Entropic Hyper-Connectomes Computation and Analysis

Michael G. Rawson*

Abstract. Brain function and connectivity is a pressing mystery in medicine related to many diseases. Neural connectomes have been studied as graphs with graph theory methods including topological methods. Work has started on hypergraph models and methods where the geometry and topology is significantly different. We define a hypergraph called the hyper-connectome with joint information entropy and total correlation. We give the pseudocode for computation from finite samples. We give the theoretic importance of this generalization's topology and geometry with respect to random variables and then prove the hypergraph can be necessary for prediction and classification. We confirm with a simulation study and computation. We prove the approximation for continuous random variables with finite samples. We compare connectome versus hyper-connectome for predicting schizophrenia in subjects based on a fMRI dataset using a linear support vector machine. The hyper-connectome achieves better performance in accuracy (up to 57%) and F1 score (up to 0.52) than the connectome.

Key words. Connectome, Hypergraph, Entropy, Machine Learning, Schizophrenia, Linear Support Vector Machine, Prediction, Classification

1. Introduction.

Functional magnetic resonance imaging (fMRI) since the 1990's has been used to investigate brain activity in a benign and noninvasive way [8]. fMRI detects oxygen usage by neurons with changing magnetic fields. fMRI is not sensitive enough to measure something as small as a neuron so a region of around 4mm by 4mm by 4mm is scanned which we call a *voxel*. A connectome is a graph describing the connections between neurons. By grouping neurons into voxels (3D cubes in space), a graph can be constructed from fMRI scan data. Each vertex of the graph is a voxel and each weighted edge is the correlation between the voxels. Correlation between voxels imply the neurons are directly or indirectly connected. This also assumes that neural activity is probabilistic. When the number of samples or measurements is low, the statistical confidence is low. To improve the confidence, voxels are often grouped into brain regions. Then each brain region has many samples and the confidence of the correlation is much greater. Additionally, the graph where the vertices correspond to just 61 brain regions is much smaller allowing for more computationally intensive graph analysis methods and this is what we shall use in subsection 5.2.

Brain fMRI scan data can be viewed as a hypergraph by using high order statistical methods. Hypergraph analysis can then yield a deeper, more informative analysis on brain connectomes. We build a brain connectome hypergraph from fMRI scans to identify indicators that are impossible to retrieve from a standard brain connectome graph. Note that the hyper-connectome is *not* a simplicial complex and thus most topological data analysis (TDA) methods, such as persistent homology, will not work here [14].

2. Literature Review.

There has been a large amount of activity around this area in the recent past. In the 2000s, Darling and Norris studied large random hypergraphs [4] which is exactly what a hyper-connectome is. Then in 2016, Munsell, Zu, Giusti, et al. considered different types of hypergraphs related to connectomes and neural data and various diseases

---

*Department of Mathematics, University of Maryland at College Park, Maryland, USA (rawson@umd.edu).

[6, 11, 19]. Sparse linear regression has been used to predict hyperedges in hypergraphs and classify disease [7, 9, 10]. In 2018, hypergraph methods were compared and contrasted to TDA, sheaf methods, point cloud methods, and others by Purvine et al. [12]. Later, Sizemore et al. created a structural hypergraph of a mouse connectome [13] and analyzed the topology. On the other hand, others learned, or optimized, a hypergraph as opposed to direct calculation [16, 18]. Banka et al. learned autoencoder embeddings by hypergraphs [3]. More broadly, Aksoy et al. performed a general comparison between graphs and hypergraphs and their methods [2]. More recently, Stolz et al. analyzed the topology of connectomes in schizophrenic subjects versus normal subjects (siblings and non-siblings) [14].

