Published in final edited form as:

Psychol Methods. 2017 December ; 22(4): 609–615. doi:10.1037/met0000169.

Bayesian Estimation and Modeling: Editorial to the Second Special Issue on Bayesian Data Analysis

Sy-Miin Chow and

Department of Health and Human Development, Pennsylvania State University, Penn State

Herbert Hoijtink

Department of Methodology and Statistics, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands

Abstract

This editorial accompanies the second special issue on Bayesian data analysis published in this journal. The emphases of this issue are on Bayesian estimation and modeling. In this editorial, we outline the basics of current Bayesian estimation techniques and some notable developments in the statistical literature, as well as adaptations and extensions by psychological researchers to better tailor to the modeling applications in psychology. We end with a discussion on future outlooks of Bayesian data analysis in psychology.

Keywords

Bayesian Estimation; Bayesian Modeling; Prior elicitation; Bayesian latent variable model; Bayesian dynamic model; Bayesian Lasso

In 2015 Psychological Methods published a call for papers for a special issue on Bayesian Data Analysis. Out of the 21 papers accepted for publication, the first 10 articles appeared in the June special issue of Psychological Methods, focusing on Bayesian data analysis in psychological research, Bayes factor, and posterior predictive p -values. This special issue presents the remaining 11 articles, grouped under the theme of Bayesian estimation and modeling.

Our key aims are to outline some basic terminology and ideas behind Bayesian modeling and estimation, summarize some helpful resources/references for the interested reader, and share some of our own thoughts on the past, present, and future use of Bayesian modeling/estimation techniques in the behavioral sciences. For more exhaustive introductions to Bayesian data analysis, the reader is referred to elsewhere (J. Albert, 2007; Daniels & Hogan, 2008; Gilks, Richardson, & Spiegelhalter, 1996; Carlin & Louis, 2008; Gelman, Carlin, Stern, & Rubin, 2003; Lynch, 2007).

Elicitation and Use of Prior Distributions

The basic idea behind Bayesian data analysis is one that is well covered in many of the articles in this special issue. Here we use a simple modeling scenario involving only fully observed data and a vector of unknown parameters to highlight some key concepts. Let θ be

a vector of parameters in the model of interest, with data array from n independent units (e.g., individuals), $\mathbf{Y} = \{y_i: i = 1, \dots, n\}$. Bayesian analysis is usually based on the joint posterior distribution of the parameters, $p(\boldsymbol{\theta}|\mathbf{Y})$, which depends on the likelihood, $p(\mathbf{Y}|\boldsymbol{\theta})$, and the prior distribution, $p(\boldsymbol{\theta})$, through the Bayes' theorem,

$$p(\boldsymbol{\theta}|\mathbf{Y}) = p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\mathbf{Y}) \propto p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (1)$$

Prior elicitation concerns the process of specifying the form and all associated hyperparameters that appear in the prior distribution, $p(\boldsymbol{\theta})$. This is a critical issue in the Bayesian modeling framework: Bayesian inference capitalizes heavily on summary statistics computed using $p(\boldsymbol{\theta}|\mathbf{Y})$; strong priors can have strong influence on inferential results, particularly in situations where insufficient information is available from the data likelihood to “overwrite” properties of the priors.

Many may regard the use of prior information as a double-edged sword. Even though various MCMC techniques routinely used in the estimation of Bayesian models have greatly expanded the repertoire of models that can be fitted using commonly available software tools, many researchers, particularly those rooted in the frequentist tradition, may feel reluctant to fully embrace the added computational advantages afforded by the use of prior distributions. Conjugate priors that lead to posterior distributions from the same family of distributions (e.g., both are normal distributions) are convenient choices but not always possible and may still involve difficult choices in the selection of hyperparameters (for examples of conjugate priors for use with parameters and other components in structural equation models, see Lee, 2007). Uninformative priors such as Jeffreys' prior (Jeffreys, 1961) and the reference prior (Bernardo, 2006) are used to express the (complete) lack of prior knowledge on particular parameters or unknowns of the model (for a review see Kass & Wasserman, 1996).

