Chapter 1

Do Calculus proofs

In Chapter **??** of Bayesuvius, we explained Do Calculus but referred to this chapter for proofs of claims that use Do Calculus. In this chapter, we've aggregated all proofs, from throughout the book, of claims that use Do Calculus.

Note that even though the 3 rules of Do Calculus are great for proving adjustment formulae for general classes of DAGs, they are sometimes overkill for proving adjustment formulae for a single specific DAG. After all, the 3 rules of Do Calculus are a consequence of the d-separation theorem. Hence, all adjustment formulae should be provable from first principles, assuming only the d-separation theorem and the standard rules of probability theory.

In this chapter, we use the following conventions.

Random variables are underlined and their values are not. For example, $\underline{a} = a$ means the random variable $\underline{a}$ takes the value a. Diagrams with nodes that are underlined represent Bayesian Networks (bnets) and the same diagram with the letters not underlined represents a specific **instantiation** of that bnet. For example $\underline{a} \to \underline{b}$ represents the bnet with conditional probability distribution $P(b|a)$, whereas $a \to b$ represents $P(b|a)$ itself.

If $\underline{a}$ is a root node, then $\sum a$ signifies a weighted sum $\sum_a P(a)$. For example, $\sum a \to b = \sum_a P(a)P(b|a)$. If $\underline{a}$ is not a root node as in $x \to \sum a \to y = \sum_a P(y|a)P(a|x)$, then $\sum a$ signifies a simple unweighted sum $\sum_a$.

Unobserved nodes are indicated by enclosing them in a dashed circle. For example, u.

Selection diagrams with selection nodes are discussed in Chapter **??**. In a selection diagram with a selection node $\underline{s} \in \{0, 1\}$, if a node $\underline{x}$ has parents $pa(\underline{x})$ where $\underline{s} \notin pa(\underline{x})$, then the TPM of $\underline{x}$ is $P(x|pa(x))$. If, on the other hand, $\underline{x}$ has parents $pa(\underline{x}) \cup \underline{s}$, then the TPM of $\underline{x}$ is $P(x|pa(x), s)$, where $P(x|pa(x), \underline{s} = 0) = P(x|pa(x))$ and $P(x|pa(x), \underline{s} = 1) = P^*(x|pa(x))$.

Some identities that are used in this chapter:

1.
$$P(y|x_1, x_2) = \sum_a P(y|a, x_1, x_2)P(a|x_1, x_2) \; . \tag{1.1}$$

$$
\begin{tikzcd}
x_1 \arrow[dr] & & & x_1 \arrow[dr] \arrow[drr] & & \\
& y & = & & \sum a \arrow[r] & y \\
x_2 \arrow[ur] & & & x_2 \arrow[ur] \arrow[urr] & &
\end{tikzcd}
\quad . \tag{1.2}
$$

One can describe this identity as "giving $\underline{y}$ a universal backdoor", because $\sum a$ is a backdoor (i.e., input) to y, and $\sum a$ is universal in the sense that it is entered by every arrow that enters y except $\sum a$ itself.

2.

$$\sum_a P(a|x_1, x_2) = 1 \tag{1.3}$$

$$
\begin{tikzcd}
x_1 \arrow[dr] & & \\
& \sum a \arrow[r, "0"'] & \quad = 1 \\
x_2 \arrow[ur] & &
\end{tikzcd}
\tag{1.4}
$$

One can describe this identity as "summing over the values of a collider node which has no emerging arrows"[1]. Eq.(1.4) can be understood as an edge case (when $\underline{y} = \emptyset$) of Eq.(1.2).

3.

$$\sum_a P(x_2|a)P(a|x_1) = P(x_2|x_1) \tag{1.5}$$

$$x_1 \longrightarrow \sum a \longrightarrow x_2 \qquad = \qquad x_1 \longrightarrow x_2 \tag{1.6}$$

One can describe this identity as "summing over the values of a mediator node".

4.

$$P(x) = \sum_a P(x|a)P(a) = \sum_b P(x|b)P(b) \tag{1.7}$$

$$P(x) \qquad = \qquad \xrightarrow[0]{} \sum a \longrightarrow x \qquad = \qquad \xrightarrow[0]{} \sum b \longrightarrow x \tag{1.8}$$

One can describe this identity as "averaging over different priors". Eq.(1.8) can be understood as an edge case of Eq.(1.6).

[1]A zeroed arrow means the same as no arrow.

A **do-adjustment formula** expresses a **do-query** (i.e., a conditional probability with do operators in its condition) by an equivalent expression without do operators. If a do-adjustment formula exists for a particular do-query, then we say the do-query is **do-identifiable**.[2] See Fig.1.1 for some simple examples of identifiable and non-identifiable do-queries.

