AppliedMath

MDPI

Article

Local Normal Approximations and Probability Metric Bounds for the Matrix-Variate *T* Distribution and Its Application to Hotelling's *T* Statistic

Frédéric Ouimet [1,2]

[1] Department of Mathematics and Statistics, McGill University, Montreal, QC H3A 0B9, Canada; frederic.ouimet2@mcgill.ca
[2] Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA

Abstract: In this paper, we develop local expansions for the ratio of the centered matrix-variate *T* density to the centered matrix-variate normal density with the same covariances. The approximations are used to derive upper bounds on several probability metrics (such as the total variation and Hellinger distance) between the corresponding induced measures. This work extends some previous results for the univariate Student distribution to the matrix-variate setting.

Keywords: asymptotic statistics; expansion; Hotelling's *T*-squared statistic; Hotelling's *T* statistic; matrix-variate normal distribution; local approximation; matrix-variate *T* distribution; normal approximation; Student distribution; *T* distribution; total variation

MSC: Primary: 62E20; Secondary: 60F99

Citation: Ouimet, F. Local Normal Approximations and Probability Metric Bounds for the Matrix-Variate *T* Distribution and Its Application to Hotelling's *T* Statistic. *AppliedMath* **2022**, 2, 446–456. https://doi.org/10.3390/appliedmath2030025

Academic Editor: Tommi Sottinen

Received: 22 June 2022
Accepted: 21 July 2022
Published: 1 August 2022

1. Introduction

For any $n \in \mathbb{N}$, define the space of (real symmetric) positive definite matrices of size $n \times n$ as follows:

$$\mathcal{S}_{++}^n := \{\mathbb{M} \in \mathbb{R}^{n\times n} : \mathbb{M} \text{ is symmetric and positive definite}\}.$$

For $d, m \in \mathbb{N}$, $\nu > 0$, $M \in \mathbb{R}^{d\times m}$, $\Sigma \in \mathcal{S}_{++}^d$ and $\Omega \in \mathcal{S}_{++}^m$, the density function of the centered (and normalized) matrix-variate *T* distribution, hereafter denoted by $T_{d,m}(\nu, \Sigma, \Omega)$, is defined, for all $X \in \mathbb{R}^{d\times m}$, by

$$K_{\nu,\Sigma,\Omega}(X) := \frac{\Gamma_d(\frac{1}{2}(\nu+m+d-1))}{\Gamma_d(\frac{1}{2}(\nu+d-1))} \frac{|\mathrm{I}_d + \nu^{-1}\Sigma^{-1}X\Omega^{-1}X^\top|^{-(\nu+m+d-1)/2}}{(\nu\pi)^{md/2}|\Sigma|^{m/2}|\Omega|^{d/2}}, \tag{1}$$

(see, e.g., (Definition 4.2.1 in [1])) where ν is the number of degrees of freedom, and

$$\begin{aligned}\Gamma_d(z) &= \int_{\mathbb{S}\in\mathcal{S}_{++}^d} |\mathbb{S}|^{z-(d+1)/2} \exp(-\mathrm{tr}(\mathbb{S}))\mathrm{d}\mathbb{S} \\ &= \pi^{d(d-1)/4} \prod_{j=1}^{d} \Gamma\left(z - \frac{j-1}{2}\right), \quad \Re(z) > \frac{d-1}{2},\end{aligned}$$

denotes the multivariate gamma function—see, e.g., (Section 35.3 in [2]) and [3]—and

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t, \quad \Re(z) > 0,$$

is the classical gamma function. The mean and covariance matrix for the vectorization of $T \sim T_{d,m}(\nu, \Sigma, \Omega)$, namely

$$\mathrm{vec}(T) := (T_{11}, T_{21}, \ldots, T_{d1}, T_{12}, T_{22}, \ldots, T_{d2}, \ldots, T_{1m}, T_{2m}, \ldots, T_{dm})^\top,$$

($\mathrm{vec}(\cdot)$ is the operator that stacks the columns of a matrix on top of each other) are known to be (see, e.g., Theorem 4.3.1 in [1], but be careful of the normalization):

$$\mathbb{E}[\mathrm{vec}(T)] = \mathbf{0}_{dm} \ \ (\text{i.e., } \mathbb{E}[T] = 0_{d\times m}),$$

and

$$\mathbb{V}\mathrm{ar}(\mathrm{vec}(T^\top)) = \frac{\nu}{(\nu-2)}\Sigma \otimes \Omega, \quad \nu > 2.$$

The first goal of our paper (Theorem 1) is to establish an asymptotic expansion for the ratio of the centered matrix-variate T density (1) to the centered matrix-variate normal (MN) density with the same covariances. According to (Gupta and Nagar [1], Theorem 2.2.1), the density of the $\mathrm{MN}_{d,m}(0_{d\times m}, \Sigma \otimes \Omega)$ distribution is

$$g_{\Sigma,\Omega}(X) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}X\Omega^{-1}X^\top\right)\right)}{(2\pi)^{md/2}|\Sigma|^{m/2}|\Omega|^{d/2}}, \quad X \in \mathbb{R}^{d\times m}. \tag{2}$$

The second goal of our paper (Theorem 2) is to apply the log-ratio expansion from Theorem 1 to derive upper bounds on multiple probability metrics between the measures induced by the centered matrix-variate T distribution and the corresponding centered matrix-variate normal distribution. In the special case $m = 1$, this gives us probability metric upper bounds between the measure induced by Hotelling's T statistic and the associated matrix-normal measure.

To give some practical motivations for the MN distribution (2), note that noise in the estimate of individual voxels of diffusion tensor magnetic resonance imaging (DT-MRI) data has been shown to be well modeled by a symmetric form of the $\mathrm{MN}_{3\times 3}$ distribution in [4–6]. The symmetric MN voxel distributions were combined into a tensor-variate normal distribution in [7,8], which could help to predict how the whole image (not just individual voxels) changes when shearing and dilation operations are applied in image wearing and registration problems; see Alexander et al. [9]. In [10], maximum likelihood estimators and likelihood ratio tests are developed for the eigenvalues and eigenvectors of a form of the symmetric MN distribution with an orthogonally invariant covariance structure, both in one-sample problems (for example, in image interpolation) and two-sample problems (when comparing images) and under a broad variety of assumptions. This work extended significantly the previous results of Mallows [11]. In [10], it is also mentioned that the polarization pattern of cosmic microwave background (CMB) radiation measurements can be represented by 2×2 positive definite matrices; see the primer by Hu and White [12]. In a very recent and interesting paper, Vafaei Sadr and Movahed [13] presented evidence for the Gaussianity of the local extrema of CMB maps. We can also mention [14], where finite mixtures of skewed MN distributions were applied to an image recognition problem.

