Bayesian meta-analytic measure

Camila B. Martins, Carlos A. de B. Pereira and Adriano Polpo

Abstract Meta-analysis is a procedure that combines results from studies (or experiments) with common interest: inferences about an unknown parameter. We present a meta-analytic measure based on a combination of the posterior density functions obtained in each of the studies. Clearly, the point of view is from a Bayesian perspective. The measure preserves both the heterogeneity between and within the studies, and it is assumed that the whole data from each study are available.

1 Introduction

The meta-analysis is the analysis that combine different studies having the same objective, inferences about a common parameter. It can be based on either the summarized results, means and variances for instance, or on the whole data, case by case results, if available. Here, it is assumed that the case by case observations of all the studies are available. A convex combination of the studies posterior densities functions is considered as the complete available information about the parameter of interest. The measure incorporates both heterogeneities types, within and between studies, keeping the important information of each experiment. Most familiar meta-analysis methods work with summarized statistics such as means or proportions. These types of data resum abandon the studies within heterogeneity, probably loosing relevant information. Hierarchical modeling that Bayesians may treat meta-analysis is questionable since its first level of uncertainty consider invisible independent observations of a leader random variable. However different locations

C.B. Martins
Federal University of Sao Paulo, e-mail: cb.martins@unifesp.br

C.A. de B. Pereira
University of Sao Paulo e-mail: cpereira@ime.usp.br

A. Polpo
Federal University of Sao Carlos e-mail: polpo@ufscar.br

may have different environments, for instance. In that way, the posterior distribution of the parameters of the leader random variable may not describe correctly the information sought.

The simple method proposed here is to combine the final Bayesian analyses obtained in each study. Considering a weighted average of the posterior densities we obtain a density that may represent better the different nuances of the studies without being restrict to any family of distributions. The proposed posterior meta-analytic measure is presented in Section 2. Section 3 illustrate the measure with an example and Section 4 is for final remarks.

2 Meta-analysis measure

Consider that we have N different studies with aim to understand some characteristic θ, $\theta \in \Theta$. Let $\boldsymbol{X} = \{\boldsymbol{X_1}, \ldots, \boldsymbol{X_N}\}$, where $\boldsymbol{X_j} = \{X_{j1}, \ldots, X_{jn_j}\}$ are the data from the j-th study, $j = 1, \ldots, N$. Consider that X_{ji} are independent random variables with density function $f(x_{ji} \mid \theta)$, $i = 1, \ldots, n_j$. The small caps $\boldsymbol{x}$ and $\boldsymbol{x_j}$ are the observed values of $\boldsymbol{X}$ and $\boldsymbol{X_j}$, respectively. The likelihood function of the j-th study is

$$L_j(\theta \mid \boldsymbol{x_j}) = \prod_{i=1}^{n_j} f(x_{ji} \mid \theta).$$

Given the prior density function $\pi(\theta)$, the posterior meta-analytic measure is defined by

$$\pi(\theta \mid \boldsymbol{x}) = \sum_{j=1}^{N} \omega_j \pi_j(\theta \mid \boldsymbol{x_j}), \tag{1}$$

where $\omega_j > 0$, $\forall j$, $\sum_{j=1}^{N} \omega_j = 1$, and

$$\pi_j(\theta \mid \boldsymbol{x_j}) = \frac{L_j(\theta \mid \boldsymbol{x_j})\pi(\theta)}{\int_{\Theta} L_j(\theta \mid \boldsymbol{x_j})\pi(\theta)\,\mathrm{d}\theta}.$$

The constant ω_j is the weight of each study. If there is a reason to consider one study more important than other, then it is possible to set a higher value for the weight of this study. We consider that the importance of each study is proportional to their sample size, that is $\omega_j = n_j / \left(\sum_{i=1}^{N} n_j \right)$.

Note that, there is only one prior $\pi(\theta)$, and there is only one parameter θ. We do not have a parameter for each study and then we combine them. We can write the proposed measure as

$$\pi(\theta \mid \boldsymbol{x}) = \pi(\theta) \left[\sum_{j=1}^{N} c_j L_j(\theta \mid \boldsymbol{x_j}) \right], \tag{2}$$

where

$$c_j = \frac{\omega_j}{\int_\Theta L_j(\theta \mid x_j)\pi(\theta)\,\mathrm{d}\theta}.$$

From Equation (1), the proposed posterior measure, $\pi(\theta \mid x)$, is a convex combination of posterior distributions of each study. By the other hand, from Equation (2), the proposed measure is the prior multiplied by a mixture of the likelihood of each study, that is a full Bayesian procedure (posterior = prior × model). Both cases results in the same posterior meta-analytic measure $\pi(\theta \mid x)$, which is a probability density function of θ given the data of all available studies.