$P(y|\mathcal{D}\underline{x} = x) = P(y|x)$ $\quad\quad$ $P(y|\mathcal{D}\underline{x} = x) \neq P(y|x)$

Figure 1.1: Examples of identifiable and non-identifiable do-queries. $\mathcal{D}\underline{x} = x$ in a do-query means we erase the arrow $\underline{z} \rightarrow \underline{x}$ and replace node $\underline{x}$ by a delta function $\delta(x, x')$. Let $Q(y|x) = \sum_z P(y|x, z)P(z)$. In both cases, $P(y|\mathcal{D}\underline{x} = x) = Q(y|x)$, but in the non-identifiable case, $Q(y|x)$ is summing over the unobserved node $\underline{z}$, so it is using data that is not available. To distinguish between these two possibilities for $Q(y|x)$, we set $Q(y|x) = P(y|x)$ in the identifiable case, and $Q(y|x) = P(y|\mathcal{D}\underline{x} = x)$ in the non-identifiable case.

The following is the simplest of do-adjustment formulae:

$$P(y|\mathcal{D}\underline{x} = x) = P(y|x) \tag{1.9}$$

By definition, since identifiability gives permission to remove $\mathcal{D}$ operators, all do-identifiable problems satisfy this simplest of adjustment formulae. However, we aim to find an adjustment formula that utilizes the full distribution of the *observed* nodes. Adjustment formulae that don't utilize the full distribution of the observed nodes are throwing away useful info from the dataset, and are less sensitive to deviations from the DAG model being hypothesized.

Given a bnet G and an adjustment formula AF for the query $P(y|\mathcal{D}\underline{x} = x, z)$, a simple identification method (SIM) that we use in this chapter to prove that AF applies to G, is the following:

1. From the bnet G, write an instantiation of G in which the arrows entering node x are amputated, node x is replaced by $\mathcal{D}\underline{x} = x$, and all observed and unobserved nodes, except x, y, z, are summed over.

2. Replace all sums over hidden nodes by sums over observed nodes.

[2] To prove that a do-query $P(y|\mathcal{D}\underline{x} = x, z)$ is do-identifiable for a graph G, just prove that $\underline{y} \perp \underline{x}|\underline{z}$ in $\mathcal{L}_{\underline{x}}G$. This is called Rule 2 of Do Calculus, but it is easy to understand just from the d-separation theorem. Info can be transmitted between $\underline{y}$ and $\underline{x}$ by either (1) paths in $\mathcal{D}_{\underline{x}}G$ or (2) paths in $\mathcal{L}_{\underline{x}}G$. $P(y|\mathcal{D}\underline{x} = x, z) = P(y|x, z)$ means the info is being transmitted only by (1). So the Rule 2 premise is checking that no info is being transmitted by (2).

3. Once all nodes in the diagram represent observed nodes only, we can replace the node $\mathcal{D}\underline{x} = x$ by x. This is justified because in the latest diagram (the one with no unobserved nodes), unavailable info that belongs to unobserved nodes is not being used. After all, that is the meaning of identifiability: that we can express a do-query using only the info that is available.

You can check that using SIM, it is not possible to find an adjustment formula for the example of a non-identifiable query in Fig.1.1. SIM does yield the backdoor adjustment formula for the identifiable query in Fig.1.1.

A **do-transport formula** expresses a do-query in terms of an equivalent do-query.

This chapter deals with both do-adjustment and do-transport formulae.

Claim 1 *(Backdoor Adjustment Formula)*

If
$$\begin{tikzcd} \underline{z} \arrow[d] \arrow[dr] & \\ \underline{x} \arrow[r] & \underline{y} \end{tikzcd}$$
then

$$P(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|x, z)P(z) \tag{1.10}$$

$$\begin{tikzcd} \mathcal{D}\underline{x} = x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \sum z \arrow[dr] & \\ x \arrow[r] & y \end{tikzcd} \tag{1.11}$$

proof:

*** proof 1:**

$P(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|x, z)P(z)$

$$\begin{tikzcd} \sum z \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \sum z \arrow[dr] & \\ x \arrow[r] & y \end{tikzcd}$$

We can replace $\mathcal{D}\underline{x} = x$ by x once all nodes in bnet are observed nodes.

