Toward Fully Bayesian Computing: Manipulating and Summarizing Posterior Simulations Using Random Variable Objects

Jouni Kerman
Statistical Methodology Group
Novartis Pharma AG
Basel, Switzerland
jouni.kerman@novartis.com

Andrew Gelman
Department of Statistics
Department of Political Science
Columbia University
New York, USA
gelman@stat.columbia.edu

March 28, 2008

Abstract

Bayesian data analysis involves Bayesian inference (model fitting), but also requires post-fitting tasks that include summarizing and manipulating inferences numerically and graphically, and doing model-checking tasks and forecasting using predictive inference. Since Bayesian inference is based on computing and summarizing probability distributions, to do Bayesian data analysis efficiently and conveniently, we need a computing environment that enables us to work with random variables as easily as we do with numerical variables.

We propose a computing environment that defines random variables as natural extensions of traditional numerical objects, which can be regarded as random variables with zero variance. Each numeric vector variable in this environment has a hidden dimension of uncertainty, which is represented by a number of simulation draws from the joint distribution of its components. The random variables can be manipulated transparently, in the same fashion as we do numeric vectors and arrays.

We present an R package, 'rv', that implements this new computing paradigm in R by introducing a new simulation-based random variable class, along with numerous mathematical, statistical, and graphical functions. By converting posterior simulations into random variable objects, they can be manipulated and summarized intuitively and efficiently. We illustrate this by several practical examples.