A likelihood ratio test for equality of natural parameters for generalized Riesz distributions

Steen A. Andersson and Jesse B. Crawford
Indiana University and Tarleton State University

Abstract

An important classical problem is testing whether several centered multivariate normal distributions have the same covariance matrix, which is equivalent to testing that certain Wishart distributions have the same natural parameter. Wishart distributions, which are supported on sets of positive definite matrices, are a special case of generalized Riesz distributions, which are supported on sets of matrices related to the Markov properties of decomposable undirected graphs. This leads to the problem of testing whether several generalized Riesz distributions have the same natural parameter. In this paper, we derive the likelihood ratio statistic for this testing problem and find its moments.

AMS (2010) *subject classification.* Primary 62H15; Secondary 62H12, 62E17, 06F99.
Keywords and phrases. Bartlett's test, Decomposable undirected graphs, Graphical models, Likelihood ratio test, Riesz distribution, Wishart distribution.

1. **Introduction**

Neyman and Pearson (1931) derived the likelihood ratio statistic for testing that several normal populations all have the same variance, and they found its moments, allowing its distribution to be approximated, cf. Box (1949). Wilks (1932) then solved this problem in the multivariate setting by finding the moments of the likelihood ratio statistic for testing homogeneity of covariance matrices. In this article, an analogous problem is solved for the generalized Riesz distributions developed by Andersson and Klein (2010). Namely, we find the moments of the likelihood ratio statistic for testing whether several generalized Riesz distributions have the same natural parameter. To begin, the classical problems described above are summarized, cf. Chapter 10 of Anderson (2003) for more details.

Suppose $(X_\alpha^{(k)} \mid \alpha = 1, \ldots, N_k)$ is a random sample from $\mathrm{N}(\mu^{(k)}, \Sigma_k)$, the multivariate normal distribution with mean $\mu^{(k)}$ and covariance matrix Σ_k, for $k = 1, \ldots, K$. The classical hypothesis of homogeneity of covariance matrices is

$$\mathrm{H}_0^c : \Sigma_1 = \cdots = \Sigma_K. \tag{1}$$

Denoting the sample covariance matrix for the kth sample by $\hat{\Sigma}_k$, the estimator for the common covariance matrix under H_0^c is

$$\hat{\Sigma} = \frac{1}{N} \sum_{k=1}^{K} N_k \hat{\Sigma}_k,$$

where $N = \sum_{k=1}^{K} N_k$, and the likelihood ratio statistic for testing H_0^c is

$$q_c = \frac{\prod_{k=1}^{K} |\hat{\Sigma}_k|^{N_k/2}}{|\hat{\Sigma}|^{N/2}}.$$

Bartlett (1937) conjectured that the test based on q_c is biased and recommended replacing each N_k with $N_k - 1$ and therefore replacing N with $N - K$. Brown (1939) confirmed that the test based on q_c is biased in the univariate case, and Pitman (1939) demonstrated the unbiasedness of Bartlett's modification in the univariate case. These results were later extended to the multivariate setting by Sugiura and Nagao (1968), Perlman (1980), and others, cf. Perlman's introduction and references.

Because we are primarily interested in the covariance structure, we will modify the problem by assuming

$$\mu^{(1)} = \cdots = \mu^{(K)} = 0. \tag{2}$$

Each sample can now be replaced by the sufficient statistic $\hat{\Sigma}_k$, which has a Wishart distribution with expectation Σ_k and N_k degrees of freedom. This distribution is denoted by $\mathbb{W}_{\Sigma_k,\lambda_k}$, where $\lambda_k = N_k/2$ is the shape parameter, cf. Section 2.

Now, under the additional assumption (2), the testing problem (1) is equivalent to testing whether the Wishart-distributed variables $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_K$ all have the same expectation parameter.

The natural parameter for a $\mathbb{W}_{\Sigma,\lambda}$ distribution is $\Delta = \lambda \Sigma^{-1}$, and in particular, the random variable $\lambda_k \hat{\Sigma}_k$ has natural parameter Σ_k^{-1}, for $k = 1, \ldots, K$. Therefore, the Wishart-distributed random variables $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_K$ all have the same expectation parameter if and only if the Wishart-distributed random variables $\lambda_1 \hat{\Sigma}_1, \ldots, \lambda_K \hat{\Sigma}_K$ all have the same natural parameter, reformulating the testing problem as a test of equality of natural parameters for Wishart distributions.

Wishart distributions have been generalized to classical Riesz distributions by Hassairi and Lajmi (2001) and to generalized Riesz distributions by Andersson and Klein (2010). Generalized Riesz distributions arise when estimating the covariance matrix of a sample from a Gaussian graphical model, cf. Proposition 3.15. Given several such samples with covariance matrices $\Sigma_1, \ldots, \Sigma_K$, a natural testing problem to consider is

$$\mathrm{H}_0^{GM} : \Sigma_1 = \cdots = \Sigma_K,$$

a generalization of the classical testing problem described above. The classical problem reduced to testing that several Wishart-distributed random variables have the same natural parameter, and as discussed in Section 3.4, testing H_0^{GM} is equivalent to testing that several generalized Riesz distributions have the same natural parameter.

Testing equality of natural parameters for several generalized Riesz distributions is examined in Section 5, culminating in the moments of the likelihood ratio statistic in Theorem 5.3. Beforehand, classical Wishart distributions are covered briefly in Section 2, and the theory of generalized Riesz distributions from Andersson and Klein (2010) is discussed in Section 3.

Generalized Riesz distributions are defined with respect to an acyclic mixed graph $\mathcal{V}$, which defines a recursive structure on a decomposable undirected graph $\mathcal{U}$, and some theorems about these distributions can be proved using induction on the number of *boxes* in $\mathcal{V}$, cf. Proposition 3.13. In fact, this is the strategy used to prove Theorem 5.3, but the induction step requires the moments of the likelihood ratio statistic of another testing problem, which are obtained in Section 4 using an invariance argument.

2. **Classical Wishart Distributions**

Let V be a finite set, and let $\mathbf{PD}(V)$ and $\mathbf{S}(V)$ be the sets of all real positive definite and symmetric $V \times V$ matrices, respectively.[1] Given a sample of size N from a centered multivariate normal distribution on $\mathbb{R}^V$ with covariance matrix $\Sigma \in \mathbf{PD}(V)$, the empirical covariance matrix $\hat{\Sigma}$ has a classical Wishart distribution with expectation Σ and shape parameter $\lambda = N/2$, denoted $\mathbb{W}_{\Sigma,\lambda}$, cf. Wishart (1928). Its density is

$$\mathrm{d}\mathbb{W}_{\Sigma,\lambda}(S) := \frac{\pi^{-\frac{V(V-1)}{4}} \lambda^{\lambda V} |S|^{\lambda - \frac{V+1}{2}}}{\prod(\Gamma(\lambda - \frac{i-1}{2}) \mid i = 1, \dots, V)|\Sigma|^{\lambda}} \exp\{-\lambda \mathrm{tr}(\Sigma^{-1} S)\}\mathrm{d}S,$$

where $|S|$ denotes the determinant of $S \in \mathbf{PD}(V)$, and $\mathrm{d}S$ is the standard Lebesgue measure on $\mathbf{S}(V)$ restricted to $\mathbf{PD}(V)$. Actually, the Wishart distribution is well defined for any $\lambda > \frac{V-1}{2}$. The parameter Σ deserves its name, since $\mathbb{E}(\mathbb{W}_{\Sigma,\lambda}) = \Sigma$, where $\mathbb{E}(\cdot)$ denotes expectation.

The statistical model[2] $(\mathbb{W}_{\Sigma,\lambda} \in \mathcal{P}(\mathbf{S}(V)) \mid \Sigma \in \mathbf{PD}(V))$ is a full regular exponential family in its expectation parameterization, and the corresponding natural parameter is $\Delta := \lambda\Sigma^{-1}$. Denoting a Wishart distribution with natural parameter Δ and shape parameter λ as $\mathrm{W}_{\Delta,\lambda}$, the density is

$$\mathrm{d}\mathrm{W}_{\Delta,\lambda}(S) = \frac{\pi^{-\frac{V(V-1)}{4}} |\Delta|^{\lambda} |S|^{\lambda - \frac{V+1}{2}}}{\prod(\Gamma(\lambda - \frac{i-1}{2}) \bigm| i = 1, \dots, V)} \exp\{-\mathrm{tr}(\Delta S)\}\mathrm{d}S.$$

If $S \sim \mathrm{W}_{\Delta,\lambda}$, and $\alpha > 0$, then

[1] A real $V \times V$ matrix is a mapping from $V \times V$ into $\mathbb{R}$, and a vector in $\mathbb{R}^V$ is a mapping from V to $\mathbb{R}$. For classical Wishart distributions, they can be regarded as ordinary $|V| \times |V|$ matrices and $|V|$-dimensional vectors, where $|V|$ is the cardinality of V. Throughout this article, the cardinality of any set A will simply be denoted by A.

[2] A statistical model is a family $(P_\theta \in \mathcal{P}(\Omega) \mid \theta \in \Theta)$ of probability measures on the same measurable space Ω, belonging to the set $\mathcal{P}(\Omega)$ of all probability measures on Ω, parameterized by a set Θ called the parameter set.

$$\mathbb{E}[|S|^\alpha] = \prod \left(\frac{\Gamma(\alpha + \lambda + \frac{1-i}{2})}{\Gamma(\lambda + \frac{1-i}{2})} \,\middle|\, i = 1, \ldots, I \right) |\Delta|^{-\alpha}. \tag{3}$$

3. Generalized Riesz Distributions

Wishart distributions provide a flexible family of probability distributions on $\mathbf{PD}(V)$. Generalized Riesz distributions generalize these distributions to $\mathbf{P}(\mathcal{U})$, a projection of the set $\mathbf{PD}(\mathcal{U})$ of covariance matrices satisfying the Markov properties for a decomposable undirected graph $\mathcal{U}$. This section presents the important properties of these distributions that will be needed later, cf. Andersson and Klein (2010).

