Diplomarbeit

Matrixwertige Gammaverteilungen und Anwendungen im Risikomanagement

Zur Erlangung des akademischen Grades

Diplom-Ingenieurin

Im Rahmen des Masterstudiums

Finanz- und Versicherungsmathematik
UE 066 405

Ausgeführt am
Forschungsbereich Risikomanagement in Finanz- und Versicherungsmathematik
Institut für Stochastik und Wirtschaftsmathematik
Fakultät für Mathematik und Geoinformation
Technische Universität Wien

Unter der Anleitung von

Univ.-Prof. Dipl.-Math. Dr. rer. nat. habil. Uwe Schmock

Durch

Karoline Vonach, BSc
Matrikelnummer 11825340

Wien, 16. Mai 2024

(Unterschrift Verfasserin)

(Unterschrift Betreuer)

Diploma Thesis

Matrix-Valued Gamma Distributions and Applications in Risk Management

Submitted for the degree of

Master of Science

Within the master's program

Financial and Actuarial Mathematics
UE 066 405

Written at the
Research Unit of Risk Management in Financial and Actuarial Mathematics
Institute of Statistics and Mathematical Methods in Economics
Faculty of Mathematics and Geoinformation
TU Wien

Under the supervision of

Univ.-Prof. Dipl.-Math. Dr. rer. nat. habil. Uwe Schmock

By

Karoline Vonach, BSc
Matriculation number 11825340

Vienna, May 16, 2024

(Signature Author) (Signature Supervisor)

Zusammenfassung

In der Versicherungsmathematik und bei der Modellierung von Kreditrisiken wird oft von Poisson-verteilten Schadensanzahlen mit gammaverteilten Intensitäten ausgegangen, da derartige Modelle Überdispersion zulassen. Diese Diplomarbeit befasst sich mit der Möglichkeit, diesen Vektor von Poissonintensitäten mithilfe einer matrixwertigen Gammaverteilung zu modellieren.

Zunächst führen wir Wishart- und matrixwertige Gammaverteilungen, einschließlich ihrer degenerierten Varianten, ein und beweisen, dass erstere eine Teilmenge von letzteren sind. Anschließend definieren wir eine allgemeinere Version von matrixwertigen Gammaverteilungen, die auch singuläre Wishart-Verteilungen einschließt, und beweisen einige grundlegende Eigenschaften dieser Verteilungsfamilie. Dabei leiten wir auch matrixwertige Versionen von Exponentialverteilungen her, von denen wir zeigen, dass sie gedächtnislos bezüglich einer geeigneten Halbordnung sind.

Weiterführend führen wir zunächst Poisson-Mischmodelle im Allgemeinen ein, um dann *Matrix-Gamma Poisson-Mischmodelle* vorzustellen. Diese haben den Vorteil, dass damit komplexere Abhängigkeitsstrukturen modelliert werden können, während die eindimensionalen Verteilungen noch immer negative Binomialverteilungen sind. Dieser Vorteil geht jedoch mit dem Nachteil einher, dass im hier beschriebenen Stand der Matrix-Gamma Poisson-Mischmodelle nur nicht-negative Korrelationen zwischen den Ausfallzahlen modelliert werden können. Wir diskutieren einige Eigenschaften dieser neuen Modellklasse und legen dabei einen besonderen Schwerpunkt auf wahrscheinlichkeitserzeugende Funktionen und den Bereich der möglichen Abhängigkeitsstrukturen. Abschließend leiten wir im Rahmen der Matrix-Gamma Poisson-Mischmodelle einige bedingte Verteilungen her, die sich für Versicherungs- und Kreditrisikomodelle als nützlich erweisen könnten.

Schlagworte: Matrixwertige Gammaverteilung, Wishart-Verteilung, Kreditrisikomodellierung, Versicherungsmathematik, Poisson-Mischmodelle.

Abstract

In the fields of actuarial science and credit risk modeling, the claim number is often assumed to have a multivariate Poisson distribution, where the random intensities follow gamma distributions, allowing for overdispersion. We propose to model this vector of random Poisson intensities in terms of a matrix-valued gamma distribution, leading to the formulation of *matrix-gamma multivariate Poisson mixture models*.

This thesis starts with discussing Wishart and matrix-valued gamma distributions, including their degenerate variants. After proving that the former are a subset of the latter via characteristic functions, we define general matrix-valued gamma distributions, which include the subclass of singular Wishart distributions, and explore and derive properties of this distribution family. This also includes deriving matrix-valued versions of exponential distributions, which we prove to be memoryless w.r.t. an appropriate partial order.

Moreover, we introduce Poisson mixture models in general before defining and discussing matrix-gamma multivariate Poisson mixture models. These models possess the advantage of capturing more sophisticated dependency structures among default numbers as compared to existing models, while preserving the property that each individual default number follows a negative binomial distribution. However, this advantage comes with the trade-off that only non-negative correlations between the default numbers can be modelled in the current state of the framework. We discuss and illustrate multiple properties of this new model, with particular emphasis given to the probability-generating functions and the range of dependency structures that can be effectively represented within this framework. Finally, we derive conditional distributions within the context of matrix-gamma multivariate Poisson mixture models, which could prove valuable for the application of this model in insurance or credit claim modeling scenarios.

Keywords: Matrix-valued gamma distribution, Wishart distribution, Credit risk modelling, Insurance mathematics, Poisson mixture distribution.

Mathematics Subject Classification (MSC2020): 60B20 (primary), 62H30, 91G05, 91G40 (secondary)

Acknowledgements

I would like to express my sincere gratitude to my supervisor Uwe Schmock, whose enthusiasm, knowledge, and countless ideas have been invaluable throughout the entire process of researching and writing this thesis. From guiding me through complex theories to solving "problems" together, he has been a constant source of inspiration and support. I am especially grateful for the opportunities he has provided beyond the scope of writing this thesis, which have significantly enhanced my academic and professional growth.

Moreover, I am also sincerely thankful to the FAM research group for welcoming me as a member of the team, including hiking trips and marathons. Special thanks go to Sandra Trenovatz for her exceptional support in all organizational matters, enjoyable chats and everything else.

To my colleagues who have become my best friends throughout the years at TU Wien, without your support, collaboration, and camaraderie this degree would have been twice as hard and half as fun.

I am deeply grateful to my family and friends for their unwavering love, encouragement, and understanding during this academic journey.

Karoline Vonach

Statement of originality

I hereby declare that I have authored the present master thesis independently and did not use any sources other than those specified. I have not yet submitted the work to any other examining authority in the same or comparable form. It has not been published yet.

Vienna, May 16, 2024

Karoline Vonach

Contents

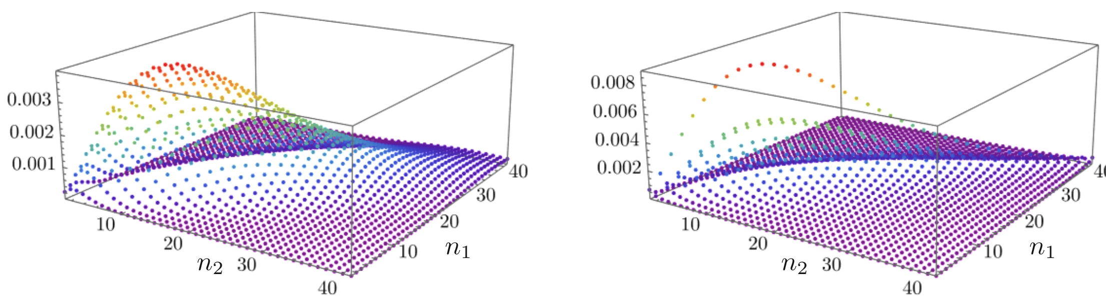
1. Introduction

In the realm of insurance mathematics and credit risk modeling, it is popular to model claim counts with Poisson mixture models due to their ability to flexibly capture heterogeneity in claim frequencies. In particular, the mixture distribution is often chosen to be a gamma distribution because it is well-suited for accounting for overdispersion, a common feature in claim count data. Additionally, the unconditional distribution of N is then given by a negative binomial distribution.

Extending this to a multivariate Poisson mixture model enables the simultaneous modeling of multiple claim counts depending on distinct, possibly interconnected, Poisson intensities. Previous works, such as [29, Theorem 7.39], have shown that this vector of Poisson intensities can be modeled to have any given covariance structure by using linear combinations of independent gamma distributed risk factors.

This thesis explores the possibility to use a matrix-valued gamma distribution, see Definition 3.24, to derive the vector of Poisson intensities for a (multivariate) Poisson mixture model. For a random matrix Y following a matrix-valued gamma distribution, the random variable $\langle v, Yv\rangle$, where $v \in \mathbb{R}^p$, follows a one-dimensional gamma distribution, as will be proven in Corollary 3.39. We utilize this property to define a *matrix-gamma (multivariate) Poisson mixture model*, where the Poisson intensities are derived from one common matrix-gamma distributed Y, potentially with different vectors $v_1, v_2, \dots, v_n$, to extract one-dimensional gamma distributions. This approach provides us with the possibility to model more sophisticated dependency structures among both the Poisson intensities and the modeled default numbers.

Figure 1.1 illustrates the difference between a matrix-gamma Poisson mixture model and a pair of negative binomial random variables, whose dependency is induced by adding independent negative binomial random variables.



(a) *Matrix-gamma Poisson mixture model* (b) *Sum of independent* NegBin *r.v.*

Figure 1.1. *Joint probability mass function of dependent r.v. (N_1, N_2) in different models with the same marginal distributions and correlation, elaborated in Remark 4.34.*

In both graphs the joint probability mass function of a pair of default numbers (N_1, N_2) with the same negative binomial marginal distributions and the same correlation $5/6$ is depicted; refer to Remark 4.34 for detailed information on the exact parameters.

We observe that a matrix-valued Poisson mixture model is able to model a high correlation more smoothly than the other model, where the "artificially" induced correlation results in a high probability mass concentration on the diagonal of the graph, where $N_1 = N_2$.

The difference of the graphs illustrates how matrix-gamma Poisson mixture models could be useful to model correlated default numbers with a smooth dependency structure. Another advantage is that within this model, all default numbers still follow a negative binomial distribution with different scale parameters p, which cannot be achieved by adding independent gamma random variables; see Remark 4.34 for further details. Furthermore, independent claim counts in a matrix-gamma Poisson mixture model can be easily modelled and detected, as their independence is equivalent to the corresponding Poisson intensities having zero correlation, which again can be traced back to a simple condition on the corresponding vectors and the parameters of the matrix-gamma distribution; refer to Theorem 4.36 for an elaboration of this property. However, it must be noted that the main limitation of matrix-gamma Poisson mixture models, as far as explored in this thesis, is that it can only model default numbers with non-negative correlation; see Lemma 4.33. Moreover, the resulting marginal negative binomial distributions must all have the identical shape parameter α, as opposed to their flexibility in the second parameter p.

The thesis is organized as follows: Chapter 2 lays the groundwork by introducing several essential tools for working with matrix-valued gamma distributions, including Definition 2.36 of the multi-dimensional gamma function and the essential Theorem 2.37. In Chapter 3, we begin with the introduction of Wishart distributions, followed by an exploration of matrix-valued gamma distributions as a generalization of the former, with both distributions extended to their degenerate cases; see Definitions 3.8, 3.24 and 3.34. Subsequent sections analyze various properties of matrix-valued gamma distributions, in particular their characteristic functions in Corollary 3.35, scaling properties in Corollary 3.39, and their distributions under biased probability measures in Theorem 3.64. Additionally, we derive the subclass of matrix-valued exponential distributions, which we prove to be memoryless w.r.t. the Loewner partial order; see Definition 2.10, Lemma 3.49. In Section 4.1 of Chapter 4, we discuss Poisson Mixture models generally before introducing matrix-gamma Poisson mixture models in Section 4.2 and the multivariate version in Section 4.3. There, we also derive several essential properties of these models, most importantly the probability-generating function in Theorem 4.20 and its alternative form given by Lemma 4.22, expectations and covariances in Lemma 4.33, and conditions on independence within the model in Theorem 4.36. Finally, Section 4.4 presents results on conditioning on observations within matrix-gamma (multivariate) Poisson mixture models.

2. Preliminaries

This chapter will introduce a variety of tools necessary to define and work with the matrix-valued gamma distribution in the following chapters. This includes several basic properties of matrices, in particular of positive definite matrices, as well as the definition of the multi-dimensional gamma function.

For this thesis, let $\mathbb{K}$ denote either the field $\mathbb{R}$ of real or the field $\mathbb{C}$ of complex numbers. Furthermore, for every $p \in \mathbb{N}$ the $p \times p$-dimensional identity matrix is denoted by I_p. For further conventions, abbreviations, symbols and notation, see *Conventions*.

2.1. Matrix Algebra and the Weinstein–Aronszajn Identity

This section demonstrates a few basic properties of matrices that will be needed for subsequent statements and proofs. Some are stated more generally for commutative rings (with one), which include fields and $\mathbb{K}$ in particular. We start with a property of the trace.

Lemma 2.1 (Cyclic permutation of matrices inside the trace)**.** *Let R denote a commutative ring. For $p, q \in \mathbb{N}$, let $A \in R^{p\times q}$ and $B \in R^{q\times p}$ be two matrices. Then*

$$\operatorname{tr}(AB) = \operatorname{tr}(BA). \tag{2.1}$$

Proof. For A, B of the given dimensions

$$\operatorname{tr}(AB) = \sum_{m=1}^{p} (AB)_{mm} = \sum_{m=1}^{p} \sum_{n=1}^{q} a_{mn} b_{nm} = \sum_{n=1}^{q} \sum_{m=1}^{p} b_{nm} a_{mn} = \operatorname{tr}(BA),$$

where we use the definition of the trace of a matrix. □

Lemma 2.2 (Representation of a determinant as a product of determinants, see [17, Eq. (0.8.5.1)])**.** *Let $\mathcal{R}$ denote a commutative ring with one. For $p \in \mathbb{N}$ and $A \in \mathcal{R}^{p\times p}$ let $\alpha \subsetneq \{1, \dots, p\}$ be a non-empty index set such that $A[\alpha]$, defined as the $|\alpha| \times |\alpha|$-dimensional submatrix of A containing all rows and columns with index in α, is invertible. Then, with $\alpha^{\mathsf{c}} := \{1, \dots, p\} \setminus \alpha$,*

$$\det A = \det(A[\alpha]) \det\bigl(A[\alpha^{\mathsf{c}}] - A[\alpha^{\mathsf{c}}, \alpha] A[\alpha]^{-1} A[\alpha, \alpha^{\mathsf{c}}]\bigr),$$

where $A[\beta, \gamma]$ is defined as the $|\beta| \times |\gamma|$-dimensional submatrix of A containing all rows with index in β and columns with index in γ for all non-empty $\beta, \gamma \subsetneq \{1, \dots, p\}$.

Proof. For every non-empty $\alpha \subsetneq \{1, \dots, p\}$ there exists a permutation matrix $P_\alpha \in \mathcal{R}^{p\times p}$ such that

$$P_\alpha A P_\alpha^{\mathsf{T}} = \begin{pmatrix} A[\alpha] & A[\alpha, \alpha^{\mathsf{c}}] \\ A[\alpha^{\mathsf{c}}, \alpha] & A[\alpha^{\mathsf{c}}] \end{pmatrix}.$$

As every permutation matrix is orthogonal, i.e. $P_\alpha^\mathsf{T} = P_\alpha^{-1}$,

$$\det P_\alpha \det P_\alpha^\mathsf{T} = \det P_\alpha \det P_\alpha^{-1} = \det(P_\alpha P_\alpha^{-1}) = \det(I_p) = 1,$$

where we use the multiplicativity of the determinant. Therefore,

$$\det A = \det(P_\alpha A P_\alpha^\mathsf{T}) = \det \begin{pmatrix} A[\alpha] & A[\alpha, \alpha^\mathsf{c}] \\ A[\alpha^\mathsf{c}, \alpha] & A[\alpha^\mathsf{c}] \end{pmatrix}.$$

Note that

$$\det \begin{pmatrix} A[\alpha]^{-1} & -A[\alpha]^{-1}A[\alpha, \alpha^\mathsf{c}] \\ 0 & I_{p-|\alpha|} \end{pmatrix} = \det(A[\alpha]^{-1}) = (\det A[\alpha])^{-1} \tag{2.2}$$

as the first matrix is an upper triangular block matrix[1] and as $A[\alpha]$ invertible is equivalent to $\det A[\alpha]$ having a multiplicative inverse element. Therefore,

$$\begin{aligned} &\det \begin{pmatrix} A[\alpha] & A[\alpha, \alpha^\mathsf{c}] \\ A[\alpha^\mathsf{c}, \alpha] & A[\alpha^\mathsf{c}] \end{pmatrix} \\ &\overset{(2.2)}{=} \det A([\alpha]) \det \begin{pmatrix} A[\alpha] & A[\alpha, \alpha^\mathsf{c}] \\ A[\alpha^\mathsf{c}, \alpha] & A[\alpha^\mathsf{c}] \end{pmatrix} \det \begin{pmatrix} A[\alpha]^{-1} & -A[\alpha]^{-1}A[\alpha, \alpha^\mathsf{c}] \\ 0 & I_{p-|\alpha|} \end{pmatrix} \\ &= \det(A[\alpha]) \underbrace{\det \begin{pmatrix} I_{|\alpha|} & 0 \\ A[\alpha^\mathsf{c}, \alpha]A[\alpha]^{-1} & A[\alpha^\mathsf{c}] - A[\alpha^\mathsf{c}, \alpha]A[\alpha]^{-1}A[\alpha, \alpha^\mathsf{c}] \end{pmatrix}}_{= \det(A[\alpha^\mathsf{c}] - A[\alpha^\mathsf{c}, \alpha]A[\alpha]^{-1}A[\alpha, \alpha^\mathsf{c}])}, \end{aligned}$$

where multiplicativity of the determinant as well as the fact, that the last matrix is a lower triangular block matrix, is used. □

Lemma 2.3 (Weinstein–Aronszajn identity, see [26, Appendix B.1])**.** *Let $\mathcal{R}$ denote a commutative ring with one. For $p, q \in \mathbb{N}$ and matrices $A \in \mathcal{R}^{p \times q}$ and $B \in \mathcal{R}^{q \times p}$,*

$$\det(I_p + AB) = \det(I_q + BA).$$

Proof. Since I_p and I_q are self-inverse,

$$\begin{aligned} \det(I_p + AB) = \det(I_q)\det(I_p - (-A)I_q^{-1}B) = \det \begin{pmatrix} I_p & -A \\ B & I_q \end{pmatrix} \\ = \det(I_p)\det(I_q - BI_p^{-1}(-A)) = \det(I_q + BA), \end{aligned}$$

where Lemma 2.2 is used for the second and third equality with $\alpha = \{p+1, \ldots, p+q\}$ and $\alpha = \{1, \ldots, p\}$, respectively. □

[1] The fact the the determinant of a upper (lower) block diagonal matrix is given by the product of the determinants of the blocks can e.g. be seen using the Leibniz formula.

Lemma 2.4 (Cauchy–Binet formula, see [1, Theorem 2.34])**.** *Let $\mathcal{R}$ denote a commutative ring with one. For $p, q \in \mathbb{N}$ with $p \geq q$ let $A \in \mathcal{R}^{q\times p}$ and $B \in \mathcal{R}^{p\times q}$ be two matrices. Then the determinant of their product is given by*

$$\det(AB) = \sum_{\substack{S\subseteq\{1,\dots,p\},\\ |S|=q}} \det(A_{\{1,\dots,q\},S}) \det(B_{S,\{1,\dots,q\}}),$$

where $A_{\{1,\dots,q\},S}$ denotes the $q \times q$ matrix, whose columns are all columns of A with their index in S. The matrix $B_{S,\{1,\dots,q\}}$ is defined analogously to be the $q \times q$ matrix, whose rows are all rows of B with their index in S.

The following theorem as well as its proof can be found in [17, Theorem 2.5.3, Theorem 2.5.6].

Theorem 2.5 (Unitary decomposition of Hermitian matrices)**.** *For $p \in \mathbb{N}$, let $A \in \mathbb{K}^{p\times p}$ be a Hermitian matrix. Then there exists a unitary matrix[2] $S \in \mathbb{K}^{p\times p}$, which is an invertible matrix satisfying $S^{\mathsf{H}} = S^{-1}$, and a diagonal matrix $D = \operatorname{diag}(\lambda_1, \dots, \lambda_p)$ with $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ such that*

$$A = SDS^{\mathsf{H}}.$$

Furthermore, the diagonal entries of D are the eigenvalues of A (with their corresponding multiplicities) and the columns of S are the corresponding normalized eigenvectors. Hence, $\operatorname{rk}(D) = \operatorname{rk}(A)$.

Remark 2.6 (Unitary decompositions of Hermitian matrices are not unique)**.** Note that the decomposition given in Theorem 2.5 is never unique for the Hermitian matrix $A \in \mathbb{K}^{p\times p}$. Take a decomposition $A = SDS^{\mathsf{H}}$ with D a diagonal matrix of eigenvalues and S unitary. Even if all of the entries of D are pairwise different, every matrix $\tilde{S}$ which is equal to S with only the sign changed for some of the columns also satisfies $A = \tilde{S}D\tilde{S}^{\mathsf{H}}$.

2.2. Positive Definite Matrices and their Cholesky Decomposition

In this section we recall the notion of positive (semi-)definite matrices along with several basic properties before stating and proving the Cholesky decomposition, see Theorem 2.12 below, which will be used to parameterize the submanifold of positive definite matrices in the subsequent sections. In particular this will enable us to be able to integrate over the set of positive definite matrices, which, in turn, will be used to define the multi-dimensional gamma function.

Definition 2.7 (Definite matrices)**.** For $p \in \mathbb{N}$ a Hermitian matrix $A \in \mathbb{K}^{p\times p}$ is called

(i) positive definite if $x^{\mathsf{H}}Ax > 0$ for all $x \in \mathbb{K}^p \setminus \{0\}$,

(ii) positive semi-definite if $x^{\mathsf{H}}Ax \geq 0$ for all $x \in \mathbb{K}^p$,

(iii) negative definite if $x^{\mathsf{H}}Ax < 0$ for all $x \in \mathbb{K}^p \setminus \{0\}$,

(iv) negative semi-definite if $x^{\mathsf{H}}Ax \leq 0$ for all $x \in \mathbb{K}^p$.

[2] In the real case the matrix S here is called an orthogonal matrix, which is an invertible matrix satisfying $S^{\mathsf{T}} = S^{-1}$.

Lemma 2.8 (Properties of positive (semi-)definite matrices)**.** *For every $p \in \mathbb{N}$, the following statements about positive (semi-)definite matrices in $\mathbb{K}^{p\times p}$ hold.*

(i) The set of all $p \times p$-dimensional positive definite matrices is a convex cone, i.e. it is closed under addition and multiplication with positive[3] scalars.

(ii) The set of all $p \times p$-dimensional positive semi-definite matrices is a convex cone, that is also closed under multiplication with non-negative scalars.

(iii) The sum of a positive definite and a positive semi-definite matrix of the same size is positive definite.

(iv) For a positive semi-definite matrix $A \in \mathbb{K}^{p\times p}$ and $x \in \mathbb{K}^p$, the product $x^{\mathsf{H}} A\, x = 0$ if and only if $Ax = 0$.

(v) A positive semi-definite matrix is positive definite if and only if it has full rank.

(vi) The inverse of a positive definite matrix is positive definite.

(vii) A Hermitian matrix $A \in \mathbb{K}^{p\times p}$ is positive definite if and only if there exists a real $\delta > 0$ such that

$$x^{\mathsf{H}} Ax \geq \delta, \quad x \in \mathbb{K}^p \text{ with } \|x\|_2 = 1, \tag{2.3}$$

where $\|\cdot\|_2$ refers to the Euclidian norm.

Proof. Item (i): Let $\alpha > 0$ and $\beta \geq 0$ be two scalars and let $A, B \in \mathbb{K}^{p\times p}$ be two positive definite matrices. Then

$$(\alpha A + \beta B)^{\mathsf{H}} = \alpha A^{\mathsf{H}} + \beta B^{\mathsf{H}} = \alpha A + \beta B$$

and the calculation $x^{\mathsf{H}}(\alpha A + \beta B)\, x = \alpha(x^{\mathsf{H}} Ax) + \beta(x^{\mathsf{H}} Bx) > 0$ holds for all $x \in \mathbb{K}^p \setminus \{0\}$. Hence, $\alpha A + \beta B$ is a positive definite matrix as well.

Item (ii): The proof can be done in analogously to Item (i), where α and β only have to be greater or equal than zero.

Item (iii): Let $A \in \mathbb{K}^{p\times p}$ be a positive definite matrix and let $B \in \mathbb{K}^{p\times p}$ be a positive semi-definite matrix. Then $A + B$ is Hermitian as well and for all $x \in \mathbb{K}^p \setminus \{0\}$

$$x^{\mathsf{H}}(A + B)\, x = \underbrace{x^{\mathsf{H}} Ax}_{>0} + \underbrace{x^{\mathsf{H}} Bx}_{\geq 0} > 0.$$

Item (iv): For the non-trivial implication take $x \in \mathbb{K}^p$ with $x^{\mathsf{H}} A\, x = 0$. Then, as A is positive semi-definite, for all $\lambda \in \mathbb{R}$ the equality

$$0 \leq (\lambda x + Ax)^{\mathsf{H}} A(\lambda x + Ax) = \lambda^2 \underbrace{x^{\mathsf{H}} A\, x}_{=0} + 2\lambda (Ax)^{\mathsf{H}} Ax + \underbrace{(Ax)^{\mathsf{H}} AAx}_{=: C \geq 0} = 2\lambda \|Ax\|_2^2 + C$$

must hold, with $\|\cdot\|_2$ referring to the Euclidean norm, which can only be fulfilled if $\|Ax\|_2 = 0$, implying $Ax = 0$.

[3] In this thesis positive refers to $x \in \mathbb{R}$ satisfying $x > 0$, whereas non-negative is written for $x \in \mathbb{R}$ fulfilling $x \geq 0$. The set of all real positive numbers is denoted by $\mathbb{R}^+$, which means $0 \notin \mathbb{R}^+$.

Item (v): This follows from combining Item (iv) with the fact that a matrix $A \in \mathbb{K}^{p\times p}$ is of full rank if and only if $Ax \neq 0$ for all $x \in \mathbb{K}^p \setminus \{0\}$.

Item (vi): Let $A \in \mathbb{K}^{p\times p}$ be a positive definite matrix. Since A is of full rank due to Item (v), its inverse A^{-1} exists and is Hermitian as

$$A^{-1} = I_p^{\mathsf{H}} A^{-1} = (AA^{-1})^{\mathsf{H}} A^{-1} = (A^{-1})^{\mathsf{H}} A^{\mathsf{H}} A^{-1} = (A^{-1})^{\mathsf{H}} A A^{-1} = (A^{-1})^{\mathsf{H}},$$

where we use that A is Hermitian. For every $x \in \mathbb{K}^p \setminus \{0\}$,

$$x^{\mathsf{H}} A^{-1} x = (AA^{-1}x)^{\mathsf{H}} A^{-1} x = (A^{-1}x)^{\mathsf{H}} A^{\mathsf{H}} (A^{-1}x) > 0$$

as $A = A^{\mathsf{H}}$ is positive definite, thus proving that A^{-1} is positive definite.

Item (vii): For the first implication, fix a positive definite $A \in \mathbb{K}^{p\times p}$. As the map $\mathbb{K}^p \ni x \mapsto x^{\mathsf{H}} A x \in \mathbb{R}^+$ is continuous and the set of all $x \in \mathbb{K}^p$ with Euclidian norm equal to one is compact, the extreme value theorem tells us that the infimum (supremum) of this function over the compact domain is attained and equal to the minimum (maximum). Hence, there exists a real δ satisfying

$$\min_{\substack{x\in\mathbb{K}^p \\ \|x\|_2=1}} x^{\mathsf{H}} A x = \delta.$$

Since A is positive definite, $\delta > 0$ and Eq. (2.3) is satisfied for this δ.

Conversely, assume that for a Hermitian $A \in \mathbb{K}^{p\times p}$ Eq. (2.3) holds. Every $y \in \mathbb{K}^p \setminus \{0\}$ can be written as $y := \alpha x$ with $\alpha := \|y\|_2 \in \mathbb{R}^+$ and $x = y/\|y\|_2 \in \mathbb{K}^p$. Hence,

$$y^{\mathsf{H}} A y = \alpha^2 x^{\mathsf{H}} A x \geq \alpha^2 \delta > 0. \qquad \square$$

The following lemma provides a characterization of positive definite matrices and can be found in [17] along with its proof.

Lemma 2.9 (Sylvester's criterion, see [17, Corollary 7.1.5, Theorem 7.2.5])**.** *For $p \in \mathbb{N}$, let $A \in \mathbb{K}^{p\times p}$ be a Hermitian matrix. Then A is positive definite if and only if all of its leading (or trailing) principal minors are positive.*

Using the definition of positive (semi-)definite matrices, we can define a partial order on the set of quadratic $p \times p$-dimensional matrices for every $p \in \mathbb{N}$.

Definition 2.10 (Loewner order)**.** For $p \in \mathbb{N}$, let $A, B \in \mathbb{K}^{p\times p}$ be two quadratic matrices. Then A is defined to be smaller or equal then B in the *Loewner order* if $B - A$ is a positive semi-definite matrix. We use the notation $A \leq_L B$. In an analogue way, $A <_L B$ if $B - A$ is positive definite.[4]

[4] Note that this relation between $\leq_L$ and $<_L$ is not the usual relation between $\leq$ and $<$ for a partial order, i.e. $x < y$ if and only if $x \leq y$ and $x \neq y$.

Lemma 2.11 (The Loewner order is a partial order)**.** *For every $p \in \mathbb{N}$ the Loewner order given in Definition 2.10 is a partial order on $\mathbb{K}^{p\times p}$.*

Proof. This lemma is proven by showing that the defined relation is reflexive, anti-symmetric and transitive. Fix any matrices $A, B, C \in \mathbb{K}^{p\times p}$.

Since $A - A = 0$ and the zero matrix is positive semi-definite, $A \leq_L A$ is proved.

If A, B satisfy $A \leq_L B$ and $B \leq_L A$, then both $B - A$ and $A - B = -(B - A)$ are positive semi-definite matrices. Hence, for all $x \in \mathbb{K}^p$,

$$x^{\mathsf{H}}(B-A)x \geq 0 \quad \text{and} \quad -x^{\mathsf{H}}(B-A)x \geq 0,$$

implying $x^{\mathsf{H}}(B - A)x = 0$. Using Item (iv) from Lemma 2.8, $(B - A)x = 0$ for all $x \in \mathbb{K}^p$, which can only be satisfied for $B - A$ being the zero matrix. Hence, $A = B$.

For transitivity, assume $A \leq_L B$ and $B \leq_L C$. Due to Item (ii) of Lemma 2.8, $C - A = (C - B) + (B - A)$ is positive semi-definite and therefore $A \leq_L C$. □

One very convenient property of positive definite matrices is that every one of them can be represented by their *Cholesky decomposition*, characterized by a corresponding unique triangular matrix.

Theorem 2.12 (Cholesky decomposition of positive definite matrices, see [24, Theorem 6.23])**.** *A matrix $A \in \mathbb{K}^{p\times p}$ is positive definite if and only if there exists a lower triangular matrix $T \in \mathbb{K}^{p\times p}$, where the diagonal entries are real and positive, such that $TT^{\mathsf{H}} = A$. For each positive definite A, this matrix T is unique.*

Proof. First, we want to prove the implication, that for all lower triangular $T \in \mathbb{K}^{p\times p}$ with real and positive diagonal entries, $A := TT^{\mathsf{H}}$ is positive definite. It can be seen that A must be Hermitian as

$$A^{\mathsf{H}} = (TT^{\mathsf{H}})^{\mathsf{H}} = (T^{\mathsf{H}})^{\mathsf{H}}T^{\mathsf{H}} = TT^{\mathsf{H}} = A.$$

Considering that T has full rank,

$$x^{\mathsf{H}}Ax = x^{\mathsf{H}}TT^{\mathsf{H}}x = \|x^{\mathsf{H}}T\|_2^2 > 0, \quad x \in \mathbb{K}^p \setminus \{0\},$$

where $\|\cdot\|_2$ is the Euclidean norm.

For the other implication, the proof is done via induction over the dimension $p \in \mathbb{N}$.

For $p = 1$, the matrix A must be equal to a positive real α. The matrix T must also be equal to a positive real t that fulfills $t^2 = \alpha$. Hence, $T = (\sqrt{\alpha})$ is the unique solution.

Now for $p \in \mathbb{N}, p \geq 2$, suppose there exists a Cholesky decomposition for all $(p-1)\times(p-1)$-dimensional positive definite matrices and let $A \in \mathbb{K}^{p\times p}$ be a positive definite matrix. Since A is Hermitian, it can be written as

$$A = \begin{pmatrix} A_{p-1} & a \\ a^{\mathsf{H}} & \alpha \end{pmatrix},$$

where $A_{p-1} \in \mathbb{K}^{(p-1)\times(p-1)}$ is Hermitian, $a \in \mathbb{K}^{p-1}$ and $\alpha \in \mathbb{R}$. Due to Lemma 2.9, A_{p-1} must be positive definite.

Suppose there exists a $p \times p$-dimensional lower triangular matrix T with real positive entries on the diagonal that fulfills $TT^{\mathsf{H}} = A$. Then T can be written as

$$T = \begin{pmatrix} T_{p-1} & 0 \\ t^{\mathsf{H}} & \theta \end{pmatrix},$$

where $T_{p-1} \in \mathbb{K}^{(p-1)\times(p-1)}$ is a lower triangular matrix with real positive entries on the diagonal, $t \in \mathbb{K}^{p-1}$ and $\theta \in \mathbb{R}^{+}$. In order for TT^{H} to be equal to A, the equality

$$TT^{\mathsf{H}} = \begin{pmatrix} T_{p-1} & 0 \\ t^{\mathsf{H}} & \theta \end{pmatrix} \begin{pmatrix} T_{p-1}^{\mathsf{H}} & t \\ 0 & \theta \end{pmatrix} = \begin{pmatrix} T_{p-1}T_{p-1}^{\mathsf{H}} & T_{p-1}t \\ (T_{p-1}t)^{\mathsf{H}} & t^{\mathsf{H}}t + \theta^2 \end{pmatrix} \overset{!}{=} \begin{pmatrix} A_{p-1} & a \\ a^{\mathsf{H}} & \alpha \end{pmatrix} = A$$

must hold. Since A_{p-1} is positive definite, there exists a unique Cholesky decomposition of $A_{p-1} = T_{p-1}T_{p-1}^{\mathsf{H}}$, where T_{p-1} is a lower triangular matrix with real positive entries on the diagonal. Now we are left to prove that the two equalities

$$T_{p-1}t = a, \quad t^{\mathsf{H}}t + \theta^2 = \alpha$$

have unique solutions $t \in \mathbb{K}^{p-1}$ and $\theta \in \mathbb{R}^{+}$. Since T_{p-1} is of full rank, the first equation has the unique solution $t := T_{p-1}^{-1}a$. We can rewrite the equation for θ as

$$\begin{aligned}\theta^2 = \alpha - t^{\mathsf{H}}t = \alpha - (T_{p-1}^{-1}a)^{\mathsf{H}}\, T_{p-1}^{-1}a = \alpha - a^{\mathsf{H}}A_{p-1}^{-1}a \\ = \begin{pmatrix} -A_{p-1}^{-1}a \\ 1 \end{pmatrix}^{\mathsf{H}} \underbrace{\begin{pmatrix} A_{p-1} & a \\ a^{\mathsf{H}} & \alpha \end{pmatrix}}_{=A} \begin{pmatrix} -A_{p-1}^{-1}a \\ 1 \end{pmatrix} > 0,\end{aligned}$$

since A is positive definite. Hence the unique solution is $\theta := \sqrt{\alpha - t^{\mathsf{H}}t} \in \mathbb{R}^{+}$. □

Remark 2.13 (Cholesky decomposition of positive semi-definite matrices, see [17, Corollary 7.2.9]). The positive semi-definite matrices can also be represented by a Cholesky decomposition in a slightly altered form: A matrix $A \in \mathbb{K}^{p\times p}$ is positive semi-definite if and only if there exists a lower triangular matrix $T \in \mathbb{K}^{p\times p}$, where the diagonal entries are real and non-negative, such that $TT^{\mathsf{H}} = A$. Such a Cholesky decomposition is not necessarily unique for positive semi-definite matrices as for the example

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

all of the matrices

$$T_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

can be used for a Cholesky decomposition of A.

Let us point out that the argumentation used in the proof of Theorem 2.12 also leads to the following lemma.

Lemma 2.14 (Mapping positive definite matrices to their Cholesky decompositions is bicontinuous)**.** *For $p \in \mathbb{N}$ let $\mathbb{T}_p^{\mathbb{K}}$ denote the set of all lower triangular matrices in $\mathbb{K}^{p\times p}$ with positive diagonal entries. Then the function ϕ that maps every $T \in \mathbb{T}_p^{\mathbb{K}}$ to the positive definite matrix $TT^{\mathsf{H}} \in \mathbb{K}^{p\times p}$ is a homeomorphism[5] between $\mathbb{T}_p^{\mathbb{K}}$ and the set of positive definite matrices in $\mathbb{K}^{p\times p}$, where both sets are equipped with the topology induced by a matrix norm on $\mathbb{K}^{p\times p}$.*

Proof. The function ϕ is continuous as multiplying and transposing matrices are continuous operations. It is bijective as a map from $\mathbb{T}_p^{\mathbb{K}}$ to the set of positive definite matrices in $\mathbb{K}^{p\times p}$ due to the uniqueness of the Cholesky decomposition of positive definite matrices shown in Theorem 2.12.

For every positive definite $A \in \mathbb{K}^{p\times p}$ the induction over the dimension $p \in \mathbb{N}$ used in the proof of Theorem 2.12 can be used to find the unique lower triangular matrix $T \in \mathbb{T}_p^{\mathbb{K}}$ satisfying $TT^{\mathsf{H}} = A$. We are going to prove by the same induction that ϕ^{-1} is continuous.

For $p = 1$, the now real T is given by $\sqrt{A}$ where $A \in \mathbb{R}^+$ in this case. As the function

$$\begin{cases} \mathbb{R}^+ & \to & \mathbb{R}^+ \\ x & \mapsto & \sqrt{x} \end{cases}$$

is continuous, this operation is continuous.

For $p \in \mathbb{N}$ with $p \geq 2$, assume ϕ^{-1} is continuous for dimension $p-1$. Due to the proof of Theorem 2.12, we can represent A and T by

$$A = \begin{pmatrix} A_{p-1} & a \\ a^{\mathsf{H}} & \alpha \end{pmatrix}, \quad T = \begin{pmatrix} T_{p-1} & 0 \\ t^{\mathsf{H}} & \theta \end{pmatrix}, \quad A_{p-1}, T_{p-1} \in \mathbb{K}^{(p-1)\times(p-1)};\ a, t \in \mathbb{K}^{p-1};\ \alpha, \theta \in \mathbb{R}^+,$$

where $T_{p-1}T_{p-1}^{\mathsf{H}}$ is the unique Cholesky decomposition of A_{p-1} from the induction hypothesis, $t = T_{p-1}^{-1}a$ and $\theta = \sqrt{\alpha - t^{\mathsf{H}}t}$. Due to the induction hypothesis, the function that maps A_{p-1} to T_{p-1} is continuous. Inversions of matrices of full rank are continuous as well, since the inequality

$$\|A^{-1} - B^{-1}\| = \|A^{-1}(B - A)B^{-1}\| \leq \|A^{-1}\|\|B - A\|\|B^{-1}\|$$

holds for all invertible A, B of the same dimensions, where $\|\cdot\|$ stands for a matrix norm, which is a submultiplicative norm, on $\mathbb{K}^{(p-1)\times(p-1)}$. Together with the fact that matrix multiplications are continuous, $(T_{p-1}, a) \mapsto T_{p-1}^{-1}a = t$ is continuous. Last, $(\alpha, t) \mapsto \sqrt{\alpha - t^{\mathsf{H}}t} = \theta$ is continuous as composition of continuous functions and as $\alpha - t^{\mathsf{H}}t \in \mathbb{R}^+$, as discussed in the end of the proof of Theorem 2.12. The extraction of submatrices is continuous as well. Therefore, ϕ^{-1} is continuous for dimension p. □

[5] A homeomorphism is a function $f\colon X \to Y$, where X, Y are topological spaces, that is bijective, continuous and where the inverse function f^{-1} is continuous as well.

The next lemma, which can be found in [17] along with its proof, states another convenient fact about positive (semi-)definite matrices.

Lemma 2.15 (The matrix square root of positive (semi-)definite matrices, see [17, Theorem 7.2.6])**.** *For $p \in \mathbb{N}$, let $A \in \mathbb{K}^{p\times p}$ be a positive semi-definite matrix. Then there exists a unique positive semi-definite $\sqrt{A} \in \mathbb{K}^{p\times p}$ such that $\sqrt{A}\sqrt{A} = A$. In case that A is positive definite, the matrix square root $\sqrt{A}$ is positive definite as well.*

2.3. Integration over the Submanifold of Positive Definite Matrices

For simplifying the notation in the following, let $\mathbb{S}_p^+$ denote the set of all real positive definite $p \times p$ matrices for $p \in \mathbb{N}$. Using the Loewner partial order given in Definition 2.10, $\mathbb{S}_p^+$ can be defined as

$$\mathbb{S}_p^+ = \{A \in \mathbb{R}^{p\times p} \mid A >_L 0\}.$$

Note that $\mathbb{S}_1^+ = \mathbb{R}^+$. The definition of the multi-dimensional gamma function denoted by $\Gamma_p(\cdot)$, that will be introduced in Section 2.5 below, uses an integral over $\mathbb{S}_p^+$. This section discusses the concept of this form of integration as used in this thesis.

As the first step we need to show that $\mathbb{S}_p^+$ is a submanifold of $\mathbb{R}^{p\cdot p} \cong \mathbb{R}^{p\times p}$.

Definition 2.16 (Submanifold of $\mathbb{R}^n$, see [10, Section 14, Satz 4])**.** For $k \leq n$ in $\mathbb{N}$ a set $M \subseteq \mathbb{R}^n$ is called a k-dimensional submanifold of $\mathbb{R}^n$ if for every point $a \in M$ there exists an open neighbourhood $V \subseteq M$ w.r.t. the relative topology[6] on M, an open set $T \subseteq \mathbb{R}^k$ and a C^1-immersion[7] $\psi\colon T \to \mathbb{R}^n$, that is a homeomorphism from T to V.

The following matrix inner product and corresponding norm will be useful for applying this concept of a submanifold to matrices.

Remark 2.17 (The Frobenius inner product and the Frobenius norm, see [28, Definition 13.9]). For $p, q \in \mathbb{N}$ and matrices $A, B \in \mathbb{K}^{p\times q}$ the Frobenius inner product is given by

$$\langle A, B\rangle_{\mathrm{F}} := \operatorname{tr}(AB^{\mathsf{H}}) = \sum_{i=1}^{p}(AB^{\mathsf{H}})_{ii} = \sum_{i=1}^{p}\sum_{j=1}^{q} A_{ij}\bar{B}_{ij}.$$

The corresponding Frobenius matrix norm for $A \in \mathbb{K}^{p\times q}$ is given by

$$\|A\|_{\mathrm{F}} = \sqrt{\langle A, A\rangle_{\mathrm{F}}} = \left(\sum_{i=1}^{p}\sum_{j=1}^{q}|A_{ij}|^2\right)^{1/2}.$$

[6] For $M \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^n$ the set $V \cap M$ is defined to be open w.r.t. the relative topology on M if V is an open set in $\mathbb{R}^n$.

[7] For $k, n \in \mathbb{N}$ and an open set $T \subseteq \mathbb{R}^k$ a continuously differentiable function $\psi\colon T \to \mathbb{R}^n$ is called C^1-immersion if the rank of the Jacobi matrix $\mathrm{d}\psi$ is equal to k for all $t \in T$.

Remark 2.18. (Vectorization of matrices, see [13, Definition 1.2.7]) Several concepts and theorems in this thesis, like the submanifolds introduced in Definition 2.16, are defined for subsets of $\mathbb{R}^n$ for $n \in \mathbb{N}$. To formally use these concepts with matrices, we define vectorization functions. Fix any $p, q \in \mathbb{N}$. For every $A \in \mathbb{K}^{p\times q}$ the vectorized version is given by

$$\operatorname{vec}(A) := \begin{pmatrix} A^1 \\ A^2 \\ \vdots \\ A^q \end{pmatrix} \in \mathbb{K}^{p\cdot q}, \tag{2.4}$$

where $A^j \in \mathbb{K}^p$ is the j-th column vector of A for all $j \in \{1, \dots, q\}$. This function is a linear bijection between $\mathbb{K}^{p\times q}$ and $\mathbb{K}^{p\cdot q}$. It is even an isomorphism of Hilbert spaces when equipping $\mathbb{K}^{p\times q}$ with the Frobenius inner product given in Remark 2.17 and $\mathbb{K}^{p\cdot q}$ with the standard inner product $\langle \cdot, \cdot \rangle$, i.e.

$$\langle A, B\rangle_{\mathrm{F}} = \operatorname{tr}(AB^{\mathsf{H}}) = \sum_{i=1}^{p}\sum_{j=1}^{q} A_{ij}\bar{B}_{ij} = \langle \operatorname{vec}(A), \operatorname{vec}(\bar{B})\rangle, \tag{2.5}$$

where the first equality follows from the definition of the Frobenius inner product given above in Remark 2.17.

Similarly, for every lower triangular (or symmetric) matrix $A \in \mathbb{K}^{p\times p}$ we define the partial vectorization

$$\operatorname{vecp}(A) = \begin{pmatrix} \tilde{A}^1 \\ \tilde{A}^2 \\ \vdots \\ \tilde{A}^p \end{pmatrix} \in \mathbb{K}^{\frac{p(p+1)}{2}}, \tag{2.6}$$

where $\tilde{A}^j = (a_{jj}, \dots, a_{pj})^{\mathsf{T}} \in \mathbb{K}^{p-j+1}$ for $j \in \{1, \dots, p\}$, are the columns of the matrix restricted to on and below the diagonal. The function $\operatorname{vecp}(\cdot)$ is a bijection when seen as a function from the set of lower triangular matrices in $\mathbb{K}^{p\times p}$ to $\mathbb{K}^{p(p+1)/2}$. When seen as a function from the set of symmetric matrices in $\mathbb{K}^{p\times p}$ to $\mathbb{K}^{p(p+1)/2}$ it is also bijective.

Remark 2.19 (Introduction of $\tilde{\mathbb{S}}_p^+$). For $p \in \mathbb{N}$, let ψ be the map that translates the positive definite, hence symmetric, matrices seen as their lower triangular version, denoted by $\tilde{\mathbb{S}}_p^+$, to the subset $\mathbb{S}_p^+$ of $\mathbb{R}^{p\cdot p} \cong \mathbb{R}^{p\times p}$,

$$\psi := \begin{cases} \tilde{\mathbb{S}}_p^+ \subseteq \mathbb{R}^{\frac{p(p+1)}{2}} & \to \quad \mathbb{S}_p^+ \subseteq \mathbb{R}^{p\cdot p} \cong \mathbb{R}^{p\times p} \\ \begin{pmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ \vdots & & & \ddots \\ a_{p1} & & \cdots & & a_{pp} \end{pmatrix} & \mapsto \begin{pmatrix} a_{11} & a_{21} & \cdots & \cdots & a_{p1} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{p2} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ a_{p1} & & \cdots & & a_{pp} \end{pmatrix} \end{cases}. \tag{2.7}$$

The lower triangular matrix in the domain is understood as an element of $\tilde{\mathbb{S}}_p^+$ via the isomorphism $\operatorname{vecp}(\cdot)$ discussed in Remark 2.18 and will be denoted by $A^{\triangle}$. So $A^{\triangle} \in \tilde{\mathbb{S}}_p^+$ if and only if $\psi(A^{\triangle}) \in \mathbb{S}_p^+$. This map ψ is a homeomorphism from $\tilde{\mathbb{S}}_p^+$ to $\mathbb{S}_p^+$ as it is bijective and the restriction of a linear map between the finite-dimensional vector spaces $\mathbb{R}^{p(p+1)/2}$ and $\mathbb{R}^{p \cdot p} \cong \mathbb{R}^{p \times p}$.

Lemma 2.20 ($\tilde{\mathbb{S}}_p^+$ is an open set)**.** *For $p \in \mathbb{N}$, let $\mathbb{T}_p$ denote the set of all lower triangular matrices in $\mathbb{R}^{p \times p}$ with positive diagonal entries,*

$$\mathbb{T}_p := \{T \in \mathbb{R}^{p \times p} \mid t_{11}, \dots, t_{pp} \in \mathbb{R}^+ \text{ and } t_{ij} = 0 \text{ for all } i < j \text{ in } \{1, \dots, p\}\}.$$

and let $\tilde{\mathbb{T}}_p$ be its vectorized version

$$\tilde{\mathbb{T}}_p = \operatorname{vecp}(\mathbb{T}_p). \tag{2.8}$$

Then both $\tilde{\mathbb{T}}_p$ and the set $\tilde{\mathbb{S}}_p^+$ are open subsets of $\mathbb{R}^{p(p+1)/2}$.

Proof. Using Eq. (2.6) and Eq. (2.8)

$$\operatorname{vecp}(\mathbb{T}_p) = \mathbb{R}^+ \times \mathbb{R}^{p-1} \times \mathbb{R}^+ \times \mathbb{R}^{p-2} \times \cdots \times \mathbb{R}^+,$$

which is an open subset of $\mathbb{R}^{p(p+1)/2}$.

To see that $\tilde{\mathbb{S}}_p^+$ is an open set, recall that ψ given in Eq. (2.7) is bijective, and by Item (vii) of Lemma 2.8, $A^{\triangle} \in \mathbb{R}^{p(p+1)/2}$ is in $\tilde{\mathbb{S}}_p^+$ if and only if there exists a $\delta > 0$ such that

$$x^{\mathsf{T}} \psi(A^{\triangle}) x \geq \delta, \quad x \in \mathbb{R}^p \text{ with } \|x\|_2 = 1,$$

where $\|\cdot\|_2$ is the Euclidian norm. Fix $A^{\triangle} \in \tilde{\mathbb{S}}_p^+$. The function ψ given in Eq. (2.7) can be extended to a linear map from $\mathbb{R}^{p(p+1)/2}$ to $\mathbb{R}^{p \cdot p}$ with the same definition. Then every vector $B^{\triangle} \in \mathbb{R}^{p(p+1)/2}$ with $\|A^{\triangle} - B^{\triangle}\|_2 < \delta/2$ satisfies

$$|x^{\mathsf{T}} \psi(B^{\triangle} - A^{\triangle}) x| \leq \|x\|_2 \underbrace{\|\psi\|}_{\leq 2} \|B^{\triangle} - A^{\triangle}\|_2 \|x\|_2 < \delta, \quad x \in \mathbb{R}^p \text{ with } \|x\|_2 = 1,$$

for all $x \in \mathbb{R}^p$ with $\|x\|_2 = 1$, where $\|\psi\|$ is the operator norm of $\psi\colon \mathbb{R}^{p(p+1)/2} \to \mathbb{R}^{p \cdot p}$ and the domain and the range are equipped with the respective Euclidian norm. Hence,

$$x^{\mathsf{T}} \psi(B^{\triangle}) x = x^{\mathsf{T}} \psi(B^{\triangle} - A^{\triangle}) x + x^{\mathsf{T}} \psi(A^{\triangle}) x > -\delta + \delta = 0.$$

As this inequality holds for all $x \in \mathbb{R}^p$ with $\|x\|_2 = 1$, it also holds for all $x \in \mathbb{R}^p \setminus \{0\}$. □

Remark 2.21 ($\mathbb{S}_p^+$ is a submanifold of $\mathbb{R}^{p \cdot p} \cong \mathbb{R}^{p \times p}$). For $p \in \mathbb{N}$, let ψ be the homeomorphism between $\tilde{\mathbb{S}}_p^+$ and $\mathbb{S}_p^+$ given in Eq. (2.7). As ψ maps lower triangular matrices seen as elements of $\mathbb{R}^{p(p+1)/2}$ to the same matrix with additional elements above the diagonal, the Jacobi matrix of ψ has rank $p(p+1)/2$. Hence, using Definition 2.16 and the fact that $\tilde{\mathbb{S}}_p^+$ is an open set in $\mathbb{R}^{p(p+1)/2}$ due to Lemma 2.20, $\mathbb{S}_p^+$ is a $p(p+1)/2$-dimensional submanifold of $\mathbb{R}^{p \cdot p} \cong \mathbb{R}^{p \times p}$.

Remark 2.22 (Integration on submanifolds of $\mathbb{R}^n$, see [10, Section 14]). For $k, n \in \mathbb{N}$ let M be a k-dimensional submanifold of $\mathbb{R}^n$. Then a C^1-immersion and homeomorphism

$$\psi\colon T \to V \subseteq M,$$

where T is an open subset of $\mathbb{R}^k$ and V is open w.r.t. the relative topology on M, is called chart. The *Gramian determinant* $g(t)$ corresponding to $\psi = (\psi_1, \dots, \psi_n)$ is defined as the determinant of the matrix $G(t) = (g_{i,j}(t))_{i,j=1}^k$, where each entry is given by

$$g_{i,j}(t) = \left\langle \frac{\partial \psi(t)}{\partial t_i}, \frac{\partial \psi(t)}{\partial t_j} \right\rangle = \sum_{v=1}^{n} \frac{\partial \psi_v(t)}{\partial t_i} \cdot \frac{\partial \psi_v(t)}{\partial t_j}, \quad t \in T \text{ and } i, j \in \{1, \dots, k\}. \tag{2.9}$$

A function $f\colon M \to \mathbb{R}$, that fulfills $f{\restriction_{M\setminus V}} = 0$ for a given chart $\psi\colon T \to V$ is called integrable over the submanifold M if the function

$$T \ni t \mapsto f\big(\psi(t)\big)\sqrt{g(t)}$$

is integrable over T w.r.t. the Lebesgue–Borel measure λ^k on $\mathbb{R}^k$ restricted to T. In this case the integral is defined to be

$$\int_M f(x)\,\mathrm{d}x = \int_T f\big(\psi(t)\big)\sqrt{g(t)}\,\lambda^k(\mathrm{d}t).$$

This definition does not depend on the choice of the chart ψ as shown in [10, Secton 14, p. 168]. This form of integration can be generalized, relaxing the condition $f{\restriction_{M\setminus V}} = 0$ for a single chart $\psi\colon T \to V$, by using a partition of unity. As this will not be required for the submanifold $\mathbb{S}_p^+$, it will not be discussed further in this thesis and the interested reader might consult [10, Secton 14, p. 168].

Remark 2.23 (Integration over $\mathbb{S}_p^+$, the set of real positive definite matrices). This remark describes how for $f\colon \mathbb{S}_p^+ \to \mathbb{K}$, satisfying certain integrability conditions, an integral

$$\int_{\mathbb{S}_p^+} f(A)\,\mathrm{d}A$$

over the submanifold $\mathbb{S}_p^+$ can be transformed to a Lebesgue–Borel integral.

First, let f take values in the real numbers. As described in Remark 2.22, we integrate over the $p(p+1)/2$-dimensional submanifold $\mathbb{S}_p^+$ of $\mathbb{R}^{p\cdot p}$ using a chart. For further calculations we use the chart ψ defined in Eq. (2.7). We refer to elements of the domain $\tilde{\mathbb{S}}_p^+$ of ψ as $A^{\triangle}$ and to those in the range $\mathbb{S}_p^+$ as A. Note that the domain $\tilde{\mathbb{S}}_p^+$ is an open subset of $\mathbb{R}^{p(p+1)/2}$, see Lemma 2.20. As the range of this chart is the entire submanifold $\mathbb{S}_p^+$, the basic criterion $f{\restriction_{M\setminus V}} = 0$, discussed in Remark 2.22, for integrating a function f over a submanifold M, is always fulfilled. Therefore,

$$\int_{\mathbb{S}_p^+} f(A)\,\mathrm{d}A = \int_{\tilde{\mathbb{S}}_p^+} f\big(\psi(A^{\triangle})\big)\,\sqrt{g(A^{\triangle})}\,\lambda^{\frac{p(p+1)}{2}}(\mathrm{d}A^{\triangle}), \tag{2.10}$$

for all $f\colon \mathbb{S}_p^+ \to \mathbb{R}$, for which the integral on the right side is well-defined, with $g(A^{\triangle})$ referring to the Gramian determinant of the chart ψ and $\lambda^{p(p+1)/2}$ being the $p(p+1)/2$-dimensional Lebesgue–Borel measure.

