Chapter 1

Do Calculus proofs

In Chapter **??**, 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. Indeed, since the 3 rules of Do Calculus are a consequence of the d-separation theorem, it follows that 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 for bnet diagrams.

Random variables are underlined and their values are not. For example, $\underline{a} = a$ means the random variable $\underline{a}$ takes the value a. A diagram with all its nodes underlined represents a Bayesian Network (bnet), whereas the same diagram with the letters not underlined represents a specific **instantiation** of that bnet. For example $\underline{a} \to \underline{b} \to \underline{c}$ represents the bnet with full probability distribution $P(c|b)P(b|a)P(a)$, whereas $a \to b \to c$ represents $P(c|b)P(b|a)$. Note that, for convenience, we define $a \to b \to c$ to exclude the priors of root nodes such as $P(a)$.

If $\underline{a}$ is a root node, then $\sum a$ signifies a weighted sum $\sum_a P(a)$. For example,

$$\sum a \to b \to c = \sum_a P(c|b)P(b|a)P(a) \tag{1.1}$$

If $\underline{a}$ is not a root node, then $\sum a$ signifies a simple unweighted sum $\sum_a$. For example,

$$x \to \sum a \to y = \sum_a P(y|a)P(a|x) \tag{1.2}$$

Two bnets are equated if their full probability distributions (i.e., their full instantiations) are equal numerically. For example,

$$\underline{a} \to \underline{b} \to \underline{c} = P(c|b)P(b|a)P(a) = \underline{a} \leftarrow \underline{b} \leftarrow \underline{c} \tag{1.3}$$

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 $pa(\underline{x}) = pa'(\underline{x}) \cup \underline{s}$, where $pa'(\underline{x}) = pa(\underline{x}) - \underline{s}$, then the TPM of $\underline{x}$ is

$$P(x|pa'(x), s) = \begin{cases} P(x|pa'(x)) & \text{if } s = 0 \\ P^*(x|pa'(x)) & \text{if } s = 1 \end{cases} \tag{1.4}$$

Some identities that are used in this chapter:

1.
$$P(y|x_1, x_2) = \sum_a P(y|a, x_1, x_2) P(a|x_1, x_2) \; . \tag{1.5}$$

$$\begin{tikzcd} x_1 \arrow[dr] & & & x_1 \arrow[dr] \arrow[drr] & & \\ & y & = & & \sum a \arrow[r] & y \\ x_2 \arrow[ur] & & & x_2 \arrow[ur] \arrow[urr] & & \end{tikzcd} \; . \tag{1.6}$$

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.
$$\sum_a P(a|x_1, x_2) = 1 \tag{1.7}$$

$$\begin{tikzcd} x_1 \arrow[dr] & & & \\ & \sum a \arrow[r, "0"'] & {} & = 1 \\ x_2 \arrow[ur] & & & \end{tikzcd} \tag{1.8}$$

One can describe this identity as "summing over the values of a collider node which has no emerging arrows"[1]. Eq.(1.8) can be understood as an edge case (when $\underline{y} = \emptyset$) of Eq.(1.6).

3.
$$\sum_a P(x_2|a) P(a|x_1) = P(x_2|x_1) \tag{1.9}$$

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

$$x_1 \longrightarrow \sum a \longrightarrow x_2 \qquad = \qquad x_1 \longrightarrow x_2 \tag{1.10}$$

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

4.
$$P(x) = \sum_a P(x|a)P(a) = \sum_b P(x|b)P(b) \tag{1.11}$$

$$P(x) \qquad = \qquad \xrightarrow[0]{} \sum a \longrightarrow x \qquad = \qquad \xrightarrow[0]{} \sum b \longrightarrow x \tag{1.12}$$

One can describe this identity as "averaging over different priors". Eq.(1.12) can be understood as an edge case of Eq.(1.10).

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. The equivalent expression must satisfy 2 constraints to be discussed below. If a do-adjustment formula exists for a particular do-query, then we say the do-query is **do-identifiable (DI)**. A **do-transport formula** is a relationship between 2 do-queries. This chapter deals with both do-adjustment and do-transport formulae.