3.1. Graph Theoretic Preliminaries. Let $\mathcal{U} = (V, F)$ be a decomposable undirected graph (DUG) with vertex set V and edge set $F \subset V \times V$, cf. Lauritzen (1996) for some basic concepts in graph theory. The undirected graph $\mathcal{U}$ is assumed to be decomposable so that it will have a representation as an acyclic mixed graph, as described in Definition 3.1. Define

$$\mathbf{S}(\mathcal{U}) := \{S \in \mathbf{S}(V) \mid S_{uv} = 0 \text{ for all } u, v \in V \text{ with } u \neq v \text{ and } (u, v) \notin F\}.$$

This is a subspace of $\mathbf{S}(V)$ with projection mapping $p_{\mathcal{U}} = p : \mathbf{S}(V) \to \mathbf{S}(\mathcal{U})$ defined by

$$p(S)_{uv} := \begin{cases} S_{uv} & \text{if } (u, v) \in F \text{ or } u = v \\ 0 & \text{if } (u, v) \notin F \text{ and } u \neq v \end{cases}, \tag{4}$$

for all $S \in \mathbf{S}(V)$. Define $\mathbf{PD}^0(\mathcal{U}) := \mathbf{S}(\mathcal{U}) \cap \mathbf{PD}(V)$ and $\mathbf{PD}(\mathcal{U}) := \mathbf{PD}^0(\mathcal{U})^{-1}$. A centered normal distribution on $\mathbb{R}^V$ with covariance matrix $\Sigma \in \mathbf{PD}(V)$ satisfies the Markov properties given by $\mathcal{U}$ if and only if $\Sigma \in \mathbf{PD}(\mathcal{U})$.

The strong, weak, and pairwise Markov properties are equivalent in this case, cf. Chapter 3 of Lauritzen (1996) for an overview of Markov properties given by undirected graphs. They are also equivalent to the LWF Markov properties from Lauritzen and Wermuth (1989) and Frydenberg (1990) and the alternative Markov property (AMP) from Andersson et al. (2001), all of which are defined for acyclic mixed graphs.

Let $\mathcal{C}$ be the set of cliques in $\mathcal{U}$, and define

$$\mathbf{P}(\mathcal{U}) := \{S \in \mathbf{S}(\mathcal{U}) \mid S_C \in \mathbf{PD}(C), \text{ for every } C \in \mathcal{C}\},$$

where S_A is the $A \times A$ submatrix[3] of S, for any $A \subseteq V$. The mapping

$$\begin{array}{rcl} \mathbf{PD}(\mathcal{U}) & \to & \mathbf{P}(\mathcal{U}) \\ S & \mapsto & p(S) \end{array} \tag{5}$$

is bijective, and the inverse mapping is denoted by

$$\begin{array}{rcl} \mathbf{P}(\mathcal{U}) & \to & \mathbf{PD}(\mathcal{U}) \\ S & \mapsto & \tilde{S} := p^{-1}(S). \end{array} \tag{6}$$

[3]The $A \times B$ submatrix is denoted by $S_{A \times B}$.

Suppose $\mathcal{V}$ is an acyclic mixed graph (AMG) with vertex set V and edge set $F \subset V \times V$. Edges in $\mathcal{V}$ can be directed or undirected, i.e., for $v_1, v_2 \in V$, $v_1 \to v_2$ means that $(v_1, v_2) \in F$, but $(v_2, v_1) \notin F$, and $v_1 - v_2$ indicates that $(v_1, v_2) \in F$ and $(v_2, v_1) \in F$. For a given vertex $v \in V$, the parents and neighbors of v are

$$\mathrm{pa}(v) \equiv \mathrm{pa}_{\mathcal{V}}(v) := \{v' \in V \mid v' \to v\},$$
$$\mathrm{nb}(v) \equiv \mathrm{nb}_{\mathcal{V}}(v) := \{v' \in V \mid v' - v\}.$$

Define an equivalence relation on V by $v \sim v'$ if $v = v'$ or if there is an undirected path from v to v' in $\mathcal{V}$. The corresponding equivalence classes are called boxes, and the set of equivalence classes is denoted by $V/\sim$. When Andersson and Klein wanted to emphasize that a box $B \in V/\sim$ is a subset of V, they used the notation $[B] := B$. If B and B' are distinct boxes such that $v \in [B]$, $v' \in [B']$, and $v \to v'$, then we write $B \to B'$, making the set of boxes into an acyclic directed graph (ADG). If $B \not\to B'$, for any $B' \in V/\sim$, the box B is called maximal.

The skeleton of $\mathcal{V}$ is the undirected graph $\mathcal{U}(\mathcal{V}) := (V, F \cup F^\circ)$, where $F^\circ := \{(v', v) \mid (v, v') \in F\}$. Given a DUG $\mathcal{U}$, a class of generalized Riesz distributions is in general only defined relative to a choice of a recursive structure on $\mathcal{U}$, i.e., a choice of an AMG $\mathcal{V}$ with skeleton $\mathcal{U}$, cf. Definition 3.7. This AMG must have the following two properties for the generalized Riesz distributions to be defined.

- (A1) The subgraphs of $\mathcal{V}$ induced by its boxes are complete.
- (A2) The graph $\mathcal{V}$ has no triplexes.[4]

Definition 3.1. If $\mathcal{U}$ is a DUG, and $\mathcal{V}$ is an AMG with skeleton $\mathcal{U}$ satisfying (A1) and (A2), then $\mathcal{V}$ is called a *representation* of $\mathcal{U}$ (as an AMG).

Every DUG can be turned into an ADG without immoralities by converting lines to arrows, so every DUG has a representation as an AMG.

The assumptions (A1) and (A2) are essential for the construction of generalized Riesz distributions. As we will see in Proposition 3.13, these distributions are constructed recursively from Wishart distributions, and in the case where $\mathcal{V}$ has only one box, a generalized Riesz distribution is a Wishart distribution, cf. Remark 3.14. Because generalized Riesz distributions are defined on $\mathbf{P}(\mathcal{U})$, this requires $\mathbf{P}(\mathcal{U}) = \mathbf{PD}(V)$ when $\mathcal{V}$ has only one box, which requires $\mathcal{U}$ to be complete in this case, as guaranteed by condition (A1).

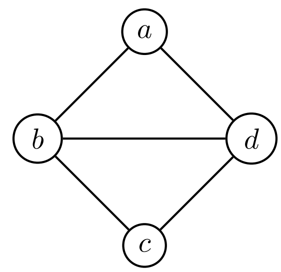
As mentioned above, the AMP and LWF properties for $\mathcal{U}$ are equivalent, and assumption (A2) implies this is also true for $\mathcal{V}$. Furthermore, condition (A2) is essential, because it guarantees that $\mathcal{U}$ and $\mathcal{V}$ are Markov equivalent (wrt. both the AMP and LWF properties).

Because $\mathcal{V}$ satisfies properties (A1) and (A2), $\mathrm{pa}(v)$ depends only on which box contains v, and we define $\langle B \rangle := \mathrm{pa}(v)$, where $v \in [B]$. Furthermore, these conditions imply that the subsets $\langle B \rangle$ and $[B] \cup \langle B \rangle$ induce AMGs with complete skeletons, making certain submatrices of S invariant under the mappings (5) and (6).

For the following DUG $\mathcal{U}$, there are 23 possible representations of $\mathcal{U}$ as an AMG $\mathcal{V}$. Ignoring labeling of the vertices, there are only eight, which are displayed in Figure 1.

---

[4]A triplex is an induced subgraph of the form $\bullet - \bullet \leftarrow \bullet$ (a *flag*) or $\bullet \to \bullet \leftarrow \bullet$ (an *immorality*).



(7)

As mentioned above, every DUG $\mathcal{U}$ has a representation $\mathcal{V}$, where $\mathcal{V}$ is an acyclic directed graph. Therefore, why is it worthwhile to define generalized Riesz distributions with respect to AMG's instead of restricting attention to ADG's?

Not only do AMG's provide a larger family of distributions, but every DUG $\mathcal{U}$ has an *intrinsic* representation as an AMG $\mathcal{V}$. That is, $\mathcal{V}$ does not depend on any arbitrary choices, such as an arbitrary ordering of the vertices, cf. Section 13 of Andersson and Klein (2010). However, the intrinsic representation may not be an ADG. For instance, the intrinsic representation of the DUG (7) is the AMG in Figure 1.d.

3.2. The Fundamental Decompositions. Let $\mathcal{U}$ be a DUG with vertex set V, and let $\mathcal{V}$ be a representation of $\mathcal{U}$ as an AMG. For $S \in \mathbf{S}(V)$ and $B \in V/\sim$, define $S_{[B]}$, $S_{\langle B\rangle}$, $S_{[B\rangle}$, and $S_{\langle B]}$ to be the $[B]\times[B]$, $\langle B\rangle\times\langle B\rangle$, $[B]\times\langle B\rangle$, and $\langle B\rangle\times[B]$ submatrices of S, respectively. Because the subgraph of $\mathcal{V}$ induced by $[B]\cup\langle B\rangle$ has a complete skeleton, $\tilde{S}_{[B]} = S_{[B]}$, $\tilde{S}_{\langle B\rangle} = S_{\langle B\rangle}$, $\tilde{S}_{[B\rangle} = S_{[B\rangle}$, and $\tilde{S}_{\langle B]} = S_{\langle B]}$, for any $S \in \mathbf{P}(\mathcal{U})$.

Let $M \in V/\sim$ be a maximal box, and let $\mathcal{V}_M$ be the AMG induced by the subset $V_M := V \setminus [M]$. Then we have $V_M/\sim\, = (V/\sim) \setminus \{M\}$. Furthermore, $[B]$, $\langle B\rangle$, $S_{[B]}$, $S_{\langle B\rangle}$, and $S_{[B\rangle}$ remain unchanged when $\mathcal{V}$ is replaced by $\mathcal{V}_M$, for $B \in \mathcal{V}_M/\sim$. The skeleton $\mathcal{U}_M$ of $\mathcal{V}_M$ is the same as the subgraph of $\mathcal{U}$ induced by V_M.