The next step is to determine the Gramian determinant of this chart ψ. According to Remark 2.22, we start by taking the partial derivatives of the components of the chart. For $A^{\triangle} \in \tilde{\mathbb{S}}_p^+$ and $i, j \in \{1, \dots, p\}$,

$$\psi_{ij}(A^{\triangle}) = 1_{\{i \geq j\}} a_{ij} + 1_{\{i<j\}} a_{ji},$$

where a_{mn} refers to the entry of $A^{\triangle}$ for all m, n in $\{1, \dots, p\}$ with $m \geq n$. Hence, for all $i, j, k, l \in \{1, \dots, p\}$ with $k \geq l$,

$$\frac{\partial \psi_{ij}(A^{\triangle})}{\partial a_{kl}} = 1_{\{i \geq j\}} 1_{\{i=k, j=l\}} + 1_{\{i<j\}} 1_{\{i=l, j=k\}}.$$

For the entries in the matrix $G(A^{\triangle})$, given in Eq. (2.9), we calculate for $k \geq l$ in $\{1, \dots, p\}$ and $m \geq n$ in $\{1, \dots, p\}$

$$\left\langle \frac{\partial \psi(A^{\triangle})}{\partial a_{kl}}, \frac{\partial \psi(A^{\triangle})}{\partial a_{mn}} \right\rangle_{\mathrm{F}} = \sum_{i,j=1}^{p} \frac{\partial \psi_{ij}(A^{\triangle})}{\partial a_{kl}} \cdot \frac{\partial \psi_{ij}(A^{\triangle})}{\partial a_{mn}} = \begin{cases} 1 & \text{if } k = l = m = n, \\ 2 & \text{if } k = m > l = n, \\ 0 & \text{otherwise}, \end{cases} \tag{2.11}$$

where the Frobenius inner product is used instead of vectorizing $\psi(A^{\triangle})$ as described in Eq. (2.5). To compute the determinant of the matrix $G(A^{\triangle})$, it is advantageous to vectorize the entries of $A^{\triangle}$ as described in Eq. (2.6) of Remark 2.18. As the Frobenius inner product in Eq. (2.11) is only non-zero if $(k, l) = (m, n)$, the matrix $G(A^{\triangle})$ is a diagonal matrix with p times entry 1 and $p(p+1)/2 - p$ times entry 2 on the diagonal. Hence, the determinant is equal to the product of the diagonal entries, resulting in

$$g(A^{\triangle}) = \det(G(A^{\triangle})) = 2^{\frac{p(p-1)}{2}}.$$

Inserting this into Eq. (2.10) results in

$$\int_{\mathbb{S}_p^+} f(A)\, \mathrm{d}A = 2^{\frac{p(p-1)}{4}} \int_{\tilde{\mathbb{S}}_p^+} f\big(\psi(A^{\triangle})\big)\, \lambda^{\frac{p(p+1)}{2}}(\mathrm{d}A^{\triangle}), \tag{2.12}$$

if $f\big(\psi(A^{\triangle})\big)$ is integrable over $\tilde{\mathbb{S}}_p^+$ w.r.t. $\lambda^{p(p+1)/2}$.

Now for a $f\colon \mathbb{S}_p^+ \to \mathbb{C}$, we can split up the function into its real and imaginary part,

$$f = f_1 + \mathrm{i} f_2,$$

where $f_1, f_2\colon \mathbb{S}_p^+ \to \mathbb{R}$. Then the integral of f over $\mathbb{S}_p^+$ is given by

$$\begin{aligned} \int_{\mathbb{S}_p^+} f(A)\, \mathrm{d}A &= \int_{\mathbb{S}_p^+} f_1(A)\, \mathrm{d}A + \mathrm{i} \int_{\mathbb{S}_p^+} f_2(A)\, \mathrm{d}A \\ &= 2^{\frac{p(p-1)}{4}} \int_{\tilde{\mathbb{S}}_p^+} f_1\big(\psi(A^{\triangle})\big)\, \lambda^{\frac{p(p+1)}{2}}(\mathrm{d}A^{\triangle}) + \mathrm{i}\, 2^{\frac{p(p-1)}{4}} \int_{\tilde{\mathbb{S}}_p^+} f_2\big(\psi(A^{\triangle})\big)\, \lambda^{\frac{p(p+1)}{2}}(\mathrm{d}A^{\triangle}). \end{aligned}$$

Hence, using the linearity of the integral, Eq. (2.12) holds for complex valued f as well, if both its real part $f_1\big(\psi(A^{\triangle})\big)$ and its imaginary part $f_2\big(\psi(A^{\triangle})\big)$ are integrable over $\tilde{\mathbb{S}}_p^+$ w.r.t. $\lambda^{p(p+1)/2}$.

2.4. One-Dimensional Gamma Distributions and their Characteristic Function

In this section one-dimensional gamma distributions will be introduced as well as their characteristic function, as a corollary to Lemma 2.33 below. We will also discuss the special case of χ^2-distributions, whose matrix-valued generalization, called Wishart distribution, will be discussed in Section 3.1 of the next chapter.

First, we will show that the one-dimensional gamma function, which is needed for the definition of the gamma distribution, is well defined, which is also a prerequisite to demonstrate that the multi-dimensional gamma function is well-defined.

Remark 2.24 (One-dimensional gamma function). The one-dimensional gamma function is given by

$$\Gamma(\alpha) := \int_0^\infty t^{\alpha-1}\, \mathrm{e}^{-t}\, \mathrm{d}t, \quad z \in \mathbb{C}, \quad \mathrm{Re}(\alpha) > 0.^{8}$$

As a first step to show that this integral is well-defined, we prove that it is finite for fixed real $\alpha > 0$. Let $n := \lceil \alpha \rceil + 1 \in \mathbb{N}$ and using this let $t_0 := \sqrt[n]{(2n)!} \in \mathbb{R}^+$. Then all $t \geq t_0$ satisfy the inequality $t^n/(2n)! \geq 1$ and therefore

$$t^{\alpha+1} \leq t^n \leq t^n \frac{t^n}{2n!} = \frac{t^{2n}}{(2n)!} \leq \mathrm{e}^t, \quad t \geq t_0,$$

due to the power series definition of the exponential function. This can be rearranged to

$$t^{\alpha-1}\, \mathrm{e}^{-t} \leq \frac{1}{t^2}, \quad t \geq t_0. \tag{2.13}$$

We split the integral into two parts at t_0,

$$\Gamma(\alpha) = \int_0^{t_0} t^{\alpha-1}\, \mathrm{e}^{-t}\, \mathrm{d}t + \int_{t_0}^\infty t^{\alpha-1}\, \mathrm{e}^{-t}\, \mathrm{d}t.$$

The first integral is finite as

$$\int_0^{t_0} t^{\alpha-1}\, \mathrm{e}^{-t}\, \mathrm{d}t \leq \int_0^{t_0} t^{\alpha-1}\, \mathrm{d}t = \frac{(t_0)^\alpha}{\alpha} < \infty$$

as $\mathrm{e}^{-t} \leq 1$ for all $t \geq 0$. Due to Eq. (2.13) the second integral can be estimated by

$$\int_{t_0}^\infty t^{\alpha-1}\, \mathrm{e}^{-t}\, \mathrm{d}t \leq \int_{t_0}^\infty \frac{1}{t^2}\, \mathrm{d}t = \frac{1}{t_0} < \infty.$$

Hence, the gamma function is well-defined for real $\alpha > 0$.

Now for $\alpha \in \mathbb{C}$ with $x := \mathrm{Re}(\alpha) > 0$, it is also well-defined as

$$|t^{\alpha-1}| = |\mathrm{e}^{(\alpha-1)\log(t)}| = \mathrm{e}^{\mathrm{Re}((\alpha-1)\log(t))} = \mathrm{e}^{(x-1)\log(t)} = t^{x-1}, \quad t > 0,$$

where $\log(\cdot)$ denotes the natural logarithm on $\mathbb{R}^+$.

[8] Note that the gamma function can be defined for all complex α except for zero and negative integers, using the analytic continuation of the integral given here, see [18, p.219] for example.

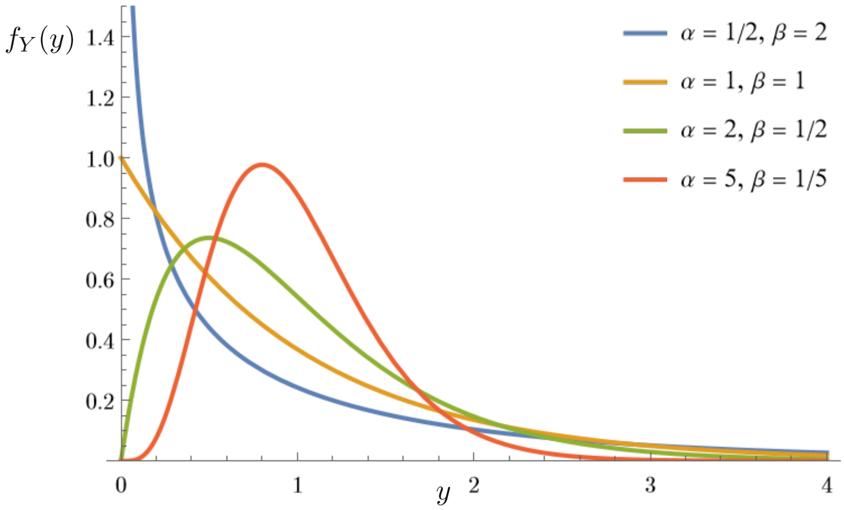


Figure 2.1. *Continuous probability density functions of one-dimensional gamma distributions with different shape and scale parameters and the same expectation, see Eq.* (2.15).

Definition 2.25 (One-dimensional gamma distributions)**.** For $\alpha \in \mathbb{R}^+$ an $\mathbb{R}^+$-valued random variable X with density

$$f_X(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \, \mathrm{e}^{-x}, \quad x \in \mathbb{R}^+,$$

is said to follow the standard one-dimensional gamma distribution with shape parameter α. For $\beta \in \mathbb{R}_0^+$ the $\mathbb{R}_0^+$-valued random variable Y with stochastic representation

$$Y \stackrel{\mathrm{d}}{=} \beta X \quad \text{(equality in distribution)}$$

is said to follow the one-dimensional gamma distribution with shape parameter α and scale parameter β. This is denoted by $Y \sim \mathrm{Gamma}(\alpha, \beta)$. We will also include the cases $\alpha = 0$ by letting $\mathrm{Gamma}(0, \beta)$ denote the degenerate distribution concentrated in zero.

There exists a more general version of one-dimensional gamma distributions, called generalized gamma distribution, which additionally involves a power $p \in \mathbb{R}^+$, which is equal to one for normal gamma distributions.

Definition 2.26 (Generalized gamma distribution, see [28, Exercise 2.37])**.** For parameters $\alpha, p \in \mathbb{R}^+$ we say that an $\mathbb{R}^+$-valued random variable X with density

$$f_X(x) = \frac{p}{\Gamma(\alpha)} x^{\alpha p-1} \, \mathrm{e}^{-x^p}, \quad x \in \mathbb{R}^+,$$

follows the standard generalized gamma distribution with shape parameter α and power p. For $\beta \in \mathbb{R}_0^+$ the $\mathbb{R}_0^+$-valued random variable Y with stochastic representation

$$Y \stackrel{\mathrm{d}}{=} \beta X \quad \text{(equality in distribution)}$$

is said to follow the generalized gamma distribution with shape parameter α, scale parameter β and power p, denoted by $Y \sim \mathrm{GenGamma}(\alpha, \beta, p)$. Again, we include the case $\alpha = 0$ by letting $\mathrm{GenGamma}(0, \beta)$ denote the degenerate distribution concentrated in zero.

Remark 2.27 (Density function of (generalized) gamma distributions). If $\beta \neq 0$, the random variable Y given in Definition 3.34 via the stochastic representation $Y \overset{\mathrm{d}}{=} \beta X$ with $X \sim \mathrm{GenGamma}(\alpha, 1, p)$ has a continuous density function. For every Borel set $B \in \mathcal{B}(\mathbb{R}^+)$,

$$\mathbb{P}[Y \in B] = \mathbb{P}[\beta X \in B] = \mathbb{P}[X \in \{b/\beta \,|\, b \in B\}] = \int_{\{b/\beta|b\in B\}} f_X(x)\,\mathrm{d}x = \int_B f_X(y/\beta)\beta\,\mathrm{d}y,$$

where the substitution $\mathbb{R}^+ \ni y := \beta x$ is used for the last equality. As this holds true for every Borel set $B \in \mathcal{B}(\mathbb{R}^+)$,

$$f_Y(y) = f_X(y/\beta)\beta = \frac{p}{\Gamma(\alpha)}\left(\frac{y}{\beta}\right)^{\alpha p-1} \mathrm{e}^{-(y/\beta)^p}\,\beta,$$

where we use the density function of X given in Definition 2.26. Hence,

$$f_Y(y) = \frac{p}{\beta^{\alpha p}\Gamma(\alpha)} y^{\alpha p-1}\,\mathrm{e}^{-(y/\beta)^p}, \quad y \in \mathbb{R}^+, \tag{2.14}$$

for $Y \sim \mathrm{GenGamma}(\alpha, \beta, p)$. By setting $p = 1$, we can derive the well-known continuous density of the one-dimensional gamma distribution with shape parameter α and scale parameter β to be

$$f_Y(y) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} y^{\alpha-1}\,\mathrm{e}^{-y/\beta}, \quad y \in \mathbb{R}^+. \tag{2.15}$$

Definition 2.28 (χ^2-distribution)**.** For $n \in \mathbb{N}_0$ let $Z_1, \ldots, Z_n$ be i.d.d. random variables following a standard normal distribution. Then a random variable

$$X \overset{\mathrm{d}}{=} Z_1^2 + \cdots + Z_n^2$$

is said to be χ^2-distributed with n degrees of freedom, with the convention that the sum over an empty set is equal to zero. This is denoted by $X \sim \chi_n^2$.

Remark 2.29 (χ^2-distributions are a subclass of gamma distributions). For $n \in \mathbb{N}$ a random variable $X \sim \chi_n^2$ follows the gamma distribution with shape parameter $\alpha = n/2$ and scale parameter $\beta = 2$. This is proven by the fact that the characteristic function determines the distribution uniquely, see Remark 3.4 for a further discussion, and Corollaries 2.34 and 2.35 below.

Remark 2.30 (One-dimensional exponential distributions). Another subclass of the one-dimensional gamma distributions is the family of exponential distributions. An $\mathbb{R}_0^+$-valued random variable Z is said to follow the exponential distribution with parameter $\beta \in \mathbb{R}_0^+$, if $Z \sim \mathrm{Gamma}(1, \beta)$. Therefore, in case $\beta > 0$, a density of Z is given by

$$f_Z(z) = \frac{1}{\beta}\,\mathrm{e}^{-z/\beta}, \quad z \in \mathbb{R}^+.$$

Exponential distributions have the powerful property of being memoryless, i.e. for all $s, t \geq 0$,

$$\mathbb{P}[Z > s + t \,|\, Z > s] = \mathbb{P}[Z > t].$$

In Lemma 3.49 below we will prove that matrix-valued exponential distributions, a subclass of matrix-valued gamma distributions given in Definition 3.47, also possess a memorylessness property using the Loewner order given in Definition 2.10.

In order to look at the characteristic function of gamma distributions we need to define complex logarithms.

Remark 2.31 (The complex logarithm and complex exponentiation, see [18, Section 4.3]). The complex logarithms of a complex number $z \in \mathbb{C} \setminus \{0\}$ are given by all $w \in \mathbb{C}$ solving

$$e^{w} = z.$$

This equation has infinitely many solutions as the functional equation of the exponential function and

$$e^{2\pi i k} = 1, \quad k \in \mathbb{Z}, \tag{2.16}$$

imply that for every complex logarithm $w \in \mathbb{C}$ of $z \in \mathbb{C} \setminus \{0\}$, the complex number $\tilde{w} = w + 2\pi i k$ for $k \in \mathbb{Z}$ also satisfies $e^{\tilde{w}} = z$. With the help of a polar form of $z \in \mathbb{C} \setminus \{0\}$ given by $z = r\, e^{i\theta}$ with $\theta \in \mathbb{R}$ and $r := |z| > 0$ all complex logarithms of z are given by

$$\log(r) + i(\theta + 2\pi k), \quad k \in \mathbb{Z},$$

where $\log(\cdot)$ refers to the natural logarithm on $\mathbb{R}^+$. The *principal value logarithm*, which extends the natural logarithm for $z \in \mathbb{R}^+$, refers to the complex logarithm of $z \in \mathbb{C} \setminus \{0\}$ with imaginary part in the interval $(-\pi, \pi]$ and will be denoted by $\operatorname{Log}(\cdot)$. Note that for $z_1, z_2 \in \mathbb{C} \setminus \{0\}$, which can be written as $z_j = r_j\, e^{i\theta_j}$ with $r_j := |z_j|$ and $\theta_j \in \mathbb{R}$ for $j \in \{1, 2\}$, their product can be represented by

$$z_1 z_2 = r_1 r_2\, e^{i(\theta_1 + \theta_2)}.$$

The sum of the principal logarithms of z_1 and z_2 satisfies

$$\begin{aligned} \operatorname{Log}(z_1) + \operatorname{Log}(z_2) &= \log(r_1) + i(\theta_1 + 2\pi k_1) + \log(r_2) + i(\theta_2 + 2\pi k_2) \\ &= \log(r_1 r_2) + i(\theta_1 + \theta_2 + 2\pi(k_1 + k_2)) = \operatorname{Log}(z_1 z_2) + 2\pi i k \end{aligned} \tag{2.17}$$

for certain $k_1, k_2, k \in \mathbb{Z}$. As $(\theta_1 + 2\pi k_1) \in (-\pi, \pi]$ and $(\theta_2 + 2\pi k_2) \in (-\pi, \pi]$ does not necessarily imply that their sum is in the interval $(-\pi, \pi]$, the integer k in Eq. (2.17) can be different from zero.

For this thesis we define complex exponentiation using the principal value logarithm, i.e. for all $\alpha \in \mathbb{C}$,

$$z^{\alpha} := e^{\alpha \operatorname{Log}(z)}, \quad z \in \mathbb{C} \setminus \{0\}. \tag{2.18}$$

Note that due to Eq. (2.17), the common power identities $a^x b^x = (ab)^x$ and $(a^x)^y = a^{xy}$, which hold for $a, b \in \mathbb{R}^+$ and $x, y \in \mathbb{R}$, do not necessarily hold if some of the parameters are complex due to Eq. (2.17) or already if a or b is negative. This can be seen by the counterexample $((-1)^2)^{1/2} = 1 \neq -1 = (-1)^1$. The properties that do hold are given in the following proposition.

Proposition 2.32 (Properties of complex exponentiation). *Note that the following equalities hold for all $a, b \in \mathbb{C} \setminus \{0\}$ and $x, y \in \mathbb{C}$ satisfying the corresponding restrictions.*

(i) $(a^x)^y = a^{xy}$ *if* $y \in \mathbb{N}_0$.

(ii) $(a^x)^y = a^{xy}$ *if* $x \in (-1, 1]$,

(iii) $a^x a^y = a^{x+y}$,

(iv) $a^x b^x = (ab)^x$ *if* a *or* b *is real and positive,*

(v) $a^x b^x = (ab)^x$ *if* $x \in \mathbb{Z}$,

(vi) $a^x b^x = \pm(ab)^x$ *if* $2x \in \mathbb{Z}$.

Proof. Item (i): If $y = 0$, then both sides of the equality are equal to one. If $y \in \mathbb{N}$, then

$$(a^x)^y = \underbrace{a^x \cdots a^x}_{y \text{ times}} = \mathrm{e}^{x\mathrm{Log}(a)} \cdots \mathrm{e}^{x\mathrm{Log}(a)} = \mathrm{e}^{(x+\cdots+x)\mathrm{Log}(a)} = \mathrm{e}^{xy\,\mathrm{Log}(a)} = a^{xy}.$$

Item (ii): Note that both $z_1 \coloneqq x\mathrm{Log}(a)$ and $z_2 \coloneqq \mathrm{Log}(a^x)$ solve $\mathrm{e}^{z_j} = a^x$, $j \in \{1,2\}$, hence they can only differ by $2\pi \mathrm{i} k$ with $k \in \mathbb{Z}$ as discussed in Remark 2.31. For a fixed $a \in \mathbb{C}$ let $\theta \in \mathbb{R}$ satisfy $\mathrm{Log}(a) = \log(|a|) + \mathrm{i}\theta$, which means that $a = |a|\,\mathrm{e}^{\mathrm{i}\theta}$ and $\theta \in (-\pi, \pi]$. Then $x\theta \in (-\pi, \pi]$ for $x \in (-1, 1]$ and therefore

$$x\mathrm{Log}(a) = \log(|a|^x) + \mathrm{i}x\theta = \mathrm{Log}(a^x).$$

Item (iii): This holds true as $a^x a^y = \mathrm{e}^{x\mathrm{Log}(a)+y\mathrm{Log}(a)} = \mathrm{e}^{(x+y)\mathrm{Log}(a)} = a^{x+y}$ due to the functional equation of the exponential function. Item (iv): W.l.o.g. assume that $a \in \mathbb{R}^+$. Then

$$a^x b^x = \mathrm{e}^{x\mathrm{Log}(a)}\,\mathrm{e}^{x\mathrm{Log}(b)} = \mathrm{e}^{x(\log(a)+\mathrm{Log}(b))} = \mathrm{e}^{x\mathrm{Log}(ab)} = (ab)^x$$

as the imaginary part of the logarithm of a is equal to zero and hence $\log(a) + \mathrm{Log}(b) = \mathrm{Log}(ab)$.

Item (v): Due to Eq. (2.17) there exists a $k \in \mathbb{Z}$ such that $\mathrm{Log}(a)+\mathrm{Log}(b) = \mathrm{Log}(ab)+2\pi\mathrm{i}k$. Hence,

$$a^x b^x = \mathrm{e}^{x\mathrm{Log}(a)}\,\mathrm{e}^{x\mathrm{Log}(b)} = \mathrm{e}^{x(\mathrm{Log}(a)+\mathrm{Log}(b))} = \mathrm{e}^{x\mathrm{Log}(ab)+x2\pi\mathrm{i}k} = (ab)^x\,\mathrm{e}^{2\pi\mathrm{i}kx} = (ab)^x,$$

where the last step follows from Eq. (2.16) as $kx \in \mathbb{Z}$.

Item (vi): The same procedure as in the Item (v) with the difference that $\mathrm{e}^{\pi\mathrm{i}2kx} = \pm 1$ for $2kx \in \mathbb{Z}$. □

The following lemma is on the one hand used for deriving the characteristic function of one-dimensional gamma distributions and χ^2-distributions and, on the other hand, useful for the proof of Theorem 2.37 below.

Lemma 2.33. *For every* $t \in \mathbb{R}$ *and* $\alpha \in \mathbb{C}$ *satisfying* $\mathrm{Re}(\alpha) > 0$,

$$2\int_{\mathbb{R}^+} x^{2\alpha-1}\,\mathrm{e}^{-x^2(1+\mathrm{i}t)}\,\mathrm{d}x = \int_{\mathbb{R}^+} z^{\alpha-1}\,\mathrm{e}^{-z(1+\mathrm{i}t)}\,\mathrm{d}z = (1+\mathrm{i}t)^{-\alpha}\,\Gamma(\alpha), \tag{2.19}$$

where the complex exponentiation is defined in Eq. (2.18).

The proof of this lemma is given after the following two corollaries, which derive well-known characteristic functions.

Corollary 2.34 (Characteristic function of one-dimensional gamma distributions)**.** *For* $\alpha, \beta \in \mathbb{R}_0^+$ *let* $Y \sim \mathrm{Gamma}(\alpha, \beta)$*. Then the characteristic function is given by*

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tY}] = (1 - \mathrm{i}\beta t)^{-\alpha}, \quad t \in \mathbb{R},$$

where the complex exponentiation is defined in Eq. (2.18)*.*

Proof. Using Eq. (2.15), the characteristic function of Y for $\alpha, \beta \in \mathbb{R}^+$ is given by

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tY}] = \int_{\mathbb{R}^+} \mathrm{e}^{\mathrm{i}ty} f_Y(y)\,\mathrm{d}y = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_{\mathbb{R}^+} y^{\alpha-1}\,\mathrm{e}^{-\frac{y}{\beta}(1-\mathrm{i}\beta t)}\,\mathrm{d}y, \quad t \in \mathbb{R}.$$

By substituting $\mathbb{R}^+ \ni z := y/\beta$,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tY}] = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_{\mathbb{R}^+} (\beta z)^{\alpha-1}\,\mathrm{e}^{-z(1-\mathrm{i}\beta t)}\,\beta\,\mathrm{d}z = (1 - \mathrm{i}\beta t)^{-\alpha}, \quad t \in \mathbb{R},$$

where we use Eq. (2.19) for the last equality. For the degenerate cases $0 \in \{\alpha, \beta\}$, the random variable Y is equal to zero with probability one, hence the characteristic function is constant and equal to one. □

Corollary 2.35 (Characteristic function of χ^2-distributions)**.** *For* $n \in \mathbb{N}$ *let* $X \sim \chi^2_n$*. Then the characteristic function is given by*

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}] = \Big(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{z^2(1-2\mathrm{i}t)}{2}}\,\mathrm{d}z\Big)^n = (1 - 2\mathrm{i}t)^{-n/2}, \quad t \in \mathbb{R}, \tag{2.20}$$

where the complex exponentiation is defined in Eq. (2.18)*.*

Proof. Due to Definition 2.28 X can be written as $X \overset{\mathrm{d}}{=} Z_1^2 + \cdots + Z_n^2$, where $Z_1, \ldots, Z_n$ are i.i.d. $\mathcal{N}(0, 1)$ random variables. Hence,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}] = \mathbb{E}[\mathrm{e}^{\mathrm{i}t(Z_1^2+\cdots+Z_n^2)}] = \big(\mathbb{E}[\mathrm{e}^{\mathrm{i}tZ_1^2}]\big)^n, \quad t \in \mathbb{R}. \tag{2.21}$$

By inserting the continuous density of the standard normal distribution each of the factors is given by

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tZ_1^2}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}z^2(1-2\mathrm{i}t)}\,\mathrm{d}z, = \frac{2}{\sqrt{\pi}} \int_{\mathbb{R}^+} \mathrm{e}^{-y^2(1-2\mathrm{i}t)}\,\mathrm{d}y = (1-2\mathrm{i}t)^{-1/2}, \quad t \in \mathbb{R}, \tag{2.22}$$

where we substitute $\mathbb{R} \ni y := z/\sqrt{2}$ and use the symmetry of the integrand for the second equation and the last equality holds due to Eq. (2.19) with $\alpha = 1/2$ combined with the fact that $\Gamma(1/2) = \sqrt{\pi}$. Now with Item (i) of Proposition 2.32, Eq. (2.20) follows from Eqs. (2.21) and (2.22). □

Proof of Lemma 2.33. We start by using integration by substitution, to integrate over $\mathbb{R}^+ \ni z := x^2$ for the left side of Eq. (2.19),

$$2\int_{\mathbb{R}^+} x^{2\alpha-1}\,\mathrm{e}^{-x^2(1+\mathrm{i}t)}\,\mathrm{d}x = 2\int_{\mathbb{R}^+} z^{\alpha-\frac{1}{2}}\,\mathrm{e}^{-z(1+\mathrm{i}t)}\,\frac{1}{2\sqrt{z}}\,\mathrm{d}z = \int_{\mathbb{R}^+} z^{\alpha-1}\,\mathrm{e}^{-z(1+\mathrm{i}t)}\,\mathrm{d}z. \tag{2.23}$$

The function

$$f := \begin{cases} D := \mathbb{C}\setminus(-\infty,0] & \to \quad \mathbb{C} \\ \qquad z & \mapsto \quad z^{\alpha-1}\,\mathrm{e}^{-z} \end{cases} \tag{2.24}$$

is holomorphic since $z^{\alpha-1} = \mathrm{e}^{(\alpha-1)\mathrm{Log}(z)}$, the principal value logarithm $\mathrm{Log}(\cdot)$ is holomorphic on D and the composition of holomorphic functions is holomorphic. Note that for every complex number $z \in D$, satisfying $\mathrm{Re}(z) > 0$, and $r \in \mathbb{R}^+$,

$$(rz)^{\alpha-1} = r^{\alpha-1}z^{\alpha-1}$$

due to Item (iv) of Proposition 2.32 as $r \in \mathbb{R}^+$. Therefore,

$$\begin{aligned}\lim_{r\to\infty}|f(rz)| &= \lim_{r\to\infty}|(rz)^{\alpha-1}\,\mathrm{e}^{-rz}| = |z^{\alpha-1}|\lim_{r\to\infty}|r^{\alpha-1}|\,|\mathrm{e}^{-rz}| \\ &= |z^{\alpha-1}|\lim_{r\to\infty} r^{\mathrm{Re}(\alpha)-1}\,\mathrm{e}^{-r\cdot\mathrm{Re}(z)} = 0,\end{aligned} \tag{2.25}$$

where we use $|r^{\alpha-1}| = \mathrm{e}^{\mathrm{Re}((\alpha-1)\mathrm{Log}(r))} = \mathrm{e}^{(\mathrm{Re}(\alpha)-1)\log(r)}$, since $r \in \mathbb{R}^+$. For $w := (1+\mathrm{i}t)/|1+\mathrm{i}t|$ and $r < R$ in $\mathbb{R}^+$, let γ_1, γ_2 denote the paths from $[r,R]$ to $\mathbb{C}$ with

$$\gamma_1(s) := s, \qquad \gamma_2(s) := ws, \quad s \in [r,R]. \tag{2.26}$$

The paths γ_3 and γ_4 are defined to be the arcs with radii r and R from the x-axis to rw and Rw, respectively. Formally, the two paths are given by

$$\gamma_3(s) := R\,w^s, \qquad \gamma_4(s) := rw^s, \quad s \in [0,1].$$

Now by Cauchy's integral theorem[9]

$$\left(\int_{\gamma_1} - \int_{\gamma_2} + \int_{\gamma_3} - \int_{\gamma_4}\right) f(z)\,\mathrm{d}z = 0. \tag{2.27}$$

For the integral over γ_4, note that the length of this curve is bounded by the length of a quarter of a circle with radius r. Hence we can write

$$\left|\int_{\gamma_4} f(z)\,\mathrm{d}z\right| = \left|\int_0^1 f(\gamma_4(s))\gamma_4'(s)\,\mathrm{d}s\right| \leq \sup_{\hat{s}\in[0,1]}|f(\gamma_4(\hat{s}))|\underbrace{\int_0^1|\gamma_4'(s)|\,\mathrm{d}s}_{\leq \pi r/2}. \tag{2.28}$$

As $|w| = 1$, the polar form of w is given by $w = \mathrm{e}^{\mathrm{i}\theta}$ for a $\theta \in (-\pi,\pi]$ and therefore the complex logarithm of w defined in Remark 2.31 is given by $\mathrm{Log}(w) = \mathrm{i}\theta$. Due to

$$\mathrm{Log}(\gamma_4(s)) = \mathrm{Log}(rw^s) = \mathrm{Log}(r\,\mathrm{e}^{s\mathrm{Log}(w)}) = \mathrm{Log}(r\,\mathrm{e}^{s\mathrm{i}\theta}) = \log(r) + \mathrm{i}s\theta,$$

[9] The theorem along with its proof can be found in [18] as Theorem 6.1.

we can write for $s \in [0,1]$

$$\operatorname{Re}\big((\alpha-1)\operatorname{Log}(\gamma_4(s))\big) = \operatorname{Re}\big((\alpha-1)(\log(r)+\mathrm{i}s\theta)\big) = (\operatorname{Re}(\alpha)-1)\log(r) - \operatorname{Im}(\alpha)s\theta,$$

where $s\theta \in (-\pi,\pi]$, and use this, with Eq. (2.24), for

$$\sup_{s\in[0,1]} |f(\gamma_4(s))| = \sup_{s\in[0,1]} |(\gamma_4(s))^{\alpha-1}|\,|\mathrm{e}^{-\gamma_4(s)}| = \sup_{s\in[0,1]} |\mathrm{e}^{(\alpha-1)\operatorname{Log}(\gamma_4(s))}|\,\underbrace{\mathrm{e}^{-\operatorname{Re}(\gamma_4(s))}}_{\le \mathrm{e}^{-r}\le 1}$$
$$\le \mathrm{e}^{(\operatorname{Re}(\alpha)-1)\log(r)+\pi|\operatorname{Im}(\alpha)|} = C\,r^{\operatorname{Re}(\alpha)-1},$$

where C refers to a real number that is independent of r. Hence the integral over γ_4 satisfies

$$\left|\int_{\gamma_4} f(z)\,\mathrm{d}z\right| \le \text{Eq.}(2.28) \le C\,r^{\operatorname{Re}(\alpha)-1}\frac{\pi r}{2} = \frac{\pi C}{2}\,r^{\operatorname{Re}(\alpha)}, \qquad r\in\mathbb{R}^+$$

and as $\operatorname{Re}(\alpha) > 0$ the integral vanishes for $r \searrow 0$. For $R\to\infty$ the integral over the arc γ_3 vanishes due to Eq. (2.25). Therefore, Eq. (2.27) leads to

$$\underbrace{\lim_{r\searrow 0}\lim_{R\to\infty}\int_{\gamma_1} f(z)\,\mathrm{d}z}_{=\int_{\mathbb{R}^+} z^{\alpha-1}\,\mathrm{e}^{-z}\,\mathrm{d}z=\Gamma(\alpha)} = \lim_{r\searrow 0}\lim_{R\to\infty}\int_{\gamma_2} f(z)\,\mathrm{d}z,$$

where the limits on the left side exist due to Remark 2.24 and hence the limits on the right side exist. Using the definition of the line integral and of γ_2, this is equivalent to

$$\Gamma(\alpha) = \int_{\mathbb{R}^+} f(wz)w\,\mathrm{d}z.$$

Now substituting $y = z/|1+\mathrm{i}t|$ and inserting the definition of w, which then fulfills $wz = y(1+\mathrm{i}t)$, and the definition of f from Eq. (2.24) leads to

$$\begin{aligned}\Gamma(\alpha) &= \int_{\mathbb{R}^+}\big(y(1+\mathrm{i}t)\big)^{\alpha-1}\,\mathrm{e}^{-y(1+\mathrm{i}t)}\,(1+\mathrm{i}t)\,\mathrm{d}y\\ &= (1+\mathrm{i}t)^{\alpha}\int_{\mathbb{R}^+} y^{\alpha-1}\,\mathrm{e}^{-y(1+\mathrm{i}t)}\,\mathrm{d}y,\end{aligned}$$

which, combined with Eq. (2.23), proves the Lemma 2.33. □

2.5. The Multi-Dimensional Gamma Function

In Definition 2.36 below the multi-dimensional gamma function will be introduced. The centerpiece of this section is Theorem 2.37, which on the one hand proves that the multi-dimensional gamma function is well-defined and on the other hand gives us a very useful Laplace transform for handling matrix-valued gamma distributions in Chapter 3 below.

Definition 2.36 (Multi-dimensional gamma function, see [13, Definition 1.4.2])**.** For $p\in\mathbb{N}$ the *p-dimensional gamma function* is defined as

$$\Gamma_p(\alpha) = \int_{\mathbb{S}_p^+} \mathrm{e}^{-\operatorname{tr}(A)}(\det A)^{\alpha-\frac{p+1}{2}}\,\mathrm{d}A \tag{2.29}$$

for all $\alpha\in\mathbb{C}$ satisfying $\operatorname{Re}(\alpha) > (p-1)/2$.

Theorem 2.37 (Computation of the multi-dimensional gamma function and a Laplace transform of a function with matrix argument, see [13, Theorem 1.4.1, Eq. (1.4.6)])**.** *For $p \in \mathbb{N}$ and $\alpha \in \mathbb{C}$ satisfying $\mathrm{Re}(\alpha) > (p-1)/2$, the value of the p-dimensional gamma function at α is given by*

$$\Gamma_p(\alpha) = (2\pi)^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\alpha - \frac{j-1}{2}\right). \tag{2.30}$$

Furthermore, for a matrix $Z = X + \mathrm{i}Y \in \mathbb{C}^{p\times p}$, where the real part $X \in \mathbb{S}_p^+$ and the imaginary part $Y \in \mathbb{R}^{p\times p}$ is symmetric, the determinant[10] *of Z satisfies $\det Z \neq 0$ and can be used for the equality*

$$\int_{\mathbb{S}_p^+} \mathrm{e}^{-\mathrm{tr}(AZ)} (\det A)^{\alpha - \frac{p+1}{2}} \,\mathrm{d}A = (\det Z)^{-\alpha}\, \Gamma_p(\alpha). \tag{2.31}$$

Note that for $p = 1$, Eq. (2.30) reduces to $\Gamma_1(\alpha) = \Gamma(\alpha)$ and Eq. (2.31) with $Z = (1 + \mathrm{i}t)$ is the last equality in Eq. (2.19).

The proof of this theorem will be given below, subsequent an additional lemma that will be used in its proof.

Remark 2.38 (Alternative formulation and interpretation of Eq. (2.30))*.* In the literature, see [13, Theorem 1.4.1] or [21, Chapter 1.0, Example 1.24] for example, the integral over the set of positive definite matrices is commonly only seen as integral over the positive definite matrices seen as a subset of $\mathbb{R}^{p(p+1)/2}$. This means instead of integrating over the submanifold $\mathbb{S}_p^+$, the integral is defined over $\tilde{\mathbb{S}}_p^+$, i.e. the entries on and below the diagonal of the positive definite matrices. Under this interpretation

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\alpha - \frac{j-1}{2}\right).$$

This equation follows from Eq. (2.12) combined with Definition 2.36 and Theorem 2.37.

The equation (2.31) is independent of the interpretation of integrals over positive definite matrices as either both sides of the equation include the factor $2^{p(p-1)/2} = 4^{p(p-1)/4}$ or both do not.

The proof of Theorem 2.37 also uses the following fundamental analysis theorem which be found in [3, Theorem 4.3.1] along with its proof.

Theorem 2.39 (Transformation theorem)**.** *Let D be an open subset of $\mathbb{R}^n$, ϕ an injective C^1-map from D to $\phi(D) \subseteq \mathbb{R}^n$ and f a measurable function from $\phi(D)$ to $\mathbb{Q}$. Let $\mathrm{d}\phi(x)$ denote the Jacobi matrix of ϕ at $x \in D$. Then*

$$\int_{\phi(D)} f(y)\, \lambda^n(\mathrm{d}y) = \int_D f\big(\phi(x)\big)\, |\det \mathrm{d}\phi(x)|\, \lambda^n(\mathrm{d}x),$$

where the existence of one integral follows from the existence of the other.

Note that as integration of vector-valued functions works component-wise, the theorem also holds for f taking values in $\mathbb{C}^p$ for $p \in \mathbb{N}$.

[10] For $Z = X + \mathrm{i}Y$ with $X \in \mathbb{S}_p^+$ and Y *not* symmetric, the determinant of Z can be equal to zero, as the example $X = I_2$ and $Y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ shows.

The proof of the next lemma will use the higher dimensional chain rule.

Remark 2.40 (Chain rule, see [19, Proposition 10.1.18]). For $l, m, n \in \mathbb{N}$ let $f\colon D_f \to \mathbb{R}^l$ with $D_f \subseteq \mathbb{R}^m$ open, and $g\colon D_g \to \mathbb{R}^m$ with $D_g \subseteq \mathbb{R}^n$ open, be two continuously differentiable functions. If $g(D_g) \subseteq D_f$, then the composition $f \circ g$ is continuously differentiable as well and

$$\mathrm{d}(f \circ g)(x) = \mathrm{d}f\big(g(x)\big)\mathrm{d}g(x), \quad x \in D_g.$$

If additionally $n = m = l$, then we can use the multiplicativity of the determinant to write

$$\det \mathrm{d}(f \circ g)(x) = \det \mathrm{d}f\big(g(x)\big) \cdot \det \mathrm{d}g(x), \quad x \in D_g. \tag{2.32}$$

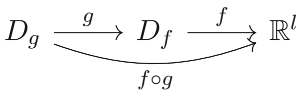


Figure 2.2. *Commutative diagram of the setting of the chain rule.*

As the calculations of $\det \mathrm{d}\phi$ needed for applications of Theorem 2.39 can be quite long, the two relevant determinants for the proof of Theorem 2.37 are given by the following lemma.

Lemma 2.41 (Determinants of relevant Jacobi matrices, see [21, Theorem 1.29, Theorem 1.20]). *For $p \in \mathbb{N}$ let $\tilde{\mathbb{S}}_p^+$ be the set of positive definite matrices seen as subset of $\mathbb{R}^{p(p+1)/2}$ and ψ be the map that extends $\tilde{\mathbb{S}}_p^+$ to $\mathbb{S}_p^+$ as defined in Eq. (2.7).*

(i) Let $\tilde{\mathbb{T}}_p$ be defined as in Eq. (2.8). Then

$$\phi := \begin{cases} \tilde{\mathbb{T}}_p & \to & \tilde{\mathbb{S}}_p^+ \\ \tilde{T} & \mapsto & A^\triangle := (TT^\intercal)^\triangle \end{cases}, \tag{2.33}$$

where $T \in \mathbb{T}_p$ is understood as the matrix version of $\tilde{T} \in \tilde{\mathbb{T}}_p \subseteq \mathbb{R}^{p(p+1)/2}$ filled up with zeros, is well-defined. The determinant of the Jacobi matrix of this map is given by

$$\det \mathrm{d}\phi(\tilde{T}) = 2^p \prod_{m=1}^{p} t_{mm}^{p-m+1}, \quad \tilde{T} \in \tilde{\mathbb{T}}_p, \tag{2.34}$$

where t_{ij} for $i \geq j$ in $\{1, \ldots, p\}$ refer to the entries of the matrix version T of $\tilde{T}$.

(ii) Let $B, C \in \mathbb{R}^{p \times p}$ be matrices let $\phi_B, \phi_C\colon \tilde{\mathbb{S}}_p^+ \to \tilde{\mathbb{S}}_p^+$ be two continuously differentiable maps satisfying $\det \mathrm{d}\phi_B \equiv (\det B)^{p+1}$ and $\det \mathrm{d}\phi_C \equiv (\det C)^{p+1}$. Then

$$\det \mathrm{d}(\phi_B \circ \phi_C) \equiv \big(\det(BC)\big)^{p+1}.$$

(iii) Let $B \in \mathbb{R}^{p \times p}$ be a matrix with full rank. Then the map

$$\phi_B := \begin{cases} \tilde{\mathbb{S}}_p^+ & \to & \tilde{\mathbb{S}}_p^+ \\ A^\triangle & \mapsto & (BAB^\intercal)^\triangle \end{cases}, \tag{2.35}$$

where $A := \psi(A^\triangle)$ is well-defined, fulfils

$$\det \mathrm{d}\phi(A^\triangle) = (\det B)^{p+1}, \quad A^\triangle \in \tilde{\mathbb{S}}_p^+. \tag{2.36}$$

(iv) Let $n \in \mathbb{N}$ and let $B \in \mathbb{R}^{p\times p}$. For the map

$$\phi := \begin{cases} \mathbb{R}^{p\times n} & \to & \mathbb{R}^{p\times n} \\ X & \mapsto & BX \end{cases}, \tag{2.37}$$

the determinant of the Jacobi matrix is given by

$$\det \mathrm{d}\phi(X) = (\det B)^n, \quad X \in \mathbb{R}^{p\times n}. \tag{2.38}$$

Proof of Item (i). In the following proof we access the entries of elements of $\tilde{\mathbb{T}}_p$ and $\tilde{\mathbb{S}}_p^+$ by understanding them as the corresponding matrices in $\mathbb{T}_p$ and $\mathbb{S}_p^+$, respectively, using two indices.

According to Theorem 2.12, $A = TT^\mathsf{T} \in \mathbb{S}_p^+$, where T is the matrix version of an element of $\tilde{\mathbb{T}}_p$. Hence, the function ϕ given in Eq. (2.33) is well-defined.

For the rest of the proof let $\tilde{T}$ denote an element of $\tilde{\mathbb{T}}_p$. In order to determine $\det \mathrm{d}\phi(\tilde{T})$, we start by looking at the partial derivatives of the entries of $\phi(\tilde{T}) \in \tilde{\mathbb{S}}_p^+$ w.r.t. the entries of $\tilde{T}$. For all $i \geq j$ in $\{1, \dots, p\}$,

$$\phi_{ij}(\tilde{T}) = \sum_{m=1}^{p} t_{im} t_{jm} 1_{\{i\geq m, j\geq m\}} = \sum_{m=1}^{j} t_{im} t_{jm},$$

where the second equality follows from $i \geq j$. Hence, for all $i, j, k, l \in \{1, \dots, p\}$ with $i \geq j$ and $k \geq l$,

$$\frac{\partial \phi_{ij}(\tilde{T})}{\partial t_{kl}} = t_{jl} 1_{\{k=i, j\geq l\}} + t_{il} 1_{\{k=j, j\geq l\}} = t_{jl} 1_{\{k=i, j\geq l\}} + t_{il} 1_{\{k=j\}},$$

where 1_I is equal to one if the conditions in I are satisfied and zero otherwise, see *Conventions*. In order to determine the matrix $\mathrm{d}\phi(\tilde{T}) \in \mathbb{R}^{p(p+1)/2\times p(p+1)/2}$, we use the $\mathrm{vecp}(\cdot)$ function given in Eq. (2.6): The matrix $\mathrm{d}\phi$ can be seen as block matrix of the Jacobi matrices of the restricted columns of the matrix version of $\phi(\tilde{T})$, defined in Remark 2.18, with respect to the restricted columns of the matrix version of $\tilde{T}$, that are denoted by $\phi(\tilde{T})^m$ and $\tilde{T}^m$ for $m \in \{1, \dots, p\}$, respectively,

$$\mathrm{d}\phi(\tilde{T}) = \begin{pmatrix} \frac{\partial \phi(\tilde{T})^1}{\partial \tilde{T}^1} & \cdots & \frac{\partial \phi(\tilde{T})^1}{\partial \tilde{T}^p} \\ \vdots & & \vdots \\ \frac{\partial \phi(\tilde{T})^p}{\partial \tilde{T}^1} & \cdots & \frac{\partial \phi(\tilde{T})^p}{\partial \tilde{T}^p} \end{pmatrix}, \quad \frac{\partial \phi(\tilde{T})^m}{\partial \tilde{T}^n} = \begin{pmatrix} \frac{\partial \phi_{mm}(T)}{\partial t_{nn}} & \cdots & \frac{\partial \phi_{mm}(T)}{\partial t_{pn}} \\ \vdots & & \vdots \\ \frac{\partial \phi_{pm}(T)}{\partial t_{nn}} & \cdots & \frac{\partial \phi_{pm}(T)}{\partial t_{pn}} \end{pmatrix}$$

for every $m, n \in \{1, \dots, p\}$ where $\partial\phi(\tilde{T})^m/\partial\tilde{T}^n \in \mathbb{R}^{(p-m+1)\times(p-n+1)}$. Now for all indices $m, n \in \{1, \dots, p\}$ satisfying $m < n$ the matrix $\partial\phi(\tilde{T})^m/\partial\tilde{T}^n$ is equal to the zero matrix, as for all $a \in \{m, \dots, p\}$ and all $b \in \{n, \dots, p\}$,

$$\frac{\partial \phi_{am}(T)}{\partial t_{bn}} = t_{mn} 1_{\{a=b, m\geq n\}} + t_{an} 1_{\{b=m\}} = 0,$$

because $b \geq n > m$. Hence, the determinant of $\mathrm{d}\phi(\tilde{T})$ is equal to the product of the determinants of the diagonal blocks. For $m \in \{1, \dots, p\}$, the entries of the matrix $\partial\phi(\tilde{T})^m/\partial\tilde{T}^m$ are given by

$$\frac{\partial\phi_{am}(\tilde{T})}{\partial t_{bm}} = t_{mm}1_{\{a=b,m\geq m\}} + t_{am}1_{\{b=m\}} = \begin{cases} 2t_{mm} & \text{if } a = b = m. \\ t_{mm} & \text{if } a = b \neq m, \\ t_{am} & \text{if } a \neq b = m, \\ 0 & \text{otherwise,} \end{cases}$$

for all $a, b \in \{m, \dots, p\}$. Hence, $\partial\phi(\tilde{T})^m/\partial\tilde{T}^m$ is a lower triangular matrix and therefore the determinant is given by the product of the diagonal entries,

$$\det\left(\frac{\partial\phi(\tilde{T})^m}{\partial\tilde{T}^m}\right) = 2t_{mm}^{p-m+1}.$$

The determinant of $\mathrm{d}\phi(\tilde{T})$ is then given as the product of the determinants of the block diagonal matrices

$$\det \mathrm{d}\phi(\tilde{T}) = \prod_{m=1}^{p} \det\left(\frac{\partial\phi(\tilde{T})^m}{\partial\tilde{T}^m}\right) = \prod_{m=1}^{p} 2t_{mm}^{p-m+1} = 2^p \prod_{m=1}^{p} t_{mm}^{p-m+1}. \qquad \square$$

Proof of Item (ii). The implication of the chain rule stated in Eq. (2.32) tells us, for every $x^\triangle \in \tilde{\mathbb{S}}_p^+$,

$$\begin{aligned}\det \mathrm{d}(\phi_B \circ \phi_C)(x^\triangle) &= \det \mathrm{d}\phi_B\big(\phi_C(x^\triangle)\big) \det \mathrm{d}\phi_C(x^\triangle) \\ &\equiv (\det B)^{p+1}(\det C)^{p+1} = \big(\det(BC)\big)^{p+1},\end{aligned}$$

where the last equality follows from the multiplicativity of the determinant. $\square$

Proof of Item (iii). Now, to see that the map given in Eq. (2.35) is well-defined, we show that BAB^T is positive definite again. For all $x \in \mathbb{R}^p \setminus \{0\}$, the product $x^\mathsf{T}BAB^\mathsf{T}x = (x^\mathsf{T}B)A(x^\mathsf{T}B)^\mathsf{T} > 0$ since $x^\mathsf{T}B \in \mathbb{R}^p \setminus \{0\}$ for $x \in \mathbb{R}^p \setminus \{0\}$ as B has full rank.

In order to determine $\det \mathrm{d}\phi$, we are using the fact that B is invertible and can therefore be written as the product of $n \in \mathbb{N}$ elementary matrices[11] E_i for $i \in \{1, \dots, n\}$, hence

$$BAB^\mathsf{T} = E_n \cdots E_2E_1AE_n \cdots E_2E_1 = E_n \cdots E_2E_1AE_1^\mathsf{T}E_2^\mathsf{T} \cdots E_n^\mathsf{T}. \tag{2.39}$$

For every $i \in \{1, \dots, n\}$, let

$$\phi_i := \left\{ \begin{array}{ccc} \tilde{\mathbb{S}}_p^+ & \to & \tilde{\mathbb{S}}_p^+ \\ X^\triangle & \mapsto & (E_iXE_i^\mathsf{T})^\triangle \end{array} \right. , \tag{2.40}$$

[11] The fact that an invertible matrix B can be written as matrix product of elementary matrices in the form discussed in *Case 1* and *Case 2* can be, for example, be found in [7, Section 5.6]. Often this decomposition is given with a third type of elementary matrices – permutation matrices. As every permutation matrix can be written as a product of elementary matrices of the other two types, considering *Case 1* and *Case 2* suffices.

where $X = \psi(X^\triangle)$. This is well-defined for every $i \in \{1, \dots, n\}$ as the function $X \mapsto E_i X E_i^\mathsf{T}$ maps every positive definite matrix argument to a positive definite matrix. Thus, we can write our function ϕ as a composition of the functions

$$\phi = \phi_n \circ \cdots \circ \phi_2 \circ \phi_1. \tag{2.41}$$

Hence, using Item (ii), it is sufficient to prove Eq. (2.36) for B equal to one of the two types of elementary matrices.

In the rest of the proof we again access the entries of elements of $\tilde{\mathbb{S}}_p^+$ by understanding them as the corresponding matrices $\mathbb{S}_p^+$, respectively, using two indices; furthermore, let $X^\triangle$ denote an element of $\tilde{\mathbb{S}}_p^+$ and $X = \psi(X^\triangle)$ the corresponding matrix from $\mathbb{S}_p^+$.

Case 1: The matrix B is a identity matrix except for one diagonal entry at position (α, α), $\alpha \in \{1, \dots, p\}$, that is equal to a real $\lambda \neq 0$. In this case, the entries of $(BXB^\mathsf{T})^\triangle$ are given by

$$\phi_B(X^\triangle)_{kl} = \begin{cases} \lambda x_{kl} & \text{if } k = \alpha \text{ or } l = \alpha,\ l \neq k, \\ \lambda^2 x_{kl} & \text{if } k = l = \alpha, \\ x_{kl} & \text{otherwise,} \end{cases} \tag{2.42}$$

for all $k \geq l$ in $\{1, \dots, p\}$. As in the proof of Item (i), the matrix $\mathrm{d}\phi_B(X^\triangle)$ can be seen as a block matrix, where the blocks correspond to the Jacobi matrices of the restricted columns, see Eq. (2.6), $\phi_B(\tilde{X}^\triangle)^m, m \in \{1, \dots, p\}$, with respect to the restricted columns $\tilde{X}^n, n \in \{1, \dots, p\}$, so

$$\mathrm{d}\phi_B(X^\triangle) = \begin{pmatrix} \frac{\partial \phi_B(\tilde{X}^\triangle)^1}{\partial \tilde{X}^1} & \cdots & \frac{\partial \phi_B(\tilde{X}^\triangle)^1}{\partial \tilde{X}^p} \\ \vdots & & \vdots \\ \frac{\partial \phi_B(\tilde{X}^\triangle)^p}{\partial \tilde{X}^1} & \cdots & \frac{\partial \phi_B(\tilde{X}^\triangle)^p}{\partial \tilde{X}^p} \end{pmatrix}, \tag{2.43}$$

where the matrix $\partial \phi_B(\tilde{X}^\triangle)^m / \partial \tilde{X}^n \in \mathbb{R}^{(p-m+1)\times(p-n+1)}$ for all $m, n \in \{1, \dots, p\}$. Now for all $m \neq n$ in $\{1, \dots, p\}$, the matrix $\partial \phi_B(\tilde{X}^\triangle)^m / \partial \tilde{X}^n$ is equal to zero as

$$\frac{\partial \phi_B(X^\triangle)_{am}}{\partial x_{bn}} = 0, \quad a \in \{m, \dots, p\}, \quad b \in \{n, \dots, p\},$$

due to Eq. (2.42) for $m \neq n$. Hence, the determinant of $\mathrm{d}\phi_B(X^\triangle)$ is equal to the product of the determinants of the diagonal blocks. For every $m \in \{1, \dots, p\}$, the entries of matrix $\partial \phi_B(\tilde{X}^\triangle)^m / \partial \tilde{X}^m$ are given by

$$\frac{\partial \phi_B(X^\triangle)_{am}}{\partial x_{bm}} = \begin{cases} 0 & \text{if } a \neq b, \\ \lambda & \text{if } a = b,\ a = \alpha \text{ or } m = \alpha,\ a \neq m, \\ \lambda^2 & \text{if } a = b = m = \alpha, \\ 1 & \text{otherwise,} \end{cases}$$

for $a, b \in \{m, \dots, p\}$. Hence $\partial \phi_B(\tilde{X}^\triangle)^m / \partial \tilde{X}^m$ is a $(p - m + 1) \times (p - m + 1)$-dimensional diagonal matrix and

$$\det \frac{\partial \phi_B(\tilde{X}^\triangle)^m}{\partial \tilde{X}^m} = \begin{cases} \lambda^2 \lambda^{p-m} = \lambda^2 \lambda^{p-\alpha} & \text{if } m = \alpha, \\ \lambda & \text{if } m < \alpha, \\ 1 & \text{otherwise,} \end{cases}$$

which leads to

$$\det \mathrm{d}\phi_B(X^\triangle) = \prod_{m=1}^{p} \det \frac{\partial \phi_B(\tilde{X}^\triangle)^m}{\partial \tilde{X}^m} = \lambda^2 \lambda^{p-\alpha}\lambda^{\alpha-1} = \lambda^{p+1} = (\det B)^{p+1},$$

where the last equation follows from the definition of B in this case.

Case 2: The matrix B is a identity matrix except for one entry at position (α, β), where $\alpha \neq \beta$ in $\{1, \ldots, p\}$, that is equal to $\lambda \in \mathbb{R} \setminus \{0\}$. Then the entries of $(BXB^\mathsf{T})^\triangle$ for all $k \geq l$ in $\{1, \ldots, p\}$ are given by

$$\phi_B(X^\triangle)_{kl} = \begin{cases} x_{kl} & \text{if } k \neq \alpha, l \neq \alpha, \\ x_{kl} + \lambda x_{\beta l} & \text{if } k = \alpha, l \neq \alpha, \\ x_{kl} + \lambda x_{k\beta} & \text{if } k \neq \alpha, l = \alpha, \\ x_{kl} + \lambda x_{\beta l} + \lambda x_{k\beta} + \lambda^2 x_{\beta\beta} & \text{if } k = l = \alpha. \end{cases} \tag{2.44}$$

To calculate $\det \mathrm{d}\phi_B(X^\triangle)$ we are again breaking down the matrix $\mathrm{d}\phi_B(X^\triangle)$ into the block matrices $\partial\phi_B(\tilde{X}^\triangle)^m/\partial\tilde{X}^n \in \mathbb{R}^{(p-m+1)\times(p-n+1)}$, for $m, n \in \{1, \ldots, p\}$, as given in Eq. (2.43).

Now we distinguish the cases $\beta > \alpha$ and $\alpha > \beta$ for the position of the λ in the elementary matrix B. First, for $\beta > \alpha$ the entries of the matrix $\partial\phi_B(\tilde{X}^\triangle)^m/\partial\tilde{X}^n$ for $m > n \in \{1, \ldots, p\}$, $a \in \{m, \ldots, p\}$ and $b \in \{n, \ldots, p\}$ are given by

$$\frac{\partial \phi_B(X^\triangle)_{am}}{\partial x_{bn}} = \lambda 1_{\{a=b,\, m=\alpha,\, n=\beta,\, a\neq\alpha\}} + \lambda^2 1_{\{a=m=\alpha,\, b=n=\beta\}} = 0, \tag{2.45}$$

where the first equation follows from Eq. (2.44), and the second one from $\beta > \alpha$ and $m > n$. Hence $\mathrm{d}\phi_B$ is a block upper diagonal matrix and the determinant is given by the product of the block diagonal matrices $\partial\phi_B(\tilde{X}^\triangle)^m/\partial\tilde{X}^m$, $m \in \{1, \ldots, p\}$.