When Bayesian modeling and estimation techniques were first gaining traction in psychology, many of the studies focus on showing the similarities between frequentist and Bayesian estimation results under the use of minimally informative but “just proper” priors. Proper priors are priors that integrate to 1. Thus, for instance, a normal prior, $p(\boldsymbol{\theta}) = N(0, \nu)$, where ν is a large positive constant, is proper but as $\nu \rightarrow \infty$, the prior is improper. Another example is the g -prior proposed by Zellner (1986) used in the paper by Haaf and Rouder (this issue) to implement a Bayes factor model-comparison strategy for assessing whether the well-known Stroop effects obey theoretically driven equality and order restrictions. Such a prior specification worked well in the modeling context considered by Haaf and Rouder. However, in many empirical applications or scenarios involving models with higher complexity, “just proper” priors may not be practical, or may quickly morph into so-called “moderately informative priors” constructed based on selected information from the data.

Well-constructed informative priors that capitalize on prior knowledge drawn from historical data, pilot data, and expert opinion can yield considerable computational advantages. Unfortunately, historical studies or data with similar characteristics to a new study may not exist. In a similar vein, incorporating expert opinion into the construction of informative

priors is challenging. In particular, how and in what ways researchers are able to justifiably elicit informative priors to truly maximize the strengths of the Bayesian computational framework are not well understood and certainly not broadly implemented. To this end, several articles in this special issue provide illustrations on ways to elicit priors that reflect researchers' preexisting beliefs.

The paper by Bolsinova, Hoijtink, Vermeulen and Beguin (this issue) introduces several empirical procedures to elicit expert knowledge to construct informative priors to enhance the quality of linking procedures in the presence of sparse data. The paper by Moeyaert, Rindskopf, Onghena, and Van den Noortgate (this issue) compares Bayesian and frequentist techniques for fitting models to single-subject experimental design (SCED) data with very small sample sizes and discusses how such modeling endeavors may benefit from Bayesian modeling/data analytic techniques. These papers provide helpful insights on how researchers may construct informative priors from previous and carefully designed studies involving test and SCED data.

The paper by Ip, Pan and Dube (this issue) considers the use of Bayesian Lasso — implemented via specialized prior distributions — as a way to regularize or shrink the small residual covariances in confirmatory factor analysis (CFA) to zero. In this case, structured prior distributions mirroring the Lasso penalty function used in the frequentist context to perform variable/model selection are used to efficiently free up non-zero residual covariances to improve model fit. Similar applications of variable selection-motivated priors in quantitative psychology have also been considered elsewhere in confirmatory factor analytic (Lu, Chow, & Loken, 2016) and structural equation modeling (Muthén & Asparouhov, 2012) contexts. Such data-driven approaches for detecting mild model misspecification have their frequentist counterparts in approaches utilizing modification indices (Saris, Satorra, & Sorbom, 1987; Sörbom, 1989). Compared to their frequentist counterparts, the Bayesian approaches provide relatively straightforward ways to quantify the uncertainty involved in the model refinement process (Lu et al., 2016). However, researchers' take on the merits and justifiability of such an approach may be mixed. Our view is that such data-driven information may help reveal potential sources of model misspecification, refine model specifications in future studies, or as in the scenario considered by Harring, McNeish, and Hancock (this issue), helps absorb other sources of unmodeled shared covariances that, if left unaccounted, may bias the inferential results of central interest to the researchers.

Harring, McNeish, and Hancock consider the issue of performing sensitivity analysis in the presence of model misspecification, specifically, when there may be omitted variables. Omitted variables are incorporated into their model collectively as a phantom variable, and sensitivity analysis is performed by fixing all parameters linked to the phantom variable at substantively sensible values; as well as by assigning prior distributions to those parameters. Indeed, with the ever-increasing complexity of many Bayesian models, few issues are as pressing as the need to evaluate the sensitivity of one's estimation results to changes in the data, prior, and sampling distributions (Berger, 1994). The need for sensitivity analysis is also clear in the work of Du, Liu, and Wang (this issue), who propose a Bayesian meta-analytic approach for adjusting publication bias and estimating population effect size by

treating studies that do not get published (due e.g., to their lack of statistical significance) as information that is not missing not at random (Rubin, 1976).

