

Prior Sensitivity Analysis for a Hierarchical Model

Junaidi, E. Stojanovski , D. Nur

The University of Newcastle, Callaghan, NSW, 2308, AUSTRALIA

Junaidi@newcastle.edu.au, Elizabeth.Stojanovski@newcastle.edu.au, Darfiana.Nur@newcastle.edu.au

Abstract

Meta-analysis can be presented in the Frequentist or Bayesian framework. Based on the model of DuMouchel, a simulation study is conducted which fixes the overall mean and variance-covariance matrix to generate estimates of the true mean effect. These estimates will be compared to the true effect to assess bias. A sensitivity analysis, to measure the robustness of results to the selection of prior distributions, is conducted by employing Uniform and Pareto distributions for the variance components, the t -distribution for the overall mean component and a combination of priors for both variance and mean components respectively. Results were more sensitive when the prior was changed only on the overall mean component.

Keywords: Sensitivity analysis, hierarchical Bayesian model, meta-analysis

1. Introduction

Meta analysis is a statistical method used to obtain an overall estimate by combining results from several individual related studies [10]. Combining results of comparable studies to obtain an overall estimate of treatment effect (e.g. odds ratio, relative risk, risk ratio) can reduce uncertainty and can be useful when the sample size used in each study is small in an attempt to increase power [16].

Meta-analyses can be presented in the Frequentist or Bayesian framework. Within the Frequentist framework, hypotheses are based on information presented within studies and results are often presented in term of 95% confidence intervals to estimate parameters [7]. Weighted averages tend to be used as the overall treatment effect from individual study estimates. One of the more common models used is the inverse of the within-study variances ([2], [11]).

Bayesian methods combine prior probability distributions that reflect a prior belief of the possible values, with the (likelihood) distributions based on the observed data, to produce the posterior probability distributions. The methods are based on the Bayesian rule for probability and can be considered an alternative approach to statistical inference. By multiplying the prior probability with the likelihood, information about the parameters which come from the observed data can be combined with information

from the prior distribution that is external to the data ([2], [3]). The posterior distribution can be explained in terms of probabilities and can be considered as borrowing strength from the other studies.

2. Methods

Hierarchical Bayesian Model

A variety of Bayesian methods have been developed in meta-analysis which include those developed by DuMouchel ([13], [14], [15]). The standard hierarchical Bayesian model proposed by DuMouchel [13] provides the following distributional assumptions:

$$Y_i \sim N(\theta_i, \sigma_Y^2 W_Y) \quad i = 1, 2, \dots, n \quad (1)$$

$$\sigma_Y^2 \sim \frac{\chi_{\nu_Y}^2}{\nu_Y}$$

$$\theta_i \sim N(\mu, \sigma_\theta^2 W_\theta) \quad (2)$$

$$\sigma_\theta^2 \sim \frac{\chi_{\nu_\theta}^2}{\nu_\theta}$$

$$\mu \sim N(0, D \rightarrow \infty)$$

The Model has 2 levels, level one indicates data from the studies, and the next level refers to study-specific parameters.

Level 1: $Y_i \sim N(\theta_i, \sigma_Y^2 W_Y)$

In the Model, n denotes the number of studies ($i = 1, 2, \dots, n$). Y_i indicates the observed statistics following the normal distribution with mean (θ_i) and covariance matrix ($\sigma_Y^2 W_Y$). Furthermore, W_Y indicates the observed precision matrices (inverse observed variance-covariance matrix) describing within-study variation. If studies are assumed independent, W_Y is to be a diagonal matrix with the individual estimates of the variance of Y_i on the diagonal. σ_Y^2 indicates the degree of uncertainty around the observed precision matrix, as expressed through the respective degree of freedom V_Y which denotes set to the average number of cases of studies ($df = n-1$). The chi-square distribution is defined by parameter V_Y to denote how well known the variance structure W_Y .