It is important to call the attention to the fact that, for proper priors, the proposed measure is always a probability density function of θ after observed the data, like any Bayesian analysis procedure.

3 Example

Table 1 presented the results of nine studies about the success in the use of the SAMe (an anti-depressant drug S-adenosylmethionine). The data are given in [2, 1].

Table 1 Use of SAMe.

j	x_j	n_j	x_j/n_j	ω_j
1	20	20	1.00	0.133
2	4	10	0.40	0.067
3	11	16	0.69	0.107
4	10	19	0.53	0.127
5	5	14	0.36	0.093
6	36	46	0.78	0.306
7	9	10	0.90	0.067
8	7	9	0.78	0.060
9	4	6	0.67	0.030
Total	106	150	0.71	–

j is for the j-th study;

x_j is the number of success;

n_j is the sample size of each study;

and $\omega_j = n_j/150$.

Let X_j be a random variable related to the number of success in the use of SAMe. We have that X_j given θ has Binomial distribution with parameters n_j as the number of trials and θ is the success rate of the use of SAMe, $j = 1, \ldots, 9$. We consider that n_j are fixed constants, and then the only parameter is θ. In this case, the likelihood function of the j-th study is

$$L_j(\theta \mid x_j, n_j) = \theta^{x_j}(1 - \theta)^{n_j - x_j}.$$

Considering that $\pi(\theta) = 1$, $\theta \in (0, 1)$, that is, the prior of θ is a Uniform distribution over $(0, 1)$. The posterior of each study is

$$\pi_j(\theta \mid x_j, n_j) = \frac{\Gamma(n_j+2)}{\Gamma(x_j+1)\Gamma(n_j-x_j+1)}\theta^{x_j}(1-\theta)^{n_j-x_j},$$

where Γ is the mathematical gamma function. In this case, $\theta \mid x_j, n_j$ has a Beta distribution with parameters $x_j + 1$ and $n_j - x_j + 1$.

The meta-analytic measure is

$$\pi(\theta \mid \boldsymbol{x}, \boldsymbol{n}) = \sum_{j=1}^{9} \omega_j \frac{\Gamma(n_j+2)}{\Gamma(y_j+1)\Gamma(n_j-x_j+1)}\theta^{x_j}(1-\theta)^{n_j-x_j},$$

where $\boldsymbol{x} = \{x_1, \ldots, x_9\}$, and $\boldsymbol{n} = \{n_1, \ldots, n_9\}$. In this example, the meta-analytic measure is a mixture of Beta distributions.
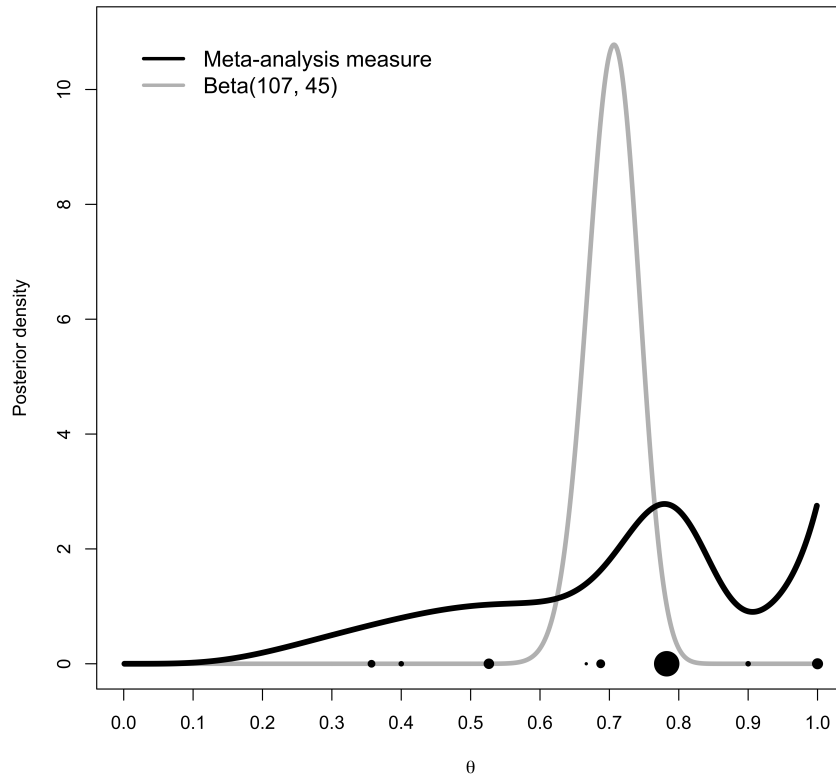