3. Theory. Given two random variables, the Pearson correlation coefficient is $corr(X_1, X_2) := \frac{\mathbb{E}[(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)]}{\sqrt{\mathbb{E}((X_1 - \mathbb{E}X_1)^2)}\sqrt{\mathbb{E}((X_2 - \mathbb{E}X_2)^2)}}$. This can be approximated given samples from distributions. The sample Pearson correlation coefficient, with n samples, is $corr(\hat{X}_1, \hat{X}_2) :=$ $\frac{\sum_i \frac{1}{n}(\hat{X}_{1,i} - \frac{1}{n}\sum_k \hat{X}_{1,k})(\hat{X}_{2,i} - \frac{1}{n}\sum_k \hat{X}_{2,k})}{\sqrt{[\sum_i \frac{1}{n}(\hat{X}_{1,i} - \frac{1}{n}\sum_k \hat{X}_{1,k})][\sum_j \frac{1}{n}(\hat{X}_{2,j} - \frac{1}{n}\sum_k \hat{X}_{2,k})]}}$. We will use this on fMRI samples to identify structure between brain regions. However, there are other statistics that we can use. The information entropy of a discrete random variable is $H(X) = -\sum_{x \in \mathbb{R}} p(x) \log p(x)$ where p is the probability that $X = x$. Note that we use the convention that $0 \cdot \log(0) = 0$. So again this can be approximated with samples. The sample information entropy, with n samples, is $H(\hat{X}) = -\sum_{x \in \mathbb{R}} \frac{1}{n} \sum_i \mathbb{1}_{\hat{X}_i = x}(\hat{X}_i) \log \left[\frac{1}{n} \sum_i \mathbb{1}_{\hat{X}_i = x}(\hat{X}_i)\right]$. Now for any collection of discrete random variables, we can define the joint information entropy as $H(X_1, ..., X_n) = -\sum_{x_1,...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \log p(x_1, ..., x_n)$. The joint information entropy can be very useful for approximating how independent brain regions are versus how closely they collaborate in a predictable way. This is captured more directly by comparing the sum of entropies with the joint entropy. This difference is the mutual information or total correlation

$$C(X_1, X_2, ..., X_n) := [\sum_i H(X_i)] - H(X_1, ..., X_n).$$

We use the following well known simplification.

Proposition 3.1.

$$C(X_1, X_2, ..., X_n) = \sum_{x_1, x_2, ..., x_n \in \mathbb{R}} p(x_1, x_2, ..., x_n) \log \frac{p(x_1, x_2, ..., x_n)}{p(x_1)p(x_2)...p(x_n)}.$$

Proof in supplemental. Next we show how discrete random variables can approximate absolutely continuous random variables.

Theorem 3.2. *Let Y_i be absolutely continuous random variables, $Y_i : \Omega \to \mathbb{R}$ measurable. The joint density p_Y can be approximated by simple functions arbitrarily close in integration. We use an approximation and we say $p_{\tilde{Y}_i}$ is a simple function. Set discrete random variables X_i to have a dirac for each term in simple function $p_{\tilde{Y}}$ with the coefficient so that the measures are the same:*

$$p_{\tilde{Y}} = \sum_k \alpha_k \mathbb{1}_{\Pi_i(a_{i,k}, b_{i,k})}, \quad p_X = \sum_k \beta_k \delta_{\Pi_i(b_{i,k} - a_{i,k})/2}, \quad \beta_k = \alpha_k \mathbb{P}_{\tilde{Y}}(\Pi_i(a_{i,k}, b_{i,k}))$$

For any $\epsilon > 0$, there is an discrete approximation with the total correlation

$$|C(Y_1, Y_2, ..., Y_n) - C(X_1, X_2, ..., X_n)| < \epsilon.$$

Proof in supplemental. With the justification of Theorem 3.2, we will approximate the total correlation from finite samples with Algorithm .1.