In general, we know that the Gaussian distribution is an attractor for sums of i.i.d. random variables with finite variance, which makes many estimators in statistics asymptotically normal. Similarly, we expect the MN distribution (2) to be an attractor for sums of i.i.d. random matrices with finite variances (Hotelling's T-squared statistic is the most natural example), thus including many estimators, such as sample covariance matrices and score statistics for matrix parameters. In particular, if a given statistic or estimator is a function of the components of a sample covariance matrix for i.i.d. observations coming from a multivariate Gaussian population, then we could study its large sample properties (such as its moments) using Theorem 1 (for example, by turning a Student-moments estimation problem into a Gaussian-moments estimation problem).

The following is a brief outline of the paper. Our main results are stated in Section 2 and proven in Section 3. Technical moment calculations are gathered in Appendix A.

Notation 1. *Throughout the paper, $a = \mathcal{O}(b)$ means that $\limsup |a/b| < C$ as $\nu \to \infty$, where $C > 0$ is a universal constant. Whenever C might depend on some parameter, we add a subscript (for example, $a = \mathcal{O}_d(b)$). Similarly, $a = \mathrm{o}(b)$ means that $\lim |a/b| = 0$, and subscripts indicate which parameters the convergence rate can depend on. If $a = (1 + \mathrm{o}(1))b$, then we write $a \sim b$. The notation $\mathrm{tr}(\cdot)$ will denote the trace operator for matrices and $|\cdot|$ their determinant. For a matrix $\mathbb{M} \in \mathbb{R}^{d\times d}$ that is diagonalizable, $\lambda_1(\mathbb{M}) \geq \cdots \geq \lambda_d(\mathbb{M})$ will denote its eigenvalues, and we let $\boldsymbol{\lambda}(\mathbb{M}) := (\lambda_1(\mathbb{M}), \ldots, \lambda_d(\mathbb{M}))^\top$.*

2. Main Results

In Theorem 1 below, we prove an asymptotic expansion for the ratio of the centered matrix-variate T density to the centered matrix-variate normal (MN) density with the same covariances. The case $d = m = 1$ was proven recently in [15] (see also [16] for an earlier rougher version). The result extends significantly the convergence in distribution result from Theorem 4.3.4 in [1].

Theorem 1. *Let $d, m \in \mathbb{N}$, $\Sigma \in \mathcal{S}^d_{++}$ and $\Omega \in \mathcal{S}^m_{++}$ be given. Pick any $\eta \in (0,1)$ and let*

$$B_{\nu,\Sigma,\Omega}(\eta) := \left\{ X \in \mathbb{R}^{d\times m} : \max_{1\leq j\leq d} \frac{\delta_{\lambda_j}}{\sqrt{\nu-2}} \leq \eta\, \nu^{-1/4} \right\}$$

denote the bulk of the centered matrix-variate T distribution, where

$$\Delta_X := \Sigma^{-1/2} X \Omega^{-1/2} \quad \text{and} \quad \delta_{\lambda_j} := \sqrt{\frac{\nu-2}{\nu}\, \lambda_j(\Delta_X \Delta_X^\top)}, \quad 1 \leq j \leq d.$$

Then, as $\nu \to \infty$ and uniformly for $X \in B_{\nu,\Sigma,\Omega}(\eta)$, we have

$$\begin{aligned}
&\log\left(\frac{[\nu/(\nu-2)]^{md/2}\, K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})} \right) \\
&\qquad = \nu^{-1} \left\{ \begin{array}{l} \frac{1}{4}\mathrm{tr}\big((\Delta_X\Delta_X^\top)^2\big) - \frac{(m+d+1)}{2}\mathrm{tr}\big(\Delta_X\Delta_X^\top\big) \\ + \frac{md(m+d+1)}{4} \end{array} \right\} \\
&\qquad + \nu^{-2} \left\{ \begin{array}{l} -\frac{1}{6}\mathrm{tr}\big((\Delta_X\Delta_X^\top)^3\big) + \frac{(m+d-1)}{4}\mathrm{tr}\big((\Delta_X\Delta_X^\top)^2\big) \\ + \frac{md}{24}(13 - 2d^2 - 3d(-3+m) + 9m - 2m^2) \end{array} \right\} \\
&\qquad + \nu^{-3} \left\{ \begin{array}{l} \frac{1}{8}\mathrm{tr}\big((\Delta_X\Delta_X^\top)^4\big) - \frac{(m+d-1)}{6}\mathrm{tr}\big((\Delta_X\Delta_X^\top)^3\big) \\ + \frac{md}{24}\left(\begin{array}{l} 26 + d^3 + 2d^2(-3+m) + 11m \\ -6m^2 + m^3 + d(11 - 9m + 2m^2) \end{array} \right) \end{array} \right\} \\
&\qquad + \mathcal{O}_{d,m,\eta}\left(\frac{1 + \mathrm{tr}\big((\Delta_X\Delta_X^\top)^5\big)}{\nu^4} \right).
\end{aligned} \tag{3}$$

Local approximations such as the one in Theorem 1 can be found for the Poisson, binomial and negative binomial distributions in [17] (based on Fourier analysis results from [18]), and [19] for the binomial distribution. Another approach, using Stein's method, is used to study the variance-gamma distribution in [20]. Moreover, Kolmogorov and Wasserstein distance bounds are derived in [21,22] for the Laplace and variance-gamma distributions.

Below, we provide numerical evidence (displayed graphically) for the validity of the expansion in Theorem 1 when $d = m = 2$. We compare three levels of approximation for various choices of $\mathbb{S}$. For any given $\mathbb{S} \in \mathcal{S}_{++}^{d}$, define

$$E_0 := \sup_{X \in B_{\nu,\Sigma,\Omega}(\nu^{-1/4})} \left| \log \left(\frac{[\nu/(\nu-2)]^{md/2} K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})} \right) \right|, \tag{4}$$

$$\begin{aligned} E_1 := \sup_{X \in B_{\nu,\Sigma,\Omega}(\nu^{-1/4})} \Bigg| & \log \left(\frac{[\nu/(\nu-2)]^{md/2} K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})} \right) \\ & - \nu^{-1} \left\{ \frac{1}{4} \mathrm{tr}\left((\Delta_X \Delta_X^\top)^2 \right) - \frac{(m+d+1)}{2} \mathrm{tr}\left(\Delta_X \Delta_X^\top \right) + \frac{md(m+d+1)}{4} \right\} \Bigg|, \end{aligned}$$

$$\begin{aligned} E_2 := \sup_{X \in B_{\nu,\Sigma,\Omega}(\nu^{-1/4})} \Bigg| & \log \left(\frac{[\nu/(\nu-2)]^{md/2} K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})} \right) \\ & - \nu^{-1} \left\{ \frac{1}{4} \mathrm{tr}\left((\Delta_X \Delta_X^\top)^2 \right) - \frac{(m+d+1)}{2} \mathrm{tr}\left(\Delta_X \Delta_X^\top \right) + \frac{md(m+d+1)}{4} \right\} \\ & - \nu^{-2} \left\{ \begin{array}{l} -\frac{1}{6} \mathrm{tr}\left((\Delta_X \Delta_X^\top)^3 \right) + \frac{(m+d-1)}{4} \mathrm{tr}\left((\Delta_X \Delta_X^\top)^2 \right) \\ + \frac{md}{24}(13 - 2d^2 - 3d(-3+m) + 9m - 2m^2) \end{array} \right\} \Bigg|. \end{aligned}$$

