

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 $\underline{a} \rightarrow \underline{b}$ represents the bnet with conditional probability distribution $P(b|a)$, whereas $a \rightarrow b$ represents $P(b|a)$ itself.

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

Unobserved nodes are indicated by enclosing them in a dashed circle. For example, (\underline{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 $P(x|pa(x), \underline{s})$, where $P(x|pa(x), \underline{s} = 0) = P(x|pa(x))$ and $P(x|pa(x), \underline{s} = 1) = P^*(x|pa(x))$.

Here is an identity that we will use frequently:

$$P(y|x_1, x_2) = \sum_a P(y|a, x_1, x_2)P(a|x_1, x_2) . \quad (1.1)$$

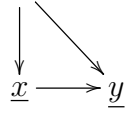
This identity can be represented graphically by

$$\begin{array}{ccc}
 x_1 & & x_1 \\
 \searrow & & \searrow \\
 & y = & \sum a \\
 \nearrow & & \nearrow \\
 x_2 & & x_2
 \end{array}
 \quad . \quad (1.2)$$

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 \underline{z} then



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

$$\begin{array}{ccc}
 \mathcal{D}\underline{x} = x \longrightarrow y & \sum z & \\
 & \searrow & \\
 & x \longrightarrow y &
 \end{array} \quad (1.4)$$

proof:

* **proof 1:**

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

$$\begin{array}{ccc}
 \sum z & & \sum z \\
 \searrow & & \searrow \\
 \mathcal{D}\underline{x} = x \longrightarrow y & = & \mathcal{D}\underline{x} = x \longrightarrow y \\
 = \sum_z P(y|x, z)P(z|\mathcal{D}\underline{x} = x) & &
 \end{array}$$

$$\begin{array}{ccc}
 \sum z & & \\
 \uparrow \circ & \searrow & \\
 \mathcal{D}\underline{x} = x & & y \\
 \nearrow & & \\
 x & & \\
 = \sum_z P(y|x, z)P(z) & & \text{by Rule 2}
 \end{array}$$

$$\begin{array}{ccc}
\sum z & & \text{by Rule 3.} \\
& \searrow & \text{No info transmission} \\
= & x \longrightarrow & y \quad \text{between } \mathcal{D}\underline{x} \text{ and } \underline{z}.
\end{array}$$

* **proof 2:**

$$\begin{aligned}
P(y|\mathcal{D}\underline{x} = x) &= \sum_z P(y|\mathcal{D}\underline{x} = x, z)P(z|\mathcal{D}\underline{x} = x) \\
&\text{by Probability Axioms} \\
&= \sum_z P(y|x, z)P(z|\mathcal{D}\underline{x} = x) \\
&\quad P(y|\mathcal{D}\underline{x} = x, z) \rightarrow P(y|x, z) \\
&\text{by Rule 2:} \quad \text{If } (\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.) \text{ in } \mathcal{L}_{\underline{a}}.\mathcal{D}_{\underline{r}}.G, \text{ then} \\
&\quad \mathcal{D}_{\underline{a}}. = a. \leftrightarrow \underline{a}. = a. \\
&\underline{y} \perp \underline{x}|\underline{z} \text{ in } \mathcal{L}_{\underline{x}}\mathcal{D}_{\emptyset}G : \begin{array}{ccc} \underline{z} & & \\ \downarrow & \searrow & \\ \underline{x} & & \underline{y} \end{array} \\
&= \sum_z P(y|x, z)P(z) \\
&\quad P(z|\mathcal{D}\underline{x} = x) \rightarrow P(z) \\
&\text{by Rule 3:} \quad \text{If } (\underline{b}. \perp \underline{a}.|\underline{r}., \underline{s}.) \text{ in } \mathcal{D}_{\underline{a}. - an(\underline{s}.)}\mathcal{D}_{\underline{r}}.G, \text{ then} \\
&\quad \mathcal{D}_{\underline{a}}. = a. \leftrightarrow 1 \\
&\underline{z} \perp \underline{x} \text{ in } \mathcal{D}_{\underline{x}}\mathcal{D}_{\emptyset}G : \begin{array}{ccc} \underline{z} & & \\ & \searrow & \\ \underline{x} & \longrightarrow & \underline{y} \end{array}
\end{aligned}$$

