χ^2 distribution are obtained. The implications are twofold: First, the null distribution of the log-likelihood ratio statistic is seen to be stochastically larger than its asymptotic χ^2 distribution. Extremely large samples apart, for the χ^2 approximation to be valid, a deflation of the log-likelihood ratio statistic is then necessary. Second, a simple adjustment to the log-likelihood ratio statistic, similar in spirit to the Bartlett adjustment, yields a conservative test.

1. Introduction. With respect to decomposable covariance selection models, in this article we study how the exact null distribution of the log-likelihood ratio statistic relates to its asymptotic χ^2 distribution. For a general discussion of covariance selection models, see Dempster (1972), Wermuth (1976a, b), Lauritzen (1982), Speed and Kiiveri (1986) and Porteous (1985a): A decomposable model is defined later in this section. However, the main objectives of the present article are first stated:

1. When testing nested decomposable covariance selection models, it is shown that the exact null distribution of the log-likelihood ratio statistic $-n \log L$ is stochastically larger than its asymptotic χ^2 distribution.
2. When testing nested decomposable covariance selection models, it is shown that a simple adjustment to the log-likelihood ratio statistic yields a conservative test.

With respect to objective 1, it is widely accepted that, for small sample sizes, the null distribution of $-n \log L$ is not well approximated by its asymptotic χ^2 distribution. Hence the extensive work on corrections to $-n \log L$ such as the Bartlett adjustment and the F-approximation of Box (1949). However, the essential point of objective 1 is that, for all sample sizes, not just those which are "small," assessing significance by comparing $-n \log L$ to its asymptotic χ^2 distribution leads to a test which rejects the null model too often.

As stated in objective 2, when testing nested decomposable covariance selection models, a simple adjustment, hereafter denoted by c, to $-n \log L$ yields a conservative test. By this we mean that, under the null model, for all n it is true that $\Pr(-n \log L/c \geq y) \leq \Pr(\chi^2 \geq y)$ for all y. Hereafter, such stochastic inequalities will be denoted as $-n \log L/c \leq \chi^2$, for example.

Received November 1987; revised October 1988.
AMS 1980 *subject classifications*. Primary 62E15; secondary 62H99.
Key words and phrases. Asymptotic distribution, Bartlett adjustment, conservative test, covariance selection, decomposability, multivariate analysis, partitioning.

For multivariate normal data, the exact null distribution of the likelihood ratio statistic has been studied extensively. Wilks (1935) and Wald and Brookner (1941) are two well known examples of this work. For particular mutual independence hypotheses, these authors express the exact null distribution function of the likelihood ratio, or log-likelihood ratio statistic, in terms of incomplete beta or gamma integrals. We emphasize that the approach of this paper is different. Our interest is in studying how the exact null distribution of the log-likelihood ratio statistic relates to its asymptotic χ^2 distribution.

At this point, it is convenient to give a short preliminary discussion of covariance selection models. These models are used to analyse the covariance structure of a multivariate normal population. We assume that a sample from $N_p(0, \Omega)$, the multivariate normal distribution in p dimensions with zero mean and covariance matrix Ω, has been taken: There is no loss of generality in assuming a zero mean. The elements of Ω^{-1} are called concentrations and, typically, one is interested in testing if any of these are zero; zero concentrations correspond to conditional independence properties of the population. A graph can be associated with any covariance selection model: The nodes of the graph correspond to the p components of X with node α connected to node β if $\omega^{\alpha\beta} = (\Omega^{-1})_{\alpha\beta} \neq 0$. The graph is useful for identifying conditional independence properties of the model and for checking if it is decomposable: A model is decomposable if and only if it does not contain a chordless cycle of length greater than or equal to 4. An important property of decomposable covariance selection models is that estimation and testing can be performed in closed form, no iterative fitting being required. For the purposes of this paper this is a necessary property and, consequently, nondecomposable models are not considered.

A covariance selection model is specified by its generating class $\{G_1, \dots, G_k\}$. Here the G_i are subsets of $\Gamma = \{1, \dots, p\}$ such that no G_i is a subset of any other G_j. A model M is specified by $M = [G_1] \cdots [G_k]$ where, under M, $\omega^{\alpha\beta} = 0$ if $\{\alpha, \beta\} \not\subseteq G_i$ for any i. For example, the model in four dimensions with only $\omega^{14} = 0$ is specified by $M = [123][234]$.

