Pure Mathematical Sciences, Vol. 1, 2012, no. 3, 137 - 149

Positive Semi-Definite Correlation Matrices: Recursive Algorithmic Generation and Volume Measure

Werner Hürlimann

FRSGlobal Switzerland, Seefeldstrasse 69, CH-8008 Zürich, Switzerland
werner.huerlimann@frsglobal.com, whurlimann@bluewin.ch

Abstract

Based on a notion of d-scaled (determinant scaled) partial correlation and two determinantal identities for correlation matrices, we derive explicit recursive closed form correlation bounds for positive semi-definite correlation matrices. As an application we obtain a new simple product formula for the volume measure of the space of positive-semi definite correlation matrices.

Mathematics Subject Classification: 15A24, 15B99, 62H20, 65F30

Keywords: correlation matrix, partial correlation, positive semi-definite property, determinantal identity, recursive algorithm

1. Introduction

The algorithmic generation of valid correlation matrices has been up to quite recently a challenging problem. A first constructive recursive optimization algorithm for this is described in Budden et al. [1]. An important improvement is [2], Corollary 4.1, who obtains recursive iterative explicit closed form bounds for correlation coefficients. Since they do not require (a priori) any optimization procedure these bounds offer a more straightforward implementation than the mentioned optimization algorithm. Through generalization of the idea expressed in [2], Example 3.1, it is possible to get rid of the previous iterative recursion and hitherto obtain very explicit recursively defined generic closed form correlation bounds (our main Theorem 3.1). The derivation of this result is based on two determinantal identities for correlation matrices, which are presented and proved in Section 2. The first one is a recursive relation for a d-scaled (determinant scaled) partial correlation, which is obtained by transforming the partial correlation appropriately. The second identity is a recursive relation for determinants of correlation matrices in terms of the squared d-scaled partial correlation. As an application we obtain a new simple product formula for the volume measure of the space of positive-semi definite correlation matrices. We note that Joe [3] has obtained previously an equivalent volume formula in a probabilistic way as a by-product of some new methods to generate random correlation matrices. The new proof uses only first-year calculus and is non-probabilistic in nature.

2. Determinantal identities for correlation matrices

For fixed $n \geq 2$ a *nxn correlation matrix* $R=(r_{ij}), 1 \leq i, j \leq n$, i.e. a symmetric matrix with unit diagonal entries, is uniquely determined by the $\frac{1}{2}(n-1)n$ upper diagonal elements $r=(r_{ij}), 1 \leq i < j \leq n$. For each $m \in \{2,...,n\}$ and any index set $s^{(m)}=(s_1,s_2,...,s_m)$ with $1 \leq s_i \leq n, i=1,...,m$, we consider the *mxm* sub-correlation matrix $R^{(m)}=(r_{s_i s_j}), 1 \leq i, j \leq m$, which is uniquely determined by $r^{(m)}=(r_{s_i s_j}), 1 \leq i < j \leq m$. It is convenient to consider own special quantities.

Definitions 2.1 (*Determinant, partial correlation and d-scaled partial correlation*) The *determinant* of the matrix $R^{(m)}$ is denoted by

$$\Delta^m(s^{(m)}) = \Delta^m(s_1,s_2,...,s_m) = \det(R^{(m)}) . \tag{2.1}$$

For $n \geq m \geq 3$ and an index set $s^{(m)}$ the *partial correlation* of (s_1,s_2) with respect to $(s_3,...,s_m)$ is recursively defined and denoted by

$$r_{s_1 s_2;s_3,...,s_m} = \frac{r_{s_1 s_2;s_3,...,s_{m-1}} - r_{s_1 s_m;s_3,...,s_{m-1}} \cdot r_{s_2 s_m;s_3,...,s_{m-1}}}{\sqrt{(1-r^2_{s_1 s_m;s_3,...,s_{m-1}}) \cdot (1-r^2_{s_2 s_m;s_3,...,s_{m-1}})}}, \tag{2.2}$$

where for $m=2$ the quantities used on the right hand side of (2.2) are by convention the correlations $r_{s_1 s_2}, r_{s_1 s_m}, r_{s_2 s_m}$. The transformed partial correlation defined and denoted by

$$N^m(s^{(m)}) = N^m(s_1, s_2; ..., s_m) = r_{s_1 s_2; s_3,...,s_m} \cdot \sqrt{\Delta^{m-1}(s_1, s_3, ..., s_m) \cdot \Delta^{m-1}(s_2, s_3, ..., s_m)} \quad (2.3)$$

is called *d-scaled* (determinant scaled) *partial correlation*. By convention we set $N^2(s_1, s_2) = r_{s_1 s_2}$.

