Article

The Sampling Distribution of the Total Correlation for Multivariate Gaussian Random Variables

Taylor Rowe and Troy Day *

Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada;
tlrowe1005@gmail.com
* Correspondence: day@queensu.ca

check for updates

Abstract: The sampling distribution of the total correlation (TC) for a d-dimensional standardized multivariate Gaussian random variable with an identity covariance matrix is derived. It is shown to be the distribution of a sum of generalized beta random variables. It is also shown that, for large dimension and sample size, a central limit theorem holds, providing a Gaussian approximation to the sampling distribution for high dimensional data.

1. Introduction

Mutual information quantifies the information shared between two random variables [1–3]. This concept can be been generalized to d variables in a variety of ways [4–7], with the most direct generalization being Watanabe's total correlation (TC),

$$T(\mathbf{X}) \equiv \sum_{i=1}^{d} h(X_i) - h(\mathbf{X}) \tag{1}$$

where $\mathbf{X}$ is a vector whose components are the d random variables $X_1, \dots, X_d$, and for continuous random variables, $h(X_i)$ is the differential entropy of X_i and $h(\mathbf{X})$ is the joint differential entropy of $\mathbf{X}$.

Total correlation is also sometimes called multivariate mutual information, and it is the Kullback–Leibler divergence between the joint density of $\mathbf{X}$ and the density obtained by taking the product of the marginal densities of the X_i. Thus, the total correlation $T(\mathbf{X})$ quantifies, in a quite general sense, the information shared among all the d random variables. The total correlation is non-negative and in the case where all d random variables are mutually independent we have $T(\mathbf{X}) = 0$ [7,8]. For the special case where $\mathbf{X}$ is multivariate Gaussian with arbitrary mean and covariance matrix Σ, the total correlation can be written explicitly as

$$T(\mathbf{X}) = \frac{1}{2} \sum_{i=1}^{d} \log \sigma_{ii}^2 - \frac{1}{2} \log |\Sigma| \tag{2}$$

where σ_{ij}^2 is the ijth entry of Σ. When the X_i are independent we have $\sigma_{ij}^2 = 0$ for all $i \neq j$ and so $\log |\Sigma| = \log \sigma_{11}^2 \sigma_{22}^2 \cdots \sigma_{dd}^2$, giving $T(\mathbf{X}) = 0$ in Equation (2) as expected.

The total correlation provides a natural way to quantify dependencies among a set of random variables. For example, often we seek to determine if a set of random variables are mutually independent because dependency among variables can indicate interesting and meaningful relationships in nature. To do so one can take a sample from the unknown distribution and compute

the total correlation from this sample. Even if the random variables *are* mutually independent, however, the total correlation measured using such a finite sample will typically be positive (rather than zero) simply because of sampling variation. Therefore, it is of interest to know the sampling distribution of the total correlation under independence. Once we have the sampling distribution we can then perform statistical tests of independence. Here we derive the sampling distribution of (2) in the case where the X_i are standardized (i.e., zero mean, unit variance), independent, Gaussian random variables.

Previous authors have proposed exact expressions for the mean and variance of the sample total correlation [9,10]. In fact, Guerrero (Section 2.1 of [9]) derived a moment generating function for the sample total correlation using the distribution of the log-determinant of a Wishart matrix (see Wilks [11,12]). Unfortunately the asymptotic approximation of Guerrero's result does not match the results of Marralec [10] suggesting that one of the two is incorrect. We will resolve this discrepancy by deriving the moment generating function directly from our expression for the probability density function of the sample total correlation. In the limit of large sample size our results match those presented in Section 4.1 of Marralec [10], suggesting that the moment generating function of [9] is incorrect.

