LETTER

Reply to Reshef et al.: Falsifiability or bust

The term "equitability" was introduced by Reshef et al. in ref. 1 to describe measures of statistical dependence that "give similar scores to equally noisy relationships of different types." Their paper also introduced a new statistic, the "maximal information coefficient" (MIC), that was said to satisfy this equitability criterion. There has since been much interest in MIC, due primarily to its claimed equitability (2, 3). However, neither the original paper (1) nor follow-up work (4) provided an unambiguous mathematical definition of equitability. In particular, the types of noise permissible in the noisy relationships used to define equitability were not described.

A recent paper of ours (5) critically examines the claim of ref. 1 that MIC is equitable. To do this, it was necessary to first pin down a precise mathematical definition of equitability. We therefore introduce a criterion, called "R^2-equitability," that is mathematically rigorous and follows naturally from the description of equitability given in the text and figures of ref. 1. We then prove that R^2-equitability cannot be satisfied by any dependence measure, including MIC. We conclude that a definition of equitability different from the one suggested by Reshef et al. is needed.

The present letter of Reshef et al. (6) disputes the relevance of R^2-equitability to the claims made in their original paper (1). They do not object to our specific mathematical definition. Rather, Reshef et al. now state that the claimed equitability of MIC was only intended to describe a qualitative tendency that they observed when analyzing some data that they themselves simulated. We find this objection of theirs troubling, as it implies that the central claim of ref. 1—that MIC is equitable—was never meant to be falsifiable.

Their letter also suggests that we would "toss out" the heuristic notion of equitability. The opposite is true. Our paper explicitly argues that equitability is an important concept in data analysis and deserves a proper formalization. After identifying fundamental problems with the R^2-equitability criterion, we propose replacing it with an alternative mathematical criterion called "self-equitability." Self-equitability uses the same definition of noise as R^2-equitability but, unlike R^2-equitability, it is satisfiable. In particular, self-equitability is satisfied by mutual information, a fundamental measure of dependence in information theory. MIC, however, violates self-equitability. Based on these mathematical results, as well as supporting simulations (5), we conclude that estimating mutual information (but not MIC) often provides a natural and practical way to equitably quantify associations in large datasets.

Finally, the letter of Reshef et al. offers additional simulation evidence to argue that estimates of MIC can sometimes approximately satisfy R^2-equitability better than do certain estimates of mutual information. The relevance of these select simulations is unclear. As proven in our paper, neither MIC nor mutual information satisfies R^2-equitability in any mathematical sense. The question of whether estimates of these quantities are approximately R^2-equitable is therefore neither well defined nor of obvious practical importance.

Justin B. Kinney[1] and Gurinder S. Atwal
Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

1 Reshef DN, et al. (2011) Detecting novel associations in large data sets. *Science* 334(6062):1518–1524.
2 Speed T (2011) Mathematics. A correlation for the 21st century. *Science* 334(6062):1502–1503.
3 Anonymous (2012) Finding correlations in big data. *Nat Biotechnol* 30(4):334–335.
4 Reshef DN, Reshef Y, Mitzenmacher M, Sabeti P (2013) Equitability analysis of the maximal information coefficient with comparisons. arXiv:1301.6314v1 [cs.LG].
5 Kinney JB, Atwal GS (2014) Equitability, mutual information, and the maximal information coefficient. *Proc Natl Acad Sci USA* 111(9): 3354–3359.
6 Reshef DN, Reshef YA, Mitzenmacher M, Sabeti PC (2014) Cleaning up the record on the maximal information coefficient and equitability. *Proc Natl Acad Sci USA* 111:E3362–E3363.