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 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, $\underline{u}$.

Here is an identity that we will use frequently:

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

This identity can be represented graphically by

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

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.

Claim 1 *(Backdoor Adjustment Formula)*

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

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

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

proof:

*** proof 1:**

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

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

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

$$
=
\begin{tikzcd}
\sum z \arrow[dr] & \\
\mathcal{D}\underline{x} = x \arrow[u, "0"] & y \\
x \arrow[ur] &
\end{tikzcd}
$$
by Rule 2

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

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

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

$\underline{y} \perp \underline{x}|\underline{z}$ 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)*

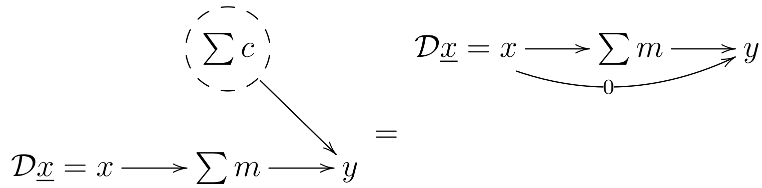
If
$$
\begin{tikzcd} & \underline{c} \arrow[dl] \arrow[dr] & \\ \underline{x} \arrow[r] & \underline{m} \arrow[r] & \underline{y} \end{tikzcd}
$$
then

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

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

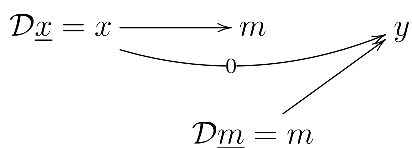
proof:
*** proof 1:**

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



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

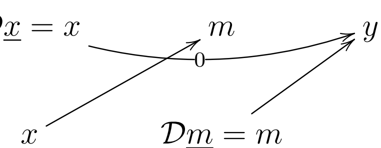
$=\sum_m$



by Rule 2

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

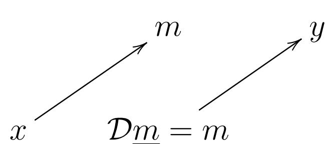
$=\sum_m$



by Rule 2

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

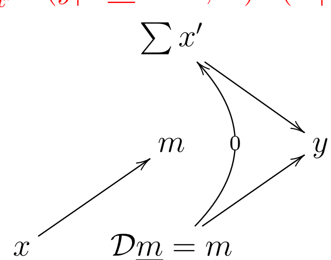
$=\sum_m$



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

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

$=\sum_m$



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

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

by Rule 2

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

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

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

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

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

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

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

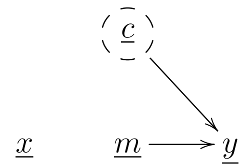
$$\begin{tikzcd} & \underline{c} \arrow[dl] \arrow[dr] & \\ \underline{x} & \underline{m} \arrow[r] & \underline{y} \end{tikzcd}$$

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

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

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



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

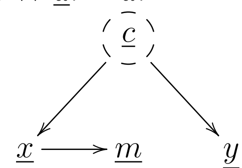
by Probability Axioms

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

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

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

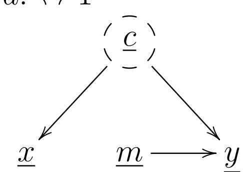


$= \sum_{x'} \sum_m P(y|m, x') P(x') P(m|x)$

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

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

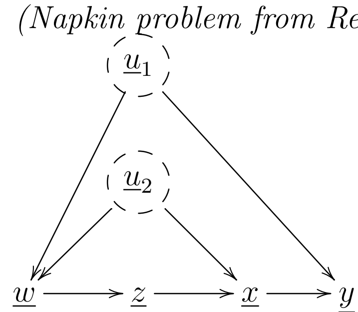
$\underline{x} \perp \underline{m}$ in $\mathcal{D}_{\underline{m}} \mathcal{D}_{\emptyset} G$:



QED

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

If



then

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

proof: coming soon

QED

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

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

$$P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{z} = z) \quad = \tag{1.8}$$

proof: coming soon
QED

Claim 5 *(from Ref.[2])*
If
$$\begin{tikzcd}
\underline{s} \arrow[r] & \underline{z} \arrow[dl] \arrow[dr] \arrow[dl, dashed, bend right] & \\
\underline{x} \arrow[rr] \arrow[rr, dashed, bend right, leftrightarrow] & & \underline{y}
\end{tikzcd}$$
then

