Rating of correlation and partial correlation matrices

Belfedal Chaima Djouhina (Algeria)

Department Probability and Statistics, Mathematical Institute

Eötvös Loránd University, Budapest, Hungary

A thesis submitted for the degree of

MSc in Mathematics

December 2022

Declaration

STATEMENT

Name: Chaima Djouhina Belfedal
Neptun ID: G12XW0

ELTE Faculty of Science: MSc

specialization: Pure Mathematics

Title of diploma work: Estimation and hypothesis testing of covariance matrices

As the author of the diploma work I declare, with disciplinary responsibility that my thesis is my own intellectual product and the result of my own work. Furthermore I declare that I have consistently applied the standard rules of references and citations.

I acknowledge that the following cases are considered plagiarism:
– using a literal quotation without quotation mark and adding citation;
– referencing content without citing the source;
– representing another person's published thoughts as my own thoughts.

Furthermore, I declare that the printed and electronical versions of the submitted diploma work are textually and contextually identical.

Budapest, 2022.11.18

Signature of Student

Acknowledgements

First and foremost, I would like to thank my supervision Mr.Prőhle Tamás for introducing me to this work, for his help through the difficulties that I faced, his patience with me, his helpful suggestions and his assistance in every step through the process. I am grateful for the many discussions that we had.

I am also grateful to every teacher who taught me all the years of my education, I would also like to thank Mr.István Ágoston for his help through the difficulties in my master.

Lastly, I would be remiss in not mentioning my family, especially my parents, and my siblings "Ikram, Alae, and Hatem". Their belief in me has kept my spirits and motivation high during this process. I would also like to thank my Friends for all the entertainment and emotional support, especially Tarek.

Abstract

In the factor analysis the well-known KMO statistics is often used. In general, the KMO is a one-dimensional representation of a multidimensional random variable's dependence structure. This depends on both the correlation and the partial correlation of the coordinates. In statistics, we represent a correlation between two random variables as a linear relationship, and also we define a partial correlation as a correlation of the dual bases of the observations.

The main result of our dissertation are the Claims 0. 1. 2. 3. and 4. in Chapter 5. These Claims are unknown in the literature. These show the exact range and the distribution of possible values for KMO statistics and a critique of Kaiser's rating.

These results have of great practical importance because they show that, in contrast to general application practice, the range of possible KMO values is only a narrower part of the $[0, 1]$ interval, and its distribution is neither even nor symmetrical.

The paper begin by introducing the correlation and the partial correlation, and their matrices. Then we deal which some theorems and properties, that help us to understand generation methods of random correlation matrices. Finally, the dissertation deals in detail with the properties of the tools needed to interpret the results concerning the KMO that is the Kaiser–Meyer–Olkin test.

The results are conjectures that are supported by simulations and calculations. Therefore, the thesis contains scripts in several places.

Contents

1

Correlation and partial correlation

1.1 The correlation

Definition 1 (The correlation).
Let ξ, η be two real random value where each one has finite variance, we denote the correlation of ξ, η by $\operatorname{cor}(\xi, \eta)$ *or $\varrho(\xi, \eta)$, is defined by:*

$$\varrho(\xi, \eta) = \frac{\mathbb{E}\big((\xi - \mathbb{E}(\xi))(\eta - \mathbb{E}(\eta))\big)}{\mathbb{D}(\xi)\mathbb{D}(\eta)} \tag{1.1}$$

with standard deviations $\mathbb{D}(\xi)$ and $\mathbb{D}(\eta)$ and expected values $\mathbb{E}(\xi)$ and $\mathbb{E}(\eta)$.

The properties of the correlation:

(1) The value of the correlation coefficient is between -1 and +1 .

(2) If ξ and η are independent then $\varrho = 0$.

(3) If $\varrho = 0$ then the ξ and η are not necessarily independent (see Example 1.1).

(4) The correlation invariant with respect to linear transformations

$$\operatorname{cor}(a\,\xi + b, c\,\eta + d) = sign(ac) \cdot \operatorname{cor}(\xi, \eta)$$

where a, b, c, d are arbitrary constants.

(5) If for all g and h functions $\operatorname{cor}(g(\xi), h(\eta)) = 0$ then the ξ and η are independent.

(6) If ξ and η are standardized random values, then $arg\min_c \mathbb{E}((\eta - c\xi)^2) = \varrho$.

(7) If $\widehat{\eta}_{|\xi}$ denotes $\varrho\xi$, the best linear approximation of η using ξ, then $\varrho^2 = \mathbb{D}^2(\widehat{\eta}_{|\xi})$ is the explained proportion of the variance of η.

(8) The positive correlation isn't always transitive. (see Example 1.2)

Example 1.1. *Let be ξ have a symmetric distribution, and let $\eta = \xi^2$, then* $\operatorname{cor}(\xi, \eta) = 0$ *but ξ and η are not independent. To demonstrate the zero correlation, enough to prove that* $\operatorname{cov}(\xi, \eta) = \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta) = 0$. *But here both terms are really zero. The second because of symmetry then $\mathbb{E}(\xi) = 0$ and the first because $\xi\eta = \xi^3$ has also a symmetric distribution.*
To show non-independence, it is sufficient to show only one pair of non-independent events defined by the two random variables. For the sake of simplicity, such a pair of events is given only in a very special case. Let $\mathcal{U}([-1, 1])$, then

$$\mathbb{P}(\xi < .5, \eta < .25) = \mathbb{P}(\eta < .25) = 5/8 \neq \mathbb{P}(\xi < .5) \cdot \mathbb{P}(\eta < .25) = .75 \cdot 5/8$$

Example 1.2. *Let A, B and C three independent variable, for which the expected values are 0 and the standard deviations are $\sqrt{2}$. Determine the values of X, Y and Z using the following equations:*

$$\begin{aligned} X &= A + B \\ Y &= C + B \\ Z &= C - A \end{aligned}$$

Then $\operatorname{cor}(X, Y) = .5 > 0$, $\operatorname{cor}(Y, Z) = .5 > 0$ *but* $\operatorname{cor}(X, Z) = -.5 < 0$.

1.2 The partial correlation

Definition 2 (the partial correlation).
Let ξ, η and $\zeta = (\zeta_1, ..., \zeta_k)$ three random variables. By definition, the partial correlation of ξ and η according of ζ is equal to the correlation of the errors of the linear predictions of ξ and η using ζ:

$$\varrho_{\xi,\eta\cdot\zeta_1,...,\zeta_k} = \varrho\big(\mathcal{E}_{\xi\sim(\zeta_1,...,\zeta_k)}, \mathcal{E}_{\eta\sim(\zeta_1,...,\zeta_k)}\big) = \varrho\big(\xi - \ell_\xi(\zeta_1, ..., \zeta_k), \eta - \ell_\eta(\zeta_1, ..., \zeta_k)\big) \quad (1.2)$$

where $\ell_\xi(\zeta_1, ..., \zeta_k)$ is the linear regression of ξ on $(\zeta_1, ..., \zeta_k)$ and
$\ell_\eta(\zeta_1, ..., \zeta_k)$ is the linear regression of η on $(\zeta_1, ..., \zeta_k)$ and
$\mathcal{E}_{\xi\sim(\zeta_1,...,\zeta_k)}$ and $\mathcal{E}_{\eta\sim(\zeta_1,...,\zeta_k)}$ are the errors of this two regressions respectively.

The properties of the partial correlation:

- The case of one dimensional predictor variable:

Lemma 1 (**linear approximation in dispersion).**
Let η and ξ two arbitrary random value.
The best linear approximation of η with X, measured in dispersion is:

$$arg\min_{a,b}\mathbb{D}\big(\eta-\ell_\eta(\xi)\big)=arg\min_{a,b}\mathbb{D}\big(\eta-(a+b\xi)\big)=(a_\eta,b_\eta) \tag{1.3}$$

where a_η = any number, and $b_\eta=\frac{\mathrm{cov}(\eta,\xi)}{\mathbb{D}^2(\xi)}$

Proof of Lemma 1.

$$\begin{aligned}\mathbb{D}^2(\eta-(a+b\xi)) &= \mathbb{D}^2\Big(\eta-(a+b\xi)-\big(\mathbb{E}(\eta)-(a+b\mathbb{E}(\xi))\big)\Big)\\ &= \mathbb{D}^2(\eta_0-b\xi_0)\\ &= \mathbb{E}\big((\eta_0-b\xi_0)^2\big)-\big(\mathbb{E}(\eta_0-b\xi_0)\big)^2 \text{ second term is zero}\\ &= \mathbb{D}^2(\eta)-2b\mathrm{cov}(\eta,\xi)+b^2\mathbb{D}^2(\xi)\end{aligned}$$

Where we used the notation $\eta_0=\eta-\mathbb{E}(\eta)$ and $\xi_0=\xi-\mathbb{E}(\xi)$, and the equivalence $\mathbb{E}(\eta_0-b\mathbb{E}(\xi_0))=0-b\cdot 0=0$.
The minimum place of this second-order expression is according to b is

$$b=\frac{-(-2\mathrm{cov}(\eta,\xi))}{2\mathbb{D}^2(\xi)}=\frac{\mathrm{cov}(\eta,\xi)}{\mathbb{D}^2(\xi)}$$

□

Lemma 2 (**linear approximation in mean squared error).**
Let η and ξ two arbitrary random value.
The best linear approximation of η with ξ, measured in mean squared error is:

$$arg\min_{a,b}\mathbb{E}\Big(\big(\eta-\ell_\eta(\xi)\big)^2\Big)=arg\min_{a,b}\mathbb{E}\Big(\big(\eta-(a+b\xi)\big)^2\Big)=(a_\eta,b_\eta) \tag{1.4}$$

where $a_\eta=\mathbb{E}(\eta)-b_\eta\mathbb{E}(\xi)$, and $b_\eta=\frac{\mathrm{cov}(\eta,\xi)}{\mathbb{D}^2(\xi)}$

Proof of Lemma 2.

$$\frac{\partial}{\partial a}\left(\mathbb{E}\Big(\big(\eta-(a+b\xi)\big)^2\Big)\right)=-2\mathbb{E}\big(\eta-(a+b\xi)\big)=0$$

With equivalent conversion: $\mathbb{E}(\eta)-a-b\mathbb{E}(\xi)=0$

That is why $a=\mathbb{E}(\eta)-b\mathbb{E}(\xi)$.

$$\begin{aligned}
\frac{\partial}{\partial b}\left(\mathbb{E}\Big(\big(\eta-(a+b\xi)\big)^2\Big)\right)=-2\mathbb{E}\Big(\xi\big(\eta-(a+b\xi)\big)\Big) &= 0\\
\mathbb{E}(\eta\xi)-a\mathbb{E}(\xi)-b\mathbb{E}(\xi^2) &= 0\\
\mathbb{E}(\eta\xi)-(\mathbb{E}(\eta)-b\mathbb{E}(\xi))\mathbb{E}(\xi)-b\mathbb{E}(\xi^2) &= 0\\
\mathbb{E}(\eta\xi)-\mathbb{E}(\eta)\mathbb{E}(\xi)-b\big(\mathbb{E}(\xi^2)-(\mathbb{E}(\xi))^2\big) &= 0\\
\operatorname{cov}(\eta,\xi)-b\mathbb{D}^2(\xi) &= 0
\end{aligned}$$

In the second line we used that $a=\mathbb{E}(\eta)-b\mathbb{E}(\xi)$, and
based on the last line it follows that indeed $b=\frac{\operatorname{cov}(\eta,\xi)}{\mathbb{D}^2(\xi)}$ is the solution. □

Let $\ell_\eta(\xi)=a+b\xi$ where $a=\mathbb{E}(\eta)-b\mathbb{E}(\xi)$, $b=\frac{\operatorname{cov}(\xi,Y)}{\mathbb{D}^2(\xi)}$

$$\mathbb{E}\big(\ell_\eta(\xi)\big)=\mathbb{E}(a+b\xi)=a+b\mathbb{E}(\xi)=(\mathbb{E}(\eta)-b\mathbb{E}(\xi))+b\mathbb{E}(\eta)=\mathbb{E}(\eta)$$

$$\mathbb{D}^2\big(\ell_\eta(\xi)\big)=\mathbb{D}^2(a+b\xi)=b^2\mathbb{D}^2(\xi)=\frac{\operatorname{cov}^2(\eta,\xi)}{\mathbb{D}^4(\xi)}\mathbb{D}^2(\xi)=\frac{\operatorname{cov}^2(\eta,\xi)}{\mathbb{D}^2(\xi)}$$