Modeling and Estimation Innovations

Bayesian inference draws heavily on the computational power offered by Markov chain Monte Carlo (MCMC; Geman & Geman, 1984; Gelfand & Smith, 1990) methods. All MCMC methods essentially involve setting up a Markov chain with certain desirable properties to ensure that the stationary distribution — the distribution reached after transient fluctuations in the estimated properties of the posterior distribution have tapered off — converges to the joint posterior distribution of interest. Consequently, after some initial samples prior to convergence (the burn-in samples) have been discarded, the empirical distribution of the remaining MCMC samples can be taken to be an approximation of all the pertinent posterior distributions, based on which Monte Carlo integration is used to obtain summary statistics for inferential purposes. For instance, sample mean or median may be used as point estimates. Uncertainty in the point estimates may be summarized using standard deviations or credible intervals based on the percentiles of the MCMC samples.

Gibbs sampling forms the basis of many MCMC techniques. Suppose the only unknowns of interest in a model are the p parameters in $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$. The joint posterior distribution of $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{\theta}|\mathbf{Y})$, may be of high dimension and computations of summary statistics of interest may involve analytically intractable integrals. The Gibbs sampling is designed to circumvent this problem by sampling iteratively, for $t = 1, \dots, T$, where T is large, from a sequence of conditional distributions:

$$\text{Sample from } p(\theta_1^t | \theta_2^{t-1}, \dots, \theta_p^{t-1}, \mathbf{Y})$$

$$\text{Sample from } p(\theta_2^t | \theta_1^t, \dots, \theta_p^{t-1}, \mathbf{Y})$$

...

$$\text{Sample from } p(\theta_p^t | \theta_1^t, \dots, \theta_{p-1}^t, \mathbf{Y}). \quad (2)$$

Each conditional distribution above is referred to as a full conditional distribution. That is, the full conditional distribution for a parameter $j \in \boldsymbol{\theta}$ is defined as the distribution of θ_j given $(\{\theta_k : k \neq j\}, \mathbf{Y})$. With sufficient burn-in iterations and under some general regularity conditions, the samples drawn using the Gibbs sampling algorithm in Equation (2) will converge to the target distribution of interest, namely, the joint posterior distribution of $\boldsymbol{\theta}$. Another useful technique in estimation using MCMC methods is data augmentation (Tanner & Wong, 1987; J. H. Albert & Chib, 1993). Under data augmentation, the sampling steps are extended to include sampling of additional latent variables (e.g., latent factors, ability,

random effects, missing data) so that the distributions from which samples are drawn reduce to tractable forms and convergence to the target joint posterior distribution can be accelerated.

The paper by Natesan and Hedges (this issue) presents an application of Bayesian change point models to answer questions in the context of data from single-subject designs. These authors used the Gibbs sampler as implemented in JAGS (Plummer, 2003) for model estimation purposes. As in the paper by Moeyaert and colleagues, who used the PROC MCMC procedure in SAS to perform Gibbs sampling estimation, Bayesian estimation techniques are a natural choice in this context given the finite (small) sample properties of data from single-subject designs.

There may be situations where sampling from one or more of the conditional distributions in Equation (2) is challenging. One possible way to circumvent this issue is to replace such steps with Metropolis-Hasting (MH) algorithms embedded within the Gibbs sampler. At the t th iteration of the Gibbs sample, suppose one is interested in sampling from the full

conditional distribution of the j th parameter, $p(\theta_j^t | \{\theta_l^{t-1} : l < j\}, \{\theta_l^{t-1} : l > j\}, \mathbf{Y})$,

abbreviated below as $p(\theta_j^t)$, but this distribution does not conform to any of the known distributional forms and it is thus impossible to sample directly from it. However, suppose one can sample relatively easily from a proposal density,

$g(\theta_j^t | \theta_j^{t-1}, \{\theta_l^{t-1} : l < j\}, \{\theta_l^{t-1} : l > j\}, \mathbf{Y})$, abbreviated below as $g(\theta_j^t | \theta_j^{t-1})$, that is constructed based on properties of the sample of θ_j^{t-1} from the previous iteration. The MH algorithm can then proceed by first simulating a potential sample value, θ_j^t , from the proposal density, and

then accepting the newly sampled value with probability $\min \left\{ 1, \frac{p(\theta_j^t)g(\theta_j^{t-1} | \theta_j^t)}{p(\theta_j^{t-1})g(\theta_j^t | \theta_j^{t-1})} \right\}$.

