A framework for Bayesian Posterior Simulation Methods in clinical practice

Razieh Bidhendi Yarandi[1], mohammad ali Mansournia[1], hojjat Zeraati[1], and Kazem Mohammad[1]

[1]Tehran University of Medical Sciences

August 28, 2020

Abstract

Purpose: Bayesian inference has become very popular science. It offers some pragmatic approaches to account for uncertainty in inference-decision making. Various estimation methods have been introduced to implement Bayesian methods but although these algorithms are powerful they are not always easy to grasp. This paper aims to provide an intuitive framework of four key Bayesian computational methods for researchers in clinical studies. We will not cover daunting mathematical discussion of these approaches, but rather offer a non-quantitative description of these algorithms and provide some illuminating examples. Materials and methods: Bayesian computational methods namely, i) Importance sampling, ii) Rejection sampling, iii) Markov-chain Monte Carlo, iv) Data augmentation were introduced. Results and conclusions: A load of literature published on Bayesian inference has proved its popularity among researches while its concept is not straightforward for amateur learners. We showed that alternative approaches which are intuitively appealing and easy-to-understand work well in case of low-dimensional problems and appropriate Prior information such as weighted prior, otherwise MCMC is a Trouble-free tool.

INTRODUCTION

The growth in the application of Bayesian analysis in the sciences has created a need to present its complex concepts in a more understandable language for non-statisticians. To many amateur users, their computationally-intensive simulation approaches has the appearance of a "black box". This article aims to describe essential approaches used in Bayesian methods for posterior simulation in an intuitive manner. Four Bayesian computational methods are presented: *Importance Sampling (IS)* , *Rejection sampling (RS), Markov Chain Monte Carlo (MCMC)* and *Data Augmentation (DA)* . In this study, we aim to provide a comprehensive, yet easy to follow explanation of these techniques to make them practical for researchers. Some illuminating examples are presented to illustrate the algorithms and the concepts they embody. R software code and some other information are available in the Supplementary file as well.

METHODS

Historically, one of the first methods of simulation by computer was Monte Carlo integration (MCI). Its computational algorithm relies on repeated random sampling from distributions of interest to obtain numerical results. For example, MCI is used to calculate the expectation of a distribution function when its estimation through integration is impossible. Its plain logic is to generate samples from a distribution function and approximate the expectation numerically by calculating the average of generated samples. In other words, empirical experimental summation is substituted for analytical (or possibly numerical) integration. This technique is conceptually simple but not always very efficient if the target probability function is not well-behaved. Attempts to deal with this problem led to a gradual evolution first to IS and next in RS; they can be regarded as improved versions of MCI. However, in some more challenging cases they have also proven inadequate to the task. Over the past 30 years or so, MCMC methods have revolutionized statistical

computing and permitted ever more complex problems to be handled. Finally, we have included DA which follows a different approach to calculate posterior estimates.

Importance Sampling: A clever substitution of sampling region

MCI as a computational method was first initiated to solve the integration problem in estimating expectations. Later it was applied to the simulation of Bayesian posterior (BP) distributions. It has a transparent algorithm: Generate random samples from a distribution function say "Target", then numerically calculate the integral by summing the values. The expectation obtained by MCI can be referred to as the *Empirical Average* (Supplementary Formula-1). For example, imagine we want to estimate the expectation of the function$h\left(x\right)=\sin\left(x\right)\sqrt{\left|\cos\left(x\right)\right|}$, where the random variable (X) follows a Normal distribution with mean=0 and SD=5 (target). Firstly, generate 10000 random samples from target distribution, then obtain the values of h(x) from each generated sample, and then calculate numerically the mean and variance of this generated sample (Supplementary R Code-1). This random sampling method is not cost-effective when the target distribution is diffuse, because a large sample size is required to obtain acceptable precision.