2. The likelihood ratio statistic for decomposable models and its moment generating function. Assume that a sample $X_1, X_2, \dots, X_n$ from $N_p(0, \Omega)$ has been taken. Let

$$S = \sum_{i=1}^{n} X_i X_i^T / n$$

denote the sample variance matrix which follows a Wishart distribution. For any $a \subseteq \Gamma = \{1, \dots, p\}$, we take $S_{(a)}$ to denote the marginal matrix obtained by deleting all those rows and columns of S which do not correspond to the elements of a. The likelihood ratio statistic for testing a decomposable model M_0 against the saturated model is $L^{n/2}$ where L has the general form

$$L = |S| \prod_{q=1}^{t-1} |S_{(b_q)}| \Bigg/ \left\{ \prod_{q=1}^{t} |S_{(a_q)}| \right\}. \tag{2.1}$$

A full discussion of this has been given in the references previously cited. For our purposes, the null moment generating function of L is of crucial importance. Using Γ to denote the gamma function, define

$$\lambda^h(p) = \prod_{i=0}^{p-1} [\Gamma\{(n-i)/2 + h\}/\Gamma\{(n-i)/2\}].$$

With $p(a_q)$ denoting the cardinality of a_q, Porteous (1985a) has shown that

$$E(L^h|M_o) = \lambda^h(p) \prod_{q=1}^{t-1} \lambda^h\{p(b_q)\} \Big/ \left[\prod_{q=1}^{t} \lambda^h\{p(a_q)\} \right]. \tag{2.2}$$

A straightforward extension of the work of Wilks (1932) can be employed to obtain (2.2). Note that when testing nested decomposable models, but where the alternative model is nonsaturated, $E(L^h|M_0)$ is again straightforward to derive and, from considerations of symmetry, is of exactly the form that one expects.

3. Preliminary results. In this section we introduce two preliminary results relating the log-likelihood ratio statistic $-n \log L$ to its null asymptotic χ^2 distribution. Both results, although elementary, are of crucial importance in the development of this article.

Consider testing the null model $M_0 = [a][b]$ against the corresponding saturated model. Then with $a' = a \setminus b$ and $b' = b \setminus a$, in the usual conditional independence notation, the corresponding null hypothesis is H_0: $X_{(a')} \perp X_{(b')}|X_{(a \cap b)}$. Here $X_{(a)}$ is the marginal vector which is obtained by deleting those components of X not indexed by a: Individual components of X will be denoted by X_α ($\alpha = 1, \dots, p$). In this case, L has the simple form

$$L = |S|\,|S_{(a \cap b)}|/\{|S_{(a)}|\,|S_{(b)}|\}.$$

LEMMA 3.1. *The log-likelihood ratio statistic* $-n \log L$ *for testing* $M_0 = [a][b]$ *against the corresponding saturated model has the following properties*:

(i) *Assuming without loss of generality that* $p(a) \geq p(b)$, *if* $p = p(a) + 1$, *and* $p(a) - p(a \cap b) = 2$, *then under* M_0,

$$-n \log L \sim -n \log \mathrm{beta}[(n-p+1)/2, 1] \sim \{1 - (p-1)/n\}^{-1}\chi_2^2 \geq \chi_2^2.$$

(ii) *If* $a \setminus b = \{\alpha\}$ *and* $b \setminus a = \{\beta\}$, *implying that one is testing the hypothesis that the single concentration* $\omega^{\alpha\beta} = 0$, *then under* M_0, $L \sim \mathrm{beta}[(n-p+1)/2, 1/2]$ *and*

$$\chi_1^2 \leq [1 - (p-1)/n]^{-1}\chi_1^2 \leq -n \log L \leq [1 - p/n]^{-1}\chi_1^2.$$

Lemma 3.1(i), which extends an earlier result of Bartlett (1938), is easily deducible from (2.2) and 3.1(ii) follows by using elementary calculus and the fact that, in this case, L is closely related to the partial correlation coefficient.