Remark 2.2 Though not relevant in the present context, we note the following stochastic interpretation. For a set of random variables $X_1, X_2, ..., X_n, n \geq 3$, with finite means and variances, the quantity $r_{12;3,...,n}$ is nothing else than the partial (linear) correlation of X_1, X_2 with respect to $X_3, ..., X_n$, which can be interpreted as the correlation between the projections of X_1, X_2 on the plane orthogonal to the space spanned by $X_3, ..., X_n$. This quantity satisfies the defining recursive formula (2.2) (see e.g. Yule and Kendall [6]).

Recall that the determinant of a correlation matrix satisfies the product representation (e.g. [2], formula (2.10), or Kurowicka and Cooke [4], Theorem 4.5, for a more general partial correlation vine specification)

$$\Delta^n(1,2,...,n) = \prod_{i=1}^{n-1}\left(1 - r_{in}^2\right) \cdot \prod_{i=1}^{n-2}\left(1 - r_{in-1;n}^2\right) \cdot \prod_{i=1}^{n-3}\left(1 - r_{in-2;n-1,n}^2\right) \cdot \prod_{k=3}^{n-2}\left\{\prod_{i=1}^{n-k-1}\left(1 - r_{in-k;n-k+1,...,n}^2\right)\right\}, \quad (2.4)$$

where an empty product equals one by convention. Our first goal is the derivation of some useful fundamental recursive identities that link determinants to d-scaled partial correlations. Without loss of generality it suffices to develop formulas for the *canonical index set* $s^{(m)} = (1,2,...,m)$ obtained by setting $s_j = j, j = 1,...,m$. General formulas are obtained by using arbitrary index sets and taking into account the permutation properties of the involved quantities. In the special case $m = 3$ we have the identity

$$N^3(1,2;3) = N^2(1,2) - N^2(1,3) \cdot N^2(2,3). \quad (2.5)$$

Proposition 2.1 (*Recursive relation for d-scaled partial correlations*) Assume that $n \geq m \geq 4$. Then we have the identity

$$\begin{aligned} N^m(1,2;3,...,m) \cdot \Delta^{m-3}(3,...,m-1) &= N^{m-1}(1,2;3,...,m-1) \cdot \Delta^{m-2}(3,...,m-1) \\ &- N^{m-1}(1,m;3,...,m-1) \cdot N^{m-1}(2,m;3,...,m-1), \end{aligned} \quad (2.6)$$

with the convention $\Delta^1(3) = 1$ in case $m = 4$.

Proof This is shown by induction. For $m = 4$ we have by the recursion (2.2) that

$$r_{12;3,4}=\frac{r_{12;3}-r_{14;3}\cdot r_{24;3}}{\sqrt{(1-r_{14;3}^2)\cdot((1-r_{24;3}^2))}}\text{, with}$$

$$r_{i4;3}=\frac{N^3(i,4;3)}{\sqrt{\Delta^2(i,3)\cdot\Delta^2(3,4)}},\quad i=1,2,\quad r_{12;3}=\frac{N^3(1,2;3)}{\sqrt{\Delta^2(1,3)\cdot\Delta^2(2,3)}}.$$

Using these relations and the defining equation (2.3) we have

$$N^4(1,2;3,4)=r_{12;3,4}\cdot\sqrt{\Delta^3(1,3,4)\cdot\Delta^3(2,3,4)}=\left(\sqrt{(1-r_{14;3}^2)\cdot((1-r_{24;3}^2))}\right)^{-1}\cdot$$
$$\frac{N^3(1,2;3)\cdot\Delta^2(3,4)-N^3(1,4;3)\cdot N^3(2,4;3)}{\Delta^2(3,4)\cdot\sqrt{\Delta^2(1,3)\cdot\Delta^2(2,3)}}\cdot\sqrt{\Delta^3(1,3,4)\cdot\Delta^3(2,3,4)}$$

Now, from the version of (2.4) for $n=3$ with index set $s^{(3)}=(s_1,s_2,s_3)=(i,3,4)$ we see that

$$\Delta^3(i,3,4)=\Delta^2(i,3)\cdot\Delta^2(3,4)\cdot(1-r_{i4;3}^2).$$

Inserted into the preceding relation shows the desired result for $m=4$. It remains to show that if the relation (2.6) holds for the dimension m then it holds for the dimension $m+1$. Proceeding similarly we observe that

$$r_{12;3,\ldots,m+1}=\frac{r_{12;3,\ldots,m}-r_{1m+1;3,\ldots,m}\cdot r_{2m+1;3,\ldots,m}}{\sqrt{(1-r_{1m+1;3,\ldots,m}^2)\cdot((1-r_{2m+1;3,\ldots,m}^2))}}\text{, with}$$

$$r_{im+1;3,\ldots,m}=\frac{N^m(i,m+1;3,\ldots,m)}{\sqrt{\Delta^{m-1}(i,3,\ldots,m)\cdot\Delta^{m-1}(3,,\ldots,m+1)}},\quad i=1,2,$$

$$r_{12;3,\ldots,m}=\frac{N^m(1,2;3,\ldots,m)}{\sqrt{\Delta^{m-1}(1,3,\ldots,m)\cdot\Delta^{m-1}(2,3,\ldots,m)}}.$$

