Algebra in probabilistic reasoning

Tobias Boege

$$\begin{bmatrix}\text{Department of Mathematics} \\ \text{KTH Royal Institute of Technology}\end{bmatrix} \mapsto \begin{bmatrix}\text{Department of Mathematics and Statistics} \\ \text{UiT The Arctic University of Norway}\end{bmatrix}$$

AIToGeLiS Day
MPI-CBG Dresden, 03 May 2024

Probabilistic reasoning

- Probabilistic reasoning deals with the representation, updating and processing of uncertain beliefs about a system of objects.
- Think: statistical models inferred from observational or interventional data.

Probabilistic reasoning

- Probabilistic reasoning deals with the representation, updating and processing of uncertain beliefs about a system of objects.
- Think: statistical models inferred from observational or interventional data.
- But also think: geometric reasoning. Objects with uncertain “positions” but certain “relations” with each other.

Probabilistic reasoning

- Probabilistic reasoning deals with the representation, updating and processing of uncertain beliefs about a system of objects.
- Think: statistical models inferred from observational or interventional data.
- But also think: geometric reasoning. Objects with uncertain "positions" but certain "relations" with each other.

In this talk: **independence relations**.

Probabilistic reasoning

- Probabilistic reasoning deals with the representation, updating and processing of uncertain beliefs about a system of objects.
- Think: statistical models inferred from observational or interventional data.
- But also think: geometric reasoning. Objects with uncertain "positions" but certain "relations" with each other.

In this talk: **independence relations**.

- Fundamental qualitative information about the system.

Probabilistic reasoning

- Probabilistic reasoning deals with the representation, updating and processing of uncertain beliefs about a system of objects.
- Think: statistical models inferred from observational or interventional data.
- But also think: geometric reasoning. Objects with uncertain “positions” but certain “relations” with each other.

In this talk: **independence relations**.

- Fundamental qualitative information about the system.
- Knowledge of independence allows more compact representation and more efficient processing.

Probabilistic reasoning

- Probabilistic reasoning deals with the representation, updating and processing of uncertain beliefs about a system of objects.
- Think: statistical models inferred from observational or interventional data.
- But also think: geometric reasoning. Objects with uncertain “positions” but certain “relations” with each other.

In this talk: **independence relations**.

- Fundamental qualitative information about the system.
- Knowledge of independence allows more compact representation and more efficient processing.
- Common assumption in geometry, statistical modeling, cryptography …

A geometric example

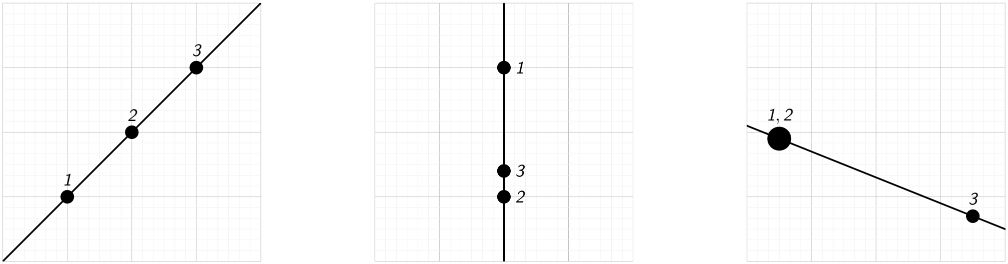
- Consider 3 points in $\mathbb{R}^2$ which lie on a line:

$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \text{ such that } |A| = 0.$$

A geometric example

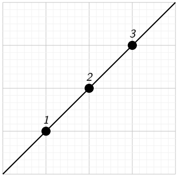
- Consider 3 points in $\mathbb{R}^2$ which lie on a line:

$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{such that } |A| = 0.$$



A geometric example

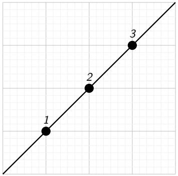
$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \text{ such that } |A| = 0.$$



- Defines a variety V. I think of a configuration $A \in V$.

A geometric example

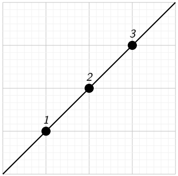
$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \text{ such that } |A| = 0.$$



- Defines a variety V. I think of a configuration $A \in V$.
- For all you know, the point $p_1 = (x_1, y_1)$ could be anywhere in $\mathbb{R}^2$.

