Computational Statistics and Data Analysis 144 (2020) 106901

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Borrowing strength and borrowing index for Bayesian hierarchical models

Ganggang Xu [a,*], Huirong Zhu [b], J. Jack Lee [c]

[a] Department of Management Science, University of Maimi, FL, 33146, USA
[b] Department of Outcome & Impact Service, Texas Children's Hospital, Houston, TX, 77030, USA
[c] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA

ARTICLE INFO

Article history:
Received 19 April 2019
Received in revised form 20 September 2019
Accepted 4 December 2019
Available online 13 December 2019

Keywords:
Bayesian hierarchical model
Borrowing index
Borrowing strength
Clinical trials
Mallow's distance

ABSTRACT

A novel borrowing strength measure and an overall borrowing index to characterize the strength of borrowing behaviors among subgroups are proposed for a given Bayesian hierarchical model. The constructions of the proposed indexes are based on the Mallow's distance and can be easily computed using MCMC samples for univariate or multivariate posterior distributions. Consequently, the proposed indexes can serve as meaningful and useful exploratory tools to better understand the roles played by the priors in a hierarchical model, including their influences on the posteriors that are used to make statistical inferences. These relationships are otherwise ambiguous. The proposed methods can be applied to both the continuous and binary outcome variables. Furthermore, the proposed approach can be easily adapted to various settings of clinical trials, where Bayesian hierarchical models are deem appropriate. The effectiveness of the proposed method is illustrated using extensive simulation studies and a real data example.

1. Introduction

Bayesian hierarchical modeling has become a powerful tool in clinical trials and data analysis due to its flexibility and adaptivity in dealing with complex settings involving subgroups. Identifying subgroups with similar or different treatment effects and estimating the treatment effects in subgroups have become the bedrock of precision medicine. Bayesian hierarchical models can be applied in the platform design or the basket design to enhance the efficiency of treatment effect estimation (Woodcock and LaVange, 2017; Chen and Lee, 2019). Despite its popularity, one of the long standing criticisms for Bayesian modeling lies in that it can be substantially affected by the choice of prior distributions, whose impacts are often not fully understood. As a result, the dilemma routinely faced by statistical practitioners is that while one wishes to maintain objectivity by reducing the impact of the priors and letting the data speak for itself (e.g., Ye and Berger, 1991; O'Hagan et al., 2006; Gelman et al., 2008; Berger et al., 2009), the statistical inference can be much improved if "correct knowledge" were effectively translated into the prior specifications (e.g., Consonni and Veronese, 1995; Evans and Sedransk, 2001; Xu et al., 2015; Jiang et al., 2016). Although the later strategy appears to be more attractive in clinical trials when the sample size is limited, such as in the pediatric trials or rare disease settings (Lewis et al., 2019), it is also difficult, if not impossible, to validate the choice of prior as the one that reflects the true knowledge. A related important problem is how to quantify the impacts of the prior on the posterior inferences so that the practitioner can be informed of how influential the prior is. This is rather challenging and there is scarce

* Corresponding author.
E-mail addresses: gangxu@bus.miami.edu (G. Xu), hxzhu@texaschildrens.org (H. Zhu), jjlee@mdanderson.org (J.J. Lee).

https://doi.org/10.1016/j.csda.2019.106901
0167-9473/

literature on the topic. For example, Evans et al. (2006) studied how to check the conflicts between the prior and the data in Bayesian modeling. Morita et al. (2008, 2012) defined an effective sample size of the prior to characterize its impacts on the posterior inference. Built upon a similar idea, Reimherr et al. (2014) proposed a new data-driven diagnostic approach to assessing prior informativeness and prior-data likelihood conflict. However, most existing approaches are model specific and only applicable to certain classes of parametric Bayes models.

Our attention is restricted to the Bayesian subgroup analysis. During the course of drug development, a well-recognized fact is that patients from different subpopulations may respond similarly or differently to the same treatment. For example, cancer patients can be classified into different subtypes based on their gene mutations or prognoses and, some times, the classification is made by subjective judgment of the clinician (Leon-Novelo et al., 2012). Identification of subgroups that are more responsive (or less responsive) to a treatment is of critical importance in the new era of precision medicine. There have been numerous proposals on this topic, e.g., Foster et al. (2011), Friede et al. (2012), Berger et al. (2014), but this will not be the focus of this paper. Our main interest is to provide more informative subsequent analysis of clinical trial data by pooling information across different subgroups that have been identified in the first stage and quantify the amount of borrowing. There has been many studies on how to construct priors to encourage information borrowing within a cluster, see, for example, Consonni and Veronese (1995), Evans and Sedransk (2001), Quintana and Iglesias (2003), Müller et al. (2011), Leon-Novelo et al. (2012). However, to the best of our knowledge, there has yet been any study on how to quantify the borrowing strength among different subgroups in cluster hierarchical models for a given set of priors; this paper aims at filling this gap.

We propose two borrowing strength measures: (a) the individual borrowing strength (InBS) and (b) the overall borrowing index (OvBI). The former serves to measure and compare the borrowing strengths across different subgroups, while the latter is a fractional number that quantifies the overall borrowing strength for the Bayesian hierarchical model. Compared to the existing literature, one distinctive feature of both InBS and OvBI is that they are completely data driven and can be easily calculated using MCMC samples from the posterior distributions for a wide range of applications. For this reason, it is straightforward to seamlessly extend our discussions on independent and hierarchical models with continuous and binary outcomes to much more complicated Bayesian hierarchical models, regardless of whether the posterior distributions have a closed form or not. An important use of InBS and OvBI is to study and quantify the roles played by hyperparameters in encouraging or discouraging borrowing from other subgroups, such that the practitioner can make more informed decision to adjust.

The rest of the paper is organized as follows. Section 2 sets up the model and notations. Section 3 gives a detailed discussion on the motivation and the construction of the borrowing strength and borrowing index. Section 4 illustrates the proposed indices with a simple example of continuous outcome in two subgroups. Simulation studies are conducted in Section 5, and a real data application is considered in Section 6 to illustrate the effectiveness of the proposed method. Section 7 gives some discussions on potential issues with the current proposal and some future research topics.

2. Model specification and notations

2.1. Continuous outcome models

We first consider the case when treatment outcome is continuous. Suppose we have J different subgroups of patients with n_j subjects in each subgroup and some measurements were taken from each subject after the treatment, denoted by $Y_{j1}, \ldots, Y_{jn_j}$, $j = 1, \ldots, J$. The first approach is to treat all subgroups separately and assume $Y_{ji} \sim N(\mu_j, \sigma_j^2)$, for all subjects in the jth group. Consequently, the inference on parameters of interest, μ_j's, will be made based only on the n_j subjects in that group for $j = 1, \ldots, J$. For example, a popular model can be specified (which we shall refer to as the "Independent model" and denote as M_{ind}) as follows:

$$Y_{ji}|\mu_j \sim N(\mu_j, \tau_{y,j}^{-1}),\ \mu_j|\tau_{\mu_j} \sim N(0, \tau_{\mu_j}^{-1}),\ \tau_{y,j} \sim Gamma(a_y, b_y),\ \tau_{\mu_j} \sim Gamma(a_\mu, b_\mu), \tag{2.1}$$

where a_y, b_y, a_μ, b_μ are pre-specified hyperparameters. Note that $N(a, b)$ represents the normal distribution with mean a and variance b and $Gamma(a, b)$ stands for the gamma distribution with the shape parameter a and the rate parameter b. However, this strategy excludes the possibility that some of the subgroups may respond similarly, in which case pooling information from similar subgroups may substantially improve the inference precision by reducing the estimation variances of model parameters.

One natural remedy is to identify clusters that consist of subgroups with similar responses and borrow "information" within the cluster to improve the efficiency of the resulting statistical inference. This approach is especially appealing in areas like oncology, where the number of subjects in some subgroups with rare conditions is typically quite small (Jones et al., 2011; Leon-Novelo et al., 2012; Berry et al., 2013). To illustrate, we consider a class of commonly used models. Let $\mathcal{S} = \{S_1, S_2, \ldots, S_K\}$ be a partition of the index set $\{1, \ldots, J\}$ consisting of K non-empty and non-overlapping clusters, the "cluster hierarchical model" can be summarized as

$$\begin{aligned} &Y_{ji}|\mu_j \sim N(\mu_j, \tau_{y,j}^{-1}), \quad \mu_j|\mathcal{S} \sim N(\theta_k, \tau_{\mu,k}^{-1}) \text{ if } j \in S_k, \quad \theta_k|\tau_\theta \sim N(0, \tau_\theta^{-1}), \\ &\tau_{y,j} \sim Gamma(a_y, b_y),\ \tau_{\mu,k} \sim Gamma(a_\mu, b_\mu), \text{ and } \mathcal{S} \sim p(\mathcal{S}|\boldsymbol{\gamma}_0), k = 1, \ldots, K, \end{aligned} \tag{2.2}$$

where $a_y, b_y, a_\mu, b_\mu, \tau_\theta, \boldsymbol{\gamma}_0$ are set of hyperparameters that are pre-specified and $p(\mathcal{S}|\boldsymbol{\gamma}_0)$ is the prior distribution of the partition, which reflects the investigator's belief on how to effectively form clusters of similar subgroups that can borrow information from each other. Throughout this paper, we assume that the number of clusters K is given.