*** proof 2:**

$P(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|\mathcal{D}\underline{x} = x, z)P(z|\mathcal{D}\underline{x} = x)$

by Probability Axioms

$= \sum_z P(y|x, z)P(z|\mathcal{D}\underline{x} = x)$

$P(y|\mathcal{D}\underline{x} = x, z) \rightarrow P(y|x, z)$

by Rule 2: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{L}_{\underline{a}.}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow \underline{a}. = a.$

$\underline{y} \perp \underline{x} | \underline{z}$ in $\mathcal{L}_{\underline{x}}\mathcal{D}_{\emptyset}G$:
$$\begin{tikzcd} \underline{z} \arrow[d] \arrow[dr] & \\ \underline{x} & \underline{y} \end{tikzcd}$$

$= \sum_z P(y|x,z)P(z)$

$P(z|\mathcal{D}\underline{x}=x) \to P(z)$

by Rule 3: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{D}_{\underline{a}.-an(\underline{s}.)}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow 1$

$\underline{z} \perp \underline{x}$ in $\mathcal{D}_{\underline{x}}\mathcal{D}_{\emptyset}G$:
$$\begin{tikzcd} & \underline{z} \arrow[d] \\ \underline{x} \arrow[r] & \underline{y} \end{tikzcd}$$

QED

Claim 2 *(Frontdoor Adjustment Formula)*

If
$$\begin{tikzcd} & \underline{c} \arrow[dl] \arrow[dr] & \\ \underline{x} \arrow[r] & \underline{m} \arrow[r] & \underline{y} \end{tikzcd}$$
then

$$P(y|\mathcal{D}\underline{x}=x) = \sum_m \left[\sum_{x'} P(y|x',m)P(x') \right] P(m|x) \tag{1.12}$$

$$\mathcal{D}\underline{x} = x \longrightarrow y \quad = \quad \begin{tikzcd} & \sum x' \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} \tag{1.13}$$

proof:

*** proof 1:**

$P(y|\mathcal{D}\underline{x}=x) = \sum_{m,c,x'} P(y|m,c)P(c|x')P(x')P(m|\mathcal{D}\underline{x}=x)$

$$\begin{tikzcd} & \sum c \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \sum x' \arrow[r] & \sum c \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd}$$

$= \sum_{m,x'} P(y|m,x')P(x')P(m|x)$

$$= \begin{tikzcd} & \sum x' \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd}$$

We can replace $\mathcal{D}\underline{x} = x$ by x once all nodes in bnet are observed nodes.

*** proof 2:**

$P(y|\mathcal{D}\underline{x} = x) = \sum_m P(y|\mathcal{D}\underline{x} = x, m)P(m|\mathcal{D}\underline{x} = x)$

by Probability Axioms

$= \sum_m P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{m} = m)P(m|\mathcal{D}\underline{x} = x)$

$P(y|\mathcal{D}\underline{x} = x, m) \to P(y|\mathcal{D}\underline{x} = x, \mathcal{D}m = m)$

by Rule 2: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{L}_{\underline{a}.}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow \underline{a}. = a.$

$\underline{y} \perp \underline{m}|\underline{x}$ in $\mathcal{L}_{\underline{m}}\mathcal{D}_{\underline{x}}G$:

$$\begin{tikzcd} & \underline{c} \arrow[dr] & \\ \underline{x} \arrow[r] & \underline{m} & \underline{y} \end{tikzcd}$$

$= \sum_m P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{m} = m)P(m|x)$

$P(m|\mathcal{D}\underline{x} = x) \to P(m|x)$

by Rule 2: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{L}_{\underline{a}.}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow \underline{a}. = a.$

$\underline{m} \perp \underline{x}$ in $\mathcal{L}_{\underline{x}}\mathcal{D}_{\emptyset}G$:

$$\begin{tikzcd} & \underline{c} \arrow[dl] \arrow[dr] & \\ \underline{x} & \underline{m} \arrow[r] & \underline{y} \end{tikzcd}$$

$= \sum_m P(y|\mathcal{D}\underline{m} = m)P(m|x)$

$P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{m} = m) \to P(y|\mathcal{D}\underline{m} = m)$

by Rule 3: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{D}_{\underline{a}.-an(\underline{s}.)}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow 1$

$\underline{y} \perp \underline{x}|\underline{m}$ in $\mathcal{D}_{\underline{x}}\mathcal{D}_{\underline{m}}G$:

$$\begin{tikzcd} & \underline{c} \arrow[dr] & \\ \underline{x} & \underline{m} \arrow[r] & \underline{y} \end{tikzcd}$$

$= \sum_{x'} \sum_m P(y|\mathcal{D}\underline{m} = m, x')P(x'|\mathcal{D}\underline{m} = m)P(m|x)$

