Rutger van Oest (2019) Unconstrained Cholesky-based parametrization of correlation matrices, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2019.1628271

Unconstrained Cholesky-based parametrization of correlation matrices

Rutger van Oest

BI Norwegian Business School

Nydalsveien 37

0484 Oslo, Norway

Email: rutger.d.oest@bi.no

Communications in Statistics – Simulation and Computation

Unconstrained Cholesky-based parametrization of correlation matrices

Abstract:

Parameter estimation is relatively complicated for models containing correlation matrices, because the elements of correlation matrices are heavily constrained. We put forward a Cholesky-based parametrization that is easy to implement and allows for unconstrained parameter estimation. To compare the new parametrization with the commonly applied spherical parametrization, we use Monte Carlo simulation in which we estimate multivariate distributions containing Gaussian copulas. We show that the new parametrization performs well, in particular as the dimensionality of the multivariate distribution increases, computing times increase, and non-convergence occurs increasingly often.

Keywords:

Correlation matrices; Cholesky decomposition; Unconstrained parametrization; Simulation

1. Introduction

Applied statistical modeling often involves estimation of the elements of a correlation matrix alongside other model parameters. For example, variables of interest co-vary and frequently have marginal distributions that do not imply an elliptical (e.g., multivariate normal) joint distribution, requiring copulas to adequately capture the underlying dependence (Nelsen, 2006; Sklar, 1959). Copulas are becoming increasingly popular, and "the so-called 'elliptical copula' have proven the most popular in applied modeling" (Danaher and Smith, 2011, p. 9). A convenient and frequently used elliptical copula is the Gaussian copula that contains a correlation matrix to link the marginal distributions (Pitt et al., 2006; Song, 2000).

Numerical maximization of the log-likelihood function to obtain estimates of the correlation matrix (and other model parameters) usually requires an extensive search in the parameter space. To avoid that infeasible parameter values cause the search algorithm to break down, the parametrization needs to become unconstrained (Pinheiro and Bates, 1996). Obviously, the correlation matrix poses a challenge, because this matrix needs to be positive definite, have all diagonal elements equal to one, and have all off-diagonal elements bounded by −1 and +1. For example, eigenspectrum decomposition may be used for unconstrained parametrization of *covariance* matrices, despite "considerable calculations" (Pinheiro and Bates, 1996, p. 292), but only captures positive definiteness and is unable to capture the additional restrictions that are present in correlation matrices.

Although unconstrained parametrizations of correlation matrices exist, such as partial correlations (Joe, 2006), these approaches tend to be relatively complicated. Therefore, "the general correlated case is typically computed by using spherical parametrizations" (Madar, 2015, p.142), which conveniently transform the Cholesky decomposition into spherical coordinates in order to make the parametrization unconstrained (e.g., Rebonato and Jäckel, 2000; Tsay and Pourahmadi, 2017).

The present paper introduces an easy Cholesky-based parametrization that has not been documented before. We use Monte Carlo simulation, with Gaussian copulas, to compare the new parametrization with the spherical parametrization. The results indicate that the relative performance of the new parametrization improves as the dimensionality of the multivariate distribution increases and the estimation problem becomes more complex.

2. Cholesky decomposition of correlation matrices

Let

$$R = (r_{i,j}) = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,M-1} & r_{1,M} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,M-1} & r_{2,M} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{M-1,1} & r_{M-1,2} & \cdots & r_{M-1,M-1} & r_{M-1,M} \\ r_{M,1} & r_{M,2} & \cdots & r_{M,M-1} & r_{M,M} \end{pmatrix}$$

denote the $M \times M$ positive definite correlation matrix, where $r_{i,j} = r_{j,i}$ due to symmetry, $r_{i,i} = 1$, $i = 1, \ldots, M$, and $r_{i,j} \in (-1, 1)$ if $i \neq j$. Because R is symmetric and positive definite, it has a Cholesky decomposition $R = LL'$, where