Let $\mathbf{D}^+(\mathcal{V})$ denote the convex cone of all $V\times V$ block diagonal matrices $E = \mathrm{Diag}(E_B \mid B \in V/\sim)$, such that $E_B \in \mathbf{PD}([B])$, for all $B \in V/\sim$, and let $\mathbf{T}_l^1(\mathcal{V})$ denote the set of all $V \times V$ matrices U such that

- $U_{vv} = 1$ for $v \in V$, and
- $U_{uv} = 0$ if $u \neq v$ and $(u,v) \notin \cup([B]\times\langle B\rangle \mid B \in V/\sim)$.

If the boxes are numbered $B_1,\ldots,B_{V/\sim}$ so that $B_i \to B_j$ implies $i<j$, then the matrices $U \in \mathbf{T}_l^1(\mathcal{V})$ are lower block-triangular matrices with identity matrices on the diagonal and possible extra zeroes below the diagonal.

With these definitions in place, we are now prepared to present several decompositions that are fundamental to the theory of generalized Riesz distributions.

Proposition 3.2. The following mapping is bijective.

$$\begin{array}{rcl} \mathbf{T}_l^1(\mathcal{V})\times\mathbf{D}^+(\mathcal{V}) & \to & \mathbf{PD}^0(\mathcal{U}) \\ (U,E) & \mapsto & U^tEU \end{array}$$

Corollary 3.3. The following mapping is bijective.

$$\begin{array}{rcl} \mathbf{T}_l^1(\mathcal{V})\times\mathbf{D}^+(\mathcal{V}) & \to & \mathbf{P}(\mathcal{U}) \\ (U,D) & \mapsto & p(U^{-1}D(U^t)^{-1}) \end{array}$$

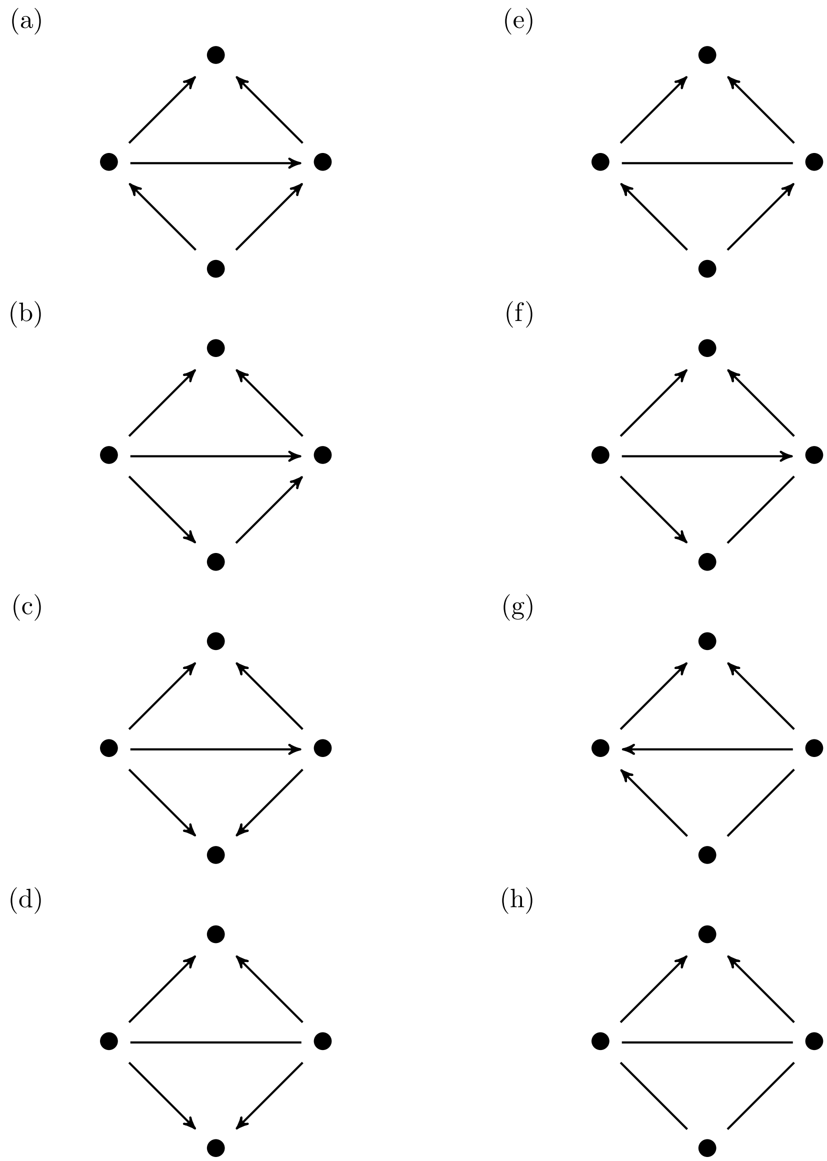


FIGURE 1. The eight representations of the DUG given in (7), ignoring labeling of the vertices. Figure 1.d shows the intrinsic representation.

For $\Delta \in \mathbf{PD}^0(\mathcal{U})$, define the matrices $\Delta_{[B]^\circ} \in \mathbf{PD}([B])$, $B \in V/\sim$, through $\mathrm{Diag}(\Delta_{[B]^\circ} \,\big|\, B \in V/\sim) := E$, where $(U, E) \in \mathbf{T}_l^1(\mathcal{V}) \times \mathbf{D}^+(\mathcal{V})$ such that $U^t E U = \Delta$.

Corollary 3.4. Let $M \in V/\sim$ be a maximal box in $\mathcal{V}$. Then, the mapping

$$\begin{aligned}
&\mathbf{PD}^0(\mathcal{U}_M)\times\mathbb{R}^{[M]\times\langle M\rangle}\times\mathbf{PD}([M])\to\mathbf{PD}^0(\mathcal{U}),\\
&(\Delta_M,\Pi_M,\Upsilon_M)\mapsto\begin{pmatrix}1_{V_M} & -\Pi^t_{M0}\\ 0 & 1_{[M]}\end{pmatrix}\begin{pmatrix}\Delta_M & 0\\ 0 & \Upsilon_M\end{pmatrix}\begin{pmatrix}1_{V_M} & 0\\ -\Pi_{M0} & 1_{[M]}\end{pmatrix}\\
&=\begin{pmatrix}\Delta_M+\Pi^t_{M0}\Upsilon_M\Pi_{M0} & -\Pi^t_{M0}\Upsilon_M\\ -\Upsilon_M\Pi_{M0} & \Upsilon_M\end{pmatrix}=:\Delta, \qquad (8)
\end{aligned}$$

is a a bijection, where $\Pi_{M0}\in\mathbb{R}^{[M]\times V_M}$ is defined by $(\Pi_{M0})_{[M]\times\langle M\rangle}=\Pi_M$, and $(\Pi_{M0})_{[M]\times(V_M\setminus\langle M\rangle)}=0$.

Proposition 3.5. Let $M\in V/\sim$ be a maximal box in $\mathcal{V}$. Then the mapping

$$\begin{aligned}
&\mathbf{P}(\mathcal{U}_M)\times\mathbb{R}^{[M]\times\langle M\rangle}\times\mathbf{PD}([M])\to\mathbf{P}(\mathcal{U}),\\
&(S_M,R_M,L_M)\mapsto p\left(\begin{pmatrix}1_{V_M} & 0\\ R_{M0} & 1_{[M]}\end{pmatrix}\begin{pmatrix}\tilde S_M & 0\\ 0 & L_M\end{pmatrix}\begin{pmatrix}1_{V_M} & R^t_{M0}\\ 0 & 1_{[M]}\end{pmatrix}\right)\\
&=p\left(\begin{pmatrix}\tilde S_M & \tilde S_M R^t_{M0}\\ R_{M0}\tilde S_M & L_M+R_{M0}\tilde S_M R^t_{M0}\end{pmatrix}\right)
\end{aligned}$$

is a a bijection, where $R_{M0}\in\mathbb{R}^{[M]\times V_M}$ is defined by $(R_{M0})_{[M]\times\langle M\rangle}=R_M$, and $(R_{M0})_{[M]\times(V_M\setminus\langle M\rangle)}=0$. The inverse mapping is

$$\begin{aligned}
&\mathbf{P}(\mathcal{U})\to\mathbf{P}(\mathcal{U}_M)\times\mathbb{R}^{[M]\times\langle M\rangle}\times\mathbf{PD}([M])\\
&S\mapsto(S_{V_M},S_{[M\rangle\bullet},S_{[M]\bullet}),
\end{aligned}$$

where $S_{[M\rangle\bullet}:=S_{[M\rangle}S^{-1}_{\langle M\rangle}$, and $S_{[M]\bullet}:=S_{[M]}-S_{[M\rangle}S^{-1}_{\langle M\rangle}S_{\langle M]}$.

In general, if $\Sigma\in\mathbf{P}(\mathcal{U})$, and $B\in V/\sim$, define $\Sigma_{[B\rangle\bullet}:=\Sigma_{[B\rangle}\Sigma^{-1}_{\langle B\rangle}$, and $\Sigma_{[B]\bullet}:=\Sigma_{[B]}-\Sigma_{[B\rangle}\Sigma^{-1}_{\langle B\rangle}\Sigma_{\langle B]}$.[5]

Corollary 3.6. Let $S\in\mathbf{P}(\mathcal{U})$. Setting $D:=\mathrm{Diag}(S_{[B]\bullet}\mid B\in V/\sim)$ and $U_{[B]\times\langle B\rangle}:=-S_{[B\rangle\bullet}$, $B\in V/\sim$, yields the unique solution $(U,D)\in\mathbf{T}^1_l(\mathcal{V})\times\mathbf{D}^+(\mathcal{V})$ to the equation $S=p(U^{-1}D(U^t)^{-1})$, cf. Corollary 3.3.

