

HELGEM DE SOUZA RIBEIRO MARTINS

**CM-GENERATOR: UMA METODOLOGIA PARA GERAÇÃO DE
MATRIZES DE CORRELAÇÃO CUSTOMIZADAS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

Orientador: Fernando Luiz Pereira de Oliveira

Coorientador: Anderson Ribeiro Duarte

**VIÇOSA - MINAS GERAIS
2020**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

M386c
2020
Martins, Helgem de Souza Ribeiro, 1987-
CM-generator : uma metodologia para geração de matrizes
de correlação customizadas / Helgem de Souza Ribeiro Martins.
– Viçosa, MG, 2020.
67 f. : il. (algumas color.) ; 29 cm.

Orientador: Fernando Luiz Pereira de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Dados aleatórios (Estatística). 2. Matrizes (Matemática).
3. Correlação (Estatística). 4. Simulação. 5. Gencor (Software).
I. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-Graduação em Estatística Aplicada e
Biometria. II. Título.

CDD 22. ed. 519.5

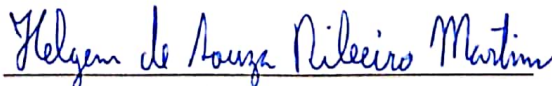
HELGEM DE SOUZA RIBEIRO MARTINS

**CM-GENERATOR: UMA METODOLOGIA PARA GERAÇÃO DE
MATRIZES DE CORRELAÇÃO CUSTOMIZADAS**

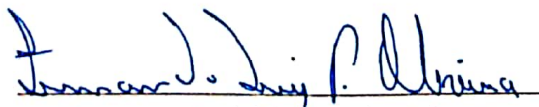
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

APROVADA: 21 de outubro de 2020.

Assentimento:



Helgem de Souza Ribeiro Martins
Autor



Fernando Luiz Pereira de Oliveira
Orientador

Dedicatória

Dedico este trabalho a meus pais e a minha namorada Rosielle, que sempre me apoiaram na busca por conhecimento.

Agradecimentos

Gostaria de agradecer inicialmente aos meus pais pela oportunidade de estudar. Muitas crianças que passaram por situações parecidas com as que passei desistiram dos estudos, mas sempre recebi deles o necessário para prosseguir.

Em segundo lugar gostaria de agradecer minha namorada Rosielle por toda a paciência durante meu doutoramento, foram muitos dias nos quais não pude lhe dar a atenção merecida.

Agradeço aos professores Anderson e Fernando, componentes de minha comissão orientadora pelos conhecimentos repassados.

Agradeço especialmente o Prof. Anderson Ribeiro Duarte, que acompanhou toda minha formação em Estatística desde a iniciação científica, passando pelo TCC, mestrado e doutorado, a quem agora tenho o prazer de chamar de amigo. Meu muito obrigado.

Agradeço meus amigos Josino e Maurício, o primeiro pela amizade de longa data e pela companhia nas viagens para Viçosa e o segundo pelo carinho com o qual sempre me recebeu em sua casa nos momentos em que precisei de guarida na cidade.

Agradeço também meus colegas de trabalho da Pró-Reitoria de Pesquisa e Pós-graduação da UFOP por todo o apoio e incentivo, além de sempre contribuir para o sucesso de minha jornada reduzida de trabalho enquanto servidor-estudante.

Agradeço aos amigos de pós-graduação pelos momentos de convivência e conhecimento compartilhado.

Agradeço aos membros da banca de avaliação pelas valiosas contribuições e aos professores do Programa de Pós-Graduação em Estatística Aplicada e Biometria pelos conhecimentos adquiridos.

Agradeço à Universidade Federal de Ouro Preto pela concessão de redução de jornada de trabalho e apoio financeiro para meu deslocamento entre as cidades de Ouro Preto e Viçosa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Epígrafe

Não há luz para aqueles que não conhecem a escuridão.

Takehiko Inoue

Resumo

MARTINS, Helgem de Souza Ribeiro, D.Sc., Universidade Federal de Viçosa, outubro de 2020. **CM-generator: uma metodologia para geração de matrizes de correlação customizadas.** Orientador: Fernando Luiz Pereira de Oliveira. Coorientador: Anderson Ribeiro Duarte

A simulação de matrizes de correlações aleatórias é um procedimento importante em diversas áreas de pesquisa. Este estudo propõe um método denominado Custom Matrix generator (CM-generator) para gerar matrizes de correlação que sempre atendem às premissas matemáticas, e um algoritmo com base na metodologia desenvolvida. A técnica proposta é capaz de gerar matrizes de correlação personalizadas, tanto em termos de intensidade das correlações quanto em relação à distribuição de proporções entre níveis de intensidade de correlação, para diversas aplicações, que podem ser usadas em vários estudos. O método produz resultados eficientes quanto ao tempo computacional e minimiza erros no processo de geração de matrizes de correlação personalizadas.

Palavras-chave: Matrizes de correlação aleatórias. Simulação. Matrizes de correlação personalizadas. Gencor.

Abstract

MARTINS, Helgem de Souza Ribeiro, D.Sc., Universidade Federal de Viçosa, October, 2020. **CM-generator: a methodology to generate customized correlation matrices.** Advisor: Fernando Luiz Pereira de Oliveira. Co-advisor: Anderson Ribeiro Duarte

The simulation of random correlation matrices is an important procedure in several areas of research. This study proposes a method called Custom Matrix generator to generate correlation matrices that always meet mathematical premises, and an algorithm based on the developed methodology. The proposed technique can generate customized correlation matrices, either in terms of the correlations intensity and regarding the distribution of proportions among levels of correlations intensity, for different applications which can be used in several studies. The method produces efficient results in terms of computational time and minimizes errors in the process of generating customized correlation matrices.

Keywords: Random correlation matrices. Simulation. Custom correlation matrices. Gencor.

Lista de Figuras

| | |
|--|----|
| 3.2.1 Algoritmo geral para geração de matrizes de correlação | 26 |
| 3.2.2 Algoritmo para geração de matrizes de correlação com limites pré-definidos | 27 |
| 3.2.3 Algoritmo para geração de matrizes de correlação com limites pré-definidos | 29 |
| 3.2.4 Algoritmo para procedimento de otimização | 31 |
| 4.1.1 Tempos de execução para matrizes gerais. | 34 |
| 4.1.2 Tempos de execução para matrizes com $0,3 < \rho \leq 0,8$ | 35 |
| 4.1.3 Tempos de execução para matrizes customizadas. | 37 |
| 4.1.4 Quantidade de simulações para otimização para matrizes customizadas. | 38 |
| 4.2.1 Distribuição de correlações através dos métodos Onion, C-vine e CM-generator com $\ell_{lower} = 0,005$ | 41 |
| 4.2.2 Distribuição de correlações através dos métodos Onion, C-vine e CM-generator com $\ell_{lower} = 0,025$ | 42 |
| 4.2.3 Evolução da distribuição de correlações através do método CM-generator para diferentes valores de ℓ_{lower} | 43 |
| 4.3.1 Histogramas para matrizes com limites nas correlações geradas através do CM-generator, (a) correlações baixas, (b) correlações médias, e (c) correlações altas. | 45 |
| 4.3.2 Histogramas para matrizes customizadas com $p\%$ de correlações baixas, $q\%$ de correlações médias, e $(100 - p - q)\%$ de correlações altas geradas através do CM-generator, (a) $p = 50\%$ e $q = 40\%$, (b) $p = 40\%$ e $q = 40\%$, (c) $p = 10\%$ e $q = 50\%$, (d) $p = 30\%$ e $q = 20\%$, (e) $p = 10\%$ e $q = 80\%$, (f) $p = 0\%$ e $q = 80\%$, (g) $p = 50\%$ e $q = 0\%$, (h) $p = 70\%$ e $q = 30\%$ | 46 |
| 4.4.1 Comparação entre acurácias obtidas através dos estimadores robustos MCD e MVE, (a) $\delta = 0.02$ e $p = 5$, (b) $\delta = 0.02$ e $p = 10$, (c) $\delta = 0.02$ e $p = 50$, (d) $\delta = 0.05$ e $p = 5$, (e) $\delta = 0.05$ e $p = 10$, (f) $\delta = 0.05$ e $p = 50$, (g) $\delta = 0.10$ e $p = 5$, (h) $\delta = 0.10$ e $p = 10$, (i) $\delta = 0.10$ e $p = 50$ | 49 |
| 5.1.1 Resultado da função plot para o objeto x. | 62 |

Lista de Tabelas

| | |
|---|----|
| 4.1.1 Tempos de execução para matrizes gerais. | 33 |
| 4.1.2 Tempos de execução para matrizes com $0,3 < \rho \leq 0,8$ | 35 |
| 4.1.3 Tempos de execução para matrizes customizadas. | 36 |
| 4.1.4 Quantidade de simulações necessárias para otimização. | 37 |
| 4.1.5 Avaliação para matrizes customizadas, tempo, volume de execuções para otimização e proporção de erros. | 39 |

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 11 |
| 1.1 | Organização da Tese | 12 |
| 2 | Revisão Bibliográfica | 14 |
| 3 | Material e Métodos | 19 |
| 3.1 | Considerações Matemáticas | 20 |
| 3.1.1 | Limites Superiores e Inferiores | 23 |
| 3.2 | Geração das Matrizes de Correlação | 25 |
| 3.2.1 | Algoritmo Geral | 25 |
| 3.2.2 | Algoritmo para Matrizes de Correlação Limitadas | 26 |
| 3.2.3 | Algoritmo para Matrizes de Correlação Customizadas | 27 |
| 4 | Resultados e Discussão | 32 |
| 4.1 | Experimentos Básicos | 32 |
| 4.1.1 | Geração de Matrizes sem Especificação Prévia | 33 |
| 4.1.2 | Geração de Matrizes com Limites Específicos para cada Medida de Correlação | 34 |
| 4.1.3 | Geração de Matrizes Customizadas | 36 |
| 4.2 | Experimentos Comparativos | 40 |
| 4.3 | Experimentos Customizados | 44 |
| 4.4 | Aplicação na Comparação entre Metodologias para Detecção de <i>Outliers</i> Multivariados | 47 |
| 5 | Pacote gencor | 51 |
| 5.1 | Utilização do pacote gencor() | 51 |
| 6 | Conclusão | 65 |

Capítulo 1

Introdução

Informações sobre as associações entre subgrupos de coordenadas de vetores aleatórios são conhecimentos relevantes em diversos problemas de interesse prático. A incerteza sobre essas associações, em geral, é explicada por meio de matrizes de correlação. Matrizes de correlação, na definição matemática, são matrizes simétricas, com diagonal unitária, demais valores pertencentes ao intervalo $[-1, 1]$, além de positivas semi-definidas, ou seja, com autovalores não negativos. Matrizes com essas propriedades são de notória relevância em investigações estatísticas. Enquanto a garantia das três primeiras características é relativamente simples de se obter, é na característica inerente aos autovalores que as metodologias de geração de matrizes de correlação existentes costumam encontrar alguma dificuldade no processo de geração.

Existe toda uma área da Estatística que depende diretamente da ocorrência de correlações significativas entre determinados grupos de variáveis, a Estatística multivariada. Nesta área, a pedra fundamental para a utilização de suas técnicas é a matriz de correlação. Tal matriz carrega toda a informação relevante com respeito a associação dos dados em estudo. De lá são extraídos elementos matemáticos que, além da já mencionada relação entre os dados, podem fornecer medidas gerais da amostra, como variância geral, por exemplo.

Dadas as informações anteriores pode-se inferir que em aplicações específicas, é necessária a utilização de matrizes de correlação específicas, tanto para a geração de novos dados aleatórios quanto para a validação de técnicas novas e existentes. Na literatura, observa-se a existência de algumas técnicas de geração de matrizes de correlação com valores fixos, bem como matrizes geradas por meio de alguma matriz de correlação já pré-existente. Entretanto, a geração de matrizes a partir de uma estrutura que inexistente, ou que seja específica o suficiente para atender aos anseios dos pesquisadores, ainda não foi explorada em estudos metodológicos. Para motivar a necessidade de geração de matrizes específicas, serão apresentados dois exemplos que deixam clara a necessidade de tais entidades matemáticas do ponto de vista metodológico.

Exemplo 1: um pesquisador desenvolveu, com base em uma técnica que pressupõe independência dos dados, um método de classificação que, além de considerar o conjunto de dados separadamente, agrega a relação

existente entre as variáveis. Entretanto, para testar a validade de sua metodologia, o pesquisador precisa gerar dados com diversas configurações de correlação para verificar a efetividade de sua metodologia frente aos dados correlacionados em diversas intensidades. O pesquisador pretende gerar dados com correlações baixas, dados com correlação de intensidade moderada, dados com alta correlação e por fim testar configurações nas quais estas três intensidades se alternam.

Exemplo 2: uma carteira de ativos econômicos pode ser criada ao considerar a rentabilidade média dos referidos papéis durante um determinado período de tempo. Entretanto, para a criação de uma carteira diversificada, um pesquisador pretende compreender como a rentabilidade de um ativo afeta a rentabilidade dos demais. Para tal, é necessário produzir diversas séries econométricas simuladas, sob algumas configurações de correlação, para então transpor o modelo teórico para aplicação no mercado financeiro.

Exemplo 3: o estudo de [Silva et al. \(2007\)](#) indica a existência de correlação significativa entre caracteres de aparência e rendimento na batata. Um pesquisador deseja gerar novas observações com base no estudo citado por meio de variações nas correlações entre os referidos caracteres para estudar as possíveis alterações na aparência e rendimento causadas por estas mudanças na intensidade de correlação entre os caracteres. Para tal é necessário produzir matrizes de correlação que sejam diferentes das matrizes observadas, porém com controle das intensidades por parte do pesquisador.

Nos três exemplos mencionados, é latente a necessidade de geração de dados correlacionados. Entretanto, as metodologias atualmente disponíveis necessitam de matrizes pré-existentes ou da definição de alguns valores base, conforme será discutido no capítulo 2. A proposta deste trabalho é, por meio de variáveis aleatórias específicas, eliminar a necessidade de especificação de entidades pré-existentes na geração de matrizes de correlação. E também, conceder um refinado controle ao usuário das correlações que serão geradas. Esta proposta inovadora certamente facilitará sobremaneira o desenvolvimento de estudos teóricos em Estatística multivariada, que poderão subsidiar o desenvolvimento de metodologias aplicáveis em dados reais e simulados.

1.1 Organização da Tese

Este tese encontra-se organizada a partir da seguinte estrutura:

- Capítulo 1: seção introdutória que delimita o problema de interesse e motiva o estudo;
- Capítulo 2: é apresentada uma revisão do estado da arte em relação ao problema de geração de matrizes de correlação;
- Capítulo 3: a proposição da técnica de geração de matrizes de correlação é completamente detalhada nesta seção, bem como os fundamentos matemáticos que possibilitam a execução da técnica e os

algoritmos desenvolvidos para executar os cálculos matemáticos e as simulações de dados. Também é apresentada uma discussão sobre os limites de geração de correlação das matrizes de correlação;

- Capítulo 4: um estudo de simulação numérica discute os aspectos de utilização da técnica, bem como os resultados alcançados quanto ao desempenho do método proposto. Além disso uma aplicação da metodologia proposta é apresentada;
- Capítulo 5: um pacote desenvolvido para o *software* R é apresentado e alguns exemplos são discutidos;
- Capítulo 6: são apresentadas conclusões sobre a metodologia proposta quanto a sua eficiência e também possíveis limitações. Além disso, discutem-se propostas de continuidade de pesquisa e demais aspectos relevantes.

Referências Bibliográficas

Silva, G. O., Pereira, A. S., Souza, V. Q., Carvalho, F. I. F., and Fritsche Neto, R. (2007). Correlações entre caracteres de aparência e rendimento e análise de trilha para aparência de batata. *Bragantia*, 66(3):381–388.

Revisão Bibliográfica

O problema da geração de matrizes de correlação através de algum mecanismo de simulação é por si só, um problema de destaque na investigação científica. Diversos estudos foram realizados nas últimas décadas com o objetivo de fornecer um mecanismo de geração que garanta os requisitos matemáticos, ou seja, que garantam que o elemento gerado seja uma matriz de correlação válida.

Entre meados da década de 70 e início da década de 80 foram realizados alguns estudos relacionados a esta temática. O estudo de [Chalmers \(1975\)](#) utiliza autovalores pré-estabelecidos na geração de matrizes de correlação aleatória, enquanto em [Bendel e Mickey \(1978\)](#), uma metodologia também baseada em autovalores é utilizada, com ênfase na amostragem de matrizes de autocorrelação no espaço de todas as matrizes que possuem os autovalores pré-estabelecidos. Já no trabalho de [Johnson e Welch \(1980\)](#), são geradas matrizes de correlação com base na distribuição Beta.

Após estes estudos iniciais, em 1984, o estudo de [Marsaglia e Olkin \(1984\)](#) discutiu com maior profundidade a geração de matrizes de correlação aleatória. Neste trabalho seminal, foram delimitadas duas vertentes básicas de pesquisa, às quais a maioria dos estudos posteriores seguiram por meio de adaptações e inovações. A primeira delas, partir de algum conjunto de dados prévio e tomar sua própria matriz de correlação ou utilizar uma matriz *template* como base. Posteriormente, a segunda abordagem consiste em produzir algum tipo de perturbação que preserve a estrutura de matriz de correlação, do ponto de vista matemático, mas que seja capaz de produzir uma nova matriz de correlação para futuras investigações.

Dentre os estudos que se enquadram na primeira vertente podemos destacar os estudos de [Ghosh e Henderson \(2003\)](#), [Joe \(2006\)](#), [Lewandowski et al. \(2009\)](#), [Simonian \(2010\)](#) e [Hardin et al. \(2013\)](#). O trabalho de [Ghosh e Henderson \(2003\)](#) estuda o comportamento da metodologia NORTA (Normal To Anything) de geração de variáveis aleatórias e a possibilidade de utilização desta metodologia para amostragem de matrizes de correlação do espaço de matrizes aleatórias. O trabalho de [Joe \(2006\)](#) propõe um método de geração de matrizes aleatórias com base em correlações parciais e estuda condições para as quais a distribuição conjunta das correlações parciais seja proporcional a uma potência do determinante da matriz de correlação.

Lewandowski et al. (2009) discutem e ampliam dois métodos de geração de matrizes aleatórias, o primeiro baseado no método discutido por Ghosh e Henderson (2003), enquanto a segunda metodologia expande o estudo apresentado por Joe (2006), além de promover um extenso estudo de simulação para verificar a validade dos métodos propostos.

O estudo de Simonian (2010) parte de matrizes de correlação inválidas, e por meio de procedimentos de álgebra linear e de aleatorização, corrige os autovalores negativos para zero e, deste modo, obtém matrizes de correlação válidas. O interesse principal deste estudo é a geração de matrizes para aplicações financeiras,

Hardin et al. (2013) apresentam um estudo que delimita a importância da simulação de matrizes de correlação em muitas áreas da estatística. Além disso, abordam-se as possíveis desvantagens da geração de dados gaussianos a partir de um modelo original. No procedimento com dados gaussianos, considerado inadequado por este trabalho, uma matriz de correlação de dados reais é adotada como *template*, geram-se amostras simuladas seguindo distribuição gaussiana multivariada com matriz de variância de acordo com a matriz de correlação dos dados reais. Posteriormente, encontram-se as matrizes de correlação das amostras simuladas. Para muitas repetições do procedimento amostral simulado, a natureza da variabilidade das correlações resultantes é semelhante entre diferentes tamanhos de amostra quando comparadas à matriz utilizada como *template*, o que não é desejável.

Uma segunda abordagem consiste em metodologias semelhantes aos estudos de Chalmers (1975) e de Bendel e Mickey (1978). Tomar como base autovalores pré-estabelecidos, sempre positivos, para a geração de matrizes de correlação. Esta metodologia é bastante eficaz uma vez que, parte de autovalores positivos e obtém diretamente o requisito mais complexo, a positividade semi-definida da matriz.

A segunda vertente é baseada na geração de matrizes por meio de autovalores não negativos. A partir desta premissa garante-se o requisito mais complexo da geração de matrizes de correlação, ser positiva semi-definida. Dentre os estudos com esta abordagem, destacam-se os trabalhos de Davies e Higham (2000) e Hüttner e Mai (2019). O primeiro revisita o algoritmo discutido em Bendel e Mickey (1978), é aprimorada a formulação inicial apresentada, o que culmina em uma redução significativa de instabilidade do algoritmo inicial.

Já no segundo trabalho os autores apresentam uma metodologia para geração de matrizes de correlações que atendam requisitos necessários às aplicações financeiras, tais como o primeiro autovalor significativamente maior que os demais, que geralmente explica sozinho 30% da variância dos dados e a propriedade de Perron-Frobenius, além de promover uma profunda revisão da metodologia de geração de matrizes aleatórias.

Além destas duas vertentes principais, outras abordagens foram testadas. Por exemplo, a geração de matrizes pelo método da rejeição. Entretanto, o estudo de Böhm e Hornik (2014) demonstrou matematicamente a inviabilidade de geração desta classe de matrizes pelo referido método para dimensões superiores a 6, o que torna tal metodologia inadequada

para problemas de ordem superior. Já em [Budden et al. \(2007\)](#) foi proposta uma metodologia para geração de matrizes de correlação de dimensão 4, por meio de operações com determinantes de matrizes 3×3 válidas. Já o trabalho de [Lewandowski et al. \(2009\)](#) generaliza o trabalho de [Budden et al. \(2007\)](#) para matrizes de dimensões quaisquer.

Uma solução mais sofisticada para esse problema é a geração de matrizes de correlação que não levam em consideração algum conjunto prévio de dados. Uma geração através de algum mecanismo independente de dados. Muitos estudos nessa direção estão diretamente ligados com algum problema específico de aplicação.

O estudo de [Odell e Feiveson \(1966\)](#) tem grande relevância na proposição de metodologia para construção de matrizes que satisfaçam a distribuição de *Wishart* que é a distribuição da matriz de covariância de amostras de vetores aleatórios normais multivariados independentes.

A motivação de estudos nessa direção decorre da grande aplicabilidade em diversos assuntos. Em muitas situações as matrizes de correlação são indispensáveis nos procedimentos de simulação. Encontram-se exemplos em diversos estudos: performance para métodos de estimação em dois estágios para modelos probit multivariados [Joe \(2005\)](#); estudos de amostragem [Bendel e Mickey \(1978\)](#); modelos de aplicações financeiras [Simonian \(2010\)](#); [Hüttner e Mai \(2019\)](#); aplicações na geração de *clusters* aleatórios [Qiu e Joe \(2006\)](#); processamento de sinais [Holmes \(1991\)](#), entre outros.

O problema de geração de matrizes aleatórias, em particular, que satisfaçam as condições de matrizes de correlação, se torna difícil no momento de garantir as características matemáticas que definem uma matriz de correlação. Os pressupostos de simetria, diagonal unitária e valores no intervalo $[-1, 1]$ são relativamente simples de satisfazer [Marsaglia e Olkin \(1984\)](#). Já garantir a concepção de matrizes positivas semi-definidas, ou seja, com autovalores não negativos, é um obstáculo considerável, sobretudo pela complexidade envolvida nas operações matriciais. Uma das lacunas alvo desse estudo é garantir que essa exigência seja atendida por meio de um mecanismo suficientemente simples.

Ao considerar os principais estudos apresentados nesta revisão, entende-se que a metodologia proposta neste trabalho é inovadora. Isso no sentido de que não há dentre os principais trabalhos dessa área, um estudo que apresente o mínimo paralelismo com a proposta que será apresentada na sequência, denominada *Custom Matrix generator* (CM-generator).