Let $\mathcal{E}_{\eta|\xi}=\eta-\ell_\eta(\xi)$ then

$$\mathbb{E}\big(\mathcal{E}_{\eta|\xi}\big)=\mathbb{E}(\eta-\ell_\eta(\xi))=\mathbb{E}(\eta)-\mathbb{E}(\ell_\eta(\xi))=\mathbb{E}(\eta)-\mathbb{E}(\eta)=0$$

$$\begin{aligned}
\mathbb{D}^2\big(\mathcal{E}_{\eta|X}\big) &= \mathbb{E}\big((\eta-\ell_\eta(\xi))^2\big)-\big(\mathbb{E}(\eta-\ell_\eta(\xi))\big)^2=\text{ second term is zero}\\
&= \mathbb{E}\big((\eta-\mathbb{E}(\eta)+\mathbb{E}(\eta)-(a+b\xi))^2\big)\\
&= \mathbb{E}\big((\eta-\mathbb{E}(\eta)+(a+b\mathbb{E}(\xi))-(a+b\xi))^2\big)\\
&= \mathbb{E}\Big(\big(\ (\eta-\mathbb{E}(\eta))-b(\xi-\mathbb{E}(\xi))\ \big)^2\Big)\\
&= \mathbb{D}^2(\eta)-2b\operatorname{cov}(Y,\xi)+b^2\mathbb{D}^2(\xi)\\
&= \mathbb{D}^2(\eta)-2\frac{\operatorname{cov}(\eta,\xi)}{\mathbb{D}^2(\xi)}\operatorname{cov}(\eta,\xi)+\frac{\operatorname{cov}^2(\eta,\xi)}{\mathbb{D}^4(\xi)}\mathbb{D}^2(\xi)\\
&= \mathbb{D}^2(\eta)-\frac{\operatorname{cov}^2(\eta,\xi)}{\mathbb{D}^2(\xi)}\\
&= \mathbb{D}^2(\eta)\big(1-\varrho^2(\eta,\xi)\big)
\end{aligned}$$

Where we used in the second line that $+\mathbb{E}(\eta) = a + b\mathbb{E}(\xi)$, and in the last the notation $\varrho = \frac{\mathrm{cov}(\xi,\eta)}{\mathbb{D}(\eta)\mathbb{D}(\xi)}$, the correltion of η and ξ.

Theorem 1 (**value of the partial correlation based on the correlations).** *Let ξ, η and ζ three random varables. Then we can calculate the partial correlation of ξ and η according ζ based on the pairwise correlations of the three variables by the following formula:*

$$\varrho_{\xi,\eta\bullet\zeta} = \frac{\varrho_{\xi,\eta} - \varrho_{\xi,\zeta}\varrho_{\eta,\zeta}}{\sqrt{1-\varrho^2_{\xi,\zeta}}\sqrt{1-\varrho^2_{\eta,\zeta}}} \tag{1.5}$$

Proof of Theorem 1.

$$\begin{aligned}
\mathrm{cov}(\mathcal{E}_{\xi|\zeta}, \mathcal{E}_{\eta|\zeta}) &= \mathbb{E}(\mathcal{E}_{\xi|\zeta}\mathcal{E}_{\eta|\zeta}) - \mathbb{E}(\mathcal{E}_{\xi|\zeta})\mathbb{E}(\mathcal{E}_{\eta|\zeta}) \\
&= \mathbb{E}\big(\xi - (a_\xi + b_\xi\zeta)\big)\big(\eta - (a_\eta + b_\eta\zeta)\big) \\
&= \mathbb{E}\Big(\big((\xi - \mathbb{E}(\xi)) - b_\xi(\zeta - \mathbb{E}(\zeta))\big)\big((\eta - \mathbb{E}(\eta)) - b_\eta(\zeta - \mathbb{E}(\zeta))\big)\Big) \\
&= \mathrm{cov}(\xi,\eta) - b_\eta\mathrm{cov}(\xi,\zeta) - b_\xi\mathrm{cov}(\eta,\zeta) + b_\xi b_\eta\mathbb{D}^2(\zeta) \\
&= \mathrm{cov}(\xi,\eta) - \frac{\mathrm{cov}(\eta,\zeta)}{\mathbb{D}^2(\zeta)}\mathrm{cov}(\xi,\zeta) - \\
&\qquad \frac{\mathrm{cov}(\xi,\zeta)}{\mathbb{D}^2(\zeta)}\mathrm{cov}(\eta,\zeta) + \frac{\mathrm{cov}(\eta,\zeta)}{\mathbb{D}^2(\zeta)}\frac{\mathrm{cov}(\xi,\zeta)}{\mathbb{D}^2(\zeta)}\mathbb{D}^2(\zeta) \\
&= \mathrm{cov}(\xi,\eta) - \frac{\mathrm{cov}(\eta,\zeta)}{\mathbb{D}^2(\zeta)}\mathrm{cov}(\xi,\zeta)
\end{aligned}$$

We used in the first row that $\mathbb{E}(\mathcal{E}_{\xi|\zeta}) = \mathbb{E}(\mathcal{E}_{\eta|\zeta}) = 0$. In the third row we wrote in place of a_ξ and a_η the $\mathbb{E}(\xi) - b_\xi\mathbb{E}(\zeta)$ and $\mathbb{E}(\eta) - b_\eta\mathbb{E}(\zeta)$.

$$\begin{aligned}
\varrho(\mathcal{E}_{\xi|\zeta}, \mathcal{E}_{\eta|\zeta}) &= \frac{\mathrm{cov}(\xi,\eta) - \frac{\mathrm{cov}(\eta,\zeta)}{\mathbb{D}^2(\zeta)}\mathrm{cov}(\xi,\zeta)}{\mathbb{D}(\xi)\sqrt{1-\varrho^2_{\xi,\zeta}}\;\mathbb{D}(\eta)\sqrt{1-\varrho^2_{\eta,\zeta}}} \\
&= \frac{\varrho_{\xi,\eta} - \varrho_{\xi,\zeta}\varrho_{\eta,\zeta}}{\sqrt{1-\varrho^2_{\xi,\zeta}}\sqrt{1-\varrho^2_{\eta,\zeta}}}
\end{aligned}$$

Which was to be proved.

□

- The case of multidimensional predictor variable:

Theorem 2 (**recursive formula of the partial correlation**).

$$\varrho_{\xi,\eta\bullet(\zeta_1,\ldots,\zeta_k,\zeta_{k+1})} = \frac{\varrho_{\xi,\eta\bullet(\zeta_1,\ldots,\zeta_k)} - \varrho_{\xi,\zeta_{k+1}\bullet(\zeta_1,\ldots,\zeta_k)}\varrho_{\eta,\zeta_{k+1}\bullet(\zeta_1,\ldots,\zeta_k)}}{\sqrt{1-\varrho^2_{\xi,\zeta_{k+1}\bullet(\zeta_1,\ldots,\zeta_k)}}\sqrt{1-\varrho^2_{\eta,\zeta_{k+1}\bullet(\zeta_1,\ldots,\zeta_k)}}} \tag{1.6}$$

Theorem 3 (**multidimensional predictor variable**).
Let $\xi = (\xi_1, \ldots, \xi_p)$ a p dimensional random value.
The value of the $\varrho_{\xi_1,\xi_p\bullet(\xi_2,\ldots,\xi_{p-1})}$ partial correlation gives the following equation:

$$\varrho_{\xi_1,\xi_p\bullet(\xi_2,\ldots,\xi_{p-1})} = \frac{r - \mathbf{r}_{1,2}\mathbf{R}^{-1}_{2,2}\mathbf{r}_{2,3}}{\sqrt{1-\mathbf{r}_{1,2}\mathbf{R}^{-1}_{2,2}\mathbf{r}_{2,1}}\sqrt{1-\mathbf{r}_{3,2}\mathbf{R}^{-1}_{2,2}\mathbf{r}_{2,3}}} \tag{1.7}$$

In the formula above the 'r'-s are submatrices of the correlation matrix of η.
Where $\eta = (\eta_1, \eta_2, \eta_3)$, here $\eta_1 = \xi_1$, $\eta_2 = (\xi_2, \ldots, \xi_{p-1})$, $\eta_3 = \xi_p$, and

$$\operatorname{cor}(\eta) = \operatorname{cor}\big((\eta_1, \eta_2, \eta_3)\big) = \begin{pmatrix} 1 & \mathbf{r}_{1,2} & r \\ \mathbf{r}_{2,1} & \mathbf{R}_{2,2} & \mathbf{r}_{2,3} \\ r & \mathbf{r}_{3,2} & 1 \end{pmatrix}$$

Definition 3 (**alternative definition of the partial correlation**).
The partial correlation of two elements of a variable set, according to the other elements of the set is the same as 'the minus correlation between the two variables own information'. For example, let $\xi_1, \ldots, \xi_k$ be a variable set, then the partial correlation of ξ_1 and ξ_k according to $\xi_2, \ldots, \xi_{k-1}$ is:

$$\varrho_{\xi_1,\xi_k\bullet\xi_2,\ldots,\xi_{k-1}} =$$
$$-\varrho\big(\mathcal{E}_{\xi_1\sim(\xi_2,\ldots,\xi_k)}, \mathcal{E}_{\xi_k\sim(\xi_1,\ldots,\xi_{k-1})}\big) = -\varrho\big(\xi_1 - \ell_{\xi_1}(\xi_2,\ldots,\xi_k), \xi_k - \ell_{\xi_k}(\xi_1,\ldots,\xi_{k-1})\big)$$

where $\ell_{\xi_1}(\xi_2, \ldots, \xi_k)$ is the regression of ξ_1 on $(\xi_2, \ldots, \xi_k)$ and
$\ell_{\xi_k}(\xi_1, \ldots, \xi_{k-1})$ is the regression of ξ_k on $(\xi_1, \ldots, \xi_{k-1})$ and
$\mathcal{E}_{\xi_1\sim(\xi_2,\ldots,\xi_k)}$ and $\mathcal{E}_{\xi_k\sim(\xi_1,\ldots,\xi_{k-1})}$ are the errors of this two regressions respectively.

Lemma 3 (**the equivalence of the two definitions of partial correlation**).
Considering the correlation pair also as an explanatory variable only changes the sign of the partial correlation. Which means that

$$\varrho\big(\mathcal{E}_{\xi_1\sim(\xi_2,\ldots,\xi_{k-1})}, \mathcal{E}_{\xi_k\sim(\xi_2,\ldots,\xi_{k-1})}\big) = -\varrho\big(\mathcal{E}_{\xi_1\sim(\xi_2,\ldots,\xi_k)}, \mathcal{E}_{\xi_k\sim(\xi_1,\ldots,\xi_{k-1})}\big) \tag{1.8}$$

A numerical example in **R** .

```
set.seed(54321)
n <- 123
p <- 5

X <- matrix(rnorm(n*p),n,p)

# the standard definition
# explanation without X1 and Xp
e_1k <- residuals(lm(X[,1]~X[,-c(1,p)]))
e_pk <- residuals(lm(X[,p]~X[,-c(1,p)]))
cor(e_1k,e_pk) # 0.02449181

# the alternative definition
# the correlation when we use the counterpart variable also
e_1m <- residuals(lm(X[,1]~X[,-1]))
e_pm <- residuals(lm(X[,p]~X[,-p]))
cor(e_1m,e_pm) # -0.02449181

# they differ only in sign
cor(e_1k,e_pk)+cor(e_1m,e_pm) # 0
```

Proof of Lemma 3:
We transform the notation so that the notation of the proof be simpler.
Let denote the target variables by $\eta_1 \equiv \xi_1$ and $\eta_2 \equiv \xi_k$ and the circumstance variables by $\xi \equiv (\xi_2, ..., \xi_{k-1})$. And denote the investigated errors by $\varepsilon_1 \equiv \mathcal{E}_{\eta_1 \sim \xi}$ and $\varepsilon_2 \equiv \mathcal{E}_{\eta_2 \sim \xi}$. And the errors of the alternative definition let denoted by $\delta_1 \equiv \mathcal{E}_{\eta_1 \sim (\eta_2, \xi)}$ and $\delta_2 \equiv \mathcal{E}_{\eta_2 \sim (\eta_1, \xi)}$. So the question of Lemma 3 is, whether the following equality:

$$\varrho(\varepsilon_1, \varepsilon_2) = -\varrho(\delta_1, \delta_2)$$

is true or not?