2. Definitions and Preliminaries

Let **X** represent a d-variate zero mean Gaussian random variable with covariance matrix $\Sigma = I_d$ where I_d is the d-dimensional identity matrix. Let $\{x_1, \ldots, x_n\}$ denote a sample of n draws from the distribution of **X**. We focus on the case where $n \geq d$. The sample covariance matrix is $\hat{\Sigma} = (1/n) \sum_{i=1}^{n} x_i x_i' = \{\hat{\sigma}_{ij}\}$ and $n\hat{\Sigma}$ is Wishart distributed with n degrees of freedom, which we denote as $n\hat{\Sigma} \sim W(\Sigma, d, n)$. From Equation (2) the sample total correlation is then also a random variable and is computed as

$$\widehat{T}_{d,n}(X) = \frac{1}{2} \sum_{i=1}^{d} \log \hat{\sigma}_{ii}{}^2 - \frac{1}{2} \log |\hat{\Sigma}| \tag{3}$$

where the subscripts d and n indicate that $\widehat{T}$ is a family of random variables indexed by dimension and sample size.

Odell and Feiveson's 1966 result [13] provides a convenient way to characterize a Wishart-distributed matrix. Suppose that $V_i^{(n)}$ $(1 \leq i \leq d)$ are independent chi-square random variables with $n - i + 1$ degrees of freedom. Suppose that N_{ij} are independent standardized normal random variables for $1 \leq i < j \leq d$, also independent of every $V_i^{(n)}$. Now construct the random variables

$$
\begin{aligned}
b_{11} &= V_1^{(n)} \\
b_{jj} &= V_j^{(n)} + \sum_{i=1}^{j-1} N_{ij}^2, & 2 \leq j \leq d \\
b_{1j} &= N_{1j} \sqrt{V_1^{(n)}} & 2 \leq j \leq d \\
b_{ij} &= N_{ij} \sqrt{V_i^{(n)}} + \sum_{k=1}^{i-1} N_{ki} N_{kj}, & 2 < i < j \leq d.
\end{aligned}
\tag{4}
$$

Then the matrix $B = \{b_{ij}\}$ (with $b_{ij} = b_{ji}$) is Wishart-distributed $W(I_d, d, n)$ and thus we have

$$n\hat{\sigma}_{ii}^2 \sim b_{ii} \sim V_i^{(n)} + A_i \qquad 1 < 1 \leq d \tag{5}$$

where A_i are independent chi-square random variables with $i - 1$ degrees of freedom and we define $A_1 = 0$. Now following [14] we can also define the lower-triangular matrix $T = \{t_{ij}\}$ as

$$
\begin{aligned}
t_{ii} &= \sqrt{V_i^{(n)}} \\
t_{ij} &= N_{ji} \qquad 1 \leq j < i \leq d \\
t_{ij} &= 0 \qquad i < j \leq d
\end{aligned}
\tag{6}
$$

and thus $B = TT'$. Furthermore, $|B| = |TT'| = |T|^2 = \prod_{i=1}^{d} t_{ii}^2 = \prod_{i=1}^{d} V_i^{(n)}$, revealing that

$$
n^d |\hat{\Sigma}| \sim \prod_{i=1}^{d} V_i^{(n)}.
\tag{7}
$$

Result (7) is a special case of results found in Wilks [11]. For analogous results involving complex matrices see Goodman [15].

3. The Sampling Distribution of the Total Correlation

With the above preliminaries the we can now state the following theorem.

Theorem 1 (The Sampling Distribution of TC)**.** *Consider a sample of size n from a set of d independent, standardized, Gaussian random variables, with $n \geq d$. The total correlation (TC) is distributed as*

$$
\widehat{T}_{d,n}(\mathbf{X}) \sim \frac{1}{2} \sum_{i=1}^{d-1} \log\left(1 + \frac{i}{n-i} F_{i,n-i}\right)
\tag{8}
$$

where $F_{i,n-i}$ are independent F-distributed random variables with i and $n - i$ degrees of freedom. Equivalently, (8) can be written as

$$
\widehat{T}_{d,n}(\mathbf{X}) \sim \sum_{i=1}^{d-1} Y_{i,n}
\tag{9}
$$

where $Y_{i,n}$ is a beta-exponential random variable with probability density

$$
f_{Y_{i,n}}(y) = \lambda \frac{(1 - e^{-\lambda y})^{\frac{i}{2}-1}(e^{-\lambda y})^{\frac{n-i}{2}}}{B(\frac{i}{2}, \frac{n-i}{2})} \qquad y > 0
$$

having parameter $\lambda = 2$.