MCI was improved by IS, a variance reduction technique which was first presented in statistical physics (1; 2). IS relaxes the procedure of treating all parts of the distribution equally, concentrating instead on those where estimation was critical. In this respect, an alternative function, say "Proposal", close to the target, is suggested by making an educated guess. Contrary to MCI on which samples are treated evenly, a "weight" which shows the importance of a sample is allocated to each generated sample through an importance function. Actually, for each sample, one calculates the likelihood of getting that sample from the target distribution proportion to the likelihood of sampling it from the proposal distribution. After the sampling process is finished, the obtained relative likelihoods were normalized in a way that they sum to one. In this way, each point has its own likelihood of occurrence as a discrete probability distribution. The expectation obtained by IS called *Weighted Average* (Supplementary Formula-2).

Consider a Normal (0, 0.05) and a t-student (dF=1) as the two different proposal distributions for Normal (0, 5), and estimate importance weights through the importance functions$\frac{target:\ N(0,5)}{proposal:\ N(0,0.05)}$ and $\frac{N(0,5)}{t(1)}$ for each generated point. After that, normalize the weights as$\frac{\frac{N(0,5)}{N(0,0.05)}}{\sum\frac{N(0,5)}{N(0,0.05)}}$and $\frac{\frac{N(0,5)}{t(1)}}{\sum\frac{N(0,5)}{t(1)}}$. Therefore, we have a discrete distribution function for which its properties such as the mean and variance are easily estimable. Estimated means for our generated samples were approximately 0 as obtained from MCI but their variances are considerably lower than the MCI approach (Supplementary R Code-2). Using alternative distributions can improve the variance of the samples, although a wide proposal distribution leads to worse estimates in terms of the variance and inefficient due to a large sampling number (Figure1). Choosing an appropriate proposal distribution that looked similar to the target would be ideal though difficult to find at times. Unbiased estimates of parameters are obtained for large samples by IS. It also works well when the importance function is not very variable. Indeed, an appropriate proposal distribution leads to lower variances and higher accuracy of approximation. Robert and Casella provided an example illustrating the use of Normal (0, 1) as a proposal to resemble sampling from a Cauchy C(0,1) target distribution caused infinite variance of importance weights (3). It leads to attach high importance to few points and provides inefficient estimates in terms of variance. By substituting heavy tails distributions like t-student rather than Normal, reasonable fitness is guaranteed.

Weighted Prior as an application of IS in BP simulation

In this section, we show how IS plays a role in BP simulation. As Bayes rule, posterior distribution is proportional to the likelihood of observed data and prior distribution of parameter (Supplementary Formula-3). It emphasizes that likelihood can play the role of importance function for priors or for a prior. Using this approach, the expectation of a posterior distribution function can be calculated. In addition, to paint a clearer picture of the posterior density and its properties- such as percentiles-it suffices to generate weighted samples from the prior.

This follows a straightforward algorithm: 1) Generate a random sample from a prior distribution. 2) Consider

the values of the likelihood of the data in this generated sample as the weight and continue this process, 3) Normalize the weights. Consequently, each point has its own probability of occurrence by the likelihood of data. This naïve sampling method which we referred to as Weighted Prior (WP) is equivalent to IS. As such IS method provides an alternative way to exact simulation from posterior (Supplementary Formula-4). It is worth mentioning that, bias estimated through this approach is of an order of $\frac{1}{n}$ so in large samples it provides unbiased estimates. It should be considered that, the contribution of generated samples to the estimation of the final posterior depends on how much it is supported by data. This means that, prior distributions should not be that much far off from the likelihood if so, most generated points receive very small importance weights. We depicted a case when this method fails (Figure 2).

Rejection Sampling: Hit the Target!

