Supplementary Material for "Error Exponents for Composite Hypothesis Testing of Markov Forest Distributions"

Vincent Y. F. Tan, Animashree Anandkumar and Alan S. Willsky

January 9, 2010

The proofs of the paper "Error Exponents for Composite Hypothesis Testing of Markov Forest Distributions" are provided. Note that all equation numbers refer to the corresponding equation in the main paper (e.g., (1) refers to equation (1) in the main paper). The same holds for theorems, propositions and lemmas.

1 Proof of Proposition 2

Proof We start from (8) which says that the acceptance region for the GLRT is given as

$$A_n(\mathrm{GLRT}) := \left\{ \mathbf{x}^n : \frac{1}{n} \log \frac{\sup_{q \in \Lambda_1} q^n(\mathbf{x}^n)}{p^n(\mathbf{x}^n)} \in \mathcal{R}_n(\alpha) \right\}, \tag{A-1}$$

since $\Lambda_0 = \{p\}$ is a simple hypothesis and $\Lambda_1 = \mathcal{D}(\mathcal{X}^d, \mathcal{T}^d \setminus \{T_0\})$. Now we examine the numerator which can be written as

$$\sup_{q \in \Lambda_1} \frac{1}{n} \sum_{k=1}^{n} \log q(\mathbf{x}_k) = \sup_{q \in \Lambda_1} \mathbb{E}_{\widehat{\mu}^n}[\log q], \tag{A-2}$$

$$= \sup_{q \in \Lambda_1} \mathbb{E}_{\widehat{\mu}^n} \left[\log \left(\prod_{i \in V} q_i \prod_{(i,j) \in E} \frac{q_{i,j}}{q_i q_j} \right) \right], \tag{A-3}$$

$$= \max_{E \neq E_0} \sum_{i \in V} \mathbb{E}_{\widehat{\mu}_i^n} [\log \widehat{\mu}_i^n] + \sum_{(i,j) \in E} \mathbb{E}_{\widehat{\mu}_{i,j}^n} \left[\log \frac{\widehat{\mu}_{i,j}^n}{\widehat{\mu}_i^n \widehat{\mu}_j^n} \right], \tag{A-4}$$

$$= - \sum_{i \in V} H(\widehat{\mu}_i^n) + \max_{E \neq E_0} \sum_{(i,j) \in E} I(\widehat{\mu}_{i,j}^n), \tag{A-5}$$

where in (A-4), we used the fact that given any tree structure E, the ML parameters on the nodes and edges are precisely given by the type, *i.e.*, $q_i^* = \widehat{\mu}_i^n$ for all $i \in V$ and $q_{i,j}^* = \widehat{\mu}_{i,j}^n$ for all $(i,j) \in E$. Thus, we have shown that the edge set of the ML distribution is given by the second-best MWST. Next, notice that

$$\frac{1}{n} \sum_{k=1}^{n} \log p(\mathbf{x}_k) = - \sum_{i \in V} H(\widehat{\mu}_i^n) + \sum_{(i,j) \in E_0} I(\widehat{\mu}_{i,j}^n). \tag{A-6}$$

Combining (A-5) and (A-6) gives the simplified form of the GLRT in Proposition 2. □

2 Proof of Theorem 3

Proof Note from the optimization for the worst-case type-II error exponent

$$J^*(p) := \inf_{q \in \mathcal{D}(\mathcal{X}^d, \mathcal{T}^d \setminus \{T_0\})} D(q \,||\, p) \tag{A-7}$$

can be written as a nested minimization:

$$J^*(p) = \min_{T \in \mathcal{T}^d \setminus \{T_0\}} \left[\inf_{q \in \mathcal{D}(\mathcal{X}^d, T)} D(p \,||\, q) \right], \tag{A-8}$$

where $\mathcal{D}(\mathcal{X}^d, T)$ is the set of distributions Markov on the tree T. We now characterize the inner optimization in (A-8). Define

$$\Psi(T; p) := \inf_{q \in \mathcal{D}(\mathcal{X}^d, T)} D(p \,||\, q). \tag{A-9}$$

