Integrating Probability Constraints into Bayesian Nets

Yun Peng[1] and Shenyong Zhang[2]

Abstract. This paper presents a formal convergence proof for E-IPFP, an algorithm that integrates low dimensional probabilistic constraints into a Bayesian network (BN) based on the mathematical procedure IPFP. It also extends E-IPFP to deal with constraints that are inconsistent with each other or with the BN structure.

1 CONVERGENCE OF E-IPFP

Let $G=(G_s,G_P)$ denote the given BN of n variables $x=(x_i,\cdots x_n)$, where $G_S=\{(x_i,\pi_i)\}$ gives the network structure and $G_P=\{P(x_i\,|\,\pi_i)\}$ is the set of conditional probability tables (CPTs). Denote JPD of x defined by G as $P(x)$. Let $R=\{R_1(y^1),$ $R_2(y^2),\cdots,R_m(y^m)\}$ be a set of probabilistic constraints, where $R_j(y^j\subseteq x)$. Our objective is to construct a new BN $G'=(G'_s,G'_P)$ with its JPD $P'(x)$ meeting the following conditions:

C1: *Constraint satisfaction*: $P'(y^j)=R_j(y^j)\ \ \forall\ R_j(y^j)\in R$;

C2: *Structural invariance*: $G_S^{'}=G_S$;

C3: *Minimality*: $P'(x)$ is as close to $P(x)$ as possible.

E-IPFP [1] is based on the mathematical procedure IPFP (iterative proportional fitting procedure) [2] which iteratively modifies the JPD by the constraints until convergence. It has been shown that the converging JPD satisfies all constraints in R (C1) and is closest to the original JPD measured by the I-divergence (C3). To satisfy the structural invariance (C2), E-IPFP extends IPFP by making the BN structure (G_S) an additional constraint

$$R_{m+1}(x)=\Pi_{i=1}^{n}Q_{k-1}(x_i\,|\,\pi_i)\,. \tag{1}$$

E-IPFP($G=(G_s,G_P)$, $R=\{R_1,R_2,\cdots R_m\}$) {

1. $Q_0(x)=\Pi_{i=1}^{n}P(x_i\,|\,\pi_i)$ where $P(x_i\,|\,\pi_i)\in G_P$;
2. Starting with k = 1, repeat the following procedure until convergence

 { 2.1. j = ((k-1) *mod* (m+1)) + 1;

 2.2. if $j < m$+1

 $Q_k(x)=Q_{k-1}(x)R_j(y^j)/Q_{k-1}(y^j)$

 2.3. else

 {extract $Q_k(x_i\,|\,\pi_i)$ from $Q_k(x)$ according to G_S ;

 $Q_k(x)=\Pi_{i=1}^{n}Q_k(x_i\,|\,\pi_i)$;}

 2.4. k = k+1;}

3. return $G'=(G_S,G_P^{'})$ with $G_P^{'}=\{Q_k(x_i\,|\,\pi_i)\}$;}

E-IPFP is exactly the same as standard IPFP except in Step 2.3 where the structural constraint applies. However, convergence proofs for IPFP's [2,3] do not apply to E-IPFP because 1) R_{m+1} changes its value in every iteration and 2) the set of all JPD satisfying G_S is not convex. We have shown in [4] IPFP with $R=\{R_1(y^1),\cdots R_m(y^m)\}$ is equivalent to IPFP with a *single* composite constraint $R'(y=y^1\cup y^2\cdots\cup y^m)$, which is computed by applying IPFP to $Q_0(y)$ with $R=\{R_1(y^1),\cdots R_m(y^m)\}$. So it suffices to prove the convergence of E-IPFP with a single constraint $R(y)$.