3.3. Generalized Riesz Distributions. As before, let $\mathcal{U}$ be a DUG with vertex set V, and let $\mathcal{V}$ be a representation of $\mathcal{U}$ as an AMG. Let $\Delta\in\mathbf{PD}^0(\mathcal{U})$, and $\lambda=(\lambda_B\mid B\in V/\sim)\in\mathbb{R}^{V/\sim}$. Define the integral

$$J_{\mathcal{V}}(\Delta,\lambda):=\int_{\mathbf{P}(\mathcal{U})}\prod(|S_{[B]\bullet}|^{\lambda_B}\mid B\in V/\sim)\exp\{-\mathrm{tr}(\Delta S)\}d\nu_{\mathcal{V}}(S),$$

with respect to the measure

$$\mathrm{d}\nu_{\mathcal{V}}(S):=\prod\left(|S_{[B]\bullet}|^{-\frac{[B]+\langle B\rangle+1}{2}}|S_{\langle B\rangle}|^{-\frac{[B]}{2}}\,\middle|\,B\in V/\sim\right)\mathrm{d}S.$$

This integral converges if and only if

$$\lambda_B>\frac{[B]+\langle B\rangle-1}{2},\text{ for all }B\in V/\sim, \qquad (9)$$

In this case,

$$J_{\mathcal{V}}(\Delta,\lambda)=c_{\mathcal{V}}(\lambda)\prod(|\Delta_{[B]\circ}|^{-\lambda_B}\mid B\in V/\sim),$$

[5] If $\langle B\rangle=\emptyset$, then $\Sigma_{[B]\bullet}=\Sigma_{[B]}$.

where

$$c_{\mathcal{V}}(\lambda) := \pi^{\frac{\dim(\mathbf{P}(\mathcal{U}))-V}{2}} \prod\left(\prod\left(\Gamma(\lambda_B - \tfrac{\langle B\rangle}{2} - \tfrac{i-1}{2}) \,\middle|\, i = 1, \ldots, [B]\right) \,\middle|\, B \in V/\sim\right).$$

These calculations lead to the following definition.

Definition 3.7. Let $\Delta \in \mathbf{PD}^0(\mathcal{U})$, and let $\lambda = (\lambda_B \mid B \in V/\sim)$ satisfy condition (9). Then the probability measure

$$\mathrm{dR}_{\Delta,\lambda}(S) := \frac{\pi^{\frac{V-\dim(\mathbf{P}(\mathcal{U}))}{2}} \prod(|\Delta_{[B]\circ}|^{\lambda_B} | B \in V/\sim) \prod(|S_{[B]\bullet}|^{\lambda_B} | B \in V/\sim)}{\prod\left(\prod\left(\Gamma(\lambda_B - \frac{\langle B\rangle}{2} - \frac{i-1}{2}) \,\middle|\, i = 1, \ldots, [B]\right) \,\middle|\, B \in V/\sim\right)} \cdot \exp\{-\mathrm{tr}(\Delta S)\} \mathrm{d}\nu_{\mathcal{V}}(S)$$

is called the *generalized Riesz distribution* on $\mathbf{P}(\mathcal{U})$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$, with *shape parameter* λ and *natural parameter* Δ.

The natural parameters and expectation parameters for generalized Riesz distributions are related by λ-inverse mappings, defined as follows. Let

$$\Sigma = p(U^{-1}\mathrm{Diag}(\Sigma_{[B]\bullet} \mid B \in V/\sim)(U^t)^{-1}) \tag{10}$$

be the decomposition of $\Sigma \in \mathbf{P}(\mathcal{U})$ as described in Corollary 3.6. For $\lambda \in \mathbb{R}_+^{V/\sim}$, the λ-inverse of Σ is defined by

$$\Sigma^{-\lambda} := U^t \mathrm{Diag}(\lambda_B \Sigma_{[B]\bullet}^{-1} \mid B \in V/\sim) U \in \mathbf{PD}^0(\mathcal{U}). \tag{11}$$

The mapping

$$\begin{array}{ccc} \mathbf{P}(\mathcal{U}) & \to & \mathbf{PD}^0(\mathcal{U}) \\ \Sigma & \mapsto & \Sigma^{-\lambda} \end{array}$$

is a bijection with inverse

$$\begin{array}{rcl} \mathbf{PD}^0(\mathcal{U}) & \to & \mathbf{P}(\mathcal{U}) \\ U^t\mathrm{Diag}(\Delta_{[B]\circ} \mid B \in V/\sim)U = \Delta & \mapsto & p(U^{-1}\mathrm{Diag}(\lambda_B \Delta_{[B]\circ}^{-1} \mid B \in V/\sim)(U^t)^{-1}) \end{array}$$

Note that

$$(\Sigma^{-\lambda})_{[B]\circ} = \lambda_B \Sigma_{[B]\bullet}^{-1}, \text{ for } B \in V/\sim . \tag{12}$$

Remark 3.8. If $\lambda = (\lambda_0 \mid B \in \mathcal{V}/\sim)$ for some constant

$$\lambda_0 > \max\left\{ \frac{[B] + \langle B\rangle - 1}{2} \,\middle|\, B \in \mathcal{V}/\sim \right\},$$

then the distribution $\mathrm{R}_{\Delta,\lambda}$ does not depend on the representation $\mathcal{V}$, cf. Remark 8.1 of Andersson and Klein (2010). Also, $\Sigma^{-\lambda} = \lambda_0 \Sigma^{-1}$ in this case.

Proposition 3.9. Let $\Delta \in \mathbf{PD}^0(\mathcal{U})$, let $\lambda = (\lambda_B \mid B \in V/\sim)$ satisfy (9), and let $\Sigma \in \mathbf{P}(\mathcal{U})$ satisfy $\Delta = \Sigma^{-\lambda}$. Then $\mathbb{E}(\mathrm{R}_{\Delta,\lambda}) = \Sigma$.

Like Wishart distributions, generalized Riesz distributions can be parameterized by their expectation parameter, leading to the following definition.

Definition 3.10. Let $\Sigma \in \mathbf{P}(\mathcal{U})$ and $\lambda = (\lambda_B \mid B \in V/\sim)$ satisfy (9). The generalized Riesz distribution on $\mathbf{P}(\mathcal{U})$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ with shape parameter λ and expectation parameter Σ is denoted by $\mathbb{R}_{\Sigma,\lambda}$.

Corollary 3.11. The maximum likelihood estimator for the statistical model

$$M : (\mathrm{R}_{\Delta,\lambda} \,|\, \Delta \in \mathbf{PD}^0(\mathcal{U}))$$

is $\hat{\Delta}(S) = S^{-\lambda}$.

One more fact about λ-inverse mappings that will be needed to prove Theorem 5.1 is presented in Proposition 3.12.

Proposition 3.12. If $\Sigma \in \mathbf{P}(\mathcal{U})$ and $\lambda \in \mathbb{R}_+^{V/\sim}$, then

$$\mathrm{tr}(\Sigma^{-\lambda}\Sigma) = \sum(\lambda_B[B] \mid B \in V/\sim).^6$$

Proof. Because $\Sigma^{-\lambda} \in \mathbf{S}(\mathcal{U})$, $\mathrm{tr}(\Sigma^{-\lambda}\Sigma) = \mathrm{tr}(\Sigma^{-\lambda}\tilde{\Sigma})$, and the proposition now follows from Equations (10) and (11). □

The following proposition is fundamental for working with generalized Riesz distributions, enabling proofs by induction on the number of boxes in $V/\sim$. Andersson and Klein used this proposition to calculate the expectation of $\mathrm{R}_{\Delta,\lambda}$, and it is essential for calculating the moments of the likelihood ratio statistic for testing equality of natural parameters.

Proposition 3.13. Let $M \in V/\sim$ be a maximal box, and let $S \in \mathbf{P}(\mathcal{U})$ be a random variable with distribution $\mathrm{R}_{\Delta,\lambda}$. Then

- The random variables $S_{[M]\bullet} \in \mathbf{PD}([M])$ and $(S_{[M\rangle\bullet}, S_{V_M}) \in \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{P}(\mathcal{U}_M)$ are independent.[7]
- The random variable $S_{[M]\bullet} \in \mathbf{PD}([M])$ has a classical Wishart distribution $\mathrm{W}_{\Delta_{[M]},\lambda_M-\frac{\langle M\rangle}{2}}$ with shape parameter $\lambda_M - \frac{\langle M\rangle}{2}$ and natural parameter $\Delta_{[M]}$.
- The distribution of $(S_{[M\rangle\bullet}, S_{V_M}) \in \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{P}(\mathcal{U}_M)$ is described as follows: The conditional distribution of $S_{[M\rangle\bullet}$ given S_{V_M} is $\mathrm{N}(\Pi_M, (2\Delta_{[M]} \otimes S_{\langle M\rangle})^{-1})$, the normal distribution on $\mathbb{R}^{[M]\times\langle M\rangle}$ with expectation Π_M (see Equation (8)) and precision matrix $2\Delta_{[M]} \otimes S_{\langle M\rangle}$; in particular, this conditional distribution depends on S_{V_M} only through $S_{\langle M\rangle}$. The distribution of S_{V_M} is the generalized Riesz distribution $\mathrm{R}_{\Delta_M,\lambda_{-M}}$ on $\mathbf{P}(\mathcal{U}_M)$ with natural parameter Δ_M (see Equation (8)) and shape parameter $\lambda_{-M} := (\lambda_B \mid B \in V_M/\sim)$.

Remark 3.14. In the special case that $\mathcal{V}$ has only one box, condition (A1) implies that $\mathcal{U}$ is complete, so $\mathbf{PD}^0(\mathcal{U}) = \mathbf{P}(\mathcal{U}) = \mathbf{PD}(V)$. Therefore, $\Delta \in \mathbf{PD}(V)$, $\lambda \in \mathbb{R}$, and $S_{[M]\bullet} = S$, which takes values in $\mathbf{PD}(V)$. Proposition 3.13 then states that $\mathrm{R}_{\Delta,\lambda} = \mathrm{W}_{\Delta,\lambda}$, that is, generalized Riesz distributions reduce to Wishart distributions when $\mathcal{V}$ has only one box.

---

[6]Here, $\lambda_B[B]$ is λ_B times the cardinality of $[B]$.

[7]For a random variable X, we write $X \in A$ when X takes values in A, not when X is an element of the set A.

Proposition 3.15 summarizes Example 17.1 of Andersson and Klein (2010), showing that a generalized Riesz distribution arises when estimating the covariance matrix of a sample from a Gaussian graphical model defined by $\mathcal{U}$.

