Journal of Statistical Planning and Inference 139 (2009) 4154–4164

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Bayesian model comparison based on expected posterior priors for discrete decomposable graphical models

Guido Consonni[a,*], Monia Lupparelli[b]

[a] *Dipartimento di Economia Politica e Metodi Quantitativi, University of Pavia, Italy*
[b] *Dipartimento di Scienze Statistiche, University of Bologna, Italy*

ARTICLE INFO

Article history:
Received 17 December 2008
Received in revised form
28 May 2009
Accepted 31 May 2009
Available online 9 June 2009

Keywords:
Bayes factor
Clique
Conjugate family
Contingency table
Decomposable model
Imaginary data
Importance sampling
Intrinsic prior
Robustness
Training sample

ABSTRACT

The implementation of the Bayesian paradigm to model comparison can be problematic. In particular, prior distributions on the parameter space of each candidate model require special care. While it is well known that improper priors cannot be routinely used for Bayesian model comparison, we claim that also the use of proper conventional priors under each model should be regarded as suspicious, especially when comparing models having different dimensions. The basic idea is that priors should not be assigned separately under each model; rather they should be related across models, in order to acquire some degree of compatibility, and thus allow fairer and more robust comparisons. In this connection, the intrinsic prior as well as the expected posterior prior (EPP) methodology represent a useful tool. In this paper we develop a procedure based on EPP to perform Bayesian model comparison for discrete undirected decomposable graphical models, although our method could be adapted to deal also with directed acyclic graph models. We present two possible approaches. One based on imaginary data, and one which makes use of a limited number of actual data. The methodology is illustrated through the analysis of a $2 \times 3 \times 4$ contingency table.

1. Introduction

Model comparison is an important area of statistics. The Bayesian view is especially suited for this purpose, see for instance the review articles by George (2005) and Berger (2005). However, its implementation can be problematic, especially when comparing models having different dimensions. In particular, prior distributions on the parameter space of each model, which are required to compute Bayes factors and posterior model probabilities, need special care, because sensitivity to prior specifications in Bayesian testing and model comparison is more critical than in Bayesian inference within a single model. In particular, the use of conventional priors is suspicious for model comparison. The problem goes much deeper than the simple realization that improper priors cannot be naively used for computing Bayes factors, because arbitrary normalizing constants do not cancel out. Indeed also proper priors are not free from difficulties when comparing hypotheses of different dimensions, as witnessed by the celebrated Jeffreys–Lindley paradox (see e.g. Robert, 2001, p. 234). The main difficulty stems from the high sensitivity of Bayes factors to the specifications of hyperparameters controlling prior-diffuseness. We claim that, when dealing simultaneously with several models, one cannot elicit priors in isolation conditionally on each single model; rather, one should take a global view and *relate* priors across models. This leads us straight into the area of *compatible* priors, see e.g. Dawid and Lauritzen (2001) and Consonni and Veronese (2008). In this connection, the intrinsic prior (IP) methodology—Berger and Pericchi (1996) and Moreno (1997)—and the expected posterior prior (EPP) methodology—Pérez and Berger (2002)—represent a useful tool. The IP and EPP

* Corresponding author.
E-mail addresses: guido.consonni@unipv.it (G. Consonni), monia.lupparelli@unibo.it (M. Lupparelli).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.05.045

methodologies, which are closely related, were both motivated by the need to use objective, typically improper, priors for model choice. However, they have a wider scope, because they can effectively deal with issues such as compatibility of priors and robustness of Bayes factors to prior elicitation.

Additionally, they embody a natural tuning coefficient, the training sample size, which represents a valuable communication device to report a range of plausible values for the Bayes factor (or posterior probability) in the light of the data; see Consonni and La Rocca (2008) for an application. The EPP methodology is somewhat more flexible in the choice of a particular mixing distribution (see Section 4.1 for details), and thus we will henceforth embed our discussion within the EPP framework.