In the R software [23], we use Equation (7) to evaluate the log-ratios inside E_0, E_1 and E_2.

Note that $X \in B_{\nu,\Sigma,\Omega}(\nu^{-1/4})$ implies $|\mathrm{tr}((\Delta_X \Delta_X^\top)^k)| \leq d$ for all $k \in \mathbb{N}$, so we expect from Theorem 1 that the maximum errors above (E_0, E_1 and E_2) will have the asymptotic behavior

$$E_i = \mathcal{O}_d(\nu^{-(1+i)}), \quad \text{for all } i \in \{0,1,2\},$$

or, equivalently,

$$\liminf_{\nu \to \infty} \frac{\log E_i}{\log(\nu^{-1})} \geq 1 + i, \quad \text{for all } i \in \{0,1,2\}. \tag{5}$$

The property (5) is verified in Figure 1 below, for $\Omega = \mathrm{I}_2$ and various choices of $\Sigma_{2\times 2}$. Similarly, the corresponding log-log plots of the errors as a function of ν are displayed in Figure 2. The simulations are limited to the range $5 \leq \nu \leq 1005$. The R code that generated Figures 1 and 2 can be found at Supplementary Material.

As a consequence of the previous theorem, we can derive asymptotic upper bounds on several probability metrics between the probability measures induced by the centered matrix-variate T distribution (1) and the corresponding centered matrix-variate normal distribution (2). The distance between Hotelling's T statistic [24] and the corresponding matrix-variate normal distribution is obtained in the special case $m = 1$.

Theorem 2 (Probability metric upper bounds). *Let $d, m \in \mathbb{N}$, $\Sigma \in \mathcal{S}_{++}^{d}$ and $\Omega \in \mathcal{S}_{++}^{m}$ be given. Assume that $X \sim T_{d,m}(\nu, \Sigma, \Omega)$, $Y \sim \mathrm{MN}_{d,m}(0_{d\times m}, \Sigma \otimes \Omega)$, and let $\mathbb{P}_{\nu,\Sigma,\Omega}$ and $\mathbb{Q}_{\Sigma,\Omega}$ be the laws of X and $Y\sqrt{\nu/(\nu-2)}$, respectively. Then, as $\nu \to \infty$,*

$$\mathrm{dist}\,(\mathbb{P}_{\nu,\Sigma,\Omega}, \mathbb{Q}_{\Sigma,\Omega}) \leq \frac{C\,(md)^{3/2}}{\nu} \qquad \text{and} \qquad \mathcal{H}(\mathbb{P}_{\nu,\Sigma,\Omega}, \mathbb{Q}_{\Sigma,\Omega}) \leq \sqrt{\frac{2C\,(md)^{3/2}}{\nu}},$$

where $C > 0$ is a universal constant, $\mathcal{H}(\cdot,\cdot)$ denotes the Hellinger distance, and $\mathrm{dist}\,(\cdot,\cdot)$ can be replaced by any of the following probability metrics: total variation, Kolmogorov (or uniform) metric, Lévy metric, discrepancy metric, Prokhorov metric.

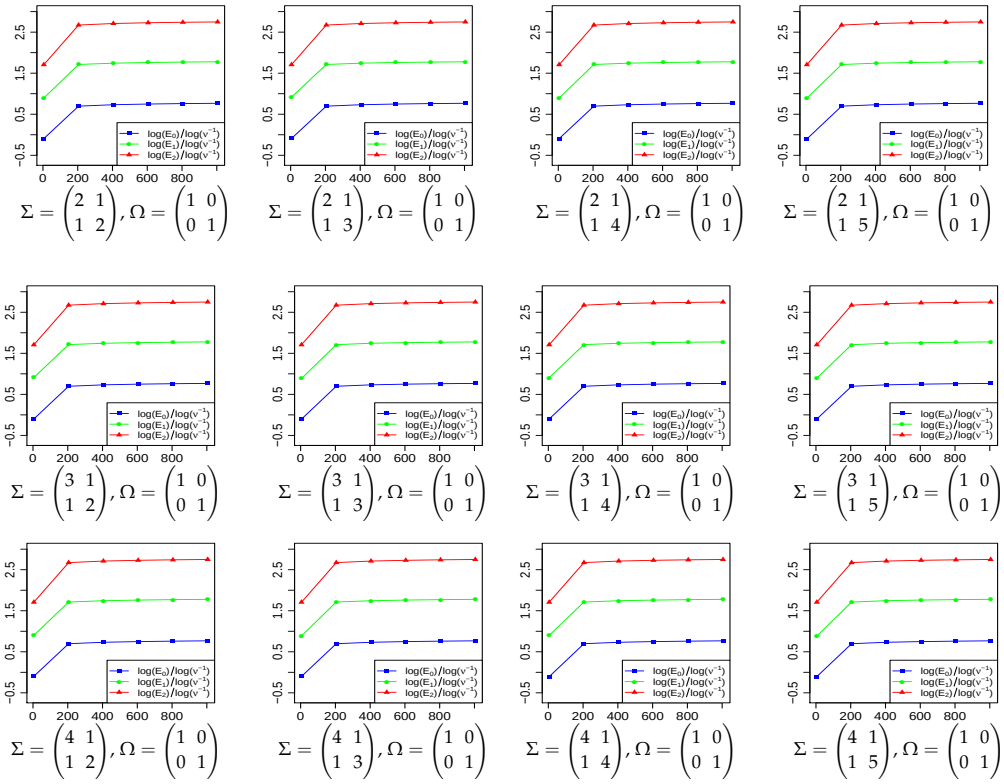

Figure 1. Plots of $\log E_i / \log(\nu^{-1})$ as a function of ν, for various choices of Σ. The plots confirm (5) for our choices of Σ and bring strong evidence for the validity of Theorem 1.