4. Simulation Study Results. First we'll consider a small example where we know the distributions. We'll setup the distributions where doing classification by hypergraph is in theory possible, which we prove. For random variables X and Y, let $X_i \sim Bernoulli(\{-1, 1\}, 1/2)$, $1 \leq i \leq 3$ and $Y = [Y_1, Y_2, Y_3] = [X_1 X_2, X_2 X_3, X_3 X_1]$. Then $\sigma_{Y_i}^2 = \mathbb{E} Y_i^2 = 1$. For $i \neq j$, let k, u, and v be such that $Y_i = X_k X_u$ and $Y_j = X_k X_v$. Then

$$(4.1) \qquad corr(Y_i, Y_j) = \frac{\mathbb{E}[Y_i Y_j]}{\sigma_{Y_i} \sigma_{Y_j}} = 0.$$

Proof in supplemental. This implies that the connectome Pearson correlation graph will not distinguish subject X from subject Y. However, the *total correlation* of Y is not 0.

Proposition 4.1.

$$C(Y_1, Y_2, Y_3) = [\sum_i H(Y_i)] - H(Y_1, Y_2, Y_3) < 0.$$

Proof in supplemental. In this case, a discriminator can distinguish subject Y from subject X with total correlation but would fail to distinguish when using Pearson correlations. If this was a neural connectome, only the hyper-connectome could classify the subjects.

We calculated the hypergraph with Algorithm .1. Then we classify the data with a linear support vector machine. We give the results in Table 1. The graph does not contain the information needed to classify the subject as a member of X versus Y. The hypergraph does contain the information and the classifier successfully distinguishes X from Y.

Table 1: Linear support vector machine classification of X versus Y from section 4. Hypergraph threshold $\epsilon = 10^{-5}$, dimension $d = 3$, and variable count 3. Subjects are 1000 subject from X and 1000 subjects from Y. There are 20 samples of each subject. The training/testing split is random 50%.

	Training Accuracy	Testing Accuracy	F1 Score
Graph	51%	49%	0.66
Hypergraph	100%	100%	1

5. Schizophrenia Dataset Results. We next compute the hyper-connectome on real data. We utilize a schizophrenia (schiz.) fMRI dataset, see [15], consisting of 104 patients with schizophrenia and 124 healthy, normal controls. Between the groups, the age and gender differences are minimal (schiz.: age 36.88 ± 14.17 with 62 males, 41 females, 1 other, and Normal: age 33.75 ± 14.22 with 61 males and 63 females). The fMRI aquisition details and preprocessing details are laid out in Adhikari et al. [1]. This dataset consists of activity in 246 regions of interest (ROI) which we call the *ROI variables* [5]. We create the connectome and hyper-connectome with a subset of these ROI variables.

5.1. Hyper-Connectome Visualization. We visualize connectomes by plotting the graph for a normal subject and schiz. subject in Figure 1a and Figure 1b. In this graph, the nodes are the brain regions and the edges are the absolute value of Pearson correlation between regions, where length and width indicate the weight of the edge. In Figure 1c and Figure 1d, we visualize the hyper-connectomes of the same two subjects. In this graph, the square nodes are the brain regions and the circles are significant hyper-edges between multiple nodes [17]. We see that the connectomes are highly clustered with few outliers versus the hyper-connectomes which are have nodes covering the connectedness spectrum.

In Figure 2, we show the corresponding adjacency matrices to the graphs in Figure 1. For each pair of ROI, we sum the weights of all common hyperedges to produce the adjacency matrices in Figure 2c and Figure 2d. We see that the magnitudes of Figure 2a and Figure 2b are similar (in 0 to 1) while the maximums of Figure 2c and Figure 2d differ significantly (5000 vs. 5500). We calculated the hypergraph with Algorithm .1. The hypergraph threshold $\epsilon = 10^{-5}$, dimension $d = 3$, and ROI variables are 30 brain regions. Samples used are the first 20 in the time series.

