Comment

The limits of pairwise correlation to model the joint entropy

Benoît Legat [1,†] and Luc Rocher [2,3†*]

1 Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM). Université catholique de Louvain, B-1348, Louvain-la-Neuve, Belgium.
2 Department of Computing, Imperial College London, London, SW7 2AZ, UK.
3 Data Science Institute, Imperial College London, London, SW7 2AZ, UK.
* Correspondence: lrocher@imperial.ac.uk
† These authors contributed equally to this work.

Citation: Legat, B.; Rocher, L. The limits of pairwise correlation to model the joint entropy. *Sensors* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Information theory is a unifying mathematical theory to measure information content, key for research in cryptography, statistical physics, and quantum computing [1–3]. A central property of information theory is the entropy, a metric quantifying the amount of information encoded in a signal [4]. In "Entropy Correlation and Its Impacts on Data Aggregation in a Wireless Sensor Network", Nga et al. propose a general entropy correlation model to study the dependence patterns between multiple spatio-temporal signals [5]. They derive lower and upper bounds on the overall information entropy from only marginal and pairwise entropies, and use these bounds to study the impact of correlation on data aggregation, compression, and clustering of signals. Replicating these findings, we however show that these bounds were incorrect, over- and under-estimating the actual association patterns depending on the data. Deriving constraints and bounds on joint entropies is still a computationally difficult task and an active field of research [1,6], and new inequalities are regularly found [7–11]. More work is likely to be needed in order to develop a simple and general entropy correlation model for spatio-temporal signals.

Nga et al. study a system of m random variables $X_1, X_2, \ldots, X_m$. They propose a normalized measure of correlation between two variables Y and Z, defined as:

$$\rho(Y,Z) = 2 - 2\frac{H(Y,Z)}{H(Y)+H(Z)} \tag{1}$$

with H the Shannon entropy [4]. The authors further denote by $\rho_{\min} = \min_{i\neq j} \rho(X_i, X_j)$ and $\rho_{\max} = \max_{i\neq j} \rho(X_i, X_j)$ the minimum and maximum correlation between pairs of variables; $H_{\min} = \min_i H(X_i)$ and $H_{\max} = \max_i H(X_i)$ the minimum and maximum individual entropies.

The general entropy correlation model proposed by the authors rely on two claims, both incorrect:

Claim 1. *In equation (13) and section 2.2.2, Nga et al. claim that higher-order correlations are bounded by pairwise correlations:*

$$\forall (i,j,k),\ \rho_{min} \le \rho(X_{ij}, X_k) \le \rho_{max}$$

Claim 2. *In equations (16) and (20), Nga et al. use Claim 1 to prove that, for any subset of m variable, its joint entropy H_m is bounded by:*

$$l_m H_{min} \le H_m \le k_m H_{max}$$

with $l_m = \frac{2-\rho_{max}}{2}(l_{m-1}+1)$, $k_m = \frac{2-\rho_{min}}{2}(k_{m-1}+1)$, *and* $l_1 = k_1 = 1$.

We propose two examples for $n = 3$ demonstrating that all four inequalities are incorrect. In our first example, we obtain $\rho_{\min} > \rho(X_{ij}, X_k)$ which contradicts the lower bound of Claim 1 and $H_3 > k_3 H_{\max}$ which contradicts the upper bound of Claim 2.

Proposition 1. *Consider the four i.i.d. discrete random variables Y_1, Y_2, Y_3, Z uniformly distributed over $\{0,1\}$. For the random variables $(X_i)_{i=1}^3 = (Y_i, Z)$, we have $\rho_{min} = 1/2$, $\rho(X_{ij}, X_k) = 2/5$ for any permutation (i,j,k) of $(1,2,3)$, $k_3 = 15/8$, $H_3 = 4$ and $H_{max} = 2$.*

Proof. As Y_1, Y_2, Y_3, Z are independent, we have $H(X_i) = H(Y_i) + H(Z) = 2$ for $i = 1,2,3$ and $H(X_{ij}) = H(Y_i) + H(Y_j) + H(Z) = 3$ for $i \neq j$. Using eq. (1), we have $\rho(X_i, X_j) = 1/2$ for $i \neq j$ hence $\rho_{\min} = 1/2$, $k_2 = 2 - \rho_{\min} = 3/2$ and $k_3 = (k_2 + 1)k_2/2 = 15/8$. For any permutation (i,j,k) of $(1,2,3)$, we have $H_3 = H(X_{ijk}) = H(Y_i) + H(Y_j) + H(Y_k) + H(Z) = 4$ hence $\rho(X_{ij}, X_k) = 2/5$. □

In our second example, we obtain $\rho_{\max} < \rho(X_{ij}, X_k)$ which contradicts the upper bound of Claim 1 and $H_3 < l_3 H_{\min}$ which contradicts the lower bound of Claim 2.

Proposition 2. *Consider three discrete random variables X_1, X_2, X_3 uniformly distributed over $\{0,1\}$ that are pairwise independent and satisfying the equation $X_1 \oplus X_2 \oplus X_3 = 0$ where $\oplus$ denotes the xor operation. We have $\rho_{max} = 0$, $\rho(X_{ij}, X_k) = 2/3$ for any permutation (i,j,k) of $(1,2,3)$, $l_3 = 3$, $H_3 = 2$ and $H_{min} = 1$.*

Proof. We have $H(X_i) = 1$ for $i = 1,2,3$ and as the variables are pairwise independent, $H(X_{ij}) = H(X_i) + H(X_j) = 2$ for $i \neq j$. Using eq. (1), we have $\rho(X_i, X_j) = 0$ for $i \neq j$ hence $\rho_{\max} = 0$, $l_2 = 2 - \rho_{\max} = 2$ and $l_3 = (l_2 + 1)l_2/2 = 3$. For any permutation (i,j,k) of $(1,2,3)$, we have $H_3 = H(X_{ijk}) = 2$ hence $\rho(X_{ij}, X_k) = 2/3$. □

Overall, the two new inequalities derived by Nga et al. for the joint entropy H_m do not appear to be correct starting at $m = 3$. The errors in the model stem from the assumption made in Claim 1 that pairwise and higher-order associations share the same minimum and maximum. The authors validate their method on a very specific dataset with $\rho_{\min} = 0.6$, $H_{\min} = 2.16$, and $H_{\max} = 2.55$, yet our examples show that different association structures yield widely different joint entropies. Bounding the joint entropy allows the authors to study the impact of correlation on data aggregation, compression, and clustering of signals. Although different bounds could potentially offer similar results, the broader conclusions of this article may not hold in practice.