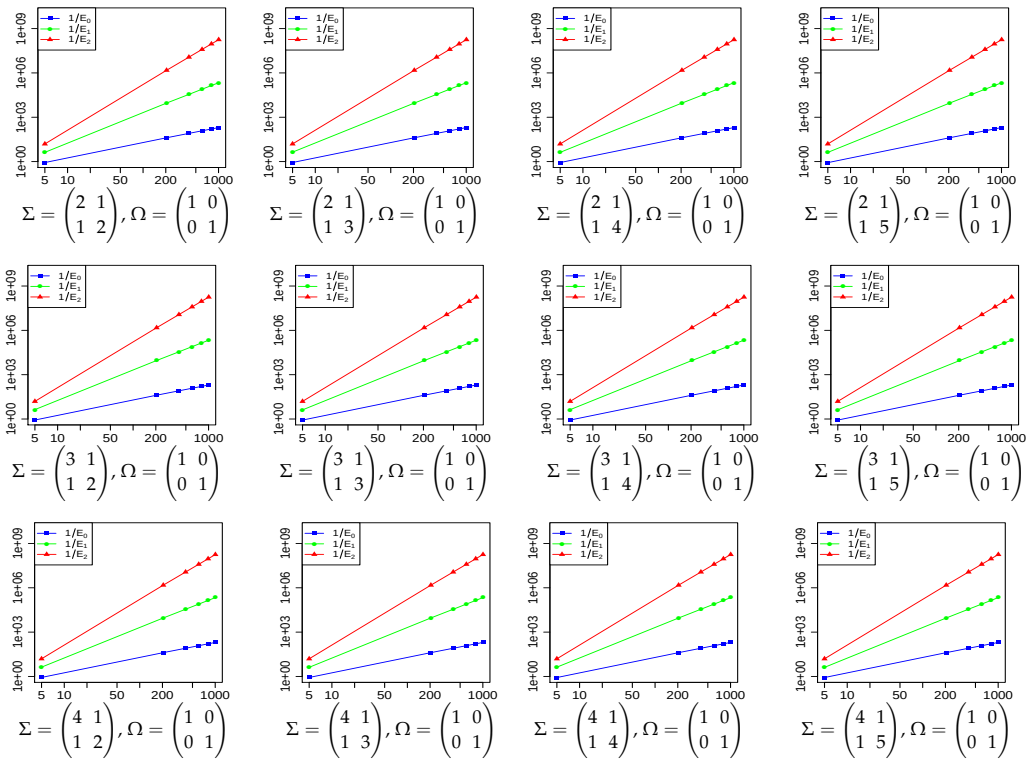

Figure 2. Plots of $1/E_i$ as a function of ν, for various choices of Σ. Both the horizontal and vertical axes are on a logarithmic scale. The plots clearly illustrate how the addition of correction terms from Theorem 1 to the base approximation (4) improves it.

3. Proofs

Proof of Theorem 1. First, we take the expression in (1) over the one in (2):

$$\begin{aligned}&\frac{[\nu/(\nu-2)]^{md/2}\,K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})}\\&\quad=\left[\frac{2}{\nu-2}\right]^{md/2}\prod_{j=1}^{d}\frac{\Gamma(\frac{1}{2}(\nu+m+d-j))}{\Gamma(\frac{1}{2}(\nu+d-j))}\\&\qquad\cdot\exp\left(\frac{(\nu-2)}{2\nu}\mathrm{tr}\left(\Delta_X\Delta_X^\top\right)\right)\left|\mathrm{I}_d+\nu^{-1}\Delta_X\Delta_X^\top\right|^{-(\nu+m+d-1)/2}.\end{aligned}\tag{6}$$

The last determinant was obtained using the fact that the eigenvalues of a product of rectangular matrices are invariant under cyclic permutations (as long as the products remain well defined). Indeed, for all $j\in\{1,2,\ldots,d\}$, we have

$$\begin{aligned}\lambda_j(\mathrm{I}_d+\nu^{-1}\Sigma^{-1}X\Omega^{-1}X^\top)&=1+\nu^{-1}\lambda_j(\Sigma^{-1}X\Omega^{-1}X^\top)\\&=1+\nu^{-1}\lambda_j(\Delta_X\Delta_X^\top)=\lambda_j(\mathrm{I}_d+\nu^{-1}\Delta_X\Delta_X^\top).\end{aligned}$$

By taking the logarithm on both sides of (6), we get

$$\begin{aligned}&\log\left(\frac{[\nu/(\nu-2)]^{md/2}\,K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})}\right)\\&\quad=-\frac{md}{2}\log\left(\frac{\nu-2}{2}\right)+\sum_{j=1}^{d}\left[\log\Gamma(\tfrac{1}{2}(\nu+m+d-j))-\log\Gamma(\tfrac{1}{2}(\nu+d-j))\right]\\&\qquad+\frac{1}{2}\sum_{j=1}^{d}\delta_{\lambda_j}^2-\frac{(\nu+m+d-1)}{2}\sum_{j=1}^{d}\log\left(1+\left(\frac{\delta_{\lambda_j}}{\sqrt{\nu-2}}\right)^2\right).\end{aligned}\tag{7}$$

By applying the Taylor expansions,

$$\begin{aligned}&\log\Gamma(\tfrac{1}{2}(\nu+m+d-j))-\log\Gamma(\tfrac{1}{2}(\nu+d-j))\\&=\frac{1}{2}(\nu+m+d-j-1)\log\left(\frac{1}{2}(\nu+m+d-j)\right)-\frac{1}{2}(\nu+d-j-1)\log\left(\frac{1}{2}(\nu+d-j)\right)\\&\quad-\frac{m}{2}+\frac{2}{12(\nu+m+d-j)}-\frac{2}{12(\nu+d-j)}\\&\quad-\frac{2^3}{360(\nu+m+d-j)^3}+\frac{2^3}{360(\nu+d-j)^3}+\mathcal{O}_{m,d}(\nu^{-4})\\&=\frac{m}{2}\log\left(\frac{\nu}{2}\right)+\frac{m(-2+2d-2j+m)}{4\nu}-\frac{m}{12\nu^2}\left\{\begin{array}{l}2+3d^2+3j^2-3j(-2+m)\\-3m+m^2+d(-6-6j+3m)\end{array}\right\}\\&\quad+\frac{m}{24\nu^3}\left\{\begin{array}{l}4d^3-4j^3-6d^2(2+2j-m)+6j^2(-2+m)+(-2+m)^2m\\-4j(2-3m+m^2)+4d(2+3j^2-3j(-2+m)-3m+m^2)\end{array}\right\}+\mathcal{O}_{m,d}(\nu^{-4}).\end{aligned}$$