$$L = (l_{i,j}) = \begin{pmatrix} l_{1,1} & 0 & \cdots & 0 & 0 \\ l_{2,1} & l_{2,2} & & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ l_{M-1,1} & l_{M-1,2} & \cdots & l_{M-1,M-1} & 0 \\ l_{M,1} & l_{M,2} & \cdots & l_{M,M-1} & l_{M,M} \end{pmatrix}$$

is a unique lower triangular matrix with all diagonal elements taking positive values. Writing out $R = LL'$ and using that L is lower triangular yields

$$r_{i,j} = \sum_{k=1}^{j} l_{i,k} l_{j,k}, \quad j \leq i, \tag{1}$$

immediately implying that

$$r_{i,i} = \sum_{k=1}^{i} {l_{i,k}}^2, \quad i = 1, \ldots, M, \tag{2}$$

It follows from $r_{i,i} = 1$ and $l_{i,i} > 0$, $i = 1, \ldots, M$, that (2) can be rewritten as

$$l_{i,i} = \sqrt{1 - \sum_{k=1}^{i-1} {l_{i,k}}^2}, \quad i = 1, \dots, M, \tag{3}$$

which shows that the diagonal elements $l_{i,i} = \sqrt{1 - \sum_{k=1}^{i-1} {l_{i,k}}^2}$ are completely determined by the off-diagonal elements $l_{i,j}$, $j = 1, \dots, i-1$, reducing the number of "free" elements in L to the number of elements below the diagonal, $M(M-1)/2$. Because $r_{i,i} = 1$ in (2), the coordinates $(l_{i,1}, \dots, l_{i,i})$ in the i-th row of L must be located on the i-dimensional unit sphere, with a (squared) Euclidean distance from the origin that is equal to one, which is indeed captured by (3).

3. New parametrization: Cholesky-based and unconstrained

The $M(M-1)/2$ "free" elements of L need to satisfy the restriction that the diagonal elements of L take positive and non-complex values; that is, $l_{i,i} = \sqrt{1 - \sum_{k=1}^{i-1} {l_{i,k}}^2} > 0$, $i = 1, \dots, M$, or equivalently, $\sum_{k=1}^{i-1} {l_{i,k}}^2 < 1$, $i = 2, \dots, M$. Thus, the $(i-1)$-dimensional subset of coordinates $(l_{i,1}, \dots, l_{i,i-1})$ should be inside the $(i-1)$-dimensional unit sphere, with a Euclidean distance from the origin that is less than one. An equivalent condition is that for any $j = 1, \dots, i-1$, the coordinates $(l_{i,1}, \dots, l_{i,j})$ should be located inside the corresponding j-dimensional unit sphere:

$$\sum_{k=1}^{j} {l_{i,k}}^2 < 1, \quad i = 2, \dots, M, \quad j = 1, \dots, i-1,$$

$$l_{i,i} = \sqrt{1 - \sum_{k=1}^{i-1} {l_{i,k}}^2}, \quad i = 1, \dots, M. \tag{4}$$

To ensure that all restrictions in (4) are satisfied, we put forward an alternative lower triangular parametrization for the elements of $L = (l_{i,j})$:

$$\Theta = (\theta_{i,j}) = \begin{pmatrix} \theta_{1,1} & 0 & \dots & 0 & 0 \\ \theta_{2,1} & \theta_{2,2} & & 0 & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \theta_{M-1,M-1} & 0 \\ \theta_{M,1} & \dots & \dots & \theta_{M,M-1} & \theta_{M,M} \end{pmatrix},$$

where $\theta_{i,i} = 1$, $i = 1, \dots, M$, and $\theta_{i,j} \in (-1, 1)$ if $j < i$. We define

$$\theta_{i,j} = \frac{l_{i,j}}{\sqrt{1 - \sum_{k=1}^{j-1} {l_{i,k}}^2}}, \quad i = 2, \dots M, \quad j = 1, \dots, i-1,$$

$$\theta_{i,i} = 1, \quad i = 1, \dots, M, \tag{5}$$

with inverse transformation

$$l_{i,j} = \theta_{i,j} \sqrt{1 - \sum_{k=1}^{j-1} {l_{i,k}}^2}, \quad i = 1, \dots M, \quad j = 1, \dots, i,$$