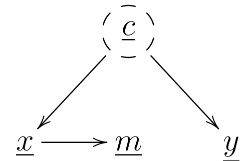
by Probability Axioms

$= \sum_{x'} \sum_m P(y|m, x')P(x'|\mathcal{D}\underline{m} = m)P(m|x)$

$P(y|\mathcal{D}\underline{m} = m, x') \to \quad P(y|m, x')$

by Rule 2: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{L}_{\underline{a}.}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow \underline{a}. = a.$

$\underline{y} \perp \underline{m}|\underline{x}$ in $\mathcal{L}_{\underline{m}}\mathcal{D}_{\emptyset}G$:

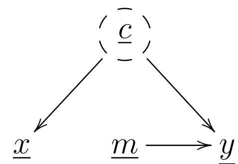


$= \sum_{x'} \sum_m P(y|m, x')P(x')P(m|x)$

$P(x'|\mathcal{D}\underline{m} = m) \to P(x')$

by Rule 3: If $(\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.)$ in $\mathcal{D}_{\underline{a}.-an(\underline{s}.)}\mathcal{D}_{\underline{r}.}G$, then $\mathcal{D}\underline{a}. = a. \leftrightarrow 1$

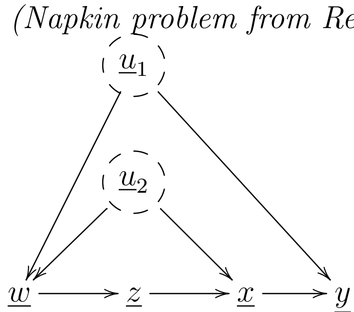
$\underline{x} \perp \underline{m}$ in $\mathcal{D}_{\underline{m}}\mathcal{D}_{\emptyset}G$:



QED

Claim 3 *(Napkin problem from Ref.[3])*

If



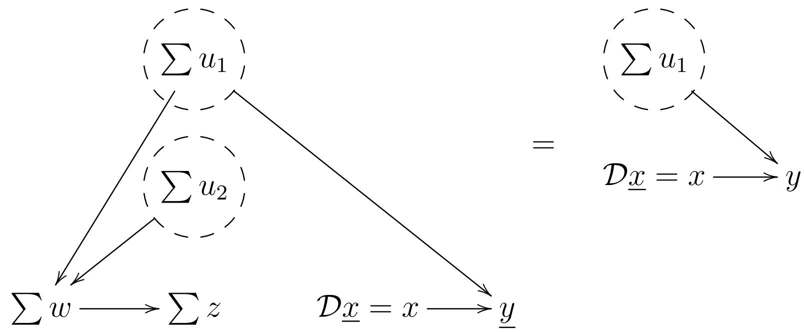
then

$$P(y|\mathcal{D}\underline{x} = x) = \sum_{w,z} P(y|x, w, z)P(w, z) \tag{1.14}$$

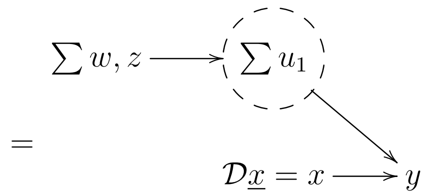
$$\begin{tikzcd} & & \sum w, z \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[r] & y \quad = & x \arrow[r] & y \end{tikzcd} \tag{1.15}$$

proof:

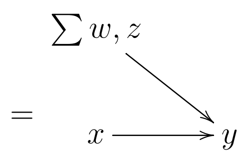
$$P(y|\mathcal{D}\underline{x} = x) = \sum_{u_1} P(y|\mathcal{D}\underline{x} = x, u_1)P(u_1)$$



$= \sum_{w,z} \sum_{u_1} P(y|\mathcal{D}\underline{x} = x, u_1) P(u_1|w,z) P(w,z)$



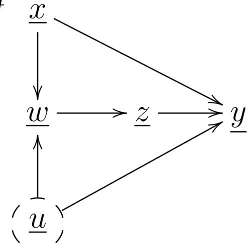
$= \sum_{w,z} P(y|x,w,z) P(w,z)$



We can replace $\mathcal{D}\underline{x} = x$ by x once all nodes in bnet are observed nodes.