(see, e.g., (Ref. [25], p. 257)) and

$$-\frac{md}{2}\log\left(\frac{\nu-2}{2}\right)+\frac{md}{2}\log\left(\frac{\nu}{2}\right)=\frac{4md}{4\nu}+\frac{12md}{12\nu^2}+\frac{32md}{24\nu^3}+\mathcal{O}_{m,d}(\nu^{-4}),$$

and

$$\log(1+y)=y-\frac{1}{2}y^2+\frac{1}{3}y^3-\frac{1}{4}y^4+\mathcal{O}_\eta(y^5),\quad |y|<\eta<1,$$

in the above equation, we obtain

$$\begin{aligned}
&\log\left(\frac{[\nu/(\nu-2)]^{md/2}\,K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})}\right)\\
&=\sum_{j=1}^{d}\frac{m(2+2d-2j+m)}{4\nu}-\sum_{j=1}^{d}\frac{m}{12\nu^2}\left\{\begin{array}{l}-10+3d^2+3j^2-3j(-2+m)\\-3m+m^2+d(-6-6j+3m)\end{array}\right\}\\
&\quad+\sum_{j=1}^{d}\frac{m}{24\nu^3}\left\{\begin{array}{l}32+4d^3-4j^3-6d^2(2+2j-m)+6j^2(-2+m)+(-2+m)^2m\\-4j(2-3m+m^2)+4d(2+3j^2-3j(-2+m)-3m+m^2)\end{array}\right\}\\
&\quad+\frac{1}{2}\sum_{j=1}^{d}\delta_{\lambda_j}^2-\frac{(\nu+m+d-1)}{2}\sum_{j=1}^{d}\left(\frac{\delta_{\lambda_j}}{\sqrt{\nu-2}}\right)^2\\
&\quad+\frac{(\nu+m+d-1)}{4}\sum_{j=1}^{d}\left(\frac{\delta_{\lambda_j}}{\sqrt{\nu-2}}\right)^4-\frac{(\nu+m+d-1)}{6}\sum_{j=1}^{d}\left(\frac{\delta_{\lambda_j}}{\sqrt{\nu-2}}\right)^6\\
&\quad+\frac{(\nu+m+d-1)}{8}\sum_{j=1}^{d}\left(\frac{\delta_{\lambda_j}}{\sqrt{\nu-2}}\right)^8+\mathcal{O}_{d,m,\eta}\left(\frac{1+\max_{1\le j\le d}|\delta_{\lambda_j}|^{10}}{\nu^4}\right).
\end{aligned}\tag{8}$$

Now,

$$\begin{aligned}
\frac{1}{2}-\frac{\nu+m+d-1}{2(\nu-2)}&=-\frac{(m+d+1)}{2\nu}-\frac{(m+d+1)}{\nu^2}-\frac{2(m+d+1)}{\nu^3}+\mathcal{O}_{m,d}(\nu^{-4}),\\
\frac{\nu+m+d-1}{4(\nu-2)^2}&=\frac{1}{4\nu}+\frac{(m+d+3)}{4\nu^2}+\frac{(m+d+2)}{\nu^3}+\mathcal{O}_{m,d}(\nu^{-4}),\\
-\frac{\nu+m+d-1}{6(\nu-2)^3}&=-\frac{1}{6\nu^2}-\frac{(m+d+5)}{6\nu^3}+\mathcal{O}_{m,d}(\nu^{-4}),\\
\frac{\nu+m+d-1}{8(\nu-2)^4}&=\frac{1}{8\nu^3}+\mathcal{O}_{m,d}(\nu^{-4}),
\end{aligned}$$

so we can rewrite (8) as

$$\begin{aligned}
&\log\left(\frac{[\nu/(\nu-2)]^{md/2}\,K_{\nu,\Sigma,\Omega}(X)}{g_{\Sigma,\Omega}(X/\sqrt{\nu/(\nu-2)})}\right)\\
&=\nu^{-1}\sum_{j=1}^{d}\left\{\frac{1}{4}\delta_{\lambda_j}^4-\frac{(m+d+1)}{2}\delta_{\lambda_j}^2+\frac{m(2+2d-2j+m)}{4}\right\}\\
&\quad+\nu^{-2}\sum_{j=1}^{d}\left\{\begin{array}{l}-\frac{1}{6}\delta_{\lambda_j}^6+\frac{(m+d+3)}{4}\delta_{\lambda_j}^4-(m+d+1)\delta_{\lambda_j}^2\\-\frac{m}{12}\left\{\begin{array}{l}-10+3d^2+3j^2-3j(-2+m)\\-3m+m^2+d(-6-6j+3m)\end{array}\right\}\end{array}\right\}\\
&\quad+\nu^{-3}\sum_{j=1}^{d}\left\{\begin{array}{l}\frac{1}{8}\delta_{\lambda_j}^8-\frac{(m+d+5)}{6}\delta_{\lambda_j}^6+(m+d+2)\delta_{\lambda_j}^4-2(m+d+1)\delta_{\lambda_j}^2\\+\frac{m}{24}\left\{\begin{array}{l}32+4d^3-4j^3-6d^2(2+2j-m)+6j^2(-2+m)+(-2+m)^2m\\-4j(2-3m+m^2)+4d(2+3j^2-3j(-2+m)-3m+m^2)\end{array}\right\}\end{array}\right\}\\
&\quad+\mathcal{O}_{d,m,\eta}\left(\frac{1+\max_{1\le j\le d}|\delta_{\lambda_j}|^{10}}{\nu^4}\right),
\end{aligned}$$

which proves (3) after some simplifications with `Mathematica`. □

Proof of Theorem 2. By the comparison of the total variation norm $\|\cdot\|$ with the Hellinger distance on page 726 of Carter [26], we already know that

$$\begin{aligned}&\|\mathbb{P}_{\nu,\Sigma,\Omega} - \mathbb{Q}_{\Sigma,\Omega}\| \\ &\leq \sqrt{2\,\mathbb{P}\Big(X \in B^c_{\nu,\Sigma,\Omega}(1/2)\Big) + \mathbb{E}\left[\log\left(\frac{\mathrm{d}\mathbb{P}_{\nu,\Sigma,\Omega}}{\mathrm{d}\mathbb{Q}_{\Sigma,\Omega}}(X)\right)\mathbb{1}_{\{X\in B_{\nu,\Sigma,\Omega}(1/2)\}}\right]}.\end{aligned} \tag{9}$$