(a) Normal Subject Connectome   (b) Schiz. Subject Connectome   (c) Normal Subject Hyper-Connectome   (d) Schiz. Subject Hyper-Connectome
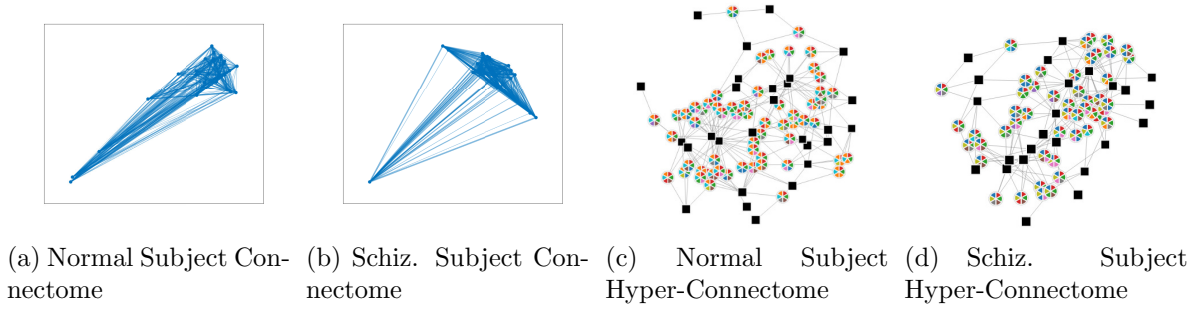
Figure 1: (a) and (b): Connectome graph of normal vs. schiz. subject with 30 ROI nodes and correlation weighted edges. (c) and (d): Hyper-Connectome of normal vs. schiz. subject, plotting significant (weight $> 2^8$) edges (circles) for 30 ROI nodes (squares).

5.2. Hyper-Connectome Classification. We showed in the previous section that the connectome and hyper-connectome are differ greatly in information content. The next question is how useful this is for distinguishing normal subjects and schizophrenic subjects. We vectorize the upper triangle of the connectome adjacency matrix and train a linear support vector machine to classify the subjects. After training, we calculate the accuracy and F1 score on the (unseen) test data. We follow the same procedure with the hyper-connectome. We report the results in Table 2. We find an increase in the testing accuracy and in the F1 score from using the hyper-connectome versus the connectome.

6. Conclusions. We have introduced the entropic hyper-connectome as a useful concept to study neuronal structure, function, and abnormalities. We have demonstrated this with fMRI data in vivo. From theory we have defined the hypergraph and proved that it can be necessary to detect various mixture distributions. We visualized the connectome and hyper-
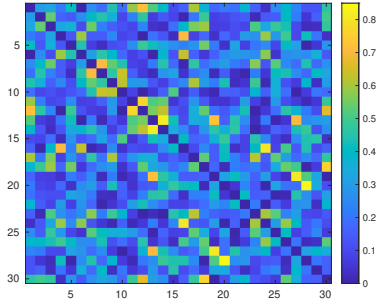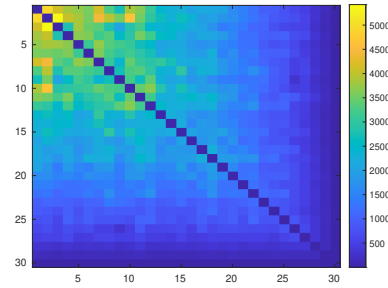
(a) Normal Subject Pearson Correlation Matrix
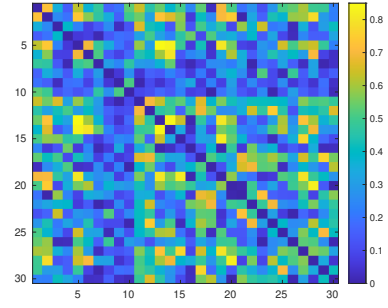
(b) Schiz. Subject Pearson Correlation Matrix
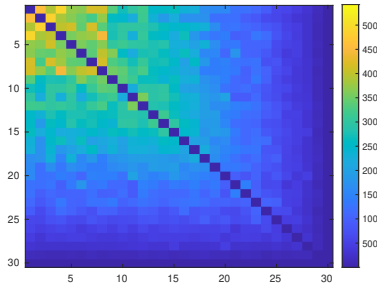
(c) Normal Subject Total Correlation Tensor

(d) Schiz. Subject Total Correlation Tensor

Figure 2: Top: Pearson correlation matrix of normal vs. schiz. subject, 30 ROIs. Bottom: Total correlation tensor of normal vs. schiz. subject, summed to matrix, 30 ROIs.