The quality of the Bayesian estimation when the MH is used depends heavily on good choices for the proposal density, $g(\theta_j^t | \theta_j^{t-1})$. Adaptations/extensions of the MH algorithm have also been proposed in the statistical literature. For instance, another common technique for handling Bayesian change-point models with multiple unknown change points is reversible jump MCMC (e.g., Johnson, Elashoff, & Harkema, 2003), a generalization of the MH algorithm that allows the dimension of the parameter space to change, thus allowing transition between models that differ in their number of parameters (Green, 1995).

The paper by Burgette (this issue) presents a model for representing the changes observed in therapy groups admitted on a rolling basis as evidenced on discrete as well as continuous outcomes. MCMC techniques offer a good alternative for handling more complex models such as the ones considered by the authors. The paper also shares new insights into the superior performance of the No-U-Turn Sampler (NUTS) in Stan (Carpenter et al., 2017; Hoffman & Gelman, 2014; Stan Development Team, 2016) compared to a relatively straightforward implementation of the random-walk MH algorithm, one variation of the MH algorithm described above. The NUTS is a variant of a class of MCMC methods called the Hamiltonian/Hybrid Monte Carlo, which includes a set of tuning parameters to be updated

in the estimation process together with other unknowns in the model. Concurrent updates of these tuning parameters improve the efficiency of the MCMC algorithm, and the NUTS enables automated selection of these tuning parameters.

Yuan and Huang (this issue) propose a Bayesian multilevel dynamic mediation model with time-varying coefficients. A nonparametric penalized spline approach is used to flexibly approximate possible over-time changes in the coefficients for direct as well as indirect (mediated) effects. A blocked Gibbs sampling approach is used to allow highly correlated parameters to be updated within the same step or block in the Gibbs sampler. Uncertainty of the indirect effect, which is typically quantified by means of summary statistics of bootstrap samples in the frequentist framework (Bollen & Stine, 1990; Shrout & Bolger, 2002), can now be deduced from summary statistics of the empirical posterior distribution.

Zhang and Wang (this issue) consider the difficulties of performing inference in the context of moderated regression models (i.e., regression models involving interaction effects between independent variables) in the presence of data that are missing completely at random and missing at random. The Gibbs sampler with adaptive rejection sampling (Gilks & Wild, 1992) in OpenBUGS was used for Bayesian estimation purposes and was found to outperform alternative techniques such as full-information maximum likelihood and multiple imputation. Several MCMC variants are available in OpenBUGS. In the context of this specific model, adaptive rejection sampling was used when sampling from any particular full conditional distribution in (2) is not possible; instead, samples are drawn from an approximation density defined by upper and lower bounds (known as envelope and squeezing functions, respectively) that encapsulate the full conditional distribution from the top and the bottom.

The paper by Levy (this issue) highlights the importance of distinguishing between indicators for latent variables and outcome variables and proposes a Bayesian approach for handling the associated computational issues and interpretational confounds. The proposed, modified Gibbs sampler approach is not fully Bayesian in the sense that the distribution used for inferential purposes is not the joint posterior distribution implied by the hypothesized model but rather, an “approximation” with particular conditional dependencies deliberately removed to eliminate the interpretational confounds. This paper showcases one heuristic use of MCMC techniques to circumvent a data analytic problem originated from a frequentist framework.

Conclusions and Outlooks

This special issue features applications of Bayesian modeling techniques to a rich array of modeling contexts: regression models, item response models, structural equation models, dynamic mediation models, dynamic group membership models, and change point models. Several software programs – such as OpenBUGS, JAGS, Stan, SAS, and in-house estimation routines – are used, with their associated MCMC algorithms. We have attempted to provide brief explanations of the basic ideas behind these MCMC algorithms to facilitate decisions on future applications, extensions, and software selection. As much as possible, we have encouraged authors to provide annotated codes and examples. These can be accessed either

in the appendix sections of the papers, or as supplementary materials on the journal website. Some publications comparing the strengths and limitations of existing MCMC software exist (e.g., Kruschke, 2015; Monnahan, Thorson, & Branch, 2017), but due to the fast-evolving nature of some of these programs, conclusions concerning each software's strengths and limitations may vary or have to be adapted in matters of weeks or months. As in the editorial for the first special issue, we do not claim the collection of articles appearing in this special issue to be exhaustive – either in terms of content or technical coverage. Nevertheless, we hope that these articles serve as useful additions to the repertoire of Bayesian tools and resources available to psychological researchers.

