

Inference using partial information

Jeff Miller, Harvard School of Public Health

As biomedical data sets grow in scale and diversity, there is an increasing need for fast and reliable inference methods for complex models. An interesting general approach is to use partial information---that is, rather than constructing a complete probabilistic model of all the data, one models certain well-chosen statistics (marginally or conditionally), relaxes constraints, and/or combines inferences from multiple sets of data. Using partial information can be advantageous in terms of computation and robustness, as well as ease-of-use, modularity, and division of labor. In the frequentist literature, partial information-based methods are ubiquitous, but they are under-utilized in the Bayesian setting, perhaps due to the challenges of coherently combining partial posteriors and ensuring the results are valid. It turns out that partial information can be used to yield generalized posteriors that are consistent and can be appropriately calibrated with some care. We will illustrate these issues, and our work towards solutions, via the problem of inferring the evolutionary history of cancer in a given patient, using whole-genome/exome sequences of tumor samples.