2.2. Binary outcome models

Instead of a continuous outcome, the treatment outcome is often categorized as binary, e.g., response/no response or success/failure. In this case, data collected only consists of J data pairs, denoted by $\{(X_1, n_1), \ldots, (X_J, n_J)\}$ with X_j being the number of subjects that respond positively in subgroup j with size n_j. To deal with such data, an "Independent model" with binary outcome can be described as

$$\begin{aligned} &X_j|p_j \sim Bin(p_j, n_j), \quad \text{logit}(p_j) = \theta_j, \quad \theta_j|\mu, \tau_\theta \sim N(\mu, \tau_\theta^{-1}), \\ &\mu \sim N(0, \tau_\mu^{-1}), \quad \tau_\theta \sim Gamma(a_\tau, b_\tau), \end{aligned} \tag{2.3}$$

where τ_μ, a_τ, b_τ are a set of pre-specified hyperparameters. Consequently, the cluster hierarchical model for binary data can be formulated as follows:

$$\begin{aligned} &X_j|p_j \sim Bin(p_j, n_j), \quad \text{logit}(p_j) = \theta_k, \text{ if } j \in S_k, \quad \theta_k|\mu, \tau_\theta \sim N(\mu, \tau_\theta^{-1}), \\ &\mu \sim N(0, \tau_\mu^{-1}), \quad \tau_\theta \sim Gamma(a_\tau, b_\tau), \text{ and } \mathcal{S} \sim p(\mathcal{S}|\boldsymbol{\gamma}_0), \quad k = 1, \ldots, K, \end{aligned} \tag{2.4}$$

where $\tau_\mu, a_\tau, b_\tau, \boldsymbol{\gamma}_0$ are a set of pre-specified hyperparameters.

In Bayesian modeling, the prior specifications have substantial impact on the posterior distributions of the parameters of interest, i.e., μ_j's or p_j's. Our main concern is to quantify the impacts of priors on the borrowing strength among different subgroups. Heuristically, conditioned on the partition $\mathcal{S}$, both cluster hierarchical models for continuous and binary outcomes assume that subgroups in a cluster share the same cluster-specific mean θ_k, with the within-cluster borrowing strength governed by the precision parameter $\tau_{\mu,k}$ (for model (2.2) only) and the borrowing strength among cluster centers controlled by the precision parameter τ_θ.

3. Borrowing strength and borrowing index

To quantify the borrowing strength, the first step is to find a baseline model where there is no borrowing at all, which we refer to as the "Independent model". The "Independent model" essentially analyzes data from different subgroups separately and can often be identified by setting the hyperparameters to some extreme values. For example, for the continuous outcome model (2.2), one version of the "Independent model" can be simply defined as in (2.1), which is equivalent to model (2.2) with $K = J$ and $\tau_\theta \to \infty$.

Let $\boldsymbol{\Theta}$ be the set of parameters of interest, we can then generate MCMC posterior samples using models (2.1) and (2.2), denoted by random variables $\boldsymbol{\Theta}^{M_{ind}}|\mathbf{Y}$ and $\boldsymbol{\Theta}^{M_h}|\mathbf{Y}$, respectively. Then the borrowing strength of model (2.2), denoted as model M_h, can be defined as the distance between distributions of $\boldsymbol{\Theta}^{M_h}|\mathbf{Y}$ and $\boldsymbol{\Theta}^{M_{ind}}|\mathbf{Y}$, that is,

$$\text{BS}(M_h, \boldsymbol{\Theta}|\mathbf{Y}) = d(F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}), \tag{3.1}$$

where $F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}$ are distribution functions of $\boldsymbol{\Theta}^{M_h}|\mathbf{Y}$ and $\boldsymbol{\Theta}^{M_{ind}}|\mathbf{Y}$, respectively, and $d(F_1, F_2)$ is a distance measure between two distribution functions $F_1(\cdot)$ and $F_2(\cdot)$. Note that the notation $\text{BS}(M_h, \boldsymbol{\Theta}|\mathbf{Y})$ stands for "the borrowing strength of the hierarchical model M_h when making posterior inferences for parameters in $\boldsymbol{\Theta}$".

Often in practice, it is more conceptually desirable to have an index that is between 0 and 1 to indicate the borrowing strength. Such an index relies on specifying another baseline model where all subgroups borrow all information available from other subgroups, which we refer to as the "Complete borrowing model". Similar to the "Independent model", the "Complete borrowing model" can be found by imposing the constraint on parameters of interest in different subgroups to take the same value. For example, in model (2.2), we can set $\mu_1 = \cdots = \mu_J = \mu$ and use the following "Complete borrowing model", denoted as M_{cmp},

$$Y_{ji}|\mu \sim N(\mu, \tau_{y,j}^{-1}), \quad \mu|\tau_\theta \sim N(0, \tau_\theta^{-1}), \quad \tau_{y,j} \sim Gamma(a_y, b_y). \tag{3.2}$$

It is straightforward to see that the model (3.2) is equivalent to the model (2.2), with $K = 1$ and $a_\mu/b_\mu \to \infty$ (so that $\tau_{\mu,1} \to \infty$ with probability 1). Then the borrowing index (BI) of the hierarchical model M_h, when making posterior inferences for parameters in $\boldsymbol{\Theta}$, can be defined as

$$\text{BI}(M_h, \boldsymbol{\Theta}|\mathbf{Y}) = \frac{d(F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}})}{d(F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}) + d(F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}})}. \tag{3.3}$$

With such a definition, we can ensure that (a) $\text{BI}(M_{ind}, \boldsymbol{\Theta}|\mathbf{Y}) = 0$, (b) $\text{BI}(M_{cmp}, \boldsymbol{\Theta}|\mathbf{Y}) = 1$ and (c) $0 \le \text{BI}(M_h, \boldsymbol{\Theta}|\mathbf{Y}) \le 1$. However, it is worth pointing out that the borrowing index $\text{BI}(M_h, \boldsymbol{\Theta}|\mathbf{Y})$ does not reflect the percentage of information in posteriors $\boldsymbol{\Theta}^{M_h}|\mathbf{Y}$ that is borrowed from other subgroups. Rather, it indicates the percentage of available information that can be borrowed from other subgroups that have actually been incorporated in the posterior distribution of $\boldsymbol{\Theta}^{M_h}|\mathbf{Y}$.

3.1. Mallow's distance

In this subsection we introduce the Mallow's distance (Mallows, 1972), which is the key building block of the proposed strength borrowing measures.

Let $\boldsymbol{X}$ and $\boldsymbol{Z}$ be two random vectors in $\mathbb{R}^q$ with distribution functions $F_{\boldsymbol{X}}(\cdot)$ and $F_{\boldsymbol{Z}}(\cdot)$, such that the pth moments $\mathbf{E}\|\boldsymbol{X}\|^p$ and $\mathbf{E}\|\boldsymbol{Z}\|^p$ exist for the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^q$. The p-Mallow's distance between $F_{\boldsymbol{X}}(\cdot)$ and $F_{\boldsymbol{Z}}(\cdot)$ is then defined as

$$d_p(F_{\boldsymbol{X}}, F_{\boldsymbol{Z}}) = \inf_{F_{\boldsymbol{X},\boldsymbol{Z}}} \left\{\mathbf{E}_{F_{\boldsymbol{X},\boldsymbol{Z}}}(\|\boldsymbol{X}-\boldsymbol{Z}\|^p)\right\}^{1/p}, \tag{3.4}$$

where the infimum is taken over all possible joint probability distributions $F_{\boldsymbol{X},\boldsymbol{Z}}$ such that marginal distribution functions for $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $F_{\boldsymbol{X}}$ and $F_{\boldsymbol{Z}}$, respectively. The Mallow's distance was first introduced in Mallows (1972) and has proven useful in many applications, see for example, Bickel and Freedman (1981), Levina and Bickel (2001), Alvarez-Esteban et al. (2008). In particular, Levina and Bickel (2001) pointed out that the Mallow's distance is the Earth Mover's Distance (EMD, Rubner et al., 2000), which is one of the most popular similarity measures in texture/image classification. The proven usefulness of the EMD in machine learning (e.g. Grauman and Darrell, 2004; Fu et al., 2006; Ren et al., 2011) gives further support for our choice of Mallow's distance.

Estimating the Mallow's distance is straightforward when $q = 1$, which takes a simple form

$$d_p(F_X, F_Z) = \left\{\int_0^1 |F_X^{-1}(u) - F_Z^{-1}(u)|^p \, du\right\}^{1/p}, \tag{3.5}$$

where $F_X^{-1}(\cdot)$ and $F_Z^{-1}(\cdot)$ are quantile functions of X and Z, respectively. However, for multivariate distributions with $q \geq 2$, the Mallow's distance does not have a closed form except for the multivariate normal distribution with $p = 2$, in which case it is also referred to as the "Fréchet distance" (Dowson and Landau, 1982). Due to this appealing connection, we shall use $p = 2$ for the rest of paper. The following lemma is a direct consequence of the theorem given in Dowson and Landau (1982).

Lemma 1. *Suppose that $\boldsymbol{X} \sim N(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}})$ and $\boldsymbol{Z} \sim N(\boldsymbol{\mu}_{\boldsymbol{Z}}, \boldsymbol{\Sigma}_{\boldsymbol{Z}})$, then the Mallow's distance between these two multivariate normal distribution functions with $p = 2$ takes the following form:*

$$d_2\{N(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}}), N(\boldsymbol{\mu}_{\boldsymbol{Z}}, \boldsymbol{\Sigma}_{\boldsymbol{Z}})\} = \left[\|\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{Z}}\|^2 + tr(\boldsymbol{\Sigma}_{\boldsymbol{X}} + \boldsymbol{\Sigma}_{\boldsymbol{Z}}) - 2\sum_{i=1}^{q}\sqrt{\lambda_i(\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{Z}})}\right]^{1/2}, \tag{3.6}$$

where $\lambda_i(\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{Z}})$ stands for the ith largest eigenvalue of the matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}$.

