Bayesian Statistics in a Nutshell

Intended for students of psychological and/or organizational sciences

Andrew Jebb and Sang Eun Woo (August 27, 2014)
http://www1.psych.purdue.edu/~sewoo/bayesian

Bayesian statistics is an approach to statistical inference (i.e., drawing conclusions about the population through sample data) that is different than the conventional *null hypothesis significance testing (NHST)* approach. Bayesian methods derive their name from *Bayes' Theorem*, a mathematical equation built off of simple probability axioms. In essence, **it allows an analyst to calculate any conditional probability of interest**.

A conditional probability is simply the probability of event *A* given that event *B* has occurred. It is a probability that is therefore "conditional" on another event. Statistical analyses are based on a collection of sample data.

Therefore, in probability terms, we know that the data have already occurred. Using Bayes' theorem, we can directly calculate the probability of various things of interest given, or conditional on, this already observed data. It is a highly intuitive approach to statistical inference that **allows for *direct* probability statements regarding things researchers are interested in, such as population values or statistical models.**

BAYESIAN VS. NULL HYPOTHESIS SIGNIFICANCE TESTING

Both Bayesian statistics and null hypothesis significance testing have their unique approaches. The basis of these differences lies in their philosophical differences about how probability should be conceived. However, **one does not have to subscribe to the Bayesian or null hypothesis significance testing notion of probability to use these statistics in practice**.

Null hypothesis significance testing involves formally testing competing specific statistical hypotheses — hypotheses related to the value of a parameter (e.g., the population value is zero or not).

NHST estimation tries to locate a *single* parameter estimate that best fits the data. It is possible to then provide a range of plausible values around that point estimate through the use of a *confidence interval*, which indicates the *precision* of the estimate.

In contrast, Bayesian statistical inference is done primarily through estimation. Each estimated parameter is represented in a probability distribution, where each potential parameter value is probabilistically weighted, allowing the analyst to see how probable each potential parameter value is.