Finally, deriving constraints and bounds on joint entropies is a computationally difficult task and an active field of research [1,6–11]. Theoretical derivations and numerical estimations have both be used to bound the joint entropy H_m, based upon research on entropic vectors. The entropic vector of the random variables $X_1, X_2, \ldots, X_m$ is the vector of the entropies of all 2^{m-1} subsets of these variables. The set of all entropic vectors is a convex cone, for which a polyhedral outer-approximation is known [12, Theorem 1]. For instance, we derive below tight[1] lower and upper bounds for H_3 in proposition 3, suggesting an alternative approach that could lead to upper bounds for $n > 3$ and lower bounds as well. This bound rely on the following inequalities [6, Theorem 2.34]:

$$H(X_I) \leq H(X_J) \tag{2}$$

which is valid for any subsets $I \subseteq J \subseteq \{1, \ldots, m\}$ and

$$H(X_I) + H(X_J) \geq H(X_{I \cap J}) + H(X_{I \cup J}) \tag{3}$$

which is valid for any subsets $I, J \subseteq \{1, \ldots, m\}$.

[1] The tightness is a consequence of the fact that eq. (2) and eq. (3) completely describe the entropic cone [12, Theorem 2].

Proposition 3. *For any three random variables X_1, X_2, X_3, the following inequalities hold:*

$$\max(H(X_{12}), H(X_{23}), H(X_{31})) \leq H_3 \leq \min(H(X_{31}) + H(X_{12}) - H(X_1), \\ H(X_{12}) + H(X_{23}) - H(X_2), H(X_{23}) + H(X_{31}) - H(X_3)).$$

Proof. For any permutation (i, j, k) of $(1, 2, 3)$, by eq. (2) with $I = \{i, j\}$ and $J = \{i, j, k\}$, we have $H(X_{ij}) \leq H(X_{ijk}) = H_3$ and by eq. (3) with $I = \{i, j\}$ and $J = \{j, k\}$, we have $H(X_{ij}) + H(X_{jk}) \geq H(X_{ijk}) + H(X_j)$ which implies that $H_3 = H(X_{ijk}) \leq H(X_{ij}) + H(X_{jk}) - H(X_j)$. □

Similar bounds can be obtained for $m > 3$ using eq. (2) and eq. (3) but their tightness is not guarantee as the entropic cone is not completely described by these inequalities for $m > 3$ [13, Theorem 6]. This gap could be reduced numerically by iteratively producing linear cuts, in order to refine the polyhedral outer-approximation of the entropic cone given by eq. (2) and eq. (3) [14]. Taken together, our findings suggest that theoretical derivations ($m \geq 3$) and numerical approximations ($m > 3$) on the entropic cone might provide future research directions towards a robust general entropy correlation model.

References

1. Yeung, R.W. The Science of Information. In *Information Theory and Network Coding*; Yeung, R.W., Ed.; Springer US: Boston, MA, 2008; pp. 1–4.
2. Lesne, A. Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Math. Struct. Comput. Sci.* **2014**, *24*.
3. Vedral, V. The role of relative entropy in quantum information theory. *Rev. Mod. Phys.* **2002**, *74*, 197–234.
4. Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* **1948**, *27*, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
5. Nguyen Thi Thanh, N.; Nguyen Kim, K.; Ngo Hong, S.; Ngo Lam, T. Entropy correlation and its impacts on data aggregation in a wireless sensor network. *Sensors* **2018**, *18*, 3118.
6. Yeung, R.W. *A First Course in Information Theory*; Springer Science & Business Media, 2012.
7. Matus, F. Infinitely Many Information Inequalities. 2007 IEEE International Symposium on Information Theory, 2007, pp. 41–44.
8. Zhen Zhang.; Jun Yang. On a new non-Shannon-type information inequality. Proceedings IEEE International Symposium on Information Theory,, 2002, pp. 235–.
9. Makarychev, K.; Makarychev, Y.; Romashchenko, A.; Vereshchagin, N. A new class of non-Shannon-type inequalities for entropies. *Commun. Inf. Syst.* **2002**, *2*, 147–166.
10. Matúš, F. Conditional Independences among Four Random Variables III: Final Conclusion. *Comb. Probab. Comput.* **1999**, *8*, 269–276.
11. Dougherty, R.; Freiling, C.; Zeger, K. Six New Non-Shannon Information Inequalities. 2006 IEEE International Symposium on Information Theory, 2006, pp. 233–236.
12. Zhang, Z.; Yeung, R.W. A non-Shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory* **1997**, *43*, 1982–1986.
13. Zhang, Z.; Yeung, R.W. On characterization of entropy function via information inequalities. *Information Theory, IEEE Transactions on* **1998**, *44*, 1440–1452.
14. Legat, B.; Jungers, R.M. Parallel optimization on the Entropic Cone. Proceedings of the 37rd Symposium on Information Theory in the Benelux, 2016, SITB '16, pp. 206–211.