Contrary to IS which has no restriction to the choice of proposal distribution, RS is pickier. Imagine that our objective is to generate samples from a function. The vital condition is that proposal distributions should cover the target. Consider $g(\theta)$ as a proposal function from which samples generated, (Supplementary 1. Rectangle), $f(\theta)$ as target and K as a known constant (Supplementary Figure 1. Zigzag triangle). The algorithm is straightforward, generate sample ϑ_0 from a proposal and calculate the likelihood of target and proposal functions at this point. If the likelihood ratio of target to proposal was greater than a random sample generated from uniform (0, 1), we acceptθ_0 as a sample from target otherwise, reject it and continue the algorithm. In a Bayesian context, this idea could extend to estimating the posterior distribution by considering K as the Maximum likelihood of ϑ and prior as proposal, the probability of each generated sample is compared to the value of maximum likelihood $\mathrm{L_{Max}}$ so the proposal resembles posterior distribution, (Supplementary Formula-5) and (Supplementary figure 2). This method fails in high dimensional space parameters due to a decrease in the acceptance rate. In addition, when $Kg(\theta)$ has considerable distance compared to $f(\theta)$, acceptance rate would decrease drastically as well. Likewise, in Bayesian context when L_{Max} is not practically possible to calculate, this method fails. Armitage presented an example in which by this approach posterior estimation of parameters of a linear regression model had been estimated via various priors as the proposals(4), that using a diffuse prior caused high rejection rate and proved the approach futile.

Markov Chain Monte Carlo method (MCMC): Transition from uncertainty to stationary state.

The emergence of MCMC approach in the 1990s led to a rapid evolution of methods to simulate posterior distributions. It can be regarded as MCI through Markov chains. Contrary to other methods that have a static mechanism, MCMC follows a dynamic mechanism whereby samples are generated via a gentle transition through a target distribution function by considering a proposal, eventually converging on a stationary distribution. There are two popular algorithms: Metropolis-Hastings firstly introduced by Metropolis in 1953 and its special case the Gibbs sampler, introduced by Geman and Geman in 1984. Recent developments in this era have provided huge literature, Armitage provided a neat catalogue of the references and summaries (Armitage, Berry, and Matthews, 2001; Armitage et al., 2005). Let us review some technical jargon in MCMC. A chain with the property of being Markov is applied to generate samples which in consequence are dependent. A transition matrix illustrates the probability of movement from one state in the chain to another. It is worth mentioning that the Markov chain should have some properties that guarantee to produce samples from a stationary distribution. By stationary, we meant that, if a sample is generated from a distribution, the next is from the same distribution, as well. Firstly, it should be *Irreducible* meaning that the Markov chain can get to any state from any state within a finite number of iterations, *Persistent or recurrent* means returning to the state at least once and *Non-Null* means finite mean number of transitions. These three conditions had led to a property referred to*ergodicity* by which one can ensure the generation of samples from a stationary distribution. Therefore, the variance of the generated samples is estimable otherwise the chain may behave badly and be effectively useless. It also proves the consistency of estimates. Failure of this method occurs when there are convergence issues. For instance, in the case of a non-persistent

chain, convergence to a stationary distribution never occurs. Being *Symmetric* is another property which influences the acceptance probability of sampling. It means that$g\left(\theta''\middle|\theta'\right) = g\left(\theta'\middle|\theta''\right)$, (Supplementary figure 3).

The Metropolis-Hastings algorithm to produce a chain of samples by iterative mechanism is defined as the following steps;

1. Generate a candidate $\theta^{(')}$ from proposal distribution g $\left(.\middle|\theta^{(i)}\right)$on which $\theta^{(i)}$ is an initial point. The next step is to see whether it is an acceptable sample with a high probability of occurrences and accept it as$\theta^{(i+1)}$ or not.
2. An acceptance rule akin to rejection sampling should be considered here. In this way, at first acceptance probability of the candidate should be estimated. It is the minimum of one and the value of the proportion of multiplication of proposal and target distribution in candidate sample condition on previous point (initial) and vice versa. When the proposal is symmetric, it reduces to the proportion of the values of target on the candidate and initial points (proposal eliminated from nominator and denominator (Supplementary Formula-6).
3. Now proceed to decide whether to accept the candidate as the next sample or not. So, generate a random number u from uniform (0, 1) distribution.
4. If u was greater than acceptance probability, choose the candidate as the next sample, otherwise put initial as the next sample and continue the process.

