# Enabling Distributed Detection with Dependent Sensors

Brian Proulx, Junshan Zhang, and Douglas Cochran
School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, 85287

***Abstract*—Computational issues affecting the feasibility of optimal distributed detection with correlated measurements are well recognized. We propose utilizing the t-cherry junction tree, an approach based on probabilistic graphical models, to approximate the joint distribution of the correlated measurements. In principle, this approach provides a sequence of progressively more efficiently represented approximations that enable tradeoff between fidelity and compactness. Practically, however, the impact of generating estimated distributions from training data can be significant as the number of parameters to estimate in a distribution grows exponentially with the number of random variables in the distribution. This limitation is quantified and the performance of this approach is illustrated via simulation studies.**

## I. Introduction

In traditional distributed detection problems, a collection of $M > 2$ distributed agents collect data and formulate decisions regarding a binary hypothesis test between common hypotheses $H_0$ and $H_1$. The decisions of the individual agents, represented as binary variates $\mathbf{U} \triangleq \{U_1, \ldots, U_M\}$, are aggregated at a fusion center, which forms a global decision regarding the two hypotheses on the basis of this data.

A prevalent assumption in addressing this problem is that the random variables $U_i$ are conditionally independent under each hypothesis [1], [2]. This assumption is justified in many settings, and it yields a number of appealing outcomes; e.g., if the binary data are conditionally independent and identically distributed, the optimal decision rule at the fusion center is a simple voting scheme. And it is generally a weighted vote even when the data values are not conditionally identically distributed provided the probabilities of detection and false alarm for the local detectors are known [3].

Despite providing highly tractable fusion rules, the performance of the global detector that assumes independence may be poor if the sensors' measurements are correlated. In this scenario, exploiting the full joint conditional distributions of $\mathbf{U}$ at the fusion center yields optimal detection performance but introduces several drawbacks. One key impediment to working with full joint conditional densities of the binary data is that they are burdensome to represent and store. For $M$ independent and identically distributed binary random variables and two hypotheses, the joint conditional densities are summarized by two parameters, which increases to $2M$ parameters if the assumption of identically distributed is removed. In the general case, $2^{M+1}$ parameters are required to catalogue a pair of joint conditional densities; with 50 sensors, this would be larger than $10^{15}$. Another complication is the need to estimate the densities from data samples.

These two issues, namely the number of model parameters to store and the amount of training data necessary to accurately estimate them, point directly to the relative strengths and weaknesses of the two approaches. The first approach, considering all $U_i$ as conditionally independent, involves storing only $2M$ parameters that can be accurately estimated using a small amount of training data. However, this may represent a very coarse approximation to the actual joint distribution of $\mathbf{U}$. The second approach is to consider the full joint distribution of the $U_i$, which is more accurate provided there is ample training data available to estimate the $2^{M+1}$ parameters, as well as storage space for these parameters.

In light of the limitations of these two approaches, we present a method that enables a controllable tradeoff between these two approaches. By constructing a t-cherry junction tree [4] for each hypothesis, approximations to the distribution of the $U_i$ can be constructed of any order, rather than only the conditionally independent (first order) case, or the case with the full joint distribution ($M^{\text{th}}$-order). This approach offers flexibility and sheds light on the impact of the choice of the order on the approximation of the full joint distribution estimated from training data.

The outline of the paper is as follows. In Section II, the t-cherry junction tree is introduced, and the impact of using estimated distributions is quantified in Section III. Simulations are used to demonstrate the performance of this approach in Section IV, and the paper concludes in Section V.

## II. T-Cherry Junction Trees

In this section, we outline the fundamentals of the k-order t-cherry junction tree, as described in [4]. Fundamentally, it represents an approximation of a joint distribution of $M$ random variables. This **full** joint distribution is given the symbol $P(\mathbf{U})$.