As an illustration of Lemma 3.1(i) consider the following example: In four dimensions one may wish to test $M_0 = [12][234]$ against the saturated model. The corresponding null hypothesis is H_0: $X_1 \perp (X_3, X_4)|X_2$ and from Lemma

3.1(i) it follows that, under M_0,

$$-n \log L \sim (1 - 3/n)^{-1}\chi_2^2.$$

It is important to note that, for these special cases, the lemma shows that the null distribution of $-n \log L$ is stochastically larger than its corresponding asymptotic distribution and also may be bounded above by a multiple of this asymptotic distribution. We shall see later that these results are, in fact, rather general.

4. Partitioning the likelihood ratio statistic. The object of this section is to introduce the technique of partitioning the likelihood ratio statistic. For our purposes, this is a powerful tool which allows one to apply the simple results of Lemma 3.1 to obtain more general stochastic inequalities. To illustrate this point we present two examples. A formal treatment, based on the work of Sundberg (1975), is given in Section 5.

EXAMPLE 4.1. In four dimensions consider testing the null model $M_0 =$ [12][34] against the saturated model: That is, test the null hypothesis $H_0\colon (X_1, X_2) \perp (X_3, X_4)$. The likelihood ratio statistic $L^{n/2}$ for testing M_0 has the form

$$L = |S|/\{|S_{(1,2)}|\,|S_{(3,4)}|\},$$

where, from now on, we abuse notation by taking $S_{(1,2)} = S_{(\{1,2\})}$. The statistic L can also be written as

$$\begin{aligned} L &= \left[|S|\,|S_{(2)}|/\{|S_{(1,2)}|\,|S_{(2,3,4)}|\}\right] \times \left[|S_{(2,3,4)}|/\{|S_{(2)}|\,|S_{(3,4)}|\}\right] \\ &= L_1 L_2, \end{aligned}$$

say. Here, the statistic L_1 corresponds to testing the null model $M_0^1 =$ [12][234] against the saturated model and L_2 to testing $M_0^2 =$ [2][34] against the marginal saturated model $M_s =$ [234]. Moreover, M_0 implies that both M_0^1 and M_0^2 are true and, by the analogous partition of (2.2), it is straightforward to deduce that, under M_0, L_1 and L_2 are independent beta$[(n-3)/2, 1]$ and beta$[(n-2)/2, 1]$ variables, respectively. From Lemma 3.1(i) it follows that

$$-n \log L \sim (1 - 3/n)^{-1}\chi_2^2 + (1 - 2/n)^{-1}\chi_2^2.$$

Once more, the log-likelihood ratio statistic is stochastically larger than its asymptotic χ^2 distribution. Note that, under M_0, $-n \log L \le (1 - 3/n)^{-1}\chi_4^2$ implying that a conservative test of M_0 is available.

EXAMPLE 4.2. In four dimensions, consider testing $M_0 =$ [1][2][3][4] against the saturated model: That is, test the mutual independence hypothesis $H_0\colon X_1 \perp X_2 \perp X_3 \perp X_4$. With L_1 and L_2 defined in Example 4.1, the likelihood ratio statistic $L^{n/2}$ for testing M_0 can be partitioned as $L = L_1 L_2 L_3$ where

$$\begin{aligned} L_3 &= \left[|S_{(1,2)}|/\{|S_{(1)}|\,|S_{(2)}|\}\right] \times \left[|S_{(3,4)}|/\{|S_{(3)}|\,|S_{(4)}|\}\right] \\ &= L_4 L_5. \end{aligned}$$

The statistics L_4 and L_5 are of the form considered in Lemma 3.1(ii). They correspond, respectively, to testing the marginal models $M_0^4 = [1][2]$ and $M_0^5 = [3][4]$ against their respective saturated models. Moreover, under M_0, M_0^4 and M_0^5 are both true and L_4 and L_5 are both distributed as beta$[(n-1)/2, 1/2]$ variables. By the analogous partition of (2.2), it can also be deduced that, under M_0, L_1, L_2, L_3 and L_4 are mutually independent. Hence, under M_0, it follows that

$$-n \log L \sim (1 - 3/n)^{-1}\chi_2^2 + (1 - 2/n)^{-1}\chi_2^2 - n \log \text{beta}[(n-1)/2, 1/2] - n \log \text{beta}[(n-1)/2, 1/2].$$

Using Lemma 3.1(ii), $-n \log L$ can be bounded stochastically as

$$(1 - 1/n)^{-1}\chi_6^2 \leq -n \log L \leq (1 - 3/n)^{-1}\chi_6^2$$

and a conservative test is again available.

