Visualizing Distributions of Covariance Matrices *

Tomoki Tokuda[a], Ben Goodrich[b], Iven Van Mechelen[a], Andrew Gelman[b],
Francis Tuerlinckx[a]

[a] *University of Leuven, Belgium*
[b] *Columbia University, New York, NY, USA*

Abstract

We present some methods for graphing distributions of covariance matrices and demonstrate them on several models, including the Wishart, inverse-Wishart, and scaled inverse-Wishart families in different dimensions. Our visualizations follow the principle of decomposing a covariance matrix into scale parameters and correlations, pulling out marginal summaries where possible and using two and three-dimensional plots to reveal multivariate structure. Visualizing a distribution of covariance matrices is a step beyond visualizing a single covariance matrix or a single multivariate dataset. Our visualization methods are available through the R package `VisCov`.

Keywords: Bayesian statistics, prior distribution, Wishart distribution, inverse-Wishart distribution, statistical graphics

1   Background

Covariance matrices and their corresponding distributions play an important role in statistics. To understand the properties of distributions, we often rely on visualization methods.

But visualizing a distribution in a high-dimensional space is a challenge, with the additional difficulty that covariance matrices must be positive semi-definite, a restriction that forces the joint distribution of the covariances into an oddly-shaped subregion of the space.

Distributions of covariance matrices show up in classical (see Anderson, 2003) and Bayesian statistics (see Daniels and Kass, 1999; Barnard et al., 2000; Yang and Berger, 1994):

- The sampling distribution of the covariance matrix of independent multivariate observations. If the data are generated according to a multivariate normal distribution, then their covariance matrix has a Wishart sampling distribution (see Wishart, 1928; Press, 1982).

- The prior for a covariance matrix in a Bayesian analysis, most simply if data are modeled as independent draws from a multivariate normal with an unknown mean and covariance matrix. In the same vein, a prior on the residual covariance matrix is needed in a multivariate linear regression model (see Box and Tiao, 1973; Zellner, 1971).

- In hierarchical regression model with varying intercepts and slopes, the vector of varying coefficients for each group is often assumed to follow a multivariate normal distributed random variable with mean zero and an unknown covariance matrix, which again needs a prior distribution for a Bayesian analysis (see Gelman and Hill, 2007).

- The covariance matrix itself may be unit-specific and drawn independently from a population distribution of covariance matrices. Such a situation is less common but can occur in both classical and Bayesian settings (see Oravecz et al., 2009).

In the remainder of this paper, we only consider models for prior distributions because, as applied Bayesians, we are often in the position of choosing a family of prior distributions or seeking to understand a prior distribution that has already been specified for an analysis. Several classes of priors for covariance matrices have been proposed in the statistical literature but many of their properties are unknown analytically and, as a profession, we have not yet acquired extensive expertise with these different models. However, it would be helpful to have a method that allows us to assess whether the distributions and the relations between the covariance parameters are in accordance with the prior information that one has about these parameters or with the way one wishes to constrain these parameters.

As an example, consider the inverse-Wishart distribution, which is used often in Bayesian modeling because it is a proper conjugate prior for an unknown covariance matrix in a multivariate normal model (see Gelman et al., 2004). Some specific analytical results for the inverse-Wishart have been derived; for example, the marginal distribution of a diagonal block submatrix of draws from an inverse-Wishart distribution is also inverse-Wishart (Press, 1982), and marginal moments of such draws have been derived as well (Von Rosen, 1988). But marginal distributions are typically not known and there is no expression for the bivariate distribution of any two covariances. As a result, our analytical knowledge of the properties of the inverse-Wishart distribution is still highly incomplete. Various alternatives to the inverse-Wishart have been proposed (see Barnard et al., 2000; O'Malley and Zaslavsky, 2008), but even fewer analytical results are known for these families, making it even more challenging to understand precisely the properties of such distributions. Consequently, our analytical understanding of these distributions falls short of providing us a full understanding of the inverse-Wishart distribution.

Since analytical results are limited, researchers need other tools to study covariance matrix distributions. Visualization provides insight into the properties of such distributions. There is a considerable literature concerning visualization of multivariate data (e.g., Valero-Mora et al., 2003; Theus and Urbanek, 2008; Cook and Swayne, 2007). For visualization of a single covariance or correlation matrix, in particular, a heatmap is often used (e.g., Friendly, 2002). Yet, these existing approaches tend to focus on the visualization of a single covariance matrix (i.e., derived from a single data set), while our approach is slightly different in that we are visualizing distributions of covariance matrices.

Visualizing a distribution of covariance matrices is different from the visualization of single covariance matrix (or correlation matrix). As such, there is a need for specialized methods because not all techniques carry over easily from the single instance case to the distribution situation. For instance, averaging a heatmap over a number of instances of correlation matrices ends up with displaying the mean correlation matrix, which does not capture the variability of the distribution. Certainly, the general principles of visualization are relevant: In particular many of the principles of dynamic graphics might well apply in our setting. On top of these existing approaches, however, we develop new methods to visualize distributions of covariance matrices, focusing on static graphics. To cope with the high dimensionality of multivariate distributions, we make the best use of symmetries in

conventional classes of prior distributions of covariance matrices and visualize key aspects of these distributions.

We propose a series of graphs to visualize covariance matrix distributions. For any joint distribution of the covariances, we construct a grid showing the marginal distribution of the scale and correlation parameters, along with two and three-dimensional scatterplots. Since a covariance matrix can be expressed as an (equiprobability or isodensity) ellipse for a multivariate normal distribution (or, more generally, for any elliptical distribution such as a multivariate t-distribution; Owen and Rabinovitch, 1983), we also display the distribution as a mixture of ellipses in a multivariate normal setting. We demonstrate for several distributions—including the Wishart, inverse-Wishart, scaled inverse-Wishart, and uniform correlation—that this series of graphs allows salient features to be visualized, analyzed, and compared. Our ultimate goal is to offer a method and tool so that we can better understand our multivariate models.

2   A four-layered visualization method

Before explaining the method in detail, let us first introduce some notation. A $k \times k$ covariance matrix Σ has variances σ_i^2 ($i = 1, \ldots, k$) on its diagonal. The typical off-diagonal element is $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$ ($i = 1, \ldots, k, j = 1, \ldots, k, i \neq j$), where σ_i is the standard deviation of the ith variable and ρ_{ij} the correlation between the ith and jth variables. We often separate covariances into standard deviations and scale-free correlations for interpretability.

We begin with the inverse-Wishart distribution:

$$\Sigma \sim \text{Inv-Wishart}_\nu(\boldsymbol{S}^{-1}) \tag{1}$$

where ν denotes the degrees of freedom and $\boldsymbol{S}$ is a positive definite $k \times k$ scale matrix. The density of Σ is proportional to:

$$p(\Sigma) \propto |\Sigma|^{-(\nu+k+1)/2} \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{S}\Sigma^{-1})\right). \tag{2}$$

The expectation of the inverse-Wishart distribution is $\boldsymbol{S}/(\nu - k - 1)$. For all illustrations of the inverse-Wishart distribution in this paper, we assume that $\boldsymbol{S}$ is a $k \times k$ identity matrix,

denoted $\boldsymbol{I_k}$, which makes the variables exchangeable. However, our method can easily be used with any other scale matrix. The distribution is proper iff $\nu \geq k$ (and the first moment only exists if $\nu > k + 1$). Furthermore, if $\boldsymbol{\Sigma}$ follows an inverse-Wishart distribution, then any submatrix among q (possibly permuted) variables is also inverse-Wishart (Eaton, 1983): $\boldsymbol{\Sigma}_D \sim \text{Inv-Wishart}_{\nu-k+q}(\boldsymbol{S}_D^{-1})$ where $\boldsymbol{S}_D$ is such a submatrix of $\boldsymbol{S}$. If we take $\nu = k + c$ (where c is a positive constant), then the degrees of freedom of the distribution of $\boldsymbol{\Sigma}_D$ are $q+c$ and do not depend on k. The degrees of freedom of the distribution of a submatrix then parallels that of the original matrix. As a consequence, we can focus on a 4×4 covariance matrix distribution to study most (but not all) properties of the inverse-Wishart distribution (because the joint marginal of two correlation coefficients ρ_{ij} and ρ_{km} without a common variable requires at least four dimensions).

In our visualization method we start with sampling L (typically $L = 1000$) covariance matrices $\boldsymbol{\Sigma}$ (or correlation matrices $\boldsymbol{R}$) from their distribution and then plot the sampling distribution of various statistics in four layers or parts. The first layer consists of univariate histograms of the (logarithm of the) standard deviations and the correlations. The second layer is a set of bivariate scatterplots of variances and/or correlations. Third, we construct three-dimensional scatterplots of either two variances and a correlation that overlays iso-density contours or three correlations. Finally, the fourth layer aims at reducing the entire covariance matrix into scalar measures called the effective variance and the effective dependence by Peña and Rodríguez (2003).

Our four-layered graphical representation reveals different aspects of the covariance distribution that can be used to compare the implications of different values of hyperparameters or different distributions. A detailed explanation of the method is given below, but Figure 1 provides an illustration of the method, as generated by our R function `VisCov` (from the package with the same name available on CRAN).