A geometric example

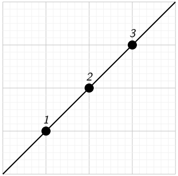
$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \text{ such that } |A| = 0.$$



- Defines a variety V. I think of a configuration $A \in V$.
- For all you know, the point $p_1 = (x_1, y_1)$ could be anywhere in $\mathbb{R}^2$.
- But if I reveal p_2 and p_3, then your uncertainty about p_1 is reduced to an $\mathbb{R}^1$!

A geometric example

$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \text{ such that } |A| = 0.$$

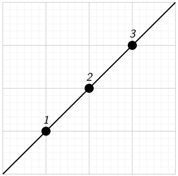


- Defines a variety V. I think of a configuration $A \in V$.
- For all you know, the point $p_1 = (x_1, y_1)$ could be anywhere in $\mathbb{R}^2$.
- But if I reveal p_2 and p_3, then your uncertainty about p_1 is reduced to an $\mathbb{R}^1$!

Functional dependence

A geometric example

$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \text{ such that } |A| = 0.$$



- Defines a variety V. I think of a configuration $A \in V$.
- For all you know, the point $p_1 = (x_1, y_1)$ could be anywhere in $\mathbb{R}^2$.
- But if I reveal p_2 and p_3, then your uncertainty about p_1 is reduced to an $\mathbb{R}^1$!

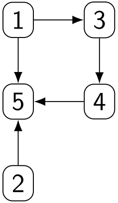
Functional dependence

- In statistics, graphical models are a direct analogue of this.

A statistical example

- A linear structural equation model defines random variables X recursively via a directed acyclic graph $G = (V, E)$ and Gaussian noise:

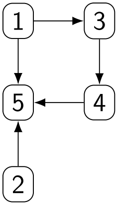
$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$



A statistical example

- A linear structural equation model defines random variables X recursively via a directed acyclic graph $G = (V, E)$ and Gaussian noise:

$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$

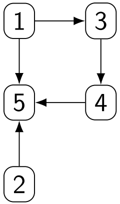


- The vector X is again Gaussian with mean zero. Since G is acyclic, we can solve for the covariance matrix $\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} \to$ model* $\mathcal{M}(G)$.

A statistical example

- A linear structural equation model defines random variables X recursively via a directed acyclic graph $G = (V, E)$ and Gaussian noise:

$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$

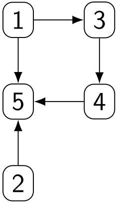


- The vector X is again Gaussian with mean zero. Since G is acyclic, we can solve for the covariance matrix $\Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1} \to$ model* $\mathcal{M}(G)$.
- If $X_{\mathrm{pa}(j)}$ are observed, then X_j is independent of its non-descendants.

A statistical example

- A linear structural equation model defines random variables X recursively via a directed acyclic graph $G = (V, E)$ and Gaussian noise:

$$X_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \omega_j).$$



- The vector X is again Gaussian with mean zero. Since G is acyclic, we can solve for the covariance matrix $\Sigma = (I - \Lambda)^{-\mathsf{T}} \Omega (I - \Lambda)^{-1} \rightarrow$ model* $\mathcal{M}(G)$.
- If $X_{\mathrm{pa}(j)}$ are observed, then X_j is independent of its non-descendants.

Conditional independence

Conditional independence $[X \perp\!\!\!\perp Y \mid Z]$

When does knowing Z make X irrelevant for Y?

When does knowing Z make X irrelevant for Y?

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

When does knowing Z make X irrelevant for Y?

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

- $[c_1 \perp\!\!\!\perp c_2]$

When does knowing Z make X irrelevant for Y?

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

- $[c_1 \perp\!\!\!\perp c_2]$
- $[c_1 \not\!\perp\!\!\!\perp c_2 \mid b]$ ...

Conditional independence $[X \perp\!\!\!\perp Y \mid Z]$

When does knowing Z make X irrelevant for Y?

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

- $[c_1 \perp\!\!\!\perp c_2]$
- $[c_1 \not\perp\!\!\!\perp c_2 \mid b]$ ...

Laws of probabilistic reasoning

Let $X_1, \dots, X_n$ be jointly distributed random variables. Assume that $X_i \perp\!\!\!\perp X_j \mid X_K$ for some choices of $i, j \in [n]$ and $K \subseteq [n] \setminus \{i, j\}$. Which other CI statements $X_r \perp\!\!\!\perp X_s \mid X_T$ also hold?

Gaussian conditional independence

Assume $X = (X_i : i \in N)$ are jointly Gaussian with covariance matrix $\Sigma \in \mathrm{PD}_N$.

Definition

The polynomial $\Sigma[K] := |\Sigma_{K,K}|$ is a principal minor of Σ and $\Sigma[ij \mid K] := |\Sigma_{iK,jK}|$ is an almost-principal minor.