Second, for $\alpha > \beta$ the entries of the matrix $\partial\phi_B(\tilde{X}^\triangle)^m/\partial\tilde{X}^n$ for $n > m \in \{1, \ldots, p\}$, $a \in \{m, \ldots, p\}$ and $b \in \{n, \ldots, p\}$ are again given by Eq. (2.45), where the second equation now follows from $\alpha > \beta$ and $n > m$. In this case, $\mathrm{d}\phi_B$ is a block lower diagonal matrix and the determinant is again given by the product of the block diagonal matrices.

So for all $\alpha \neq \beta$, where $\alpha, \beta \in \{1, \ldots, p\}$, the determinant is given by the product of the block diagonal matrices $\partial\phi_B(\tilde{X}^\triangle)^m/\partial\tilde{X}^m$ for $m \in \{1, \ldots, p\}$. For $a, b \in \{m, \ldots, p\}$ the entries of this matrix are given by

$$\frac{\partial \phi_B(X^\triangle)_{am}}{\partial x_{bm}} = \begin{cases} 1 & \text{if } a = b, \\ \lambda & \text{if } a = \alpha, b = \beta, \\ 0 & \text{otherwise.} \end{cases}$$

Since this is a lower/ upper diagonal matrix with only ones on the diagonal,

$$\det \frac{\partial \phi_B(\tilde{X}^\triangle)^m}{\partial \tilde{X}^m} = 1$$

for every $m \in \{1, \ldots, p\}$, which leads to

$$\det \mathrm{d}\phi_B(X^\triangle) = \prod_{m=1}^{p} \det \frac{\partial \phi_B(\tilde{X}^\triangle)^m}{\partial \tilde{X}^m} = 1 = 1^{p+1} = (\det B)^{p+1},$$

where the last equation follows from the fact that B in this case is an identity matrix except for one entry at (α, β), where $\alpha \neq \beta$ in $\{1, \dots, p\}$. □

Proof of Item (iv). For the rest of the proof let X denote an element of $\mathbb{R}^{p \times n}$. In order to determine the matrix $\mathrm{d}\phi(X) \in \mathbb{R}^{pn \times pn}$, both $\phi(X)$ and X need to be vectorized using the vec(·) function given in Eq. (2.4), as $\phi(X)$ and X are elements of $\mathbb{R}^{p \times n}$. Therefore, the matrix $\mathrm{d}\phi$ can be seen as block matrix of the Jacobi matrices of the columns of $\phi(X)$ with respect to the columns of X,

$$\mathrm{d}\phi(X) = \begin{pmatrix} \frac{\partial \phi(X)^1}{\partial X^1} & \dots & \frac{\partial \phi(X)^1}{\partial X^n} \\ \vdots & & \vdots \\ \frac{\partial \phi(X)^n}{\partial X^1} & \dots & \frac{\partial \phi(X)^n}{\partial X^n} \end{pmatrix},$$

where $\partial \phi(X)^i / \partial X^j \in \mathbb{R}^{p \times p}$ for every $i, j \in \{1, \dots, n\}$. Due to the definition of ϕ in Eq. (2.37),

$$\phi(X)^i = BX^i, \quad i \in \{1, \dots, n\}.$$

Therefore, for all $i, j \in \{1, \dots, n\}$

$$\frac{\partial \phi(X)^i}{\partial X^j} = \begin{cases} B & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Hence, $\mathrm{d}\phi(X)$ is a block diagonal matrix with

$$\det \mathrm{d}\phi(X) = \prod_{i=1}^{n} \det \frac{\partial \phi(X)^i}{\partial X^i} = \prod_{i=1}^{n} \det B = (\det B)^n,$$

thus proving Eq. (2.38). □

Proof of Theorem 2.37. The equality

$$\int_{\mathbb{S}_p^+} \mathrm{e}^{-\mathrm{tr}(AZ)} (\det A)^{\alpha - \frac{p+1}{2}} \,\mathrm{d}A = (\det Z)^{-\alpha} \,(2\pi)^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\alpha - \frac{i-1}{2}\right) \tag{2.46}$$

holds true if and only if both equalities of the theorem, Eq. (2.30) and Eq. (2.31) hold true. Hence, we will prove Eq. (2.46) in the following five steps.

Step 1. As described in Remark 2.23, we integrate over the $\frac{p(p+1)}{2}$-dimensional submanifold $\mathbb{S}_p^+$ of $\mathbb{R}^{p \cdot p}$ using the chart ψ defined in Eq. (2.7). By inserting

$$f := \begin{cases} \mathbb{S}_p^+ & \to \quad \mathbb{C} \\ A & \mapsto \quad \mathrm{e}^{-\mathrm{tr}(AZ)} (\det A)^{\alpha - \frac{p+1}{2}} \end{cases}$$

into Eq. (2.12) and using the decomposition of Z into its real and imaginary part $Z = X + \mathrm{i}Y$,

$$\int_{\mathbb{S}_p^+} \mathrm{e}^{-\mathrm{tr}(AZ)} (\det A)^{\alpha - \frac{p+1}{2}} \,\mathrm{d}A$$
$$= 2^{\frac{p(p-1)}{4}} \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\mathrm{tr}(\psi(A^\triangle)(X + \mathrm{i}Y))} \big(\det \psi(A^\triangle)\big)^{\alpha - \frac{p+1}{2}} \lambda^{\frac{p(p+1)}{2}} (\mathrm{d}A^\triangle). \tag{2.47}$$

From now on we will write $\mathrm{d}A^{\triangle}$ for $\lambda^{\frac{p(p+1)}{2}}(\mathrm{d}A^{\triangle})$.

Step 2. In this step we transform the integral to simplify the real part X of the matrix Z. Since the matrix X is positive definite, Lemma 2.15 tells us that there exists a unique positive definite, hence symmetric, matrix B such that $BB = BB^{\mathsf{T}} = X$. Instead of integrating over $A^{\triangle}$, we now want to integrate over $\Lambda^{\triangle} = \psi^{-1}(BAB)$ using the transformation formula from Theorem 2.39. Therefore, we define the map

$$\phi = \left\{ \begin{array}{ccc} \tilde{\mathbb{S}}_p^+ & \to & \tilde{\mathbb{S}}_p^+ \\ \Lambda^{\triangle} & \mapsto & \psi^{-1}(B^{-1}\Lambda B^{-1}) \end{array} \right. . \tag{2.48}$$

Since B^{-1} is symmetric and invertible, the results from Item (iii) of Lemma 2.41 tell us that the map is well defined and

$$\det \mathrm{d}\phi = (\det B^{-1})^{p+1}.$$

This can be rearranged to

$$|\det \mathrm{d}\phi| = |\det B^{-1}|^{p+1} = ((\det B)^2)^{-\frac{p+1}{2}} = (\det X)^{-\frac{p+1}{2}}.$$

As $\tilde{\mathbb{S}}_p^+$ is open due to Lemma 2.20, inserting this into the transformation theorem applied to Eq. (2.47) yields

$$\begin{aligned}(2.47) &= 2^{\frac{p(p-1)}{4}}(\det X)^{-\frac{p+1}{2}} \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(\psi(\phi(\Lambda^{\triangle}))(X+\mathrm{i}Y))} (\det \psi(\phi(\Lambda^{\triangle})))^{\alpha-\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle} \\ &= 2^{\frac{p(p-1)}{4}}(\det X)^{-\frac{p+1}{2}} \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(B^{-1}\Lambda B^{-1}(X+\mathrm{i}Y))} (\det(B^{-1}\Lambda B^{-1}))^{\alpha-\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle}, \end{aligned} \tag{2.49}$$

where the definition of ψ given in Eq. (2.48) is used for the second equality. Since $BB = X$, the equation $\det(B^{-1}\Lambda B^{-1}) = \det \Lambda (\det BB)^{-1} = \det \Lambda (\det X)^{-1}$ holds for every $\Lambda \in \mathbb{S}_p^+$. Hence, this integral can be rewritten as

$$\begin{aligned}(2.49) &= 2^{\frac{p(p-1)}{4}}(\det X)^{-\frac{p+1}{2}} \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(B^{-1}\Lambda B^{-1}(X+\mathrm{i}Y))} (\det \Lambda)^{\alpha-\frac{p+1}{2}} (\det X)^{-\alpha+\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle} \\ &= (\det X)^{-\alpha}\, 2^{\frac{p(p-1)}{4}} \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(B^{-1}\Lambda B^{-1}(X+\mathrm{i}Y))} (\det \Lambda)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle}. \end{aligned} \tag{2.50}$$

Using the linearity of the trace, Lemma 2.1 and $X = BB$,

$$\begin{aligned}\operatorname{tr}(B^{-1}\Lambda B^{-1}(X+\mathrm{i}Y)) &= \operatorname{tr}(B^{-1}\Lambda B^{-1}X) + \mathrm{i}\cdot\operatorname{tr}(B^{-1}\Lambda B^{-1}Y) \\ &= \operatorname{tr}(\Lambda) + \mathrm{i}\cdot\operatorname{tr}(\Lambda B^{-1}YB^{-1}) = \operatorname{tr}(\Lambda(I_p + \mathrm{i}B^{-1}YB^{-1})).\end{aligned}$$

Combining this with the equations before leads to

$$\begin{aligned}&\int_{\mathbb{S}_p^+} \mathrm{e}^{-\operatorname{tr}(AZ}(\det A)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}A = (2.50) \\ &\qquad = (\det X)^{-\alpha}\, 2^{\frac{p(p-1)}{4}} \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(\Lambda(I_p+\mathrm{i}B^{-1}YB^{-1}))} (\det \Lambda)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle}. \end{aligned} \tag{2.51}$$

Step 3. The next step is to transform the integral in Eq. (2.51) to simplify the complex part Y of the matrix Z. For this and the next steps we introduce

$$\mathcal{I} := \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(\Lambda(I_p+\mathrm{i}B^{-1}YB^{-1}))} (\det \Lambda)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle} \tag{2.52}$$

and use this to rewrite Eq. (2.51) to

$$\int_{\mathbb{S}_p^+} \mathrm{e}^{-\operatorname{tr}(AZ)} (\det A)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}A = (\det X)^{-\alpha}\, 2^{\frac{p(p-1)}{4}} \cdot \mathcal{I}. \tag{2.53}$$

Since both the matrices B^{-1} and Y are symmetric, the product

$$B^{-1}YB^{-1} = (B^{-1})^{\mathsf{T}} Y^{\mathsf{T}} (B^{-1})^{\mathsf{T}} = (B^{-1}YB^{-1})^{\mathsf{T}}$$

is symmetric as well. Due to Theorem 2.5, there exists an orthogonal matrix $S \in \mathbb{R}^{p\times p}$, and a diagonal matrix $D \in \mathbb{R}^{p\times p}$, such that

$$B^{-1}YB^{-1} = SDS^{\mathsf{T}} = SDS^{-1}. \tag{2.54}$$

Since $\psi(\Lambda^{\triangle}) = \Lambda$ we can rewrite $\mathcal{I}$ from Eq. (2.52) to

$$\mathcal{I} = \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(\psi(\Lambda^{\triangle})(I_p+\mathrm{i}SDS^{\mathsf{T}}))} (\det \psi(\Lambda^{\triangle}))^{\alpha-\frac{p+1}{2}} \,\mathrm{d}\Lambda^{\triangle}. \tag{2.55}$$

Next, we want to transform $\mathcal{I}$ using the map

$$\eta := \begin{cases} \tilde{\mathbb{S}}_p^+ & \to \quad \tilde{\mathbb{S}}_p^+ \\ M^{\triangle} & \mapsto \quad \psi^{-1}(SMS^{\mathsf{T}}) \end{cases}, \tag{2.56}$$

where $M = \psi(M^{\triangle})$ for all $M^{\triangle} \in \tilde{\mathbb{S}}_p^+$. Since S is invertible, the determinant $|\det \mathrm{d}\eta|$ is given by Item (iii) of Lemma 2.41. Hence,

$$|\det \mathrm{d}\eta| = |\det S|^{p+1} = 1^{p+1} = 1,$$

since the absolute value of the determinant of every orthogonal matrix is equal to one. As $\tilde{\mathbb{S}}_p^+$ is an open set, we can apply Theorem 2.39 to Eq. (2.55) with η given in Eq. (2.56),

$$\mathcal{I} = \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(SMS^{\mathsf{T}}(I_p+\mathrm{i}SDS^{\mathsf{T}}))} \big(\det(SMS^{\mathsf{T}})\big)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}M^{\triangle}. \tag{2.57}$$

Using Lemma 2.1 and the fact that S is orthogonal,

$$\begin{aligned} \operatorname{tr}\big(SMS^{\mathsf{T}}\big(I_p+\mathrm{i}SDS^{\mathsf{T}}\big)\big) &= \operatorname{tr}\big(MS^{-1}\big(I_p+\mathrm{i}SDS^{-1}\big)S\big) \\ &= \operatorname{tr}\big(MS^{-1}S+\mathrm{i}MS^{-1}SDS^{-1}S\big) = \operatorname{tr}(M(I_p+\mathrm{i}D)). \end{aligned}$$

Hence, the integral in Eq. (2.57) can be rearranged to

$$\mathcal{I} = \int_{\tilde{\mathbb{S}}_p^+} \mathrm{e}^{-\operatorname{tr}(M(I_p+\mathrm{i}D))} (\det M)^{\alpha-\frac{p+1}{2}} \,\mathrm{d}M^{\triangle}, \tag{2.58}$$

where the fact that $\det(SMS^{\mathsf{T}}) = \det S \cdot \det M \cdot \det S^{\mathsf{T}} = \det M$, as S is orthogonal, is used.

Step 4. The next step is to transform $\mathcal{I}$ using the Cholesky decomposition. Let $\tilde{\mathbb{T}}_p$ be as in Eq. (2.8) and let the map ϕ be defined by $\phi(\tilde{T}) = (TT^\mathsf{T})^\triangle \in \tilde{\mathbb{S}}_p^+$ for all $\tilde{T} \in \tilde{\mathbb{T}}_p$, where T refers to the matrix version of $\tilde{T}$. Using Item (i) of Lemma 2.41, this map is well-defined and the determinant of its Jacobi matrix is given by

$$|\det \mathrm{d}\phi(\tilde{T})| = 2^p \prod_{k=1}^{p} t_{kk}^{p-k+1}, \quad \tilde{T} \in \tilde{\mathbb{T}}_p,$$

where t_{ij} for $i \geq j$ in $\{1, \dots, p\}$ refer to the entries of the matrix version T of $\tilde{T}$ and will be used in this format for the rest of the proof. As both $\tilde{\mathbb{S}}_p^+$ and $\tilde{\mathbb{T}}_p$ are open sets due to Lemma 2.20, we may apply the transformation formula from Theorem 2.39 with ϕ to Eq. (2.58). As $M = \psi(M^\triangle)$ for a unique $M^\triangle \in \tilde{\mathbb{S}}_p^+$,

$$\mathcal{I} = \int_{\tilde{\mathbb{T}}_p} \mathrm{e}^{-\mathrm{tr}(\psi(\phi(\tilde{T}))(I_p + \mathrm{i}D))} (\det \psi(\phi(\tilde{T})))^{\alpha - \frac{p+1}{2}}\, 2^p \prod_{k=1}^{p} t_{kk}^{p-k+1} \,\mathrm{d}\tilde{T}. \tag{2.59}$$

For the matrix version T of $\tilde{T} \in \tilde{\mathbb{T}}_p$, the identity $\det \psi(\phi(\tilde{T})) = \det(TT^\mathsf{T}) = (\det T)^2 = \prod_{k=1}^{p} t_{kk}^2$ holds and

$$\mathrm{tr}(\psi(\phi(\tilde{T}))(I_p + \mathrm{i}D)) = \mathrm{tr}(TT^\mathsf{T}(I_p + \mathrm{i}D)) = \sum_{\substack{k,l=1 \\ l \leq k}}^{p} t_{kl}^2 (1 + \mathrm{i}\lambda_k),$$

where λ_k refers to the k-th entry of the diagonal matrix D. Therefore, Eq. (2.59) is equal to

$$\mathcal{I} = \int_{\tilde{\mathbb{T}}_p} \exp\Bigg(- \sum_{\substack{k,l=1 \\ l \leq k}}^{p} t_{kl}^2 (1 + \mathrm{i}\lambda_k)\Bigg) \Bigg(\prod_{k=1}^{p} t_{kk}^2\Bigg)^{\alpha - \frac{p+1}{2}} 2^p \prod_{k=1}^{p} t_{kk}^{p-k+1} \,\mathrm{d}\tilde{T}. \tag{2.60}$$

As $\tilde{\mathbb{T}}_p$ denotes the vectorized set of lower triangular matrices with positive diagonal entries, we can use Fubini's theorem for the Lebesgue–Borel measure to rearrange integrals. Hence, Eq. (2.60) simplifies to

$$\mathcal{I} = 2^p \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}_{\frac{p(p-1)}{2}\,\text{times}} \Bigg(\underbrace{\int_{\mathbb{R}^+} \cdots \int_{\mathbb{R}^+}}_{p\,\text{times}} \exp\Bigg(- \sum_{\substack{k,l=1 \\ l \leq k}}^{p} t_{kl}^2 (1 + \mathrm{i}\lambda_k)\Bigg) \Bigg(\prod_{k=1}^{p} (t_{kk}^2)^{\alpha - \frac{k}{2}}\Bigg) \prod_{k=1}^{p} \mathrm{d}t_{kk} \Bigg) \prod_{\substack{k,l=1 \\ k > l}}^{p} \mathrm{d}t_{kl}.$$

Further rearranging leads to

$$\mathcal{I} = \Bigg(\prod_{\substack{k,l=1 \\ k > l}}^{p} \int_{\mathbb{R}} \exp\left(-t_{kl}^2 (1 + \mathrm{i}\lambda_k)\right) \mathrm{d}t_{kl}\Bigg) \prod_{k=1}^{p} 2 \int_{\mathbb{R}^+} \exp\left(-t_{kk}^2 (1 + \mathrm{i}\lambda_k)\right) (t_{kk}^2)^{\alpha - \frac{k}{2}} \,\mathrm{d}t_{kk}, \tag{2.61}$$

so the integral can be broken down into the product of multiple one-dimensional complex integrals which can be determined separately.

Step 5. As the last step we put the results of the prior steps together to arrive at the claimed equality.

By using Eq. (2.19) of Lemma 2.33 with $\alpha - \frac{k-1}{2}$ for every $k \in \{1, \dots, p\}$

$$\int_{\mathbb{R}^+} \exp\left(-t_{kk}^2(1+\mathrm{i}\lambda_k)\right)(t_{kk}^2)^{\alpha-\frac{k}{2}}\,\mathrm{d}t_{kk} = (1+\mathrm{i}\lambda_k)^{\frac{k-1}{2}-\alpha}\,\Gamma\left(a - \frac{k-1}{2}\right). \tag{2.62}$$

As we assumed $\mathrm{Re}(\alpha) > (p-1)/2$, the requirement $\mathrm{Re}(\alpha) - \frac{k-1}{2} > 0$ is fulfilled for all $k \in \{1, \dots, p\}$. With the help of Eq. (2.20) with $n = 1$ and $2t = \lambda_k$ for all $k > l$ in $\{1, \dots, p\}$

$$\int_{\mathbb{R}} \exp\left(-t_{kl}^2(1+\mathrm{i}\lambda_k)\right)\mathrm{d}t_{kl} = 2\int_{\mathbb{R}^+} \exp\left(-t_{kl}^2(1+\mathrm{i}\lambda_k)\right)\mathrm{d}t_{kl} = \frac{\sqrt{\pi}}{\sqrt{1+\mathrm{i}\lambda_k}}, \tag{2.63}$$

where the first equality follows from the symmetry of the integrand.

Inserting Eq. (2.62) and Eq. (2.63) into Eq. (2.61) results in

$$\mathcal{I} = \left(\prod_{\substack{k,l=1\\k>l}}^{p} \sqrt{\pi}\,(1+\mathrm{i}\lambda_k)^{-\frac{1}{2}}\right)\prod_{k=1}^{p}(1+\mathrm{i}\lambda_k)^{\frac{k-1}{2}-a}\,\Gamma\left(\alpha - \frac{k-1}{2}\right).$$

Using the fact that the determinant of the diagonal matrix $I_p + \mathrm{i}D$ is given as product of its diagonal entries $(1+\mathrm{i}\lambda_k)$, $k \in \{1, \dots, p\}$, its determinant is different from zero and the integral can be rearranged to

$$\mathcal{I} = \left(\det(I_p + \mathrm{i}D)\right)^{-\alpha}\pi^{\frac{p(p-1)}{4}}\prod_{k=1}^{p}\Gamma\left(\alpha - \frac{k-1}{2}\right). \tag{2.64}$$

Using the properties of the determinant and Eq. (2.54) where $S \in \mathbb{R}^{p\times p}$ is orthogonal,

$$\begin{aligned}\det(I_p + \mathrm{i}D) &= \det S \cdot \det(I_p + \mathrm{i}D)\cdot \det S^{-1} = \det(SS^{-1} + \mathrm{i}SDS^{-1})\\ &= \det(I_p + \mathrm{i}B^{-1}YB^{-1}) = \det(I_p + \mathrm{i}YB^{-1}B^{-1}) = \det(I_p + \mathrm{i}YX^{-1}),\end{aligned}$$

where Lemma 2.1, the Weinstein–Aronszajn identity given in Lemma 2.3 and the fact that $X = BB$ is used. Note that this implies that the determinant of $I_p + \mathrm{i}YX^{-1}$ is not equal to zero. Combining this and Eq. (2.64) into Eq. (2.53), we arrive at

$$\begin{aligned}&\int_{\mathbb{S}_p^+} \mathrm{e}^{-\mathrm{tr}(\Lambda Z)}(\det\Lambda)^{\alpha-\frac{p+1}{2}}\,\mathrm{d}\Lambda\\ &\qquad = (\det X)^{-\alpha}\left(\det(I_p + \mathrm{i}YX^{-1})\right)^{-\alpha}(2\pi)^{\frac{p(p-1)}{4}}\prod_{k=1}^{p}\Gamma\left(\alpha - \frac{k-1}{2}\right). \end{aligned} \tag{2.65}$$

Since

$$(I_p + \mathrm{i}YX^{-1})X = X + \mathrm{i}Y = Z,$$

the determinant of Z is given as the product of the determinants of X and $I_p + \mathrm{i}YX^{-1}$, is therefore different from zero and Eq. (2.65) can be simplified to

$$\int_{\mathbb{S}_p^+} \mathrm{e}^{-\mathrm{tr}(\Lambda Z)}(\det\Lambda)^{\alpha-\frac{p+1}{2}}\,\mathrm{d}\Lambda = (\det Z)^{-\alpha}(2\pi)^{\frac{p(p-1)}{4}}\prod_{k=1}^{p}\Gamma\left(\alpha - \frac{k-1}{2}\right),$$

which completes the proof. □

3. Matrix-Valued Gamma Distributions and their Properties

3.1. Wishart Distributions and their Basic Properties

Matrix-valued gamma distributions can be seen as the generalized form of the family of Wishart distributions, see Definition 3.8 below, which, in turn, is the matrix-valued version of the class of χ^2-distributions. In order to define these distributions, we start by introducing a matrix-valued version of normal distributions, which uses the Kronecker product.

Definition 3.1 (Kronecker product, see [17, Definition 4.2.1])**.** For $m, n, p, q \in \mathbb{N}$ the Kronecker product of $A \in \mathbb{R}^{p\times q}$ and $B \in \mathbb{R}^{m\times n}$ is defined to be the block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1q}B \\ \vdots & & \vdots \\ a_{p1}B & \dots & a_{pq}B \end{pmatrix} \in \mathbb{R}^{mp\times nq}. \tag{3.1}$$

Definition 3.2 (Matrix-valued normal distributions, see [13, Definition 2.2.1, Theorem 2.3.10])**.** Let $n, p \in \mathbb{N}$.

(i) An $\mathbb{R}^{p\times n}$-valued random matrix Z is said to follow the *standard matrix-valued normal distribution* if $\operatorname{vec}(Z^\top)$ follows the $(p \cdot n)$-dimensional normal distribution with mean vector zero and covariance matrix $I_p \otimes I_n = I_{pn}$, where $\operatorname{vec}(\cdot)$ is defined in Eq. (2.4) and $\otimes$ refers to the Kronecker product defined in Eq. (3.1).

(ii) For positive semi-definite matrices $\Sigma \in \mathbb{R}^{p\times p}$ and $\Psi \in \mathbb{R}^{n\times n}$, let S and P denote the positive semi-definite matrix square roots satisfying $SS = \Sigma$ and $PP = \Psi$ respectively, see Lemma 2.15. Let Z be a random matrix as described in Item (i). For a constant matrix $M \in \mathbb{R}^{p\times n}$, a random matrix X satisfying $X \stackrel{\mathrm{d}}{=} M + SZP$ is said to follow the *matrix-valued normal distribution with mean matrix M and covariance matrix*[1] $\Sigma \otimes \Psi$. In this case we write $X \sim \operatorname{MNormal}(M, \Sigma \otimes \Psi)$.

Remark 3.3 (Relationship between matrix-valued and multivariate normal distributions). Note that due to Definition 3.2 and Eqs. (3.52) and (3.53), a random matrix $X \sim \operatorname{MNormal}(M, \Sigma \otimes \Psi)$ if and only if $\operatorname{vec}(X^\top)$ follows a multivariate normal distribution with mean $\operatorname{vec}(M^\top)$ and covariance matrix $\Sigma \otimes \Psi$ – thus justifying the name of the parameter. Hence, not every random matrix whose vectorized version follows a normal distribution also follows a matrix-valued normal distribution.

[1] The fact that this input parameter $\Sigma \otimes \Psi$ is actually describes the covariance matrix will be elaborated in Remark 3.3.

Remark 3.4 (Characteristic function of a matrix-valued distribution, see [13, Section 1.9]). Let $n, p \in \mathbb{N}$. For an $\mathbb{R}^n$-valued random variable X the characteristic function is given as

$$\varphi_X(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle z, X\rangle}], \quad z \in \mathbb{R}^n,$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product. This characteristic function determines the distribution of x uniquely, as shown in [28, Theorem 2.30].

For an $\mathbb{R}^{p\times n}$-valued random matrix X we use the vectorization of matrices discussed in Remark 2.18 for the characteristic function,

$$\varphi_X(z) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle \mathrm{vec}(z), \mathrm{vec}(X)\rangle}], \quad z \in \mathbb{R}^{p\times n}.$$

This characteristic function then also uniquely determines the distribution of random matrix X. Since, using the definition and property of the Frobenius inner product given in Eq. (2.5),

$$\langle \mathrm{vec}(z), \mathrm{vec}(X)\rangle = \langle z, X\rangle_{\mathrm{F}} = \mathrm{tr}(z^\top X), \quad z \in \mathbb{R}^{p\times n},$$

the characteristic function of X can be written as

$$\varphi_X(z) = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top X)}], \quad z \in \mathbb{R}^{p\times n}. \tag{3.2}$$

Remark 3.5 (Density, characteristic function and affine transformations of matrix-valued normal distributions, see [13, Theorem 2.3.1, Theorem 2.2.1, Theorem 2.3.10]). In the setting of Definition 3.2, let $X \sim \mathrm{MNormal}(M, \Sigma \otimes \Psi)$.

(i) We claim that the characteristic function of X is given by

$$\varphi_X(z) = \mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top M - \frac{1}{2} z^\top \Sigma z \Psi)}, \quad z \in \mathbb{R}^{p\times n}.$$

The standard case $\Sigma = I_p$ and $\Psi = I_n$ follows from [13, Theorem 2.3.1]. For the general case, note that X can be represented by $X \overset{\mathrm{d}}{=} M + SZP$, where $Z \sim \mathrm{MNormal}(M, I_p \otimes I_n)$ and $\Sigma = SS$ as well as $\Psi = PP$. Hence, for every $z \in \mathbb{R}^{p\times n}$,

$$\varphi_X(z) = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top X)}] = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top (M+SZP))}] = \mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top M)}\, \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}Pz^\top SZ)}],$$

where Lemma 2.1 is used for the last equality. Using the characteristic function of the standard case Z,

$$\mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}Pz^\top SZ)}] = \varphi_Z((Pz^\top S)^\top) = \mathrm{e}^{\mathrm{tr}(-\frac{1}{2}Pz^\top S(Pz^\top S)^\top)} = \mathrm{e}^{\mathrm{tr}(-\frac{1}{2}z^\top SS^\top zP^\top P)},$$

where again Lemma 2.1 as well as the symmetry of S and P are used. Hence,

$$\varphi_X(z) = \mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top M)}\, \mathrm{e}^{\mathrm{tr}(-\frac{1}{2}z^\top SS^\top zP^\top P)} = \mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\top M - \frac{1}{2}z^\top \Sigma z \Psi)}.$$

(ii) If Σ and Ψ are of full rank, i.e. are positive definite, a density of X w.r.t. the $p \cdot n$-dimensional Lebesgue–Borel measure is given by

$$f_X(x) = (2\pi)^{-\frac{np}{2}} (\det \Sigma)^{-\frac{n}{2}} (\det \Psi)^{-\frac{p}{2}}\, \mathrm{e}^{-\mathrm{tr}\left(\frac{1}{2}\Sigma^{-1}(x-M)\Psi^{-1}(x-M)^\top\right)}, \quad x \in \mathbb{R}^{p\times n}, \tag{3.3}$$

see [13, Theorem 2.2.1]

(iii) For $q, m \in \mathbb{N}$ let $C \in \mathbb{R}^{q\times p}$ and $D \in \mathbb{R}^{n\times m}$. Then the random matrix $CXD \sim \mathrm{MNormal}(CMD, (C\Sigma C^\top) \otimes (D\Psi D^\top))$, see [13, Theorem 2.3.10].

Using matrix-valued normal distributions, the family of Wishart distributions can be defined, where we distinguish between a standard and a general case.

Definition 3.6 (Standard Wishart distributions)**.** For $n, p \in \mathbb{N}$, let Z be an $\mathbb{R}^{p\times n}$-valued random matrix following the standard matrix-valued normal distribution given in Definition 3.2, denoted by $Z \sim \mathrm{MNormal}(0, I_p \otimes I_n)$. Then an $\mathbb{R}^{p\times p}$-valued random matrix X satisfying

$$X \overset{\mathrm{d}}{=} ZZ^{\mathsf{T}} \quad \text{(equality in distribution)}$$

is said to follow the *standard Wishart distribution* with n degrees of freedom. This is denoted by $X \sim \mathrm{Wishart}(n, I_p)$. In case that $n < p$, the distribution of X is called *standard singular Wishart distribution.*

Lemma 3.7 (Independent entries of standard Wishart distributions)**.** *For $n, p \in \mathbb{N}$ let $X \sim \mathrm{Wishart}(n, I_p)$. Then the elements of every collection $X_{i_1,j_1}, \dots, X_{i_n,j_n}$ with $n \in \mathbb{N}$ of entries of X, where $\{i_k, j_k\} \cap \{i_l, j_l\} = \varnothing$ for all $k, l \in \{1, \dots, n\}$, are independent. This implies, in particular, that all diagonal elements of a standard Wishart are independent.*

Proof. Using Definition 3.6, let $Z \sim \mathrm{MNormal}(0, I_p \otimes I_n)$ be the random matrix satisfying $ZZ^{\mathsf{T}} \overset{\mathrm{d}}{=} X$. The entries of Z are independent per definition. Therefore,

$$X_{i_k,j_k} = \sum_{m=1}^{n} Z_{i_k m} Z_{j_k m} \quad k \in \{1, \dots, n\},$$

hence, if the conditions of the lemma are satisfied, $X_{i_1,j_1}, \dots, X_{i_n,j_n}$ are independent. □

Definition 3.8 (Wishart distributions, see [13, Definition 3.2.1], [20, Section 2.3])**.** For $n, p \in \mathbb{N}$ let $\Sigma \in \mathbb{R}^{p\times p}$ be a positive semi-definite matrix. A random matrix Y is said to follow the *Wishart distribution* with n degrees of freedom and covariance matrix Σ if there exists a dimension $q \in \mathbb{N}$, a matrix $A \in \mathbb{R}^{p\times q}$ such that $\Sigma = AA^{\mathsf{T}}$, and a random $\mathbb{R}^{q\times q}$-valued matrix $X \sim \mathrm{Wishart}(n, I_q)$ as given in Definition 3.6, such that

$$Y \overset{\mathrm{d}}{=} AXA^{\mathsf{T}} \quad \text{(equality in distribution).} \tag{3.4}$$

This is denoted by $Y \sim \mathrm{Wishart}(n, \Sigma)$. In case $\Sigma \neq 0$ and $n < \mathrm{rk}(\Sigma)$, and in case $\Sigma = 0$, the distribution of X is called *singular Wishart distribution.*

Lemma 3.9 (Existence of Wishart distributions)**.** *Let $n, p \in \mathbb{N}$ and $\Sigma \in \mathbb{R}^{p\times p}$. Then the* $\mathrm{Wishart}(n, \Sigma)$ *distribution exists if and only if Σ is positive semi-definite. In this case there exists an $A \in \mathbb{R}^{p\times q}$ such that $\Sigma = AA^{\mathsf{T}}$, where $q = \max\{1, \mathrm{rk}(\Sigma)\}$.*

The uniqueness of a Wishart distribution given Σ and n is demonstrated via the characteristic function in Theorem 3.14 below.

Proof of Lemma 3.9. For every matrix $A \in \mathbb{R}^{p\times q}$, the quadratic matrix $\Sigma = AA^\top$ is symmetric because

$$\Sigma^\top = (AA^\top)^\top = (A^\top)^\top A^\top = AA^\top = \Sigma$$

and positive semi-definite as

$$x^\top \Sigma x = x^\top AA^\top x = \|x^\top A\|_2^2 \geq 0, \quad x \in \mathbb{R}^p.$$

Conversely, for every positive semi-definite matrix Σ we claim there exists an $A \in \mathbb{R}^{p\times q}$ such that $\Sigma = AA^\top$ where $q = \max\{1, \mathrm{rk}(\Sigma)\}$. For $\Sigma = 0$ the matrix A can be chosen to be the zero column vector in $\mathbb{R}^p$. Now let $\Sigma \neq 0$. Since Σ is symmetric, there exists an orthogonal decomposition, see Theorem 2.5, given by $\Sigma = UDU^\top$, where $U \in \mathbb{R}^{p\times p}$ is orthogonal and $D \in \mathbb{R}^{p\times p}$ is a diagonal matrix with the eigenvalues of Σ as diagonal entries with the corresponding multiplicity. As Σ is positive semi-definite, these eigenvalues are non-negative and therefore have a real square root. For $D = \mathrm{diag}(d_1, \dots, d_n)$, let

$$\sqrt{D} := \mathrm{diag}(\sqrt{d_1}, \dots, \sqrt{d_n}).$$

In case that Σ does not have full rank, only $\mathrm{rk}(\Sigma)$ of the eigenvalues $d_1, \dots, d_n$ are not equal to zero. Let $\tilde{D}$ denote the $p \times \mathrm{rk}(\Sigma)$-dimensional matrix that is left when removing all columns equal to the zero vector from $\sqrt{D}$. Finally, the matrix $A := U\tilde{D} \in \mathbb{R}^{p\times \mathrm{rk}(\Sigma)}$ fulfills $AA^\top = \Sigma$ and is of rank $\mathrm{rk}(\Sigma)$. □

Lemma 3.10 (Alternative definition of Wishart distributions)**.** *For $n, p \in \mathbb{N}$, let $\Sigma \in \mathbb{R}^{p\times p}$ be a positive semi-definite matrix and let $X \sim \mathrm{MNormal}(0, \Sigma \otimes I_n)$. Then $XX^\top \sim \mathrm{Wishart}(n, \Sigma)$.*

Proof. Due to Definition 3.2 (ii), the random matrix X given here can be written as $X \overset{\mathrm{d}}{=} SZ$, where positive semi-definite $S \in \mathbb{R}^{p\times p}$ satisfies $SS = \Sigma$ and $Z \sim \mathrm{MNormal}(0, I_p \otimes I_n)$. Hence,

$$XX^\top \overset{\mathrm{d}}{=} SZ(SZ)^\top = SZZ^\top S^\top,$$

which, using Eq. (3.4) and Definition 3.8, follows a Wishart distribution with n degrees of freedom and covariance matrix $SS^\top = SS = \Sigma$. □

Remark 3.11 (The family of Wishart distribution takes values in the positive semi-definite matrices)*.* For every $n, p \in \mathbb{N}$, the random matrix $X \sim \mathrm{Wishart}(n, I_p)$ takes values in the positive semi-definite matrices a.s. since it has the stochastic representation $X \overset{\mathrm{d}}{=} ZZ^\top$ with $Z \sim \mathrm{MNormal}(0, I_p \otimes I_n)$, see Definition 3.6.

Now for a positive semi-definite $\Sigma \neq 0$, let A denote the $\mathbb{R}^{p\times \mathrm{rk}(\Sigma)}$ dimensional matrix satisfying $AA^\top = \Sigma$ as discussed in Lemma 3.9. For $X \sim \mathrm{Wishart}(n, I_{\mathrm{rk}(\Sigma)})$ the values of the random matrix $Y \overset{\mathrm{d}}{=} AXA^\top \sim \mathrm{Wishart}(n, \Sigma)$ are symmetric a.s. as well as $(AXA^\top)^\top = AXA^\top$ and positive semi-definite a.s. as

$$x^\top Y x \overset{\mathrm{d}}{=} x^\top AXA^\top x = x^\top AX(x^\top A)^\top \geq 0, \quad x \in \mathbb{R}^p.$$

If $\Sigma = 0$ then Y is deterministic and equal to the zero matrix, which is positive semi-definite as well.

Remark 3.12 (Sampling from a Wishart distribution). Combining the definition of standard Wishart distributions directly derived from i.i.d. standard normal variables[2] in Definition 3.6 with the definition of general Wishart distributions via Eq. (3.4) above provides a method to sample from $\mathrm{Wishart}(n, \Sigma)$ for all $n \in \mathbb{N}$ and positive semi-definite Σ.

Remark 3.13 (Characteristic function of a symmetric matrix-valued distribution [20, Remark 1]). We will demonstrate that for an $\mathbb{R}^{p\times p}$-valued random matrix X, that is a.s. symmetric, the characteristic function $\varphi_X(z)$ is already determined uniquely by its values for symmetric $z \in \mathbb{R}^{p\times p}$. Since X is symmetric a.s., using Lemma 2.1 and the fact that the trace of a matrix is equal to the trace of the transpose of that matrix,

$$\mathrm{tr}(z^\mathsf{T} X) \overset{\text{a.s.}}{=} \frac{1}{2}\big(\mathrm{tr}(z^\mathsf{T} X) + \mathrm{tr}(z^\mathsf{T} X^\mathsf{T})\big) = \frac{1}{2}\big(\mathrm{tr}(z^\mathsf{T} X) + \mathrm{tr}(zX)\big) = \mathrm{tr}\Big(\frac{z^\mathsf{T} + z}{2} X\Big), \quad z \in \mathbb{R}^{p\times p}.$$

Therefore it is sufficient to just look at the values for symmetric $z \in \mathbb{R}^{p\times p}$. This also leads to

$$\varphi_X(z) = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}z^\mathsf{T} X)}] = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zX)}] \tag{3.5}$$

for every symmetric $z \in \mathbb{R}^{p\times p}$.

Theorem 3.14 (Characteristic function of Wishart distributions, see [20, Proposition 3.2]). *For $p, q, n \in \mathbb{N}$ and $A \in \mathbb{R}^{p\times q}$ defining $\Sigma = AA^\mathsf{T} \in \mathbb{R}^{p\times p}$ let $Y \sim \mathrm{Wishart}(n, \Sigma)$.*

(i) For $w \in \mathbb{C}^{p\times p}$, let $\tilde{w} := (w + w^\mathsf{T})/2$ with representation $\tilde{w} = u + \mathrm{i}v$ with $u, v \in \mathbb{R}^{p\times p}$ symmetric. If[3] $I_q + 2A^\mathsf{T} uA \in \mathbb{S}_q^+$, then the determinant of $I_p + 2\tilde{w}\Sigma$ is different from zero and

$$\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(wY)}] = \big(\det(I_p + 2\tilde{w}\Sigma)\big)^{-n/2}, \tag{3.6}$$

where the complex square root here and in Eq. (3.7) below is understood as defined in Eq. (2.18).

(ii) In particular, the characteristic function of Y is given by

$$\varphi_Y(z) = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)}] = \big(\det(I_p - 2\mathrm{i}z\Sigma)\big)^{-n/2} \tag{3.7}$$

for all symmetric $z \in \mathbb{R}^{p\times p}$.

This theorem proves that a Wishart distribution is uniquely determined by the parameters n and Σ.

[2] The standard normal variables can, for example, be sampled using the Box–Muller method [4]: Let U, V be two random variables following the uniform distribution on unit interval $(0, 1)$. Then $Z_1 = \sqrt{-2\log(U)}\cos(2\pi)$ and $Z_2 = \sqrt{-2\log(U)}\sin(2\pi)$ are independent and follow the standard normal distribution.

[3] The condition on $w \in \mathbb{C}^{p\times p}$ is in particular satisfied if the real part of w or of $\tilde{w}$ is positive semi-definite, especially if it is equal to the zero matrix.

Proof. This proof of Item (i) is split into seven steps. Item (ii) follows from Item (i) by inserting $w = \mathrm{i}z$ for a symmetric $z \in \mathbb{R}^{p\times p}$.

Step 1. First, this theorem will be proven for standard Wishart distributions, so let the $\mathbb{R}^{p\times p}$-valued random matrix X follow the Wishart distribution with n degrees of freedom and covariance matrix I_p. Fix $w \in \mathbb{C}^{p\times p}$ satisfying the above-mentioned conditions. With an analogue argument to Remark 3.13, $\operatorname{tr}(wY) = \operatorname{tr}(\tilde{w}Y)$ a.s.

Using Eq. (3.4), there exists an $\mathbb{R}^{p\times n}$-valued random matrix $Z \sim \mathrm{MNormal}(0, I_p \otimes I_n)$ satisfying $ZZ^\mathsf{T} \overset{\mathrm{d}}{=} X$. Therefore,

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\operatorname{tr}(wX)}] = \mathbb{E}[\mathrm{e}^{-\operatorname{tr}(\tilde{w}ZZ^\mathsf{T})}] &= (2\pi)^{-\frac{np}{2}} \int_{\mathbb{R}^{p\times n}} \mathrm{e}^{-\operatorname{tr}(\tilde{w}zz^\mathsf{T})}\, \mathrm{e}^{-\frac{1}{2}\operatorname{tr}(zz^\mathsf{T})}\, \mathrm{d}z \\ &= (2\pi)^{-\frac{np}{2}} \int_{\mathbb{R}^{p\times n}} \mathrm{e}^{-\frac{1}{2}\operatorname{tr}((I_p+2u+2\mathrm{i}v)zz^\mathsf{T})}\, \mathrm{d}z, \end{aligned} \tag{3.8}$$

where we use the density function of Z given in Eq. (3.3). As $I_p + 2u \in \mathbb{S}_p^+$ there exists a matrix $B \in \mathbb{S}_p^+$ such that

$$BB = I_p + 2u \tag{3.9}$$

due to Lemma 2.15. Due to Item (vi) of Lemma 2.8, the matrix B^{-1} is positive definite and therefore the map

$$\phi := \begin{cases} \mathbb{R}^{p\times n} & \to \quad \mathbb{R}^{p\times n} \\ \quad \tilde{z} & \mapsto \quad B^{-1}\tilde{z} \end{cases},$$

satisfies $\det \mathrm{d}\phi = (\det B^{-1})^n$ due to Item (iv) of Lemma 2.41. Inserting this map and Eq. (3.8) into Theorem 2.39,

$$\mathbb{E}[\mathrm{e}^{-\operatorname{tr}(wX)}] = \underbrace{(\det B^{-1})^n}_{=(\det(I_p+2u))^{-\frac{n}{2}}}(2\pi)^{-\frac{np}{2}} \int_{\mathbb{R}^{p\times n}} \mathrm{e}^{-\frac{1}{2}\operatorname{tr}((I_p+2u+2\mathrm{i}v)B^{-1}\tilde{z}(B^{-1}\tilde{z})^\mathsf{T})}\, \mathrm{d}\tilde{z}. \tag{3.10}$$

Due to Lemma 2.1, Eq. (3.9) and the fact that B^{-1} is symmetric,

$$\operatorname{tr}((I_p + 2u + 2\mathrm{i}v)B^{-1}\tilde{z}(B^{-1}\tilde{z})^\mathsf{T}) = \operatorname{tr}((I_p + 2\mathrm{i}B^{-1}vB^{-1})\tilde{z}\tilde{z}^\mathsf{T}).$$

Hence, Eq. (3.10) can be rewritten as

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\operatorname{tr}(wX)}] &= (\det(I_p + 2u))^{-\frac{n}{2}}(2\pi)^{-\frac{np}{2}} \int_{\mathbb{R}^{p\times n}} \mathrm{e}^{-\frac{1}{2}\operatorname{tr}((I_p+2\mathrm{i}B^{-1}vB^{-1})\tilde{z}\tilde{z}^\mathsf{T})}\, \mathrm{d}\tilde{z} \\ &= (\det(I_p + 2u))^{-\frac{n}{2}}\mathbb{E}[\mathrm{e}^{-\operatorname{tr}(\mathrm{i}B^{-1}vB^{-1}\tilde{Z}\tilde{Z}^\mathsf{T})}], \end{aligned} \tag{3.11}$$

where $\tilde{Z} \sim \mathrm{MNormal}(0, I_p \otimes I_n)$ due to Eq. (3.3).

Step 2. Now for every symmetric $v \in \mathbb{R}^{p\times p}$, the matrix $B^{-1}vB^{-1}$ must be symmetric as well as B^{-1} is symmetric. Due to Theorem 2.5 there exists a diagonal matrix $D = \operatorname{diag}(\lambda_1, \dots, \lambda_p)$ with $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ and an orthogonal matrix $S \in \mathbb{R}^{p\times p}$ such that $B^{-1}vB^{-1} = S^\mathsf{T}DS$, see Theorem 2.5. Hence,

$$\mathbb{E}[\mathrm{e}^{-\operatorname{tr}(\mathrm{i}B^{-1}vB^{-1}\tilde{Z}\tilde{Z}^\mathsf{T})}] = \mathbb{E}[\mathrm{e}^{-\operatorname{tr}(S^\mathsf{T}DS\tilde{Z}\tilde{Z}^\mathsf{T})}] = \mathbb{E}[\mathrm{e}^{-\operatorname{tr}(D(S\tilde{Z})(S\tilde{Z})^\mathsf{T})}], \tag{3.12}$$

where the last equality follows from Lemma 2.1. Since affine transformations of normal distributions are normal distributions, see Item (iii) from Remark 3.5, $S\tilde{Z} \sim \mathrm{MNormal}(0, SI_pS^\mathsf{T} \otimes I_n)$. As S is orthogonal, $\tilde{Z}$ and $S\tilde{Z}$ follow the same distribution and therefore

$$(\tilde{Z}S)(\tilde{Z}S)^\mathsf{T} \overset{\mathrm{d}}{=} \tilde{Z}\tilde{Z}^\mathsf{T}. \tag{3.13}$$

Step 3. Note that due to the definition of the matrix multiplication

$$\tilde{Z}\tilde{Z}^\mathsf{T} = \sum_{j=1}^{n} \tilde{Z}_j^c(\tilde{Z}_j^c)^\mathsf{T}, \tag{3.14}$$

where $\tilde{Z}_j^c \in \mathbb{R}^p$ denotes the j-th column of $\tilde{Z}$ for every $j \in \{1, \dots, n\}$. Since all entries of $\tilde{Z}$ are independent and identically distributed normal variables, the same holds true for its columns and $\tilde{Z}_j^c \sim \mathrm{Normal}(0, I_p)$ for every $j \in \{1, \dots, n\}$. Hence, using Eq. (3.12), Eq. (3.13) and Eq. (3.14),

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}B^{-1}vB^{-1}\tilde{Z}\tilde{Z}^\mathsf{T})}] &= \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}\tilde{Z}^\mathsf{T})}] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\sum_{j=1}^{n}\tilde{Z}_j^c(\tilde{Z}_j^c)^\mathsf{T})}] \\ &= \prod_{j=1}^{n} \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}_j^c(\tilde{Z}_j^c)^\mathsf{T})}] = \big(\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}^c(\tilde{Z}^c)^\mathsf{T})}]\big)^n \end{aligned} \tag{3.15}$$

where $\tilde{Z}^c \sim \mathrm{Normal}(0, I_p)$.

Step 4. Each of the factors can now be computed as

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}^c(\tilde{Z}^c)^\mathsf{T})}] &= (2\pi)^{-\frac{p}{2}} \int_{\mathbb{R}^p} \mathrm{e}^{-\mathrm{tr}(\mathrm{i}Dxx^\mathsf{T})}\, \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(xx^\mathsf{T})}\, \mathrm{d}x \\ &= (2\pi)^{-\frac{p}{2}} \int_{\mathbb{R}^p} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}((I_p+2\mathrm{i}D)xx^\mathsf{T})}\, \mathrm{d}x, \end{aligned} \tag{3.16}$$

where the first equality follows from the definition of a multivariate standard normal distribution, as it can also be seen in Eq. (3.3). Let

$$G := I_p + 2\mathrm{i}D = \mathrm{diag}(1 + 2\mathrm{i}\lambda_1, \dots, 1 + 2\mathrm{i}\lambda_p) \in \mathbb{C}^{p\times p}.$$

Then, for $x \in \mathbb{R}^p$

$$\mathrm{tr}((I_p + 2\mathrm{i}D)xx^\mathsf{T}) = \mathrm{tr}(Gxx^\mathsf{T}) = \sum_{k=1}^{p}\sum_{l=1}^{p} G_{kl}x_lx_k = \sum_{k=1}^{p} G_{kk}x_k^2,$$

where the last equality follows from the fact that G is a diagonal matrix. Combining this with Eq. (3.16),

$$\mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}D\tilde{Z}^c(\tilde{Z}^c)^\mathsf{T})}] = (2\pi)^{-\frac{p}{2}} \int_{\mathbb{R}^p} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^{p} G_{kk}x_k^2}\, \mathrm{d}x = \prod_{k=1}^{p} (2\pi)^{-\frac{1}{2}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}G_{kk}x_k^2}\, \mathrm{d}x_k, \tag{3.17}$$

using Fubini's theorem for the last equality. Considering Definition 2.28, each factor of this product corresponds to the characteristic function of a χ_1^2 distribution at $t := 2\lambda_k$. Hence, using Corollary 2.35 with $n = 1$,

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}G_{kk}x_k^2}\, \mathrm{d}x_k = \frac{1}{\sqrt{G_{kk}}}.$$

Step 5. Combining this with Eq. (3.15) and Eq. (3.17),

$$\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}\tilde{Z}^\mathsf{T})}] = \Big(\prod_{k=1}^{p} \frac{1}{\sqrt{G_{kk}}}\Big)^n = \pm\Big(\prod_{k=1}^{p} G_{kk}\Big)^{-\frac{n}{2}} = \pm(\det G)^{-\frac{n}{2}} = \pm\big(\det(I_p + 2\mathrm{i}D)\big)^{-\frac{n}{2}}, \tag{3.18}$$

where the second equality follows from Items (i) and (vi) of Proposition 2.32, the third equality holds as G is a diagonal matrix and the last equality follows from the definition of G. Since $\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}\tilde{Z}^\mathsf{T})}]$ is a continuous function of D and $\big(\det(I_p + 2\mathrm{i}D)\big)^{-n/2}$ is a continuous function of D taking values in $\mathbb{C} \setminus \{0\}$,

$$\frac{\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}\tilde{Z}^\mathsf{T})}]}{\big(\det(I_p + 2\mathrm{i}D)\big)^{-n/2}}$$

is a continuous function of D taking values in $\{-1, 1\}$. As for $D = 0$ both the numerator and the denominator of the function are equal to one

$$\frac{\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}\tilde{Z}^\mathsf{T})}]}{\big(\det(I_p + 2\mathrm{i}D)\big)^{-n/2}} = 1$$

for all diagonal matrices D and therefore Eq. (3.18) holds without the $\pm$ sign.

Step 6. Due to Eq. (3.15) and as for every orthogonal matrix S the equality $\det S \det S^\mathsf{T} = \det(SS^\mathsf{T}) = \det(SS^{-1}) = 1$ holds,

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}B^{-1}vB^{-1}\tilde{Z}\tilde{Z}^\mathsf{T})}] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}D\tilde{Z}\tilde{Z}^\mathsf{T})}] &= \big(\det S^\mathsf{T} \det(I_p + 2\mathrm{i}D)\det S\big)^{-\frac{n}{2}} \\ &= \big(\det(S^\mathsf{T}S + 2\mathrm{i}S^\mathsf{T}DS)\big)^{-\frac{n}{2}} = \big(\det(I_p + 2\mathrm{i}B^{-1}vB^{-1})\big)^{-\frac{n}{2}},\end{aligned}$$

where $S^\mathsf{T}S = I_p$ and $S^\mathsf{T}DS = B^{-1}vB^{-1}$ were used for the last equality. Hence,

$$\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\mathrm{i}B^{-1}vB^{-1}\tilde{Z}\tilde{Z}^\mathsf{T})}] = \big(\det(I_p + 2\mathrm{i}B^{-1}vB^{-1})\big)^{-\frac{n}{2}} = \big(\det(I_p + 2\mathrm{i}v\underbrace{B^{-1}B^{-1}}_{=(I_p+2u)^{-1}})\big)^{-\frac{n}{2}},$$

where the second equality holds due to Lemma 2.3. Combining this with Eq. (3.11)

$$\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(wX)}] = (\det(I_p + 2u))^{-\frac{n}{2}}\big(\det(I_p + 2\mathrm{i}v(I_p + 2u)^{-1})\big)^{-\frac{n}{2}} = (\det(I_p + 2\tilde{w}))^{-\frac{n}{2}}, \tag{3.19}$$

thus proving the standard case.

Step 7. For $Y \sim \mathrm{Wishart}(n, \Sigma)$ with $\Sigma = AA^\mathsf{T}$, where $A \in \mathbb{R}^{p\times q}$, using Definition 3.8, we know that Y has he stochastic representation $Y \overset{\mathrm{d}}{=} AXA^\mathsf{T}$, where X follows the standard Wishart distribution of dimension $q \in \mathbb{N}$. Then, for every $w \in \mathbb{C}^{p\times p}$ satisfying the conditions of this theorem and the corresponding $\tilde{w} = u + \mathrm{i}v$,

$$\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(wY)}] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\tilde{w}AXA^\mathsf{T})}] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}((A^\mathsf{T}(u+\mathrm{i}v)A)X)}],$$

where we are using Lemma 2.1 for the second equation. As $I_p + 2A^\mathsf{T}uA$ is positive definite and $A^\mathsf{T}vA$ is symmetric for symmetric z, we can insert the standard case derived in Eq. (3.19),

$$\mathbb{E}[\mathrm{e}^{-\mathrm{tr}(wY)}] = (\det(I_q + 2A^\mathsf{T}\tilde{w}A))^{-\frac{n}{2}} = (\det(I_p + 2\tilde{w}\underbrace{AA^\mathsf{T}}_{=\Sigma}))^{-n/2},$$

where we use the Weinstein–Aronszajn identity (Lemma 2.3) for the second equality. □

Some more practical applications of Wishart distributions will be discussed in the following two remarks.

Remark 3.15 (The sample covariance matrix of a multivariate normal sample is following a Wishart distribution, see [13, Section 3.3.2]). One of the main applications of Wishart distributions is the fact that the distribution of the sample covariance matrix of a sample following the multivariate normal distribution has a Wishart distribution.

For $n, p \in \mathbb{N}$ let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d. random vectors following an $\mathbb{R}^p$-valued multivariate normal distribution with mean vector $\mu \in \mathbb{R}^p$ and covariance matrix Σ. Then the sample covariance matrix

$$\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathsf{T}},$$

where $\bar{\mathbf{x}}$ refers to the sample mean, takes values in $\mathbb{R}^{p \times p}$ and follows the Wishart$(n-1, \Sigma)$ distribution as proven in [13, Theorem 3.3.6].

Remark 3.16 (Further applications of Wishart distributions). Wishart distributions also come up in the following settings:

- In continuation to Remark 3.15, the class of Wishart distributions is used for maximum-likelihood estimation of the covariance matrix of the multivariate normal distribution, see [5, Chapter 6.6].
- In Bayesian statistics Wishart distributions are used for the prior probability distribution of the precision matrix, which is the inverse of the covariance matrix, of a multivariate normal distribution with known mean, see [16, Chapter 7.3].

3.2. Definition and Basic Properties of Matrix-Valued Gamma Distributions

In this section we first introduce standard matrix-valued gamma distributions via density functions before extending them to the commonly known matrix-valued gamma distributions in an analogue way to Definition 3.8 of Wishart distributions. Then, after showing that the family of matrix-valued gamma distributions is an extension to the class of non-singular Wishart distributions, we define *general matrix-valued gamma distributions* by also taking into account singular Wishart cases. For this distribution family we derive several basic properties.

Remark 3.17 (The measure on $\mathbb{S}_p^+$ implicitly given by the $p(p+1)/2$-dimensional Lebesgue–Borel measure). As discussed in Remark 2.22, for $k, n \in \mathbb{N}$ we define an integral over a k-dimensional submanifold of $\mathbb{R}^n$ by transforming it to an integral of the k-dimensional Lebesgue–Borel measure via one/multiple charts, where the value of the integral does not depend on the choice of the chart. Hence, this the k-dimensional Lebesgue–Borel measure implicitly defines a measure on the k-dimensional submanifold.

In the case of the submanifold $\mathbb{S}_p^+$ we will denote the measure implicitly given by the $p(p+1)/2$-dimensional Lebesgue–Borel measure with μ.