There is obviously more than one way of partitioning L. However, the null distribution of $-n \log L$ is clearly independent of this choice. We emphasize that the partitioning technique is only a tool which aids one in studying the null distribution of $-n \log L$. However, motivated by the simple relationship Lemma 3.1(i), we always partition L into as many beta$[\cdot, 1]$ components as possible. Note that whenever the first argument of a beta variable is not specified, from now on it shall be implicit that this is of the form $(n - t)/2$, for some positive integer t.

5. General stochastic inequalities. In this section we prove that, for decomposable covariance selection models, the previously discussed stochastic inequalities hold quite generally. Given Lemma 3.1 this result follows directly from

THEOREM 5.1. *Let* $-n \log L$ *denote the log-likelihood ratio statistic for testing nested decomposable covariance selection models. Then the null distribution of* L *can be expressed as a product of mutually independent beta*$[\cdot, 1]$ *and beta*$[\cdot, 1/2]$ *variables.*

The proof of this result, which relies heavily on the work of Sundberg (1975), is given in the Appendix. For completeness, we now state in full the main result of this article.

THEOREM 5.2. *Let* $-n \log L$ *denote the log-likelihood ratio statistic for testing nested decomposable covariance selection models. Then,*

(i) *The null distribution of* $-n \log L$ *is stochastically larger than its asymptotic* χ^2 *distribution.*

(ii) *An adjustment* c *can be found such that the null distribution of* $-n \log L/c$ *is stochastically smaller than the corresponding asymptotic* χ^2 *distribution, thus yielding a conservative test.*

That the null distribution of L can be expressed as a product of mutually independent beta$[\cdot, 1/2]$ and beta$[\cdot, 1]$ variables is established by Theorem 5.1, but this result is of little help in obtaining such an expression. However, if one tries to formalise the procedure of partitioning L, the simplicity of the technique is obscured. Consequently, no such treatment is attempted here. In particular cases, it is a relatively simple matter to find a partition and, in this respect, the graphical representation is useful (see Section 7).

Since any beta$[(n - t)/2, 1]$ variable can be expressed as a product of independent beta$[(n - t)/2, 1/2]$ and beta$[(n - t + 1)/2, 1/2]$ variables, clearly L can also be partitioned into a product of mutually independent beta$[\cdot, 1/2]$ variables; if $-n \log L$ is on d degrees of freedom, then this product will consist of d such variables.

Theorem 5.2(i) indicates that, even for large sample sizes, it may be dangerous to approximate the null distribution of $-n \log L$ by χ^2. For decomposable covariance selection models, the Bartlett adjustment is available [see Porteous (1985b)]. This adjustment of $-n \log L$ is known to improve substantially the χ^2 approximation: For empirical evidence supporting this assertion, see Section 6 and Porteous (1985a, b). For decomposable covariance selection models, the null distribution of $-n \log L$ is χ^2 plus terms of $O(n^{-1})$ whereas, from Box's (1949) results, for the Bartlett adjusted statistic, the error term is $O(n^{-2})$.

6. An empirical study. In all that follows, the Bartlett and conservative adjusted statistics are denoted by $-n \log L/b$ and $-n \log L/c$, respectively. In this section, when testing two simple null models M_0 against the saturated model, the exact null densities of $-n \log L$, $-n \log L/b$ and $-n \log L/c$ are compared numerically to their asymptotic χ^2 distribution. Both tests are of hypotheses in four dimensions. The Bartlett adjustment b can be calculated using the general formula of Porteous (1985b) whereas, for both of the tests considered, the conservative adjustment is $c = (1 - 3/n)^{-1}$.

Null Model 1: $M_0[1][234]$. The hypothesis which corresponds to testing M_0 against the saturated model is H_0: $X_1 \perp (X_2, X_3, X_4)$. Under M_0, the L-statistic for testing H_0 follows a beta$[(n - 3)/2, 3/2]$ distribution. This fact was used in calculating the following numerical results.