$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|\mathcal{D}\underline{x} = x, z)P^*(z) \tag{1.9}$$

$$\begin{tikzcd}
\underline{s} = 1 \arrow[drr] & & \\
\mathcal{D}\underline{x} = x \arrow[rr] & & y
\end{tikzcd}
=
\begin{tikzcd}
\underline{s} = 1 \arrow[r] & \sum z \arrow[dr] & \\
\mathcal{D}\underline{x} = x \arrow[rr] & & y
\end{tikzcd} \tag{1.10}$$

where $P(\cdot) = P(\cdot|\underline{s} = 0)$*, and* $P^*(\cdot) = P(\cdot|\underline{s} = 1)$*. We won't draw an implicit root node* $\underline{s} = 0$ *with arrows pointing to all nodes not pointed to by* $\underline{s} = 1$*.*

proof:

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

$$\begin{tikzcd}
\underline{s} = 1 \arrow[r] & \sum z \arrow[dr] & \\
\mathcal{D}\underline{x} = x \arrow[rr] & & y
\end{tikzcd}
=
\begin{tikzcd}
\underline{s} = 1 \arrow[r] \arrow[drr, "0"] & \sum z \arrow[dr] & \\
\mathcal{D}\underline{x} = x \arrow[rr] \arrow[ur, "0"] & & y
\end{tikzcd}$$

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

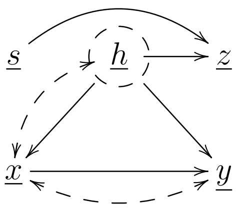
$$=
\begin{tikzcd}
\underline{s} = 1 \arrow[r] \arrow[drr, "0"] & \sum z \arrow[dr] & \\
\mathcal{D}\underline{x} = x \arrow[rr] & & y
\end{tikzcd}$$

$$= \textstyle\sum_z P(y|\mathcal{D}\underline{x} = x, z, \underline{s} = 0)P(z|\underline{s} = 1)$$

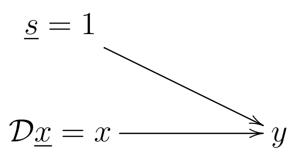
$$=
\begin{tikzcd}
\underline{s} = 1 \arrow[r] & \sum z \arrow[dr] & \\
\mathcal{D}\underline{x} = x \arrow[rr] & & y
\end{tikzcd}$$

QED

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

If  *then*

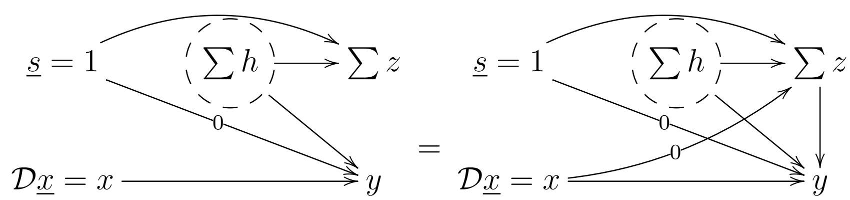
$$P^*(y|\mathcal{D}\underline{x} = x) = P(y|\mathcal{D}\underline{x} = x) \tag{1.11}$$

 $= \mathcal{D}\underline{x} = x \longrightarrow y$ (1.12)

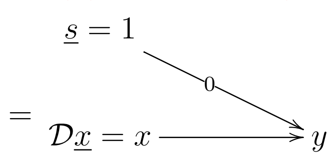
where $P(\cdot) = P(\cdot|\underline{s} = 0)$ *and* $P^*(\cdot) = P(\cdot|\underline{s} = 1)$*. We won't draw an implicit root node* $\underline{s} = 0$ *with arrows pointing to all nodes not pointed to by* $\underline{s} = 1$*.*

proof:

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



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



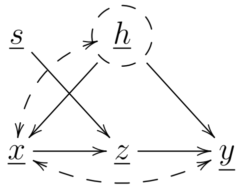
$= P(y|\mathcal{D}\underline{x} = x, \underline{s} = 0)$