Example: Accepted candidate

Step 1:$r(\beta_{\text{new}},\ \beta_{t-1}) = \frac{Posterior(\beta_{\text{new}})}{Posterior(\beta_{t-1})} = \frac{Beta(1,1,0.4)\times Binomial(10,4,0.4)}{Beta(1,1,0.5)\times Binomial(10,4,0.5)} = 1.19$

Step 2: Acceptance probabilityα $(\beta_{\text{new}},\ \beta_{t-1}) = min(1,\ r(\beta_{\text{new}},\ \beta_{t-1})) = min(1,\ 1.19) = 1$

Step3: Draw a random number, u , from a Uniform (0, 1), here u=0.345

Step4: If u is less than the acceptance probability, the proposed value of β_{new} will be accepted. Otherwise, we rejectβ_{new} and keep, β_{t-1}. Here we accept it.

Example: Rejected candidate

Step 1: $r(\beta_{\text{new}},\ \beta_{t-1}) = \frac{Posterior(\beta_{\text{new}})}{Posterior(\beta_{t-1})} = \frac{Beta(1,1,0.2)\times Binomial(10,4,0.2)}{Beta(1,1,0.3)\times Binomial(10,4,0.3)} = 0.43$

Step 2: Acceptance probabilityα $(\beta_{\text{new}},\ \beta_{t-1}) = min(1,\ r(\beta_{\text{new}},\ \beta_{t-1})) = min(1,\ 0.43) = 0.43$

Step3: Draw a random number, u , from a Uniform (0, 1), here$u$ =0.675

Step4: Since u > r, we reject it with the probability 43%.

MCMC by Metropolis-Hastings Algorithm: An Intuitive illustration

How samples are generated by MCMC, is the main question of interest. Imagine we have a uniform distribution (0, 5) as proposal, (Dashed line) and a target, (Zigzag pattern) (Supplementary figure 4). To show how to estimate this transition matrix, we considered a discrete target distribution as well as two numbers of states involved. Therefore, we need a 2×2 transition matrix exists whose elements are the probability of movement. To explain clearly, four possible transitions for generating samples exist. If the first sample generated from state 1, what the probabilities of being in the next state 1 or 2 would be, and so when the first in 2. According to the proposal, the probabilities of being in state 1 and 2 are$\frac{1}{5}$ and $\frac{4}{5}$, respectively. Therefore, estimated Transition Matrix is P=

$\begin{bmatrix} \frac{13}{15} & amp; \frac{1}{5} \\ \frac{2}{15} & amp; \frac{4}{5} \end{bmatrix}$ (Table1). If the process of sampling is repeated n times, based on stochastic process the n^{th}-step transition matrix will be $P^n = \frac{1}{\alpha+p}$

$\begin{bmatrix} \alpha & amp;\alpha \\ p & amp;p \end{bmatrix}$

$\begin{bmatrix} \frac{\mathbf{3}}{\mathbf{5}} & amp;\frac{\mathbf{3}}{\mathbf{5}} \\ \frac{\mathbf{2}}{\mathbf{5}} & amp;\frac{\mathbf{2}}{\mathbf{5}} \end{bmatrix}$. It is proven this sampling mechanism converges to a stationary form which shows target distribution (Supplementary Figure 4).

According to the stochastic process theorem, from this convergence time onwards each sample generated is independent of previous states(5). From the Bayesian point of view, considering proposal as prior distribution and target as prior × likelihood, this process of sampling is defined as Metropolis-Hasting which eventually converges to a stationary posterior.

Data Augmentation: Translation of Common Sense into Reality

Methods introduced in the past followed the Monte Carlo approach for computing posterior distribution. On the contrary, DA allows approximate Bayesian analysis with a standard maximum likelihood function. Its philosophy is to translate prior information as equivalent data and add this external information to the observed study data then conventional methods of frequentist can be applied. No specific tools are required to compute posterior mean and variance; inverse-variance weighted averaging is a rule of thumb for estimation (6). This technique provides an effective remedy to treat bias estimation caused by data sparseness (7-12). In fact, it considers prior information as a penalty for maximum likelihood estimates and approximates posterior mode and variance.

Learn Posterior Estimation by Heart: An illustration in pharmacology

DA is a remedial tool for sparse data issues to provide unbiased estimation of parameters. To illustrate DA mechanism, we considered how to estimate posterior properties via inverse-variance weighting, and then show the influence of prior and likelihood components on estimating posterior distributions. Finally, we depict how to construct data from prior to observe its role as equivalent data augmented to actual data.