Level 2: $\theta_i \sim N(\mu, \sigma_\theta^2 W_\theta)$

θ_i denotes study-specific parameters following the normal distribution with mean ($\mu =$ an overall mean) and covariance matrix ($\sigma_\theta^2 W_\theta$). W_θ is the prior precision matrix describing between-study variation. Independence is assumed between studies, so the precision matrices are all diagonal. σ_θ^2 indicates the degree of uncertainty around the prior precision matrix, as expressed through the respective degree of freedom V_θ which denotes set to equal to the number of studies ($df = n - 1$). The chi-square distribution is defined by parameter V_θ to denote how well known the variance structure W_θ . μ is an overall mean following the normal distribution with mean (0) and variance ($D \rightarrow \infty$). $D \rightarrow \infty$ indicates that elements of D are very large and tend to infinity.

Prior Sensitivity Analysis

The prior distribution plays a crucial role in Bayesian analysis. The conclusion obtained using the Bayesian approach is dependent on the prior distributions. The choice of prior(s) distribution must be determined with care, particularly when the likelihood doesn't dominate the prior. The prior distribution will be less important when the number of studies is large. The non-informative prior distribution will be very useful in the situation when prior information, expectations and beliefs are minimal or not available. In a multi-parameter setting, the specification or elicitation of prior beliefs is not an easy task. Uniform priors or Jeffrey's prior are assumed non-informative priors. The use of vague priors can be problematic due to small amount of data. Hence choosing a vague prior

distribution is heavily dependent on the situation ([5], [12]).

DuMouchel [15] stated that results can be affected by different specifications of prior distributions. Sensitivity analysis, to measure the robustness of results regarding selection of prior distributions, should always be carried out. The final results in terms of posterior distributions in meta-analysis will be more robust if the results obtained are unchanged via a sensitivity analysis [12].

Using different prior distributions, for variance components for the within and between studies standard deviation σ , were specified on the Model. However it should be realized that specification of a prior distribution on the standard deviation scale, implies a distribution on the variance and precision scales. The parameterisations for the different prior distributions are described in WinBugs. The prior distributions used on the model are presented in Table 2.1.

Variance components	$\sigma^2 \sim$ Uniform(1/1000, 1000) $1/\sigma^2 \sim$ Pareto(1, 0.25)
Overall mean	$\mu \sim t$ -distribution
Combination	$\sigma^2 \sim$ Uniform(1/1000, 1000) $1/\sigma^2 \sim$ Pareto(1, 0.25) $\mu \sim t$ -distribution

Table 2.1. Prior distributions used on the model due to sensitivity analysis.

3. Results

A simulation study for the model is presented. By employing 1,000 random samples in each of 30 studies, the R code program was created to simulate from the multivariate normal distribution. By fixing the overall mean and variance-covariance matrix, we generate estimates of the true mean effect. These estimates will be compared to the true effect to assess the bias. Steps used for the simulation study will be as follow.

- Step 1.* We fix the value of an overall mean (μ).
- Step 2.* We generate θ_i based on the μ (where n be the number of studies, $i = 1, 2, \dots, n$); Σ_θ indicates symmetric, positive definite $n \times n$ variance-covariance matrix.
- Step 3.* The value of observed statistics (Y_i) will be obtained based on θ_i . Σ_Y denotes a symmetric, positive definite $n \times n$ variance-covariance matrix.

Furthermore, by using the risk ratios Y_1, Y_2, \dots, Y_{30} wobtained from simulation and weighted matrix (inverse of variance), we calculate the value an overall mean in WinBugs based on the DuMouchel model presented in Section 2. We obtained an overall mean value of 2.554 associated with credible interval 2.233-2.878 which was close to the fixed true effect (2.560) confirming the setup of the simulation study.