Null Model 2: $M_0 = [12][34]$. The hypothesis which corresponds to testing M_0 against the saturated model is H_0: $(X_1, X_2) \perp (X_3, X_4)$. The null distribution of the L-statistic for testing M_0 against the saturated model is known to be related to a beta$[(n - 3), 2]$ variable; this fact was used in calculating the following numerical results.

With $n = 5$, 25 and 50, the χ^2 density and the exact null densities of $-n \log L$, $-n \log L/b$ and $-n \log L/c$ are plotted in Figures 1 and 2 for null models 1 and 2, respectively. It can be seen that $-n \log L/b$ is always approximated extremely well by χ^2: In fact, for $n = 25, 50$, the exact null density of $-n \log L/b$ is indistinguishable from that of χ^2. Provided that the sample size is

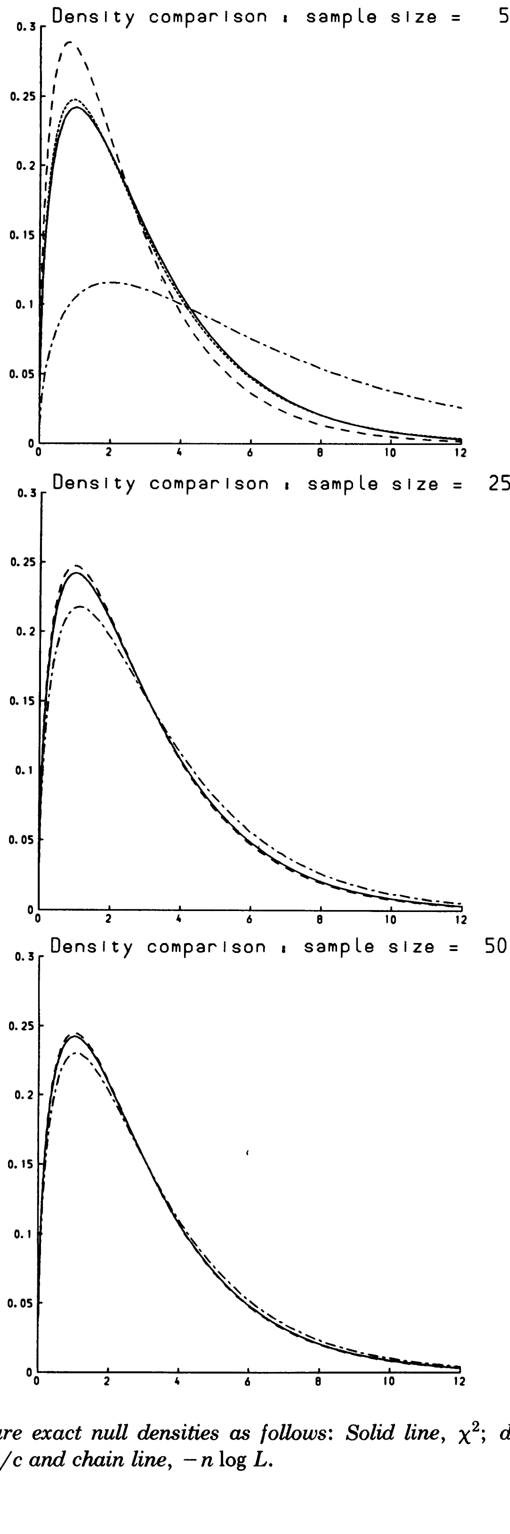

FIG. 1. *The curves are exact null densities as follows: Solid line,* χ^2; *dotted line,* $-n \log L/b$; *dashed line,* $-n \log L/c$ *and chain line,* $-n \log L$.