See Fig.1.1 for some simple examples of of bnets for which the do-query $P(y|\mathcal{D}\underline{x} = x)$ is DI and non-DI.

$(a)P(y|\mathcal{D}\underline{x} = x)$ is DI $\qquad$ $(b)P(y|\mathcal{D}\underline{x} = x)$ is non-DI

Figure 1.1: Examples of bnets for which the do-query $P(y|\mathcal{D}\underline{x} = x)$ is DI and non-DI.

Let G be a graph before amputation of the arrows entering node $\underline{x}$, and let $G_x = \mathcal{D}_{\underline{x}=x}G$ be the same graph after amputation. Also let $P_G() = P()$ be the full probability distribution for graph G, and $P_{G_x}()$ be that for G_x. In general, the following is always true, whether it applies to a bnet with or without hidden nodes:[2]

$$P(y|\mathcal{D}\underline{x} = x) = P_{G_x}(y|x) \tag{1.13}$$

However, the right hand side of this equation is not a valid adjustment formula for this query because it's not expressed in terms of $P()$. We define a valid adjustment formula for query $P(y|\mathcal{D}\underline{x} = x)$ to be a bnet instantiation diagram that satisfies the following 2 constraints:

[2]Note that $P_{G_x}(y|x) \neq P_G(y|x) = P(y|x)$. In fact, $P_{G_x}(y|x) = P(y|x)$ iff there is no confounding, so $P_{G_x}(y|x) \neq P(y|x)$ indicates confounding.

1. **(structural constraint)**

   The adjustment formula must be representable by a bnet instantiation that has a DAG structure identical to the DAG structure of G, except that arrows entering node $\underline{x}$ have been amputated. All nodes of that instatiation, except nodes x and y, must be summed over.

2. **(probabilitistic constraint)**

   - If G has hidden nodes, these must be renamed and assigned a TPM that can be constructed from the *observable* TPMs of G.
   - The observable nodes of G with hidden parents, must also be assigned a TPM that can be constructed from the *observable* TPMs of G.
   - The observable nodes of G with no hidden parents, must be assigned the same TPM as they have in G.

The reason for these 2 constraints is that we want an adjustment formula to show exactly how it is calculated from the full probability distribution $P()$. If we don't show exactly how it is calculated (i.e., whether with or without the amputated arrows), it is impossible to distinguish between marginals of $P()$ and $P_{G_x}()$ such as $P(y|x)$ and $P_{G_x}(y|x)$.

Based on these 2 constraints, we can easily see why the query $P(y|\mathcal{D}\underline{x} = x)$ is DI (resp., non-DI) for bnet (a) (resp., bnet (b)) of Fig.1.1. For bnet (a), after amputating arrow $z \to x$ and summing over node z, we get

$$P(y|\mathcal{D}\underline{x} = x) \quad = \quad \begin{tikzcd} \sum z \arrow[dr] & \\ x \arrow[r] & y \end{tikzcd} \tag{1.14}$$

The right hand side of Eq.(1.14) is a valid adjustment formula because it satisfies both constraints. For bnet (b), if we amputate arrow $z \to x$ and sum over node z, we get

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

The right hand side of Eq.(1.15) is not a valid adjustment formula because it violates the second constraint. Furthermore, try as we may, there is no way to replace the sum over hidden node z by a sum over an observed node, such that constraint 2 is satisfied.

Claim 1 *(Backdoor Adjustment Formula)*

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

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

$$= \sum \begin{tikzcd} & z \arrow[dr] & \\ & x \arrow[r] & y \end{tikzcd} \tag{1.17}$$

proof:

*** proof 1:**

$$P(y|\mathcal{D}\underline{x} = x) = \sum \begin{tikzcd} & z \arrow[dr] & \\ & x \arrow[r] & y \end{tikzcd} \tag{1.18}$$