Although the Mallow's distance generally does not have a closed form for $q \geq 2$, its estimation is straightforward. Suppose we have independent samples $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_B\}$ and $\{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_B\}$ from distribution $F_{\boldsymbol{X}}(\cdot)$ and $F_{\boldsymbol{Z}}(\cdot)$, respectively. Let $\widehat{F}_{\boldsymbol{X},B}$, $\widehat{F}_{\boldsymbol{Z},B}$ be the corresponding empirical distribution functions by assigning each data point an equal weight $1/B$. Then the Mallows distance between these two empirical distributions (Levina and Bickel, 2001) takes the form

$$d_2(\widehat{F}_{\boldsymbol{X},B}, \widehat{F}_{\boldsymbol{Z},B}) = \left\{\min_{(l_1,\ldots,l_B)} \frac{1}{B}\sum_{i=1}^{B}\|\boldsymbol{X}_i - \boldsymbol{Z}_{l_i}\|^2\right\}^{1/2}, \tag{3.7}$$

where the minimum is taken over all possible permutations of indices $1, \ldots, B$. The computation of $d_2(\widehat{F}_{\boldsymbol{X},B}, \widehat{F}_{\boldsymbol{Z},B})$ can be efficiently carried out using the Hungarian algorithm (Kuhn, 1955), which costs about $O(B^3)$ floating operations. By the triangle inequality, we have that

$$|d_2(F_{\boldsymbol{X}}, F_{\boldsymbol{Z}}) - d_2(\widehat{F}_{\boldsymbol{X},B}, \widehat{F}_{\boldsymbol{Z},B})| \leq d_2(F_{\boldsymbol{X}}, \widehat{F}_{\boldsymbol{X},B}) + d_2(F_{\boldsymbol{Z}}, \widehat{F}_{\boldsymbol{Z},B}) \to 0 \text{ almost surely, as } B \to \infty,$$

where the last convergence result stems from the almost sure convergence of the empirical distribution functions $\widehat{F}_{\boldsymbol{X},B}$ and $\widehat{F}_{\boldsymbol{Z},B}$, e.g., del Barrio et al. (1999). Therefore, for a sufficiently large B, (3.7) provides a good estimator for $d_2(F_{\boldsymbol{X}}, F_{\boldsymbol{Z}})$. In the case when the dimensions of random variables $\boldsymbol{X}$ and $\boldsymbol{Z}$ are large, the necessary sample size B for a sufficiently good estimate may be too large to use the Hungarian algorithm. More computationally efficient algorithms can be used to compute (3.7), e.g., Varadarajan (1998).

3.2. Estimating the InBS and OvBI

Using definitions (3.1)–(3.3), we can study the borrowing strength of hierarchical models for both the continuous and binary outcomes. We shall only use the hierarchical model (2.2) for continuous outcome to illustrate the idea, where parameters of interest are subgroup means μ_j's. There are two types of borrowing strength we are interested in: (1) the individual borrowing strength for each subgroup; (2) the overall borrowing index for all subgroups. The former refers to the amount of strength borrowing that occurs in the posterior distribution of a particular μ_j under the hierarchical model (2.2) and the latter refers to the overall amount of strength borrowing among posteriors of μ_j's. While the individual

borrowing strength reveals how much a particular group benefits from the hierarchical model, the overall borrowing index quantifies how effective the hierarchical structure is in encouraging borrowing information across subgroups.

We start by studying the individual borrowing strength, denoted as InBS, of the posterior distribution of the mean μ_j, in which case the target parameter $\boldsymbol{\Theta} = \mu_j$ is univariate. For univariate distributions, following (3.5), the estimated Mallow's distance between two posterior distributions under model M_1, M_2, denoted by $F^{M_1}_{\mu_j|\mathbf{Y}}$ and $F^{M_2}_{\mu_j|\mathbf{Y}}$, takes the following simple form:

$$\widehat{d}_2\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right) = d_2\left(\widehat{F}^{M_1}_{\mu_j|\mathbf{Y}}, \widehat{F}^{M_2}_{\mu_j|\mathbf{Y}}\right) = \left\{\frac{1}{B}\sum_{i=1}^{B}\left|\mu^{M_1}_{j,(i)} - \mu^{M_2}_{j,(i)}\right|^2\right\}^{1/2}, \tag{3.8}$$

where $\mu^{M_k}_{j,(1)} \le \mu^{M_k}_{j,(2)} \le \cdots \mu^{M_k}_{j,(B)}$, $k = 1, 2$, are ordered MCMC samples. Note that the number of simulated samples B can be made arbitrarily large and thus the estimation error can be controlled at a very low level without much computational cost. Using the above estimated Mallow's distance, the InBS for subgroup j under the hierarchical model can be estimated as

$$\widehat{\text{InBS}}_p(M_h, \mu_j|\mathbf{Y}) = \widehat{d}_2\left(F^{M_{ind}}_{\mu_j|\mathbf{Y}}, F^{M_h}_{\mu_j|\mathbf{Y}}\right), \tag{3.9}$$

for $j = 1, \ldots, J$. Roughly speaking, $\widehat{\text{InBS}}_p(M_h, \mu_j|\mathbf{Y})$ indicates how much information in $F^{M_h}_{\mu_j|\mathbf{Y}}$ is borrowed from other subgroups. Comparing $\widehat{\text{InBS}}_p(M_h, \mu_j|\mathbf{Y})$'s across all subgroups gives a better idea of which groups are borrowing more information from others.

Our second interest lies in quantifying the overall borrowing strength of the Bayesian hierarchical model. One natural way to characterize the overall borrowing strength is to compare joint posterior distributions of $(\mu_1, \mu_2, \ldots, \mu_J)$ under the "Independent model" (2.1), the "Complete borrowing model" (3.2), and the hierarchical model (2.2). In this case, the target parameter $\boldsymbol{\Theta}$ becomes multivariate, i.e., $\boldsymbol{\Theta} = (\mu_1, \ldots, \mu_J)$. The building block for estimating $\text{BI}(M_h, \boldsymbol{\Theta}|\mathbf{Y})$ as defined in (3.3) is the Mallow's distance between two multivariate empirical distribution functions (3.7). However, when the number of groups J is large, using (3.7) requires a large number of MCMC samples for a desired precision. Although we can generate as many MCMC samples as needed in practice, the computation of (3.7) may not be feasible when the sample size B gets extremely large. For this reason, we propose an alternative distance measure between two multivariate distributions, denoted by $F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}$ under model M_1 and M_2, as follows:

$$\widehat{d}^*_2\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) = \max\left[d_2\left\{N(\widehat{\boldsymbol{\mu}}_{M_1}, \widehat{\boldsymbol{\Sigma}}_{M_1}), N(\widehat{\boldsymbol{\mu}}_{M_2}, \widehat{\boldsymbol{\Sigma}}_{M_2})\right\}, \sqrt{\sum_{j=1}^{J}\widehat{d}_2{}^2\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right)}\right], \tag{3.10}$$

where $\widehat{\boldsymbol{\mu}}_{M_k}$ and $\widehat{\boldsymbol{\Sigma}}_{M_k}$ are the sample mean and sample covariance matrix of the posterior samples of $\boldsymbol{\Theta}|\mathbf{Y}$ generated under model M_k, $k = 1, 2$ and $\widehat{d}_2\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right)$'s are defined in (3.8) with $p = 2$. It is straightforward to check that $\widehat{d}^*_2\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$ is a valid distance measure. Furthermore, the following theorem indicates that $\widehat{d}^*_2\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$ defines a lower bound of the Mallow's distance estimator (3.7) for multivariate distribution functions.

Theorem 1. *For posterior distributions of $\boldsymbol{\Theta} = (\mu_1, \ldots, \mu_J)|\mathbf{Y}$ under model M_1 and M_2, the empirical distribution functions $\widehat{F}^{M_k}_{\boldsymbol{\Theta}|\mathbf{Y}}$, $k = 1, 2$, based on MCMC samples of size B satisfy that*

$$\widehat{d}^*_2\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) \le d_2\left(\widehat{F}^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, \widehat{F}^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right),$$

where the equality holds when both $F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}(\cdot)$ and $F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}(\cdot)$ are multivariate normal.

The proof is given in Appendix.

Note that Theorem 1 only applies to the case with $p = 2$ for the Mallow's distance, which is the most popular choice in practice. One way to interpret $\widehat{d}^*_2(\cdot, \cdot)$ is that it quantifies the distance between $F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}$ by the discrepancy between their Gaussian approximations. The major advantage of $\widehat{d}^*_2\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$ over $d_2\left(\widehat{F}^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, \widehat{F}^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$ is that it is computationally much more efficient when the MCMC sample size needed is extremely large. With the newly defined distance measure (3.10), the overall borrowing index (OvBI) can be estimated as follows:

$$\widehat{\text{OvBI}}^*(M_h, \boldsymbol{\Theta}|\mathbf{Y}) = \frac{\widehat{d}^*_2\left(F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)}{\widehat{d}^*_2\left(F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) + \widehat{d}^*_2\left(F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)}. \tag{3.11}$$

When the number of subgroups J is not large or when it is more desirable for other choices of p, one can simply replace $\widehat{d}^*_2(\cdot, \cdot)$ in (3.11) with the Mallow's distance between empirical distributions as defined in (3.7) whenever its computation is feasible.

3.3. Why Mallow's distance?

The key in the definition of Borrowing Strength (3.1) and Borrowing Index (3.3) is the use of Mallow's distance measure between two probability distributions. Although, in principle, it appears that any well-defined distance measure can be used, the Mallow's distance has some unique advantages. The first advantage is that it can be efficiently computed using MCMC samples, as given in (3.8) and (3.10). On the contrary, other distance measures such as the Kullback–Leibler divergence and the Hellinger distance (Pollard, 2002) typically require density functions for computations. For example, the Hellinger distance of two posterior distributions, say $F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}$, can be computed as

$$H\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) = \sqrt{\frac{1}{2}\int_{\mathbb{R}^q}\left(\sqrt{f^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}} - \sqrt{f^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}}\right)^2 \mathrm{d}\boldsymbol{\Theta}},$$

where $f^{M_j}_{\boldsymbol{\Theta}|\mathbf{Y}}$'s are probability density functions of $F^{M_j}_{\boldsymbol{\Theta}|\mathbf{Y}}$, $j = 1, 2$. The computation of $H\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$ requires numerical integrations on $\mathbb{R}^q$ and nonparametric density estimators based on MCMC samples, both of which can be difficult even for a moderate dimension q.

The second advantage of the Mallow's distance is that it is well defined for degenerate distributions, which is critical for our definition of the overall borrowing index (OvBI). The OvBI requires the distance between $F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}$, where the design of M_{cmp} in (3.2) suggests that $F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}$ is degenerate, in the sense that all components in $\boldsymbol{\Theta} = (\mu_1, \ldots, \mu_J)^T$ are restricted to be identical under (3.2). When $F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}$ is degenerate, the Kullback–Leibler divergence between $F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}$ is not well defined and the corresponding Hellinger distance has rather undesirable properties. For example, when $F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}$ is non-degenerate multivariate normal and $F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}$ is degenerate multivariate normal, it is straightforward to show that

$$H\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) \equiv 1.$$

Consequently, when both $F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}$ and $F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}$ are non-degenerate multivariate normal, they are always equally "close" to $F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}$, which is not intuitively desirable.