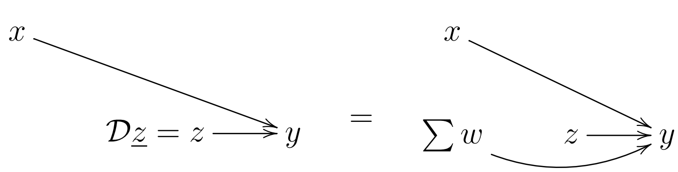
QED

Claim 4 *(from Ref.[3])*
If



then

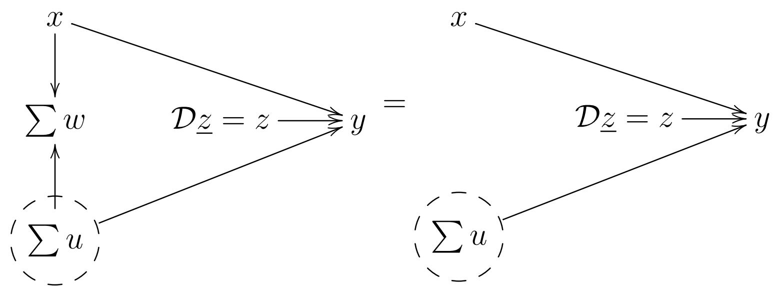
$$P(y|\mathcal{D}\underline{z} = z, x) = \sum_{w} P(y|z,x,w) P(w) \tag{1.16}$$



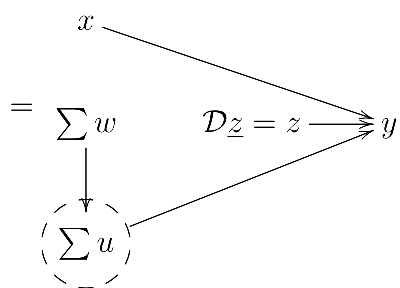
(1.17)

proof:

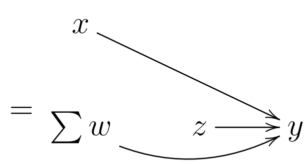
$P(y|\mathcal{D}\underline{z}=z,x) = \sum_u P(y|\mathcal{D}\underline{z}=z,x,u)P(u)$



$= \sum_w \sum_u P(y|\mathcal{D}\underline{z}=z,x,u)P(u|w)P(w)$



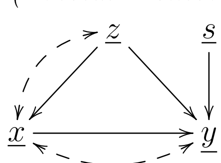
$= \sum_w P(y|z,x,w)P(w)$



We can replace $\mathcal{D}\underline{z}=z$ by z once all nodes in bnet are observed nodes.

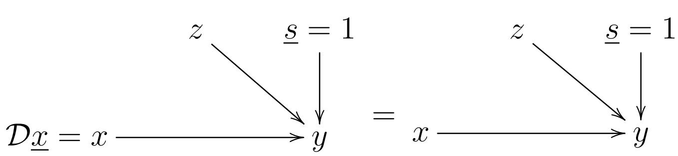
QED

Claim 5 *(Trivial Memoryless Transportability, from Ref.[2])*
If



where $\underline{s} \in \{0,1\}$ *is a selection node, then*

$$P^*(y|\mathcal{D}\underline{x}=x,z) = P^*(y|x,z) \quad \textit{(replace } \mathcal{D} \textit{ by } 1\textit{, keep } P^*\textit{)} \tag{1.18}$$



(1.19)

proof:

$$P(y|\mathcal{D}\underline{x}=x,z,\underline{s}=1) = P(y|x,z,\underline{s}=1)$$

$$\begin{tikzcd} & z \arrow[dr] & \underline{s}=1 \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} = \begin{tikzcd} & z \arrow[dr] & \underline{s}=1 \arrow[d] \\ x \arrow[rr] & & y \end{tikzcd}$$

QED

Claim 6 *(Direct Transportability, a.k.a. External Validity, from Ref.[2])*
If
$$\begin{tikzcd} \underline{s} \arrow[r] & \underline{z} \arrow[dl] \arrow[dr] \arrow[dl, dashed, bend right, leftrightarrow] & \\ \underline{x} \arrow[rr] \arrow[rr, dashed, bend right, leftrightarrow] & & \underline{y} \end{tikzcd}$$
where $\underline{s}\in\{0,1\}$ is a selection node, then

$$P^*(y|\mathcal{D}\underline{x}=x,z) = P(y|\mathcal{D}\underline{x}=x,z) \quad \textit{(replace } P^* \textit{ by } P\textit{, keep } \mathcal{D}\textit{)} \tag{1.20}$$

$$\begin{tikzcd} \underline{s}=1 \arrow[r] & z \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} = \begin{tikzcd} & z \arrow[dr] \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} \tag{1.21}$$

Furthermore,

$$P^*(y|\mathcal{D}\underline{x}=x) = \sum_z P(y|\mathcal{D}\underline{x}=x,z)P^*(z) \tag{1.22}$$

$$\begin{tikzcd} \underline{s}=1 \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \underline{s}=1 \arrow[r] & \sum z \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} \tag{1.23}$$

proof:

$$P(y|\mathcal{D}\underline{x}=x,z,\underline{s}=1) = P(y|\mathcal{D}\underline{x}=x,z)$$

$$\begin{tikzcd} \underline{s}=1 \arrow[r] & z \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} = \begin{tikzcd} & z \arrow[dr] \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd}$$