Denote by $\mathcal{P}_H$ the orthogonal projection on the space spanned by the variable set H.

$$\begin{aligned}
\delta_1 &= \eta_1 - \mathcal{P}_{\eta_2,\xi}(\eta_1) \\
&= \eta_1 - \mathcal{P}_{\eta_2,\xi}(\mathcal{P}_\xi(\eta_1) + \varepsilon_1) \\
&= \eta_1 - \big(\mathcal{P}_{\eta_2,\xi}(\mathcal{P}_\xi(\eta_1)) + \mathcal{P}_{\eta_2,\xi}(\varepsilon_1)\big) \\
&= \eta_1 - \big(\mathcal{P}_\xi(\eta_1) + \mathcal{P}_{\eta_2,\xi}(\varepsilon_1)\big) \qquad \text{because } \mathcal{P}_\xi(\eta_1) \in \{\eta_2, \xi\} \\
&= \eta_1 - \big(\mathcal{P}_\xi(\eta_1) + \mathcal{P}_\xi(\varepsilon_1) + \mathcal{P}_{\varepsilon_2}(\varepsilon_1)\big) \quad \text{because } \xi \perp \varepsilon_2 \text{ thus } \mathcal{P}_{\eta_2,\xi} = \mathcal{P}_\xi + \mathcal{P}_{\varepsilon_2} \\
&= \eta_1 - \mathcal{P}_\xi(\eta_1) - \mathcal{P}_{\varepsilon_2}(\varepsilon_1) \qquad \text{because } \varepsilon_1 \perp \xi \\
&= \varepsilon_1 - \mathcal{P}_{\varepsilon_2}(\varepsilon_1) \qquad \text{because } \eta_1 = \mathcal{P}_\xi(\eta_1) + \varepsilon_1 \\
&= \mathcal{P}_{\varepsilon_2^\perp}(\varepsilon_1)
\end{aligned}$$

It can be proved in the same way that $\delta_2 = \mathcal{P}_{\varepsilon_1^\perp}(\varepsilon_2)$. So δ_1 is the projection of ε_1 onto the subspace perpendicular to ε_2, and δ_2 is the projection of ε_2 onto the subspace perpendicular to ε_1. Considering that the correlation is equal to the cosine of the angle of the vectors, the figure below interprets the truth of the statement. [1, 2]

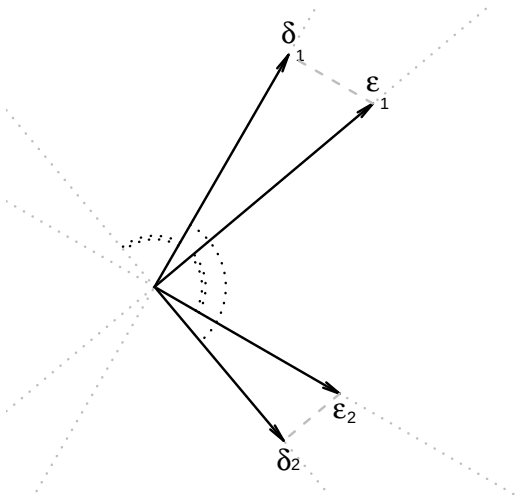

Figure 1.1: The $\angle(\varepsilon_2, \varepsilon_1) = \pi - \angle(\delta_2, \delta_1)$.

Remark 1.

The partial correlation is not necessarily less than the correlation. (see Example 1.3)

Example 1.3. *We take two cases used the Figure 1.2.*

A case when the partial correlation is less then the correlation:

```
rxy <- ryx <- .5
rxz <- rzx <- .5
ryz <- rzy <- .5
pxy.z <- (rxy-rxz*ryz)/(sqrt(1-rxz^2)*sqrt(1-ryz^2))
pxy.z #  1/3  .33333
```

If the $\mathrm{cor}(\xi,\eta) = .5$ *and* $\mathrm{cor}(\xi,\zeta) = \mathrm{cor}(\eta,\zeta) = .5$ *then the* $\mathrm{cor}(\xi,\eta\cdot\zeta) = 1/3 < \mathrm{cor}(\xi,\eta)$.

A case when the partial correlation is greather then the correlation:

```
rxy <- ryx <- .5
rxz <- rzx <- - 1/3
ryz <- rzy <- 1/3
pxy.z <- (rxy-rxz*ryz)/(sqrt(1-rxz^2)*sqrt(1-ryz^2))
pxy.z #  11/16 = .6875
```

If the $\mathrm{cor}(\xi,\eta) = .5$ *and* $-\mathrm{cor}(\xi,\zeta) = \mathrm{cor}(\eta,\zeta) = \frac{1}{3}$ *then the* $\mathrm{cor}(\xi,\eta\cdot\zeta) = \frac{11}{16} > \mathrm{cor}(\xi,\eta)$.

The figure created by **R** we illustrates, that the situation mentioned in Remark 1 isn't a rare case.

```
lg.plot <- function(rXY,res=99){
  title <- paste("Order of magnitude of correlation and",
                 "partial correlation to each other",
                 "\nCorr(X,Y)=.5,",
                 "partial: smaller='green', larger='blue'",
                 "\nimpossible correlation pairs = 'white'")
  rXZ <- as.vector(rep(-res:res/(res+1),2*res+1))
  n   <- length(rXZ)
  rYZ <- sort(rXZ)
  rYX <-rXY;rZX<-rXZ;rZY<-rYZ
  rXY.Z <- (rXY-rXZ*rYZ)/(sqrt(1-rXZ^2)*sqrt(1-rYZ^2))
  col <- rep("black",n)
  col[abs(rXY.Z) < abs(rXY)] <- "green"
```

```
  col[abs(rXY.Z) > abs(rXY)] <- "blue"
  non.pos.def <- rep(FALSE,n)
  for(k in 1:n)
   { R <- matrix(c(1,rYX,rZX[k],rXY,1,rZY[k],rXZ[k],rYZ[k],1),3,3)
     if(det(R) <= 0) non.pos.def[k] <- TRUE }
  col[non.pos.def] <- "white"
  par(mar=c(5,5,4,.5))
  plot(rXZ,rYZ,pch=20,col=col,las=1,asp=1,
    main=title,xlab="Corr(X,Z)",ylab="Corr(Y,Z)")
  abline(h=c(-1,1),v=c(-1,1),col="gray",lty=3)
  return(paste("the conditions 'pos.def' and '|parc.cor|<1'",
               "are equivalent:",
               all.equal(abs(rXY.Z)>1,non.pos.def)))
}
lg.plot(.5)
# "the conditions 'pos.def' and '|parc.cor|<1' are equivalent: TRUE"
```

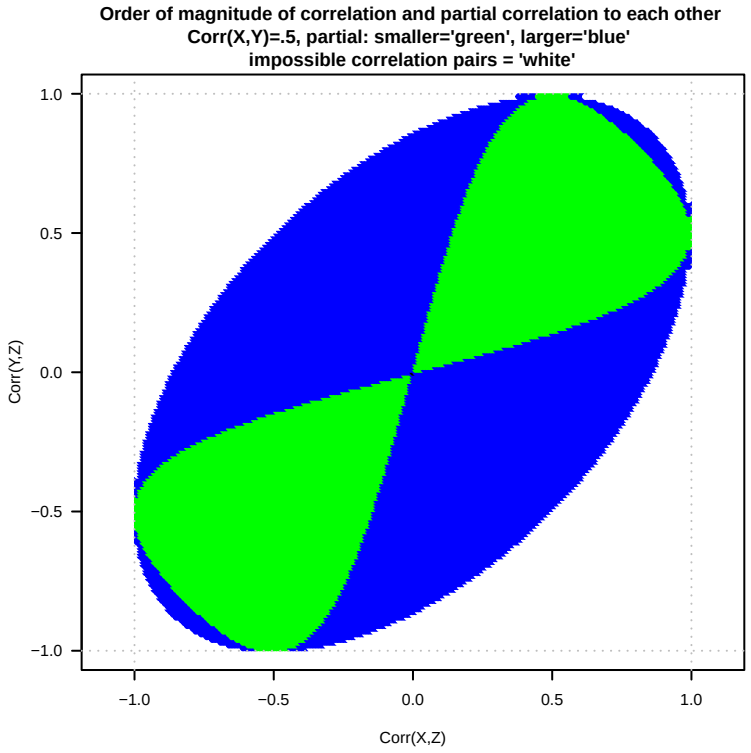

Figure 1.2: The relation of the correlation and the partial correlation

The two coordinates are $\text{cor}(X, Z)$ and $\text{cor}(Z, Y)$, and $\text{cor}(X, Y)$ is fixed at .5.

Colors indicate whether the $\mathrm{cor}(X, Y \bullet Z)$ partial correlation is greater or less as the $\mathrm{cor}(X, Y) = .5$ correlation.

The program also demostrate the validity of the Lemma 4 "feasibility of correlations".

Lemma 4 (**feasibility of correlations).**
The triad of three correlations $(\varrho_{\xi,\eta}, \varrho_{\xi,\zeta}, \varrho_{\eta,\zeta})$ *is a valid correlations triad of three random values* (ξ, η, ζ) *if and only if the partial correlation* $\varrho_{\xi,\eta\bullet\zeta}$ *calculated by the above formula 1.5 (see Theorem 1) has a value in* $[-1, 1]$.

Proof of Lemma 4.
A matrix with diagonal equals 1 can be a correlation matrix if and only if it is positive semi-definite. According to the classical Sylvester condition the matrix M

$$M = \begin{pmatrix} 1 & \varrho_{\xi,\eta} & \varrho_{\xi,\zeta} \\ \varrho_{\eta,\xi} & 1 & \varrho_{\eta,\zeta} \\ \varrho_{\zeta,\xi} & \varrho_{\zeta,\eta} & 1 \end{pmatrix}$$

is positive semi-definite than and only then, if the determinant of all of its principal minors are non-negative. The determinant of the symmetric sub-matrices of size 1×1 and 2×2 are necessarily non-negative. The only problem is the non-negativeness of the determinant of the whole matrix.
By definition the determinant of M is:

$$\begin{aligned} \det(M) &= 1 \cdot 1 \cdot 1 + \varrho_{\xi,\eta} \cdot \varrho_{\eta,\zeta} \cdot \varrho_{\zeta,\xi} + \varrho_{\xi,\zeta} \cdot \varrho_{\eta,\xi} \cdot \varrho_{\zeta,\eta} \\ &\quad - \varrho_{\xi,\zeta} \cdot 1 \cdot \varrho_{\zeta,\xi} - 1 \cdot \varrho_{\eta,\zeta} \cdot \varrho_{\zeta,\eta} - \varrho_{\xi,\eta} \cdot \varrho_{\eta,\xi} \cdot 1 \\ &= 1 + 2 \cdot \varrho_{\xi,\eta} \cdot \varrho_{\eta,\zeta} \cdot \varrho_{\zeta,\xi} - \varrho^2_{\xi,\zeta} - \varrho^2_{\eta,\zeta} - \varrho^2_{\xi,\eta} \end{aligned}$$

Using the condition that the result of the formula of the partial correlation 1.5 (see Theorem 1) gives a value in $[-1, 1]$, by equivalent transformations gives:

$$
\begin{aligned}
\varrho_{\xi,\eta\bullet\zeta} = \frac{\varrho_{\xi,\eta} - \varrho_{\xi,\zeta}\varrho_{\eta,\zeta}}{\sqrt{1-\varrho_{\xi,\zeta}^2}\sqrt{1-\varrho_{\eta,\zeta}^2}} &\in [-1,1] \\
\frac{(\varrho_{\xi,\eta} - \varrho_{\xi,\zeta}\varrho_{\eta,\zeta})^2}{(1-\varrho_{\xi,\zeta}^2)(1-\varrho_{\eta,\zeta}^2)} &\leq 1 \\
(\varrho_{\xi,\eta} - \varrho_{\xi,\zeta}\varrho_{\eta,\zeta})^2 &\leq (1-\varrho_{\xi,\zeta}^2)(1-\varrho_{\eta,\zeta}^2) \\
\varrho_{\xi,\eta}^2 + \varrho_{\xi,\zeta}^2\varrho_{\eta,\zeta}^2 - 2\varrho_{\xi,\eta}\varrho_{\xi,\zeta}\varrho_{\eta,\zeta} &\leq 1 - \varrho_{\xi,\zeta}^2 - \varrho_{\eta,\zeta}^2 + \varrho_{\xi,\zeta}^2\varrho_{\eta,\zeta}^2 \\
\varrho_{\xi,\eta}^2 - 2\varrho_{\xi,\eta}\varrho_{\xi,\zeta}\varrho_{\eta,\zeta} &\leq 1 - \varrho_{\xi,\zeta}^2 - \varrho_{\eta,\zeta}^2 \\
0 &\leq 1 + 2\varrho_{\xi,\eta}\varrho_{\xi,\zeta}\varrho_{\eta,\zeta} - \varrho_{\xi,\zeta}^2 - \varrho_{\eta,\zeta}^2 - \varrho_{\xi,\eta}^2
\end{aligned}
$$

And that's exactly what we're looking for. □

2

Correlation and partial correlation matrices

2.1 Correlation matrices

Definition 4 (covariance matrix).