Translational Abstracts

Using Phantom Variables in Structural Equation Modeling to Assess Model Sensitivity to External Misspecification

Jeffrey R. Harring, Daniel M. McNeish and Gregory R. Hancock

Missing translational abstract

Distinguishing outcomes from indicators via Bayesian modeling

Roy Levy

Latent variables are commonly used to acknowledge and model the presence of measurement error, and are often used to predict outcome variables of interest. A problem arises in the common use of such latent variable models, wherein the outcome variables a researcher is interested in predicting exert unwanted influence on the definition of the latent variables. This study proposes and studies new methods to preclude this unwanted influence, including methods to detect when such an unwanted influence may be present. Through analyses of simulated and real data, the new methods are shown to outperform existing approaches in the situations studied here. The new methods can be implemented in freely available software, and it is argued that researchers should consider using them in situations when they wish to prevent the outcomes from influencing the latent variables that predict them.

Moderation Analysis with Missing Data in the Predictors

Qian Zhang and Lijuan Wang

Missing translational abstract

Bayesian Dynamic Mediation Analysis

Ying Yuan and Jing Huang

Mediation analysis is a statistical method that help researchers understand how the effect of an independent variable (e.g., an intervention program) on the dependent variable (e.g., severity of depression) is mediated through a mediating variable (e.g., improvement in coping skill). Most existing methods for mediation analysis assume that mediation is a stationary, time-invariant process, which overlooks the inherently dynamic nature of many

human psychological processes and behavioral activities. In this article, we consider mediation as a dynamic process that continuously changes over time, and propose novel Bayesian statistical models to describe and estimate such dynamic mediation effects. The proposed method is flexible and capable of faithfully reflecting the true nature of the mediation process. It provides a valuable tool to help researchers obtain a more complete understanding of the dynamic nature of the mediation process underlying psychological and behavioral phenomena.

An Alternative to Post-Hoc Model Modification in Confirmatory Factor Analysis: The Bayesian Lasso

Edward Ip, Junhao Pan and Laurette Dube

Confirmatory factor analysis (CFA) is a commonly used tool by psychologists to determine the underlying dimensions of a survey. The “Big Five”, which were 5 key traits that were thought to characterize one’s personality, were identified using CFA. CFA requires strong model assumptions, which sometimes are not realistic because the restrictions could lead to poor fit between real data and the model. Existing literature tends to use ad hoc methods to fix the problem. Instead of using a patching-up approach, this study describes a scientifically justifiable way to relax model assumptions. Based on sound statistical principles, the proposed method adopts analytic techniques from the machine learning field to solve the potentially complex interrelatedness problem between survey items. By using both simulation experiments and a real data set regarding emotion measurement, it is shown that the method can be used to help researchers implement CFA in more flexible ways.

Using expert knowledge for test equating

Maria Bolsinova, Herbert Hoijtink, Jorine Adinda Vermeulen and Anton Beguin

If each year a new version of an educational test is used then the results of the new version are not directly comparable to the results of the reference test version due to the difference in difficulty of the two tests and the differences in the ability of the new and reference populations of students. Typically extra data (other than the examination data) are collected to compare the difficulty of the items in the two tests. However, in high stakes testing, where the amount of data that are available to link the reference and the new test are limited due to security reasons, these extra data do not provide enough information to obtain the desired levels of certainty. In this article we argue that linking data are not the only source of information about the difference in the difficulty of the two test forms. Experts may also provide information about this difference. In the study we propose and evaluate two methods for elicitation of the prior knowledge about the difference in difficulty of two tests from subject-matter experts. The results prove the utility of the proposed methodology, since the precision of the linking results increases without the increase in bias.