Table 2: Linear support vector machine prediction of schizophrenia. Hypergraph threshold $\epsilon = 10^{-5}$, dimension $d = 3$, and 61 ROI variables considered. Samples over time are 30. The training/testing split is random 50%. Calculations are average of 10 independent trials.

	Training Accuracy	Testing Accuracy	F1 Score
Graph	100%	51%	0.44
Hypergraph	100%	57%	0.52

connectome and see significant differences. Finally, we trained a classifier to show that the hyper-edges can improve classification.

REFERENCES

[1] B. M. Adhikari, L. E. Hong, H. Sampath, J. Chiappelli, N. Jahanshad, P. M. Thompson, L. M. Rowland, V. D. Calhoun, X. Du, S. Chen, and P. Kochunov, *Functional network connectivity impairments and core cognitive deficits in schizophrenia*, Human Brain Mapping, 40 (2019), pp. 4593–4605, https://doi.org/10.1002/hbm.24723.

[2] Aksoy, Sinan G., Joslyn, Cliff, Ortiz Marrero, Carlos, Praggastis, Brenda, and Purvine, Emilie, *Hypernetwork science via high-order hypergraph walks*, EPJ Data Sci., 9 (2020), p. 16, https://doi.org/10.1140/epjds/s13688-020-00231-0.

[3] A. Banka, I. Buzi, and I. Rekik, *Multi-view brain hyperconnectome autoencoder for brain state classification*, in Predictive Intelligence in Medicine, I. Rekik, E. Adeli, S. H. Park, and M. d. C. Valdés Hernández, eds., Cham, 2020, Springer International Publishing, pp. 101–110.

[4] R. W. R. Darling and J. R. Norris, *Structure of large random hypergraphs*, The Annals of Applied Probability, 15 (2005), pp. 125 – 152, https://doi.org/10.1214/105051604000000567.

[5] L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. Yu, and T. Jiang, *The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture*, Cerebral Cortex, 26 (2016), pp. 3508–3526, https://doi.org/10.1093/cercor/bhw157.

[6] C. Giusti, R. Ghrist, and D. S. Bassett, *Two's company, three (or more) is a simplex*, Journal of Computational Neuroscience, 41 (2016), pp. 1–14, https://doi.org/10.1007/s10827-016-0608-6.

[7] H. Guo, Y. Li, Y. Xu, Y. Jin, J. Xiang, and J. Chen, *Resting-state brain functional hyper-network construction based on elastic net and group lasso methods*, Frontiers in Neuroinformatics, 12 (2018), https://doi.org/10.3389/fninf.2018.00025.

[8] S. A. Huettel, A. W. Song, G. McCarthy, et al., *Functional magnetic resonance imaging*, vol. 1, Sinauer Associates Sunderland, MA, 2004.

[9] B. Jie, C.-Y. Wee, D. Shen, and D. Zhang, *Hyper-connectivity of functional networks for brain disease diagnosis*, Medical Image Analysis, 32 (2016), pp. 84–100, https://doi.org/10.1016/j.media.2016.03.003.

[10] Y. Li, J. Liu, X. Gao, B. Jie, M. Kim, P.-T. Yap, C.-Y. Wee, and D. Shen, *Multimodal hyper-connectivity of functional networks using functionally-weighted lasso for mci classification*, Medical Image Analysis, 52 (2019), pp. 80–96, https://doi.org/10.1016/j.media.2018.11.006.

[11] B. C. Munsell, G. Wu, Y. Gao, N. Desisto, and M. Styner, *Identifying relationships in functional and structural connectome data using a hypergraph learning method*, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Cham, 2016, Springer International Pub., pp. 9–17.