Inserting these relations into (2.3) we obtain that $N^{m+1}(1,2;3,\ldots,m+1)$ is equal to

$$\sqrt{\Delta^m(1,3,\ldots,m+1)\cdot\Delta^m(2,3,\ldots,m+1)}\cdot\left(\sqrt{(1-r_{1m+1;3,\ldots,m}^2)\cdot((1-r_{2m+1;3,\ldots,m}^2))}\right)^{-1}\cdot$$
$$\frac{N^m(1,2;3,\ldots,m)\cdot\Delta^{m-1}(3,\ldots,m)-N^m(1,m+1;3,\ldots,m)\cdot N^m(2,m+1;3,\ldots,m)}{\Delta^{m+1}(3,,\ldots,m+1)\cdot\sqrt{\Delta^{m-1}(1,3,\ldots,m)\cdot\Delta^{m-1}(2,3,\ldots,m)}}$$

Now, Proposition 2.2 below remains true when replacing the canonical index set by any other index set. In particular (2.8) is valid for the index set $s^{(m)}=(s_1,s_2,\ldots,s_m)$

with $s_1 = i \in \{1,2\}, s_2 = m+1, s_j = j, j = 3,...,m$, which yields (taking into account (2.3) and the invariance of determinants under permutation of indices)

$$\Delta^m(i,3,...,m+1)\cdot\Delta^{m-2}(3,...,m) = \Delta^{m-1}(i,3,...,m)\cdot\Delta^{m-1}(3,...,m)\cdot(1-r^2_{im+1;3,...,m}), \quad i=1,2 .$$

Inserted into the preceding relation shows (2.6) for the dimension $m+1$. ◊

Through successive iteration of (2.6) we obtain an explicit finite representation.

Corollary 2.1 (*Series expansion for d-scaled partial correlations*) Assume that $n \geq m \geq 4$. Then we have the finite series representation

$$\begin{aligned} & N^m(1,2;3,...,m)/\Delta^{m-2}(3,...,m-1) \\ & = N^3(1,2;3) - \sum_{k=2}^{m-2}\frac{N^{k+1}(1,k+2;3,...,k+1)\cdot N^{k+1}(2,k+1;3,...,k+1)}{\Delta^{k-1}(3,...,k+1)\cdot\Delta^k(3,...,k+2)} \end{aligned} \tag{2.7}$$

<u>Proof</u>. This is left as exercise to the interested reader. ◊

While (2.6) allows recursive evaluation of the d-scaled partial correlations, the next identity performs the same for determinants.

Proposition 2.2 (*Recursive relation for determinants of correlation matrices*) Assume that $n \geq m \geq 3$. Then we have the identity

$$\Delta^m(1,2,3,...,m)\cdot\Delta^{m-2}(3,...,m) = \Delta^{m-1}(1,3,...,m)\cdot\Delta^{m-1}(2,3,...,m) - N^m(1,2;3,...,m)^2 \tag{2.8}$$

with the convention $\Delta^1(3)=1$ in case $m=3$.

Proof The dimension three version of the product representation (2.4) yields

$$\Delta^3(1,2,3) = \Delta^2(1,3)\cdot\Delta^2(2,3)\cdot(1-r^2_{12;3}) .$$

Using (2.3) shows (2.8) for $m=3$. Now, we have to show that if (2.8) holds for the dimension m then it holds for the dimension $m+1$. By induction assumption and with appropriate choice of the index set, we have for $i=1,2$ the relations

$$\Delta^m(i,,3,...,m+1)\cdot\Delta^{m-2}(4,...,m+1) = \Delta^{m-1}(i,4,...,m+1)\cdot\Delta^{m-1}(3,4,...,m+1)\cdot(1-r^2_{i3;4,...,m+1})$$

With this the right hand side of (2.8) for the dimension $m+1$ can be rewritten as

$$\Delta^{m}(1,3,\ldots,m+1)\cdot\Delta^{m}(2,3,\ldots,m+1)\cdot(1-r_{12;3,4,\ldots,m+1}^{2})=$$
$$\Delta^{m-1}(3,\ldots,m+1)\cdot\frac{\prod_{i=1}^{3}\Delta^{m-1}(i,4,\ldots,m+1)}{\Delta^{m-2}(4,\ldots,m+1)^{2}}\cdot\left[\prod_{i=1}^{2}(1-r_{i3;4,\ldots,m+1}^{2})\right]\cdot(1-r_{12;3,4,\ldots,m+1}^{2})$$