O principal resultado que será fornecido pela metodologia CM-generator é garantir a obtenção de matrizes de correlação de qualquer ordem, com ou sem especificações de forma. Isso sem a necessidade de qualquer matriz de correlação original como *template*. Também não é necessário o pré-estabelecimento de autovalores associados às matrizes de correlação para execução do procedimento. Visto de outra forma, não se faz necessário qualquer conjunto de dados prévio para iniciar o procedimento de geração de matrizes de correlação. Além disso, o CM-generator é capaz de fornecer matrizes de correlação com estrutura pré-fixada pelo usuário do método. É possível estabelecer tanto um percentual de variáveis cuja correlação é alta, média, baixa; quanto limites específicos que delimitam o que

seria uma correlação alta, ou então baixa, por exemplo. Todas essas escolhas serão estabelecidas de acordo com as necessidades do problema sob investigação.

Referências Bibliográficas

- Bendel, R. B. e Mickey, M. R. (1978). Population correlation matrices for sampling experiments. Communications in Statistics - Simulation and Computation, 7(2):163–182.
- Budden, M., Hadavas, P., Hoffman, L. e Pretz, C. (2007). Generating valid 4×4 correlation matrices. Applied Mathematics E-Notes, 7:53–59.
- Böhm, W. e Hornik, K. (2014). Generating random correlation matrices by the simple rejection method: Why it does not work. Statistics & Probability Letters, 87:27–30.
- Chalmers, C. P. (1975). Generation of correlation matrices with a given eigen-structure. Journal of Statistical Computation and Simulation, 4(2):133–139.
- Davies, P. I. e Higham, N. J. (2000). Numerically stable generation of correlation matrices and their factors. BIT Numerical Mathematics, 40(4):640–651.
- Ghosh, S. e Henderson, S. G. (2003). Behavior of the nortia method for correlated random vector generation as the dimension increases. ACM Transactions on Modeling and Computer Simulation (TOMACS), 13(3):276–294.
- Hardin, J., Garcia, S. R. e Golan, D. (2013). A method for generating realistic correlation matrices. The Annals of Applied Statistics, 7(3):1733–1762.
- Holmes, R. B. (1991). On random correlation matrices. SIAM Journal on Matrix Analysis and Applications, 12(2):239–272.
- Hüttner, A. e Mai, J. (2019). Simulating realistic correlation matrices for financial applications: correlation matrices with the perron–frobenius property. Journal of Statistical Computation and Simulation, 89(2):315–336.
- Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis, 94(2):401–419.
- Joe, H. (2006). Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis, 97(10):2177–2189.
- Johnson, D. G. e Welcht, W. J. (1980). The generation of pseudo-random correlation matrices. Journal of Statistical Computation and Simulation, 11(1):55–69.

- Lewandowski, D., Kurowicka, D. e Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis, 100:1989–2001.
- Marsaglia, G. e Olkin, I. (1984). Generating correlation matrices. SIAM Journal on Scientific and Statistical Computing, 5(2):470–475.
- Odell, P. L. e Feiveson, A. H. (1966). A numerical procedure to generate a sample covariance matrix. Journal of the American Statistical Association, 61(313):199–203.
- Qiu, H. e Joe, H. (2006). Separation index and partial membership for clustering. Computational Statistics & Data Analysis, 50:585–603.
- Simonian, J. (2010). The most simple methodology to create a valid correlation matrix for risk management and option pricing purposes. Applied Economics Letters, 17(18):1767–1768.

Material e Métodos

O método CM-generator não enfrenta dificuldades em contornar o problema mais significativo, garantir que a matriz gerada possua apenas autovalores não negativos. Trata-se de procedimento que garante esse requisito e todos os demais requisitos matemáticos através de sua própria estratégia de construção.

A metodologia proposta parte da geração de uma amostra multivariada simulada da distribuição Normal. Uma amostra que possua uma matriz de correlação amostral capaz de atender os requisitos previamente estabelecidos pelo usuário. As matrizes geradas são, na prática, matrizes de correlação amostral, ou seja, atendem em sua completude todos os requisitos matemáticos de matrizes de correlação. A matriz é obtida por meio de um vetor de variáveis aleatórias com medidas de variabilidade pré-fixadas. Além disso, permitem um nível de customização bastante refinado das matrizes geradas.

Além da possibilidade de geração de matrizes de correlação com controle da intensidade de correlações existentes entre cada uma de suas variáveis, um segundo objetivo também pode ser alcançado, uma customização específica na estrutura da matriz de correlação gerada. Customizações específicas atendem às mais diversas aplicações de interesse prático. A possibilidade de estabelecer níveis específicos de correlação (baixa, média, alta, entre outras) e os respectivos percentuais de correlações entre variáveis incluídas em cada um desses níveis, é uma abordagem inovadora e de grande utilidade no meio científico.

Suponha uma aplicação específica, na qual necessariamente devam ser gerados dados com medidas de correlação pertencentes estritamente ao intervalo $[-p, p]$, para algum p específico. Em algumas aplicações exige-se p próximo de 1, em outras, na vizinhança de 0. Um método eficaz seria capaz de definir, com precisão, a distribuição das correlações. Além das respectivas intensidades dentro de uma matriz de correlação de acordo com o cenário necessário para atender às necessidades do pesquisador. Por exemplo, gerar um banco de dados com 20% das correlações entre variáveis no intervalo $[-0,25; 0,25]$, 30% no intervalo $[-0,55; -0,25) \cup (0,25; 0,55]$ e 50% no intervalo $[-1; -0,55) \cup (0,55; 1]$.

O objetivo específico do método CM-generator é fornecer ao pesquisador o completo controle da matriz de correlação que será gerada. Dessa

forma, atender às peculiaridades inerentes às metodologias envolvidas no desenvolvimento de pesquisas que utilizarão a metodologia proposta. A construção da técnica será baseada em um conjunto de considerações matemáticas que garantirão a possibilidade de controle de intensidade no eventual cenário de geração de matrizes de correlação. Inicialmente essas considerações matemáticas serão abordadas.

3.1 Considerações Matemáticas

Sejam Z, X, Y variáveis aleatórias gaussianas independentes, tais que:

$$Z \sim N(0, 1), X \sim N(0, a^2) \text{ e } Y \sim N(0, b^2)$$

em que $a, b > 0$.

Sejam ainda, U e V as seguintes variáveis aleatórias:

$$U = X \pm Z \implies U \sim N(0, 1 + a^2)$$

$$V = Y \pm Z \implies V \sim N(0, 1 + b^2)$$

Lema 3.1.1 A correlação entre Z e U é dada por $\rho_{Z,U} = \pm \frac{1}{\sqrt{1 + a^2}}$.

Prova do Lema 3.1.1

$$\text{Cov}(Z, U) = \text{Cov}(Z, X \pm Z) = \text{Cov}(Z, X) \pm \text{Cov}(Z, Z)$$

por independência,

$$\text{Cov}(Z, X) \pm \text{Cov}(Z, Z) = 0 \pm \text{Var}(Z) = \pm 1$$

logo, $\rho_{Z,U}$ é dada por

$$\rho_{Z,U} = \frac{\text{Cov}(Z, U)}{\sqrt{\text{Var}(Z)}\sqrt{\text{Var}(U)}} = \pm \frac{1}{\sqrt{1 + a^2}}$$

□

analogamente, $\rho_{Z,V} = \pm \frac{1}{\sqrt{1 + b^2}}$.

Lema 3.1.2 A correlação entre U e V é dada por $\rho_{U,V} = \pm \frac{1}{\sqrt{1 + a^2}\sqrt{1 + b^2}}$

Prova do Lema 3.1.2

$$\begin{aligned} \text{Cov}(U, V) &= \text{Cov}(X \pm Z, Y \pm Z) \\ &= \text{Cov}(X, Y) \pm \text{Cov}(X, Z) \pm \text{Cov}(Z, Y) \pm \text{Cov}(Z, Z) \end{aligned}$$

como Z , X e Y são independentes entre si:

$$\text{Cov}(U, V) = 0 \pm 0 \pm 0 \pm \text{Cov}(Z, Z) = \pm \text{Var}(Z) = \pm 1$$

portanto

$$\rho_{U,V} = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)}\sqrt{\text{Var}(V)}} = \frac{\pm 1}{\sqrt{1+a^2}\sqrt{1+b^2}}$$

□

Determinados os valores das correlações $\rho_{Z,U}$, $\rho_{Z,V}$ e $\rho_{U,V}$, deseja-se obter números reais positivos a e b que garantam a definição dos limites ℓ_{lower} e ℓ_{upper} . Os valores ℓ_{lower} e ℓ_{upper} funcionam simultaneamente como cotas inferiores e superiores para as correlações $\rho_{Z,U}$, $\rho_{Z,V}$ e $\rho_{U,V}$. A existência dessas cotas satisfaz o seguinte conjunto de inequações:

$$\ell_{lower} < \frac{1}{\sqrt{1+a^2}} < \ell_{upper} \quad (3.1a)$$

$$\ell_{lower} < \frac{1}{\sqrt{1+b^2}} < \ell_{upper} \quad (3.1b)$$

$$\ell_{lower} < \frac{1}{\sqrt{1+a^2}\sqrt{1+b^2}} < \ell_{upper} \quad (3.1c)$$

em que $0 < \ell_{lower} < \ell_{upper} < 1$.

Teorema 3.1.1 *Existem valores de a e b e condições específicas tais que as inequações 3.1a, 3.1b e 3.1c são satisfeitas simultaneamente.*

Prova do Teorema 3.1.1

Seja $\ell_{upper}^* = \sqrt{\ell_{upper}}$, note que $\ell_{upper}^* > \ell_{upper}$ pois $0 < \ell_{upper} < 1$, e ainda, suponha:

$$\frac{1}{\sqrt{1+a^2}} < \ell_{upper} < \ell_{upper}^* \quad (3.2a)$$

$$\frac{1}{\sqrt{1+b^2}} < \ell_{upper} < \ell_{upper}^* \quad (3.2b)$$

Note que números reais a e b que satisfazem as inequações 3.2a e 3.2b garantem a validade da cota superior da inequação 3.1c, basta multiplicar as inequações 3.2a e 3.2b.

$$\frac{1}{\sqrt{1+a^2}} \frac{1}{\sqrt{1+b^2}} < \ell_{upper}^*{}^2 = \ell_{upper}$$

Em outras palavras, a e b que validam as inequações 3.2a e 3.2b validam também as cotas superiores das inequações 3.1a, 3.1b e 3.1c.

Defina agora $\ell_{lower}^* = \sqrt{\ell_{lower}}$, note que $\ell_{lower}^* > \ell_{lower}$ pois $0 < \ell_{lower} < 1$ e suponha obter valores reais para a e b tais que:

$$\ell_{lower} < \ell_{lower}^* < \frac{1}{\sqrt{1+a^2}} < \ell_{upper}^* \quad (3.3a)$$

$$\ell_{lower} < \ell_{lower}^* < \frac{1}{\sqrt{1+b^2}} < \ell_{upper}^* \quad (3.3b)$$

Note que a existência desses valores a e b satisfazendo as inequações 3.2a e 3.2b está condicionada a $\ell_{lower}^* < \ell_{upper}$. Se essa condição não for atendida para algum caso particular, os intervalos limites deverão ser reajustados, sendo essa a única condição específica para garantir a validade das inequações 3.1a, 3.1b e 3.1c. Agora multiplicando as inequações 3.3a e 3.3b:

$$\ell_{lower} = \ell_{lower}^*{}^2 < \frac{1}{\sqrt{1+a^2}} \frac{1}{\sqrt{1+b^2}} < \ell_{upper}^*{}^2 = \ell_{upper} .$$

Logo, qualquer escolha de valores a e b satisfazendo as inequações 3.3a e 3.3b satisfaz simultaneamente as inequações 3.1a, 3.1b e 3.1c.

$$\begin{aligned} \ell_{lower}^* < \frac{1}{\sqrt{1+a^2}} < \ell_{upper}^* &\implies \frac{1}{\ell_{upper}^*} < \sqrt{1+a^2} < \frac{1}{\ell_{lower}^*} \implies \\ \implies \frac{1}{\ell_{upper}} < 1+a^2 < \frac{1}{\ell_{lower}} &\implies \frac{1}{\ell_{upper}} - 1 < a^2 < \frac{1}{\ell_{lower}} - 1 \implies \\ \implies \sqrt{\frac{1}{\ell_{upper}} - 1} < \mathbf{a} < \sqrt{\frac{1}{\ell_{lower}} - 1} . \end{aligned}$$

De forma análoga,

$$\sqrt{\frac{1}{\ell_{upper}} - 1} < \mathbf{b} < \sqrt{\frac{1}{\ell_{lower}} - 1}$$

logo, os valores reais de a e b que garantem as cotas inferior e superior de cada uma das inequações 3.1a, 3.1b e 3.1c devem ser escolhidos simultaneamente ao intervalo de medida não nula:

$$I = \left(\sqrt{\frac{1}{\ell_{upper}} - 1}, \sqrt{\frac{1}{\ell_{lower}} - 1} \right) \quad (3.4)$$

em que $0 < \ell_{lower} < \ell_{upper} < 1$. □

De posse das condições matemáticas adequadas para escolha dos valores de a e b um algoritmo com os passos de execução para a geração das

matrizes de correlação pode ser construído.

3.1.1 Limites Superiores e Inferiores

Conforme discutido anteriormente, o método de geração é baseado na definição de limites inferiores e superiores para as medidas de correlação. Mesmo nos casos em que não se definem os referidos limites, como por exemplo em uma geração completamente aleatorizada, existe a necessidade de definir os valores de desvios padrão a e b para a obtenção de correlações pertencentes ao intervalo $(\ell_{lower}, \ell_{upper})$, tais que $\ell_{lower} = 0$ e $\ell_{upper} = 1$.

Para $\ell_{lower}, \ell_{upper} \in (0, 1)$, não encontra-se limitações na obtenção de a e b . Entretanto, ao se buscar $\ell_{lower}, \ell_{upper} \in [0, 1]$, já ocorrem restrições matemáticas, uma vez que

$$\lim_{\ell_{lower} \rightarrow 0} \sqrt{\frac{1}{\ell_{lower}} - 1} \rightarrow \infty \quad (3.5)$$

Deste modo, não se pode definir $\ell_{lower} = 0$, mas deve-se buscar por valores próximos a zero. Por outro lado, também existem restrições ao se definir $\ell_{upper} = 1$, pois

$$\lim_{\ell_{upper} \rightarrow 1} \sqrt{\frac{1}{\ell_{upper}} - 1} \rightarrow 0 \quad (3.6)$$

e neste caso, seria obtido $a = 0$. Porém, como a é um desvio padrão, por definição, geraria-se uma variável aleatória degenerada por ausência de variabilidade e, por este motivo, deve-se buscar por valores de ℓ_{upper} suficientemente próximos a 1, mas não iguais a 1.

Conforme supracitado, as escolhas de ℓ_{lower} e ℓ_{upper} próximas às bordas do intervalo unitário $(0, 1)$ devem ser realizadas de forma a se aproximar dos limites do intervalo, mas jamais assumindo os valores limitantes propriamente ditos. Neste caso, busca-se por valores suficientemente próximos a 0 e 1. Porém, o quão próximos devem ser os limites? Existirão diferenças entre as correlações geradas conforme se aproxima mais ou menos das bordas do intervalo?

Como a geração de correlações é baseada em desvios padrão gerados de sorteios aleatórios de variáveis aleatórias uniformes pertencentes ao intervalo definido por ℓ_{lower} e ℓ_{upper} e o intervalo descrito em 3.4, as alterações nos limites inferior e superior influenciam sensivelmente na geração de correlações.

Suponha definir $\ell_{lower} = 0,01$ e $\ell_{upper} = 0,99$. Nesta configuração, o seguinte cenário ocorreria:

$$I = \left(\sqrt{\frac{1}{0,99} - 1}, \sqrt{\frac{1}{0,01} - 1} \right) = (0,1005038; 9,949874),$$

ou seja, a e b seriam pertencentes ao intervalo $(0,1005038; 9,949874)$. Para

avaliar o impacto destas medidas, será analisado o comportamento das seguintes probabilidades: $P(\rho < 0,1)$ e $P(\rho > 0,9)$. Ao considerar a estrutura proposta pela metodologia apresentada, os desvios padrão que definem as seguintes correlações são $a = 0,333333$ e $b = 3$, ou seja,

$$\begin{aligned} P(\rho < 0,1) &= P(u > 3) \\ P(\rho > 0,9) &= P(u < 0,333333) \end{aligned}$$

em que $u \sim U(0,1005038, 9,949874)$.

Logo,

$$P(\rho < 0,1) = P(u > 3) = \frac{9,949874 - 3}{9,949874 - 0,1005038} = 0,7056161$$

$$P(\rho > 0,9) = P(u < 0,333333) = \frac{0,333333 - 0,1005038}{9,949874 - 0,1005038} = 0,02363902$$

Suponha agora a escolha de limites ℓ_{lower} e ℓ_{upper} mais próximos a 0 e 1, respectivamente, digamos $\ell_{lower} = 0,001$ e $\ell_{upper} = 0,999$. Para tais limites, obter-se-ia o seguinte intervalo:

$$I = \left(\sqrt{\frac{1}{0,999} - 1}, \sqrt{\frac{1}{0,001} - 1} \right) = (0,0316386; 31,60696).$$

Definindo os valores a e b neste novo intervalo, seriam obtidas as seguintes probabilidades:

$$P(\rho < 0,1) = P(u > 3) = \frac{31,60696 - 3}{31,60696 - 0,0316386} = 0,9059911$$

$$P(\rho > 0,9) = P(u < 0,333333) = \frac{0,333333 - 0,0316386}{31,60696 - 0,0316386} = 0,009554753$$

Qualquer alteração nos limites altera as bordas do intervalo. Observa-se que, um acréscimo de 0,009 em ℓ_{upper} impactou em um aumento de aproximadamente 20% da probabilidade de serem gerados valores inferiores a 0,1. Enquanto uma redução de 0,009 impactou em uma queda na probabilidade de geração de correlações superiores a 0,9 de aproximadamente 1,3%. Tais cálculos indicam dois fatos específicos:

- (i) a escolha de limites superiores ℓ_{upper} próximos a 1, que levam a limites inferiores a dos desvios padrão gerados a se aproximarem de zero, o que pode acarretar uma alta concentração de dados gerados na vizinhança de zero;
- (ii) a escolha de limites inferiores ℓ_{lower} próximos a 0 elevam o limite superior b dos desvios padrão, porém isso não afeta de forma tão significativa a dispersão das correlações geradas.

Durante a implementação computacional da metodologia, controlar o parâmetro de limite inferior do intervalo do qual serão gerados os desvios padrão influenciarão sobremaneira a dispersão da distribuição. No capítulo 4 serão apresentados os resultados simulados da metodologia proposta e um experimento específico explorará a variação dos limites inferiores do intervalo I no formato da distribuição das correlações geradas.

3.2 Geração das Matrizes de Correlação

Inicialmente, são definidas as variáveis aleatórias para compor o vetor aleatório que será utilizado na geração da matriz de correlação. Posteriormente, são definidos os limites capazes de controlar a intensidade das correlações. A metodologia proposta aqui, parte do pressuposto que a tarefa de gerar observações de modelos probabilísticos específicos e conhecidos, como por exemplo as distribuições gaussianas e uniformes pode ser executada sem transtornos, logo tal tarefa não será tratada como objeto de investigação nesse estudo. Em particular, será admitido que qualquer método eficaz de geração dessas variáveis pode ser utilizado para este estudo.

3.2.1 Algoritmo Geral

Considere uma amostra de n observações da distribuição normal padrão $Z \sim N(0, 1)$ representado pelo vetor coluna \tilde{Z} . Seja \mathbf{X} uma matriz composta por k vetores coluna com n observações provenientes das variáveis aleatórias X_1, X_2, \dots, X_k tais que $X_i \sim N(0, a_i^2)$, e \mathbf{Z} uma matriz com k colunas, em que todas as colunas são iguais a \tilde{Z} . Defina \mathbf{U} tal que:

$$\mathbf{U} = \mathbf{X} \pm \mathbf{Z} = [U_1 \ U_2 \ \dots \ U_i \ \dots \ U_k]. \quad (3.7)$$

Note que U_i é um vetor com n observações da variável aleatória gaussiana de média 0 e variância $1 + a_i^2$. A correlação entre U_i e U_j , é dada por:

$$\rho_{U_i, U_j} = \pm \frac{1}{\sqrt{1 + a_i^2} \sqrt{1 + a_j^2}}.$$

A amostra de n observações de $Z \sim N(0, 1)$ e também as observações das variáveis X_1, X_2, \dots, X_k com $X_i \sim N(0, a_i^2)$ são obtidas por meio de mecanismos de simulação computacional. Posteriormente, as amostras geradas são somadas de acordo com a equação 3.7 e é obtida a matriz de correlação desse conjunto de dados. O procedimento assegura que todos os pressupostos matemáticos para garantir que a matriz gerada é uma matriz de correlação estejam completamente atendidos, pois trata-se de uma matriz de correlação amostral. Além disso, o conjunto de restrições previamente imposto na escolha das medidas de variabilidade das variáveis garante medidas de correlação dentro dos intervalos pré-estabelecidos.