QED

Claim 2 (*Frontdoor Adjustment Formula*)

$$\begin{array}{ccc}
\text{If} & \begin{array}{c} (\underline{c}) \\ \swarrow \quad \searrow \\ \underline{x} \longrightarrow \underline{m} \longrightarrow \underline{y} \end{array} & \text{then}
\end{array}$$

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

$$\begin{aligned}
\mathcal{D}\underline{x} = x \longrightarrow y & \quad \sum x' \\
& = x \longrightarrow \sum m \longrightarrow y
\end{aligned} \quad (1.6)$$

proof:

* proof 1:

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

$$\begin{array}{c} \text{Diagram: } \sum c \text{ (dashed circle) points to } y. \quad \mathcal{D}\underline{x} = x \xrightarrow{\quad} \sum m \xrightarrow{\quad} y \\ \text{with a curved arrow from } \mathcal{D}\underline{x} = x \text{ to } y \text{ labeled } 0. \end{array}$$

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

$$\begin{aligned} &= \sum_m \text{Diagram: } \mathcal{D}\underline{x} = x \xrightarrow{\quad} m \xrightarrow{\quad} y, \text{ with a curved arrow from } \mathcal{D}\underline{x} = x \text{ to } y \text{ labeled } 0. \text{ by Rule 2} \\ &= \sum_m P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{m} = m)P(m|x) \end{aligned}$$

$$\begin{aligned} &= \sum_m \text{Diagram: } \mathcal{D}\underline{x} = x \xrightarrow{\quad} m \xrightarrow{\quad} y, \text{ with a curved arrow from } \mathcal{D}\underline{x} = x \text{ to } y \text{ labeled } 0. \text{ by Rule 2} \\ &= \sum_m P(y|\mathcal{D}\underline{m} = m)P(m|x) \end{aligned}$$

by Rule 3.

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

$$\begin{aligned} &= \sum_m \text{Diagram: } x \xrightarrow{\quad} m \xrightarrow{\quad} y, \text{ with a curved arrow from } x \text{ to } y \text{ labeled } 0. \\ &= \sum_{m, x'} P(y|\mathcal{D}\underline{m} = m, x')P(x'|\mathcal{D}\underline{m} = m)P(m|x) \end{aligned}$$

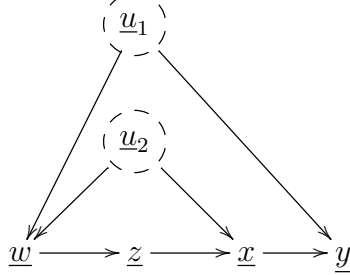
$$\begin{aligned} &= \sum_m \text{Diagram: } x \xrightarrow{\quad} m \xrightarrow{\quad} y, \text{ with a curved arrow from } x \text{ to } y \text{ labeled } 0. \\ &= \sum_{m, x'} P(y|m, x')P(x'|\mathcal{D}\underline{m} = m)P(m|x) \end{aligned}$$