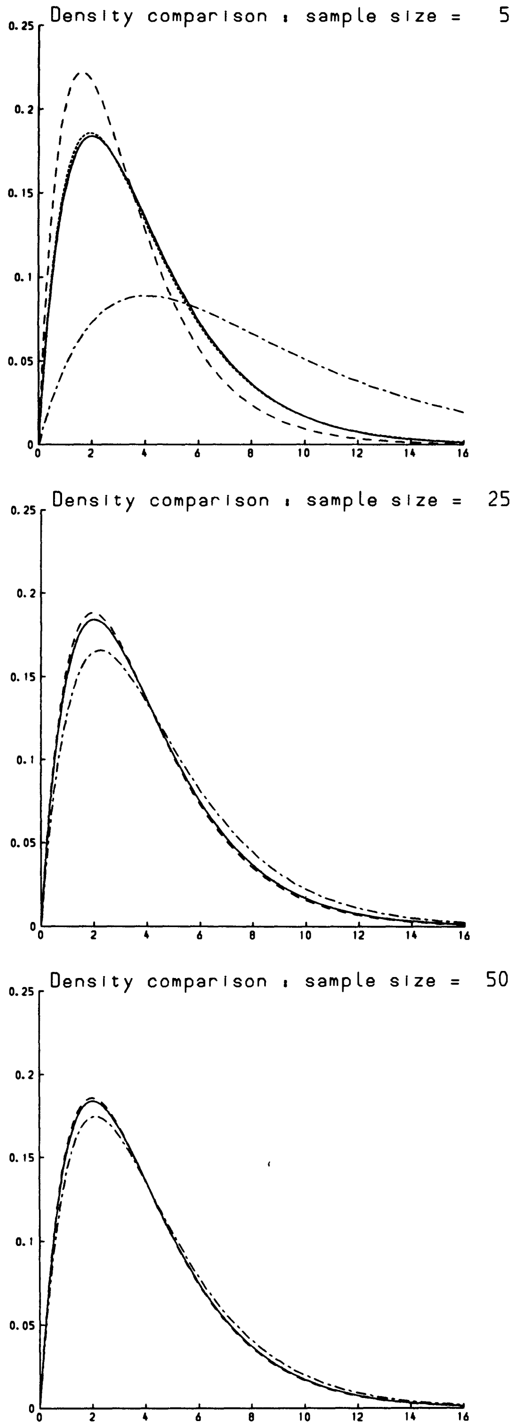

FIG. 2. *The curves are exact null densities as follows: Solid line,* χ^2; *dotted line,* $-n \log L/b$; *dashed line,* $-n \log L/c$ *and chain line,* $-n \log L$.

TABLE 1

Exact rejection rates corresponding to (a) *Figure* 1 *and* (b) *Figure* 2. *The numbers in row* (*i*) *are the exact null probabilities that* $-n \log L$ *exceeds the appropriate upper percentage point of* χ^2. *Corresponding rates for* $-n \log L/b$ *and* $-n \log L/c$ *are given in rows* (*ii*) *and* (*iii*), *respectively.*

		(a) M_0 = [1][234]				(b) M_0 = [12][34]			
		10%	5%	1%	0.1%	10%	5%	1%	0.1%
$n = 5$	(i)	0.3972	0.2972	0.1510	0.0574	0.4392	0.3337	0.1736	0.0668
	(ii)	0.1010	0.0521	0.0116	0.0014	0.1006	0.0515	0.0111	0.0013
	(iii)	0.0651	0.0300	0.0052	0.0004	0.0555	0.0245	0.0038	0.0003
$n = 25$	(i)	0.1316	0.0711	0.0170	0.0022	0.1360	0.0739	0.0178	0.0023
	(ii)	0.1000	0.0500	0.0100	0.0010	0.1000	0.0500	0.0100	0.0010
	(iii)	0.0942	0.0463	0.0089	0.0009	0.0934	0.0459	0.0088	0.0008
$n = 50$	(i)	0.1147	0.0596	0.0130	0.0015	0.1167	0.0608	0.0133	0.0015
	(ii)	0.1000	0.0500	0.0100	0.0010	0.1000	0.0500	0.0100	0.0010
	(iii)	0.0972	0.0482	0.0095	0.0009	0.0968	0.0480	0.0094	0.0009

TABLE 2

Model types M_0 *for which, when tested against the saturated model, the null distribution of the log-likelihood ratio statistic has a simple form.*