On the other hand, from the version of (2.4) with arbitrary index set we obtain

$$\Delta^{m-1}(i,4,\ldots,m+1)=(1-r_{im+1}^{2})\cdot(1-r_{im;m+1}^{2})\cdot\ldots\cdot(1-r_{i4;\ldots,m+1}^{2})\cdot\Delta^{m-2}(4,\ldots,m+1),\quad i=1,2,3,$$
$$\Delta^{m-2}(4,\ldots,m+1)=\prod_{i=4}^{m}(1-r_{im+1}^{2})\cdot\prod_{i=4}^{m-1}(1-r_{im;m+1}^{2})\cdot\prod_{k=2}^{m-4}\left\{\prod_{i=4}^{m-k}\left(1-r_{im-k+1;m-k+2,\ldots,m+1}^{2}\right)\right\},$$

with the convention that an empty product equals one. Inserted into the preceding relation and using again the representation (2.4) we see that this identifies with

$$\Delta^{m-1}(3,\ldots,m+1)\cdot\prod_{i=1}^{m}(1-r_{im+1}^{2})\cdot\prod_{i=1}^{m-1}(1-r_{im;m+1}^{2})\cdot\prod_{k=2}^{m-1}\left\{\prod_{i=4}^{m-k}\left(1-r_{im-k+1;m-k+2,\ldots,m+1}^{2}\right)\right\}$$
$$=\Delta^{m-1}(3,\ldots,m+1)\cdot\Delta^{m+1}(1,2,3,\ldots,m+1),$$

which shows the identity (2.8) for the dimension $m+1$. ◊

3. Algorithmic recursive generation of correlation matrices

We are ready for our main result on positive semi-definite correlation matrices.

Theorem 3.1 (*Recursive generation of valid correlation matrices*) A correlation matrix parameterized by $r=(r_{ij}), 1\le i<j\le n$, is positive semi-definite if, and only if, the following conditions are fulfilled:

$$r_{in}\in[-1,1],\quad i=1,\ldots,n-1,\quad n\ge 2, \tag{3.1}$$

$$r_{in-1}\in\left[{}_{n}B_{in-1}^{-},{}_{n}B_{in-1}^{+}\right],\quad i=1,\ldots,n-2,\quad n\ge 3,$$
$${}_{n}B_{in-1}^{\pm}=N^{2}(i,n)\cdot N^{2}(n-1,n)\pm\sqrt{\Delta^{2}(i,n)\cdot\Delta^{2}(n-1,n)} \tag{3.2}$$

$$r_{in-2}\in\left[{}_{n}B_{in-2}^{-},{}_{n}B_{in-2}^{+}\right],\quad i=1,\ldots,n-3,\quad n\ge 4,$$
$${}_{n}B_{in-2}^{\pm}=N^{2}(i,n-1)\cdot N^{2}(n-2,n-1) \tag{3.3}$$
$$+\frac{N^{3}(i,n;n-1)\cdot N^{3}(n-2,n;n-1)}{\Delta^{2}(n-1,n)}\pm\frac{\sqrt{\Delta^{3}(i,n-1,n)\cdot\Delta^{3}(n-2,n-1,n)}}{\Delta^{2}(n-1,n)}$$

$$r_{in-k} \in \left[{}_nB^{-}_{in-k}, {}_nB^{+}_{in-k} \right], \quad i=1,...,n-k-1, \quad k=3,...,n-2, \quad n\ge 5,$$
$$\begin{aligned} {}_nB^{\pm}_{in-k} &= N^2(i,n-k+1)\cdot N^2(n-k,n-k+1) \\ &+ \sum_{j=2}^{k} \left\{ \begin{array}{l} N^{j+1}(i,n-k+j;n-k+1,...,n-k+j-1)\cdot \\ N^{j+1}(n-k,n-k+j;n-k+1,...,n-k+j-1)/ \\ \Delta^{j-1}(n-k+1,...,n-k+j-1)\cdot\Delta^{j}(n-k+1,...,n-k+j) \end{array} \right\} \\ &\pm \frac{\sqrt{\Delta^{k+1}(i,n-k+1,...,n)\cdot\Delta^{k+1}(n-k,n-k+1,...,n)}}{\Delta^{k}(n-k,n-k+1,...,n)} \end{aligned} \tag{3.4}$$