$$\theta_{i,i} = 1, \quad i = 1, \dots, M. \tag{6}$$

The absolute value of $\theta_{i,j}$ in (5) expresses the absolute value of $l_{i,j}$ as a fraction of the maximum absolute value of $l_{i,j}$ in order to stay within the unit sphere, given the previous $(l_{i,1}, \dots, l_{i,j-1})$. The sign of $\theta_{i,j}$ captures whether $l_{i,j}$ is positive or negative. Note that (6) implies that ${l_{i,j}}^2 < 1 - \sum_{k=1}^{j-1} {l_{i,k}}^2$, and thus $\sum_{k=1}^{j} {l_{i,k}}^2 < 1$, for all $j = 1, \dots, i-1$, and that $l_{i,i} = \sqrt{1 - \sum_{k=1}^{i-1} {l_{i,k}}^2}$. Thus, the transformation from $l_{i,j}$ to $\theta_{i,j}$ satisfies all restrictions in (4).

4. Spherical parametrization: Cholesky-based and unconstrained

The typical approach to obtain unconstrained correlation parameters is the spherical parametrization that transforms the Cholesky decomposition $R = LL'$ into *spherical* coordinates:

$$l_{i,j} = \cos(\theta_{i,j}) \prod_{k=1}^{j-1} \sin(\theta_{i,k}), \quad i = 1, \dots, M, \quad j = 1, \dots, i,$$

$$\theta_{i,i} = 0, \quad i = 1, \dots, M, \tag{7}$$

where $\theta_{i,j} \in (0, \pi)$ for all $i = 2, \dots, M$ and $j = 1, \dots, i-1$ (e.g., Rebonato and Jäckel, 2000; Tsay and Pourahmadi, 2017). For instance, for $M = 2$ dimensions, (7) reduces to $l_{1,1} = 1$, $l_{2,1} = \cos(\theta_{2,1})$, and $l_{2,2} = \sin(\theta_{2,1})$.

Whereas (7) imposes the unit sphere restrictions in spherical coordinates, the new parametrization (6) can be regarded as doing so in *radial-based* coordinates; that is, after applying the radial-based transformation by Bauwens et al. (2004).

5. Monte Carlo simulation

To compare the performance of the new parametrization with the spherical parametrization, we simulate data from non-elliptical distributions that do not necessarily imply marginal distributions from the same distributional family; we use the Gaussian copula to capture the dependence structure (e.g., Danaher and Smith, 2011; Song, 2000).

For $i = 1, \dots, M$, let $f_i(x_{l,i}|\phi_i)$ denote the marginal density of variable x_i, with $x_{l,i}$ being the *l*-th observation of x_i, and let $F_i(x_{l,i}|\phi_i)$, denote the corresponding cumulative distribution function, with parameters ϕ_i. The joint density of $(x_{l,1}, \dots, x_{l,M})$ is the product of the copula density and the marginal densities:

$$g(x_{l,1}, \dots, x_{l,M}|\phi_1, \dots \phi_M, R) = c(x_{l,1}, \dots, x_{l,M}|\phi_1, \dots \phi_M, R) \prod_{i=1}^{M} f_i(x_{l,i}|\phi_i), \qquad (8)$$

where

$$c(x_{l,1}, \dots, x_{l,M}|\phi_1, \dots \phi_M, R)$$
$$= |R|^{-1/2} \exp\left(-\frac{1}{2}\left[\Phi^{-1}\left(F_i(x_{l,i}|\phi_i)\right)\right]'(R^{-1} - I_M)\left[\Phi^{-1}\left(F_i(x_{l,i}|\phi_i)\right)\right]\right) \quad (9)$$

is the Gaussian copula density, Φ^{-1} denotes the inverse of the standard normal cumulative distribution function, and I_M is the $M \times M$ identity matrix (e.g., Danaher and Smith, 2011;

Song, 2000). If $R = I_M$, (9) reduces to $c(x_{l,1}, \dots, x_{l,M} | \phi_1, \dots \phi_M, R) = 1$, which is the independence copula density.