Gaussian conditional independence

Assume $X = (X_i : i \in N)$ are jointly Gaussian with covariance matrix $\Sigma \in \mathrm{PD}_N$.

Definition

The polynomial $\Sigma[K] := |\Sigma_{K,K}|$ is a principal minor of Σ and $\Sigma[ij \mid K] := |\Sigma_{iK,jK}|$ is an almost-principal minor.

Algebraic statistics proves:

- Σ is PD if and only if $\Sigma[K] > 0$ for all $K \subseteq N$.
- $[i \perp\!\!\!\perp j \mid K]$ holds if and only if $\Sigma[ij \mid K] = 0$.
- $\mathbb{E}[X] = \mu$ is irrelevant.

Gaussian CI models

Definition

A CI constraint is a CI statement $[i \perp\!\!\!\perp j \mid K]$ or its negation $[i \not\perp\!\!\!\perp j \mid K]$.
The model of a set of CI constraints is the set of all PD matrices which satisfy them.

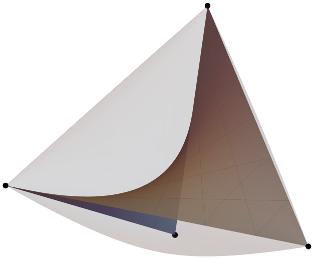


Figure: Model of $\Sigma[12 \mid 3] = a - bc = 0$ in the space of 3×3 correlation matrices.

Models and implication

Implication problem for Gaussian conditional independence

Given a clause $\bigwedge \mathcal{P} \implies \bigvee \mathcal{Q}$, where $\mathcal{P}$ and $\mathcal{Q}$ are sets of CI statements over N, decide if it is valid for all N-variate Gaussians.

$$\bigwedge \mathcal{P} \implies \bigvee \mathcal{Q}$$

Models and implication

Implication problem for Gaussian conditional independence

Given a clause $\bigwedge \mathcal{P} \implies \bigvee \mathcal{Q}$, where $\mathcal{P}$ and $\mathcal{Q}$ are sets of CI statements over N, decide if it is valid for all N-variate Gaussians.

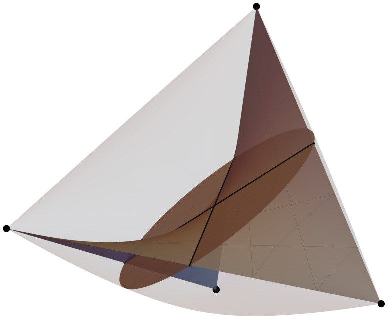
$$\underset{\text{is not valid}}{\bigwedge \mathcal{P} \implies \bigvee \mathcal{Q}} \quad \iff \quad \underset{\text{has a point}}{\mathcal{M}(\mathcal{P} \cup \neg \mathcal{Q})}$$

Example of CI implication

$$\Sigma = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}$$

- If $\Sigma[12\,|\,] = a$ and $\Sigma[12\,|\,3] = a - bc$ vanish, then $bc = \Sigma[13\,|\,] \cdot \Sigma[23\,|\,]$ must vanish:

$$[12\,|\,] \wedge [12\,|\,3] \implies [13\,|\,] \vee [23\,|\,].$$



Normal form for proofs and refutations

Let $f_i \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Tarski's transfer principle)

If a polynomial system $\{f_i \bowtie_i 0\}$, $\bowtie_i \in \{=, \neq, <, \leq, \geq, >\}$, has a solution over $\mathbb{R}$, then it has a solution in a finite real extension of $\mathbb{Q}$.

Normal form for proofs and refutations

Let $f_i \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Tarski's transfer principle)

If a polynomial system $\{f_i \bowtie_i 0\}$, $\bowtie_i \in \{=, \neq, <, \leq, \geq, >\}$, has a solution over $\mathbb{R}$, then it has a solution in a finite real extension of $\mathbb{Q}$.

$\rightarrow$ If $\bigwedge \mathcal{P} \implies \bigvee \mathcal{Q}$ is false, there is a counterexample matrix Σ with $\overline{\mathbb{Q}}$ entries.

$[12 \,|\,] \wedge [12 \,|\, 3] \implies [13 \,|\,]$ is false and a counterexample is

$$\begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1 \end{pmatrix}.$$

Normal form for proofs and refutations

Let $f_i, g_j, h_k \in \mathbb{Z}[t_1, \dots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Positivstellensatz)

A polynomial system $\{f_i = 0, g_j \geq 0, h_k \neq 0\}$ *is infeasible if and only if there exist* $f \in \text{ideal}(f_i)$, $g \in \text{cone}(g_j)$ *and* $h \in \text{monoid}(h_k)$ *such that* $g + h^2 = f$.