Because $\underline{s} \perp \underline{y}|\underline{z}$

Furthermore,

$$P(y|\mathcal{D}\underline{x}=x,\underline{s}=1) = \sum_z P(y|\mathcal{D}\underline{x}=x,z)P(z|\underline{s}=1)$$

$$\begin{tikzcd} \underline{s}=1 \arrow[r] & \sum z \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd}$$

QED

Claim 7 *(S-Admisssible Transportability, from Ref.[2])*

If

$$\begin{tikzcd} \underline{s} \arrow[r] & \underline{z} \arrow[r] \arrow[dl] \arrow[dl, dashed, bend right, leftrightarrow] & \underline{a} \arrow[d] \\ \underline{x} \arrow[rr] \arrow[rr, dashed, bend right, leftrightarrow] & & \underline{y} \end{tikzcd}$$

where $\underline{s} \in \{0, 1\}$ is a selection node, then

$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_a P(y|\mathcal{D}\underline{x} = x, a)P^*(a) \tag{1.24}$$

$$\begin{tikzcd} \underline{s}=1 \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \underline{s}=1 \arrow[r] & \sum a \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} \tag{1.25}$$

proof:

$$P(y|\mathcal{D}\underline{x} = x, \underline{s} = 1) = \sum_a P(y|\mathcal{D}\underline{x} = x, a)P(a|\underline{s} = 1)$$

$$\begin{tikzcd} \underline{s}=1 \arrow[r] & \sum z \arrow[r] & \sum a \arrow[d] \\ & \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \underline{s}=1 \arrow[r] & \sum a \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd}$$

QED

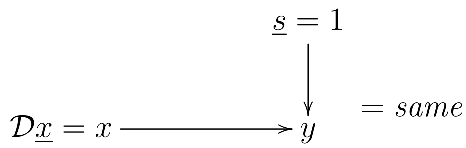
Claim 8 *(Non-transportability, from Ref.[2])*

If

$$\begin{tikzcd} & \underline{h} \arrow[dl] \arrow[dr] & \underline{s} \arrow[d] \\ \underline{x} \arrow[rr] & & \underline{y} \end{tikzcd}$$

where $\underline{s} \in \{0, 1\}$ is a selection node, then

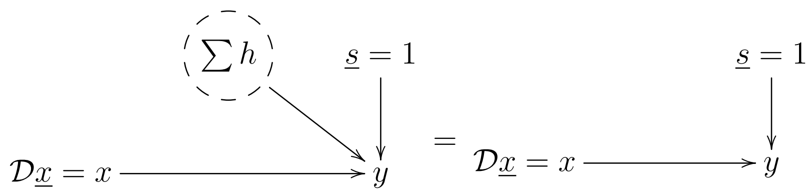
$$P^*(y|\mathcal{D}\underline{x} = x) = P^*(y|\mathcal{D}\underline{x} = x) \tag{1.26}$$



(1.27)

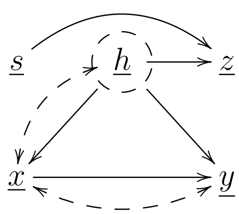
proof:

$P^*(y|\mathcal{D}\underline{x}=x) = P^*(y|\mathcal{D}\underline{x}=x)$

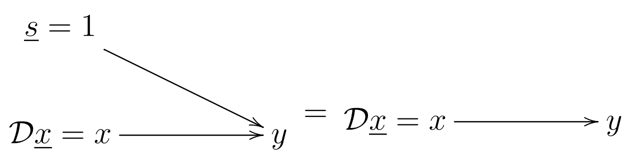


Can't replace $\mathcal{D}\underline{x}=x$ by x because $\underline{y} \not\perp \underline{x}$ in $\mathcal{L}_{\underline{x}}G$. Hence, Rule 2 not satisfied.
QED

Claim 9 *(from Ref.[2])*

If  *where* $\underline{s}\in\{0,1\}$ *is a selection node, then*

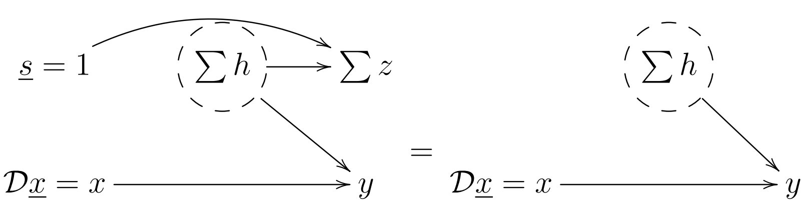
$$P^*(y|\mathcal{D}\underline{x}=x) = P(y|\mathcal{D}\underline{x}=x) \tag{1.28}$$