Fig. 1 Meta-analytic measure and posterior distribution considering all samples from a unique study, Beta(107, 45); the dots represent the proportion of each study, and the size of the dot is proportional to the sample size of the study.
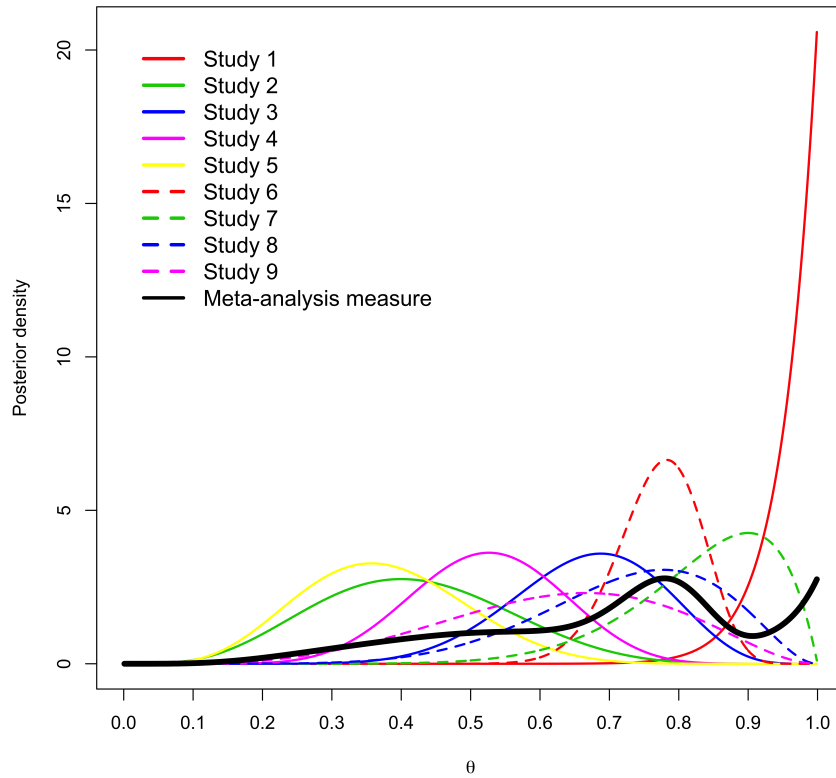
Fig. 2 Meta-analytic measure and posterior distributions of all studies.

For the hierarchical model, we cannot have a direct analysis of θ as a population parameter. As discussed in the introduction, at first prior level, we will need $\theta_1, \ldots, \theta_9$, the success rates of each study. At second level, we will have a prior for the hyper-parameter of θ_j, which will not lead to the analysis of θ. [1] did a hierarchical analysis of these data from the perspective of multicenter analysis, his main interest lies on the estimation of θ_j, and the difference between them.

By other hand, if we had considered that we have a unique study, our data should be 106 success in 150 trials. Considering the same uniform prior for θ, our posterior $\theta \mid x = 106, n = 150$ would have Beta distribution with parameters $x + 1 = 107$ and $n - x + 1 = 45$. The meta-analytic measure and the Beta(107, 45) distribution are in Figure 1. As expected the meta-analytic measure preserve the characteristic of the data, and the Beta(107, 45) distribution does not represent the different results of the studies.

If our interest lies on the comparison of the studies, we can draw the posterior of each study together with the meta-analytic measure (Figure 2). From the meta-analytic measure we have that the median is 0.72, the mode is 0.78, the mean is 0.69 and the 95% high posterior density credible interval is $(0.31; 1.00]$. We can see a meta-analytic measure with three modes in 0.50, 0.78 and 1.00. This may suggest that we have three groups in the studies.

4 Final remarks

The proposed method is a posterior distribution, called meta-analytic measure. The results show that we are not doing inference over means (the usual method for meta-analysis), and the proposed measure provides an complete inferential framework. We are able to evaluate posterior mean, mode, median, variance, credible interval, or even perform a hypothesis test. We have a measure that represents the observed data, the heterogeneity of the studies, and it the analysis can be performed as any traditional Bayesian method.

References

1. Berry, D.A.: A bayesian approach to multicenter trials and meta-analysis. Tech. Rep. ED325480, National Science Foundation, Washington, DC. (1989)
2. Janicak, P., Lipinski, J., Davis, J., Coinaty, J., Waternaux, C., Cohen, B., Altman, E., Sharma, R.: S-adenosyl-methionine (same) in depresion: a literature review and preliminary data report. Alabama Journal of Medical Sciences **25**(3), 306–313 (1988)