Marginalization in Bayesian Networks: Integrating Exact and Approximate Inference

Fritz Bayer
19.12.2021
Example Classification of Cancer Subgroups

- Assume data is clustered into groups, e.g. cancer subgroups
- Classification of data against the clusters
Introduction to Bayesian Networks
(Categorical Case)

- Most popular causal model
- Allows graphical interpretation
- Challenges
  - Learning the graph structure (NP-hard)
  - Marginalization (NP-hard)
- Missing data requires marginalization
Introduction to Bayesian Networks

- DAG $\mathcal{G} = (V, E)$ with nodes $V$ and edges $E$
- Nodes $V$ are associated with variables $X_V$ with probability distribution $P(X_V)$
- Factorization (Markov conditions)

$$P(X_V) = \prod_{i \in V} P(X_i \mid X_{pa(i)})$$
Marginalization in Bayesian Networks  
(Categorical Case)

- Let $e \subseteq V$ be evidence nodes, e.g. observed variables
- Marginal probability distribution

\[ P(X_e) = \sum_{X_{V'}} P(X_{V'}, X_e) \]

by summing over $V' = V \setminus e$

⇒ Problem is NP-hard
Example of Highdimensional Bayesian Network

Approximate inference in blue
A node \( i \in V \) in a DAG \( G = (V, E) \) over \( X_V \) is irrelevant w.r.t. a set of nodes \( e \) if \( (\{i\} \cup de(i)) \cap e = \emptyset \).
Reduction of Sampled Variables

Definition (Relevant Subgraph)

The relevant subgraph $G'$ of a DAG $G$ w.r.t. a set of nodes $e$ is the remaining graph after removal of all irrelevant nodes and their edges.
Reduction of Sampled Variables

**Proposition (Marginalization over Relevant Subnetwork)**

Let $G'$ be the relevant subnetwork of a DAG $G$ w.r.t. a set of variables $x_e$ and let $p_{G'}$ and $p_G$ be the respective probability distributions that satisfy the Markov properties. Then $p_{G'}(x_e) = p_G(x_e)$.

![Diagram showing the reduction of sampled variables and subnetworks](image-url)
Definition (Conditionally Independent Subset)

Let $U \subset V$. A set of variables $X = \{X_u : u \in U\}$ is a conditionally independent subset w.r.t. a set of variables $x_e$, if

- all variables in the subset are $d$-connected, i.e. $X_i$ is $d$-connected to $X_j$ w.r.t. $e$ $\forall i, j \in U$, and

- all variables in the subset are $d$-separated from the remaining variables, i.e. $X_i$ is $d$-separated from $X_j$ w.r.t. $e$ $\forall i \in U, j \in V \setminus \{U \cup e\}$. 
Example for Complexity Reduction
In Junction-Tree Algorithm

Get Moral Graph of a DAG: 1. Moralization, 2. Triangulation

- Original DAG
- Moral Graph
- Moral Graph with SGS

→ Five additional edges
→ One additional edge
Proposition (Marginalization in Subsets)

Let $G'$ be the relevant subnetwork of a DAG $G$ w.r.t. a set of nodes $e$. Let $S = \{S_1, ..., S_n\}$ be the conditionally independent subsets of the relevant subnetwork. Then

$$P(X_e) = P(X_{e'}) \prod_{S_i \in S_{\text{exact}}} \sum_{X_{S_i}} P(X_{S_i}, X_{e_{ch}^i} | X_{e_{mb}^i \setminus e_{ch}^i}) \prod_{S_j \in S_{\text{approx}}} \mathbb{E}_{P\left(X_{S_j} \mid X_{e_{mb}^j}\right)} \frac{P\left(X_{e_{ch}^j} | X_{S_j}\right)}{Q\left(X_{S_j}\right)}$$

where $e_{mb}^i = e \cap \{mb(u) : u \in S_i\}$, $e_{ch}^i = e \cap \{ch(u) : u \in S_i\}$ and $e' = e \setminus \{e_{ch}^i \forall i\}$. 

December, 19th 2021 11/19
Example of Highdimensional Bayesian Network
Subgroup Separation

Approximate inference in blue
Benchmark Results
Over Varying Dimensions

- Simulated DAGs
  (100 DAGs, 10 iterations)
- Evidence at random

\[ NRMSE = \sqrt{\frac{\sum_{i=1}^{n} (P(X_e) - \mathbb{E}_i [P(X_e)])^2}{n}} \cdot P(X_e)^{-1} \]
Benchmark Results
Over Varying Dimensions

- Simulated DAGs
  (100 DAGs, 10 iterations)
- Evidence at random

\[ NRMSE = \sqrt{\frac{\sum_{i=1}^{n}(P(X_e) - \mathbb{E}_i[P(X_e)])^2}{n}} \cdot P(X_e)^{-1} \]
Application
Classification of Cancer Subtypes

- Determine the cancer subtype of kidney cancer samples
- Patient samples from Korean population study
- Diagnosed with renal cell carcinoma (RCC)
  - Clear cell RCC (ccRCC)
  - Papillary RCC (pRCC)
Application Results
Classification of Cancer Subtypes

Ratios of correctly assigned cancer type
- 68 % without marginalization
  (cluster 26 genes, classify 26 genes)
- 76 % with marginalization
  (cluster 70 genes, classify 26 genes)
- 83 % with complete data from TCGA
  (cluster 70 genes, classify 70 genes)
Application
Classification of Cancer Subtypes

Patient Samples

Probability of Cancer Subtypes

Cancer Type
- ccRCC
- pRCC
- Difference

FPR
TPR
With Marg. (AUC=0.86)
Without Marg. (AUC=0.83)

December, 19th 2021
ETH Zürich
Standard Inference Methods

Standard approximate inference problem
Find probability of a single variable
\[ P(X_i | X_e) \]

Marginal probability distribution
Find probability of multiple variables
\[ P(X_1, \ldots, X_n | X_e) \text{ (or } P(X_e) \text{)} \]

Not easy to unify because
\[ P(X_1, \ldots, X_n | X_e) \neq \prod_i P(X_i | X_e, X_{pa(i)}) \]
Conclusion

• Marginalization in Bayesian networks
  – Present efficient method
  – Allows to handle missing data
  – R package SubGroupSeparation

• Separation to subgroups can be generalized
to other approximate inference schemes
Thank you for your attention!

Preprint: https://arxiv.org/pdf/2112.09217.pdf
Code: https://github.com/cbg-ethz/SubGroupSeparation

Fritz Bayer      Niko Beerenwinkel
frbayer@ethz.ch  Giusi Moffa        Jack Kuipers