2.1 Layer 1: Histograms of $\log(\sigma_i)$ and ρ_{ij}

For the $L = 1000$ draws, we display the histogram of the logarithm of the standard deviations and the histogram of a correlation ρ_{ij} (typically ρ_{12}). Given the assumption of exchangeability, the histogram for each standard deviation is the same, as is the histogram for each correlation; therefore, it is only necessary to display one of each.
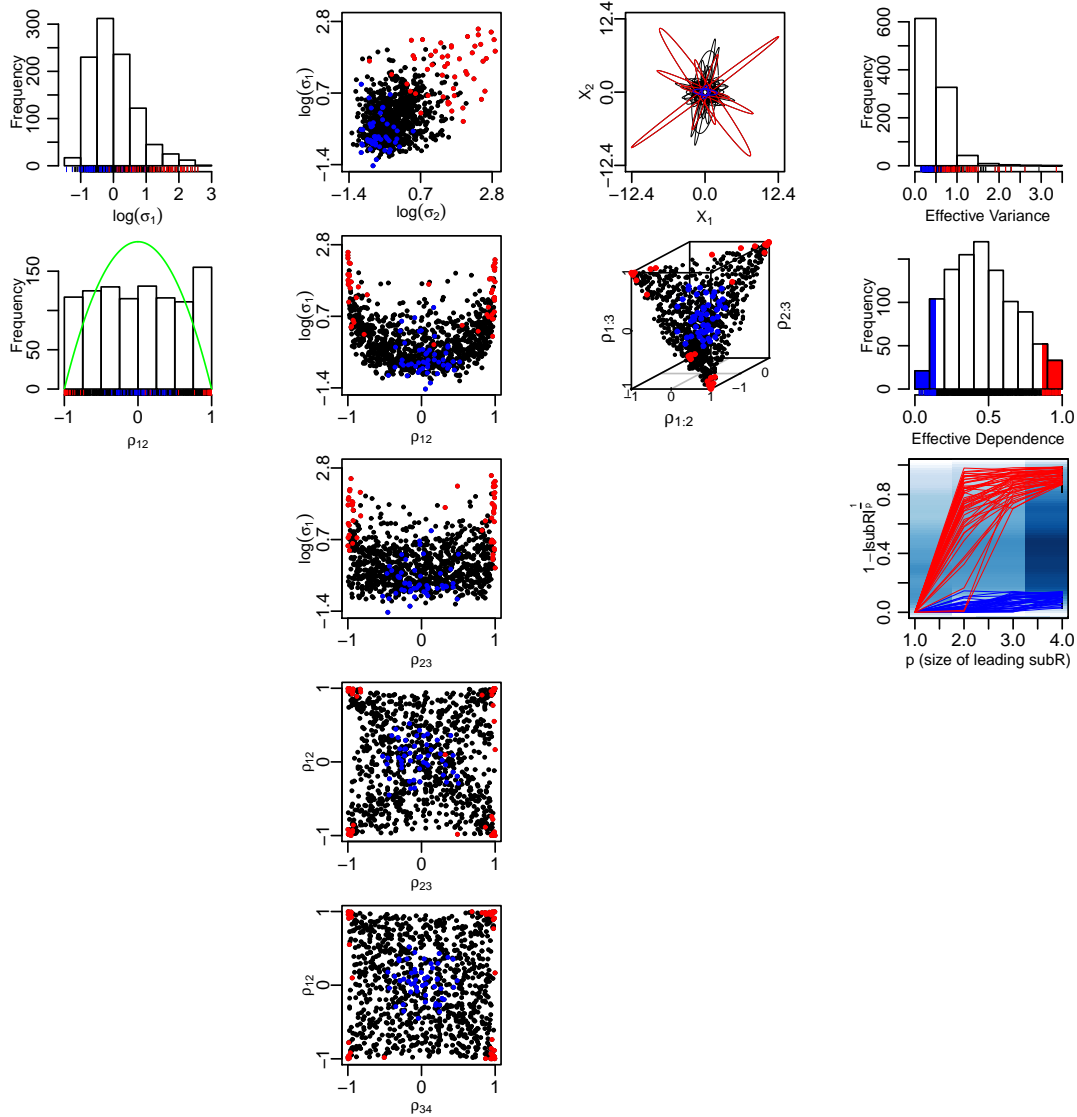
Figure 1: Visualization of an inverse-Wishart distribution with dimension $k = 4$, degrees of freedom $k + 1$ and an identity scale matrix, as generated by our R function `VisCov`. To construct this plot, 1000 covariance matrices were sampled from the inverse-Wishart distribution. Each column of plots refers to a different layer of the visualization method (univariate, bivariate, trivariate and multivariate). The first column shows two histograms (of a log standard deviation and of correlation). A green reference line is added to the correlation histogram, which is the density of a correlation of a correlation matrix that is uniformly distributed. In the second column, various scatterplots are shown. In the third column, the first plot shows 100 50% equiprobability ellipses of normal distribution centered at the origin (based on a random subsample of the 1000 covariance matrices). In the last column, the effective variance and dependence are shown. The covariance matrices with extreme effective dependences are colored blue or red (for small and large effective dependence, respectively). The points in the other plots based on these extreme effective dependency matrices are also colored blue and red. The final plot in the last column shows effective dependence as a function of a growing dimension of the leading principal submatrix. See the text for more explanation.

2.2   Layer 2: scatterplots

From the L draws, we can construct $\binom{k(k+1)/2}{2}$ scatterplots for pairs of covariances, but a smaller number of plots will suffice. The scatterplots can be found in the second column of Figure 1, where a scatterplot of variable y versus x will be denoted as (x, y).

If $k = 2$, the two plots (σ_1, σ_2) and (σ_1, ρ_{12}) exhaust the relevant information due to the exchangeability assumption. If $k = 3$, there are four non-redundant plots: (σ_1, σ_2), (σ_1, ρ_{12}), (σ_1, ρ_{23}), (ρ_{12}, ρ_{23}). The difference between (σ_1, ρ_{12}) and (σ_1, ρ_{23}) is that there is an overlap of variables in the former. Finally, if $k > 3$, five scatterplots are necessary when exchangeability is assumed: (σ_1, σ_2), (σ_1, ρ_{12}), (σ_1, ρ_{23}), (ρ_{12}, ρ_{23}), (ρ_{12}, ρ_{34}). In the last plot, the correlations have no variable in common, which illustrates out earlier claim that covariance matrices with $k = 4$ are sufficient to reveal most of the interesting information about the inverse-Wishart distribution.

2.3   Layer 3: Contour plot and three-dimensional scatterplots

Here we look at three elements of the covariance matrix simultaneously in two different ways.

Contour plot. In the contour plot approach (see third column, first plot of Figure 1), we focus on the distribution of a specific 2×2 marginal sub-matrix of $\boldsymbol{\Sigma}$. For instance, we can look at the sub-matrix formed by variables i and j, defined as

$$\begin{pmatrix} \sigma_i^2 & \sigma_i \sigma_j \rho_{ij} \\ \sigma_i \sigma_j \rho_{ij} & \sigma_j^2 \end{pmatrix}. \tag{3}$$

Given the exchangeability assumption, we take $i = 1$ and $j = 2$ and further assume the two variables are bivariate normal with mean vector zero. Thus, the 50% equiprobability ellipse (the contour plot in which 50% of the bivariate normal density lies) can be plotted (see Johnson and Wichern, 2007). It would be straightforward to use another bivariate distribution if desired.

Each ellipse represents an idealized cloud of points in two dimensions and gives information about the orientation and spread of the points along both axes. To avoid clutter, we usually show fewer contour plots, and about 100 seem sufficient to visualize the pattern of isodensity contours.

Three-dimensional scatterplot. We also include a three-dimensional scatterplot of three

correlations: ρ_{ij}, $\rho_{i,j+1}$ and $\rho_{i+1,j+1}$. The triplet $(\rho_{ij}, \rho_{i,j+1}, \rho_{i+1,j+1})$ corresponds to a 3×3 correlation submatrix from the $k \times k$ correlation matrix. Other triplets of correlations can be shown, but in order not to overload the picture, we usually take $i = 1$ and $j = 2$.

The two-dimensional scatterplots of the correlations already suggest some of the intricate relations among the correlations under the positive semi-definiteness constraint, and the three-dimensional scatterplot goes a step further. It has been shown by Rousseeuw and Molenberghs (1994) that the support of the distribution of three correlations is a convex body (called a elliptical tetrahedron). Any cross-section parallel to one of the two-dimensional coordinate planes forms an ellipse, implying the support of the conditional distribution of two correlation coefficients given the third one is elliptical. Hence, the general pattern in the second plot of the third column of Figure 1 will often occur throughout this paper: For a large enough sample, the convex hull of the points in the scatterplot will approximately coincide with the elliptical tetrahedron, but the way the points are distributed in this volume differs from one distribution to another.

2.4 Layer 4: Effective variance and dependence

Visualization in four or more dimensions is difficult, but visualization in fewer dimensions cannot capture all the relevant information in a covariance distribution. Thus, we also analyze scalar statistics that are a function of the entire covariance matrix.

Peña and Rodríguez (2003) have defined the effective variance V_e of a $k \times k$ covariance matrix $\boldsymbol{\Sigma}$ to be

$$V_e = |\boldsymbol{\Sigma}|^{\frac{1}{k}} \tag{4}$$

and the effective dependence as

$$D_e = 1 - |\boldsymbol{R}|^{\frac{1}{k}} \tag{5}$$

where $\boldsymbol{R}$ is the correlation matrix derived from $\boldsymbol{\Sigma}$.

