

CS262a Bayesian Networks

© Tingfeng Xia @ UCLA

Winter Quarter, 2022

This work is licensed under a [Creative Commons “Attribution-NonCommercial-ShareAlike 4.0 International”](#) license.



Preface

This document is intentionally kept short, and is not a complete summary of what is covered in this course. Excerpt from online:

The objective of this class is to provide an in-depth exposition of knowledge representation, reasoning, and machine learning under uncertainty using the framework of Bayesian networks. Both theoretical underpinnings and practical considerations will be covered, with a special emphasis on constructing and learning graphical models, and on various exact and approximate inference algorithms. Additional topics include logical approaches to probabilistic inference, compilation techniques, sensitivity analysis, undirected graphical models, and statistical relational learning.

Instructor: Professor Adnan Darwiche

Book: Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press 2009.

Contents

1	Propositional Logic	4
1.1	Principle Logical Forms	4
1.2	Equivalent Forms	4
1.3	Instantiation Agreement	4
1.4	Information Theory	4

2	Probability Calculus	5
2.1	Bayesian Conditioning	5
2.2	Independence and Notations	5
3	Bayesian Networks	6
3.1	Soft Evidence	6
3.1.1	All Things Considered Method	6
3.1.2	Nothing-else Considered Method	6
3.2	Noisy Sensors	7
3.3	Markov Assumptions	7
3.4	Graphoid Axioms	7
3.5	Positive Graphoid Axioms	7
3.6	D-seperation Linear Prune Theorem	8
3.7	D-seperation Properties	8
4	Inference by Variable Elimination	8
4.1	Interaction Graphs	8
4.2	Elimination Order	8
4.2.1	Minimum Degree Ordering	8
4.2.2	Minimum Fill Ordering	8
5	Inference by Factor Elimination	8
5.1	Factor Elimination	8
5.2	Elimination Trees	9
5.3	Message Passing Formulation	9
6	Inference by Conditioning	9
6.1	Run time Comparison	9
6.1.1	Variable Elimination (VE)	9
6.1.2	Message Passing	10
6.1.3	Polytree / Belief Propagation	10
6.1.4	Cut Set Conditioning	10
6.1.5	Any Space Recursive Cut Set	11
7	Compiling Bayesian Networks	11
7.1	Network Polynomials	11
7.2	AC Properties	11
7.3	AC Derivative Probabilistic Implications	12
7.4	Compilation via Variable Elimination	12

8	Causality - Interventions	12
8.1	Notations	12
8.2	Types of Causal Graphs	12
8.2.1	Markovian Model	13
8.2.2	Semi-Markovian Model	13
8.3	Identifiability Criterion	13
8.3.1	Causal Effect Rule	13
8.3.2	Backdoor Criteria	13
8.3.3	Frontdoor Criteria	14
8.3.4	Exogenous X	14
8.4	The Do-Calculus	14
8.4.1	Graph Alterations	15
9	Counterfactual Reasoning	15
9.1	The Information Hierarchy	15
9.2	Counterfactual Queries	15
9.2.1	Probability of Necessity (PN)	15
9.2.2	Probability of Sufficiency (PS)	15
9.2.3	Probability of Necessity and Sufficiency (PNS)	16
9.2.4	Probability of Disablement	16
9.2.5	Probability of Enablement	16
9.3	Structural Causal Models (SCM)	16
9.4	Evenets	16
9.4.1	Observational Event	16
9.4.2	Interventional Event	16
9.4.3	Counterfactual Event	17
9.5	Satisfaction	17
9.5.1	Observational Event	17
9.5.2	Interventional Event	17
9.5.3	Counterfactual Event	17
9.6	Queries	17
9.7	Exogeneity and Monotonicity	18

1 Propositional Logic

1.1 Principle Logical Forms

Inconsistent. Something that never holds; $\text{Mods}(\cdot) = \emptyset$; $\Pr(\alpha) = 0$

Valid. Something that always holds; $\text{Mods}(\cdot) = \Omega$; $\Pr(\alpha) = 1$

Equivalent. $\text{Mods}(\alpha) = \text{Mods}(\beta)$

Mutual Exclusive. $\text{Mods}(\alpha) \cap \text{Mods}(\beta) = \emptyset$

Exhaustive. $\text{Mods}(\alpha) \cup \text{Mods}(\beta) = \Omega$

Entailment / Implication. $\alpha \models \beta \triangleq \text{Mods}(\alpha) \subseteq \text{Mods}(\beta)$