Proof As a starting point we recall Corollary 4.1 in [2]. The formulation depends upon the use of the multivariate iterated function defined recursively through

$$\begin{aligned} &w^{(k)}(x_1,y_1,...,x_k,y_k,z) = w(x_1,y_1,w^{(k-1)}(x_2,y_2,...,x_k,y_k,z)), \\ &x_i,y_i,z\in[-1,1], \quad i=1,...,k, \quad k=1,...,n-2, \\ &w^{(1)}(x,y,z) = w(x,y,z) = xy + z\cdot\sqrt{(1-x^2)(1-y^2)}, \quad x,y,z\in[-1,1]. \end{aligned} \tag{3.5}$$

Then, the correlation matrix $r=(r_{ij}), 1\le i,j\le n$, is positive semi-definite if, and only if, the following conditions are fulfilled:

$$r_{in} \in [-1,1], \quad i=1,...,n-1, \quad n\ge 2, \tag{3.6}$$

$$\begin{aligned} &r_{in-1} \in [B_{i,n-1;n}(-1), B_{i,n-1;n}(1)], \quad i=1,...,n-2, \quad n\ge 3, \\ &B_{i,n-1;n}(z) = w(r_{in}, r_{n-1n}, z) \end{aligned} \tag{3.7}$$

$$\begin{aligned} &r_{in-2} \in [B_{i,n-2;n-1,n}(-1), B_{i,n-2;n-1,n}(1)], \quad i=1,...,n-3, \quad n\ge 4, \\ &B_{i,n-2;n-1,n}(z) = w^{(2)}(r_{in-1}, r_{n-2n-1}, r_{in;n-1}, r_{n-2n;n-1}, z) \end{aligned} \tag{3.8}$$

$$\begin{aligned} &r_{in-k} \in [B_{i,n-k;n-k+1,...,n}(-1), B_{i,n-k;n-k+1,...,n}(1)], \\ &i=1,...,n-k-1, \quad k=3,...,n-2, \quad n\ge 5, \\ &B_{i,n-k;n-k+1,...,n}(z) = \\ &w^{(k)}(r_{in-k+1}, r_{n-kn-k+1}, r_{in-k+2;n-k+1}, r_{n-kn-k+2;n-k+1}, ..., r_{in;n-k+1,...,n-1}, r_{n-kn;n-k+1,...,n-1}, z) \end{aligned} \tag{3.9}$$

It suffices to show that the bounds (3.7)-(3.9) for r_{in-k}, $i=1,...,n-k-1$, $k=1,...,n-2$ coincide with the explicit bounds stated in (3.2)-(3.4). First, it is clear that the bounds in (3.7) coincides with those in (3.2) by definition of the function $w(\cdot)$ and the Definitions 2.1 of the mathematical symbols $N^2(\cdot), \Delta^2(\cdot)$. Now, let $i=1,...,n-k-1,\ k=2,...,n-2,\ n\ge 4.$ The expansion of (3.8)-(3.9) in a finite series yields the bounds

$$\begin{aligned}
{}_nB^{\pm}_{in-k} &= r_{in-k+1}\cdot r_{n-kn-k+1} \\
&+ \sum_{j=2}^{k}\left\{\begin{array}{l} r_{in-k+j;n-k+1,...,n-k+j-1}\cdot r_{n-kn-k+j;n-k+1,...,n-k+j-1}\cdot\sqrt{(1-r^2_{in-k+1})\cdot(1-r^2_{n-kn-k+1})} \\ \cdot\prod_{s=2}^{j-1}\sqrt{(1-r^2_{in-k+s;n-k+1,...,n-k+s-1})\cdot(1-r^2_{n-kn-k+s;n-k+1,...,n-k+s-1})}\end{array}\right\} \\
&\pm\left\{\begin{array}{l}\sqrt{(1-r^2_{in-k+1})\cdot(1-r^2_{n-kn-k+1})}\cdot \\ \prod_{s=2}^{k}\sqrt{(1-r^2_{in-k+s;n-k+1,...,n-k+s-1})\cdot(1-r^2_{n-kn-k+s;n-k+1,...,n-k+s-1})}\end{array}\right\}
\end{aligned} \tag{3.10}$$

Clearly, the first terms in (3.10) coincide with the corresponding first terms in (3.3)-(3.4). The identification of the other terms is shown by induction on k. For $k=2$ we have by definition (2.3) the partial correlation formula

$$r_{in;n-1} = \frac{N^3(i,n;n-1)}{\sqrt{\Delta^2(i,n-1)\cdot\Delta^2(n-1,n)}},\quad i=1,...,n-2, \tag{3.11}$$

which implies that

$$\begin{aligned}
&r_{in;n-1}\cdot r_{n-2n;n-1}\cdot\sqrt{(1-r^2_{in-1})\cdot(1-r^2_{n-2n-1})} \\
&= \frac{N^3(i,n;n-1)\cdot N^3(n-2,n;n-1)}{\Delta^2(n-1,n)},\quad i=1,...,n-3
\end{aligned} \tag{3.12}$$