[12] E. Purvine, S. Aksoy, C. Joslyn, K. Nowak, B. Praggastis, and M. Robinson, *A topological approach to representational data models*, in Human Interface and the Management of Information. Interaction, Visualization, and Analytics, S. Yamamoto and H. Mori, eds., Cham, 2018, Springer International Publishing, pp. 90–109.

[13] A. E. Sizemore, J. E. Phillips-Cremins, R. Ghrist, and D. S. Bassett, *The importance of the whole: Topological data analysis for the network neuroscientist*, Network Neuroscience, 3 (2019), pp. 656–673, https://doi.org/10.1162/netn_a_00073.

[14] B. J. Stolz, T. Emerson, S. Nahkuri, M. A. Porter, and H. A. Harrington, *Topological data analysis of task-based fMRI data from experiments on schizophrenia*, Journal of Physics: Complexity, 2 (2021), p. 035006, https://doi.org/10.1088/2632-072x/abb4c6.

[15] Q. Wu, X. Huang, A. J. Culbreth, J. A. Waltz, L. E. Hong, and S. Chen, *Extracting brain disease-related connectome subgraphs by adaptive dense subgraph discovery*, Biometrics, (2021), pp. 1–13, https://doi.org/10.1111/biom.13537.

[16] L. Xiao, J. Wang, P. H. Kassani, Y. Zhang, Y. Bai, J. M. Stephen, T. W. Wilson, V. D. Calhoun, and Y.-P. Wang, *Multi-hypergraph learning-based brain functional connectivity analysis in fmri data*, IEEE Transactions on Medical Imaging, 39 (2020), pp. 1746–1758.

[17] Y. Zhou, A. Rathore, E. Purvine, and B. Wang, *Topological simplifications of hypergraphs*, 2021, https://arxiv.org/abs/2104.11214.

[18] C. Zu, Y. Gao, B. Munsell, M. Kim, Z. Peng, J. R. Cohen, D. Zhang, and G. Wu, *Identifying disease-related subnetwork connectome biomarkers by sparse hypergraph learning*, Brain Imaging and

Behavior, 13 (2019), pp. 879–892, https://doi.org/10.1007/s11682-018-9899-8.

[19] C. Zu, Y. Gao, B. Munsell, M. Kim, Z. Peng, Y. Zhu, W. Gao, D. Zhang, D. Shen, and G. Wu, *Identifying high order brain connectome biomarkers via learning on hypergraph*, in Machine Learning in Medical Imaging, Cham, 2016, Springer International Publishing, pp. 1–9.

Supplementary Materials.

Proof of Proposition 3.1.

$$C(X_1, X_2, ..., X_n) = [\sum_i H(X_i)] - H(X_1, ..., X_n)$$

$$= [\sum_{x_1,...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \log p(x_1, ..., x_n)] - [\sum_i \sum_{x_i} p(x_i) \log p(x_i)]$$

$$= [\sum_{x_1,...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \log p(x_1, ..., x_n)] - [\sum_i \sum_{x_i} \sum_{x_1,...,x_{i-1},x_{i+1},...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \log p(x_i)]$$

$$= [\sum_{x_1,...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \log p(x_1, ..., x_n)] - [\sum_{x_1,...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \sum_i \log p(x_i)]$$

$$= \sum_{x_1,...,x_n \in \mathbb{R}} [p(x_1, ..., x_n) \log p(x_1, ..., x_n) - p(x_1, ..., x_n) \sum_i \log p(x_i)]$$

$$= \sum_{x_1,...,x_n \in \mathbb{R}} p(x_1, ..., x_n) \log \frac{p(x_1, ..., x_n)}{p(x_1)...p(x_n)} \qquad \blacksquare$$

Proof of Proposition 4.1.

