Errata in "Elements of Causal Inference: Foundations and Learning Algorithms"

Below, you find a collection of all typos and mistakes from our book that we know of. The part in blue is correct (hopefully!). We thank all readers who kindly sent us comments to any of these typos.

Zurich, September 16, 2024
Jonas Peters, Dominik Janzing, and Bernhard Schölkopf

1 Not yet corrected in a new print

- page 51
  We write

  "only linear functions f".

  This should read

  only linear functions f_Y

- page 76
  We write

  "to which we assign the label zero.".

  This should read

  to which we assign the label one.

- page 85
  We write

  "starting from source nodes, then nodes with at most one parent and so on".

This should read

using the causal order and starting from a source node

- page 92
  We write

  "the following statements are equivalent:".

  This should read

  the following statements satisfy $(iv) \Rightarrow (i) \Leftrightarrow (ii) \Rightarrow (iii)$.

- page 107
  We write

  "the opposite of the global Markov condition".

  This should read

  the converse of the global Markov condition

- page 115
  We write

  "any node on a directed path from X to Y (except for descendants of X that are not on a directed path from X to Y)".

  This should read

  "node $V \in \mathbf{X}$ if $V \neq X$ and V lies on a directed path from X to Y".

- page 118
  We write

  "more formally, s we have".

  This should read

  more formally, we have

- page 148
  We write

  "$X - \mathbb{E}[X \mid Z] = N_X$".

  It should read:

  $$X - \mathbb{E}[X \mid Z] = N_X - \mu_{N_X}$$

- page 151
  We write

  "but also that the scoring function".

  It should read:

  but sometimes also that the scoring function

- page 171
  We write

  "if there is no hidden common cause $C \notin \mathbf{X}$ that is causing more than one variable in $\mathbf{X}$".

  This should read

  if there is no hidden common cause $C \notin \mathbf{X}$ that is causing at least two variables in $\mathbf{X}$ (and the causing paths go only through nodes that are not in $\mathbf{X}$)

- page 181
  We write (four times)

  "seperation".

  This should read (four times)

  separation

- page 181
  We write:

$$\text{“}P_{\mathbf{O},\mathbf{V}}\text{”}.$$

It should read

$$P_{\mathbf{O},\mathbf{H}}$$

- page 187
  We write

$$\begin{aligned}\text{“}&\mathbb{E}[XY|S=-1,T=-1]+\mathbb{E}[XY|S=-1,T=1]\\&+\mathbb{E}[XY|S=1,T=-1]+\mathbb{E}[XY|S=1,T=1]\leq 2\text{”}.\end{aligned}$$

  This should read

$$\begin{aligned}&\mathbb{E}[XY|S=-1,T=-1]+\mathbb{E}[XY|S=-1,T=1]\\&-\mathbb{E}[XY|S=1,T=-1]+\mathbb{E}[XY|S=1,T=1]\leq 2\text{”}\end{aligned}$$

- page 217
  We write

$$\text{“non-parametric independent test”}.$$

  This should read

$$\text{non-parametric independence test}$$

- page 188
  We write

$$\text{“}H(S\cup\{X_j,X_k\})\leq\text{”}.$$

  This should read

$$H(S)+H(S\cup\{X_j,X_k\})\leq$$

- page 226
  We write

$$\text{“To simplify notation (...) the negation of a statement.”}.$$

  This should be replaced with the content of the document ‘Correction to Proposition 6.13 in ‘Elements of Causal Inference’ ”, which is on my (Jonas Peters’) website.

2 Already corrected in a new print

- page 40
  We write

  "where $N_X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $N_Y \sim \mathcal{N}(\mu_X, \sigma_Y^2)$".

  This should read

  where $N_X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $N_Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

- page 44
  We write

  "$F_{Y|x}^{-1}(n_Y) := \inf\{x \in \mathbb{R} : F_{Y|x}(x) \geq n_Y\}$.".

  The correct definition for the inverse cdf is

  $F_{Y|x}^{-1}(n_Y) := \inf\{y \in \mathbb{R} : F_{Y|x}(y) \geq n_Y\}$.