Proof. Writing Equation (3) as

$$
\widehat{T}_{d,n}(\mathbf{X}) = \frac{1}{2} \log \frac{\prod_{i=1}^{d} \hat{\sigma}_{ii}^2}{|\hat{\Sigma}|}
\tag{10}
$$

and using result (5) and (7) one obtains

$$
\begin{aligned}
\widehat{T}_{d,n}(\mathbf{X}) &\sim \frac{1}{2} \log \frac{\prod_{i=1}^{d}\left(V_i^{(n)} + A_i\right)}{\prod_{i=1}^{d} V_i^{(n)}} \\
&\sim \frac{1}{2} \log \prod_{i=1}^{d}\left(1 + \frac{A_i}{V_i^{(n)}}\right) \\
&\sim \frac{1}{2} \sum_{i=1}^{d} \log\left(1 + \frac{A_i}{V_i^{(n)}}\right).
\end{aligned}
$$

Scaling each chi-square by their corresponding degrees of freedom and re-indexing, yields (8). Equivalently, if we define $Y_{i,n} = \frac{1}{2}\log\left(1 + \frac{i}{n-i}F_{i,n-i}\right)$ then $\widehat{T}_{d,n}(\mathbf{X}) \sim \sum_{i=1}^{d-1} Y_{i,n}$, and using standard techniques it be can shown that the random variable $Y_{i,n}$ has probability density

$$f_{Y_{i,n}}(y) = 2\frac{(1 - e^{-2y})^{\frac{i}{2}-1}(e^{-2y})^{\frac{n-i}{2}}}{B(\frac{i}{2}, \frac{n-i}{2})} \qquad y > 0$$

where $B(x, y)$ is the beta function. $\square$

Corollary 1. *The moment generating function for $\widehat{T}_{d,n}(\mathbf{X})$ is*

$$M_{d,n}(t) = \left[\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-t}{2})}\right]^{d-1} \prod_{i=1}^{d-1} \frac{\Gamma(\frac{n-i-t}{2})}{\Gamma(\frac{n-i}{2})} \tag{11}$$

where $\Gamma(x)$ is the gamma function. The mean and variance of $\widehat{T}_{d,n}(\mathbf{X})$ are therefore

$$\mu_{d,n} = \frac{d-1}{2}\psi(n/2) - \frac{1}{2}\sum_{i=1}^{d-1}\psi\left(\frac{n-i}{2}\right)$$

$$\sigma_{d,n}^2 = -\frac{d-1}{4}\psi^{(1)}(n/2) + \frac{1}{4}\sum_{i=1}^{d-1}\psi^{(1)}\left(\frac{n-i}{2}\right) \tag{12}$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function and $\psi^{(k)}(x)$ denotes its kth derivative.

Proof. Taking $Y_{i,n} = \frac{1}{2}\log\left(1 + \frac{i}{n-i}F_{i,n-i}\right)$, the moment generating function for $Y_{i,n}$ is

$$\phi_{i,n}(t) = \mathbb{E}[e^{tY_{i,n}}] = \frac{\Gamma[\frac{n}{2}]\Gamma[\frac{n-i-t}{2}]}{\Gamma[\frac{n-i}{2}]\Gamma[\frac{n-t}{2}]}.$$

The random variables in the sum $\sum_{i=1}^{d-1} Y_{i,n}$ are independent, and therefore the moment generating function $M_{d,n}(t)$ for $\widehat{T}_{d,n}(\mathbf{X})$ is the appropriate product of the functions $\phi_{i,n}(t)$. Equation (12) then follow directly from the properties of moment generating functions. $\square$

Guerrero [9] obtained a formula for the mean and variance of $\widehat{T}_{d,n}(\mathbf{X})$ (except for a typo in the variance) using Wilks' [12] moment generating function for the generalized variance. These are remarkably close to (12), but the proposed moment generating function for the sample total correlation information provided in Guerrero [9] appears to be incorrect.

4. A Central Limit Theorem for the Total Correlation

Girkos central limit theorem [16] implies asymptotic normality of the sample log-determinant, as seen in the work of Bao et al., and Cai et al. [17,18]. This suggests the existence of a central limit theorem result for $\widehat{T}_{d,n}(\mathbf{X})$ when the dimension d and sample size n are large. Here we provide such a result.