*** 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

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

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

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

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

$$P(y|\mathcal{D}\underline{x}=x) = \sum_m \left[\sum_{x'} P(y|x',m)P(x')\right] P(m|x) \tag{1.19}$$

$$= \begin{tikzcd} & \sum x' \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} \tag{1.20}$$

proof:
*** proof 1:**

$$P(y|\mathcal{D}\underline{x}=x) = \begin{tikzcd} & \sum c \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} \tag{1.21}$$

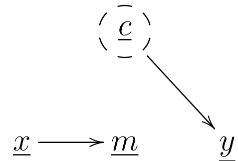
$$= \begin{tikzcd} \sum x' \arrow[r] & \sum c \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} \tag{1.22}$$

$$= \begin{tikzcd} & \sum x' \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} \tag{1.23}$$

*** 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 $\mathcal{L}_{\underline{m}} \mathcal{D}_{\underline{x}} G$:

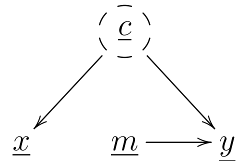


$= \sum_m P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{m} = m) P(m|x)$

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

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

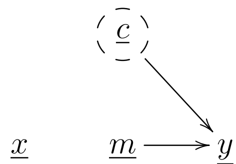


$= \sum_m P(y|\mathcal{D}\underline{m} = m) P(m|x)$

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

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



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

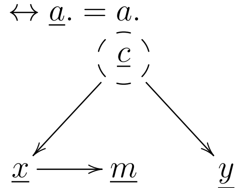
by Probability Axioms

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

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

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

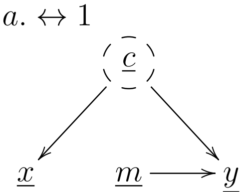


$= \sum_{x'} \sum_m P(y|m, x') P(x') P(m|x)$

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

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.[4])*
If

$$\begin{tikzcd}
& \underline{u}_1 \arrow[ddl] \arrow[ddrr] & & \\
& \underline{u}_2 \arrow[dl] \arrow[dr] & & \\
\underline{w} \arrow[r] & \underline{z} \arrow[r] & \underline{x} \arrow[r] & \underline{y}
\end{tikzcd}$$

then

$$P(y|\mathcal{D}\underline{x} = x) = \sum_{w,z,z',x'} P(y|x,z')P(w|z',x')P(z')P(x')P(z|w) \tag{1.24}$$

$$= \begin{tikzcd}
& \sum z' \arrow[ddl] \arrow[ddrrr] & & & \\
& \sum x' \arrow[dl] & & & \\
\sum w \arrow[r] & \sum z & & x \arrow[r] & \underline{y}
\end{tikzcd} \tag{1.25}$$

Note that x' and z' can be swapped, and we still get a valid adjustment formula. So there can be more that one adjustment formula!

proof:

$$P(y|\mathcal{D}\underline{x}=x) = \begin{tikzcd} & \sum u_1 \arrow[ddl] \arrow[ddrr] & & \\ & \sum u_2 \arrow[dl] & & \\ \sum w \arrow[r] & \sum z & x \arrow[r] & \underline{y} \end{tikzcd} \tag{1.26}$$

$$= \begin{tikzcd} & \sum u_1 \arrow[ddl] \arrow[ddrr] & & \sum z' \arrow[ll] \\ & \sum u_2 \arrow[dl] & & \sum x' \arrow[ll] \\ \sum w \arrow[r] & \sum z & x \arrow[r] & \underline{y} \end{tikzcd} \tag{1.27}$$

$$= \begin{tikzcd} & \sum z' \arrow[ddl] \arrow[ddrr] & & \\ & \sum x' \arrow[dl] & & \\ \sum w \arrow[r] & \sum z & x \arrow[r] & \underline{y} \end{tikzcd} \tag{1.28}$$