$$\begin{aligned}
& \underline{y} \perp \underline{x} | \underline{m} \text{ in } \mathcal{D}_{\underline{x}} \mathcal{D}_{\underline{m}} G : \quad \begin{array}{c} (\underline{c}) \\ \searrow \\ \underline{x} \quad \underline{m} \longrightarrow \underline{y} \end{array} \\
&= \sum_{x'} \sum_m P(y | \mathcal{D}_{\underline{m}} = m, x') P(x' | \mathcal{D}_{\underline{m}} = m) P(m | x) \\
&\quad \text{by Probability Axioms} \\
&= \sum_{x'} \sum_m P(y | m, x') P(x' | \mathcal{D}_{\underline{m}} = m) P(m | x) \\
&\quad P(y | \mathcal{D}_{\underline{m}} = m, x') \rightarrow P(y | m, x') \\
&\quad \text{by Rule 2: If } (\underline{b} \perp \underline{a} | \underline{r}, \underline{s}) \text{ in } \mathcal{L}_{\underline{a}} \mathcal{D}_{\underline{r}} G, \text{ then} \\
&\quad \quad \mathcal{D}_{\underline{a}} = a. \leftrightarrow a. = a. \\
& \underline{y} \perp \underline{m} | \underline{x} \text{ in } \mathcal{L}_{\underline{m}} \mathcal{D}_{\emptyset} G : \quad \begin{array}{c} (\underline{c}) \\ \swarrow \quad \searrow \\ \underline{x} \longrightarrow \underline{m} \quad \underline{y} \end{array} \\
&= \sum_{x'} \sum_m P(y | m, x') P(x') P(m | x) \\
&\quad P(x' | \mathcal{D}_{\underline{m}} = m) \rightarrow P(x') \\
&\quad \text{by Rule 3: If } (\underline{b} \perp \underline{a} | \underline{r}, \underline{s}) \text{ in } \mathcal{D}_{\underline{a} - an(\underline{s})} \mathcal{D}_{\underline{r}} G, \text{ then} \\
&\quad \quad \mathcal{D}_{\underline{a}} = a. \leftrightarrow 1 \\
& \underline{x} \perp \underline{m} \text{ in } \mathcal{D}_{\underline{m}} \mathcal{D}_{\emptyset} G : \quad \begin{array}{c} (\underline{c}) \\ \swarrow \quad \searrow \\ \underline{x} \quad \underline{m} \longrightarrow \underline{y} \end{array}
\end{aligned}$$

QED

Claim 3 (*Napkin problem from Ref.[2]*)

If then

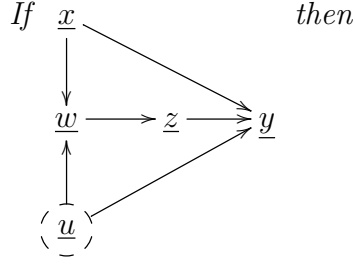


$$P(y | \mathcal{D}_{\underline{x}} = x) = \tag{1.7}$$

proof: coming soon

QED

Claim 4 (*from Ref.[2]*)



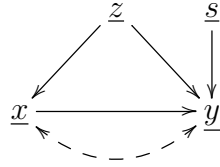
$$P(y|\mathcal{D}\underline{x} = x, \mathcal{D}\underline{z} = z) = \quad (1.8)$$

proof: coming soon

QED

Claim 5 (*Trivial Transportability, from Ref.[1]*)

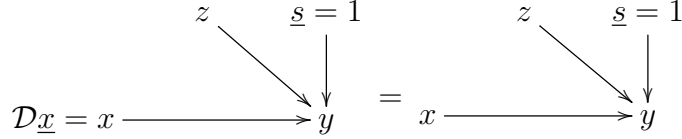
If $\underline{s} \in \{0, 1\}$ is a selection node, then



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

proof:

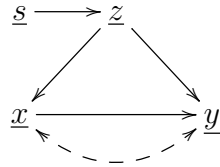
$$P(y|\mathcal{D}\underline{x} = x, z, \underline{s} = 1) = P(y|x, z, \underline{s} = 1)$$



QED

Claim 6 (*Direct Transportability, a.k.a. External Validity, from Ref.[1]*)

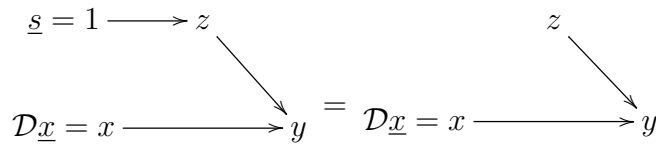
If $\underline{s} \in \{0, 1\}$ is a selection node, then



$$P^*(y|\mathcal{D}\underline{x} = x, z) = P(y|\mathcal{D}\underline{x} = x, z) \quad (1.10)$$

proof:

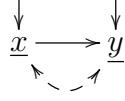
$$P(y|\mathcal{D}\underline{x} = x, z, \underline{s} = 1) = P(y|\mathcal{D}\underline{x} = x, z, \underline{s} = 0)$$



QED

Claim 7 (*S-Admisssible Transportability, from Ref.[1]*)

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



$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_a P(y|\mathcal{D}\underline{x} = x, a)P^*(a) \quad (1.11)$$

$$\begin{array}{ccc} \underline{s} = 1 & & \underline{s} = 1 \longrightarrow \sum a \\ & \searrow & \downarrow \\ \mathcal{D}\underline{x} = x \longrightarrow y & = & \mathcal{D}\underline{x} = x \longrightarrow y \end{array} \quad (1.12)$$

proof:

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

$$\begin{array}{ccc} \underline{s} = 1 \longrightarrow \sum z \longrightarrow \sum a & & \underline{s} = 1 \longrightarrow \sum a \\ & \searrow & \swarrow \quad \searrow \\ \mathcal{D}\underline{x} = x \longrightarrow y & = & \mathcal{D}\underline{x} = x \longrightarrow y \end{array}$$

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

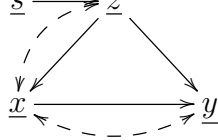
$$\begin{array}{ccc} \underline{s} = 1 \longrightarrow \sum a & & \\ & \searrow & \\ = \mathcal{D}\underline{x} = x \longrightarrow y & & \\ = \sum_a P(y|\mathcal{D}\underline{x} = x, a, \underline{s} = 0)P(a|\underline{s} = 1) & & \end{array}$$

$$\begin{array}{ccc} \underline{s} = 1 \longrightarrow \sum a & & \\ & \searrow & \\ = \mathcal{D}\underline{x} = x \longrightarrow y & & \end{array}$$

QED

Claim 8 (*from Ref.[1]*)

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



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

$$\begin{array}{ccc}
\underline{s} = 1 & & \underline{s} = 1 \longrightarrow \sum z \\
& \searrow & \searrow \\
\mathcal{D}\underline{x} = x \longrightarrow y & = & \mathcal{D}\underline{x} = x \longrightarrow y
\end{array} \tag{1.14}$$

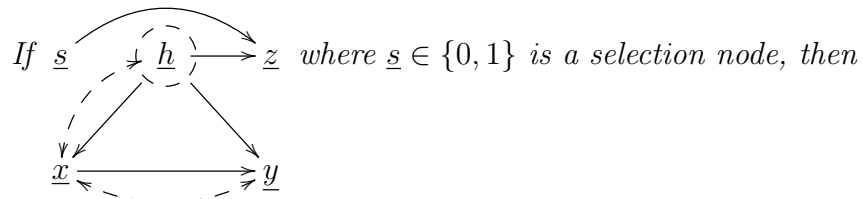
proof:

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

$$\begin{array}{ccc}
\underline{s} = 1 \longrightarrow \sum z & & \underline{s} = 1 \longrightarrow \sum z \\
& \searrow & \searrow \\
\mathcal{D}\underline{x} = x \longrightarrow y & = & \mathcal{D}\underline{x} = x \longrightarrow y \\
& \nearrow 0 \quad \nearrow 0 & \\
= \sum_z P(y|\mathcal{D}\underline{x} = x, z, \underline{s} = 1)P(z|\underline{s} = 1) & & \\
& \searrow 0 & \\
= \mathcal{D}\underline{x} = x \longrightarrow y & & \\
= \sum_z P(y|\mathcal{D}\underline{x} = x, z, \underline{s} = 0)P(z|\underline{s} = 1) & & \\
& \searrow & \\
= \mathcal{D}\underline{x} = x \longrightarrow y & &
\end{array}$$

QED

Claim 9 (from Ref.[1])