A execução do procedimento parte da escolha da dimensão da matriz de correlação desejada pelo usuário e da especificação dos limites inferiores

e superiores. Baseado nessa escolha, os desvios padrão que serão utilizados na geração das variáveis aleatórias $X_i \sim N(0, a_i^2)$ são estabelecidos. A matriz \mathbf{U} é obtida como descrito anteriormente e fornece a matriz de correlação de interesse. A figura 3.2.1 apresenta o pseudo-código do algoritmo geral.

```

algoritmo
  leia  $dim, nrep$ 
  se  $dim = null$ 
     $dim \leftarrow 10$ 
  fim se
  se  $nrep = null$ 
     $nrep \leftarrow 1000$ 
  fim se
   $a \leftarrow 0,01$ 
   $b \leftarrow 99,99$ 
  gere  $z \sim N(0, 1)$  de tamanho  $nrep$ 
  faça  $\mathbf{Z}_{nrep \times dim} = (z_1, z_2, \dots, z_{dim}), z_i = z$ 
  para  $i = 1$  até  $dim$  faça
    gere  $sd_i \sim U(a, b)$ 
    gere  $x_i \sim N(0, sd_i)$ 
    gere  $sig_i$  de  $C = \{-1, 1\}$ 
  fim para
   $\mathbf{U} \leftarrow (sig_1 \cdot (x_1 + z_1), sig_2 \cdot (x_2 + z_2), \dots, sig_{dim} \cdot (x_{dim} + z_{dim}))$ 
   $gencor \leftarrow Correlation(\mathbf{U})$ 
  imprima  $gencor$ 
fim algoritmo

```

Figura 3.2.1: Algoritmo geral para geração de matrizes de correlação

3.2.2 Algoritmo para Matrizes de Correlação Limitadas

Uma extensão óbvia da tarefa desempenhada pelo algoritmo descrito anteriormente é delimitar os valores de correlação em faixas escolhidas pelo usuário. Suponha simular uma matriz de correlação $k \times k$ cujos valores absolutos estejam sempre entre limites específicos. O intervalo definido na equação 3.4 fornece a escolha adequada dos valores de desvio padrão. Por exemplo, para obter uma matriz de correlações cujas correlações tenham valores absolutos entre 0,35 e 0,75, nesse caso, os desvios padrão devem ser escolhidos no intervalo $I = (0,57735; 1,36277)$. O conjunto de escolhas dos desvios padrão no intervalo anterior é suficiente para garantir que a matriz gerada apresentará correlações dentro dos intervalos pré-especificados. O pseudo-código do algoritmo para correlações limitadas é apresentado de forma resumida na figura 3.2.2.

```

algoritmo
  leia  $dim$ ,  $nrep$ ,  $\ell_{lower}$ ,  $\ell_{upper}$ ,  $method$ 
  se  $dim = null$  então  $dim \leftarrow 10$ 
  fim se
  se  $nrep = null$  então  $nrep \leftarrow 1000$ 
  fim se
  se  $\ell_{lower} = null$  então  $\ell_{lower} \leftarrow 0,3$ 
  fim se
  se  $\ell_{upper} = null$  então  $\ell_{upper} \leftarrow 0,6$ 
  fim se
  se  $method = "low"$ 
     $a \leftarrow 0,01$ 
     $b \leftarrow [(1/\ell_{lower}) - 1]^{0,5}$ 
  fim se
  se  $method = "medium"$ 
     $a \leftarrow [(1/\ell_{upper}) - 1]^{0,5}$ 
     $b \leftarrow [(1/\ell_{lower}) - 1]^{0,5}$ 
  fim se
  se  $method = "high"$ 
     $a \leftarrow [(1/\ell_{upper}) - 1]^{0,5}$ 
     $b \leftarrow 99,99$ 
  fim se
  gere  $z \sim N(0, 1)$  de tamanho  $nrep$ 
  faça  $Z_{nrep \times dim} = (z_1, z_2, \dots, z_{dim})$ ,  $z_i = z$ 
  para  $i = 1$  até  $dim$  faça
    gere  $sd_i \sim U(a, b)$ 
    gere  $x_i \sim N(0, sd_i)$ 
    gere  $sig_i$  de  $C = \{-1, 1\}$ 
  fim para
   $U \leftarrow (sig_1 \cdot (x_1 + z_1), sig_2 \cdot (x_2 + z_2), \dots, sig_{dim} \cdot (x_{dim} + z_{dim}))$ 
  gencor  $\leftarrow Correlation(U)$ 
  imprima gencor
fim algoritmo

```

Figura 3.2.2: Algoritmo para geração de matrizes de correlação com limites pré-definidos

3.2.3 Algoritmo para Matrizes de Correlação Customizadas

Além da delimitação dos valores de correlação em níveis escolhidos pelo usuário, um controle mais refinado pode ser proposto. Suponha a possibilidade de gerar matrizes customizadas em três aspectos:

- (i) definir mais que um nível de valores de correlação;
- (ii) estabelecer limites específicos para cada um desses níveis;
- (iii) determinar a proporção de correlações que serão geradas em cada um dos níveis especificados.

Admita, por exemplo, estabelecer medidas de correlação tais que $0 < |\rho| \leq 0,3$ sejam ditas correlações baixas. Medidas de correlação tais que $0,3 < |\rho| \leq 0,6$ sejam ditas correlações médias. E ainda, medidas de correlação tais que $|\rho| > 0,6$ sejam ditas correlações altas. Além disso, suponha obter uma matriz de correlações com 40% de correlações baixas, 30% de correlações médias e 30% de correlações altas, obviamente com a diagonal principal excluída. Este seria um problema de obtenção de uma matriz de correlação que atende condições bastante específicas, ou seja, uma matriz de correlação customizada.

Um raciocínio simplista seria definir os intervalos em termos dos limites inferior e superior com a proporcionalidade desejada preservada. Em outras palavras, limites tais que:

- (i) para 40% das correlações os valores sejam ditos baixos;
- (ii) para 30% das correlações os valores sejam ditos médios;
- (iii) para 30% das correlações os valores sejam ditos altos.

Nessa proposta, a escolha seria a proposição de 40% dos desvios padrão obtidos tais que $\ell_{lower} = 0$ e $\ell_{upper} = 0,3$, a proposição de 30% dos desvios padrão obtidos tais que $\ell_{lower} = 0,3$ e $\ell_{upper} = 0,6$ e por fim, 30% dos desvios padrão obtidos para $\ell_{lower} = 0,6$ e $\ell_{upper} = 1,0$.

Apesar de parecer bastante natural, tal proposta não garante o sucesso na escolha pré-fixada de percentual entre as faixas. Mais que isso, certamente essa escolha não levará a matriz desejada. Não é possível especificar, de maneira exata, a intensidade de correlação existente entre dois vetores colunas pertencentes à faixas de intensidade distintas. De acordo com a metodologia proposta, dois vetores em um mesmo intervalo de intensidade $[\ell_1, \ell_2]$ sempre possuirão correlação pertencente a este intervalo. Por outro lado, o efeito de correlação de um vetor pertencente ao intervalo $[\ell_1, \ell_2]$ com outro vetor pertencente a um intervalo distinto $[\ell_3, \ell_4]$, pode pertencer a um dos intervalos ou até mesmo a nenhum dos dois.

Para esse propósito específico, suponha gerar uma matriz seguindo a especificação descrita anteriormente. Provavelmente as proporções resultantes em cada faixa de correlação não serão exatamente como pré-fixadas. Por exemplo, desejava-se $\gamma\%$ em uma determinada faixa e o resultado obtido foi de $(\gamma + \varepsilon)\%$ para a determinada faixa. Nessa situação, talvez fosse necessário escolher previamente no procedimento algum percentual inferior a $\gamma\%$ para resultar em uma matriz com proporção $\gamma\%$ para a determinada faixa. Entretanto, determinar quanto inferior deve ser o valor para γ não é uma tarefa trivial.

No método aqui proposto para geração de matrizes de correlação, qualquer procedimento ingênuo de otimização pode ser utilizado para vasculhar as mudanças necessárias nos percentuais prévios para garantir que a matriz gerada convirja na direção da matriz com a especificação desejada pelo usuário do método. Parece claro que esta especificidade do método exige algum esforço computacional maior que o caso geral. Entretanto, não se trata de algo proibitivo visando obter uma matriz de correlação para algum objetivo extremamente específico.

Na figura 3.2.3 é apresentado o pseudo-código para implementação do método customizado para geração de matrizes aleatórias de correlação.

```

algoritmo
  leia dim, nrep, nsim
  leia custom_limits, custom_prop
  method  $\leftarrow$  custom
  se dim = null então dim  $\leftarrow$  10
  fim se
  se nrep = null então nrep  $\leftarrow$  1000
  fim se
  se nsim = null então nsim  $\leftarrow$  1000
  fim se
  se custom_limits = null
    custom_limits  $\leftarrow$  (llower, lupper)
  fim se
  se custom_prop = null
    k  $\leftarrow$  #( custom_limits )
    /*k é o número de níveis de correlação -1 */
    pi  $\leftarrow$  1/(k + 1)
    custom_prop  $\leftarrow$  (p1, p2, \dots, pk+1)
  fim se
  opt_precision  $\leftarrow$  0
  cont  $\leftarrow$  0
  gere z  $\sim$   $N(0, 1)$  de tamanho nrep
  faça  $\mathbf{Z}_{nrep \times dim} = (z_1, z_2, \dots, z_{dim})$ , zi = z
  enquanto (opt_precision > precision) ou (cont = nsim) faça
  para i = 1 até dim faça
    gere sdi  $\sim$   $U(a, b)$ 
    gere xi  $\sim$   $N(0, sd_i)$ 
    gere sigi de  $C = \{-1, 1\}$ 
  fim para
   $\mathbf{U} \leftarrow (sig_1 \cdot (x_1 + z_1), sig_2 \cdot (x_2 + z_2), \dots, sig_{dim} \cdot (x_{dim} + z_{dim}))$ 
  calcule (α1, α2, \dots, αk+1)
  /*αi é a proporção de correlações no i-nível para  $\mathbf{U}^*$ */
  /*rotina de otimização*/
    leia  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{k+1})$ 
    retorne new_α = (α1, α2, \dots, αk+1)
  /*fim da rotina de otimização*/
  calcule opt_precision
  cont  $\leftarrow$  cont + 1
  fim enquanto
  gencor  $\leftarrow$  Correlation( $\mathbf{U}$ )
  imprima gencor
fim algoritmo

```

Figura 3.2.3: Algoritmo para geração de matrizes de correlação com limites pré-definidos

A proposta utilizada nesse estudo se mostrou bastante eficiente, trata-se de uma estratégia simples e com conceitos similares ao algoritmo *Simulated Annealing*, uma descrição básica das premissas do algoritmo *Simulated Annealing* pode ser obtida em Cruz et al. (2018). Para definir o procedimento, seja $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)$ um vetor com as proporções desejadas para as correlações em cada uma das faixas de interesse. Seja ainda $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ um vetor com as proporções obtidas na matriz gerada para as correlações em cada uma das faixas de interesse, os passos de execução da estratégia de otimização são:

1. defina $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)$;
2. gere uma matriz P_1 com desvios padrão escolhidos por faixas seguindo as proporções de Γ ;
3. para P_1 obtenha $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$;
4. obtenha i^* tal que $i^* = \arg \max(\gamma_i - \theta_i)$, j^* tal que $j^* = \arg \min(\gamma_j - \theta_j)$ e defina $w_1 = \max(|\gamma_i - \theta_i|)$;
5. faça $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_{i^*} - w_1, \dots, \gamma_{j^*} + w_1, \dots, \gamma_k)$;
6. gere uma P_2 com desvios padrão escolhidos por faixas seguindo as novas proporções de Γ ;
7. para P_2 obtenha $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ e defina $w_2 = \max(|\gamma_i - \theta_i|)$;
8. verifique se $w_1 > w_2$, se sim faça $P_1 = P_2$ e retorne ao passo (3);
9. se $w_1 \leq w_2$, seja $u \sim U(0, 1)$;
10. se $u > \frac{n}{N}$ faça $P_1 = P_2$ e retorne ao passo (3);
11. se $u \leq \frac{n}{N}$ retorne ao passo (2).

Na sequência de passos anteriores, o inteiro n representa quantas vezes o procedimento foi executado. Esse procedimento deverá ser repetido enquanto $n \leq N$ ou $|w_1 - w_2| > K$ para N e K pré-fixados pelo usuário. As escolhas desses critérios de parada são de grande importância. Vale ressaltar que para uma matriz de correlação $k \times k$, existem $k(k-1)/2$ medidas de correlação, e não é possível garantir que sempre existirá uma partição inteira que atenda exatamente as proporções originalmente indicadas. Por outro lado, o método é capaz de produzir a matriz especificada na maioria das situações. Quando o propósito não é alcançado, o método fornece ao usuário uma matriz que esteja suficientemente próxima do desejado, isso de acordo com os critérios de parada N e K previamente estabelecidos. O pseudo-código do mecanismo de otimização pode ser visualizado na figura 3.2.3

```

algoritmo
  leia custom_prop, custom_rep,  $\alpha$ , Z
  enquanto nrep < custom_nrep faça
    temp_gencor  $\leftarrow$  Correlation(Z)
    calcule new_α
    /*new_αi é a proporção de correlações no i-nível*/
    /*new_αi se refere a matriz temp_gencor*/
    dist  $\leftarrow$  |custom_prop - new_α|
    maxdist  $\leftarrow$  argmax(dist)
    kmin  $\leftarrow$  posição argmin(dist) no vetor dist
    kmax  $\leftarrow$  posição argmax(dist) no vetor dist
    gere  $u \sim U(0,1)$ 
    se maxdist > custom_precision e  $u > \frac{nrep}{custom_nrep}$ 
      custom_prop[kmax]  $\leftarrow$  custom_prop[kmax] - maxdist
      custom_prop[kmin]  $\leftarrow$  custom_prop[kmin] + maxdist
    fim se
    se maxdist > custom_precision e  $u < \frac{nrep}{custom_nrep}$ 
      nrep  $\leftarrow$  nrep + 1
    fim se
    se maxdist < custom_precision
      nrep  $\leftarrow$  custom_nrep
    fim se
  fim enquanto
  retorne custom_prop
fim algoritmo

```

Figura 3.2.4: Algoritmo para procedimento de otimização

Referências Bibliográficas

Cruz, F. R. B., Duarte, A. R. e Souza, G. L. (2018). Multi-objective performance improvements of general finite single-server queueing networks. *Journal of Heuristics*, 24(5):757–781.

Resultados e Discussão

Todos os códigos utilizados neste estudo, para gerar as matrizes descritas foram implementados através do software estatístico R, eles estão disponíveis, mediante solicitação aos autores, e devem ser usados apenas para fins educacionais e de pesquisa. As execuções foram conduzidas em um equipamento Intel(R) Core(TM) i5-5200U CPU @2.20GHz, 8 GB de memória RAM, com o sistema operacional Windows 10 e a versão 3.5.0 do software estatístico R.

Diversos experimentos foram executados com interesse em validar a metodologia proposta. Os testes executados visam verificar a estrutura de distribuição das correlações geradas e avaliar os tempos computacionais para execução do procedimento. Testes comparativos com respeito ao método CM-generator e também metodologias bastante difundidas como o método aqui nominado Onion e o método denominado C-vine. Estes métodos são bem detalhados por [Lewandowski et al. \(2009\)](#) e foram adotados por já possuírem implementações em linguagem R disponíveis.

O método Onion é uma estratégia iterativa que parte de uma matriz unidimensional e amplia a dimensão matricial sucessivamente através do acréscimo de linhas extras até a dimensão desejada. Este acréscimo ocorre por meio da geração de novas observações condicionadas às observações previamente obtidas. O procedimento parte da premissa da possibilidade de partição para a matriz de correlação em submatrizes de correlação descrito em [Joe \(2006\)](#).

A ideia central do método C-vine parte da proposta de gerar uma matriz de correlação $k \times k$ através de amostragem de $k(k-1)/2$ valores de correlações parciais escolhidas sob critérios específicos. A distribuição de uma correlação parcial é uma distribuição Beta de parâmetros $(k-n)/2$ e $(k-n)/2$ em que k é a cardinalidade do conjunto de variáveis condicionantes de acordo com [Ghosh e Henderson \(2003\)](#).

4.1 Experimentos Básicos

Inicialmente, foram executados testes gerais, sem pré-fixação de níveis limites para os valores de correlação para diversas dimensões matriciais. Em um segundo momento, os testes verificaram o desempenho da geração

de matrizes para limites específicos mínimo e máximo para os valores de correlação. Um terceiro teste contou com matrizes completamente customizados em relação aos níveis de intensidade das correlações. Por fim, foi realizado um teste no qual os limites das correlações eram pré-fixados e, para cada execução, um vetor com percentuais de correlações em cada um dos níveis de correlação construídos foi sorteado, ou seja, um desenho bem específico de matriz.

Os experimentos são dependentes de alguns parâmetros específicos que devem ser definidos previamente. A construção dos algoritmos de execução detalhada no capítulo 3 exige a escolha do número de observações em cada vetor coluna X_i . Nesse experimento os vetores são compostos por 1000 observações. Para gerações de matrizes que utilizam o mecanismo de otimização o experimento executado utilizou $N = 1000$ e $K = 0,03$. Vale lembrar que N representa número máximo de tentativas de otimização e K representa limiar máximo de aceitação da distância euclidiana entre γ_i e θ_i definidos na subseção 3.2.3. Os experimentos aqui conduzidos visam atestar a eficácia do método proposto em gerar aleatoriamente matrizes de correlação. .

4.1.1 Geração de Matrizes sem Especificação Prévia

Nesse contexto de simulação o interesse está nos tempos de execução do algoritmo. Foram realizados experimentos com matrizes de dimensão 3; 10; 20; 50; 100; 200; 500; 750 e 1000; em cada cenário foram geradas 1000 matrizes, os resultados do tempo de execução com as 1000 execuções podem ser visualizados na tabela 4.1.1 e na figura 4.1.1.

Tabela 4.1.1: Tempos de execução para matrizes gerais.

| dimensão | Tempo (segundos) | | | |
|----------|-----------------------|---------|----------|---------|
| | média (desvio padrão) | mínimo | perc. 95 | máximo |
| 3 | 0,00111 (0,00317) | 0,00000 | 0,00301 | 0,07399 |
| 10 | 0,00238 (0,00316) | 0,00000 | 0,00801 | 0,03109 |
| 20 | 0,00554 (0,00726) | 0,00000 | 0,01569 | 0,14100 |
| 50 | 0,01797 (0,01412) | 0,00000 | 0,04105 | 0,20145 |
| 100 | 0,05710 (0,03002) | 0,01362 | 0,11331 | 0,24753 |
| 250 | 0,21116 (0,09796) | 0,10031 | 0,36526 | 1,25385 |
| 500 | 0,69312 (0,31657) | 0,41607 | 1,24016 | 3,02652 |
| 750 | 1,49685 (0,63012) | 0,97060 | 2,61086 | 6,19616 |
| 1000 | 2,59961 (1,02785) | 1,72086 | 4,45336 | 9,29481 |

Nesse teste específico não existe a necessidade do mecanismo de otimização. Como já mencionado, por uma questão de construção, todas as matrizes geradas atendem à condição matemática de serem matrizes de correlação e, como não existem especificações prévias para as matrizes que serão geradas, uma única execução é suficiente para o objetivo desse teste.

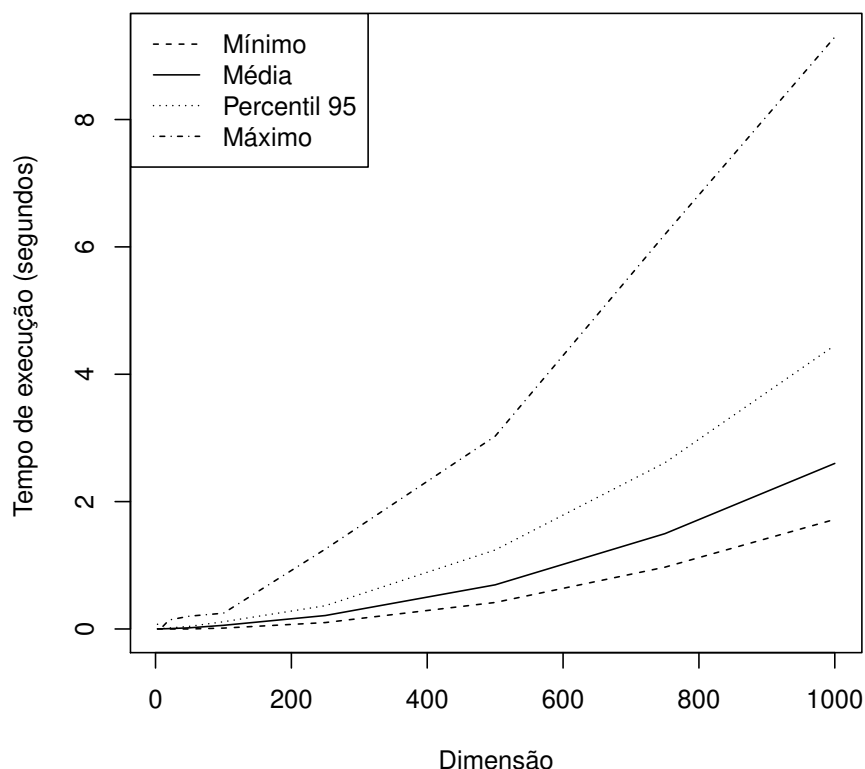


Figura 4.1.1: Tempos de execução para matrizes gerais.

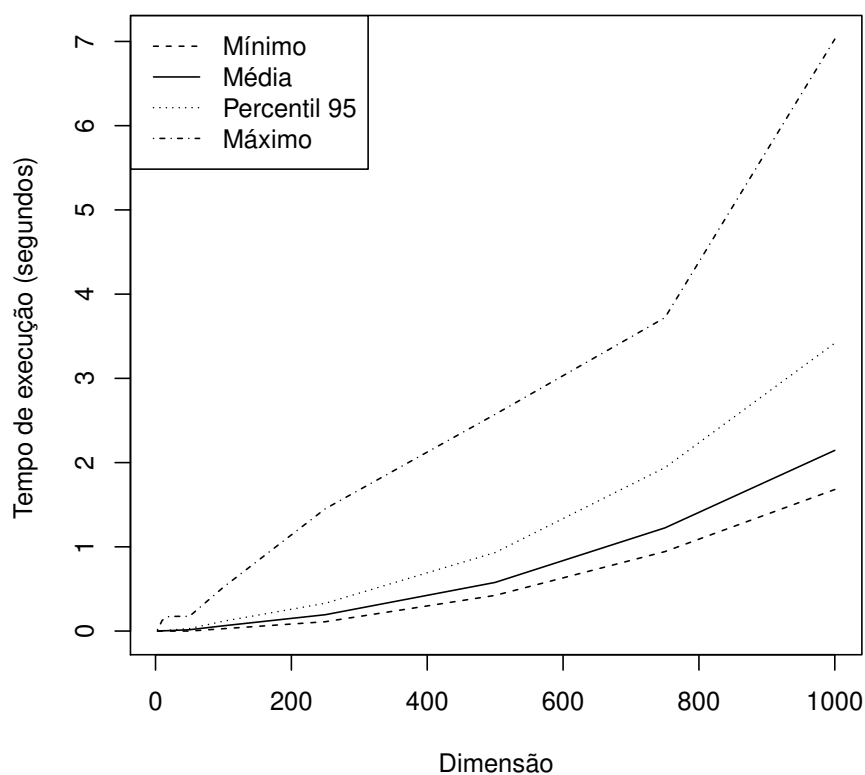
A constatação central através da tabela 4.1.1 e da figura 4.1.1 diz respeito à relação entre os tempos de execução e a dimensão das matrizes geradas. A figura 4.1.1 ilustra que o crescimento de tempo computacional com respeito ao aumento das dimensões de matrizes geradas é bem comportado, não existem sinais de aumentos proibitivos, mesmo para matrizes suficientemente grandes como o caso 1000×1000 . O limite máximo de tempo entre execuções mostra que em casos fortuitos o tempo se amplia, mas na média o tempo de execução é bem comportado. Além disso, os valores de desvio padrão apresentados na tabela 4.1.1, bem como o percentil 95 deixam claro que os valores de tempo máximo elevados são casos raros e não recorrentes.

4.1.2 Geração de Matrizes com Limites Específicos para cada Medida de Correlação

Novamente foram realizados experimentos com matrizes de dimensões 3; 10; 20; 50; 100; 200; 500; 750 e 1000, para cada um dos cenários, foram geradas 1000 matrizes. Em todas as situações, os limites escolhidos foram sempre selecionados com interesse em garantir $0,3 < |\rho| \leq 0,8$. Os resultados do tempo de execução podem ser visualizados na tabela 4.1.2 e na figura 4.1.2.

Tabela 4.1.2: Tempos de execução para matrizes com $0,3 < |\rho| \leq 0,8$.

| dimensão | Tempo (segundos) | | | |
|----------|------------------------|---------|----------|---------|
| | média (desvio padrão) | mínimo | perc. 95 | máximo |
| 3 | 0,0008226156 (0,00133) | 0,00000 | 0,00199 | 0,01805 |
| 10 | 0,0024255824 (0,00454) | 0,00000 | 0,00500 | 0,12852 |
| 20 | 0,0051925387 (0,00705) | 0,00000 | 0,01099 | 0,17400 |
| 50 | 0,0155297747 (0,01081) | 0,00000 | 0,03003 | 0,17498 |
| 100 | 0,0625075152 (0,03393) | 0,02799 | 0,11699 | 0,52075 |
| 250 | 0,1940926809 (0,09111) | 0,11097 | 0,32992 | 1,45217 |
| 500 | 0,5773629847 (0,22484) | 0,42330 | 0,93207 | 2,57274 |
| 750 | 1,2249448297 (0,39301) | 0,94357 | 1,94005 | 3,72111 |
| 1000 | 2,1450527096 (0,67797) | 1,68032 | 3,41809 | 7,02723 |

Figura 4.1.2: Tempos de execução para matrizes com $0,3 < |\rho| \leq 0,8$.

