# Conditions For Hyper-EM And Large Graphical Modelling

Seong-Ho Kim and Sung-Ho Kim[1]

ABSTRACT

We propose an improved version of Kim (2000) to the effect that in principle we may deal with a graphical model of any size. Kim (2000) proposed a method of estimating parameters for a model of categorical variables which is too large to handle as a single model. We applied the proposed method to a simulated data of 158 binary variables.

*Keywords:* between-submodel equivalence; d-split; EM; hyper-EM; marginal restructuring; removable nodes

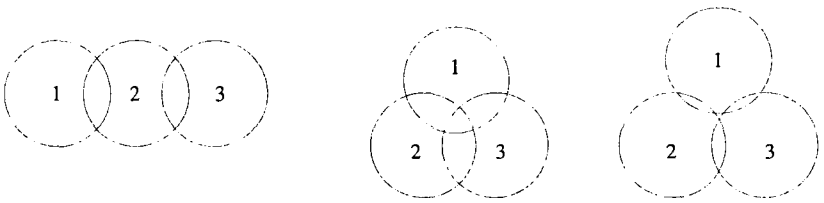
## 1. Introduction

Kim (2000) expands the range of application of EM into large Bayesian networks of categorical variables, where the model is too large to handle as a single model. Suppose a graphical model is d-split (to be defined later) into a set of submodels. If a set of neighboring submodels share a set $S$ of variables and the estimates for the variables in $S$ are the same whether they are obtained from one of the neighboring submodels or another, then we say that *between-submodel equivalence* (BSE for short) holds at $S$. The BSE is not guaranteed in the E-step of the hyper-EM as proposed by Kim (2000), but it is attained in the M-step. In this paper, we propose a new version of the hyper-EM so that the BSE may be attained in both of E-step and M-step and under the specific condition the M-step may become simpler than that of the hyper-EM as derived in Kim (2000).

## 2. D-splitting

We denote a graph $\mathcal{G}$ by $\mathcal{G} = (V, E)$, where $V$ is a set of the nodes involved in $\mathcal{G}$ and $E$ a set of the directed edges or arrows between nodes in $V$ of the form $(i, j)$. The directed edge $(i, j)$ stands for the arrow from node $i$ to $j$. The nodes are labelled in accordance with the direction of arrows, i.e., the nodes are well-ordered. If a pair of sets of nodes share a non-empty set of nodes, then we call the pair *neighboring* sets. If there is an arrow $(a, b)$, then we will say that node $a$ is a parent of node $b$. We will be

[1]Division of Applied Mathematics, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, South Korea. e-mail: mathan@mail.kaist.ac.kr and shkim@mail.kaist.ac.kr




(1) a series mode (2) a ring mode (3) a ring mode

Figure 3.1: Three different modes of submodel arrangement. Circles symbolize submodels

denoted by $pa(a)$ the set of the parents of node $a$ and by $ch(a)$ the set of the children. Let $fa(a) = \{a\} \cup pa(a)$. The node which does not have any child node will be called a *terminal* node, and the node which does not have any parent node will be called a *root* node. A subset is *complete* if it induces a complete subgraph. A complete subset that is maximal is called a *clique*.

**Definition 2.1.** *A chain $C$ from $\alpha$ to $\beta$ in a directed acyclic graph $\mathcal{G}$ is said to be* blocked *by $S$, if it contains a node $\gamma \in C$ such that either*

*(i) $\gamma \in S$ and arrows of $C$ do not meet head to head at $\gamma$, or*

*(ii) $\gamma \notin S$, nor has $\gamma$ any descendants in $S$, and arrows of $C$ do meet head to head at $\gamma$.*

We will denote by $V$ the index set of all the variables involved in a model and by $V_i$ for submodel $i$, and we will use the lowercase $x$ to denote the cell location of a contingency table, with $x_A$ and $x^i$ for the contingency table of the variables indexed in $A$ and in $V_i$, respectively. We will denote by $i \wedge j$ the index set of the variables that are involved in both submodels $i$ and $j$. If $V_i$ and $V_j$ are the index sets of the variables involved in submodels $i$ and $j$, respectively, then we have $i \wedge j = V_i \cap V_j$.