*Definition 2.1:* A **junction tree** is a tree structure over a variable set $U_1, \ldots, U_M$ with the following properties:

1) Each node of the junction tree is a subset of random variables, denoted $\mathbf{U}_C$, and is called a cluster. Associated with each cluster is the distribution $p(\mathbf{U}_C)$. These clusters represent the maximal cliques of the associated Markov random field.
2) Every edge connecting two clusters in the junction tree contains a separator, a subset of random variables containing the intersection of the two clusters being

linked: $\mathbf{U}_S = \mathbf{U}_{C_1} \cap \mathbf{U}_{C_2}$. As with clusters, each separator has the distribution $p(\mathbf{U}_S)$ associated with it.

3) If a random variable is contained in two different clusters, it is also contained in every cluster on the path between those two. This is called the running intersection property.
4) The union of all clusters is the entire set of random variables $U_1, \ldots, U_M$.

One key aspect differentiates a *k-order t-cherry* junction tree from a regular junction tree. Namely, all clusters contain exactly $k$ random variables, making the **order** of the tree $k$. Thus, each separator contains $k-1$ random variables. Denoting the set of all clusters as $\mathcal{C}$, and likewise the set of all separators $\mathcal{S}$, it can be shown [4] that $|\mathcal{C}| = M-k+1$ and $|\mathcal{S}| = M-k$.

Two key results conclude the theory behind the k-order t-cherry junction tree. The first is that the approximate joint distribution encoded by a junction tree $P_{JT}(\mathbf{U})$ is [5]

$$P_{JT}(\mathbf{U}) = \frac{\prod_{C\in\mathcal{C}} p(\mathbf{U}_C)}{\prod_{S\in\mathcal{S}} p(\mathbf{U}_S)}. \tag{1}$$

Secondly, the Kullback-Leibler (KL) divergence between the true distribution and the approximation can be written as

$$\begin{aligned} D_{KL}(P(\mathbf{U})||P_{JT}(\mathbf{U})) = &- H(\mathbf{U}) - \sum_{C\in\mathcal{C}} I(\mathbf{U}_C) \\ &+ \sum_{S\in\mathcal{S}} I(\mathbf{U}_S) + \sum_{i=1}^{M} H(U_i), \end{aligned} \tag{2}$$

where $I(\cdot)$ represents the mutual information and $H(\cdot)$ represents the entropy. Thus the weight of a junction tree is

$$w_T \triangleq \sum_{C\in\mathcal{C}} I(\mathbf{U}_C) + \sum_{S\in\mathcal{S}} I(\mathbf{U}_S). \tag{3}$$

The best approximation to $P(\mathbf{U})$ is found by maximizing $w_T$, which however, is an NP hard problem [6].

## III. Data-Driven Distribution Approximation

While the divergence between a t-cherry junction tree approximation and the actual distribution is understood when all distributions are known perfectly, we quantify the divergence between the full and junction tree distributions when the junction tree is constructed using distributions estimated from training data. The true distribution $P(\mathbf{U})$ remains the same, but now the t-cherry junction tree is constructed by first estimating all the necessary marginal distributions. Thus cluster and separator distributions are represented as $\hat{p}(\mathbf{U}_C)$ and $\hat{p}(\mathbf{U}_S)$, resulting in a joint distribution over all $M$ random variables denoted $\hat{P}_{JT}(\mathbf{U})$. In this section, we calculate the scaling behavior of the increased KL divergence resulting from estimated distributions as a function of both the order of the t-cherry junction tree and the number of data samples.

To begin, we calculate the impact of using estimated distributions on the KL divergence between the true full distribution $P(\mathbf{U})$ and the junction tree approximation using estimated distributions $\hat{P}_{JT}(\mathbf{U})$.

*Theorem 3.1:* The KL divergence between the true distribution $P(\mathbf{U})$ and the junction tree constructed using distributions estimated from training data $\hat{P}_{JT}(\mathbf{U})$ is

$$\begin{aligned} D_{KL}(P(\mathbf{U})||\hat{P}_{JT}(\mathbf{U})) = &-H(\mathbf{U}) + \sum_{i=1}^{M} H(U_i) \\ &- \sum_{C\in\mathcal{C}} I(\mathbf{U}_C) + \sum_{S\in\mathcal{S}} I(\mathbf{U}_S) + \sum_{C\in\mathcal{C}} D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) \\ &- \sum_{S\in\mathcal{S}} D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)). \end{aligned} \tag{4}$$