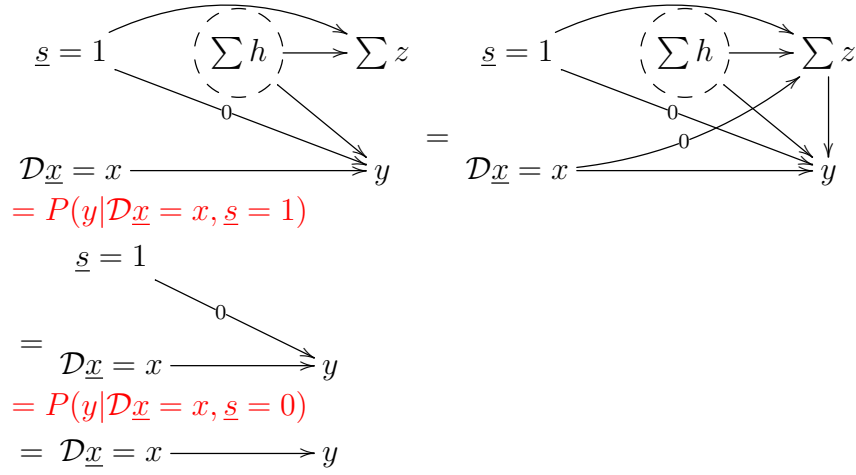


$$P^*(y|\mathcal{D}\underline{x} = x) = P(y|\mathcal{D}\underline{x} = x) \tag{1.15}$$

$$\begin{array}{ccc}
\underline{s} = 1 & & \\
& \searrow & \\
\mathcal{D}\underline{x} = x \longrightarrow y & = & \mathcal{D}\underline{x} = x \longrightarrow y
\end{array} \tag{1.16}$$

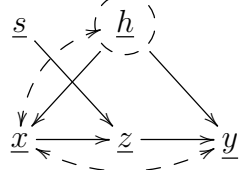
proof:

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

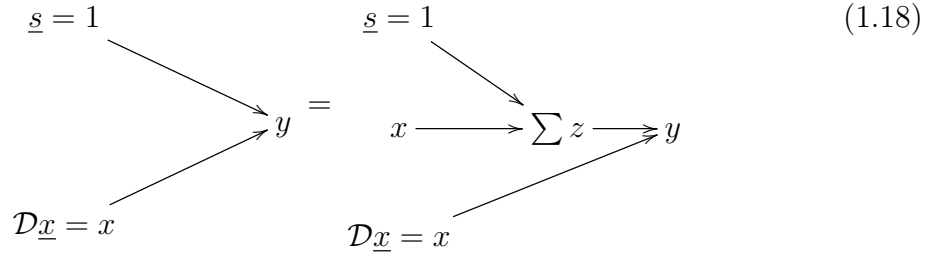


QED

Claim 10 (from Ref.[1])

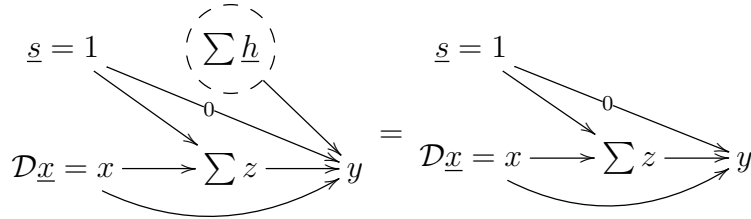
If \underline{s}  where $\underline{s} \in \{0, 1\}$ is a selection node, then

$$P^*(y|\mathcal{D}\underline{x} = x) = \sum_z P(y|\mathcal{D}\underline{x} = x, z)P^*(z|x) \quad (1.17)$$



proof:

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



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

$$= \begin{array}{c} \underline{s} = 1 \\ \searrow \\ \mathcal{D}\underline{x} = x \longrightarrow \sum z \longrightarrow y \\ \quad \quad \quad \curvearrowright \end{array}$$

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

$$= \begin{array}{c} \underline{s} = 1 \\ \searrow \\ x \longrightarrow \sum z \longrightarrow y \\ \mathcal{D}\underline{x} = x \nearrow \end{array}$$

QED

Bibliography

- [1] 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>.
- [2] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic Books, 2018.