These two statistics facilitate comparisons over different values of k or different-sized submatrices of $\boldsymbol{\Sigma}$. The first two plots of the last column of Figure 1 give the histograms of the effective variance and dependence under the inverse-Wishart distribution. In a third

plot, the effective dependence of a leading $i \times i$ submatrix is given on the y-axis as a function of i (with $i = 1, \ldots, k$). Each $k \times k$ correlation matrix defines a line (by letting the leading $i \times i$ submatrix grow) and the collection of lines is smoothed and shown as the light blue background.

The most extreme matrices with respect to effective dependence are indicated by the colored tails (blue for low values and red for large values). The matrices with very low or very high effective dependence can be identified in the histograms via the rug, via the same blue and red color scheme in the scatterplots and contour plots, which illustrates how the effective dependence relates to, for instance, the bivariate distribution of the correlations.

3 Illustration of the visualization method for several distributions

In this section, we apply the four-layered visualization method for various distributions of covariance matrices, draw implications from the plots, and compare across distributions. Due to space limitations, we do not present the complete graphical display (such as in Figure 1), but focus our attention on plots that facilitate comparisons. The full four-layered plot can be recreated using our `VisCov` function.

3.1 The inverse-Wishart distribution

We first reexamine several important features of the inverse-Wishart distribution with $k = 4$ dimensions and $\nu = k+1 = 5$ degrees of freedom from Figure 1. First, the univariate marginal distribution of a correlation is uniform, which is known analytically when $\nu = k+1$. Second, the scatterplots reveal a strong positive relationship between the (log) standard deviations, which is also reflected in the contour plot (third layer) where ellipses stretching along only one of the main axes are rare. In this distribution, large ellipses tend to be oriented along one of the two principal diagonals. Third, if two variables are extremely correlated, their standard deviations tend to be large, which shows up in the contour plot as well. Fourth, both the bivariate and trivariate scatterplots reveal a pattern where some correlations tend to be similar in magnitude, while the corners of the bivariate and trivariate scatterplots are quite dense. Fifth, covariance matrices with a large degree of effective dependence (colored

in red) tend to have large (log) standard deviations and extreme correlations, which is also apparent in the contour plots where extreme effective dependence coincides with small volume in the metric space. Covariance matrices with little effective dependence tend to have smaller ellipses and (log) standard deviations.

The right column of Figure 2 uses $\nu = k+50$, rather than $\nu = k+1$ in the left column with $k = 4$ in both cases. This figure compares two types of scatterplots (containing log standard deviations and correlations), the contour plots, and the histogram of effective dependencies. When $\nu = k+50$, there is less dependence among the log standard deviations and among the correlations, but the marginal distributions of both log standard deviations and correlations become heavily concentrated. A similar pattern can be seen in the contour plots: The length of the major and minor axes decreases, as well as the eccentricity. For a large number of degrees of freedom, the effective dependence becomes very small, which is in line with the fact that the marginal correlations tend to be close to zero.

Both Figures 1 and 2 support the observation in Gelman et al. (2004) that the inverse-Wishart is quite restrictive, in part due to the fact that it only has one degrees of freedom parameter, ν. When the degrees of freedom are set to $\nu = k+1$, the marginal distribution of the correlation is uniform, but the joint distribution of the correlations is far from uniform. There tends to be an abundance of mass in the extreme corners of its support and often severe effective dependence. Increasing the degrees of freedom concentrates the distribution around its expectation (here $\boldsymbol{I}_k/(\nu - k - 1)$), which is a strong prior that may not be appropriate in a particular research situation.

We also compare different values of dimensionality, k, which does not lead to qualitatively different conclusions when looking at the first three layers of our visualization method. However, the distribution of effective dependences is pushed toward the upper bound of one when k increases (and $\nu = k+1$). Figure 3 compares two histograms of effective dependence for the inverse-Wishart distributions with $k = 4$ and $k = 100$. Proof 1 in Appendix shows that $\lim_{k \to \infty} 1 - E\left(|\boldsymbol{R}|^{\frac{1}{k}}\right) = 1$, which is also apparent from the histogram when $k = 100$. The intuition behind this property is that the average proportion of explained variability of the variables increases with the number of variables, just as a simple R^2 increases in a regression if more predictors are added.

In order to mitigate the restrictiveness of the inverse-Wishart distribution, separation strategies have been proposed where the covariance matrix is decomposed so that different
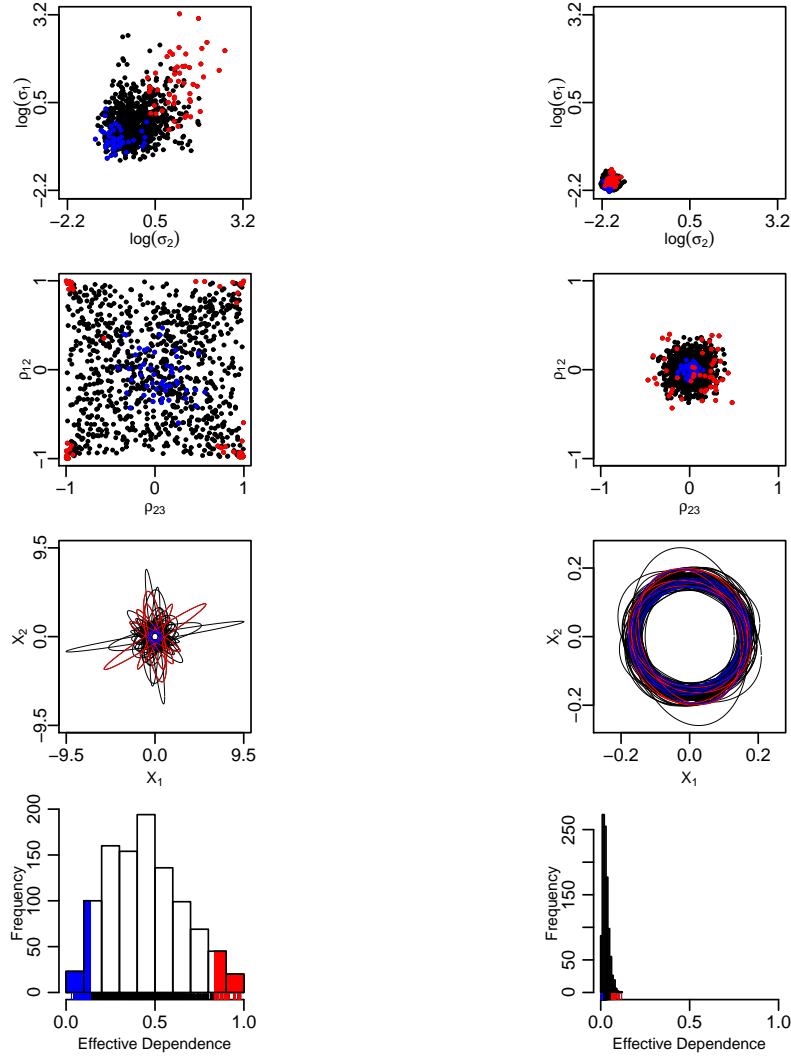
Figure 2: Visualization of an inverse-Wishart distribution (via 1000 simulations) with dimension $k = 4$, degrees of freedom $k + 1$ and $k + 50$, and an identity scale matrix. The plots in the left and right columns use $\nu = k + 1$ and $\nu = k + 50$ respectively. The first row represents the scatterplot of two log standard deviations, the second row depicts the scatterplot of two correlations (that share a common variable), the third row shows 100 contour plots, and the last row contains histograms of the effective dependences. See the text for more explanation.
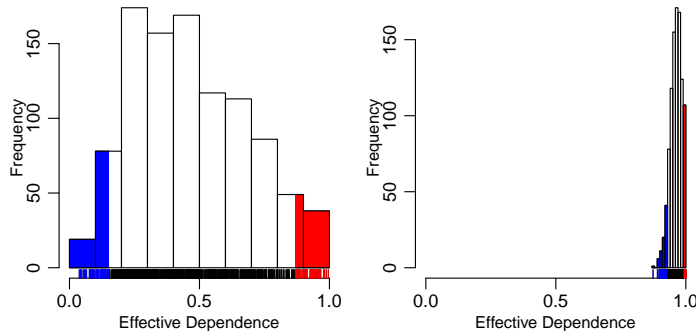
Figure 3: The distribution of the effective dependence of an inverse-Wishart distribution with dimension $k = 4$ (left panel) and $k = 100$ (right panel) based on 1000 simulations with $\nu = k + 1$ and an identity scale matrix.

priors can be specified on the factors of the covariance matrix. The next three subsections will be devoted to three separation strategies based on different decompositions of $\boldsymbol{\Sigma}$.