**Definition 2.2.** *We will say that the graph $\mathcal{G}$ of a recursive model is split in the sense of d-separation if, for any pair of subgraphs, say $\mathcal{G}_1 = (V_1, E_1)$ and $\mathcal{G}_2 = (V_2, E_2)$, $V_1 \setminus V_2$ is d-separated from $V_2 \setminus V_1$ by $V_1 \cap V_2$. We will call the splitting in the sense of d-separation* d-splitting.

## 3. Family And Hyper-EM Conditions

After a d-splitting, the whole model can be arranged in the form of submodel-chain. For instance, in Figure 3.1, the graphs of submodels are linked in series in panel (1), while they are linked in a ring in the other panels where a circle stands for a set of the

variables involved in a submodel and the intersection of the two circles is the set of the variables that are shared by the two submodels. In panel (2) all the submodels share a set of variables, while there is no such set in any of the panels (1) and (3).

**Definition 3.1.** *Let a node $v$ be in $V$ with graph $\mathcal{G} = (V, E)$. If the marginal of $V \setminus \{v\}$ of the MLE of the joint probability function is the same to the MLE of the marginal of $V \setminus \{v\}$, i.e.,*

$$\sum_{x \,:\, x_{\{v\}} \in \mathcal{X}_{\{v\}}} \hat{P}_V(x) = \widehat{\Big( \sum_{x \,:\, x_{\{v\}} \in \mathcal{X}_{\{v\}}} P_V(x) \Big)}, \tag{3.1}$$

*then we will call a node $v \in V$ a* removable *node.*

**Theorem 3.1.** *Consider a node $v$ in $V$ with graph $\mathcal{G} = (V, E)$. (i) $fa(ch(v))$ is a unique clique which includes a node $v$ or (ii) $v$ is a terminal node where $|ch(v)|$ is the number of nodes of $ch(v)$ if and only if a node $v$ is removable.*

Suppose $A \subseteq V$ and $|A| \geq 2$. If all the nodes in $A$ can be removed one after another, we will say that $A$ is *sequentially removable.* If a pair of submodels of a model shares a non-empty set of variables, we will call the submodels of the pair *neighboring submodels* and call one of them a neighbor submodel of the other.

**Definition 3.2.** *Suppose a recursive model is d-split into $k$ submodels. When $\mathcal{G}_i$ and $\mathcal{G}_j$ are neighbors, we obtain the structure for $V_i$, $V_i \cap V_j$, and $V_j$ in the expression $E_i \cup E_i^{i \wedge j}$, $E^{i \wedge j} (= E_i^{i \wedge j} \cup E_j^{i \wedge j})$, and $E_j \cup E_j^{i \wedge j}$, respectively. We will call this procedure the* marginal restructuring *for $V_i \cap V_j$ where $E_i^{i \wedge j}$ is the subset of $E_i$ confined to $V_i \cap V_j$.*

> **Family Condition:** Consider the graphs of submodels $1, 2, \cdots, k$ of a recursive model. If there is a node $v$ which is in the graphs of $m$ $(m \leq k)$ submodels,
>
> $$\mathcal{G}_{i_1} = (V_{i_1}, E_{i_1}), \mathcal{G}_{i_2} = (V_{i_2}, E_{i_2}), \cdots, \mathcal{G}_{i_m} = (V_{i_m}, E_{i_m}),$$
>
> then $pa(v) \subseteq V_j$ for some $j \in \{i_1, i_2, \cdots, i_m\}$.
>
> **Hyper-EM Condition:** Consider the graphs of submodels $1, 2, \cdots, k$ of a recursive model. Suppose the marginal restructuring is made on $V_i \cap V_j$ for all $i < j$. $V_j \setminus V_i$ is sequentially removable where $i < j$.

We can show that the hyper-EM works with submodels d-split under the Hyper-EM condition.

## 4. Proposed Hyper-EM

Consider neighboring submodels $i$ and $j$ of a recursive model d-slit under Hyper-EM conditon and the marginal restructuring is made on $V_i \cap V_j$. Denote the E-step operation on submodel $i$ by $\epsilon(i)$ and the estimate-transfusion from submodel $i$ into submodel $j$ by $\tau(i,j)$.

