Positive semi-definite correlation matrix completion

Peter Jäckel[*] Christian Kahl[†]

First version: 6th May 2009
This version: 10th May 2009

Abstract

We give an intuitive derivation for the correlation matrix completion algorithm suggested in [KG06]. This leads us to a more general formula for the completion. The presented extension is positive semi-definite by construction, but we also give a simplified algebraic proof for its universal validity.

1 Introduction

Since the nature of this note is to present an extension to [KG06], we skip the general motivation and background of the problem and refer the reader to the references [KG06, Kah07].

Given a set of $2n$ standard normal variates $x_1, \cdots, x_n$, and $y_1, \cdots, y_n$, and the constraint that the pairwise correlations

$$\langle x_i x_j \rangle = r_{ij} \tag{1}$$
$$\langle x_i y_i \rangle = \eta_i \tag{2}$$

are pre-specified, we seek a completion of the as yet under-specified (auto-)correlation matrix of

$$z := (x_1, \cdots, x_n, y_1, \cdots, y_n)^\top \tag{3}$$

which has the structure

$$\langle z \cdot z^\top \rangle = \begin{pmatrix} r_{11} & \dots & r_{1n} & \eta_1 & & ? \\ \vdots & \ddots & \vdots & & \ddots & \\ r_{n1} & \dots & r_{nn} & ? & & \eta_n \\ \eta_1 & & ? & 1 & & ? \\ & \ddots & & & \ddots & \\ ? & & \eta_n & ? & & 1 \end{pmatrix} = \begin{pmatrix} R & B \\ B^\top & C \end{pmatrix} \tag{4}$$

with $r_{ii} = 1$.

2 Pairwise Cholesky construction

We start our intuition with the suggestion that each of the y_i can be represented as a linear combination of x_i and a further standard normal variate ϵ_i which is independent from all the x_j, as given by a pairwise Cholesky decomposition:

$$y_i = \eta_i x_i + \eta_i' \epsilon_i \ , \tag{5}$$
$$\langle x_i \epsilon_j \rangle = 0 \ , \tag{6}$$

with

$$\eta_i' := \sqrt{1 - \eta_i^2} \ . \tag{7}$$

This immediately yields

$$\langle x_i y_j \rangle = r_{ij} \eta_j \tag{8}$$

whence we choose $B := RH$ with

$$H := \mathrm{diag}(\eta_1, \cdots, \eta_n) \tag{9}$$

as in [KG06]. Further, we have

$$c_{ij} = \langle y_i y_j \rangle = \eta_i r_{ij} \eta_j + \eta_i' \langle \epsilon_i \epsilon_j \rangle \eta_j' \ . \tag{10}$$

We note that for $\langle \epsilon_i \epsilon_j \rangle = 0$ we obtain the structure given in [KG06].

3 Positive semi-definiteness

Since the matrices B and C given in the previous section are derived from linear combinations of standard normal variates, the completed matrix

$$A := \langle z \cdot z^\top \rangle = \begin{pmatrix} R & B \\ B^\top & C \end{pmatrix} \tag{11}$$

is *by construction* symmetric positive semi-definite, which we denote as

$$A \succeq 0 \ . \tag{12}$$

However, for the sake of completeness, we provide below a simple algebraic proof.

Given two matrices $R, E \in \mathbb{R}^{n \times n}$, with $R \succeq 0, E \succeq 0$, we set

$$A := \begin{pmatrix} R & RH \\ HR & C \end{pmatrix} \tag{13}$$

[*]OTC Analytics
[†]ABN AMRO

with

$$C := HRH + H'EH' \ , \tag{14}$$
$$H' := \mathrm{diag}(\eta'_1, \cdots, \eta'_n) \ . \tag{15}$$

Since the spectrum of A is invariant with respect to the addition of one of its (scaled) rows to any other, and likewise for columns, Gaussian elimination gives us

$$\begin{pmatrix} R & RH \\ HR & C \end{pmatrix} \succeq 0 \tag{16}$$
$$\begin{pmatrix} R & RH \\ 0 & C - HRH \end{pmatrix} \succeq 0 \tag{17}$$
$$\begin{pmatrix} R & 0 \\ 0 & C - HRH \end{pmatrix} \succeq 0 \ . \tag{18}$$

Since $R \succeq 0$, equation (18) holds if $C - HRH \succeq 0$. This, however, follows trivially since

$$C - HRH = HRH + H'EH' - HRH \tag{19}$$
$$= H'EH' \succeq 0 \ . \tag{20}$$

4 Summary

We showed how, given a correlation structure R for n standard normal variates $x_1, \cdots, x_n$, and given the correlations $\langle x_i y_i \rangle = \eta_i$ to a second set of standard normal variates $y_1, \cdots, y_n$, one can *constructively* arrive at

$$b_{ij} = \langle x_i y_j \rangle = r_{ij} \eta_i \tag{21}$$
$$c_{ij} = \langle y_i y_j \rangle = \eta_i r_{ij} \eta_j + \eta'_i e_{ij} \eta'_j \tag{22}$$

for an arbitrary correlation matrix $E \in \mathbb{R}^{n \times n}, E \succeq 0$, as a possible choice for the completed correlation matrix $A = \begin{pmatrix} R & B \\ B^\top & C \end{pmatrix}$. We also proved

$$\begin{pmatrix} R & RH \\ HR & HRH + H'EH' \end{pmatrix} \succeq 0 \tag{23}$$

by the aid of straightforward Gaussian elimination of rows and columns.

It remains to be said that in practice one may wish to use the homogenous parametric form

$$e_{ij} = \beta + (1 - \beta)\delta_{ij} \tag{24}$$

for E, with $\beta \in \left[-\frac{1}{n-1}, 1\right]$ and $\delta_{(\cdot,\cdot)}$ being the Kronecker symbol, for the sake of simplicity.

References

[Kah07] C. Kahl. *Modeling and simulation of stochastic volatility in finance.* PhD thesis, Bergische Universität Wuppertal and ABN AMRO, 2007. Published by www.dissertation.com, www.amazon.com/Modelling-Simulation-Stochastic-Volatility-Finance/dp/1581123833/, ISBN-10: 1581123833.

[KG06] C. Kahl and M. Günther. Complete the Correlation Matrix. Working paper, Bergische Universität Wuppertal, 2006. www.math.uni-wuppertal.de/~kahl/publications/CompleteTheCorrelationMatrix.pdf.