Meta-analysis of Bayesian analyses

Paul Blomstedt, Diego Mesquita and Samuel Kaski
Helsinki Institute for Information Technology HIIT,
Department of Computer Science, Aalto University
{paul.blomstedt, diego.mesquita, samuel.kaski}@aalto.fi

Abstract

Meta-analysis aims to combine results from multiple related statistical analyses. While the natural outcome of a Bayesian analysis is a posterior distribution, meta-analyses traditionally combine analyses summarized as point estimates, often with limiting distributional assumptions. In this extended abstract, we develop a framework for combining posterior distributions, which builds on standard Bayesian inference, but using distributions instead of data points as observations. The resulting framework preserves basic theoretical properties, such as order-invariance in successive updates and posterior concentration. In addition to providing a consensus analysis for multiple Bayesian analyses, we highlight the benefit of being able to reuse posteriors from computationally costly analyses and update them post-hoc without having to rerun the analyses themselves. Wide applicability of the framework is illustrated with examples of combining results from likelihood-free Bayesian analyses, which would be difficult to carry out using standard methodology.

1 Introduction

Meta-analysis comprises a variety of statistical methods to combine results from similar studies or analyses. Traditionally, the results of each study $j = 1 \dots J$ are taken in as summary statistics D_j, which are supposed to provide information about the quantity (or *effect*) of interest. These statistics usually assume the form of moment estimates for the effect size of some treatment, for which case there is a vast body of work and well-established Bayesian methodology [5].

In the standard setting of Bayesian *random effects meta-analysis*, to reflect the general idea of the studies being non-identical but related, it is common to regard them as exchangeable [4], with the study-specific effect represented by some parameter, say θ_j, and the overall effect by another parameter, say φ, leading to the following hierarchical model:

$$\begin{aligned} \varphi &\sim Q \\ \theta_j &\sim P_\varphi \\ D_j &\sim F_{\theta_j}, \end{aligned}$$

where F_{θ_j} is typically modeled as $\mathcal{N}(\theta_j, \hat{\sigma}_j^2)$, with $\hat{\sigma}_j^2$ estimated from data. One of the primary goals of the above model, and of meta-analysis itself, is to estimate φ, for which the marginal posterior density is given by

$$q(\varphi|D_1, \dots, D_J) \propto \prod_{j=1}^{J} \left[\int f(D_j|\theta_j) p(\theta_j|\varphi) d\theta_j \right] q(\varphi), \tag{1}$$

where $f(D_J|\theta_j)$ is the density function associated with F_{θ_j}.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Surprisingly, while the natural outcome of a Bayesian analysis is a posterior distribution, the analogous task of combining posterior distributions has received little attention.

There are many compelling reasons for reporting analysis results as posterior distributions instead of data summaries. First, a distribution describes the analyst's uncertainty in the obtained results. Second, a posterior distribution can be directly specified on a quantity of interest, whereas summary statistics are most commonly computed directly from observations, providing only indirect information about the quantity. Furthermore, the posterior distribution may also include prior knowledge not present in the data, but possibly obtained by expert elicitation [1]. This is particularly important in problems where not enough data is available to inform a model about some of its parameters. For example, models describing complex biological phenomena may have so many parameters that they cannot be estimated without the use of informative priors [6].

In this extended abstract, we introduce a novel framework for the Meta-analysis of Bayesian Analyses (MBA). We illustrate our framework with an example combining results from likelihood-free inference of moving average models. A more detailed treatment of MBA, comprising theoretical guarantees and an additional illustration on real-world data can be found in the supplementary material.