The procedure of submodel E-step and estimate-transfusion(SEET procedure) is described below. Once an E-step is implemented within submodel $i$, the resulting estimates are transfused into neighboring submodels and the transfusion continues until the estimate-transfusion takes place into all the submodels except submodel $i$. If the submodels are arranged as in panel (1) of Figure 3.1 and the SEET procedure begins with $\epsilon(1)$, $\epsilon(2)$, and $\epsilon(3)$, then the SEET procedure proceeds as in

$$\epsilon(1) \rightarrow \tau(1,2) \rightarrow \tau(2,3); \tag{4.1}$$

$$\text{either} \quad \epsilon(2) \rightarrow \tau(2,3) \rightarrow \tau(2,1) \quad \text{or} \quad \epsilon(2) \rightarrow \tau(2,1) \rightarrow \tau(2,3); \tag{4.2}$$

$$\epsilon(3) \rightarrow \tau(3,2) \rightarrow \tau(2,1), \tag{4.3}$$

respectively. In general, when the arrangement is in series mode and the SEET procedure begins at a submodel $i$ which is not located at an end of the submodel-chain, the estimate-transfusion may be made in any order as long as the transfusion proceeds outward from submodel $i$. When $k$ submodels are arranged in a ring, the submodels are symmetric in terms of location in the whole model. Suppose $k$ submodels are arranged in a ring and denote by $T$ the index set of the variables that are involved in at least two submodels. Then each of the $k$ submodels is d-separated from the other submodels by $T$. We will consider $T$ as a submodel in a ring-mode arrangement. So, if an E-step is done in submodel $i$, then the transfusion proceeds into the $T$ set and then into the submodels other than submodel $i$ from the $T$ set one after another. Re-expressing this process yields

$$\epsilon(i) \rightarrow \tau(i,T) \rightarrow \tau(T,j_1) \rightarrow \tau(T,j_2) \rightarrow \cdots \tau(T,j_{k-1}), \tag{4.4}$$

where $j_1, \cdots, j_{k-1}$ are a permutation of $1, 2, \cdots, i-1, i+1, \cdots, k$. We do the same for all the $k$ submodels one after another. And we repeat the same round of SEET procedures until convergence takes place. We will call by a *full* SEET procedure such processes as (4.1) through (4.4) where the estimate-transfusion is made throughout the whole model. As a matter of fact we may not do such full SEETs. We will describe this in detail using the ring-shape arrangement in Figure 3.1. Consider the following SEET procedures which make up a cycle of IPF and will iterate until convergence:

$$\epsilon(1) \rightarrow \tau(1,T) \rightarrow \tau(T,2) \rightarrow \tau(T,3) \tag{4.5}$$

$$\epsilon(2) \rightarrow \tau(2,T) \rightarrow \tau(T,3) \rightarrow \tau(T,1) \tag{4.6}$$

$$\epsilon(3) \rightarrow \tau(3,T) \rightarrow \tau(T,1) \rightarrow \tau(T,2) \tag{4.7}$$

Assuming that a cycle of the IPF runs in the order of the expressions, (4.5), (4.6), and (4.7), we may skip the three operations, $\tau(T,3)$, $\tau(T,1)$, and $\tau(T,2)$, at the end of