Denote the set of JPD of x that satisfy $R(y)$ as $\mathbf{P}_{R(y)}$ and the set of JPD that satisfy structural constraint as $\mathbf{P}_{G_S}$. Let $Q_0(x)=\Pi_{x_i\in x}P(x_i\,|\,\pi_i)$ be the JPD of the given BN; $Q_1(x)=Q_0(x)R(y)/Q_0(y)$ the I-Projection of $Q_0(x)$ to $\mathbf{P}_{R(y)}$; $Q_2(x)=\Pi_{x_i\in x}Q_1(x_i\,|\,\pi_i)$ the structural constraint; and $Q_3(x)=Q_2(x)R(y)/Q_2(y)$ be the I-Projection of $Q_2(x)$ back to $\mathbf{P}_{R(y)}$.

Points of Q_0 through Q_3 are depicted in Figure 1 below. Note that Q_1 is obtained from Q_0 by Step 2.2, Q_2 from Q_1 by Step 2.3, and Q_3 from Q_2 by Step 2.2 in the next iteration of E-IPFP.

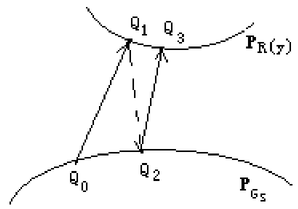

Figure 1. Successive JPDs from E-IPFP

The convergence of E-IPFP can be established by showing $I(Q_1\,||\,Q_0)\geq I(Q_3\,||\,Q_2)$, i.e., the I-divergence between the two end-points of I-projection to $\mathbf{P}_{R(y)}$ is monotonically decreasing in successive iterations. Since $Q_1,Q_3\in\mathbf{P}_{R(y)}$, and Q_3 is an I-Projection of Q_2, we have $I(Q_1\,||\,Q_2)\geq I(Q_3\,||\,Q_2)$. So E-IPFP converges if

$$\Delta(x)=I(Q_1\,||\,Q_0)-I(Q_1\,||\,Q_2) \tag{2}$$

Is non-negative

Theorem 1. *For any given BN* $G=(G_s,G_P)$ *and* $R(y)$, $\Delta(x)\geq 0$.

Proof. By induction on $|x|$, the number of variables in G.

Base case: $|x|$ = 1, $x=(x_1)$, the constraint is $R(x_1)$. It is trivial that $Q_2(x_1)=Q_1(x_1)=R(x_1)$. Then by (8)

$$\Delta(x)=\sum R(x_1)\log\frac{R(x_1)}{Q_0(x_1)}=I(R(x_1)\,||\,Q_0(x_1))\geq 0,$$

Inductive assumption: $\Delta(x_1,x_2,...,x_n)\geq 0$ for any $n\geq 1$.

Inductive proof: show that $\Delta(x_0,x_1,x_2,...,x_n)\geq 0$. Without loss of generality, let x_0 be a root node of the BN. For clarity, let $x=(x_1,x_2,...,x_n)$. By (2),

$$\Delta(x_0,x)=\sum_{x_0,x}Q_1(x_0,x)\log\frac{Q_2(x_0,x)}{Q_0(x_0,x)}=\sum_{x_0,x}Q_1(x_0,x)\log\frac{Q_1(x_0)Q_2(x\,|\,x_0)}{Q_0(x_0)Q_0(x\,|\,x_0)}$$

$$=\sum_{x_0,x}Q_1(x_0,x)\log\frac{Q_1(x_0)}{Q_0(x_0)}+\sum_{x_0,x}Q_1(x_0,x)\log\frac{Q_2(x\,|\,x_0)}{Q_0(x\,|\,x_0)}\ =\Delta_1(x_0,x)+\Delta_2(x_0,x)$$

$$\Delta_1(x_0,x)=\sum_{x_0,x}Q_1(x_0,x)\log\frac{Q_1(x_0)}{Q_0(x_0)}=\sum_{x_0}(\sum_{x}Q_1(x_0,x))\log\frac{Q_1(x_0)}{Q_0(x_0)}$$

$$=\sum_{x_0}Q_1(x_0)\log\frac{Q_1(x_0)}{Q_0(x_0)}=I(Q_1(x_0)\,||\,Q_0(x_0))\geq 0 \tag{3}$$

Now consider Δ_2.