1.2 Equivalent Forms

- $\text{Mods}(\alpha \wedge \beta) = \text{Mods}(\alpha) \cap \text{Mods}(\beta)$
- $\text{Mods}(\alpha \vee \beta) = \text{Mods}(\alpha) \cup \text{Mods}(\beta)$
- $\text{Mods}(\neg \alpha) = \overline{\text{Mods}(\alpha)}$

1.3 Instantiation Agreement

Two instantiation, each of which can cover a subset of different variables, are said to be compatible with each other if they agree on all common variables. Denoted as $\mathbf{x} \sim \mathbf{y}$.

1.4 Information Theory

Entropy.

$$\text{ENT}(X) = - \sum_x \Pr(x) \log \Pr(x) \quad (1.1)$$

where $0 \log 0 = 0$ by convention. With a higher entropy, we say that it is more chaotic.

Conditional Entropy.

$$\text{ENT}(X|Y) = \sum_y \Pr(y) \text{ENT}(X|y) \quad \text{where} \quad \text{ENT}(X|y) = - \sum_x \Pr(x|y) \log \Pr(x|y) \quad (1.2)$$

Conditioning never increases the entropy, i.e.

$$\text{ENT}(X|Y) \leq \text{ENT}(X) \quad (1.3)$$

Mutual Information

$$MI(X; Y) = \sum_{x,y} \Pr(x, y) \log \frac{\Pr(x, y)}{\Pr(x)\Pr(y)} \quad (1.4)$$

$$= ENT(X) - ENT(X|Y) \quad (1.5)$$

$$= ENT(Y) - ENT(Y|X) \quad (1.6)$$

Conditional Mutual Information

$$MI(X; Y|Z) = \sum_{x,y,z} \Pr(x, y, z) \log \frac{\Pr(x, y|z)}{\Pr(x|z)\Pr(y|z)} \quad (1.7)$$

$$= ENT(X|Z) - ENT(X|Y, Z) \quad (1.8)$$

$$= ENT(Y|Z) - ENT(Y|X, Z) \quad (1.9)$$

2 Probability Calculus

2.1 Bayesian Conditioning

Bayesian Condition is specified by the formula

$$\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)} \quad (2.1)$$

In particular, not to be confused with Bayesian inference (to be added later).

2.2 Independence and Notations

Independence.

$$\alpha \perp\!\!\!\perp \beta \iff \Pr(\alpha|\beta) = \Pr(\alpha) \vee \Pr(\beta) = 0 \quad (2.2)$$

$$\iff \Pr(\alpha \wedge \beta) = \Pr(\alpha)\Pr(\beta) \quad (2.3)$$

Conditional Independence.

$$(\alpha \perp\!\!\!\perp \beta)|\gamma \iff \Pr(\alpha|\beta \wedge \gamma) = \Pr(\alpha|\gamma) \vee \Pr(\beta \wedge \gamma) = 0 \quad (2.4)$$

$$\iff \Pr(\alpha \wedge \beta|\gamma) = \Pr(\alpha|\gamma)\Pr(\beta|\gamma) \vee \Pr(\gamma) = 1 \quad (2.5)$$

Set Independence

$$I_{Pr}(X, Z, Y) \iff (x \perp\!\!\!\perp y)|z, \quad \forall x, y, z \in X, Y, Z \quad (2.6)$$

3 Bayesian Networks

3.1 Soft Evidence

3.1.1 All Things Considered Method

We normalize / rescale w according to new evidence.

$$\Pr'(w) = \begin{cases} \frac{\Pr'(\beta)}{\Pr(\beta)} \Pr(w) & \text{if } w \models \beta \\ \frac{\Pr'(\neg\beta)}{\Pr(\neg\beta)} \Pr(w) & \text{if } w \models \neg\beta \end{cases} \quad (3.1)$$

The closed form is called the Jefferey's Rule.