*Proof:* We start with considering the definition of KL divergence between two absolutely continuous discrete distributions, which can be thought of as vectors of probabilities of length $|P(\mathbf{U})|$. Note that each cluster consists of $k$ random variables and is written $\mathbf{U}_C = \{U_{C_1}, \ldots, U_{C_k}\}$ and likewise for separators. The KL divergence is then

$$D_{KL}(P(\mathbf{U})||\hat{P}_{JT}(\mathbf{U})) = \sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{P(\mathbf{U})}{\hat{P}_{JT}(\mathbf{U})}\right). \tag{5}$$

This divergence can be related to the divergence between the true distribution and the distribution from the junction tree created with the true, not estimated, marginal distributions, denoted $P_{JT}(\mathbf{U})$. This can be done by multiplying and dividing to yield

$$\begin{aligned} &D_{KL}(P(\mathbf{U})||\hat{P}_{JT}(\mathbf{U})) = \\ &\qquad D_{KL}(P(\mathbf{U})||P_{JT}(\mathbf{U})) + \sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{P_{JT}(\mathbf{U})}{\hat{P}_{JT}(\mathbf{U})}\right). \end{aligned} \tag{6}$$

Observe that $D_{KL}(P(\mathbf{U})||P_{JT}(\mathbf{U}))$ follows directly from Equation (2); so the term $\sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{P_{JT}(\mathbf{U})}{\hat{P}_{JT}(\mathbf{U})}\right)$ remains to be quantified. Using the definition of the probability density function of a junction tree, this can be rewritten as

$$\begin{aligned} \sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{P_{JT}(\mathbf{U})}{\hat{P}_{JT}(\mathbf{U})}\right) = &\sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{\prod_{C\in\mathcal{C}} p(\mathbf{U}_C)}{\prod_{S\in\mathcal{S}} p(\mathbf{U}_S)}\right) \\ &- \sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{\prod_{C\in\mathcal{C}} \hat{p}(\mathbf{U}_C)}{\prod_{S\in\mathcal{S}} \hat{p}(\mathbf{U}_S)}\right). \end{aligned} \tag{7}$$

Consider the part of this expression with the estimated terms: $-\sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{\prod_{C\in\mathcal{C}} \hat{p}(\mathbf{U}_C)}{\prod_{S\in\mathcal{S}} \hat{p}(\mathbf{U}_S)}\right)$. As each random variable appears within a cluster exactly one more time than it appears in a separator [4], adding and subtracting the term

$\sum_{\mathbf{U}} \log\left(\prod_{C\in\mathcal{C}} \hat{p}(U_{C_1})\cdots\hat{p}(U_{C_k})\right)$ yields that

$$-\sum_{\mathbf{U}} P(\mathbf{U}) \log\left(\frac{\prod_{C\in\mathcal{C}} \hat{p}(\mathbf{U}_C)}{\prod_{S\in\mathcal{S}} \hat{p}(\mathbf{U}_S)}\right) = \sum_{\mathbf{U}} P(\mathbf{U})\Bigg[ -\log\left(\frac{\prod_{C\in\mathcal{C}} \hat{p}(\mathbf{U}_C)}{\prod_{C\in\mathcal{C}} \hat{p}(U_{C_1})\cdots\hat{p}(U_{C_k})}\right) + \log\left(\frac{\prod_{S\in\mathcal{S}} \hat{p}(\mathbf{U}_S)}{\prod_{S\in\mathcal{S}} \hat{p}(U_{S_1})\cdots\hat{p}(U_{S_{k-1}})}\right) - \log\left(\prod_{i=1}^{M} \hat{p}(U_i)\right)\Bigg] \quad (8)$$

To simplify notation, for a set of random variables $\mathbf{X} \triangleq \{X_1,\ldots,X_k\}$, define

$$I_p\left(\hat{p}(\mathbf{X})\right) \triangleq \sum_{\mathbf{X}} p(\mathbf{X}) \log \frac{\hat{p}(\mathbf{X})}{\hat{p}(X_1)\cdots\hat{p}(X_k)}. \quad (9)$$