Similarly, with changed index set the relationship (2.8) yields

$$1-r^2_{in;n-1} = \frac{\Delta^3(i,n-1;n)}{\Delta^2(i,n-1)\cdot\Delta^2(n-1,n)},\quad i=1,...,n-2, \tag{3.13}$$

which implies that

$$\begin{aligned}
&\sqrt{(1-r^2_{in-1})\cdot(1-r^2_{n-2n-1})}\cdot\sqrt{(1-r^2_{in;n-1})(1-r^2_{n-2n;n-1})} \\
&= \frac{\sqrt{\Delta^3(i,n-1;n)\cdot\Delta^3(n-2,n-1;n)}}{\Delta^2(n-1,n)},\ i=1,...,n-3.
\end{aligned} \tag{3.14}$$

Together this shows the formula (3.3) for arbitrary $n \geq 4$. Now, let $k = 3,...,n-2, n \geq 5$, and assume that (3.10) with k replaced by $k-1$ coincide with the corresponding terms in (3.4) for arbitrary $n \geq 5$. We must show that (3.10) coincides with (3.4). We observe that the summands for $j = 2,...,k-1$ in the sum of (3.10) coincide by induction assumption with the corresponding terms in (3.3)-(3.4) of the bounds ${}_{n-1}B^{\pm}_{in-1-(k-1)}$, which obviously coincide with the corresponding terms of the bounds ${}_nB^{\pm}_{in-k}$, $i = 1,...,n-k-1$. Therefore, it remains to show that the summand with $j = k$ in the sum of (3.10) and the last term in (3.10) coincide with the corresponding terms in (3.4) for arbitrary $n \geq 5$. For the summand $j = k$ in the sum of (3.10), we have by definition (2.3) that

$$r_{in;n-k+1,...,n-1} = \frac{N^{k+1}(i,n;n-k+1,...,n-1)}{\sqrt{\Delta^k(i,n-k+1,n-1) \cdot \Delta^k(n-k,n-k+1,...,n)}}, \quad i = 1,...,n-k . \tag{3.15}$$

Using the induction assumption on the last term of the series expansion for ${}_{n-1}B^{\pm}_{in-1-(k-1)}$, we see that

$$\begin{aligned} &\sqrt{(1-r^2_{in-k+1}) \cdot (1-r^2_{n-kn-k+1})} \cdot \\ &\prod_{s=2}^{k-1} \sqrt{(1-r^2_{in-k+s;n-k+1,...,n-k+s-1}) \cdot (1-r^2_{n-kn-k+s;n-k+1,...,n-k+s-1})} \\ &= \frac{\sqrt{\Delta^k(i,n-k+1,...,n-1) \cdot \Delta^k(n-k,n-k+1,...,n-1)}}{\Delta^{k-1}(n-k,n-k+1,...,n-1)}. \end{aligned} \tag{3.16}$$

Using (3.15)-(3.16) it follows that the summand $j = k$ in the sum of (3.10) equals

$$\begin{aligned} &r_{in;n-k+1,...,n-1} \cdot r_{n-kn;n-k+1,...,n-1} \cdot \sqrt{(1-r^2_{in-k+1}) \cdot (1-r^2_{n-kn-k+1})} \\ &\cdot \prod_{s=2}^{k-1} \sqrt{(1-r^2_{in-k+s;n-k+1,...,n-k+s-1}) \cdot (1-r^2_{n-kn-k+s;n-k+1,...,n-k+s-1})} \\ &= \frac{N^{k+1}(i,n;n-k+1,...,n-1) \cdot N^{k+1}(n-k,n;n-k+1,...,n-1)}{\Delta^{k-1}(n-k,n-k+1,...,n-1) \cdot \Delta^k(n-k,n-k+1,...,n)}, \\ &i = 1,...,n-k-1, \end{aligned} \tag{3.17}$$

and coincides with the corresponding term in (3.4). On the other hand, with changed index set the relationship (2.8) yields

$$1-r^2_{in;n-k+1,\dots,n-1}=\frac{\Delta^{k+1}(i,n-k+1,\dots,n)\cdot\Delta^{k-1}(n-k,n-k+1,\dots,n-1)}{\Delta^{k}(i,n-k+1,\dots,n-1)\cdot\Delta^{k}(n-k,n-k+1,\dots,n-1)},\quad i=1,\dots,n-k. \tag{3.18}$$

Using this and the relationship (3.16) above, we see that the last term in (3.10) coincides with the last term in (3.4). The result is shown. ◊

4. Volume measure for the convex set of valid correlation matrices

Consider the set of positive semi-definite *nxn* correlation matrices parameterized by the set $S_n=\{r=(r_{ij}),1\le i<j\le n\}, n\ge 2$, which is known to be a convex subset of the m-dimensional cube $[-1,1]^m$ with $m=\frac{1}{2}(n-1)n$. The explicit recursive generation of the correlation bounds is used to calculate the volume V_n of this set. The derivation depends upon the determinantal identities introduced in Section 2.

