Reply to Murrell et al.: Noise matters

The concept of statistical "equitability" plays a central role in the 2011 paper by Reshef et al. (1). Formalizing equitability first requires formalizing the notion of a "noisy functional relationship," that is, a relationship between two real variables, X and Y, having the form

$$Y = f(X) + \eta,$$

where f is a function and η is a noise term. Whether a dependence measure satisfies equitability strongly depends on what mathematical properties the noise term η is allowed to have: the narrower one's definition of noise, the weaker the equitability criterion becomes. Unfortunately, the allowed mathematical properties of η is an issue on which the paper by Reshef et al. is silent.

Our paper (2) adopts a broad definition of noise. Essentially, we require only that

$$X \leftrightarrow f(X) \leftrightarrow Y,$$

be a Markov chain. This requirement is sensible: if this Markov chain condition is not satisfied, it would be difficult to interpret Y as being a measurement of $f(X)$ as opposed to a measurement of X itself. However, we place no other constraints on η.

Using this definition of noise, we define R^2-equitability to mean exactly the notion of equitability used by Reshef et al (1). We then prove that no nontrivial dependence measure (including the maximal information coefficient) satisfies this criterion. The same broad definition of noise is then used to define an alternative notion of equitability called "self-equitability." We show that self-equitability is closely related to the Data Processing Inequality, is satisfied by the well-known mutual information measure, and is not satisfied by the maximal information coefficient.

In their letter, Murrell et al. (3) consider the consequences of adopting a different and narrower definition of noise, one that requires that η have zero mean at all X values. Murrell et al. then show that, when using this restricted definition of noise, R^2-equitability becomes satisfiable (although not by the maximal information coefficient). We welcome this discussion: it matters how one defines noise, and the letter by Murrell et al. correctly highlights this important fact, as well as its consequences. We doubt the general utility of their alternative definition of R^2-equitability, but at the moment this is just a matter of opinion.

However, we do dispute Murrell et al. on two points. First, the title of their letter incorrectly suggests that our paper is wrong about R^2-equitability being unsatisfiable. Murrell et al. come to different conclusions about R^2-equitability only because they define it differently. It is unsurprising that changing a mathematical definition leads to different results.

We also disagree with their assertion that our definition of noise is inappropriate. First, our definition is entirely in line with standard information theory arguments (4, 5). Second, our definition of noise leads to valuable mathematical results, explicitly connecting self-equitability to well-established concepts in information theory. Finally, there is the indisputable fact that the noise in many real experiments does not have zero mean. For example, this is manifestly true in the experiments of Kinney et al. (6).

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 Kinney JB, Atwal GS (2014) Equitability, mutual information, and the maximal information coefficient. *Proc Natl Acad Sci USA* 111(9): 3354–3359.
3 Murrell B, Murrell D, Murrell H (2014) R^2-equitability is satisfiable. *Proc Natl Acad Sci USA* 111:E2160.
4 Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication* (Univ of Illinois, Urbana, IL).
5 Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
6 Kinney JB, Murugan A, Callan CG, Jr., Cox EC (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci USA* 107(20): 9158–9163.