Using the relative topology on $\mathbb{S}_p^+$ as subset of $\mathbb{R}^{p\times p}$, we can define the Borel σ-algebra[4] $\mathcal{B}(\mathbb{S}_p^+)$. For every Borel-set $B \in \mathcal{B}(\mathbb{S}_p^+)$ and every chart ψ, the set $\psi(B)$ is measurable w.r.t. the $p(p+1)/2$-dimensional Lebesgue–Borel measure, see [2, Chapter XII, Remark 1.1, Theorem 1.2], hence B is measurable w.r.t. μ.

For any random variable X taking values in $\mathbb{S}_p^+$ with density function

$$f_X \colon \mathbb{S}_p^+ \to \mathbb{R}_0^+,$$

w.r.t. the measure μ, the function $x^\triangle := \psi^{-1}(X)$ defines a random variable since ψ, defined in Eq. (2.7), is a homeomorphism. Using Remark 2.23, a density function of $X^\triangle$ is given by

$$f_{X^\triangle} := \begin{cases} \tilde{\mathbb{S}}_p^+ & \to \quad \mathbb{R}_0^+ \\ x^\triangle & \mapsto \quad 2^{\frac{p(p-1)}{4}} f_X(\psi(x^\triangle)) = 2^{\frac{p(p-1)}{4}} f_X(x) \end{cases} \tag{3.20}$$

w.r.t. the $p(p+1)/2$-dimensional Lebesgue–Borel measure.

Definition 3.18 (Standard matrix-valued gamma distributions)**.** For $p \in \mathbb{N}$ a $\mathbb{S}_p^+$-valued random variable X with density function

$$f_X(x) = \frac{1}{\Gamma_p(\alpha)} \mathrm{e}^{-\operatorname{tr}(x)} (\det x)^{\alpha - \frac{p+1}{2}}, \quad x \in \mathbb{S}_p^+, \tag{3.21}$$

w.r.t. the measure μ on $\mathbb{S}_p^+$, that is implicitly given by the $p(p+1)/2$-dimensional Lebesgue–Borel measure, as discussed in Remark 3.17, is said to be standard matrix-valued gamma distributed with shape parameter $\alpha > (p-1)/2$, denoted by $X \sim \mathrm{MGamma}(\alpha, I_p)$.

Example 3.19 (Density function of a $\mathbb{S}_2^+$-valued standard matrix-valued gamma distribution)**.** As the $\mathrm{MGamma}(\alpha, I_p)$ distribution is matrix-valued, visualizing it via a density function poses a challenge. For $p = 2$ it is possible to illustrate the different level lines of the density function with an appropriate parameterization: Using Remark 2.19 and Eq. (2.7) in particular,

$$f_X(x) = f_X(\psi(x^\triangle)) = f_X(\psi(x_{11}, x_{12}, x_{22})), \quad x = \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \in \mathbb{S}_2^+.$$

Note that due to Sylvester's criterium, see [17, Corollary 7.1.5, Theorem 7.2.5], a matrix is positive definite if and only if its principal minors are positive, hence $\psi(x_{11}, x_{12}, x_{22}) \in \mathbb{S}_2^+$ if and only if $x_{11}, x_{22} \in \mathbb{R}^+$ and $\det(\psi(x_{11}, x_{12}, x_{22})) = x_{11}x_{22} - x_{12}^2 > 0$.

For $\alpha = 7/2$ some level lines are illustrated in Figure 3.1. The maximum point of the density function with these parameters is at $(x_{11}, x_{12}, x_{22}) = (2, 0, 2)$ where $f(\psi(2, 0, 2)) = 0.293$.

Remark 3.20 (Standard Wishart distributions are well-defined). We can see that the function f_X from Eq. (3.21) is in fact a probability density as $x \in \mathbb{S}_p^+$ leads to $\det x \geq 0$ and, by

[4] For a topological space the Borel σ-algebra is defined to be the smallest σ-algebra containing all open subsets of the topological space.

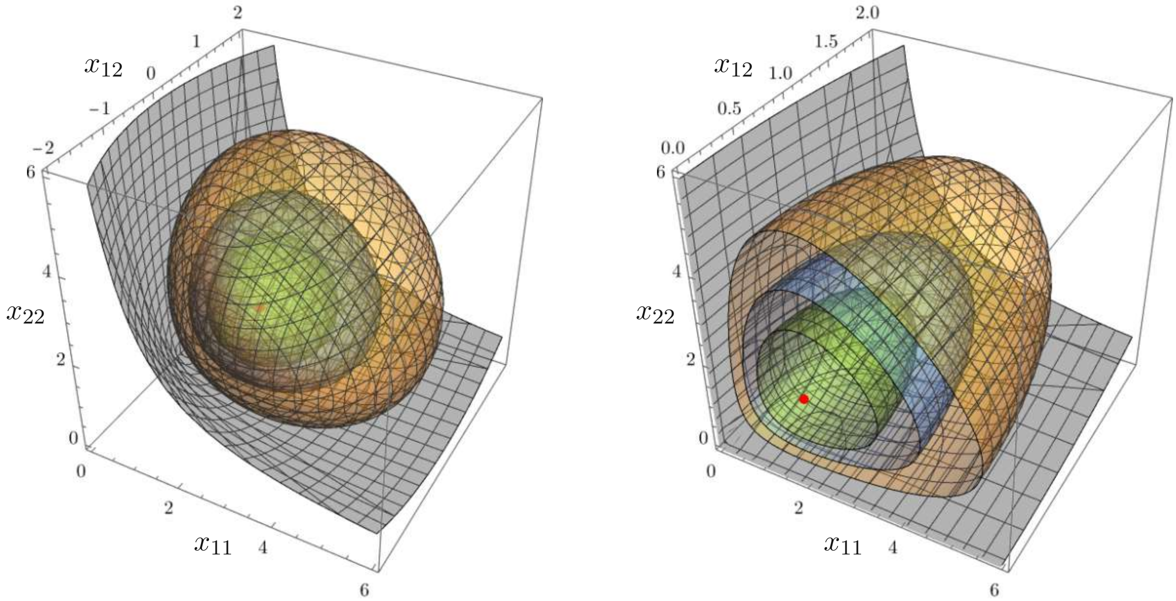


Figure 3.1. *Illustration of a density function of* $X \sim \mathrm{MGamma}(7/2, I_2)$ *by its contour lines, see Example 3.19. The shown levels are* $f_X(\psi(x_{11}, x_{12}, x_{22})) = 0.012$ *in green,* $f_X(\psi(x_{11}, x_{12}, x_{22})) = 0.008$ *in blue and* $f_X(\psi(x_{11}, x_{12}, x_{22})) = 0.004$ *in orange. The maximum point of this function is shown by the red point. The grey area shows the border to where* $f_X(\psi(x_{11}, x_{12}, x_{22})) = 0$ *due to the conditions on permitted* (x_{11}, x_{12}, x_{22}) *such that* $\psi(x_{11}, x_{12}, x_{22}) \in \mathbb{S}_2^+$.

using Eq. (2.31) with $Z = I_p$,

$$\begin{aligned}\int_{\mathbb{S}_p^+} f_X(x)\,\mathrm{d}x &= \frac{1}{\Gamma_p(\alpha)} \int_{\mathbb{S}_p^+} \mathrm{e}^{-\operatorname{tr}(x)} (\det x)^{\alpha - \frac{p+1}{2}}\,\mathrm{d}x \\ &= \frac{1}{\Gamma_p(\alpha)} \Gamma_p(\alpha) (\det I_p)^{-\alpha} = 1.\end{aligned}$$

Example 3.21 (Joint distribution of $(\det(X), \operatorname{tr}(X))$ for $X \sim \mathrm{MGamma}(\alpha, I_2)$)**.** For this example, we consider a $\mathbb{S}_2^+$-valued $X \sim \mathrm{MGamma}(\alpha, I_2)$, where $\alpha > 1/2$. Then

$$\det(X) = X_{11}X_{22} - X_{12}^2, \quad \operatorname{tr}(X) = X_{11} + X_{22}.$$

Let $X^\triangle$ be the corresponding $\tilde{\mathbb{S}}_2^+$-valued random matrix, which then, due to Eqs. (3.20) and (3.21), has a density

$$\begin{aligned}f_{X^\triangle}(x^\triangle) = \sqrt{2} f_X(x) &= \frac{\sqrt{2}}{\Gamma_2(\alpha)} \mathrm{e}^{-\operatorname{tr}(x)} (\det x)^{\alpha - \frac{3}{2}} \\ &= \frac{\sqrt{2}}{\Gamma_2(\alpha)} \mathrm{e}^{-(x_{11}+x_{22})} (x_{11}x_{22} - x_{12}^2)^{\alpha - \frac{3}{2}}, \quad x^\triangle = (x_{11}, x_{12}, x_{22}) \in \tilde{\mathbb{S}}_2^+, \qquad (3.22)\end{aligned}$$

where $x = \psi(x^\triangle)$ and $\Gamma_2(\alpha)$ can be expressed with Eq. (2.30).

Consider the continuous function

$$\phi := \begin{cases} \tilde{\mathbb{S}}_2^+ & \to & R \subset \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \\ (a,b,c) & \mapsto & (ac - b^2, a+c, b) =: (d,t,z) \end{cases},$$

which satisfies $\phi(X^\triangle) = (\det(X), \mathrm{tr}(X), X_{12})$. As, in the setting of this function, the inequality $d + z^2 = ac \le (a+c)^2/4 = t^2/4$ holds, the range of ϕ can be restricted to

$$R := \{(d,t,z) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \mid t \ge 2\sqrt{d}, z^2 \le t^2/4 - d\}, \tag{3.23}$$

making it surjective. Now, for $(d,t,z) \in R$, the equality $\phi(a,b,c) = (d,t,z)$ has two solutions for (a,b,c) in $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}$:

$$\begin{aligned} (a_1,b_1,c_1) &= \Big(\frac{1}{2}\big(t + \sqrt{t^2 - 4(z^2+d)}\big), \frac{1}{2}\big(t - \sqrt{t^2 - 4(z^2+d)}\big), z\Big), \\ (a_2,b_2,c_2) &= \Big(\frac{1}{2}\big(t - \sqrt{t^2 - 4(z^2+d)}\big), \frac{1}{2}\big(t + \sqrt{t^2 - 4(z^2+d)}\big), z\Big), \end{aligned} \tag{3.24}$$

which are only equal if $t^2 = 4(z^2+d)$, meaning $a_1 = c_1 = a_2 = c_2$. Hence, we split up the domain of ϕ in order to make the function injective and let $\phi_i := \phi\restriction_{D_i}$ for $i \in \{1,2\}$, where

$$D_1 := \{(a,b,c) \in \tilde{\mathbb{S}}_2^+ \,|\, a > c\}, \quad D_2 := \{(a,b,c) \in \tilde{\mathbb{S}}_2^+ \,|\, a < c\},$$

hence ϕ_i is bijective when restricting R to the same set given in Eq. (3.23) with strict inequalities, with $\phi_i^{-1}(d,t,z)$ given by (a_i,b_i,c_i) from Eq. (3.24) for $i \in \{1,2\}$. The Jacobi matrix of ϕ_1^{-1} is given by

$$d\phi_1^{-1}(d,t,z) = \begin{pmatrix} \frac{\partial a_1}{\partial d} & \frac{\partial a_1}{\partial t} & \frac{\partial a_1}{\partial z} \\ \frac{\partial b_1}{\partial d} & \frac{\partial b_1}{\partial t} & \frac{\partial b_1}{\partial z} \\ \frac{\partial c_1}{\partial d} & \frac{\partial c_1}{\partial t} & \frac{\partial c_1}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{-1}{C} & \frac{1}{2} + \frac{t}{2C} & * \\ \frac{1}{C} & \frac{1}{2} - \frac{t}{2C} & * \\ 0 & 0 & 1 \end{pmatrix} \text{ with } C := \sqrt{t^2 - 4(z^2+d)},$$

hence, using the Laplace expansion of the third row,

$$|\det d\phi_1^{-1}(d,t,z)| = \left| \frac{-1}{C}\Big(\frac{1}{2} - \frac{t}{2C}\Big) - \frac{1}{C}\Big(\frac{1}{2} + \frac{2}{2C}\Big) \right| = \frac{1}{\sqrt{t^2 - 4(z^2+d)}}. \tag{3.25}$$

The determinant of $d\phi_2^{-1}$ can be computed analogously and its absolute value also given by the right hand-side of Eq. (3.25).

Now for every open set $B \in \mathcal{B}(R)$,

$$\mathbb{P}[(\det(X), \mathrm{tr}(X), X_{12}) \in B] = \mathbb{P}[\phi(X^\triangle) \in B] = \mathbb{P}[X^\triangle \in \phi^{-1}(B)].$$

Note that $\phi^{-1}(B) = \phi^{-1}(B) \cap D_1 \,\dot\cup\, \phi^{-1}(B) \cap D_2 \,\dot\cup\, \phi^{-1}(B) \cap \tilde{\mathbb{S}}_2^+ \backslash (D_1 \cup D_2)$ and

$$\mathbb{P}[\phi(X^\triangle) \in \tilde{\mathbb{S}}_2^+ \backslash (D_1 \cup D_2)] = \mathbb{P}[X_{11} = X_{22}] = 0,$$

as X_{11} and X_{22} are independent and both follow absolutely continuous probability distributions, see Corollaries 3.23 and 3.53 below. As additionally $\phi^{-1}(B) \cap D_i = \phi_i^{-1}(B)$ for $i \in \{1,2\}$,

$$\mathbb{P}[(\det(X), \mathrm{tr}(X), X_{12}) \in B] = \mathbb{P}[X^\triangle \in \phi_1^{-1}(B)] + \mathbb{P}[X^\triangle \in \phi_2^{-1}(B)]. \tag{3.26}$$

For $i \in \{1,2\}$, we now use the transformation formula of Theorem 2.39 with ϕ_i^{-1} and the set B, inserting the determinant given in Eq. (3.25),

$$\mathbb{P}[X^{\triangle} \in \phi_i^{-1}(B)] = \int_{\phi_i^{-1}(B)} f_{X^{\triangle}}(a,b,c)\,\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}c = \int_B \frac{f_{X^{\triangle}}(\phi_i^{-1}(d,t,z))}{\sqrt{t^2-4(z^2+d)}}\,\mathrm{d}d\,\mathrm{d}t\,\mathrm{d}z$$
$$= \int_B \frac{\sqrt{2}\,\mathrm{e}^{-t}\,d^{\alpha-\frac{3}{2}}}{\Gamma_2(\alpha)\sqrt{t^2-4(z^2+d)}}\,\mathrm{d}d\,\mathrm{d}t\,\mathrm{d}z,$$

where the last equation follows from Eq. (3.22). Combining this with Eq. (3.26) and the fact that B was an arbitrary open set[5] of $\mathcal{B}(R)$, a density of $(\det(X), \operatorname{tr}(X), X_{12})$ is given by

$$f_{\det(X),\operatorname{tr}(X),X_{21}}(d,t,z) = 2\frac{\sqrt{2}\,\mathrm{e}^{-t}\,d^{\alpha-\frac{3}{2}}}{\Gamma_2(\alpha)\sqrt{t^2-4(z^2+d)}}$$

for every $(d,t,z) \in R$. The joint density of the trace and the determinant is then, due to the fact that $z^2 \leq t^2/4 - d$ if and only if $z \in [\sqrt{d-t^2/4}, \sqrt{t^2/4-d}]$ for $(d,t,z) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}$ with $t \geq 2\sqrt{d}$, given by

$$f_{\det(X),\operatorname{tr}(X)}(d,t) = \int_{\sqrt{d-t^2/4}}^{\sqrt{t^2/4-d}} 2\frac{\sqrt{2}\,\mathrm{e}^{-t}\,d^{\alpha-\frac{3}{2}}}{\Gamma_2(\alpha)\sqrt{t^2-4(z^2+d)}}\,\mathrm{d}z$$
$$= 2\frac{\sqrt{2}\,\mathrm{e}^{-t}\,d^{\alpha-\frac{3}{2}}}{\Gamma_2(\alpha)} 2 \underbrace{\int_0^{\sqrt{t^2/4-d}} \frac{1}{\sqrt{t^2-4(z^2+d)}}\,\mathrm{d}z}_{=:\mathcal{I}} \tag{3.27}$$

for all $(d,t) \in \mathbb{R}^+ \times \mathbb{R}^+$ with $t \geq 2\sqrt{d}$, where the second equality follows from the symmetry of the integrand. The integral $\mathcal{I}$ can now be computed using the substitution $y := 2z/\sqrt{t^2-4d}$,

$$\mathcal{I} = \int_0^{\sqrt{t^2/4-d}} \frac{1}{\sqrt{t^2-4d-(2z)^2)}}\,\mathrm{d}z = \int_0^1 \frac{1}{\sqrt{t^2-4d-(t^2-4d)y^2}} \frac{\sqrt{t^2-4d}}{2}\,\mathrm{d}y$$
$$= \frac{1}{2}\int_0^1 \frac{1}{\sqrt{1-y^2}}\,\mathrm{d}y = \frac{\arcsin(1)}{2} = \frac{\pi}{4},$$

which we can insert back into Eq. (3.27) to arrive at

$$f_{\det(X),\operatorname{tr}(X)}(d,t) = \frac{\sqrt{2}\pi}{\Gamma_2(\alpha)}\,\mathrm{e}^{-t}\,d^{\alpha-\frac{3}{2}}\,1_{\{t\geq 2\sqrt{d}\}}, \quad (d,t) \in \mathbb{R}^+ \times \mathbb{R}^+, \tag{3.28}$$

which could have been expected when looking at the density of X given in Eq. (3.21).

In Remark 3.29 and Corollary 3.36 below the marginal distributions of $\det X$ and $\operatorname{tr} X$ are discussed in a more general setting and can be used together with Eq. (3.28) to derive conditional densities for the determinant and trace.

The density function given in Eq. (3.28) is visualized in Figure 3.2 for $\alpha \in \{7/2, 9/2, 11/2\}$.

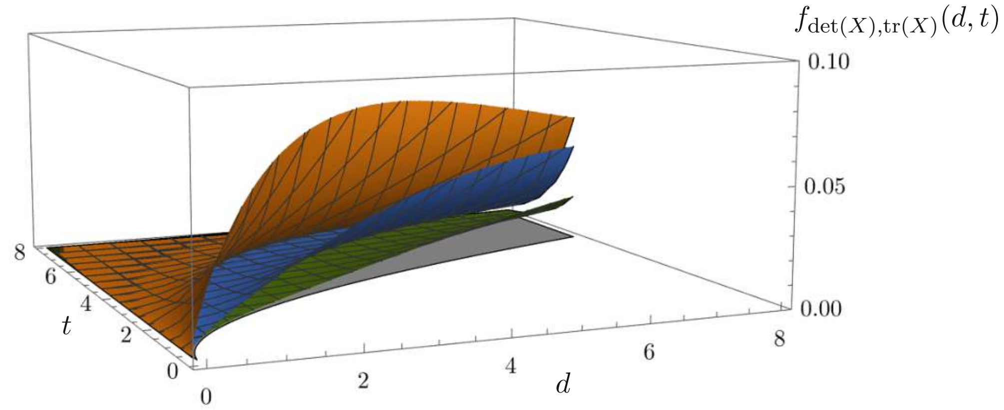


Figure 3.2. *Illustration of joint density functions of the determinant and the trace of* $X \sim \mathrm{MGamma}(\alpha, I_2)$ *for* $\alpha = 7/2$ *in orange,* $\alpha = 9/2$ *in blue and* $\alpha = 11/2$ *in green. These densities are greater than zero on the grey area* $\{(d,t) \in \mathbb{R}^+ \times \mathbb{R}^+ \mid t^2 > 4d\}$*. Elaborated in Example 3.21.*

The standard matrix-valued gamma distribution can also be characterized via the entries of its Cholesky decomposition.

Theorem 3.22 (Characterization of standard matrix-valued gamma distributions, see [33, Section 4.1(i), Property 1])**.** *Let* $p \in \mathbb{N}$ *and* $\alpha > (p-1)/2$*. Let* $\tilde{\mathbb{T}}_p$ *be as defined in Eq. (2.8). Then* X *is a* $\mathbb{S}_p^+$*-valued standard matrix-valued gamma distributed random variable with shape parameter* α[6] *if and only if the vector version* $\tilde{T} \in \tilde{\mathbb{T}}_p$ *of the lower triangular matrix* T*, that is uniquely defined by the Cholesky decomposition of* $X = TT^\mathsf{T}$*, is a random matrix, where the non-trivial entries* T_{ij} *of* T *with* $i \geq j$ *for* $i,j \in \{1,\dots,p\}$*, are independent and their distributions are given by:*

(i) $T_{ij} \sim \mathcal{N}(0, \frac{1}{2})$ *for all* $i > j$ *in* $\{1,\dots,p\}$*,*

(ii) $T_{ii} \sim \mathrm{GenGamma}(\alpha - \frac{i-1}{2}, 1, 2)$ *for all* $i \in \{1,\dots,p\}$*, see Definition 2.26.*

This leads to $T_{ii}^2 \sim \mathrm{Gamma}(\alpha - \frac{i-1}{2}, 1)$ *for all* $i \in \{1,\dots,p\}$ *.*

Proof. The proof of this theorem is split into five steps. Throughout the proof we access the entries of elements of $\tilde{\mathbb{T}}_p$ and of $\tilde{\mathbb{T}}_p$-valued $\tilde{T}$ by understanding them as the corresponding matrices in $\mathbb{T}_p$ using two indices.

Step 1. First we will derive a density function of $\tilde{T}$ that is equivalent to Items (i) and (ii) holding true: As all of the entries of $\tilde{T}$ are independent and given by the lower triangular entries of T, a probability density function of $\tilde{T}$ is given by

$$f_{\tilde{T}}(\tilde{t}) = \prod_{\substack{i,j=1\\ i\leq j}}^{p} f_{T_{ij}}(t_{ij}) = \Bigg(\prod_{\substack{i,j=1\\ i<j}}^{p} \frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-t_{ij}^2}\Bigg)\Bigg(\prod_{i=1}^{p} \frac{2}{\Gamma(\alpha - \frac{i-1}{2})}\,\mathrm{e}^{-t_{ii}^2}\,t_{ii}^{2\alpha-i}\Bigg), \quad \tilde{t} \in \tilde{\mathbb{T}}_p,$$

[5] Note that the open sets of a Borel σ-algebra form an intersection stable collection generating the Borel σ-algebra.

[6] This theorem can be slightly generalized for $X \sim \mathrm{MGamma}(\alpha, \beta I_p)$ with $\beta \in \mathbb{R}^+$, see Remark 3.28.

where we are using that the continuous density of a $\mathcal{N}(0, \frac{1}{2})$-distributed variable is given by $x \mapsto (1/\sqrt{\pi})\,\mathrm{e}^{-x^2}$ for $x \in \mathbb{R}$ and Eq. (2.14) for the diagonal entries following a generalized gamma distribution. Using Eq. (2.30), this density of $\tilde{T}$ can be rearranged to

$$\begin{aligned} f_{\tilde{T}}(\tilde{t}) &= \frac{2^{\frac{p(p-1)}{4}}}{\Gamma_p(\alpha)} \exp\Bigg(-\sum_{\substack{i,j=1\\ i\leq j}}^{p} t_{ij}^2\Bigg) \prod_{i=1}^{p} 2\,(t_{ii}^2)^{\alpha-\frac{i}{2}} \\ &= \frac{2^{\frac{p(p-1)}{4}}}{\Gamma_p(\alpha)}\, \mathrm{e}^{-\mathrm{tr}(tt^\mathsf{T})} (\det(tt^\mathsf{T}))^{\alpha-\frac{p+1}{2}}\, 2^p \prod_{i=1}^{p} t_{ii}^{p-i+1}, \quad \tilde{t} \in \tilde{\mathbb{T}}_p, \end{aligned} \tag{3.29}$$

where t is the matrix version of $\tilde{t}$ and where the second equality follows from $\det(tt^\mathsf{T}) = \prod_{i=1}^{p} t_{ii}^2$ and $\mathrm{tr}(tt^\mathsf{T}) = \sum_{i\leq j} t_{ij}^2$.

Step 2. Next, we will show that X is a random matrix if and only if $\tilde{T}$ is a random matrix: Lemma 2.14 states that the function that maps $T \in \mathbb{T}_p$ to TT^T is a homeomorphism. As discussed in Remark 3.17, $\mathbb{S}_p^+$-valued X is a random matrix if and only if $\tilde{\mathbb{S}}_p^+$-valued $X^\triangle := \psi^{-1}(X)$ is a random matrix, where ψ is defined in Eq. (2.7). Using the continuous and bijective function ϕ, given in Eq. (2.33) of Lemma 2.41, that maps every lower triangular matrix $\tilde{t} \in \tilde{\mathbb{T}}_p$ to the positive definite matrix $\psi^{-1}(tt^\mathsf{T}) \in \tilde{\mathbb{S}}_p^+$, we can see that

$$X^\triangle = \psi^{-1}(X) = \psi^{-1}(TT^\mathsf{T}) = \phi(\tilde{T})$$

is a $\tilde{\mathbb{S}}_p^+$-valued random matrix and hence X is a $\mathbb{S}_p^+$-valued random matrix if $\tilde{T}$ is a random matrix. Conversely, $\tilde{T} = \phi^{-1}(X^\triangle)$ is a random matrix if $X^\triangle$ is a random matrix.

Step 3. Assume $\tilde{T}$ is the random matrix described by Items (i) and (ii). For every open set B of $\mathcal{B}(\tilde{\mathbb{S}}_p^+)$,

$$\mathbb{P}[X^\triangle \in B] = \mathbb{P}[\phi(\tilde{T}) \in B] = \mathbb{P}[\tilde{T} \in \phi^{-1}(B)] = \int_{\phi^{-1}(B)} f_{\tilde{T}}(\tilde{t})\,\mathrm{d}\tilde{t}, \tag{3.30}$$

where we write $\mathrm{d}\tilde{t}$ for $\lambda^{\frac{p(p+1)}{2}}(\mathrm{d}\tilde{t})$. Inserting the determinant of ϕ given in Eq. (2.34) into Eq. (3.29),

$$f_{\tilde{T}}(\tilde{t}) = \underbrace{\frac{2^{\frac{p(p-1)}{4}}}{\Gamma_p(\alpha)}\, \mathrm{e}^{-\mathrm{tr}(tt^\mathsf{T})} (\det(tt^\mathsf{T}))^{\alpha-\frac{p+1}{2}}}_{=:\, g(\psi(\phi(\tilde{t})))} |\det \mathrm{d}\phi(\tilde{t})|, \quad \tilde{t} \in \tilde{\mathbb{T}}_p, \tag{3.31}$$

as $\psi(\phi(\tilde{t})) = tt^\mathsf{T}$, where t is the matrix version of $\tilde{t}$. Inserting this into Eq. (3.30),

$$\mathbb{P}[X^\triangle \in B] = \int_{\phi^{-1}(B)} g(\psi(\phi(\tilde{t})))|\det \mathrm{d}\phi(\tilde{t})|\,\mathrm{d}\tilde{t} = \int_B g(\psi(x^\triangle))\,\mathrm{d}x^\triangle, \tag{3.32}$$

where we are using the transformation formula from Theorem 2.39 with the function ϕ and the open set B for the second equality. As this is true for all open sets B a density of $X^\triangle$ is given by

$$f_{X^\triangle}(x^\triangle) = g(\psi(x^\triangle)), \quad x^\triangle \in \tilde{\mathbb{S}}_p^+.$$

Therefore, using Eq. (3.20),

$$f_X(x) = 2^{\frac{-p(p-1)}{4}} f_{X^\triangle}(x^\triangle) = 2^{\frac{-p(p-1)}{4}} g(x) = \frac{1}{\Gamma_p(\alpha)} \mathrm{e}^{-\mathrm{tr}(x)} (\det(x))^{\alpha - \frac{p+1}{2}}, \quad x \in \mathbb{S}_p^+,$$

where we are using the definition of $g(\cdot)$ in Eq. (3.31). As this is a density of the $\mathrm{MGamma}(\alpha, I_p)$ distribution, as given in Definition 3.18, this proves one implication of the theorem.

Step 4. Now assume X follows the standard matrix-valued gamma distribution and hence, using Eq. (3.20),

$$f_{X^\triangle}(x^\triangle) = \frac{2^{\frac{p(p-1)}{4}}}{\Gamma_p(\alpha)} \mathrm{e}^{-\mathrm{tr}(x)} (\det(x))^{\alpha - \frac{p+1}{2}} = g(\psi(x^\triangle)), \quad x^\triangle \in \tilde{\mathbb{S}}_p^+,$$

where $x = \psi(x^\triangle)$ and $g(\cdot)$ as in Eq. (3.31). Now as in the step above but in the other direction, we can see that for every open set $B \in \mathcal{B}(\tilde{\mathbb{T}}_p)$,

$$\mathbb{P}[\tilde{T} \in B] = \mathbb{P}[\phi^{-1}(X^\triangle) \in B] = \mathbb{P}[X^\triangle \in \phi(B)] = \int_{\phi(B)} g(\psi(x^\triangle))\,\mathrm{d}x^\triangle.$$

Furthermore, with the transformation of the second equality of Eq. (3.32) with the set $\phi(B)$ instead of the B there and the definition of $g(\cdot)$ in Eq. (3.31)

$$\mathbb{P}[\tilde{T} \in B] = \int_B g(\psi(\phi(\tilde{t})))|\det \mathrm{d}\phi(\tilde{t})|\,\mathrm{d}\tilde{t} = \int_B \frac{2^{\frac{p(p-1)}{4}}}{\Gamma_p(\alpha)} \mathrm{e}^{-\mathrm{tr}(tt^\intercal)} (\det(tt^\intercal))^{\alpha - \frac{p+1}{2}} 2^p \prod_{i=1}^p t_{ii}^{p-i+1}\,\mathrm{d}\tilde{t},$$

which, together with Eq. (3.29) and the first step of this proof proves that $\tilde{T}$ follows the desired distribution.

Step 5. Finally, we will show that Item (ii) leads to the entries T_{ii}^2 being gamma distributed for every $i \in \{1, \dots, p\}$. For the function $\xi(x) \coloneqq x^2$ for $x \in \mathbb{R}^+$, and every Borel set B of $\mathcal{B}(\mathbb{R}^+)$,

$$\mathbb{P}[T_{ii}^2 \in B] = \mathbb{P}[T_{ii} \in \xi(B)] = \int_{\xi(B)} f_{T_{ii}}(x)\,\mathrm{d}x = \frac{2}{\Gamma(\alpha - \frac{i-1}{2})} \int_{\xi(B)} \mathrm{e}^{-x^2} (x^2)^{\alpha - \frac{i}{2}}\,\mathrm{d}x,$$

where Eq. (2.14) is used for the last equality. By substituting $y \coloneqq \xi(x)$ this is equal to

$$\mathbb{P}[T_{ii}^2 \in B] = \frac{2}{\Gamma(\alpha - \frac{i-1}{2})} \int_B \mathrm{e}^{-y} y^{\alpha - \frac{i}{2}} \frac{1}{2\sqrt{y}}\,\mathrm{d}y.$$

Since this is true for every Borel set B, a density of T_{ii}^2 is given by

$$f_{T_{ii}^2}(y) = \frac{1}{\Gamma(\alpha - \frac{i-1}{2})} \mathrm{e}^{-y} y^{\alpha - \frac{i+1}{2}}, \quad y \in \mathbb{R}^+,$$

which is a density of the one-dimensional gamma distribution with shape parameter $\alpha - i - 1/2$ and scale parameter $\beta = 1$, see Eq. (2.15). □

Corollary 3.23 (Independent entries of standard matrix-valued gamma distributions)**.** *For $p \in \mathbb{N}$ and $\alpha > (p-1)/2$, let $X \sim \mathrm{MGamma}(\alpha, I_p)$. Then the elements of every collection $X_{i_1,j_1}, \ldots, X_{i_n,j_n}$ with $n \in \mathbb{N}$ of entries of X, where $\{i_k, j_k\} \cap \{i_l, j_l\} = \varnothing$ for all $k, l \in \{1, \ldots, n\}$, are independent. This implies, in particular, that all diagonal elements of a standard matrix-valued gamma distribution are independent.*

Proof. Using the previous theorem, let T denote the $\mathbb{T}_p$-valued matrix satisfying $TT^\mathsf{T} = X$, where the non-trivial entries are independent. Then

$$X_{i_k,j_k} = \sum_{m=1}^{j} T_{i_k m} T_{j_k m}, \quad k \in \{1, \ldots, n\},$$

hence, if the conditions of this corollary are satisfied, $X_{i_1,j_1}, \ldots, X_{i_n,j_n}$ are independent. □

Definition 3.24 (Matrix-valued gamma distributions)**.** For $p \in \mathbb{N}$ an $\mathbb{R}^{p\times p}$-valued random matrix Y is said to be matrix-gamma distributed with shape parameter $\alpha \in \mathbb{R}^+$ and scale parameter matrix $C \in \mathbb{R}^{p\times p}$, if there exists a dimension $q \in \mathbb{N}$, a matrix $A \in \mathbb{R}^{p\times q}$, such that $C = AA^\mathsf{T}$, and a random $\mathbb{S}_q^+$-valued matrix X following the standard matrix-valued gamma distribution with parameter $\alpha > (q-1)/2$, as given in Definition 3.18, such that

$$Y \overset{\mathrm{d}}{=} AXA^\mathsf{T} \quad \text{(equality in distribution).} \tag{3.33}$$

We use the notation $Y \sim \mathrm{MGamma}(\alpha, C)$.

Remark 3.25 (One-dimensional gamma distributions are a subclass of matrix-valued gamma distributions). Setting $p = 1$ and assuming C is positive, we can see that matrix-valued gamma distributions are in fact a generalization of one-dimensional gamma distributions. The density function given in Eq. (3.21) simplifies to the density of the $\mathrm{Gamma}(\alpha, 1)$ distribution, see Eq. (2.15). In the one-dimensional case Eq. (3.33) is equivalent to multiplying the standard gamma distributed variable with the scale parameter $C \in \mathbb{R}^+$, leading to the $\mathrm{Gamma}(\alpha, C)$ distribution.

Remark 3.26 (Sampling from matrix-valued gamma distributions). Combining Theorem 3.22 with the definition of matrix-valued gamma distributions given in Eq. (3.33) provides a method to sample from $\mathrm{MGamma}(\alpha, C)$.

Remark 3.27 (Existence of $\mathrm{MGamma}(\alpha, C)$). The $\mathrm{MGamma}(\alpha, C)$ distribution exists for every positive semi-definite scale parameter matrix C and shape parameter $\alpha > (\mathrm{rk}(C)-1)/2$: In an analogue argumentation to Lemma 3.9 we can see that for every positive semi-definite $C \neq 0$ there exists a matrix $A \in \mathbb{R}^{p\times \mathrm{rk}(C)}$ such that $AA^\mathsf{T} = C$. For $C = 0$, we can chose A to be the p-dimensional zero vector. Hence, the $\mathrm{MGamma}(\alpha, C)$ distribution can be defined for every $\alpha > (\mathrm{rk}(C) - 1)/2$.

The uniqueness of the distribution given C and α follows from the characteristic function in Theorem 3.32 below.

Remark 3.28. The characterization of the standard matrix-valued gamma distribution given in Theorem 3.22 via its Cholesky decomposition can be generalized for scale parameter matrices of the form βI_p for $\beta \in \mathbb{R}^+$:

For $p \in \mathbb{N}$ let $\alpha > (p-1)/2$ and $\beta > 0$. Let $\tilde{\mathbb{T}}_p$ be as defined in Eq. (2.8). Then $Y \sim \mathrm{MGamma}(\alpha, \beta I_p)$ if and only if the vector version $\tilde{T} \in \tilde{\mathbb{T}}_p$ of the lower triangular matrix T, that is uniquely defined by the Cholesky decomposition of $Y = TT^\mathsf{T}$, is a random matrix, where the non-trivial entries T_{ij} of T with $i \geq j$ for $i, j \in \{1, \dots, p\}$, are independent and their distributions are given by:

(i) $T_{ij} \sim \mathcal{N}(0, \frac{\beta}{2})$ for all $i > j$ in $\{1, \dots, p\}$,

(ii) $T_{ii} \sim \mathrm{GenGamma}(\alpha - \frac{i-1}{2}, \beta, 2)$ for all $i \in \{1, \dots, p\}$.

This leads to $T_{ii}^2 \sim \mathrm{Gamma}(\alpha - \frac{i-1}{2}, \beta)$ for all $i \in \{1, \dots, p\}$.

Remark 3.29 (Distribution of the determinant of some matrix-valued gamma distributions, see [13, Theorem 3.3.22(i)]). For $p \in \mathbb{N}$ let $\alpha > (p-1)/2$ and let positive semi-definite C be of the form $C = \beta I_p$ for a $\beta \geq 0$. Using the corresponding $\mathbb{T}_p$-valued T from Remark 3.28, we can see that

$$\det Y = \det(TT^\mathsf{T}) = \prod_{i=1}^{p} T_{ii}^2$$

where $T_{ii} \sim \mathrm{Gamma}(\alpha - (i-1)/2, \beta)$ for every $i \in \{1, \dots, p\}$.

Hence, the distribution of the determinant of Y is the product distribution of p independent gamma distributions with the same scale parameter β. For $p = 2$ and $\beta = 1$, a joint density of $\det(X)$ and $\mathrm{tr}(X)$ is given in Example 3.21.

Remark 3.30 (The family of matrix-valued gamma distribution takes values in the positive semi-definite matrices). For $p \in \mathbb{N}$, let $C \in \mathbb{R}^{p \times p}$ be positive semi-definite and let $\alpha > (\mathrm{rk}(C) - 1)/2$. Since X following a standard matrix-valued gamma distributions only takes positive definite values, the values of the $\mathrm{MGamma}(\alpha, C)$ distributed random matrix $Y \overset{\mathrm{d}}{=} AXA^\mathsf{T}$ are positive semi-definite a.s., see Remark 3.11. In the case that A has rank p, or equivalently C has full rank, which implies that C is positive definite due to Item (vi) of Lemma 2.8, the values of Y are even positive definite a.s.

Remark 3.31. Matrix-valued gamma distributions and Wishart distributions in Definition 3.8 are commonly only defined for positive definite scale matrices/ covariance matrices, as for example in [13, Definition 3.6.1] or [20, Eq. 2.1]. In this case the matrix-valued gamma distributions are often defined via the densities given in Theorem 3.45 below.

Theorem 3.32 (Joint Laplace transform and characteristic function of matrix-valued gamma distributions, see [13, Theorem 3.3.7])**.** *Let $p, q \in \mathbb{N}$, $\alpha > (q-1)/2$, and let $C \in \mathbb{R}^{p \times p}$ be given as matrix product $C = AA^\mathsf{T}$, with $A \in \mathbb{R}^{p \times q}$. Let $Y \sim \mathrm{MGamma}(\alpha, C)$ as in Definition 3.24 have the stochastic representation $Y \overset{\mathrm{d}}{=} AXA^\mathsf{T}$ with $X \sim \mathrm{MGamma}(\alpha, I_q)$.*

(i) For $w \in \mathbb{C}^{p \times p}$, let $\tilde{w} := (w + w^\mathsf{T})/2$ with representation $\tilde{w} = u + \mathrm{i}v$ with $u, v \in \mathbb{R}^{p \times p}$ symmetric. If[7] $I_q + A^\mathsf{T} u A \in \mathbb{S}_q^+$, then the determinant of $I_p + 2\tilde{w}C$ is different from

[7] The condition on $w \in \mathbb{C}^{p \times p}$ is in particular satisfied if the real part of w, or even of $\tilde{w}$, is positive semi-definite, in particular if it is equal to the zero matrix.

zero and for every constant scalar $\beta > (p-1)/2 - \alpha$ and every constant matrix $M \in \mathbb{R}^{p\times p}$,

$$\mathbb{E}\left[(\det MY)^{\beta}\,\mathrm{e}^{-\mathrm{tr}(wY)}\right] = \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)}(\det MC)^{\beta}\big(\det(I_p+\tilde{w}C)\big)^{-(\alpha+\beta)}. \tag{3.34}$$

(ii) The characteristic function of Y, as defined in Eq. (3.5), is given by

$$\varphi_Y(z) = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)}] = \big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha} \tag{3.35}$$

for all symmetric $z \in \mathbb{R}^{p\times p}$. It is sufficient to consider symmetric z due to Remark 3.13, considering Y takes values in the symmetric matrices as discussed in Remark 3.30.

Proof. In the following we will prove Item (i). Item (ii) follows from Item (i) by setting $M = I_p$, $\beta = 0$, $w = \mathrm{i}z$ for $z \in \mathbb{R}^{p\times p}$ symmetric.

For Item (i), we first consider the case of a standard matrix-valued gamma distributed random variable, denoted by X, a real $\beta > (p-1)/2-\alpha$ and a matrix w with corresponding symmetric matrix $\tilde{w} = u+\mathrm{i}v \in \mathbb{C}^{p\times p}$ with u satisfying $I_p+u \in \mathbb{S}_p^+$. With the same argument as in Remark 3.13, $\mathrm{tr}(wY) = \mathrm{tr}(\tilde{w}Y)$ a.s. Hence, using the density of the distribution given in Eq. (3.21),

$$\begin{aligned}\mathbb{E}\left[(\det MX)^{\beta}\,\mathrm{e}^{-\mathrm{tr}(wX)}\right] &= \int_{\mathbb{S}_p^+}(\det Mx)^{\beta}\,\mathrm{e}^{-\mathrm{tr}(\tilde{w}x)}\,\frac{1}{\Gamma_p(\alpha)}\,\mathrm{e}^{-\mathrm{tr}(x)}(\det x)^{\alpha-\frac{p+1}{2}}\,\mathrm{d}x\\ &= \frac{(\det M)^{\beta}}{\Gamma_p(\alpha)}\int_{\mathbb{S}_p^+}\mathrm{e}^{-\mathrm{tr}(((I_p+u)+\mathrm{i}v)x)}(\det x)^{\alpha+\beta-\frac{p+1}{2}}\,\mathrm{d}x.\end{aligned}$$

As I_p+u is positive definite, $v \in \mathbb{R}^{p\times p}$ is symmetric and $\alpha+\beta > (p-1)/2$, we can apply Eq. (2.31) from Theorem 2.37, telling us that $\det(I_p+\tilde{w}) \neq 0$ and leading to

$$\begin{aligned}\mathbb{E}\left[(\det MX)^{\beta}\,\mathrm{e}^{-\mathrm{tr}(wX)}\right] &= \frac{(\det M)^{\beta}}{\Gamma_p(\alpha)}\Gamma_p(\alpha+\beta)\big(\det(I_p+\tilde{w})\big)^{-(\alpha+\beta)}\\ &= \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)}(\det M)^{\beta}\big(\det(I_p+\tilde{w})\big)^{-(\alpha+\beta)},\end{aligned} \tag{3.36}$$

which proves the lemma for the standard case.

For $Y \sim \mathrm{MGamma}(\alpha, C)$, using the stochastic representation $Y \overset{\mathrm{d}}{=} AXA^{\mathsf{T}}$, where X follows the standard matrix-valued gamma distribution of dimension $q \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}[(\det MY)^{\beta}\,\mathrm{e}^{-\mathrm{tr}(wY)}] &= \mathbb{E}[(\det MAXA^{\mathsf{T}})^{\beta}\,\mathrm{e}^{-\mathrm{tr}(\tilde{w}AXA^{\mathsf{T}})}]\\ &= (\det \underbrace{AA^{\mathsf{T}}}_{=C})^{\beta}\,\mathbb{E}[(\det MX)^{\beta}\,\mathrm{e}^{-\mathrm{tr}((A^{\mathsf{T}}\tilde{w}A)X)}],\end{aligned}$$

where we are using Lemma 2.1 for the second equation. As $A^{\mathsf{T}}\tilde{w}A$ is symmetric for symmetric $\tilde{w}$ and as $I_q + \mathrm{Re}(A^{\mathsf{T}}\tilde{w}A) = I_q + A^{\mathsf{T}}uA$ is positive definite, we can insert the standard case derived in Eq. (3.36),

$$\begin{aligned}\mathbb{E}[(\det MY)^{\beta}\,\mathrm{e}^{-\mathrm{tr}(wY)}] &= (\det C)^{\beta}\frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)}(\det M)^{\beta}\big(\det(I_q+A^{\mathsf{T}}\tilde{w}A)\big)^{-(\alpha+\beta)}\\ &= \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)}(\det MC)^{\beta}\big(\det(I_p+\tilde{w}\underbrace{AA^{\mathsf{T}}}_{=C})\big)^{-(\alpha+\beta)},\end{aligned}$$

where we are using the Weinstein–Aronszajn identity given in Lemma 2.3 for the second equality. □

For $p \in \mathbb{N}$ the characteristic function of a Wishart distribution with parameters $n \in \mathbb{N}$ and positive semi-definite $\Sigma \in \mathbb{R}^{p\times p}$ satisfying $n > \mathrm{rk}(\Sigma)$, given in Eq. (3.7) of Theorem 3.14, has the same form as the characteristic function of a $\mathrm{MGamma}(n/2, 2\Sigma)$ distribution given via Eq. (3.35).

Corollary 3.33 (Matrix-valued gamma distributions are a generalization of non-singular Wishart distributions)**.** *For $n, p \in \mathbb{N}$ let the positive semi-definite matrix $\Sigma \in \mathbb{R}^{p\times p}$ satisfy $n > \mathrm{rk}(\Sigma)$. Then*

$$\mathrm{Wishart}(n, \Sigma) = \mathrm{MGamma}(n/2, 2\Sigma). \tag{3.37}$$

This corollary and the similar form of the characteristic functions leads to the definition of a more general version of matrix-valued gamma distributions incorporating the singular Wishart distributions.

Definition 3.34 (General matrix-valued gamma distributions)**.** For $p \in \mathbb{N}$ and positive semi-definite $C \in \mathbb{R}^{p\times p}$, let

$$\mathcal{G}_C := \left\{0, \frac{1}{2}, 1, \frac{3}{2}, \dots, \frac{\mathrm{rk}(C)-2}{2}\right\} \cup \left[\frac{(\mathrm{rk}(C)-1)^+}{2}, \infty\right). \tag{3.38}$$

The general matrix-valued gamma distribution family is now defined for every $\alpha \in \mathcal{G}_C$, hence combining Definitions 3.8 and 3.24. An $\mathbb{R}^{p\times p}$-valued random matrix Y following this distribution satisfies

(i) $Y \sim \mathrm{MGamma}(\alpha, C)$, if $\alpha > \frac{(\mathrm{rk}(C)-1)^+}{2}$,

(ii) $Y \sim \mathrm{Wishart}(2\alpha, \frac{1}{2}C)$, if $\alpha \in \left\{\frac{1}{2}, 1, \frac{3}{2}, 2, \dots, \frac{\mathrm{rk}(C)-1}{2}\right\}$, which is a singular Wishart distribution,

(iii) Y follows the degenerate distribution concentrated in $0 \in \mathbb{R}^{p\times p}$ if $\alpha = 0$.

This will also be denoted by $Y \sim \mathrm{MGamma}(\alpha, C)$.

Corollary 3.35 (Characteristic function of general matrix-valued gamma distributions)**.** *For $p \in \mathbb{N}$, let $Y \sim \mathrm{MGamma}(\alpha, C)$ with $C \in \mathbb{R}^{p\times p}$ positive semi-definite and $\alpha \in \mathcal{G}_C$. Then the characteristic function is given by*

$$\varphi_Y(z) = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)}] = \big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha} \tag{3.39}$$

for all symmetric $z \in \mathbb{R}^{p\times p}$.

Corollary 3.36 (Distribution of the trace of matrix-valued gamma distributions, see [13, Theorem 3.3.23])**.** *For $p \in \mathbb{N}$ let $C \in \mathbb{R}^{p\times p}$ be a positive semi-definite scale parameter matrix and let $\alpha \in \mathcal{G}_C$. Then the trace of $Y \sim \mathrm{MGamma}(\alpha, C)$ is a $\mathbb{R}_0^+$-valued random variable with characteristic function*

$$\varphi_{\mathrm{tr}(Y)}(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}t\mathrm{tr}(Y)}] = \big(\det(I_p - \mathrm{i}tC)\big)^{-\alpha}, \quad t \in \mathbb{R}.$$

If $C = \mathrm{diag}(\beta_1, \dots, \beta_p)$, then $\mathrm{tr}(Y)$ follows the same distribution as the sum of p independent random variables[8] *$Y_1, \dots, Y_p$ with $Y_i \sim \mathrm{Gamma}(\alpha, \beta_i)$ for $i \in \{1, \dots, p\}$. Note that C*

[8] This distribution corresponds to a finite gamma convolution, see [27].

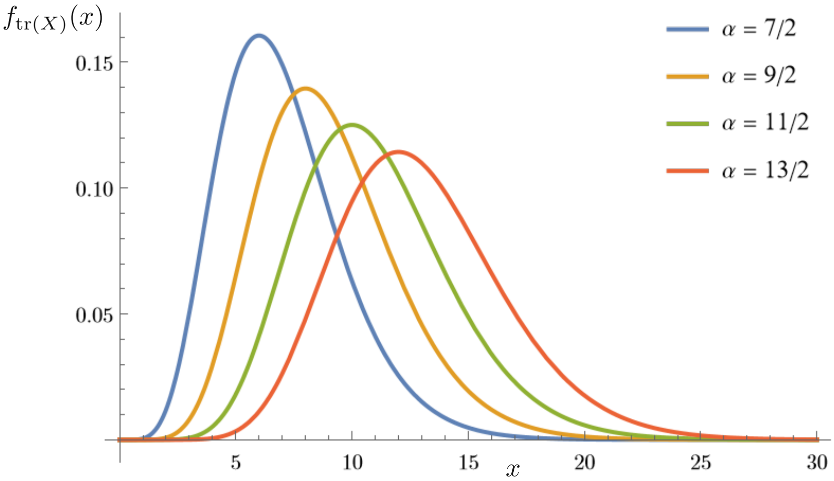


Figure 3.3. *Continuous probability density function of the trace of a standard matrix gamma distribution, see Corollary 3.36, with $p = 3$ and different α.*

positive semi-definite implies $\beta_1, \dots, \beta_p \in \mathbb{R}_0^+$. If C is even a multiple of the identity matrix, i.e. $C = \beta I_p$, then $\mathrm{tr}(Y) \sim \mathrm{Gamma}(\alpha p, \beta)$.

This is illustrated in the standard case $\beta = 1$ in Figure 3.3. For $p = 2$ and $C = I_2$, a joint density of $\det(X)$ and $\mathrm{tr}(X)$ is given in Example 3.21.

Proof. In case $\alpha = 0$, all of statements are true since $Y = 0$ in this case. For the rest of the proof let $\alpha \neq 0$.

The characteristic function of $\mathrm{tr}(Y)$ is given by

$$\varphi_{\mathrm{tr}(Y)}(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}t\mathrm{tr}(Y)}] = \mathbb{E}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}(tI_p)Y)}] = \varphi_Y(tI_p), \quad t \in \mathbb{R}.$$

Inserting the characteristic function of Y given in Corollary 3.35 proves the first statement of the corollary.

If C is a diagonal matrix, then it can be represented by $C = \mathrm{diag}(\beta_1, \dots, \beta_p)$ with $\beta_1, \dots, \beta_p \in \mathbb{R}_0^+$ as C is positive semi-definite. Then

$$\varphi_{\mathrm{tr}(Y)}(t) = \big(\det(I_p - \mathrm{i}tC)\big)^{-\alpha} = \left(\prod_{j=1}^{p}(1-\mathrm{i}\beta_j)\right)^{-\alpha} = \exp\left(-\alpha \,\mathrm{Log}\left(\prod_{j=1}^{p}(1-\mathrm{i}t\beta_j)\right)\right), \quad t \in \mathbb{R},$$

where we use the definition of complex exponentiation given in Eq. (2.18). Now, due to Eq. (2.17) there exists a $k \in \mathbb{Z}$ such that

$$\mathrm{Log}\left(\prod_{j=1}^{p}(1-\mathrm{i}t\beta_j)\right) = 2\pi\mathrm{i}k + \sum_{j=1}^{p}\mathrm{Log}(1-\mathrm{i}t\beta_j), \quad t \in \mathbb{R}.$$

Hence, for every $t \in \mathbb{R}$,

$$\varphi_{\mathrm{tr}(Y)}(t) = \exp\left(-\alpha\left(2\pi\mathrm{i}k + \sum_{j=1}^{p}\mathrm{Log}(1-\mathrm{i}t\beta_j)\right)\right) = \mathrm{e}^{-2\pi\mathrm{i}k\alpha}\prod_{j=1}^{p}(1-\mathrm{i}t\beta_j)^{-\alpha}.$$

As $\varphi_{\mathrm{tr}(Y)}(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}t\mathrm{tr}(Y)}]$ is continuous in t, this $k \in \mathbb{Z}$ must be the same for all $t \in \mathbb{R}$ as $\alpha \neq 0$. Since $\varphi_{\mathrm{tr}(Y)}(0) = 1$, we can conclude that $k = 0$ and therefore

$$\varphi_{\mathrm{tr}(Y)}(t) = \prod_{j=1}^{p} (1 - \mathrm{i}t\beta_j)^{-\alpha}, \quad t \in \mathbb{R}.$$

Since $\mathbb{R} \ni t \mapsto (1 - \mathrm{i}t\beta_j)^{-\alpha}$ is the characteristic function of the $\mathrm{Gamma}(\alpha, \beta_j)$ distribution due to Corollary 2.34, this proves the second statement of the corollary.

Finally, if $C = \beta I_p$ for some $\beta \in \mathbb{R}^+$, then

$$\varphi_{\mathrm{tr}(Y)}(t) = \prod_{j=1}^{p} (1 - \mathrm{i}t\beta)^{-\alpha} = \big((1 - \mathrm{i}t\beta)^{-\alpha}\big)^p = (1 - \mathrm{i}t\beta)^{-\alpha p}, \quad t \in \mathbb{R},$$

which is the characteristic function of the $\mathrm{Gamma}(\alpha p, \beta)$ distribution, where the last equality holds due to Item (i). □

Remark 3.37 (A further extension of the definition of matrix-valued gamma distributions using the characteristic function is not possible). For $p \in \mathbb{N}$ and $C \in \mathbb{R}^{p \times p}$, the general matrix-valued gamma distribution has been defined for every $\alpha \in \mathcal{G}_C$. The characteristic function given in Eq. (3.39) might lead to the question whether a function of this form also describes a probability distribution for α not fulfilling this criterion. In fact it has been proven in [25] that only for $\alpha \in \mathcal{G}_C$ the right side of Eq. (3.39) describes a characteristic function of a probability distribution.

Remark 3.38 (Matrix-valued gamma distributions are not infinitely divisible). The characteristic function of a matrix-valued gamma distribution exists if and only if $\alpha \in \mathcal{G}_C$ due to Remark 3.37. So for a counterexample to infinite divisibility of matrix-valued gamma distributions let C be positive semi-definite, let $\alpha \in \mathcal{G}_C$ and let $n = 2$ be the factor we want to divide the distribution by. If the distribution were infinitely divisible, then there should exist a characteristic function f such that the characteristic function of $Y \sim \mathrm{MGamma}(\alpha, C)$ can be written as

$$\varphi_Y(z) = \big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha} = (f(z))^2.$$

The two possibilities for $f(z)$ are

$$f_1(z) = \big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha/2} \quad \text{and} \quad f_2(z) = -\big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha/2},$$

where $f_1(z)$ is not a characteristic function for all α as $\alpha \in \mathcal{G}_C$ does not imply $\alpha/2 \in \mathcal{G}_C$, which would be necessary due to Remark 3.37. The second option $f_2(z)$ is not a characteristic function since $f_2(0) = -1 \neq 1$.

The family of general matrix-valued gamma distributions with the same shape parameter α has the following property.

Corollary 3.39 (Scaling property of matrix-valued gamma distributions)**.** *For $p, q \in \mathbb{N}$, let $Y \sim \mathrm{MGamma}(\alpha, C)$ with $C \in \mathbb{R}^{p \times p}$ positive semi-definite and $\alpha \in \mathcal{G}_C$ and let L be a constant matrix in $\mathbb{R}^{q \times p}$. Then the $\mathbb{R}^{q \times q}$-valued random matrix $LYL^\top$ follows the $\mathrm{MGamma}(\alpha, LCL^\top)$ distribution.*

Proof. This proof is done via the characteristic function of LYL^T. For every symmetric $z \in \mathbb{R}^{q\times q}$

$$\varphi_{LYL^\mathsf{T}}(z) = \mathbb{E}[\mathrm{e}^{\operatorname{tr}(\mathrm{i}zLYL^\mathsf{T})}] = \mathbb{E}[\mathrm{e}^{\operatorname{tr}(\mathrm{i}L^\mathsf{T}zLY)}] = \varphi_Y(L^\mathsf{T}zL),$$

where Lemma 2.1 is used for the second equation. Inserting the characteristic function given in Eq. (3.39),

$$\varphi_{LYL^\mathsf{T}}(z) = \varphi_Y(L^\mathsf{T}zL) = \big(\det(I_p - \mathrm{i}L^\mathsf{T}zLC)\big)^{-\alpha} = \big(\det(I_q - \mathrm{i}zLCL^\mathsf{T})\big)^{-\alpha}$$

for all symmetric $z \in \mathbb{R}^{q\times q}$, where the last equation follows from the Weinstein–Aronszajn identity given in Lemma 2.3. Since $\operatorname{rk}(LCL^\mathsf{T}) \leq \operatorname{rk}(C)$ and hence $\alpha \in \mathcal{G}_{LCL^\mathsf{T}}$, this is the characteristic function of $\operatorname{MGamma}(\alpha, LCL^\mathsf{T})$ at z. As discussed in Remark 3.4 and Remark 3.13 the characteristic function of the symmetric random matrix LYL^T at symmetric $z \in \mathbb{R}^{q\times q}$ determines its distribution uniquely, hence $LYL^\mathsf{T} \sim \operatorname{MGamma}(\alpha, LCL^\mathsf{T})$. □

Each permutation π of $\{1,\dots,p\}$ can be identified with its permutation matrix $P_\pi := (\delta_{i,\pi(j)})_{i,j=1}^p$, where $\delta_{i,j}$ is the Kronecker delta. Hence, $P_\pi Y P_\pi^\mathsf{T} = (Y_{\pi^{-1}(i),\pi^{-1}(j)})_{i,j=1}^p$ for every matrix $Y \in \mathbb{R}^{p\times p}$.