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

QED

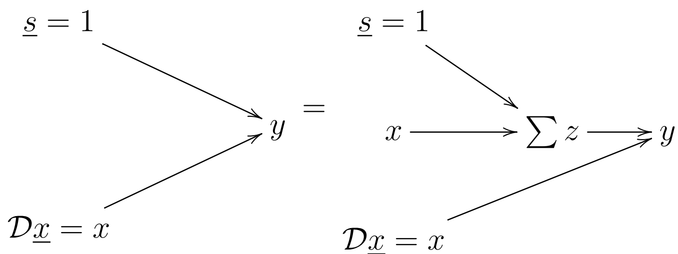
Claim 7 *(from Ref.[2])*

If



then

$$P^*(y|\mathcal{D}\underline{x}=x)=\sum_z P(y|\mathcal{D}\underline{x}=x,z)P^*(z|x) \tag{1.13}$$

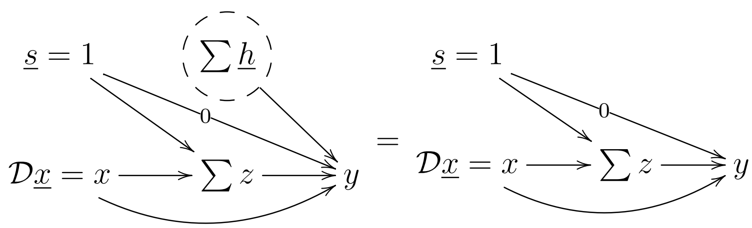


(1.14)

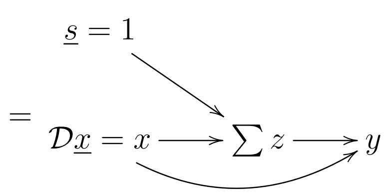
where $P(\cdot)=P(\cdot|\underline{s}=0)$ *and* $P^*(\cdot)=P(\cdot|\underline{s}=1)$. *We won't draw an implicit root node* $\underline{s}=0$ *with arrows pointing to all nodes not pointed to by* $\underline{s}=1$.

proof:

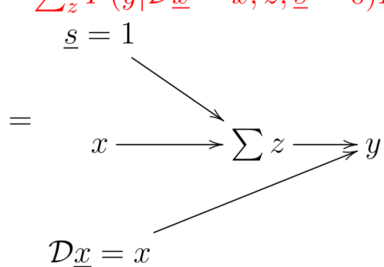
$$P(y|\mathcal{D}\underline{x}=x,\underline{s}=1)=\textstyle\sum_z P(y|\mathcal{D}\underline{x}=x,z,\underline{s}=1)P(z|\mathcal{D}\underline{x}=x,\underline{s}=1)$$



$$=\textstyle\sum_z P(y|\mathcal{D}\underline{x}=x,z,\underline{s}=0)P(z|\mathcal{D}\underline{x}=x,\underline{s}=1)$$



$$=\textstyle\sum_z P(y|\mathcal{D}\underline{x}=x,z,\underline{s}=0)P(z|x,\underline{s}=1)$$



QED

Claim 8 *(from Ref.[2])*
If
$$\begin{tikzcd} \underline{s} \arrow[r] & \underline{z} \arrow[r] \arrow[d] & \underline{a} \arrow[d] \\ & \underline{x} \arrow[r] \arrow[r, bend right, dashed, leftrightarrow] & \underline{y} \end{tikzcd}$$
then

$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_a P(y|\mathcal{D}\underline{x} = x, a)P^*(a) \tag{1.15}$$

$$\begin{tikzcd} \underline{s}=1 \arrow[dr] & \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \underline{s}=1 \arrow[r] & \sum a \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} \tag{1.16}$$

proof:

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

$$\begin{tikzcd} \underline{s}=1 \arrow[r] \arrow[drr] & \sum z \arrow[r] & \sum a \arrow[d] \\ & \mathcal{D}\underline{x}=x \arrow[r] & y \end{tikzcd} = \begin{tikzcd} \underline{s}=1 \arrow[r] \arrow[dr, bend right] & \sum a \arrow[d] \\ \mathcal{D}\underline{x}=x \arrow[r] \arrow[ur, "0"] & y \end{tikzcd}$$

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

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

$= \sum_a P(y|\mathcal{D}\underline{x} = x, a, \underline{s} = 0)P(a|\underline{s} = 1)$

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

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 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.