On Bayesian Transformation Selection: Problem Formulation and Preliminary Results

E. Charitidou[1], D. Fouskakis[2], I. Ntzoufras[3]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 8

[1] National Technical University of Athens, Department of Mathematics, Athens, Greece
`echarit@central.ntua.gr`
[2] National Technical University of Athens, Department of Mathematics, Athens, Greece
`fouskakis@math.ntua.gr`
[3] Athens University of Economics and Business, Department of Statistics, Athens, Greece
`ntzoufras@aueb.gr`

Abstract

The problem of transformation selection is thoroughly treated from a Bayesian perspective. Several families of transformations are considered with a view to achieving normality: the *Box-Cox*, the *Modulus*, *Yeo & Johnson* and the *Dual* transformation. Markov Chain Monte Carlo algorithms have been constructed in order to sample from the posterior distribution of the transformation parameter λ_T associated with each competing family T. We investigate different approaches to constructing compatible prior distributions for λ_T over alternative transformation families, using the power-prior and the unit-information prior approaches. In order to distinguish between different transformation families, posterior model probabilities have been calculated. Using simulated datasets, we show the usefulness of our approach.

Keywords: Bayesian transformation selection; MCMC; Power-prior; Prior compatibility.

1 Introduction

In the literature, the term *transformation selection* so far pertains to the choice of an optimal value for the transformation parameter within a given family. We introduce a two-step approach where a transformation family is selected at an initial level while at a second level the value of the transformation parameter is specified given the family. Working within the Bayesian context requires careful choice of prior distributions. In our case, this becomes even more complex since the prior distribution for the transformation parameter λ_T under each family T need to be compatible to account for the different interpretation of λ_T given T.

2 Bayesian formulation

Four uniparametric families of transformations are considered and compared with each other: *Box-Cox* [1], *Modulus* [4], *Yeo & Johnson* [6] and *Dual* [5].

Each family is indexed by T and involves a transformation parameter λ_T. Let us denote by $\mathbf{y} = (y_1, \dots, y_n)^T$ the observed data and by $\mathbf{y}^{(\lambda_T)} = (y_1^{(\lambda_T)}, \dots, y_n^{(\lambda_T)})^T$ the transformed data for a given value of the parameter λ_T within a particular transformation family T. We aim for $\mathbf{y}^{(\lambda_T)}$ to be a sample from a Normal distribution $N(\mu_T, \sigma_T^2)$ with unknown parameter vector $\left(\mu_T, \sigma_T^2\right)$ under some appropriate value of λ_T.

Table 1: Posterior model probabilities and log-marginal likelihood values for each trasformation family T along with Monte Carlo estimates for the posterior median (sd) of λ_T, all estimated using Chib's approximation method for the Student simulated dataset.

		Prior[1]	Modulus	Box-Cox	Dual	Id	YJ	Log
$n = 100$	$P(T\|y)$	Prior A	0.99	0.01	< 0.01	< 0.01	< 0.01	< 0.01
		Prior B	0.99	0.01	< 0.01	< 0.01	< 0.01	< 0.01
	$\log f(\mathbf{y}\|T)$	Prior A	-250.59	-255.89	-256.39	-258.28	-259.36	-292.14
		Prior B	-251.92	-257.01	-259.77	-258.28	-260.48	-292.14
	λ_T	Prior A	0.36 (0.15)	1.60 (0.22)	1.62 (0.22)	-	1.08 (0.08)	-
		Prior B	0.34 (0.15)	1.62 (0.23)	1.63 (0.21)	-	1.09 (0.08)	-

		Prior	Modulus	Box-Cox	Dual	YJ	Id	Log
$n = 1000$	$P(T\|y)$	Prior A	1.00	0.00	0.00	0.00	0.00	0.00
		Prior B	1.00	0.00	0.00	0.00	0.00	0.00
	$\log f(\mathbf{y}\|T)$	Prior A	-3733.00	-3960.67	-3961.52	-4027.56	-4125.02	-4601.98
		Prior B	-3732.75	-3960.92	-3963.83	-4027.63	-4125.02	-4601.98
	λ_T	Prior A	-0.01 (0.04)	2.93 (0.12)	2.93 (0.12)	1.23 (0.01)	-	-
		Prior B	-0.01 (0.04)	2.94 (0.12)	2.94 (0.12)	1.23 (0.01)	-	-

[1] Prior A: Unit-information Normal prior; Prior B: Power-prior.