Sensitivity analysis, to measure the robustness of results to the selection of prior distributions, is conducted. The DuMouchel model utilised the Chi-square distribution on the variance parameters, σ_Y^2 and σ_θ^2 . Based on Lambert [12], the Uniform and Pareto distribution will be used here for the variance parameters. The Normal distribution will be utilised initially for the overall mean (μ), consistent with that used by Dumouchel. Furthermore, the normal distribution was imposed for an overall mean on the DuMouchel model will be changed by t -distribution without changing the value of variances. In addition, combining the priors on variances and an overall mean will be done. However, some changing on prior cannot be done potentially due to range of distribution not fit. For example, when we change both variances using Uniform (1/1000,1000). Simulation data based on 1,000 random samples for 30 studies here will be generated. The true overall mean (μ) used for this demonstration is 2.560.

Prior distribution for variance components

Spiegelhalter [6] investigates the uniform prior distribution on the variance.

$$\sigma^2 \sim \text{Uniform}(1/1000, 1000)$$

By using this distribution for parameters σ_Y^2 as well as σ_θ^2 , burn-in for 10,000 iterations on the model, the results show an estimated overall mean are 2.558 and 2.564, respectively. These are close to the true effect (2.560). From this preliminary analysis, the use of these others prior distributions on the model do not appear to have a substantial effect on the true study estimate.

Risks Ratios	Mean	S.D	2.5%	97.5%
$\sigma_Y^2 \sim \text{Uniform}(1/1000, 1000)$				
μ	2.558	0.144	2.283	2.826
$\sigma_\theta^2 \sim \text{Uniform}(1/1000, 1000)$				
μ	2.564	0.089	2.417	2.714
$1/\sigma_\theta^2 \sim \text{Pareto}(1, 0.25)$				
μ	2.459	0.539	1.406	3.497
$1/\sigma_Y^2 \sim \text{Pareto}(1, 0.25) \ \& \ 1/\sigma_\theta^2 \sim \text{Pareto}(1,0.25)$				
μ	2.412	0.624	1.173	3.615

Table 3.1. Summary statistics for overall mean (μ) by changing the prior variance components using Uniform and Pareto distributions on the model.

$$1/\sigma^2 \sim \text{Pareto}(1, 0.25)$$

This is equivalent to a uniform prior distribution for variance in the range (0, 4). By changing parameter variance at σ_θ^2 , we obtained the overall mean 2.459 (1.406 – 3.497). The overall mean 2.412 (1.173 – 3.615) was obtained when σ_Y^2 and σ_θ^2 changed using this distribution. Table 3.1 shows summary statistics by changing prior distribution variance components on the Model.

Prior distribution for the overall mean

$$\mu \sim t\text{-distribution}(0, k= \text{df})$$

The t -distribution will be employed for the overall mean of the model. Density of the t -distribution for degree of freedom 2, 3, 5, 10, 30 and 50 will be compared to the normal distribution ($\mu = 2.554$). The overall estimated mean using the t -distribution is presented in Table 3.2. This shows the results of overall mean to be very close to the true parameter

Risks ratios	Mean	S.D	2.5%	97.5%
df = 2				
μ	2.556	0.166	2.232	2.875
df = 3				
μ	2.554	0.166	2.226	2.871
df = 5				
μ	2.555	0.165	2.233	2.874
df = 10				
μ	2.554	0.168	2.224	2.874
df = 30				
μ	2.552	0.167	2.228	2.881
df = 50				
μ	2.555	0.167	2.234	2.881

value.

Table 3.2. Summary statistics the overall mean (μ) by changing the prior of mean using t -distribution (0, k =df) on the model.

Prior distribution for both variance and overall mean

Prior distributions for both the variance (Uniform and Pareto) and overall mean (t -distribution) simultaneously were employed for the model. By changing the overall mean using t -distribution (0, $k=2$), $\sigma_Y^2 \sim \text{Uniform}(1/1000,1000)$ obtained the overall mean is 2.455 (1.402 – 3.489). When the overall mean was changed using t -distribution (0, $k=2$), $1/\sigma_Y^2$ and $1/\sigma_\theta^2 \sim \text{Pareto}(1, 0.25)$ the result was 2.406 (1.164 – 3.598). These show the overall mean to be reasonably close to the true effect. Summary results by changing the priors on the variances and mean can be seen in Table 3.3.

