Using informative priors for approximate inference in non-linear Gaussian state space models

Øyvind Hoveid[1]
[1] Norwegian Agricultural Economics Research Institute, Oslo

A basic problem for exact Bayesian inference in non-linear Gaussian state space models, is the presence of several variances, at least one of measurement errors and at least one of state transitions. Unless the variance priors are sufficiently informative, the joint likelihood of global parameters, θ, state variables, x, and data y, $\pi(y, x, \theta)$, will be multi-modal with one infinite mode for each variance that can turn zero. All these modes should in principle be explored as part of an exhaustive Bayesian analysis, but this will not necessarily happen with versions of Metropolis-Hastings algorithm.

With linear Gaussian models the problem is avoided by calculation of the likelihood of data and global parameters, $\pi(y, \theta) = \int \pi(y, x, \theta)dx$, by means of integration over state variables. Integration is analytic with the Kalman filter and smoother, and semi-analytic with INLA. This likelihood with its single finite mode can be explored if the dimension of θ is not too large. The procedure is not straightforward for non-linear models where integration need be conducted numerically, however. The accuracy of integration need be sufficient to establish an approximate posterior likelihood that can be applied for inference. Numerical integration in high-dimensional space is a challenging task, in particular because convenient properties of symmetry and pervasive conditional independence, fade when models turn non-linear.

As an alternative strategy for non-linear models I propose least-informative priors which rule out infinite modes of the joint likelihood at zero variances. The priors can be Wishart distributed with sufficient degrees of freedom. The finite mode(s) can then be easily found. Approximate Bayesian inference can be achieved with importance sampling from the Laplace-approximation at the mode(s). Cases of zero variance can be approached with additional models with zero variance imposed. Model selection (with Bayes factors or DIC) on the various models will then reveal whether a zero variance is a sensible assumption or not. A decisive test in favor of the non-zero variance model rules out a positive probability of zero variance, since the informative prior has not benefitted the performance of the associated model. Weak tests on the other hand suggest that a zero variance should be preferred.

The procedure will be illustrated with an estimated model of farm behavior according to stochastic dynamic optimization. Observed behavior (less measurement errors), x_t, is assumed to maximize a program conditional on stocks (also less errors) X_t:

$$\max_x \{U(x, X_t; \theta) + \beta \mathrm{E}_{\pi(X_{t+1}|x_t, X_t; \theta)} U^*(X_{t+1}; \theta)\} = U^*(X_t; \theta)$$

where U is temporal utility, U^* is expected present utility of all future events and optimized behaviors, β is a discount factor and π is a contingent distribution. Optimization induces a first order condition and an Euler equation that identifies the optimal responses, $x^*(X_t)$, and the functional relationships between the three functions. The global parameters comprise now the parameters of the functions, θ, the discount factor, the variances of state variables, (x_t, X_t), the variances of measurement errors and at last the errors of optimization.

Keywords: State space models, Non-linear models, Gaussian models, Informative priors