Let ξ a p-dim, η a q-dim random value, then the $p \times q$ size C matrix, with elements

$$C_{i,j} = cov(\xi_i, \eta_j) = \mathbb{E}\big((\xi_i - \mathbb{E}(\xi_i))(\eta_j - \mathbb{E}(\eta_j))\big) \tag{2.1}$$

is the covariance matrix of ξ and η

Proposition 1.

If C is a covariance matrix of a random vector ξ (the covariance of ξ with ξ, i.e. the covariance of ξ with itself), then for any constant vector $\vec{\mu}$ we have

$$\vec{\mu}^T C \vec{\mu} \geq 0 \tag{2.2}$$

That is, C satisfies the property of being a positive semi-definite matrix.

Definition 5 (**correlation matrix).**

Let ξ a p-dim, η a q-dim random value, then the $p \times q$ size R matrix, with elements

$$R_{i,j} = C_{i,j}/(\mathbb{D}(\xi_i)\mathbb{D}(\eta_j)) \tag{2.3}$$

is the correlation matrix of ξ and η

In the following, we are forced to give three not significantly different definitions of the partial correlation matrix, because unfortunately there is no uniform definition of the partial correlation matrices used in practice.

2.2 Partial correlation matrices

Definition 6 (**Partial Correlation Matrix** (the negative definit variant))**.**

The (i, j)-th off-diagonal element of the partial correlation matrix of a p dimensional ξ random value equals by the partial correlation of the i-th and j-th coordinates given all of the other coordinates:

$$\varrho_{\xi_i,\xi_j \bullet \zeta_{k\in\mathfrak{I}}} \quad \textit{for all } i=1,...,p \textit{ and } j = 1, ..., p \quad \textit{where } \mathfrak{I} = \{1, ..., p\}\backslash\{i, j\}$$

The diagonal of the partial correlation matrix is equal to -1.

Let denote this version of the partial correlation matrix by P^- to distinguish it from other versions of the definitions.

Remark 2.

There are two important alternative definitions of the partial correlation matrix:

- *marked as P^*, formed by changing the sign of the diagonal elements of P^-*
- *marked as P^+, formed by changing the sign of each element of the matrix P^-*

That is, the alternative partial correlation matrices are defined according to the following equations:

$$\begin{aligned} P^* &= P^- + 2 \cdot I \\ P^+ &= -P^- \end{aligned}$$

Comment

The P^* is the version for which all elements, including the diagonal, are equal to the partial correlations of the corresponding variables, given all other variables as circumstances.

The `corpcor` [3] and `randcorr` [4] **R** supplementary packages – which was used several times in our simlation study – works with the P^* type partial correlation matrices.

Notation mode

For a square matrix M the $Diag^{-1/2}(M)$ is a diagonal matrix, which off-diagonal elements are zeros and the diagonal elements are the reciprocal of the square root of the diagonal elements of M.

Lemma 5 (the scaled and negated inverse of the Correlation matrix).

The P^- identical to the scaled and negated version of the inverse of the correlation matrix:

$$P^- = -Diag^{-1/2}(R^{-1})\ R^{-1}\ Diag^{-1/2}(R^{-1}) \tag{2.4}$$

Proof of Lemma 5.

Using the alternative definition of the partial correlation (see Definition 3), the partial correlation is the negated correlation of $\mathcal{E}_{\xi_1 \sim (\xi_2,...,\xi_k)}$ and $\mathcal{E}_{\xi_k \sim (\xi_1,...,\xi_{k-1})}$ (see Lemma 3).

But the $\mathcal{E}_{\xi_1 \sim (\xi_2,...,\xi_k)}$ is nothing else as an orthogonal vector to the subspace generated by the variables $\xi_2, ..., \xi_k$ and the $\mathcal{E}_{\xi_k \sim (\xi_1,...,\xi_{k-1})}$ in the same way.

Let $(\xi_1^*, ..., \xi_k^*)$ the biorthogonal base for the vector set $(\xi_1^*, ..., \xi_k^*)$. In other words, let $\text{cor}(\xi_i^*, \xi_j) = \delta(i,j)$ where $\delta(i,j) = 1$ or 0 must be fulfilled, according to conditions $i = j$ and $i \neq j$, respectively. Then $\mathcal{E}_{\xi_1 \sim (\xi_2,...,\xi_k)} = \lambda_1 \xi_1^*$ and $\mathcal{E}_{\xi_k \sim (\xi_1,...,\xi_{k-1})} = \lambda_k \xi_k^*$. And the partial correlation of ξ_1 and ξ_k equal to the negated correlation of ξ_1^* and ξ_k^*, because we can drop the λ_1 and λ_k due to standardization.

By the definition of the ξ^* variables $\text{cov}(\xi^*, \xi) = I$ and for a suitable B linear transformation $\xi^* = B\xi$. This means, that if $\text{cov}(\xi) = \Sigma$ then $\text{cov}(\xi^*, \xi) = B\Sigma$, therefore

$I = B\Sigma$ and follows that $B = \Sigma^{-1}$. The statement follows directly from the fact that $\mathrm{cov}(\xi^*) = \mathrm{cov}(B\Sigma) = \mathrm{cov}(\Sigma^{-1}\Sigma\Sigma^{-T}) = \mathrm{cov}(\Sigma^{-T}) = \mathrm{cov}(\Sigma^{-1})$ □

A short example in **R** :

An P^* type partial correlation matrix generated in three different ways.

1) according to the classical definition (omitting two variables).

2) by alternative definition (only one variable omitted)

3) by inversion method

```
library("ppcor")
set.seed(123)
p <- 5
n <- 12
X <- matrix(rnorm(n*p),n,p)

# the partial correlation matrix by the package 'ppcor'
P_ref <- pcor(X)[[1]]


# the classical definition
P1 <- matrix(0,p,p)
for(i in 1:p) for(j in 1:p)
  P1[i,j] <-  cor(residuals(lm(X[,i]~X[,-c(i,j)])),
                  residuals(lm(X[,j]~X[,-c(i,j)])))

# the alternative definition
P2 <- matrix(0,p,p)
for(i in 1:p) for(j in 1:p)
  P2[i,j] <-  cor(residuals(lm(X[,i]~X[,-i])),
                  residuals(lm(X[,j]~X[,-j])))
P2[outer(1:p,1:p,'!=')] <- -P2[outer(1:p,1:p,'!=')]

# by inversion
isd <- function(M) diag(1/sqrt(diag(M)))
P3 <- solve(cor(X))
P3 <- - isd(P3) %*% P3 %*% isd(P3)
diag(P3) <- 1 # diag -1 change to 1

# the result of the four calculations is equal
all.equal(P_ref,P1) # TRUE
all.equal(P_ref,P2) # TRUE
all.equal(P_ref,P3) # TRUE
```

Theorem 4 (**properties of the partial covariance matrices [5]**).

1. *The P^- matrix is negative semi-definite.*
2. *The eigenvalues of the P^* partial correlation matrix are less then 2.*
3. *The P^+ matrix is a positive semi-definite.*

Proof of the first sentence (1.), the other two are trivial consequences.
The partial correlation matrix P^-, by the definition equals to $-D^T R^{-1} D$ for $D = Diag^{-1/2}(R^{-1})$, and for the positive definite R. If the nonzero eigenvalues of R are the positive $\lambda_1, ..., \lambda_k$, then the eigenvalues of R^{-1} are the also positive $1/\lambda_1, ..., 1/\lambda_k$ real values. And what's more the R^{-1} and the $D^T R^{-1} D$ matrices are *congruent*, which means that this two matrices have the same number of positive, negative and zero eigenvalues. So the P^- is really negative definite. □

Proof of the second sentence (2.)
If v an eigenvector of P^- with eigenvalue λ, then

$$P^* v = (P^- + 2I)v = P^- v + 2Iv = \lambda v + 2v = (\lambda + 2)v$$

For this reason, the set of eigenvectors of P^* is the same as that of P^-, and all the eigenvalues of P^* are equal to the eigenvalues of P^- plus 2. So since P^- is negative definite, all the eigenvalues of P^* are less then or equal to 2. □

Proof of the third sentence (3.)
Seeing that P^- is negative semidefinite, it is obvious that the $P^+ = -P^-$ is positive semi-definite. □

Lemma 6 (**a bijective volume preserving transformation**).
There exists a bijective and volume preserving mapping between the set of correlation and the set of partial correlation matrices.

If R is a valid correlation matrix, then the $P = -D_R R^{-1} D_R$

where $D_R = Diag^{-1/2}(R^{-1})$, is a valid partial correlation matrix.

If P is a valid partial correlation matrix, then the $R = -D_P P^{-1} D_P$

where $D_P = Diag^{-1/2}(P^{-1})$, is a valid correlation matrix.

This mappings are measure preserving transformations.

Corollary 1 ('false-method' for generating a partial correlation matrix). *We obtain a valid partial correlation matrix by generating a random correlation matrix and simply changing the sign of its off-diagonal elements. The result of this generation and transformation is a P^*-like partial correlation matrix.*

But obviously this matrix is not equal to any of the types of partial correlation matrices associated with the generated correlation matrix. That is why we call this very simple partial correlation generation method as 'false-method'.

3

Generating pseudo random correlation matrices

In the previous chapter, we discussed correlation matrices, So now we will look at some methods we used to generate random correlation matrices. Our programming background is the **R** programming language with two supplementary packages '`randcorr`' and '`clusterGeneration`'. The **R** is an open-source environment for statistical computing, supported by the **R** Core Team and the **R** Foundation for Statistical Computing. It was initiated in 1991 by statisticians Ross Ihaka and Robert Gentleman, in New-Zealand [6]. Users have since created countless packages to extend the functionality of the **R** language. We use the two above additional packages to generate a random correlation matrix. The packages which we use are the work of the statisticians who first published the methods used.

3.1 The Brute Force – a "trial and error" type algorithm

The Brute Force – a "trial and error" type algorithm :
Let generated an matrix with diagonal which is equal to 1 and with off-diagonal elements which are random, independent and on $[-1, 1]$ uniformly distributed. And repeat this action until the result isn't positive definite.

But, this extremely simple "trial and error" algorithm does not work [7].