Returning to Equation (7), it can be rewritten as

$$\sum_{C\in\mathcal{C}} \left[I(p(\mathbf{U}_C)) - I_p(\hat{p}(\mathbf{U}_C))\right] - \sum_{S\in\mathcal{S}} \left[I(p(\mathbf{U}_S)) - I_p(\hat{p}(\mathbf{U}_S))\right] - \sum_{i=1}^{M} H(p(U_i)) + \sum_{\mathbf{U}} p(\mathbf{U}) \log\left(\prod_{i=1}^{M} \hat{p}(U_i)\right). \quad (10)$$

Note that for every cluster

$$I(p(\mathbf{U}_C)) - I_p(\hat{p}(\mathbf{U}_C)) = D_{KL}\left(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)\right) + \sum_{i=1}^{k} D_{KL}(p(U_{C_i})||\hat{p}(U_{C_i})), \quad (11)$$

and similarly for separators. Plugging this into Equation (10) results in

$$\sum_{C\in\mathcal{C}} D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) - \sum_{S\in\mathcal{S}} D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)) - \sum_{C\in\mathcal{C}}\sum_{i=1}^{k} D_{KL}(p(U_{C_i})||\hat{p}(U_{C_i})) + \sum_{S\in\mathcal{S}}\sum_{i=1}^{k-1} D_{KL}(p(U_{S_i})||\hat{p}(U_{S_i})) - \sum_{i=1}^{M} H(U_i) + \sum_{\mathbf{U}} p(\mathbf{U}) \log\left(\prod_{i=1}^{M} \hat{p}(U_i)\right). \quad (12)$$

Using Lemma 1 from [4] and the fact that

$$\sum_{\mathbf{U}} p(\mathbf{U}) \log\left(\prod_{i=1}^{M} \hat{p}(U_i)\right) = \sum_{i=1}^{M} H(U_i) + D_{KL}(p(U_i)||\hat{p}(U_i)), \quad (13)$$

Equation (12) becomes

$$\sum_{C\in\mathcal{C}} D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) - \sum_{S\in\mathcal{S}} D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)). \quad (14)$$

Plugging this into Equation (6) completes the proof. ■

The weight of the estimation terms is defined as $w_E$ and is

$$w_E \triangleq \sum_{C\in\mathcal{C}} D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) - \sum_{S\in\mathcal{S}} D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)). \quad (15)$$

An important property of $w_E$ is that $w_E \geq 0$; that is, using estimated instead of true distributions increases the KL divergence, or at least does not decrease it.

*Lemma 3.1:* Using distributions estimated from data, as opposed to the true distributions, results in an increase (or at least not a decrease) in KL divergence between the true distribution and the junction tree distribution, i.e.,

$$\sum_{C\in\mathcal{C}} D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) - \sum_{S\in\mathcal{S}} D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)) \geq 0. \quad (16)$$

*Proof:* Without loss of generality, assume that for any connected cluster and separator pair, the separator is comprised of $S = \{U_{C_1},\ldots,U_{C_{k-1}}\} \triangleq \mathbf{U}_{C\setminus k}$. Then utilizing the chain rule of conditional divergence [7],

$$D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) = D_{KL}\left(p(U_{C_k}|\mathbf{U}_{C\setminus k})||\hat{p}(U_{C_k}|\mathbf{U}_{C\setminus k})\right) + D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)). \quad (17)$$

Assume, again without loss of generality, that $\mathcal{C}$ and $\mathcal{S}$ are ordered such that separator $S_1$ is on the edge between clusters $C_1$ and $C_2$, $S_2 = C_2 \cap C_3$, and so forth. Then, as there are $M-k$ cluster-separator pairs and one remaining cluster,

$$\sum_{C\in\mathcal{C}} D_{KL}(p(\mathbf{U}_C)||\hat{p}(\mathbf{U}_C)) - \sum_{S\in\mathcal{S}} D_{KL}(p(\mathbf{U}_S)||\hat{p}(\mathbf{U}_S)) = \sum_{C\in\{C_1,\ldots,C_{M-k}\}} D_{KL}\left(p(U_{C_k}|\mathbf{U}_{C\setminus k})||\hat{p}(U_{C_k}|\mathbf{U}_{C\setminus k})\right) + D_{KL}(p(\mathbf{U}_{C_{M-k+1}})||\hat{p}(\mathbf{U}_{C_{M-k+1}})) \geq 0. \quad (18)$$