Jeffery's Rule

$$\Pr'(\alpha) = q\Pr(\alpha|\beta) + (1 - q)\Pr(\alpha|\neg\beta) \quad (3.2)$$

Jeffery's Rule - General Case

$$\Pr'(\alpha) = \sum_{i=1}^n \Pr'(\beta_i) \Pr(\alpha|\beta_i) \quad (3.3)$$

3.1.2 Nothing-else Considered Method

Odds.

$$O(\beta) = \frac{\Pr(\beta)}{\Pr(\neg\beta)} \quad (3.4)$$

Bayes Factor

$$k = \frac{O'(\beta)}{O(\beta)} = \frac{\Pr'(\beta)/\Pr'(\neg\beta)}{\dots} \quad (3.5)$$

from where we can expand and organize

$$\Pr'(\beta) = \frac{k\Pr(\beta)}{k\Pr(\beta) + \Pr(\neg\beta)} \quad (3.6)$$

Closed Form Solution.

$$\Pr'(\alpha) = \frac{k\Pr(\alpha \wedge \beta) + \Pr(\alpha \wedge \neg\beta)}{k\Pr(\beta) + \Pr(\neg\beta)} \quad (3.7)$$

3.2 Noisy Sensors

$$O'(\beta) = \underbrace{\frac{1-f_n}{f_p}}_{k^+} O(\beta) \quad O'(\beta) = \underbrace{\frac{f_n}{1-f_p}}_{k^-} O(\beta) \quad (3.8)$$

3.3 Markov Assumptions

$$\text{Markov}(G) = \{I_{\text{Pr}}(V, \text{Parents}(V), \text{ND}(V))\}_V \quad (3.9)$$

where ND means non-descendants, and includes all nodes except for $V, \text{Parents}(V)$ and $\text{Descendants}(V)$ (all the way till leaf)

3.4 Graphoid Axioms

Symmetry.

$$I_{\text{Pr}}(X, Z, Y) \iff I_{\text{Pr}}(Y, Z, X) \quad (3.10)$$

Decomposition.

$$I_{\text{Pr}}(X, Z, Y \cup W) \implies I_{\text{Pr}}(X, Z, Y) \wedge I_{\text{Pr}}(X, Z, W) \quad (3.11)$$

Weak Union.

$$I_{\text{Pr}}(X, Z, Y \cup W) \implies I_{\text{Pr}}(X, Z \cup Y, W) \quad (3.12)$$

Contraction.

$$I_{\text{Pr}}(X, Z, Y) \wedge I_{\text{Pr}}(X, Z \cup Y, W) \implies I_{\text{Pr}}(X, Z, Y \cup W) \quad (3.13)$$

Triviality.

$$I_{\text{Pr}}(X, Z, \emptyset) \quad (3.14)$$

3.5 Positive Graphoid Axioms

... includes everything from Graphoid Axioms (Section 3.4) and in addition has

Intersection.

$$I_{\text{Pr}}(X, Z \cup W, Y) \wedge I_{\text{Pr}}(X, Z \cup Y, W) \implies I_{\text{Pr}}(X, Z, Y \cup W) \quad (3.15)$$

3.6 D-seperation Linear Prune Theorem

3.7 D-seperation Properties

Soundness.

$$\text{dsep}_G(X, Z, Y) \implies I_{\text{Pr}}(X, Z, Y) \quad (3.16)$$

(Weak) Completeness. There exists a parametrization Θ that for every DAG G such that

$$I_{\text{Pr}}(X, Z, Y) \iff \text{dsep}_G(X, Z, Y) \quad (3.17)$$

4 Inference by Variable Elimination

4.1 Interaction Graphs

add to here

4.2 Elimination Order

Choosing an optimal elimination order in VE is NP-hard. We resort to heuristics.

4.2.1 Minimum Degree Ordering

When choosing node to eliminate, choose the one that has the smallest number of neighbours in the interaction graph.

4.2.2 Minimum Fill Ordering

5 Inference by Factor Elimination

5.1 Factor Elimination

The elimination of factor f_i from a set of factors S is a two step process. We first eliminate all variables V that appear only in factor f_i and then multiply the result $\sum_V f_i$ by some other factor $f_j \in S$.

Projection. Factor projection operation is defined as

$$\text{project}(f, Q) \triangleq \sum_{\text{vars}(f) \setminus Q} f \quad (5.1)$$

which is essentially summing out all variables not in Q . In the vanilla factor elimination algorithm (FE1), the projection is used in the return statement. This is essentially a normalization step (onto variables specified only).