3.2 Separation strategy with marginal correlation matrix from the inverse-Wishart

Barnard et al. (2000) propose the following decomposition:

$$\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_k) \cdot \boldsymbol{R} \cdot \operatorname{diag}(\sigma_1, \ldots, \sigma_k), \tag{6}$$

where $\boldsymbol{R}$ has the marginal distribution of the correlation matrix in the inverse-Wishart distribution. The notation $\operatorname{diag}(\sigma_1, \ldots, \sigma_k)$ refers to a diagonal matrix obtained by placing the σ_i's on the diagonal. In other words, the standard deviations are integrated out, leaving the marginal distribution of the correlation matrix, and then replacement of standard deviations can be taken from any marginal distribution to form a new joint distribution of the covariance matrix. As derived by Barnard et al. (2000) (see also the first part of Proof 1 in the Appendix), the kernel of the density function of $\boldsymbol{R}$ is:

$$p(\boldsymbol{R}) \propto |\boldsymbol{R}|^{\frac{(\nu-1)(k-1)}{2}-1} \left(\prod_{i=1}^{k} |\boldsymbol{R}_{ii}| \right)^{-\frac{\nu}{2}} \tag{7}$$

where $\boldsymbol{R}_{ii}$ is the ith principal sub-matrix of $\boldsymbol{R}$ (obtained from $\boldsymbol{R}$ by removing row and

column i). There are several options for the prior distribution on each σ_i (see O'Malley and Zaslavsky, 2008). In this paper, we assume each σ_i is distributed as a folded standard normal:

$$\sigma_i \overset{i.i.d}{\sim} N^+(0, 1) \qquad (8)$$

for $i = 1, \ldots, k$.

Thus, the standard deviations are not affected by the degrees of freedom by construction, as is reflected in the first row of Figure 4. Similarly, the correlations are independent of the standard deviations, as is reflected in the second row. However, the univariate and joint distributions of the correlations are similar to those from the previous inverse-Wishart distribution because they depend on the same degrees of freedom parameter, ν. Increasing ν would move the marginal distribution of a correlation from a uniform distribution toward a peaked distribution around zero. Again, we see a star-like pattern in the bivariate scatterplots. Thus, despite some increased flexibility, this distribution has many of the same problems as the previous inverse-Wishart distribution.

3.3 Separation strategy with a uniform prior distribution on the correlation matrix

An alternative distribution for $\boldsymbol{\Sigma}$ is obtained with a separation strategy in which the matrix $\boldsymbol{R}$ in Equation 6 has a joint uniform distribution (Barnard et al., 2000) on its support, the space of all $k \times k$ correlation matrices. Here we again assume each σ_i has a folded standard normal distribution.

Investigating the properties of such distribution requires an efficient algorithm to draw a correlation matrix uniformly, which is not straightforward for $k \geq 3$ due to the positive semi-definiteness restriction. Recently, Joe (2006) proposed such an algorithm that was further refined by Lewandowski et al. (2009) where the correlation matrix $\boldsymbol{R}$ is a bijective function of $(k-1)(k-2)/2$ partial correlations. Partial correlations are correlations between the residuals of two variables when each is regressed on some subset of the other variables. The simplest approach is to condition on all variables before the ith when defining the partial correlation between the ith and jth variables where $i < j$. In that case, each partial correlation can be drawn independently from a symmetric Beta distribution spread over the $(-1, 1)$ interval with both shape parameters equal to $\alpha_i = \eta + (k - i - 1)/2$ where $\eta > 0$ is
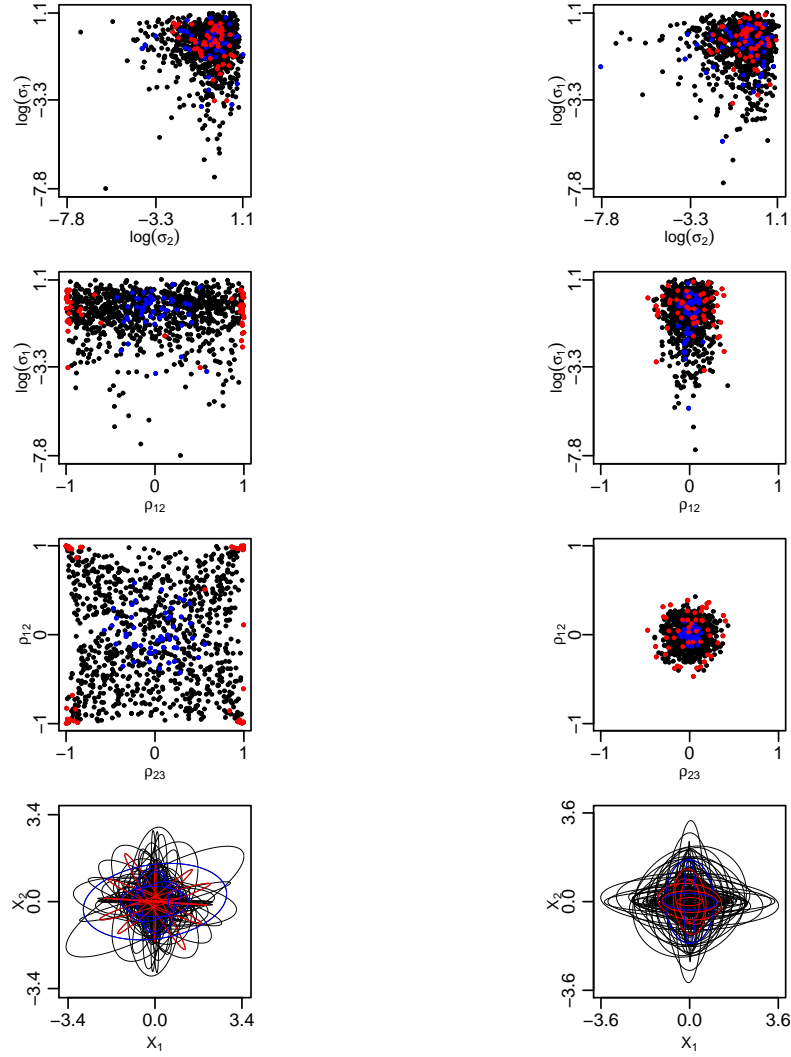
Figure 4: Visualizing a covariance matrix distribution based on decomposing an inverse-Wishart distributed covariance matrix into a correlation matrix and standard deviations. In the left column, $\nu = k + 1$ while $\nu = k + 50$ in the right column and the standard deviations have a folded standard normal distribution. The number of realizations is 1000 (only 100 50% equiprobability ellipses in the final row are shown) and the dimension is $k = 4$ in both cases.

a hyperparameter. Lewandowski et al. (2009) prove that

$$p(\boldsymbol{R}|\eta) = c(k,\eta)|\boldsymbol{R}|^{\eta-1} \tag{9}$$

where $c(k,\eta)$ is a normalization constant first given in Joe (2006). Thus, iff $\eta = 1$, then $\boldsymbol{R}$ has a joint uniform distribution, and Lewandowski et al. (2009) prove that the marginal distribution of each correlation is symmetric Beta over the $(-1,1)$ interval with both shape parameters $\alpha = k/2$. Thus, the marginal distribution of each correlation becomes more concentrated around zero as k increases in order to satisfy the positive semi-definiteness constraint on the correlations jointly. However, the distribution of $\boldsymbol{\Sigma}$ is not uniform because its density also depends on the realizations of the standard deviations.

We visualize the implied distribution of $\boldsymbol{\Sigma}$ in Figure 5 with $k = 4$ and $k = 50$. Again, the correlations are independent of the standard deviations by construction. However, unlike the distributions that we have seen thusfar, the scatterplot of pairs of correlations does not follow a star-like shape. Rather, the correlations seem to be less dependent on each other. Also, a comparison of the evolution of $1 - |\boldsymbol{R}_{p,k}|^{\frac{1}{p}}$ as a function of p reveals two important points. First, the range of the effective dependencies is large for small k and very narrow for large k. In fact, the distribution of the effective dependence converges to $1 - \exp(-1) = 0.631$ as $k \to \infty$ (see Proof 2 in Appendix). Second, for large k, the effective dependence of the leading submatrix increases almost linearly but increases more erratically for small k.

3.4 The scaled inverse-Wishart distribution

As a final separation strategy, the scaled inverse-Wishart distribution makes use of an over-parametrized distribution (in which there is a trade-off between sets of parameters). The covariance matrix is decomposed as follows (see O'Malley and Zaslavsky, 2008):

$$\boldsymbol{\Sigma} = \text{diag}(\xi_1,\ldots,\xi_k) \cdot \boldsymbol{Q} \cdot \text{diag}(\xi_1,\ldots,\xi_k). \tag{10}$$