Given that $\Delta_X = \Sigma^{-1/2}X\Omega^{-1/2} \sim T_{d,m}(\nu, \mathrm{I}_d, \mathrm{I}_m)$ by Theorem 4.3.5 in [1], we know, by Theorem 4.2.1 in [1], that

$$\Delta_X \overset{\text{law}}{=} (\nu^{-1}\mathbb{S})^{-1/2}Z,$$

for $\mathbb{S} \sim \text{Wishart}_{d\times d}(\nu + d - 1, \mathrm{I}_d)$ and $Z \sim \text{MN}_{d\times m}(0_{d\times m}, \mathrm{I}_d \otimes \mathrm{I}_m)$ that are independent, so that, by Theorems 3.3.1 and 3.3.3 in [1], we have

$$\Delta_X\Delta_X^\top \,|\, \mathbb{S} \sim \text{Wishart}_{d\times d}(m, \nu\,\mathbb{S}^{-1}). \tag{10}$$

Therefore, by conditioning on $\mathbb{S}$, and then by applying the sub-multiplicativity of the largest eigenvalue for nonnegative definite matrices, and a large deviation bound on the maximum eigenvalue of a Wishart matrix (which is sub-exponential), we get, for ν large enough,

$$\begin{aligned}\mathbb{P}\Big(X \in B^c_{\nu,\Sigma,\Omega}(1/2)\Big) &\leq \mathbb{E}\left[\mathbb{P}\left(\lambda_1(\Delta_X\Delta_X^\top) > \frac{\nu^{1/2}}{4}\,\middle|\,\mathbb{S}\right)\right] \\ &\leq \mathbb{E}\left[\mathbb{P}\left(\lambda_1((\nu^{-1}\mathbb{S})^{-1/2})\lambda_1(ZZ^\top)\lambda_1((\nu^{-1}\mathbb{S})^{-1/2}) > \frac{\nu^{1/2}}{4}\,\middle|\,\mathbb{S}\right)\right] \\ &= \mathbb{E}\left[\mathbb{P}\left(\lambda_1(ZZ^\top) > \frac{\lambda_d(\mathbb{S})}{4\,\nu^{1/2}}\,\middle|\,\mathbb{S}\right)\right] \\ &\leq C_{m,d}\,\exp\left(-\frac{\nu^{1/2}}{10^4 md}\right),\end{aligned} \tag{11}$$

for some positive constant $C_{m,d}$ that depends only on m and d. By Theorem 1, we also have

$$\begin{aligned}&\mathbb{E}\left[\log\left(\frac{\mathrm{d}\mathbb{P}_{\nu,\Sigma,\Omega}}{\mathrm{d}\mathbb{Q}_{\Sigma,\Omega}}(X)\right)\mathbb{1}_{\{X\in B_{\nu,\Sigma,\Omega}(1/2)\}}\right] \\ &= \nu^{-1}\left\{\begin{array}{l}\frac{1}{4}\cdot\mathbb{E}\big[\mathrm{tr}\big((\Delta_X\Delta_X^\top)^2\big)\big] \\ -\frac{(m+d+1)}{2}\cdot\mathbb{E}\big[\mathrm{tr}(\Delta_X\Delta_X^\top)\big] + \frac{md(m+d+1)}{4}\end{array}\right\} \\ &\quad + \nu^{-1}\left\{\begin{array}{l}\mathcal{O}\Big(\mathbb{E}\Big[\mathrm{tr}\big((\Delta_X\Delta_X^\top)^2\big)\mathbb{1}_{\{X\in B_{\nu,\Sigma,\Omega}(1/2)\}}\Big]\Big) \\ +(m+d)\,\mathcal{O}\Big(\mathbb{E}\Big[\mathrm{tr}(\Delta_X\Delta_X^\top)\mathbb{1}_{\{X\in B_{\nu,\Sigma,\Omega}(1/2)\}}\Big]\Big) + \mathcal{O}(m(m+d))\end{array}\right\} \\ &\quad + \nu^{-2}\left\{\begin{array}{l}\mathcal{O}\big(\mathbb{E}\big[\mathrm{tr}\big((\Delta_X\Delta_X^\top)^3\big)\big]\big) + (m+d)\,\mathcal{O}\big(\mathbb{E}\big[\mathrm{tr}\big((\Delta_X\Delta_X^\top)^2\big)\big]\big) \\ +(m+d)\,\mathcal{O}\big(\mathbb{E}\big[\mathrm{tr}(\Delta_X\Delta_X^\top)\big]\big) + \mathcal{O}(md(m+d)^2)\end{array}\right\}.\end{aligned} \tag{12}$$

On the right-hand side, the first line is estimated using Lemma A1, and the second line is bounded using Lemma A2. We find

$$\mathbb{E}\left[\log\left(\frac{\mathrm{d}\mathbb{P}_{\nu,\Sigma,\Omega}}{\mathrm{d}\mathbb{Q}_{\Sigma,\Omega}}(X)\right)\mathbb{1}_{\{X\in B_{\nu,\Sigma,\Omega}(1/2)\}}\right] = \mathcal{O}(m^3d^3\nu^{-2}).$$

Putting (11) and (12) together in (9) gives the conclusion. □

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/appliedmath2030025/s1.

Funding: F.O. is supported by postdoctoral fellowships from the NSERC (PDF) and the FRQNT (B3X supplement and B3XR).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The R code for the simulations in Section 2 is in Supplementary Material.

Acknowledgments: We thank the three referees for their comments.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Technical Computations

Below, we compute the expectations for some traces of powers of the matrix-variate Student distribution. The lemma is used to estimate some trace moments and the $\asymp \nu^{-2}$ errors in (12) of the proof of Theorem 2, and also as a preliminary result for the proof of Lemma A2.

