Objective Bayes for everyone

Some thoughts on Walker et al. and Rue et al.

Robin J. Ryder

Centre de Recherche en Mathématiques de la Décision, Université Paris-Dauphine

Objective Bayes, Warwick
29 June 2019

Two fascinating on specifying priors in an objective (or objectively subjective) way.

If we want Bayes to be competitive against traditional frequentist methods, and newer Machine Learning methods, we need it to be easy for practitioners.

Computational aspects of inference: R packages, STAN, INLA...

Priors: if there is a prior set by default in the software, use that one.

Otherwise, tradition, rule of thumb...

Objective Bayes is hard!

Questions asked by a practitioner about their prior:

1. Is it easy to use?
2. Is it easy to defend?
3. Is it easy to break?
4. Is there a theoretical grounding?

- Easy to use: conjugate; can compute the value point-wise (for Metropolis-Hastings); can simulate from it (for ABC).
- Easy to defend: others have used it; based on sensible subjective beliefs...
- Easy to break: can try variations on the prior and check sensitivity/robustness.
- Theoretical grounding: objective (for some definition).

An O'Bayes success story

Penalising model component complexity: A principled, practical approach to constructing priors

D Simpson, H Rue, A Riebler, TG Martins… - Statistical …, 2017 - projecteuclid.org

In this paper, we introduce a new concept for constructing prior distributions. We exploit the natural nested structure inherent to many model components, which defines the model component to be a flexible extension of a base model. Proper priors are defined to penalise …

List some sensible principles and desiderata, and deduce prior, leaving some latitude to the user.

Principles

1. Simpliciy
2. Complexity measured by KL
3. Constant rate decay, expressed by λ
4. Flexible scaling

Principle 1: avoid "forced overfitting", which happens when $\pi(\xi) = 0$.
These are enough to choose your prior.

Objective in the sense that it is principled, but leaves some flexibility to the user if necessary.
Based on some intuitive parameterization.
Start with a base model, defined as $\xi = 0$, and builds the prior by penalizing the distance to the base model.
In a way, prior defined on a summary; a prior counterpart to ABC. Makes it invariant to parameterization.

A few questions

The choice of λ looks critical (but apparently isn't?) -> family of distributions.

Impact of the order in which you build your model?

The PC prior relies on orthogonal parameters. How do we handle global dependence structure of the model, if it cannot be rewritten? Or parameters which make no sense taken individually?

Prior induced by a score function

Score function

$$S(\theta, p) = -\log p(\theta) + \frac{p''(\theta)}{p(\theta)} - \frac{1}{2}\left(\frac{p'(\theta)}{p(\theta)}\right)^2$$

which is proper and 2-local.
From this, derive a prior, with potentially some user-defined settings c and $u(0)$.

Prior induced by a score function

Constraints can take the form of specific values ($p(0) = p(1) = 0$), or be on the shape of the prior (p convex).

From a score function and some constraints, derive a universal (ie model-free) prior over Θ, of a more general class than maximum entropy.

Could use a different score function (weighted, or completely different form): general principles stand.

Allows your prior to be properly defined in the M-open case.

Implicit solution

$$p(\theta) \propto \exp(-u(\theta))$$

$$u'(\theta) = \pm\sqrt{ce^{u(\theta)} - 2(1 + u(\theta))}$$

which is solved numerically.

Pretty property: the same prior is derived by minimizing

$$\int \frac{1}{2}\frac{p'(\theta)^2}{p(\theta)} + p(\theta)\log p(\theta)\, d\theta$$

A question on computation

In practice, MCMC needs

$$\frac{p(\theta)}{p(\theta')} = \exp(u(\theta) - u(\theta'))$$

which can be approximated for small $|\theta - \theta'|$ by a Taylor expansion.
BUT errors will accumulate and propagate, so after a large number of iterations you are using a different, undefined prior (which varies from one iteration to the next).
Can some correction be computed? It's OK if it is expensive: only run it every $K >> 1$ iterations.
Going back to the universality of the solution: worth finding a computational solution, as it can then be reused as long as Θ and constraints are unchanged.

A few general questions

- Multivariate parameters
- Use for testing
- Priors which are independent of the model (which is good for misspecification/model choice...) make special sense when the parameter is interpretable. What do we do with nuisance parameters, which are also those for which we have little intuition?
- Propriety
- Invariance

Base models

Havard: Complicated model is an extension of a base model, as parameterized by ξ. Comparing (but not quite testing) $\xi = 0$ vs $\xi > 0$; want $\pi(0) > 0$.
Stephen: Base model is at the centre of the complicated model (is it the median? the mode?).
Natural interpretation as univariate/bivariate test, if we view $\xi = 0$ as a null model.
Unclear how to specify when there is no single obvious choice of a base model.
Impact of the order in which you build your model?

Two approaches to building priors in a systematic fashion.
Different robustnesses: Robust to non-identifiability; Robust to misspecification
Make subjectivity explicit , and make it objective from there.
Big advances in usability of objective priors.

"All models are wrong, but some have well-implemented R packages, so I just use those."
(Anonymous biostatistician on Twitter.)

We are two steps closer to automatic objective priors: they are easier to use, easier to defend, robust and theoretically grounded.