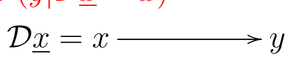
(1.29)

proof:

$P(y|\mathcal{D}\underline{x}=x,\underline{s}=1) = \sum_h P(y|\mathcal{D}\underline{x}=x,h)P(h)$

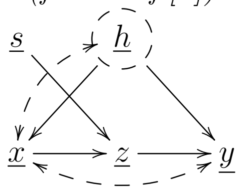


$= P(y|\mathcal{D}\underline{x}=x)$

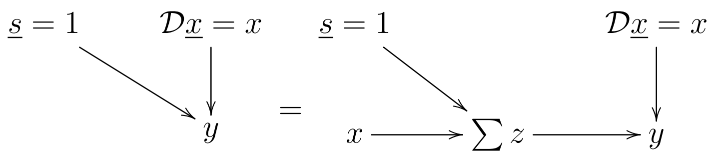
$=$ 

QED

Claim 10 *(from Ref.[2])*

If  *where* $\underline{s} \in \{0,1\}$ *is a selection node, then*

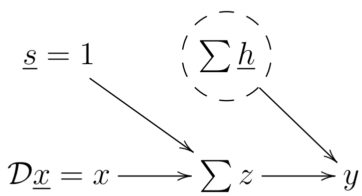
$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|\mathcal{D}\underline{x} = x, z)P^*(z|x) \tag{1.30}$$



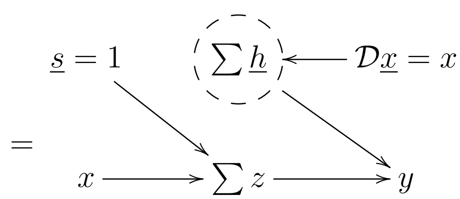
(1.31)

proof:

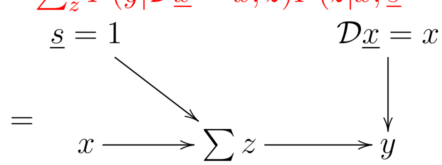
$P(y|\mathcal{D}\underline{x} = x, \underline{s} = 1) = \sum_h \sum_z P(y|h,z)P(h)P(z|\mathcal{D}\underline{x} = x, \underline{s} = 1)$



$= \sum_h \sum_z P(y|h,z)P(h|\mathcal{D}\underline{x} = x)P(z|x, \underline{s} = 1)$



$= \sum_z P(y|\mathcal{D}\underline{x} = x, z)P(z|x, \underline{s} = 1)$



QED

Claim 11 *(Unconfounded Mediation, from Ref.[1])*

If
$$\begin{tikzcd}
& \underline{m} \arrow[dr] & \\
\underline{d} \arrow[ur] \arrow[rr] & & \underline{y}
\end{tikzcd}$$
then

$$P(y|\mathcal{D}\underline{d}=d,\mathcal{I}_{\underline{m}}\underline{d}=d') = \sum_m P(y|d,m)P(m|d') \tag{1.32}$$

$$\begin{tikzcd}
\mathcal{I}\underline{d}=d' \arrow[dr] & \\
\mathcal{D}\underline{d}=d \arrow[r] & y
\end{tikzcd}
=
\begin{tikzcd}
\mathcal{I}\underline{d}=d' \arrow[r] & \sum m \arrow[dr] & \\
\mathcal{D}\underline{d}=d \arrow[rr] & & y
\end{tikzcd} \tag{1.33}$$

proof:

$$P(y|\mathcal{D}\underline{d}=d,\mathcal{I}\underline{d}=d') = \textstyle\sum_m P(y|d,m)P(m|d')$$

$$\begin{tikzcd}
\mathcal{I}\underline{d}=d' \arrow[r] & \sum m \arrow[dr] & \\
\mathcal{D}\underline{d}=d \arrow[rr] & & y
\end{tikzcd}$$

QED

Claim 12 *(Mediation with universal prior* $\underline{\xi}$ *and universal confounder* $\underline{u}$*, from Ref.[1])*