Define the mean and variance of $Y_{i,n}$ as $m_{i,n} = \mathbb{E}[Y_{i,n}]$ and $s_{i,n}^2 = \mathbb{E}[(Y_{i,n} - \mu_{i,n})^2]$, and the mean-centered random variables $Y_{i,n}^* = Y_{i,n} - m_{i,n}$. Note that $\sigma_{d,n}^2 = \sum_{i=1}^{d-1} s_{i,n}^2$.

Theorem 2 (Asymptotic normality of TC)**.** *Suppose $n \to \infty$ and $d \to \infty$ in such a way that $n/d \to k$ where $1 \le k < \infty$. Then*

$$\frac{1}{\sigma_{d,n}^2}\sum_{i=1}^{d-1} Y_{i,n}^* \to N(0,1) \tag{13}$$

where convergence is in distribution. Thus, for large n and d (with $n \geq d$) the total correlation $\widehat{T}_{d,n}(X)$ is approximately normally distributed with mean and variance given by $\mu_{d,n}$ and $\sigma_{d,n}^2$ in Equations (12).

Proof. The $Y_{i,n}^*$ are a triangular array of random variables such that, for any fixed n the $Y_{i,n}^*$ ($1 \leq i \leq d-1$) are independent. Thus, (13) will hold provided that the Lyapunov condition is satisfied [19]; namely, that there exists a $\delta > 0$ such that

$$\lim_{d,n \to \infty} \frac{1}{\sigma_{d,n}^{2+\delta}} \sum_{i=1}^{d-1} \mathbb{E}[|Y_{i,n}^*|^{2+\delta}] = 0.$$

For $\delta = 2$ the entries in Lyapunov's summation represent each $Y_{i,n}$'s fourth central moment, for which the generating function is $C_{i,n}(t) = e^{-m_{i,n}t}\phi_{i,n}(t)$. The summation therefore becomes

$$\sum_{i=1}^{d-1} \mathbb{E}[(Y_{i,n}^*)^4] = \sum_{i=1}^{d-1} \frac{1}{16}\left(3\left(\psi^{(1)}(\frac{n-i}{2}) - \psi^{(1)}(n/2)\right)^2 + \psi^{(3)}(\frac{n-i}{2}) - \psi^{(3)}(n/2)\right)$$

$$= \frac{3}{16}\sum_{i=1}^{d-1}\left(\psi^{(1)}(\frac{n-i}{2}) - \psi^{(1)}(n/2)\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\psi^{(3)}(\frac{n-i}{2}) - \frac{d-1}{16}\psi^{(3)}(n/2)$$

while the denominator in Lyapunov's condition is

$$\sigma_{d,n}^4 = \left(\frac{1}{4}\sum_{i=1}^{d-1}\psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4}\psi^{(1)}(n/2)\right)^2.$$

In Appendix A we show that

$$0 \leq \frac{3}{16}\sum_{i=1}^{d-1}\left(\psi^{(1)}(\frac{n-i}{2}) - \psi^{(1)}(n/2)\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\psi^{(3)}(\frac{n-i}{2}) - \frac{d-1}{16}\psi^{(3)}(n/2) \leq \frac{48}{n-d+1}$$

and, for any fixed $1 \leq k < \infty$, and for sufficiently large d and n with n/d sufficiently close to k,

$$\frac{1}{4}\left(\ln\left(\frac{n}{n-d+1}\right) + \frac{d-1}{n(n-d+1)} - \frac{d-1}{2}\left(\frac{2}{n} + \frac{4}{n^2}\right)\right)^2 \leq \left(\frac{1}{4}\sum_{i=1}^{d-1}\psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4}\psi^{(1)}(n/2)\right)^2.$$

Therefore, for any fixed $1 \leq k < \infty$, and for sufficiently large d and n with n/d sufficiently close to k, we have

$$0 \leq \frac{1}{\sigma_{d,n}^4}\sum_{i=1}^{d-1}\mathbb{E}[|Y_{i,n}^*|^4] \leq \frac{\frac{48}{n-d+1}}{\frac{1}{4}\left(\ln\left(\frac{n}{n-d+1}\right) + \frac{d-1}{n(n-d+1)} - \frac{d-1}{2}\left(\frac{2}{n} + \frac{4}{n^2}\right)\right)^2}. \tag{14}$$