Corollary 3.40 (Permutation property of matrix-valued gamma distributions)**.** *For $p \in \mathbb{N}$, $C \in \mathbb{R}^{p\times p}$ positive semi-definite and $\alpha \in \mathcal{G}_C$, let $Y \sim \operatorname{MGamma}(\alpha, C)$ and let π be a permutation of $\{1,\dots,p\}$. Then $P_\pi Y P_\pi^\mathsf{T}$ follows the $\operatorname{MGamma}(\alpha, P_\pi C P_\pi^\mathsf{T})$ distribution, where P_π refers to the permutation matrix corresponding to π.*

Remark 3.41 (Matrix-valued gamma distributions cannot be easily characterized by derived one-dimensional distributions)*.* Multidimensional normal distributions can be characterized by derived one-dimensional distributions, see [28, Lemma 2.33]: an $\mathbb{R}^n$-valued random vector X follows an n-dimensional normal distribution if and only if $\langle t, X\rangle$ follows a one-dimensional normal distribution for all $t \in \mathbb{R}^n$.

Corollary 3.39 shows that one implication of this characterization holds for matrix-valued gamma distributions as well: For $p \in \mathbb{N}$ let $C \in \mathbb{R}^{p\times p}$ be positive semi-definite and $\alpha \in \mathcal{G}_C$. For $Y \sim \operatorname{MGamma}(\alpha, C)$ the random variable $\langle t, Yt\rangle$ follows a one-dimensional gamma distribution for all $t \in \mathbb{R}^p$. The other implication, however, does not hold true. For $Y_0 \sim \operatorname{Gamma}(\alpha_0, \beta_0)$ with shape parameter $\alpha_0 > 0$ and scale parameter $\beta_0 > 0$, the random matrix

$$Y := \begin{pmatrix} Y_0 & 0 \\ 0 & Y_0 \end{pmatrix}$$

satisfies the condition, that for every $t \in \mathbb{R}^2$

$$\langle t, Yt\rangle = \|t\|_2^2 Y_0$$

follows the one-dimensional gamma distribution with shape parameter α_0 and scale parameter $\|t\|_2^2\beta_0$. Using the expectations and variances of the entries of a matrix-valued gamma distribution, that will be proven in Theorem 3.59, we can see that the random matrix Y does not follow a two-dimensional gamma distribution: The expectation $\mathbb{E}[Y] \overset{!}{=} \alpha C$ implies that a possible scale parameter matrix $C \in \mathbb{R}^{2\times 2}$ must have entries $c_{12} = c_{21} = 0$ and $c_{11} \neq 0$, $c_{22} \neq 0$, which contradicts $\operatorname{Var}[Y_{12}] = 0 \overset{!}{=} \alpha/2\,(c_{11}c_{22} + c_{12}^2)$ for any $\alpha > (p-1)/2$.

With the help of the characteristic function, we can prove that the summation property of one-dimensional gamma distributions also holds for their matrix-valued version.

Lemma 3.42 (Summation property of matrix-valued gamma distributions)**.** *For $n \in \mathbb{N}$, let $Y_k \sim \mathrm{MGamma}(\alpha_k, C)$ for $k \in \{1, \dots, n\}$ be independent random matrices of dimension $p \times p$, where C is a positive semi-definite scale matrix and $\alpha_1, \dots, \alpha_n \in \mathcal{G}_C$. Then*

$$\sum_{k=1}^{n} Y_k \sim \mathrm{MGamma}(\alpha_1 + \dots + \alpha_n, C).$$

Proof. For a proof by induction, it suffices to consider the essential case $n = 2$. As Y_1 and Y_2 are independent and their characteristic function is given in Eq. (3.39),

$$\begin{aligned}\varphi_{Y_1+Y_2}(z) = \varphi_{Y_1}(z)\,\varphi_{Y_2}(z) &= \big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha_1}\big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha_2} \\ &= \big(\det(I_p - \mathrm{i}zC)\big)^{-(\alpha_1+\alpha_2)},\end{aligned}$$

for every symmetric $z \in \mathbb{R}^{p \times p}$, where the last equality holds due to Item (iii) of Proposition 2.32. As discussed in Remark 3.4 and Remark 3.13, the characteristic function at symmetric matrices determines the distribution of a symmetric random matrix uniquely, hence $Y_1 + Y_2 \sim \mathrm{MGamma}(\alpha_1 + \alpha_2, C)$. □

Remark 3.43 (The matrix-valued gamma distributions with a given C form a convolution semigroup)*.* For a fixed positive semi-definite scale parameter matrix C, the matrix-valued gamma distributions with any shape parameter $\alpha \in \mathcal{G}_C$ and scale parameter matrix C form a convolution semigroup due to Lemma 3.42. With $\mathrm{MGamma}(0, C)$ as a neutral element this set of distributions is even a monoid.

Corollary 3.44 (Combined scaling and summation property of matrix-valued gamma distributions)**.** *For $n, q \in \mathbb{N}$ and $p_k \in \mathbb{N}$ for $k \in \{1, \dots, n\}$, let $L_k \in \mathbb{R}^{q \times p_k}$ and let $Y_k \sim \mathrm{MGamma}(\alpha_k, C_k)$ for $k \in \{1, \dots, n\}$ be independent random matrices of dimension $p_k \times p_k$, where C_k is positive semi-definite and $\alpha_k \in \mathcal{G}_{C_k}$ for all $k \in \{1, \dots, n\}$. If there exists a positive semi-definite matrix $C \in \mathbb{R}^{q \times q}$ such that $L_k C_k L_k^\top = C$ for all $k \in \{1, \dots, n\}$, then*

$$\sum_{k=1}^{n} L_k Y_k L_k^\top \sim \mathrm{MGamma}(\alpha_1 + \ldots + \alpha_n, C).$$

Proof. Combine Lemma 3.42 with Corollary 3.39. □

Theorem 3.45 (Density of matrix-valued gamma distributions)**.** *For $p \in \mathbb{N}$, $C \in \mathbb{S}_p^+$ and $\alpha > (p-1)/2$, let $Y \sim \mathrm{MGamma}(\alpha, C)$. Then a density of Y w.r.t. the measure μ on $\mathbb{S}_p^+$, discussed in Remark 3.17, is given by*

$$f_Y(y) = \frac{1}{(\det C)^{\alpha}\Gamma_p(\alpha)}\,\mathrm{e}^{-\mathrm{tr}(C^{-1}y)}(\det y)^{\alpha - \frac{p+1}{2}}, \quad y \in \mathbb{S}_p^+. \tag{3.40}$$

Proof. We are going to prove this theorem by using the uniqueness of the characteristic function. Let $\tilde{Y}$ be a random $\mathbb{S}_p^+$-valued matrix distributed according to the density function given in Eq. (3.40). Fix a symmetric $z \in \mathbb{R}^{p\times p}$. Then the characteristic function of $\tilde{Y}$ at z is given by

$$\varphi_{\tilde{Y}}(z) = \mathbb{E}\,[\mathrm{e}^{\mathrm{tr}(\mathrm{i}z\tilde{Y})}] = \frac{1}{(\det C)^{\alpha}\Gamma_p(\alpha)} \int_{\mathbb{S}_p^+} \mathrm{e}^{-\mathrm{tr}((C^{-1}-\mathrm{i}z)y)} (\det y)^{\alpha - \frac{p+1}{2}}\, \mathrm{d}y.$$

As C^{-1} is positive definite and $z \in \mathbb{R}^{p\times p}$ is symmetric, we can apply Eq. (2.31) from Theorem 2.37 in combination with Lemma 2.1 and arrive at

$$\varphi_{\tilde{Y}}(z) = \frac{1}{(\det C)^{\alpha}\Gamma_p(\alpha)} \Gamma_p(\alpha)\big(\det(C^{-1} - \mathrm{i}z)\big)^{-\alpha} = \left(\frac{\det(C^{-1} - \mathrm{i}z)}{\det(C^{-1})}\right)^{-\alpha}.$$

As the determinant is multiplicative,

$$\varphi_{\tilde{Y}}(z) = \det\big((C^{-1} - \mathrm{i}z)C\big)^{-\alpha} = \big(\det(C^{-1}C - \mathrm{i}zC)\big)^{-\alpha} = \big(\det(I_p - \mathrm{i}zC)\big)^{-\alpha},$$

which is the characteristic function of a random matrix, that follows a matrix-valued gamma distribution with parameters α and C, as proven in Theorem 3.32. As this is true for every symmetric $z \in \mathbb{R}^{p\times p}$, the random matrix $\tilde{Y}$ must follow a matrix-valued gamma distribution due to Remark 3.4 and Remark 3.13. □

Remark 3.46 (Complex matrix-valued gamma distributions)*.* In a similar manner to real matrix-valued gamma distributions discussed in this thesis, their complex version can be defined, taking values in the set of complex positive (semi-)definite matrices. This distribution is discussed in [23, Section 5.2a], for the complex case of the Wishart distributions, see [31].

3.3. Matrix-Valued Exponential Distributions

As a generalization to the one-dimensional exponential distributions discussed in Remark 2.30, we will define matrix-valued exponential distributions as a subclass of the matrix-valued gamma distributions and prove that they are memoryless w.r.t. the Loewner partial order given in Definition 2.10.

Definition 3.47 (Matrix-valued exponential distributions, see [22, Section 2.1])**.** For $p \in \mathbb{N}$ an $\mathbb{R}^{p\times p}$-valued random matrix Z is said to have a matrix-valued exponential distribution with positive semi-definite scale parameter matrix $C \in \mathbb{R}^{p\times p}$, if $Z \sim \mathrm{MGamma}((p+1)/2, C)$. This is denoted by $Z \sim \mathrm{MExp}(C)$.

Remark 3.48 (Characteristic function and density of matrix-valued exponential distributions)*.* For positive semi-definite $C \in \mathbb{R}^{p\times p}$ let $Z \sim \mathrm{MExp}(C)$.

For every symmetric $z \in \mathbb{R}^{p\times p}$, the characteristic function of the matrix-valued exponential function is given by

$$\varphi_Z(z) = (\det(I_p - \mathrm{i}zC))^{-\frac{p+1}{2}},$$

using Eq. (3.39).

Due to Theorem 3.45, for $C \in \mathbb{S}_p^+$ a density of this matrix-valued exponential distribution is given by

$$f_Z(x) = \frac{\mathrm{e}^{-\mathrm{tr}(C^{-1}x)}}{(\det C)^{\frac{p+1}{2}}\Gamma_p\big(\frac{p+1}{2}\big)}, \quad x \in \mathbb{S}_p^+. \tag{3.41}$$

One-dimensional exponential distributions have the powerful property of memorylessness. This result can be generalized to its matrix-valued version using the Loewner partial order.

Lemma 3.49 (Matrix-valued exponential dsitributions are memoryless w.r.t. the Loewner order)**.** *For $p \in \mathbb{N}$ and $C \in \mathbb{S}_p^+$, let $Z \sim \mathrm{MExp}(C)$. Then*

$$\mathbb{P}[Z >_L s + t \,|\, Z >_L s] = \mathbb{P}[Z >_L t]$$

for all positive semi-definite $s, t \in \mathbb{R}^{p\times p}$, where $<_L$ refers to the Loewner partial order given in Definition 2.10.

Proof. For every positive semi-definite $t \in \mathbb{R}^{p\times p}$, we can write

$$\mathbb{P}[Z >_L t] = \mathbb{P}[Z - t \in \mathbb{S}_p^+] = \int_{\{x\in\mathbb{S}_p^+ : x-t\in\mathbb{S}_p^+\}} f_Z(x)\,\mathrm{d}x = \int_{\mathbb{S}_p^+} \mathbb{1}_{\{x-t\in\mathbb{S}_p^+\}} f_Z(x)\,\mathrm{d}x$$

where the second equality follows from Definition 2.10 and the fact that Z takes values in the positive definite matrices. Using Remark 2.22 and Eq. (2.7) this integral is equal to

$$\mathbb{P}[Z >_L t] = \int_{\tilde{\mathbb{S}}_p^+} \mathbb{1}_{\{\psi(x^\triangle)-t\in\mathbb{S}_p^+\}} f_Z(\psi(x^\triangle))\,\mathrm{d}x^\triangle = \int_{\{x^\triangle\in\tilde{\mathbb{S}}_p^+ : x-t\in\mathbb{S}_p^+\}} f_Z(x)\,\mathrm{d}x^\triangle,$$

where $x = \psi(x^\triangle)$. The map

$$\phi := \begin{cases} D := \{x^\triangle \in \tilde{\mathbb{S}}_p^+ : x - t \in \mathbb{S}_p^+\} & \to \quad \tilde{\mathbb{S}}_p^+ \\ \qquad x^\triangle & \mapsto \quad (x-t)^\triangle \end{cases}$$

is well-defined as $x = \psi(x^\triangle) \in \mathbb{S}_p^+$ if and only if $x^\triangle \in \tilde{\mathbb{S}}_p^+$, see Remark 2.19. Furthermore, $\psi(D) = \tilde{\mathbb{S}}_p^+$ as for every $y^\triangle \in \tilde{\mathbb{S}}_p^+$ and corresponding $y = \psi(y^\triangle) \in \mathbb{S}_p^+$, the matrix $y + t \in \mathbb{S}_p^+$ due to Item (iii) of Lemma 2.8 and hence $(y+t)^\triangle \in D$ with $\phi((y+t)^\triangle) = y^\triangle$. As $(\cdot)^\triangle$ is the inverse function of the homeomorphism ψ given in Eq. (2.7), the map ϕ here is a homeomorphism as well, hence $D = \phi^{-1}(\tilde{\mathbb{S}}_p^+)$ is an open set as $\tilde{\mathbb{S}}_p^+$ is open due to Lemma 2.20. Furthermore, due to the linearity of $(\cdot)^\triangle$, the Jacobi matrix $\mathrm{d}\phi = I_{p(p+1)/2}$, leading to $\det \mathrm{d}\phi = 1$. Hence, we can use Theorem 2.39 to arrive at

$$\mathbb{P}[Z >_L t] = \int_D f_Z(x-t+t)\,\mathrm{d}x^\triangle = \int_D f_Z(\psi((x-t)^\triangle)+t)\,\mathrm{d}x^\triangle = \int_{\tilde{\mathbb{S}}_p^+} f_Z(y+t)\,\mathrm{d}y^\triangle, \tag{3.42}$$

where $y = \psi(y^\triangle)$. As the density of Z given in Eq. (3.41) satisfies

$$f_Z(y+t) = \frac{\mathrm{e}^{-\mathrm{tr}(C^{-1}(y+t))}}{(\det C)^{\frac{p+1}{2}}\Gamma_p\left(\frac{p+1}{2}\right)} = \mathrm{e}^{-\mathrm{tr}(C^{-1}t)}\,\frac{\mathrm{e}^{-\mathrm{tr}(C^{-1}y)}}{(\det C)^{\frac{p+1}{2}}\Gamma_p\left(\frac{p+1}{2}\right)} = \mathrm{e}^{-\mathrm{tr}(C^{-1}t)}\,f_Z(y),$$

where we use the linearity of the trace and the matrix multiplication, Eq. (3.42) can be rewritten as

$$\mathbb{P}[Z >_L t] = \mathrm{e}^{-\mathrm{tr}(C^{-1}t)} \int_{\tilde{\mathbb{S}}_p^+} f_Z(y)\,\mathrm{d}y^\triangle = \mathrm{e}^{-\mathrm{tr}(C^{-1}t)},$$

where the last equality holds due to the fact that we integrate a density of Z over all of the values Z can take as $y \in \mathbb{S}_p^+$ if and only if $y^\triangle \in \tilde{\mathbb{S}}_p^+$. For any positive semi-definite

$s, t \in \mathbb{R}^{p\times p}$, using the definition of conditional probabilities and the fact that $Z >_L t+s$ implies $Z >_L s$,

$$\mathbb{P}[Z >_L s+t \,|\, Z >_L s] = \frac{\mathbb{P}[Z >_L s+t]}{\mathbb{P}[Z >_L s]} = \frac{\mathrm{e}^{-\operatorname{tr}(C^{-1}(s+t))}}{\mathrm{e}^{-\operatorname{tr}(C^{-1}s)}} = \mathrm{e}^{-\operatorname{tr}(C^{-1}t)} = \mathbb{P}[Z >_L t]$$

holds, which proves the lemma. □

3.4. Marginal Distributions

For a better understanding of the structure of a matrix-valued gamma distribution we partition the values it can take into a block matrix and derive the distribution of these blocks in Theorem 3.52. Since the proof of this theorem requires the density function for most cases it is limited to distributions with positive definite scale parameter matrix C and non-singular shape parameter α. It is, however, possible to derive a similar block notation for singular Wishart distributions, see Theorem 3.54 below.

Lemma 3.50. *Let $\mathcal{K}$ denote a field. For $p \in \mathbb{N}$ let $A, B \in \mathcal{K}^{p\times p}$ be symmetric and consider their partitions*

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{pmatrix}, \quad \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^\top & B_{22} \end{pmatrix}, \tag{3.43}$$

with $A_{11}, B_{11} \in \mathcal{K}^{r\times r}$, where $r \in \{1, \dots, p-1\}$. If A_{11} and B_{22} are invertible, then

$$\begin{aligned} \operatorname{tr}(AB) &= \operatorname{tr}(A_{11}B_{11\cdot 2}) + \operatorname{tr}(A_{22\cdot 1}B_{22}) \\ &\quad + \operatorname{tr}\big(A_{11}(B_{12} + A_{11}^{-1}A_{12}B_{22})B_{22}^{-1}(B_{12} + A_{11}^{-1}A_{12}B_{22})^\top\big), \end{aligned} \tag{3.44}$$

where $B_{11\cdot 2} := B_{11} - B_{12}B_{22}^{-1}B_{12}^\top$ and $A_{22\cdot 1} := A_{22} - A_{12}^\top A_{11}^{-1}A_{12}$.

Proof. Using the partitions,

$$\operatorname{tr}(AB) = \operatorname{tr}\left(\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{12}^\top & B_{22} \end{pmatrix}\right)$$

can be rewritten as

$$\begin{aligned} \operatorname{tr}(AB) &= \operatorname{tr}(A_{11}B_{11}) + \operatorname{tr}(A_{12}B_{12}^\top) + \operatorname{tr}(A_{12}^\top B_{12}) + \operatorname{tr}(A_{22}B_{22}) \\ &= \operatorname{tr}(A_{11}(B_{11} - B_{12}B_{22}^{-1}B_{12}^\top + B_{12}B_{22}^{-1}B_{12}^\top)) + \operatorname{tr}(A_{12}B_{12}^\top) + \operatorname{tr}(A_{12}^\top B_{12}) \\ &\quad + \operatorname{tr}\big((A_{22} - A_{12}^\top A_{11}^{-1}A_{12} + A_{12}^\top A_{11}^{-1}A_{12})B_{22}\big). \end{aligned}$$

Using the definition of the blocks $B_{11\cdot 2}$ and $A_{22\cdot 1}$

$$\begin{aligned} \operatorname{tr}(AB) &= \operatorname{tr}(A_{11}B_{11\cdot 2}) + \operatorname{tr}(A_{22\cdot 1}B_{22}) + \operatorname{tr}\big(A_{11}(B_{12}B_{22}^{-1}B_{12}^\top)\big) \\ &\quad + \operatorname{tr}(A_{12}B_{12}^\top) + \operatorname{tr}(A_{12}^\top B_{12}) + \operatorname{tr}\big((A_{12}^\top A_{11}^{-1}A_{12})B_{22}\big). \end{aligned}$$

As the last four terms of this sum satisfy

$$\begin{aligned} &\operatorname{tr}\big(A_{11}(B_{12}B_{22}^{-1}B_{12}^\top)\big) + \operatorname{tr}(A_{12}B_{12}^\top) + \operatorname{tr}(A_{12}^\top B_{12}) + \operatorname{tr}\big((A_{12}^\top A_{11}^{-1}A_{12})B_{22}\big) \\ &\qquad = \operatorname{tr}\big(A_{11}(B_{12} + A_{11}^{-1}A_{12}B_{22})B_{22}^{-1}(B_{12} + A_{11}^{-1}A_{12}B_{22})^\top\big), \end{aligned}$$

Eq. (3.44) follows. □

Lemma 3.51 (Characterization of positive definite matrices, see [17, Eq. (7.7.5)]). *For $p \in \mathbb{N}$ let $A \in \mathbb{K}^{p\times p}$ be Hermitian and consider the same partition as in Eq. (3.43) with $A_{11} \in \mathbb{K}^{r\times r}$ for an $r \in \{1, \ldots, p-1\}$. Then A is positive definite if and only if A_{22} and $A_{11\cdot 2} := A_{11} - A_{12}A_{22}^{-1}A_{12}^{\mathsf{H}}$ are positive definite.*

Proof. Due to Lemma 2.9, A_{22} is positive definite if A is positive definite. Hence, A_{22} is invertible in both cases, see Item (v) of Lemma 2.8, and $A_{11\cdot 2}$ is well-defined. The matrix

$$H := \begin{pmatrix} I_r & -A_{12}A_{22}^{-1} \\ 0 & I_{p-r} \end{pmatrix} \in \mathbb{K}^{p\times p}$$

is invertible and satisfies

$$HAH^{\mathsf{H}} = \begin{pmatrix} I_r & -A_{12}A_{22}^{-1} \\ 0 & I_{p-r} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^{\mathsf{H}} & A_{22} \end{pmatrix} \begin{pmatrix} I_r & 0 \\ -A_{22}^{-1}A_{12}^{\mathsf{H}} & I_{p-r} \end{pmatrix} = \begin{pmatrix} A_{11\cdot 2} & 0 \\ 0 & A_{22} \end{pmatrix}. \tag{3.45}$$

For the first implication, let A be positive definite. Then A_{11} and A_{22} are positive definite due to Lemma 2.9 and therefore $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{12}^{\mathsf{H}}$ must be Hermitian as well as A_{22}^{-1} is positive definite due to Item (vi) of Lemma 2.8. For $x \in \mathbb{K}^r \setminus \{0\}$, let $\tilde{x} := (x_1, \ldots, x_r, 0, \ldots, 0) \in \mathbb{K}^p \setminus \{0\}$. Then

$$x^{\mathsf{H}} A_{11\cdot 2}\, x = \tilde{x}^{\mathsf{H}} \begin{pmatrix} A_{11\cdot 2} & 0 \\ 0 & A_{22} \end{pmatrix} \tilde{x} = (H^{\mathsf{H}}\tilde{x})^{\mathsf{H}} A H^{\mathsf{H}}\tilde{x} > 0,$$

as H is of full rank and hence $H^{\mathsf{H}}\tilde{x} \in \mathbb{K}^p \setminus \{0\}$, thus proving that $A_{11\cdot 2}$ is positive definite.

For the other implication, let $A_{11\cdot 2}$ and A_{22} be positive definite. For a fixed $x \in \mathbb{K}^p \setminus \{0\}$, the vector $y := (H^{\mathsf{H}})^{-1}x \in \mathbb{K}^p \setminus \{0\}$ can be partitioned into $y = (y_1, y_2)^{\mathsf{T}}$ with $y_1 \in \mathbb{K}^r \setminus \{0\}$ and $y_2 \in \mathbb{K}^{p-r} \setminus \{0\}$. Then, using Eq. (3.45),

$$x^{\mathsf{H}}Ax = y^{\mathsf{H}}HAH^{\mathsf{H}}y = y_1^{\mathsf{H}}A_{11\cdot 2}\, y_1 + y_2^{\mathsf{H}}A_{22}\, y_2 > 0.$$

As this holds true for an arbitrary $x \in \mathbb{K}^p \setminus \{0\}$, it must be true for all $x \in \mathbb{K}^p \setminus \{0\}$, hence A is positive definite. □

Theorem 3.52 (Marginal and conditional distributions of matrix-valued gamma distributions partitioned into blocks, see [20, Proposition 2.3]). *For $p \in \mathbb{N}$, positive semi-definite $C \in \mathbb{R}^{p\times p}$ and $\alpha \in \mathcal{G}_C$ consider the partition of $Y \sim \mathrm{MGamma}(\alpha, C)$ and the scale parameter matrix C into the blocks*

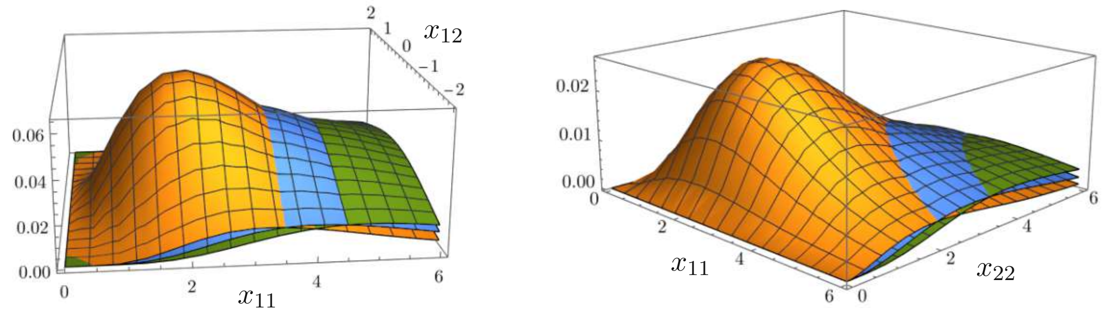
$$Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^{\mathsf{T}} & Y_{22} \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}^{\mathsf{T}} & C_{22} \end{pmatrix}$$

with Y_{11}, C_{11} of size $r \times r$, where $r \in \{1, \ldots, p-1\}$. Then

(i) $Y_{11} \sim \mathrm{MGamma}(\alpha, C_{11})$,

(ii) $Y_{22} \sim \mathrm{MGamma}(\alpha, C_{22})$.

If $C \in \mathbb{S}_p^+$ and $\alpha > (p-1)/2$, then the following statements hold true as well, where $C_{11\cdot 2}$ denotes $C_{11} - C_{12}C_{22}^{-1}C_{12}^{\mathsf{T}}$ and $C_{22\cdot 1} := C_{22} - C_{12}^{\mathsf{T}}C_{11}^{-1}C_{12}$ with the same notation applying to Y:



(a) $f_{X_{11},X_{12}}(x_{11},x_{12})$ (b) $f_{X_{11},X_{22}}(x_{11},x_{22})$

Figure 3.4. *Illustration of the joint density functions of* (X_{11}, X_{12}) *and* (X_{11}, X_{22}) *of* $X \sim \mathrm{MGamma}(\alpha, I_2)$ *for* $\alpha = 7/2$ *in orange,* $\alpha = 9/2$ *in blue and* $\alpha = 11/2$ *in green. Note that these can also be the density functions of two-dimensional marginal distributions of singular Wishart distributions, see Example 3.56.*

(iii) $Y_{12}|Y_{22}^{\triangle} \sim \mathrm{MNormal}(C_{12}C_{22}^{-1}Y_{22}, \frac{1}{2}C_{11\cdot2} \otimes Y_{22})$ *a.s.,*

(iv) $Y_{12}^{\mathsf{T}}|Y_{11}^{\triangle} \sim \mathrm{MNormal}(C_{12}^{\mathsf{T}}C_{11}^{-1}Y_{11}, \frac{1}{2}C_{22\cdot1} \otimes Y_{11})$ *a.s.,*

(v) $Y_{11\cdot2} \sim \mathrm{MGamma}(\alpha - \frac{p-r}{2}, C_{11\cdot2})$,

(vi) $Y_{22\cdot1} \sim \mathrm{MGamma}(\alpha - \frac{r}{2}, C_{22\cdot1})$,

(vii) (Y_{12}, Y_{22}) *and* $Y_{11\cdot2}$ *are independent,*

(viii) (Y_{12}, Y_{11}) *and* $Y_{22\cdot1}$ *are independent.*

Proof. For Item (i) we define the matrix

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{r \times p}.$$

Inserting this matrix into Corollary 3.39,

$$Y_{11} = LXL^{\mathsf{T}} \sim \mathrm{MGamma}(\alpha, LCL^{\mathsf{T}}) = \mathrm{MGamma}(\alpha, C_{11}).$$

Item (ii) follows by the permutation property given in Corollary 3.40.

Items (iii), (v) and (vii) will be proven simultaneously in three steps.

Step 1. As C is required to be positive definite in the case of Item (iii), Item (v) and Item (vii), the matrices C_{11}, C_{22}, $C_{11\cdot2}$ and $C_{22\cdot1}$ are positive definite as well due to Lemma 3.51. Hence, they are all invertible and the inverse of C is given by inverse is given by

$$C^{-1} = \begin{pmatrix} C_{11\cdot2}^{-1} & -C_{11}^{-1}C_{12}C_{22\cdot1}^{-1} \\ -C_{22}^{-1}C_{12}^{\mathsf{T}}C_{11\cdot2}^{-1} & C_{22\cdot1}^{-1} \end{pmatrix},$$

which can be proven by verifying $CC^{-1} = I_p$. Combining this with Eq. (3.44) from Lemma 3.50 above leads to

$$\begin{aligned}\operatorname{tr}(C^{-1}y) &= \operatorname{tr}(C_{11\cdot 2}^{-1}y_{11\cdot 2}) + \operatorname{tr}(C_{22}^{-1}y_{22}) \\ &\quad + \operatorname{tr}\big(C_{11\cdot 2}^{-1}(y_{12} - C_{12}C_{22}^{-1}y_{22})y_{22}^{-1}(y_{12} - C_{12}C_{22}^{-1}y_{22})^{\mathsf{T}}\big), \quad y \in \mathbb{S}_p^+. \end{aligned} \tag{3.46}$$

Using Lemma 2.2 with index set $\{r+1, \dots, p\}$,

$$\det C = \det(C_{22})\det(C_{11} - C_{12}C_{22}^{-1}C_{12}^{\mathsf{T}}) = \det(C_{22})\det(C_{11\cdot 2}), \tag{3.47}$$

where we use the fact that C_{22} is positive definite and invertible as a consequence of $C \in \mathbb{S}_p^+$. For every value $y \in \mathbb{S}_p^+$, that Y might take, as discussed in Remark 3.27, we can also write $\det y = \det y_{22} \det y_{11\cdot 2}$. Using the representation of $\Gamma_p(\alpha)$ given in Eq. (2.30),

$$\Gamma_p(\alpha) = (2\pi)^{\frac{r(p-r)}{2}}\Gamma_r\left(\alpha - \frac{p-r}{2}\right)\Gamma_{p-r}(\alpha), \tag{3.48}$$

where we use the equality

$$r(r-1) + (p-r)(p-r-1) + 2r(p-r) = p(p-1) \tag{3.49}$$

for the exponent of 2π. Due to Eq. (3.20) and Eq. (3.40), a density function of $Y^{\triangle}$ is given by

$$f_{Y^{\triangle}}(y^{\triangle}) = 2^{\frac{p(p-1)}{4}} f_Y(y) = \frac{2^{p(p-1)/4}}{(\det C)^{\alpha}\Gamma_p(\alpha)} \mathrm{e}^{-\operatorname{tr}(C^{-1}y)}(\det y)^{\alpha - \frac{p+1}{2}}, \quad y^{\triangle} \in \tilde{\mathbb{S}}_p^+,$$

where we recall the notation from Remark 2.19, where each $y^{\triangle} \in \tilde{\mathbb{S}}_p^+$ can be bijectively mapped to $y = \psi(y^{\triangle}) \in \mathbb{S}_p^+$. Inserting Eqs. (3.46), (3.47) and (3.48) into this density function together with some rearranging results in

$$\begin{aligned} f_{Y^{\triangle}}(y^{\triangle}) = &\frac{2^{r(r-1)/4}}{\Gamma_r(\alpha - \frac{p-r}{2})(\det C_{11\cdot 2})^{\alpha - \frac{p-r}{2}}} \mathrm{e}^{-\operatorname{tr}(C_{11\cdot 2}^{-1}y_{11\cdot 2})}(\det y_{11\cdot 2})^{\alpha - \frac{p+1}{2}}(\det C_{11\cdot 2})^{-\frac{p-r}{2}} \\ &\times \frac{2^{(p-r)(p-r-1)/4}}{\Gamma_{p-r}(\alpha)(\det C_{22})^{\alpha}}(\det y_{22})^{\alpha - \frac{(p-r)+1}{2}} \mathrm{e}^{-\operatorname{tr}(C_{22}^{-1}y_{22})}(\det y_{22})^{-\frac{r}{2}} \\ &\times \underbrace{2^{\frac{2r(p-r)}{4}}(2\pi)^{-\frac{r(p-r)}{2}}}_{=\pi^{-\frac{r(p-r)}{2}}} \mathrm{e}^{-\operatorname{tr}\left(C_{11\cdot 2}^{-1}(y_{12} - C_{12}C_{22}^{-1}y_{22})y_{22}^{-1}(y_{12} - C_{12}C_{22}^{-1}y_{22})^{\mathsf{T}}\right)}, \quad y^{\triangle} \in \tilde{\mathbb{S}}_p^+, \end{aligned} \tag{3.50}$$

where $y = \psi(y^{\triangle})$, and where we use Eq. (3.49) for the exponent of 2 in Eq. (3.50).

Step 2. Consider the function

$$\phi := \begin{cases} \tilde{\mathbb{S}}_p^+ & \to \quad R \subseteq \mathbb{R}^{p(p+1)/2} \\ \begin{pmatrix} y_{11}^{\triangle} & \\ y_{12}^{\mathsf{T}} & y_{22}^{\triangle} \end{pmatrix} & \mapsto \quad \begin{pmatrix} y_{11\cdot 2}^{\triangle} & \\ y_{12}^{\mathsf{T}} & y_{22}^{\triangle} \end{pmatrix} \end{cases},$$

with

$$R \coloneqq \{\mathrm{vecp}(x) \,|\, x \in \mathbb{R}^{p\times p},\, x = x^\mathsf{T},\, x_{11} \in \mathbb{S}_r^+,\, x_{22} \in \mathbb{S}_{p-r}^+\},$$

which makes ϕ well-defined and bijective as a symmetric $A \in \mathbb{R}^{p\times p}$ is positive definite if and only if $y_{11\cdot 2}$ and y_{22} are positive definite, see Lemma 3.51. Note that ϕ and ϕ^{-1} are continuous, as $y_{11\cdot 2} = y_{11} - (y_{12}y_{22}^{-1}y_{12}^\mathsf{T})$ and $(\cdot)^\triangle = \psi^{-1}(\cdot)$ is a homeomorphism as elaborated in Remark 2.19.

Using ϕ we can transform the density given in Eq. (3.50) to a joint density of $\tilde{Y}^\triangle \coloneqq (Y_{11\cdot 2}^\triangle, Y_{12}^\mathsf{T}, Y_{22}^\triangle)$. In order to apply the transformation formula from Theorem 2.39, we need to compute $|\det \mathrm{d}\phi(\cdot)|$: At each point of $\tilde{\mathbb{S}}_p^+$ the matrix $\mathrm{d}\phi$ can be seen as block matrix

$$\mathrm{d}\phi(y^\triangle) = \begin{pmatrix} \partial y_{11\cdot 2}^\triangle/\partial y_{11}^\triangle & \partial y_{11\cdot 2}^\triangle/\partial y_{12}^\mathsf{T} & \partial y_{11\cdot 2}^\triangle/\partial y_{22}^\triangle \\ \partial y_{12}^\mathsf{T}/\partial y_{11}^\triangle & \partial y_{12}^\mathsf{T}/\partial y_{12}^\mathsf{T} & \partial y_{12}^\mathsf{T}/\partial y_{22}^\triangle \\ \partial y_{22}^\triangle/\partial y_{11}^\triangle & \partial y_{22}^\triangle/\partial y_{12}^\mathsf{T} & \partial y_{22}^\triangle/\partial y_{22}^\triangle \end{pmatrix}, \quad y^\triangle \in \tilde{\mathbb{S}}_p^+,$$

where each block refers to the matrix of partial derivatives of the vectorized versions of the respective matrices. Since all of the partial derivatives matrices below the block diagonal are equal to zero matrices, the determinant of $\mathrm{d}\phi$ is equal to the product of the determinants of the diagonal block matrices. Since

$$\partial y_{11\cdot 2}^\triangle/\partial y_{11}^\triangle = \partial(y_{11}^\triangle - (y_{12}y_{22}^{-1}y_{12}^\mathsf{T})^\triangle)/\partial y_{11}^\triangle = \underbrace{\partial y_{11}^\triangle/\partial y_{11}^\triangle}_{\equiv I_{\frac{r(r+1)}{2}}} - \underbrace{\partial(y_{12}y_{22}^{-1}y_{12}^\mathsf{T})^\triangle/\partial y_{11}^\triangle}_{\equiv 0},$$

we can write

$$\det \mathrm{d}\phi = \det(\partial y_{11\cdot 2}^\triangle/\partial y_{11}^\triangle)\det(\partial y_{12}^\mathsf{T}/\partial y_{12}^\mathsf{T})\det(\partial y_{22}^\triangle/\partial y_{22}^\triangle) \equiv 1\cdot 1\cdot 1 = 1,$$

as the determinant of the identity matrix of any dimension is equal to one. Now for every open set B of $\mathcal{B}(R)$,

$$\mathbb{P}[\tilde{Y}^\triangle \in B] = \mathbb{P}[\phi(Y^\triangle) \in B] = \mathbb{P}[Y^\triangle \in \phi^{-1}(B)] = \int_{\phi^{-1}(B)} f_{Y^\triangle}(y^\triangle)\,\mathrm{d}y^\triangle,$$

where we write $\mathrm{d}y^\triangle$ for $\lambda^{\frac{p(p+1)}{2}}(\mathrm{d}y^\triangle)$. As ϕ is bijective and as $\det \mathrm{d}\phi = 1$,

$$\begin{aligned}\mathbb{P}[\tilde{Y}^\triangle \in B] = \int_{\phi^{-1}(B)} f_{Y^\triangle}(y^\triangle)\,\mathrm{d}y^\triangle &= \int_{\phi^{-1}(B)} f_{Y^\triangle}(\phi^{-1}(\phi(y^\triangle)))|\det \mathrm{d}\phi(y^\triangle)|\,\mathrm{d}y^\triangle \\ &= \int_B f_{Y^\triangle}(\phi^{-1}(\tilde{y}^\triangle))\,\mathrm{d}\tilde{y}^\triangle,\end{aligned}$$

where the transformation formula from Theorem 2.39 with the function ϕ is used for the second equality. As this is true for every Borel set B a density of $\tilde{Y}^\triangle$ is given by

$$f_{\tilde{Y}^\triangle}(\tilde{y}^\triangle) = f_{Y^\triangle}(\phi^{-1}(\tilde{y}^\triangle)) = f_{Y^\triangle}\left(\begin{pmatrix} \tilde{y}_{11} + \tilde{y}_{12}\tilde{y}_{22}^{-1}\tilde{y}_{12}^\mathsf{T} & \tilde{y}_{12} \\ \tilde{y}_{12}^\mathsf{T} & \tilde{y}_{22}\end{pmatrix}^\triangle\right), \quad \tilde{y}^\triangle \in R,$$

where $\tilde{y} = \psi(\tilde{y}^\triangle)$ by extending the definition of the domain of ψ in Eq. (2.7) to R.

Step 3. Combining this with Eq. (3.50), we can see that the density can be written as

$$f_{\tilde{Y}^{\triangle}}(\tilde{y}^{\triangle}) = f_{(Y_{11\cdot 2}^{\triangle}, Y_{12}^{\mathsf{T}}, Y_{22}^{\triangle})}(\tilde{y}^{\triangle}) = f_{Y_{11\cdot 2}^{\triangle}}(\tilde{y}_{11}^{\triangle}) \cdot f_{(Y_{12}^{\mathsf{T}}, Y_{22}^{\triangle})}(\tilde{y}_{12}^{\mathsf{T}}, \tilde{y}_{22}^{\triangle})$$

for all $\tilde{y}^{\triangle} \in R$, or, equivalently for all $\tilde{y}_{11}^{\triangle} \in \tilde{\mathbb{S}}_r^+$, $\tilde{y}_{12} \in \mathbb{R}^{r\times(p-r)}$ and $\tilde{y}_{22}^{\triangle} \in \tilde{\mathbb{S}}_{p-r}^+$ with

$$f_{Y_{11\cdot 2}^{\triangle}}(\tilde{y}_{11}^{\triangle}) = \frac{2^{r(r-1)/2}}{\Gamma_r(\alpha - \frac{p-r}{2})(\det C_{11\cdot 2})^{\alpha - \frac{p-r}{2}}} \, \mathrm{e}^{-\operatorname{tr}(C_{11\cdot 2}^{-1}\tilde{y}_{11})} (\det \tilde{y}_{11})^{(\alpha - \frac{p-r}{2}) - \frac{r+1}{2}},$$

where $\tilde{y}_{11} = \psi(\tilde{y}_{11}^{\triangle}) \in \mathbb{S}_r^+$ and noting that $(\tilde{y}_{11} + \tilde{y}_{12}\tilde{y}_{22}^{-1}\tilde{y}_{12}^{\mathsf{T}}) - \tilde{y}_{12}\tilde{y}_{22}^{-1}\tilde{y}_{12}^{\mathsf{T}} = \tilde{y}_{11}$, and

$$\begin{aligned} f_{(Y_{12}^{\mathsf{T}}, Y_{22}^{\triangle})}(\tilde{y}_{12}^{\mathsf{T}}, \tilde{y}_{22}^{\triangle}) &= \frac{2^{(p-r)(p-r-1)/4}}{\Gamma_{p-r}(\alpha)(\det C_{22})^{\alpha}} (\det \tilde{y}_{22})^{\alpha - \frac{(p-r)+1}{2}} \, \mathrm{e}^{-\operatorname{tr}(C_{22}^{-1}\tilde{y}_{22})} (\det \tilde{y}_{22})^{-\frac{r}{2}} \\ &\quad \times (\det C_{11\cdot 2})^{-\frac{p-r}{2}} \pi^{-\frac{r(p-r)}{2}} \, \mathrm{e}^{-\operatorname{tr}\left(C_{11\cdot 2}^{-1}(\tilde{y}_{12} - C_{12}C_{22}^{-1}\tilde{y}_{22})\tilde{y}_{22}^{-1}(\tilde{y}_{12} - C_{12}C_{22}^{-1}\tilde{y}_{22})^{\mathsf{T}}\right)}, \end{aligned}$$

thus proving the independence of $Y_{11\cdot 2}^{\triangle}$ and $(Y_{12}^{\mathsf{T}}, Y_{22}^{\triangle})$ and, equivalently, the independence of $Y_{11\cdot 2}$ and (Y_{12}, Y_{22}). Furthermore, we can see that $f_{Y_{11\cdot 2}}(\tilde{y}_{11}) = 2^{-r(r-1)/2} f_{Y_{11\cdot 2}}^{\triangle}(\tilde{y}_{11}^{\triangle})$, where $\tilde{y}_{11} \in \mathbb{S}_r^+$, corresponds to the r-dimensional matrix-valued gamma distribution with shape parameter $\alpha - (p-r)/2$ and scale parameter matrix $C_{11\cdot 2}$, hence proving Item (v). We already know the distribution of Y_{22} from Item (ii), hence a density of $Y_{22}^{\triangle}$, using Eqs. (3.20) and (3.40), is given by

$$f_{Y_{22}^{\triangle}}(\tilde{y}_{22}^{\triangle}) = \frac{2^{(p-r)(p-r-1)/4}}{\Gamma_{p-r}(\alpha)(\det C_{22})^{\alpha}} (\det \tilde{y}_{22})^{\alpha - \frac{(p-r)+1}{2}} \, \mathrm{e}^{-\operatorname{tr}(C_{22}^{-1}\tilde{y}_{22})}, \quad \tilde{y}_{22}^{\triangle} \in \tilde{\mathbb{S}}_{p-r}^+,$$

where $\tilde{y}_{22} = \psi(\tilde{y}_{22}^{\triangle})$. Therefore, we can calculate a conditional density of Y_{12}^{T} given $Y_{22}^{\triangle}$

$$\begin{aligned} f_{Y_{12}^{\mathsf{T}}|Y_{22}^{\triangle}}(\tilde{y}_{12}^{\mathsf{T}}|\tilde{y}_{22}^{\triangle}) &= \frac{f_{(Y_{12}^{\mathsf{T}}, Y_{22}^{\triangle})}(\tilde{y}_{12}^{\mathsf{T}}, \tilde{y}_{22}^{\triangle})}{f_{Y_{22}^{\triangle}}(\tilde{y}_{22}^{\triangle})} \\ &= (\det C_{11\cdot 2})^{-\frac{p-r}{2}} (\det \tilde{y}_{22})^{-\frac{r}{2}} \pi^{-\frac{r(p-r)}{2}} \, \mathrm{e}^{-\operatorname{tr}\left(C_{11\cdot 2}^{-1}(\tilde{y}_{12} - C_{12}C_{22}^{-1}\tilde{y}_{22})\tilde{y}_{22}^{-1}(\tilde{y}_{12} - C_{12}C_{22}^{-1}\tilde{y}_{22})^{\mathsf{T}}\right)} \\ &= f_{Y_{12}|Y_{22}^{\triangle}}(\tilde{y}_{12}|\tilde{y}_{22}^{\triangle}) \end{aligned}$$

for all $\tilde{y}_{12} \in \mathbb{R}^{r\times(p-r)}$ and $\tilde{y}_{22}^{\triangle} \in \tilde{\mathbb{S}}_{p-r}^+$, where the last equality follows from the the fact that $\mathbb{P}[Y_{12} \in B] = \mathbb{P}[Y_{12}^{\mathsf{T}} \in \{b^{\mathsf{T}} | b \in B\}]$ for every Borel set $B \in \mathcal{B}(\mathbb{R}^{r\times(p-r)})$. This function is, using Eq. (3.3) and

$$\left(\det \frac{1}{2} C_{11\cdot 2}\right)^{-\frac{p-r}{2}} = (\det C_{11\cdot 2})^{-\frac{p-r}{2}} \left(\frac{1}{2}\right)^{-\frac{r(p-r)}{2}},$$

a density of the $\mathrm{MNormal}(C_{12}C_{22}^{-1}Y_{22}, \frac{1}{2}C_{11\cdot 2} \otimes Y_{22})$ distribution, thus proving Item (iii).

Items (iv), (vi) and (viii) follow by the permutation property given in Corollary 3.40.

□

Corollary 3.53 (The diagonal entries of a matrix-valued gamma distribution follow a one-dimensional gamma distribution)**.** *Let* $Y \sim \mathrm{MGamma}(\alpha, C)$*. Then every diagonal entry* Y_{ii}*, with* $i \in \{1, \dots, p\}$*, of the matrix* Y *follows the one-dimensional gamma distribution with shape parameter* α *and scale parameter* c_{ii}*, where* c_{ii} *refers to the* i*-th diagonal entry of the scale matrix* C*.*

We can describe singular Wishart distributions using a similar block notation as given in Theorem 3.52. This notation is also used in literature to introduce singular Wishart distributions, e.g. see [20, Section 2.3].

Theorem 3.54 (Characterization of standard singular Wishart distributions)**.** *For* $n, p \in \mathbb{N}$ *with* $n < p$*, let* $X \sim \mathrm{Wishart}(n, I_p)$*. Then the partition of* X *into the blocks*

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{12}^\mathsf{T} & X_{22} \end{pmatrix},$$

where X_{11} *is of size* $r \times r$*, satisfies*

$$X_{11} \sim \mathrm{MGamma}(\tfrac{n}{2}, 2I_n), \quad X_{12}^\mathsf{T} | X_{11} \sim \mathrm{MNormal}(0, I_{p-n} \otimes X_{11}) \text{ a.s.}, \quad X_{22} \overset{\text{a.s.}}{=} X_{12}^\mathsf{T} X_{11}^{-1} X_{12}.$$

Note that we can also write $X_{11} \sim \mathrm{Wishart}(n, I_n)$ *due to Corollary 3.33.*

Remark 3.55. Note that this distribution of the blocks X_{11} and X_{12} is the same as the distribution of the respective blocks of a standard matrix-valued gamma distribution, see Theorem 3.52.

Example 3.56 (Two-dimensional marginal distributions of singular Wishart distributions)**.** For $n, p \in \mathbb{N}$ with $n < p$, let $X \sim \mathrm{Wishart}(n, I_p)$, which corresponds to the $\mathrm{MGamma}(n/2, 2I_p)$ distribution.

Combining Theorem 3.54 with Corollary 3.40 implies that for every $i \in \{1, \dots, n-1\}$ the $\mathbb{R}$-valued entries of X satisfy

$$X_{ii} \sim \mathrm{Gamma}(n/2, 2), \quad X_{i,i+1} | X_{ii} \sim \mathcal{N}(0, 1) \text{ a.s.}$$

Hence, for sufficiently large $p \in \mathbb{N}$, Figure 3.4a shows a density function of the joint distribution of $(X_{ii}/2, X_{i,i+1}/2)$.

Due to Lemma 3.7 and Corollary 3.53, the diagonal entries of X are i.i.d. Therefore, for sufficiently large $p \in \mathbb{N}$, Figure 3.4b shows a density function of the joint distribution of two diagonal entries of X scaled by factor $1/2$.

Proof of Theorem 3.54. For $X \sim \mathrm{Wishart}(n, I_p)$, let Z be the $\mathbb{R}^{p \times n}$-valued random matrix following the standard matrix-valued normal distribution, satisfying $ZZ^\mathsf{T} \overset{\text{d}}{=} X$, see Definition 3.6. Let Z_1 denote the first n rows of N, and Z_2 the last $p - n$ rows, hence X can be partitioned into

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{12}^\mathsf{T} & X_{22} \end{pmatrix} \overset{\text{d}}{=} ZZ^\mathsf{T} = \begin{pmatrix} Z_1 Z_1^\mathsf{T} & Z_1 Z_2^\mathsf{T} \\ Z_2 Z_1^\mathsf{T} & Z_2 Z_2^\mathsf{T} \end{pmatrix}.$$

Since the dimensions of X_{11} and $Z_1 Z_1^\mathsf{T}$ are equal, the blocks of the two partitions correspond to each other. Since $Z_1 \sim \mathrm{MNormal}(0, I_n \otimes I_n)$, Definition 3.8 implies that $Z_1 Z_1^\mathsf{T} = X_{11} \sim \mathrm{Wishart}(n, I_n)$ or, using Corollary 3.33, $X_{11} \sim \mathrm{MGamma}(\frac{n}{2}, 2I_n)$.

As $Z_2 \sim \text{MNormal}(0, I_{p-n} \otimes I_n)$ and affine transformations of normal distributions are normal distributions, see Item (iii) form Remark 3.5,

$$Z_2 Z_1^\top \mid Z_1 \sim \text{MNormal}(0, I_{p-n} \otimes (Z_1^\top)^\top I_n Z_1^\top) = \text{MNormal}(0, I_{p-n} \otimes X_{11}) \text{ a.s.}$$

Since the distribution of $Z_2 Z_1^\top \mid Z_1$ only depends on the Z_1 measurable random matrix X_{11}, we can conclude that $Z_2 Z_1^\top \mid X_{11} \sim \text{MNormal}(0, I_{p-n} \otimes X_{11})$ a.s.

For the distribution of X_{22}, note that

$$X_{12}^\top X_{11}^{-1} X_{12} \overset{\text{d}}{=} Z_2 Z_1^\top (Z_1 Z_1^\top)^{-1} Z_1 Z_2^\top = Z_2 Z_1^\top (Z_1^\top)^{-1} (Z_1)^{-1} Z_1 Z_2^\top = Z_2 Z_2^\top \overset{\text{d}}{=} X_{22},$$

where the second equality follows from the fact that Z_1 as well as $Z_1^\top$ are quadratic and have full rank a.s. since it is a non-degenerate normally distributed random matrix. □

3.5. Expectations and Covariances

In this section the expectation and covariance matrix of general matrix-valued gamma distributions are derived using the characteristic function.

Remark 3.57 (The commutation matrix K_p). For $p \in \mathbb{N}$ the commutation matrix K_p refers to the $p^2 \times p^2$ matrix that consists of the $p \times p$ block matrices S_{ij} for $i, j \in \{1, \ldots, p\}$ that are defined by

$$(S_{ij})_{kl} = \begin{cases} 1 & \text{if } i = l \text{ and } j = k \\ 0 & \text{otherwise.} \end{cases} \tag{3.51}$$

for all $k, l \in \{1, \ldots, p\}$. The commutation matrix has the property that it transforms the vectorized version of any matrix $A \in \mathbb{R}^{p \times p}$ into the vectorized version of its transpose, i.e.

$$K_p \operatorname{vec}(A) = \operatorname{vec}(A^\top).$$

Remark 3.58 (Properties of the Kronecker product, see [11, Eq. (2.11), Eq.(2.13), Eq. (2.14), Section 2.5]). Let p, q, n, m, k, l be natural numbers greater than zero.

(i) For matrices $A \in \mathbb{R}^{p \times q}$, $B \in \mathbb{R}^{q \times m}$ and $C \in \mathbb{R}^{m \times n}$

$$\operatorname{vec}(ABC) = (C^\top \otimes A)\operatorname{vec}(B) \tag{3.52}$$

holds.

(ii) For all matrices $A \in \mathbb{R}^{p \times q}$, $B \in \mathbb{R}^{q \times m}$, $C \in \mathbb{R}^{n \times k}$ and $D \in \mathbb{R}^{k \times l}$

$$(A \otimes C)(B \otimes D) = AB \otimes CD. \tag{3.53}$$

(iii) For the the commutation matrices K_p and K_q given in Remark 3.57 and $A, B \in \mathbb{R}^{p \times q}$

$$(A \otimes B)K_q = K_p(B \otimes A). \tag{3.54}$$

Theorem 3.59 (Expectation and covariance matrix of matrix-valued gamma distributions)**.** *For $p \in \mathbb{N}$, let $Y \sim \mathrm{MGamma}(\alpha, C)$ with positive semi-definite scale parameter matrix $C \in \mathbb{R}^{p\times p}$ and shape parameter $\alpha \in \mathcal{G}_C$. Then the expectation and covariance of the entries of the matrix Y are given by*

$$\mathbb{E}[Y_{ij}] = \alpha\, c_{ij}, \tag{3.55}$$
$$\mathrm{Cov}(Y_{ij}, Y_{kl}) = \frac{\alpha}{2}(c_{ik}c_{jl} + c_{il}c_{jk}) \tag{3.56}$$

for all $i, j, k, l \in \{1, \dots, p\}$. In matrix notation we can also write

$$\mathbb{E}[Y] = \alpha C, \tag{3.57}$$
$$\mathrm{Cov}(\mathrm{vec}(Y)) = \frac{\alpha}{2}(C \otimes C)(K_p + I_{p^2}), \tag{3.58}$$

where K_p refers to the commutation matrix defined in Remark 3.57.

Remark 3.60 (Identifiability of matrix-valued gamma distributions)*.* The expectations and covariances given in Theorem 3.59 above tell us that a matrix-valued gamma distribution is identifiable if we assume $\alpha > 0$ and $C \neq 0$, i.e. it is possible to determine the parameters of the distribution after obtaining an infinite number of observations from it.[9]

Remark 3.61 (Equivalence of the notations for the covariance)*.* In order to prove that Eq. (3.56) and Eq. (3.58) are equivalent, fix two entries Y_{ij}, Y_{kl} of Y with $i, j, k, l \in \{1, \dots, p\}$. Using the definition of the function $\mathrm{vec}(\cdot)$, given in Eq. (2.4), we can see the covariance matrix $\mathrm{Cov}(\mathrm{vec}(Y))$ as a block matrix out of the blocks $C^{\mathrm{b}}_{jl} \in \mathbb{R}^{p\times p}$, that are the covariance matrices between the j-th and the l-th column of Y for all $j, l \in \{1, \dots, p\}$. Considering the right side of Eq. (3.58), the matrix C^{b}_{jl} is given by

$$C^{\mathrm{b}}_{jl} = \frac{\alpha}{2}\left(\sum_{m=1}^{p} \underbrace{(C \otimes C)_{\mathbf{jm}}}_{=c_{jm}C} \big(\underbrace{(K_p)_{\mathbf{ml}}}_{=S_{ml}} + I_p\, 1_{\{m=l\}}\big)\right) = \frac{\alpha}{2}\left(c_{jl}C + \sum_{m=1}^{p} c_{jm} C S_{ml}\right),$$

where bold indices are used for accessing the blocks of the the block matrix and the definition of the Kronecker product given in Eq. (3.1) and of the commutation matrix given in Remark 3.57 are used for the small equalities.

For the covariance of Y_{ij} and Y_{kl} we have to consider the entry at position (i, k) of C^{b}_{jl}. Using the definition of S_{ij} given in Eq. (3.51),

$$\begin{aligned}\mathrm{Cov}(Y_{ij}, Y_{kl}) = (C^{\mathrm{b}}_{jl})_{ik} &= \frac{\alpha}{2}\left(c_{jl}c_{ik} + \sum_{m=1}^{p} c_{jm}(CS_{ml})_{ik}\right)\\ &= \frac{\alpha}{2}\left(c_{jl}c_{ik} + \sum_{m=1}^{p} c_{jm} \sum_{n=1}^{p} c_{in} \underbrace{(S_{ml})_{nk}}_{=1_{\{m=k, l=n\}}}\right) = \frac{\alpha}{2}(c_{jl}c_{ik} + c_{jk}c_{il}),\end{aligned}$$

which is equal to the covariance given in Eq. (3.56).