Theorem 4.1 (*Volume measure for the space of valid correlation matrices*) The volume of the space $S_n, n\ge 2$, is given by

$$V_n=Vol(S_n)=\prod_{k=1}^{n-1}(J_k)^k,\quad J_k=\int_{-1}^{1}(1-t^2)^{\frac{1}{2}(k-1)}dt. \tag{4.1}$$

Proof Since $V_2=J_1=2, V_3=J_1\cdot(J_2)^2=2\cdot(\frac{\pi}{2})^2$, we assume that $n\ge 4$. A close look at the recursive bounds of Theorem 3.1 shows that the volume is obtained by integrating backwards successively over the packages of variables $r_{12},(r_{i3}),\dots,(r_{in-1}),(r_{in})$ in this order. First of all, we note that

$$\int_{-{}_nB_{12}^-}^{{}_nB_{12}^+}1\cdot dr_{12}=2\cdot\frac{\sqrt{\Delta^{n-1}(1,3,\dots,n)\cdot\Delta^{n-1}(2,3,\dots,n)}}{\Delta^{n-2}(3,\dots,n)}. \tag{4.2}$$

As next step, the integral of (4.2) with respect to the variables $r_{i3}, i=1,2$, only depends upon the square root expressions and separates. Therefore, we have to evaluate the integrals defined by

$${}_nI_{i3}:=\int_{-{}_nB_{i3}^-}^{{}_nB_{i3}^+}\sqrt{\Delta^{n-1}(i,3,\dots,n)}dr_{i3},\quad i=1,2. \tag{4.3}$$

With appropriate choice of the index set the identity (2.8) yields the expressions

$$\frac{\Delta^{n-1}(i,3,...,n)}{\Delta^{n-3}(4,...,n)} = \frac{\Delta^{n-2}(i,4,...,n)\cdot\Delta^{n-2}(3,4,...,n)}{\Delta^{n-3}(4,...,n)^2} - \left[\frac{N^{n-1}(i,3;4,...,n)}{\Delta^{n-3}(4,...,n)}\right]^2, \quad i=1,2. \qquad (4.4)$$

Moreover, the use of the d-scaled partial correlation expansion (2.7) yields the following expression for the square bracket quantity in (4.4) (with changed index set)

$$\begin{aligned} &\frac{N^{n-1}(i,3;4,...,n)}{\Delta^{n-3}(4,...,n)} = N^3(i,3;4) \\ &-\sum_{k=2}^{n-3}\frac{N^{k+1}(i,k+3;4,...,k+2)\cdot N^{k-1}(3,k+3;4,...,k+2)}{\Delta^{k-1}(4,...,k+2)\cdot\Delta^k(4,...,k+3)}, \end{aligned} \qquad (4.5)$$

where for $n=4$ the sum is obviously empty and set equal to zero. Using (2.5) and the convention $N^2(i,3)=r_{i3}$ we note that $N^3(i,3;4)=r_{i3}-N^2(i,4)\cdot N^2(3,4)$. Setting $k=n-3$ in (3.4) we see through comparison with (4.5) that

$${}_nB_{i3}^{\pm} = r_{i3} - \frac{N^{n-1}(i,3;4,...,n)}{\Delta^{n-3}(4,...,n)} \pm \frac{\sqrt{\Delta^{n-2}(i,4,...,n)\cdot\Delta^{n-2}(3,4,...,n)}}{\Delta^{n-3}(3,4,...,n)}. \qquad (4.6)$$

Since the last expression in (4.6) does not depend on the integrating variable and (4.5) depends on it only through its first term $N^3(i,3;4)$, it is natural to make the change of variables

$$x_i = x_i(r_{i3}) = N^{n-1}(i,3;4,...,n)/\Delta^{n-3}(4,...,n), \quad dx_i = dr_{i3}, \quad i=1,2. \qquad (4.7)$$