If M_n is the set of positive definite matrices of size n, then

$$\mathbb{P}(\text{the trial is positive semi-definite}) = \frac{\text{Volume}(M_n)}{\text{Volume}\left([-1,1]^{\frac{n(n-1)}{2}}\right)} < \left(\frac{\sqrt{\pi}}{2}\right)^{\frac{n(n-1)}{2}}$$

This cause that the rejection rate is growing extremely rapidly with the size of the matrix:

Size	Probability of rejections
2	0.00000000000000000
3	0.38314972493191513
4	0.81722954812797488
5	0.97799554763240115
6	0.99905047980881456
7	0.99998671616120716
8	0.99999994457736563
9	0.99999999993580357
10	0.99999999999998057

3.2 The Pourahmadi-Wang method

The Pourahmadi-Wang method of `randcorr` package [8], [4]:

The method of M. Pourahmadi and X. Wang use the Cholesky decomposition of the correlation matrix. This technique necessity a random generation of the lower triangular Cholesky factor of the correlation matrix.
For this it uses a hyperspherical parameterization of this factor, and a random generation of the necessary angles with the Makalic-Schmidt sample-rejection algorithm. Makalic-Schmidt algorithm generate the random angles from the intervall $[0, \pi]$ with distribution proportional to $\sin x^k$.

The hyperspherical parametrization of the lower triangular Cholesky-factor of a positive

definite correlation matrix, has the following structure:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ c_{2,1} & s_{2,1} & 0 & 0 & \dots & 0 \\ c_{3,1} & c_{3,2}s_{3,1} & s_{3,1}s_{3,2} & 0 & \dots & 0 \\ c_{4,1} & c_{4,2}s_{4,1} & c_{4,3}s_{4,1}s_{4,2} & s_{4,1}s_{4,2}s_{4,3} & \dots & 0 \\ c_{5,1} & c_{5,2}s_{5,1} & c_{5,3}s_{5,1}s_{5,2} & c_{5,4}s_{5,1}s_{5,2}s_{5,3} & \dots & 0 \\ \vdots & & & & & \\ c_{p,1} & c_{p,2}s_{p,1} & c_{p,3}s_{p,1}s_{p,2} & c_{p,4}s_{p,1}s_{p,2}s_{p,3} & \dots & \prod_{k=1}^{p-1} s_{p,k} \end{pmatrix}$$

Where the letters 'c' represent the function 'cos' and the letters 's' represent the function 'sin' and the indices refers to the elements of the Θ matrix:

$$\Theta = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \vartheta_{2,1} & 0 & 0 & \dots & 0 \\ \vartheta_{3,1} & \vartheta_{3,2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vartheta_{p,1} & \vartheta_{p,2} & \vartheta_{p,3} & \dots & 0 \end{pmatrix}$$

The elements of Θ are random angles $\vartheta_{i,j}$, whose distribution is proportional to $\sin^k(x)$ on the interval $[0, \pi]$, in the j-th column with the parameter $k = p - j$, where p is the size of the requested random correlation matrix.

The algorithm of E. Makalic and DF. Schmidt for generate the random ϑ values founded on two independent random values $U \sim \mathcal{U}([0, 1])$ and $X \sim \pi \cdot \mathcal{B}eta(k+1, k+1)$ it accepts the X if

$$\log(U)/k \leq \log\left(\frac{\pi^2 \sin X}{4X(\pi - X)}\right)$$

So, this technique is a sampling-rejection algorithm where the $\mathcal{B}eta(k+1, k+1)$ distribution is the envelope distribution.

Three examples of generating random correlation matrices:

```
# install.packages("randcorr")
library("randcorr")
randcorr(5)
```

```
  S <- randcorr(1000)
  class(S) # "matrix" "array"
  dim(S) # 1000 1000
  attributes(S) # dim 1000 1000

  set.seed(123)
  S <- randcorr(12)
  round(eigen(S)$val,3) # $2.786 2.292 1.897 ... 0.152 0.032 0.02
```

The following is a list of the '`rand.theta`' and '`rcorr_ch`' functions. These are substantially simplified and faster versions of '`randcorr::randcorr.sample.sink()`' and '`randcorr::randcorr()`' functions:

```
rand.theta <- function (k)
{
    th <- rep(0    , k)
    OK <- rep(FALSE, k)
    while (!all(OK)) {
        i <- !OK
        m <- sum(i)
        th[i] <- pi * stats::rbeta(m, k + 1, k + 1)
        OK[i] <- log(stats::runif(m))/k <
                      2*log(pi/2)+log(sin(th[i]))-log(th[i])-log(pi-th[i])
    }
    return(th)
}

# ---

rcorr_ch <- function(p)
 {
   theta <- matrix(0, p, p)
   for (j in 1:(p - 1))
     theta[(j + 1):p, j] <- rand.theta(p - j)
   S <- cbind(1,(t(apply(sin(theta),1,cumprod))[,-p]))
   C <- cos(theta)
   L <- S*C
   R <- L%*%t(L)
   return(R)
 }
rcorr(4)
```

3.3 The Joe's method

The Joe's method of 'clusterGeneration' package [9],[10]:

The outline of Henry Joe's method is as follows.
If we want to generate a correlation matrix of $p \times p$ size, we first take an identity matrix of $p \times p$ size. Then fill the diagonals parallel to the main diagonal with independent $\mathcal{B}eta$ random quantities. Let the two parameters of these $\mathcal{B}eta$ distributions in the case of the d-th parallel diagonal $(p+1-d)/2$.

From the symmetric parameter matrix prepared in this way, a series of transformations modifying only the corner elements of the sub-matrices gives the desired correlation matrix.

First, the 3×3, then the 4×4, etc., and finally the symmetric sub-matrices of $p \times p$ size must be taken in this consecutive order. For each sub-matrix, only the pair of elements in the two corners of the sub-matrix must be modified in the matrix. The new value of the corner cells corresponds to the inverse of the transformation of Theorem 3:

$$r = \mathbf{r}_{1,2}\mathbf{R}_{2,2}^{-1}\mathbf{r}_{2,3} + \varrho_{\xi_1,\xi_p\bullet(\xi_2,\ldots,\xi_{p-1})}\sqrt{1-\mathbf{r}_{1,2}\mathbf{R}_{2,2}^{-1}\mathbf{r}_{2,1}}\sqrt{1-\mathbf{r}_{3,2}\mathbf{R}_{2,2}^{-1}\mathbf{r}_{2,3}}$$

Here, the $\mathbf{R}$ matrix represents the current sub-matrix.

Notes

The diagonal of the generated matrix is therefore identically 1, and the values in the first diagonal below and above the main diagonal are identical to the generated $\mathcal{B}eta$ distribution values. All $\mathcal{B}eta$ distributions used in this algorithm are to be understood on the $[-1, 1]$ interval.

Example 3.1.

Generating a correlation matrix using '`clusterGeneration::rcorrmatrix()`':

```
# install.packages("clusterGeneration")
library("clusterGeneration")

set.seed(123)
S <- rcorrmatrix(4)
class(S) # "matrix" "array"
round(eigen(S)$val,3) # 1.879 1.371 0.716 0.034$
```

The following is a list of '`rcorr_pc()`' *and* '`corner()`' *functions which are substantially simplified and faster versions of the functions* '`clusterGeneration::rcorrmatrix()`' *and* '`clusterGeneration::rjm()`' *respectively:*

```
rcorr_pc <- function (d)
  {
    if (d == 1) return(matrix(1,1,1))
    if (d == 2) return(matrix(c(1,rho<-runif(1,-1,1),rho,1),2,2))
    rr <- diag(d)
    for (j in 1:(d - 1))
        rr[j,j+1] <- rr[j+1,j] <- 2*rbeta(1,d/2,d/2)-1
    for (m in 2:(d - 1))
        for (j in 1:(d - m))
            rr[j,j+m] <-rr[j+m,j] <-corner(rr[j:(j+m),j:(j+m)],(d+1-m)/2)
    return(rr)
  }

# ---

corner <- function(subM,alp)
  {
    rcond <- 2*rbeta(1,alp,alp)-1
    b <- nrow(subM)
    M1 <- subM[2:(b-1),1]
    M3 <- subM[2:(b-1),b]
    M2 <- solve(subM[2:(b-1),2:(b-1)])
    r13 <- t(M1)%*%M2%*%M3
    r11 <- t(M1)%*%M2%*%M1
    r33 <- t(M3)%*%M2%*%M3
    return(r13+rcond*sqrt((1-r11)*(1-r33)))
  }

set.seed(123);rcorr_pc(4);set.seed(123);rcorrmatrix(4)
```

4

Adequacy tests of the factor models

4.1 The factor analysis

We will discuss factor models in this chapter and we focus on the `KMO` test.

Introduction:

In this chapter, we assume that we want to model a p-dimensional variable $\eta = (\eta_1, ..., \eta_p)$. And our objective is to model *the common information* contained in the variables $\eta_1, ..., \eta_p$. In the case of the *factor model*, we assume that there exists a $q < p$ dimensional $\xi = (\xi_1, ..., \xi_q)$ *common variable*, and a p-dimensional $\delta = (\delta_1, ..., \delta_p)$ *uniqueness* with uncorrelated coordinates, and an $\varepsilon = (\varepsilon_1, ..., \varepsilon_p)$ *error*, as little as possible, for which:

$$\eta = \mathbf{L}\xi + \delta + \varepsilon$$

where $\mathbf{L}$ is a linear transformation $\mathbb{R}^q \to \mathbb{R}^p$, that is a matrix of size $p \times q$. In this decomposition the ξ is the common variable, which coordinates are *the factors*. The coordinates of δ are *the unique factors*, each of which explains the standard deviation of only one of the coordinates $\eta_1, ..., \eta_p$. The ε is the error term, which completes the approximation $\mathbf{L}\xi + \delta$.

We assume that the three right-hand side terms are uncovaried. We want the model to *explain the covariance matrix* of the observed variable η. Therefore, we can assume without introducing further restrictions that the expected value of all elements of the

model is zero. Under these conditions, the equation of the covariance matrices is:

$$\mathbf{\Sigma}_\eta = \mathbf{L}\mathbf{\Sigma}_\xi\mathbf{L}^T + \mathbf{D} + \mathbf{\Sigma}_\varepsilon$$

where $\mathbf{D} = \mathbf{\Sigma}_\delta$, the covariance matrix of δ, a diagonal matrix with non-negative diagonal elements. The $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\xi$ and $\mathbf{\Sigma}_\varepsilon$ are the covariance matrices of η, ξ and ε respectively.

We will take the advantage of the fact that $\mathbf{L}$ is arbitrary. If $\mathbf{\Sigma}_\xi = \mathbf{R}^T\mathbf{R}$ is the Cholesky decomposition of $\mathbf{\Sigma}_\xi$, then $\mathbf{L}\xi = \mathbf{L}\mathbf{R}^T\mathbf{R}^{-T}\xi$. So with the notation $\mathbf{L}_I = \mathbf{L}\mathbf{R}^T$ and instead of ξ with the q dimensional variable $\xi_I = \mathbf{R}^{-T}\xi$ whose covariance matrix is an identity, since $cov(\xi_I) = \mathrm{cov}(\mathbf{R}^{-T}\xi) = \mathbf{R}^{-T}\Sigma_\xi\mathbf{R}^{-1} = \mathbf{R}^{-T}(\mathbf{R}^T\mathbf{R})\mathbf{R}^{-1} = \mathbf{I}_\xi$:

$$cov(\mathbf{L}\xi) = cov(\mathbf{L}R^TR^{-T}\xi) = cov(\mathbf{L}_I\xi_I) = \mathbf{L}_I\mathbf{I}_\xi\mathbf{L}_I^T = \mathbf{L}_I\mathbf{L}_I^T$$

As a result, without further limitation can be assumed, that the covariance matrix $\mathbf{\Sigma}_\xi$ is an identity matrix. In other words, it follows from the above formula and can be expected to hold that:

$$\mathbf{\Sigma}_\eta = \mathbf{L}\mathbf{L}^T + \mathbf{D} + \mathbf{\Sigma}_\varepsilon$$

The requirement for ε to be small means that the following approximation must be met:

$$\mathbf{\Sigma}_\eta \approx \mathbf{L}\mathbf{L}^T + \mathbf{D}$$

and the question is the optimal value of q, the count of the common factors, and the value of the $\mathbf{L}$, the transformation between the unobserved ξ factors and the observed η vector.