Proposition 3.15. Let $\mathcal{U} = (V, F)$ be a DUG, and suppose $(X_\alpha \mid \alpha = 1, \dots, N)$ is a random sample from $\mathrm{N}(0, \Sigma)$, where $\Sigma \in \mathbf{PD}(\mathcal{U})$.

- The MLE $\hat{\Sigma}$ for Σ exists with probability one if and only if N is greater than or equal to the cardinality of every clique in $\mathcal{U}$.
- In that case, if $\mathcal{V}$ is an arbitrary representation of $\mathcal{U}$ as an AMG, then $p(\hat{\Sigma}) \sim \mathbb{R}_{p(\Sigma),\lambda}$, where $\lambda = (\frac{N}{2} \mid B \in \mathcal{V}/\sim)$.[8]

Once again, if $\mathcal{V}$ has only one box, condition (A1) implies that $\mathcal{U}$ is complete, so it induces vacuous Markov properties. More precisely, the covariance matrix $\Sigma \in \mathbf{PD}(\mathcal{U}) = \mathbf{PD}(V)$, so $(X_\alpha \mid \alpha = 1, \dots, N)$ is just an ordinary sample from a centered multivariate normal distribution. Also, the projection mapping p is the identity mapping, and Proposition 3.15 states that

$$\hat{\Sigma} = p(\hat{\Sigma}) \sim \mathbb{R}_{p(\Sigma),\lambda} = \mathbb{W}_{\Sigma,\lambda},$$

where $\lambda = \frac{N}{2}$, as required by the classical theory.

3.4. Homogeneity of Covariance Matrices for Gaussian Graphical Models. This section generalizes the test of homogeneity of covariance matrices discussed in the introduction to the setting where each multivariate normal sample satisfies the Markov properties given by a DUG $\mathcal{U}$. This problem is then shown to be equivalent to testing that several generalized Riesz distributions have the same natural parameter.

Suppose that $(X_\alpha^{(k)} \mid \alpha = 1, \dots, N_k)$ is a random sample from $\mathrm{N}(0, \Sigma_k)$, where $\Sigma_k \in \mathbf{PD}(\mathcal{U})$, for $k = 1, \dots, K$. Each of these normal distributions satisfy the Markov properties given by $\mathcal{U}$, because their covariance matrices are elements of $\mathbf{PD}(\mathcal{U})$.

Furthermore, assume that N_k is greater than or equal to the cardinality of every clique in $\mathcal{U}$, so that the MLE $\hat{\Sigma}_k$ for Σ_k exists with probability one, for $k = 1, \dots, K$. By Proposition 3.15, if $\mathcal{V}$ is an arbitrary representation of $\mathcal{U}$ as an AMG, then $p(\hat{\Sigma}_k) \sim \mathbb{R}_{p(\Sigma_k),\lambda_k}$, where $\lambda_k = (\frac{N_k}{2} \mid B \in \mathcal{V}/\sim)$. Note that the components of λ_k do not depend on the boxes $B \in \mathcal{V}/\sim$.

A natural testing problem to consider in this setting is that these covariance matrices are equal,

$$\mathrm{H}_0^{GM} : \Sigma_1 = \cdots = \Sigma_K, \tag{13}$$

which is a generalization of the classical test of homogeneity of covariance matrices presented in the introduction. Because $p : \mathbf{PD}(\mathcal{U}) \to \mathbf{P}(\mathcal{U})$ is bijective, this is equivalent to testing that the random matrices $p(\hat{\Sigma}_1), \dots, p(\hat{\Sigma}_K)$ have the same expectation parameters. By Proposition 3.9, the natural parameter of $p(\hat{\Sigma}_k)$ is $\Sigma_k^{-\lambda_k} = \frac{N_k}{2}\Sigma_k^{-1}$, cf. Remark 3.8. Similarly, $\frac{N_k}{2} p(\hat{\Sigma}_k)$ has natural parameter Σ_k^{-1}.

Therefore, the original testing problem (13) is equivalent to testing whether the random matrices

[8] The distribution does not depend on the representation $\mathcal{V}$ because the components of λ do not depend on the boxes $B \in \mathcal{V}/\sim$, cf. Remark 3.8.

$$\tfrac{N_1}{2}p(\hat{\Sigma}_1), \ldots, \tfrac{N_K}{2}p(\hat{\Sigma}_K),$$

whose distributions are $\mathrm{R}_{\Sigma_1^{-1},\lambda_1}, \ldots, \mathrm{R}_{\Sigma_K^{-1},\lambda_K}$, have the same natural parameter.

It should be emphasized that the shape parameters in this testing problem are known, determined by the sample sizes $N_1, \ldots, N_K$. Although the shape parameters $\lambda_k = (\frac{N_k}{2} \mid B \in \mathcal{V}/\sim)$ have components that do not depend on the boxes $B \in \mathcal{V}/\sim$, Section 5 considers the more general setting where the λ_k's are arbitrary known shape parameters satisfying condition (9). Shape parameters whose components are not identical result from incomplete observations, as discussed in Example 17.2 of Andersson and Klein (2010).

4. A Preliminary Testing Problem

Testing equality of natural parameters for generalized Riesz distributions is discussed in Section 5. The proof of the main theorem in that section, Theorem 5.3, requires the moments of the likelihood ratio statistic for a different testing problem, which are obtained in this section.

Assume I and N are finite index sets, and let $\Phi_k \in \mathbf{PD}(N)$, $l_k > \frac{I+N-1}{2}$, $\xi_k \in \mathbb{R}^{I\times N}$, and $\delta_k \in \mathbf{PD}(I)$, for $k = 1, \ldots, K$.[9] Based on statistically independent observable random variables $x_k \sim \mathrm{N}(\xi_k, (\delta_k \otimes \Phi_k)^{-1})$ and $s_k \sim \mathrm{W}_{\delta_k, l_k - \frac{N}{2}}$, $k = 1, \ldots, K$, we would like to test the hypothesis

$$\mathrm{H}_0' : \xi_1 = \cdots = \xi_K, \delta_1 = \cdots = \delta_K$$

versus the hypothesis H' that the ξ_k's and δ_k's are arbitrary elements of $\mathbb{R}^{I\times N}$ and $\mathbf{PD}(I)$, respectively. The observation space for both H_0' and H' is $(\mathbb{R}^{I\times N} \times \mathbf{PD}(I))^K$, which is also the parameter space for H'. The parameter space for H_0' is $\mathbb{R}^{I\times N} \times \mathbf{PD}(I)$.

Using the notation for statistical models introduced in Section 2, the testing problem is

$$\mathrm{H}_0' : \left(\bigotimes_{k=1}^{K} \mathrm{N}(\xi, (\delta \otimes \Phi_k)^{-1}) \otimes \mathrm{W}_{\delta, l_k - \frac{N}{2}} \;\middle|\; \xi \in \mathbb{R}^{I\times N}, \delta \in \mathbf{PD}(I) \right) \text{ vs.}$$

$$\mathrm{H}' : \left(\bigotimes_{k=1}^{K} \mathrm{N}(\xi_k, (\delta_k \otimes \Phi_k)^{-1}) \otimes \mathrm{W}_{\delta_k, l_k - \frac{N}{2}} \;\middle|\; \xi_k \in \mathbb{R}^{I\times N}, \delta_k \in \mathbf{PD}(I), k = 1, \ldots, K \right),$$

where $\otimes$ denotes product measure.

Our goal for this section is to find the moments of the likelihood ratio statistic q' for testing H_0' vs. H'. We begin by finding the maximum likelihood estimators under H_0' and H', which requires the following lemma.

Lemma 4.1. Let $x \in \mathbb{R}^{I\times N}$, $s \in \mathbf{PD}(I)$, $\Phi \in \mathbf{PD}(N)$, and $l > \frac{I-1}{2}$. Assume that Y is a subspace of $\mathbb{R}^{1\times N}$, so that Y^I is a subspace of $\mathbb{R}^{I\times N}$, and let $P_Y \in \mathbb{R}^{N\times N}$ be the matrix for the Φ-orthogonal projection onto Y.[10] The maximum value of the function

---

[9] The parameters Φ_k and l_k are assumed to be known, and ξ_k and δ_k are unknown.

[10] The elements of $\mathbb{R}^{1\times N}$ are row vectors indexed by N, and for any $u \in \mathbb{R}^{1\times N}$, the projection of u onto Y is uP_Y. Vectors $u, v \in \mathbb{R}^{1\times N}$ are Φ-orthogonal if $u\Phi v^t = 0$.

$$f : Y^I \times \mathbf{PD}(I) \to \mathbb{R}_+$$
$$f(\xi, \delta) = \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\delta(x-\xi)\Phi(x-\xi)^t\right)\right\} |\delta|^l \exp\{-\mathrm{tr}(\delta s)\}$$

is attained when $\xi = xP_Y$ and $\delta = l(\frac{1}{2}(x-xP_Y)\Phi(x-xP_Y)^t+s)^{-1}$. The maximum value is $l^{lI}|\frac{1}{2}(x-xP_Y)\Phi(x-xP_Y)^t+s|^{-l}\exp\{-lI\}$.

Proof. For any $\xi \in Y^I$ and $\delta \in \mathbf{PD}(I)$,

$$f(\xi,\delta) = \exp\{-\frac{1}{2}\mathrm{tr}(\delta(xP_Y-\xi)\Phi(xP_Y-\xi)^t)\}$$
$$\cdot|\delta|^l \exp\left\{-\mathrm{tr}\left(\delta\left(\frac{1}{2}(x-xP_Y)\Phi(x-xP_Y)^t+s\right)\right)\right\}.$$

Because $\delta \otimes \Phi$ is positive definite, the maximum value of the first factor is attained when $\xi = xP_Y$, regardless of the value of $\delta \in \mathbf{PD}(I)$. By Corollary 3.11 applied to the classical Wishart distribution, the second factor is maximized when $\delta = l(\frac{1}{2}(x-xP_Y)\Phi(x-xP_Y)^t+s)^{-1}$. □

Theorem 4.2. The maximum likelihood estimators for H' are $\hat{\xi}_k = x_k$ and $\hat{\delta}_k = l_k s_k^{-1}$, $k = 1, \ldots, K$. The MLEs for H'_0 are

$$\hat{\xi} = \sum_{k=1}^K x_k\Phi_k\Phi^{-1}, \text{ and}$$
$$\hat{\delta} = l\left(\frac{1}{2}\sum_{k=1}^K x_k\Phi_k x_k^t - \frac{1}{2}\left(\sum_{k=1}^K x_k\Phi_k\right)\Phi^{-1}\left(\sum_{k=1}^K x_k\Phi_k\right)^t + s\right)^{-1},$$

where $\Phi = \sum_{k=1}^K \Phi_k$, $s = \sum_{k=1}^K s_k$, and $l = \sum_{k=1}^K l_k$.