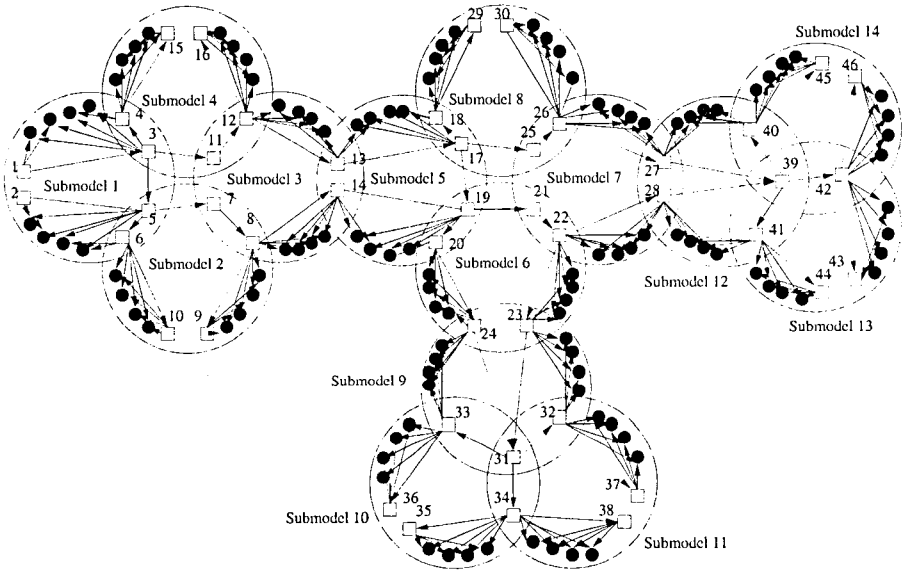



Figure 5.1: Mixed model

the three expressions since the submodels are updated only through the $T$ set and so we may only do $\tau(T, i)$ right before $\epsilon(i)$. By applying the same argument to a series mode arrangement, we can reduce the number of operations during a SEET procedure. For instance, as for panel (1) in Figure 3.1, we may do the SEET and repeat it until convergence following the sequence of SEET procedures as given in

$$\epsilon(1) \rightarrow \tau(1,2) \rightarrow \epsilon(2) \rightarrow \tau(2,3) \rightarrow \epsilon(3) \rightarrow \tau(3,2) \rightarrow \epsilon(2) \rightarrow \tau(2,1)$$

instead of the sequence as given in (4.1), (4.2), and (4.3).

The M-step of the proposed hyper-EM is the same as that of the hyper-EM by Kim (2000).

## 5. Simulation Experiment

We will consider a simulation experiment about a mixed model. This model combines with ring-type models as a series. The detailed structure of the model is the same as Figure 5.1. The model is an large model compared with series-type model and ring-type model. The total number of variables is 158 and 46 variables of these variables are the latent variables and the remainder 112 are the observed variables. From Figure 5.1 we can see that the model is d-split into all 14 submodels with 4 transfusion boxes having structures. All the variables in transfusion boxes of the model are latent variables. All

Table 5.1: Goodness-of-fit levels for mixed model.

| Submodel(d.f.) | 1st(213) | 2nd(212) | 3rd(212) | 4th(212) |
|---|---|---|---|---|
| $\chi^2$ | 262.5 | 191.7 | 257.3 | 209.1 |
| P-value | 0.012 | 0.838 | 0.018 | 0.543 |
| Submodel(d.f.) | 5th(212) | 6th(212) | 7th(212) | 8th(212) |
| $\chi^2$ | 238.3 | 250.1 | 212.5 | 227.3 |
| P-value | 0.104 | 0.037 | 0.477 | 0.224 |
| Submodel(d.f.) | 9th(212) | 10th(214) | 11th(214) | |
| $\chi^2$ | 239.8 | 202.5 | 200.2 | |
| P-value | 0.093 | 0.704 | 0.742 | |
| Submodel(d.f.) | 12th(212) | 13th(214) | 14th(214) | |
| $\chi^2$ | 240.7 | 218.6 | 251.9 | |
| P-value | 0.086 | 0.400 | 0.039 | |

Sample size = 100000, Threshold = 1.0.

the variables involved are binary. After d-split, the graph structures of submodels 3, 5, 7, 9, and 12 are modify in parts for marginal restructuring. Edges, (7,11), (13,14), (21,25), (23,24), and (27,28) in Figure 5.1 are to be added after d-splitting. The goodness-of-fit results are good for the 14 submodels. Estimation of such a large model as in Figure 5.1 is almost impossible if we deal with it as a single model.

## REFERENCES

Asmussen, S. and Edwards, D. (1983). "Collapsibility and response variables in contingency tables," *Biometrika.* **70**, 67-578.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MA.: MIT Press.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *JRSS B*, **39**, 1-38.

Dawid, A. P. (1979). "Conditional independence in statistical theory," *J. R. Statist. Soc. B*, **41**(1), 1-31

Fienberg S. E. and S.-H. Kim, S.-H. (1988). "Calibration and refinement for classification trees," *J. Statist. Plann. Inference*, **70**, 241-254.

Kim, S.-H. (2000). "Hyper-EM for large recursive models of categorical variables," *Computational Statistics & Data Analysis*, **33**(4), 401-424.

Lauritzen, S. L. (1996). *Graphical Models*, Oxford: Clarendon Press.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, England: John Wiley & Sons Ltd.