4. An illustrative example with continuous outcomes

In this section, we give an illustrative example for the proposed InBS and OvBI for a simplified version of the hierarchical model (2.2) for continuous outcome. Suppose we have two groups of samples from the following Bayesian hierarchical model:

$$\textbf{Model } M_h\text{:}\quad Y_{ji} \sim N(\mu_j, \tau_y^{-1}),\quad \mu_j \sim N(\mu_0, \frac{1}{2}\tau_\mu^{-1}),\quad \mu_0 \sim N(0, \tau_{\mu_0}^{-1}),\quad i = 1, \ldots, n_j,\quad j = 1, 2,$$

where $\tau_{\mu_0} \to 0$ and τ_y and τ_μ are some pre-specified hyperparameters. The full posterior distribution of (μ_1, μ_2) can then be written as proportional to the following:

$$\tau_y^{\frac{n_1+n_2}{2}}\tau_\mu^{\frac{1}{2}}\tau_{\mu_0}^{\frac{1}{2}}\exp\left\{-\frac{\tau_y}{2}\sum_{i=1}^{n_1}(y_{1,i}-\mu_1)^2 - \frac{\tau_y}{2}\sum_{i=1}^{n_2}(y_{2,i}-\mu_2)^2 - \tau_\mu\sum_{j=1}^{2}(\mu_j-\mu_0)^2 - \frac{\tau_{\mu_0}}{2}\mu_0^2\right\}.$$

By integrating out μ_0 and letting $\tau_{\mu_0} \to 0$, we have the marginal posterior (μ_1, μ_2), given $\mathbf{Y}$, as

$$P(\mu_1, \mu_2|\mathbf{Y}) \propto \tau_y^{\frac{n_1+n_2}{2}}\tau_\mu^{\frac{1}{2}}\exp\left\{-\frac{n_1\tau_y+\tau_\mu}{2}\mu_1^2 - \frac{n_2\tau_y+\tau_\mu}{2}\mu_2^2 + n_1\tau_y\overline{Y}_1\mu_1 + n_2\tau_y\overline{Y}_2\mu_2 + \tau_\mu\mu_1\mu_2\right\},$$

where $\overline{Y}_j = n_j^{-1}\sum_{i=1}^{n_j} Y_{ji}$, $j = 1, 2$. It is then straightforward to show that the joint posterior distribution of $(\mu_1, \mu_2)^T|\mathbf{Y}$ is bivariate normal as follows:

$$\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}|\mathbf{Y} \sim N\left[\begin{pmatrix} w\overline{Y}_{\cdot\cdot} + (1-w)\overline{Y}_1 \\ w\overline{Y}_{\cdot\cdot} + (1-w)\overline{Y}_2\end{pmatrix}, \tau_y^{-1}\begin{pmatrix} \frac{1-w}{n_1} + \frac{w}{n_1+n_2}, & \frac{w}{n_1+n_2} \\ \frac{w}{n_1+n_2}, & \frac{1-w}{n_2} + \frac{w}{n_1+n_2}\end{pmatrix}\right]$$

where $w = \frac{\tau_\mu/\tau_y}{\frac{n_1n_2}{n_1+n_2}+\tau_\mu/\tau_y}$, $\overline{Y}_{\cdot\cdot} = \frac{n_1\overline{Y}_1+n_2\overline{Y}_2}{n_1+n_2}$. For this simple model, the "Independent model" can be easily identified by setting $w = 0$ and the joint posterior becomes

$$\textbf{Independent model } M_{ind}\text{:}\quad \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}|\mathbf{Y}_1, \mathbf{Y}_2 \sim N\left[\begin{pmatrix}\overline{Y}_1\\ \overline{Y}_2\end{pmatrix}, \tau_y^{-1}\begin{pmatrix}\frac{1}{n_1} & 0\\ 0 & \frac{1}{n_2}\end{pmatrix}\right].$$

On the other hand, the "Complete borrowing model" can be found by setting $w = 1$ and the joint posterior becomes a degenerate multivariate normal

$$\textbf{Complete borrowing model } M_{cmp}: \quad \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} |\mathbf{Y}_1, \mathbf{Y}_2 \sim N\left[\begin{pmatrix} \overline{Y}_{\cdot\cdot} \\ \overline{Y}_{\cdot\cdot} \end{pmatrix}, \frac{1}{(n_1+n_2)\tau_y} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right].$$

A straightforward application of Lemma 1 gives the Mallow's distance between $F^{M_{ind}}_{\mu_1|\mathbf{Y}}$ and $F^{M_h}_{\mu_1|\mathbf{Y}}$ with $p = 2$ as follows:

$$d_2(F^{M_{ind}}_{\mu_1|\mathbf{Y}}, F^{M_h}_{\mu_1|\mathbf{Y}}) = \sqrt{\frac{1}{n_1\tau_y}\left\{\frac{n_2 w^2}{(n_1+n_2)}T^2 + \left(\sqrt{1-\frac{n_2}{n_1+n_2}w} - 1\right)^2\right\}},$$

$$d_2(F^{M_h}_{\mu_1|\mathbf{Y}}, F^{M_{cmp}}_{\mu_1|\mathbf{Y}}) = \sqrt{\frac{1}{n_1\tau_y}\left\{\frac{n_2(1-w)^2}{(n_1+n_2)}T^2 + \left(\sqrt{1-\frac{n_2}{n_1+n_2}w} - \sqrt{\frac{n_1}{n_1+n_2}}\right)^2\right\}},$$

where $T^2 = \frac{(\overline{Y}_1-\overline{Y}_2)^2\tau_y}{n_1^{-1}+n_2^{-1}}$. Consequently, the individual borrowing strength for group 1, i.e., μ_1, can be computed as

$$\text{InBS}(M_h, \mu_1|\mathbf{Y}) = \sqrt{\frac{1}{n_1\tau_y}\left\{\frac{n_2 w^2}{(n_1+n_2)}T^2 + \left(\sqrt{1-\frac{n_2}{n_1+n_2}w} - 1\right)^2\right\}}. \tag{4.1}$$

Similarly, $\text{InBS}(M_h, \mu_2|\mathbf{Y})$ can be obtained by switching n_1 and n_2 in the above equation. As expected, the borrowing strength is controlled by the quantity $w = \frac{\tau_\mu/\tau_y}{\frac{n_1 n_2}{n_1+n_2}+\tau_\mu/\tau_y}$. As w increases, the borrowing strength of μ_1 becomes stronger through two channels: (1) the posterior mean was shrunk towards the overall sample mean $\overline{Y}_{\cdot\cdot}$ from the group mean $\overline{Y}_1$; (2) the variance was reduced from the group variance $n_1^{-1}\tau_y^{-1}$ to the pooled variance $(n_1+n_2)^{-1}\tau_y^{-1}$. Note that the quantity T^2 quantifies the difference between two subgroup sample means. From a frequentist point of view, T^2 follows a non-central χ^2 distribution with 1 degree of freedom, i.e., $T^2 \sim \chi^2(\lambda, 1)$ with the non-centrality parameter $\lambda = \frac{(\mu_{10}-\mu_{20})^2\tau_y}{n_1^{-1}+n_2^{-1}}$, where μ_{10} and μ_{20} are "true values" of μ_1 and μ_2, respectively. More specifically, recall Eq. (3.6) in Lemma 1, where the first part of the Mallow's distance corresponds to the difference in distribution means (related to mean borrowing) while the second part compares the difference in covariance matrices (related to variance borrowing). Therefore, in the individual borrowing strength (4.1), the relative magnitude of T^2 compared to the second term inside the bracket determines whether the major source of borrowing strength is from the "mean borrowing" or from the "variance borrowing".

For better illustrations, Fig. 1 gives some examples of InBS with various values of T^2 and $\tau_y = 1$. Sample sizes of two subgroups were set to be $n_1 = 10$ and $n_2 = 20$. The left panel shows that the individual borrowing strength appears to be a linear increasing function of w, with the smaller group borrowing more information than the larger group. The right panel directly compares the InBS between two subgroups, which indicates that as T^2 increases, the ratio of individual borrowing strength (smaller subgroup/larger subgroup) decreases. This is because when $T^2 = 0$, the only source of borrowing strength is from "variance borrowing", which is determined by the sample sizes n_1 and n_2. In this case, the first group with a smaller sample size tends to borrow more strength from the second group, rather than the other way around. As T^2 increases, the major source of borrowing strength shifts from the "variance borrowing" to "mean borrowing", which is less impacted by group sample sizes. Therefore, the amount of strength borrowing starts to become similar for the two groups when T^2 is large.