	M_0	Exact null distribution
3 dimensions	[12][3]	$(1-2/n)^{-1}\chi_2^2$
4 dimensions	[123][34]	$(1-3/n)^{-1}\chi_2^2$
	[12][34], [12][23][4]	$(1-2/n)^{-1}\chi_2^2 + (1-3/n)^{-1}\chi_2^2$
5 dimensions	[1234][345]	$(1-4/n)^{-1}\chi_2^2$
	[123][234][45], [123][234][35], [123][345]	$(1-3/n)^{-1}\chi_2^2 + (1-4/n)^{-1}\chi_2^2$
	[1234][5]	$(1-2/n)^{-1}\chi_2^2 + (1-4/n)^{-1}\chi_2^2$
	[123][34][5], [123][45]	$(1-2/n)^{-1}\chi_2^2 + (1-3/n)^{-1}\chi_2^2 + (1-4/n)^{-1}\chi_2^2$
	[12][23][4][5], [12][34][5]	$(1-2/n)^{-1}\chi_4^2 + (1-3/n)^{-1}\chi_2^2 + (1-4/n)^{-1}\chi_2^2$

not too small, $-n \log L/c$ is approximated adequately by χ^2. However, for the sample sizes considered, $-n \log L$ is never near to χ^2.

The rejection rates which correspond to Figures 1 and 2 are presented in Table 1, and similar conclusions apply. Note that, even for $n = 50$, approximating $-n \log L$ by χ^2 is unsatisfactory.

7. Discussion. The main results of this paper are stated in Theorem 5.2. It is now natural to ask the question: Does Theorem 5.2(i) hold more generally? For most families of distributions, this question will be far more difficult to address than it was for the models of this article. However, the serious implications of Theorem 5.2(i) justify that, in general, further study should be devoted to obtaining a better understanding of how the null distribution of $-n \log L$

approaches χ^2. This point is particularly pertinent when corrections , such as the Bartlett and conservative adjustments, are not available.

Throughout this paper it has been indicated that, for certain tests, the exact null distribution of $-n \log L$ has a particularly simple form; this occurs if L can be partitioned into a product of mutually independent beta$[\cdot, 1]$ variables. When testing a null model M_0 against the saturated model in three to five dimensions, Table 2 presents a comprehensive list of all such tests. Note also that, in three, four and five dimensions the corresponding conservative tests are of the respective forms $(1 - 2/n)(-n \log L)$, $(1 - 3/n)(-n \log L)$ and $(1 - 4/n)(-n \log L)$. In fact, it has been pointed out by one of the referees that when testing any null decomposable model M_0 not of the form discussed in Lemma 3.1(ii) against the saturated model, this is indeed the general pattern, although a formal treatment is not attempted here. For tests involving a single concentration, from Lemma 3.1(ii) it can be seen that, in p dimensions, $(1 - p/n)(-n \log L)$ is conservative. Hence a slightly larger deflation is appropriate.

In fact, since the null distribution of L can be partitioned into a product of mutually independent beta$[(n - t), 1/2]$ variables, it follows from Lemma 3.1(ii) and the fact that, in p dimensions, each $t \leq p - 1$, that for all nested decomposable models, $(1 - p/n)(-n \log L)$ is conservative. But note that, in most cases, as above, a smaller deflation will be appropriate. However, $(1 - p/n)(-n \log L)$ is both simple to use and widely applicable.

To conclude, for decomposable covariance selection models, it has been demonstrated in this article that if significance is to be assessed by referring $-n \log L$ to its asymptotic distribution, then an adjustment of $-n \log L$ is appropriate. Both Bartlett and conservative adjustments are available and their relative merits have been discussed.

APPENDIX

Together with Sundberg's (1975) results, Theorem 5.1 follows from

Lemma A1. *Let $-n \log L$ denote the log-likelihood ratio statistic for testing the null model $M_0 = [a][b]$ against the corresponding saturated model. Then the null distribution of L can be expressed as a product of mutually independent beta$[\cdot, 1]$ and beta$[\cdot, 1/2]$ variables.*

Proof. We use induction on d, the degrees of freedom corresponding to $-n \log L$.

(i) $d = 1$. The test involves a single concentration and is of the form discussed in Section 3. In this case, $L \sim$ beta$[(n - p + 1)/2, 1/2]$.

(ii) $d = 2$. There are two cases of consider and we discuss these in terms of concentrations and the corresponding null hypotheses: (a) For distinct α, β and δ, the null hypothesis is H_0: $\omega^{\alpha\beta} = \omega^{\alpha\delta} = 0$. Then $a = \Gamma \setminus \{\alpha\}$, $b = \Gamma \setminus \{\beta, \delta\}$ and by Lemma 3.1(i), $L \sim$ beta$[(n - p + 1)/2, 1]$. (b) For distinct α, β, δ and γ the null hypothesis is H_0: $\omega^{\alpha\beta} = \omega^{\delta\gamma} = 0$. But in this case, the interaction graph of the null model contains the chordless cycle of length 4, $\alpha \sim \gamma \sim \beta \sim \delta \sim \alpha$.