Normal form for proofs and refutations

Let $f_i, g_j, h_k \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Positivstellensatz)

A polynomial system $\{f_i = 0, g_j \geq 0, h_k \neq 0\}$ *is infeasible if and only if there exist* $f \in \mathrm{ideal}(f_i)$, $g \in \mathrm{cone}(g_j)$ *and* $h \in \mathrm{monoid}(h_k)$ *such that* $g + h^2 = f$.

→ If $\bigwedge \mathcal{P} \implies \bigvee \mathcal{Q}$ is true, there exists an algebraic proof for it with $\mathbb{Z}$ coefficients.

$[12\,|\,] \wedge [12\,|\,3] \implies [13\,|\,] \vee [23\,|\,]$ is true and a proof is the final polynomial

$$\Sigma[13\,|\,] \cdot \Sigma[23\,|\,] = \Sigma[3] \cdot \Sigma[12\,|\,] - \Sigma[12\,|\,3].$$

A 5×5 final polynomial

The following implication is valid for all positive-definite 5×5 matrices:

$$[12 \mid\,] \wedge [14 \mid 5] \wedge [23 \mid 5] \wedge [35 \mid 1] \wedge [45 \mid 2] \wedge [15 \mid 23] \wedge [34 \mid 12] \wedge [24 \mid 135] \implies [25 \mid\,] \vee [34 \mid\,].$$

A 5 × 5 final polynomial

The following implication is valid for all positive-definite 5×5 matrices:

$$[12 \,|\,] \wedge [14 \,|\, 5] \wedge [23 \,|\, 5] \wedge [35 \,|\, 1] \wedge [45 \,|\, 2] \wedge [15 \,|\, 23] \wedge [34 \,|\, 12] \wedge [24 \,|\, 135] \implies [25 \,|\,] \vee [34 \,|\,].$$

$$[25 \,|\,][34 \,|\,] \cdot [1][2][3][15] =$$
$$\Big(cd^2egr + bd^2fgr - ad^2grh - 2cd^2e^2i - 2bd^2efi - 2pdfgri + 2ad^2ehi + 2pdefi^2 - 2pdqhi^2 + 2pcqi^3 +$$
$$2pdqrij - 2pbqi^2j - pcegrt + pbfgrt + pagrht + 2pce^2it - 2pcqrit + 2pbqhit - 2paehit\Big) \cdot [12 \,|\,] +$$
$$\Big(pdqer + pbqgr - 2pbqei\Big) \cdot [14 \,|\, 5] - \Big(pcdqr + p^2fgr - 2pbcqi + 2pb^2qj - 2p^2qrj\Big) \cdot [23 \,|\, 5] +$$
$$\Big(cdqgr - 2cdqei + 2pqghi - 2pqfi^2 - pqgrj + 2pqeij - 2pe^2ft + 2pqfrt\Big) \cdot [35 \,|\, 1] +$$
$$\Big(pd^2er - 2pbdei + p^2gri + 2pb^2et - 2p^2ert\Big) \cdot [45 \,|\, 2] - \Big(2pdfi - 2pbft\Big) \cdot [15 \,|\, 23] -$$
$$\Big(d^2gr - 2d^2ei - pgrt + 2peit\Big) \cdot [34 \,|\, 12] - 2pqi \cdot [24 \,|\, 135].$$

A 5 × 5 final polynomial

```
R = QQ[p,a,b,c,d, q,e,f,g, r,h,i, s,j, t];
X = genericSymmetricMatrix(R,p,5);
I = ideal(
  det X_{0}^{1}, det X_{0,3}^{2,3}, det X_{0,4}^{3,4},
  det X_{1,4}^{2,4}, det X_{2,0}^{4,0}, det X_{3,1}^{4,1},
  det X_{0,1,2}^{4,1,2}, det X_{2,0,1}^{3,0,1},
  det X_{1,0,2,4}^{3,0,2,4}
);
U = g*h*p*q*r*(p*t-d^2); -- [25 | ][34 | ] · [1][2][3][15] ∈ monoid(𝒱)
U % I --> 0, meaning monoid(𝒱) ∩ ideal(𝒱) ≠ ∅ in ℚ[X]
-- Get a proof that U is in I:
G = gens I; -- the equations generating ideal(𝒱)
H = U // G; -- linear combinators for U from G
U == G*H --> true
```

General proofs and refutations

Theorem (Tarski's transfer principle)

If an implication is wrong, there exists a counterexample to it with real algebraic probabilities.