2 Meta-analysis of Bayesian analyses

We consider the setting where instead of summary statistics, we have posterior distributions with densities $\pi_j(\theta_j)$ available from each of J studies, based on which we wish to update our prior knowledge about the global effect φ, in analogy with Equation (1). We build on the interpretation that each $\pi_j(\theta_j)$ is a probabilistic representation of belief [2], which reflects our uncertainty about the value of the corresponding local effect θ_j. In this work, we directly marginalize away the uncertainty in the observations as they appear in the likelihood, which leads us to update $q(\varphi)$ as

$$q^*(\varphi) \propto \prod_{j=1}^{J} \left[\int p(\theta_j|\varphi)\pi_j(\theta_j)d\theta_j \right] q(\varphi). \tag{2}$$

Note that Equations (1) and (2) differ only in the local likelihood $f(D_j|\theta_j)$ being replaced with the density $\pi_j(\theta_j)$. Intuitively, the former carries information provided by the data only, while the latter carries information provided by both the data and additional prior knowledge. While Equation (2) does not characterize a typical (marginal) posterior, we prove that it retains some basic properties of standard Bayesian inference, such as order-invariance in successive updates (in contrast with Jeffrey's rule) and posterior concentration as $J \to \infty$.

Additionally, updating $q(\varphi)$ into $q^*(\varphi)$ in Equation (2) is equivalent to propagating beliefs from leaf nodes to the root node in a tree-structured graphical model, with node potentials $q(\varphi)$, $\pi_j(\theta_j)$ and edge potentials $\psi_j(\theta_j, \varphi) := p(\theta_j|\varphi)$, $j = 1, \ldots, J$, see Figure 1. In a similar fashion, we may update any $\pi_j(\theta_j)$ into $\pi_j^*(\theta_j)$ by propagating beliefs to the jth leaf node from the remaining nodes in the model. This yields a way of updating study-specific posteriors *post-hoc*, borrowing strength from posteriors obtained in other studies.

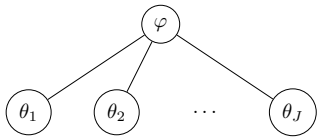

Figure 1: Tree-structured graphical model with study-specific posteriors providing the initial beliefs for $\theta_1, \ldots, \theta_J$.

A major advantage of our meta-analysis framework is its flexibility. In particular, the study-specific inferences resulting in $\pi_j(\theta_j)$ are independent of the combination model, $q(\varphi)\prod_{j=1}^{J} p(\theta_j|\varphi)$, which is imposed by the meta-analyst. This means that, unlike in hierarchical models, all study-level complexities are hidden 'under the hood' and need not explicitly be included in the meta-analysis. For instance, in likelihood-free models [7, 8], the data can typically be summarized by a number of different statistics but there is no closed-form likelihood to relate these to the parameter of interest. In our framework, likelihood-free inferences can be conducted separately for each study using *approximate Bayesian computation*, after which the resulting posteriors are directly combined in a meta-analysis. We highlight the benefit of being able to reuse results from computationally costly analyses, such as likelihood-free inferences, and update them without having to rerun the analyses themselves. In the current paper, we propose a straightforward

computational strategy for our framework, where we first impose parametric approximations on the observed posteriors.

3 Numerical illustration: MA(q) process

We consider the problem of combining results from the analyses of moving average models conducted using likelihood-free models. In likelihood-free models, there is no closed-form likelihood to relate the quantity or effect of interest to data, which poses a challenge for traditional formulations of meta-analysis. In our framework, we directly utilize the inferred posteriors to build a joint model. In addition to modeling the shared central tendency of the inferred model parameters, we demonstrate that weakly informative or poorly identifiable posteriors for individual studies can be updated post-hoc through joint modeling.

The MA(q) process is a standard example in the literature on likelihood-free inference due to its simple structure but fairly complex likelihood and non-trivial relationship between parameters and observed data. Assuming zero mean, the MA$(q = 2)$ process $(y_t)_{t\in\mathbb{N}^+}$ is defined as

$$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}, \tag{3}$$

where $(\theta_1, \theta_2) \in \mathbb{R}^2$ and $\epsilon_s \sim \mathcal{N}(0, 1)$, $s \in \mathbb{Z}$. The quantity of interest for which we conduct inference is $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Following Marin *et al.* [8], we use as a prior for $\boldsymbol{\theta}$ an uniform distribution over the set $\mathcal{T} \subset \mathbb{R}^2 \triangleq \{(\theta_1, \theta_2) \in \mathbb{R}^2 | -(\theta_2 + 1) < \theta_1 < \theta_2 + 1,\ -1 < \theta_2 < 1\}$, which, by restriction of the parameter space, imposes a general identifiability condition for MA(q) processes. Inference for $\boldsymbol{\theta}$ is then conducted using approximate Bayesian computation (ABC) with rejection sampling, taking as summary statistics the empirical autocovariances of lags one and two. Furthermore, a Euclidean distance of 0.1 is used as acceptance threshold.