A figura 4.1.2 ilustra o crescimento de tempo computacional com o aumento das dimensões de matrizes geradas. Não existem sinais de aumentos proibitivos, ou seja, aumento no tempo de execução ocorrendo de forma descontrolada com respeito ao aumento das dimensão da matriz de correlação a ser gerada. A evolução do tempo médio com o aumento da dimensão da matriz ocorre de forma não-linear, porém não se revela um

crescimento em tempo de execução que seja proibitivo. Na comparação entre as tabelas 4.1.1 e 4.1.2 um claro padrão de similaridade é verificado. Isso se deve ao fato de que acrescentar limites para os valores de correlação, de acordo com o Teorema 3.1.1, não afeta o tempo computacional para geração das matrizes.

4.1.3 Geração de Matrizes Customizadas

Para execução desse experimento, é preciso estabelecer os limites para os níveis de correlação e um vetor de probabilidades para determinar o volume de medidas de correlações que serão incluídas em cada um dos níveis de correlação. Para tanto, foi adotada uma estratégia simples para estabelecer os limites dos níveis de uma forma aleatória.

Um valor aleatório \mathcal{L} é gerado uniformemente no intervalo $(0,1;1)$ e os seguintes níveis para valores de correlação são estabelecidos:

- (1) $0 < |\rho| \leq \mathcal{L}^2$;
- (2) $\mathcal{L}^2 < |\rho| \leq \mathcal{L}$ e
- (3) $\mathcal{L} < |\rho| < 1$.

A escolha do intervalo $(0,1;1)$ com limite inferior fixado em 0,1 se deve ao fato de que valores abaixo desse limiar poderiam se tornar muito próximos de 0 quando tomados ao quadrado. Isso produziria níveis muito estreitos e pouco úteis em termos de avaliação de desempenho. Estes limites são aplicados no algoritmo customizado, com proporções iguais em cada nível de correlação.

Novamente foram realizados experimentos com matrizes de dimensão 3; 10; 20; 50; 100; 200; 500; 750 e 1000, e em cada cenário foram geradas 1000 matrizes com resultados apresentados nas tabelas 4.1.3 e 4.1.4 e nas figuras 4.1.3 e 4.1.4.

Tabela 4.1.3: Tempos de execução para matrizes customizadas.

| dimensão | Tempo (segundos) | | | |
|----------|-----------------------|---------|------------|------------|
| | média (desvio padrão) | mínimo | perc. 95 | máximo |
| 3 | 0,1753 (0,3301) | 0,00000 | 0,74710 | 2,22699 |
| 10 | 0,8181 (1,0873) | 0,00199 | 3,05680 | 3,71402 |
| 20 | 0,9120 (1,6449) | 0,00499 | 4,70700 | 5,82358 |
| 50 | 2,0880 (4,3176) | 0,02299 | 11,44100 | 20,87895 |
| 100 | 6,3418 (11,6312) | 0,08100 | 29,14029 | 33,85034 |
| 250 | 33,5348 (57,5129) | 0,29599 | 135,57460 | 193,00202 |
| 500 | 99,6911 (202,5429) | 1,45200 | 535,82239 | 850,98490 |
| 750 | 252,0071 (477,6025) | 2,51297 | 1227,92619 | 1446,84853 |
| 1000 | 446,9882 (855,5721) | 5,88199 | 2179,02340 | 3690,74851 |

A tabela 4.1.4 apresenta resultados sobre a quantidade de rodadas do procedimento de otimização utilizado.

Tabela 4.1.4: Quantidade de simulações necessárias para otimização.

| dimensão | Tempo (segundos) | | | |
|----------|-----------------------|--------|------------|------------|
| | média (desvio padrão) | mínimo | perc. 95 | máximo |
| 3 | 78,51242 (160,22580) | 0 | 368,20000 | 1000,00000 |
| 10 | 261,46727 (361,06390) | 0 | 1000,00000 | 1000,00000 |
| 20 | 191,57336 (354,80100) | 0 | 1000,00000 | 1000,00000 |
| 50 | 179,09029 (372,82910) | 1 | 1000,00000 | 1000,00000 |
| 100 | 219,56433 (405,50480) | 2 | 1000,00000 | 1000,00000 |
| 250 | 250,09932 (428,90480) | 1 | 1000,00000 | 1000,00000 |
| 500 | 191,44921 (388,69480) | 2 | 1000,00000 | 1000,00000 |
| 750 | 213,73138 (405,76380) | 1 | 1000,00000 | 1000,00000 |
| 1000 | 211,69752 (404,05120) | 2 | 1000,00000 | 1000,00000 |

A figura 4.1.3 apresenta o padrão de crescimento do tempo computacional com respeito a dimensão das matrizes geradas, na qual novamente não se detectam sinais de aumentos proibitivos.

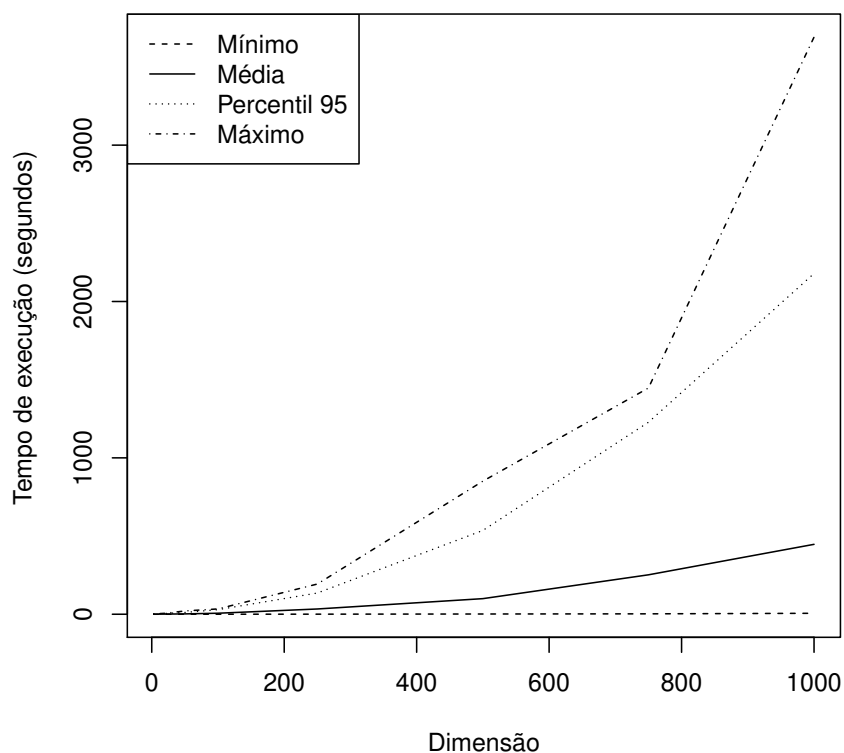


Figura 4.1.3: Tempos de execução para matrizes customizadas.

O tempo médio de execução dos procedimentos de geração de matrizes customizadas também apresenta um crescimento não-linear. Observa-se que existe uma distância significativa entre o percentil 95 e os valores má-

ximos, indicando que estes se tratam de *outliers*. Em média, o percentil 95 representa 68% do valor máximo. Mesmo para matrizes de dimensões superiores, como as de dimensão 1000, que possui 499500 correlações simuladas, observa-se tempo médio de geração de aproximadamente 7 minutos, o que foi considerado satisfatório ao se considerar o número de operações envolvido no procedimento de geração, bem como o nível de customização atingido pelo método.

A figura 4.1.4 ilustra o cenário do número de execuções do procedimento de otimização para distribuir as correlações entre os vários níveis de correlação.

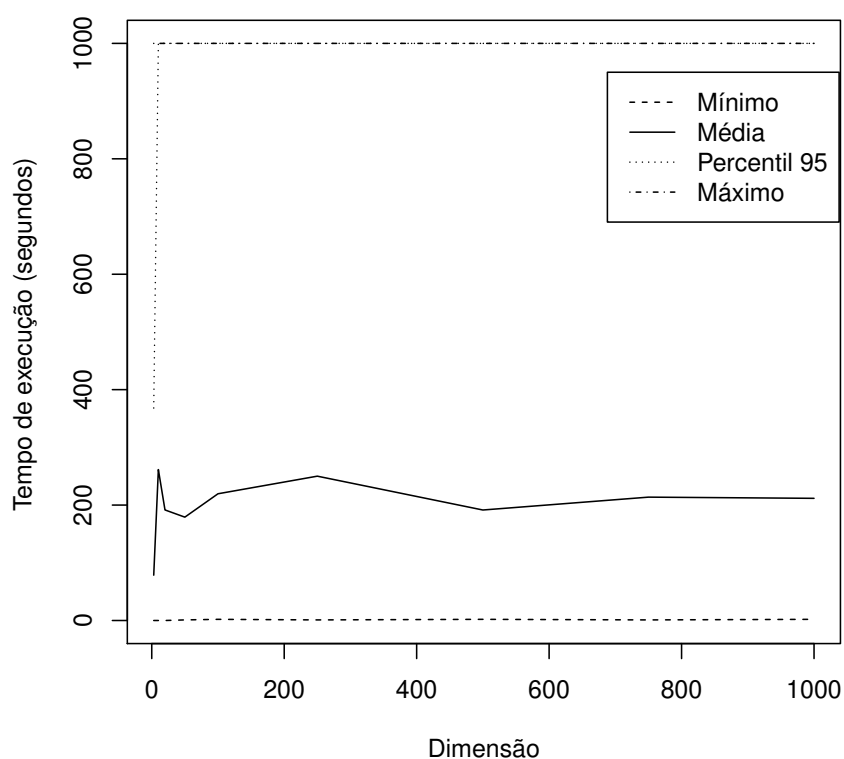


Figura 4.1.4: Quantidade de simulações para otimização para matrizes customizadas.

Uma particularidade está associada a esse estudo, para matrizes com dimensão entre 3 e 10. Ao dividir o volume total de medidas de correlação em matrizes de dimensão k , dado por $k(k-1)/2$, por três proporções igualitárias nos três níveis, para os casos de dimensão 5 e 8 não existem combinações de valores inteiros para atingir a divisão proporcional exata entre os níveis. Essa peculiaridade faz com que o processo de otimização necessite de mais rodadas. Não se trata de uma característica impeditiva, mas uma constatação pontual relevante nesse estudo. Em particular, o procedimento de otimização tende a ser mais custoso em casos de matrizes pequenas. A busca por um exato vetor de proporções em um problema de

variáveis inteiras já é bastante sofisticado, e se o espaço de busca dessas variáveis inteiras for muito restritivo (caso específico de matrizes pequenas) o problema fica ainda mais difícil.

Em relação ao número de simulações necessárias à convergência das matrizes customizadas para as proporções desejadas, com margem de erro máxima de 3%, observa-se que para todos os casos testados, a convergência ocorre em média próxima à 200 execuções do procedimento de otimização. Tal fato indica que o crescimento do tempo computacional é função exclusiva da dimensão da matriz gerada e que o procedimento de otimização por si não apresenta dificuldades em convergir, independentemente da dimensão da matriz de interesse.

Em todas as dimensões, ocorreram gerações que necessitaram das 1000 iterações para a obtenção dos valores que estivessem o mais próximos possíveis dos requeridos. Observa-se que neste caso, o máximo e o percentil 95 coincidem. No experimento customizado, observou-se que em aproximadamente 83% dos casos o algoritmo alcançou a convergência, ou seja, conseguiu gerar as matrizes customizadas conforme solicitado dentro da margem de erro estipulada, o que indica que o mecanismo de otimização, embora apresente casos nos quais não se atinge a otimização plena, fornece a matriz customizada desejada na ampla maioria dos casos analisados.

Além dos experimentos anteriores, um outro experimento específico, também customizado, foi conduzido. Foram geradas 1000 matrizes 20×20 e 50×50 com níveis de correlação: (1) $0 < |\rho| \leq 0,3$; (2) $0,3 < |\rho| \leq 0,6$ e (3) $|\rho| > 0,6$. Para cada execução, um vetor com as proporções de correlações em cada um dos níveis era gerado aleatoriamente. A matriz era obtida com intuito de atender tal customização. Os valores de tempo de execução, número de rodadas para otimização e o erro máximo entre as proporções observadas e esperadas são apresentados na 4.1.5

Tabela 4.1.5: Avaliação para matrizes customizadas, tempo, volume de execuções para otimização e proporção de erros.

| dimensão | Tempo (segundos) | | |
|----------|----------------------------------|---------|------------|
| | média (desvio padrão) | mínimo | máximo |
| 20 | 0,27471 (0,66573) | 0,00400 | 4,32861 |
| 50 | 0,12240 (0,36323) | 0,00999 | 7,49732 |
| dimensão | Otimização (número de execuções) | | |
| | média (desvio padrão) | mínimo | máximo |
| 20 | 67,58600 (171,67137) | 1,00000 | 1000,00000 |
| 50 | 10,60900 (46,75888) | 1,00000 | 1000,00000 |
| dimensão | Erro máximo de proporção | | |
| | média (desvio padrão) | mínimo | máximo |
| 20 | 0,02162 (0,01119) | 0,00158 | 0,13368 |
| 50 | 0,01959 (0,00713) | 0,00122 | 0,06126 |

Mesmo em situações bastante específicas, com interesse em atender problemas bastante singulares o método é capaz de fornecer as matriz em tempos computacionais desprezíveis. A convergência no procedimento de otimização é bastante veloz. A tabela 4.1.5 apresenta ainda uma avaliação da diferença entre a proporção desejada para distribuição entre os níveis de correlação e a proporção alcançada pela estratégia de otimização. Para os três níveis de correlação em cada matriz gerada, a tabela 4.1.5 ilustra estatísticas para diferença máxima entre a proporção desejada e a proporção alcançada.

Para o erro limite $K = 0,03$ pré-fixado no procedimento de otimização, para matrizes de dimensão 20, apenas em 1,7% dos casos as 1000 rodadas de otimização foram utilizadas antes do critério de parada do erro limite K . Já para matrizes de dimensão 50, apenas em 0,1% dos casos as 1000 rodadas de otimização foram utilizadas antes do critério de parada do erro limite K , confirmando a robustez do método de geração de matrizes customizadas proposto neste estudo.

4.2 Experimentos Comparativos

Aqui serão apresentados experimentos comparativos para a metodologia CM-generator com respeito a outras metodologias para geração de matrizes de correlação. Em particular, duas metodologias já consolidadas na literatura são discutidas, o método denominado Onion proposto por Joe (2006) e o método denominado C-vine proposto por Ghosh e Henderson (2003). Inicialmente foram executados experimentos cujo interesse era visualizar a estrutura da distribuição das correlações geradas por meio das três metodologias.

É importante mencionar a condição matemática sobre o valor do limite inferior ℓ_{lower} na geração de correlações aleatórias. Uma atribuição de valor para o limite ℓ_{lower} que esteja extremamente próxima de zero pode possibilitar a construção de intervalos demasiadamente largos para escolha dos valores dos desvios-padrão. Em algumas situações, essa situação pode acarretar na geração de um grande volume de correlações próximas de zero. Os estudos experimentais que foram realizados, são capazes de atestar que este efeito somente ocorre para $\ell_{lower} < 10^{-4}$, ou seja, bastante próximo de zero.

Essa imposição, para ℓ_{lower} , tende a afetar, de forma suave, a estrutura da distribuição das medidas de correlações observadas. Os experimentos foram executados para várias dimensões matriciais, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200. Para cada dimensão matricial fixada, foram geradas 1000 matrizes através de cada uma das metodologias em estudo (Onion, C-vine e CM-generator).

A figura 4.2.1 apresenta uma comparação entre matrizes de correlação geradas pelos três métodos, em particular, o método CM-generator com limite inferior fixado em $\ell_{lower} = 0,005$. São apresentadas densidades ajustadas através das correlações obtidas nas 1000 matrizes geradas para cada um dos métodos.

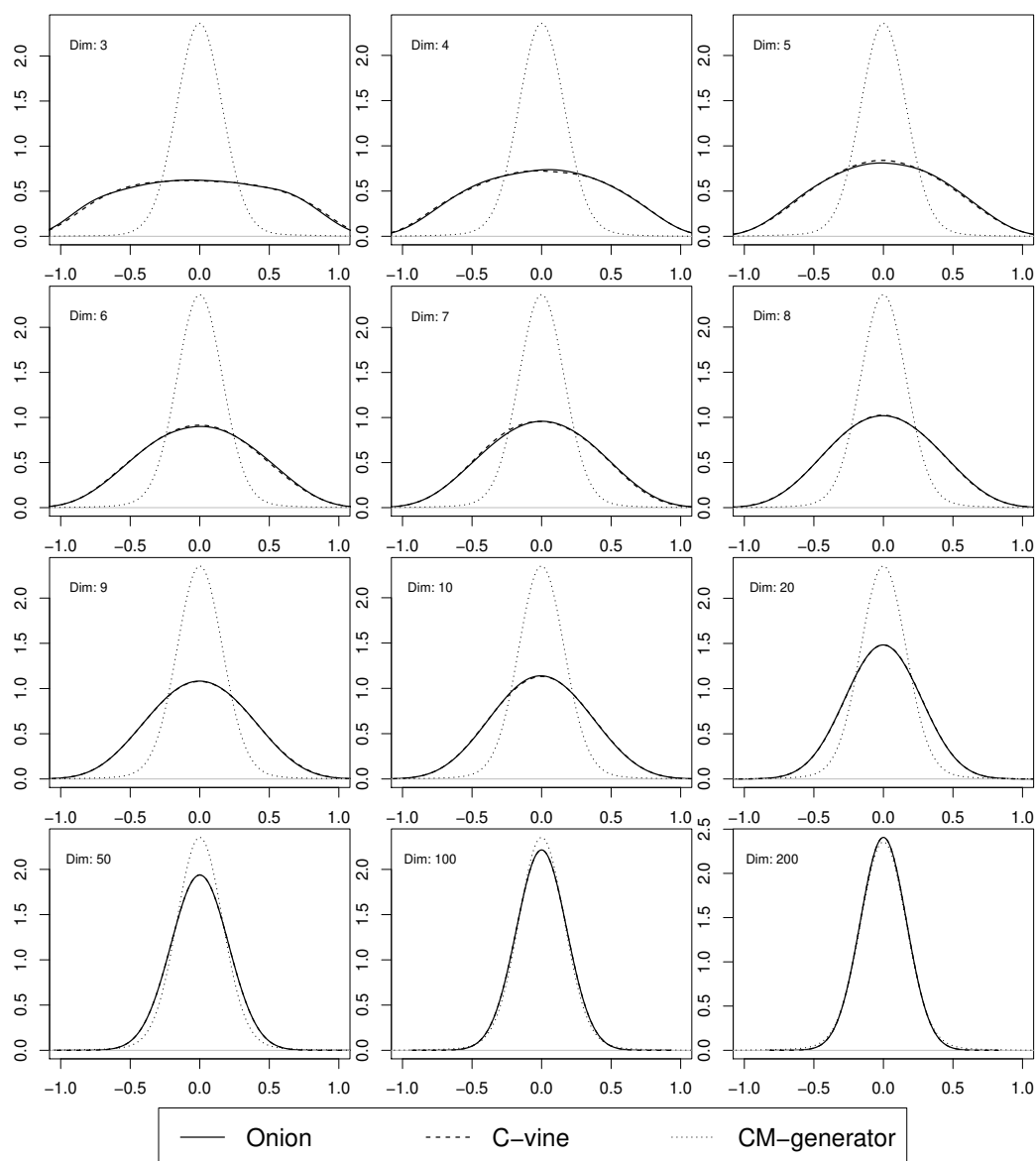


Figura 4.2.1: Distribuição de correlações através dos métodos Onion, C-vine e CM-generator com $\ell_{lower} = 0,005$.

Por meio da figura 4.2.1, verifica-se que a distribuição das medidas de correlações nos métodos Onion e C-vine é bastante influenciada pela dimensão da matriz de correlação gerada. Para dimensões pequenas, os métodos Onion e C-vine distribuem medidas de correlações altas com probabilidades elevadas, esse efeito vai se dissipando com o aumento da dimensão matricial.

Por outro lado, a estrutura da distribuição das correlações é sempre bastante estável para as matrizes de correlação geradas pelo método CM-generator. O aumento dimensional também ilustra que os métodos Onion e C-vine convergem em estrutura da distribuição para a resposta obtida através do método CM-generator com o aumento dimensional das matrizes de correlação.

As conclusões obtidas por meio da figura 4.2.1 podem deixar algum

receio quanto a capacidade do método CM-generator em gerar valores de correlações mais elevados. Este efeito está diretamente ligado à escolha do limite inferior ℓ_{lower} . Uma pequena elevação no valor do limite inferior ℓ_{lower} pode alterar de forma significativa o peso das caudas da distribuição. Essa elevação aumenta sensivelmente a probabilidade associada à geração de correlações mais altas.

Este efeito pode ser verificado na figura 4.2.2, produzida com as mesmas dimensões matricias da figura 4.2.1, porém com $\ell_{lower} = 0,025$. A figura 4.2.2 novamente mostra a distribuição das correlações nos métodos Onion e C-vine sendo influenciada pela dimensão da matriz gerada. Novamente a estrutura da distribuição das correlações é bastante estável no método CM-generator.