$\boldsymbol{Q}$ has an inverse-Wishart distribution with degrees of freedom ν and identity scale matrix:

$$\boldsymbol{Q} \sim \text{Inv-Wishart}_{\nu}(\boldsymbol{I}). \tag{11}$$
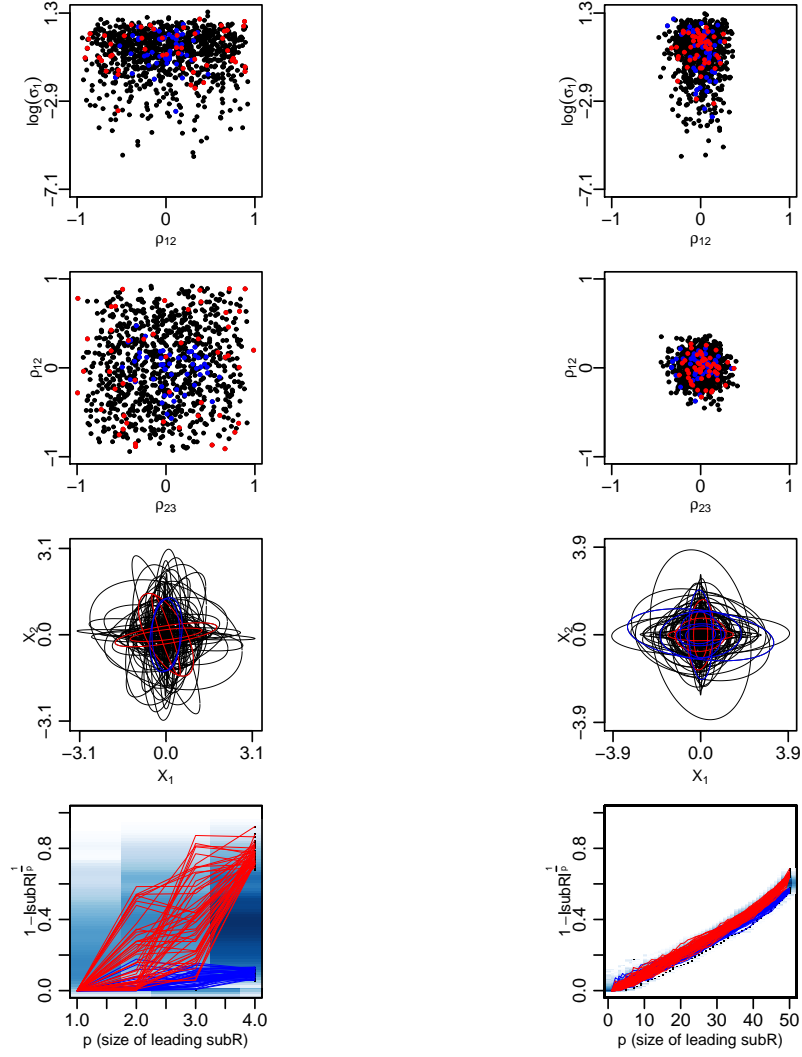
Figure 5: Visualizing a covariance matrix distribution based on a separation strategy with a uniformly distributed correlation matrix and folded standard normal standard deviations. In the left column, $k = 4$ and $k = 50$ in the right column. The number of realizations is 1000 (only 100 50% equiprobability ellipses in the third row are shown).

In this case, ξ_i is not a standard deviation because $\boldsymbol{Q}$ does not have ones on its diagonal. In other words, the ith standard deviation is $\sqrt{\xi_i^2 Q_{i,i}}$. Nevertheless, we take the distribution for each ξ_i to be folded standard normal.
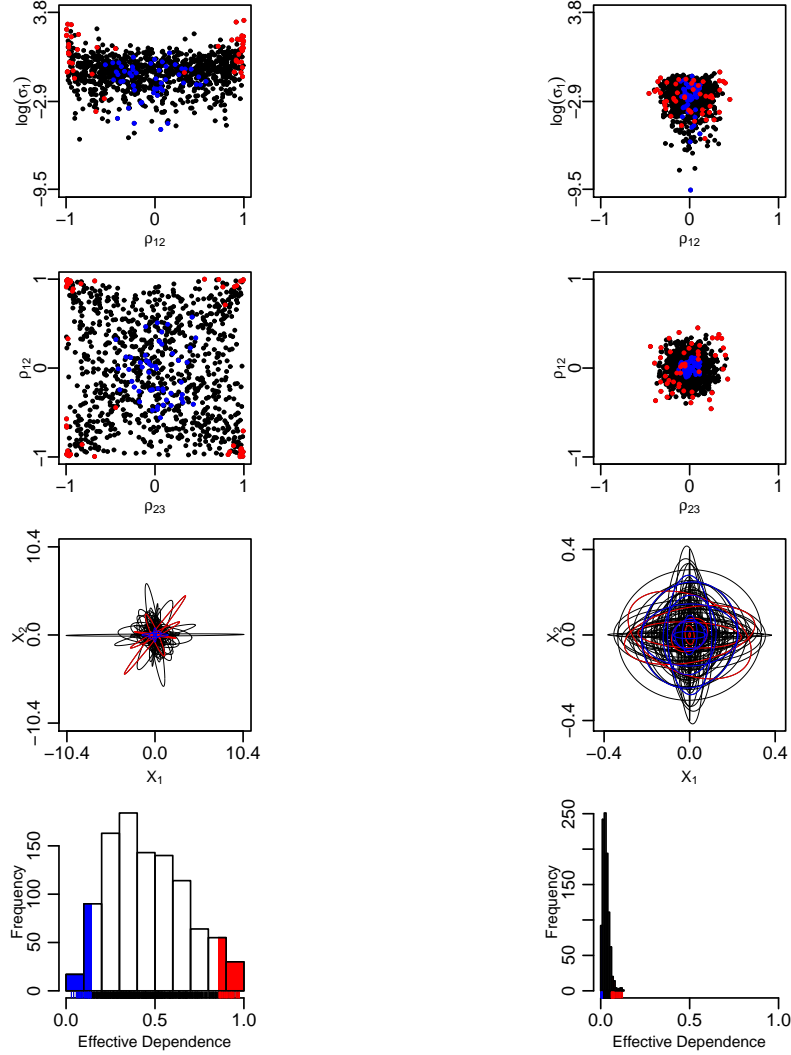

Figure 6: The scaled inverse-Wishart with $\nu = k + 1$ (left column) and $\nu = k + 50$ (right column) with $k = 4$. The number of realizations is 1000 (only 100 50% equiprobability ellipses in the third row are shown).

The scaling operation has no effect on the correlations, so the correlational properties of the scaled inverse-Wishart distribution are the same as those of the unscaled inverse-Wishart

distribution. One motivation for the scaled inverse Wishart distribution is to mitigate the dependence between the standard deviations and the correlations that plagues the unscaled inverse-Wishart distribution. However, the scaling does not completely eliminate this dependence.

In Figure 6, similar patterns as for the unscaled inverse-Wishart are seen in the scatter and contour plots for $\nu = k + 1$. The dependence between the standard deviation and correlation is weaker, confirming the aforementioned property. When $\nu = k + 100$, the plots are quite different from those of the unscaled inverse-Wishart because the distribution of $\boldsymbol{\Sigma}$ is dominated by $\boldsymbol{\xi}$ (with large variability) as the correlations tends toward zero.

3.5   The Wishart distribution

The Wishart distribution, prominent in multivariate statistics (see Johnson and Wichern, 2007; Press, 1982), is the sampling distribution of $(n-1)\boldsymbol{S}$, where $\boldsymbol{S}$ is the $k \times k$ sample covariance matrix calculated from a sample of size n normal observations on k variables with mean vector $\boldsymbol{\mu}_0$ and population covariance matrix $\boldsymbol{\Sigma}_0$ (see Wishart, 1928). As can be seen from Figure 7, the marginal distribution of a covariance matrix depends on k (when $\nu = k + 1$). As the dimension increases, there is less variability in standard deviations, correlations and effective dependence. Moreover, Proof 3 in the Appendix shows that the effective dependence again converges to $1 - \exp(-1)$ as k diverges. If one seeks similar marginal distributions for the standard deviations as k varies, different values of ν must be used (and not always $\nu = k + 1$).

3.6   A new distribution

Our visualization tool may help in the evaluation of customized distributions. For example, assume that we define the following prior distribution on covariance matrices:

$$\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_k) \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{D} \cdot \boldsymbol{\Lambda}' \cdot \operatorname{diag}(\sigma_1, \ldots, \sigma_k). \tag{12}$$

where $\boldsymbol{\Lambda}$ is a $k \times k$ randomly generated orthogonal matrix (draw a $k \times k$ matrix $\boldsymbol{W}$ with standard normal deviates, calculate the singular value decomposition $\boldsymbol{W} = \boldsymbol{UDV}'$ and then set $\boldsymbol{\Lambda} = \boldsymbol{UV}'$), $\boldsymbol{D}$ is a diagonal matrix of eigenvalues whose marginal distribution can
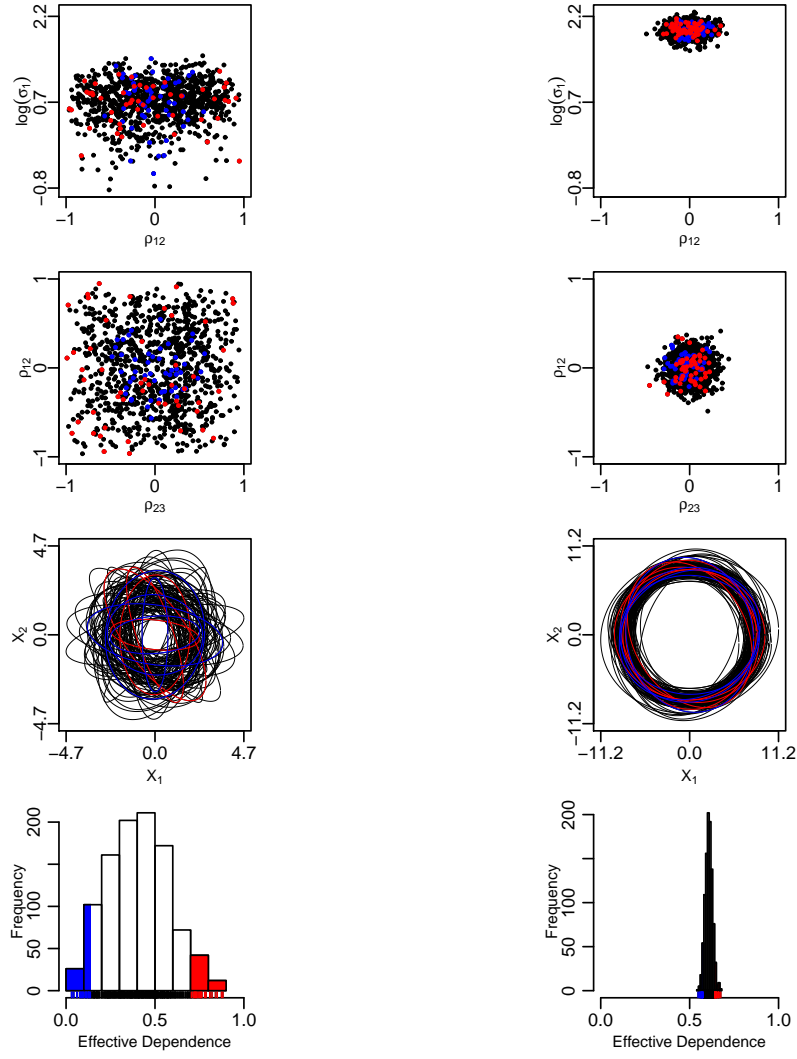
Figure 7: Wishart distribution with $\nu = k + 1$ where $k = 4$ (left column) and $k = 50$ (right column). The number of realizations is 1000 (only 100 50% equiprobability ellipses in the third row are shown).