- page 51
  We write

  " $p(x, y) = p_{N_X}(x)p_{N_Y}(y - f_Y(x))$. ".

  It should read

  $p(x, y) = p_X(x)p_{N_Y}(y - f_Y(x))$.

- page 51
  We write

  "Thus, f_X and p_{N_E}".

  It should read

  Thus, f_X and p_{N_Y}

- page 57
  We write

"$\mathbf{Y} = A\mathbf{X} + N_{\mathbf{X}}, \quad N_{\mathbf{X}} \perp\!\!\!\perp \mathbf{X}$,".

It should read

$$\mathbf{Y} = A\mathbf{X} + N_{\mathbf{Y}}, \quad N_{\mathbf{Y}} \perp\!\!\!\perp \mathbf{X},$$

- page 58
  We write

  "$A_{\mathbf{X}}$ for the model from $\mathbf{X}$ to $\mathbf{Y}$ and $A_{\mathbf{Y}}$ for the model from $\mathbf{Y}$ to $\mathbf{X}$.".

  It should read

  $A_{\mathbf{X}}$ for the model regressing $\mathbf{X}$ on $\mathbf{Y}$ and $A_{\mathbf{Y}}$ for the model regressing $\mathbf{Y}$ on $\mathbf{X}$.

- page 67
  In 4.2.2., the first inequality on page 67 reads

  "$H(X) \leq H(Y)$,".

  It should read

  $$H(X) \geq H(Y),$$

- page 69
  In Problem 4.16, part (a) reads:

  "Prove that $f(x) = \mathbb{E}[Y \mid X = x]$.".

  It should read:

  Prove that $f(x) = \mathbb{E}[Y \mid X = x] - \mu_{N_Y}$.

- page 83
  In Definition 6.1, we write:

  "neither i_k nor any of its descendants is in $\mathbf{S}$ and".

  It should read:

neither i_k nor any of its descendants is in $\mathbf{S}$, i.e.,
$(\{i_k\} \cup \mathbf{DE}_{i_k}) \cap \mathbf{S} = \emptyset$, and

(This is important for the case $\mathbf{DE}_{i_k} = \emptyset$.)

- page 84
  We write

  "An SCM $\mathfrak{C}$ defines a unique distribution over the variables $\mathbf{X} = (X_1, \ldots, X_d)$ such that $X_j = f_j(\mathbf{PA}_j, N_j)$, in distribution, for $j = 1, \ldots, d$.".

  It should read:

  An SCM $\mathfrak{C}$ defines a unique distribution over the variables $X_1, \ldots, X_d$: any $X_1, \ldots, X_d, N_1, \ldots, N_d$ satisfying $X_j = f_j(\mathbf{PA}_j, N_j)$ almost surely, where $(N_1, \ldots, N_d)$ has the desired distribution, induce the same distribution over $\mathbf{X} = (X_1, \ldots, X_d)$.

  (This is, admittedly, a less confusing formulation. Formally, we defined an SCM as a pair of structural equations and a d-dimensional noise distribution. An SCM does not include any $(X_1, \ldots, X_d, N_1, \ldots, N_d)$, which 'enter' only as a solution to the SCM. See [Bongers et al., 2016] for more details on SCMs including cycles and hidden variables.)

- page 134
  We write:

  "converges in distribution against $\mathbf{X} := (I - B)^{-1}\mathbf{N}$".

  It should read

  converges almost surely against $\mathbf{X} := (I - B)^{-1}\mathbf{N}$

- page 174
  We write:

  "We have seen that there is no solely graphical criteria for".

  It should read

We have seen that there is no solely graphical criterion for

- page 175
  We write:

  "Although A is the more effective drug, we propose to use B.".

  It should read

  Although A is the more effective treatment, we propose to use B.

- page 181
  We write:

  "that induces a distribution $P_{\mathbf{O},\mathbf{V}}$.".

  It should read

  that induces a distribution $P_{\mathbf{O},\mathbf{H}}$.

References

S. Bongers, J. Peters, B. Schölkopf, and J. M. Mooij. Structural causal models: Cycles, marginalizations, exogenous reparametrizations and reductions. *ArXiv e-prints (1611.06221)*, 2016.