QED

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

$$\begin{tikzcd} \underline{x} \arrow[d] \arrow[drr] & & \\ \underline{w} \arrow[r] & \underline{z} \arrow[r] & \underline{y} \\ \underline{u} \arrow[u] \arrow[urr] & & \end{tikzcd}$$

then

$$P(y|\mathcal{D}\underline{z}=z,x) = \sum_w P(y|z,x,w)P(w) \tag{1.29}$$

$$= \begin{tikzcd} x \arrow[drr] & & \\ \sum w \arrow[rr, bend right] & z \arrow[r] & y \end{tikzcd} \tag{1.30}$$

proof:

$$
\begin{aligned}
P(y|\mathcal{D}\underline{z}=z,x) &= \begin{tikzcd} x \arrow[d] \arrow[drr] & & \\ \sum w & z \arrow[r] & y \\ \sum u \arrow[u] \arrow[urr] & & \end{tikzcd} && (1.31) \\
&= \begin{tikzcd} x \arrow[d] \arrow[drr] & & \\ \sum w & z \arrow[r] & y \\ \sum u \arrow[u] \arrow[urr] & \sum w' \arrow[l] & \end{tikzcd} && (1.32) \\
&= \begin{tikzcd} x \arrow[d] \arrow[drr] & & \\ \sum w & z \arrow[r] & y \\ \sum w' \arrow[u] \arrow[urr] & & \end{tikzcd} && (1.33) \\
&= \begin{tikzcd} x \arrow[drr] & & \\ \sum w \arrow[rr, bend right] & z \arrow[r] & y \end{tikzcd} && (1.34)
\end{aligned}
$$

QED

Claim 5 *(Trivial Memoryless Transportability, from Ref.[3])*
If $\begin{tikzcd} & \underline{z} \arrow[dl] \arrow[dr] \arrow[dl, dashed, bend right, leftrightarrow] & \underline{s} \arrow[d] \\ \underline{x} \arrow[rr] \arrow[rr, dashed, bend right, leftrightarrow] & & \underline{y} \end{tikzcd}$ *where* $\underline{s} \in \{0,1\}$ *is a selection node, then*

$$P^*(y|\mathcal{D}\underline{x}=x,z) = P^*(y|x,z) \quad \textit{(replace } \mathcal{D} \textit{ by } 1\textit{, keep } P^*\textit{)} \tag{1.35}$$

$$\begin{tikzcd} & z \arrow[dr] & \underline{s}=1 \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} = \begin{tikzcd} & z \arrow[dr] & \underline{s}=1 \arrow[d] \\ x \arrow[rr] & & y \end{tikzcd} \tag{1.36}$$

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.[3])*
If $\begin{tikzcd} \underline{s} \arrow[r] & \underline{z} \arrow[dl] \arrow[dr] \arrow[dl, dashed, bend right] & \\ \underline{x} \arrow[rr] \arrow[rr, dashed, bend right, leftrightarrow] & & \underline{y} \end{tikzcd}$ *where* $\underline{s}\in\{0,1\}$ *is a selection node, then*

$$P^*(y|\mathcal{D}\underline{x}=x,z)=P(y|\mathcal{D}\underline{x}=x,z)\quad \textit{(replace } P^* \textit{ by } P, \textit{ keep } \mathcal{D}) \tag{1.37}$$

$$\begin{tikzcd} & \underline{s}=1 \arrow[r] & z \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[rr] & & y \end{tikzcd} = \begin{tikzcd} & z \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} \tag{1.38}$$

Furthermore,

$$P^*(y|\mathcal{D}\underline{x}=x)=\sum_z P(y|\mathcal{D}\underline{x}=x,z)P^*(z) \tag{1.39}$$

$$\begin{tikzcd} \underline{s}=1 \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \underline{s}=1 \arrow[r] & \sum z \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} \tag{1.40}$$

proof:

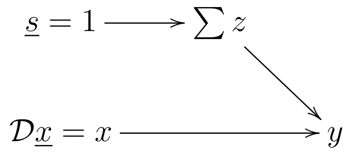
$$P(y|\mathcal{D}\underline{x}=x,z,\underline{s}=1)=P(y|\mathcal{D}\underline{x}=x,z)$$

