Online supplement for article "Cumulants of multiinformation density in the case of a multivariate normal distribution"

1 Mutual independence, i_d and multiinformation

We here show the equivalence between the following

1. The $\boldsymbol{X}_n$'s are mutually independent;
2. $i_d \equiv 0$;
3. $I(\boldsymbol{X}_1; \ldots; \boldsymbol{X}_N) = 0$;
4. $\text{Var}(i_d) = 0$.

The equivalence between (1) and (2) is straightforward by definition of i_d. If $i_d \equiv 0$, then all its moments of i_d are equal to 0, including its expectation (multiinformation) and its variance, leading to $I(\boldsymbol{X}_1; \ldots; \boldsymbol{X}_N) = 0$ and $\text{Var}(i_d) = 0$.

Since multiinformation is a Kullback-Leibler divergence, $I(\boldsymbol{X}_1; \ldots; \boldsymbol{X}_N) = 0$ entails that we have

$$f(\boldsymbol{x}) = \prod_{n=1}^{N} f_n(\boldsymbol{x}_n)$$

for all $\boldsymbol{x}$, i.e., the $\boldsymbol{X}_n$'s are mutually independent.

Finally, if $\text{Var}(i_d) = 0$, then i_d is a constant, that is,

$$f(\boldsymbol{X}) = k \prod_{n=1}^{N} f_n(\boldsymbol{x}_n).$$

The fact that f and the f_n's are distributions, and therefore must norm to 1, entails that we must have $k = 1$, that is, $i_d \equiv 0$.

2 Positive-definiteness of $\mathbf{\Sigma}^{-1} - t\mathbf{\Phi}$

We need to show that $\mathbf{\Sigma}^{-1} - t\mathbf{\Phi}$ is a symmetric positive definite matrix in a neighborhood of $t = 0$. This matrix can be expressed as

$$\mathbf{\Sigma}^{-1} - t\mathbf{\Phi} = (1+t)\mathbf{\Sigma}^{-1} - t\,\mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{NN})^{-1}.$$

As a difference of two symmetric matrices, it is also symmetric. Furthermore, since the two matrices in the right-hand side of the equation are positive definite they are diagonalizable in the same basis, i.e., there exists a nonsingular matrix $\boldsymbol{F}$ such that (Anderson, 2003, Theorem A.2.2)

$$\boldsymbol{F}^{\mathrm{t}}\mathbf{\Sigma}^{-1}\boldsymbol{F} = \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_d^2 \end{pmatrix}$$

and

$$\boldsymbol{F}^{\mathrm{t}}\mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{NN})^{-1}\boldsymbol{F} = \boldsymbol{I}.$$

Since $\mathbf{\Sigma}^{-1}$ is positive definite, we furthermore have $\lambda_i^2 > 0$. $\mathbf{\Sigma}^{-1} - t\mathbf{\Phi}$ is therefore diagonalizable as well, with eigenvalues given by $(1+t)\lambda_i^2 - t = (\lambda_i^2 - 1)t + \lambda_i^2$, which is positive in a neighborhood of $t = 0$. $\mathbf{\Sigma}^{-1} - t\mathbf{\Phi}$ is therefore positive definite in a neighborhood of $t = 0$.

3 Alternative expression of multiinformation

For a decomposition of a multidimensional normal variable into several subvectors, multiinformation reads

$$I(\boldsymbol{X}_1; \ldots; \boldsymbol{X}_N) = \frac{1}{2}\ln\frac{\prod_{n=1}^{N}|\mathbf{\Sigma}_{nn}|}{|\mathbf{\Sigma}|}.$$

By comparison, we calculate

$$\begin{aligned} \boldsymbol{I}_d + \mathbf{\Gamma} &= \boldsymbol{I}_d + \mathbf{\Sigma}\mathbf{\Phi} \\ &= \boldsymbol{I}_d + \mathbf{\Sigma}\left[\mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{NN})^{-1} - \mathbf{\Sigma}^{-1}\right] \\ &= \mathbf{\Sigma}\,\mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{NN})^{-1}, \end{aligned}$$

leading to

$$\begin{aligned} |\boldsymbol{I}_d + \mathbf{\Gamma}| &= |\mathbf{\Sigma}\,\mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{NN})^{-1}| \\ &= \frac{|\mathbf{\Sigma}|}{\prod_{n=1}^{N}|\mathbf{\Sigma}_{nn}|}, \end{aligned}$$

and, finally,

$$-\frac{1}{2}\ln|\boldsymbol{I}_d + \mathbf{\Gamma}| = \frac{1}{2}\ln\frac{\prod_{n=1}^{N}|\mathbf{\Sigma}_{nn}|}{|\mathbf{\Sigma}|}.$$

4 Checking asymptotic normality

Let the correlation matrix $\boldsymbol{R}_d$ be a d-by-d homogeneous matrix with parameter ρ, i.e., a matrix with 1s on the diagonal and all off-diagonal elements equal to ρ. such a matrix has two eigenvalues: $1+(d-1)\rho$ with multiplicity 1 (associated with the vector composed only of 1s) and $1-\rho$ with multiplicity $d-1$ (associated with the subspace of vectors with a zero mean). Such a matrix is positive definite for

$$-\frac{1}{d-1} \leq \rho < 1.$$

The expectation of i_d is given by

$$-\frac{1}{2}\left\{(d-1)\ln(1-\rho)+\ln[1+(d-1)\rho]\right\}$$

To compute the higher cumulants of i_d, let $\boldsymbol{U}_d$ the d-by-d matrix with all elements equal to 1. Using the fact that $\boldsymbol{\Gamma} = \rho(\boldsymbol{U}_d - \boldsymbol{I}_d)$ together with $\boldsymbol{U}_d^l = d^{l-1}\boldsymbol{U}_d$ for $l \geq 2$, we obtain

$$\begin{aligned}
\boldsymbol{\Gamma}^l &= \rho^l\left[(-1)^l\boldsymbol{I}_d + \frac{(d-1)^l-(-1)^l}{d}\boldsymbol{U}_d\right] \\
\mathrm{tr}(\boldsymbol{\Gamma}^l) &= \rho^l\left[(-1)^l d + (d-1)^l - (-1)^l\right] \\
\kappa_l(i_d) &= \frac{(l-1)!}{2}\rho^l\left[(-1)^l d + (d-1)^l - (-1)^l\right].
\end{aligned}$$

In particular, we have $\mathrm{Var}(i_d) = \rho^2 d(d-1)/2$. For large d, we have $\kappa_l(i_d) \sim \frac{(l-1)!}{2}\rho^l d^l$ for $l \geq 2$ and, in particular, $\mathrm{Var}(i_d) \sim \rho^2 d^2/2$. To investigate the asymptotic normality of i_d, we classically consider $u = [i_d - \mathrm{E}(i_d)]/\sqrt{\mathrm{Var}(i_d)}$. Using the fact that the cumulant of order l is homogeneous of degree l, we obtain $\kappa_l(u) \sim 2^{l/2-1}(l-1)! = \mathrm{cste}$. If u were asymptotically normal, $\kappa_l(u)$ for $l \geq 3$ would tend to 0 as $d \to \infty$, which is not the case. As a consequence, u is not asymptotically normal.

References

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis. Wiley Series in Probability and Mathematical Statistics. 3rd ed., John Wiley and Sons, New York.