Bayesian Parameter Estimation for Protein Contact Potentials

Scott C. Schmidler*
Duke University, Durham, NC, USA schmidler@stat.duke.edu

Xiang Zhou
University of Chicago, Chicaco, IL, USA xz7@uchicago.edu

Protein contact potentials are Gibbs random fields widely used for problems in structural bioinformatics and computational biophysics, including protein structure prediction, protein-protein interactions, and protein-ligand docking. Parameters for these potentials are commonly extracted from structural databases using the so-called ``quasi-chemical approximation". We show that potentials obtained in this manner are inconsistent, and describe an alternative Bayesian approach based on asymptotically unbiased pseudolikelihood estimation. Empirical studies indicate that the resulting parameter estimates converge quickly to the true values for moderate sample sizes. We demonstrate this approach in a simple fold recognition experiment, our estimators significantly outperform the quasi-chemical contact potentials, achieving performance comparable to potentials specifically optimized for discrimination of native and non-native folds without requiring discriminative training or decoy sets. Our results, which apply to more general contact potentials, indicate that evolutionary models can perform competitively in prediction tasks when proper statistical inference is applied.

Key Words: Bayesian, protein structure prediction, MCMC, Markov random fields