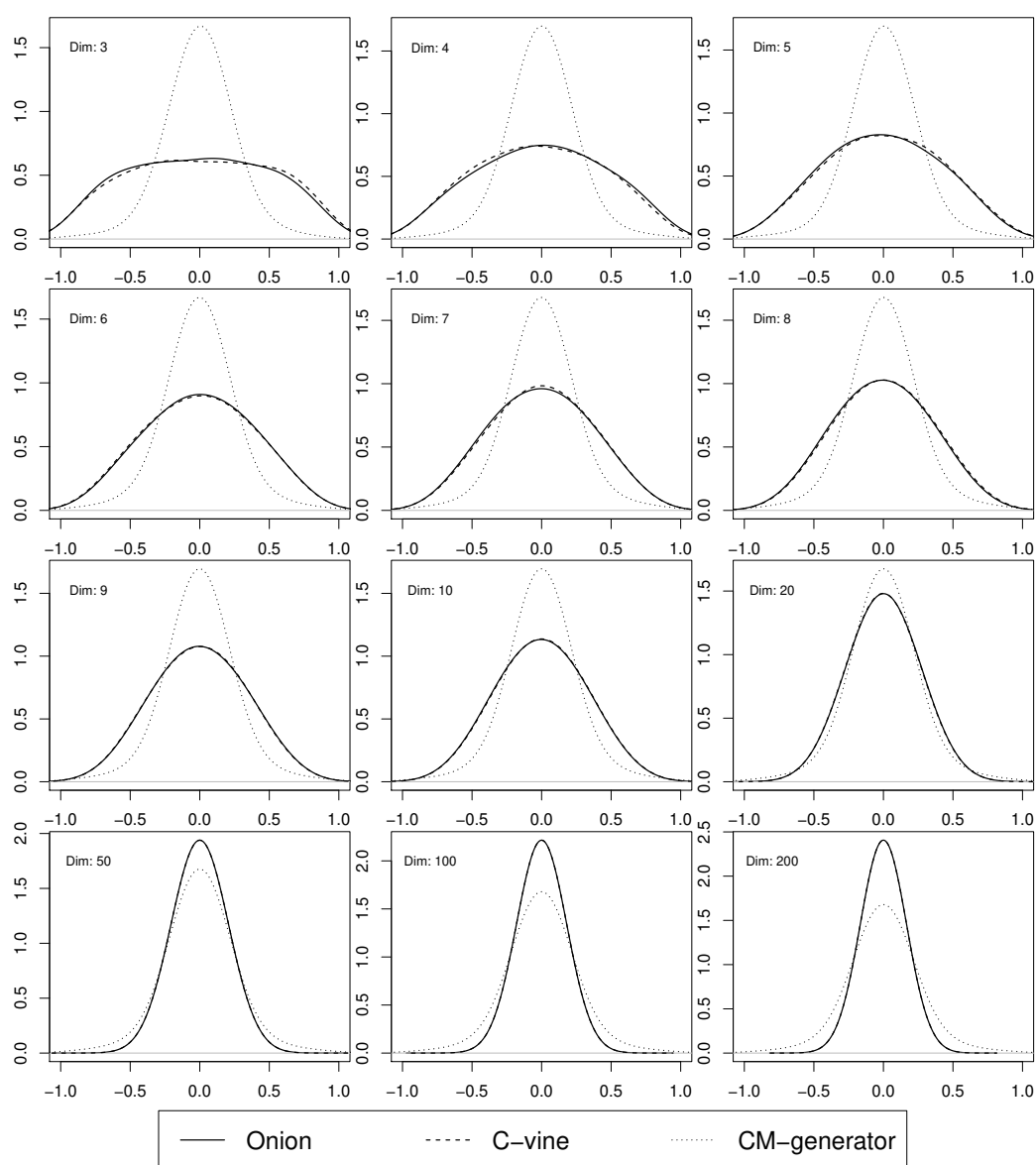


Figura 4.2.2: Distribuição de correlações através dos métodos Onion, C-vine e CM-generator com $\ell_{lower} = 0,025$.

As caudas das distribuições de correlações geradas através dos métodos

CM-generator ampliam o peso probabilístico para correlações mais elevadas. Isso quando comparadas às caudas das distribuições apresentadas na figura 4.2.1.

O aumento dimensional das matrizes de correlação faz com que os métodos Onion e C-vine se tornem mais restritivos para geração de correlações elevadas que o próprio método CM-generator. O usuário do CM-generator tem uma distribuição de correlações mais estável. É possível influenciar no peso das caudas dessa distribuição através de escolhas prévias para o valor de ℓ_{lower} . Valores para o limite ℓ_{lower} que estejam muito próximos de zero gerarão matrizes de correlação cuja distribuição apresentará caudas mais leves. Por outro lado, valores mais distantes de zero garantirão matrizes de correlação cuja distribuição apresentará caudas mais pesadas.

A figura 4.2.3 ilustra claramente este efeito para o método CM-generator em experimentos com uma dimensão matricial fixada em 30. Nota-se que uma pequena elevação no valor de ℓ_{lower} afeta substancialmente o peso nas caudas da distribuição das correlações. Essa possibilidade de escolha introduz para o usuário do método um relevante grau de liberdade no procedimento de geração de matrizes de correlação. Por meio de escolhas do pesquisador, as matrizes geradas podem ser mais adequadas ao problema real sob investigação.

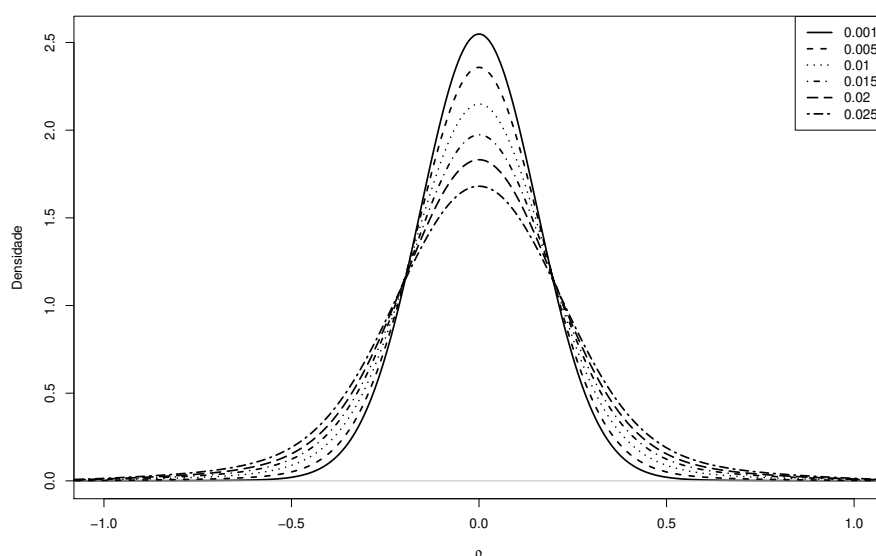


Figura 4.2.3: Evolução da distribuição de correlações através do método CM-generator para diferentes valores de ℓ_{lower} .

Os experimentos anteriores atestam a habilidade do método em gerar matriz de correlação sem especificações previamente fixadas. Em particular, os tempos computacionais para execução desta tarefa não apresentam restrições impactantes. Ao contrário, são mais rápidos que os métodos comparados. Na geração de 1000 matrizes de dimensão 200, o método C-vine executou as gerações, em média, em 0,774 seg. com desvio padrão de 0,223 seg. Já o método Onion repetiu o mesmo procedimento, em média,

em 0,368 seg. com desvio padrão de 0,202 seg. E por fim, o CM-generator executou a tarefa, em média, em 0,080 seg. com desvio padrão de 0,046 seg.

4.3 Experimentos Customizados

Apesar das discussões comparativas da seção anterior, a grande contribuição do método proposto, não é gerar matrizes de correlação gerais, mas sim discutir a possibilidade da geração de matrizes customizadas. Diversas estratégias de customização podem ser aventadas. Inicialmente serão discutidas customizações mais simples.

Um formato possível de customização seria estabelecer um valor máximo para as correlações. Por exemplo, suponha definir como medida de correlação baixa, correlações pertencentes ao intervalo $[-0,3; 0,3]$. Uma outra customização seria estabelecer o valor mínimo para as correlações. Por exemplo, suponha produzir uma matriz de correlação com todas as correlações contidas no conjunto real $[-1,0; -0,6) \cup (0,6; 1,0]$. Um terceiro formato, já mais específico, seria estabelecer limites máximo e mínimo para as correlações. Por exemplo, suponha produzir uma matriz de correlação com todas as correlações contidas no conjunto real $[-0,6; -0,3) \cup (0,3; 0,6]$.

Experimentos com os formatos abordados anteriormente são discussões voltadas para aplicações mais específicas. O interesse é permitir investigações científicas em aplicações em que diferentes níveis de correlação podem impactar substancialmente. A figura 4.3.1 apresenta o resultado para a geração de 1000 matrizes de dimensão 20 através de histogramas referentes às correlações obtidas nas matrizes geradas. A figura 4.3.1(a) apresenta a distribuição das correlações baixas, pertencentes ao intervalo $[-0,3; 0,3]$. A estrutura de distribuição é semelhante a estrutura visualizada nas matrizes gerais apesar de restrita ao intervalo de interesse. A figura 4.3.1(b) mostra a distribuição das correlações médias, contidas no conjunto real $[-0,6; -0,3) \cup (0,3; 0,6]$, e a figura 4.3.1(c) mostra a distribuição das correlações altas, inclusas no conjunto real $[-1,0; -0,6) \cup (0,6; 1,0]$. Em todas as situações a distribuição permanece semelhante aos formatos visualizados para matrizes de correlação sem especificações prévias.

A capacidade de gerar matrizes de correlação com limites específicos para os valores das correlações é muito particular ao método CM-generator. O procedimento é realizado com o mesmo esforço computacional que o procedimento de geração sem especificação. Resultados dessa natureza trazem ao pesquisador o poder de escolha de níveis de correlação para a execução de testes de natureza variada.

As customizações apresentadas nas figuras 4.3.1(a), 4.3.1(b) e 4.3.1(c) já são bastante específicas. A literatura não apresenta metodologias capazes de gerar matrizes tão particulares para investigações científicas. Obviamente, os limites para definir correlações altas, médias e baixas podem ser alterados de acordo com a conveniência do usuário do método. Alterações desse tipo não afetam a execução do procedimento, apenas limitam as correlações geradas na matriz resultante.

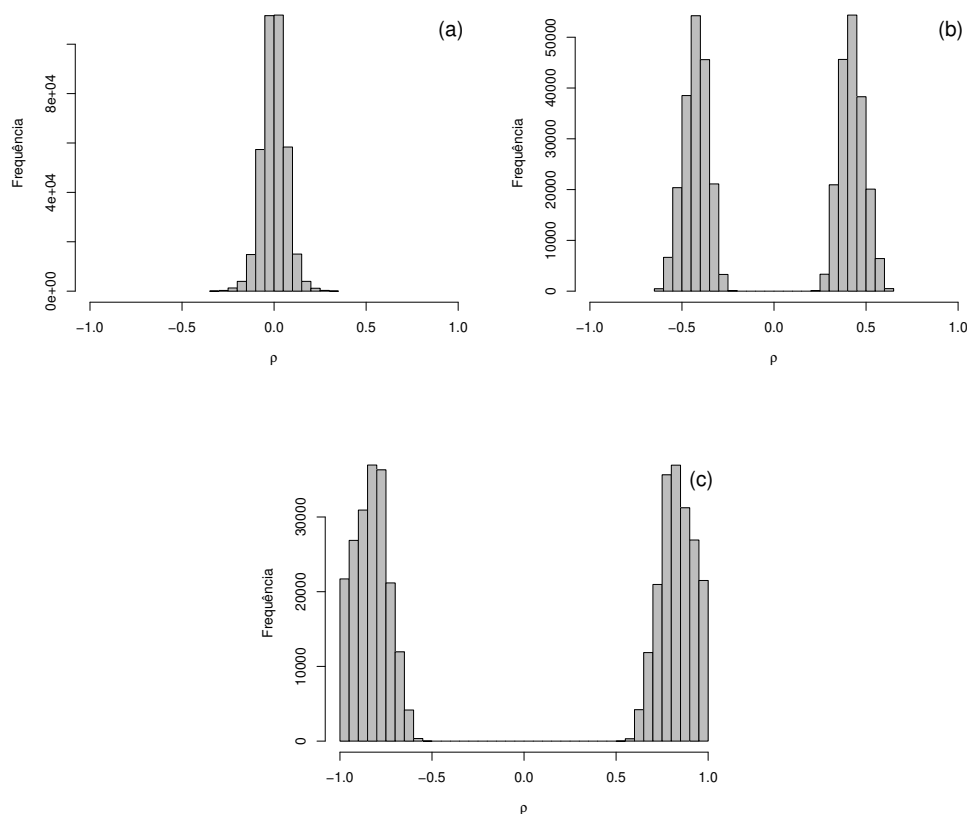


Figura 4.3.1: Histogramas para matrizes com limites nas correlações geradas através do CM-generator, (a) correlações baixas, (b) correlações médias, e (c) correlações altas.

Níveis de customização ainda mais específicos podem ser abordados. Suponha os mesmos limiares anteriores para estabelecer o que são correlações altas, médias e baixas. Mas agora, suponha produzir uma única matriz de correlação com correlações nos três níveis simultaneamente, por exemplo $p\%$ de correlações baixas, $q\%$ de correlações médias, e $(100 - p - q)\%$ de correlações altas. Um procedimento dessa ordem deverá seguir a estratégia de otimização mencionada na sub-seção 3.2.3.

É importante observar que a existência de uma partição inteira exata entre as correlações que satisfaça a divisão $p\%$, $q\%$ e $(100 - p - q)\%$ pode não existir. Em uma situação de impossibilidade, a customização deve fornecer ao pesquisador uma matriz que se aproxime ao máximo da partição desejada. Este estudo apresentará exemplos experimentais de matrizes geradas com três níveis de correlação inclusos na mesma matriz. Entretanto, a metodologia proposta é eficaz para trabalhar com outras quantidades de níveis de correlação. Em outras palavras, o experimento que será apresentado, não possui nenhuma perda de generalidade quanto a capacidade de produzir matrizes com diversos níveis de intensidade de correlação. A figura 4.3.2 apresenta histogramas para diferentes partições em três níveis de correlação.

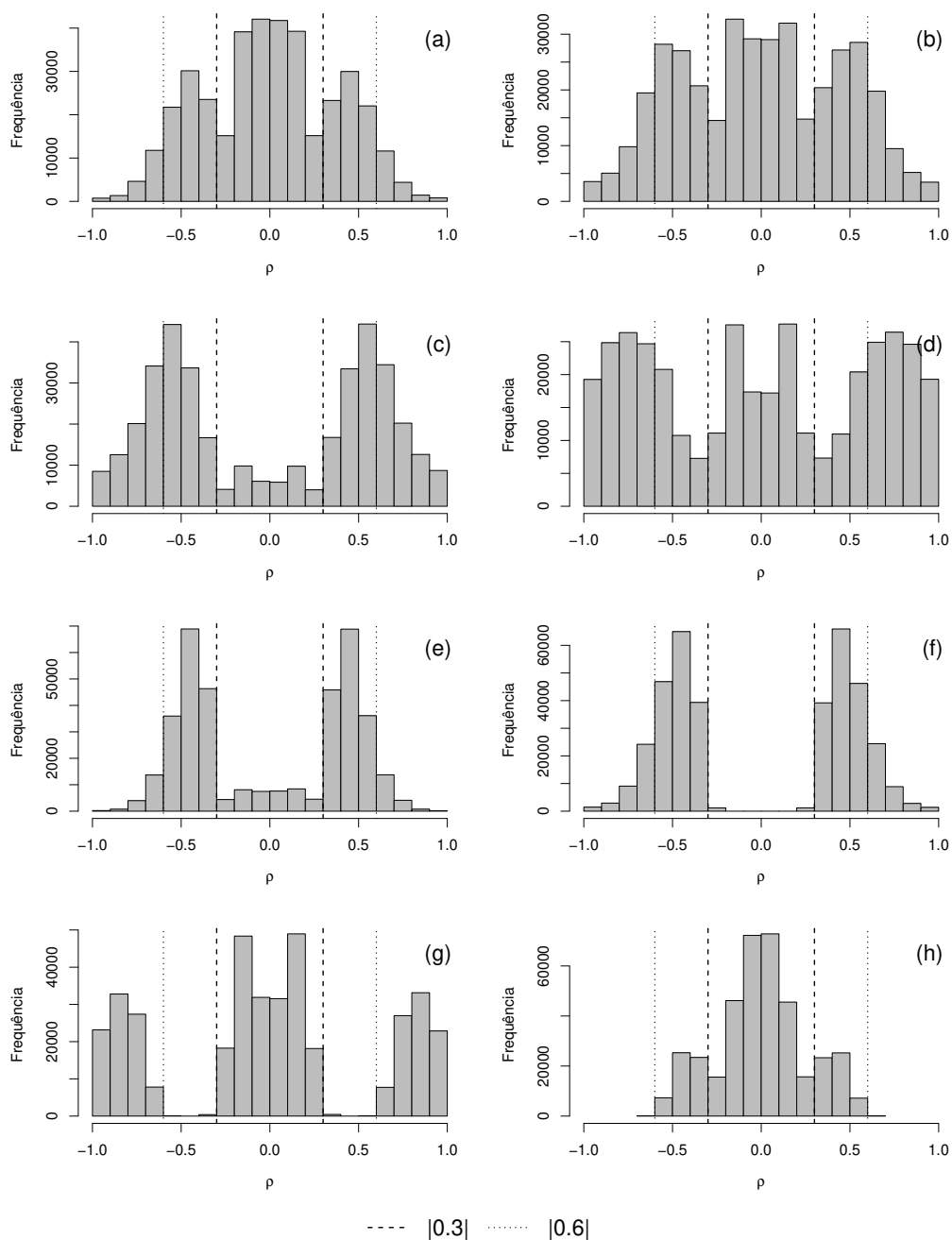


Figura 4.3.2: Histogramas para matrizes customizadas com $p\%$ de correlações baixas, $q\%$ de correlações médias, e $(100 - p - q)\%$ de correlações altas geradas através do CM-generator, (a) $p = 50\%$ e $q = 40\%$, (b) $p = 40\%$ e $q = 40\%$, (c) $p = 10\%$ e $q = 50\%$, (d) $p = 30\%$ e $q = 20\%$, (e) $p = 10\%$ e $q = 80\%$, (f) $p = 0\%$ e $q = 80\%$, (g) $p = 50\%$ e $q = 0\%$, (h) $p = 70\%$ e $q = 30\%$.

Os níveis de correlação foram novamente fixados em correlações baixas pertencentes ao intervalo $[-0,3; 0,3]$, correlações médias no conjunto real $[-0,6; -0,3) \cup (0,3; 0,6]$, e correlações altas, inclusas no conjunto real $[-1,0; -0,6) \cup (0,6; 1,0]$. As matrizes geradas em todos os experimentos possuem dimensão 20.

Os histogramas apresentados na figura 4.3.2 apresentam a capacidade do método CM-generator em produzir matrizes com correlações seguindo uma distribuição em níveis pré-fixados. Como mencionado anteriormente, trata-se de um procedimento de otimização para garantir a partição mais adequada. O método não garante a partição exata e o gasto em tempo computacional é mais elevado que os verificados na geração puramente aleatória.

Mesmo em situações bastante específicas, com interesse em atender problemas bastante singulares, a convergência no procedimento de otimização é bastante veloz. O pior cenário, quanto ao tempo computacional, foi para produzir a partição com 30% de correlações baixas, 20% de correlações médias e 50% de correlações altas, ilustrada na figura 4.3.2(d). O tempo médio para a produção de cada uma das matrizes foi de 0,098 seg. O CM-generator executa uma simulação similar de mesma dimensão, porém sem customização em tempo inferior a 0,020 seg.

Além da agilidade no procedimento de execução, uma relevante métrica de qualidade seria avaliar uma comparação entre a partição desejada e a partição alcançada. Para as 1000 matrizes geradas em cada um dos experimentos, a proporção média de correlações baixas médias e altas gerada foi avaliada. O pior cenário foi verificado na partição com 50% de correlações baixas, 0% de correlações médias e 50% de correlações altas, ilustrada na figura 4.3.2(g). Em média, a partição obtida nas matrizes geradas foi de 51,83% de correlações baixas, 0,32% de correlações médias e 47,85% de correlações altas. A melhor situação foi verificada na partição com 40% de correlações baixas, 40% de correlações médias e 20% de correlações altas, ilustrada na figura 4.3.2(b). Em média, a partição obtida nas matrizes geradas foi de 40,04% de correlações baixas, 40,05% de correlações médias e 19,91% de correlações altas.

É importante ressaltar que trata-se de um procedimento de otimização. Não existe expectativa prévia de alcançar a proporcionalidade exata. O objetivo central é fornecer ao pesquisador uma matriz com correlações que se aproximem ao máximo dos níveis desejados. Isso, por si só, é um resultado inovador, que as metodologias já existentes não são capazes de produzir.

4.4 Aplicação na Comparação entre Metodologias para Detecção de *Outliers* Multivariados

A avaliação da existência de *outliers* em conjuntos de dados multivariados é um problema de grande interesse prático em diversas áreas de estudo. Muitos pesquisadores elaboram metodologias inovadoras para este tipo de verificação. As metodologias propostas precisam ser testadas pre-

viamente, através de conjuntos de dados simulados e controlados para ter sua validação garantida.

Filzmoser et al. (2005) apresentam um método para detecção de *outliers* multivariados capaz de distinguir valores extremos de uma distribuição Normal. A forma de conjuntos de dados multivariados é explicada por sua matriz de covariância. As estimativas do efeito de variabilidade são obtidas por meio de estimadores robustos. Dentre os estimadores robustos, pode-se destacar o Minimum Covariance Determinant, aqui nominado MCD, proposto por Rousseeuw e Driessen (1999) e o Minimum Volume Ellipsoid, aqui dito MVE, proposto por Rousseeuw e Zomeren (1990).

Muitos estudos nessa área utilizam procedimentos de simulação de dados como os descritos por Filzmoser (2005). O procedimento em questão gera $n - n_{out}$ pontos de uma distribuição normal p-variada $N_p \sim (\mathbf{0}, I)$ e n_{out} observações da distribuição *outlier* $N_p \sim (\eta \mathbf{1}, I)$. Um experimento dessa natureza parte da premissa de completa independência entre as p variáveis envolvidas. Um experimento mais realista deveria agregar efeitos de dependência entre as variáveis envolvidas, este efeito pode afetar sobremaneira a estrutura de forma dos dados simulados. Este formato de simulação de dados seria capaz de produzir um mecanismo de teste muito mais realista. Para tanto, é necessário um mecanismo de geração de matrizes de correlação válidas para serem adotadas no procedimento de simulação.

Um estudo comparativo para avaliação da qualidade do procedimento de detecção de *outliers* multivariados por meio de procedimentos baseados nos estimadores robustos MCD e MVE foi realizado. A figura 4.4.1 apresenta os resultados para escolhas distintas de $\delta \in \{0,02, 0,05, 0,10\}$, aqui denominada taxa de mistura, e número de variáveis $p \in \{5, 10, 50\}$. Para cada combinação δ e p foram executadas $N = 1000$ simulações.

Para a execução do procedimento comparativo, foram produzidos dados simulados de distribuições Normais p-variadas, com probabilidade $(1 - \delta)$, os dados são de uma distribuição $N_p \sim (\mathbf{0}, \Sigma)$. Com probabilidade δ , os dados são de uma distribuição $N_p \sim (\eta \mathbf{1}, \Sigma)$. As simulações foram executadas com $\eta = 2$ e uma matriz de correlação Σ customizada, garantido 20% de correlações altas, com $|\rho| > 0,6$, 30% de correlações médias tais que $0,3 \leq |\rho| \leq 0,6$ e o restante de correlações baixas, com $|\rho| < 0,3$, obtida através do CM-generator. A métrica de comparação entre os dois procedimentos de detecção foi a acurácia dentre as execuções.