Therefore the null model is not decomposable and this case need not be considered.

(iii) Assume that Lemma A1 holds for all tests on $d \le k$ degrees of freedom, where $k \ge 2$, and take $d = k + 1$. Without loss of generality we may assume that a has dimension at most $p - 2$. Take $\alpha \in b \setminus a$ and, dispensing with the usual set notation for the singleton α, partition L as

$$\begin{aligned} L &= |S|\,|S_{(a \cap b)}|/\{|S_{(a)}|\,|S_{(b)}|\} \\ &= \big[|S|\,|S_{(\{a \cup \alpha\} \cap b)}|/\{|S_{(a \cup \alpha)}|\,|S_{(b)}|\}\big]\big[|S_{(a \cup \alpha)}|\,|S_{(a \cap b)}|/\{|S_{(a)}|\,|S_{(\{a \cup \alpha\} \cap b)}|\}\big] \\ &= L_1 L_2. \end{aligned}$$

The statistic L_1 corresponds to testing $M_0^1 = [a \cup \alpha][b]$ against the saturated model and, in the marginal distribution of $X_{(a \cup \alpha)}$, L_2 corresponds to testing $M_0^2 = [a][\{a \cup \alpha\} \cap b] = [a][\{a \cap b\} \cup \alpha]$ against the saturated model. Moreover, M_0 implies that both M_0^1 and M_0^2 are true. Let d_1 and d_2 be the degrees of freedom corresponding to L_1 and L_2. Then since $d_1 \ge 1$, $d_2 \ge 1$ and $d = d_1 + d_2$, it follows that $d_1 \le k$, $d_2 \le k$ and, by the induction hypothesis, the null distributions of L_1 and L_2 can be expressed as products of the desired form. Finally, the analogous partition of (2.2) implies that, under M_0, L_1 and L_2 are independent. Hence the null distribution of L is a product of the correct form.

This completes the proof of the lemma. □

By routine application of the arguments used by Sundberg in proving his Lemma's 1, 2, 3 and 4, it can be deduced that the L-statistic for testing nested decomposable covariance selection models can be factorised into a product of L-statistics, each one corresponding to a test of the form specified in the statement of Lemma A1. Note that Sundberg calls the corresponding hypothesis a "general independence" hypothesis. By symmetry, there is an analogous factorisation of the corresponding moment generating function and it then follows that the L-statistics of the factorisation are mutually independent. The proof of Theorem 5.1 now follows from Lemma A1.

Acknowledgments. I am grateful to both referees and the Editor for very constructive comments which have led to an improvement in the presentation of the results.

REFERENCES

BARTLETT, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Cambridge Philos. Soc.* **34** 33–40.

BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36** 317–346.

DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

LAURITZEN, S. L. (1982). *Lectures on Contingency Tables*, 2nd ed. Univ. Aalborg Press, Denmark.

PORTEOUS, B. T. (1985a). Properties of log linear and covariance selection models. Ph.D. dissertation, Univ. Cambridge.

PORTEOUS, B. T. (1985b). Improved likelihood ratio statistics for covariance selection models. *Biometrika* **72** 97–101.

SPEED, T. P. and KIIVERI, H. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138–150.

SUNDBERG, R. (1975). Some results about decomposable (or Markov-type) models for multi-dimensional contingency tables: Distribution of marginals and partitioning of tests. *Scand. J. Statist.* **2** 71–79.

WALD, A. and BROOKNER, R. J. (1941). On the distribution of Wilks' statistic for testing the independence of several groups of variates. *Ann. Math. Statist.* **12** 137–152.

WERMUTH, N. (1976a). Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics* **32** 95–108.

WERMUTH, N. (1976b). Model search among multiplicative models. *Biometrics* **32** 253–263.

WILKS, S. S. (1932). Certain generalisations in the analysis of variance. *Biometrika* **24** 471–494.

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* **3** 309–326.

58 PALMERSTON PLACE (1ST FLAT)
EDINBURGH EH12 5AY
UNITED KINGDOM