Bayesian Models of Inductive Learning

Thomas L. Griffiths (tom_griffiths@brown.edu)
Department of Cognitive and Linguistic Sciences
Brown University, Providence RI 02912 USA

Charles Kemp (ckemp@mit.edu)
Joshua B. Tenenbaum (jbt@mit.edu)
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, Cambridge MA 02139 USA

Many of the central problems of cognitive science are problems of induction, calling for uncertain inferences from limited data. How can people learn the meaning of a new word from just a few examples? What makes a set of examples more or less representative of a concept? What makes two objects seem more or less similar? Why are some generalizations apparently based on all-or-none rules while others appear to be based on gradients of similarity? How do we infer the existence of hidden causal properties or novel causal laws? This tutorial will introduce an approach to explaining these everyday inductive leaps in terms of Bayesian statistical inference, drawing upon tools from statistics (Bernardo & Smith, 1994; Gelman, Carlin, Stern, & Rubin, 1995), machine learning (Duda, Hart, & Stork, 2000; Mackay, 2003), and artificial intelligence (Pearl, 1988; Russell & Norvig, 2002).

In Bayesian models, learning and reasoning are explained as probability computations over a hypothesis space of possible concepts, word meanings, or causal laws. The structure of the learner's hypothesis space reflects their domain-specific prior knowledge, while the nature of the probability computations depends on domain-general statistical principles. Bayesian models of cognition thus pull together two approaches that have historically been kept separate, providing a way to combine structured representations and domain-specific knowledge with domain-general statistical learning.

We will demonstrate how this approach can be used to model natural tasks where people draw on considerable prior knowledge, including abstract domain theories and structured relational systems (e.g., biological taxonomies, causal networks). Formalizing aspects of these knowledge structures will be critical to specifying reasonable prior probabilities for Bayesian inference. Specifically, we will show how key principles in people's intuitive theories of natural domains can be formalized as probabilistic generative systems, generating plausible hypotheses to guide Bayesian learning and reasoning (Tenenbaum, Griffiths, & Kemp, 2006).

Bayesian inference has become an increasingly popular component of formal models of human cognition (Chater, Tenenbaum, & Yuille, 2006). This full-day tutorial aims to prepare students to use these modeling methods intelligently: to understand how they work, the advantages they offer over alternative approaches, and their limitations. The tutorial will assume minimal background in Bayesian statistics and a level of mathematical sophistication appropriate for an audience with general interests in computational modeling.

The tutorial will begin with a discussion of the how Bayesian models fit into the general project of developing formal models of cognition. We will then outline some of the basic principles of Bayesian statistics that are of relevance to modeling cognition (Griffiths & Yuille, 2006), before turning to a series of case studies illustrating these methods, contrasting multiple models both within the Bayesian approach and across different modeling approaches. Topics will include graphical models and causal induction, property induction, Monte Carlo methods, and probabilistic modeling of some basic aspects of language. Through considering these case studies, we will also discuss how to relate the abstract computations of Bayesian models to more traditional models framed in terms of cognitive processing or neurocomputational mechanisms.

References

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* New York: Wiley.

Chater, N., Tenenbaum, J. B., & Yuille, A. (Eds.) (2006). Special issue on "Probabilistic models of cognition". *Trends in Cognitive Sciences*, *10*(7).

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification.* New York: Wiley.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* New York: Chapman & Hall.

Griffiths, T. L., & Yuille, A. (2006). A primer on probabilistic inference. *Trends in Cognitive Sciences*, *10*(7).

Mackay, D. J. C. (2003). *Information theory, inference, and learning algorithms.* Cambridge: Cambridge University Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Francisco, CA: Morgan Kaufmann.

Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: A modern approach* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models for inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7).