[9] The term identifiability is commonly used in statistics, see e.g. [12].

Proof of Theorem 3.59. Due to Definition 3.34 and therefore Definitions 3.8 and 3.24, a random matrix Y following a general matrix-valued gamma distribution with $C \neq 0$ and $\alpha \neq 0$ has the stochastic representation $Y \overset{\mathrm{d}}{=} AXA^\intercal$ with $A \in \mathbb{R}^{p\times \mathrm{rk}(C)}$, such that $AA^\intercal = C$, where $X \sim \mathrm{MGamma}(\alpha, I_{\mathrm{rk}(C)})$. For $C = 0$ or $\alpha = 0$ both expectations and covariances are trivial.

The proof of Theorem 3.59 is now split into three steps: In the first step we derive the expectation of this random matrix X following a q-dimensional general standard matrix-valued gamma distribution for every $q \in \mathbb{N}$ and in the second step its covariance matrix. In the third step we generalize both results for all positive semi-definite scale parameter matrices $C \in \mathbb{R}^{p\times p}$ using properties of vectorization and the Kronecker product.

Step 1. Due to Corollary 3.35, the characteristic function of X is given by

$$\varphi_X(z) = \big(\det(I_q - \mathrm{i}z)\big)^{-\alpha}$$

for every symmetric $z \in \mathbb{R}^{q\times q}$. Combined with Remark 3.13, the characteristic function of X for a general, not necessarily symmetric, $z \in \mathbb{R}^{q\times q}$ is given by

$$\varphi_X(z) = \varphi_X\left(\frac{z+z^\intercal}{2}\right) = \left(\det\left(I_q - \mathrm{i}\,\frac{z+z^\intercal}{2}\right)\right)^{-\alpha}. \tag{3.59}$$

Throughout this proof we will denote

$$M(z) \coloneqq I_q - \mathrm{i}\,\frac{z+z^\intercal}{2} \in \mathbb{R}^{q\times q}, \tag{3.60}$$

and hence $\varphi_X(z) = (\det M(z))^{-\alpha}$. We can derive the expectation of the entries of X by differentiating the characteristic function

$$\mathbb{E}[X_{ij}] = (-\mathrm{i})\frac{\partial \varphi_X}{\partial z_{ij}}(0), \quad i,j \in \{1,\dots,q\}. \tag{3.61}$$

As X takes values in the symmetric matrices it is sufficient to determine the expectations for the indices $i \geq j$ in $\{1,\dots,q\}$.

Inserting Eq. (3.59) and Eq. (3.60),

$$\frac{\partial \varphi_X}{\partial z_{ij}}(z) = -\alpha(\det M(z))^{-(\alpha+1)}\frac{\partial}{\partial z_{ij}}(\det M(z)) \tag{3.62}$$

For $q = 1$ and $q = 2$, the expectations can be easily computed using the simple representation of the determinant in these cases. Hence, we can restrict to $q \geq 3$ for the rest of this step.

In order to differentiate the determinant of the matrix $M(z)$ we use the Laplace expansion[10] along the i-th row for $i \in \{1,\dots,q\}$

$$\det M(z) = \sum_{m=1}^{q}(-1)^{i+m}\left(1_{\{m=i\}} - \mathrm{i}\,\frac{z_{im}+z_{mi}}{2}\right)\det(M(z)_{im}), \tag{3.63}$$

where the notation $M(z)_{im}$ refers to the matrix $M(z)$ without the i-th row and m-th column.

[10] See [14, Satz 7.4.7]

In case $i = j$, we can now differentiate the determinant of this matrix w.r.t. z_{ij},

$$\frac{\partial}{\partial z_{ii}} \det M(z) = (-1)^{i+i}(-\mathrm{i})\,\frac{1+1}{2}\,\det(M(z)_{ii}), \tag{3.64}$$

which is equal to $-\mathrm{i}$ for $z = 0$ for every $i \in \{1, \dots, q\}$.

If $i > j$ in $\{1, \dots, q\}$, then we apply the Laplace expansion a second time to Eq. (3.63), now by the j-th row:

$$\det M(z) = \sum_{m=1}^{q} (-1)^{i+m} \left(1_{\{m=i\}} - \mathrm{i}\,\frac{z_{im} + z_{mi}}{2}\right) \times \underbrace{\sum_{\substack{n=1\\ n\neq m}}^{q} (-1)^{j+n-1_{\{n>m\}}} \left(1_{\{n=j\}} - \mathrm{i}\,\frac{z_{jn} + z_{nj}}{2}\right) \det\big(M(z)_{\{i,j\}\{m,n\}}\big)}_{=\det(M(z)_{im})},$$

where $M(z)_{\{i,j\}\{m,n\}}$ stands for the matrix without the i-th and j-th row and m-th and n-th column. Now, using the product rule, we can differentiate this determinant w.r.t. z_{ij} with $i > j$ in $\{1, \dots, q\}$,

$$\begin{aligned}\frac{\partial}{\partial z_{ij}} \det M(z) &= (-1)^{i+j}(-\mathrm{i})\frac{1+0}{2} \det(M(z)_{ij}) \\ &+ (-1)^{j+i}(-\mathrm{i})\frac{0+1}{2} \underbrace{\sum_{\substack{m=1\\ m\neq i}}^{q} (-1)^{i+m}(-1)^{-1_{\{i>m\}}} \left(1_{\{m=i\}} - \mathrm{i}\,\frac{z_{im} + z_{mi}}{2}\right) \det\big(M(z)_{\{i,j\}\{m,i\}}\big)}_{=\det(M(z)_{ji})},\end{aligned} \tag{3.65}$$

where the last small equality follows from a Laplace expansion along the i-th row of $M(z)_{ji}$ using $i > j$. The terms in Eq. (3.65) are equal to zero for $z = 0$ since $\det(M(0)_{ij}) = \det((I_q)_{ij}) = 0$ for $i \neq j$ in $\{1, \dots, q\}$. Combining this with Eq. (3.61), Eq. (3.62) and Eq. (3.64),

$$\mathbb{E}[X_{ij}] = (-\mathrm{i})(-\alpha) \cdot 1 \cdot 1_{\{i=j\}}(-\mathrm{i}) = \alpha 1_{\{i=j\}}, \quad i, j \in \{1, \dots, q\},$$

where we are using $\det(M(0)) = 0$. Hence, the claimed expected values have been proven for the standard case.

Step 2. For the covariances take $i, j, k, l \in \{1, \dots, q\}$ with $i \geq j$ and $k \geq l$. Then the covariance between X_{ij} and X_{kl} is given by

$$\mathrm{Cov}(X_{ij}, X_{kl}) = \mathbb{E}[X_{ij}X_{kl}] - \mathbb{E}[X_{ij}]\mathbb{E}[X_{kl}] = (-\mathrm{i})^2 \frac{\partial^2 \varphi_X}{\partial z_{ij} \partial z_{kl}}(0) - \alpha^2 1_{\{i=j\}} 1_{\{k=l\}}, \tag{3.66}$$

using the expectations derived in the first step. Again, for $q = 1$ and $q = 2$, this can be easily computed using the simple representation of the determinant in these cases. Hence, we can restrict to $q \geq 3$ for the rest of this step.

For all other cases we can use the product rule to arrive at

$$\frac{\partial^2 \varphi_X}{\partial z_{ij}\partial z_{kl}}(z) = -\alpha\Big(-(\alpha+1)(\det M(z))^{-(\alpha+2)}\frac{\partial}{\partial z_{kl}}(\det M(z))\frac{\partial}{\partial z_{ij}}(\det M(z)) + (\det M(z))^{-(\alpha+1)}\frac{\partial^2}{\partial z_{ij}\partial z_{kl}}(\det M(z))\Big)$$

for all $i,j,k,l \in \{1,\dots,q\}$ with $i \geq j$, $k \geq l$. Apart from the last second partial derivative of $\det M(z)$, we know the values of the terms at $z=0$ from the first step, see Eq. (3.64) and Eq. (3.65), so

$$\frac{\partial^2 \varphi_X}{\partial z_{ij}\partial z_{kl}}(0) = -\alpha\Bigg(\underbrace{-(\alpha+1)(-\mathrm{i})^2}_{=\alpha+1}1_{\{i=j\}}1_{\{k=l\}} + \frac{\partial^2(\det M(\cdot))}{\partial z_{ij}\partial z_{kl}}(0)\Bigg). \tag{3.67}$$

Now for the missing partial derivative we again differentiate for $i=j$ and $i>j$ and also for the different possibilities of $k \geq l$ in $\{1,\dots,q\}$:

Case 1: The indices $i,j,k,l \in \{1,\dots,q\}$ satisfy $i=j$ and $k=l$. If $i \neq k$, then

$$\begin{aligned}\frac{\partial^2}{\partial z_{ii}\partial z_{kk}}(\det M(z)) &= (-\mathrm{i})\frac{\partial}{\partial z_{kk}}\det(M(z)_{ii}) \\ &= (-\mathrm{i})^2\underbrace{(-1)^{k-1_{\{k>i\}}+k-1_{\{k>i\}}}}_{=1}\det\big(M(z)_{\{i,k\}\{i,k\}}\big),\end{aligned}$$

where we are using Eq. (3.64) for the first equality and the Laplace expansion for the k-th row of $M(z)$ for the second equality. If $i=k$, then this second derivative has to be equal to zero as the variable $z_{kk}=z_{ii}$ is not part of $\det(M(z)_{ii})$.

Case 2: The indices $i,j,k,l \in \{1,\dots,q\}$ satisfy $i=j$ and $k>l$. Then, due to Eq. (3.64),

$$\frac{\partial^2}{\partial z_{ii}\partial z_{kl}}(\det M(z)) = (-\mathrm{i})\frac{\partial}{\partial z_{kl}}\det(M(z)_{ii}) = 0$$

because, using the Laplace expansion of the k-th and l-th row, the determinants of $M(z)_{\{i,k\}\{i,l\}}$ and $(M(z)_{\{i,l\}\{i,k\}})$ for $z=0$ must be equal to zero as $M(0)=I_q$ and $k \neq l$.

Case 3: The indices $i,j,k,l \in \{1,\dots,q\}$ satisfy $i>j$ and $k=l$. As second derivatives are symmetric, this is the same as *Case 2*.

Case 4: The indices $i,j,k,l \in \{1,\dots,q\}$ satisfy $i>j$ and $k>l$. Then, using Eq. (3.65),

$$\frac{\partial^2}{\partial z_{ij}\partial z_{kl}}(\det M(z)) = \frac{(-\mathrm{i})(-1)^{i+j}}{2}\frac{\partial}{\partial z_{kl}}\Big(\det(M(z)_{ij}) + \det(M(z)_{ji})\Big). \tag{3.68}$$

We will split this case into several subcases.

Case 4(i): The indices additionally satisfy $i=k$ and $j=l$. Consider the Laplace expansion of $\det(M(z)_{ij})$ along the l-th row,

$$\det(M(z)_{ij}) = \sum_{\substack{m=1\\m\neq j}}^{q}(-1)^{l+m-1_{\{m>j\}}}\left(1_{\{l=m\}} - \mathrm{i}\,\frac{z_{lm}+z_{ml}}{2}\right)\det(M(z)_{\{i,l\}\{j,m\}}).$$

Then, considering that $k > j$,

$$\frac{\partial}{\partial z_{kl}} \det(M(z)_{ij}) = (-1)^{l+k}(-\mathrm{i})\frac{1}{2} \underbrace{\det\big(M(z)_{\{i,l\}\{j,k\}}\big)}_{=1 \text{ for } z=0}, \tag{3.69}$$

where the last small equality follows form the fact that $M(z)_{\{i,l\}\{j,k\}} = I_{q-2}$ in this case. For $\det(M(z)_{ji})$ we look at the Laplace expansion along the k-th row and arrive, using the analogue procedure at

$$\frac{\partial}{\partial z_{kl}} \det(M(z)_{ji}) = (-1)^{k+l}(-\mathrm{i})\frac{1}{2} \underbrace{\det\big(M(z)_{\{j,k\}\{i,l\}}\big)}_{=1 \text{ for } z=0}.$$

Combining this with Eq. (3.68),

$$\frac{\partial^2(\det M(\cdot))}{\partial z_{ij}\partial z_{kl}}(0) = \frac{(-\mathrm{i})(-1)^{i+j}}{2}\left[\frac{-\mathrm{i}}{2}(-1)^{l+k} + \frac{-\mathrm{i}}{2}(-1)^{l+k}\right] = -\frac{(-\mathrm{i})^2(-1)^{i+j+k+l}}{2} = \frac{1}{2}.$$

Case 4(ii): The indices additionally satisfy $i = k$, implying $j \neq k$, and $j \neq l$. Considering the Laplace expansion of $\det(M(z)_{ij})$ along the l-th row results in, using Eq. (3.69)

$$\frac{\partial}{\partial z_{kl}} \det(M(z)_{ij}) = (-1)^{l+k}(-\mathrm{i})\frac{1}{2} \underbrace{\det\big(M(z)_{\{i,l\}\{j,k\}}\big)}_{=0 \text{ for } z=0}.$$

Consider the Laplace expansion of $\det(M(z)_{ji})$ along the k-th row, which is also the i-th row,

$$\det(M(z)_{ji}) = \sum_{\substack{m=1\\ m\neq i}}^{q} (-1)^{k+m-1_{\{m>i\}}} \left(1_{\{k=m\}} - \mathrm{i}\,\frac{z_{km}+z_{mk}}{2}\right) \det\big(M(z)_{\{j,k\}\{i,m\}}\big).$$

In case $q = 3$, it can be easily seen that the determinant of $M(z)_{\{j,k\}\{i,m\}}$ for $z = 0$ must be equal to zero as $i = k$ and $j \neq l$. For $q \geq 4$ additionally expanding along the l-th row results in

$$\begin{aligned}\det(M(z)_{ji}) = &\sum_{\substack{m=1\\ m\neq i}}^{q} (-1)^{k+m-1_{\{m>i\}}} \left(1_{\{k=m\}} - \mathrm{i}\,\frac{z_{km}+z_{mk}}{2}\right) \\ &\times \sum_{\substack{n=1\\ n\neq i, n\neq m}}^{q} (-1)^{l+n-1_{\{n>i\}}-1_{\{n>m\}}} \left(1_{\{l=n\}} - \mathrm{i}\,\frac{z_{ln}+z_{nl}}{2}\right) \det\big(M(z)_{\{j,k,l\}\{i,m,n\}}\big)\end{aligned}$$

Now differentiating w.r.t. z_{kl} gives us

$$\begin{aligned}&\frac{\partial}{\partial z_{kl}} \det(M(z)_{ji}) \\ &= (-1)^{k+l-1_{\{l>i\}}}(-\mathrm{i})\frac{1}{2} \underbrace{\det\big(M(z)_{\{j,k\}\{i,l\}}\big)}_{=0 \text{ for } z=0} + (-1)^{l+k-1_{\{k>i\}}}(-\mathrm{i})\frac{1}{2} \underbrace{\det\big(M(z)_{\{j,l\}\{i,k\}}\big)}_{=0 \text{ for } z=0}.\end{aligned} \tag{3.70}$$

Hence the derivative of both $\det(M(z)_{ij}$ and $\det(M(z)_{ji}$ w.r.t. z_{kl} is equal to zero, resulting in, using Eq. (3.68),

$$\frac{\partial^2(\det M(\cdot))}{\partial z_{ij}\partial z_{kl}}(0) = 0.$$

Case 4(iii): The indices additionally satisfy $i = l$, which implies both $j \neq l$ and $j \neq k$. This case works the same as *Case 4(ii)* above with k and l switched.

Case 4(iv): The indices satisfy $i \neq k$ and $i \neq l$. If either $j = l$ or $j = k$, then this case works in the same way as *Case 4(ii)* above with j and i switched. If $j \neq k$ and $j \neq l$, then Eq. (3.70) holds and for i and j switched Eq. (3.70) holds as well, thus also implying

$$\frac{\partial^2(\det M(\cdot))}{\partial z_{ij}\partial z_{kl}}(0) = 0.$$

with Eq. (3.68).

Summarizing all of these cases for $i, j, k, l \in \{1, \dots, q\}$ with $i \geq j$ and $k \geq l$

$$\frac{\partial^2(\det M(\cdot))}{\partial z_{ij}\partial z_{kl}}(0) = \begin{cases} (-\mathrm{i})^2 = -1 & \text{if } i = j \neq k = l, \\ \frac{1}{2} & \text{if } i = k > j = l, \\ 0 & \text{otherwise.} \end{cases}$$

Combining this with Eq. (3.66) and Eq. (3.67)

$$\operatorname{Cov}(X_{ij}, X_{kl}) = \begin{cases} (-\mathrm{i})^2(-\alpha)(\alpha+1) - \alpha^2 = \alpha & \text{if } i = j = k = l, \\ (-\mathrm{i})^2(-\alpha)\big((\alpha+1) - 1\big) - \alpha^2 = 0 & \text{if } i = j \neq k = l, \\ (-\mathrm{i})^2(-\alpha)\frac{1}{2} - 0 = \frac{\alpha}{2} & \text{if } i = k > j = l, \\ 0 & \text{otherwise,} \end{cases}$$

for all $i, j, k, l \in \{1, \dots, q\}$ with $i \geq j$ and $k \geq l$, corresponding to Eq. (3.56) for $C = I_q$. Using matrix notation and the symmetry of X this is equivalent to

$$\operatorname{Cov}(\operatorname{vec}(X)) = \frac{\alpha}{2}(K_q + I_{q^2}) \tag{3.71}$$

for X following the standard q-dimensional matrix-valued gamma distribution with parameter α as discussed in Remark 3.61.

Step 3. As discussed in the beginning, every matrix $Y \sim \operatorname{MGamma}(\alpha, C)$ has a stochastic representation $Y \stackrel{\mathrm{d}}{=} AXA^{\mathsf{T}}$, where $X \sim \operatorname{MGamma}(\alpha, I_{\operatorname{rk}(C)})$ either follows a standard matrix-valued gamma distribution or a singular Wishart distribution. As the expectation is linear,

$$\mathbb{E}[Y] = \mathbb{E}[AXA^{\mathsf{T}}] = A\,\mathbb{E}[X]\,A^{\mathsf{T}} = A\,\alpha I_p\,A^{\mathsf{T}} = \alpha C,$$

which proves Eq. (3.55). For the covariance we can write

$$\operatorname{Cov}(\operatorname{vec}(Y)) = \operatorname{Cov}(\operatorname{vec}(AXA^{\mathsf{T}})) = \operatorname{Cov}((A \otimes A)\operatorname{vec}(X)),$$

where the last equality follows from Eq. (3.52). Since the covariance is bilinear, we can rewrite this equation to

$$\operatorname{Cov}(\operatorname{vec}(Y)) = (A \otimes A)\operatorname{Cov}(\operatorname{vec}(X)) \underbrace{(A \otimes A)^{\mathsf{T}}}_{= A^{\mathsf{T}} \otimes A^{\mathsf{T}}},$$

where the last small equality follows from the Definition 3.1 of the Kronecker product. Using the result Eq. (3.71) from the second step,

$$\operatorname{Cov}(\operatorname{vec}(Y)) = (A \otimes A)\frac{\alpha}{2}\big(K_{\operatorname{rk}(C)} + I_{(\operatorname{rk}(C))^2}\big)(A^{\mathsf{T}} \otimes A^{\mathsf{T}}).$$

With the help of Eq. (3.54) and Eq. (3.53) we arrive at

$$\begin{aligned}\operatorname{Cov}(\operatorname{vec}(Y)) = \frac{\alpha}{2}(K_p + I_{p^2})(A \otimes A)(A^{\mathsf{T}} \otimes A^{\mathsf{T}}) &= \frac{\alpha}{2}(K_p + I_{p^2})(AA^{\mathsf{T}} \otimes AA^{\mathsf{T}}) \\ &= \frac{\alpha}{2}(K_p + I_{p^2})(C \otimes C),\end{aligned}$$

which proves the theorem. □

3.6. Identification of some Biased Distributions

This section explores how matrix-valued gamma distributions change under certain biased measures. The one-dimensional version of the results given in Theorem 3.64 can also be found in [29, Lemma 4.35].

Definition 3.62 (Biased probability measure, see [29, Definition 2.10])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X be a non-negative random variable on this probability space, satisfying $0 < \mathbb{E}[X] < \infty$. Then the X-biased probability measure $\mathbb{P}_X$ on the measurable space $(\Omega, \mathcal{F})$ is defined by

$$\mathbb{P}_X[A] := \frac{\mathbb{E}[X \mathbb{1}_A]}{\mathbb{E}[X]}, \quad A \in \mathcal{F}. \tag{3.72}$$

Remark 3.63. In the context of Definition 3.62 above, for $c \in \mathbb{R}^+$ the cX-biased probability measure $\mathbb{P}_{cX}$ is the same as $\mathbb{P}_X$ as the expectation $\mathbb{E}[cX] = c\mathbb{E}[X]$ satisfies $0 < \mathbb{E}[cX] < \infty$ and

$$\mathbb{P}_{cX}[A] = \frac{\mathbb{E}[cX \mathbb{1}_A]}{\mathbb{E}[cX]} = \frac{\mathbb{E}[X \mathbb{1}_A]}{\mathbb{E}[X]} = \mathbb{P}_X[A], \quad A \in \mathcal{F}.$$

Theorem 3.64 (Biased matrix-valued gamma distributions)**.** *Assume that the random matrix $Y \sim \operatorname{MGamma}(\alpha, C)$ with shape parameter $\alpha \in \mathcal{G}_C$ and positive semi-definite scale parameter matrix $C = AA^{\mathsf{T}}$, where $A \in \mathbb{R}^{p \times q}$ for $q = \max\{1, \operatorname{rk}(C)\}$ is as discussed in Remark 3.27. Let $D \in \mathbb{R}^{p \times p}$ be a symmetric matrix fulfilling*[11] *$I_p + A^{\mathsf{T}}DA \in \mathbb{S}_p^+$. Let the scalar β fulfill one of the two conditions*

[11] This is in particular fulfilled by all positive semi-definite matrices D.

(i) $\beta > \frac{p-1}{2} - \alpha$ *and* $\alpha > \frac{p-1}{2}$ *and* C *is positive definite or*

(ii) $\beta = 0$.

Then, using the biased probability measure given in Definition 3.62, the distribution of Y *under the* $(\det Y)^\beta \exp(-\mathrm{tr}(DY))$*-biased measure* $\mathbb{P}$ *is the matrix-valued gamma distribution with shape parameter* $\alpha + \beta$ *and the positive definite scale parameter matrix* $A^\top(I_p + A^\top DA)^{-1}A$*, formally*

$$\mathbb{P}_{(\det Y)^\beta \exp(-\mathrm{tr}(DY))} Y^{-1} = \mathrm{MGamma}(\alpha + \beta, A^\top(I_p + A^\top DA)^{-1}A). \tag{3.73}$$

If C *has full rank this scale parameter matrix can be rewritten as* $(C^{-1} + D)^{-1}$.

Proof. Note that

$$\mathbb{E}[(\det Y)^\beta \, \mathrm{e}^{-\mathrm{tr}(DY)}] = \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)} (\det C)^\beta (\det(I_p + DC))^{-(\alpha+\beta)},$$

where the equality follows from Theorem 3.32 or Theorem 3.14, in case $\alpha \le (p-1)/2$ implying $\beta = 0$ by the conditions of this theorem, and the fact that $I_p + A^\top DA$ is positive definite. Note that this expectation is greater than zero for either of the conditions β can fulfill, using the convention $0^0 = 1$.

Combining this expectation with the definition of the biased probability measure given in Eq. (3.72) a density of the $(\det Y)^\beta \exp(-\mathrm{tr}(DY))$-biased probability measure w.r.t. $\mathbb{P}$ is given by

$$\frac{\mathrm{d}\mathbb{P}_{(\det Y)^\beta \exp(-\mathrm{tr}(DY))}}{\mathrm{d}\mathbb{P}} = \frac{\Gamma_p(\alpha)}{\Gamma_p(\alpha+\beta)} \frac{(\det(I_p + DC))^{\alpha+\beta}}{(\det C)^\beta} (\det Y)^\beta \, \mathrm{e}^{-\mathrm{tr}(DY)}. \tag{3.74}$$

To determine the distribution of Y w.r.t. $\mathbb{P}_{(\det Y)^\beta \exp(-\mathrm{tr}(DY))}$ we calculate its characteristic function

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_{(\det Y)^\beta \exp(-\mathrm{tr}(DY))}}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)}] &= \mathbb{E}_{\mathbb{P}}\left[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)} \frac{\mathrm{d}\mathbb{P}_{(\det \mathbb{Y})^\beta \exp(-\mathrm{tr}(\mathbb{D}\mathbb{Y}))}}{\mathrm{d}\mathbb{P}}\right] \\ &= \frac{\Gamma_p(\alpha)}{\Gamma_p(\alpha+\beta)} \frac{(\det(I_p + DC))^{\alpha+\beta}}{(\det C)^\beta} \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)} (\det Y)^\beta \, \mathrm{e}^{-\mathrm{tr}(DY)}]\end{aligned}$$

for $z \in \mathbb{R}^{p\times p}$ symmetric. Since $I_p + A^\top DA$ is positive definite, we can apply Theorem 3.32 with $w = D - \mathrm{i}z$ and $M = I_p$ for the integral and arrive at

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_{(\det Y)^\beta \exp(-\mathrm{tr}(DY))}}[\mathrm{e}^{\mathrm{tr}(\mathrm{i}zY)}] \\ &= \frac{\Gamma_p(\alpha)}{\Gamma_p(\alpha+\beta)} \frac{(\det(I_p + DC))^{\alpha+\beta}}{(\det C)^\beta} \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)} (\det C)^\beta (\det(I_p + (D - \mathrm{i}z)C))^{-(\alpha+\beta)} \\ &= \left(\frac{\det(I_p + DC)}{\det(I_p + (D - \mathrm{i}z)C)}\right)^{\alpha+\beta} = \left(\frac{\det(I_p + A^\top DA)}{\det(I_p + A^\top DA - \mathrm{i}A^\top zA)}\right)^{\alpha+\beta},\end{aligned} \tag{3.75}$$

where the fact that $C = AA^\top$ and the Weinstein–Aronszajn identity, given in Lemma 2.3, is applied to the numerator and denominator in the last step. Since $I_p + A^\top DA$ is invertible,

$$(I_p + A^\top DA - \mathrm{i}A^\top zA)(I_p + A^\top DA)^{-1} = I_p - \mathrm{i}A^\top zA(I_p + A^\top DA)^{-1}.$$

Hence, Eq. (3.75) can be rewritten as

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_{(\det Y)^\beta \exp(-\operatorname{tr}(DY))}}[\mathrm{e}^{\operatorname{tr}(\mathrm{i}zY)}] &= \big(\det(I_p - \mathrm{i}A^\top zA(I_p + A^\top DA)^{-1})\big)^{-(\alpha+\beta)}\\ &= \big(\det(I_q - \mathrm{i}zA(I_p + A^\top DA)^{-1}A^\top)\big)^{-(\alpha+\beta)},\end{aligned}$$

where the Weinstein–Aronszajn identity is again used for the last step. As discussed in Remark 3.4, the characteristic function determines the distribution uniquely, hence the distribution of Y under $\mathbb{P}_{(\det Y)^\beta \exp(-\operatorname{tr}(DY))}$ is $\mathrm{MGamma}(\alpha + \beta, A^\top(I_p + A^\top DA)^{-1}A)$.

In the case that C is invertible, $A \in \mathbb{R}^{p\times p}$ is invertible as well and the scale parameter matrix can be rewritten as

$$\begin{aligned}A^\top(I_p + A^\top DA)^{-1}A &= (A^{-1}(I_p + A^\top DA)(A^\top)^{-1})^{-1}\\ &= ((AA^\top)^{-1} + D)^{-1} = (C^{-1} + D)^{-1}.\end{aligned}$$

□

4. Matrix-Valued Gamma Distributions in Poisson Mixture Models

4.1. Poisson Mixture Models

A mixture distribution is a probability distribution of one or multiple random variables, whose parameters are determined by other random variables. One of the most prominent examples is a gamma-mixed Poisson distribution, where the random variable $\Lambda \sim \Gamma(\alpha, \beta)$ and the conditional distribution of N given Λ is Poisson(Λ). Then N is said to follow a gamma-mixed Poisson distribution. Subsequently, we will demonstrate the well-known fact that such a random variable N follows a negative binomial distribution.

In order to achieve this, we start by defining the probability-generating function, a very important tool for discrete (mixture) distributions.

Definition 4.1 (Probability-generating function)**.** Fix $d \in \mathbb{N}$. For an $\mathbb{N}_0^d$-valued random vector $N = (N_1, \dots, N_d)$ the probability-generating function is defined by

$$\varphi_N(s) = \mathbb{E}\bigg[\prod_{i=1}^{d} s_i^{N_i}\bigg], \quad s = (s_1, \dots, s_d) \in \mathbb{C}^d,$$

where the expectation exists at least for all s with $\|s\|_\infty \le 1$. The probability-generating function belongs to the distribution of N and not to the random vector N itself but for a simpler notation we will still denote it with φ_N.

Remark 4.2 (The probability-generating function determines a distribution uniquely, see [29, Eq. (4.1), Eq. (4.15)]). Note that for an $\mathbb{N}_0^d$-valued random vector $N = (N_1, \dots, N_n)$

$$\varphi_N^{(n)}(0, \dots, 0) = n_1! \cdots n_d!\, \mathbb{P}[N = n], \quad n = (n_1, \dots, n_d) \in \mathbb{N}_0^d \tag{4.1}$$

holds, where $\varphi_N^{(n)}$ refers to the function φ_N partially differentiated n_i times in the i-th component[1] for all $i \in \{1, \dots, d\}$. Hence, the probability-generating function describes one unique distribution. Conversely, the probability-generating function is uniquely defined by Definition 4.1 for each $\mathbb{N}_0^d$-valued random vector N.

Definition 4.3 (Poisson distributions, see [29, Definition 3.1])**.** For $\lambda \ge 0$ an $\mathbb{N}_0$-valued random variable N is said to follow the Poisson distribution with intensity λ if

$$\mathbb{P}[N = n] = \frac{\lambda^n}{n!}\, \mathrm{e}^{-\lambda}, \quad n \in \mathbb{N}_0,$$

using the convention $0^0 = 1$. This is denoted by $N \sim \text{Poisson}(\lambda)$.

[1] Note that the order in which the partial derivatives are taken does not matter, since the zero vector is part of the interior of the set for which φ_N is well-defined and C^∞.

Definition 4.4 (Negative binomial distributions, see [29, Eq. (4.60)])**.** An $\mathbb{N}_0$-valued random variable N follows the negative binomial distribution with parameters $\alpha > 0$ and $p \in (0,1)$ if

$$\mathbb{P}[N = n] = \binom{n+\alpha-1}{n}(1-p)^\alpha p^n, \quad n \in \mathbb{N}_0.$$

We will use the notation $N \sim \mathrm{NegBin}(\alpha, p)$. We will also include the cases $\alpha = 0$ and $p = 0$ by letting $\mathrm{NegBin}(0, p)$ and $\mathrm{NegBin}(\alpha, 0)$ denote the degenerate distribution concentrated in zero.

The negative multinomial distribution, a generalization of the negative binomial distribution, is given in Definition 4.23 below.

Example 4.5 (Some probability-generating functions, see [29, Example 4.3, Eq. (4.63)])**.** Some examples of probability-generating functions for $\mathbb{N}_0$-valued random variables include

(i) *Poisson distribution:* For $N \sim \mathrm{Poisson}(\lambda)$ with parameter $\lambda \geq 0$,

$$\varphi_N(s) = \mathrm{e}^{\lambda(s-1)}, \quad s \in \mathbb{C}. \tag{4.2}$$

(ii) *Negative binomial distribution:* For $N \sim \mathrm{NegBin}(\alpha, p)$ with parameters $\alpha \geq 0$ and $p \in [0, 1)$,

$$\varphi_N(s) = \left(\frac{1-p}{1-ps}\right)^\alpha, \quad s \in \mathbb{C} \text{ with } p|s| < 1, \tag{4.3}$$

where the complex exponentiation is understood as discussed in Remark 2.31.

Lemma 4.6 (Affine transformations of probability-generating functions, see [29, Lemma 4.6])**.** *For $c, d \in \mathbb{N}$, let N be an $\mathbb{N}_0^d$-valued random vector and let $A \in \mathbb{N}_0^{c\times d}$ be a matrix. Then the probability generating function of the $\mathbb{N}_0^c$-valued random vector AN is given by*

$$\varphi_{AN}(s_1, \ldots, s_c) = \varphi_N(t_1, \ldots, t_d), \quad \textit{where} \quad t_j := \prod_{i=1}^{c} s_i^{a_{i,j}}, \; j \in \{1, \ldots, d\}$$

for all $s \in \mathbb{C}^c$ satisfying $\|s\|_\infty \leq 1$.

Corollary 4.7 (Properties of the probability-generating functions, see [29, Example 4.7])**.** *For $d \in \mathbb{N}$ let N be an $\mathbb{N}_0^d$-valued random vector. Then the followings properties hold*

(i) For $c \in \{1, \ldots, n\}$ the probability-generating function of the first c components of N is given by

$$\varphi_{(N_1,\ldots,N_c)}(s_1, \ldots, s_c) = \varphi_N((s_1, \ldots, s_c, 1, \ldots, 1), \quad s \in \mathbb{C}^c \text{ with } \|s\|_\infty \leq 1. \tag{4.4}$$

(ii) If the last $d-c+1$ of N are aggregated for $c \in \{2, \ldots, d\}$, then the probability-generating function is given by

$$\varphi_{(N_1,\ldots,N_{c-1},N_c+\cdots+N_d)}(s_1, \ldots, s_c) = \varphi_N(s_1, \ldots, s_{c-1}, s_c, \ldots, s_c) \tag{4.5}$$

for all $s \in \mathbb{C}^c$ with $\|s\|_\infty \leq 1$.

(iii) For every permutation σ of $\{1, \ldots, d\}$,

$$\varphi_{(N_{\sigma(1)},\ldots,N_{\sigma(d)})}(s_1, \ldots, s_d) = \varphi_N(s_{\sigma^{-1}(1)}, \ldots, s_{\sigma^{-1}(d)}), \quad s \in \mathbb{C}^d \text{ with } \|s\|_\infty \leq 1. \tag{4.6}$$

Lemma 4.8 (Characterization of independence using probability-generating functions, see [29, Lemma 4.14])**.** *For $d_1, \dots, d_n \in \mathbb{N}$ let $N^1, \dots, N^n$ be random vectors, where N^i takes values in $\mathbb{N}_0^{d_i}$ for every $i \in \{1, \dots, n\}$. Then $N^1, \dots, N^n$ are pairwise independent if and only if*

$$\varphi_{(N^1,\dots,N^n)}(s^1, \dots, s^n) = \prod_{i=1}^{n} \varphi_{N^i}(s^i), \quad s^i \in \mathbb{C}^{d_i} \textit{ with } \|s^i\|_\infty \le 1 \text{ for all } i \in \{1, \dots, n\}.$$

Lemma 4.9 (Summation property of probability-generating functions for independent random variables, see [29, Theorem 4.15])**.** *For $d \in \mathbb{N}$ let $N^1, \dots, N^n$ be independent $\mathbb{N}_0^d$-valued random vectors. Then*

$$\varphi_{N^1+\dots+N^n}(s) = \prod_{i=1}^{n} \varphi_{N^i}(s), \quad s \in \mathbb{C}^d \textit{ with } \|s\|_\infty \le 1.$$

Corollary 4.10 (Summation property of negative binomial distributions)**.** *For $n \in \mathbb{N}$, let $N_k \sim \mathrm{NegBin}(\alpha_k, p)$ for $k \in \{1, \dots, n\}$ be independent random variables, where $p \in [0, 1)$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}_0^+$. Then*

$$\sum_{k=1}^{n} N_k \sim \mathrm{NegBin}(\alpha_1 + \ldots + \alpha_n, p).$$

Proof. For a proof by induction, it is sufficient to consider the case $n = 2$. As N_1 and N_2 are independent and their probability-generating functions are given in Eq. (4.3), Lemma 4.9 leads to

$$\varphi_{N_1+N_2}(s) = \varphi_{N_1}(s)\varphi_{N_2}(s) = \left(\frac{1-p}{1-ps}\right)^{\alpha_1}\left(\frac{1-p}{1-ps}\right)^{\alpha_2} = \left(\frac{1-p}{1-ps}\right)^{\alpha_1+\alpha_2}$$

for every $s \in \mathbb{C}$ with $p|s| < 1$. As discussed in Remark 4.2, the probability-generating function determines the distribution uniquely, hence $N_1 + N_2 \sim \mathrm{NegBin}(\alpha_1 + \alpha_2, p)$. □

Using the probability-generating function, we can prove the fact that gamma-mixed Poisson distributions are part of the class of negative binomial distributions.

Lemma 4.11 (Gamma-mixed Poisson distributions)**.** *For $\alpha, \beta > 0$, let the random variable Λ follow the one-dimensional gamma distribution with shape parameter α and scale parameter β. Let the random variable*

$$N|\Lambda \sim \mathrm{Poisson}(\Lambda).$$

Then N follows a negative binomial distribution with parameter α and $p = \frac{\beta}{1+\beta}$.

Proof. For every $s \in \mathbb{C}$ satisfying $|s| \le 1$ the probability-generating function of N is given by

$$\varphi_N(s) = \mathbb{E}[s^N] = \mathbb{E}[\mathbb{E}[s^N|\Lambda]] = \mathbb{E}[\mathrm{e}^{\Lambda(s-1)}] = (1 + (1-s)\beta)^{-\alpha}$$

where the third equality follows from Eq. (4.2) and the fourth one from the exponential moments of gamma distributions. This can be, for example seen from Theorem 3.32 with

$p = 1$, $M = 1$, $w = (1-s)$, the β from the lemma equal to zero and $Y \sim \text{Gamma}(\alpha, \beta)$. Since

$$(1 + (1-s)\beta)^{-\alpha} = \left(\frac{1}{1+\beta-s\beta}\right)^{\alpha} = \left(\frac{\frac{1}{1+\beta}}{\frac{1+\beta-s\beta}{1+\beta}}\right)^{\alpha} = \left(\frac{1-p}{1-sp}\right)^{\alpha},$$

Eq. (4.3) and the fact that the probability-generating function determines a distribution uniquely as discussed in Remark 4.2, $N \sim \text{NegBin}(\alpha, p)$. □

Definition 4.12 (Multivariate Poisson distributions, see [29, Definition 3.40])**.** For $m \in \mathbb{N}$, let the set $G \subseteq \mathcal{P}(\{1, \ldots, m\})$ fulfil $\varnothing \notin G$. For every $g \in G$ let N_g denote a random variable following the Poisson distribution with parameter $\lambda_g \in [0, \infty)$. If the set $(N_g)_{g \in G}$ is independent, then the $\mathbb{N}_0^m$-valued random variable

$$N := \sum_{g \in G} c_g N_g,$$

where the entries of the vector $c_g = (c_{g,1}, \ldots, c_{g,m})$ are given by $c_{g,i} = \mathbb{1}_g(i)$ for every $i \in \{1, \ldots, m\}$ and every $g \in G$, is said to follow the *m-variate Poisson distribution* $\text{MPoisson}(G, \lambda, m)$, where $\lambda = (\lambda_g)_{g \in G}$.

Using this distribution family we can define general multivariate Poisson mixture models.

Definition 4.13 (General multivariate Poisson mixture models, see [29, Section 3.6])**.** Let $(\Lambda_g)_{g \in G}$ denote a collection of $[0, \infty)$-valued random variables. In the setting of Definition 4.12, let N be distributed with a m-variate Poisson distribution where the parameters $(\lambda_g)_{g \in G} = \lambda$ are given by $\Lambda := (\Lambda_g)_{g \in G}$, i.e.

$$N | \Lambda \sim \text{MPoisson}(G, \Lambda, m),$$

with the additional assumptions that for each $g \in G$

$$\mathbb{P}[N_g = n_g \,|\, (\Lambda_h)_{h \in G}] \overset{\text{a.s.}}{=} \mathbb{P}[N_g = n_g \,|\, \Lambda_g]$$

for every $(n_g)_{g \in G} \in \mathbb{N}_0^G$ and that the variables $(N_g)_{g \in G}$ are conditionally independent, i.e.

$$\mathbb{P}[N_g = n_g \text{ for all } g \in G \,|\, (\Lambda_h)_{h \in G}] \overset{\text{a.s.}}{=} \prod_{g \in G} \mathbb{P}[N_g = n_g \,|\, (\Lambda_h)_{h \in G}].$$

The rest of this section discusses some biased distributions of negative binomial distributions, where we first prove that the concept of the biased probability measure defined in Definition 3.62 can be applied iteratively, which will be used subsequently.

Lemma 4.14 (Iterative biased probability measures)**.** *Let X and Y be non-negative random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where X satisfies $0 < \mathbb{E}[X] < \infty$. Furthermore suppose that $0 < \mathbb{E}[XY] < \infty$, which in this setting is equivalent to assuming $0 < \mathbb{E}_{\mathbb{P}_X}[Y] < \infty$. Then the XY-biased probability measure $\mathbb{P}_{XY}$ on the measurable space $(\Omega, \mathcal{F})$ is given by*

$$\mathbb{P}_{XY}[A] = \frac{\mathbb{E}_{\mathbb{P}_X}[Y \mathbb{1}_A]}{\mathbb{E}_{\mathbb{P}_X}[Y]}, \quad A \in \mathcal{F},$$

where $\mathbb{P}_X$ refers to the X-biased probability measure. Equivalently, a density of $\mathbb{P}_{XY}$ w.r.t. $\mathbb{P}$ is given by

$$\frac{\mathrm{d}\mathbb{P}_{XY}}{\mathrm{d}\mathbb{P}} \overset{\text{a.s.}}{=} \frac{\mathrm{d}(\mathbb{P}_X)_Y}{\mathrm{d}\mathbb{P}} \overset{\text{a.s.}}{=} \frac{\mathrm{d}(\mathbb{P}_X)_Y}{\mathrm{d}\mathbb{P}_X}\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}}.$$

Proof. Using Eq. (3.72) a density of the X-biased probability measure $\mathbb{P}_X$ w.r.t. $\mathbb{P}$ is given by

$$\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}} = \frac{X}{\mathbb{E}[X]}. \tag{4.7}$$

Using the same definition a density of the biased probability measure $(\mathbb{P}_X)_Y$ w.r.t. $\mathbb{P}_X$ is given by

$$\frac{\mathrm{d}(\mathbb{P}_X)_Y}{\mathrm{d}\mathbb{P}_X} = \frac{Y}{\mathbb{E}_{\mathbb{P}_X}[Y]}. \tag{4.8}$$

Using Eq. (4.7),

$$\mathbb{E}_{\mathbb{P}_X}[Y] = \mathbb{E}\left[Y\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}}\right] = \frac{\mathbb{E}[XY]}{\mathbb{E}[X]}, \tag{4.9}$$

which also proves the equivalence of assuming $0 < \mathbb{E}[XY] < \infty$ and assuming $0 < \mathbb{E}_{\mathbb{P}_X}[Y] < \infty$. Furthermore, this leads to

$$\frac{\mathrm{d}(\mathbb{P}_X)_Y}{\mathrm{d}\mathbb{P}} \overset{\text{a.s.}}{=} \frac{\mathrm{d}(\mathbb{P}_X)_Y}{\mathrm{d}\mathbb{P}_X}\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}} = \frac{XY}{\mathbb{E}[XY]},$$

where the second equation follows from combining Eqs. (4.7), (4.8) and (4.9). As the last term is equal to a density of the XY-biased probability measure $\mathbb{P}_{XY}$ w.r.t. $\mathbb{P}$ per Definition 3.62, this proves the lemma. □

Lemma 4.15 (Biased negative binomial distributions). *For $\alpha > 0$ and $p \in [0,1)$ let $N \sim \mathrm{NegBin}(\alpha, p)$. Then the following statements hold:*

(i) For every $k \in \mathbb{N}_0$,

$$\mathbb{E}[(N+\alpha)(N+\alpha+1)\cdots(N+\alpha+k)] = \frac{\alpha(\alpha+1)\cdots(\alpha+k)}{(1-p)^{k+1}} \tag{4.10}$$

and

$$\mathbb{P}_{(N+\alpha)(N+\alpha+1)\cdots(N+\alpha+k)}N^{-1} = \mathrm{NegBin}(\alpha+k+1, p). \tag{4.11}$$

(ii) Furthermore, if $p > 0$, then for every $k \in \mathbb{N}_0$,

$$\mathbb{E}[N(N-1)\cdots(N-k)\,] = p^{k+1}\alpha(\alpha+1)\cdots(\alpha+k), \tag{4.12}$$

and

$$\mathbb{P}_{N(N-1)\cdots(N-k)}N^{-1} = \mathcal{L}(\tilde{N}+k+1), \text{ where } \tilde{N} \sim \mathrm{NegBin}(\alpha+k+1, p). \tag{4.13}$$

(iii) Finally, if $p > 0$, then for every $\tilde{p} \in (0,1)$,

$$\mathbb{E}\left[\left(\frac{\tilde{p}}{p}\right)^N\right] = \left(\frac{1-p}{1-\tilde{p}}\right)^\alpha, \tag{4.14}$$

and

$$\mathbb{P}_{(\tilde{p}/p)^N} N^{-1} = \mathrm{NegBin}(\alpha, \tilde{p}). \tag{4.15}$$

Proof. Item (i): For $k \in \mathbb{N}_0$ and $N \sim \mathrm{NegBin}(\alpha, p)$ we can compute the expectation

$$\begin{aligned}&\mathbb{E}[s^N (N+\alpha)(N+\alpha+1)\cdots(N+\alpha+k)]\\ &\qquad = \sum_{n=0}^\infty s^n (n+\alpha)(n+\alpha+1)\cdots(n+\alpha+k) \underbrace{\binom{n+\alpha-1}{n}(1-p)^\alpha p^n}_{=\mathbb{P}[N=n]},\end{aligned}$$

where $s \in \mathbb{C}$ with $|s| \le 1$. Using the definition of the binomial coefficient this expectation can be rewritten as

$$\begin{aligned}&\mathbb{E}[s^N (N+\alpha)(N+\alpha+1)\cdots(N+\alpha+k)\,]\\ &\quad = \sum_{n=0}^\infty s^n \alpha(\alpha+1)\cdots(\alpha+k)\binom{n+(\alpha+k+1)-1}{n}(1-p)^\alpha p^n\\ &\quad = \frac{\alpha(\alpha+1)\cdots(\alpha+k)}{(1-p)^{k+1}} \underbrace{\sum_{n=0}^\infty s^n \binom{n+(\alpha+k+1)-1}{n}(1-p)^{\alpha+k+1}p^n}_{=\varphi_{\mathrm{NegBin}(\alpha+k+1,p)}(s)},\end{aligned}$$

where the last small equality follows from Definition 4.4. Inserting $s = 1$ leads to Eq. (4.10) as the value of a probability-generating function for $s = 1$ is always equal to one. Combining this with Definition 4.1 and Definition 3.62, we can see that the probability-generating function of N under $\mathbb{P}_{(N+\alpha)\cdots(N+\alpha+k)}$ is equal to the probability-generating function of the $\mathrm{NegBin}(\alpha+k+1, p)$ distribution, which proves Eq. (4.11) due to Remark 4.2.

Item (ii): Let $k \in \mathbb{N}_0$ and $N \sim \mathrm{NegBin}(\alpha, p)$. Then for $s \in \mathbb{C}$ with $|s| \le 1$

$$\mathbb{E}[N(N-1)\cdots(N-k)\, s^N\,] = \sum_{n=0}^\infty n(n-1)\cdots(n-k)\, s^n \binom{n+\alpha-1}{n}(1-p)^\alpha p^n.$$

Since the first $k+1$ summands are equal to zero we can do an index shift of the sum,

$$\begin{aligned}&\mathbb{E}[N(N-1)\cdots(N-k)\, s^N\,]\\ &\quad = \sum_{n=0}^\infty (n+k+1)(n+k)\cdots(n+1)\, s^{n+k+1}\binom{n+k+1+\alpha-1}{n+k+1}(1-p)^\alpha p^{n+k+1}\\ &\quad = p^{k+1}\alpha(\alpha+1)\cdots(\alpha+k)\underbrace{\sum_{n=0}^\infty s^{n+k+1}\binom{n+(\alpha+k+1)-1}{n}(1-p)^\alpha p^n}_{=\mathbb{E}[s^{\tilde{N}+k+1}]},\end{aligned}$$

where the second equality follows from the properties of the binomial coefficient and $\tilde{N}$ in the last equality follows the $\mathrm{NegBin}(\alpha+k+1,p)$ distribution. Inserting $s=1$ leads to Eq. (4.12) as $\mathbb{E}[1^{\tilde{N}+k+1}]=1$. Combining this with Theorem 4.20 and Definition 3.62, the probability-generating function of N under $\mathbb{P}_{N(N-1)\cdots(N-k)}$ is equal to the the probability-generating function of $\tilde{N}+k+1$, which proves Eq. (4.13) using Remark 4.2.

Item (iii): For every $\tilde{p}\in(0,1)$ and every $s\in\mathbb{C}$ with $|s|\le 1$

$$\mathbb{E}\Big[\Big(\frac{s\tilde{p}}{p}\Big)^N\Big]=\Big(\frac{1-p}{1-p(s\tilde{p}/p)}\Big)^\alpha=\Big(\frac{1-p}{1-s\tilde{p}}\Big)^\alpha,$$

where the first equation follows from Eq. (4.3). Inserting $s=1$ leads to Eq. (4.14). Hence, we can see that the probability-generating function of N under $\mathbb{P}_{(\tilde{p}/p)^N}$ for every $s\in\mathbb{C}$ with $|s|\le 1$ as

$$\mathbb{E}_{\mathbb{P}_{(\tilde{p}/p)^N}}[s^N]=\frac{\mathbb{E}\big[s^N\big(\frac{\tilde{p}}{p}\big)^N\big]}{\mathbb{E}\big[\big(\frac{\tilde{p}}{p}\big)^N\big]}=\frac{\big(\frac{1-p}{1-s\tilde{p}}\big)^\alpha}{\big(\frac{1-p}{1-\tilde{p}}\big)^\alpha}=\Big(\frac{1-\tilde{p}}{1-s\tilde{p}}\Big)^\alpha,$$

which proves Eq. (4.15) due to Remark 4.2. □

Corollary 4.16. *For $\alpha>0$ and $p\in(0,1)$ let $N\sim\mathrm{NegBin}(\alpha,p)$. For $k,l\in\mathbb{N}_0\cup\{-1\}$ and $\tilde{p}\in(0,1)$, let*

$$\Lambda:=\Big(\prod_{i=0}^{k}(N-i)\Big)\Big(\prod_{j=0}^{l}(N+\alpha+j)\Big)(\tilde{p}/p)^N.$$

Then

$$\mathbb{P}_\Lambda N^{-1}=\mathcal{L}(\tilde{N}+k+1),\ \textit{where}\ \tilde{N}\sim\mathrm{NegBin}(\alpha+k+l+2,\tilde{p}). \tag{4.16}$$

Proof. This corollary follows from combining Lemma 4.15 with Lemma 4.14: First, note that due to Eq. (4.15) N follows the $\mathrm{NegBin}(\alpha,\tilde{p})$ distribution under $\mathbb{P}_{(\tilde{p}/p)^N}$. Hence, if $l\ge 0$, then

$$\mathbb{E}_{\mathbb{P}_{(\tilde{p}/p)^N}}\Big[\prod_{j=0}^{l}(N+\alpha+j)\Big]\in(0,\infty)$$

due to Item (i). Hence, with Eq. (4.11) and Lemma 4.14

$$\mathbb{P}_{\big(\prod_{j=0}^{l}(N+\alpha+j)\big)(\tilde{p}/p)^N}N^{-1}=\mathrm{NegBin}(\alpha+l+1,\tilde{p}),$$

which also holds true for $l=-1$. Thus, Eq. (4.16) is proven for $k=-1$. Using Item (ii), if $k\ge 0$,

$$\mathbb{E}_{\mathbb{P}_{(\prod_{j=0}^{l}(N+\alpha+j))(\tilde{p}/p)^N}}\Big[\prod_{i=0}^{k}(N-i)\Big]\in(0,\infty)$$

and therefore with Eq. (4.13) and Lemma 4.14, Eq. (4.16) follows. □

4.2. Matrix-Gamma Poisson Mixture Models

The following definition introduces a gamma-mixed Poisson model, where the gamma distributed components are derived from one common matrix-valued gamma distribution. Throughout this section several properties of this distribution family will be discussed.

Definition 4.17 (Matrix-gamma Poisson mixture models)**.** For $n, p \in \mathbb{N}$, let the random matrix $Y \sim \mathrm{MGamma}(\alpha, C)$ with positive semi-definite scale parameter matrix $C \in \mathbb{R}^{p\times p}$ and shape parameter $\alpha \in \mathcal{G}_C$, see Definition 3.34. For $v_i \in \mathbb{R}^p$, let the random variable $\Lambda_i \coloneqq \langle v_i, Yv_i\rangle$ for every $i \in \{1, \dots, n\}$. Let $N_1, \dots, N_n$, where $N_i|\Lambda_i \sim \mathrm{Poisson}(\Lambda_i)$ for every $i \in \{1, \dots, n\}$, be conditionally independent given Y. Then the distribution of $(N_1, \dots, N_n)$ will be denoted by $\mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$.

Remark 4.18 (Matrix-gamma Poisson mixture models are not identifiable). A given matrix-gamma Poisson mixture model does not determine the parameters $\alpha, C, v_1, \dots, v_n$ uniquely as will be discussed below in Lemma 4.21. Hence, matrix-gamma Poisson mixture models are non-identifiable[2] as multiple parameterizations can lead to identical observations.

Remark 4.19. In the context of Definition 4.17, Corollary 3.39 can be used to conclude that the random variable $\Lambda_i \coloneqq \langle v_i, Yv_i\rangle$ follows a one-dimensional gamma distribution with shape parameter α and scale parameter $\langle v_i, Cv_i\rangle$ for every $i \in \{1, \dots, n\}$.

Theorem 4.20 (Probability-generating function of matrix-gamma Poisson mixture models)**.** *Let $(N_1, \dots, N_n) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ according to Definition 4.17. Then the probability-generating function of $(N_1, \dots, N_n)$, see Definition 4.1, is given by*

$$\varphi_{(N_1,\dots,N_n)}(s) = \mathbb{E}\Big[\prod_{i=1}^{n} s_i^{N_i}\Big] = \Big(\det\Big(I_p + \sum_{i=1}^{n}(1-s_i)v_i v_i^{\mathsf{T}} C\Big)\Big)^{-\alpha} \tag{4.17}$$

for at least all $s \in \mathbb{C}^n$ satisfying $\|s\|_\infty \le 1$.

Proof. Fix any $s \in \mathbb{C}^n$ satisfying $\|s\|_\infty \le 1$. Using the conditional independence of the random variables $N_1, \dots, N_n$ given Y, the probability-generating function can be rewritten as

$$\varphi_{(N_1,\dots,N_n)}(s) = \mathbb{E}\Big[\prod_{i=1}^{n} s_i^{N_i}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i=1}^{n} s_i^{N_i}\,\Big|\, Y\Big]\Big] = \mathbb{E}\Big[\prod_{i=1}^{n} \mathbb{E}[s_i^{N_i}|Y]\Big]. \tag{4.18}$$

Since $N_i|Y \sim \mathrm{Poisson}(\langle v_i, Yv_i\rangle)$ for every $i \in \{1, \dots, n\}$, its probability-generating function, conditioned on Y, is given by

$$\mathbb{E}[s_i^{N_i}|Y] \overset{\text{a.s.}}{=} \mathrm{e}^{\langle v_i, Yv_i\rangle(s_i-1)}$$

due to Eq. (4.2). Using this and

$$\langle v_i, Yv_i\rangle = v_i^{\mathsf{T}} Y v_i = \mathrm{tr}(v_i v_i^{\mathsf{T}} Y), \quad i \in \{1, \dots, n\},$$

[2] The term identifiability is commonly used in statistics, see e.g. [12].

where the second equality follows from Lemma 2.1, the probability-generating function from Eq. (4.18) can be rewritten as

$$\begin{aligned}\varphi_{(N_1,\dots,N_n)}(s) &= \mathbb{E}\left[\prod_{i=1}^{n} \mathrm{e}^{\mathrm{tr}(v_i v_i^\mathsf{T} Y)(s_i-1)}\right] \\ &= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} \mathrm{tr}(v_i v_i^\mathsf{T} Y)(s_i-1)\right)\right] \\ &= \mathbb{E}\left[\exp\left(-\mathrm{tr}\left(\sum_{i=1}^{n}(1-s_i)v_i v_i^\mathsf{T} Y\right)\right)\right], \qquad (4.19)\end{aligned}$$

where the last step follows from the linearity of the trace. Due to $\|s\|_\infty \leq 1$, it follows that every component $|s_i| \leq 1$, hence $\mathrm{Re}(1-s_i) \geq 0$ for all $i \in \{1,\dots,n\}$. Furthermore, the matrix $v_i v_i^\mathsf{T}$ is positive semi-definite for all $i \in \{1,\dots,n\}$. Combining this with Lemma 2.8 Item (ii), the matrix

$$\mathrm{Re}\left(\sum_{i=1}^{n}(1-s_i)v_i v_i^\mathsf{T}\right) = \sum_{i=1}^{n} \mathrm{Re}(1-s_i)v_i v_i^\mathsf{T}$$

is positive semi-definite and

$$\mathrm{Im}\left(\sum_{i=1}^{n}(1-s_i)v_i v_i^\mathsf{T}\right) = \sum_{i=1}^{n} \mathrm{Im}(1-s_i)v_i v_i^\mathsf{T}$$

is symmetric as linear combination of symmetric matrices. Therefore, Eq. (3.34) of Theorem 3.32 or Eq. (3.6) of Theorem 3.14, depending on the α, can be applied to Eq. (4.19) with $w = \sum_{i=1}^{n}(1-s_i)v_i v_i^\mathsf{T}$, resulting in Eq. (4.17). □

Lemma 4.21 (The MGammaPoisson$(\alpha, C, v_1, \dots, v_n)$ distribution does not uniquely determine its parameters)**.** *Let the parameters $\alpha, C, v_1, \dots, v_n$ fulfil the conditions of Definition 4.17. Then for every orthogonal matrix $S \in \mathbb{R}^{p\times p}$,*

$$\mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n) = \mathrm{MGammaPoisson}(\alpha, SCS^\mathsf{T}, Sv_1, \dots, Sv_n). \qquad (4.20)$$

Proof. As C is a positive semi-definite matrix, SCS^T is positive semi-definite as well and as $\mathrm{rk}(SCS^\mathsf{T}) = \mathrm{rk}(C)$, shape parameter $\alpha \in \mathcal{G}_{SCS^\mathsf{T}}$, hence the distribution on the right side of Eq. (4.20) is well-defined.