Next, we proceed to study the overall borrowing index (OvBI) for Model M_h. In this example, the parameter of interest $\boldsymbol{\Theta} = (\mu_1, \mu_2)$. Using Lemma 1, the Mallow's distance from posteriors of $\boldsymbol{\Theta}$ under the "Independent model" or "Complete borrowing model" to the model M_h, respectively, can be shown to have the following forms:

$$d_2^2\left(F^{M_{ind}}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) = C_n\left\{\frac{(n_1^2+n_2^2)w(wT^2-1)}{2(n_1+n_2)^2} + 1 - \frac{\sqrt{n_1 n_2}}{n_1+n_2}\sqrt{\frac{n_1^2+n_2^2}{n_1 n_2}(1-w)+w+2\sqrt{1-w}}\right\},$$

$$d_2^2\left(F^{M_h}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_{cmp}}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) = C_n\left\{\frac{(n_1^2+n_2^2)}{2(n_1+n_2)^2}\left\{(1-w)^2T^2-1-w\right\} + 1 - \frac{\sqrt{n_1 n_2}}{n_1+n_2}\sqrt{1-\frac{(n_1-n_2)^2}{(n_1+n_2)^2}w}\right\},$$

where $C_n = 2\left(n_1^{-1}+n_2^{-1}\right)\tau_y^{-1}$. Then the OvBI is of the form

$$\text{OvBI}(M_h, \boldsymbol{\Theta}|\mathbf{Y}) = \frac{1}{1+\sqrt{\frac{(n_1^2+n_2^2)\{(1-w)^2T^2-1-w\}+2\left\{(n_1+n_2)^2-\sqrt{n_1 n_2}\sqrt{(n_1+n_2)^2-(n_1-n_2)^2 w}\right\}}{(n_1^2+n_2^2)(w^2T^2-w)+2\left\{(n_1+n_2)^2-(n_1+n_2)\sqrt{(n_1^2+n_2^2)(1-w)+n_1 n_2(w+2\sqrt{1-w})}\right\}}}}.$$

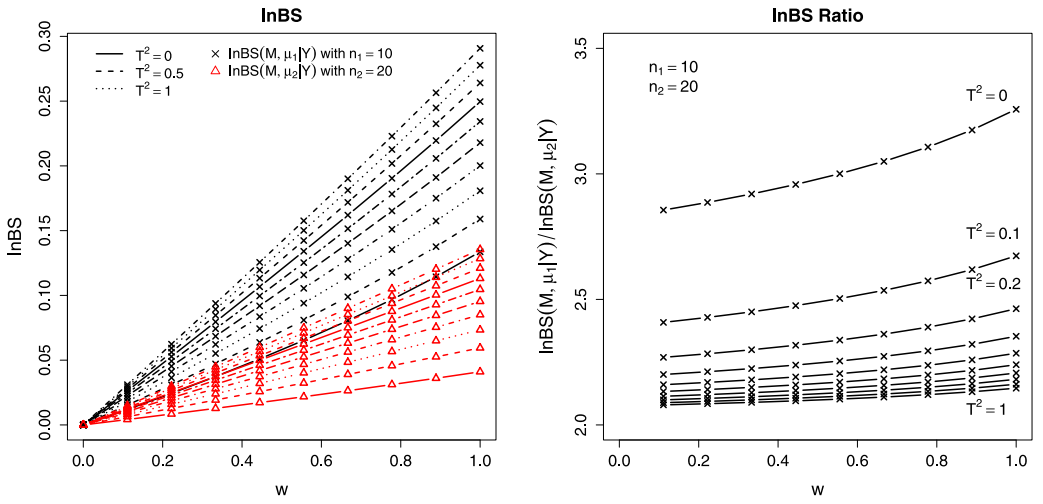

Fig. 1. Illustrations of individual borrowing strength (InBS).

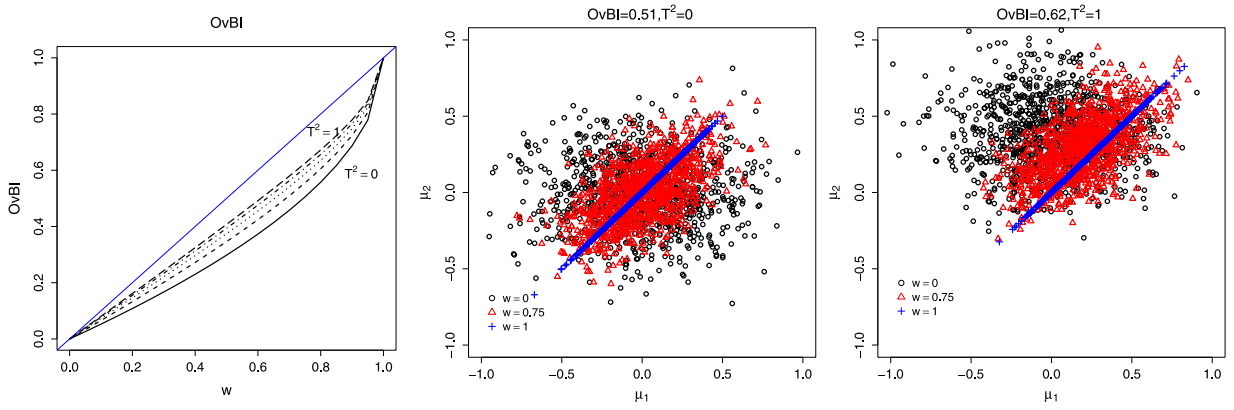

Fig. 2. Illustrations of overall borrowing index (OvBI).

Fig. 2 gives some examples of OvBI with various values of T^2 under the same setup of Fig. 1. The left panel shows that the OvBI is still an increasing function of w, indicating as the w increases the overall borrowing strength induced by model M becomes stronger. The middle and right panels illustrate two sets of realizations with $T^2 = 0, 1$ from three models with the $w = 0, 0.75, 1$. The OvBI is stronger for the case with $T^2 = 1$ (0.62) than the case with $T^2 = 0$ (0.51) when $w = 0.75$, which correctly reflects the similarities among observed distribution patterns of posterior MCMC samples generated from three Bayesian hierarchical models.

5. Simulation studies

In this section, we apply the proposed InBS and OvBI measures to more complicated hierarchical models for continuous as well as binary outcome data.

5.1. Continuous outcome models

We first consider the continuous outcome hierarchical model (2.2), where the partition prior distribution is specified as the finite mixture model (Diebolt and Robert, 1994). More specifically, we assume that

$$Y_{ij}|\mu_j \sim N(\mu_j, \tau_{y,j}^{-1}),\ \mu_j \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\theta_k, \tau_{\mu,k}^{-1}),\ (\pi_1, \ldots, \pi_K) \sim \mathit{Dirichlet}(\alpha, \ldots, \alpha), \tag{5.1}$$

where $\mathcal{N}(\mu_k, \tau_{\mu,k}^{-1})$ is the density function of the normal distribution $N(\mu_k, \tau_{\mu,k}^{-1})$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Other parameters such as θ_k's and $\tau_{\mu,k}$'s follow the same settings as described in model (2.2). In particular, the hyperparameters were set as $a_y = b_y = \tau_\theta = 10^{-6}$, $\alpha = 1/K$, $b_\mu = 0.01$ while a_μ and the number of clusters K may vary from case to case.

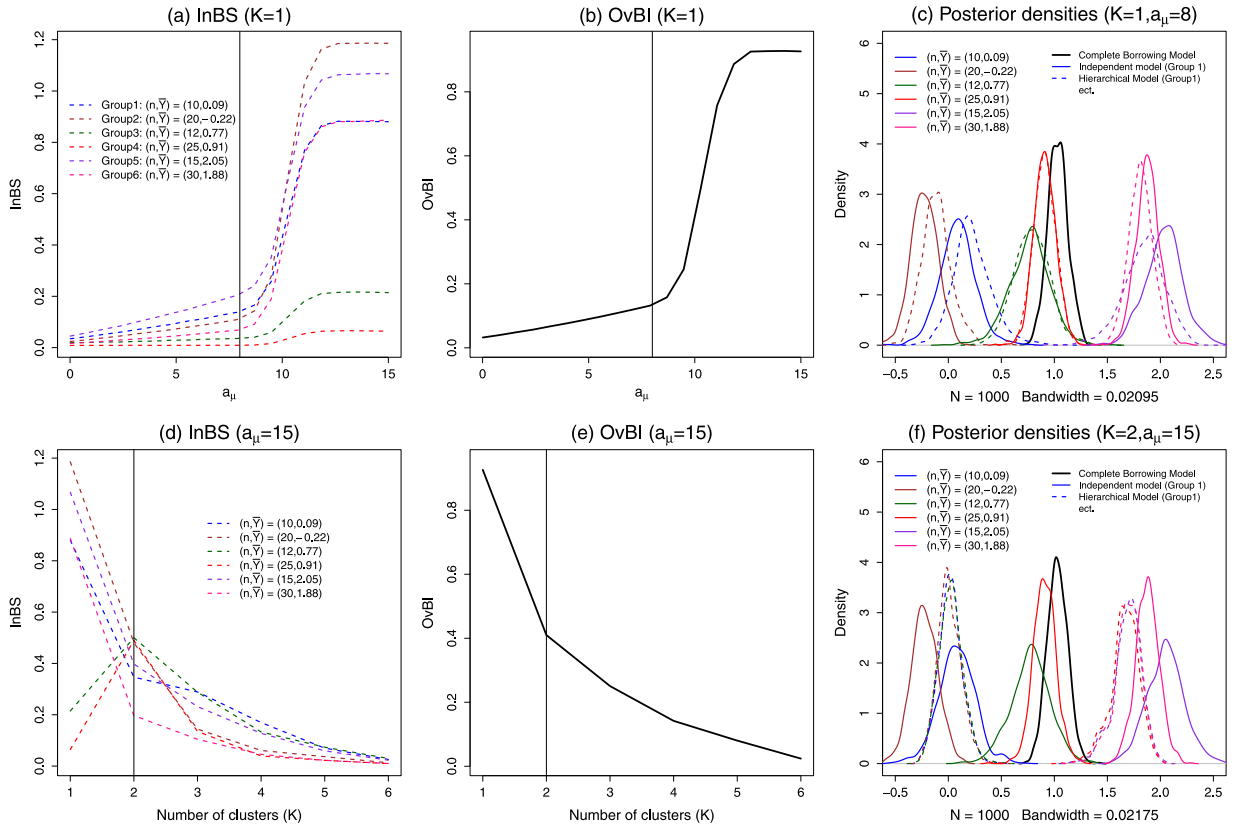

Fig. 3. Illustrations of Individual BS and Overall BI for the continuous outcome.

The goal of the simulation study is to illustrate the impacts of a_μ and K on the borrowing strength of the finite normal mixture model (5.1) using our proposed InBS and OvBI measures.