Lemma A1. *Let $d, m \in \mathbb{N}$, $\Sigma \in \mathcal{S}_{++}^{d}$ and $\Omega \in \mathcal{S}_{++}^{m}$ be given. If $X \sim T_{d,m}(\nu, \Sigma, \Omega)$ according to (1), then*

$$\mathbb{E}\Big[\mathrm{tr}\Big(\Delta_X \Delta_X^\top\Big)\Big] = \frac{md\,\nu}{\nu-2}, \tag{A1}$$

$$\mathbb{E}\Big[\mathrm{tr}\Big((\Delta_X \Delta_X^\top)^2\Big)\Big] = \frac{md\,\nu^2\,\{(m+d)(\nu-2)+\nu+md\}}{(\nu-1)(\nu-2)(\nu-4)}, \tag{A2}$$

where we recall $\Delta_X := \Sigma^{-1/2} X \Omega^{-1/2}$. In particular, as $\nu \to \infty$, we have

$$\mathbb{E}\Big[\mathrm{tr}\Big(\Delta_X \Delta_X^\top\Big)\Big] \sim md \quad \textit{and} \quad \mathbb{E}\Big[\mathrm{tr}\Big((\Delta_X \Delta_X^\top)^2\Big)\Big] \sim md(m+d+1).$$

Proof of Lemma A1. For $\mathbb{W} \sim \mathrm{Wishart}_{d\times d}(n, \mathbb{V})$ with $n > 0$ and $\mathbb{V} \in \mathcal{S}_{++}^{d}$, we know from (Ref. [1], p. 99) (alternatively, see (Ref. [27], p. 66) or (Ref. [28], p. 308)) that

$$\mathbb{E}[\mathbb{W}] = n\,\mathbb{V} \quad \text{and} \quad \mathbb{E}[\mathbb{W}^2] = n\,\{(n+1)\,\mathbb{V} + \mathrm{tr}(\mathbb{V})\,\mathrm{I}_d\}\,\mathbb{V},$$

and from (Ref. [1], pp. 99–100) (alternatively, see [29] and ([28], p. 308), or ([30], pp. 101–103)) that

$$\mathbb{E}[\mathbb{W}^{-1}] = \frac{\mathbb{V}}{n-d-1}, \quad \text{for } n-d-1>0,$$

$$\mathbb{E}[\mathbb{W}^{-2}] = \frac{\mathrm{tr}(\mathbb{V}^{-1})\,\mathbb{V}^{-1} + (n-d-1)\,\mathbb{V}^{-2}}{(n-d)(n-d-1)(n-d-3)}, \quad \text{for } n-d-3>0,$$

and from (Corollary 3.1 in [30]) that

$$\mathbb{E}[\mathrm{tr}(\mathbb{W}^{-1})\,\mathbb{W}^{-1}] = \frac{(n-d-2)\,\mathrm{tr}(\mathbb{V}^{-1})\,\mathbb{V}^{-1} + 2\,\mathbb{V}^{-2}}{(n-d)(n-d-1)(n-d-3)}, \quad \text{for } n-d-3>0.$$

Therefore, by combining the above moment estimates with (10), we have

$$\mathbb{E}\Big[\Delta_X \Delta_X^\top\Big] = \mathbb{E}\Big[\mathbb{E}[\Delta_X \Delta_X^\top \,|\, \mathbb{S}]\Big] = \mathbb{E}[m\,(\nu\,\mathbb{S}^{-1})] = m\,\nu\,\mathbb{E}[\mathbb{S}^{-1}] = \frac{m\,\nu}{\nu-2}\,\mathrm{I}_d,$$

$$\begin{aligned}\mathbb{E}\Big[(\Delta_X \Delta_X^\top)^2\Big] &= \mathbb{E}\Big[\mathbb{E}[(\Delta_X \Delta_X^\top)^2 \,|\, \mathbb{S}]\Big] = \mathbb{E}\Big[m\,\Big\{(m+1)\,(\nu\,\mathbb{S}^{-1}) + \mathrm{tr}(\nu\,\mathbb{S}^{-1})\,\mathrm{I}_d\Big\}\,(\nu\,\mathbb{S}^{-1})\Big] \\ &= m\,\nu^2\,\Big\{(m+1)\,\mathbb{E}[\mathbb{S}^{-2}] + \mathbb{E}[\mathrm{tr}(\mathbb{S}^{-1})\,\mathbb{S}^{-1}]\Big\}\end{aligned}$$

$$= \frac{m\,\nu^2\,\{(m+1)\,(\nu+d-2)+(\nu-3)\,d+2\}}{(\nu-1)(\nu-2)(\nu-4)}\,\mathrm{I}_d,$$

By linearity, the trace of an expectation is the expectation of the trace, so (A1) and (A2) follow from the above equations. □

We can also estimate the moments of Lemma A1 on various events. The lemma below is used to estimate the $\asymp \nu^{-1}$ errors in (12) of the proof of Theorem 2.

Lemma A2. *Let $d, m \in \mathbb{N}$, $\Sigma \in \mathcal{S}^d_{++}$ and $\Omega \in \mathcal{S}^m_{++}$ be given, and let $A \in \mathscr{B}(\mathbb{R}^{d\times m})$ be a Borel set. If $X \sim T_{d,m}(\nu, \Sigma, \Omega)$ according to (1), then, for ν large enough,*

$$\left|\mathbb{E}\left[\mathrm{tr}(\Delta_X\Delta_X^\top)\mathbb{1}_{\{X\in A\}}\right]\right| \le 2\,md^{3/2}\,(\mathbb{P}(X\in A^c))^{1/2}, \tag{A3}$$

$$\left|\mathbb{E}\left[\mathrm{tr}\left((\Delta_X\Delta_X^\top)^2\right)\mathbb{1}_{\{X\in A\}}\right] - \frac{md\,\nu^2\{(m+d)(\nu-2)+\nu+md\}}{(\nu-1)(\nu-2)(\nu-4)}\right| \le 100\,m^2d^{5/2}\,(\mathbb{P}(X\in A^c))^{1/2}, \tag{A4}$$

where we recall $\Delta_X := \Sigma^{-1/2}X\Omega^{-1/2}$.

Proof of Lemma A2. By Lemma A1, the Cauchy–Schwarz inequality and Jensen's inequality,

$$(\mathrm{tr}(\Delta_X\Delta_X^\top))^2 \le d\cdot\mathrm{tr}((\Delta_X\Delta_X^\top)^2),$$

we have

$$\begin{aligned}\left|\mathbb{E}\left[\mathrm{tr}(\Delta_X\Delta_X^\top)\mathbb{1}_{\{X\in A\}}\right]\right| &= \left|\mathbb{E}\left[\mathrm{tr}(\Delta_X\Delta_X^\top)\mathbb{1}_{\{X\in A^c\}}\right]\right| \\ &\le \left(\mathbb{E}\left[(\mathrm{tr}(\Delta_X\Delta_X^\top))^2\right]\right)^{1/2}(\mathbb{P}(X\in A^c))^{1/2} \\ &\le \left(d\cdot\mathbb{E}\left[\mathrm{tr}((\Delta_X\Delta_X^\top)^2)\right]\right)^{1/2}(\mathbb{P}(X\in A^c))^{1/2} \\ &\le 2\,md^{3/2}\,(\mathbb{P}(X\in A^c))^{1/2},\end{aligned}$$

which proves (A3). Similarly, by Lemma A1, Holder's inequality and Jensen's inequality,