Unfortunately, there is no closed form solution to this problem.

If the $\mathbf{D}$ were known, then the problem is only the approximation of $\mathbf{\Sigma}_\eta - \mathbf{D}$, which is apparently a principal component task, with an explicit solution. But in the case of factor analysis the diagonal matrix $\mathbf{D}$ is not known and in the same time the full diagonal need not be approximated, which fact essentially changes the task.

Examples:

We present in the following the factor model of four artificial data sets of size $N = 10000$. We assume four different dependencies between the coordinates. In other words, we show four examples where the assumed data generation mechanism is significantly different. Let all the analyzed variable be 6 dimensional.

The **R** script that initializes the examples:

```
rm(list=ls())
p <- 6 # dimension of examples datasets
N <- 10000 # the size of the datasets

kmo <- function(x)
  {
    R <- if(nrow(x)==ncol(x)) x else cor(x)
    P <- corpcor::cor2pcor(R)
    sR <- sum(R^2)-ncol(R)
    sP <- sum(P^2)-ncol(P)
    return(sR/(sR+sP))
   }
```

Example 4.1.

Now we create and analyze a 6-dimensional variable whose coordinates are independent, but with different $(1.5, 1.3, 1.1, 0.9, 0.7, 0.5)$ *deviations:*

```
# a six independent variable with different standard deviations
set.seed(321)
Y1 <- matrix(rnorm(N*p),N,p)
Y1 <- Y1%*% diag(sqrt(seq(1.5,.5,length=6)))
R <- cor(Y1)
round(R,3) # the off-diagonal is approximately zero
round(corpcor::cor2pcor(R),3) # off-diag approx 0
kmo(R) #  0.4988599
factanal(Y1,3)
#                 Factor1 Factor2 Factor3
# Cumulative Var   0.078   0.085   0.091
# Uniquenesses: 0.993 0.535 0.997 0.966 0.999 0.961
```

The high values of the 'Uniquenesses' i.e the variances of the δ show that the factor model explains a very small fraction of the information encoded in the covariances of

η. This is justified because the coordinates are uncorrelated, so there is actually no common information in the coordinates.

Example 4.2.

Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ and $(\varepsilon_1, .., \varepsilon_6)$ be independent random variables, and let

$$\eta = (\eta_1, ..., \eta_6) = (\mathcal{U}, \mathcal{U}, \mathcal{U}, \mathcal{V}, \mathcal{V}, \mathcal{V}) + (0, 0, \mathcal{W}, \mathcal{W}, 0, 0) + (\varepsilon_1, ..., \varepsilon_6)$$

That is, let η be a variable formed from three factors, observed with error. The first two background variables explain the first three and the second three variables on the one hand. The third background variable is present in the fourth and fifth variables.

```
uv <- 4/5;w <- 4/7
set.seed(123)
E <- matrix(rnorm(6*N),N,6)
U <- matrix(rnorm(N),N,1)
V <- matrix(rnorm(N),N,1)
W <- matrix(rnorm(N),N,1)
Y2 <- uv*cbind(U,U,U,V,V,V)+sqrt(1-uv^2)*E
Y2[,3:4] <- sqrt(1-w^2)*Y2[,3:4]+w*cbind(W,W)
kmo(Y2) # 0.6228279
factanal(Y2,3)
# Common variables   xi_1    xi_2    xi_3
# Cumulative Var    0.277   0.549   0.679
# Loadings
#           eta_1:          0.804
#           eta_2:          0.789
#           eta_3:          0.602   0.621
#           eta_4:  0.617           0.615
#           eta_5:  0.808
#           eta_6:  0.792
# Uniquenesses: 0.351 0.374 0.251 0.239 0.343 0.370
```

The factor model clearly reconstructs the background variables. The uniquenesses are small and the cumulative explained variance is also high. In other words, this factor model is good from every point of view.

Example 4.3.

Six consecutive observations of an `AR(1)` *process.*

$$\eta_t = \varphi \cdot \eta_{t-1} + \varepsilon_t$$

```
# a 6 observations length section of an AR(1) process
phi <- 2/3
set.seed(123)
Y3 <- matrix(rnorm(N*p),N,p)
for(k in 2:p)
  Y3[,k] <- phi*Y3[,k-1] + sqrt(1-phi^2)*Y3[,k]
R <- cor(Y3)
round(R,3) # a Toeplitz matrix
round(corpcor::cor2pcor(R),3) # approx tridiagonal
kmo(R) # 0.7292996
factanal(Y3,3)

#                Factor1 Factor2 Factor3
# Cumulative Var   0.277   0.549   0.679
# Uniquenesses 0.541 0.027 0.378 0.005 0.005 0.559
```

The correlation matrix of such a process is necessarily a Toeplitz matrix, and its partial correlation matrix is tridiagonal. Because the theoretical auto-correlations are time-invariant, and the higher-order partial correlations are zero. The uniquenesses shows a great variety, because we only analyzed a short section of the stationary process.

Example 4.4.

In this example, the same information is loaded with independent errors.

$$\eta = (\eta_1, ..., \eta_6) = (\xi, \xi, \xi, \xi, \xi, \xi) + (\varepsilon_1, ..., \varepsilon_6)$$

```
# six identical coordinates + independent noise
a <- 3/4
set.seed(123)
X <- matrix(rep(rnorm(N),6),N,6)
E <- matrix(rnorm(N*6),N,6)
Y4 <- a*X +sqrt(1-a^2)*E
R <- cor(Y4)
round(R,3) # the out-diagonal is a constant
corpcor::cor2pcor(R) # constant out-diag
kmo(R) #    0.9132102
factanal(Y4,3)
#                Factor1 Factor2 Factor3
# Cumulative Var   0.332   0.634   0.637
# Uniquenesses: 0.005 0.432 0.439 0.425 0.445 0.432
```

The common information ξ is obviously present in these coordinates, loaded with errors. The out-diagonal correlations and the out-diagonal partial correlations are necessarily

constants. The standard deviations were chosen so that the partial correlation are smaller. The uniqueness is only good for the first variable. Because this is an overfitted model. For these data, the correct number of factors would be 1.

A natural question is:

How can we foresee whether a data set can be modeled well with the factor model?

If the correlation matrix is a unity matrix, then all eigenvalues of the correlation matrix are one. In a factor analysis, this case would correspond to such a solution where the highest number of observable variables – that strongly correlate with a calculated factor – is only 1. Thus, if the out-diagonal elements of the correlation matrix is a zero, then there is no common information, and the factor analysis solutions cannot explain nothing for the whole information encoded in the observed variables.

The partial correlation measures the relationship strength of that parts of the two variables which cannot be explained by the other variables. This means that a good factor model is not possible if the partial correlation values are too large.

These considerations show that a good factor model can be constructed if the correlations are large and the partial correlations are relatively small.

A kind of test of these aspects is the use of `KMO` statistic.

4.2 The Kaiser–Meyer–Olkin (`KMO`) statistic

Let $\mathbf{R}$ be the correlation matrix for the observations and $\mathbf{P}$ the partial correlation matrix for the same observations. Denote the elements of $\mathbf{R}$ by $r_{i,j}$ and the elements of $\mathbf{P}$ by $p_{i,j}$. Then the `KMO` statistics of the observations is

$$\texttt{KMO} = \frac{\sum_{i \neq j} {r_{i,j}}^2}{\sum_{i \neq j} {r_{i,j}}^2 + \sum_{i \neq j} {p_{i,j}}^2}$$

It is clear that the value of `KMO` is certain in the interval $[0, 1]$. At first glance, it may seem that the possible values for `KMO` are the full $[0, 1]$ interval.

That is each correlation can be 0, and in the same way each partial correlation can be 0. But as we will show in the next chapter, this is not true.

The literature [11] recommends that the KMO value for factor analysis be evaluated according to the following table.

KMO	interpretation
in the .90s,	marvelous
in the .80s,	meritorious
in the .70s,	middling
in the .60s,	mediocre
in the .50s,	miserable
below .50s,	unacceptable

This recommendation is widely used in practice. In the next section, we show that in light of the frequency of occurrence of possible values of KMO, this assessment is questionable. Especially for low dimensions.

The KMO is an abbreviation derived from initials of mathematicians: HF. Kaiser, GJ. Meyer and I. Olkin. They have published several articles using this statistic.

Ratings of the four above examples according to the KMO :

The following script, using the befors defined 'kmo()' function calculate the KMO in the case of the four examples:

```
kmo(Y1) # 0.4988599  'unacceptable'
kmo(Y2) # 0.6228279  'mediocre'
kmo(Y3) # 0.7292996  'middling'
kmo(Y4) # 0.9132102  'marvelous'
```

So, in the previous four examples, the KMO is 0.5, 0.62, 0.73 and 0.91 respectively.

It can be seen that, in the case of the first example, when the sample has independent coordinates and the 'uniquenesses' of the factor model are also large, we obtained an 'unacceptable' KMO value according to Kaiser's classification.

In the second case, when the variables are builded from 3 common factors, the factor model is also better and the `KMO` rating is 'mediocre'.

The third example is a special case. Here, the data is a section of an `AR1` process. It is then known that the higher order partial correlations are zero. In other words, it is necessary for the `KMO` qualification to be better. It can be seen that the factor model is also better: higher cumulative variance and lower uniquenesses.

The fourth example above shows that the classification of the sample according to `KMO` can be even 'marvelous' if the off-diagonal of the correlation matrix is constant.

5

Study of the KMO statistic

5.1 A simulation study of the KMO statistics

The next generator and KMO calculator function will be used

```
KMO <- function(R)
  {
    P <- corpcor::cor2pcor(R)
    R2 <- sum(R^2) - ncol(R)
    P2 <- sum(P^2) - ncol(P)
    return(KMO=R2 / (R2+P2))
  }
randKMO <- function(k)
  {
    R <- randcorr::randcorr(k)
    return(randKMO=KMO(R))
  }
```

We studied the distribution of KMO statistics based on samples of 100,000 sizes for random matrices of different sizes. Thus, the random correlation matrices were generated using an implementation of the Pourahmadi-Wang ([8]) algorithm.

The three tests that laid the foundation for the rest :

- The estimated range (minimum and maximum) of the possible KMO values (5.1).
- The quantiles of the distribution of the KMO statistics (5.2).
- The kernel smoothed density function estimate of the KMO statistics (5.1).
- The sizes of the examined matrices were $k = 3, 4, 5, 6, 7$ and 8 (5.2).

matrix size	minKMO	maxKMO
3x3	0.20038760	0.79462910
4x4	0.10638300	0.86166370
5x5	0.07099476	0.84957787
6x6	0.06393129	0.81668437
7x7	0.05950654	0.77532830
8x8	0.05170802	0.74238502

Table 5.1: The estimated range of the random KMO values by a sample of size 100 thousand

matrix size	0%	25%	50%	75%	100%
3x3	20.0	35.8	46.4	54.3	79.5
4x4	10.6	30.5	41.1	50.8	86.2
5x5	7.1	26.9	36.6	46.4	85.0
6x6	6.4	24.3	33.1	42.3	81.7
7x7	6.0	22.1	30.2	38.7	77.5
8x8	5.2	20.3	27.6	35.6	74.2

Table 5.2: The estimated quantiles of the random KMO values by a sample of size 100 thousand

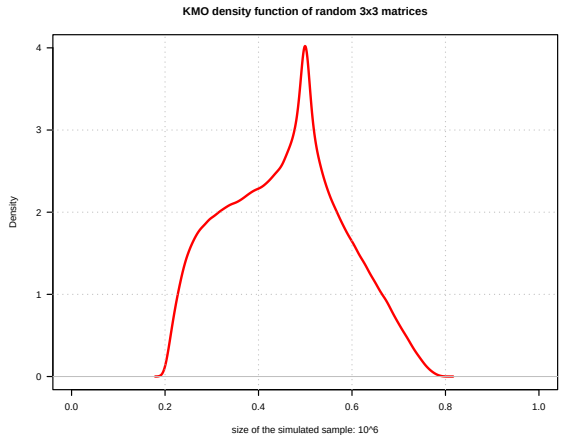

Figure 5.1: The frequency distribution of the KMO statistic for a 3×3 random matrix