5.2 Elimination Trees

Variables. $vars(i)$ denotes the variables mentioned at node i . $vars(i, j)$ denotes all variables mentioned in nodes to the i -side of the graph (inclusive). Hence, it holds that $vars(i) \subseteq vars(i, j)$.

Separators.

$$S_{ij} \triangleq vars(i, j) \cap vars(j, i) \quad (5.2)$$

Clusters.

$$C_i \triangleq vars(i) \cup \bigcup_j S_{ij} \quad (5.3)$$

5.3 Message Passing Formulation

The Message that i sends to j is

$$M_{ij} \triangleq \text{project} \left(\phi_i \prod_{k \neq j} M_{ki}, S_{ij} \right) \quad (5.4)$$

where the product is off all the messages from edges except for the output direction and S_{ij} is the separator defined in Section 5.2.

6 Inference by Conditioning

6.1 Run time Comparison

6.1.1 Variable Elimination (VE)

Let w be the width of the tree, n denote the number of variables, and $|Q|$ as query variable size.

Time Complexity. $O(n \exp(w))$

Space Complexity.

$$O(n \exp(w) + n \exp(|Q|)) \equiv O(n \exp(w)), \quad \text{if } |Q| < \infty \quad (6.1)$$

6.1.2 Message Passing

One Specific Message. The cost to pass one specific message is

$$O(\exp(|C_i|)) \quad (6.2)$$

where $|C_i|$ is the cluster size.

Every Message. Since cluster sizes are bounded above by tree width, we can say that passing of every message will have an upper-bound runtime of

$$O(\exp(w)) \quad (6.3)$$

where w is the width of the elimination tree.

Amount of Messages. The total amount of messages is

$$O(2(m-1)) \quad \text{where } m = |V| \quad (6.4)$$

since we have a tree structure, meaning we have exactly $(m-1)$ edges and each edge can have forward backward each once.

All Messages - All Cluster Marginals. The total time to pass all messages and compute all cluster marginals is

$$O(m \exp(w)) \quad \text{or} \quad O(n \exp(w)), \text{ for } O(n) \text{ edges.} \quad (6.5)$$

6.1.3 Polytree / Belief Propagation

Runtime. Define k as the max number of parents in the poly tree, then k is the same as the width of elimination tree. Let, also, n denote the number of nodes in the polytree. The algorithm has runtime

$$O(n \exp(k)) \quad (6.6)$$

6.1.4 Cut Set Conditioning

Time and Space (Total) Complexity.

$$O(n \exp(k)) \quad \text{where } n = |N| \text{ and } k \text{ is width} \quad (6.7)$$

6.1.5 Any Space Recursive Cut Set

	no cache	all cache	$\Delta no \rightarrow all$
space	$O(wn)$	$O(n \exp(w))$	\uparrow
time	$O(n \exp(w \log n))$	$O(n \exp(w))$	\downarrow

7 Compiling Bayesian Networks

7.1 Network Polynomials

The network polynomial is a summation over all instantiations of a network,

$$f \triangleq \sum_z \prod_{\theta_{x|u} \sim z} \theta_{x|u} \prod_{\lambda_x \sim z} \lambda_x \quad (7.1)$$

7.2 AC Properties

AC Size. of an AC is defined as the number of edges in the circuit.

AC Complexity. is the size of smallest AC that represents the network polynomial.

Decomposable. At each \star node, we need

$$\text{vars}(AC_A) \cap \text{vars}(AC_B) = \emptyset \quad (7.2)$$

Deterministic. At each $+$ node, we require at most one positive input is non-zero for all *complete instantiation*.

Smooth. At each $+$ node, we require

$$\text{vars}(AC_A) = \text{vars}(AC_B) \quad (7.3)$$

AC for Marginals. requires decomposable and smooth. This guarantees that sub-circuits are of complete variable instantiations.

AC for Marginals and MPE. requires all three above: decomposable, deterministic, and smooth. The additional determinism guarantees a 1-to-1 mapping between sub-circuits and complete variable instantiations.