Risks Ratios	Mean	S.D	2.5%	97.5%
$\mu \sim t\text{-distribution}(0, k= 2)$ $\sigma_Y^2 \sim \text{Uniform}(1/1000, 1000)$ $\sigma_\theta^2 \sim \chi^2$				
μ	2.455	0.530	1.402	3.489
$\mu \sim t\text{-distribution}(0, k= 2)$ $1/\sigma_Y^2 \sim \text{Pareto}(1, 0.25)$ $1/\sigma_\theta^2 \sim \text{Pareto}(1, 0.25)$				
μ	2.406	0.618	1.1647	3.5944

Table 3.3. Summary statistics for μ by changing the prior on mean using t -distribution and variance components using Uniform/Pareto.

4. Conclusions

The simulation study on the model showed the overall estimated mean to be close to the true effect, indicating the estimator as consistent and unbiased. While the prior distribution was imposed on the overall mean only, a change in prior showed results to be consistent. A change in prior on the variance components only and on the combination of variance and mean components simultaneously are more sensitive compared to when modifying the prior on only the mean component.

References

- [1] A.J. Sutton, K.R. Abrams, D.R. Jones, "Methods for meta-analysis in medical research", 2004.
- [2] A.J. Sutton, K.R. Abrams, "Bayesian methods in meta-analysis and evidence synthesis", *statistical methods in medical research*, 10 277-303, 2001.
- [3] C. DiMaggio, S. Galea, "substance use and misuse in the aftermath of terrorism, A Bayesian meta-analysis", *Addiction*, 104 (6) 894-904, 2009.
- [4] D.A. Berry, D.K. Stangl, "Meta-analysis in medicine and health policy", Marcel Dekker, New York, 2000.
- [5] D. Gamerman, "Markov chain monte carlo : sthocastics simulation for Bayesian inference", 1997.
- [6] D.J. Spiegelhalter, K.R. Abrams, "Bayesian approach to clinical trials and health-care evaluation", 2004.
- [7] F. Bullard, "A brief introduction to Bayesian statistics", 2001.
- [8] K. Honoki, E. Stojanovski, "Prognostic significance of P16INK4a alteration for ewing sarcoma", *American cancer society*, 1351-1360, 2007.
- [9] K.R. Abrams, C.L. Gillies, P.C. Lambert, "Meta-analysis of heterogeneously reported trials assessing change from baseline", *Statistics in Medicine*, 24 3823-3844, 2005.
- [10] M. Borenstein, L.V. Hedges, J.P.T. Higgins, "Introduction to meta-analysis", 2009.
- [11] M.J. Batterham, "Investigating heterogeneity in studies of resting energy expenditure in persons with HIV/AIDS: a meta-analysis", *American Journal of Clinical Nutrition*, 81 (3) 702-713, 2005.
- [12] P.C. Lambert, A.J. Sutton, "How vague is vague? A simulation study of the impact of the use of vague prior distribution in MCMC using WinBUGS", *Statistics in Medicine*, 24 2401-2428, 2005.
- [13] W. DuMouchel, "Predictive cross-validation of Bayesian meta-analysis", *Bayesian Statistics*, 5 107-128, 1996.
- [14] W. DuMouchel, D.A. Berry, "Meta-analysis for response models", *Statistics in Medicine*, 14 679-685, 1995.
- [15] W. DuMouchel, "Bayesian meta-analysis", *Statistical Methodology in the Pharmaceutical Sciences*, 509-529, 1990.
- [16] Y.Y. Jung, "Identifying differentially expressed genes in meta analysis via Bayesian model clustering", *Biometrical Journal*, 48 (3) 435-450, 2006.