

Errata, complements, and comments for  
*Algorithmic Learning in a Random World*  
(second edition, 2022)

Vladimir Vovk, Alexander Gammernan, and Glenn Shafer

January 30, 2025

## 1 Errata

- **Page 94, last line of Remark 3.15:** We should have set  $\phi(p) := \log_2 p$ , since we assumed that  $\phi$  is an increasing function (and if we do set  $\phi(p) := -\log_2 p$ , the criterion should prefer small rather than large values of (3.37)).
- **Page 142, the line following (4.38):**  $(X, Y)$  is not used and can be removed.
- **Page 302, 2nd line up from Lemma 9.14:** Remove “and any  $\epsilon > 0$ ”. When we say “p-variable” at the end of this sentence, we, of course, mean a p-variable with respect to  $P$ .
- **Page 359, 3rd line up from bottom:** “Asarin [8]” should be “Asarin [9]”.
- **Page 375, 3rd line from bottom:** “cliques” should be “maximal cliques”.
- **Page 449, reference 2:** Remove “Johnson, W. E.” from the author list and add “W. E. Johnson and” in front of the title. (This was a very bold edit made by the Springer copy editor, which the authors missed.)
- **Page 449, reference 8:** Replace “ $\delta$ -random” by “ $\Delta$ -random”.
- The  $\square$  symbol is often missing at the ends of proofs (e.g., in Propositions 2.9 and 2.15, Lemmas 2.16–2.18, and Corollary 2.20).

Some of these deficiencies (such as those on pp. 142 and 302) might not be errors from the point of view of formal logic, but they are superfluous and confusing.

## 2 Complements

- **Page xxv, section “Other Notations”:** We should have mentioned that  $\ln$  stands for natural log.
- **Page 68, first paragraph of Sect. 2.9.8:** Proposition 1 in [265] does not involve unspecified constants.
- **Page 94, the end of Remark 3.15:** The statement  $\mathcal{O}(S) = \mathcal{R}(\text{CP})$  of Theorem 3.1 can also be generalised to the criterion  $S_\phi$  preferring large values of

$$\frac{1}{k} \sum_{i=l+1}^{l+k} \sum_y \phi(p_i^y) \text{ or } \mathbb{E}_{x,\tau} \sum_y \phi(p(x, y)),$$

where  $\phi(p) := -\log_2 p$  is Greenland’s [140] S-value corresponding to a p-value of  $p$ . Namely,  $\mathcal{O}(S_\phi) = \mathcal{R}(\text{CP})$ . Indeed,

$$\begin{aligned} \sum_{y \in \mathbf{Y}} \phi(p(x, y)) &= \sum_{y \in \mathbf{Y}} \int_0^\infty 1_{\phi(p(x, y)) < a} da = \int_0^\infty \sum_{y \in \mathbf{Y}} 1_{p(x, y) > \phi^{-1}(a)} da \\ &= \int_0^\infty \left| \Gamma^{\phi^{-1}(a)}(x) \right| da = \int_1^0 |\Gamma^\epsilon(x)| \phi'(\epsilon) d\epsilon = \int_0^1 \frac{|\Gamma^\epsilon(x)|}{\epsilon \ln 2} d\epsilon, \end{aligned}$$

where  $\phi'$  is the derivative of  $\phi$ , and then the same argument applies.

- **Page 138, first line after the proof of Proposition 4.8:** We refer to “[385, proof of Theorem 1]”. That proof is unnecessarily complicated; namely, Lemma 1 in [385] (in the book’s bibliography) is easily obtained by the method of coupling [Dubhashi and Panconesi, 2009, Sect. 7.4].
- **Page 358, Sect. 11.6:** Here we should have mentioned that generalized conformal prediction was introduced in Vovk [2003].
- **Pages 359–360, Sect. 11.6.1:** The earliest description of Kolmogorov’s programme may have been given in his popular talk [Kolmogorov, 2001, pp. 135–136].
- **Page 362, Sect. 11.6.4:** Gibbs and Maxwell worked in the 3D Euclidean space, of course. In Bourbaki’s [Bourbaki, 1969, Chap. 24] notation, we consider a gas made up of  $N$  molecules of mass  $m$  at (absolute) temperature  $T$ , and  $v_1, \dots, v_N$  are their velocities. The kinetic energy of the system is

$$\frac{m}{2} \left( \|v_1\|^2 + \dots + \|v_N\|^2 \right) = 3NkT, \quad (1)$$

where  $k$  is Boltzmann’s constant. If  $T$  is given (this is our summary of the system), the components of  $v_1, \dots, v_N$  lie on the sphere (1). Gibbs’s model is that the actual velocities are chosen from the uniform distribution on that sphere. Maxwell’s law of velocities is that the velocities have a Gaussian distribution (with given variances,  $2kT/m$  for each component, which are independent).