To illustrate our meta-analysis framework, we first sample $J = 12$ realizations of $\boldsymbol{\theta}$ using the following generating process:

$$\theta_1 \sim \mathrm{Unif}(0.4, 0.8), \quad \theta_2 \sim \mathcal{N}(-0.4 + \theta_{1j}, 0.04^2). \tag{4}$$

Given each realization $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j})$, $j = 1, \ldots, J$, we then generate a series of 10 data points, $(y_{j1}, \ldots, y_{j10})$, according to Equation (3). For each time-series, we independently perform ABC as described above, generating 1000 samples from each posterior. It can be seen that the very limited information given by the data in each of the analyses leaves the posteriors with a considerable amount of uncertainty.

For meta-analysis, we first specify a model for the study-specific effects $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J$ *as if* they were observed quantities from an exchangeable sequence. As the true generating mechanism of the effects is typically unknown, the model must be specified according to the analyst's judgment. To reflect this, we model the generating process as a Gaussian distribution with parameters $\varphi = (\boldsymbol{\mu}, \Sigma_0)$,

$$\boldsymbol{\theta}_j \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma_0). \tag{5}$$

For $\boldsymbol{\mu}$ and the covariance matrix Σ_0, we use Gaussian and inverse Wishart prior, respectively,

$$\boldsymbol{\mu} \sim \mathcal{N}_2(\boldsymbol{m}, V) \quad \text{and} \quad \Sigma_0 \sim \mathcal{W}^{-1}(\nu, \Psi)\,, \tag{6}$$

with

$$\boldsymbol{m} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}, \quad V = \begin{bmatrix} 0.4 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}, \quad \text{and} \quad \nu = 4, \quad \Psi = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}.$$

The above values were chosen to provide reasonable coverage of $\mathcal{T}$, the constrained support of $\boldsymbol{\theta}$. Furthermore, ν was chosen as $\dim(\boldsymbol{\theta}) + 2$ to directly yield Ψ as the mean of the inverse Wishart prior on Σ_0.

After specifying the assumed generative model for $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J$ according to Equations (5)–(6), the following step in MBA is to incorporate the observed beliefs for each $\boldsymbol{\theta}_j$ into the inference. For computational convenience, we initially approximate the study-specific posteriors using a suitable parametric family. In our current example, we fit a bivariate normal distribution to each of the $J = 12$ posteriors. Inference for the joint model was carried out using Hamiltonian Monte Carlo implemented in the Stan software [3], and finally, the results (both overall effect and updated local effects) were refined using sampling/importance resampling [9] (further details in the appendix).

We compare MBA against results obtained using traditional random-effects meta-analysis (REMA), for which we use the conditional sum of squares estimate and a bootstrap covariance matrix as summary statistics. The hierarchical distribution on $\boldsymbol{\theta}_j$ follows Equations (5) and (6) above. Therefore, the essential difference between REMA and MBA is whether we combine this distribution with a likelihood function or with the observed posterior on $\boldsymbol{\theta}_j$. The results of the comparison are presented in Figures 2–4.[1]

Figure 2 shows the posterior distribution for the overall mean effect φ, obtained using four different models: in addition to MBA and REMA, we used fixed-effects meta-analysis (FEMA, which is a special case of REMA), and a 'naive' model corresponding to ordinary Bayesian inference using the means of the observed posteriors on $\boldsymbol{\theta}_j$ as observed data. As expected, FEMA is clearly inappropriate in this situation, and results in a heavily biased and overly confident posterior. The naive model is less biased than FEMA as it makes use of information contained in the observed posteriors, but compared to MBA, it does not properly account for the uncertainty contained in them. Both REMA and MBA result in posteriors with more spread, still assigning reasonably high probability mass to the neighbourhood around the true parameter value.