7.3 AC Derivative Probabilistic Implications

$$\frac{\partial f}{\partial \lambda_x}(\mathbf{e}) = \Pr(\mathbf{x}, \mathbf{e} - X) \quad (7.4)$$

and

$$\theta_{x|u} \frac{\partial f}{\partial \theta_{x|u}}(\mathbf{e}) = \Pr(\mathbf{x}, \mathbf{u}, \mathbf{e}) \quad (7.5)$$

7.4 Compilation via Variable Elimination

Circuit Factors. “In a circuit factor, each variable instantiation is mapped to a circuit node instead of a number.”

Operations. We use $+(n_1, n_2)$ to denote an addition node that has n_1 and n_2 as its children. Similarly, $\star(n_1, n_2)$ denotes a multiplication node. An operation (multiplication or addition) of two circuit factors $f(X)$ and $f(Y)$ is a factor over variables $Z = X \cup Y$,

$$f(z) = [\star \text{ or } +](f(x), f(y)), \quad \text{where } x \sim z \quad \text{and} \quad y \sim z \quad (7.6)$$

Procedure.

1. **Made nodes for each CPT.** For each family $X|U$, construct nodes $\star(\lambda_x, \theta_{x|u})$ for each instantiation xu of XU .
2. **Eliminate Everything.** We apply VE to eliminate all variables in the network to reach trivial instantiation \top (corresponds to root).

8 Causality - Interventions

8.1 Notations

Causal Effect (CE). of $X = x$ on $Y = y$ can be written as

$$\Pr(Y = y | do(X = x)) \equiv \Pr(y | do(x)) \equiv \Pr(y_x) \quad (8.1)$$

Interventional Distribution. For $\Pr(X, Y, Z)$, the interventional distribution for $do(X = x)$ is denoted as

$$\Pr_{X=x}(Y, Z) \quad (8.2)$$

8.2 Types of Causal Graphs

8.1

^{8.1}Hidden variables are roots.

8.2.1 Markovian Model

Each hidden variable in a Markovian Model has at most one child. It has an alternative name of “no hidden confounders”. In this case, causal effects are always identifiable.

8.2.2 Semi-Markovian Model

Some hidden variable has more than one child. In this case, causal effects are not always identifiable.

8.3 Identifiability Criterion

8.3.1 Causal Effect Rule

The Causal Effect Rule links together association and intervention. It states the following: if \mathbf{Z} are the parents of X , then

$$\Pr(y|do(x)) = \Pr(y_x) = \sum_{\mathbf{z}} \Pr(y|x, \mathbf{z})\Pr(\mathbf{z}) \quad (8.3)$$

The catch to this formulation is one have to know the parents - meaning that we need to have a correct causal structure prior to using this formula. This is a strong assumption. Often, the structure is exactly what we are after.^{8.2}

8.3.2 Backdoor Criteria

A path between X and Y is *blocked* by Z iff

- some collider is not in Z , or
- some non-collider is in Z .

where a collider node is simply a convergent valve defined earlier ($\rightarrow W \leftarrow$). Here we distinguish only between colliders and non-colliders. The Backdoor Criteria states the following: Consider a causal graph G and causal effect $\Pr(y_x)$. A set of variables \mathbf{Z} satisfies the backdoor criteria iff

- no node in \mathbf{Z} is a descendant of X ,
- \mathbf{Z} *blocks* every path between X and Y that contains an arrow into X .

Then, if \mathbf{Z} is a backdoor, then

$$\Pr(y_x) = \sum_{\mathbf{z}} \Pr(y|x, \mathbf{z})\Pr(\mathbf{z}) \quad (8.4)$$

^{8.2}Recall that different causal structures can generate the same distribution, and data alone is not enough.

Incompleteness. The backdoor criteria is incomplete. When it identifies that a causal effect has no backdoor, the causal effect can be either identifiable or not identifiable (inconclusive).