The data y_{ji}'s were simulated as six groups of sizes $n_1 = 10$, $n_2 = 20$, $n_3 = 12$, $n_4 = 25$, $n_5 = 15$, $n_6 = 30$, respectively. Groups 1 and 2 were simulated from $N(0, 0.5^2)$, Groups 3 and 4 were from $N(\sqrt{6/7}, 0.5^2)$ and Groups 5 and 6 were from $N(2\sqrt{6/7}, 0.5^2)$. Using calculations similar to those in Diebolt and Robert (1994), it is straightforward to show that the posterior distributions of μ_j's are completely determined by sample group means $\overline{Y}_j$'s and sample group second moments $\overline{Y_j^2}$'s (the average of y_{ji}^2's). In this simulation, we used a realization of y_{ji}'s with $(\overline{Y}_1, \overline{Y_1^2}) = (0.09, 0.22)$, $(\overline{Y}_2, \overline{Y_2^2}) = (-0.22, 0.32)$, $(\overline{Y}_3, \overline{Y_3^2}) = (0.77, 0.92)$, $(\overline{Y}_4, \overline{Y_4^2}) = (0.91, 1.07)$, $(\overline{Y}_5, \overline{Y_5^2}) = (2.05, 4.58)$, and $(\overline{Y}_6, \overline{Y_6^2}) = (1.88, 3.85)$. Using this data set, we compute the individual borrowing strength (InBS) and the overall borrowing index (OvBI) using MCMC samples of μ_j's collected from model (5.1), the "Independent model" (2.1) and the "Complete borrowing model" (3.2). For each MCMC run, a sample of size 1,000 was collected from 20,000 iterations by taking one point out of every 20 iterations, after a burning period of 2,000 iterations. Based on this sample, one set of estimates of InBS and OvBI can be calculated. To reduce the variability due to MCMC sampling, we took averages of InBS and OvBI computed using 200 independent MCMC posterior samples as our final outputs. The results are summarized in Fig. 3.

The intuitive understanding of the roles played by hyperparameters a_μ and K on the borrowing strengths within model (5.1) are quite clear: a bigger value of a_μ encourages the borrowing within the same cluster, while a larger value of K reduces the borrowing strength by creating more clusters. However, these qualitative insights do not paint a complete picture. For example, in Fig. 3(a), we can see that, when $K = 1$, there is a steep jump around $a_\mu = 10$ in terms of individual borrowing strengths. This phenomenon is also echoed by the sudden jump of the overall borrowing index around $a_\mu = 10$. While statistical practitioners tend to choose a very small value for a_μ in practice so that the prior is "non-informative", these surprising patterns in Fig. 3(a)–(b) suggest that not much of difference would be made as long as a_μ is less than 10. This observation can be explained by the fact that the sample means of groups 1–6 are quite different from each other, thus a small value of a_μ is not enough to force borrowing among different groups, see posterior densities of μ_j's in Fig. 3(c) for an illustration for the case with $a_\mu = 8$.

In Fig. 3(d)–(e), we study the impact of K, which controls the number of clusters. To force strong borrowing within the same cluster, a_μ is set to a high value 15. Generally speaking, the borrowing strength of the model (5.1) decreases as K increases, which is as expected. However, an interesting observation in Fig. 3(d) is that the InBS of group 3 and group 4

actually increases when K goes from 1 to 2, while the OvBI in Fig. 3(e) continues to decrease steadily as K increases. This can be explained from Fig. 3(f). When $K = 1$, posterior distributions of groups 3–4 under the "Independent model" are already quite close to those of the "Complete borrowing model" (which is also the poster of groups 3–4 under model (5.1) with $K = 1$ and $a_\mu = 15$), hence the InBS is low in this case. However, when $K = 2$, groups 3–4 are forced to be split up into two different clusters, steering away from their posterior distributions under the "Independent model". Hence, groups 3–4 borrow more information from other groups when $K = 2$, yielding higher InBS scores.

5.2. Binary outcome models

We now consider the following binary outcome hierarchical model:

$$X_j|p_j \sim Bin(p_j, n_j), \quad \text{logit}(p_j) \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \tau_{\mu,k}^{-1}), \quad \mu_k \sim N(0, \tau_\mu^{-1}),$$
$$\tau_{\mu,k} \sim Gamma(a_\mu, b_\mu), \text{ and } (\pi_1, \ldots, \pi_K) \sim Dirichlet(\alpha, \ldots, \alpha), \tag{5.2}$$

where all notations are the same as those defined in (5.1). The hyperparameters were set as $\alpha = 1/K$, $b_\mu = 0.01$, while a_μ and the number of clusters K may vary from case to case.

In this simulation, we used a realization of six groups of binomial outcomes with $(n_1, X_1) = (30, 5)$, $(n_2, X_2) = (14, 2)$, $(n_3, X_3) = (15, 7)$, $(n_4, X_4) = (25, 11)$, $(n_5, X_5) = (15, 5)$ and $(n_6, X_6) = (35, 10)$. Using this data set, we compute the individual borrowing strength (InBS) and the overall borrowing index (OvBI) using MCMC samples of p_j's collected from model (5.2), the "Independent model" (2.3) and the "complete borrowing" model by replacing (n_j, X_j) in (2.3) with $(\sum_{j=1}^{J} n_j, \sum_{j=1}^{J} X_j)$. For the "Independent model" and the "complete borrowing" model, we fix hyperparameters $\tau_\mu = a_\tau = 10^{-6}$ and $b_\tau = 0.01$. For each MCMC run, a sample of size 1,000 was collected from 20,000 iterations by taking one point out of every 20 iterations, after a burning period of 2,000 iterations. Based on this sample, one set of estimates of InBS and OvBI can be calculated. To reduce the variability due to MCMC sampling, we took averages of InBS and OvBI computed using 200 independent MCMC posterior samples as our final outputs. The results are summarized in Fig. 4.

The messages from Fig. 4 are similar to those from Fig. 3, except that Fig. 4(a)–(b) do not have a sudden increase of borrowing strength as a_μ increases. And the smallest OvBI in Fig. 4(b) is around 80%, as opposed to 0% in Fig. 3(b). This suggests that the tuning parameter a_μ may have a smaller role in enhancing strength of borrowing in the binary outcome models, compared to its role in the continuous outcome models.

To sum up, our simulation studies clearly demonstrate that the proposed InBS and OvBI measures can effectively help facilitate our understandings of the hyperparameters in a Bayesian hierarchical model, as well as discover irregularities when using certain choices of hyperparameters.

6. Real data analysis

In this section, the proposed InBS/OvBI measures were used to study the nonexchangeable product partition model (NEPPM) proposed in Leon-Novelo et al. (2012) with an application to a clinical study for sarcoma. Sarcoma is a rare form of cancer that affects soft and connective tissues of human body and many subtypes have been identified based on histology. In this clinical trial at MD Anderson Cancer Center, 179 patients with various subtypes of sarcoma were recruited to receive a treatment with irinotecan, a chemotherapy agent. Of these 179 patients, 10 subtypes of sarcoma were identified and classified by physicians into two categories (Intermediate prognosis and Good prognosis) based on the patients' characteristics. (Note: The data and the naming of the two prognosis groups were taken from Leon-Novelo et al. (2012).) The treatment efficacy of a patient is evaluated using the tumor shrinkage at the end of the second treatment cycle, with at least 30% shrinkage reported as a Success, and a 20% or more increase of tumor size as a failure. If a patient did not meet either of above criteria, he/she would be evaluated at the fourth treatment cycle, and again would be declared as a failure only if a 20% or more increase of tumor size was reported. The final results are summarized in Table 1. More details about the clinical trial can be found in Leon-Novelo et al. (2012).

Let X_j and n_j denote the number of successes and trials in the group of patients with subtype j, respectively, $j = 1, \ldots, 10$. Let z_j be the prognosis category of subtype j, which can take two possible values: "Intermediate prognosis" and "Good prognosis". The NEPPM model proposed in Leon-Novelo et al. (2012) aims at analyzing this data set using the model structure described in (2.4) with a specially designed partition prior $p(\mathcal{S}|\boldsymbol{\gamma}_0)$ as follows:

$$p\left(\mathcal{S} = \{S_1, \ldots, S_K\}|\boldsymbol{\gamma}_0\right) \propto \prod_{k=1}^{K} \underbrace{\{\alpha(\#S_k - 1)!\}}_{c_D(S_k)} \underbrace{\left\{\frac{\prod_{c=1}^{C} m_{kc}!}{(\#S_k + C - 1)!}\right\}^{\gamma}}_{d(S_k)}, \quad \alpha \sim Gamma(a_\alpha, b_\alpha), \tag{6.1}$$

where $\#S_k$ is the number of subtypes in cluster S_k, $C = 2$ is the number of prognosis categories, m_{kc} is the number of subtypes in cluster S_k that belong to prognosis category c, and $\boldsymbol{\gamma}_0 = (a_\alpha, b_\alpha, \gamma)$ is a set of positive numbers. In (6.1), $c_D(S_k)$

