**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.

København, April 8, 2021
Jonas Peters, Dominik Janzing, and Bernhard Schölkopf

# 1 Not yet corrected in a new print

- page 92
  We write

  "the following statements are equivalent:".

  This should read

  the following statements (i), (ii), and (iv) are equivalent (and each of them implies (iii)):

- page 227
  We write

  "We further have $(ii) \overset{\text{(trivial)}}{\Longrightarrow} (iii)$ and that (...) the negation of a statement.".

  This should read

  We further have $(ii) \overset{\text{(trivial)}}{\Longrightarrow} (iii)$.

## 2 Already corrected in a new print

- page 40
  We write

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

  This should read

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

- page 44
  We write

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

  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

  $$\text{“ } p(x, y) = p_{N_X}(x) p_{N_Y}(y - f_Y(x)). \text{ ”.}$$

  It should read

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

- page 51
  We write

  $$\text{“Thus, } f_X \text{ and } p_{N_E}\text{”.}$$

  It should read

  $$\text{Thus, } f_X \text{ and } p_{N_Y}$$

- 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.