■

The behavior of the estimation weight $w_E$ is essential to quantify the tradeoff between allowing more complicated correlation structures (thus a more accurate approximation to the true distribution) and the difficulty inherent in estimating distributions that contain more random variables. As the number of entries to be estimated in a distribution scales exponentially with the number of random variables in the joint distribution, an exponential increase in $w_E$ is expected, and Theorem 3.2 demonstrates that this is the case.

*Theorem 3.2:* A) An approximation to the expectation of the KL divergence between a discrete distribution and its

estimate, using $N$ samples, is

$$\mathbb{E}\left[D_{KL}(p(\mathbf{X})||\hat{p}(\mathbf{X}))\right] \sim \frac{1}{2}\sum_{i=1}^{|p|}\frac{1-p_i}{N}. \quad (19)$$

B) An approximation to expected value of the estimation error weight for the binary variable case is

$$\mathbb{E}[w_E] \approx \frac{M-k+1}{N}\left(2^{k-1}-2^{-k}\right) - \frac{M-k}{N}\left(2^{k-2}-2^{-(k-1)}\right). \quad (20)$$

*Proof:* A local approximation of the KL divergence between two distributions $p$ and $q$, $D_{KL}(p||q)$, can be constructed by considering $q$ as a perturbed form of $p$. Specifically, let $q = p(1+v)$, where $v$ is a vector that redistributes probability mass of $p$. In order to ensure that $\sum p(1+v) = 1$, it is necessary that $\sum_{i=1}^{|p|} p_i v_i = 0$. This yields the approximation [8]

$$D_{KL}(p||p(1+v)) \sim \frac{1}{2}\sum_{i=1}^{|p|} p_i v_i^2. \quad (21)$$

For this proof, consider the divergence of a distribution $p$ over a set of random variables $\mathbf{X} \triangleq \{X_1, \ldots, X_M\}$: $D_{KL}(p(\mathbf{X})||\hat{p}(\mathbf{X}))$. With enough samples $N$, this divergence should be small, and Equation (21) will represent a good approximation to the true divergence. To understand how the divergence between the true distribution and estimated distribution scales as a function of the number of random variables contained in each distribution (the order of the junction tree), consider an entry in the estimated distribution, $\hat{p}_i$. Every $\hat{p}_i$ is considered to be independent of every other $\hat{p}_j$. Each $\hat{p}_i$ is modeled as a binomial random variable with probability of success $p_i$ and $N$ trials. A good approximation when $N$ is large is to consider each $\hat{p}_i$ as as a normal random variable centered at the true value $p_i$. That is, $\hat{p}_i \sim \mathcal{N}(p_i, \frac{p_i(1-p_i)}{N})$.

Let $Y_i \triangleq v_i\sqrt{p_i}$, then $Y_i \sim \mathcal{N}(0, \frac{1-p_i}{N})$. This implies that $\frac{N}{1-p_i}Y_i^2 \sim \chi_1^2$ and $\mathbb{E}[Y^2] = \frac{1-p_i}{N}$. Plugging this into the expression for the divergence yields

$$\mathbb{E}\left[D_{KL}(p(\mathbf{X})||\hat{p}(\mathbf{X}))\right] \sim \mathbb{E}\left[\frac{1}{2}\sum_{i=1}^{|p|} p_i v_i^2\right] = \mathbb{E}\left[\frac{1}{2}\sum_{i=1}^{|p|} Y_i^2\right] = \frac{1}{2}\sum_{i=1}^{|p|}\frac{1-p_i}{N}, \quad (22)$$

which proves the first part of the theorem.