Result of a hypothetical data was reported Ln(RR)=Ln (6.3) =1.82, Variance(Ln(RR)) = 0.84 and 95% limits RR= (1.02 , 37.3) which was subjected to sparse data due to wide 95% limits . Suppose that, in a meta-analysis for side effect of a drug study we found prior information for RR with 95% limits between $\frac{\mathbf{1}}{\mathbf{3}}$ and 3. Mean and variance of prior for Ln (RR) are estimated as Prior mean for ln (RR)=$\frac{(\mathrm{Ln}\left(\frac{1}{3}\right)+\mathrm{Ln}(3))}{2}$ =0, and Prior Variance for ln (RR)= $(\frac{\left|\mathrm{Ln}\left(\frac{1}{3}\right)-Ln(3)\right|}{2*1.96})^{2} = 0.10$. Inverse variances equaling $\frac{1}{0.1} = 10$ and$\frac{1}{0.84} = 1.2$ illustrating prior information dominated data information by nearly 8 times. Posterior mean and variance for Ln (RR) could be estimated as the following weighted averaging rule of thumb; Posterior mean for $\ln(\mathrm{RR}) = \frac{\frac{0}{0.10}+\frac{1.82}{0.84}}{\frac{1}{0.10}+\frac{1}{0.84}} =$ 0.19and Posterior variance for$\ln(\mathrm{RR}) \approx \frac{1}{\frac{1}{0.10}+\frac{1}{0.84}} = 0.09$. Posterior RR and its 95%CI through DA provided unbiased estimation of RR with more reasonable values of RR and narrower 95%CI. In addition, the value of posterior mean which is closer to prior means showed the influence of the prior as well. Various prior ranges for Ln (RR), estimated posterior 95% CI, as well as the influence of prior and data, were illustrated in (Table 2). It was depicted that, for prior ($\frac{\mathbf{1}}{\mathbf{6}}$, 6) data and prior had the same influence (equal weights) while for ($\frac{\mathbf{1}}{\mathbf{10}}$, 10) it was data dominated.

Here, *Compatibility* of prior and data is a great issue as well. DA fails in case of incompatibility causes misleading results(13). For our example, $\frac{(Ln(6.3)-0)}{(0.84+0.10)^{\frac{1}{2}}} = 1.9 P_{\text{value}} = 0.057$showed that compatibility hypothesis is not rejected.

Summary of pros and cons of the approaches

Supplementary table 2 presents advantages, failures and their remedies of the approaches.

DISCUSSION

A load of literature published on Bayesian inference has proved its popularity among researches while its

concept is not straightforward for amateur learners (14). The purpose of our paper was to provide a comprehensive framework with illuminating examples to shed light upon the concept of some of Bayesian mechanisms of sampling. We showed that alternative approaches which are intuitively appealing and easy-to-understand work well in case of low-dimensional problems and appropriate Prior information such as weighted prior, otherwise MCMC is a Trouble-free tool. Although its concept is not an intuitively realizable advanced method of MCMC tackles most complex issues. Different studies tried to cover Bayesian statistical approach as a need specifically for many sciences (15-26). Also, DA method as an alternative approach gives researchers more tangible sense in the role of prior and data for inference making, the posterior calculation is simple using this method. We tried to cover the sufficient methods of Bayesian simulation approaches with some clear examples and provide an introductory work of Bayesian foundation; R software codes are available in the Supplementary as well.

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material: All data generated or analyzed during the current study are

included in this published article.

Competing interests: The authors declare that they have no conflict of interest.

Funding: This research received no specific grant from any funding agency in the public,

commercial, or not-for-profit sectors.

Authors' contributions: Dr. RBY, Prof. KM, Prof. MAM and Prof. HZ had significant contributions to the conception, design, acquisition, analysis and interpretation of the information. Methodological concepts were considered Dr. RBY, Prof. KM, Prof. MAM, Prof. HZ. All authors worked on the drafting and agreed on final approval of the version to be published. Also, the agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

1. **Hammersley JM & Morton KW**(1954) Poor Man's Monte Carlo. *Journal of the Royal Statistical Society. Series B* **16** , 23-38.