$$(\mathrm{tr}((\Delta_X\Delta_X^\top)^2))^2 \le d\,\mathrm{tr}((\Delta_X\Delta_X^\top)^4),$$

we have, for ν large enough,

$$\begin{aligned}&\left|\mathbb{E}\left[\mathrm{tr}((\Delta_X\Delta_X^\top)^2)\mathbb{1}_{\{X\in A\}}\right] - \frac{md\,\nu^2\,\{(m+d)(\nu-2)+\nu+md\}}{(\nu-1)(\nu-2)(\nu-4)}\right| \\ &= \left|\mathbb{E}\left[\mathrm{tr}((\Delta_X\Delta_X^\top)^2)\mathbb{1}_{\{X\in A^c\}}\right]\right| \le \left(\mathbb{E}\left[(\mathrm{tr}((\Delta_X\Delta_X^\top)^2))^2\right]\right)^{1/2}(\mathbb{P}(X\in A^c))^{1/2} \\ &\le \left(d\,\mathbb{E}\left[\mathrm{tr}((\Delta_X\Delta_X^\top)^4)\right]\right)^{1/2}(\mathbb{P}(X\in A^c))^{1/2} \\ &\le \left(d\,10^4(md)^4\right)^{1/2}(\mathbb{P}(X\in A^c))^{1/2} \le 100\,m^2d^{5/2}\,(\mathbb{P}(X\in A^c))^{1/2},\end{aligned}$$

which proves (A4). This ends the proof. □

References

1. Gupta, A.K.; Nagar, D.K. *Matrix Variate Distributions*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 1999; p. 384.
2. Olver, F.W.J.; Lozier, D.W.; Boisvert, R.F.; Clark, C.W. (Eds.) *NIST Handbook of Mathematical Functions*; U.S. Department of Commerce, National Institute of Standards and Technology: Washington, DC, USA; Cambridge University Press: Cambridge, UK, 2010; p. xvi+951.
3. Nagar, D.K.; Roldán-Correa, A.; Gupta, A.K. Extended matrix variate gamma and beta functions. *J. Multivar. Anal.* **2013**, *122*, 53–69. [CrossRef]
4. Pajevic, S.; Basser, P.J. Parametric description of noise in diffusion tensor MRI. In Proceedings of the 7th Annual Meeting of the ISMRM, Philadelphia, PA, USA, 22–28 May 1999; p. 1787.
5. Basser, P.J.; Jones, D.K. Diffusion-tensor MRI: Theory, experimental design and data analysis—A technical review. *NMR Biomed.* **2002**, *15*, 456–467. [CrossRef] [PubMed]
6. Pajevic, S.; Basser, P.J. Parametric and non-parametric statistical analysis of DT-MRI data. *J. Magn. Reson.* **2003**, *161*, 1–14. [CrossRef]
7. Basser, P.J.; Pajevic, S. A normal distribution for tensor-valued random variables: Applications to diffusion tensor MRI. *IEEE Trans. Med. Imaging* **2003**, *22*, 785–794. [CrossRef] [PubMed]
8. Gasbarra, D.; Pajevic, S.; Basser, P.J. Eigenvalues of random matrices with isotropic Gaussian noise and the design of diffusion tensor imaging experiments. *SIAM J. Imaging Sci.* **2017**, *10*, 1511–1548. [CrossRef] [PubMed]
9. Alexander, D.C.; Pierpaoli, C.; Basser, P.J.; Gee, J.C. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans. Med. Imaging* **2001**, *20*, 1131–1139. [CrossRef] [PubMed]
10. Schwartzman, A.; Mascarenhas, W.F.; Taylor, J.E. Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann. Statist.* **2008**, *36*, 2886–2919. [CrossRef]
11. Mallows, C.L. Latent vectors of random symmetric matrices. *Biometrika* **1961**, *48*, 133–149. [CrossRef]
12. Hu, W.; White, M. A CMB polarization primer. *New Astron.* **1997**, *2*, 323–344. [CrossRef]
13. Vafaei Sadr, A.; Movahed, S.M.S. Clustering of local extrema in Planck CMB maps. *MNRAS* **2021**, *503*, 815–829. [CrossRef]
14. Gallaugher, M.P.B.; McNicholas, P.D. Finite mixtures of skewed matrix variate distributions. *Pattern Recognit.* **2018**, *80*, 83–93. [CrossRef]
15. Ouimet, F. Refined normal approximations for the Student distribution. *J. Classical Anal.* **2022**, *20*, 23–33. [CrossRef]
16. Shafiei, A.; Saberali, S.M. A simple asymptotic bound on the error of the ordinary normal approximation to the Student's *t*-distribution. *IEEE Commun. Lett.* **2015**, *19*, 1295–1298. [CrossRef]
17. Govindarajulu, Z. Normal approximations to the classical discrete distributions. *Sankhyā Ser. A* **1965**, *27*, 143–172.
18. Esseen, C.G. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **1945**, *77*, 1–125. [CrossRef]
19. Cressie, N. A finely tuned continuity correction. *Ann. Inst. Statist. Math.* **1978**, *30*, 435–442. [CrossRef]
20. Gaunt, R.E. Variance-gamma approximation via Stein's method. *Electron. J. Probab.* **2014**, *19*, 1–33. [CrossRef]
21. Gaunt, R.E. New error bounds for Laplace approximation *via* Stein's method. *ESAIM Probab. Stat.* **2021**, *25*, 325–345. [CrossRef]
22. Gaunt, R.E. Wasserstein and Kolmogorov error bounds for variance-gamma approximation via Stein's method I. *J. Theoret. Probab.* **2020**, *33*, 465–505. [CrossRef]
23. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
24. Hotelling, H. The generalization of Student's ratio. *Ann. Math. Statist.* **1931**, *2*, 360–378. [CrossRef]
25. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; National Bureau of Standards Applied Mathematics Series, For sale by the Superintendent of Documents; U.S. Government Printing Office: Washington, DC, USA, 1964; Volume 55, p. xiv+1046.
26. Carter, A.V. Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.* **2002**, *30*, 708–730. [CrossRef]
27. de Waal, D.J.; Nel, D.G. On some expectations with respect to Wishart matrices. *South African Statist. J.* **1973**, *7*, 61–67.
28. Letac, G.; Massam, H. All invariant moments of the Wishart distribution. *Scand. J. Statist.* **2004**, *31*, 295–318. [CrossRef]
29. Haff, L.R. An identity for the Wishart distribution with applications. *J. Multivar. Anal.* **1979**, *9*, 531–544. [CrossRef]
30. von Rosen, D. Moments for the inverted Wishart distribution. *Scand. J. Statist.* **1988**, *15*, 97–109. [CrossRef]