For the second part of the theorem, a more specific approach is taken, though it should be noted this approach is easily generalized to any collection of discrete random variables. Specifically, consider all random variables $X_j$, $j = 1, \ldots, M$, to be binary random variables. In order to provide an upper bound on the approximation to the divergence, consider each $X_j$ to be uniformly distributed and assume that this uniformity holds as the number of random variables in the distribution, denoted $k$, increases. For example, when $k = 1$, $p(\mathbf{X}) = [0.5, 0.5]$ and for $k = 2$, $p(\mathbf{X}) = [0.25, 0.25, 0.25, 0.25]$, and so forth. Note that here these entries in $p(\mathbf{X})$ are in lexicographical order. These assumptions results in the upper bound

$$\mathbb{E}\left[D_{KL}(p(\mathbf{X})||\hat{p}(\mathbf{X}))\right] \approx \frac{1}{N}\left(2^{k-1}-2^{-k}\right). \quad (23)$$

Using Equation (23) and noting that there are $M-k+1$ clusters and $M-k$ separators completes the proof. ■

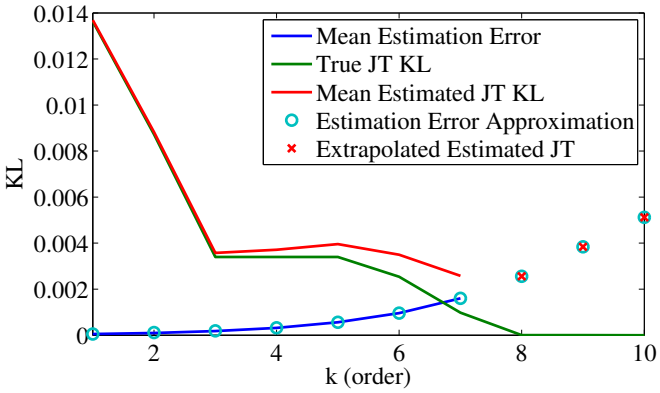


Fig. 1. Divergence using estimated distributions

To demonstrate these concepts, the behavior of the junction tree constructed for hypothesis $H_0$ as explained in Section IV is presented in Figure 1. Note that the approximation from Equation (23) captures the scaling behavior of $w_E$. Additionally, the true junction tree does not exactly model the joint distribution until the order is eight, and thus combined with the increasing $w_E$, the third order junction tree represents a local minimum of the divergence. Also interesting in this figure is the lack of entries for $w_E$ after the seventh order tree. In this case, there were distributions in the junction tree that were estimated from the training data that resulted in at least one $\hat{p}_i = 0$ where the corresponding $p_i \neq 0$. This represents essentially an absolute continuity error that causes the KL divergence to be $\infty$. This occurred in an eighth order tree despite having $10^5$ samples, which would initially seem adequate to estimate distributions containing eight binary random variables, and thus 256 parameters.

## IV. Numerical Examples

In order to evaluate the approach developed above, utilizing junction trees, as well as examine the impact of constructing a junction tree from training data, we present a simulation study. Consider a sensor deployment of $M = 10$ sensors. The measurement data taken by these sensors is correlated, and the structure of the correlation is captured by a Markov random field. The specific structure of this Markov random field is seen in Figure 2. Note that the maximum clique size is three, so an optimal t-cherry junction tree of order three will be exactly equal to the true distribution.

For each hypothesis, $H_0$ and $H_1$, there are $N = 10^5$ training samples of the true distribution $p(\mathbf{U})$. Using this training data, two t-cherry junction trees are constructed, one for

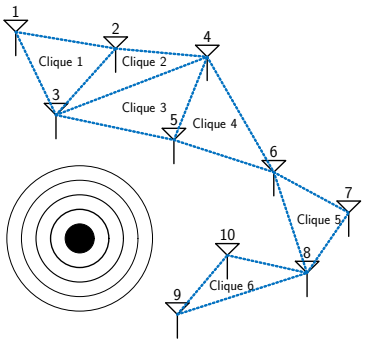


Fig. 2. Markov random field structure

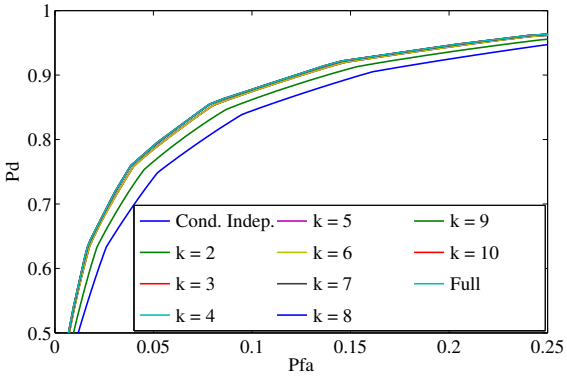


Fig. 3. ROC curves with true distributions

each hypothesis. This process is averaged over 1,000 different realizations of the sample data.