2. **Rosenbluth MN & Rosenbluth AW** (1955) Monte Carlo Calculation of the Average Extension of Molecular Chains. *The Journal of Chemical Physics* **23** , 356-59.

3. **Robert C & Casella G**(2010) *Introducing Monte Carlo Methods with R* : Springer-Verlag New York.

4. **Armitage P, Berry G & Matthews JNS** (2001) *Statistical Methods in Medical Research* . 4th Edition ed: Wiley-Blackwell.

5. **T.J.Bailey N** (1966)*Elements of stochastic processes with applications to the Natural Sciences* : Wiley.

6. **Greenland S** (2006) Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology* **35** , 765-75.

7. **Greenland S, Mansournia MA & Altman DG** (2016) Sparse data bias: a problem hiding in plain sight.*BMJ* **352** .

8. **Greenland S & Christensen R** (2001) Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression.*Statistics in Medicine* **20** , 2421-28.

9. **Bedrick EJ, Christensen R & Johnson W** (1996) A New Perspective on Priors for Generalized Linear Models. *Journal of the American Statistical Association***91** , 1450-60.

10. **Bedrick EJ, Christensen R & Johnson W** (1997) Bayesian Binomial Regression: Predicting Survival at a Trauma Center. *The American Statistician* **51** , 211-18.

11. **Greenland S & Mansournia MA** (2015) Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in medicine* **34** , 3133-43.

12. **Mansournia MA, Heinze G, Geroldinger A & Greenland S** (2017) Separation in Logistic Regression: Causes, Consequences, and Control. *American Journal of Epidemiology* **187** , 864-70.

13. **George EPB** (1980) Sampling and Bayes' Inference in Scientific Modelling and Robustness.*Journal of the Royal Statistical Society. Series A (General)***143** , 383-430.

14. **Albert J** (1997) Teaching Bayes' Rule: A Data-Oriented Approach. *The American Statistician***51** , 247-53.

15. **Turner BM & Van Zandt T**(2012) A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology* **56** , 69-85.

16. **Etz A & Vandekerckhove J** (2018) Introduction to Bayesian Inference for Psychology.*Psychonomic Bulletin & Review* **25** , 5-34.

17. **Matzke D, Boehm U & Vandekerckhove J** (2018) Bayesian inference for psychology, part III: Parameter estimation in nonstandard models. *Psychonomic Bulletin & Review* **25** , 77-101.

18. **Wagenmakers E-J, Love J, Marsman M, et al.** (2018) Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review***25** , 58-76.

19. **Wagenmakers E-J, Marsman M, Jamil T, et al.** (2018) Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review* **25** , 35-57.

20. **Zhang L, Pfister M & Meibohm B** (2008) Concepts and challenges in quantitative pharmacology and model-based drug development. *The AAPS journal* **10** , 552-59.

21. **Racine A, Grieve A, Fluhler H & Smith A** (1986) Bayesian methods in practice: experiences in the pharmaceutical industry. *Applied Statistics* , 93-150.

22. **Barrett JS, Fossler MJ, Cadieu KD & Gastonguay MR** (2008) Pharmacometrics: a multidisciplinary field to facilitate critical thinking in drug development and translational research settings. *The Journal of Clinical Pharmacology* **48** , 632-49.

23. **Grieve AP** (2007) 25 years of Bayesian methods in the pharmaceutical industry: a personal, statistical bummel. *Pharmaceutical Statistics* **6** , 261-81.

24. **Morgan D** (2018) Bayesian applications in pharmaceutical statistics. *Pharmaceutical Statistics* **17** , 298-300.

25. **Miočević M** (2019) A Tutorial in Bayesian Mediation Analysis With Latent Variables.*Methodology* .

26. **Natesan P** (2019) Fitting Bayesian Models for Single-Case Experimental Designs.*Methodology* .