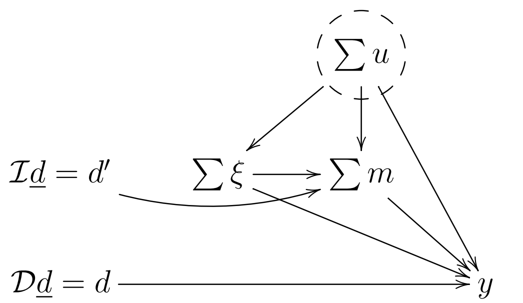
If
$$\begin{tikzcd}
& \underline{u} \arrow[dl] \arrow[d] \arrow[ddl] \arrow[ddr] & \\
\underline{\xi} \arrow[r] \arrow[d] \arrow[drr] & \underline{m} \arrow[dr] & \\
\underline{d} \arrow[ur] \arrow[rr] & & \underline{y}
\end{tikzcd}$$
then

$$P(y|\mathcal{D}\underline{d}=d,\mathcal{I}_{\underline{m}}\underline{d}=d') = \sum_\xi \sum_m P(y|d,m,\xi)P(m|d',\xi)P(\xi) \tag{1.34}$$

$$\begin{tikzcd}
\mathcal{I}\underline{d}=d' \arrow[dr] & \\
\mathcal{D}\underline{d}=d \arrow[r] & y
\end{tikzcd}
=
\begin{tikzcd}
\mathcal{I}\underline{d}=d' \arrow[rr, bend right] & \sum \xi \arrow[r] \arrow[drr] & \sum m \arrow[dr] & \\
\mathcal{D}\underline{d}=d \arrow[rrr] & & & y
\end{tikzcd} \tag{1.35}$$

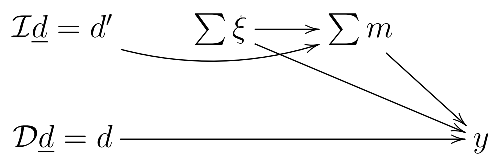
proof:

$$P(y|\mathcal{D}\underline{d}=d, \mathcal{I}\underline{d}=d') = \sum_{\xi,u}\sum_m P(y|d,m,\xi,u)P(m|d',\xi,u)\underbrace{P(\xi|u)P(u)}_{P(\xi,u)}$$



$$= \sum_\xi \sum_m P(y|d,m,\xi)P(m|d',\xi)P(\xi)$$

$=$

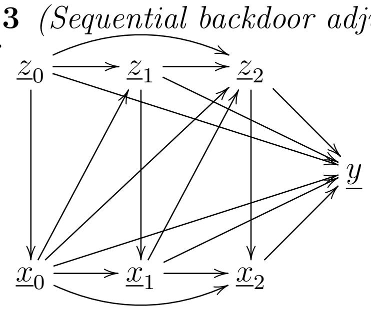


We switch from averaging over the prior of ξ, u to averaging over the prior of ξ.

QED

Claim 13 *(Sequential backdoor adjustment formula)*

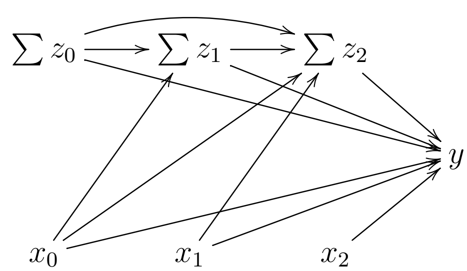
If



then

$$P(y|\mathcal{D}\underline{x}^3 = x^3) = \mathcal{Q}(y|x^3) \tag{1.36}$$

$=$

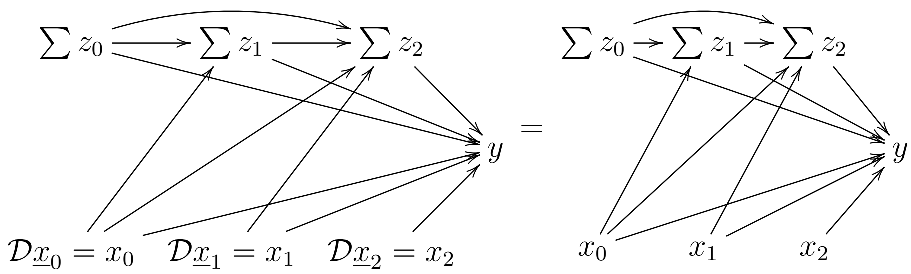


(1.37)

The result shown here for $n=3$ is true for any integer $n \geq 1$.

proof:

$P(y|\mathcal{D}\underline{x}^3 = x^3) = \mathcal{Q}(y|x^3)$



We can replace $\mathcal{D}\underline{x}_i = x_i$ by x_i once all nodes in bnet are observed nodes.

QED

Bibliography

[1] Judea Pearl. Causal and counterfactual inference. *The Handbook of Rationality*, pages 1–41, 2019. https://ftp.cs.ucla.edu/pub/stat_ser/r485.pdf.

[2] Judea Pearl and Elias Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25, 2011. https://ojs.aaai.org/index.php/AAAI/article/view/7861.

[3] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic Books, 2018.