

ReBaStaBa: handling Bayesian networks with R

Jean-Baptiste Denis^{1,*}, Marie-Laure Delignette-Muller^{2,3,4}, Régis Pouillot⁵

1. Mathématiques et Informatique Appliquées - INRA - Jouy

2. University of Lyon, Lyon, France

3. CNRS UMR5558, Villeurbanne, France

4. National veterinary school of Lyon, Marcy l'Etoile, France

5. 4515 Willard Ave., Chevy Chase, MD 20815, U.S.

* Contact author: Jean-Baptiste.Denis@Jouy.Inra.Fr

Keywords: Bayesian network, R

Bayesian networks [BN] are an increasing used tool in many applications. There are several strong reasons for such a success: (1) the use of directed acyclic graph to define the structure of a BN is an attractive, efficient and easy way of formalization, (2) it exists nowadays powerful and convenient softwares to apply BN in real applications. Surprisingly enough for a statistician, BN have been mainly promoted by scientists of artificial intelligence and the main softwares dealing with BN in a statistical perspective [Plummer, 2009] do not mention them as such and do not propose specific outputs related to their underlying existence.

For some studies performed in food borne disease assessment or in human physiology, we constructed BN involving categorical and continuous variables (Figure 1). If some were attainable with OpenBugs or Jags, those based in empirical distributions required a direct programming, that we did with R. Soon, rather than doing it specifically, we undertook the writing of a collection of generic R functions, under the name of ReBaStaBa (*REseaux BAYésiens traités par STATistique BAYésienne*) [Denis, 2009]; it could become one day an R package.

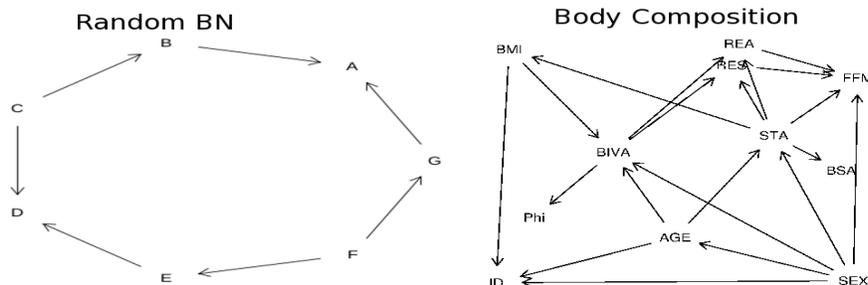


Figure 1: BN produced with rebastaba (Body Composition is a joint work with L. Mioche, Inra).

The aim of rebastaba is not to replace existing tools but (i) to give the possibility of handling (under the R environment) very general BNs, (ii) to provide from them useful outputs about their properties, (iii) to use them in a basic way and (iv) to offer interfaces with other applications (like deal, jags, grappa). Even if rebastaba is still evolving according to the points found when using it, the main concepts are stabilized and the necessary functions to facilitate inputs and to produce outputs are available.

In introduction, a distinction will be made between BN (basically considered as defining a joint probability distribution for a set of random variables) and Bayesian statistics (used to extract by means of the Bayes' theorem information from data). Then based on some cases, the main possibilities of rebastaba will be exemplified. Finally some hints will be given on the retained choices and the main different S4 classes introduced to answer the challenge.

References

J.-B. Denis (2009). jbd tools,

<http://w3.jouy.inra.fr/unites/miaj/public/matrixq/jbdenis/outils/welcome.html>.

M. Plummer (2009). Jags, Just Another Gibbs Sampler,

<http://www-fis.iarc.fr/~martyn/software/jags/>.