In this paper we perform Bayesian model determination for discrete decomposable (undirected) graphical models using the EPP methodology. Specifically, Section 2 contains background material on graphical models and notation; Section 3 presents useful results originally developed by Consonni and Massam (2007): an efficient parameterization of discrete decomposable graphical models, a class of conjugate priors, as well as a reference prior. Sections 4 and 5, with their specific focus on discrete graphical models, constitute the innovative part of the paper: the former develops a 'base-model', as well as an 'empirical distribution', version of expected posterior prior; while the latter presents an EPP-based Bayesian model comparison methodology. Section 6 applies the methodology to a $2 \times 3 \times 4$ contingency table representing the classification of 491 subjects according to three categorical variables, namely hypertension, obesity, and alcohol intake, with the objective of identifying the most promising models for the explanation of these data. Finally, Section 7 presents some concluding remarks.

2. Background and notation

We briefly recall some basic facts about undirected graphical models. Let V be a finite set of vertices; and define E to be a subset of $V \times V$ containing unordered pairs $\{\gamma, \delta\}$, $\gamma \in V$, $\delta \in V$, $\gamma \neq \delta$. An undirected graph G is the pair (V, E). An undirected graph is *complete* if all pairs of vertices are joined by an edge. For further details, and in particular the notions of *decomposable* graph and *clique* we refer to Lauritzen (1996).

For a given ordering $C_1, ..., C_k$ of the cliques of a decomposable undirected graph G, we will use the following notation:

$$H_l = \bigcup_{j=1}^{l} C_j, \quad l = 1, \ldots, k, \quad S_l = H_{l-1} \cap C_l, \quad l = 2, \ldots, k, \quad R_l = C_l \backslash S_l, \quad l = 2, \ldots, k.$$

The set H_l is called the l-th *history*, S_l the l-th *separator* and R_l the l-th *residual*. The ordered sequence of the cliques is said to be *perfect* if for any $l > 1$ there is an $i < l$ such that $S_l \subseteq C_i$.

Given a random vector $A = (A_\gamma, \gamma \in V)$, a *graphical model*, Markov with respect to an undirected graph G, is a family of joint probability distributions on A such that $A_\delta \perp\!\!\!\perp A_\gamma \mid A_{V \backslash \{\delta, \gamma\}}$, for any pair $\{\delta, \gamma\} \notin E$. We assume A to be a discrete random vector, with each element A_γ taking values in the finite set $\mathscr{I}_\gamma$. For a given undirected decomposable graph G, we use for simplicity the same symbol G also to denote a discrete graphical model, Markov with respect to the graph G.

The Cartesian product $\mathscr{I} = \times_{\gamma \in V} \mathscr{I}_\gamma$ defines a table whose generic element

$$i = (i_\gamma, \gamma \in V)$$

is called a *cell* of the table. Consider N units, and assume that each one can be classified into one and only one of the $|\mathscr{I}|$ cells. Let $y(i)$ be the i-th cell-count; then the collection of cell-counts

$$y = (y(i), i \in \mathscr{I}), \quad \sum_{i \in \mathscr{I}} y(i) = N,$$

defines a contingency table. Conditionally on the probability $p(i)$ that a randomly chosen unit belongs to cell $i \in \mathscr{I}$, y is distributed according to a multinomial model $\mathscr{M}u(y|p, N)$, with

$$p = (p(i), i \in \mathscr{I}), \quad p(i) \geq 0, \quad \sum_{i \in \mathscr{I}} p(i) = 1.$$

Clearly p belongs to the $|\mathscr{I}|$ dimensional simplex.

For every non-empty set $E \subseteq V$, let

$$i_E = (i_\gamma, \gamma \in E), \quad i_E \in \mathscr{I}_E = \times_{\gamma \in E} \mathscr{I}_\gamma$$

denote the cell in the E-marginal table; further denote with $p(i_E)$ and $y(i_E)$ the corresponding marginal probability and observed cell-count

$$p(i_E) = \sum_{j \in \mathscr{I} | j_E = i_E} p(j), \quad y(i_E) = \sum_{j \in \mathscr{I} | j_E = i_E} y(j).$$

For every C_l, let

$$p^{C_l} = (p(i_{C_l}), i_{C_l} \in \mathscr{I}_{C_l}), \quad y^{C_l} = (y(i_{C_l}),\ i_{C_l} \in \mathscr{I}_{C_l}), \quad l = 1, \ldots, k$$