

1032-62-206

Ilya M. Nemenman* (nemenman@lanl.gov), CCS-3, MS B256, LANL, Los Alamos, NM 87545.

Entropy Estimation: Coincidences, Additivity, and Uninformative Priors. Preliminary report.

Modern biological data analysis often requires estimation of information-theoretic (entropic) quantities. Unfortunately, their estimation suffers strongly from insufficient sample size and the consequent bias. Correspondingly, development of better entropy estimators has been a hot topic lately. As a result, we now understand that, in the worst case, estimating an entropy of a variable is only marginally simpler than estimating its entire probability distribution. However, for limited classes of probability distributions, entropy estimation can be much simpler, sometimes requiring about a square-root-fewer samples than the worst case result. One particular way to achieve this improvement can be derived by re-examining standard Bayesian "uninformative" priors, relating them to coincidence counting methods (known since the 1930s as the capture-recapture technique for estimation of population sizes), and using the additivity of entropy to control the bias. In this talk, I will describe in detail this method, which we introduced in 2002. I will pay a particular attention to some of its novel developments, illustrating them on the problem of estimation of configurational entropies of lattice proteins. (Received August 21, 2007)