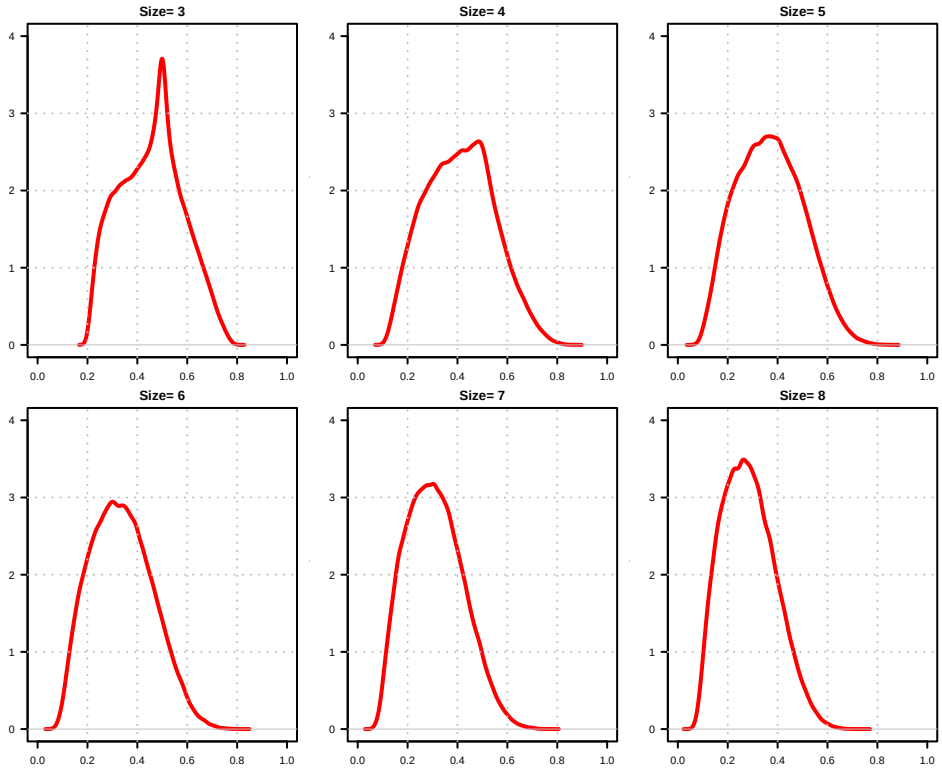

Figure 5.2: Density function of KMO for different matrix sizes.

The density function drawings above are a bit misleading. Because the range of possible values for KMO statistics is growing in reality. But this is not reflected in the density function estimates made by the kernel function smoothing method, because the probability of occurrence of extreme KMO values is very small.

5.2 Five different Claimes for KMO statistics

5.2.1 **Claime 0** - the property of the false pair of a correlation matrix

The application of the volume preserving bijection of the Lemma 6, allows an alternative way to investigate the distribution of the KMO statistical.

Take a random correlation matrix and by negating its off-diagonal elements we catch a random partial correlation matrix.
It is important that the resulting matrix is not the partial correlation matrix's counterpart to the original correlation matrix.

The script below demonstrates, that the 'false' partial correlation pair of a correlation matrix (see the Lemma 6) has a KMO statistic, that is the complement of the KMO of the original correlation matrix.

```
k <- 5
set.seed(123)
X <- randcorr::randcorr(k)
R1 <- X
P1 <- corpcor::cor2pcor(R1)
kmo1 <- (sum(R1^2) - k) / (sum(R1^2) + sum(P1^2) - 2*k)
P2 <- -X
diag(P2) <- 1
R2 <- corpcor::pcor2cor(P2)
kmo2 <- (sum(R2^2) - k) / (sum(R2^2) + sum(P2^2) - 2*k)

c(kmo1,kmo2,kmo1+kmo2) 0.2111204 0.7888796 1
```

This shows, taking into account the asymmetry of the distribution of the KMO statistic, that from the point of view of the distribution of the KMO statistic, it does not matter whether the correlation or partial correlation matrix is considered uniformly distributed, among all possible matrices.

5.2.2 **Claime 1** - increasing the KMO

If we change the correlations of a correlation matrices to a constant, which is equals to the squared mean of the correlations, then
- the matrix remains positive semi-definite
- the KMO value does not decrease

The following simulation shows the truth of the Claime 1.

```
PR_eq <- function(k=3)
  {
    R0 <- randcorr(k)
    i <- 1:k; j <- 1:k;
```

```
    r0 <- R0[outer(i,j,"<")]; r1 <- sqrt(mean(r0^2))
    R1 <- matrix(r1,k,k); diag(R1) <- 1
    pos.def <- all(eigen(R1)$val>0)
    P0 <- cor2pcor(R0); P1 <- cor2pcor(R1);
    sR0 <- sum(R0^2)-k; sP0 <- sum(P0^2)-k
    sR1 <- sum(R1^2)-k; sP1 <- sum(P1^2)-k
    kmo0 <- sR0/(sR0+sP0);kmo1 <- sR1/(sR1+sP1);
    log <- c(pos.def=pos.def,less=sum(P0^2)>sum(P1^2),better=kmo0<kmo1)
    return(list(c(sR0=sum(R0^2),sR1=sum(R1^2),sP0=sum(P0^2),sP1=sum(P1^2)),
                kmo=c(KMO0=kmo0,KMO1=kmo1),
                log=log))
} # $
```

5.2.3 **Claime 2** - exact range of the KMO

The exact range of the KMO value of an matrix of size $k \times k$ is :

$$(\ 1/(k^2 - 2 * k + 2) \ , \ (k-1)^2/(k^2 - 2 * k + 2) \ )$$

The plot of the exact ranges of the possible KMO values:

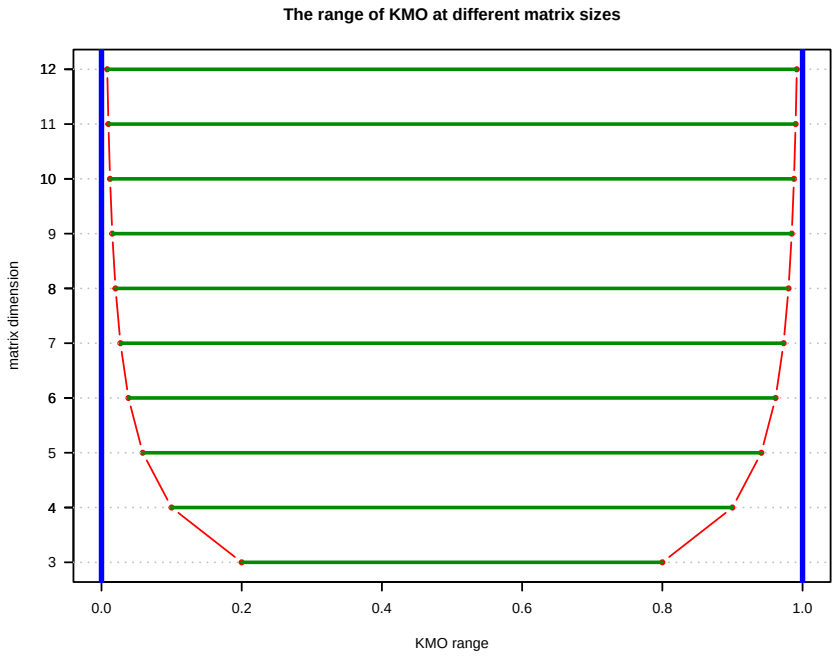

Figure 5.3: The exact KMO range for different matrix sizes.

Proof of Claime 2 :

First of all, we calculate the value of the partial correlations as a function of the size of the matrix using the formula 1.5 of Theorem 1 and then using the recursive formula 1.6 of Theorem 2.

The off-diagonal elements of the partial correlation matrix of size 3×3 matrices, if the matrix belongs to the correlation matrix of size 3×3 with constant r correlations:

$$p_3 = p_{1,3} = \frac{r_{1,3} - r_{2,1}r_{2,3}}{\sqrt{1 - r_{2,1}^2}\sqrt{1 - r_{2,3}^2}} = \frac{r - r^2}{\sqrt{1 - r^2}\sqrt{1 - r^2}} = \frac{r - r^2}{1 - r^2} = \frac{r}{1 + r}$$

We know that $p_k = \frac{r}{1+(k-2)\cdot r}$ for $k = 3$.

We show that the same is true for $k = k + 1$

$$p_{k+1} = \frac{p_k}{1 + p_k} = \frac{\frac{r}{1+(k-2)\cdot r}}{1 + \frac{r}{1+(k-2)\cdot r}} = \frac{r}{1 + (k + 1 - 2)r}$$

In a matrix of size $k \times k$ there are $k(k-1)$ off-diagonal elements.

So the KMO value of a correlation matrix with constant r correlations is :

$$\mathtt{KMO}_k = \frac{k(k-1)\cdot r^2}{k(k-1)\cdot r^2 + k(k-1)\cdot p_k^2} = \frac{r^2}{r^2 + \left(\frac{r}{1+(k-2)\cdot r}\right)^2} =$$

$$= \frac{1}{1 + \frac{1}{(1+(k-2)\cdot r)^2}} = \frac{(1 + (k-2)\cdot r)^2}{1 + (1 + (k-2)\cdot r)^2}$$

□

The exact ranges of KMO statistics depending on matrix size:

	exact range of KMO	
matrix size	inf	sup
3x3	1/5 = 0.200	4/5 = 0.800
4x4	1/10 = 0.100	9/10 = 0.900
5x5	1/17 = 0.059	16/17 = 0.941
6x6	1/26 = 0.038	25/26 = 0.962
7x7	1/37 = 0.027	36/37 = 0.973
8x8	1/50 = 0.020	49/50 = 0.980

5.2.4 **Claime 3** - the extreme point of the KMO parametrized by the correlations

The minimum and maximum of the possible values for the KMO is accessible. Both the minimum and the maximum are taken by a correlation matrix whose off-diagonal elements are constant.

If the off-diagonal elements of a matrix of size $k \times k$ are $-1/(k-1)$ then the KMO of this matrix is minimal. If the off-diagonal elements of a matrix of size $k \times k$ approximate the 1 from below, then the KMO of this matrix approximate his available maximum value.

The KMO could apparently be 0. After all, if the correlation matrix is an identity matrix, the numerator of the KMO is zero. However, the definition of partial correlations is uncertain. It does not matter whether the definition of the correlation of residuals or the method of calculation from correlations is considered logical.
In any case, with both views, it is acceptable to consider the partial correlations to be zero. But under the specified conditions, the denominator of KMO is also zero.
However, there is no definition under which the value of KMO is continuous at this point. The same problem with the correlation matrix in which all correlations are 1.

In short, the KMO is not continuous in the points $-1/(k-1)$ and 1, that is, in the environment of the extremal points of the domain.

5.2.5 **Claime 4** - the extreme point of the KMO by the partial correlations

The extreme values of the KMO can also be achieved by the appropriate selection of the partial correlation matrix. Again, a partial correlation matrix whose off-diagonal elements are constants must be chosen. Only the minimum and maximum locations will be reversed. The KMO is nearly minimal when the off-diagonal partial correlations approach 1 from below. KMO is nearly maximal when the off-diagonal partial correlations approach $-1/(k-1)$ from above.

The situation is equivalent as for the correlation matrix. This can be explained on the basis of the relationship between the correlation matrix and the partial correlation matrix, which was explained in detail in the Chapter 2, based on the article [5].

5.3 Relation between different matrix generators in the distribution of KMO statistics

The aim of this last study, documented in the following figure is to illustrate the role of different matrix generators in the distribution of KMO statistics.

This study is important and interesting because several generators are known to choose from possible matrices according to an undocumented, unknown distribution.
Moreover, there are some for which not all possible correlation matrices are possible.

We examine the case of matrices of only 3×3. This is because we also wanted to include the *brute force* method in the study, and their efficiency deteriorates significantly as the size increases.

Based on experiments, we found that the examined distributions are sufficiently accurate for 100 thousand samples. The distributions were estimated by kernel smoothing based on the generated samples. The obtained density functions were compared only visually, according to our objectives.