8.3.3 Frontdoor Criteria

Consider a causal graph G and causal effect $\Pr(y_x)$. A set of variables \mathbf{Z} satisfies the front-door criteria iff Then if \mathbf{Z} is a frontdoor, then,

$$\Pr(y_x) = \sum_{\mathbf{z}} \Pr(\mathbf{z}|x) \sum_{x'} \Pr(y|x', \mathbf{z}) \Pr(x') \quad (8.5)$$

8.3.4 Exogenous X

For the query $\Pr(y_x)$, when X is an exogenous variable, meaning that it has no parents,

$$\Pr(y_x) = \Pr(y|x) \quad (8.6)$$

8.4 The Do-Calculus

The key idea is to apply a series of rules until we get a formula that is comprised of solely associational quantities. There are three re-write rules in total. $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ are disjoint sets of variables,

Rule 1. Ignoring Observations.

$$\Pr(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = \Pr(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) \quad \text{if} \quad dsep_{G_{\bar{\mathbf{x}}}}(\mathbf{Y}, \mathbf{XW}, \mathbf{Z}) \quad (8.7)$$

where we perform $dsep$ test on an altered graph $G_{\bar{\mathbf{x}}}$, rather than the original causal graph G . (Detailed in Section 8.4.1).

Rule 2. Action / Observation Exchange.

$$\Pr(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = \Pr(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) \quad \text{if} \quad dsep_{G_{\bar{\mathbf{x}}\bar{\mathbf{z}}}}(\mathbf{Y}, \mathbf{XW}, \mathbf{Z}) \quad (8.8)$$

There are several weakened versions of this rule. First, we consider the case of $\mathbf{X} = \emptyset$, then

$$\Pr(\mathbf{y}|do(\mathbf{z}), \mathbf{w}) = \Pr(\mathbf{y}|\mathbf{z}, \mathbf{w}) \quad \text{if} \quad dsep_{G_{\bar{\mathbf{z}}}}(\mathbf{Y}, \mathbf{W}, \mathbf{Z}) \quad (8.9)$$

Simplifying even further, we can remove the common “kept” part \mathbf{W} , and get

$$\Pr(\mathbf{y}|do(\mathbf{z})) = \Pr(\mathbf{y}|\mathbf{z}) \quad \text{if} \quad dsep_{G_{\bar{\mathbf{z}}}}(\mathbf{Y}, \emptyset, \mathbf{Z}) \quad (8.10)$$

notice that this Exogenous X rule (also stated in Section 8.3.4) follows directly as a corollary of Rule 2.

Rule 3. Ignoring Actions.

$$\Pr(\mathbf{y}|\text{do}(\mathbf{x}), \text{do}(\mathbf{z}), \mathbf{w}) = \Pr(\mathbf{y}|\text{do}(\mathbf{x}), \mathbf{w}) \quad \text{if} \quad \text{dsep}_{G_{\overline{\mathbf{XZ}}(\mathbf{W})}}(\mathbf{Y}, \mathbf{XW}, \mathbf{Z}) \quad (8.11)$$

where encounter a special notation $\overline{\mathbf{Z}}(\mathbf{W})$ that means “not all variables in \mathbf{Z} , but only those variables in \mathbf{Z} that do not have ancestor in \mathbf{W} ”.

8.4.1 Graph Alterations

The calculus rules we specified earlier performs dsep tests on altered graphs, where

- $G_{\bar{\mathbf{x}}}$ is obtained via removing edges pointing into variables \mathbf{X} from G .
- $G_{\mathbf{x}}$ is obtained via removing edges pointing away from variables \mathbf{X} from G .

9 Counterfactual Reasoning

Causal effects are on the second level: interventions. In counterfactual reasoning, we step up onto level three: what-if’s.

9.1 The Information Hierarchy

To add to what we have said above,

- Associational reasoning: Bayesian Network
- Interventional reasoning: Causal Bayesian network (causal graph)
- Counterfactual reasoning: Functional Bayesian network (functional dependencies)

9.2 Counterfactual Queries

9.2.1 Probability of Necessity (PN)

Probability that y would not have occurred in the absence of x ($\text{do}(\bar{x})$), given that x and y did in fact occur, i.e.,

$$PN = \Pr(\bar{y}_{\bar{x}}|x, y) \quad (9.1)$$

9.2.2 Probability of Sufficiency (PS)

Probability that setting x would produce y in a situation where both x and y are absent, i.e.,

$$PS = \Pr(y_x|\bar{y}, \bar{x}) \quad (9.2)$$

9.2.3 Probability of Necessity and Sufficiency (PNS)

Probability that y responds to x both ways (measures the necessity and sufficiency for x to produce y):

$$PNS = \Pr(y_x, \bar{y}_{\bar{x}}) = \Pr(x, y)PN + \Pr(\bar{x}, \bar{y})PS \quad (9.3)$$

9.2.4 Probability of Disablement

Probability that y would have been prevented if it were not for x ,

$$PD = \Pr(\bar{y}_{\bar{x}}|y) \quad (9.4)$$

9.2.5 Probability of Enablement

Probability that y would have been realized if it were not for absence of x ,

$$PE = \Pr(y_x|\bar{y}) \quad (9.5)$$

9.3 Structural Causal Models (SCM)

World. In SCM, a world is defined slightly different from what we had in associational graphs. A world is an instantiation of exogenous variables. This is because fixing exogenous variables fully specifies the entire network.