Regarding the prior probability of each of the six transformation families (including the identical and the log transformation), no special prior weight is assigned to any family, i.e. $f(T) = \frac{1}{|\mathcal{T}|} = \frac{1}{6}$. As to the prior on the transformation parameters, it has a hierarchical form: $f(\boldsymbol{\theta}_T|T) = f(\mu_T, \sigma_T^2|\lambda_T, T)f(\lambda_T|T)$. The main parameter of interest within a family is λ_T while (μ_T, σ_T^2) are regarded as nuisance parameters; therefore we employ an independent Jeffreys prior (reference prior) for those.

On the grounds of the different interpretation of λ_T among families, the concept of the power-prior [3] is adopted in order to construct compatible prior distributions. The power-prior for λ_T is formed as the posterior distribution of a set of imaginary data $\mathbf{y}^*$ under a reference baseline prior $\pi_0(\lambda_T|T) \propto 1$:

$$\pi\left(\lambda_T|\mathbf{y}^*, T\right) = \left(\frac{f\left(\mathbf{y}^*|\lambda_T, T\right)^{1/n^*}}{\int f\left(\mathbf{y}^*|\lambda_T, T\right)^{1/n^*} \mathrm{d}\lambda_T}\right). \tag{1}$$

In addition, a unit-information Normal prior setting is used, based on the same imaginary data $\mathbf{y}^*$, which theoretically approximates the former power-prior setting. The variance of the latter prior is determined through the observed Fisher information of $\mathbf{y}^*$. Approximations of the intractable integrals included in the process are achieved through Chib's estimator [2] incorporating the output of a random-walk Metropolis-Hastings algorithm simulating from the marginal posterior distribution of λ_T.

3 Results

In order to illustrate our approach, we have simulated data from a variety of distributions. The Student distribution t_2, having non-centrality parameter equal to -1, is an example of particular interest since symmetry is accompanied by fat tails. The latter characteristic usually induces failure of transformation to normality under most families. Looking at the figures in Table 1, we observe that the supremacy of the Modulus family for this distribution is unquestionable for both medium and large sample sizes ($n = 100$ & $n = 1000$) under both power-prior and unit-information Normal prior.

4 Conclusions

The compatibility issues in transformation selection have been addressed through the power prior approach. By and large, there is more than adequate convergence of results under both prior settings. The fat tailed Student distribution is optimally associated to the Modulus transformation. The latter result has been verified for other fat tailed distributions such as the Laplace.

Aknowledgements

This work has been partially funded by the Research Committee of the National Technical University of Athens (Π.Ε.Β.Ε. 2010 Scheme).

References

[1] Box, G.E.P. and Cox, D.R. **An Analysis of Transformations (with discussion)**. *Journal of the Royal Statistical Society, Series B*; 1964; 26(2); pp. 211-252.

[2] Chib, S. and Jeliazkov, I. **Marginal likelihood from the Metropolis-Hastings output**. *Journal of the American Statistical Association*; 2001; 96; pp. 270-281.

[3] Ibrahim, J.G. and Chen, M.H. **Power-Prior Distributions for Regression Models**. *Statistical Science*; 2000; 15; pp. 46-60.

[4] John, J.A. and Draper, N.R. **An alternative family of transformations**. *Applied Statistics*; 1980; 29; pp. 190-197.

[5] Yang, Z. **A modified family of power transformations**. *Economic Letters*; 2006; 92(1); pp. 14-19.

[6] Yeo, I.K. and Johnson, R.A. **A new family of power transformations to improve normality or symmetry**. *Biometrika*; 2000; 87(4); pp. 954-959.