Now first consider the case where $n = d$ (and therefore $k = 1$). Then (14) simplifies to

$$0 \leq \frac{1}{\sigma_{d,n}^4}\sum_{i=1}^{d-1}\mathbb{E}[|Y_{i,n}^*|^4] \leq \frac{48}{\frac{1}{4}\left(\ln n + \frac{n-1}{n} - \frac{n-1}{2}\left(\frac{2}{n} + \frac{4}{n^2}\right)\right)^2}.$$

Taking the limit $n \to \infty$ yields zero on the right-hand side, verifying Lyapunov's condition for $k = 1$. Next, consider the case where $n > d$. Taking the limit in (14) as $n \to \infty$ and $d \to \infty$ in such a way that $n/d \to k$ where $1 < k < \infty$, again we see that the right-hand side is zero. This verifies Lyapunov's condition in the case where $k > 1$, thereby completing the proof. $\quad\square$

5. Conclusions

The total correlation of a multivariate random variable (sometimes called multivariate mutual information) is the Kullback–Leibler divergence between the joint density of the random variable and the product of its marginal densities. It therefore provides a natural measure of the degree of independence of a set of random variables. In this paper we derived the sampling distribution of the total correlation for a d-dimensional standardized multivariate Gaussian random variable with identity covariance matrix, and showed that it is the distribution of a sum of generalized beta random variables. We also proved that, for large dimension and sample size, a central limit theorem holds, providing a Gaussian approximation to the sampling distribution for high dimensional data.

Author Contributions: conceptualization, T.R. and T.D.; methodology, T.R. and T.D.; formal analysis, T.R. and T.D.; investigation, T.R. and T.D.; writing–original draft preparation, T.R.; writing–review and editing, T.D.

Appendix A

The proof of the central limit theorem result makes use of the following two lemmas. Both are based on an inequality for the digamma function found in [20] (where $m \geq 1$ is an integer)

$$\frac{(m-1)!}{x^m} + \frac{m!}{2x^{m+1}} \leq (-1)^{m+1}\psi^{(m)}(x) \leq \frac{(m-1)!}{x^m} + \frac{m!}{x^{m+1}}. \tag{A1}$$

Lemma A1. *Suppose $d \leq n$. Then the following inequality holds*

$$0 \leq \frac{3}{16}\sum_{i=1}^{d-1}\left(\psi^{(1)}(\frac{n-i}{2}) - \psi^{(1)}(n/2)\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\psi^{(3)}(\frac{n-i}{2}) - \frac{d-1}{16}\psi^{(3)}(n/2) \leq \frac{48}{n-d+1}.$$

Proof. The left-hand inequality follows from the fact that $\psi^{(1)}(x)$ and $\psi^{(3)}(x)$ are both monotonically decreasing functions and so $\psi^{(1)}(\frac{n-i}{2}) \geq \psi^{(1)}(\frac{n}{2})$ and $\psi^{(3)}(\frac{n-i}{2}) \geq \psi^{(3)}(\frac{n}{2})$ for all $1 \leq i \leq d-1$. For the right-hand inequality we have

$$\frac{3}{16}\sum_{i=1}^{d-1}\left(\psi^{(1)}(\frac{n-i}{2}) - \psi^{(1)}(n/2)\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\psi^{(3)}(\frac{n-i}{2}) - \frac{d-1}{16}\psi^{(3)}(n/2)$$

$$\leq \frac{3}{16}\sum_{i=1}^{d-1}\left(\psi^{(1)}(\frac{n-i}{2})\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\psi^{(3)}(\frac{n-i}{2})$$

$$\leq \frac{3}{16}\sum_{i=1}^{d-1}\left(\frac{2}{n-i} + \frac{4}{(n-i)^2}\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\left(\frac{16}{(n-i)^3} + \frac{96}{(n-i)^4}\right)$$

$$\leq \frac{3}{16}\sum_{i=1}^{d-1}\left(\frac{8}{n-i}\right)^2 + \frac{1}{16}\sum_{i=1}^{d-1}\frac{192}{(n-i)^3}$$