be any distribution whose support is positive. Here we draw each diagonal element of $\boldsymbol{D}$ from a $\text{Beta}(0.5, 5)$ distribution, which tends to yield a few eigenvalues close to one and many eigenvalues close to zero. Again, each σ_i is a standard deviation and has a folded standardized normal distribution.

We investigate the properties of this customized distribution in Figure 8. The correlations are concentrated around zero but not as strongly as we have seen with other covariance distributions. The effective dependence is a function of the eigenvalues only and is centered around 0.69 (a value close to the limiting value for the uniformly distributed correlation matrices as k diverges).

4   Comparison of distributions

In order to select a prior in an empirical application, researchers must know the properties of the various choices for a covariance distribution. In this section, we compare these four distributions with regard to the marginal and joint distributions of the covariances and the dependence on k.

First, for the inverse-Wishart distribution, as ν gets larger, the correlations are concentrated around zero. For the scaled inverse-Wishart, a similar pattern is evident for the correlation when $\nu = k + 1$, but the standard deviations depend heavily on the prior for each ξ_i. In the previous section, we drew these scale parameters independently from a folded standardized normal distribution, so the dependence among the standard deviations is small and driven by $\boldsymbol{Q}$. When $\boldsymbol{R}$ is the marginal correlation matrix of the inverse-Wishart distributed covariance matrix, the correlations are more dependent than when $\boldsymbol{R}$ is given a joint uniform distribution.

Second, the size of the covariance matrix affects some of its properties in different ways across distributions. For the distributions derived from the inverse-Wishart, we can invoke an exchangeability property so only a few plots are needed to understand univariate and bivariate properties of the covariances, even if k is large. However, the effective dependence is a function of k and its sampling distribution depends on the distribution of the covariance matrix. In particular, if $\boldsymbol{R}$ is given a joint uniform distribution, then the effective dependence is bounded away from units as k diverges, which is not the case for the inverse-Wishart distribution. If a researcher wants to use a prior that is dependent on the dimensionality,
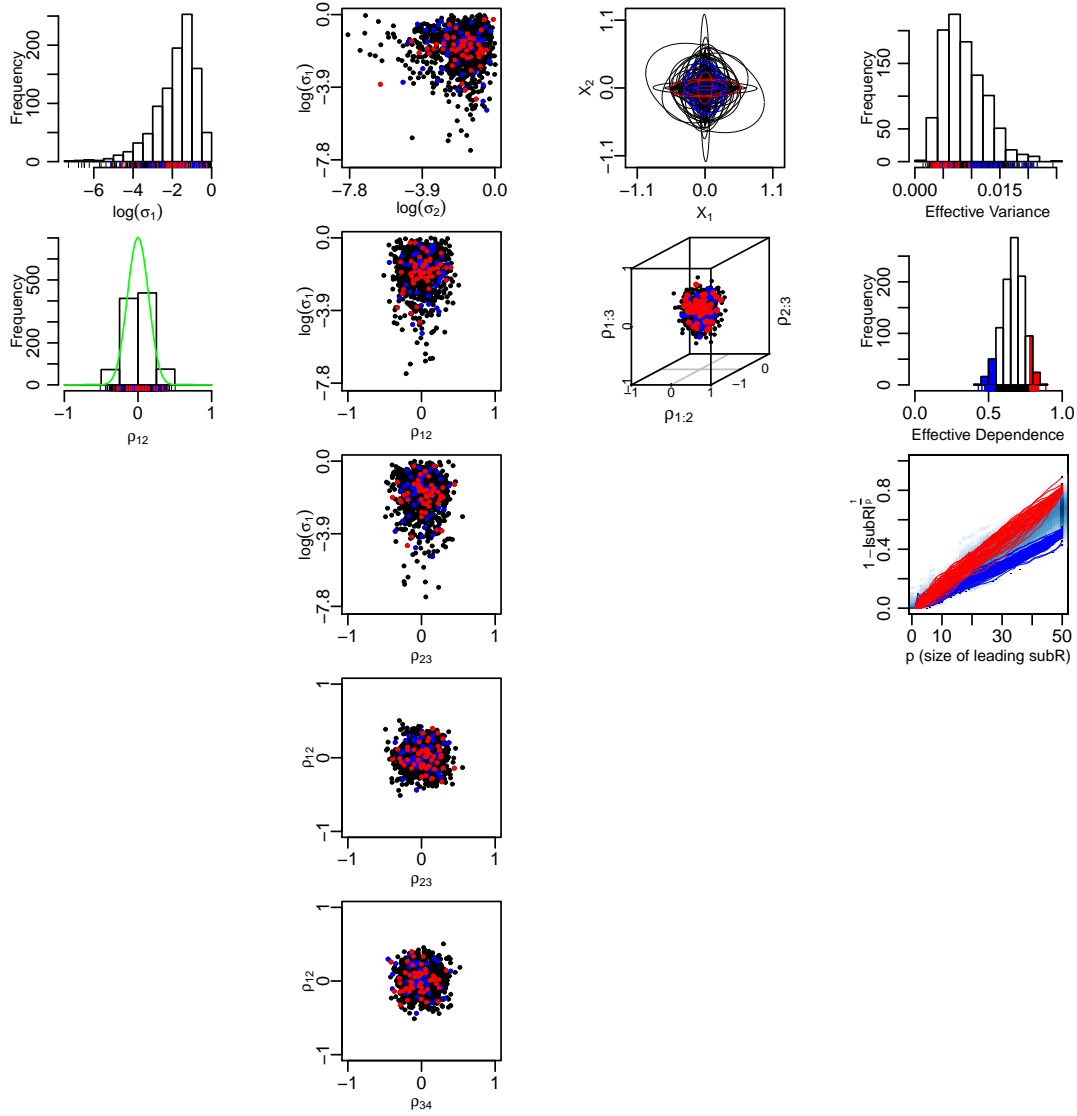
Figure 8: Visualization of a customized distribution of covariance matrices. To construct the plot, 1000 covariance matrices (dimension $k = 50$) were simulated and plotted.

our visualization tool can be used to gauge its properties.

5 Conclusion

We have introduced a four-layered visualization method for a distribution of covariance matrices by using histograms, scatterplots and contour plots. The four layers of plots complement each other, enabling a researcher to visualize a distribution of covariance matrix from different perspectives. As we have seen in the examples, this novel method of visualization effectively reveals important properties of distributions, which are not known from analytical results.

This method can be applied to any proper distribution for covariance matrices. Thus, the method is useful not only to deepen our understanding of existing statistical models, but also in understanding newly proposed distributional families.

An obvious limitation of the method is that it cannot visualize at once a family of distributions that is dependent on the dimensionality of the variable set; in that case, it is necessary to draw plots for at least several values of k.

Appendix

A   Analytical results of the effective dependence

A1   Proof 1: $\lim_{k\to\infty} 1 - E\left(|\boldsymbol{R}|^{\frac{1}{k}}\right) = 1$ for an inverse-Wishart distribution

We consider the marginal distribution of the correlation matrix $\boldsymbol{R}$ derived from $\boldsymbol{\Sigma}$ that has an inverse-Wishart distribution with scale matrix $\boldsymbol{I}_k$. We need to consider a transformation of variables from $\boldsymbol{\Sigma}$ to the correlation matrix $\boldsymbol{R}$ and the standard deviations $\boldsymbol{S} = \mathrm{diag}(\xi_1, \ldots, \xi_k)$. Since the Jacobian of this transformation is given by $2^k(\prod_{i=1}^{k} \xi_i)^k$

(Barnard et al., 2000), we have:

$$f(\boldsymbol{R}, \boldsymbol{S}) = f(\boldsymbol{\Sigma}) \times \text{Jacobian}$$

$$= c_1^{-1} |\boldsymbol{\Sigma}|^{-\frac{\nu+k+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1})\right) 2^k (\prod_{i=1}^{k} \xi_i)^k$$

$$= c_1^{-1} 2^k |\boldsymbol{R}|^{-\frac{\nu+k+1}{2}} \prod_{i=1}^{k} \xi_i^{-(\nu+1)} \exp\left(-\frac{r^{ii}}{2\xi_i^2}\right)$$

where r^{ii} is the ith diagonal element of $\boldsymbol{R}^{-1}$ and c_1 is the normalizing constant for the inverse-Wishart. Integrating out $\boldsymbol{S}$ gives:

$$f(\boldsymbol{R}) = \int_0^\infty f(\boldsymbol{R}, \boldsymbol{S}) d\xi_1 \ldots d\xi_k$$

$$= 2^{\frac{k\nu}{2}} c_1^{-1} \Gamma\left(\frac{\nu}{2}\right)^k |\boldsymbol{R}|^{\frac{1}{2}(\nu-1)(k-1)-1} \prod_{i=1}^{k} |\boldsymbol{R}_{ii}|^{-\frac{\nu}{2}}$$

where $\boldsymbol{R}_{ii}$ is the ith principal submatrix of $\boldsymbol{R}$.