Table1. Estimation of a Two- State Transition Matrix for a discreet target and a uniform (0, 5) proposal distribution

PS	**NS**	**@Probability of transition**
1	1	$\frac{1}{5} \times 1 + \frac{10}{15}$ *(repeat the sampling in case of rejecting)*
1	2	$\frac{4}{5} \times \frac{1}{6}$
2	1	$\frac{1}{5} \times 1$
2	2	$\frac{4}{5} \times 1$

Abbreviations: Previous State (PS), Next State (NS),

[@]Probability of transition: probability of state * (probability of target distribution in NS ÷ probability of target distribution in PS)

Table 2. Posterior 95% CI for the Range of Prior Information via Data Augmentation Method

Prior95% CI for Ln (RR)	**Posterior 95% CI for Ln (RR)**	**Prior Information weight:** $(\frac{1}{\text{prior variance (ln(RR))}})$
($\frac{1}{2}$, 2)	(-0.42, 0.88)	8.3
($\frac{1}{3}$, 3)	(-0.44, 1.43)	3.2
($\frac{1}{4}$, 4)	(-0.41, 1.77)	2.0
($\frac{1}{5}$, 5)	(-0.38, 2.01)	1.5
($\frac{1}{6}$, 6)	(-0.36, 2.18)	1.2
($\frac{1}{7}$, 7)	(-0.33, 2.30)	1.0
($\frac{1}{10}$, 10)	(-0.28, 2.56)	0.7

Hosted file

`image1.emf` available at https://authorea.com/users/354031/articles/477715-a-framework-for-bayesian-posterior-simulation-methods-in-clinical-practice

Figure1 . Comparison of target and proposals distribution

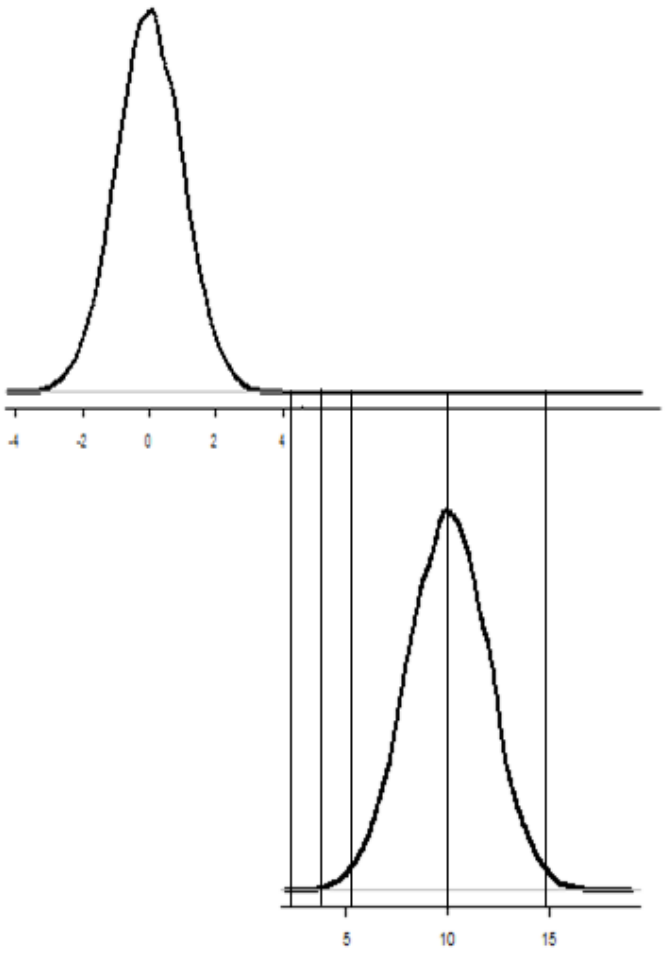

Figure2. As an example, which prior is not supported by data, imagine prior (lower) as Normal (?=10, SD=2) and likelihood (upper) Normal (0,1) is far from each other, most of the drawn samples from prior get very small weights. The probability of sampling from-3σ and lower (the parts that get higher weights) is nearly 1% so out of 10000 we expected 100 non-zero weights. Posterior mean estimated equals 1.5 while we expected 0.01 (Posterior ˜ Normal (0.01, 0.01). Therefore, this approach proved inefficient in terms of accuracy of estimate and number of sampling.