Theorem (Positivstellensatz)

If an implication is correct, there exists a proof of it in the form of a single polynomial identity with integer coefficients.

General proofs and refutations

Theorem (Tarski's transfer principle)

If an implication is wrong, there exists a counterexample to it with real algebraic probabilities.

Theorem (Positivstellensatz)

If an implication is correct, there exists a proof of it in the form of a single polynomial identity with integer coefficients.

- These geometric theorems apply to probabilistic reasoning!

General proofs and refutations

Theorem (Tarski's transfer principle)

If an implication is wrong, there exists a counterexample to it with real algebraic probabilities.

Theorem (Positivstellensatz)

If an implication is correct, there exists a proof of it in the form of a single polynomial identity with integer coefficients.

- These geometric theorems apply to probabilistic reasoning!
- They give theoretical guarantees and exact certificates.

General proofs and refutations

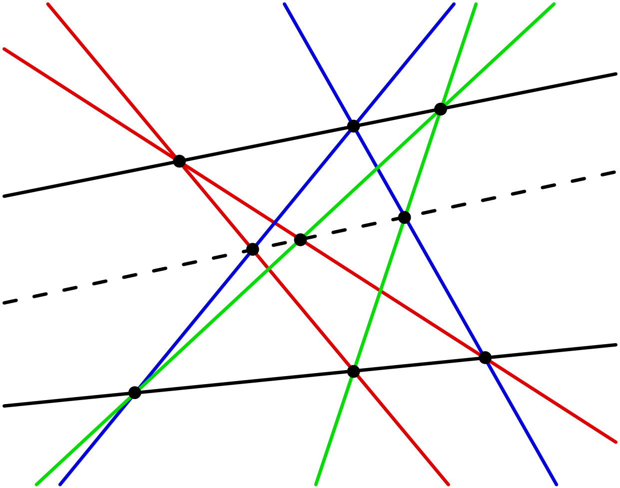
Theorem (Tarski's transfer principle)

If an implication is wrong, there exists a counterexample to it with real algebraic probabilities.

Theorem (Positivstellensatz)

If an implication is correct, there exists a proof of it in the form of a single polynomial identity with integer coefficients.

- These geometric theorems apply to probabilistic reasoning!
- They give theoretical guarantees and exact certificates.
- In practice, few things work symbolically. Require robust numerical non-linear algebra tools like HomotopyContinuation.jl to experiment and form conjectures.



Thank you for your attention!

Problem 1: Gaussian CI implication

Let Σ be the covariance matrix of a regular Gaussian distribution. (Thus Σ is strictly positive definite!) Then $[i \perp\!\!\!\perp j \mid K]$ holds if and only if $|\Sigma_{iK,jK}| = 0$.

(a) For a three Gaussian random variables $1, 2, 3$ show that

$$[1 \perp\!\!\!\perp 2 \mid 3] \wedge [1 \perp\!\!\!\perp 3 \mid 2] \implies [1 \perp\!\!\!\perp 2] \wedge [1 \perp\!\!\!\perp 3].$$

(b) For four Gaussian random variables $1, 2, 3, 4$ show that

$$[1 \perp\!\!\!\perp 3] \wedge [1 \perp\!\!\!\perp 4] \wedge [1 \perp\!\!\!\perp 4 \mid 2, 3] \wedge [2 \perp\!\!\!\perp 3 \mid 1, 4] \implies [1 \perp\!\!\!\perp 4].$$

(Hint: Primary decomposition.)

Problem 2: Graphical models

The Gaussian graphical model $\mathcal{M}_G$ of a directed acyclic graph $G = (V, E)$ consists of all positive definite $V \times V$ matrices Σ which satisfy

$$[i \perp\!\!\!\perp j \mid \mathrm{pa}(j)] \quad \text{for all } i < j \text{ such that } i \to j \notin E.$$

Here $<$ is a topological ordering on G and pa denotes the parent set.

(a) Show that the two DAGs $1 \to 2 \to 3$ and $1 \leftarrow 2 \leftarrow 3$ define the same model. What is its dimension? Which dimension did you expect?

(b) For any directed acyclic graph G show that if $i \to j$ is an edge, then $[i \perp\!\!\!\perp j \mid \mathrm{pa}(j)]$ does not hold for a generic $\Sigma \in \mathcal{M}_G$.

(c) What do you think is the right Bayesian network to represent the causal relationships between "Summer", "Rain barrel is full", "Ground is wet", "It rained", "Sprinkler was on" and "Umbrella is wet"? Compare your models.