Case 1. $x_0\in y$ .Let $y'=y\setminus\{x_0\}$, then $R(y)=R(x_0,y')$. Since

[1] Univ. Maryland, Baltimore County, USA, email: ypeng@umbc. edu
[2] University of Science and Technology of China, China, email: syzhang@mail.ustc.edu.cn

$Q_1(x_0, x) = Q_1(x_0) \cdot Q_1(x \mid x_0)$ and

$$Q_1(x_0, x) = Q_0(x_0)\frac{R(x_0)}{Q_0(x_0)} Q_0(x \mid x_0)\frac{R(y' \mid x_0)}{Q_0(y' \mid x_0)}$$

and $Q_1(x_0) = R(x_0)$, then $Q_1(x \mid x_0) = Q_0(x \mid x_0) R(y' \mid x_0) / Q_0(y' \mid x_0)$.

Note that, for any particular state x_0^* of variable x_0, $Q_0(x \mid x_0^*) = \Pi_{x_i \in x} Q_0^*(x_i \mid \pi_i)$ is a BN of *x*, where

$$Q_0^*(x_i \mid \pi_i) = \begin{cases} Q_0(x_i \mid \pi_i, x_0 = x_0^*) & \text{if } x_i \text{ is a child of } x_0; \\ Q_0(x_i \mid \pi_i) & \text{otherwise.} \end{cases} \quad (4)$$

Therefore, $Q_1(x \mid x_0^*)$ is an I-Projection of $Q_0(x \mid x_0^*)$ to $\mathbf{P}_{R(y' \mid x_0)}$ from which CPTs of $Q_2(x \mid x_0^*)$ are extracted, so

$$\sum_x Q_1(x \mid x_0^*) \log \frac{Q_2(x \mid x_0^*)}{Q_0(x \mid x_0^*)} = \Delta(x \mid x_0^*) \geq 0;$$

by inductive assumption, and

$$\Delta_2(x_0, x) = \sum_{x_0} Q_1(x_0) \sum_x Q_1(x_0 \mid x) \log \frac{Q_2(x \mid x_0)}{Q_0(x \mid x_0)} \geq 0 \quad (5)$$

Case 2. $x_0 \notin y$. By definition of Q_1, we have

$$Q_1(x \mid x_0) = Q_0(x \mid x_0) \frac{R(y) / Q_1(x_0)}{Q_0(y) / Q_0(x_0)}.$$

Since $Q_0(y) / Q_0(x_0) = Q_0(y \mid x_0) / Q_0(x_0 \mid y)$, then

$$Q_1(x \mid x_0) = Q_0(x \mid x_0) \frac{R^*(y)}{Q_0(y \mid x_0)} \quad (6)$$

where $R^*(y) = R(y) Q_0(x_0 \mid y) / Q_1(x_0)$.

It can be shown easily that $R^*(y)$ is a PD of *y*. Therefore, for any given x_0^*, by (6), $Q_1(x \mid x_0^*)$ is an I-Projection of $Q_0(x \mid x_0^*)$ to $\mathbf{P}_{R^*(y)}$. Then by inductive assumption and analogous to (5),

$$\Delta_2(x_0, x) = \sum_{x_0, x} Q_1(x_0, x) \log \frac{Q_2(x \mid x_0)}{Q_0(x \mid x_0)} \geq 0. \quad \square$$

2 INCONSISTENT CONSTRAINTS

When constraints $R = \{R_1(y^1), \cdots R_m(y^m)\}$ are inconsistent either with each other or with the BN structure, E-IPFP (and IPFP) will not converge to a single point but rather oscillates between some JPDs. We have developed an algorithm SMOOTH to deal with inconsistent constraints for IPFP with general JPD [9]. Now we adopt it to E-IPFP. The basic idea of SMOOTH is to make the modification *bi-directional*: at each iteration, not only the JPD is pulled closer to the constraint but also the constraint is pulled towards the current JPD. By doing so, the inconsistency among the constraints is gradually reduced or *smoothened*.