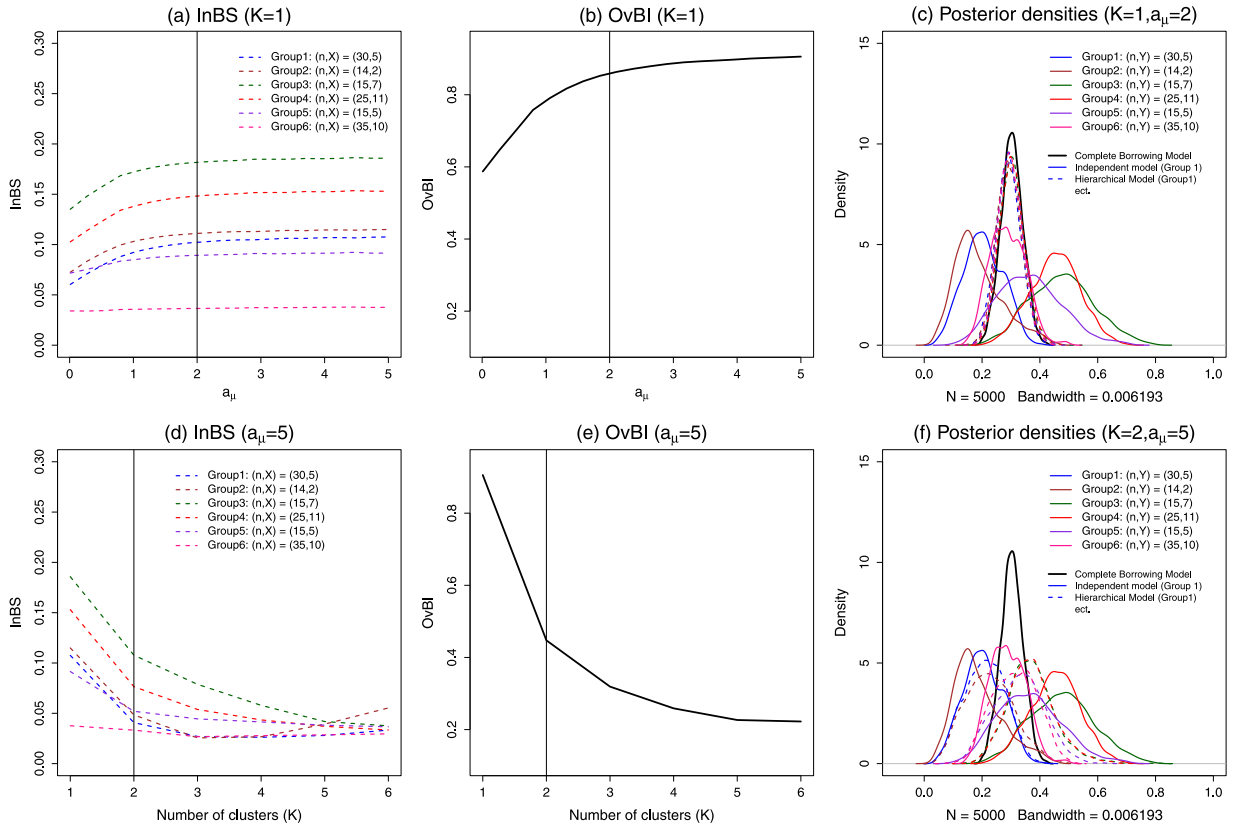

Fig. 4. Illustrations of Individual BS and Overall BI for the binary outcome.

Table 1
Reported number of Successes/Trials.

Subtype	Successes/Trials	Success Rate
Intermediate prognosis		
1. Leiomyosarcoma	6/28	0.214
2. Liposarcoma	7/29	0.241
3. Malignant fibrous histiocytoma	3/29	0.103
4. Osteosarcoma	5/26	0.192
5. Synovial	3/20	0.150
6. Angiosarcoma	2/15	0.133
7. Malignant peripheral nerve sheath tumor	1/5	0.200
8. Fibrosarcoma	1/12	0.083
Good prognosis		
9. Ewing's	0/13	0
10. Rhabdo	0/2	0

can be viewed as the cohesion measure induced by a Dirichlet process prior and $d(S_k)$ serves as a similarity function to encourage subtypes from the same prognosis category to form a cluster.

One major advantage of the NEPPM model (6.1) lies in that the number of clusters is stochastic, thus no pre-specified K is needed as required by the finite mixture model. However, the strength borrowing patterns of the NEPPM model are also much less apparent. A first look at (2.4) suggests that the hyperparameters a_τ, b_τ control the magnitude of τ_θ, which in turn controls the borrowing strengths among different clusters. The formulation of the partition prior (6.1) also suggests that the magnitude of α, controlled by hyperparameters a_α and b_α, determines the number of clusters formed in each iteration, which also should have some impact on the borrowing strengths of the model.

To quantify the borrowing strength patterns of the NEPPM model, we fix $\tau_\mu = 0.001, b_\tau = 0.1$ in (2.4) and $\gamma = 1, b_\alpha = 0.5$ in (6.1), as suggested in Leon-Novelo et al. (2012). The goal is to study the changing patterns of the proposed InBS and OvBI measures for various values of a_τ and a_α. In order to compute the InBS and OvBI, the posterior success rates from the "Independent model" is obtained by applying model (2.3) to each subtype group independently, while those from the "Complete borrowing model" are obtained by applying model (2.3) to the pooled success/trials

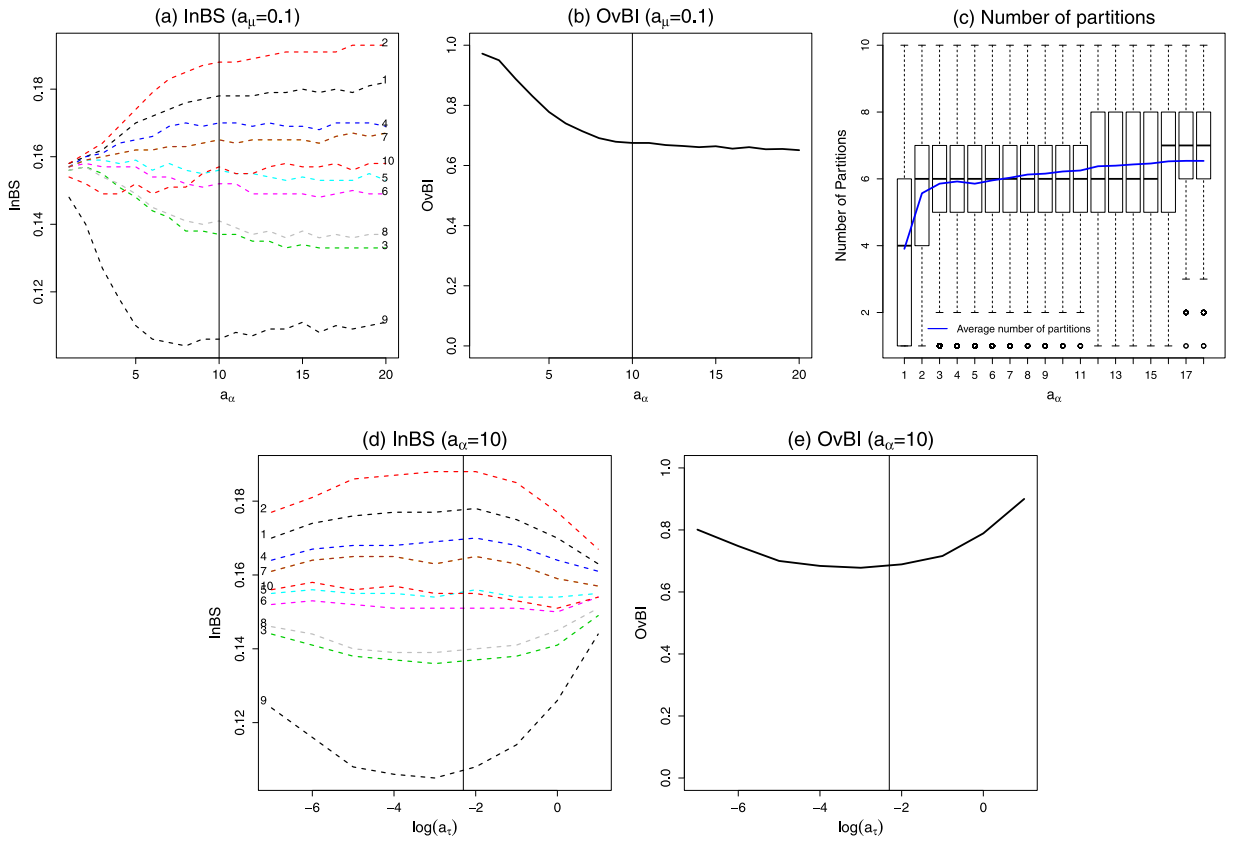

Fig. 5. Illustrations of InBS and OvBI for NEPPM model. The vertical lines indicate the hyperparameters used in Leon-Novelo et al. (2012) with $a_\alpha = 10$ and $a_\tau = 1/10$.

(i.e., 28/179) data by ignoring the subtype-classifications among patients. The InBS and OvBI were computed by averaging over 20 independent MCMC runs, and the results were summarized in Fig. 5.

The first noticeable feature in Fig. 5(a) is Group 9, whose individual borrowing strength decrease drastically as a_α increases from 1 to 5 and then stabilizes. This is because when a_α is small, we can see from Fig. 5(c) that the number of partitions is small, and hence all groups are forced to borrow from each other. This results in Fig. 5(a), where all groups have similar borrowing strength when $a_\alpha \leq 1$. As a_α increases, the average number of partitions increases, and the NEPPM model (6.1) encourages groups in the same category to be in the same partition. Therefore, when a_α exceeds 5 or so, Groups 9 and 10 are increasingly more likely to form their own partition, and therefore they will only borrow information from each other. This explains why both groups have stabilized InBS scores for $a_\alpha \geq 5$. Another interesting feature in Fig. 5(a) is that Groups 1,2,4,7 all have success rates greater than the overall success rate $28/179 = 0.156$, and their InBS scores all display increasing patterns as a_α increases. On the contrary, the InBS scores of Groups 8,3,9, whose success rates are less than the overall success rate, decreases as a_α increases. Similar phenomena are observed in 5(d) when a_τ increases. From 5(b), we can see that the overall borrowing strength of the NEPPM model (6.1) decreases steadily as a_α increases, and then stabilizes after a_α exceeds approximately 7. This echoes the observed pattern in Fig. 5(c), where the average number of partitions increase as a_α increases, yielding less overall borrowing across different subgroups. Lastly, from Fig. 5(d)–(e), we can see that the role played by hyperparameter a_τ, with regard to the borrowing strength, is quite different from that of a_α. Both InBS and OvBI scores are non-monotone functions of a_τ, which contradicts the intuition that a_τ should always encourage strength borrowing across different partitions. However, one should also notice that the magnitude of changes in OvBI scores caused by a_τ is relatively smaller than a_α, suggesting that the choice of a_α plays a more important role in determining the borrowing strength of the NEPPM model.

7. Discussion

We proposed an approach to quantify borrowing strength in Bayesian hierarchical models when applied to subgroup analysis. The InBS score measures the individual borrowing strength for each subgroup, and the OvBI score quantifies the overall borrowing strength of the model. Graphical representations of InBS and OvBI against different hyperparameters

paint a more complete picture of their roles in the Bayesian model, and therefore provide the practitioner a more quantitative understandings of the models under consideration. Such a graphical tool has the potential to help the practitioner to make more informed decisions on choosing a sensible model, and to reduce the level of subjectivity in choosing hyperparameters in data analysis.