Because $c_1 = 2^{\frac{k\nu}{2}} \Gamma_k(\frac{\nu}{2})$, the density function of $\boldsymbol{R}$ is:

$$f(\boldsymbol{R}) = \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} |\boldsymbol{R}|^{\frac{1}{2}(\nu-1)(k-1)-1} \prod_{i=1}^{k} |\boldsymbol{R}_{ii}|^{-\frac{\nu}{2}} \tag{A1}$$

where $\Gamma_k(\cdot)$ is the k-dimensional multivariate gamma function. By definition,

$$E(|\boldsymbol{R}|^{1/k}) = \int |\boldsymbol{R}|^{\frac{1}{k}} f(\boldsymbol{R}) d\boldsymbol{R}$$

$$= \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \int |\boldsymbol{R}|^{\frac{1}{2}(\nu-1)(k-1)-1+\frac{1}{k}} \prod_{i=1}^{k} |\boldsymbol{R}_{ii}|^{-\frac{\nu}{2}} d\boldsymbol{R}$$

Now, we obtain the upper bound of $E(|\boldsymbol{R}|^{\frac{1}{k}})$ in terms of k by applying Hölder's inequality,

letting $p = k^2/2$ and $q = k^2/(k^2 - 2)$.

$$E(|\boldsymbol{R}|^{\frac{1}{k}}) = \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \int |\boldsymbol{R}|^{-\frac{1}{k^2}} (|\boldsymbol{R}|^{\frac{1}{2}(\nu q - 1)(k-1)-1} \prod_{i=1}^{k} |\boldsymbol{R}_{ii}|^{-\frac{\nu q}{2}})^{\frac{1}{q}} d\boldsymbol{R}$$

$$\leq \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \left\{ \int |\boldsymbol{R}|^{-\frac{1}{k^2}\frac{k^2}{2}} d\boldsymbol{R} \right\}^{\frac{2}{k^2}} \left\{ \int |\boldsymbol{R}|^{\frac{1}{2}(\nu q - 1)(k-1)-1} \prod_{i=1}^{k} |\boldsymbol{R}_{ii}|^{-\frac{\nu q}{2}} d\boldsymbol{R} \right\}^{1/q}$$

Evaluating the first integral by the normalizing constant in Joe (2006) and the second by Equation A1,

$$E(|\boldsymbol{R}|^{\frac{1}{k}}) \leq \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \left\{ 2^{\sum_{i=1}^{k-1}(k-i-1)(k-i)} \right\}^{\frac{2}{k^2}} \left\{ \prod_{i=1}^{k-1} \mathrm{Beta}\left(\frac{k-i}{2}, \frac{k-i}{2}\right)^{k-i} \right\}^{\frac{2}{k^2}} \frac{\Gamma_k\left(\frac{\nu k^2}{2(k^2-2)}\right)^{\frac{k^2-2}{k^2}}}{\Gamma\left(\frac{\nu k^2}{2(k^2-2)}\right)^{k \times \frac{k^2-2}{k^2}}}.$$

It can be shown that:

$$\lim_{k \to \infty} \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \frac{\Gamma_k\left(\frac{\nu k^2}{2(k^2-2)}\right)^{\frac{k^2-2}{k^2}}}{\Gamma\left(\frac{\nu k^2}{2(k^2-2)}\right)^{k \times \frac{k^2-2}{k^2}}} = 1.$$

Further, by using the approximation $\mathrm{Beta}(x,x) \sim \sqrt{2\pi}\frac{x^{2x-1}}{(2x)^{2x-\frac{1}{2}}}$ (derived by applying the Stirling's formula), we get:

$$\log\left\{ 2^{\sum_{i=1}^{k-1}(k-i-1)(k-i)} \prod_{i=1}^{k-1} \mathrm{Beta}\left(\frac{k-i}{2}, \frac{k-i}{2}\right)^{k-i} \right\}^{\frac{2}{k^2}}$$

$$\approx \frac{2}{k^2} \sum_{i=1}^{k-1}(k-i)\left\{ \frac{1}{2}\log 2\pi - \frac{1}{2}\log(k-i) \right\}.$$

As $k \to \infty$, the right-hand side of the inequality goes to $-\infty$. In sum, it has been shown that the upper bound of $E(|\boldsymbol{R}|^{\frac{1}{k}})$ goes to 0 as $k \to \infty$. Since $E(|\boldsymbol{R}|^{\frac{1}{k}}) \geq 0$, it implies that $\lim_{k \to \infty} E(|\boldsymbol{R}|^{\frac{1}{k}}) = 0$, leading to $\lim_{k \to \infty}(1 - E(|\boldsymbol{R}|^{\frac{1}{k}})) = 1$.

A2 Proof 2: For a uniformly distributed correlation matrix, the effective dependence converges to $1 - \exp(-1)$

We consider the correlation matrix $\boldsymbol{R}$, which follows a uniform distribution. By definition,

$$E(|\boldsymbol{R}|^{\frac{1}{k}}) = \int |\boldsymbol{R}|^{\frac{1}{k}} f(\boldsymbol{R}) d\boldsymbol{R}$$

$$= c_2^{-1} \int |\boldsymbol{R}|^{\frac{1}{k}} d\boldsymbol{R}$$

where $c_2 = 2^{\sum_i^{k-1} i^2} \prod_{i=1}^{k-1} \{\text{Beta}(\frac{1}{2}(i+1), \frac{1}{2}(i+1))\}^i$ (Joe, 2006).

Using the normalizing constant to evaluate the integral above, we get:

$$\int |\boldsymbol{R}|^{\frac{1}{k}} d\boldsymbol{R} = 2^{\sum_i^{k-1} (\frac{2}{k}+i)i} \prod_{i=1}^{k-1} \left\{ \text{Beta}\left(\frac{1}{2}(i+1) + \frac{1}{k}, \frac{1}{2}(i+1) + \frac{1}{k} \right) \right\}^i$$

As a consequence, $E(|\boldsymbol{R}|^{\frac{1}{k}})$ becomes:

$$E(|\boldsymbol{R}|^{\frac{1}{k}}) = \frac{2^{\sum_i^{k-1} (\frac{2}{k}+i)i} \prod_{i=1}^{k-1} \{\text{Beta}(\frac{1}{2}(i+1) + \frac{1}{k}, \frac{1}{2}(i+1) + \frac{1}{k})\}^i}{2^{\sum_i^{k-1} i^2} \prod_{i=1}^{k-1} \{\text{Beta}(\frac{1}{2}(i+1), \frac{1}{2}(i+1))\}^i}$$

Next, we approximate the logarithm of $E(|\boldsymbol{R}|^{\frac{1}{k}})$ by a first-order Taylor expansion (a prime $'$ refers to a first derivative):

$$\log(E(|\boldsymbol{R}|^{\frac{1}{k}})) = (k-1)\log 2 + \sum_{i=1}^{k-1} i \left\{ \log \text{Beta}\left(\frac{1}{2}(i+1) + \frac{1}{k}, \frac{1}{2}(i+1) + \frac{1}{k} \right) - \right.$$

$$\left. \log \text{Beta}\left(\frac{1}{2}(i+1), \frac{1}{2}(i+1) \right) \right\}$$

$$\approx (k-1)\log 2 + \sum_{i=1}^{k-1} \frac{i}{k} \log \text{Beta}\left(\frac{1}{2}(i+1), \frac{1}{2}(i+1) \right)'$$

$$= (k-1)\log 2 + \sum_{i=1}^{k-1} \frac{i}{k} \frac{\text{Beta}\left(\frac{1}{2}(i+1), \frac{1}{2}(i+1) \right)'}{\text{Beta}\left(\frac{1}{2}(i+1), \frac{1}{2}(i+1) \right)}$$

$$= (k-1)\log 2 + \frac{2}{k} \sum_{i=1}^{k-1} i \left(\psi(\frac{1}{2}(i+1)) - \psi(i+1) \right)$$

where $\psi(\cdot)$ is the digamma function, defined as $\psi(x) = \frac{\Gamma(x)'}{\Gamma(x)}$.