9.4 Evenets

9.4.1 Observational Event

- Form: \mathbf{x} where \mathbf{X} is a set of endogenous variables
- Meaning: variables \mathbf{X} took the value \mathbf{x}
- Examples: x, \bar{y}, z

9.4.2 Interventional Event

- Form: $\mathbf{y}_{\mathbf{x}}$ where \mathbf{X} and \mathbf{Y} are sets of endogenous variables
- Meaning: variables \mathbf{Y} took the value \mathbf{y} after setting $\mathbf{X} = \mathbf{x}$
- Examples: $y_x, y_{\bar{x}z}, (x\bar{y})_{zw}$

9.4.3 Counterfactual Event

- Form: η_1, \dots, η_n where η_i is an observational or interventional event
- Meaning: $\bigwedge_i \eta_i$
- Examples: mix of above two event examples

9.5 Satisfaction

Recalling what we had earlier, we denote \mathbf{u} satisfies event η in SCM \mathcal{M} as

$$\mathbf{u} \models_{\mathcal{M}} \eta \quad (9.6)$$

9.5.1 Observational Event

We write

$$\mathbf{u} \models_{\mathcal{M}} \mathbf{x} \quad (9.7)$$

iff world \mathbf{u} fixes (endogenous) variables \mathbf{X} to \mathbf{x} in SCM \mathcal{M} .

9.5.2 Interventional Event

$$\mathbf{u} \models_{\mathcal{M}} \mathbf{y}_{\mathbf{x}} \iff \mathbf{u} \models_{\mathcal{M}_{\mathbf{x}}} \mathbf{y} \quad (9.8)$$

9.5.3 Counterfactual Event

$$\mathbf{u} \models_{\mathcal{M}} \eta_1, \dots, \eta_n \iff \bigwedge_i \mathbf{u} \models_{\mathcal{M}} \eta_i \quad (9.9)$$

9.6 Queries

The computation of observational events η in SCM is easy. First we compute the table with world probabilities. The table should have a row for each instantiation of the exogenous variables $\mathbf{u} = [U_1, \dots, U_n]$. Endogenous variables, e.g. X, Y, Z will be computed using the factors, for example $f_X(\dots) = \dots$. Then, to compute probability of observational event $\Pr(\eta)$, we simply select rows that satisfies η and add up their respective $\Pr(\mathbf{u})$ values. To compute an interventional event η , we have to first transform the table. This is a two step process,

1. Force the endogenous variable(s) inside $do(\cdot)$ to take value of 1. For example if we want y_x , then we have $do(x)$ and thus we set all the values of X in the table to 1.

2. Now we need to fix the table by adjusting values y . We do so by simplifying the relationship, for example

$$f_Y(X, U_y, U_x) = xu_r + \bar{x}u_yu_r + \bar{x}\bar{u}_y\bar{u}_r \quad (9.10)$$

simplifies to

$$f_Y(X = x, U_x, U_r) = u_r \quad (9.11)$$

Then, the values of y should be computed using this simplified formula (on the altered table).

9.7 Exogeneity and Monotonicity

Exogeneity. X is exogenous relative to Y iff events y_x and x are independent and so are $y_{\bar{x}}$ and x ,

$$\Pr(y_x|x) = \Pr(y_x) \quad \text{and} \quad \Pr(y_{\bar{x}}|x) = \Pr(y_{\bar{x}}) \quad (9.12)$$

Monotonicity. Y is monotonic relative to X iff the event $\bar{y}_x, y_{\bar{x}}$ is unsatisfiable. This notion is general, x and y does not have to be inside the same family. A more general notation is to condition on probability of event $\bar{y}_x, y_{\bar{x}}$ being zero.