Journal of Data Science 23 (1), 56–58
January 2025
DOI: 10.6339/25-JDS1161F
Discussion

Discussion of "Power Priors for Leveraging Historical Data: Looking Back and Looking Forward"☆

Guohui Wu[1,*]

[1]*Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA, U.S.A.*

Borrowing from historical data has become increasingly popular in clinical trials due to its potential of improving the statistical power of a study and reducing sample size. To enable Bayesian information borrowing from historical data, some ubiquitous approaches include prior based approach and model based approach, among a wide array of approaches (e.g., see Viele et al., 2014). As a prior based approach, the power prior developed by Ibrahim and Chen (2000) introduces a discounting parameter, often denoted by a_0, to control the amount of borrowing from historical data. Chen et al. (2025) provides a comprehensive and valuable review on the use of various power priors for leveraging historical data.

My discussion will cover two aspects of power priors. First, I will focus on the issue of treating the discounting parameter a_0 either as a fixed or a random parameter in models. Second, I will discuss the use of variational Bayes (VB) methods to accelerate Bayesian inference for models that incorporate power priors.

Should a_0 Be Fixed or Random? When treated as fixed, the discounting parameter $a_0 \in [0, 1]$ is typically specified based on the similarity between the current and historical data. To measure the similarity between these two data, some metrics need to be introduced. With a chosen similarity metric, one can fix the value of a_0 in the power prior according to the assessed similarity between the current and historical data. In the two extreme cases, fixing $a_0 = 1$ leads to "full borrowing" (i.e., pooling historical data with current data), whereas setting $a_0 = 0$ amounts to "no borrowing" (i.e., ignoring historical data). When choosing a value for a_0, one can resort to empirical Bayes power prior (Gravestock et al., 2017), which estimates a_0 via empirical Bayes and accounts for discrepancies between the current and historical data. In practice, a value of α_0 may also be elicited based on prior knowledge and justified on a case-by-case basis.

Alternatively, a_0 can be random, which results in normalized power prior. In this case, one would expect the estimate of a_0 to adapt to the congruence between the current and historical data. Nevertheless, Pawel et al. (2023) show that normalized power prior always discounts the historical data for both normal and binomial models with beta priors for a_0, even when current data is a duplicate of the historical data and two sample sizes are large. As a result, "full borrowing" is not achievable when using normalized power priors. This counterintuitive yet important finding for normalized power priors motivates the two key questions that are not discussed in the paper by Chen et al. (2025).

- How to interpret a_0 in normalized power prior in relation to the discrepancies between the current and historical data?
- Can normalized power prior allow the "right amount" of information borrowing that is commensurate with the discrepancies between the current and historical data?

☆Main article: https://doi.org/10.6339/24-JDS1161.

* Email: gwu03@amgen.com.

Accelerate Model Fitting with Power Priors For Bayesian inference of models that incorporate power priors, computational challenges can be of practical concern. As an example, when using normalized power prior, the normalizing constant can be intractable and difficult to compute. To enable efficient sampling for a_0, well-customized MCMC sampling algorithms are essential, e.g., see Carvalho and Ibrahim (2021) and Han et al. (2023). Despite various software packages that were developed to facilitate the use of power priors, there is still a need to develop more efficient computational algorithms for fitting models with power priors. Here, I will concentrate on the use of VB methods to achieve efficient model fitting with power priors. As an alternative to MCMC, variational Bayes methods have been widely used in statistical research, e.g., see Blei et al. (2017) and the references therein for a review.

To ease subsequent discussion, we will first introduce some notations. Let $\boldsymbol{y}$ be the observed data and $\boldsymbol{\Theta}$ be the set of model parameters. Given the prior distribution $p(\boldsymbol{\Theta})$, the posterior distribution $p(\boldsymbol{\Theta}|\boldsymbol{y})$ takes the form of

$$p(\boldsymbol{\Theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}, \boldsymbol{\Theta})}{p(\boldsymbol{y})} \propto p(\boldsymbol{y}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}), \tag{1}$$

where $p(\boldsymbol{y}|\boldsymbol{\Theta})$ is the likelihood function; $p(\boldsymbol{y}, \boldsymbol{\Theta})$ is the joint likelihood of the data $\boldsymbol{y}$ and parameters $\boldsymbol{\Theta}$; and $p(\boldsymbol{y})$ is the marginal likelihood. For Bayesian inference via MCMC methods, samples drawn from $p(\boldsymbol{\Theta}|\boldsymbol{y})$ are used to learn about the target posterior distribution.