Proof. For fixed observations, the likelihood function for H' is proportional to

$$\prod_{k=1}^K \exp\left\{-\frac{1}{2}\mathrm{tr}(\delta_k(x_k-\xi_k)\Phi_k(x_k-\xi_k)^t)\right\}|\delta_k|^{l_k}\exp\{-\mathrm{tr}(\delta_k s_k)\}.$$

Applying Lemma 4.1 to the kth factor with $Y = \mathbb{R}^{1\times N}$ and $P_Y = 1_N$ immediately yields $\hat{\xi}_k = x_k$ and $\hat{\delta}_k = l_k s_k^{-1}$.[11]

Under the null hypothesis H'_0, $\xi := \xi_1 = \cdots = \xi_K$, and $\delta := \delta_1 = \cdots = \delta_K$. Let $x = (x_1, \ldots, x_K) \in \mathbb{R}^{I\times NK}$, $\tilde{\xi} = (\xi, \ldots, \xi) \in \mathbb{R}^{I\times NK}$, and $\tilde{\Phi} = \mathrm{Diag}(\Phi_1, \ldots, \Phi_K)$. For fixed observations, the likelihood function for H'_0 is proportional to

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\delta(x-\tilde{\xi})\tilde{\Phi}(x-\tilde{\xi})^t)\right\}|\delta|^l\exp\{-\mathrm{tr}(\delta s)\}.$$

Lemma 4.1 will now be applied to find the MLE's for H'_0. In this case, $\tilde{\xi}$ can be any element of the subspace $Y^I \subseteq \mathbb{R}^{I\times NK}$, where $Y = \{(\beta, \ldots, \beta) \,|\, \beta \in \mathbb{R}^{1\times N}\}$. Because Y is a regression subspace with design matrix $D = (1_N, \ldots, 1_N) \in \mathbb{R}^{N\times NK}$, the projection matrix is

$$P_Y = \tilde{\Phi}D^t(D\tilde{\Phi}D^t)^{-1}D.$$

[11] 1_N is the $N \times N$ identity matrix.

Therefore,

$$\hat{\xi} = x\tilde{\Phi}D^t(D\tilde{\Phi}D^t)^{-1} = \sum_{k=1}^{K} x_k\Phi_k\Phi^{-1},$$

and the expression for $\hat{\delta}$ is obtained easily thereafter. □

Continuing the proof of Theorem 4.2, a final application of Lemma 4.1 leads to an expression for the likelihood ratio statistic.

Theorem 4.3. The likelihood ratio statistic for testing H_0' vs. H' is

$$q' = \frac{l^{lI}\prod_{k=1}^{K} l_k^{-l_k I}|s_k|^{l_k}}{\left|\frac{1}{2}\sum_{k=1}^{K} x_k\Phi_k x_k^t - \frac{1}{2}\left(\sum_{k=1}^{K} x_k\Phi_k\right)\Phi^{-1}\left(\sum_{k=1}^{K} x_k\Phi_k\right)^t + s\right|^l} = \frac{|\hat{\delta}|^l}{\prod_{k=1}^{K}|\hat{\delta}_k|^{l_k}}. \tag{14}$$

Now, an invariance argument based on the following lemma is used to show that $\hat{\delta}$ and q' are independent under H_0', which will allow the central moments of q' to be determined afterwards.

Lemma 4.4. Let G be a locally compact group, which acts properly on a locally compact space X and properly and transitively on a locally compact space Y. Let furthermore $t : X \to Y$ be a continuous equivariant[12] map, and let $\pi : X \to X/G$ denote the orbit projection. Then the map (t, π) is proper. If ν is an invariant measure on X, and ν_0 is an invariant measure on Y, then there exists a unique measure κ on the locally compact space X/G such that $(t, \pi)(\nu) = \nu_0 \otimes \kappa$.

Proof. This is Lemma 3 of Andersson *et al.* (1983). □

Theorem 4.5. Under the null hypothesis H_0', $\hat{\xi}$, $\hat{\delta}$, and q' are independent, and

- $\hat{\xi} \sim \mathrm{N}(\xi, (\delta \otimes \Phi)^{-1})$
- $\hat{\delta}^{-1} \sim \mathrm{W}_{l\delta, l-\frac{N}{2}}$.

Proof. We will begin with a group invariance argument based on Lemma 4.4. Let $X = \Theta = (\mathbb{R}^{I\times N} \times \mathbf{PD}(I))^K$, and define $\Theta_0 = \mathbb{R}^{I\times N} \times \mathbf{PD}(I)$. Let $G = \mathbb{R}^{I\times N} \times \mathbf{GL}(I)$, and define a group operation on G by $(\eta_1, A_1)(\eta_2, A_2) = (\eta_1 + A_1\eta_2, A_1A_2)$. The model H' is invariant under the following actions of G on its observation space and parameter space.

$$G \times X \to X$$
$$[(\eta, A), (x_k, s_k \mid k = 1, \ldots, K)] \mapsto (\eta + Ax_k, As_kA^t \mid k = 1, \ldots, K)$$
$$G \times \Theta \to \Theta$$
$$[(\eta, A), (\xi_k, \delta_k \mid k = 1, \ldots, K)] \mapsto (\eta + A\xi_k, (A^{-1})^t\delta_kA^{-1} \mid k = 1, \ldots, K).$$

The submodel H_0' is invariant under the same action on its observation space and the following action on its parameter space

$$G \times \Theta_0 \to \Theta_0$$

[12] Commuting with the action of G.

$$[(\eta, A), (\xi, \delta)] \mapsto (\eta + A\xi, (A^{-1})^t \delta A^{-1}).$$

The testing problem H_0' vs. H' is also invariant under these actions. More specifically, the embedding map

$$\begin{aligned}\Theta_0 &\to \Theta \\ (\xi, \delta) &\mapsto (\xi, \delta \mid k = 1, \ldots, K)\end{aligned}$$

is equivariant, and G acts transitively on Θ_0. This follows from the fact that $(A, \delta) \mapsto (A^{-1})^t \delta A^{-1}$ is a transitive action of $\mathbf{GL}(I)$ on $\mathbf{PD}(I)$. This action is also proper, and one easily deduces that all of the above actions of G are proper. For $A \in \mathbf{GL}(I)$, the Jacobian of the mapping

$$\begin{array}{rcl}\mathbb{R}^{I \times N} & \to & \mathbb{R}^{I \times N} \\ x & \mapsto & \eta + Ax\end{array}$$

is $|A|^N$, and the Jacobian of the mapping

$$\begin{array}{rcl}\mathbf{PD}(I) & \to & \mathbf{PD}(I) \\ S & \mapsto & ASA^t\end{array}$$

is $|A|^{I+1}$. Therefore, the measure

$$\mathrm{d}\nu(x_1, s_1, \ldots, x_K, s_K) = \prod_{k=1}^{K} |s_k|^{-\frac{I+N+1}{2}} \mathrm{d}x_k \mathrm{d}s_k$$

is invariant under the action of G on X. The probability measures defined by H' have continuous densities with respect to ν, and ν assigns finite positive mass to any compact neighborhood. Therefore, invariance of the testing problem implies equivariance of the maximum likelihood estimators and invariance of the likelihood ratio statistic, q'.

Note that X, Θ_0, and G are locally compact Hausdorff spaces, and let $\pi : X \to X/G$ be the orbit projection. Applying Lemma 4.4 with $Y = \Theta_0$ and $t = (\hat{\xi}, \hat{\delta})$, there exists a measure κ on X/G such that

$$(t, \pi)(\nu) = \nu_0 \otimes \kappa,$$

where $\mathrm{d}\nu_0(\xi, \delta) = |\delta|^{-\frac{I-N+1}{2}} \mathrm{d}\xi \mathrm{d}\delta$ is an invariant measure on Θ_0.

Now, we will find the distribution of $(\hat{\xi}, \hat{\delta}, \pi)$ under H_0' through a change of variables. Using the notation from the proof of Theorem 4.2, for suitable constants c_1 and c_2, we have

$$\begin{aligned}&\mathrm{d}\left[\bigotimes_{k=1}^{K} \mathrm{N}(\xi, (\delta \otimes \Phi_k)^{-1}) \otimes \mathrm{W}_{\delta, l_k - \frac{N}{2}}\right](x_1, s_1, \ldots, x_K, s_K) \\ =& c_1 \exp\left\{-\mathrm{tr}\left(\delta\left(\frac{1}{2}(x - xP_Y)\tilde{\Phi}(x - xP_Y)^t + s\right)\right)\right. \\ &\left. -\frac{1}{2}\mathrm{tr}(\delta(xP_Y - \tilde{\xi})\tilde{\Phi}(xP_Y - \tilde{\xi})^t)\right\} \prod_{k=1}^{K} |s_k|^{l_k - \frac{I+N+1}{2}} \mathrm{d}x_k \mathrm{d}s_k \\ =& c_2 \exp\left\{-\mathrm{tr}(l\delta\hat{\delta}^{-1}) - \frac{1}{2}\mathrm{tr}(\delta(\hat{\xi} - \xi)\Phi(\hat{\xi} - \xi)^t)\right\} q' |\hat{\delta}|^{-l} \mathrm{d}\nu.\end{aligned}$$

Because q' is invariant, there exists a function $q'_\pi : X/G \to (0,1]$ such that $q' = q'_\pi \circ \pi$. Transforming the above measure by the mapping (t, π) yields the central distribution of $(\hat{\xi}, \hat{\delta}, \pi)$, which is the following measure on $\Theta_0 \times X/G$.

$$\begin{aligned} d\mu(\xi_0, \delta_0, \pi_0) =& c_2 \exp\left\{ -\mathrm{tr}(l\delta\delta_0^{-1}) - \frac{1}{2}\mathrm{tr}(\delta(\xi_0 - \xi)\Phi(\xi_0 - \xi)^t) \right\} \\ & \cdot q'_\pi(\pi_0)|\delta_0|^{-l} \mathrm{d}(\nu_0 \otimes \kappa)(\xi_0, \delta_0, \pi_0) \\ =& c_2 \exp\left\{ -\mathrm{tr}(l\delta\delta_0^{-1}) - \frac{1}{2}\mathrm{tr}(\delta(\xi_0 - \xi)\Phi(\xi_0 - \xi)^t) \right\} \\ & \cdot q'_\pi(\pi_0)|\delta_0|^{-l-\frac{I-N+1}{2}} \mathrm{d}\xi_0 \mathrm{d}\delta_0 \mathrm{d}\kappa(\pi_0), \end{aligned}$$

proving that the maximum likelihood estimators are independent of the orbit projection under the null hypothesis. Since the likelihood ratio statistic is a function of the orbit projection, it is also independent of the maximum likelihood estimators under the null hypothesis.