E-IPFP-SMOOTH($G = (G_s, G_P)$, $R = \{R_1, R_2, \cdots R_m\}$) {

1. $Q_0(x) = \Pi_{i=1}^n P(x_i \mid \pi_i)$ where $P(x_i \mid \pi_i) \in G_P$;
2. Starting with *k* = 1, repeat the following procedure until convergence

   { 2.1. *j* = ((*k*-1) *mod* (*m*+1)) + 1;

   2.2. if *j* < *m*+1

   $\{R_j(y^j) = \alpha R_j(y^j) + (1-\alpha) Q_{k-1}(y^j);$

   $Q_k(x) = Q_{k-1}(x) \cdot \frac{R_j(y^j)}{Q_{k-1}(y^j)};\}$

   2.3. else

   {extract $Q_k(x_i \mid \pi_i)$ from $Q_k(x)$ according to G_S;

   $Q_k(x) = \Pi_{i=1}^n Q_k(x_i \mid \pi_i)$;}

   2.4. $k = k+1$;}
3. return $G' = (G_S, G_P')$ with $G_P' = \{Q_k(x_i \mid \pi_i)\}$;}

Note that this algorithm differs from E-IPFP only in Step 2.2 where it modifies the constraint before the I-projection is performed. The convergence of E-IPFP-SMOOTH is given in the theorem below. Here we only deal with the situation that the constraints are inconsistent with the BN structure (the convergence for situations in which constraints are inconsistent with each other has been established in our earlier work [4]). Similar to Theorem 1, we only show the convergence with a single (possibly composite) constraint.

Theorem 2. *For any given BN* $G = (G_s, G_P)$ *and constraint R*(*y*) *inconsistent with* G_S*, E-IPFP-SMOOTH converges to Q* consistent with* G_S.

Recall that from Theorem 1 we have $I(Q_1 \| Q_0) \geq I(Q_3 \| Q_2)$, where, as shown in Figure 1, Q_3 is an I-projection of Q_2 to $\mathbf{P}_{R(y)}$ if E-IPFP is used. Now with E-IPFP-SMOOTH, *R*(*y*) is modified in Step 2.2 to

$$R'(y) = \alpha R(y) + (1-\alpha) Q_2(y) \quad (7)$$

Let Q_3' be the I-projection of Q_2 to $\mathbf{P}_{R'(y)}$ using $R'(y)$. To show the convergence of E-IPFP-SMOOTH, we only need to show that $I(Q_1 \| Q_0) \geq I(Q_3' \| Q_2)$. This can be done by showing that

$$I(Q_3 \| Q_2) \geq I(Q_3' \| Q_2) \quad (8)$$

We proof (8) by showing that when α moves from 0 toward 1, $I(Q_3' \| Q_2)$ strictly increases from 0 toward $I(Q_3 \| Q_2)$. Due to the page limit, the actual proof of Theorem 2 is omitted.

Experiments with BN of different size and with different sets of constraints (both marginal and conditional) have shown that E-IPFP and its E-IPFP-SMOOTH work as expected with time complexity exponential to the BN size. The computation can be significantly speed-up if the constraint set R can be decomposed and the update is allowed to be localized (see D-IPFP in [1]). Further speed-up can be achieved for the SMOOTH versions by allowing the smooth factor α to gradually decreasing toward 0 (see [4]).

SMOOTH modifies the constraints to fit the BN structure when they are not consistent. It is more challenging to change the structure to fit the constraints. We are actively working on this problem and have some leads that are interesting and promising.

ACKNOWLEDGEMENT

This work was supported in part by NIST award 60NANB6-D6206 and the China Scholarship Council (CSC).

REFERENCES

[1] Y. Peng and Z. Ding, 'Modifying Bayesian Networks by Probability Constraints', *Proc. 21st Conference on Uncertainty in Artificial Intelligence*, July 2005, Edinburgh.

[2] J. Vomlel, *Methods of Probabilistic Knowledge Integration*, PhD Thesis, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Dec 1999.

[3] I. Csiszár, 'Information-type measures of difference of probability distributions and indirect observation', *Studia Sci. Math. Hungar*., **2**, 229-318 (1967).

[4] Peng, Y., Zhang, S., Pan, R.: "Bayesian Network Reasoning with Uncertain Evidences", to appear *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 2010.