Bayesian models for semicontinuous outcomes in rolling admission therapy groups

Lane F Burgette and Susan M. Paddock

Alcohol and other drug abuse are frequently treated in a group therapy setting. If two clients attend the some of the same sessions, we might expect that $\hat{\theta}$ on average $\hat{\theta}$ their post-treatment outcomes would be more similar than if they had not attended any sessions together. Hence, if participants are allowed to enroll in therapy on a rolling basis, irregular patterns of session attendance can induce complex relationships between participant outcomes. Statistical methods have been developed previously to account for rolling admission group therapy when the outcomes are normally-distributed. In the case of alcohol and other drug use interventions, however, a substantial fraction of participants often report zero use after treatment. We extend previous work to build models that accommodate semicontinuous outcomes, which are a mixture of continuous and discrete distributions, for such situations. We find that modern Bayesian statistical methods and software allow users to efficiently estimate non-standard models such as these. We illustrate our methods using data from a group-based intervention to treat substance abuse and depression, focusing on the outcome of average number of drinks per day. We find that the intervention is associated with a drop in the probability of any drinking, but find no evidence of a change in the amount of drinking, conditional on some drinking.

Bayesian Unknown Change-Point Models to Investigate Immediacy in Single Case Designs

Prathiba Natesan and Larry V. Hedges

An immediate change in the dependent variable following introduction of treatment is called immediacy. Immediacy is a necessary condition to show strong evidence of causality in single case designs (SCDs). Currently no objective and inferential statistical tools are used to investigate immediacy. We propose a Bayesian unknown change-point model to investigate and quantify immediacy in SCD analysis. This model assumes that the change-point between baseline and treatment phases is unknown and estimates the change-point. Unlike visual analysis that considers only 3–5 observations in consecutive phases to investigate immediacy, this model considers all data points. Immediacy is indicated when the change-point is estimated accurately and with less uncertainty. This model can accommodate delayed effects where immediacy may not be possible due to the nature of some treatments taking longer to take effect. The use of Bayesian methods helps overcome some of the most common drawbacks of single case designs such as small sample sizes and lack of independence of observations. The model performs well for data with at least 8 observations per phase and a mean difference of at least 3 standard deviations between the baseline and treatment phases. This method is illustrated with real data.

Multilevel Modeling of Single-Case Data: A Comparison of Maximum Likelihood and Bayesian Estimation

Mariola Moeyaert, David Rindskopf, Patrick Onghena and Wim Van den Noortgate

Variance component estimates in contexts of two-level modeling of single-case experimental data tend to be biased and imprecisely estimated. In this study, Bayesian estimation is proposed as an alternative to a maximum likelihood estimation procedure. Prior selection methods and prior construction for the variance components in the Bayesian framework are described. The purpose is to compare parameter recovery of maximum likelihood estimation

and Bayesian estimation using Monte Carlo simulation methods. Both maximum likelihood and Bayesian estimation procedures result in biased and less precise variance estimates when the number of participants is small (i.e., 3). By increasing the number of participants to 5 or 7, unbiased and more precise estimates are obtained for all priors, except for the inverse-Wishart prior using the identity matrix. When a more informative prior is added, more precise estimates were obtained, even when only three participants were included. An empirical example is given to illustrate the Bayesian approach. We recommend that applied researchers use a variety of different prior distributions (following the proposed prior selection and constriction) and discuss to what extent and in what sense the results depend on the prior used.

Developing Constraint in Bayesian Mixed Models

Julia M. Haaf and Jeffrey N. Rouder

A common question in psychological science is whether there is an effect for a certain task. With “an effect”, researchers usually refer to an average effect across trials and participants. For example, a Stroop effect is the difference between the average reaction time in the congruent condition where the word “red” is presented in red, and the incongruent condition where the word “green” is presented in red. Here, we investigate the question whether everyone shows such an effect. To do so, we develop four different Bayesian models each representing different constraints on individual participants’ effects: 1. That no one has an effect; 2. that everyone has the same effect; 3. That everyone has an effect in the same direction; and 4. that effects can vary in size and direction. We apply these four models to seven data sets from Stroop, Simon and Eriksen tasks. We find that everyone shows a Stroop effect, but not everyone shows a Simon effect in the same direction, and, for the data set analyzed here, no one shows an Eriksen effect. The Bayesian modeling framework developed here can be useful for researchers in the fields of cognitive and social psychology, in particular when studying individual differences. The analysis can serve as a check for whether observed individual differences in a data set may be meaningful or not, and whether cognitive processing may differ across people.

A Bayesian “fill in” method for correcting for publication bias in meta-analysis

Han Du, Fang Liu and Lijuan Wang

Missing translational abstract

Acknowledgments

The first author is supported by NSF grant BCS-0826844, and NIH grant R01GM105004.