Finally, we transform the above measure by the map $(\xi_0, \delta_0, \pi_0) \mapsto (\xi_0, \delta_0^{-1}, \pi_0)$, whose Jacobian is $|\delta_0|^{-(I+1)}$, obtaining

$$\begin{aligned} & c_2 |\delta_0^{-1}|^{l-\frac{N}{2}-\frac{I+1}{2}} \exp\{-\mathrm{tr}(l\delta\delta_0^{-1})\} \\ & \cdot \exp\left\{ -\frac{1}{2}\mathrm{tr}(\delta(\xi_0 - \xi)\Phi(\xi_0 - \xi)^t) \right\} q'_\pi(\pi_0) d\xi_0 d\delta_0^{-1} d\kappa(\pi_0). \end{aligned}$$

Therefore, $(\hat{\xi}, \hat{\delta}^{-1}) \sim \mathrm{N}(\xi, (\delta \otimes \Phi)^{-1}) \otimes \mathrm{W}_{l\delta, l-\frac{N}{2}}$ under H'_0. □

We are now prepared to find the central moments of q'. These moments and those obtained in Section 5 will both be expressed in terms of the following constants.

Definition 4.6. Suppose I and N are nonnegative integers, and let $l > \frac{I+N-1}{2}$. Define the constant

$$c(\alpha, l, I, N) = l^{-\alpha l I} \prod \left[\frac{\Gamma(\alpha l + l - \frac{N+i-1}{2})}{\Gamma(l - \frac{N+i-1}{2})} \,\middle|\, i = 1, \ldots, I \right].$$

Theorem 4.7. For any $\alpha > 0$, the αth moment of q' under H'_0 is

$$\mathbb{E}[(q')^\alpha] = \frac{\prod_{k=1}^K c(\alpha, l_k, I, N)}{c(\alpha, l, I, N)},$$

where $l = l_1 + \cdots + l_K$. In particular, the central moments of q' do not depend on the unknown parameter, so q' is ancillary under H'_0.

Proof. Assume that H'_0 is true. The random variables $\hat{\delta}_k$, $k = 1, \ldots, K$ are independent, and by Theorem 4.5, $\hat{\delta}$ and q' are independent. Therefore, equation (14) yields

$$\mathbb{E}[|\hat{\delta}^{-1}|^{\alpha l}] \mathbb{E}[(q')^\alpha] = \prod_{k=1}^K \mathbb{E}[|\hat{\delta}_k^{-1}|^{\alpha l_k}].$$

Since $\hat{\delta}^{-1} \sim \mathrm{W}_{l\delta, l-\frac{N}{2}}$, and $\hat{\delta}_k^{-1} = l_k^{-1} s_k \sim \mathrm{W}_{l_k\delta, l_k-\frac{N}{2}}$, the theorem follows from the moments of the generalized variance, given in equation (3).

□

5. Testing Equality of Natural Parameters for Generalized Riesz Distributions

In this section, we find the moments of the likelihood ratio statistic for testing that several generalized Riesz distributions have the same natural parameter. Let $\mathcal{V}$ be a representation of the DUG $\mathcal{U} = (V, F)$, and suppose $\lambda_k \in \mathbb{R}^{V/\sim}$ satisfies (9), for $k = 1, \ldots, K$. Based on independent observable random variables $S_k \sim \mathrm{R}_{\Delta_k, \lambda_k}$, $k = 1, \ldots, K$, we would like to test the hypothesis

$$\mathrm{H}_0 : \Delta_1 = \cdots = \Delta_K$$

versus the hypothesis H that the Δ_k's are arbitrary elements of $\mathbf{PD}^0(\mathcal{U})$.[13] The observation space for both H_0 and H is $\mathbf{P}(\mathcal{U})^K$, the parameter space for H_0 is $\mathbf{PD}^0(\mathcal{U})$, and the parameter space for H is $\mathbf{PD}^0(\mathcal{U})^K$. In summary, the testing problem is

$$\mathrm{H}_0 : \left(\bigotimes_{k=1}^{K} \mathrm{R}_{\Delta, \lambda_k} \in \mathcal{P}(\mathbf{P}(\mathcal{U})^K) \mid \Delta \in \mathbf{PD}^0(\mathcal{U}) \right) \text{ vs.}$$
$$\mathrm{H} : \left(\bigotimes_{k=1}^{K} \mathrm{R}_{\Delta_k, \lambda_k} \in \mathcal{P}(\mathbf{P}(\mathcal{U})^K) \mid \Delta_k \in \mathbf{PD}^0(\mathcal{U}),\ k = 1, \ldots, K \right).$$

We begin by finding the maximum likelihood estimator for Δ under H_0. Let L_{H_0} and L_M be the likelihood functions for models H_0 and M (see Corollary 3.11), respectively, let $S_1, \ldots, S_K \in \mathbf{P}(\mathcal{U})$, and define $S = S_1 + \cdots + S_K$ and $\lambda = \lambda_1 + \cdots + \lambda_K$. In what follows, $\propto$ denotes equality up to a factor not depending on Δ.

$$\begin{aligned} L_{H_0}(\Delta; S_1, \ldots, S_K, \lambda_1, \ldots, \lambda_K) &= \prod_{k=1}^{K} L_M(\Delta; S_k, \lambda_k) \\ &\propto \prod_{k=1}^{K} J_{\mathcal{V}}(\Delta, \lambda_k)^{-1} \exp\{-\mathrm{tr}(\Delta S_k)\} \\ &\propto J_{\mathcal{V}}(\Delta, \lambda)^{-1} \exp\{-\mathrm{tr}(\Delta S) \\ &\propto L_M(\Delta; S, \lambda). \end{aligned}$$

By Corollary 3.11, the value of Δ maximizing $L_M(\Delta; S, \lambda)$ is $S^{-\lambda}$, so the maximum likelihood estimator for Δ under H_0 is

$$\hat{\Delta} = S^{-\lambda}.$$

Again, by Corollary 3.11, the maximum likelihood estimator for Δ_k under H is

$$\hat{\Delta}_k = S_k^{-\lambda_k}, \text{ for } k = 1, \ldots, K.$$

[13] The shape parameters λ_k are assumed to be known in this testing problem.

Therefore, the likelihood ratio statistic is

$$\begin{aligned} q(S_1,\ldots,S_K) &= \prod_{k=1}^{K} \frac{L_M(\hat{\Delta};S_k,\lambda_k)}{L_M(\hat{\Delta}_k;S_k,\lambda_k)} \\ &= \prod \left(\frac{|\hat{\Delta}_{[B]\circ}|^{\lambda_B}}{\prod_{k=1}^{K} |\hat{\Delta}_{k[B]\circ}|^{\lambda_{kB}}} \,\middle|\, B \in V/\sim \right) \frac{\prod_{k=1}^{K} \exp\{\operatorname{tr}(\hat{\Delta}_k S_k)\}}{\exp\{\operatorname{tr}(\hat{\Delta} S)\}}, \end{aligned}$$

where $\hat{\Delta}_{k[B]\circ} = (\hat{\Delta}_k)_{[B]\circ}$, and $\lambda_{kB} = (\lambda_k)_B$. By Proposition 3.12, the exponentials cancel, and the following theorem has been proved.

Theorem 5.1. The likelihood ratio statistic for testing H_0 vs. H is

$$\begin{aligned} q : \mathbf{P}(\mathcal{U})^K &\to (0,1] \\ q(S_1,\ldots,S_K) &= \prod \left(\frac{|\hat{\Delta}_{[B]\circ}|^{\lambda_B}}{\prod_{k=1}^{K} |\hat{\Delta}_{k[B]\circ}|^{\lambda_{kB}}} \,\middle|\, B \in V/\sim \right) \\ &= \prod \left(\frac{\lambda_B^{\lambda_B [B]}}{|S_{[B]\bullet}|^{\lambda_B}} \prod_{k=1}^{K} \frac{|S_{k[B]\bullet}|^{\lambda_{kB}}}{\lambda_{kB}^{\lambda_{kB}[B]}} \,\middle|\, B \in V/\sim \right), \end{aligned}$$

where $S = S_1 + \cdots + S_K$ and $\lambda = \lambda_1 + \cdots + \lambda_K$.

The central moments of q are found by inducting on the number of boxes in $\mathcal{V}$, so we begin with the case where $\mathcal{V}$ has only one box and then proceed to the general case.

Theorem 5.2. If $\mathcal{V}$ has only one box and $\alpha > 0$, the αth moment of q under H_0 is

$$\mathbb{E}[q^\alpha] = \frac{\prod_{k=1}^{K} c(\alpha, \lambda_{kB}, [B], 0)}{c(\alpha, \lambda_B, [B], 0)},$$

where $B = V$ is the single box in $\mathcal{V}$.