Then, since the tree structure E is fixed, by using the tree decomposition of q, Ψ can be expressed as

$$\Psi(T; p) = \inf_{q \in \mathcal{D}(\mathcal{X}^d, T)} \left\{ -H(p) - \sum_{i \in V} \mathbb{E}_{p_i}[\log q_i] - \sum_{(i,j) \in E} \mathbb{E}_{p_{i,j}} \left[\log \frac{q_{i,j}}{q_i q_j} \right] \right\}, \tag{A-10}$$

$$= -H(p) + \sum_{i \in V} H(p_i) - \sum_{(i,j) \in E} I(p_{i,j}), \tag{A-11}$$

$$= \sum_{(i,j) \in E_0} I(p_{i,j}) - \sum_{(i,j) \in E} I(p_{i,j}), \tag{A-12}$$

where (A-11) follows from the fact that the optimizing distribution $q_i^* = p_i$ for all $i \in V$ and $q_{i,j}^* = p_{i,j}$ for all $(i, j) \in E$ and (A-12) follows from the tree decomposition of p as in (1). From (A-12), we observe that for calculating $J^*(p)$ in (A-8), we only need to optimize over the second term, *i.e.*, over the structure E. Thus, the problem reduces to

$$J^*(p) = \sum_{e \in E_0} I(p_e) - \max_{T=(V,E) \in \mathcal{T}^d \setminus \{T_0\}} \sum_{e \in E} I(p_e). \tag{A-13}$$

The optimization in (A-13) is a MWST problem with the constraint that the optimizing tree structure cannot be equal to E_0. Thus, the second best MWST is chosen. It is well known that [1] that the second best MWST differs from the MWST by a single edge. Hence,

$$J^*(p) = \min_{e' \notin E_0} \min_{e \in \mathrm{Path}(e'; E_0)} I(p_e) - I(p_{e'}). \tag{A-14}$$

Finally, by using the data-processing inequality [2, Ch. 1], we observe that only non-edges $e' = (i, j)$ such that $L(i, j) = 2$ need to be considered since for nodes i and l with $L(i, l) > 2$, $I(p_{i,l}) \le I(p_{i,j})$ for any node j along the path connecting nodes i and l (see Fig. 2). Hence (11) holds.

The existence of q^* also follows because for fixed p, the relative entropy $D(p \,||\, \cdot)$ is lower semi-continuous[1] over $\mathcal{P}(\mathcal{X}^d)$ and the constraint set $\mathcal{D}(\mathcal{X}^d, \mathcal{T}^d \setminus \{T_0\})$ is compact. □

3 Proof of Corollary 4

Proof The mutual information between jointly Gaussian random variables $(X_i, X_j) \sim p_e = p_{i,j}$ is [2, Ch. 7]

$$I(p_e) = I(X_i; X_j) = \frac{1}{2} \log \left[\frac{1}{1 - \rho_e^2} \right], \tag{A-15}$$

where ρ_e is the correlation coefficient of X_i and X_j. Furthermore, for a non-edge e' spanning two edges e_1 and e_2, the correlation coefficient is given by $\rho_{e'} = \rho_{e_1} \rho_{e_2}$. This follows from the Markov property. This completes the proof in light of (11) in Theorem 3. □

4 Proof of Theorem 5

Proof This follows along exactly the same lines as Theorem 3. □

[1] This follows from the positivity of the distributions.

5 Proof of Proposition 6

Proof ($\Rightarrow$) Suppose there exists such a $\delta > 0$. Then all $I(p_e) > 0$ for all edges $e \in E_0$. Then all the differences in mutual information quantities $I(p_e) - I(p_{e'})$ (where $e' \in E_1$ could possibly replaced $e \in E_0$) would be strictly positive and hence $J^*(p) > 0$ from (11) .
($\Leftarrow$) Assume that $J^*(p) > 0$. Suppose, to the contrary, such a $\delta > 0$ does not exist. This means that there exists a $e' \in E_1$ and a $p \in \Lambda_0$ such that $I(p_e) = 0$ for some $e \in \mathrm{Path}(e'; E_0)$. Then $I(p_e) = I(p_{e'})$ and the error exponent $J^*(p) = 0$, a contradiction. $\square$

6 Proof of Proposition 7

Proof We start from the result in Corollary 4, which says that the worst-case type-II error exponent is given as

$$J^*(\Lambda_0, \Lambda_1) = \min_{e_1, e_2 \in E_0 : e_1 \sim e_2} \frac{1}{2} \log \left[\frac{1 - \rho_{e_1}^2 \rho_{e_2}^2}{1 - \rho_{e_1}^2} \right]. \tag{A-16}$$

Thus, the optimization reduces to

$$J^*(\Lambda_0, \Lambda_1) = \min \left\{ \frac{1}{2} \log \left[\frac{1 - \rho_{e_1}^2 \rho_{e_2}^2}{1 - \rho_{e_1}^2} \right] : \eta_1 \le |\rho_{e_i}| \le 1 - \eta_2, i = 1, 2 \right\}. \tag{A-17}$$

Now, it is easy to verify that the function $f : [\eta_1, 1 - \eta_2]^2 \to \mathbb{R}$ given by the recipe $f(x, y) := (1 - x^2 y^2)/(1 - x^2)$ is minimized when x and y attain the lower and upper boundary points respectively. This completes the proof. $\square$

References

[1] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. McGraw-Hill Science/Engineering/Math, 2003.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.