O método baseado no estimado robusto MCD se mostrou sistematicamente mais efetivo que o método baseado no estimador MVE. Todos os gráficos apresentados na figura 4.4.1 confirmam a superioridade com respeito à acurácia no procedimento de detecção. Neste experimento, a acurácia é a razão total de acertos na detecção entre *outliers* e não *outliers*. As figuras 4.4.1 (a), (b) e (c) ilustram que as duas metodologias apresentam perda de qualidade com o aumento excessivo do número de variáveis envolvidas no problema em investigação. O aumento dimensional de $p = 5$ para $p = 10$ não é suficiente para ilustrar essa perda, mas a mudança de $p = 10$ para $p = 50$ ocasiona uma queda substancial da medida de acurácia. O mesmo efeito pode claramente ser observado quando são avaliadas

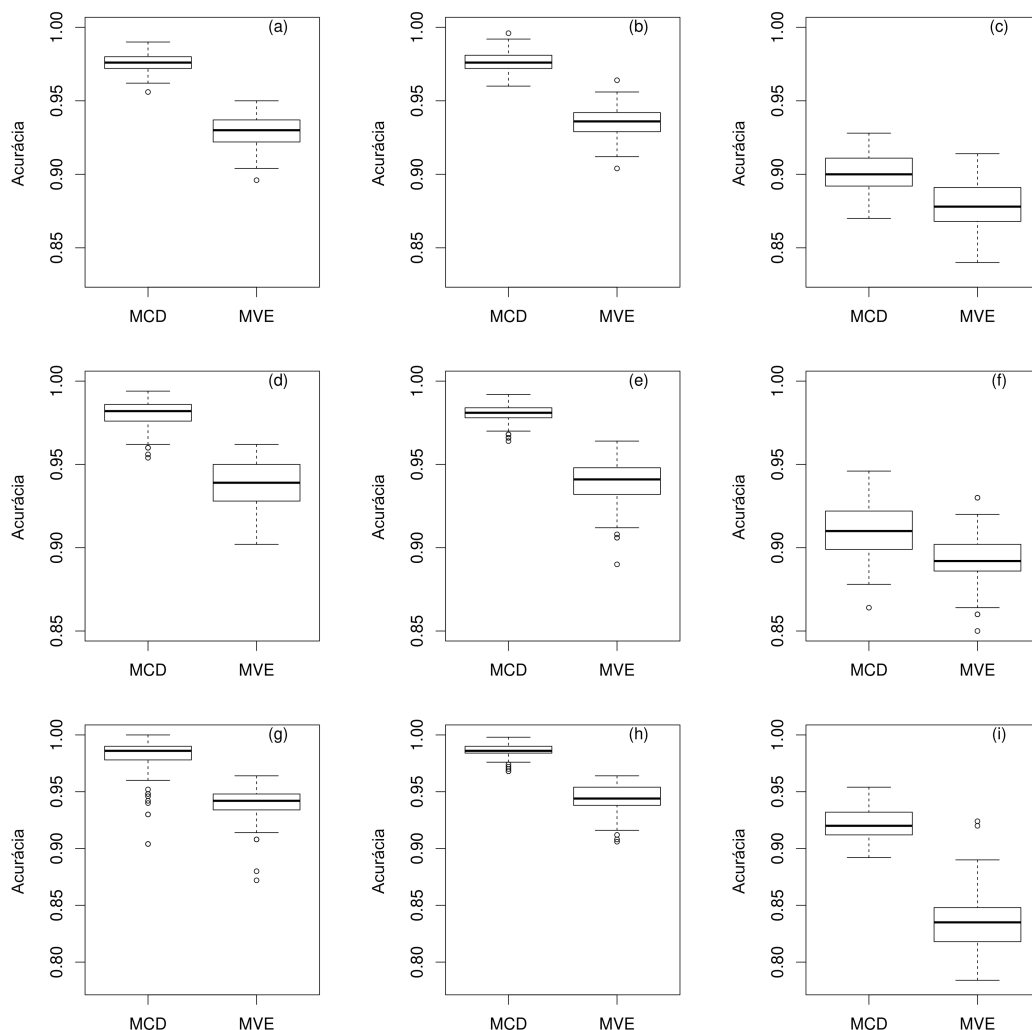


Figura 4.4.1: Comparação entre acurácias obtidas através dos estimadores robustos MCD e MVE, (a) $\delta = 0.02$ e $p = 5$, (b) $\delta = 0.02$ e $p = 10$, (c) $\delta = 0.02$ e $p = 50$, (d) $\delta = 0.05$ e $p = 5$, (e) $\delta = 0.05$ e $p = 10$, (f) $\delta = 0.05$ e $p = 50$, (g) $\delta = 0.10$ e $p = 5$, (h) $\delta = 0.10$ e $p = 10$, (i) $\delta = 0.10$ e $p = 50$.

as figuras 4.4.1 (d), (e) e (f) e também as figuras 4.4.1 (g), (h) e (i). Um outro foco de avaliação considera o aumento da taxa de mistura δ . O aumento do valor δ ocasiona um aumento do volume de *outliers* gerados na amostra simulada. Esse aumento ocasiona também uma suave perda de acurácia nos dois métodos em avaliação.

Referências Bibliográficas

Filzmoser, P. (2005). Identification of multivariate outliers: a performance study. *Austrian Journal of Statistics*, 34(2):127–138.

Filzmoser, P., Garrett, R. e Reimann, C. (2005). Multivariate outlier detec-

- tion in exploration geochemistry. Computers & Geosciences, 31(5):579–587.
- Ghosh, S. e Henderson, S. G. (2003). Behavior of the norta method for correlated random vector generation as the dimension increases. ACM Transactions on Modeling and Computer Simulation (TOMACS), 13(3):276–294.
- Joe, H. (2006). Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis, 97(10):2177–2189.
- Lewandowski, D., Kurowicka, D. e Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis, 100(9):1989–2001.
- Rousseeuw, P. J. e Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3):212–223.
- Rousseeuw, P. J. e Zomeren, B. C. V. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85(411):633–639.

Pacote gencor

Uma relevante contribuição além do método em si é a apresentação de um pacote para utilização do método CM-generator, apresentado no capítulo 3. O software R é bastante importante para o progresso dessas investigações, pois já contém funcionalidades usadas na implementação e na proposição original do método. O objetivo do Pacote gencor é fornecer a capacidade de gerar matrizes de correlação realísticas, com ou sem especificações sobre os níveis de correlação e percentuais de medidas nesses níveis. Uma das principais motivações foi a carência de pacotes R e *software* de código aberto em geral, capazes de gerar matrizes de correlação customizadas pelo usuário.

A flexibilidade de utilização da linguagem R é um item motivador para a construção desse pacote. A automatização do método de geração de matrizes de correlação pode ajudar pesquisadores e profissionais a criar evidências para seus estudos e o seu procedimento de tomada de decisão. Se essa etapa de processamento é ágil e confiável, os pesquisadores podem dedicar mais tempo no seu problema específico e não na concepção de códigos para ferramentas a serem utilizadas.

5.1 Utilização do pacote gencor()

Serão apresentados diversos exemplos de utilização da função gencor(), todos os casos aparecem acompanhados da sintaxe adequada de utilização através do software R. Todos os procedimentos foram realizados em sequência e antes do primeiro procedimento a semente para geração de números “pseudo-aleatórios” através do software R foi fixada através do comando `set.seed(1234)`.

A execução do comando básico sem determinação de parâmetros quaisquer leva à geração de uma matriz de correlação com dimensão 10:

```
R> gencor()
```

```
Generated Matrix
```

```
          [,1]          [,2]          [,3]          [,4]          [,5]
[1,]  1.000000000 -0.053636201  0.468000406 -0.84201308  0.069185793
```

```

[2,] -0.053636201  1.000000000 -0.035222423  0.04199028  0.009373827
[3,]  0.468000406 -0.035222423  1.000000000 -0.41026671  0.048648842
[4,] -0.842013075  0.041990283 -0.410266715  1.00000000  -0.098309776
[5,]  0.069185793  0.009373827  0.048648842 -0.09830978  1.000000000
[6,] -0.010204938  0.050585445 -0.009625704  0.01786045 -0.037206277
[7,]  0.241909704 -0.049437808  0.104019015 -0.20668108 -0.007317307
[8,] -0.299471737 -0.003909536 -0.156266086  0.28482324  0.014555970
[9,] -0.007074192 -0.007820787  0.008817396  0.00325618 -0.002008594
[10,] -0.349705095  0.020368103 -0.159910760  0.28664369  0.001161471
      [,6]      [,7]      [,8]      [,9]     [,10]
[1,] -0.010204938  0.241909704 -0.299471737 -0.007074192 -0.349705095
[2,]  0.050585445 -0.049437808 -0.003909536 -0.007820787  0.020368103
[3,] -0.009625704  0.104019015 -0.156266086  0.008817396 -0.159910760
[4,]  0.017860455 -0.206681075  0.284823237  0.003256180  0.286643692
[5,] -0.037206277 -0.007317307  0.014555970 -0.002008594  0.001161471
[6,]  1.000000000 -0.077441887  0.020676827  0.009449644  0.017989274
[7,] -0.077441887  1.000000000 -0.130563615  0.003460042 -0.078360984
[8,]  0.020676827 -0.130563615  1.000000000 -0.011994554  0.148791481
[9,]  0.009449644  0.003460042 -0.011994554  1.000000000  0.038649997
[10,] 0.017989274 -0.078360984  0.148791481  0.038649997  1.000000000

```

O mesmo procedimento pode ser executado com especificação do parâmetro `method` como segue:

```
R> gencor(method = "random")
```

Generated Matrix

```

      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.00000000 0.04815158 -0.072786835 -0.259043613 0.35385075
[2,] 0.04815158 1.00000000 -0.012368275 -0.049907617 0.09289568
[3,] -0.07278684 -0.01236827 1.000000000 -0.008324524 -0.04429468
[4,] -0.25904361 -0.04990762 -0.008324524 1.000000000 -0.27672802
[5,] 0.35385075 0.09289568 -0.044294680 -0.276728018 1.000000000
[6,] 0.61382154 0.10438763 -0.080113825 -0.499043564 0.57016062
[7,] 0.01514422 -0.01661991 0.019554588 0.058818058 0.03119533
[8,] -0.58421915 -0.07978716 0.087234965 0.464935081 -0.52138570
[9,] -0.53032676 -0.08932683 0.055172466 0.412587239 -0.48976856
[10,] -0.26092846 -0.05178339 0.021745562 0.230277435 -0.28495105
      [,6]      [,7]      [,8]      [,9]     [,10]
[1,] 0.61382154 0.015144215 -0.584219148 -0.53032676 -0.260928459
[2,] 0.10438763 -0.016619906 -0.079787158 -0.08932683 -0.051783390
[3,] -0.08011382 0.019554588 0.087234965 0.05517247 0.021745562
[4,] -0.49904356 0.058818058 0.464935081 0.41258724 0.230277435
[5,] 0.57016062 0.031195332 -0.521385701 -0.48976856 -0.284951047
[6,] 1.000000000 -0.016160128 -0.913842176 -0.81465667 -0.472996083
[7,] -0.01616013 1.000000000 -0.007667281 0.01469851 -0.001211056
[8,] -0.91384218 -0.007667281 1.000000000 0.76465680 0.444669492
[9,] -0.81465667 0.014698511 0.764656796 1.00000000 0.404377586
[10,] -0.47299608 -0.001211056 0.444669492 0.40437759 1.000000000

```

Uma matriz de correlação de outra dimensão pode ser produzida através da especificação do parâmetro `d` como segue:

```
R> gencor(d = 15)
```

```
Generated Matrix
```

| | [,1] | [,2] | [,3] | [,4] | [,5] |
|-------|---------------|--------------|--------------|--------------|---------------|
| [1,] | 1.00000000 | 0.31640787 | 0.019985442 | 0.202366228 | 0.339795366 |
| [2,] | 0.31640787 | 1.00000000 | 0.043851434 | 0.455535365 | 0.717297215 |
| [3,] | 0.01998544 | 0.04385143 | 1.000000000 | 0.031979342 | 0.038747290 |
| [4,] | 0.20236623 | 0.45553537 | 0.031979342 | 1.000000000 | 0.513112965 |
| [5,] | 0.33979537 | 0.71729721 | 0.038747290 | 0.513112965 | 1.000000000 |
| [6,] | 0.16718671 | 0.35674626 | 0.025700887 | 0.246310531 | 0.406467540 |
| [7,] | 0.09239711 | 0.15935881 | -0.020719709 | 0.102726440 | 0.210982532 |
| [8,] | -0.06046142 | -0.06904687 | 0.056654668 | -0.063497001 | -0.103941261 |
| [9,] | -0.14857688 | -0.33836652 | -0.009924318 | -0.244245496 | -0.379376550 |
| [10,] | 0.09458618 | 0.09704572 | 0.005705375 | 0.053139239 | 0.100006541 |
| [11,] | -0.03342247 | 0.03014846 | 0.027674026 | -0.021705184 | -0.004766696 |
| [12,] | -0.01255738 | -0.01460699 | 0.053233984 | -0.003027838 | -0.027151540 |
| [13,] | -0.36453335 | -0.77755616 | -0.055508766 | -0.572198002 | -0.899907235 |
| [14,] | 0.15885168 | 0.36266657 | 0.063861447 | 0.261244704 | 0.391645500 |
| [15,] | 0.36751391 | 0.78735548 | 0.053535666 | 0.575045681 | 0.915295289 |
| | [,6] | [,7] | [,8] | [,9] | [,10] |
| [1,] | 0.16718671 | 0.092397106 | -0.060461416 | -0.148576878 | 0.094586180 |
| [2,] | 0.35674626 | 0.159358812 | -0.069046868 | -0.338366522 | 0.097045719 |
| [3,] | 0.02570089 | -0.020719709 | 0.056654668 | -0.009924318 | 0.005705375 |
| [4,] | 0.24631053 | 0.102726440 | -0.063497001 | -0.244245496 | 0.053139239 |
| [5,] | 0.40646754 | 0.210982532 | -0.103941261 | -0.379376550 | 0.100006541 |
| [6,] | 1.00000000 | 0.126847833 | 0.026757233 | -0.164226399 | 0.093590439 |
| [7,] | 0.12684783 | 1.000000000 | 0.045218000 | -0.071867372 | 0.006887894 |
| [8,] | 0.02675723 | 0.045218000 | 1.000000000 | 0.037168500 | 0.026636647 |
| [9,] | -0.16422640 | -0.071867372 | 0.037168500 | 1.000000000 | -0.025966319 |
| [10,] | 0.09359044 | 0.006887894 | 0.026636647 | -0.025966319 | 1.000000000 |
| [11,] | 0.01494650 | 0.006149877 | 0.018563502 | -0.016715158 | -0.017483259 |
| [12,] | -0.05356201 | 0.046290346 | -0.040959209 | 0.054192023 | -0.002008924 |
| [13,] | -0.43409776 | -0.222600051 | 0.088503988 | 0.389156292 | -0.114539894 |
| [14,] | 0.22828856 | 0.079819477 | -0.003995619 | -0.212199893 | 0.049332675 |
| [15,] | 0.44245145 | 0.223596705 | -0.092246444 | -0.408856775 | 0.123325588 |
| | [,11] | [,12] | [,13] | [,14] | [,15] |
| [1,] | -0.0334224746 | -0.012557383 | -0.364533351 | 0.158851683 | 0.3675139091 |
| [2,] | 0.0301484616 | -0.014606992 | -0.777556162 | 0.362666571 | 0.7873554769 |
| [3,] | 0.0276740258 | 0.053233984 | -0.055508766 | 0.063861447 | 0.0535356656 |
| [4,] | -0.0217051837 | -0.003027838 | -0.572198002 | 0.261244704 | 0.5750456809 |
| [5,] | -0.0047666956 | -0.027151540 | -0.899907235 | 0.391645500 | 0.9152952888 |
| [6,] | 0.0149464975 | -0.053562012 | -0.434097761 | 0.228288556 | 0.4424514464 |
| [7,] | 0.0061498767 | 0.046290346 | -0.222600051 | 0.079819477 | 0.2235967046 |
| [8,] | 0.0185635022 | -0.040959209 | 0.088503988 | -0.003995619 | -0.0922464437 |
| [9,] | -0.0167151577 | 0.054192023 | 0.389156292 | -0.212199893 | -0.4088567749 |
| [10,] | -0.0174832593 | -0.002008924 | -0.114539894 | 0.049332675 | 0.1233255876 |
| [11,] | 1.0000000000 | 0.023957178 | 0.006466623 | -0.027069538 | -0.0008978391 |
| [12,] | 0.0239571785 | 1.000000000 | 0.022648131 | 0.009975348 | -0.0252486589 |
| [13,] | 0.0064666231 | 0.022648131 | 1.000000000 | -0.437029085 | -0.9772766966 |
| [14,] | -0.0270695383 | 0.009975348 | -0.437029085 | 1.000000000 | 0.4459471404 |

```
[15,] -0.0008978391 -0.025248659 -0.977276697 0.445947140 1.000000000
```

O mesmo procedimento anterior pode ser executado com a especificação dos parâmetros `d` e `method` como segue:

```
R> gencor(d = 15, method = "random")
```

Generated Matrix

```

      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.000000000 -0.04475657 -0.04980265 -0.085867610 -0.008333466
[2,] -0.044756566 1.000000000 0.60993937 -0.038977401 0.143113649
[3,] -0.049802652 0.60993937 1.000000000 -0.032696857 0.093790656
[4,] -0.085867610 -0.03897740 -0.03269686 1.000000000 -0.006238106
[5,] -0.008333466 0.14311365 0.09379066 -0.006238106 1.000000000
[6,] 0.005198280 0.31495791 0.26404019 -0.004056238 0.072062765
[7,] -0.077236258 0.40928357 0.32148352 -0.044589517 0.076511455
[8,] 0.048040518 -0.83240693 -0.66892372 0.063376321 -0.159303492
[9,] -0.036029982 -0.01452877 -0.06881218 0.062861463 -0.034725701
[10,] 0.055611484 -0.80229155 -0.64174673 0.069912559 -0.150404343
[11,] 0.085110329 -0.31312589 -0.24871368 0.016590871 -0.004877283
[12,] -0.062941552 0.03751477 0.05493218 0.004280902 0.029744872
[13,] 0.031514146 -0.47127211 -0.38179153 0.013975323 -0.105634418
[14,] 0.013432467 -0.42410018 -0.33105981 0.051312143 -0.080719271
[15,] -0.004688684 -0.33623434 -0.27031941 0.062263091 -0.022791542
      [,6]      [,7]      [,8]      [,9]      [,10]
[1,] 0.005198280 -0.077236258 0.04804052 -0.0360299815 0.055611484
[2,] 0.314957915 0.409283566 -0.83240693 -0.0145287697 -0.80229155
[3,] 0.264040185 0.321483520 -0.66892372 -0.0688121788 -0.64174673
[4,] -0.004056238 -0.044589517 0.06337632 0.0628614629 0.06991256
[5,] 0.072062765 0.076511455 -0.15930349 -0.0347257011 -0.15040434
[6,] 1.000000000 0.150710709 -0.32690519 -0.0315759499 -0.30381266
[7,] 0.150710709 1.000000000 -0.45010942 -0.0352368460 -0.43886124
[8,] -0.326905188 -0.450109415 1.000000000 0.0375864458 0.90226539
[9,] -0.031575950 -0.035236846 0.03758645 1.0000000000 0.03996380
[10,] -0.303812664 -0.438861240 0.90226539 0.0399637970 1.00000000
[11,] -0.082285109 -0.165490187 0.36340658 -0.0007432459 0.36545229
[12,] 0.038813411 -0.007488154 -0.02085172 -0.0043836432 -0.01633885
[13,] -0.193102487 -0.248607115 0.52517238 0.0049823174 0.49523964
[14,] -0.190448069 -0.246192393 0.46624587 0.0055095741 0.47763972
[15,] -0.089903290 -0.194672870 0.40909122 0.0435146130 0.39517290
      [,11]      [,12]      [,13]      [,14]      [,15]
[1,] 0.0851103288 -0.062941552 0.031514146 0.013432467 -0.004688684
[2,] -0.3131258949 0.037514770 -0.471272113 -0.424100183 -0.336234337
[3,] -0.2487136774 0.054932184 -0.381791532 -0.331059808 -0.270319409
[4,] 0.0165908706 0.004280902 0.013975323 0.051312143 0.062263091
[5,] -0.0048772833 0.029744872 -0.105634418 -0.080719271 -0.022791542
[6,] -0.0822851094 0.038813411 -0.193102487 -0.190448069 -0.089903290
[7,] -0.1654901866 -0.007488154 -0.248607115 -0.246192393 -0.194672870
[8,] 0.3634065831 -0.020851724 0.525172382 0.466245873 0.409091215
[9,] -0.0007432459 -0.004383643 0.004982317 0.005509574 0.043514613
[10,] 0.3654522873 -0.016338849 0.495239639 0.477639719 0.395172899

```

```
[11,] 1.000000000 0.026615773 0.222986976 0.196145003 0.195803756
[12,] 0.0266157729 1.000000000 0.027039457 0.044979887 0.044715682
[13,] 0.2229869765 0.027039457 1.000000000 0.244882763 0.251759847
[14,] 0.1961450027 0.044979887 0.244882763 1.000000000 0.188178508
[15,] 0.1958037562 0.044715682 0.251759847 0.188178508 1.000000000
```

Uma matriz de correlação de dimensão 15, com medidas de correlação restritas ao intervalo $[-0,5;0,5]$ pode ser produzida através da especificação dos parâmetros `d`, `method` e `lim_low` como segue:

```
R> gencor(d = 15, method = "low", lim_low = 0.5)
```

Generated Matrix

```

          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.000000000 0.007284129 0.002999055 0.054431018 0.004901336
[2,] 0.007284129 1.000000000 -0.021389155 -0.008184552 -0.049195251
[3,] 0.002999055 -0.021389155 1.000000000 0.077786681 0.057972201
[4,] 0.054431018 -0.008184552 0.077786681 1.000000000 0.060212299
[5,] 0.004901336 -0.049195251 0.057972201 0.060212299 1.000000000
[6,] 0.032806742 -0.092357587 0.104796931 0.105047280 0.096144961
[7,] -0.020364866 -0.047654836 0.022741468 0.018084249 -0.010769381
[8,] -0.026804467 0.045504838 -0.166917397 -0.168884807 -0.074475481
[9,] -0.057698409 0.065033079 -0.154497070 -0.113295497 -0.040046494
[10,] -0.015358730 0.025492706 0.018403744 0.053363376 0.003634369
[11,] -0.072922424 0.060144156 -0.107488943 -0.094886831 -0.021398221
[12,] 0.071468161 -0.095756201 0.119074663 0.083819146 0.023184437
[13,] 0.038802257 0.008824660 0.056011253 0.011613310 -0.005605738
[14,] 0.089475903 -0.015082222 0.081144435 0.111144840 0.069338104
[15,] -0.016317073 -0.055390184 -0.017018917 0.007719759 0.010277412
          [,6]      [,7]      [,8]      [,9]      [,10]
[1,] 0.03280674 -0.02036487 -0.02680447 -0.05769841 -0.015358730
[2,] -0.09235759 -0.04765484 0.04550484 0.06503308 0.025492706
[3,] 0.10479693 0.02274147 -0.16691740 -0.15449707 0.018403744
[4,] 0.10504728 0.01808425 -0.16888481 -0.11329550 0.053363376
[5,] 0.09614496 -0.01076938 -0.07447548 -0.04004649 0.003634369
[6,] 1.00000000 0.06822349 -0.18664172 -0.15992335 0.059926138
[7,] 0.06822349 1.00000000 -0.11141945 -0.03403126 0.035100118
[8,] -0.18664172 -0.11141945 1.00000000 0.16916183 -0.094772548
[9,] -0.15992335 -0.03403126 0.16916183 1.00000000 -0.056180142
[10,] 0.05992614 0.03510012 -0.09477255 -0.05618014 1.000000000
[11,] -0.12217423 -0.04652230 0.23130894 0.15140990 -0.048909719
[12,] 0.10390791 0.04525346 -0.17039468 -0.07315337 0.013194294
[13,] 0.06111811 0.05248507 -0.05652878 -0.07956008 0.010638385
[14,] 0.07541630 0.02657407 -0.08111261 -0.04032610 0.029721167
[15,] 0.03194410 0.03020546 0.00919204 0.02110702 -0.028333205
          [,11]      [,12]      [,13]      [,14]      [,15]
[1,] -0.072922424 0.07146816 0.038802257 0.08947590 -0.016317073
[2,] 0.060144156 -0.09575620 0.008824660 -0.01508222 -0.055390184
[3,] -0.107488943 0.11907466 0.056011253 0.08114443 -0.017018917
[4,] -0.094886831 0.08381915 0.011613310 0.11114484 0.007719759
```

```

[5,] -0.021398221  0.02318444 -0.005605738  0.06933810  0.010277412
[6,] -0.122174232  0.10390791  0.061118111  0.07541630  0.031944102
[7,] -0.046522304  0.04525346  0.052485067  0.02657407  0.030205464
[8,]  0.231308945 -0.17039468 -0.056528781 -0.08111261  0.009192040
[9,]  0.151409898 -0.07315337 -0.079560084 -0.04032610  0.021107021
[10,] -0.048909719  0.01319429  0.010638385  0.02972117 -0.028333205
[11,]  1.000000000 -0.11333523 -0.016427245 -0.06089118  0.001490003
[12,] -0.113335234  1.00000000  0.101715617  0.05134070 -0.029004883
[13,] -0.016427245  0.10171562  1.000000000  0.03910378 -0.010578040
[14,] -0.060891178  0.05134070  0.039103777  1.00000000  0.025503671
[15,]  0.001490003 -0.02900488 -0.010578040  0.02550367  1.000000000