- **Page 405, line 2:** We mention that Fisher regarded his verification protocol as somewhat unnatural but do not mention that the protocol given in [405, Sect. 9 of the arXiv report] is completely natural (since it involves testing predictions based on all the available data).
- **Page 473:** We should have included  
 Probability integral transformation, 307, 327  
 in the index.

### 3 Comments

#### 3.1 Lemma 9.6 on p. 269

Lemma 9.6 is a special case of a result on the optimality of the likelihood ratio. Let  $P$  and  $Q$  be probability measures on the same sample space, playing the role of the null and alternative hypotheses, respectively. The *e-power* of an  $e$ -variable  $E$  w.r. to  $P$  is then defined as  $\int \log E \, dQ$ .

**Lemma A.** *For given null and alternative hypotheses  $P$  and  $Q$ , respectively, such that  $Q \ll P$ , the largest  $e$ -power is attained by the likelihood ratio  $dQ/dP$ : for any  $e$ -variable  $E$ ,*

$$\int \log E \, dQ \leq \int \log \frac{dQ}{dP} \, dQ.$$

*And if  $Q \ll P$  is violated, the largest  $e$ -power is  $\infty$ .*

The likelihood ratio  $dQ/dP$  in Lemma A is understood to be the Radon–Nikodym derivative of  $Q$  w.r. to  $P$ . For a simple proof, see, e.g., [Shafer, 2021, Sect. 2.2.1] or [Vovk and Wang, 2024, Lemma 1].

Lemma 9.6 in the book is a special case of Lemma A in which  $P$  is the uniform probability measure on  $[0, 1]$  and  $\rho = dQ/dP$ .

#### 3.2 Section 9.3 on pp. 294–298

The behaviour of the test martingales and related processes in this section appears anomalous; e.g., the mean in the majority of rows in Table 9.1 lies outside the interquartile range. In particular, the final values of our test martingales have heavy-tailed distributions. The reason is that we consider multiplicative processes, in the terminology of [Nair et al., 2022, Chap. 6].

### References

Nicolas Bourbaki. *Éléments de mathématique, Livre VI, Intégration, Chapitre IX, Intégration sur les espaces topologiques séparé, Note historique*, volume 1343 of *Actualités scientifiques et industrielles*, pages 113–123. Hermann, Paris, 1969.

- Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, Cambridge, 2009.
- Andrei N. Kolmogorov. Понятие “информация” и основы теории вероятностей: Стенограмма доклада. In D. A. Pospelov and Ya. I. Fet, editors, *Колмогоров и кибернетика*, pages 118–138. ИВМиМГ СО РАН, Novosibirsk, Russia, 2001. Институт философии АН СССР, Moscow, 5 April 1965.
- Jayakrishnan Nair, Adam Wierman, and Bert Zwart. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge, Cambridge University Press, 2022.
- Glenn Shafer. The language of betting as a strategy for statistical and scientific communication (with discussion). *Journal of the Royal Statistical Society A*, 184:407–478, 2021.
- Vladimir Vovk. Well-calibrated predictions from on-line compression models. In Ricard Gavaldà, Klaus P. Jantke, and Eiji Takimoto, editors, *Proceedings of the Fourteenth International Conference on Algorithmic Learning Theory*, volume 2842 of *Lecture Notes in Artificial Intelligence*, pages 268–282, Berlin, 2003. Springer. Extended version published in *Theoretical Computer Science* (Special Issue devoted to ALT 2003) **364**, 10–26 (2006).
- Vladimir Vovk and Ruodu Wang. Nonparametric e-tests of symmetry. *New England Journal of Statistics in Data Science*, 2:261–270, 2024.