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

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

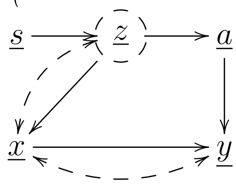
Furthermore,

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

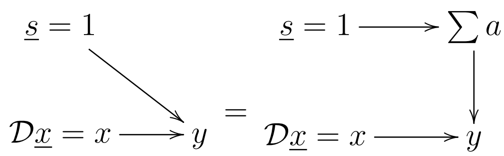


QED

Claim 7 *(S-Admisssible Transportability, from Ref.[3])*

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

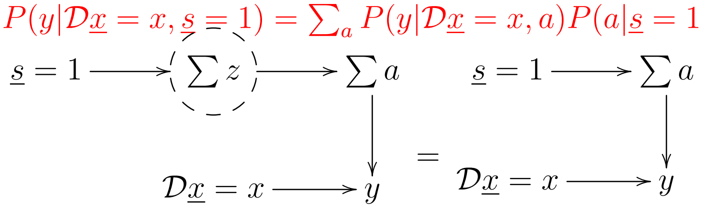
$$P^*(y|\mathcal{D}\underline{x}=x)=\sum_a P(y|\mathcal{D}\underline{x}=x,a)P^*(a) \tag{1.41}$$



(1.42)

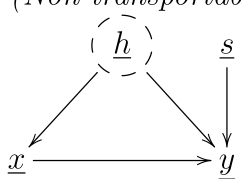
proof:

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



QED

Claim 8 *(Non-transportability, from Ref.[3])*

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

$$P^*(y|\mathcal{D}\underline{x}=x)=P^*(y|\mathcal{D}\underline{x}=x) \tag{1.43}$$

$$\begin{tikzcd} & \underline{s}=1 \arrow[d] & \\ \mathcal{D}\underline{x}=x \arrow[r] & y & = \textit{same} \end{tikzcd} \tag{1.44}$$

proof:

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

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

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.[3])*

$$\textit{If} \quad \begin{tikzcd} \underline{s} \arrow[rr, bend left] & \underline{h} \arrow[r] \arrow[dl] \arrow[dr] & \underline{z} \\ \underline{x} \arrow[rr] \arrow[ur, dashed, bend left, leftrightarrow] \arrow[rr, dashed, bend right, leftrightarrow] & & \underline{y} \end{tikzcd} \quad \textit{where } \underline{s} \in \{0,1\} \textit{ is a selection node, then}$$

$$P^*(y|\mathcal{D}\underline{x}=x) = P(y|\mathcal{D}\underline{x}=x) \tag{1.45}$$

$$\begin{tikzcd} \underline{s}=1 \arrow[dr] & & & & \\ \mathcal{D}\underline{x}=x \arrow[r] & y & = & \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} \tag{1.46}$$

proof:

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

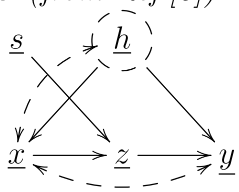
$$\begin{tikzcd} \underline{s}=1 \arrow[rr, bend left] & \sum h \arrow[r] \arrow[d] & \sum z & & & \sum h \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] & y & & = & \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd}$$

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

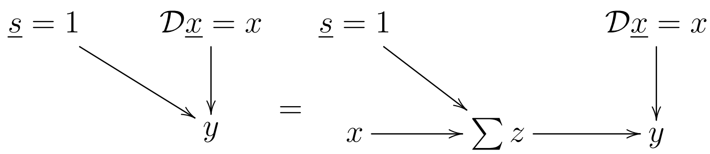
$$= \mathcal{D}\underline{x}=x \longrightarrow y$$

QED

Claim 10 *(from Ref.[3])*

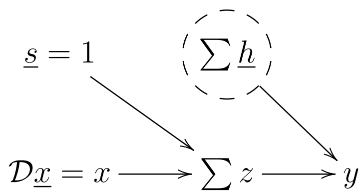
If  *where* $\underline{s} \in \{0, 1\}$ *is a selection node, then*

$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|\mathcal{D}\underline{x} = x, z)P^*(z|x) \tag{1.47}$$

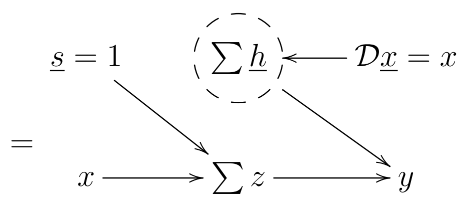
 (1.48)

proof:

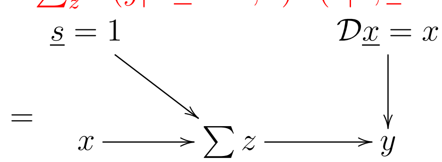
$$P(y|\mathcal{D}\underline{x} = x, \underline{s} = 1) = \sum_h \sum_z P(y|h, z)P(h)P(z|\mathcal{D}\underline{x} = x, \underline{s} = 1)$$



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

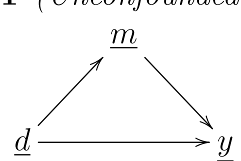


$$= \sum_z P(y|\mathcal{D}\underline{x} = x, z)P(z|x, \underline{s} = 1)$$



QED

Claim 11 *(Unconfounded Mediation, from Ref.[2])*

If  *then*

$$P(y|\mathcal{D}\underline{d}=d,\mathcal{I}_{\underline{m}}\underline{d}=d')=\sum_m P(y|d,m)P(m|d') \tag{1.49}$$

$$\begin{tikzcd} \mathcal{I}\underline{d}=d' \arrow[dr] & \\ \mathcal{D}\underline{d}=d \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \mathcal{I}\underline{d}=d' \arrow[r] & \sum m \arrow[d] \\ \mathcal{D}\underline{d}=d \arrow[r] & y \end{tikzcd} \tag{1.50}$$

proof:

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

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

QED

Claim 12 *(Mediation with universal prior $\underline{\xi}$ and universal confounder $\underline{u}$, from Ref.[2])*

If

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

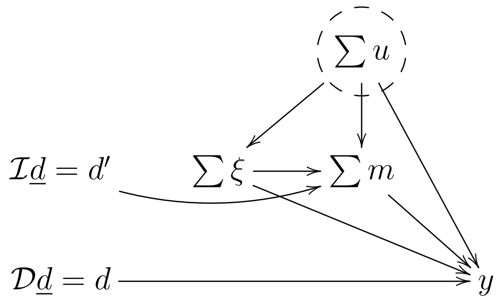
then

$$P(y|\mathcal{D}\underline{d}=d,\mathcal{I}_{\underline{m}}\underline{d}=d')=\sum_\xi\sum_m P(y|d,m,\xi)P(m|d',\xi)P(\xi) \tag{1.51}$$

$$\begin{tikzcd} \mathcal{I}\underline{d}=d' \arrow[dr] & \\ \mathcal{D}\underline{d}=d \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \mathcal{I}\underline{d}=d' \arrow[rr, bend right] & \sum \xi \arrow[r] \arrow[drr] & \sum m \arrow[dr] & \\ \mathcal{D}\underline{d}=d \arrow[rrr] & & & y \end{tikzcd} \tag{1.52}$$

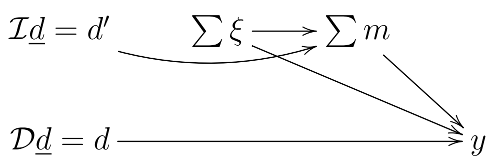
proof:

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



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

=

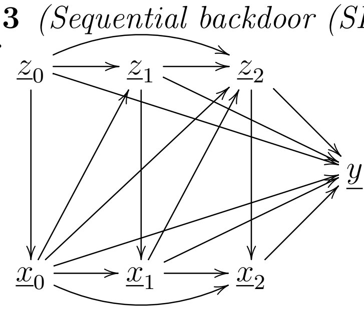


We switch from averaging over the prior of ξ, u to averaging over the prior of ξ.

