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.

We use the following conventions. Random values 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 $\underline{a} \rightarrow \underline{b}$ represents the bnet with conditional probability distribution $P(b|a)$, whereas $a \rightarrow b$ represents $P(b|a)$ itself. Thus, the identity

$$P(y|x) = \sum_a P(y|a,x)P(a|x) \tag{1.1}$$

can be represented graphically by

$$\begin{tikzcd} x \arrow[r] & y \end{tikzcd} \quad = \quad \begin{tikzcd} \sum a \arrow[dr] & \\ x \arrow[u] \arrow[r] & y \end{tikzcd} \tag{1.2}$$

where $\sum a$ means the random variable $\underline{a}$ is summed over.

If $\underline{a}$ is a root node, then $\sum a$ signifies a weighted sum $\sum_a P(a)$. For example, $\sum a \rightarrow b = \sum_a P(a)P(b|a)$. If $\underline{a}$ is not a root node as in Eq.(1.2), then $\sum a$ signifies an unweighted sum $\sum_a$.

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

Claim 1 *(Backdoor Adjustment Formula)*

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

$$P(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|x,z)P(z) \tag{1.3}$$

$$\begin{tikzcd} \mathcal{D}\underline{x} = x \arrow[r] & y \end{tikzcd} \quad = \quad \begin{tikzcd} \sum z \arrow[dr] & \\ x \arrow[r] & y \end{tikzcd} \tag{1.4}$$

proof:

*** proof 1:**

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

$$\begin{tikzcd} \sum z \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \sum z \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[u, "0"] \arrow[r] & y \end{tikzcd}$$

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

$$= \begin{tikzcd} \sum z \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[u, "0"] & y \\ x \arrow[ur] & \end{tikzcd}$$

by Rule 2

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

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

by Rule 3.
No info transmission
between $\mathcal{D}\underline{x}$ and $\underline{z}$.

*** 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 $\mathcal{D}\underline{a}. = a. \leftrightarrow \underline{a}. = a.$

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

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

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

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 $\mathcal{D}\underline{a}. = a. \leftrightarrow 1$

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

.

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

$$P(y|\mathcal{D}\underline{x} = x) = \sum_m \left[\sum_{x'} P(y|x', m)P(x') \right] P(m|x) \tag{1.5}$$

$$\mathcal{D}\underline{x} = x \longrightarrow y \quad = \quad \begin{tikzcd} & \sum x' \arrow[dr] & \\ x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} \tag{1.6}$$

proof:

*** proof 1:**

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

$$\begin{tikzcd} & \sum c \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[r] & \sum m \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \mathcal{D}\underline{x} = x \arrow[r] \arrow[rr, bend right, "0"'] & \sum m \arrow[r] & y \end{tikzcd}$$

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

$$= \sum_m \begin{tikzcd} \mathcal{D}\underline{x}=x \arrow[r] \arrow[rr, bend right, "0"'] & m & y \\ & \mathcal{D}\underline{m}=m \arrow[ur] & \end{tikzcd}$$ by Rule 2

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

$$= \sum_m \begin{tikzcd} \mathcal{D}\underline{x}=x \arrow[rr, bend right, "0"'] & m & y \\ x \arrow[ur] & \mathcal{D}\underline{m}=m \arrow[ur] & \end{tikzcd}$$ by Rule 2

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

$$= \sum_m \begin{tikzcd} & m & y \\ x \arrow[ur] & \mathcal{D}\underline{m}=m \arrow[ur] & \end{tikzcd}$$

by Rule 3.
No info transmission
between $\mathcal{D}\underline{x}$ and $\underline{y}$
if condition on $\mathcal{D}\underline{m}$.

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

$$= \sum_m \begin{tikzcd} & \sum x' \arrow[dr] & \\ & m & y \\ x \arrow[ur] & \mathcal{D}\underline{m}=m \arrow[ur] \arrow[uu, bend right, "0"'] & \end{tikzcd}$$

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

$$= \sum_m \begin{tikzcd} & \sum x' \arrow[dr] & \mathcal{D}\underline{m}=m \arrow[l, "0"] \\ & m \arrow[r] & y \\ x \arrow[ur] & & \end{tikzcd}$$ by Rule 2

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

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

by Rule 3.
No info transmission
between $\mathcal{D}\underline{m}$ and $\underline{x}'$.

* **proof 2:**

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

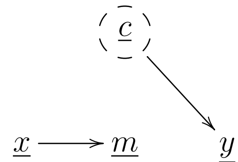
by Probability Axioms

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

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

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

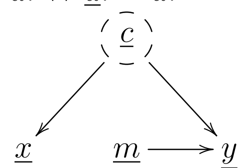


$= \sum_m P(y|\mathcal{D}\underline{x}=x, \mathcal{D}\underline{m}=m)P(m|x)$

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

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

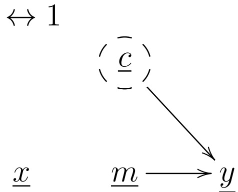


$= \sum_m P(y|\mathcal{D}\underline{m}=m)P(m|x)$

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

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



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

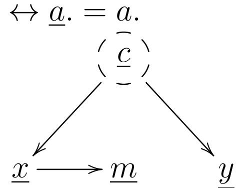
by Probability Axioms

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

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

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 $\mathcal{D}\underline{a}. = a. \leftrightarrow 1$

$$\underline{x} \perp \underline{m} \text{ in } \mathcal{D}_{\underline{m}}\mathcal{D}_{\emptyset}G: \quad \begin{tikzcd} & \underline{c} \arrow[dl] \arrow[dr] & \\ \underline{x} & \underline{m} \arrow[r] & \underline{y} \end{tikzcd}$$

.