We will show Eq. (4.20) by showing the equality of the characteristic functions of $N \sim \mathrm{MGammaPoisson}(\alpha, C, Sv_1, \dots, Sv_n)$ and $M \sim \mathrm{MGammaPoisson}(\alpha, SCS^\mathsf{T}, Sv_1, \dots, Sv_n)$. As S orthogonal implies

$$Sv_i(Sv_i)^\mathsf{T} SCS^\mathsf{T} = Sv_i v_i^\mathsf{T} CS^\mathsf{T}, \quad i \in \{1,\dots,n\},$$

Eq. (4.17) tells us, for every $s \in \mathbb{C}^n$ satisfying $\|s\|_\infty \leq 1$,

$$\begin{aligned}\varphi_M(s) &= \left(\det\left(I_p + \sum_{i=1}^{n}(1-s_i)Sv_i(Sv_i)^\mathsf{T} SCS^\mathsf{T}\right)\right)^{-\alpha} \\ &= \left(\det\left(I_p + \sum_{i=1}^{n}(1-s_i)v_i v_i^\mathsf{T} C \underbrace{S^\mathsf{T} S}_{=I_p}\right)\right)^{-\alpha} = \varphi_N(s),\end{aligned}$$

where Lemma 2.3 is used for the second equality. As the probability-generating function determines the distribution uniquely, see Remark 4.2, Eq. (4.20) follows. □

The following lemma can be used for an alternative representation of the probability-generating function of matrix-gamma Poisson mixture models given in Theorem 4.20.

Lemma 4.22. *For $n, p \in \mathbb{N}$, let $v_i \in \mathbb{R}^p$ for every $i \in \{1, \dots, n\}$ and let $C \in \mathbb{R}^{p \times p}$ be positive semi-definite. For all $s_1, \dots, s_n \in \mathbb{C}$,*

$$\det\Big(I_p + \sum_{i=1}^{n} (1 - s_i) v_i v_i^\top C\Big) = 1 + \sum_{\substack{I \subseteq \{1,\dots,n\} \\ 0 < |I| \le p}} \det(V_I^\top C V_I) \prod_{i \in I} (1 - s_i), \tag{4.21}$$

where V_I denotes a $(p \times |I|)$-matrix consisting of the column vectors v_i with $i \in I$, where the order is not relevant.

Proof. This proof is split into five steps.

Step 1. We start by using the fact, that the positive semi-definite matrix C can be represented as the product $AA = C$ with its positive semi-definite matrix square root A, see Lemma 2.15, to define

$$u_i := A v_i \in \mathbb{R}^p, \quad \tilde{u}_i := (1 - s_i) A v_i \in \mathbb{C}^p.$$

Since

$$u_i \tilde{u}_i^\top = (1 - s_i) A v_i v_i^\top A,$$

where the last step follows from the fact that A is symmetric, the equation

$$\begin{aligned} \det\Big(I_p + \sum_{i=1}^{n} u_i \tilde{u}_i^\top\Big) &= \det\Big(I_p + \sum_{i=1}^{n} (1 - s_i) A v_i v_i^\top A\Big) \\ &= \det\Big(I_p + \sum_{i=1}^{n} (1 - s_i) v_i v_i^\top A A\Big) = \det\Big(I_p + \sum_{i=1}^{n} (1 - s_i) v_i v_i^\top C\Big) \end{aligned} \tag{4.22}$$

must hold, where the Weinstein–Aronszajn identity given in Lemma 2.3 is used for the second equality. For every non-empty $I \subseteq \{1, \dots, n\}$, let U_I and $\tilde{U}_I$ denote the $(p \times |I|)$-matrices consisting of the column vectors u_i or $\tilde{u}_i$, respectively, with $i \in I$ and V_I the corresponding $(p \times |I|)$-matrix consisting of the column vectors v_i in the same order. Then

$$\det(U_I^\top \tilde{U}_I) = \det\big((A V_I)^\top A V_I\big) \prod_{i \in I} (1 - s_i) = \det(V_I^\top C V_I) \prod_{i \in I} (1 - s_i), \tag{4.23}$$

where the first equality follows from the fact that the determinant is linear in each column and the second equality holds as $A^\top A = AA = C$.

Combining Eq. (4.22) and Eq. (4.23), it is sufficient to prove

$$\det\Big(I_p + \sum_{i=1}^{n} u_i \tilde{u}_i^\top\Big) = 1 + \sum_{\substack{I \subseteq \{1,\dots,n\} \\ 0 < |I| \le p}} \det(U_I^\top \tilde{U}_I) \tag{4.24}$$

instead of Eq. (4.21).

Step 2. The k-th column of the matrix $I_p + \sum_{i=1}^n u_i \tilde{u}_i^\mathsf{T}$ is given by

$$e_k + \sum_{i=1}^n u_i (\tilde{u}_i)_k \in \mathbb{C}^p$$

for every $k \in \{1, \dots, p\}$, where $(\tilde{u}_i)_k \in \mathbb{C}$ refers to the k-th entry in the column vector $\tilde{u}_i$ for $i \in \{1, \dots, n\}$. Using the fact that the determinant is linear in each column, the determinant can be written as the sum

$$\det\Big(I_p + \sum_{i=1}^n u_i \tilde{u}_i^\mathsf{T}\Big) = \sum_{i_1=0}^n \cdots \sum_{i_p=0}^n \det(c_{i_1,1}, \dots, c_{i_p,p}), \tag{4.25}$$

where $c_{j,k} \in \mathbb{C}^p$ is given by

$$c_{j,k} \coloneqq \begin{cases} e_k & \text{for } j = 0, \\ u_j (\tilde{u}_j)_k & \text{for } j \in \{1, \dots, n\}, \end{cases} \tag{4.26}$$

for every $k \in \{1, \dots, p\}$.

As the determinant is equal to zero for every matrix with linearly dependent columns, all summands containing a multiple of a vector u_i in more than one column vanish. Hence, we can group the summands of Eq. (4.25) by which vectors u_i, $i \in \{1, \dots, n\}$, they contain scalar multiples of. For every $I \subseteq \{1, \dots, n\}$ with $|I| \le p$, using the vectors defined in Eq. (4.26), consider the set

$$M_I \coloneqq \big\{(c_{j_1,1}, \dots, c_{j_p,p}) \,\big|\, j_k \in I \cup \{0\} \text{ for all } k \in \{1, \dots, p\}, \\ \text{for all } i \in I \text{ there exists exactly one } k \in \{1, \dots, p\} \text{ with } j_k = i\big\}.$$

For $I = \varnothing$ this means $M_\varnothing = \{(c_{0,1}, \dots, c_{0,p})\} = \{(e_1, \dots, e_p)\}$. Then Eq. (4.25) can be rewritten as

$$\det\Big(I_p + \sum_{i=1}^n u_i \tilde{u}_i^\mathsf{T}\Big) = \sum_{\substack{I \subseteq \{1,\dots,n\} \\ |I| \le p}} \sum_{B \in M_I} \det B = 1 + \sum_{\substack{I \subseteq \{1,\dots,n\} \\ 0 < |I| \le p}} \sum_{B \in M_I} \det B. \tag{4.27}$$

Step 3. Now for a fixed non-empty $I \subseteq \{1, \dots, n\}$ with $|I| \le p$, we want to determine the inner sum on the right side of Eq. (4.27). Note that every matrix in M_I has exactly $p - |I|$ columns that are unit vectors. For every $S \subseteq \{1, \dots, p\}$ with $|S| = p - |I|$, we can define

$$(M_I)_S \coloneqq \{B = (c_{j_1,1}, \dots, c_{j_p,p}) \in M_I \mid j_s = 0 \text{ for every } s \in S\},$$

which denotes the set of matrices containing the unit vectors e_s in the respective columns for all $s \in S$ and multiples of the vectors u_i for every $i \in I$ in the other columns. Note that in case $|I| = p$ the only possible S is the empty set. Using this we can write

$$\sum_{B \in M_I} \det B = \sum_{\substack{S \subseteq \{1,\dots,p\} \\ |S| = p - |I|}} \sum_{B \in (M_I)_S} \det B. \tag{4.28}$$

Step 4. Fix $S \subseteq \{1,\dots,p\}$ with $|S| = p - |I|$. Let $S^c := \{1,\dots,p\}\backslash S$ as well as the set I be ordered (arbitrarily), so we can write $S^c = \{S_1^c,\dots,S_{|I|}^c\} \subseteq \{1,\dots,p\}$ and $I = \{I_1,\dots,I_{|I|}\} \subseteq \{1,\dots,p\}$. Then the columns of every matrix $B = (c_{j_1,1},\dots,c_{j_p,p}) \in (M_I)_S$ can be written as

$$c_{j_s,s} = \begin{cases} e_s & \text{for } s \in S, \\ u_{I_{\sigma(i)}}(\tilde{u}_{I_{\sigma(i)}})_s & \text{for } s = S_i^c \text{ for a } i \in \{1,\dots,|I|\}, \end{cases} \tag{4.29}$$

for a permutation $\sigma \in \mathrm{Sym}(\{1,\dots,|I|\})$. The other way around every $\sigma \in \mathrm{Sym}(\{1,\dots,|I|\})$ can be used to define a $B \in (M_I)_S$ with Eq. (4.29), which we will denote by B^σ. Therefore,

$$\sum_{B\in(M_I)_S} \det B = \sum_{\sigma\in\mathrm{Sym}(\{1,\dots,|I|\})} \det B^\sigma. \tag{4.30}$$

For a fixed $\sigma \in \mathrm{Sym}(\{1,\dots,|I|\})$, using the Laplace expansion[3] in every column with index $s \in S$, which contains the unit vector e_s,

$$\det B^\sigma = \prod_{s\in S}(-1)^{s+s} \det B^\sigma_{S^c,S^c} = \det B^\sigma_{S^c,S^c},$$

where $B^\sigma_{S^c,S^c}$ refers to the $|I| \times |I|$ matrix that results from removing the s-th column and the s-th row from B^σ for all $s \in S$.

Combining the original definition of $B \in (M_I)_S$ given in Eq. (4.26) with the fact that the determinant is linear in each column,

$$\begin{aligned}\det B^\sigma_{S^c,S^c} &= \det\big((u_{I_{\sigma(1)}})_{S^c}(\tilde{u}_{I_{\sigma(1)}})_{S_1^c},\dots,(u_{I_{\sigma(|I|)}})_{S^c}(\tilde{u}_{I_{\sigma(|I|)}})_{S^c_{|I|}}\big)\\ &= \left(\prod_{i=1}^{|I|}(\tilde{u}_{I_{\sigma(i)}})_{S_i^c}\right)\det\big((u_{I_{\sigma(1)}})_{S^c},\dots,(u_{I_{\sigma(|I|)}})_{S^c}\big),\end{aligned}$$

where $(u_i)_{S^c} \in \mathbb{C}^{|I|}$ refers to the vector u_i with only the entries with index in S^c for all $i \in \{1,\dots,n\}$.

Next we want to reorder the columns of $B^\sigma_{S^c,S^c}$ for all $\sigma \in \mathrm{Sym}(\{1,\dots,|I|\})$, such that column $i \in \{1,\dots,|I|\}$ contains a multiple of the vector $(u_{I_i})_{S^c}$. Using the fact that for a quadratic matrix any permutation of the columns multiplies the determinant by the sign of the permutation,

$$\det B^\sigma_{S^c,S^c} = \left(\prod_{i=1}^{|I|}(\tilde{u}_{I_{\sigma(i)}})_{S_i^c}\right)\underbrace{\mathrm{sgn}(\sigma^{-1})}_{=\mathrm{sgn}(\sigma)}\det\big((u_1)_{S^c},\dots,(u_{|I|})_{S^c}\big).$$

Inserting this back into Eq. (4.30),

$$\begin{aligned}\sum_{B\in(M_I)_S}\det B &= \sum_{\sigma\in\mathrm{Sym}(\{1,\dots,|I|\})}\det B_\sigma\\ &= \det\big((u_1)_{S^c},\dots,(u_{|I|})_{S^c}\big)\underbrace{\sum_{\sigma\in\mathrm{Sym}(\{1,\dots,|I|\})}\mathrm{sgn}(\sigma)\prod_{i=1}^{|I|}(\tilde{u}_{I_{\sigma(i)}})_{S_i^c}}_{=\det((\tilde{u}_1)_{S^c},\dots,(\tilde{u}_{|I|})_{S^c})},\end{aligned}$$

[3] See [14, Satz 7.4.7]

where the last equation follows from Leibniz' formula for determinants[4]. Using the notation of Lemma 2.4,

$$\begin{aligned}\sum_{B\in(M_I)_S} \det B &= \det\big((u_1)_{S^c},\dots,(u_{|I|})_{S^c}\big)\det\big((\tilde u_1)_{S^c},\dots,(\tilde u_{|I|})_{S^c}\big)\\ &= \det\big((U_I^\top)_{\{1,\dots,|I|\},S^c}\big)\det\big((\tilde U_I)_{S^c,\{1,\dots,|I|\}}\big),\end{aligned} \tag{4.31}$$

where we also use the fact that the determinant of a quadratic matrix is equal to the determinant of the same matrix transposed.

Step 5. In case $|I| = p$ and therefore $S = \varnothing$ and $S^c = \{1,\dots,p\}$, Eqs. (4.28) and (4.31) tell us

$$\sum_{B\in M_I} \det B = \sum_{B\in(M_I)_\varnothing} \det B = \det(U_I^\top \tilde U_I).$$

In all other cases where $0 < |I| < p$, inserting Eq. (4.31) back into Eq. (4.28),

$$\begin{aligned}\sum_{B\in M_I} \det B &= \sum_{\substack{S\subseteq\{1,\dots,p\}\\ |S|=p-|I|}} \det\big((U_I^\top)_{\{1,\dots,|I|\},S^c}\big)\det\big((\tilde U_I)_{S^c,\{1,\dots,|I|\}}\big)\\ &= \sum_{\substack{S\subseteq\{1,\dots,p\}\\ |S|=|I|}} \det\big((U_I^\top)_{\{1,\dots,|I|\},S}\big)\det\big((\tilde U_I)_{S,\{1,\dots,|I|\}}\big) = \det(U_I^\top\tilde U_I),\end{aligned}$$

where we are using the Cauchy–Binet formula given in Lemma 2.4 for the last equality. Combining these results with Eq. (4.27) proves Eq. (4.24) and hence the lemma. □

Definition 4.23 (Negative multinomial distributions, see [29, Definition 4.51])**.** For $n \in \mathbb{N}$ let $p_1,\dots,p_n \in [0,1)$ satisfy $q := 1-\sum_{i=1}^n p_i \in (0,1]$ and let $\alpha > 0$ denote a shape parameter. Then the $\mathbb{N}_0^n$-valued random vector $N = (N_1,\dots,N_n)$ is said to follow a negative multinomial distribution if

$$\mathbb{P}[N = (k_1,\dots,k_n)] = \frac{\Gamma(\alpha+k_1+\dots+k_n)}{\Gamma(\alpha)} q^\alpha \prod_{i=1}^d \frac{p_i^{k_i}}{k_i!}$$

for all $(k_1,\dots,k_n) \in \mathbb{N}_0^n$. We denote this by $(N_1,\dots,N_n) \sim \mathrm{NegMult}(\alpha,p_1,\dots,p_n)$.

Note that a negative multinomial distribution with $n = 1$ follows a negative binomial distribution, see Definition 4.4.

Remark 4.24 (Probability-generating function of the negative multinomial distributions, see [29, Eq. (4.89)]). For $(N_1,\dots,N_n) \sim \mathrm{NegMult}(\alpha,p_1,\dots,p_n)$ according to Definition 4.23 the probability-generating function is given by

$$\varphi_{(N_1,\dots,N_n)}(s) = \left(\frac{1-\sum_{i=1}^n p_i}{1-\sum_{i=1}^n p_i s_i}\right)^\alpha \tag{4.32}$$

for all $s \in \mathbb{C}^n$ with $|\sum_{i=1}^n p_i s_i| < 1$.

[4] See [14, Satz 7.2.5, Remark 7.2.9]

Corollary 4.25 (Negative multinomial distributions are a subclass of matrix-gamma Poisson mixture distributions)**.** *Let* $(N_1, \dots, N_n)$ *follow the* $\mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ *distribution in the context of Definition 4.17 with* $\alpha \neq 0$*. If the vectors* $v_1, \dots, v_n$ *are pairwise linearly dependent, then the vector* $(N_1, \dots, N_n)$ *follows a negative multinomial distribution. If* C *is of full rank, i.e.* $C \in \mathbb{S}_p^+$*, then the reverse direction holds true as well. In both of these cases* $(N_1, \dots, N_n) \sim \mathrm{NegMult}(\alpha, p_1, \dots, p_n)$*, where*

$$p_i = \frac{\langle v_i, Cv_i \rangle}{c}, \quad i \in \{1, \dots, n\} \text{ with } c \coloneqq 1 + \sum_{j=1}^{n} \langle v_j, Cv_j \rangle. \tag{4.33}$$

Proof. Combining Theorem 4.20 with Lemma 4.22, the probability-generating function of $(N_1, \dots, N_n)$ is given by

$$\varphi_{(N_1,\dots,N_n)}(s) = \Bigg(1 + \sum_{\substack{I \subseteq \{1,\dots,n\} \\ 0<|I|\le p}} \det(V_I^\mathsf{T} C V_I) \prod_{i\in I} (1 - s_i)\Bigg)^{-\alpha}, \quad s \in \mathbb{C}^n, \|s\|_\infty \le 1, \tag{4.34}$$

where V_I denotes a $(p \times |I|)$-matrix consisting of the column vectors v_i with $i \in I$, where the order is not relevant. If the vectors $v_1, \dots, v_n$ are pairwise linearly dependent, then the matrix $V_I^\mathsf{T} C V_I$ does not have full rank for every $I \subseteq \{1, \dots, n\}$ satisfying $|I| \ge 2$, hence the determinant is zero and the summand vanishes. Hence, we can rewrite Eq. (4.34) to

$$\varphi_{(N_1,\dots,N_n)}(s) = \Bigg(\frac{1}{1 + \sum_{i=1}^{n} \underbrace{\det(v_i^\mathsf{T} C v_i)}_{=\langle v_i, Cv_i\rangle}(1 - s_i)}\Bigg)^{\alpha} = \Bigg(\frac{1 - \sum_{i=1}^{n} \frac{\langle v_i, Cv_i\rangle}{c}}{1 - \sum_{i=1}^{n} \frac{\langle v_i, Cv_i\rangle}{c} s_i}\Bigg)^{\alpha},$$

which is the probability-generating function of the $\mathrm{NegMult}(\alpha, p_1, \dots, p_n)$ distribution, where $p_1, \dots, p_n$ are given by Eq. (4.33), see Eq. (4.32). Together with the fact, that the probability-generating function determines a distribution uniquely, this proves the first part of the corollary.

For the other direction, note that Eq. (4.34) can only be equal to the probability-generating function of a negative multinomial distribution if all summands with $|I| \ge 2$ disappear.[5] Since the probability-generating function is defined for every $s \in \mathbb{C}^n$ with $\|s\|_\infty \le 1$, this is equivalent to $V_I^\mathsf{T} C V_I \in \mathbb{R}^{|I|\times|I|}$ having determinant zero, i.e. not having full rank for all $I \subseteq \{1, \dots, n\}$ with $|I| \ge 2$.

Under the condition that C has full rank,

$$\mathrm{rk}(V_I^\mathsf{T} C V_I) = \mathrm{rk}(C V_I V_I^\mathsf{T}) = \mathrm{rk}(V_I^\mathsf{T} V_I) = \mathrm{rk}(V_I),$$

for every non-empty $I \subseteq \{1, \dots, n\}$, where Lemma 2.1 is used for the first equality.[6] Since $\mathrm{rk}(V_I) < |I|$ for all $I \subseteq \{1, \dots, n\}$ with $|I| \ge 2$ is only achieved by all vectors $v_1, \dots, v_n$ being linearly dependent, the corollary holds true. □

[5] This can be e.g. seen by taking both Eqs. (4.32) and (4.34) to the power of $-1/\alpha$ and comparing the coefficients.

[6] Note that for every $A \in \mathbb{K}^{m\times n}$, the equality $\mathrm{rk}(A^\mathsf{H} A) = \mathrm{rk}(A)$ holds, since for every $x \in \mathbb{K}^m$ with $A^\mathsf{H} A x = 0$ is equivalent to $Ax = 0$ because $x^\mathsf{H} A^\mathsf{H} A x = |Ax|^2$.

In order to derive the individual point probabilities of a matrix-gamma Poisson mixture model, we need to differentiate the probability-generating function, see Eq. (4.1). We will do this with the help of a multivariate version of the Faà di Bruno formula, which itself is a generalization of the chain rule, that has been derived in [6]. For the formulation of this generalization we need the following total order on vectors with entries in $\mathbb{N}_0$.

Remark 4.26 (The total order $\prec$ on $\mathbb{N}_0^n$). For two vectors $\mathbf{u} = (u_1, \dots, u_n), \mathbf{v} = (v_1, \dots, v_n) \in \mathbb{N}_0^n$, let $|\mathbf{u}| := u_1 + \dots + u_n$ and $|\mathbf{v}| := v_1 + \dots + v_n$. We say $\mathbf{u} \prec \mathbf{v}$ if one of the following three conditions is satisfied:

(i) $|\mathbf{u}| < |\mathbf{v}|$,

(ii) $|\mathbf{u}| = |\mathbf{v}|$ and $u_1 < v_1$,

(iii) $|\mathbf{u}| = |\mathbf{v}|$, $u_1 = v_1, \dots u_m = v_m$ and $u_{m+1} < v_{m+1}$ for some $m \in \{1, \dots, n-1\}$.

Theorem 4.27 (Multivariate Faà di Bruno formula, see [6, Corollary 2.10])**.** *For $n \in \mathbb{N}$, let $\boldsymbol{k} = (k_1, \dots, k_n)^\mathsf{T} \in \mathbb{N}_0^n$ and let $|\boldsymbol{k}| := k_1 + \dots + k_n$. Let $f\colon D_f \subseteq \mathbb{R} \to \mathbb{R}$ with D_f open be $|\boldsymbol{k}|$-times continuously differentiable and let $g\colon D_g \subseteq \mathbb{R}^n \to \mathbb{R}$, where D_g is open, be partially differentiable k_i times in the i-th component for all $i \in \{1, \dots, n\}$. Then the $\boldsymbol{k}$-th derivative of $h(s) := f(g(s_1, \dots, s_n))$ for $s = (s_1, \dots, s_n) \in D_g$ is given by*

$$h^{(\boldsymbol{k})}(s) = \sum_{r=1}^{|\boldsymbol{k}|} f^{(r)}(g(s)) \sum_{p(\boldsymbol{k},r)} \boldsymbol{k}! \prod_{j=1}^{|\boldsymbol{k}|} \frac{\left(g^{(\boldsymbol{l}_j)}(s)\right)^{m_j}}{m_j!(\boldsymbol{l}_j!)^{m_j}},$$

where the factorial $\boldsymbol{k}! := \prod_{i=1}^n k_i!$ and the set

$$\begin{aligned} p(\boldsymbol{k}, r) := \Big\{ m_1, \dots, m_{|\boldsymbol{k}|} \in \mathbb{N}_0, \boldsymbol{l}_1, \dots, \boldsymbol{l}_{|\boldsymbol{k}|} \in \mathbb{N}_0^n \;\Big|\; & \text{there exists a } s \in \{1, \dots, |\boldsymbol{k}|\} \text{ such that} \\ & m_i = 0 \text{ and } \boldsymbol{l}_i = 0 \text{ for all } i \in \{1, \dots, |\boldsymbol{k}| - s\}, \\ & m_i > 0 \text{ for all } i \in \{|\boldsymbol{k}| - s + 1, \dots, |\boldsymbol{k}|\} \text{ and } 0 \prec \boldsymbol{l}_{|\boldsymbol{k}|-s+1} \prec \dots \prec \boldsymbol{l}_{|\boldsymbol{k}|}, \\ & \sum_{i=|\boldsymbol{k}|-s+1}^{|\boldsymbol{k}|} m_i = r \text{ and } \sum_{i=|\boldsymbol{k}|-s+1}^{|\boldsymbol{k}|} m_i \boldsymbol{l}_i = \boldsymbol{k} \Big\}, \end{aligned} \tag{4.35}$$

where the total order $\prec$ is defined in Remark 4.26.

Lemma 4.28 (Partial derivatives of the probability-generating function of a matrix-gamma Poisson mixture model)**.** *Using the setting of Definition 4.17, let $N = (N_1, \dots, N_n)$ follow the* MGammaPoisson$(\alpha, C, v_1, \dots, v_n)$ *distribution with probability-generating function $\varphi_N(\cdot)$. Let*

$$\mathcal{P}_{n,p} := \{I \subseteq \{1, \dots, n\} \mid 0 < |I| \leq p\}$$

and $\lambda_I := \det(V_I^\mathsf{T} C V_I)$ for every $I \in \mathcal{P}_{n,p}$, where V_I denotes a $(p \times |I|)$-matrix consisting of the column vectors v_i with $i \in I$, where the order is not relevant. With this notation the probability-generating function can be written as $\varphi_N(s) = (g(s))^{-\alpha}$ with

$$g(s) = 1 + \sum_{I \in \mathcal{P}_{n,p}} \lambda_I \prod_{i \in I} (1 - s_i) \tag{4.36}$$

for all $s \in \mathbb{C}^n$ *with* $\|s\|_\infty \leq 1$.

For every $\boldsymbol{k} = (k_1, \ldots, k_n)^\mathsf{T} \in \mathbb{N}_0^n$, *the* $\boldsymbol{k}$*-th derivative of the probability-generating function* $\varphi_N(s) = (g(s))^{-\alpha}$, *now seen as a function from* $\{s \in \mathbb{R}^n \mid \|s\|_\infty < 1\} \to \mathbb{R}$, *is given by*

$$\varphi_N^{(k)}(s) = \sum_{r=1}^{|\boldsymbol{k}|} (-1)^r \alpha^{\overline{r}} (g(s))^{-(\alpha+r)} \sum_{\hat{p}(\boldsymbol{k},r)} \prod_{j=1}^{|\boldsymbol{k}|} \frac{\boldsymbol{k}!}{m_j!} \Bigg((-1)^{|\boldsymbol{l}_j|} \sum_{\substack{I \in \mathcal{P}_{n,p} \\ \zeta(\boldsymbol{l}_j) \subsetneq I}} \lambda_I \prod_{i \in I \setminus \zeta(\boldsymbol{l}_j)} (1 - s_i) \Bigg)^{m_j}, \tag{4.37}$$

where $\alpha^{\overline{r}}$ *denotes the rising factorial* $\alpha \cdots (\alpha + r - 1)$ *and the index set*

$$\hat{p}(\boldsymbol{k}, r) := \big\{ (m_1, \ldots, m_{|\boldsymbol{k}|}; \boldsymbol{l}_1, \ldots, \boldsymbol{l}_{|\boldsymbol{k}|}) \in p(\boldsymbol{k}, r) \bigm| \boldsymbol{l}_1, \ldots, \boldsymbol{l}_{|\boldsymbol{k}|} \in \{0,1\}^n \big\}, \tag{4.38}$$

where $p(\boldsymbol{k}, r)$ *is defined in Eq.* (4.35) *and*

$$\zeta := \left\{ \begin{array}{ccc} \{0,1\}^n & \to & \mathcal{P}(\{1, \ldots, n\}) \\ \boldsymbol{l} & \mapsto & \{i \in \{1, \ldots n\} \mid \boldsymbol{l}_i = 1\} \end{array} \right. . \tag{4.39}$$

Proof. Due to Lemma 4.22, the probability-generating function of N can be written $\varphi_N(s) = (g(s))^{-\alpha}$, where $g(\cdot)$ is defined in Eq. (4.36). We will apply the multivariate Faà di Bruno formula, stated in Theorem 4.27 above, to this function. For every $r \in \{1, \ldots, n\}$ the r-th derivative of

$$f := \left\{ \begin{array}{ccc} \mathbb{R}^+ & \to & \mathbb{R}^+ \\ x & \mapsto & x^{-\alpha} \end{array} \right.$$

is given by

$$f^{(r)}(x) = (-1)^r \alpha^{\overline{r}} (x)^{-(\alpha+r)}, \quad x \in \mathbb{R}.$$

Now for $\mathbf{l} \in \mathbb{N}_0^n$, the $\mathbf{l}$-th derivative of $g(\cdot)$ as defined in Eq. (4.36) must be equal to zero if any entry of $\mathbf{l}$ is greater than one as this would mean differentiating $g(\cdot)$ twice w.r.t. the same component. Hence we can restrict the set $p(\mathbf{k}, r)$ in Theorem 4.27 to $\hat{p}(\mathbf{k}, r)$ defined in Eq. (4.38). As both $0! = 1$ and $1! = 1$, the factorial $\mathbf{l}! = 1$ for every $\mathbf{l} \in \{0,1\}^n$. Furthermore, for a $\mathbf{l} \in \{0,1\}^n$

$$g^{(\mathbf{l})}(s) = (-1)^{|\mathbf{l}|} \sum_{\substack{I \in \mathcal{P}_{n,p} \\ \zeta(\mathbf{l}) \subsetneq I}} \lambda_I \prod_{i \in I \setminus \zeta(\mathbf{l})} (1 - s_i), \quad s \in \mathbb{R}^d \text{ with } \|s\|_\infty \leq 1,$$

using ζ defined in Eq. (4.39), thus proving Eq. (4.37). □

Corollary 4.29 (Individual probabilities of matrix-gamma Poisson mixture models)**.** *Using the notation of Lemma 4.28 the individual probability that* N *is equal to* $\boldsymbol{k} = (k_1, \ldots, k_n)^\mathsf{T} \in \mathbb{N}_0^n$ *is given by*

$$\mathbb{P}[N = \boldsymbol{k}] = \sum_{r=1}^{|\boldsymbol{k}|} (-1)^r \alpha^{\overline{r}} \Bigg(1 + \sum_{I \in \mathcal{P}_{n,p}} \lambda_I \Bigg)^{-(\alpha+r)} \sum_{\hat{p}(\boldsymbol{k},r)} \prod_{j=1}^{|\boldsymbol{k}|} \frac{1}{m_j!} \Bigg((-1)^{|\boldsymbol{l}_j|} \sum_{\substack{I \in \mathcal{P}_{n,p} \\ \zeta(\boldsymbol{l}_j) \subsetneq I}} \lambda_I \Bigg)^{m_j}, \tag{4.40}$$

where $\alpha^{\overline{r}}$ *denotes the rising factorial* $\alpha \cdots (\alpha + r - 1)$ *and* $\hat{p}(\boldsymbol{k}, r)$ *and* ζ *are defined in Eqs.* (4.38) *and* (4.39)*, respectively.*

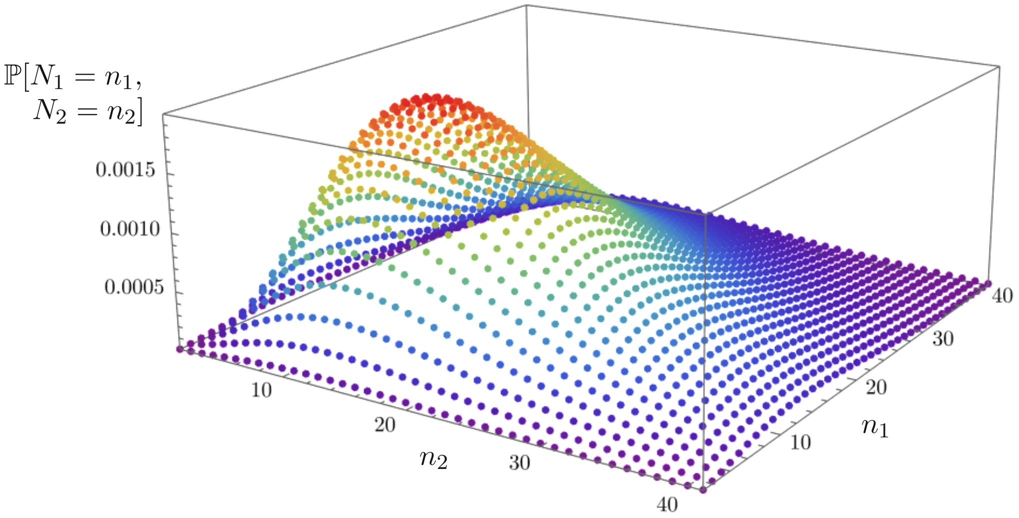


Figure 4.1. *Joint probability mass function of* $(N_1, N_2) \sim \text{MGammaPoisson}(\alpha, C, v_1, v_2)$ *with* $\alpha = 7/2$*, scale matrix* $C = I_2$ *and vectors* $v_1 = (\sqrt{5}, 0)^\mathsf{T}$ *and* $v_2 = (0, \sqrt{5})^\mathsf{T}$*. The random variables* N_1 *and* N_2 *are uncorrelated in this case, using Eq.* (4.43) *and both* N_1 *and* N_2 *follow the* $\text{NegBin}(7/2, 5/6)$ *distribution. See Figure 4.2 below for an example with correlated* N_1, N_2*.*

Proof. For every $\mathbf{k} \in \mathbb{N}_0^n$ the probability that N is equal to $\mathbf{k}$ using the probability-generating function can be derived from Eq. (4.1), resulting in

$$\mathbb{P}[N = \mathbf{k}] = \frac{\varphi_N^{(\mathbf{k})}(0, \dots, 0)}{k_1! \cdots k_n!} = \frac{\varphi_N^{(\mathbf{k})}(0, \dots, 0)}{\mathbf{k}!}.$$

Combining this with Eq. (4.37) of Lemma 4.28 and the fact that $g(0) = 1 + \sum_{I \in \mathcal{P}_{n,p}} \lambda_I$ proves the corollary. □

Remark 4.30 (Numerically instable computation of the point probabilities). Note that the computation of the point probabilities of a matrix-gamma Poisson mixture model is not numerically stable, which is caused by a changing sign of the summands in Eq. (4.40).

Lemma 4.31 (Summation property of matrix-gamma Poisson mixture models)**.** *Fix* $k \in \mathbb{N}$*. Using the setting of Definition 4.17, let* $N^j := (N_1^j, \dots, N_n^j)$ *for every* $j \in \{1, \dots, k\}$ *follow the* $\text{MGammaPoisson}(\alpha_j, C^j, v_1^j, \dots, v_n^j)$ *distribution. Assume there exist a positive definite* $C \in \mathbb{R}^{p \times p}$ *and vectors* $v_1, \dots, v_n \in \mathbb{R}^p$ *such that for every* $j \in \{1, \dots, k\}$ *there exists an orthogonal matrix* $S^j \in \mathbb{R}^{p \times p}$ *with* $S^j C (S^j)^\mathsf{T} = C^j$ *and* $S^j v_i = v_i^j$ *for every* $i \in \{1, \dots, n\}$*.*[7] *If additionally* $N^1, \dots, N^k$ *are independent, then*

$$\sum_{j=1}^{k} N^j \sim \text{MGammaPoisson}(\alpha_1 + \cdots + \alpha_k, C, v_1, \dots, v_n).$$

[7] This condition is in particular satisfied for an N^j with $j \in \{1, \dots, k\}$, that follows a $\text{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ distribution.

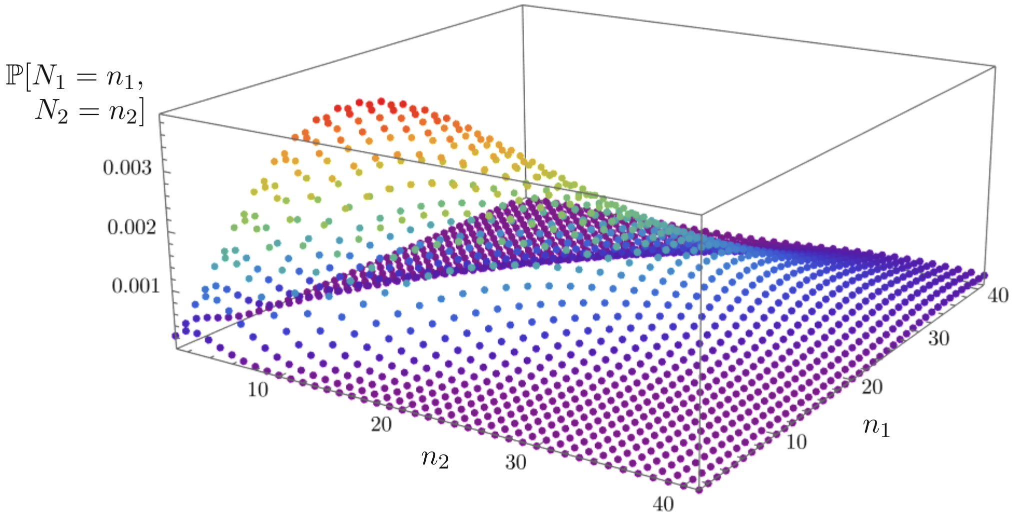


Figure 4.2. *Joint probability mass function of* $(N_1, N_2) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, v_2)$ *with* $\alpha = 7/2$*, scale matrix* $C = I_2$ *and vectors* $v_1 = (1,2)^\mathsf{T}$ *and* $v_2 = (1,2)^\mathsf{T}$*. The random variables* N_1 *and* N_2 *have correlation coefficient* $5/6$*, see Eq.* (4.43) *and both* N_1 *and* N_2 *follow the* $\mathrm{NegBin}(7/2, 5/6)$ *distribution. See Figure 4.1 above for an example with uncorrelated* N_1, N_2*. See Figure 4.3 below for an example for* N_1, N_2 *with the same marginal distributions and correlation, resulting from adding appropriate independent negative binomial random variables.*

Proof. Due to Lemma 4.21 $N^j \sim \mathrm{MGammaPoisson}(\alpha_j, C, v_1, \dots, v_n)$. Because of Lemma 4.9 the probability-generating function of the sum of the independent random vectors $N^1, \dots N^n$ is given as the product of the individual probability-generating functions. Hence, for all $s \in \mathbb{C}^n$, satisfying $\|s\|_\infty \le 1$,

$$\varphi_{N^1+\dots+N^k}(s) = \prod_{j=1}^k \varphi_{N^j}(s) = \prod_{j=1}^k \left(\det\left(I_p + \sum_{i=1}^n (1-s_i) v_i v_i^\mathsf{T} C\right)\right)^{-\alpha_j}$$
$$= \left(\det\left(I_p + \sum_{i=1}^n (1-s_i) v_i v_i^\mathsf{T} C\right)\right)^{-\sum_{j=1}^k \alpha_j}.$$

Since the probability-generating function determines the distribution uniquely, the claim follows. □

Remark 4.32. For $N = (N_1, \dots, N_n) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ with $\alpha \neq 0$ and $Cv_i \neq 0$ for at least one $i \in \{1, \dots n\}$ and a natural number $m \in \mathbb{N}$, the random vector defined by $m \cdot N$ does not follow a matrix-gamma Poisson distribution for $m \neq 1$. This can be seen by using Lemma 4.6, the probability-generating function is given by

$$\varphi_{mN}(s) = \varphi_{(mI_p)N}(s) = \left(\det\left(I_p + \sum_{i=1}^n (1-s_i^m) v_i v_i^\mathsf{T} C\right)\right)^{-\alpha}$$

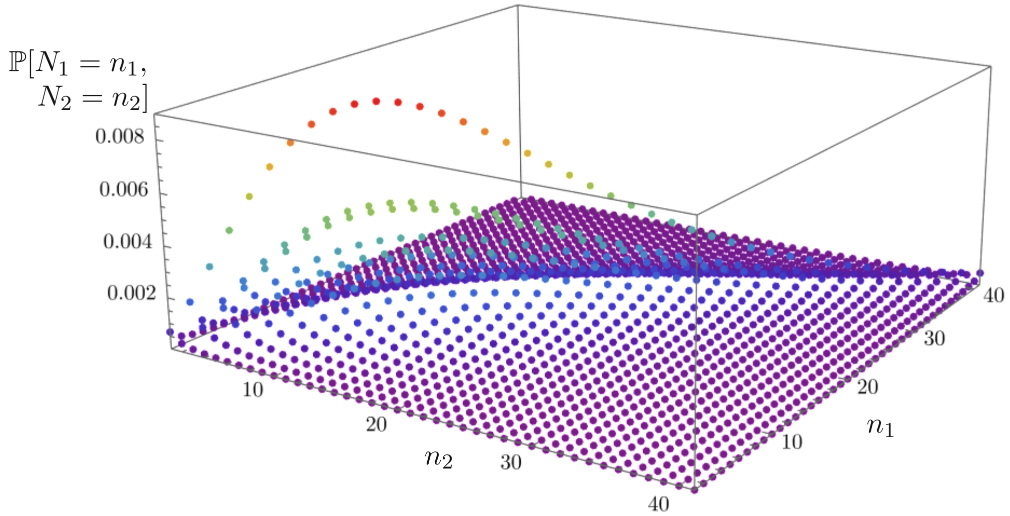


Figure 4.3. *Joint probability mass function of* $(N_1, N_2) \coloneqq (M_1 + M_2, M_1 + M_3)$*, where* M_1, M_2, M_3 *are independent,* $M_1 \sim \mathrm{NegBin}(35/12, 5/6)$ *and* $M_2, M_3 \sim \mathrm{NegBin}(7/12, 5/6)$*. The random variables* N_1 *and* N_2 *have correlation coefficient* $5/6$ *and both* N_1 *and* N_2 *follow the* $\mathrm{NegBin}(7/2, 5/6)$ *distribution. Compare with Figure 4.2, where* N_1, N_2 *follow a* MGammaPoisson *distribution with the same marginal distributions and correlation as here.*

$s \in \mathbb{C}^n$, satisfying $\|s\|_\infty \le 1$, which can only be the probability-generating function of a matrix-gamma Poisson distribution if $m = 1$.

Lemma 4.33 (Properties of the matrix-gamma Poisson mixture models)**.** *In the context of Definition 4.17, let* $(N_1, \dots, N_n)$ *follow the* $\mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ *distribution.*

(i) For every $i, j \in \{1, \dots, n\}$ *satisfying* $i \neq j$*, the expectation, variance and covariance are given by*

$$\mathbb{E}[N_i] = \mathbb{E}[\Lambda_i] = \alpha\langle v_i, Cv_i\rangle, \tag{4.41}$$

$$\mathrm{Var}(N_i) = \mathbb{E}[\Lambda_i] + \mathrm{Var}(\Lambda_i) = \alpha\big(\langle v_i, Cv_i\rangle + \langle v_i, Cv_i\rangle^2\big), \tag{4.42}$$

$$\mathrm{Cov}(N_i, N_j) = \mathrm{Cov}(\Lambda_i, \Lambda_j) = \alpha\langle v_i, Cv_j\rangle^2. \tag{4.43}$$

(ii) Marginal distributions: For every $i \in \{1, \dots, n\}$*,*

$$(N_1, \dots, N_i) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_i).$$

(iii) Permutation property: For every permutation σ *of* $\{1, \dots, n\}$*,*

$$(N_{\sigma(1)}, \dots, N_{\sigma(n)}) \sim \mathrm{MGammaPoisson}(\alpha, C, v_{\sigma(1)}, \dots, v_{\sigma(n)}).$$

(iv) One-dimensional marginal distribution: For every $i \in \{1, \dots, n\}$*,*

$$N_i \sim \mathrm{NegBin}\Big(\alpha, \frac{\langle v_i, Cv_i\rangle}{1 + \langle v_i, Cv_i\rangle}\Big).$$

Proof. Item (i): Recall that $\Lambda_i \sim \text{Gamma}(\alpha, \langle v_i, Cv_i\rangle)$ for every $i \in \{1,\dots,n\}$, see Remark 4.19. The definition of N_i for an $i \in \{1,\dots,n\}$ and the properties of the conditional expectation are used to conclude that

$$\mathbb{E}[N_i] = \mathbb{E}[\mathbb{E}[N_i|Y]] = \mathbb{E}[\Lambda_i] = \alpha\langle v_i, Cv_i\rangle.$$

For every $i \in \{1,\dots,p\}$ the law of total variance can be used to derive

$$\text{Var}(N_i) = \mathbb{E}[\text{Var}(N_i|Y)] + \text{Var}\big(\mathbb{E}[N_i|Y]\big) = \mathbb{E}[\Lambda_i] + \text{Var}(\Lambda_i) = \alpha\langle v_i, Cv_i\rangle + \alpha\langle v_i, Cv_i\rangle^2.$$

For every $i, j \in \{1,\dots,n\}$ with $i \neq j$

$$\begin{aligned}
\text{Cov}(N_i, N_j) &= \mathbb{E}[N_i N_j] - \mathbb{E}[N_i]\,\mathbb{E}[N_j] \\
&= \mathbb{E}[\mathbb{E}[N_i N_j|Y]] - \mathbb{E}[\mathbb{E}[N_i|Y]]\,\mathbb{E}[\mathbb{E}[N_j|Y]] \\
&= \mathbb{E}[\mathbb{E}[N_i|Y]\mathbb{E}[N_j|Y]] - \mathbb{E}[\mathbb{E}[N_i|Y]]\,\mathbb{E}[\mathbb{E}[N_j|Y]] \\
&= \mathbb{E}[\Lambda_i\Lambda_j] - \mathbb{E}[\Lambda_i]\,\mathbb{E}[\Lambda_j] \\
&= \text{Cov}(\Lambda_i, \Lambda_j),
\end{aligned}$$

where the third equality follows from the conditional independence of N_i and N_j given Y. The covariance of Λ_i and Λ_j is given by

$$\begin{aligned}
\text{Cov}(\Lambda_i, \Lambda_j) &= \text{Cov}(\langle v_i, Yv_i\rangle, \langle v_j, Yv_j\rangle) \\
&= \text{Cov}\Bigg(\sum_{k,l=1}^{p} (v_i)_k(v_i)_l Y_{kl}, \sum_{m,n=1}^{p} (v_j)_m(v_j)_n Y_{mn}\Bigg) \\
&= \sum_{k,l,m,n=1}^{p} (v_i)_k(v_i)_l(v_j)_m(v_j)_n \,\text{Cov}(Y_{kl}, Y_{mn}),
\end{aligned}$$

where the last equation follows from the bilinearity of the covariance. Using the covariance of the components of Y derived in Eq. (3.56), this can be rewritten as

$$\begin{aligned}
\text{Cov}(\Lambda_i, \Lambda_j) &= \sum_{k,l,m,n=1}^{p} (v_i)_k(v_i)_l(v_j)_m(v_j)_n \,\frac{\alpha}{2}(c_{km}c_{ln} + c_{kn}c_{lm}) \\
&= \frac{\alpha}{2}\Bigg(\underbrace{\sum_{k,m=1}^{p}(v_i)_k(v_j)_m c_{km}}_{=\langle v_i, Cv_j\rangle}\underbrace{\sum_{l,n=1}^{p}(v_i)_l(v_j)_n c_{ln}}_{=\langle v_i, Cv_j\rangle} + \underbrace{\sum_{k,n=1}^{p}(v_i)_k(v_j)_n c_{kn}}_{=\langle v_i, Cv_j\rangle}\underbrace{\sum_{l,m=1}^{p}(v_i)_l(v_j)_m c_{lm}}_{=\langle v_i, Cv_j\rangle}\Bigg) \\
&= \alpha\,\langle v_i, Cv_j\rangle^2,
\end{aligned}$$

thus proving Item (i).

Item (ii): Eq. (4.4) is used to determine the probability-generating function of $(N_1,\dots,N_i)$:

For every $s \in \mathbb{C}^i$, satisfying $\|s\|_\infty \leq 1$,

$$\begin{aligned}\varphi_{(N_1,\dots,N_i)}(s_1,\dots,s_i) &= \varphi_{(N_1,\dots,N_n)}(s_1,\dots,s_i,1,\dots,1)\\ &= \left(\det\left(I_p + \sum_{j=1}^{i}(1-s_j)v_jv_j^\top C + \sum_{j=i+1}^{n}(1-1)v_jv_j^\top C\right)\right)^{-\alpha}\\ &= \left(\det\left(I_p + \sum_{j=1}^{i}(1-s_j)v_jv_j^\top C\right)\right)^{-\alpha},\end{aligned}$$

where we are using Eq. (4.17) from Theorem 4.20 for the second equation. Hence, since the probability-generating function determines the distribution uniquely, $(N_1,\dots,N_i) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1,\dots,v_i)$.

Item (iii): Due to Eq. (4.6), the probability-generating function is given by

$$\begin{aligned}\varphi_{(N_{\sigma(1)},\dots,N_{\sigma(n)})}(s_1,\dots,s_n) &= \varphi_{(N_1,\dots,N_n)}(s_{\sigma^{-1}(1)},\dots,s_{\sigma^{-1}(n)})\\ &= \left(\det\left(I_p + \sum_{i=1}^{n}(1-s_{\sigma^{-1}(i)})v_iv_i^\top C\right)\right)^{-\alpha}\\ &= \left(\det\left(I_p + \sum_{j\in\sigma^{-1}(\{1,\dots,n\})}(1-s_j)v_{\sigma(j)}v_{\sigma(j)}^\top C\right)\right)^{-\alpha}\end{aligned}$$

for every $s \in \mathbb{C}^n$ satisfying $\|s\|_\infty \leq 1$, where we are again using Eq. (4.17) from Theorem 4.20 for the second equation. Since $\sigma^{-1}(\{1,\dots,n\}) = \{1,\dots,n\}$, this is the probability-generating function of the $\mathrm{MGammaPoisson}(\alpha, C, v_{\sigma(1)},\dots,v_{\sigma(n)})$ distribution.

Item (iv): This follows from the fact that $N_i|\Lambda_i \sim \mathrm{Poisson}(\Lambda_i)$, $\Lambda_i \sim \mathrm{Gamma}(\alpha, \langle v_i, Cv_i\rangle)$ and Lemma 4.11. □

Remark 4.34 (Modelling dependent negative binomial random variables). With the help of Corollary 4.10, we can obtain correlated negative binomial random variables by adding independent negative binomial random variables with the same parameter $p \in [0,1)$. For example, for independent $M_i \sim \mathrm{NegBin}(\alpha_i, p)$ with $\alpha_i \in \mathbb{R}_0^+$ for $i = 1,2,3$, the random variables $(N_1, N_2) := (M_1 + M_2, M_1 + M_3)$ have

$$\mathrm{Cov}(N_1,N_2) = \mathrm{Cov}(M_1+M_2, M_1+M_3) = \mathrm{Var}(M_1)^{8} = \alpha_1\frac{p}{(1-p)^2}$$

and marginal distributions $N_1 \sim \mathrm{NegBin}(\alpha_1+\alpha_2, p)$ and $N_2 \sim \mathrm{NegBin}(\alpha_1+\alpha_3, p)$.

With matrix-gamma Poisson mixture models, on the other hand, it is possible to model dependent negative binomial random variables with different parameter p but with the same parameter α. As an example let $(\tilde{N}_1, \tilde{N}_2) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, v_2)$, where $C \in \mathbb{R}^{p\times p}$ is positive semi-definite, $\alpha \in \mathcal{G}_C$ and $v_1, v_2 \in \mathbb{R}^p$. Then, due to Eq. (4.43),

$$\mathrm{Cov}(\tilde{N}_1, \tilde{N}_2) = \alpha\langle v_1, Cv_2\rangle^2$$

[8] The variance of the negative binomial distribution is e.g. derived in [29, Eq.(4.63)].

and the marginal distributions are given by $\tilde{N}_1 \sim \mathrm{NegBin}(\alpha, \langle v_1, Cv_1\rangle/(1+\langle v_1, Cv_1\rangle))$ as well as $\tilde{N}_2 \sim \mathrm{NegBin}(\alpha, \langle v_2, Cv_2\rangle/(1+\langle v_2, Cv_2\rangle))$ due to Item (iv) of Lemma 4.33.

In Figures 4.2 and 4.3 the probability mass functions of these two versions of correlated negative binomial random variables are illustrated in an example, where now both the parameters α and p are the same for the two dependent random variables.

Remark 4.35 (Matrix-gamma Poisson mixture models do not have the aggregation property). Note that for $(N_1, \dots, N_n) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ in the context of Definition 4.17, the random vector

$$(N_1, \dots, N_i + \dots + N_n)$$

does not necessarily follow a matrix-gamma Poisson mixture distribution for every $i \in \{1, \dots, n-1\}$.

This can be seen by the counterexample with parameters $n = 2$, $v_1 = (1, 0)^\mathsf{T}$ and $v_1 = (0, 1)^\mathsf{T}$. The probability-generating function of $N_1 + N_2$ would then be given by

$$\begin{aligned}\varphi_{N_1+N_2}(s) = \varphi_{(N_1,N_2)}(s, s) &= \big(\det(I_2 + (1-s)v_1v_1^\mathsf{T}C + (1-s)v_2v_2^\mathsf{T}C)\big)^{-\alpha}\\ &= \big(\det(I_2 + (1-s)\underbrace{(v_1v_1^\mathsf{T} + v_2v_2^\mathsf{T})}_{=I_2}C)\big)^{-\alpha},\end{aligned}$$

where the first equation follows from Eq. (4.5) and the second from Eq. (4.17) of Theorem 4.20. This cannot be the probability-generating function of a one-dimensional matrix-gamma Poisson mixture model, which would be a negative binomial distribution due to Item (iv) from Lemma 4.33, as it can further be computed to

$$\begin{aligned}\varphi_{N_1+N_2}(s) &= \big((1+(1-s)c_{11})(1+(1-s)c_{22}) - ((1-s)c_{12})^2\big)^{-\alpha}\\ &= \big(1 + (1-s)(c_{11}+c_{22}) + (1-s)^2(c_{11}c_{22} - c_{12}^2)\big)^{-\alpha},\end{aligned}$$

which can only be the probability-generating function of a negative binomial distribution as given in Eq. (4.3) if $c_{11}c_{22} - c_{12}^2 = 0$, which, for example, does not hold true if C is of full rank.

Theorem 4.36 (Independent variables in matrix-gamma Poisson mixture models)**.** *In the context of Definition 4.17, let $(N_1, \dots, N_n)$ follow a $\mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ distribution with $\alpha > 0$. Then the following statements are equivalent:*

(i) $N_1, \dots, N_n$ are independent.

(ii) The covariances $\mathrm{Cov}(N_i, N_j) = 0$ for all $i, j \in \{1, \dots, n\}$ with $i \neq j$.

(iii) The covariances $\mathrm{Cov}(\Lambda_i, \Lambda_j) = 0$ for all $i, j \in \{1, \dots, n\}$ with $i \neq j$.

(iv) The vectors v_i, v_j satisfy $\langle v_i, Cv_j\rangle = 0$ for all $i, j \in \{1, \dots, n\}$ with $i \neq j$.

Every statement leads to the following representation of the probability-generating function:

$$\varphi_{(N_1,\dots,N_n)}(s) = \prod_{i=1}^{n}\big(\det(I_p + (1-s_i)v_iv_i^\mathsf{T}C)\big)^{-\alpha} = \prod_{i=1}^{n}(1 + (1-s_i)\langle v_i, Cv_i\rangle)^{-\alpha}$$

for all $s \in \mathbb{C}^n$ satisfying $\|s\|_\infty \leq 1$, which is the joint probability-generating function of n independent negative binomial random variables as stated in Lemma 4.33 Item (iii).

Proof. The implication Item (i) $\Rightarrow$ Item (ii) follows from the fact that independent random variables have covariance zero.

Since for every $i, j \in \{1, \dots, n\}$ with $i \neq j$, due to Eq. (4.43) of Lemma 4.33, $\operatorname{Cov}(N_i, N_j) = \operatorname{Cov}(\Lambda_i, \Lambda_j) = \alpha \langle v_i, C v_j \rangle^2$, Item (ii) and Item (iii) are equivalent. Since $\alpha > 0$ is assumed, $\operatorname{Cov}(\Lambda_i, \Lambda_j) = 0$ if and only if $\langle v_i, C v_j \rangle = 0$, hence both Item (ii) and Item (iii) are equivalent to Item (iv) as well.