For an arbitrary density function $q(\boldsymbol{\Theta})$, the logarithm of the marginal likelihood in (1) can be decomposed as (Ormerod and Wand, 2010)

$$\begin{aligned}\ln p(\boldsymbol{y}) &= \ln \frac{p(\boldsymbol{y}, \boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\boldsymbol{y})} \underbrace{\int q(\boldsymbol{\Theta}) d\boldsymbol{\Theta}}_{=1} \\ &= \underbrace{\int q(\boldsymbol{\Theta}) \ln \frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\boldsymbol{y})} d\boldsymbol{\Theta}}_{D_{\mathrm{KL}}(q||p)} + \underbrace{\int q(\boldsymbol{\Theta}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{\Theta})}{q(\boldsymbol{\Theta})} d\boldsymbol{\Theta}}_{\text{evidence lower bound (ELBO)}},\end{aligned}$$

where $D_{\mathrm{KL}}(q||p)$ denotes the Kullback–Leibler divergence between $q(\boldsymbol{\Theta})$ and $p(\boldsymbol{\Theta}|\boldsymbol{y})$. The optimal VB posterior $q^*(\boldsymbol{\Theta})$ is obtained by solving the optimization problem as follows:

$$q^*(\boldsymbol{\Theta}) = \underset{q \in \mathcal{Q}}{\arg\max}\, \mathrm{ELBO}, \tag{2}$$

where $\mathcal{Q}$ denotes some family of distributions.

Unlike MCMC methods that use sampling to learn about $p(\boldsymbol{\Theta}|\boldsymbol{y})$, VB methods find $q^*(\boldsymbol{\Theta})$ that "best" approximates $p(\boldsymbol{\Theta}|\boldsymbol{y})$ by solving an optimization problem. By turning a sampling problem into an optimization problem, variational inference is generally faster than MCMC. For model fitting with power priors, developing VB algorithms to achieve efficient and accurate Bayesian inference can be beneficial and a topic for future research.

References

Blei DM, Kucukelbir A, McAuliffe JD (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518): 859–877. https://doi.org/10.1080/01621459.2017.1285773

Carvalho LM, Ibrahim JG (2021). On the normalized power prior. *Statistics in Medicine*, 40(24): 5251–5275. https://doi.org/10.1002/sim.9124

Chen MH, Guan Z, Lin M, Sun M (2025). Power priors for leveraging historical data: Looking back and looking forward. *Journal of Data Science*, 23(1): 1–30. https://doi.org/10.6339/24-JDS1161.

Gravestock I, Held L (CN Consortium) (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics*, 16(5): 349–360. https://doi.org/10.1002/pst.1814

Han Z, Zhang Q, Wang M, Ye K, Chen MH (2023). On efficient posterior inference in normalized power prior Bayesian analysis. *Biometrical Journal*, 65(5): 2200194. https://doi.org/10.1002/bimj.202200194

Ibrahim JG, Chen MH (2000). Power prior distributions for regression models. *Statistical Science*, 15(1): 46–60. https://doi.org/10.1214/ss/1009212673

Ormerod JT, Wand MP (2010). Explaining variational approximations. *American Statistician*, 64(2): 140–153. https://doi.org/10.1198/tast.2010.09058

Pawel S, Aust F, Held L, Wagenmakers EJ (2023). Normalized power priors always discount historical data. *Stat*, 12(1): e591. https://doi.org/10.1002/sta4.591

Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, 13(1): 41–54. https://doi.org/10.1002/pst.1589