Using (4.4) the integrals (4.3) transform to

$$\begin{aligned} &{}_nI_{i3} := \sqrt{\Delta^{n-3}(4,...,n)}\cdot\int_{-c_i}^{c_i}\sqrt{c_i^2-x_i^2}\,dx_i, \\ &c_i = \sqrt{\Delta^{n-2}(i,4,...,n)\cdot\Delta^{n-2}(3,4,...,n)}\,/\,\Delta^{n-3}(3,4,...,n), \quad i=1,2. \end{aligned} \qquad (4.8)$$

Setting $x_i = c_i\cdot t$ one gets ${}_nI_{i3}=c_i^2J_2$, $i=1,2$. Using this and (4.2) we see that the new integrand over the remaining space of integration is

$$2\cdot{}_nI_{13}\cdot{}_nI_{23}/\Delta^{n-2}(3,4,...,n) = J_1(J_2)^2\prod_{i=1}^{3}\Delta^{n-2}(i,4,...,n)/\Delta^{n-3}(3,4,...,n)^3. \qquad (4.9)$$

We proceed now in the same way. In general, the calculation of the formula (4.1) for the dimension $n \geq 4$ requires $n-2$ steps, numbered $k = 1,...,n-2$. We claim that in the first $n-3$ steps we have to calculate integrals defined by

$$ {}_nI_{ik+2} := \int_{-{}_nB^-_{ik+2}}^{{}_nB^+_{ik+2}} \Delta^{n-k}(i,k+2,...,n)^{\frac{1}{2}k} dr_{ik+2}, \quad i = 1,...,k+1, \tag{4.10} $$

and that the result of this integration leads in the next step to the new integrand

$$ \prod_{j=1}^{k+1}(J_j)^j \cdot \prod_{i=1}^{k+2} \Delta^{n-k-1}(i,k+3,...,n)^{\frac{1}{2}(k+1)} / \Delta^{n-k-2}(k+3,...,n)^{\frac{1}{2}(k+1)(k+2)}, \tag{4.11} $$

which for the last step $k = n-2$ reduces to the integrand

$$ \prod_{j=1}^{n-2}(J_j)^j \cdot \prod_{i=1}^{n-1} \Delta^2(i,n)^{\frac{1}{2}(n-2)} / \Delta^1(n)^{\frac{1}{2}(n-2)(n-1)}, \tag{4.12} $$

because by convention $\Delta^1(n) = 1$. A final integration yields the desired formula as

$$ \begin{aligned} V_n &= \prod_{k=1}^{n-2}(J_k)^k \cdot \prod_{i=1}^{n-1} \left\{ \int_{-1}^{1} \Delta^2(i,n)^{\frac{1}{2}(n-2)} dr_{in} \right\} \\ &= \prod_{k=1}^{n-2}(J_k)^k \cdot \prod_{i=1}^{n-1} \left\{ \int_{-1}^{1} (1-t^2)^{\frac{1}{2}(k-1)} dt \right\} = \prod_{k=1}^{n-1}(J_k)^k. \end{aligned} \tag{4.13} $$

The remaining assertion is proved by induction, where the induction step for $k = 1$ has been shown above. Showing that the assertion holds for $k+1$ when it holds for k is performed similarly and is left to the reader. ◊

Remark 4.1 Numerical values for V_3, V_4 had been mentioned in [5]. More recently, Joe [3] derives a volume formula in a probabilistic way as a by-product of some new methods to generate random correlation matrices. The new proof requires only elementary first-year calculus and is non-probabilistic in nature. It is not difficult to recover Joe's formula from the following recursions (use partial integration)

$$ J_{2(k+1)} = \frac{2k+1}{2k+2} J_{2k}, \quad J_2 = \frac{\pi}{2}, \quad J_{2k+1} = \frac{2k}{2k+1} J_{2k-1}, \quad J_1 = 2, \quad k = 1,2,... \tag{4.14} $$

References

[1] M. Budden, P. Hadavas and L. Hoffman, On the generation of correlation matrices, *Applied Mathematics E-Notes*, **8** (2008), 279-282.

[2] W. Hürlimann, Compatibility conditions for the multivariate normal copula with given rank correlation matrix, *Pioneer Journal of Theoretical and Applied Statistics*, (2012), to appear.

[3] H. Joe, Generating random correlation matrices based on partial correlations, *Journal of Multivariate Analysis*, **97** (2006), 2177-2189.

[4] D. Kurowicka and M.R. Cooke, *Uncertainty analysis with high dimensional dependence modelling*, J. Wiley, Chichester, 2006.

[5] P.J. Rousseuw and G. Molenberghs, The shape of correlation matrices, *The American Statistician*, **48**(4) (1994), 276-279.

[6] G. Yule and M. Kendall, *An Introduction to the Theory of Statistics* (14^{th} ed.), Charles Griffin & Company, Belmont, California, 1965.

Received: March, 2012