However, the proposed approach does not tackle the problem of how to encourage "smart borrowing" among subgroups. In other words, the InBS and the OvBI scores do not reveal whether borrowed information is correct or incorrect. Various approaches have been proposed on this topic, e.g., Evans and Sedransk (2001), Quintana and Iglesias (2003), Müller et al. (2011). One interesting future research topic is how to incorporate the InBS and OvBI scores into such existing models, such that one can choose the hyperparameters that produce the largest amount of "correct" or "optimal" information borrowing under the proper context.

Acknowledgments

Ganggang Xu's research was supported by National Science Foundation Award SES-1902195. JJL's work was supported in part by grant CA016672, 1P50CA221703 from the National Cancer Institute, USA and RP160668 from the Cancer Prevention and Research Institute of Texas (CPRIT), USA. We appreciate the editorial assistance from Jessica Swann.

Appendix

Proof of Theorem 1. For any permutation $(l_1, \ldots, l_B)$ of $(1, \ldots, B)$, the definitions of $\widehat{d}_2\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right)$ in (3.8) ensure that

$$\widehat{d}_2^{\,2}\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right) \le \frac{1}{B}\sum_{i=1}^{B} |\,\Theta^{M_1}_{i,j} - \Theta^{M_2}_{l_i,j}\,|^2, \quad j = 1, \ldots, J,$$

with $\Theta^{M_k}_1, \ldots, \Theta^{M_k}_B$ being posterior samples of $\boldsymbol{\Theta}|\mathbf{Y}$ under model M_k, whose jth elements are $\mu^{M_k}_{j,i}$'s for $k = 1, 2$ and $j = 1, \ldots, J$. Therefore, we have

$$\sum_{j=1}^{J}\widehat{d}_2^{\,2}\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right) \le \frac{1}{B}\sum_{j=1}^{J}\sum_{i=1}^{B} |\,\Theta^{M_1}_{i,j} - \Theta^{M_2}_{l_i,j}\,|^2 = \frac{1}{B}\sum_{i=1}^{B} \|\,\Theta^{M_1}_{i} - \Theta^{M_2}_{l_i}\,\|^2$$

for any permutation $l_1, \ldots, l_B$, which further implies that

$$\sum_{j=1}^{J}\widehat{d}_2^{\,2}\left(F^{M_1}_{\mu_j|\mathbf{Y}}, F^{M_2}_{\mu_j|\mathbf{Y}}\right) \le d_2\left(\widehat{F}^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, \widehat{F}^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right),$$

where the inequality follows from the definition of $d_2\left(\widehat{F}^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, \widehat{F}^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$ in (3.7).

Next, by the inequality (6) in Dowson and Landau (1982), we have that

$$d_2\left\{N(\widehat{\boldsymbol{\mu}}_{M_1}, \widehat{\boldsymbol{\Sigma}}_{M_1}), N(\widehat{\boldsymbol{\mu}}_{M_2}, \widehat{\boldsymbol{\Sigma}}_{M_2})\right\} \le \mathbf{E}\|\boldsymbol{X}_{M_1} - \boldsymbol{X}_{M_2}\|^2,$$

where $\boldsymbol{X}_{M_k}$'s are random variables with mean $\widehat{\boldsymbol{\mu}}_{M_k}$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}_{M_k}$, $k = 1, 2$, and the expectation is taken over all joint distributions of $\boldsymbol{X}_{M_1}$ and $\boldsymbol{X}_{M_2}$. By Dowson and Landau (1982), the above equality holds when $\boldsymbol{X}_{M_k}$'s are normally distributed. Since by definition, the empirical distributions $\widehat{F}^{M_k}_{\boldsymbol{\Theta}|\mathbf{Y}}$'s have means $\widehat{\boldsymbol{\mu}}_{M_k}$'s and covariance matrices $\widehat{\boldsymbol{\Sigma}}_{M_k}$'s, we have that

$$d_2\left\{N(\widehat{\boldsymbol{\mu}}_{M_1}, \widehat{\boldsymbol{\Sigma}}_{M_1}), N(\widehat{\boldsymbol{\mu}}_{M_2}, \widehat{\boldsymbol{\Sigma}}_{M_2})\right\} \le d_2\left(\widehat{F}^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, \widehat{F}^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right),$$

which completes the proof of inequality $\widehat{d}^{*}_2\left(F^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, F^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right) \le d_2\left(\widehat{F}^{M_1}_{\boldsymbol{\Theta}|\mathbf{Y}}, \widehat{F}^{M_2}_{\boldsymbol{\Theta}|\mathbf{Y}}\right)$. □

References

Alvarez-Esteban, P.C., Del Barrio, E., Cuesta-Albertos, J.A., Matran, C., 2008. Trimmed comparison of distributions. J. Amer. Statist. Assoc. 103 (482), 697–704.

del Barrio, E., Giné, E., Matrán, C., 1999. Central limit theorems for the wasserstein distance between the empirical and the true distributions. Ann. Probab. 1009–1071.

Berger, J.O., Bernardo, J.M., Sun, D., 2009. The formal definition of reference priors. Ann. Statist. 905–938.

Berger, J.O., Wang, X., Shen, L., 2014. A Bayesian approach to subgroup identification. J. Biopharmaceutical Stat. 24 (1), 110–129.

Berry, S.M., Broglio, K.R., Groshen, S., Berry, D.A., 2013. BayesIan hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. Clinical Trials 1740774513497539.

Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Ann. Statist. 1196–1217.

Chen, N., Lee, J.J., 2019. BayesIan hierarchical classification and information sharing for clinical trials with subgroups and binary outcomes. Biom. J. 61 (5), 1219–1231.

Consonni, G., Veronese, P., 1995. A Bayesian method for combining results from several binomial experiments. J. Amer. Statist. Assoc. 90 (431), 935–944.
Diebolt, J., Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian sampling. J. R. Stat. Soc. Ser. B Stat. Methodol. 363–375.
Dowson, D., Landau, B., 1982. The Fréchet distance between multivariate normal distributions. J. Multivariate Anal. 12 (3), 450–455.
Evans, M., Moshonov, H., et al., 2006. Checking for prior-data conflict. Bayesian Anal. 1 (4), 893–914.
Evans, R., Sedransk, J., 2001. Combining data from experiments that may be similar. Biometrika 643–656.
Foster, J.C., Taylor, J.M., Ruberg, S.J., 2011. Subgroup identification from randomized clinical trial data. Stat. Med. 30 (24), 2867–2880.
Friede, T., Parsons, N., Stallard, N., 2012. A conditional error function approach for subgroup selection in adaptive clinical trials. Stat. Med. 31 (30), 4309–4320.
Fu, A.Y., Wenyin, L., Deng, X., 2006. Detecting phishing web pages with visual similarity assessment based on earth mover's distance (EMD). IEEE Trans. Dependable Secure Comput. 3 (4).
Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S., 2008. A weakly informative default prior distribution for logistic and other regression models. Ann. Appl. Stat. 1360–1383.
Grauman, K., Darrell, T., 2004. Fast contour matching using approximate earth mover's distance. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, vol. 1. IEEE, p. I.
Jiang, Y., He, Y., Zhang, H., 2016. Variable selection with prior information for generalized linear models via the prior Lasso method. J. Amer. Statist. Assoc. 111 (513), 355–376.
Jones, H.E., Ohlssen, D.I., Neuenschwander, B., Racine, A., Branson, M., 2011. BayesIan models for subgroup analysis in clinical trials. Clinical Trials 8 (2), 129–143.
Kuhn, H.W., 1955. The hungarian method for the assignment problem. Naval Res. Logistics Quart. 2 (1–2), 83–97.
Leon-Novelo, L., Bekele, B.N., Müller, P., Quintana, F., Wathen, K., 2012. Borrowing strength with nonexchangeable priors over subpopulations. Biometrics 68 (2), 550–558.
Levina, E., Bickel, P., 2001. The earth mover's distance is the mallows distance: Some insights from statistics. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, vol. 2. IEEE, pp. 251–256.
Lewis, C.J., Sarkar, S., Zhu, J., Carlin, B.P., 2019. Borrowing from historical control data in cancer drug development: A cautionary tale and practical guidelines. Stat. Biopharmaceutical Res. 11 (1), 67–78.
Mallows, C., 1972. A note on asymptotic joint normality. Ann. Math. Stat. 508–515.
Morita, S., Thall, P.F., Müller, P., 2008. Determining the effective sample size of a parametric prior. Biometrics 64 (2), 595–602.
Morita, S., Thall, P.F., Müller, P., 2012. Prior effective sample size in conditionally independent hierarchical models. Bayesian Anal. (Online) 7 (3).
Müller, P., Quintana, F., Rosner, G.L., 2011. A product partition model with regression on covariates. J. Comput. Graph. Statist. 20 (1), 260–278.
O'Hagan, A., et al., 2006. Science, subjectivity and software (comment on articles by Berger and by Goldstein). Bayesian Anal. 1 (3), 445–450.
Pollard, D., 2002. A User's Guide to Measure Theoretic Probability, vol. 8. Cambridge University Press.
Quintana, F.A., Iglesias, P.L., 2003. BayesIan clustering and product partition models. J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (2), 557–574.
Reimherr, M., Meng, X.-L., Nicolae, D.L., 2014. Being an informed Bayesian: Assessing prior informativeness and prior likelihood conflict. ArXiv preprint, arXiv:1406.5958.
Ren, Z., Yuan, J., Zhang, Z., 2011. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In: Proceedings of the 19th ACM International Conference on Multimedia. ACM, pp. 1093–1096.
Rubner, Y., Tomasi, C., Guibas, L.J., 2000. The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. 40 (2), 99–121.
Varadarajan, K.R., 1998. A divide-and-conquer algorithm for min-cost perfect matching in the plane. In: Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on. IEEE, pp. 320–329.
Woodcock, J., LaVange, L.M., 2017. Master protocols to study multiple therapies, multiple diseases, or both. New Engl. J. Med. 377 (1), 62–70.
Xu, G., Liang, F., Genton, M.G., 2015. A bayesian spatio-temporal geostatistical model with an auxiliary lattice for large datasets. Statist. Sinica 61–79.
Ye, K., Berger, J.O., 1991. Noninformative priors for inferences in exponential regression models. Biometrika 645–656.