For different generation methods, we will examine how changes the distribution of KMO depending on whether the we choose at random the correlation or the partial correlation matrix.

In both cases, we also examine the distribution under the brute force method. The result of this method is a kind of reference. This is because according to this method we choose at first a matrix with elements distributed uniform and independent in $[-1, 1]$. That is, a random point from the $k(k-1)/2$ dimensional parameter space of $k \times k$ symmetric matrices for a matrix with a diagonal of 1 and -1, respectively.

For the correlation matrix, this is repeated until a positive definite matrix is obtained. For a partial correlation a matrix with each eigenvalue less than 2 is necessary.

That is, the ‘brute force’ method as an "accept-reject type algorithm" take a random matrix slowly but with obviously uniform distribution from all the possible matrices.

The simulation shows that the brute force and the two more efficient generation methods give a similarly distributed KMO value. Thus, according to the theoretical results, the Pourahmadi-Wang method of parameterizing Cholesky and the method H. Joe of parameterizing the correlation matrix with partial correlation indeed generates a correlation matrix with a uniform distribution.

If the partial correlation is chosen instead of the correlation matrix according to a uniform distribution, the distribution of KMO will be the reflection of the original distribution.

The latter remark, from an analytical point of view, is a very important result in the evaluation of KMO in the case of empirical data sets.

5.3.1 Distribution of KMO statistic of a 3×3 matrix

Our plot below consists of six sub-plot.

The two columns differ in which matrix has a uniform distribution
- on the left side it is seen the distribution of KMO when the correlation matrix taken at random
- on the right side is the distribution of KMO when the partial correlation matrix taken at random

The rows show the results obtained when different random generators are used
- in the first row there are the results using the brute force method
- in the second row there are the results using the Cholesky parametrization method
- in the last row there are the results using the Joe's method

The sample size was in all cases 100 thousand.

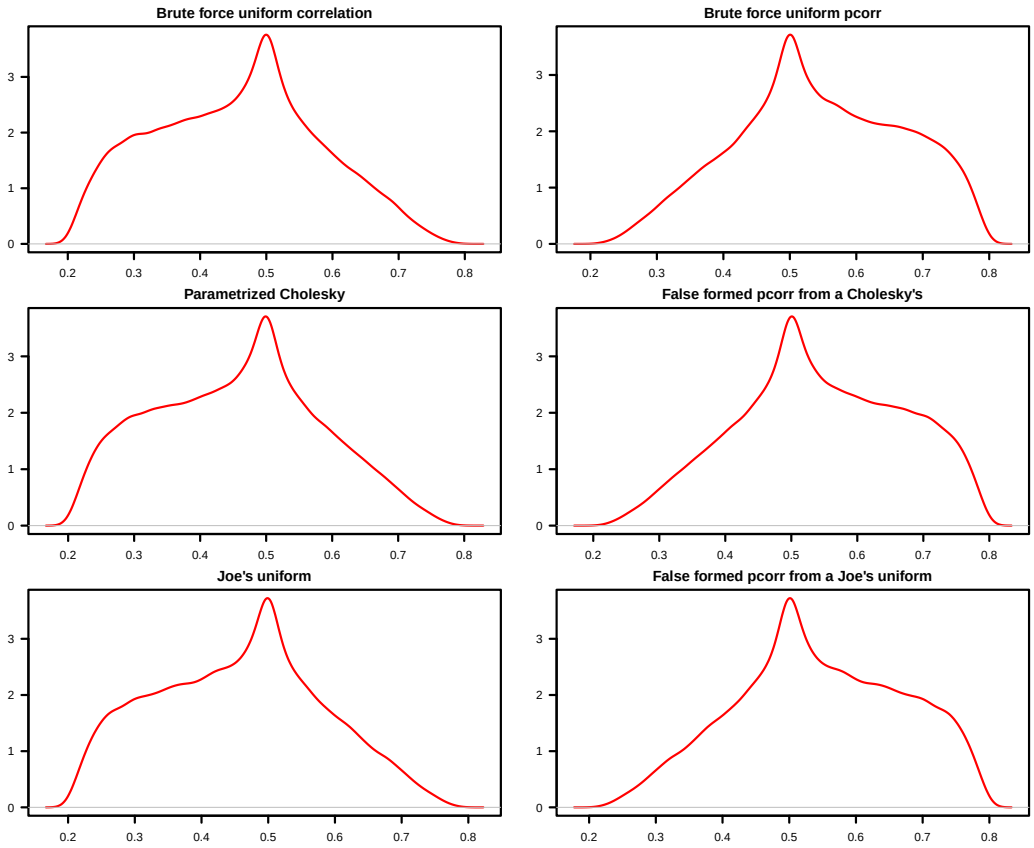

Figure 5.4: Density function of KMO depending on the random generator used.

The previous study shows that the distribution of KMO strongly depends on which of the correlation and partial correlation matrix is chosen with a uniform distribution from the all possible ones.

This also means that the classification of a random sample according to KMO strongly depends on how we model the formation of the particular sample.

5.3.2 Distribution of KMO statistic of a 7×7 matrix

The table 5.3 shows the estimated occurrence probabilities of KMO ratings for 7×7 matrices, and also it shows that the good rating is extremely rare for random correlation matrices. In a sample of 1 million items, the KMO was only one time greater than .8.

Despite, as we have seen, that for 7×7 matrices, the set of theoretical values of KMO is the same as the interval [0.027027031,0.9729729]. For the simulated data the range was much shorter. In the case of uniform correlation matrix [0.05950654, 0.77532830], and in the case of the uniform partial correlation matrix [0.2154118, 0.9450869].

But the theoretical extremes are approachable because e.g
Consider two special off-diagonal constant matrices.
One that is close to minimum and another that is close to maximum:

$$\mathbf{KMO}\left(\begin{pmatrix} 1 & r_0 & r_0 & r_0 & r_0 & r_0 & r_0 \\ r_0 & 1 & r_0 & r_0 & r_0 & r_0 & r_0 \\ r_0 & r_0 & 1 & r_0 & r_0 & r_0 & r_0 \\ r_0 & r_0 & r_0 & 1 & r_0 & r_0 & r_0 \\ r_0 & r_0 & r_0 & r_0 & 1 & r_0 & r_0 \\ r_0 & r_0 & r_0 & r_0 & r_0 & 1 & r_0 \\ r_0 & r_0 & r_0 & r_0 & r_0 & r_0 & 1 \end{pmatrix}\right) = 0.0270286 \qquad \text{when } r_0 = -\frac{1}{6} + 10^{-6}$$

$$\mathbf{KMO}\left(\begin{pmatrix} 1 & r_1 & r_1 & r_1 & r_1 & r_1 & r_1 \\ r_1 & 1 & r_1 & r_1 & r_1 & r_1 & r_1 \\ r_1 & r_1 & 1 & r_1 & r_1 & r_1 & r_1 \\ r_1 & r_1 & r_1 & 1 & r_1 & r_1 & r_1 \\ r_1 & r_1 & r_1 & r_1 & 1 & r_1 & r_1 \\ r_1 & r_1 & r_1 & r_1 & r_1 & 1r_1 & r_1 \\ r_1 & r_1 & r_1 & r_1 & r_1 & r_1 & 1 \end{pmatrix}\right) = 0.9729729 \qquad \text{when } r_1 = 1 - 10^{-6}$$

In other words,extreme values can be approached to any extent.
So the occurrence of a "marvelous" sample is not ruled out either, But we didn't have that.
The estimated probabilities of possible ratings – recommended by Tukey, – based on the simulated size 1 million sample are as follows:

The empirical probabilities were calculated from a sample of 1 million items.

KMO of a matrix of size 7×7		% probability of occurrence at uniform R	uniform P
marvelous	1 - .9	0.00	0.76
meritorious	.9 - .8	0.00	18.49
middling	.8 - .7	0.03	30.51
mediocre	.7 - .6	0.66	28.23
miserable	.6 - .5	4.99	16.28
unacceptable	.5 - 0	94.32	5.73

Table 5.3: The probability of occurrence of Kaiser's KMO types in 7×7 matrices

The random correlation matrices was obtained by the Cholesky method (see Pourahmadi-Wang [8], [4]) and the partial correlation matrices was formed by the 'false-method' from random correlation matrices (see Artner and al. [5]).
This result, if we assume that our observed data set follows a random correlation matrix, or a random partial correlation matrix, casts doubt on Kaiser's assessment of the KMO statistic.

6

General conclusion

Our work consists of the following main elements and results:

1. Properties of the correlation matrices are presented.
2. Several alternative definitions for the partial correlation.
3. Magnitude relationship of the partial correlation and the correlation:
   it is not necessarily less than or greater.
4. Calculations of the partial correlation matrix based on the correlation matrix.
5. -We briefly described the factor analysis model.
   -We focus on the role of `KMO` test.
6. The main results are the consequences of the simulation studies of the `KMO` test.
7. We presented three different methods for generating random correlation matrices.
8. Joe's method based on the corner partial correlations is faster and more efficient compared with the two other methods:
   - the Brute Force – a "trial and error" type algorithm and
   - the Pourahmadi-Wang method based on Cholesky decomposition.
9. Simulations shows that the `KMO` is certain in a sub-interval of the interval $[0, 1]$.
10. The test result shows
    a refutation of the classical evaluation method of `KMO` statistics.

11. The simulation shows that the nature of the distribution does not depend on which random generator is used.

12. At the same time, the distribution of KMO strongly depends on whether the correlation or partial correlation matrix is chosen according to a uniform distribution.

13. However, the non-symmetrical distribution obtained according to the two different generation methods examined is a mirror image of each other.

Future work

In our dissertation, we have summarized some interesting results regarding the KMO statistic used to evaluate the result of a factor analysis. These results seem to be unknown not only in the applied but also in the theoretical literature. But it is certainly true that these facts are ignored when the results of factor models are evaluated.

Our results are partly based on calculations, but on several points based on simulations only. Since these results show that the Kaiser's standard assessment of the KMO statistics is not unconditionally well-founded, it would be justified to examine these results in more detail, not only on a simulation basis.

List of Figures

Bibliography

[1] A. Kanto; S. Puntanen. **A connection between the partial correlation coefficient and the correlation coefficient of certain residuals**. Communications in Statistics - Simulation and Computation, pages 639–641, 1983. (8)

[2] Guy Der Megreditchian; Pierre Cazes. **Nouvelle démonstration et généralisation de la formule de Kanto et Puntanen sur les corrélations entre résidus des régressions linéaires**. Annales de l'ISUP, Publications de l'Institut de Statistique de l'Université de Paris, pages 25–39, 1988, XXXIII (1). ffhal-03672284. (8)

[3] Juliane Schafer, Rainer Opgen-Rhein, Verena Zuber, Miika Ahdesmaki, A. Pedro Duarte Silva, and Korbinian Strimmer. corpcor: Efficient Estimation of Covariance and (Partial) Correlation, 2021. R package version 1.6.10. (15)

[4] E. Makalic; DF. Schmidt. **An efficient algorithm for sampling from $sin^k(x)$ for generating random correlation matrices**. arXiv:1809.05212, 2018. (15), (20), (44)

[5] R. Artner; PP. Wellingerhof; G. Lafit; T. Loossens; W.Vanpaemel; F. Tuerlinckx. **The shape of partial correlation matrices**. Communications in Statistics - Theory and Methods, 2020m09d03. (17), (39), (44)

[6] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2021. (19)

[7] W. Böhm; K. Hornik. **Generating random correlation matrices by the simple rejection method: Why it does not work?** Statistics and Probability Letters, pages 27–30, 2014. (19)

[8] M. Pourahmadi; Xiao Wang. **Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor**. Statistics and Probability Letters, pages 5–12, 2015. (20), (33), (44)

[9] H. Joe. **Generating Random Correlation Matrices Based on Partial Correlations**. Journal of Multivariate Analysis, page 2177–2189, 2006. (23)

[10] Weiliang Qiu and Harry Joe. clusterGeneration: Random Cluster Generation (with Specified Degree of Separation), 2020. R package version 1.3.7. (23)

[11] HF. Kaiser. **An index of factorial simplicity**. Psychometrika, pages 31–36, 1974. (31)