$$C(Y_1, Y_2, Y_3) = [\sum_i H(Y_i)] - H(Y_1, Y_2, Y_3)$$

$$= [\sum_i \mathbb{P}(Y_i = 1) \log \mathbb{P}(Y_i = 1) + \mathbb{P}(Y_i = -1) \log \mathbb{P}(Y_i = -1)]$$

$$+ \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1) \log \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1)$$

$$+ \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = -1) \log \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = -1)$$

$$+ \mathbb{P}(Y_1 = 1, Y_2 = -1, Y_3 = 1) \log \mathbb{P}(Y_1 = 1, Y_2 = -1, Y_3 = 1)$$

$$+ \mathbb{P}(Y_1 = 1, Y_2 = -1, Y_3 = -1) \log \mathbb{P}(Y_1 = 1, Y_2 = -1, Y_3 = -1)$$

$$+ \mathbb{P}(Y_1 = -1, Y_2 = 1, Y_3 = 1) \log \mathbb{P}(Y_1 = -1, Y_2 = 1, Y_3 = 1)$$

$$+ \mathbb{P}(Y_1 = -1, Y_2 = 1, Y_3 = -1) \log \mathbb{P}(Y_1 = -1, Y_2 = 1, Y_3 = -1)$$

$$+ \mathbb{P}(Y_1 = -1, Y_2 = -1, Y_3 = 1) \log \mathbb{P}(Y_1 = -1, Y_2 = -1, Y_3 = 1)$$

$$+ \mathbb{P}(Y_1 = -1, Y_2 = -1, Y_3 = -1) \log \mathbb{P}(Y_1 = -1, Y_2 = -1, Y_3 = -1)$$

$$= [\sum_i 1/2 \log 1/2 + 1/2 \log 1/2]$$

$$+ \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1) \log \mathbb{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1)$$
$$+ \mathbb{P}(Y_1 = 1, Y_2 = -1, Y_3 = -1) \log \mathbb{P}(Y_1 = 1, Y_2 = -1, Y_3 = -1)$$
$$+ \mathbb{P}(Y_1 = -1, Y_2 = 1, Y_3 = -1) \log \mathbb{P}(Y_1 = -1, Y_2 = 1, Y_3 = -1)$$
$$+ \mathbb{P}(Y_1 = -1, Y_2 = -1, Y_3 = 1) \log \mathbb{P}(Y_1 = -1, Y_2 = -1, Y_3 = 1)$$
$$= [3 \log 1/2] + 1/4 \log 1/4 + 1/4 \log 1/4 + 1/4 \log 1/4 + 1/4 \log 1/4$$
$$= 3 \log 1/2 + \log 1/4$$
$$< 0$$

∎

Proof of Theorem 3.2. Recall

$$p_{\tilde{Y}} = \sum_k \alpha_k \mathbb{1}_{\Pi_i(a_{i,k}, b_{i,k})}, \quad p_X = \sum_k \beta_k \delta_{\Pi_i(b_{i,k} - a_{i,k})/2}, \quad \beta_k = \alpha_k \mathbb{P}_{\tilde{Y}}(\Pi_i(a_{i,k}, b_{i,k})).$$

The total correlation

$$C(\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_n) = \int_{y_1, y_2, ..., y_n \in \mathbb{R}} p_{\tilde{Y}}(y_1, y_2, ..., y_n) \log \frac{p_{\tilde{Y}}(y_1, y_2, ..., y_n)}{p_{\tilde{Y}}(y_1) p_{\tilde{Y}}(y_2) ... p_{\tilde{Y}}(y_n)} dy_1 ... dy_n.$$

Write compactly as the integral of the simple function $\Phi_{\tilde{Y}}(y) = \sum_{k \in \mathbb{N}} \alpha_k \mathbb{1}_{\Pi_i(a_{i,k}, b_{i,k})}(y)$

$$C(\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_n) = \int_{y \in \mathbb{R}^n} \Phi_{\tilde{Y}}(y) dy = \int_{y \in \mathbb{R}^n} \sum_{k \in \mathbb{N}} \alpha_k \mathbb{1}_{\Pi_i(a_{i,k}, b_{i,k})}(y) dy$$