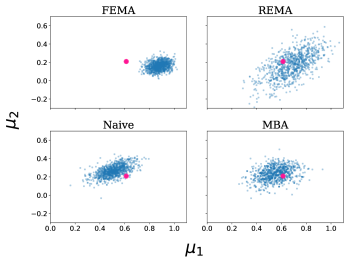

Figure 2: Posterior for the overall mean effect $\boldsymbol{\mu}$ using FEMA, REMA, a 'naive' model, and MBA. The red dot denotes the actual mean of the distribution used to generate the $\boldsymbol{\theta}_j$ values for the MA(2) series.

Figure 3 shows updated beliefs for the local effects plotted on top of the original ABC posteriors. The updated beliefs exhibit shrinkage towards the overall mean effect and, in this case, concentrate more accurately around the actual local effect values. On the other hand, many of the REMA posteriors for the local effects, shown in Figure 4, are biased and concentrate in regions away from true values.

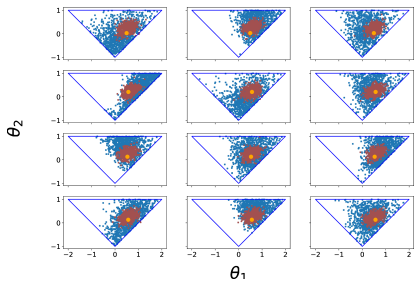

Figure 3: Updated posteriors for study-specific effects obtained using MBA (red) on top of the ones originally obtained using independent ABC's (blue). Yellow dots denote the true generating parameter values.

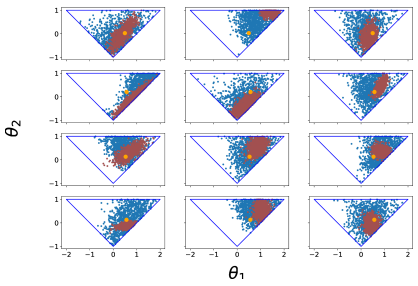

Figure 4: Posterior marginals for study-specific effects obtained using REMA (red) on top of the ones originally obtained using independent ABC's (blue). Yellow dots denote the true generating parameter values.

4 Conclusion

In this extended abstract, we have presented a framework which uses estimated posterior distributions to conduct meta-analysis. In many fields, it has become common practice to store data sets in dedicated repositories to be reused for the benefit of the entire research community. Given the view taken in this work, that posterior distributions can be seen as data observed with uncertainty, we believe that in many cases it would be equally beneficial to make full posterior distributions available for reuse. This would enable posteriors from potentially time-consuming and costly Bayesian analyses to be used as a basis for new studies. Indeed, even if the original data, the model and the code implementing it were available, reproducing posterior distributions could require a substantial computational effort. In addition to making posteriors publicly available, more research is needed

[1]The experiment was repeated with multiple random seeds, yielding similar results.

on developing methods to make appropriate use of the information they provide. The current work is a first step in this direction and our hope is that it will inspire other researchers to make further advances to this end.

Acknowledgments

The authors were funded by the Academy of Finland (grants 319264, 292334 and Flagship programme: Finnish Center for Artificial Intelligence FCAI).

References

[1] Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., and Rousseau, J. (2012). Combining expert opinions in prior elicitation. *Bayesian Anal.*, **7**(3), 503–532.

[2] Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons, Inc., Chichester.

[3] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *J. Statist. Soft.*, **76**(1).

[4] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, London, third edition.

[5] Higgins, P. T., Thompson, S. G., and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *J. R. Statist. Soc. A*, **172**(1), 137–159.

[6] Kuikka, S., Vanhatalo, J., Pulkkinen, H., Mäntyniemi, S., and Corander, J. (2014). Experiences in Bayesian inference in Baltic salmon management. *Statist. Sci.*, **29**(1), 42–49.

[7] Lintusaari, J., Gutmann, M. U., Dutta, R., Kaski, S., and Corander, J. (2017). Fundamentals and recent developments in approximate Bayesian computation. *System. Biol.*, **66**(1), e66–e82.

[8] Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statist. Comp.*, **22**(6), 1167–1180.

[9] Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling–resampling perspective. *The American Statistician*, **46**(2), 84–88.