5.1 Simulation setup

We obtain the scenarios by varying the number of marginal distributions in (8) and (9); we take $M = 6$, 12, 18, or 24. In a first set of scenarios, we consider gamma marginal distributions for all variables (e.g., Song, 2000), so that all marginal distributions are from the same family. In a second set of scenarios, we take gamma distributions for one half of the variables, and we take beta distributions for the other half; thus, the marginal distributions are no longer from the same family.

Within each scenario we simulate 1000 data sets, each containing 500 observations. We obtain each data set by first simulating the parameters $(\phi_1, \dots \phi_M, R)$ and then using the algorithm in Danaher and Smith (2011, p. 11) to draw the data $(x_{l,i})$ from (8) and (9). For each gamma distribution, we draw both the shape parameter and the scale parameter from the standard lognormal distribution. Similarly, we draw the two shape parameters of each beta distribution from the standard lognormal distribution. To simulate realistic correlation matrices *R*, we use the algorithm in Madar (2015, p. 145).

For each simulated data set, we estimate the correlation matrix *R* and the parameters of the marginal distributions, $\phi_1, \dots \phi_M$, by maximizing the log-likelihood function:

$$\ln L = \sum_l \left[\ln \left(c(x_{l,1}, \dots, x_{l,M} | \phi_1, \dots \phi_M, R) \right) + \sum_{i=1}^{M} \ln \left(f_i(x_{l,i} | \phi_i) \right) \right], \qquad (10)$$

where $R = LL'$. We use (6) to express *L* in terms of the transformed parameters $\Theta = (\theta_{i,j})$ for the new parametrization; we use (7) for the spherical parametrization. Because $\theta_{i,j} \in (-1, 1)$ in the new parametrization, we write $\theta_{i,j} = (\exp(\tilde{\theta}_{i,j}) - 1)/(\exp(\tilde{\theta}_{i,j}) + 1)$ and estimate unbounded $\tilde{\theta}_{i,j}$. Similarly, $\theta_{i,j} \in (0, \pi)$ in the spherical parametrization; we write $\theta_{i,j} =$

$\pi/(1+\exp(-\tilde{\theta}_{i,j}))$ and estimate unbounded $\tilde{\theta}_{i,j}$ when applying the spherical transformation. For both parametrizations and each simulated data set, we start the numerical maximization of (10) from $\tilde{\theta}_{i,j}=0$. These starting values correspond to the independence copula, with $R=I_M$, which is a natural starting point when the dependence structure is unknown.

< INSERT TABLE 1 ABOUT HERE >

< INSERT TABLE 2 ABOUT HERE >

5.2 Simulation results

We implement the simulation in the programming language Ox (Doornik, 2007); we use Ox's default settings for the search algorithm. Table 1 compares the new parametrization with the spherical parametrization by considering five criteria: the percentage of simulated data sets (i.e., simulation runs) for which the search algorithm stopped without reporting convergence, the average time needed to reach convergence, the upper 10% quantile for time needed to reach convergence, the upper 5% quantile, and the upper 1% quantile. The reported computing times are based on simulation runs for which both the new parametrization and the spherical parametrization resulted in convergence. Thus, the corresponding statistics are not distorted by computing times without convergence; we report the frequency of non-convergence separately.

Table 1 shows that convergence is usually reached for the new parametrization in all scenarios. As the dimensionality *M* increases, the percent non-convergence increases only slowly when using the new parametrization. In contrast, the percent non-convergence increases quickly when using the spherical parametrization. Importantly, non-converge occurs less often for the new parametrization than for the spherical parametrization in all scenarios. Furthermore, the relative time efficiency of the two parametrizations depends on the dimensionality M, with the new parametrization becoming more preferable as *M*

increases. The spherical parametrization is faster for $M = 6$, but the new parametrization is faster for $M = 12$, $M = 18$, and $M = 24$.