$$= \sum_{k \in \mathbb{N}} \alpha_k \int_{\Pi_i(a_{i,k}, b_{i,k})} dy$$

$$= \sum_{k \in \mathbb{N}} \beta_k \int \delta_{\Pi_i(b_{i,k} - a_{i,k})/2}$$

$$= \sum_{x \in \mathbb{R}^n} \Phi_X(x)$$

$$= \sum_{x_1, x_2, ..., x_n \in \mathbb{R}} p_X(x_1, x_2, ..., x_n) \log \frac{p_X(x_1, x_2, ..., x_n)}{p_X(x_1) p_X(x_2) ... p_X(x_n)}$$

$$= C(X_1, X_2, ..., X_n).$$

∎

Proof of Equation (4.1).

$$corr(Y_i, Y_j) = \frac{\mathbb{E}[Y_i Y_j]}{\sigma_{Y_i} \sigma_{Y_j}}$$

$$= \mathbb{P}(Y_i = 1, Y_j = 1) - \mathbb{P}(Y_i = -1, Y_j = 1) - \mathbb{P}(Y_i = 1, Y_j = -1) + \mathbb{P}(Y_i = -1, Y_j = -1)$$

$$= \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) + \mathbb{P}(X_1 = -1, X_2 = -1, X_3 = -1)$$

$$\quad - 2\mathbb{P}(Y_i = -1, Y_j = 1) + \mathbb{P}(X_k = -1, X_u = 1, X_v = 1) + \mathbb{P}(X_k = 1, X_u = -1, X_v = -1)$$

$$= 1/8 + 1/8 - 2\mathbb{P}(X_k = -1, X_u = 1, X_v = -1) - 2\mathbb{P}(X_k = 1, X_u = -1, X_v = 1) + 1/8 + 1/8$$

$$= 0 \qquad\qquad\qquad \blacksquare$$

Algorithm .1 Total Correlation

Input:
    $M \in \mathbb{N}$: variables
    $N \in \mathbb{N}$: samples
    $d \in \mathbb{N}$: dimensions
    $X \in \mathbb{R}^{M \times N}$: measurements
    $\epsilon \in \mathbb{R}$: threshold
Output:
    $T \in \mathbb{R}^{M^d}$: total correlation
Begin:
$T = 0 \in \mathbb{R}^{M^d}$
for $1 \leq i_1 \leq M$ **do**
  for $i_1 \leq i_2 \leq M$ **do**

    ...
    for $i_{d-1} \leq i_d \leq M$ **do**
      for $1 \leq s_1 \leq N$ **do**
        $\omega_1 = X(i_1, s_1)$
        $p_1(\omega_1) = \frac{1}{N} \sum_j \mathbb{1}_{|X(i_1,j)-\omega_1|<\epsilon}(X)$
        for $1 \leq s_2 \leq N$ **do**
          $\omega_2 = X(i_2, s_2)$
          $p_2(\omega_2) = \frac{1}{N} \sum_j \mathbb{1}_{|X(i_2,j)-\omega_2|<\epsilon}(X)$
          ...
          for $1 \leq s_d \leq N$ **do**
            $\omega_d = X(i_d, s_d)$
            $p_d(\omega_d) = \frac{1}{N} \sum_j \mathbb{1}_{|X(i_d,j)=\omega_d|<\epsilon}(X)$
            $p(\omega_1, \omega_2, ..., \omega_d) = \frac{1}{N} \sum_j \mathbb{1}_{|X(i_1,j)-\omega_1|<\epsilon,...,|X(i_d,j)-\omega_d|<\epsilon}(X)$
            $T(i_1, i_2, ..., i_d) = T(i_1, i_2, ..., i_d) + p(\omega_1, \omega_2, ..., \omega_d) \log \frac{p(\omega_1,\omega_2,...,\omega_d)}{p_1(\omega_1)p_2(\omega_2)...p_d(\omega_d)}$
          end for

          ...
        end for
      end for
    end for

    ...
  end for
end for
return T