```

Uma matriz de correlação de dimensão 15, com medidas de correlação restritas ao intervalo $(-0,7; -0,3) \cup (0,3; 0,7)$ pode ser produzida através da especificação dos parâmetros `d`, `method`, `lim_low` e `lim_medium` como segue:

```
R> gencor(d = 15, method = "medium", lim_low = 0.3, lim_medium = 0.7)
```

Generated Matrix

```

          [,1]      [,2]      [,3]      [,4]      [,5]
[1,]  1.0000000  0.3398226 -0.3268535  0.3337473  0.4245043
[2,]  0.3398226  1.0000000 -0.3507868  0.3553574  0.4141074
[3,] -0.3268535 -0.3507868  1.0000000 -0.3542065 -0.4321948
[4,]  0.3337473  0.3553574 -0.3542065  1.0000000  0.4288170
[5,]  0.4245043  0.4141074 -0.4321948  0.4288170  1.0000000
[6,]  0.2887903  0.3106612 -0.3528603  0.3611993  0.3903427
[7,] -0.4367409 -0.4554947  0.4636289 -0.5091261 -0.5501535
[8,] -0.3207860 -0.3620017  0.4021087 -0.3796219 -0.4501290
[9,] -0.3154015 -0.3099797  0.3961100 -0.3226069 -0.4192044
[10,] -0.3817374 -0.4393799  0.4657351 -0.4428300 -0.5027103
[11,]  0.4265867  0.4646251 -0.5159471  0.4991705  0.5331165
[12,]  0.4203007  0.4506620 -0.5006308  0.4464460  0.5290928
[13,]  0.3325137  0.3284521 -0.3919820  0.3391904  0.3806582
[14,]  0.3545250  0.3927760 -0.4537000  0.4491250  0.4718191
[15,] -0.3303432 -0.3460673  0.3791168 -0.3622185 -0.4241134
          [,6]      [,7]      [,8]      [,9]      [,10]
[1,]  0.2887903 -0.4367409 -0.3207860 -0.3154015 -0.3817374
[2,]  0.3106612 -0.4554947 -0.3620017 -0.3099797 -0.4393799
[3,] -0.3528603  0.4636289  0.4021087  0.3961100  0.4657351
[4,]  0.3611993 -0.5091261 -0.3796219 -0.3226069 -0.4428300
[5,]  0.3903427 -0.5501535 -0.4501290 -0.4192044 -0.5027103
[6,]  1.0000000 -0.4335037 -0.3766860 -0.3003193 -0.3728778
[7,] -0.4335037  1.0000000  0.4960079  0.4198929  0.5590979
[8,] -0.3766860  0.4960079  1.0000000  0.3684406  0.4297567
[9,] -0.3003193  0.4198929  0.3684406  1.0000000  0.4257020
[10,] -0.3728778  0.5590979  0.4297567  0.4257020  1.0000000
[11,]  0.4240669 -0.6104328 -0.5185849 -0.4365377 -0.5679992
[12,]  0.4256730 -0.5555661 -0.4755868 -0.4350259 -0.5270472
[13,]  0.3220445 -0.4267109 -0.3841448 -0.3419279 -0.3831333
[14,]  0.3964939 -0.5173005 -0.4426785 -0.3813877 -0.4654951

```

```

[15,] -0.3387218  0.4513906  0.4065570  0.3557235  0.4305820
      [,11]      [,12]      [,13]      [,14]      [,15]
[1,]  0.4265867  0.4203007  0.3325137  0.3545250 -0.3303432
[2,]  0.4646251  0.4506620  0.3284521  0.3927760 -0.3460673
[3,] -0.5159471 -0.5006308 -0.3919820 -0.4537000  0.3791168
[4,]  0.4991705  0.4464460  0.3391904  0.4491250 -0.3622185
[5,]  0.5331165  0.5290928  0.3806582  0.4718191 -0.4241134
[6,]  0.4240669  0.4256730  0.3220445  0.3964939 -0.3387218
[7,] -0.6104328 -0.5555661 -0.4267109 -0.5173005  0.4513906
[8,] -0.5185849 -0.4755868 -0.3841448 -0.4426785  0.4065570
[9,] -0.4365377 -0.4350259 -0.3419279 -0.3813877  0.3557235
[10,] -0.5679992 -0.5270472 -0.3831333 -0.4654951  0.4305820
[11,]  1.0000000  0.5981219  0.4202314  0.5361927 -0.5017613
[12,]  0.5981219  1.0000000  0.3823948  0.5052892 -0.4189860
[13,]  0.4202314  0.3823948  1.0000000  0.3716720 -0.2898962
[14,]  0.5361927  0.5052892  0.3716720  1.0000000 -0.4276703
[15,] -0.5017613 -0.4189860 -0.2898962 -0.4276703  1.0000000

```

Uma matriz de correlação de dimensão 30, com medidas de correlação restritas ao intervalo $(-1; -0,75) \cup (0,75; 1)$ pode ser produzida através da especificação dos parâmetros `d`, `method` e `lim_medium` como segue:

```
R> gencor(d = 30, method = "high", lim_medium = 0.75)
```

Generated Matrix

```

      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,]  1.0000000  0.7810099  0.7827667  0.9022679  0.7834227  0.9032560
[2,]  0.7810099  1.0000000  0.7508270  0.8804749  0.7726832  0.8827153
[3,]  0.7827667  0.7508270  1.0000000  0.8562212  0.7521673  0.8590370
[4,]  0.9022679  0.8804749  0.8562212  1.0000000  0.8696898  0.9966990
[5,]  0.7834227  0.7726832  0.7521673  0.8696898  1.0000000  0.8720608
[6,]  0.9032560  0.8827153  0.8590370  0.9966990  0.8720608  1.0000000
[7,] -0.9016135 -0.8802590 -0.8555173 -0.9939735 -0.8709143 -0.9943240
[8,]  0.8901674  0.8675095  0.8489811  0.9807508  0.8515596  0.9807597
[9,] -0.8698599 -0.8471032 -0.8300554 -0.9611705 -0.8395109 -0.9621008
[10,] -0.8931783 -0.8729453 -0.8473705 -0.9858793 -0.8578789 -0.9867383
[11,] -0.9030478 -0.8821334 -0.8579394 -0.9961484 -0.8713391 -0.9966977
[12,]  0.8944352  0.8738751  0.8530404  0.9855849  0.8588501  0.9858740
[13,] -0.8831745 -0.8596486 -0.8387098 -0.9754507 -0.8522465 -0.9760288
[14,]  0.8933801  0.8640713  0.8401986  0.9745419  0.8587197  0.9770163
[15,]  0.9052064  0.8827884  0.8577866  0.9972851  0.8715603  0.9978835
[16,] -0.8803222 -0.8612656 -0.8415849 -0.9722936 -0.8471883 -0.9730196
[17,] -0.8321771 -0.8121417 -0.7738047 -0.9110848 -0.8106425 -0.9125477
[18,]  0.8230624  0.8019065  0.7820256  0.9081290  0.7874001  0.9072316
[19,] -0.8998836 -0.8826263 -0.8574691 -0.9920074 -0.8711878 -0.9922691
[20,] -0.7882924 -0.7822746 -0.7472083 -0.8734552 -0.7571975 -0.8744280
[21,]  0.8488022  0.8280615  0.8133462  0.9384166  0.8214247  0.9386989
[22,]  0.8380071  0.8092799  0.8026334  0.9178706  0.8018398  0.9194447
[23,] -0.8761091 -0.8472792 -0.8360901 -0.9685123 -0.8526907 -0.9685798
[24,] -0.8746970 -0.8566894 -0.8382224 -0.9659286 -0.8412098 -0.9666837

```

| | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|
| [25,] | 0.8244390 | 0.7926085 | 0.7714743 | 0.9049996 | 0.7889331 | 0.9060253 |
| [26,] | 0.8989303 | 0.8798661 | 0.8527539 | 0.9924037 | 0.8666749 | 0.9930275 |
| [27,] | -0.8542461 | -0.8356055 | -0.8196112 | -0.9490289 | -0.8353212 | -0.9478795 |
| [28,] | 0.8995273 | 0.8778016 | 0.8573091 | 0.9934113 | 0.8682647 | 0.9938541 |
| [29,] | 0.8769277 | 0.8507627 | 0.8321334 | 0.9550863 | 0.8336138 | 0.9549524 |
| [30,] | -0.9031394 | -0.8818355 | -0.8544073 | -0.9943021 | -0.8672489 | -0.9949017 |
| | [,7] | [,8] | [,9] | [,10] | [,11] | [,12] |
| [1,] | -0.9016135 | 0.8901674 | -0.8698599 | -0.8931783 | -0.9030478 | 0.8944352 |
| [2,] | -0.8802590 | 0.8675095 | -0.8471032 | -0.8729453 | -0.8821334 | 0.8738751 |
| [3,] | -0.8555173 | 0.8489811 | -0.8300554 | -0.8473705 | -0.8579394 | 0.8530404 |
| [4,] | -0.9939735 | 0.9807508 | -0.9611705 | -0.9858793 | -0.9961484 | 0.9855849 |
| [5,] | -0.8709143 | 0.8515596 | -0.8395109 | -0.8578789 | -0.8713391 | 0.8588501 |
| [6,] | -0.9943240 | 0.9807597 | -0.9621008 | -0.9867383 | -0.9966977 | 0.9858740 |
| [7,] | 1.0000000 | -0.9785173 | 0.9591594 | 0.9837671 | 0.9939189 | -0.9836292 |
| [8,] | -0.9785173 | 1.0000000 | -0.9447334 | -0.9714587 | -0.9807646 | 0.9705329 |
| [9,] | 0.9591594 | -0.9447334 | 1.0000000 | 0.9518234 | 0.9617999 | -0.9508317 |
| [10,] | 0.9837671 | -0.9714587 | 0.9518234 | 1.0000000 | 0.9863499 | -0.9742657 |
| [11,] | 0.9939189 | -0.9807646 | 0.9617999 | 0.9863499 | 1.0000000 | -0.9852185 |
| [12,] | -0.9836292 | 0.9705329 | -0.9508317 | -0.9742657 | -0.9852185 | 1.0000000 |
| [13,] | 0.9726105 | -0.9596257 | 0.9388928 | 0.9662985 | 0.9748480 | -0.9633179 |
| [14,] | -0.9738861 | 0.9621952 | -0.9388380 | -0.9655536 | -0.9753669 | 0.9660061 |
| [15,] | -0.9952993 | 0.9819742 | -0.9624636 | -0.9873170 | -0.9973606 | 0.9865845 |
| [16,] | 0.9697317 | -0.9571004 | 0.9360249 | 0.9637547 | 0.9721699 | -0.9593634 |
| [17,] | 0.9103806 | -0.9013322 | 0.8862912 | 0.9055938 | 0.9124584 | -0.9008516 |
| [18,] | -0.9052084 | 0.8904552 | -0.8781960 | -0.8938339 | -0.9063590 | 0.8993450 |
| [19,] | 0.9897335 | -0.9764679 | 0.9559214 | 0.9824788 | 0.9924200 | -0.9812892 |
| [20,] | 0.8726664 | -0.8564482 | 0.8496103 | 0.8680185 | 0.8738096 | -0.8642568 |
| [21,] | -0.9371023 | 0.9242098 | -0.9050888 | -0.9266434 | -0.9374653 | 0.9268561 |
| [22,] | -0.9160730 | 0.9034654 | -0.8844485 | -0.9074364 | -0.9180146 | 0.9068276 |
| [23,] | 0.9656782 | -0.9526355 | 0.9335634 | 0.9584090 | 0.9676552 | -0.9577539 |
| [24,] | 0.9625566 | -0.9530551 | 0.9310353 | 0.9567510 | 0.9663521 | -0.9567789 |
| [25,] | -0.9059470 | 0.8864192 | -0.8753799 | -0.8985360 | -0.9072504 | 0.8955268 |
| [26,] | -0.9902185 | 0.9773294 | -0.9565650 | -0.9821315 | -0.9925723 | 0.9822090 |
| [27,] | 0.9475790 | -0.9312371 | 0.9167761 | 0.9387945 | 0.9485499 | -0.9389246 |
| [28,] | -0.9917651 | 0.9769907 | -0.9597600 | -0.9823222 | -0.9931317 | 0.9826419 |
| [29,] | -0.9532549 | 0.9403751 | -0.9187374 | -0.9460252 | -0.9553396 | 0.9443161 |
| [30,] | 0.9925109 | -0.9787288 | 0.9588183 | 0.9845317 | 0.9945150 | -0.9835366 |
| | [,13] | [,14] | [,15] | [,16] | [,17] | [,18] |
| [1,] | -0.8831745 | 0.8933801 | 0.9052064 | -0.8803222 | -0.8321771 | 0.8230624 |
| [2,] | -0.8596486 | 0.8640713 | 0.8827884 | -0.8612656 | -0.8121417 | 0.8019065 |
| [3,] | -0.8387098 | 0.8401986 | 0.8577866 | -0.8415849 | -0.7738047 | 0.7820256 |
| [4,] | -0.9754507 | 0.9745419 | 0.9972851 | -0.9722936 | -0.9110848 | 0.9081290 |
| [5,] | -0.8522465 | 0.8587197 | 0.8715603 | -0.8471883 | -0.8106425 | 0.7874001 |
| [6,] | -0.9760288 | 0.9770163 | 0.9978835 | -0.9730196 | -0.9125477 | 0.9072316 |
| [7,] | 0.9726105 | -0.9738861 | -0.9952993 | 0.9697317 | 0.9103806 | -0.9052084 |
| [8,] | -0.9596257 | 0.9621952 | 0.9819742 | -0.9571004 | -0.9013322 | 0.8904552 |
| [9,] | 0.9388928 | -0.9388380 | -0.9624636 | 0.9360249 | 0.8862912 | -0.8781960 |
| [10,] | 0.9662985 | -0.9655536 | -0.9873170 | 0.9637547 | 0.9055938 | -0.8938339 |
| [11,] | 0.9748480 | -0.9753669 | -0.9973606 | 0.9721699 | 0.9124584 | -0.9063590 |
| [12,] | -0.9633179 | 0.9660061 | 0.9865845 | -0.9593634 | -0.9008516 | 0.8993450 |
| [13,] | 1.0000000 | -0.9551642 | -0.9764943 | 0.9523610 | 0.8949073 | -0.8899892 |

| | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|
| [14,] | -0.9551642 | 1.0000000 | 0.9775380 | -0.9516425 | -0.8976331 | 0.8910670 |
| [15,] | -0.9764943 | 0.9775380 | 1.0000000 | -0.9734503 | -0.9136054 | 0.9089701 |
| [16,] | 0.9523610 | -0.9516425 | -0.9734503 | 1.0000000 | 0.8948183 | -0.8835121 |
| [17,] | 0.8949073 | -0.8976331 | -0.9136054 | 0.8948183 | 1.0000000 | -0.8303864 |
| [18,] | -0.8899892 | 0.8910670 | 0.9089701 | -0.8835121 | -0.8303864 | 1.0000000 |
| [19,] | 0.9712509 | -0.9715387 | -0.9934306 | 0.9692988 | 0.9105015 | -0.9044980 |
| [20,] | 0.8519491 | -0.8577159 | -0.8749871 | 0.8521710 | 0.8097855 | -0.7978923 |
| [21,] | -0.9177625 | 0.9223487 | 0.9398395 | -0.9143336 | -0.8560643 | 0.8509643 |
| [22,] | -0.9021038 | 0.9002885 | 0.9185144 | -0.8975996 | -0.8433664 | 0.8393532 |
| [23,] | 0.9465117 | -0.9470455 | -0.9685714 | 0.9425459 | 0.8850976 | -0.8835124 |
| [24,] | 0.9478375 | -0.9445947 | -0.9673530 | 0.9393701 | 0.8921209 | -0.8812989 |
| [25,] | -0.8884535 | 0.8927090 | 0.9078954 | -0.8824985 | -0.8341188 | 0.8275361 |
| [26,] | -0.9732736 | 0.9717003 | 0.9936223 | -0.9682257 | -0.9103727 | 0.9020469 |
| [27,] | 0.9271258 | -0.9296868 | -0.9495897 | 0.9267131 | 0.8642588 | -0.8636764 |
| [28,] | -0.9723207 | 0.9723361 | 0.9949058 | -0.9692905 | -0.9117307 | 0.9056812 |
| [29,] | -0.9326713 | 0.9362131 | 0.9561189 | -0.9316243 | -0.8753542 | 0.8629441 |
| [30,] | 0.9729421 | -0.9742963 | -0.9957405 | 0.9710369 | 0.9132452 | -0.9032929 |
| | [,19] | [,20] | [,21] | [,22] | [,23] | [,24] |
| [1,] | -0.8998836 | -0.7882924 | 0.8488022 | 0.8380071 | -0.8761091 | -0.8746970 |
| [2,] | -0.8826263 | -0.7822746 | 0.8280615 | 0.8092799 | -0.8472792 | -0.8566894 |
| [3,] | -0.8574691 | -0.7472083 | 0.8133462 | 0.8026334 | -0.8360901 | -0.8382224 |
| [4,] | -0.9920074 | -0.8734552 | 0.9384166 | 0.9178706 | -0.9685123 | -0.9659286 |
| [5,] | -0.8711878 | -0.7571975 | 0.8214247 | 0.8018398 | -0.8526907 | -0.8412098 |
| [6,] | -0.9922691 | -0.8744280 | 0.9386989 | 0.9194447 | -0.9685798 | -0.9666837 |
| [7,] | 0.9897335 | 0.8726664 | -0.9371023 | -0.9160730 | 0.9656782 | 0.9625566 |
| [8,] | -0.9764679 | -0.8564482 | 0.9242098 | 0.9034654 | -0.9526355 | -0.9530551 |
| [9,] | 0.9559214 | 0.8496103 | -0.9050888 | -0.8844485 | 0.9335634 | 0.9310353 |
| [10,] | 0.9824788 | 0.8680185 | -0.9266434 | -0.9074364 | 0.9584090 | 0.9567510 |
| [11,] | 0.9924200 | 0.8738096 | -0.9374653 | -0.9180146 | 0.9676552 | 0.9663521 |
| [12,] | -0.9812892 | -0.8642568 | 0.9268561 | 0.9068276 | -0.9577539 | -0.9567789 |
| [13,] | 0.9712509 | 0.8519491 | -0.9177625 | -0.9021038 | 0.9465117 | 0.9478375 |
| [14,] | -0.9715387 | -0.8577159 | 0.9223487 | 0.9002885 | -0.9470455 | -0.9445947 |
| [15,] | -0.9934306 | -0.8749871 | 0.9398395 | 0.9185144 | -0.9685714 | -0.9673530 |
| [16,] | 0.9692988 | 0.8521710 | -0.9143336 | -0.8975996 | 0.9425459 | 0.9393701 |
| [17,] | 0.9105015 | 0.8097855 | -0.8560643 | -0.8433664 | 0.8850976 | 0.8921209 |
| [18,] | -0.9044980 | -0.7978923 | 0.8509643 | 0.8393532 | -0.8835124 | -0.8812989 |
| [19,] | 1.0000000 | 0.8698813 | -0.9334024 | -0.9120048 | 0.9629863 | 0.9622416 |
| [20,] | 0.8698813 | 1.0000000 | -0.8234975 | -0.8132777 | 0.8452609 | 0.8482455 |
| [21,] | -0.9334024 | -0.8234975 | 1.0000000 | 0.8614653 | -0.9075101 | -0.9076367 |
| [22,] | -0.9120048 | -0.8132777 | 0.8614653 | 1.0000000 | -0.8893142 | -0.8927947 |
| [23,] | 0.9629863 | 0.8452609 | -0.9075101 | -0.8893142 | 1.0000000 | 0.9366443 |
| [24,] | 0.9622416 | 0.8482455 | -0.9076367 | -0.8927947 | 0.9366443 | 1.0000000 |
| [25,] | -0.9018422 | -0.7983508 | 0.8546231 | 0.8407709 | -0.8824133 | -0.8751308 |
| [26,] | -0.9884961 | -0.8720566 | 0.9346535 | 0.9140881 | -0.9626464 | -0.9628226 |
| [27,] | 0.9446280 | 0.8252007 | -0.8955880 | -0.8674414 | 0.9209027 | 0.9216633 |
| [28,] | -0.9889341 | -0.8732194 | 0.9355599 | 0.9138024 | -0.9646350 | -0.9633351 |
| [29,] | -0.9520518 | -0.8360794 | 0.9024741 | 0.8827849 | -0.9276546 | -0.9262718 |
| [30,] | 0.9907003 | 0.8727518 | -0.9368252 | -0.9153372 | 0.9651698 | 0.9647412 |
| | [,25] | [,26] | [,27] | [,28] | [,29] | [,30] |
| [1,] | 0.8244390 | 0.8989303 | -0.8542461 | 0.8995273 | 0.8769277 | -0.9031394 |
| [2,] | 0.7926085 | 0.8798661 | -0.8356055 | 0.8778016 | 0.8507627 | -0.8818355 |

```

[3,]  0.7714743  0.8527539 -0.8196112  0.8573091  0.8321334 -0.8544073
[4,]  0.9049996  0.9924037 -0.9490289  0.9934113  0.9550863 -0.9943021
[5,]  0.7889331  0.8666749 -0.8353212  0.8682647  0.8336138 -0.8672489
[6,]  0.9060253  0.9930275 -0.9478795  0.9938541  0.9549524 -0.9949017
[7,] -0.9059470 -0.9902185  0.9475790 -0.9917651 -0.9532549  0.9925109
[8,]  0.8864192  0.9773294 -0.9312371  0.9769907  0.9403751 -0.9787288
[9,] -0.8753799 -0.9565650  0.9167761 -0.9597600 -0.9187374  0.9588183
[10,] -0.8985360 -0.9821315  0.9387945 -0.9823222 -0.9460252  0.9845317
[11,] -0.9072504 -0.9925723  0.9485499 -0.9931317 -0.9553396  0.9945150
[12,]  0.8955268  0.9822090 -0.9389246  0.9826419  0.9443161 -0.9835366
[13,] -0.8884535 -0.9732736  0.9271258 -0.9723207 -0.9326713  0.9729421
[14,]  0.8927090  0.9717003 -0.9296868  0.9723361  0.9362131 -0.9742963
[15,]  0.9078954  0.9936223 -0.9495897  0.9949058  0.9561189 -0.9957405
[16,] -0.8824985 -0.9682257  0.9267131 -0.9692905 -0.9316243  0.9710369
[17,] -0.8341188 -0.9103727  0.8642588 -0.9117307 -0.8753542  0.9132452
[18,]  0.8275361  0.9020469 -0.8636764  0.9056812  0.8629441 -0.9032929
[19,] -0.9018422 -0.9884961  0.9446280 -0.9889341 -0.9520518  0.9907003
[20,] -0.7983508 -0.8720566  0.8252007 -0.8732194 -0.8360794  0.8727518
[21,]  0.8546231  0.9346535 -0.8955880  0.9355599  0.9024741 -0.9368252
[22,]  0.8407709  0.9140881 -0.8674414  0.9138024  0.8827849 -0.9153372
[23,] -0.8824133 -0.9626464  0.9209027 -0.9646350 -0.9276546  0.9651698
[24,] -0.8751308 -0.9628226  0.9216633 -0.9633351 -0.9262718  0.9647412
[25,]  1.0000000  0.9037666 -0.8595722  0.9055985  0.8716416 -0.9044559
[26,]  0.9037666  1.0000000 -0.9438316  0.9896275  0.9514184 -0.9908100
[27,] -0.8595722 -0.9438316  1.0000000 -0.9474479 -0.9061833  0.9464344
[28,]  0.9055985  0.9896275 -0.9474479  1.0000000  0.9504569 -0.9913206
[29,]  0.8716416  0.9514184 -0.9061833  0.9504569  1.0000000 -0.9541044
[30,] -0.9044559 -0.9908100  0.9464344 -0.9913206 -0.9541044  1.0000000