$$\leq 12\sum_{j=n-d+1}^{n-1}\left(\frac{1}{j^2} + \frac{1}{j^3}\right)$$

$$= 12\left(\frac{1}{(n-d+1)^2} + \frac{1}{(n-d+1)^3} + \sum_{j=n-d+2}^{n-1}\left(\frac{1}{j^2} + \frac{1}{j^3}\right)\right)$$

$$\leq 12\left(\frac{1}{(n-d+1)^2} + \frac{1}{(n-d+1)^3} + \int_{j=n-d+1}^{n-1}\left(\frac{1}{x^2} + \frac{1}{x^3}\right)dx\right)$$

$$= 12 \left(\frac{1}{(n-d+1)^2} + \frac{1}{(n-d+1)^3} + \frac{1}{n-d+1} - \frac{1}{n-1} + \frac{1}{2(n-d+1)} - \frac{1}{2(n-1)} \right)$$

$$\leq 12 \left(\frac{1}{(n-d+1)^2} + \frac{1}{(n-d+1)^3} + \frac{1}{n-d+1} + \frac{1}{2(n-d+1)} \right)$$

$$\leq \frac{48}{n-d+1}.$$

□

Lemma A2. *Suppose $d \leq n$. Then, for any fixed $1 \leq k \leq \infty$, and for sufficiently large d and n with n/d sufficiently close to k, the following inequality holds*

$$\frac{1}{4} \left(\ln \left(\frac{n}{n-d+1} \right) + \frac{d-1}{n(n-d+1)} - \frac{d-1}{2} \left(\frac{2}{n} + \frac{4}{n^2} \right) \right)^2 \leq \left(\frac{1}{4} \sum_{i=1}^{d-1} \psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4} \psi^{(1)}(n/2) \right)^2. \quad \text{(A2)}$$

Proof. First note that the quantity in the parentheses on the right-hand side is positive because $\psi^{(1)}(x)$ is a monotonically decreasing function and so $\psi^{(1)}(\frac{n-i}{2}) \geq \psi^{(1)}(\frac{n}{2})$ for all $1 \leq i \leq d-1$. Thus, if for some quantity A we have $0 \leq A \leq \frac{1}{4} \sum_{i=1}^{d-1} \psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4} \psi^{(1)}(n/2)$ then $A^2 \leq (\frac{1}{4} \sum_{i=1}^{d-1} \psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4} \psi^{(1)}(n/2))^2$. We construct such a quantity A as follows. First consider the summation term on the right-hand side of (A2). Using (A1) we have

$$\frac{1}{4} \sum_{i=1}^{d-1} \psi^{(1)}(\frac{n-i}{2}) \geq \frac{1}{4} \sum_{i=1}^{d-1} \left(\frac{2}{n-i} + \frac{2}{(n-i)^2} \right)$$

$$= \frac{1}{2} \sum_{j=n-d+1}^{n-1} \left(\frac{1}{j} + \frac{1}{j^2} \right)$$

$$\geq \frac{1}{2} \int_{n-d+1}^{n} \left(\frac{1}{x} + \frac{1}{x^2} \right) dx$$

$$= \frac{1}{2} \frac{d-1}{n(n-d+1)} + \frac{1}{2} \ln \left(\frac{n}{n-d+1} \right).$$

Using (A1) for the second term in parentheses on the right-hand side of (A2) gives

$$\frac{d-1}{4} \psi^{(1)}(n/2) \leq \frac{d-1}{4} \left(\frac{2}{n} + \frac{4}{n^2} \right).$$

Thus we have

$$\frac{1}{2} \ln \left(\frac{n}{n-d+1} \right) + \frac{1}{2} \frac{d-1}{n(n-d+1)} - \frac{d-1}{4} \left(\frac{2}{n} + \frac{4}{n^2} \right) \leq \frac{1}{4} \sum_{i=1}^{d-1} \psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4} \psi^{(1)}(n/2). \quad \text{(A3)}$$