The second author is supported by the Consortium on Individual Development (CID) which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). We would like to thank Dr. Lisa Harlow, Dr. Fred L. Oswald, and Meleah Ladd for providing us with many helpful comments and assistance throughout the review and editorial processes.

References

- Albert, J. Bayesian computation with r. New York: Springer; 2007.
- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993; 88:669–679.
- Berger JO. An overview of robust bayesian analysis. *Test*. 1994; 3:5–58.
- Bernardo, JM. Handbook of statistics. In: Dey, D., Rao, C., editors. Reference analysis. Elsevier; 2006. p. 17-90.
- Bollen KA, Stine R. Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*. 1990; 20:115–140.
- Carlin, BP., Louis, TA. Bayesian methods for for data analysis. Boca Raton, FL: Chapman & Hall/CRC; 2008.
- Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Riddell A. Stan: A probabilistic programming language. *Journal of Statistical Software*. 2017; 76doi: 10.18637/jss.v076.i01
- Daniels, MJ., Hogan, JW. Missing data in longitudinal studies: Strategies for bayesian modeling and sensitivity analysis. London: Chapman and Hall; 2008.
- Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 1990; 85:389–409.
- Gelman, A., Carlin, JB., Stern, H., Rubin, DB. Bayesian data analysis, second edition. Boca Raton, FL: Chapman & Hall/CRC; 2003.
- Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1984; 6:721–741.
- Gilks, WR. Richardson, S., Spiegelhalter, DJ., editors. Markov chain Monte Carlo in practice. London: Chapman & Hall/CRC Press; 1996.
- Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*. 1992; 41:337–348.
- Green PJ. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*. 1995; 82:711–732.
- Hoffman MD, Gelman A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research*. 2014; 15(1):1593–1623.
- Jeffreys, H. Theory of probability. 3rd. Oxford University Press; 1961.
- Johnson T, Elashoff R, Harkema S. A bayesian change-point analysis of electromyographic data: detecting muscle activation patterns and associated applications. *Biostatistics*. 2003; 4(1):143–164. [PubMed: 12925335]
- Kass RE, Wasserman L. The selection of prior distributions by formal rules (corr: 1998v93 p412. *Journal of the American Statistical Association*. 1996; 91:1343–1370.
- Kruschke, J. Doing bayesian data analysis: A tutorial with R, JAGS, and Stan. Second. Boston: Academic Press; 2015. doi: <https://doi.org/10.1016/B978-0-12-405888-0.09999-2>
- Lee, SY. Structural equation modeling: A bayesian approach. West Sussex, England: Wiley & Sons, Ltd; 2007.
- Lu ZH, Chow SM, Loken E. Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*. 2016; 51(4):519–539. Retrieved from <http://dx.doi.org/10.1080/00273171.2016.1168279>. DOI: 10.1080/00273171.2016.1168279 [PubMed: 27314566]
- Lynch, SM. Introduction to applied bayesian statistics and estimation for social scientists. New York, NY: Springer; 2007.
- Monnahan CC, Thorson JT, Branch TA. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*. 2017; 8:339–348.
- Muthén BO, Asparouhov T. Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*. 2012; 17(1):313–335. [PubMed: 22962886]
- Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*. 2003

- Rubin D. Inference and missing data. *Biometrika*. 1976; 63(3):581–592.
- Saris WE, Satorra A, Sorbom D. The detection and correction of specification errors in structural equation models. *Sociological Methodology*. 1987; 17:105–129. Retrieved from <http://dx.doi.org/10.2307/271030>. DOI: 10.2307/271030
- Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*. 2002; 7:422–445. [PubMed: 12530702]
- Sörbom D. Model modification. *Psychometrika*. 1989 Sep 27; 54(3):371–384. Retrieved from <http://dx.doi.org/10.1007/BF02294623>. DOI: 10.1007/BF02294623
- Stan Development Team. RStan: the R interface to Stan. 2016. Retrieved from <http://mc-stan.org/> (R package version 2.14.1)
- Tanner M, Wong W. The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*. 1987; 82:528–550.
- Zellner, A. On assessing prior distributions and bayesian regression analysis with g -prior distribution. In: Goel, PK., Zellner, A., editors. *Bayesian inference and decision techniques: Essays in honour of Bruno de Finetti*. Amsterdam: North Holland; 1986. (p. 233â243)