```

Uma matriz de correlação de dimensão 20, com 30% das medidas de correlação restritas ao intervalo $(-0,2; -0,2)$, com 30% das medidas de correlação restritas ao intervalo $(-0,5; -0,2) \cup (0,2; 0,5)$, com 20% das medidas de correlação restritas ao intervalo $(-0,8; -0,5) \cup (0,5; 0,8)$ e com 20% das medidas de correlação restritas ao intervalo $(-1; -0,8) \cup (0,8; 1)$ pode ser produzida através da especificação dos parâmetros `d`, `method`, `custom_lim` e `custom_prop` como segue:

```
R> gencor(d = 20, method = "custom", custom_lim = c(0.2, 0.5,
+ 0.8), custom_prop = c(0.3, 0.3, 0.2, 0.2))
```

Generated Matrix

```

          [,1]      [,2]      [,3]      [,4]      [,5]
[1,]  1.0000000  0.02210416  0.02109844  0.01935097  0.04857111
[2,]  0.02210416  1.00000000 -0.42733863 -0.05351704 -0.35384178
[3,]  0.02109844 -0.42733863  1.00000000  0.01273470  0.30016136
[4,]  0.01935097 -0.05351704  0.01273470  1.00000000  0.04560603
[5,]  0.04857111 -0.35384178  0.30016136  0.04560603  1.00000000
[6,]  0.02257483 -0.71015746  0.59338484  0.06410992  0.49605874
[7,] -0.01967226  0.70613277 -0.58656111 -0.06800368 -0.49838079
[8,]  0.02403446 -0.70780901  0.58954691  0.06689320  0.48976320

```

| | | | | | |
|-------|--------------|--------------|--------------|--------------|-------------|
| [9,] | 0.02251618 | -0.71194297 | 0.59248279 | 0.06414703 | 0.49608519 |
| [10,] | 0.03127248 | -0.68033912 | 0.56425944 | 0.07043806 | 0.46269068 |
| [11,] | -0.01899677 | 0.70999953 | -0.59396591 | -0.06450936 | -0.49523980 |
| [12,] | -0.01311497 | 0.67284509 | -0.56344905 | -0.05196070 | -0.46606685 |
| [13,] | -0.01495589 | 0.51910409 | -0.43025986 | -0.06657027 | -0.32842614 |
| [14,] | 0.01737884 | 0.04477609 | -0.05078379 | -0.05430235 | 0.01359478 |
| [15,] | -0.01037398 | 0.40029661 | -0.30329962 | 0.04086748 | -0.23846491 |
| [16,] | -0.02351759 | 0.35003582 | -0.29156868 | -0.04690721 | -0.27376493 |
| [17,] | -0.01867695 | 0.61124159 | -0.50026957 | -0.02541687 | -0.41924852 |
| [18,] | -0.02141769 | 0.71035002 | -0.59448533 | -0.06431152 | -0.49427107 |
| [19,] | -0.02186981 | 0.70947686 | -0.59304125 | -0.06225369 | -0.49530205 |
| [20,] | -0.01494213 | -0.31098797 | 0.23338194 | 0.07897320 | 0.23544003 |
| | [,6] | [,7] | [,8] | [,9] | [,10] |
| [1,] | 0.022574833 | -0.019672261 | 0.024034463 | 0.02251618 | 0.03127248 |
| [2,] | -0.710157458 | 0.706132770 | -0.707809009 | -0.71194297 | -0.68033912 |
| [3,] | 0.593384836 | -0.586561110 | 0.589546912 | 0.59248279 | 0.56425944 |
| [4,] | 0.064109917 | -0.068003680 | 0.066893199 | 0.06414703 | 0.07043806 |
| [5,] | 0.496058738 | -0.498380793 | 0.489763203 | 0.49608519 | 0.46269068 |
| [6,] | 1.000000000 | -0.989591945 | 0.994246429 | 0.99762502 | 0.94270669 |
| [7,] | -0.989591945 | 1.000000000 | -0.985650766 | -0.98844042 | -0.93377609 |
| [8,] | 0.994246429 | -0.985650766 | 1.000000000 | 0.99311074 | 0.94006218 |
| [9,] | 0.997625018 | -0.988440421 | 0.993110743 | 1.000000000 | 0.94051036 |
| [10,] | 0.942706692 | -0.933776094 | 0.940062178 | 0.94051036 | 1.000000000 |
| [11,] | -0.998507451 | 0.989545087 | -0.993961294 | -0.99710252 | -0.94114783 |
| [12,] | -0.950659478 | 0.943315682 | -0.946368686 | -0.95051352 | -0.89962952 |
| [13,] | -0.709559035 | 0.698879177 | -0.703315903 | -0.70800521 | -0.65988691 |
| [14,] | 0.004771566 | -0.005022924 | 0.003191447 | 0.00215922 | -0.01604667 |
| [15,] | -0.518006524 | 0.514888735 | -0.514330491 | -0.51737712 | -0.48522478 |
| [16,] | -0.493123119 | 0.483794272 | -0.488752059 | -0.49348950 | -0.48500329 |
| [17,] | -0.831593973 | 0.819432429 | -0.829868120 | -0.83256837 | -0.79445903 |
| [18,] | -0.998712626 | 0.989474924 | -0.994239262 | -0.99747751 | -0.94140264 |
| [19,] | -0.997512187 | 0.988207370 | -0.992548840 | -0.99621968 | -0.94040652 |
| [20,] | 0.440197034 | -0.443397712 | 0.439694681 | 0.43681922 | 0.42432690 |
| | [,11] | [,12] | [,13] | [,14] | [,15] |
| [1,] | -0.018996773 | -0.013114967 | -0.01495589 | 0.017378841 | -0.01037398 |
| [2,] | 0.709999529 | 0.672845090 | 0.51910409 | 0.044776087 | 0.40029661 |
| [3,] | -0.593965909 | -0.563449045 | -0.43025986 | -0.050783794 | -0.30329962 |
| [4,] | -0.064509358 | -0.051960705 | -0.06657027 | -0.054302352 | 0.04086748 |
| [5,] | -0.495239799 | -0.466066851 | -0.32842614 | 0.013594785 | -0.23846491 |
| [6,] | -0.998507451 | -0.950659478 | -0.70955903 | 0.004771566 | -0.51800652 |
| [7,] | 0.989545087 | 0.943315682 | 0.69887918 | -0.005022924 | 0.51488873 |
| [8,] | -0.993961294 | -0.946368686 | -0.70331590 | 0.003191447 | -0.51433049 |
| [9,] | -0.997102516 | -0.950513521 | -0.70800521 | 0.002159220 | -0.51737712 |
| [10,] | -0.941147832 | -0.899629522 | -0.65988691 | -0.016046674 | -0.48522478 |
| [11,] | 1.000000000 | 0.951346222 | 0.70928926 | -0.005577118 | 0.51819668 |
| [12,] | 0.951346222 | 1.000000000 | 0.67007654 | 0.001358622 | 0.48831972 |
| [13,] | 0.709289256 | 0.670076537 | 1.000000000 | -0.019285560 | 0.40310821 |
| [14,] | -0.005577118 | 0.001358622 | -0.01928556 | 1.000000000 | 0.02600161 |
| [15,] | 0.518196677 | 0.488319718 | 0.40310821 | 0.026001608 | 1.000000000 |
| [16,] | 0.494146668 | 0.470847787 | 0.35879967 | -0.009335888 | 0.24895451 |
| [17,] | 0.833710513 | 0.790419604 | 0.57392958 | -0.009009068 | 0.40970110 |

```

[18,] 0.998455880 0.951370708 0.70894496 -0.002789772 0.51784422
[19,] 0.997128321 0.950266626 0.71020716 -0.004492407 0.51391467
[20,] -0.439107993 -0.419937903 -0.28550099 -0.049661818 -0.25256624
      [,16]      [,17]      [,18]      [,19]      [,20]
[1,] -0.023517594 -0.018676947 -0.021417692 -0.021869805 -0.01494213
[2,] 0.350035818 0.611241586 0.710350018 0.709476861 -0.31098797
[3,] -0.291568680 -0.500269570 -0.594485334 -0.593041254 0.23338194
[4,] -0.046907209 -0.025416874 -0.064311525 -0.062253693 0.07897320
[5,] -0.273764927 -0.419248518 -0.494271067 -0.495302046 0.23544003
[6,] -0.493123119 -0.831593973 -0.998712626 -0.997512187 0.44019703
[7,] 0.483794272 0.819432429 0.989474924 0.988207370 -0.44339771
[8,] -0.488752059 -0.829868120 -0.994239262 -0.992548840 0.43969468
[9,] -0.493489501 -0.832568370 -0.997477513 -0.996219681 0.43681922
[10,] -0.485003290 -0.794459029 -0.941402639 -0.940406517 0.42432690
[11,] 0.494146668 0.833710513 0.998455880 0.997128321 -0.43910799
[12,] 0.470847787 0.790419604 0.951370708 0.950266626 -0.41993790
[13,] 0.358799670 0.573929583 0.708944964 0.710207162 -0.28550099
[14,] -0.009335888 -0.009009068 -0.002789772 -0.004492407 -0.04966182
[15,] 0.248954506 0.409701097 0.517844217 0.513914674 -0.25256624
[16,] 1.000000000 0.380360175 0.493629131 0.492294467 -0.21213287
[17,] 0.380360175 1.000000000 0.832606135 0.832764163 -0.37418750
[18,] 0.493629131 0.832606135 1.000000000 0.997385322 -0.43792179
[19,] 0.492294467 0.832764163 0.997385322 1.000000000 -0.43394920
[20,] -0.212132866 -0.374187499 -0.437921791 -0.433949197 1.000000000

```

Além da função principal, foram desenvolvidas duas funções auxiliares, `plot.gencor` e também `summary.gencor`, que podem ser executadas por meio da execução dos códigos básicos `plot` e `summary`, uma vez que são direcionadas para objetos da classe `gencor`. Um exemplo da função `plot` pode ser observado na figura 5.1.1. gerada através das diretivas:

```

R> x <- gencor(10, "random")
R> plot.gencor(x)

```

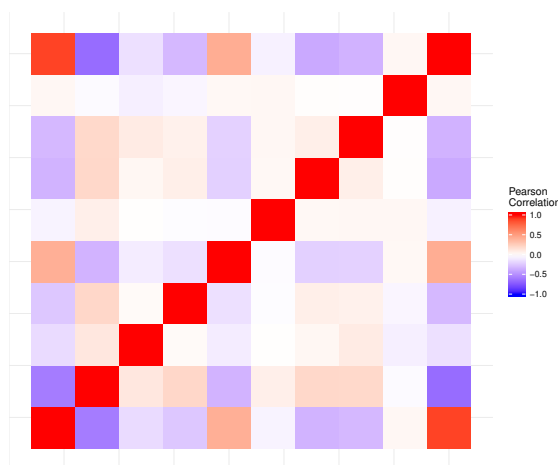


Figura 5.1.1: Resultado da função `plot` para o objeto `x`.

A função `summary` exibe um resumo do objeto gerado pela função `gencor`. É apresentado o método escolhido para a geração, a matriz gerada pelo método, as proporções obtidas dentro de cada faixa limite de correlação, para uma verificação da qualidade dos resultados obtidos, bem como o tempo de execução em segundos, o número de elementos que compõe os vetores de variáveis aleatórias utilizados na geração, a precisão do método e o número de simulações utilizadas até a obtenção do resultado esperado.

```
R> x <- gencor(12, custom_prop = c(.2, .5, .3), custom_lim =
+ c(.2, .7), custom_precision = 0.01)
R> summary(x)
```

Used Method: Custom

Generated Matrix

| | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] |
|-------|------------|-------------|------------|------------|------------|-------------|
| [1,] | 1.0000000 | 0.13633919 | 0.6895186 | -0.6858201 | 0.3663963 | 0.31555096 |
| [2,] | 0.1363392 | 1.0000000 | 0.1797907 | -0.1664551 | 0.1023553 | 0.09319282 |
| [3,] | 0.6895186 | 0.17979069 | 1.0000000 | -0.9636472 | 0.5222100 | 0.47570750 |
| [4,] | -0.6858201 | -0.16645510 | -0.9636472 | 1.0000000 | -0.5174546 | -0.47678581 |
| [5,] | 0.3663963 | 0.10235531 | 0.5222100 | -0.5174546 | 1.0000000 | 0.26573889 |
| [6,] | 0.3155510 | 0.09319282 | 0.4757075 | -0.4767858 | 0.2657389 | 1.0000000 |
| [7,] | 0.5881370 | 0.16727293 | 0.8409024 | -0.8219610 | 0.4409581 | 0.40698361 |
| [8,] | 0.6524248 | 0.16909493 | 0.9303373 | -0.9195087 | 0.5036495 | 0.44277866 |
| [9,] | -0.6844910 | -0.18041903 | -0.9761021 | 0.9628705 | -0.5134569 | -0.47830764 |
| [10,] | -0.6468983 | -0.17407132 | -0.9368873 | 0.9242041 | -0.4937149 | -0.45871274 |
| [11,] | -0.5862029 | -0.12989449 | -0.8118106 | 0.8028284 | -0.4505915 | -0.41012898 |
| [12,] | -0.2033556 | -0.06244596 | -0.3118208 | 0.3119409 | -0.1402732 | -0.13171077 |
| | [,7] | [,8] | [,9] | [,10] | [,11] | [,12] |
| [1,] | 0.5881370 | 0.6524248 | -0.6844910 | -0.6468983 | -0.5862029 | -0.20335555 |
| [2,] | 0.1672729 | 0.1690949 | -0.1804190 | -0.1740713 | -0.1298945 | -0.06244596 |
| [3,] | 0.8409024 | 0.9303373 | -0.9761021 | -0.9368873 | -0.8118106 | -0.31182078 |
| [4,] | -0.8219610 | -0.9195087 | 0.9628705 | 0.9242041 | 0.8028284 | 0.31194093 |
| [5,] | 0.4409581 | 0.5036495 | -0.5134569 | -0.4937149 | -0.4505915 | -0.14027319 |
| [6,] | 0.4069836 | 0.4427787 | -0.4783076 | -0.4587127 | -0.4101290 | -0.13171077 |
| [7,] | 1.0000000 | 0.8012751 | -0.8375299 | -0.8009812 | -0.6931537 | -0.26953347 |
| [8,] | 0.8012751 | 1.0000000 | -0.9266994 | -0.8905712 | -0.7684144 | -0.30725529 |
| [9,] | -0.8375299 | -0.9266994 | 1.0000000 | 0.9352744 | 0.8165281 | 0.30856590 |
| [10,] | -0.8009812 | -0.8905712 | 0.9352744 | 1.0000000 | 0.7724899 | 0.27973349 |
| [11,] | -0.6931537 | -0.7684144 | 0.8165281 | 0.7724899 | 1.0000000 | 0.23840985 |
| [12,] | -0.2695335 | -0.3072553 | 0.3085659 | 0.2797335 | 0.2384099 | 1.0000000 |

Limits and Obtained Proportions

```
(0,0.2] (0.2,0.7] (0.7,1]
0.1969697 0.5000000 0.3030303
```

```
Elapsed Time: 1.432665 secs
Random variables size: 1000
Custom Precision of method: 0.01
Simulations to convergence: 473
```

A função `gencor` gera um objeto da classe `gencor` composto dos seguintes elementos:

- `Matrix` - Matriz de correlações gerada;
- `Method` - Método de geração selecionado;
- `Proportions` - Proporções observadas em cada faixa de correlação da matriz gerada;
- `Runtime` - Tempo de execução da geração;
- `Nsim` - Número de iterações do procedimento de otimização;
- `Precision` - Precisão do procedimento de otimização;
- `Dimension` - Dimensão da matriz gerada;
- `SDev` - Vetor de desvios padrão utilizados na geração de matrizes de correlação;
- `Custom_prop` - Proporções utilizadas na geração de matrizes de correlação customizadas por faixas;
- `Custom_lim` - Limites utilizados na geração de matrizes de correlação customizadas por faixas;
- `Signal` - Método de distribuição dos sinais utilizados na geração de matrizes de correlação;
- `Nrep` - Tamanho da variável aleatória normal geradora;
- `Sort_intensity` - Tipo de distribuição de sinais utilizado na geração de matrizes de correlação.

Conforme apresentado neste capítulo, o pacote desenvolvido fornece aos potenciais usuários uma interface simples e intuitiva em linguagem R, explorando todo o potencial da metodologia proposta neste trabalho, constando de todas as funcionalidades desejáveis aos pesquisadores que porventura tenham a necessidade de gerar matrizes de correlação, sejam elas completamente aleatorizadas ou mesmo matrizes customizadas que visem o atendimento de interesses mais sofisticados.

Conclusão

O objetivo central deste trabalho foi desenvolver um método capaz de gerar matrizes de correlação de quaisquer dimensões e com variados valores e características para as medidas de correlação na referida matriz. Em particular, a metodologia proposta, em sua contribuição mais inovativa, se mostrou hábil para gerar matrizes de correlação customizadas de diversas dimensões, que se ajustem à situações específicas, sem a geração de vícios que possam invalidar os estudos de diversos pesquisadores. O método aqui apresentado teve sua construção apoiada em fortes preceitos matemáticos. Algoritmos para execução do procedimento foram apresentados de forma detalhada. Informações foram apresentadas para que usuários potenciais possam utilizar o método com total liberdade.

Com interesse em assegurar aos pesquisadores interessados na utilização dessa metodologia, um estudo de simulação computacional foi executado. O comportamento da metodologia proposta se mostrou bastante estável quanto a qualidade dos resultados. A agilidade no procedimento computacional foi atestada experimentalmente. Além disso, a avaliação de qualidade foi analisada em comparação com duas outras metodologias, quanto à estrutura de distribuição das correlações geradas, e também quanto ao tempo computacional para execução. O método proposto apresentou robustez em diversos níveis experimentais, seja gerando matrizes de correlação quaisquer, ou com diferentes customizações.

Adicionalmente foi apresentado um pacote estatístico em linguagem R, cujo interesse é promover uma maior disseminação da metodologia proposta, com foco na customização da metodologia. O procedimento de otimização para obtenção das matrizes com uma customização mais específica é bastante simples e de fácil compreensão para qualquer usuário da ferramenta.

Como propostas de continuidade para essa investigação, pretende-se reduzir o tempo de execução do algoritmo customizado, estudar a geração de variáveis aleatórias com base em outras distribuições estatísticas no procedimento de geração das matrizes de correlação. Bem como a possibilidade de utilização de cópulas e a aplicação desta metodologia em estudos anteriores que careciam de matrizes de correlação realísticas, além da avaliação de novas possibilidades de customização, como matrizes exclusivamente positivas, proporção de sinais positivos e negativos, dentre outras.

Referências Bibliográficas

- Bendel, R. B. and Mickey, M. R. (1978). Population correlation matrices for sampling experiments. Communications in Statistics - Simulation and Computation, 7(2):163–182.
- Budden, M., Hadavas, P., Hoffman, L., and Pretz, C. (2007). Generating valid 4×4 correlation matrices. Applied Mathematics E-Notes, 7:53–59.
- Böhm, W. and Hornik, K. (2014). Generating random correlation matrices by the simple rejection method: Why it does not work. Statistics & Probability Letters, 87:27–30.
- Chalmers, C. P. (1975). Generation of correlation matrices with a given eigen-structure. Journal of Statistical Computation and Simulation, 4(2):133–139.
- Cruz, F. R. B., Duarte, A. R., and Souza, G. L. (2018). Multi-objective performance improvements of general finite single-server queueing networks. Journal of Heuristics, 24(5):757–781.
- Davies, P. I. and Higham, N. J. (2000). Numerically stable generation of correlation matrices and their factors. BIT Numerical Mathematics, 40(4):640–651.
- Filzmoser, P. (2005). Identification of multivariate outliers: a performance study. Austrian Journal of Statistics, 34(2):127–138.
- Filzmoser, P., Garrett, R., and Reimann, C. (2005). Multivariate outlier detection in exploration geochemistry. Computers & Geosciences, 31(5):579–587.
- Ghosh, S. and Henderson, S. G. (2003). Behavior of the norta method for correlated random vector generation as the dimension increases. ACM Transactions on Modeling and Computer Simulation (TOMACS), 13(3):276–294.
- Hardin, J., Garcia, S. R., and Golan, D. (2013). A method for generating realistic correlation matrices. The Annals of Applied Statistics, 7(3):1733–1762.
- Holmes, R. B. (1991). On random correlation matrices. SIAM Journal on Matrix Analysis and Applications, 12(2):239–272.

- Hüttner, A. and Mai, J. (2019). Simulating realistic correlation matrices for financial applications: correlation matrices with the perron–frobenius property. Journal of Statistical Computation and Simulation, 89(2):315–336.
- Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis, 94(2):401–419.
- Joe, H. (2006). Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis, 97(10):2177–2189.
- Johnson, D. G. and Welcht, W. J. (1980). The generation of pseudo-random correlation matrices. Journal of Statistical Computation and Simulation, 11(1):55–69.
- Lewandowski, D., Kurowicka, D., and Joe, H. (2009a). Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis, 100:1989–2001.
- Lewandowski, D., Kurowicka, D., and Joe, H. (2009b). Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis, 100(9):1989–2001.
- Marsaglia, G. and Olkin, I. (1984). Generating correlation matrices. SIAM Journal on Scientific and Statistical Computing, 5(2):470–475.
- Odell, P. L. and Feiveson, A. H. (1966). A numerical procedure to generate a sample covariance matrix. Journal of the American Statistical Association, 61(313):199–203.
- Qiu, H. and Joe, H. (2006). Separation index and partial membership for clustering. Computational Statistics & Data Analysis, 50:585–603.
- Rousseeuw, P. J. and Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3):212–223.
- Rousseeuw, P. J. and Zomeren, B. C. V. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85(411):633–639.
- Silva, G. O., Pereira, A. S., Souza, V. Q., Carvalho, F. I. F., and Fritsche Neto, R. (2007). Correlações entre caracteres de aparência e rendimento e análise de trilha para aparência de batata. Bragantia, 66(3):381–388.
- Simonian, J. (2010). The most simple methodology to create a valid correlation matrix for risk management and option pricing purposes. Applied Economics Letters, 17(18):1767–1768.