The last step is to show the implication Item (iv) $\Rightarrow$ Item (i): Due to Lemma 4.8 we know that $N_1, \dots, N_n$ are independent if and only if their joint probability-generating function is equal to the product of their individual probability-generating functions. Using Theorem 4.20,

$$\begin{aligned}\prod_{i=1}^{n} \varphi_{N_i}(s_i) &= \prod_{i=1}^{n} \big(\det(I_p + (1 - s_i) v_i v_i^\top C)\big)^{-\alpha} \\ &= \Bigg(\det\Bigg(I_p + \sum_{\varnothing \neq I \subseteq \{1,\dots,n\}} \prod_{i \in I} (1 - s_i) v_i v_i^\top C\Bigg)\Bigg)^{-\alpha} \qquad (4.44)\end{aligned}$$

for all $s \in \mathbb{C}^n$ satisfying $\|s\|_\infty \leq 1$, where multiplicativity of the determinant is used for the second equality. Under the assumption that Item (iv) is satisfied,

$$(1 - s_i) v_i \underbrace{v_i^\top C (1 - s_j) v_j}_{= (1 - s_j)\langle v_i, C v_j \rangle = 0} v_j^\top C = 0, \quad i \neq j \text{ in } \{1, \dots, n\},$$

hence Eq. (4.44) simplifies to

$$\prod_{i=1}^{n} \varphi_{N_i}(s_i) = \Bigg(\det\Bigg(I_p + \sum_{i=1}^{n} (1 - s_i) v_i v_i^\top C\Bigg)\Bigg)^{-\alpha} = \varphi_{(N_1,\dots,N_n)}(s),$$

where we again use Theorem 4.20, thus proving the independence of $N_1, \dots, N_n$. □

4.3. Matrix-Gamma Multivariate Poisson Mixture Models

Combining the findings from Section 4.2 with Definition 4.13 of general multivariate Poisson mixture models, we can define a multivariate Poisson mixture model, with Poisson intensities derived from a matrix-valued gamma distribution.

Definition 4.37 (Matrix-gamma multivariate Poisson mixture models)**.** For $m \in \mathbb{N}$ let the set $G \subseteq \mathcal{P}(\{1, \dots, m\})$ fulfil $\varnothing \notin G$. For $p \in \mathbb{N}$ let $Y \sim \operatorname{MGamma}(\alpha, C)$ with positive semi-definite scale parameter matrix $C \in \mathbb{R}^{p \times p}$ and shape parameter $\alpha \in \mathcal{G}_C$, see Definition 3.34, and let $v_g \in \mathbb{R}^p$ for every $g \in G$.

A matrix-gamma multivariate Poisson mixture model is now defined to be the special case of a general multivariate Poisson mixture model given in Definition 4.13 with parameter $\Lambda := (\Lambda_g)_{g \in G}$. An $\mathbb{R}^m$-valued random vector N following this distribution is denoted by $N \sim \operatorname{MGammaMPoisson}(\alpha, C, G, (v_g)_{g \in G}, m)$.

A given matrix-gamma multivariate Poisson mixture model does not determine the parameters uniquely, see Lemma 4.41 below.

Remark 4.38 (Matrix-gamma multivariate Poisson mixture models are a generalization of matrix-gamma Poisson mixture models)*.* A matrix-gamma Poisson mixture model, as given in Definition 4.17, is a special case of a matrix-gamma multivariate Poisson mixture model with $G = \{\{1\}, \{2\}, \dots, \{n\}\}$ for $n \in \mathbb{N}$.

Conversely, the random variables $(N_g)_{g \in G}$ from the definition of the general multivariate Poisson mixture model, see Definition 4.13, used for the definition of our MGammaMPoisson model, follow a matrix-gamma Poisson mixture distribution, if G is ordered (arbitrarily).

For every component $i \in \{1, \dots, m\}$, let the set

$$G_i \coloneqq \{g \in G | i \in g\}. \tag{4.45}$$

Lemma 4.39 (Expectations and covariances of matrix-gamma multivariate Poisson mixture models)**.** *Let $N = (N_1, \dots, N_m)^\intercal$ be distributed according to a matrix-gamma multivariate Poisson mixture model in the setting of Definition 4.37. For every $i, j \in \{1, \dots, m\}$ the following equations hold, where G_i is defined in Eq. (4.45) for every component $i \in \{1, \dots, m\}$:*

$$\mathbb{E}[N_i] = \alpha \sum_{g \in G_i} \langle v_g, C v_g \rangle, \tag{4.46}$$

$$\mathrm{Cov}(N_i, N_j) = \alpha \Bigg(\sum_{g \in G_i \cap G_j} \langle v_g, C v_g \rangle + \sum_{g \in G_i, h \in G_j} \langle v_g, C v_h \rangle^2 \Bigg), \tag{4.47}$$

$$\mathrm{Var}(N_i) = \alpha \sum_{g \in G_i} \Bigg(\langle v_g, C v_g \rangle + \sum_{h \in G_i} \langle v_g, C v_h \rangle^2 \Bigg). \tag{4.48}$$

Proof. Eq. (4.46) holds due to

$$\mathbb{E}[N_i] = \Bigg[\sum_{g \in G_i} N_g \Bigg] = \sum_{g \in G_i} \mathbb{E}[N_g] = \alpha \sum_{g \in G_i} \langle v_g, C v_g \rangle,$$

where the first equality follows from Definition 4.12 of N, the second equality from the linearity of the expectation and the last equality from Eq. (4.41).

Since the covariance is linear,

$$\mathrm{Cov}(N_i, N_j) = \sum_{g \in G_i, h \in G_j} \mathrm{Cov}(N_g, N_h) = \sum_{g \in G_i \cap G_j} \mathrm{Var}(N_g) + \sum_{\substack{g \in G_i, h \in G_j, \\ g \neq h}} \mathrm{Cov}(N_g, N_h).$$

Inserting Eq. (4.42) and Eq. (4.43) leads to

$$\begin{aligned} \mathrm{Cov}(N_i, N_j) &= \alpha \sum_{g \in G_i \cap G_j} \langle v_g, C v_g \rangle + \langle v_g, C v_g \rangle^2 + \alpha \sum_{\substack{g \in G_i, h \in G_j, \\ g \neq h}} \langle v_g, C v_h \rangle^2 \\ &= \alpha \Bigg(\sum_{g \in G_i \cap G_j} \langle v_g, C v_g \rangle + \sum_{g \in G_i, h \in G_j} \langle v_g, C v_h \rangle^2 \Bigg). \end{aligned}$$

As this argument holds true for $i = j$ as well, Eq. (4.48) follows from $\mathrm{Var}(N_i) = \mathrm{Cov}(N_i, N_i)$. □

Lemma 4.40 (Probability-generating function of matrix-gamma multivariate Poisson mixture models)**.** *Let the random vector N be distributed according to a matrix-gamma multivariate Poisson mixture model given in Definition 4.37. Then for at least every $s \in \mathbb{C}^m$ satisfying $\|s\|_\infty \leq 1$ its probability-generating function is given by*

$$\varphi_N(s) = \left(\det\left(I_p + \sum_{g\in G}(1 - s^{c_g})v_g v_g^\top C\right)\right)^{-\alpha}, \tag{4.49}$$

where

$$s^{c_g} := \prod_{i\in g} s_i. \tag{4.50}$$

Proof. Fix $s \in \mathbb{C}^m$ satisfying $\|s\|_\infty \leq 1$. Due to the definition of N

$$\varphi_N(s) = \mathbb{E}[s^N] = \mathbb{E}\left[\prod_{g\in G}^{n}(s^{c_g})^{N_g}\right],$$

where s^{c_g} is given in Eq. (4.50).

Using the conditional independence of the random variables N_g for $g \in G$, the probability-generating function can be rewritten as

$$\varphi_N(s) = \mathbb{E}\left[\mathbb{E}\left[\prod_{g\in G}^{n}(s^{c_g})^{N_g}\,\middle|\,Y\right]\right] = \mathbb{E}\left[\prod_{g\in G}\mathbb{E}[\,(s^{c_g})^{N_g}\,|\,Y\,]\right].$$

Since $N_g \sim \text{Poisson}(\langle v_g, Y v_g\rangle)$ for every $g \in G$,

$$\mathbb{E}[\,(s^{c_g})^{N_g}\,|\,Y\,] = \mathrm{e}^{\langle v_g, Y v_g\rangle (s^{c_g}-1)}$$

due to Eq. (4.2).

Since $\|s\|_\infty \leq 1$ particularly implies $|s^{c_g}| \leq 1$, an analogue procedure to the proof of Theorem 4.20 can be used to derive the probability-generating function of N to be

$$\varphi_N(s) = \left(\det\left(I_p + \sum_{g\in G}(1 - s^{c_g})v_g v_g^\top C\right)\right)^{-\alpha}.$$

□

Lemma 4.41 (A matrix-gamma multivariate Poisson mixture model does not uniquely determine its parameters)**.** *Let N be distributed according to a matrix-gamma multivariate Poisson mixture model given in Definition 4.37 with corresponding parameters $\alpha, C, G, v_1, \ldots, v_g, m$ and let $S \in \mathbb{R}^{p\times p}$ be any orthogonal matrix. Then*

$$\begin{aligned}&\text{MGammaMPoisson}(\alpha, C, G, (v_g)_{g\in G}, m)\\ &\qquad = \text{MGammaMPoisson}(\alpha, SCS^\top, G, (Sv_g)_{g\in G}, m).\end{aligned} \tag{4.51}$$

The proof of this lemma can be done in the analogue way to the proof of Lemma 4.21 via the probability-generating function.

Lemma 4.42 (Alternative representation of the probability-generating function of matrix-gamma multivariate Poisson mixture models)**.** *In the setting of Definition 4.37, let N follow a matrix-gamma multivariate Poisson mixture distribution. For all $s \in \mathbb{C}^m$ with $\|s\|_\infty \leq 1$ the probability-generating function given in Eq. (4.49) can be written as*

$$\varphi_N(s) = \left(\det\left(I_p + \sum_{g\in G}(1-s^{c_g})v_g v_g^\mathsf{T} C\right)\right)^{-\alpha} = \left(1 + \sum_{\substack{I\subseteq G\\ |I|\leq p}} \det(V_I^\mathsf{T} C V_I)\prod_{g\in I}(1-s^{c_g})\right)^{-\alpha},$$

where V_I denotes a $(p \times |I|)$-matrix consisting of the column vectors v_g with $g \in I$, where the order is not relevant.

Proof. Follows from Lemma 4.22 with G instead of $\{1,\dots,n\}$ and s^{c_g} instead of s_i. □

Lemma 4.43 (Summation property of matrix-gamma multivariate Poisson mixture models)**.** *Fix $k \in \mathbb{N}$. Using the setting of Definition 4.37, let $N^j := (N_1^j,\dots,N_n^j)$ follow the $\mathrm{MGammaMPoisson}(\alpha_j, C^j, G, (v_g^j)_{g\in G}, m)$ for every $j \in \{1,\dots,k\}$ distribution. Assume there exist a positive definite $C \in \mathbb{R}^{p\times p}$ and vectors $v_g \in \mathbb{R}^p$ for every $g \in G$ such that for every $j \in \{1,\dots,k\}$ there exists an orthogonal matrix $S^j \in \mathbb{R}^{p\times p}$ with $S^j C (S^j)^\mathsf{T} = C^j$ and $S^j v_g = v_g^j$ for every $g \in G$.*[9] *If additionally $N^1,\dots,N^k$ are independent, then*

$$\sum_{j=1}^k N^j$$

follows the matrix-gamma multivariate Poisson mixture model with the same parameters $G, C, (v_g)_{g\in G}$ and $\alpha = \sum_{j=1}^k \alpha_j$.

Proof. Because of Lemma 4.41, $N^j \sim \mathrm{MGammaMPoisson}(\alpha_j, C, G, (v_g)_{g\in G}, m)$. Due to the independence requirement and Lemma 4.9, the probability-generating function of the sum of $N^1,\dots N^k$ is given as the product of the individual probability-generating functions. Hence, using Lemma 4.40, for all $s \in \mathbb{C}^m$ with $\|s\|_\infty \leq 1$,

$$\begin{aligned}\varphi_{N^1+\dots+N^k}(s) = \prod_{j=1}^k \varphi_{N^j}(s) &= \prod_{j=1}^k \left(\det\left(I_p + \sum_{g\in G}(1-s^{c_g})v_g v_g^\mathsf{T} C\right)\right)^{-\alpha_j}\\ &= \left(\det\left(I_p + \sum_{g\in G}(1-s^{c_g})v_g v_g^\mathsf{T} C\right)\right)^{-\sum_{j=1}^k \alpha_j}.\end{aligned}$$

Since the probability-generating function determines the distribution uniquely, this proves the lemma. □

[9] This condition is in particular satisfied for for a N^j with $j \in \{1,\dots,k\}$, that follows a $\mathrm{MGammaMPoisson}(\alpha_j, C, G, (v_g)_{g\in G}, m)$ distribution.

Lemma 4.44 (Properties of matrix-gamma multivariate Poisson mixture models)**.** *In the setting of Definition 4.37, let N follow a matrix-gamma multivariate Poisson mixture distribution. Then N has the following properties.*

(i) Marginal distributions: For every $i \in \{1, \dots, m\}$ the random vector $(N_1, \dots, N_i)$ follows a matrix-gamma multivariate Poisson mixture model with

$$\tilde{G} := \{g \cap \{1, \dots, i\} \,|\, g \in G, g \cap \{1, \dots, i\} \neq \varnothing\}$$

and

$$\tilde{c}_{\tilde{g}} := (c_g)_{\{1,\dots,i\}} \in \{0,1\}^i$$

for every $\tilde{g} \in \tilde{G}$,where g denotes the corresponding subset in G. For every $\tilde{g} \in \tilde{G}$ the vector $v_{\tilde{g}}$ is given by v_g with the corresponding subset $g \in G$. The parameters α, C and m stay the same.

(ii) Permutation property: For every permutation σ of $\{1, \dots, n\}$ the random vector $(N_{\sigma(1)}, \dots, N_{\sigma(n)})$ follows the matrix-gamma multivariate Poisson mixture model with the same α, C and m but with $\hat{g} := \{\sigma(i) \,|\, i \in g\}$ for every $g \in G$ and $\hat{G} = \cup_{g \in G} \hat{g}$ and $v_{\hat{g}} = v_g$.

(iii) One-dimensional marginal distribution: For every $i \in \{1, \dots, m\}$ with $|G_i| = 1$

$$N_i \sim \operatorname{NegBin}\Big(\alpha, \frac{\langle v_g, C v_g \rangle}{1 + \langle v_g, C v_g \rangle}\Big),$$

where $G_i = \{g\}$.

Proof. Item (i): With Eq. (4.4) we can determine the probability-generating function of $(N_1, \dots, N_i)$: For every $s \in \mathbb{C}^i$, satisfying $\|s\|_\infty \leq 1$,

$$\begin{aligned} \varphi_{(N_1,\dots,N_i)}(s_1, \dots, s_i) &= \varphi_{(N_1,\dots,N_n)}(s_1, \dots, s_i, 1, \dots, 1) \\ &= \bigg(\det\Big(I_p + \sum_{g \in G}\Big(1 - (s_1, \dots, s_i, 1, \dots, 1)^{c_g}\Big) v_g v_g^\mathsf{T} C\Big)\bigg)^{-\alpha} \\ &= \bigg(\det\Big(I_p + \sum_{g \in G}(1 - \prod_{k \in g, k \leq i} s_k) v_g v_g^\mathsf{T} C\Big)\bigg)^{-\alpha}, \end{aligned}$$

where Eq. (4.50) is used for the last equality. Hence, using the notation of Item (i) and the fact that the g-th summand vanishes if $\varnothing = g \cap \{1, \dots, i\} \notin \tilde{G}$,

$$\begin{aligned} \varphi_{(N_1,\dots,N_i)}(s_1, \dots, s_i) &= \bigg(\det\Big(I_p + \sum_{\tilde{g} \in \tilde{G}}(1 - \prod_{k \in \tilde{g}} s_k) v_{\tilde{g}} v_{\tilde{g}}^\mathsf{T} C\Big)\bigg)^{-\alpha} \\ &= \bigg(\det\Big(I_p + \sum_{\tilde{g} \in \tilde{G}}(1 - s^{\tilde{c}_g}) v_{\tilde{g}} v_{\tilde{g}}^\mathsf{T} C\Big)\bigg)^{-\alpha}. \end{aligned}$$

Hence, since the probability-generating function determines the distribution uniquely, Item (i) holds.

Item (ii): Due to Eq. (4.6), the probability-generating function is given by

$$\begin{aligned}\varphi_{(N_{\sigma(1)},\ldots,N_{\sigma(n)})}(s_1,\ldots,s_n) &= \varphi_{(N_1,\ldots,N_n)}(s_{\sigma^{-1}(1)},\ldots,s_{\sigma^{-1}(n)}) \\ &= \left(\det\left(I_p + \sum_{g\in G}(1-(s_{\sigma^{-1}(1)},\ldots,s_{\sigma^{-1}(n)})^{c_g})v_g v_g^{\mathsf{T}} C\right)\right)^{-\alpha} \\ &= \left(\det\left(I_p + \sum_{g\in G}(1-s^{c_{\hat{g}}})v_g v_g^{\mathsf{T}} C\right)\right)^{-\alpha} \\ &= \left(\det\left(I_p + \sum_{\hat{g}\in \hat{G}}(1-s^{c_{\hat{g}}})v_{\hat{g}} v_{\hat{g}}^{\mathsf{T}} C\right)\right)^{-\alpha}\end{aligned}$$

for every $s \in \mathbb{C}^m$ satisfying $\|s\|_\infty \leq 1$, where we use the fact that $v_g = v_{\hat{g}}$ for the corresponding $\hat{g}$ for every $g \in G$.

Item (iii): This follows from Item (iv) of Lemma 4.33. □

4.4. Conditioning in the Derived Poisson Mixture Models

Multivariate Poisson mixture models arising from matrix-valued gamma distributions, discussed in the previous section, might be used to describe joint default numbers of a portfolio of creditors in a credit risk model or claim counts of a collective model in actuarial mathematics. In this section we explore how this model is able to incorporate information on observed default or claim numbers.

For example, assume the claim counts $(N_1,\ldots,N_n)$ follow a matrix-gamma Poisson mixture distribution, as in Definition 4.17. Let N_1 represent the claim count of the first quarter of the year. If we already know the value N_1 takes, then we can model how this alters the distribution of the claim count N_2 of the second quarter. So we aim to find

$$\mathcal{L}(N_2|N_1 = k), \quad k \in \mathbb{N}_0.$$

or, extending this to a general $i \in \{1,\ldots,n-1\}$,

$$\mathcal{L}(N_{i+1}|N_1 = k_1,\ldots,N_i = k_i), \quad k \in \mathbb{N}_0^i.$$

We first explore this topic for simple gamma-mixed Poisson models before moving onto matrix-gamma Poisson mixture models models and ultimately to matrix-gamma multivariate Poisson mixture models.

4.4.1. Simple Gamma-Mixed Poisson Models

Similarly to Lemma 4.11, a simple gamma Poisson mixture model is described by

$$N|\Lambda \sim \text{Poisson}(\lambda\Lambda), \quad \Lambda \sim \text{Gamma}(\alpha,\beta) \tag{4.52}$$

with parameters $\alpha,\beta \in \mathbb{R}^+$ and a real non-negative λ. Let us look how the distribution of Λ changes if we have additional knowledge on the number of events – let it be claims or defaults, depending on the application.

Lemma 4.45 (Conditioning in simple gamma-mixed Poisson models, see [9, Theorem 3.16]). *For N, Λ given in Eq. (4.52) and $k \in \mathbb{N}_0$ such that $\mathbb{P}[N = k] > 0$,*

$$\mathcal{L}(\Lambda | N = k) = \mathbb{P}_{\Lambda^k \,\mathrm{e}^{-\lambda\Lambda}}\, \Lambda^{-1} = \mathrm{Gamma}(\alpha + k, (\beta^{-1} + \lambda)^{-1}), \tag{4.53}$$

where $\mathbb{P}_{\Lambda^k \,\mathrm{e}^{-\lambda\Lambda}}$ refers to the $(\Lambda^k \,\mathrm{e}^{-\lambda\Lambda})$-biased measure as defined in Definition 3.62. Furthermore, if $\mathbb{P}[N \geq 1] > 0$,

$$\mathcal{L}(\Lambda | N \geq 1) = \mathbb{P}_{(1-\mathrm{e}^{-\lambda\Lambda})}\Lambda^{-1} = \frac{1}{\mathbb{P}[N \geq 1]}\,\mathcal{L}(\Lambda) - \frac{\mathbb{P}[N = 0]}{\mathbb{P}[N \geq 1]}\,\mathcal{L}(\Lambda | N = 0) \tag{4.54}$$

Proof. For $k \in \mathbb{N}_0$,

$$\mathbb{P}[N = k] = \mathbb{E}[\mathbb{P}[N = k\,|\,\Lambda]] = \mathbb{E}\Big[\frac{(\lambda\Lambda)^k}{k!}\,\mathrm{e}^{-\lambda\Lambda}\Big].$$

Using Remark 3.63,

$$\mathbb{P}_{\frac{(\lambda\Lambda)^k}{k!}\,\mathrm{e}^{-\lambda\Lambda}} = \mathbb{P}_{\Lambda^k\,\mathrm{e}^{-\lambda\Lambda}},$$

hence the argumentation of [9, Theorem 3.16] can be used for the first equality of Eq. (4.53). The second equality follows from Eq. (3.73) for $p = 1$.

The proof of Eq. (4.54) can also be found in [9, Theorem 3.16]. □

Corollary 4.46 (Conditioning in simple gamma-mixed Poisson models with multiple known events). *For $\alpha, \beta \in \mathbb{R}^+$ let $\Lambda \sim \mathrm{Gamma}(\alpha, \beta)$. For $n \in \mathbb{N}$ let $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ be non-negative and let*

$$N_i | \Lambda \sim \mathrm{Poisson}(\lambda_i \Lambda), \quad i \in \{1, \dots, n\},$$

be independent when conditioned on Λ. Then for $i \in \{1, \dots, n-1\}$ and $k \in \mathbb{N}_0^i$ with $\mathbb{P}[N_j = k_j] > 0$ for every $j \in \{1, \dots, i\}$,

$$\mathcal{L}(\Lambda | N_1 = k_1, \dots, N_i = k_i) = \mathrm{Gamma}\big(\alpha + k_1 + \dots + k_i, (\beta^{-1} + \lambda_1 + \dots + \lambda_i)^{-1}\big) \tag{4.55}$$

and therefore, using the conditional independence of $N_1, \dots, N_n$ and Lemma 4.11,

$$\mathcal{L}(N_{i+1} | N_1 = k_1, \dots, N_i = k_i) = \mathrm{NegBin}\Big(\alpha + k_1 + \dots + k_i, \frac{\lambda_{i+1}}{\beta^{-1} + \lambda_1 + \dots + \lambda_i + \lambda_{i+1}}\Big).$$

Proof. Due to the conditional independence of $N_1, \dots, N_i$ given Λ,

$$\begin{aligned}\mathbb{P}[N_1 = k_1, \dots, N_i = k_i] &= \mathbb{E}[\mathbb{P}[N_1 = k_1, \dots, N_i = k_i | \Lambda]] \\ &= \mathbb{E}[\mathbb{P}[N_1 = k_1 | \Lambda] \cdots \mathbb{P}[N_i = k_i | \Lambda]] = \mathbb{E}\Big[\frac{\lambda_1^{k_1} \cdots \lambda_i^{k_i}}{k_1! \dots k_i!} \Lambda^{k_1 + \dots + k_i}\, \mathrm{e}^{-(\lambda_1 + \dots + \lambda_i)\Lambda}\Big]\end{aligned}$$

for every $k \in \mathbb{N}_0^i$. For N_j with $\mathbb{P}[N_j = k_j] > 0$ for every $j \in \{1, \dots, i\}$ this expectation is greater then zero and hence, with Remark 3.63 and analogous arguments to the proof of Lemma 4.45, this corollary holds true. □

Remark 4.47 (The conditional distributions in a simple gamma-mixed Poisson model only depend on the sum of the events and the intensities, see e.g. [8, Property 3.1]). Note that the conditional distribution in Eq. (4.55) only depends on the total number of events and the sum of the intensities of the random variables $N_1, \dots, N_n$. If we assume that each N_i represents the claim count of a certain time period, e.g. quarter or month, then this implies that the arrangement of past claim counts across time periods is not significant, only the overall sum and the total time period matter.

4.4.2. Matrix-Gamma Poisson Mixture Models

Moving on to matrix-gamma Poisson mixture models, we can see that Lemma 4.45 and Corollary 4.46 can be generalized to these models with the restriction that we can only condition on whether the number of events is equal to or strictly greater than zero. Per Definition 4.17, the number of events N only depends on the mixing matrix-valued gamma distribution Y via the one-dimensional distribution $\langle v, Yv\rangle$, hence it is a very interesting finding that Y still follows a matrix-valued gamma distribution under the condition $N = 0$.

Lemma 4.48 (Conditioning in matrix-gamma Poisson mixture models)**.** *Let N follow the* $\mathrm{MGammaPoisson}(\alpha, C, v)$ *distribution according to Definition 4.17 with $n = 1$ based on* $Y \sim \mathrm{MGamma}(\alpha, C)$ *and* $C = AA^\top$ *with* $A \in \mathbb{R}^{p\times q}$ *where* $q = \max\{1, \mathrm{rk}(C)\}$ *as discussed in Remark 3.27. Then*

$$\begin{aligned}\mathcal{L}(Y|N=0) = \mathbb{P}_{\mathrm{e}^{-\mathrm{tr}(vv^\top Y)}}\, Y^{-1} &= \mathrm{MGamma}(\alpha, A^\top(I_p + A^\top vv^\top A)^{-1}A)\\ &= \mathrm{MGamma}(\alpha, (C^{-1} + vv^\top)^{-1}),\end{aligned} \tag{4.56}$$

where the second equality only holds if C is invertible or, equivalently, positive definite. Furthermore, if $\mathbb{P}[N \geq 1] > 0$, *then*

$$\mathcal{L}(Y|N\geq 1) = \mathbb{P}_{1-\mathrm{e}^{-\mathrm{tr}(vv^\top Y)}}\, Y^{-1} = \frac{1}{\mathbb{P}[N\geq 1]}\mathcal{L}(Y) - \frac{\mathbb{P}[N=0]}{\mathbb{P}[N\geq 1]}\mathcal{L}(Y|N=0). \tag{4.57}$$

Proof. As $N|Y \sim \mathrm{Poisson}(\langle v, Y, v\rangle)$

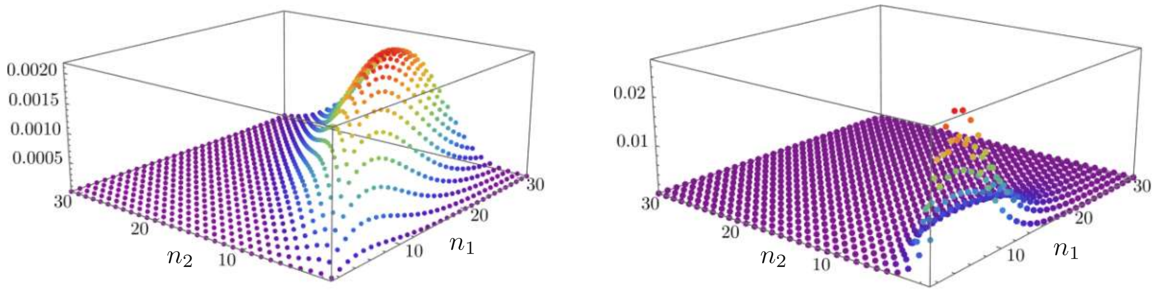
$$\mathbb{P}[N=0] = \mathbb{E}[\mathbb{P}[N=0|Y]] = \mathbb{E}[\mathrm{e}^{-\langle v,Y,v\rangle}] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(vv^\top Y)}].$$

With an analogue argument to Lemma 4.45, the first equality of Eq. (4.56) follows. Due to Eq. (3.73) of Theorem 3.64 the second and third equality hold. The existence of the inverse of $I_p + A^\top vv^\top A$ is also justified in Theorem 3.64.

The argument for Eq. (4.57) is the same as for Eq. (4.54) and follows from [9, Theorem 3.16]. □

Corollary 4.49 (Conditioning in matrix-gamma Poisson mixture models with multiple known events)**.** *Let* $N = (N_1, \dots, N_n) \sim \mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$ *according to Definition 4.17 based on* $Y \sim \mathrm{MGamma}(\alpha, C)$, *where* $C = AA^\top$ *with* $A \in \mathbb{R}^{p\times q}$ *where* $q = \max\{1, \mathrm{rk}(C)\}$ *as discussed in Remark 3.27. Then for every* $i \in \{1, \dots, n\}$

$$\begin{aligned}\mathcal{L}(Y|N_1=0,\dots,N_i=0) &= \mathbb{P}_{\mathrm{e}^{-\mathrm{tr}((v_1v_1^\top+\cdots+v_iv_i^\top)Y)}}\, Y^{-1}\\ &= \mathrm{MGamma}(\alpha, A^\top(I_p + A^\top(v_1v_1^\top + \cdots + v_iv_i^\top)A)^{-1}A)\\ &= \mathrm{MGamma}(\alpha, (C^{-1} + v_1v_1^\top + \cdots + v_iv_i^\top)^{-1}),\end{aligned}$$



(a) $\mathbb{P}[\tilde{N}_1 = n_1, \tilde{N}_2 = n_2]$ (b) $\mathbb{P}[\tilde{N}_1 = n_1, \tilde{N}_2 = n_2 \,|\, N_1 = 0, N_2 = 0]$

Figure 4.4. *Illustration of the probability mass functions of Example 4.50 of conditioning in a simple matrix-gamma mixture model.*

where the last equality only holds if $C \in \mathbb{S}_p^+$.

Proof. Due to the conditional independence of $N_1, \dots, N_i$,

$$\mathbb{P}[N_1 = 0, \dots, N_i = 0] = \mathbb{E}[\mathbb{P}[N_1 = 0|Y] \cdots \mathbb{P}[N_i = 0|Y]] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}((v_1 v_1^\mathsf{T} + \cdots + v_i v_i^\mathsf{T})Y)}].$$

Hence, with the same argument as in the lemma above this corollary must hold. □

Example 4.50 (Conditional distributions of claim numbers in a matrix-gamma Poisson mixture model)**.** In the setting of Definition 4.37, consider two $\mathbb{R}^2$ valued random vectors (N_1, N_2) and $(\tilde{N}_1, \tilde{N}_2)$ that follow a MGammaPoisson(α, C, v_1, v_2) distribution based on $Y \sim$ MGamma(α, C). We assume that (N_1, N_2) and $(\tilde{N}_1, \tilde{N}_2)$ are conditionally independent given Y. We want to explore how the distribution of $(\tilde{N}_1, \tilde{N}_2)$ changes conditioned on the event $\{N_1 = 0, N_2 = 0\}$. For simplicity assume C is positive definite.

Due to Item (iv) of Lemma 4.33, the marginal distributions of N_1 and $\tilde{N}_1$ are given by NegbBin$(\alpha, \langle v_i, Cv_i\rangle/(1 + \langle v_1, Cv_1\rangle)$ and $N_2, \tilde{N}_2 \sim$ NegbBin$(\alpha, \langle v_2, Cv_2\rangle/(1 + \langle v_2, Cv_2\rangle))$. Using Eq. (4.42) and Eq. (4.43), the correlation coefficient

$$\rho_{N_1,N_2} = \frac{\mathrm{Cov}(N_1, N_2)}{\sqrt{\mathrm{Var}(N_1)}\sqrt{\mathrm{Var}(N_2)}} = \frac{\langle v_1, Cv_2\rangle^2}{\sqrt{(\langle v_1, Cv_1\rangle + \langle v_1, Cv_1\rangle^2)(\langle v_2, Cv_2\rangle + \langle v_2, Cv_2\rangle^2)}}.$$

The correlation between $\tilde{N}_1$ and $\tilde{N}_2$ is given by the same formula.

Now, conditioned on $N_1 = 0, N_2 = 0$, the distribution of Y is the MGamma$(\alpha, (C^{-1} + v_1^\mathsf{T} v_1 + v_2^\mathsf{T} v_2)^{-1})$ distribution. Hence,

$$\mathcal{L}(\tilde{N}_1, \tilde{N}_2 \,|\, N_1 = 0, N_2 = 0) \sim \mathrm{MGammaPoisson}(\alpha, (C^{-1} + v_1^\mathsf{T} v_1 + v_2^\mathsf{T} v_2)^{-1}, v_1, v_2).$$

The marginal distributions and the correlation of $\tilde{N}_1$ and $\tilde{N}_2$ under this condition change accordingly.

If we assume $\alpha = 7/2$, scale matrix $C = 9I_2$ and vectors $v_1 = (1/3, 2/3)^\mathsf{T}$ and $v_2 = (1/3, 1)^\mathsf{T}$, then $\mathcal{L}(N_1) = \mathcal{L}(\tilde{N}_1) =$ NegBin$(7/2, 5)$ and $\mathcal{L}(N_2) = \mathcal{L}(\tilde{N}_2) =$ NegBin$(7/2, 10)$. The correlation coefficient of $\tilde{N}_1$ and $\tilde{N}_2$ is equal to 0.853. Conditional on $N_1 = 0, N_2 = 0$, the marginal distributions change to $(\tilde{N}_1|N_1 = 0, N_2 = 0) \sim \mathrm{NegBin}(7/2, 20/31)$ and $(\tilde{N}_2|N_1 = 0, N_2 = 0) \sim \mathrm{NegBin}(7/2, 70/31)$. The correlation coefficient between $\tilde{N}_1$ and $\tilde{N}_2$ given $N_1 = 0, N_2 = 0$ is equal to 0.510.

This example is illustrated in Figure 4.4.

4.4.3. Matrix-Gamma Multivariate Poisson Mixture Models

Finally, the results from the section before can also be applied to the more general matrix-gamma multivariate Poisson mixture models, leading to Lemma 4.51 below. Again, we condition on whether or not events have occurred and not on the number of events.

Lemma 4.51 (Conditioning in matrix-gamma multivariate Poisson mixture models)**.** *Let $N = (N_1, \dots, N_m)^\top$ follow a matrix-gamma multivariate Poisson mixture distribution in the setting of Definition 4.37 based on $Y \sim \mathrm{MGamma}(\alpha, C)$. For $i \in \{1, \dots, m\}$ let G_i as in Eq. (4.45). Then*

$$\begin{aligned}\mathcal{L}(Y|N_i = 0) &= \mathbb{P}_{\mathrm{e}^{-\mathrm{tr}(\sum_{g\in G_i} v_g (v_g)^\top Y)}\, Y}^{-1} \\ &= \mathrm{MGamma}(\alpha, A^\top (I_p + A^\top \sum_{g\in G_i} v_g (v_g)^\top A)^{-1} A) \\ &= \mathrm{MGamma}(\alpha, (C^{-1} + \sum_{g\in G_i} v_g (v_g)^\top)^{-1}),\end{aligned}$$

where the last equality only holds if $C \in \mathbb{S}_p^+$. Furthermore, if $\mathbb{P}[N_i \geq 1] > 0$, then

$$\mathcal{L}(Y|N_i \geq 1) = \mathbb{P}_{1-\mathrm{e}^{-\mathrm{tr}(\sum_{g\in G_i} v_g (v_g)^\top Y)}\, Y}^{-1} = \frac{1}{\mathbb{P}[N_i \geq 1]}\mathcal{L}(Y) - \frac{\mathbb{P}[N_i = 0]}{\mathbb{P}[N_i \geq 1]}\mathcal{L}(Y|N_i = 0).$$

Proof. Due to Definition 4.37 of matrix-gamma Poisson mixture models and Definition 4.13 of the general multivariate Poisson mixture models

$$N_i = \sum_{g\in G_i} N_g$$

and hence, as $N_i = 0$ if and only if $N_g = 0$ for all $g \in G_i$,

$$\mathbb{P}[N_i = 0] = \mathbb{P}[N_g = 0 \text{ for all } g \in G_i] = \mathbb{E}\Big[\prod_{g\in G_i} \mathbb{P}[N_g = 0|Y]\Big],$$

where we use the conditional independence of the N_g for $g \in G$ given Y for the second equality. As $N_g|Y \sim \mathrm{Poisson}(\langle v_g, Y v_g\rangle)$,

$$\mathbb{P}[N_i = 0] = \mathbb{E}\Big[\prod_{g\in G_i} \mathrm{e}^{-\mathrm{tr}(v_g (v_g)^\top Y)}\Big] = \mathbb{E}[\mathrm{e}^{-\mathrm{tr}(\sum_{g\in G_i} v_g (v_g)^\top Y)}].$$

Thus, with the analogue argumentation of Corollary 4.49, the first statement holds true.

As

$$\mathbb{P}[N_i \geq 1] = 1 - \mathbb{P}[N_i = 0] = \mathbb{E}[1 - \mathrm{e}^{-\mathrm{tr}(\sum_{g\in G_i} v_g (v_g)^\top Y)}]$$

the argument for the second statement is the same as for Eq. (4.54) and follows from [9, Theorem 3.16]. □

A. Open Questions and Ideas

Several of the mentioned concepts have potential for further exploration. Some incomplete tasks and future ideas are collected in the list below.

- This thesis only looks at the matrix-valued generalization of one-dimensional gamma distributions given in Definition 2.25. It would be interesting to investigate whether there is also a reasonable matrix-valued extension of generalized gamma distributions given in Definition 2.26.

- Complex matrix-valued gamma distributions could be explored further, see Remark 3.46. While some work has been done on this topic, see e.g. [23, Section 5.2a], it would be interesting to find out which of the properties discussed here can be adjusted to also hold true for the complex version.

- Example 3.21 explores the joint distribution of the trace and the determinant of standard matrix-valued gamma distributions taking values in $\mathbb{S}_2^+$. The function

$$f_{\det(X),\operatorname{tr}(X)}(d,t) = \frac{\sqrt{2}\pi}{\Gamma_2(\alpha)}\, \mathrm{e}^{-t}\, d^{\,\alpha-\frac{3}{2}}\, 1_{\{t\geq 2\sqrt{d}\}}, \quad (d,t)\in\mathbb{R}^+\times\mathbb{R}^+.$$

  is a density of $(\det(X),\operatorname{tr}(X))$ for $X\sim \operatorname{MGamma}(\alpha, I_2)$. Due to the form of a density of X given in Eq. (3.21), it is likely that this result can be generalized to all $p\in\mathbb{N}$ and possibly in a similar form for scale matrices $C\in\mathbb{S}_p^+$.

- In the context of singular Wishart distributions, singular matrix-valued beta distributions are often mentioned, see [30, 32]. It could be explored further, which properties of the non-singular matrix-valued beta distributions, as for example listed in [13, Chapter 5], also hold true for the singular versions. This could possibly be done via characteristic functions as done for singular Wishart distributions in this thesis.

- The distribution of $N\sim \operatorname{NegBin}(\alpha,p)$ under biased probability measures was explored in Lemma 4.15 and Corollary 4.16. It would be interesting to find out if there are similar biased distributions of negative multinomial distributions, as given in Definition 4.23, which are also a subset of matrix-gamma Poisson mixture models as proven in Corollary 4.25.

- Under which conditions is the $\operatorname{MGammaPoisson}(\alpha, C, v_1,\ldots,v_n)$ distribution, using the notation of Definition 4.17, fully determined by the parameter α and the values of $\langle v_i, Cv_j\rangle$ for $i,j\in\{1,\ldots,n\}$? Item (i) of Lemma 4.33 motivates this question, an answer using the probability-generating function, see Theorem 4.20, might be possible.

- With the matrix-gamma multivariate Poisson mixture model, as discussed in this thesis, only non-negative correlations between claim counts can be modelled, see Eq. (4.43). In [29, Theorem 7.39] it has been shown that a multivariate Poisson mixture model can be modeled to have every given covariance structure by using linear combinations of independent gamma distributed risk factors for the vector of Poisson intensities. It would be interesting to see if these ideas could be combined to arrive at an extended matrix-gamma multivariate Poisson mixture model, possibly driven by multiple matrix-valued gamma distributions, that allows negative correlations between the modelled claim counts.

B. Main Points of the Thesis

This appendix lists the main points and contributions of the thesis, categorized into literature-based insights and novel findings introduced within this work.

Literature-Based Main Points:

(i) The content of Sections 2.1, 2.2, 2.3 and 2.4 is based mainly on cited literature.

(ii) The Laplace transform of $f(A) := (\det A)^{a-\frac{p+1}{2}}$ for positive definite matrices $A \in \mathbb{R}^{p \times p}$ given in Eq. (2.31) of Theorem 2.37. Was found in [13, Eq. (1.4.6)] or [15, Eq. (1.1)] with partially incomplete proof.

(iii) Definition and properties of (singular) Wishart distributions, see Definition 3.8, Corollary 3.33, Theorem 3.54 and Theorem 3.14, found in [20, Section 2.3] and [13, Chapter 3] with partially incomplete proof of Theorem 3.14.

(iv) Definition and several basic properties of matrix-valued gamma distributions, such as its characteristic function given in Theorem 3.32, resulting in Corollary 3.39 and Corollary 3.40, the summation property given in Lemma 3.42, moments given in Theorem 3.59. Was mainly found in [20, Chapter 2] and [13, Chapter 3.6].

(v) Equivalence of the distribution of a standard matrix-valued gamma distribution and its Cholesky decomposition given in Theorem 3.22. Was partially found in [33, Section 4.1(i), Property 1].

(vi) Marginal distributions of matrix-valued gamma distributions given in Theorem 3.52, found in [20, Proposition 2.3] and [13, Theorem 3.3.9].

(vii) An introduction to (Poisson) mixture models as well as their usage in credit risk modelling is given in Section 4.1, mainly taken from [29]. In particular the one-dimensional gamma-mixed Poisson model, which is (unconditionally) equal to a negative binomial distribution is discussed, see Lemma 4.11.

(viii) Section 4.4.1 on conditioning in a simple gamma-mixed Poisson model has been taken mainly from [9].

Main Points Introduced in this Thesis:

(i) Seeing the set of positive matrices as an integral domain as a $p(p-1)/2$-dimensional submanifold of $\mathbb{R}^{p \cdot p}$ instead of a subset of $\mathbb{R}^{p(p-1)/2}$ in the context of matrix-valued distributions, see Remark 2.23 and Remark 2.38.

(ii) The joint distribution of the determinant and the trace of a standard matrix-valued gamma distribution for $p = 2$, see Example 3.21.

(iii) The equivalence of introducing Wishart distributions via ZZ^T with Z following a matrix-valued normal distribution and via its density or the block notation given in Theorem 3.54 has been proven using the characteristic function here in Theorem 3.14, which has not been done this way before (as far as we know).

(iv) Extending and reformulating the definition of matrix-valued gamma distributions found in the literature such as [13] or [23] as well as Wishart distributions to include degenerated cases both in terms of the scale parameter matrix and the shape parameter, see Definition 3.34, and proving several properties given in Theorem 3.32, Corollary 3.39, Corollary 3.40, Lemma 3.42, Theorem 3.52 Items (i) and (ii) and Theorem 3.59 for this extended definition.

(v) Extending the Laplace transform of the matrix valued gamma function given in Theorem 3.32 for $M \neq 0$ and $\beta \neq 0$ and not necessarily symmetric w.

(vi) The fact that a matrix-valued gamma distribution cannot be easily characterized by one-dimensional distributions, see Remark 3.41.

(vii) The fact that a matrix-valued exponential distribution, as defined in [22, Section 1.2] as well as Definition 3.47 in this thesis, is memoryless w.r.t. the Loewner partial order is given in Lemma 3.49.

(viii) The distribution of a matrix-valued gamma distribution under a biased probability-biased measure given in Theorem 3.64.

(ix) The distribution of a negative binomial distribution under a biased probability-biased measure given in Lemma 4.15.

(x) The application of matrix-valued gamma distributions to Poisson mixture models, leading to the introduction of matrix-gamma Poisson mixture models given in Section 4.2, see Definition 4.17.

(xi) Several basic properties of matrix-gamma Poisson mixture models are given in Lemmas 4.31 and 4.33. The probability-generating function of these models is derived in Theorem 4.20 as well as an alternative representation, see Lemma 4.22. Conditions on when a matrix-gamma Poisson mixture model is equal to a negative multinomial distribution are given in Corollary 4.25 based on Lemma 4.22. Conditions for independent components of a matrix-gamma Poisson mixture model are given in Theorem 4.36.

(xii) The introduction and the properties of matrix-gamma multivariate Poisson mixture models described in Section 4.3, see Definition 4.37, Lemmas 4.39, 4.40, 4.42, 4.43 and 4.44.

(xiii) Section 4.4 on conditioning in matrix-gamma (multivariate) Poisson mixture models; here the most interesting finding is given and described in Lemma 4.48 and the text before.

Conventions, Abbreviations, Symbols and Notation

Conventions

- In this thesis *positive* and *negative* are meant in the strict sense, i.e. they do not include zero.
- When writing displayed formulas, we follow in most cases the custom to omit the words *for all* or the symbol $\forall$ when quantifying variables.
- The union over an empty index set is the empty set: $\bigcup_{i\in\varnothing} A_i := \varnothing$.
- The intersection over an empty index set is the full set (which depends on the context).
- The sum over an empty index set equals 0.
- The product over an empty index set equals 1.

Abbreviations

- a.s., almost surely (with respect to a probability measure)
- e.g., Latin 'exempli gratia' (translation: for example)
- i.e., Latin 'id est' (translations: that means, in other words, in this case)
- i.i.d., independent and identically distributed
- w.r.t., with respect to

Symbols

- $\subsetneq$, strict subset
- $\subseteq$, subset (equality is possible)
- $\supsetneq$, strict superset (i.e. contains as a strict subset)
- $\supseteq$, superset (i.e. contains as a subset, equality is possible)
- $\prec$, in this thesis only used for the total order defined in Remark 4.26
- $\leq_L$, Loewner partial order, see Definition 2.10
- $\cup$, $\bigcup$, union of sets
- $\cap$, $\bigcap$, intersection of sets
- $\times$, $\bigtimes$, Cartesian product of sets

- $|\cdot|$, absolute value on $\mathbb{R}$ and $\mathbb{C}$, cardinality of a set, occasionally used as 1-norm, also called Manhattan norm, over $\mathbb{N}_0^d$ with $d \in \mathbb{N}$
- $\|\cdot\|$, norm on a vector space, specified further when used
- $\|\cdot\|_2$, Euclidian norm on $\mathbb{R}^d$ and $\mathbb{C}^d$ with $d \in \mathbb{N}$
- $\|\cdot\|_{\mathrm{F}}$, Frobenius norm, see Remark 2.17
- $(\cdot)^{\mathsf{c}}$, complement of a set
- $\langle\cdot,\cdot\rangle$, standard inner product on $\mathbb{R}^d \times \mathbb{R}^d$ and $\mathbb{C}^d \times \mathbb{C}^d$ with $d \in \mathbb{N}$
- $\langle\cdot,\cdot\rangle_{\mathrm{F}}$, Frobenius inner product, see Remark 2.17
- $\sim$, distributed according to
- $\circ$, composition of functions
- $\otimes$, Kronecker product, see Definition 3.1
- $\mathbb{1}_A$, indicator function which is 1 on A and 0 on the complement
- 1_I, function which is equal to 1 when all conditions of the set I are fulfilled and equal to 0 otherwise
- $(\cdot)^{\mathsf{H}}$, Hermitian transpose of a vector or matrix
- $(\cdot)^{\mathsf{T}}$, transpose of a vector or matrix
- $(\cdot)^{\triangle}$, lower triangular version of a positive semi-definite matrix, see Remark 2.19
- $\square$, end of a proof

Notation

Greek letters are ordered alphabetically according to their spelling in the Latin alphabet.

- $\overset{\text{a.s.}}{=}$, almost surely equal
- $\mathcal{B}_S$, Borel σ-algebra of the topological space S
- $\mathbb{C}$, the field of complex numbers
- χ_n^2, χ^2-distribution with $n \in \mathbb{N}$ degrees of freedom, see Definition 2.28
- $\mathrm{Cov}(X, Y)$, covariance matrix of the random vectors X and Y viewed as column vectors
- $\mathrm{Cov}(X)$, covariance matrix $\mathrm{Cov}(X, X)$
- $\overset{\text{d}}{=}$, equality in distribution
- det, determinant (of a square matrix)
- $\mathrm{diag}(v)$, diagonal matrix with the entries of the vector $v \in \mathbb{K}^d$ on the diagonal.
- e, Euler's number 2.71828 . . ., also used for the exponential function
- $\mathbb{E}[X\mathbb{1}_A] = \int_A X \,\mathrm{d}\mathbb{P} = \int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$, expectation, Lebesgue integral
- $\mathbb{E}[X \,|\, \mathcal{G}]$, conditional expectation of X given the sub-σ-algebra $\mathcal{G}$

- $\int f\,\mathrm{d}\mu$, integral for the function f w.r.t. the measure μ
- exp, exponential function
- f', derivative of the function f
- $f^{(n)}$, n-th derivative of the function f for $n \in \mathbb{N}$
- $\mathcal{G}_C$, where C is a positive semi-definite matrix, see Eq. (3.38)
- $\Gamma(\cdot)$, one-dimensional gamma function, see Remark 2.24
- $\Gamma_p(\cdot)$, p-dimensional gamma function for $p \in \mathbb{N}$, see Eq. (2.29)
- $\mathrm{Gamma}(\alpha, \beta)$, one-dimensional gamma distribution, see Definition 2.25
- $\mathrm{GenGamma}(\alpha, \beta, p)$, generalized gamma distribution, see Definition 3.34
- i, imaginary unit in the field $\mathbb{C}$ of complex numbers
- $\mathbb{1}_A$, indicator function of the set A
- I_d, the $(d \times d)$-identity matrix
- $\mathbb{K}$, either the field $\mathbb{R}$ of real or the field $\mathbb{C}$ of complex numbers
- $\lambda^k(\cdot)$, k-dimensional Lebesgue–Borel measure on $\mathbb{R}^k$ for $k \in \mathbb{N}$
- $\mathcal{L}(X)$, the law (or probability distribution) of the random variable $X\colon \Omega \to S$, i.e. the pushforward measure $\mathbb{P}X^{-1}$ of $\mathbb{P}$ on the measurable space $(S, \mathcal{S})$
- $\log(\cdot)$, natural logarithm on $\mathbb{R}^+$
- $\mathrm{Log}(\cdot)$, principal value logarithm on $\mathbb{C}\backslash\{0\}$, see Remark 2.31
- $\mathrm{MExp}(C)$, matrix-valued exponential distribution, see Definition 3.47
- $\mathrm{MGamma}(\alpha, C)$, matrix-valued gamma distribution, see Definition 3.24
- $\mathrm{MGammaMPoisson}(\alpha, C, G, (v_g)_{g\in G}, m)$, matrix-gamma Poisson mixture model, see Definition 4.37
- $\mathrm{MGammaPoisson}(\alpha, C, v_1, \dots, v_n)$, matrix-gamma Poisson mixture model, see Definition 4.17
- $\mathrm{MNormal}(M, \Sigma \otimes \Psi)$, matrix-valued normal distribution, see Definition 3.2
- $\mathrm{MPoisson}(G, \lambda, m)$, multivariate Poisson distribution, see Definition 4.12
- $\mathbb{N} = \{1, 2, 3, \dots\}$, natural numbers (without zero)
- $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, the additive semi-group of natural numbers including zero
- $\mathrm{NegBin}(\alpha, p)$, negative binomial distribution, see Definition 4.4
- $\mathrm{NegMult}(\alpha, p_1, \dots, p_n)$, negative multinomial distribution, see Definition 4.23
- $\mathcal{N}(\mu, \sigma^2)$, normal distribution with expectation $\mu \in \mathbb{R}$ and variance $\sigma^2 \geq 0$
- $\mathcal{N}(\mu, C)$, multivariate normal distribution with expectation vector $\mu \in \mathbb{R}^d$ and covariance matrix $C \in \mathbb{R}^{d\times d}$
- Ω, sample space
- $\mathbb{P}$, probability measure on the measurable space $(\Omega, \mathcal{F})$

- $\mathbb{P}_X$, X-biased probability measure, see Definition 3.62
- $\mathcal{P}(\Omega)$, power set, i.e. the set of all subsets of Ω
- π, length of the perimeter of a circle with diameter 1
- Poisson(λ), Poisson distribution, see Definition 4.3
- $\psi(\cdot)$, if not specified otherwise it refers to Eq. (2.7)
- $\mathbb{Q}$, the field of rational numbers or a probability measure, depending on the context
- $\mathbb{R}$, the field of real numbers
- $\mathbb{R}^+ = (0, \infty)$, set of positive real numbers
- $\mathbb{R}_0^+ = [0, \infty)$, set of non-negative real numbers
- $\mathbb{S}_p^+$, the set of positive definite matrices in $\mathbb{R}^{p \times p}$
- $\tilde{\mathbb{S}}_p^+$, the set of positive definite matrices in $\mathbb{R}^{p \times p}$ transformed to their lower triangular version and vectorized, see Remark 2.19
- Sym(Ω), symmetric group on the set Ω
- $\mathbb{T}_p^{\mathbb{K}}$, the set of all $\mathbb{K}^{p \times p}$-valued lower triangular matrices with positive entries on the diagonal
- $\tilde{\mathbb{T}}_p$, the set $\mathbb{T}_p^{\mathbb{R}}$ seen as a subset of $\mathbb{R}^{p(p+1)/2}$, only considering entries on and below the diagonal
- $\operatorname{tr}(A)$, trace of a square matrix A
- $\operatorname{Var}(X)$, variance of a real-valued random variable X
- $\operatorname{vec}(\cdot)$, vectorization function, see Eq. (2.4)
- $\operatorname{vecp}(\cdot)$, partial vectorization function, see Eq. (2.6)
- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the commutative ring of integers

List of Figures

Bibliography

[1] W. Adkins and S. Weintraub. *Algebra: An Approach via Module Theory*. Graduate Texts in Mathematics. Springer, 2012. 5

[2] H. Amann and J. Escher. *Analysis III*. Grundstudium Mathematik. Birkhäuser, Basel, 2008. 44

[3] M. Bluemlinger. *Analysis 3*. 2019. TU Wien. 24

[4] G. Box and M. Muller. A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, 29(2):610–611, June 1958. 39

[5] C. Chatfied and A. Collins. *Introduction to Multivariate Analysis*. Springer, US, 2013. 43

[6] G. Constantine and T. Savits. A multivariate Fàa di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996. 93

[7] O. Deiser and C. Lasser. *Erste Hilfe in Linearer Algebra: Überblick und Grundwissen mit vielen Abbildungen und Beispielen*. Springer, Berlin, Heidelberg, 2015. 27

[8] M. Denuit et al. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. John Wiley & Sons, 2007. 108

[9] L. Fabrykovski. Extended Credit Risk$^+$ with Guarantees. Master's thesis, TU Wien, Vienna, 2017. 107, 108, 110, 113

[10] O. Forster. *Analysis 3: Maß- und Integrationstheorie, Integralsätze im* $\mathbb{R}^n$ *und Anwendungen*. Springer, Wiesbaden, 8. edition, 2017. 11, 14

[11] A. Graham. *Kronecker Products and Matrix Calculus with Applications*. Dover Publications, 2018. 68

[12] J. Guillaume et al. Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, 119:418–432, 2019. 69, 86

[13] A. Gupta and D. Nagar. *Matrix Variate Distributions*. Taylor & Francis, 1999. 12, 23, 24, 35, 36, 37, 43, 52, 54, 111, 113, 114

[14] H. Havlicek. *Lineare Algebra für Technische Mathematiker*. Heldermann Verlag, 2012. 70, 90, 91

[15] C. Herz. Bessel functions of matrix argument. *Annals of Mathematics*, 61(3):474–523, 1955. 113

[16] P. Hoff. *A First Course in Bayesian Statistical Methods*. Springer, New York, 2009. 43

[17] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985. 3, 5, 7, 9, 11, 35, 44, 62

[18] J. Howie. *Complex Analysis*. Springer, London, 2005. 16, 19, 22

[19] M. Kaltenbäck. *Fundamental Analysis*. Heldermann Verlag, 2015. 25

[20] T. Kozubowski, S. Mazur, and K. Podgórski. *Matrix Gamma Distributions and Related Stochastic Processes*. Örebro University School of Business, 2022. 37, 39, 52, 62, 67, 113

[21] A. Mathai. *Jacobians of Matrix Transformations and Functions of Matrix Argument*. World Scientific Pub., 1997. 24, 25

[22] A. Mathai and S. Provost. On the singular gamma, Wishart, and beta matrix-variate density functions. *Canadian Journal of Statistics*, 50(4):1143–1165, 2022. 59, 114

[23] A. Mathai, S. Provost, and H. Haubold. *Multivariate Statistical Analysis in the Real and Complex Domains*. Springer, 2022. 59, 111, 114

[24] L. Nannen. Numerische Mathematik B. TU Wien, 2019. 8

[25] S. Peddada and D. Richards. Proof of a conjecture of M. L. Eaton on the characteristic function of the Wishart distribution. *The Annals of Probability*, 19(2):868–874, 1991. 56

[26] C. Pozrikidis. *An Introduction to Grids, Graphs, and Networks*. Oxford University Press, USA, 2014. 4

[27] C. Rudolph. A generalization of Panjer's recursion for dependent claim numbers and an approximation of Poisson mixture models, 2014. TU Wien. 54

[28] U. Schmock. Lecture Notes in Stochastic Analysis for Financial and Actuarial Mathematics. TU Wien, June 2023. 11, 17, 36, 57

[29] U. Schmock. Modelling Dependent Credit Risks with Extensions to CreditRisk$^+$ and Application to Operational Risk (Lecture Notes). TU Wien, June 2023. 1, 75, 79, 80, 81, 82, 91, 99, 112, 113

[30] M. Srivastava. Singular Wishart and multivariate beta distributions. *The Annals of Statistics*, 31(5):1537 – 1560, 2003. 111

[31] M. S. Srivastava. On the Complex Wishart Distribution. *The Annals of Mathematical Statistics*, 36(1):313–315, 1965. 59

[32] H. Uhlig. On singular Wishart and singular multivariate beta distributions. *The Annals of Statistics*, pages 395–405, 1994. 111

[33] K. Viraswami. On multivariate gamma distributions. Master's thesis, McGill University, Montreal, 1991. 48, 113

Index