Report on the manuscript of "Bayesian Restricted Likelihood Methods: Conditioning on Insufficient Statistics in Bayesian Regression"

In order to handle outliers in regression analysis, this paper suggests the Bayesian inference conditioning on some robust estimator $T(y_{obs})$ of the regression parameter instead of the full data, since the impact of outliers is usually mitigated in the insufficient statistic $T(y_{obs})$. Metropolis-within-Gibbs sampler is used to simulate from the augmented target density $f(\theta, y \mid T(y) = T(y_{obs}))$, and the key issue boils down to how to simulate from $f(y \mid \theta, T(y) = T(y_{obs}))$. Given some special properties of commonly used $T(y_{obs})$, which are conditions C5–C8, main contributions of the paper are in two-fold. First, it constructs a one-to-one and onto mapping $h(\cdot)$ from some linear subspace of $\mathbb{R}^n$ to the manifold $\{y \in \mathbb{R}^n : T(y) = T(y_{obs})\}$, so that any probability density in the linear subspace can be transformed to a probability density on the manifold which can be used as the proposal density in the Metropolis-Hasting(MH) algorithm targeting $f(y \mid \theta, T(y) = T(y_{obs}))$. Second, the jacobian of $h^{-1}(\cdot)$ is derived analytically in order to evaluate the proposal density. I think this paper provides intersting contributions on the important topic of summary statistic-based inference. But the paper didn't given enough discussion on pratical implementation of its algorithm.

Comments

1. The authors propose a broad class of the proposal density, but didn't discuss the choice and tuning within the class. Tuning the proposal is crucial to the mixing and efficiency of the MH algorithm, epecially in the high-dimension space, $n-p$-dimensional, that the paper is dealing with. There are two difficulties in tuning within the proposed class. First, since the proposal density is a complicated transformation of $p(z^*)$ in $\mathbb{R}^{n-p-1}$, even if $p(z^*)$ belongs to a standard parametric family, $p(y_p|\theta)$ does not and it is not straightforward to assess the properties that are usually used in tuning MCMC algorithm, e.g. mode and

tails behaviour. Second, the proposed algorithm (11) is restrictive because it only works for independent proposal, not the random walk proposal. Random walk is more able to explore a non-standard parameter space like the manifold here, while it seems tricky to design the independent proposal that well covers the probability mass of the target density, which again due to the transformation.

2. The simulation studies didn't report the choice of $p(z^*)$, acceptance rates of the MH, and mixing of the overall MCMC algorithm. Due to point 1, I think these issues would be of interest to readers.

3. The models studied in both the simulation and real data examples are just simple linear regression which is not realistic. Multiple linear regression with at least three independent variables should be studied.

4. What are the benefits of the proposed method over the classical robust point estimators, like Huber or Tukey estimators compared in the examples, to offset the additional computational cost and tuning efforts? From the simulation studies, whether it has improvement or not depends on the prior distribution, which seems to be from comparing Bayesian and Frequentist methods. Actually, literatures on the asymptotic properties of $f(\theta|T(y_{obs}))$ shows that, if $T(y_{obs})$ is asymptotically unbiased, $T(y_{obs})$ and $E(\theta \mid T(y_{obs}))$ have the same asymptotic variance. Hence if the prior doesn't matter, both methods have similar performance.

5. I think $f(\theta|T(y_{obs}))$ might perform better in the following scenario: when $T(y_{obs})$ is asymptotically biased, $E(\theta|T(y_{obs}))$ can still be asymptotically unbiased as long as the true parameter is identifiable conditioning on the summary statistic, i.e the value of $E(T(y_{obs}))$ is unique on the true parameter. One possible example is the robust ridge-regression estimator, but it needs further investigation whether this estimator satisfies C1-C8.

6. In page 14, should $C^{\perp}(X)$ $n-p$-dimensional instead of $n-p-1$-dimensional?