Omri Tal ISHPSSB 2013

Is Shannon's Information Theory Applicable to Genetic Data?

Absract

Shannon famously remarked that a single concept of information could not satisfactorily account for the numerous possible applications of the general field of communication theory. I employ some basic principles from Shannon's work on information theory (Shannon 1948) to develop a measure of information for describing 'population structure' using genetic data. This sense of information is somewhat less abstract than *entropy* or *Kolmogorov Complexity* and is utility-oriented. Specifically, given a collection of genotypes sampled from known multiple populations I would like to quantify the potential for correct classification of genotypes of unknown origin. Motivated by Shannon's axiomatic approach in deriving a unique information measure for communication, I first identify a set of intuitively justifiable criteria that any such quantitative information measure should satisfy. I will show that standard information-theoretic measures such as mutual information or relative entropy cannot satisfactorily account for this sense of information, necessitating a decision-theoretic approach. I will also review very recent empirical work of biologists to assess the 'population signal' from genetic samples.