The Receiver Operating Characteristic (ROC) curves of the detector are presented in Figure 3 for the cases in which considering the sensors as conditionally independent, second through ninth order t-cherry junction trees, and using the full distribution. Notice that when the sensors are considered conditionally independent, the performance is the worst. Even using a second order junction tree results in a significant improvement. Also, the third through sixth order junction trees all result in essentially identical performance. These ROC curves are not the same as the true distribution curve because the t-cherry junction trees constructed were not optimal. From the seventh order onward, the junction trees have managed to exactly replicate the true distribution. In conclusion, using a second and third order t-cherry junction tree offered significant gains over the conditionally independent case, and higher order t-cherry junction trees resulted in diminishing returns.

The area under the ROC curve (AUC) is another metric used to evaluate the performance of a detector. Figure 4 plots the performance of t-cherry junction trees when using the true distributions and distributions estimated from training data. Notice that the performances are similar until the ninth order t-cherry junction tree. In this case, the estimated distributions are too inaccurate due to an insufficiency of training data, and the performance is greatly degraded. This emphasizes the importance of correctly trading off the gain resulting

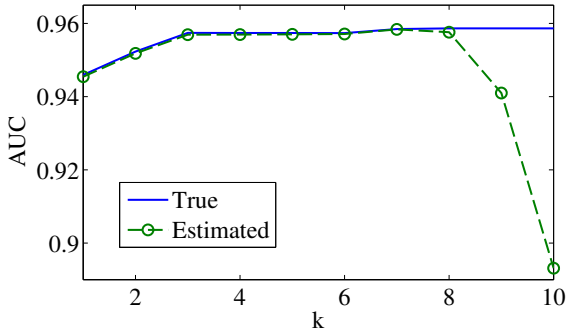


Fig. 4. AUC of the detectors

from using a (more accurate) higher order t-cherry junction tree against the increased "noise" that results from imprecise estimation of the "larger" distributions involved.

## V. Conclusion

In this paper, we presented a novel approach to improve the performance of a distributed detection system: the t-cherry junction tree. This tool allows for a smooth tradeoff between treating each sensor as conditionally independent and utilizing the full joint distribution of the individual sensor decisions. Using the t-cherry junction tree allows for some correlation structure in the sensors' decisions to be preserved while simultaneously requiring dramatically fewer parameters to store than storing the full joint distribution.

Additionally, the closed form KL divergence of a junction tree approximation from the true distribution when distributions estimated from data was calculated. This degradation in the ability to closely approximate is exponential in increasing junction tree order, which implies that in many circumstances, there is a maximum order beyond which the error from estimating distributions containing more random variables outpaces any potential gain from allowing more complicated correlation structures.

## References

[1] J. N. Tsitsiklis, "Problems in decentralized decision making and computation." Ph.D. dissertation, Massachusetts Institute of Technology, 1984.

[2] ——, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals, and Systems*, 1988.

[3] C. W. Helstrom, *Elements of Signal Detection and Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[4] T. Szantai and E. Kovacs, "Hypergraphs as a mean of discovering the dependence structure of a discrete multivariate probability distribution," *Annals of Operations Research*, vol. 193, no. 1, pp. 71–90, 2012.

[5] S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 50, no. 2, pp. pp. 157–224, 1988.

[6] B. Proulx and J. Zhang, "Modeling social network relationships via t-cherry junction trees," in *Proceedings of IEEE INFOCOM*, April 2014, pp. 2229–2237.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[8] E. Abbe and L. Zheng, "Linear universal decoding for compound channels," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 5999–6013, Dec 2010.