As $\psi(x) = \log x - \frac{1}{2x} + o(\frac{1}{x^2})$ (Abramowitz and Stegun, 1972), we find that:

$$\log(E(|\boldsymbol{R}|^{\frac{1}{k}})) \approx (k-1)\log 2 + \frac{2}{k}\sum_{i=1}^{k-1} i\left\{\log\frac{1}{2}(i+1) - \frac{1}{i+1} - \log(i+1) + \frac{1}{2(i+1)} + o(\frac{1}{i^2})\right\}$$

$$= -\frac{k-1}{k} + \frac{1}{k}\sum_{i=1}^{k-1} o\left(\frac{1}{i}\right)$$

Consequently, this leads to:
$$\lim_{k\to\infty} \log(E(|\boldsymbol{R}|^{\frac{1}{k}})) = -1$$

and also:
$$\lim_{k\to\infty} E(|\boldsymbol{R}|^{\frac{1}{k}}) = \exp(-1)$$

In the same way, it can be shown that $\lim_{k\to\infty} E(|\boldsymbol{R}|^{\frac{2}{k}}) = \exp(-2)$. Thus,

$$\lim_{k\to\infty} \text{Var}(|\boldsymbol{R}|^{\frac{1}{k}}) = \exp(-2) - (\exp(-1))^2 = 0$$

These results imply that the effective dependence converges to $1 - \exp(-1) = 0.631$ in probability.

A3 Proof 3: For a Wishart distributed correlation matrix, the effective dependence converges to $1 - \exp(-1)$

Again, we consider the marginal distribution of the correlation matrix $\boldsymbol{R}$ derived from $\boldsymbol{\Sigma}$ that has a Wishart distribution with scale matrix $\boldsymbol{I}_k$. We need to consider a transformation of variables from $\boldsymbol{\Sigma}$ to the correlation matrix $\boldsymbol{R}$ and the standard deviations $\boldsymbol{S} = \text{diag}(\xi_1, \ldots, \xi_k)$. Since the Jacobian of this transformation is given by $2^k(\prod_{i=1}^k \xi_i)^k$ (Barnard et al., 2000, see also Proof 1),

$$f(\boldsymbol{R}, \boldsymbol{S}) = f(\boldsymbol{\Sigma}) \times \text{Jacobian}$$

$$= c_3^{-1} |\boldsymbol{\Sigma}|^{\frac{\nu-k-1}{2}} \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma})\right) 2^k \left(\prod_{i=1}^{k} \xi_i\right)^k$$

$$= 2^k c_3^{-1} |\boldsymbol{R}|^{\frac{\nu-k-1}{2}} \prod_{i=1}^{k} \xi_i^{\nu-1} \exp\left(-\frac{1}{2}\xi_i^2\right)$$

where c_3 is the normalizing constant for the Wishart. The density function of $\boldsymbol{R}$ is then:

$$f(\boldsymbol{R}) = \int_0^\infty f(\boldsymbol{R}, \boldsymbol{S}) d\xi_1 \dots d\xi_k$$

$$= 2^{\frac{k\nu}{2}} c_3^{-1} \Gamma\left(\frac{\nu}{2}\right)^k |\boldsymbol{R}|^{\frac{\nu-k-1}{2}}$$

As $c_3 = 2^{\frac{k\nu}{2}} \Gamma_k(\frac{\nu}{2})$, we get:

$$f(\boldsymbol{R}) = \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} |\boldsymbol{R}|^{\frac{\nu-k-1}{2}}$$

where $\Gamma_k(\cdot)$ is the k-dimensional multivariate gamma function. By definition, this leads to:

$$E(|\boldsymbol{R}|^{1/k}) = \int |\boldsymbol{R}|^{\frac{1}{k}} f(\boldsymbol{R}) d\boldsymbol{R}$$

$$= \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \int |\boldsymbol{R}|^{\frac{\nu-k-1}{2}+\frac{1}{k}} d\boldsymbol{R}$$

Using the normalizing constant to evaluate the integral above (Joe, 2006), the result is:

$$E(|\boldsymbol{R}|^{1/k}) = \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} 2^{\sum_{i=1}^{k-1}(\nu-k-1+\frac{2}{k}+i)i} \tag{A2}$$

$$\prod_{i=1}^{k-1} \left\{ \text{Beta}\left(\frac{\nu-k-1}{2} + \frac{1}{k} + \frac{i+1}{2}, \frac{\nu-k-1}{2} + \frac{1}{k} + \frac{i+1}{2}\right) \right\}^i$$

As $f(\boldsymbol{R})$ is a density function, $\int f(\boldsymbol{R}) d\boldsymbol{R} = 1$ and it follows:

$$\int \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} |\boldsymbol{R}|^{\frac{\nu-k-1}{2}} d\boldsymbol{R} = \frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} \int |\boldsymbol{R}|^{\frac{\nu-k-1}{2}} d\boldsymbol{R} = 1$$

Again, by using the results of Joe (2006), we get:

$$\frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} 2^{\sum_{i=1}^{k-1}(\nu-k-1+i)i} \prod_{i=1}^{k-1} \left\{ \text{Beta}(\frac{\nu-k-1}{2} + \frac{i+1}{2}, \frac{\nu-k-1}{2} + \frac{i+1}{2}) \right\}^i = 1$$

This implies that:

$$\frac{\Gamma(\frac{\nu}{2})^k}{\Gamma_k(\frac{\nu}{2})} = \frac{1}{2^{\sum_{i=1}^{k-1}(\nu-k-1+i)i} \prod_{i=1}^{k-1} \left\{ \text{Beta}(\frac{\nu-k-1}{2} + \frac{i+1}{2}, \frac{\nu-k-1}{2} + \frac{i+1}{2}) \right\}^i}$$

Substituting the previous result in Equation A2 gives

$$E(|\boldsymbol{R}|^{\frac{1}{k}}) = \frac{2^{\sum_i^{k-1}(c-1+\frac{2}{k}+i)i} \prod_{i=1}^{k-1} \left\{ \text{Beta}(\frac{1}{2}(i+c) + \frac{1}{k}, \frac{1}{2}(i+c) + \frac{1}{k}) \right\}^i}{2^{\sum_i^{k-1}(c-1+i)i} \prod_{i=1}^{k-1} \left\{ \text{Beta}(\frac{1}{2}(i+c), \frac{1}{2}(i+c)) \right\}^i}$$

In the same way as in Proof 2, it can be shown that:

$$\log(E(|\boldsymbol{R}|^{\frac{1}{k}})) \approx (k-1)\log 2 + \frac{2}{k} \sum_{i=1}^{k-1} i \left\{ \log\frac{1}{2}(i+c) - \frac{1}{i+c} - \log(i+c) + \frac{1}{2(i+c)} + o\left(\frac{1}{i^2}\right) \right\}$$

$$= -\frac{k-1}{k} + \frac{1}{k} \sum_{i=1}^{k-1} o\left(\frac{1}{i}\right)$$

Thus, it follows that:
$$\lim_{k\to\infty} \log(E(|\boldsymbol{R}|^{\frac{1}{k}})) = -1$$

and also that:
$$\lim_{k\to\infty} E(|\boldsymbol{R}|^{\frac{1}{k}}) = \exp(-1)$$

In the same way, it can be shown that $\lim_{k\to\infty} E(|\boldsymbol{R}|^{\frac{2}{k}}) = \exp(-2)$. Thus,

$$\lim_{k\to\infty} \text{Var}(|\boldsymbol{R}|^{\frac{1}{k}}) = \exp(-2) - (\exp(-1))^2 = 0$$

These results imply that the effective dependence for a Wishart distribution converges to $1 - \exp(-1) = 0.631$ in probability.

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs and mathematical tables.* New York, Dover Publications.

Anderson, T. (2003). *An introduction to multivariate statistical analysis.* New York, Wiley.

Barnard, J., McCulloch, R., and Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10:1281–1311.

Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis.* New York, Wiley.

Cook, D. and Swayne, D. (2007). *Interactive and dynamic graphics for data analysis.* Springer.

Daniels, M. and Kass, M. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94:1254–1263.

Eaton, M. (1983). *Multivariate statistics: a vector space approach.* New York, Wiley.

Friendly, M. (2002). Corrgrams: exploratory displays for correlation matrices. *The American Statistician*, 56:316–324.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian data analysis.* London, CRC press.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97:2177–2189.

Johnson, R. and Wichern, D. (2007). *Applied multivariate statistical analysis.* Saddle River, NJ, Prentice Hall.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100:1989–2001.

O'Malley, A. and Zaslavsky, A. (2008). Domain-level covariance analysis for survey data with structured nonresponse. *Journal of the American Statistical Association*, 103:1405–1418.

Oravecz, Z., Tuerlinckx, F., and Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, 74:395–418.

Owen, J. and Rabinovitch, R. (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance*, 38:745–752.

Peña, D. and Rodríguez, J. (2003). Descriptive measures of multivariate scatter and linear dependence. *Journal of Multivariate Analysis*, 85:361–374.

Press, S. (1982). *Applied multivariate analysis: using Bayesian and frequentist methods of inference.* Malabar, Fla., Krieger.

Rousseeuw, P. and Molenberghs, G. (1994). The shape of correlation matrices. *The American Statistician*, 48:276–279.

Theus, M. and Urbanek, S. (2008). *Interactive graphics for data analysis: principles and examples.* Chapman & Hall/CRC.

Valero-Mora, P., Young, F., and Friendly, M. (2003). Visualizaing categorical data in ViSta. *Computational Statistics & Data Analysis*, 43:495–508.

Von Rosen, D. (1988). Moments for the inverted Wishart distribution. *Scandinavian Journal of Statistics*, 15:97–109.

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–52.

Yang, R. and Berger, J. (1994). Estimation of a covariance matrix using the reference prior. *Annals of Statistics*, 22:1195–1211.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* New York, Wiley.