QED

Claim 13 *(Sequential backdoor (SBD) adjustment formula, from Ref.[5])*

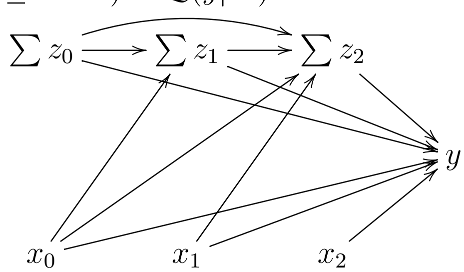
If



then

$$P(y|\mathcal{D}\underline{x}^3 = x^3) = \mathcal{Q}(y|x^3) \tag{1.53}$$

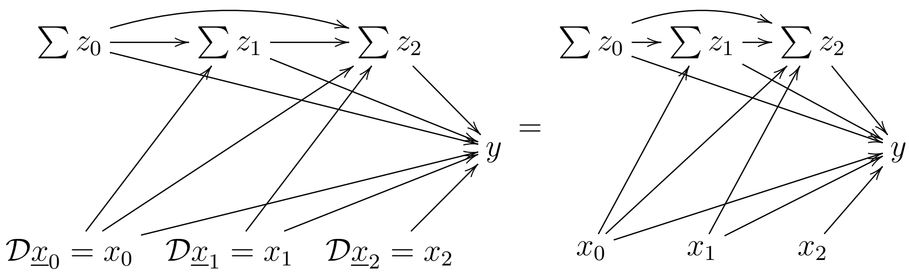
$$\mathcal{D}\underline{x}^3 = x^3 \longrightarrow y \quad = \tag{1.54}$$



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

proof:

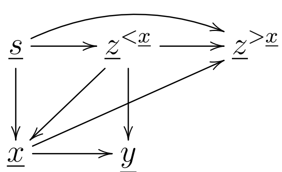
$P(y|\mathcal{D}\underline{x}^3 = x^3) = \mathcal{Q}(y|x^3)$



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

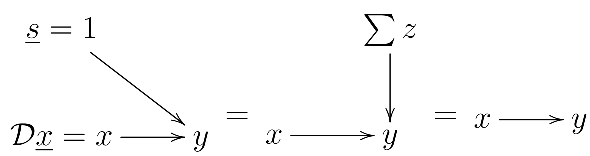
QED

Claim 14 *(Selection Bias (SB) Backdoor Adjustment Formula, from Ref.[1])*

If  *where* $\underline{s} \in \{0,1\}$ *is a selection node and* $\underline{z} = (\underline{z}^{<\underline{x}}, \underline{z}^{>\underline{x}})$*,*

then

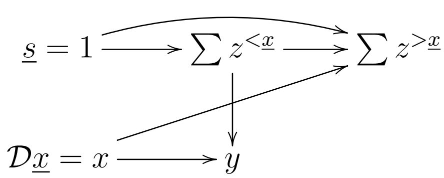
$$P(y|\mathcal{D}\underline{x} = x, \underline{s} = 1) = \sum_z P(y|x,z)P(z) = P(y|x) \tag{1.55}$$



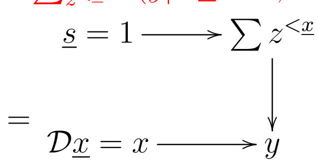
(1.56)

proof:

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



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



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

$$= \begin{array}{ccc} \underline{s}=1 & \longrightarrow & \sum z \\ & & \downarrow \\ x & \longrightarrow & y \end{array}$$

$\mathcal{D}$ can be removed because there are no sums over unobserved nodes.

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

$$= \begin{array}{ccc} & & \sum z \\ & & \downarrow \\ x & \longrightarrow & y \end{array}$$

$\underline{s} = 1$ node can be removed because this expression must equal $P(y|x, \underline{s} = 1)$. Furthermore, $\underline{y} \perp \underline{s}|(\underline{x}, \underline{z})$ in the hypothesis bnet. Hence, this expression must also equal $P(y|x)$.

QED

Bibliography

[1] Elias Bareinboim, Jin Tian, and Judea Pearl. Recovering from selection bias in causal and statistical inference. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014. `https://ftp.cs.ucla.edu/pub/stat_ser/r425.pdf`.

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

[3] 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`.

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

[5] Judea Pearl and James M Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. *arXiv preprint arXiv:1302.4977*, 2013. `https://arxiv.org/abs/1302.4977`.