QED

Claim 3 *(from Ref.[1])*

If

$$\begin{tikzcd} & \underline{w} \arrow[d] \arrow[ddr] \arrow[ddr, dashed, bend left, leftrightarrow, "\underline{u}"] & \\ & \underline{h} \arrow[dr] & \\ \underline{x} \arrow[uur] \arrow[rr] \arrow[dr, dashed, bend right, leftrightarrow] & & \underline{y} \\ & \underline{e} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

then

$$P(y|\mathcal{D}\underline{x} = x) = \sum_e P(y|x, e)P(e) \tag{1.7}$$

$$\mathcal{D}\underline{x} = x \longrightarrow y \quad = \quad \begin{tikzcd} x \arrow[rr] & & y \\ & \sum e \arrow[ur] & \end{tikzcd} \tag{1.8}$$

proof:

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

$$\begin{tikzcd} & \sum w \arrow[d] \arrow[ddr] \arrow[ddr, dashed, bend left, leftrightarrow, "\sum u"] & \\ & \sum h \arrow[dr] & \\ \mathcal{D}\underline{x} = x \arrow[uur] \arrow[rr] & & y \\ & \sum e \arrow[ur] & \end{tikzcd} = \begin{tikzcd} \mathcal{D}\underline{x} = x \arrow[rr] & & y \\ & \sum e \arrow[ur] & \end{tikzcd}$$

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

$$= \begin{tikzcd} \mathcal{D}\underline{x}=x \arrow[rr] \arrow[dr, "0"'] & & y \\ & \sum e \arrow[ur] & \end{tikzcd}$$

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

by Rule 2

$$= \begin{tikzcd} x \arrow[drr] & & \\ \mathcal{D}\underline{x}=x \arrow[dr, "0"'] & & y \\ & \sum e \arrow[ur] & \end{tikzcd}$$

$= \sum_e P(y|x,e)P(e)$

by Rule 3.
No information transmission between $\mathcal{D}\underline{x}$ and $\underline{e}$.

$$= \begin{tikzcd} x \arrow[rr] & & y \\ & \sum e \arrow[ur] & \end{tikzcd}$$

.
QED

Claim 4 *(from Ref.[1])*
If
$$\begin{tikzcd} & \underline{w} \arrow[dl] \arrow[d] & \underline{s} \arrow[l] \\ \underline{z} \arrow[r] & \underline{x} \arrow[r] & \underline{y} \end{tikzcd}$$
then

$$P(y|\mathcal{D}\underline{x}=x) = \sum_z P(y|x,z,w,\underline{s}=1)P(z|w,\underline{s}=1) \tag{1.9}$$

$$\mathcal{D}\underline{x}=x \longrightarrow y \quad = \quad \begin{tikzcd} & \underline{w}=w \arrow[dl] \arrow[dr] & \underline{s}=1 \arrow[dll] \arrow[d] \\ \sum z \arrow[rr, bend right] & \underline{x}=x \arrow[r] & y \end{tikzcd} \tag{1.10}$$

proof:

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

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

by Rule 1

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

$$= \begin{tikzcd} & \underline{w}=w \arrow[dl] \arrow[dr] & \underline{s}=1 \arrow[l] \arrow[dll] \arrow[d] \\ \sum z \arrow[rr, bend right] & \mathcal{D}\underline{x}=x \arrow[l, "0"'] \arrow[r] & y \end{tikzcd}$$

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

$$= \begin{tikzcd} & \underline{w}=w \arrow[dl] \arrow[dr] & \underline{s}=1 \arrow[l] \arrow[dll] \arrow[d] \\ \sum z \arrow[rr, bend right] & \mathcal{D}\underline{x}=x \arrow[l, "0"'] & y \\ & x \arrow[ur] & \end{tikzcd}$$

by Rule 2

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

$$= \begin{tikzcd} & \underline{w}=w \arrow[dl] \arrow[dr] & \underline{s}=1 \arrow[dll] \arrow[d] \\ \sum z \arrow[rr, bend right] & x \arrow[r] & y \end{tikzcd}$$

by Rule 3.
No info transmission between $\mathcal{D}\underline{x}$ and $\underline{z}$.

Claim 5 *(Napkin problem from Ref.[2])*
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

$$P(y|\mathcal{D}\underline{x}=x) = \tag{1.11}$$

proof: coming soon
QED

Claim 6 *(from Ref.[2])*

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

$$P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{z} = z) \quad = \tag{1.12}$$

proof: coming soon
QED

Bibliography

[1] Paul Hunermund and Elias Bareinboim. Causal inference and data fusion in econometrics. https://arxiv.org/abs/1912.09104, 2021.

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