In: Jaeger D., Jung R. (Ed.) *Encyclopedia of Computational Neuroscience*. Springer, 2013.

Bayesian Behavioral Data Analysis

Samuel J. Gershman
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Definition

Bayesian analysis is a method for reasoning probabilistically about parameters of a model given some observed data. Applied to behavioral data, Bayesian analysis can be used to fit and compare models of cognition. This approach to behavioral data analysis offers a number of advantages over classical (frequentist) analysis, including a coherent representation of uncertainty, flexibility to handle complex models and missing data, and an avoidance of pathologies inherent to significance testing based on *p*-values.

Detailed Description

Bayesian analysis starts with the specification of a joint distribution on the data *D* and parameters (or hidden variables) *H*. This joint distribution can be broken down into two components: P(*D*,*H*) = P(*D*|*H*) P(*H*). The first component, P(*D*|*H*), is known as the *likelihood*, and the second component, P(*H*), is known as the *prior.* Given some data, we are interested in the conditional distribution, P(*H*|*D*), commonly known as the *posterior*; this distribution specifies everything we know about *H* given the observed data and our prior beliefs. Bayes' rule stipulates how to form the posterior from the prior and likelihood:

$$P(H|D) = \frac{P(D|H)P(H)}{\sum_H P(D|H)P(H)}$$

The denominator is sometimes known as the *marginal likelihood.* When *H* is continuous, the summation is replaced by integration. A practical overview of Bayesian analysis can be found in Kruschke (2010). Below we discuss some of the relevant issues that arise in applying Bayesian analysis to behavioral data.

Posterior Probabilities vs. *P*-values

The classical approach to behavioral data analysis is based on the null hypothesis significance testing (NHST) framework, in which the significance of a summary statistic is evaluated by calculating the probability of obtaining an equally or more extreme value if the statistic was drawn from a null distribution. If the resulting *p*-value is below some conventional threshold (e.g., 0.05), the null hypothesis is rejected and the result is considered significant. This framework is related to Bayesian analysis: the *p*-value is the tail probability of the likelihood when *H* corresponds to the null hypothesis.

In contrast to *p*-values, posterior probabilities quantify the evidential support for all possible values of *H*. Often different values of *H* can plausibly have generated the data, and the posterior captures this uncertainty. In contrast, the confidence intervals of NHST express the range of *H* that would not be rejected by a significance test.

Under NHST, when multiple tests are applied to a data set, the probability of erroneously declaring a null effect significant increases. This necessitates some kind of correction for multiple comparisons. Bayesian analysis does not require such corrections, since there is a single posterior distribution that does not depend on how many ways you examine it.

Hierarchical Models

A powerful use of Bayesian analysis is the construction of hierarchical models in which uncertainty about the parameters of the prior is captured by placing a "hyper-prior" on these parameters (Lee 2011). An important example of this is modeling individual differences: each subject in a study has a separate set of parameters, but these parameters are coupled by virtue of being drawn from a common group-level distribution. One advantage of hierarchical models is that they allow sharing of "statistical strength" between subjects, such that the parameter estimates of one subject are informative about the parameter estimates of another subject. At the same time, subjects are allowed to express idiosyncratic parameters. Hierarchical models can thus strike a balance between forcing all subjects to have the same parameters and modeling each subject separately.

Model Comparison

An important problem in cognitive modeling is model comparison. The Bayesian approach to this problem centers on calculating the *Bayes factor* (Kass & Raftery 1995), which is the ratio of marginal likelihoods for two models. The Bayes factor rewards models that fit the data better, but penalizes model complexity. Intuitively, complex models are able to fit a given data set better than simple models, but complex models also spread their probability mass over many data sets, and consequently the probability of any given data set will tend to be smaller.

Whereas the posterior will often be relatively insensitive to the prior when the amount of data is large, Bayes factors can be exquisitely sensitive to the choice of prior (Kass & Raftery 1995).

This has sometimes been considered a drawback of Bayesian model comparison. However, others (e.g., Vanpaemel 2010) have argued that prior sensitivity is reasonable, since the prior should be considered part of a fully specified cognitive model.

Computation

For all but a small class of models, employing Bayes' rule exactly is computationally intractable. This happens because the marginal likelihood computation involves summing over all possible hypotheses, which is infeasible for large hypothesis spaces. Consequently, most algorithms for Bayesian analysis use approximations. Most commonly, these algorithms approximate the posterior with a set of stochastic samples—so-called *Monte Carlo* approximations (Robert & Casella 2004). As the number of samples increase, the approximation becomes increasingly accurate. There are several software packages that compute Monte Carlo approximations automatically given a probabilistic model (see Kruschke 2010 for an introduction), making Bayesian methods widely accessible for behavioral data analysis.

References

Kass RE, Raftery AE (1995) Bayes factors. J Amer Stat Association 90:773-795

Kruschke JK (2010) Doing Bayesian data analysis: A tutorial introduction with R and BUGS. Academic Press, Burlington

Lee MD (2011) How cognitive modeling can benefit from hierarchical Bayesian models. J Math Psych 55:1-7

Robert CP, Casella G (2004) Monte Carlo Statistical Methods. Springer, New York

Vanpaemel W (2010) Prior sensitivity in theory testing: An apologia for the Bayes factor. J Math Psych 54:491-498