Table 2 contains the 10%, 5%, and 1% upper quantiles for computing time. It confirms that the new parametrization's relative performance increases as the dimensionality M increases. Furthermore, the quantiles indicate that the new parametrization becomes more preferable when moving farther into the tail of the distribution of computing times; that is, when increased complexity makes it harder to estimate the model parameters. In all scenarios with $M = 12$, $M = 18$, and $M = 24$, the quantiles are lower (and thus better) for the new parametrization than for the spherical parametrization. Furthermore, when considering $M = 6$, the spherical parameterization provides the lowest 10% and 5% quantiles, but the new parametrization provides the lowest 1% quantiles.

6. Conclusion

We presented an easy Cholesky-based parametrization of correlation matrices that allows for unconstrained estimation. The simulation results for Gaussian copulas indicated that, compared with the commonly applied spherical parametrization, the new parametrization reduces non-convergence and is better able to curb long computing times. Whereas we focused on copula estimation, we encourage future research to evaluate the performance of the new parametrization in other applications.

Declarations of interest: none

References

Bauwens, L., Bos, C. S., Van Dijk, H. K., and Van Oest, R. D. (2004). Adaptive radial-based direction sampling: Some flexible and robust Monte Carlo integration methods. *Journal of Econometrics*, 123(2), 201–225.

Danaher, P. J., and Smith, M. S. (2011). Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science*, 30(1), 4-21.

Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox*, 3rd edition. London: Timberlake Consultants Press.

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

Madar, V. (2015). Direct formulation to Cholesky decomposition of a general nonsingular correlation matrix. *Statistics and Probability Letters*, 103, 142-147.

Nelsen, R. B. (2006). *An introduction to copulas*, 2nd edition. New York: Springer.

Pinheiro, J., and Bates, D. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3), 289-296.

Pitt, M., Chan, D., and Kohn, R. (2006). Efficient Bayesian Estimation for Gaussian copula regression models. *Biometrika*, 93(3), 537-554.

Rebonato, R., and Jäckel, P. (2000). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk*, 2, 17-27.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institute de Statistique de l'Universite de Paris*, 8, 229-231.

Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, 27(2), 305-320.

Tsay, R. S., and Pourahmadi, M. (2017). Modelling structured correlation matrices. *Biometrika*, 104(1), 237-242.

Table 1. Performance comparison of new parametrization and spherical parametrization in terms of percent non-convergence and average computing time across simulation runs.

		Gamma marginals		Gamma-beta marginals	
Criterion	*M*	New	Spherical	New	Spherical
% Non-convergence	6	0.0	2.9	0.0	2.1
	12	0.1	22.1	0.2	14.6
	18	1.5	51.7	0.9	40.2
	24	5.9	80.4	3.5	67.5
Average time	6	5.77	5.49	7.18	6.68
	12	52.12	53.68	64.49	64.71
	18	190.20	202.50	229.75	242.67
	24	497.35	547.66	594.72	643.43

Note. Computing time is measured in seconds; *M* is the dimensionality of the multivariate distribution.

Table 2. Performance comparison of new parametrization and spherical parametrization in terms of upper 10%, 5%, and 1% quantiles for computing time across simulation runs.

		Gamma marginals		Gamma-beta marginals	
Criterion	*M*	New	Spherical	New	Spherical
Upper 10% quantile of time	6	7.02	6.71	8.77	8.16
	12	63.19	65.11	78.03	78.93
	18	229.74	248.46	280.38	300.18
	24	603.23	681.36	741.09	802.50
Upper 5% quantile of time	6	7.62	7.13	9.50	8.86
	12	66.68	69.26	82.84	84.05
	18	244.39	267.60	299.06	319.49
	24	633.35	732.97	802.24	867.36
Upper 1% quantile of time	6	9.24	9.29	10.91	11.20
	12	79.98	87.67	98.85	103.56
	18	273.80	307.68	329.32	357.31
	24	708.50	815.20	895.16	945.95

Note. Computing time is measured in seconds; *M* is the dimensionality of the multivariate distribution.