It remains to be shown that the left-hand side of (A3) is non-negative. Taking the limit of the left-hand side of (A3) as d and n get large, and assuming $n/d \to k$ where $1 \leq k < \infty$, we obtain

$$\frac{1}{2} \ln \left(\frac{k}{k-1} \right) - \frac{1}{2k}$$

which is strictly positive for any fixed k. Thus, for any fixed $1 \leq k < \infty$ there exists values d^* and n^* such that for all $d > d^*$ and $n > n^*$ with n/d sufficiently close to k we have

$$0 \leq \frac{1}{2} \ln \left(\frac{n}{n-d+1} \right) + \frac{1}{2} \frac{d-1}{n(n-d+1)} - \frac{d-1}{4} \left(\frac{2}{n} + \frac{4}{n^2} \right). \quad \text{(A4)}$$

As a result, for any fixed $1 \le k < \infty$, and for all $d > d^*$ and $n > n^*$ with n/d sufficiently close to k, we have

$$\frac{1}{4}\left(\ln\left(\frac{n}{n-d+1}\right) + \frac{d-1}{n(n-d+1)} - \frac{d-1}{2}\left(\frac{2}{n} + \frac{4}{n^2}\right)\right)^2 \le \left(\frac{1}{4}\sum_{i=1}^{d-1}\psi^{(1)}(\frac{n-i}{2}) - \frac{d-1}{4}\psi^{(1)}(n/2)\right)^2. \quad \text{(A5)}$$

$\square$

References

1. Linfoot, E.H. An informational measure of correlation. *Inform. Contr.* **1957**, *1*, 85–89. [CrossRef]
2. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Champaign, IL, USA, 1949.
3. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 2012.
4. Watanabe, S. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* **1960**, *4*, 66–82. [CrossRef]
5. Garner, W.R. *Uncertainty and Structure as Psychological Concepts*; JohnWiley & Sons: New York, NY, USA, 1962.
6. Studený, M.; Vejnarová, J. The multiinformation function as a tool for measuring stochastic dependence. In *Learning in Graphical Models*; Springer: Berlin, Germany, 1998; pp. 261–297.
7. Joe, H. Relative entropy measures of multivariate dependence. *J. Am. Stat. Assoc.* **1989**, *84*, 157–164. [CrossRef]
8. Kullback, S. *Information Theory and Statistics*; Dover: New York, NY, USA, 1968.
9. Guerrero, J.L. Multivariate mutual information: Sampling distribution with applications. *Commun. Stat. Theory Methods* **1994**, *23*, 1319–1339. [CrossRef]
10. Marrelec, G.; Benali, H. Large-sample asymptotic approximations for the sampling and posterior distributions of differential entropy for multivariate normal distributions. *Entropy* **2011**, *13*, 805–819. [CrossRef]
11. Wilks, S.S. Moment-generating operators for determinants of product moments in samples from a normal system. *Ann. Math.* **1934**, *35*, 312–340. [CrossRef]
12. Wilks, S. Certain Generalizations in the Analysis of Variance. *Biometrika* **1932**, *24*, 471–494. [CrossRef]
13. Odell, P.L.; Feiveson, A.H. A numerical procedure to generate a sample covariance matrix. *J. Am. Stat. Assoc.* **1966**, *61*, 199–203. [CrossRef]
14. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed.; Wiley: New York, NY, USA, 2003.
15. Goodman, N.R. The distribution of the determinant of a complex Wishart distributed matrix. *Ann. Math. Stat.* **1963**, *34*, 178–180. [CrossRef]
16. Girko, V.L. A refinement of the central limit theorem for random determinants. *Theor. Probab. Appl.* **1998**, *42*, 121–129. [CrossRef]
17. Bao, Z.; Pan, G.; Zhou, W. The logarithmic law of random determinant. *Bernoulli* **2015**, *21*, 1600–1628. [CrossRef]
18. Cai, T.T.; Liang, T.; Zhou, H.H. Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. *J. Multivariate Anal.* **2015**, *137*, 161–172. [CrossRef]
19. Billiingsley, P. *Probability and Measure*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1995.
20. Guo, B.-N.; Qi, F. An extension of an inequality for ratios of gamma functions. *J. Approx. Theor.* **2011**, *163*, 1208–1216. [CrossRef]