Proof. Assume that H_0 is true. Because $\mathcal{V}$ has only one box,

$$\begin{aligned} S_{k[B]\bullet} = S_k &\sim \mathrm{R}_{\Delta,\lambda_k} = \mathrm{W}_{\Delta,\lambda_{kB}}, k = 1,\ldots,K, \\ S_{[B]\bullet} = S &\sim \mathrm{R}_{\Delta,\lambda} = \mathrm{W}_{\Delta,\lambda_B}, \text{ and} \\ |\lambda_B^{-1} S|^{\lambda_B} q &= \prod_{k=1}^{K} |\lambda_{kB}^{-1} S_k|^{\lambda_{kB}}, \end{aligned}$$

cf. Remark 3.14. It now suffices to show that S and q are independent under H_0, which follows from an invariance argument like the one used to prove Theorem 4.5. □

Theorem 5.3. Given $\alpha > 0$, the αth moment of q under H_0 is

$$\mathbb{E}[q^\alpha] = \prod \left(\frac{\prod_{k=1}^{K} c(\alpha, \lambda_{kB}, [B], \langle B \rangle)}{c(\alpha, \lambda_B, [B], \langle B \rangle)} \,\middle|\, B \in V/\sim \right),$$

where $\lambda = \lambda_1 + \cdots + \lambda_K$.

Proof. We proceed by induction on the number of boxes. If $\mathcal{V}$ has only one box B, $\langle B\rangle = \emptyset$, and the result follows from Theorem 5.2.

On the other hand, assume that $\mathcal{V}$ has at least two boxes and that the theorem has been proved for graphs with fewer boxes. Let M be a maximal box, define $V_M := V \setminus [M]$, and let $\mathcal{U}_M$ be the subgraph of $\mathcal{U}$ induced by V_M. By Corollary 3.4, there exist matrices $\Delta_M \in \mathbf{PD}^0(\mathcal{U}_M)$, $\Delta_{[M]} \in \mathbf{PD}([M])$, and $\Pi_M \in \mathbb{R}^{[M]\times\langle M\rangle}$ such that

$$\Delta = \begin{pmatrix} 1_{V_M} & -\Pi_{M0}^t \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} \Delta_M & 0 \\ 0 & \Delta_{[M]} \end{pmatrix} \begin{pmatrix} 1_{V_M} & 0 \\ -\Pi_{M0} & 1_{[M]} \end{pmatrix}.$$

For $k = 1, \ldots, K$, assume $S_k \sim \mathrm{R}_{\Delta,\lambda_k}$, so that H_0 is true. By Theorem 5.1,

$$q = \prod (q_B \mid B \in V/\sim), \text{ where}$$
$$q_B = \frac{\lambda_B^{\lambda_B[B]}}{|S_{[B]\bullet}|^{\lambda_B}} \prod_{k=1}^{K} \frac{|S_{k[B]\bullet}|^{\lambda_{kB}}}{\lambda_{kB}^{\lambda_{kB}[B]}}.$$

Therefore,

$$q = q_M q_{-M}, \text{ where}$$
$$q_{-M} = \prod (q_B \mid B \in V_M/\sim),$$

and the notation suppresses the dependence on S_k, $k = 1, \ldots, K$. Below, we will show that

$$\mathbb{E}[q_M^\alpha \parallel S_{1V_M}, \ldots, S_{KV_M}] = \frac{\prod_{k=1}^K c(\alpha, \lambda_{kM}, [M], \langle M\rangle)}{c(\alpha, \lambda_M, [M], \langle M\rangle)}.^{14} \tag{15}$$

The other factor q_{-M} depends only on $S_{1V_M}, \ldots, S_{KV_M}$, and it is the likelihood ratio statistic for testing H_0 vs. H when the graph is $\mathcal{U}_M$. By Proposition 3.13, $S_{kV_M} \sim \mathrm{R}_{\Delta_M,(\lambda_k)_{-M}}$, so $S_{1V_M}, \ldots, S_{KV_M}$ satisfy H_0 for the graph $\mathcal{U}_M$. Therefore, by the induction hypothesis,

$$\mathbb{E}[q_{-M}^\alpha] = \prod \left(\frac{\prod_{k=1}^K c(\alpha, \lambda_{kB}, [B], \langle B\rangle)}{c(\alpha, \lambda_B, [B], \langle B\rangle)} \,\middle|\, B \in V_M/\sim \right).$$

Since q_{-M} is a measurable function of the S_{kV_M}'s, and $\mathbb{E}[q_M^\alpha \parallel S_{1V_M}, \ldots, S_{KV_M}]$ is a constant,

$$\begin{aligned}\mathbb{E}[q^\alpha] &= \mathbb{E}[q_{-M}^\alpha]\mathbb{E}[q_M^\alpha \parallel S_{1V_M}, \ldots, S_{KV_M}] \\ &= \prod \left(\frac{\prod_{k=1}^K c(\alpha, \lambda_{kB}, [B], \langle B\rangle)}{c(\alpha, \lambda_B, [B], \langle B\rangle)} \,\middle|\, B \in V/\sim \right).\end{aligned}$$

To prove the theorem, it only remains to establish Equation (15), which follows from the results for the testing problem H_0' vs. H' discussed in Section 4. For $k = 1, \ldots, K$, define $x_k := S_{k[M\rangle\bullet}$ and $s_k := S_{k[M]\bullet}$. By Proposition 3.13, the conditional distribution of $(x_1, s_1, \ldots, x_K, s_K)$ given $S_{1V_M}, \ldots, S_{KV_M}$ is

[14] For random variables X and Y, $\mathbb{E}[X \parallel Y]$ denotes the conditional expectation of X given Y.

$$\bigotimes_{k=1}^{K} \mathrm{N}(\xi, (\delta \otimes \Phi_k)^{-1}) \otimes \mathrm{W}_{\delta, l_k - \frac{N}{2}},$$

where $\xi = \Pi_M$, $\delta = \Delta_{[M]}$, $\Phi_k = 2S_{k\langle M\rangle}$, $l_k = \lambda_{kM}$, and $N = \langle M\rangle$. In particular, the conditional distribution of $(x_1, s_1, \ldots, x_K, s_K)$ given $S_{1V_M}, \ldots, S_{KV_M}$ is a probability measure in the model H_0'. Furthermore, the following calculations show that $q_M(S_1, \ldots, S_K) = q'(x_1, s_1, \ldots, x_K, s_K)$.

$$\frac{1}{2}\sum_{k=1}^{K} x_k \Phi_k x_k^t - \frac{1}{2}\left(\sum_{k=1}^{K} x_k \Phi_k\right)\Phi^{-1}\left(\sum_{k=1}^{K} x_k \Phi_k\right)^t + s =$$

$$\sum_{k=1}^{K} S_{k[M\rangle\bullet} S_{k\langle M\rangle} S_{k[M\rangle\bullet}^t + S_{k[M]\bullet} - \left(\sum_{k=1}^{K} S_{k[M\rangle\bullet} S_{k\langle M\rangle}\right) S_{\langle M\rangle}^{-1} \left(\sum_{k=1}^{K} S_{k[M\rangle\bullet} S_{k\langle M\rangle}\right)^t$$

Noting that $S_{k[M]} = S_{k[M]\bullet} + S_{k[M\rangle\bullet} S_{k\langle M\rangle} S_{k[M\rangle\bullet}^t$, and $S_{k[M\rangle} = S_{k[M\rangle\bullet} S_{k\langle M\rangle}$, the above expression is equal to

$$S_{[M]} - S_{[M\rangle} S_{\langle M\rangle}^{-1} S_{[M\rangle}^t = S_{[M]\bullet}.$$

Because $q_M(S_1, \ldots, S_K) = q'(x_1, s_1, \ldots, x_K, s_K)$, and because the conditional distribution of $(x_1, s_1, \ldots, x_K, s_K)$ given $S_{1V_M}, \ldots, S_{KV_M}$ is a probability measure in the model H_0', the conditional moments of q_M given $S_{1V_M}, \ldots, S_{KV_M}$ are equal to the moments of q' from Theorem 4.7. This establishes Equation (15) and completes the proof.

□

References

ANDERSON, T.W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.). John Wiley and Sons, New Jersey.

ANDERSSON, S.A., BRØNS, H.K., and JENSEN, S.T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics*, **11(2)**, 392–415.

ANDERSSON, S.A. and KLEIN, T. (2010). On Riesz and Wishart distributions associated with decomposable undirected graphs. *Journal of Multivariate Analysis*, **101(3)**, 789–810.

ANDERSSON, S.A., MADIGAN, D., and PERLMAN, M.D. (2001). Alternative Markov properties for chain graphs. *Scandinavian Journal of Statistics*, **28(1)**, 33–85.

BARTLETT, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London*, **160(A)**, 268-282.

BARTLETT, M.S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, **34**, 33-40.

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

BROWN, G.W. (1939). On the power of the L_1 test for equality of several variances. *Ann. Math. Statist.*, **10**, 119-128.

FRYDENBERG, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics*, **17**, 333–353.

HASSAIRI, A. and LAJMI, S. (2001). Riesz exponential families on symmetric cones. *Journal of Theoretical Probability*, **14(4)**, 927-948.

LAURITZEN, S.L. (1996). *Graphical Models.* Clarendon Press, Oxford.

LAURITZEN, S.L. and WERMUTH, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, **17**, 31–57.

NEYMAN, J. and PEARSON, E.S. (1931). On the problem of k samples. *Bull. Acad. Polonaise Sci. Let.*.

PERLMAN, M. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *Ann. Statist.*, **8(2)**, 247-263.

PITMAN, E.J.G. (1939). Tests of hypotheses concerning location and scale parameters. *Biometrika*, **31**, 200-215.

SUGIURA, N. and NAGAO, N. (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *Ann. Math. Statist.*, **39**, 1686-1692.

SUGIURA, N. and NAGAO, N. (1969). On Bartletts test and Lehmanns test for homogeneity of variances. *Ann. Math. Statist.*, **40**, 2018-2032.

WILKS, S.S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24(3/4)**, 471-494.

WISHART, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, **20A(1/2)**, 32-52.

S.A. Andersson
Department of Statistics
Indiana University
309 N. Park Ave
Bloomington, IN 47408

standers@indiana.edu

J.B. Crawford
Department of Mathematics
Tarleton State University
1333 W. Washington
Stephenville, TX 76402

jcrawford@tarleton.edu