# Bayesian Concentration Ratio and Dissonance[*]

Wei Shi[†], Ming-Hui Chen[‡], Lynn Kuo[§], and Paul O. Lewis[¶]

**Abstract.** We propose two new classes of Bayesian measure to investigate conflict among data sets from multiple studies. The first ("concentration ratio") is used to quantify the amount of information provided by a single data set through the comparison of the prior and its posterior distribution, or two data sets according to their corresponding posterior distributions. The second class ("dissonance") quantifies the extent of contradiction between two data sets. Both classes are based on volumes of highest density regions. They are well calibrated, supported by simulation, and computational algorithms are provided for their calculation. We illustrate these two classes in three real data applications: a benchmark dose toxicology study, a missing data study related to health effects of pollution, and a pediatric cancer study leveraging adult data.

**Keywords:** area under the curve, data compatibility, highest density regions, Monte Carlo methods, power prior.

## 1 Introduction

With the development of modern information technology, the need to compare and combine data across multiple studies has never been more important. For example, the ability to pool data and perform integrative data analysis is particularly important and timely in the addiction sciences (Conway et al., 2014). In phylogenomics, there is a need for both measuring the amount of information provided by sequence data from individual genes and quantifying the degree of gene tree conflict (Lewis et al., 2016). In meta-analysis, it is important to quantify the extent of heterogeneity among a collection of studies (Higgins and Thompson, 2002; Higgins et al., 2003). In pediatric drug development, difficulties arise in clinical trials due to a low incidence of the disease in children and many substantial challenges, including economic, logistical, technical, and ethical barriers. Recent attention has increasingly focused on extrapolation; that is, leveraging available data from adults or older age groups to draw conclusions for the pediatric population (Schoenfeld et al., 2009; Gamalo-Siebers et al., 2017; Lakshminarayanan and Natanegara, 2019).

When drawing inferences from various data sources, it is crucial to judge whether the data are compatible. Divergence is often used to establish the distance of one probability distribution from another as a means of assessing "incompatibility" in statistics and information geometry. Many kinds of divergence have been proposed. For exam-

---

[†]Department of Statistics, University of Connecticut, wei.shi@uconn.edu

[‡]Department of Statistics, University of Connecticut, ming-hui.chen@uconn.edu

[§]Department of Statistics, University of Connecticut, lynn.kuo@uconn.edu

[¶]Department of Ecology and Evolutionary Biology, University of Connecticut, paul.lewis@uconn.edu

ple, f-divergence (Ali and Silvey, 1966; Csiszár and Shields, 2004; Morimoto, 1963) (including the Kullback-Leibler divergence (Kullback and Leibler, 1951) and Hellinger distance (Hellinger, 1909)), Rényi's divergence (Rényi, 1961), and Bregman divergence (Bregman, 1967). Many divergence measures are not easy to compute empirically and it is hard to tell when the conflict is severe. In the Bayesian setting, Shi (2017) defined a partition-based measure to quantify the compatibility of two data sets based on their posterior distributions. Methods for detecting and measuring conflict using $p$-values include Marshall and Spiegelhalter (2007), Gåsemyr and Natvig (2009), Evans and Moshonov (2006), Presanis et al. (2013), and Nott et al. (2020). While $p$-value is a single summary, to extend its scope, we propose two new classes of measures: concentration ratio measure focusing on scale change and dissonance measure focusing on location shift. Both measures are easy to interpret and can be computed for any multi-modal, skewed, or complex distribution.

Lindley (1956) defined the information provided by an experiment as the difference between the posterior and prior differential entropies. Information is the resolution of uncertainty (Shannon, 1948); it reduces the set of values deemed plausible. A $100(1-\alpha)\%$ Bayesian credible region provides one way to define a set of plausible values after the data are collected. When a distribution is not symmetric or unimodal, a $100(1-\alpha)\%$ highest density region (HDR) is more desirable than other regions, as every point inside the HDR has higher density value than every point outside.

We propose two novel classes of Bayesian measures based on HDRs. $LCR(\alpha)$ (Logarithm of the Concentration Ratio) is defined as the log-ratio of the volume of two HDRs at level $\alpha$. $LCR(\alpha)$ indicates whether one distribution is more dispersed than the other, however, it cannot detect conflict caused by a location shift. Therefore $\mathcal{D}(\alpha)$ (dissonance) is introduced as the fraction of the volume of the smaller HDR that is not overlapping with the larger HDR at the $\alpha$ level.

Computing HDR volume is required for both measures. In a one-dimensional case, the HDR consists of the highest density interval and thus the volume is measured by the interval length. Chen and Shao (1999) provided a Monte Carlo method to estimate the HD interval for a unimodal distribution. For two or more dimensions, the computation of the HDR is more complicated, so is the estimation of its volume. Tanner (1996) provided an algorithm to calculate the content and boundary of the HDR using a normal approximation. However, Tanner's algorithm does not work well when the distribution is multi-modal or skewed, and computation of the volume of the HDR remains unaddressed. To overcome such difficulties, we propose two two-stage Monte Carlo algorithms to first estimate the HDR and then compute its volume. The algorithms are then applied to calculating $LCR(\alpha)$ and $\mathcal{D}(\alpha)$. Simulation studies are provided in Appendix C in Supplementary Material (Shi et al., 2021) to examine the empirical performance of the algorithms.

The article is organized as follows. Section 2 presents a detailed development of $LCR(\alpha)$ and $\mathcal{D}(\alpha)$. We propose two two-stage Monte Carlo algorithms to estimate the volume of a HDR in Section 3.1 and use them to calculate $LCR(\alpha)$ and $\mathcal{D}(\alpha)$ in Section 3.2. To gain further insight, we compare our dissonance measure to conflict $p$-value

approaches in Evans and Moshonov (2006) and Presanis et al. (2013) for normal location and binomial problems in Section 4. We display the results of concentration ratio measure and dissonance measure for several scenarios, including location shift, scale change, tail change, multi-modal and skewed distributions in Section 5. In Section 6, the proposed $LCR(\alpha)$ measure and the integrated $\mathcal{D}(\alpha)$ measure are applied to three studies: (1) a benchmark dose toxicology study, (2) evaluating missing data information in a health effect study on respiratory function for children, and (3) investigating how to leverage adult data to analyze pediatric data. HDR has been criticized for being not invariant under reparametrization, so we explore an invariant HDR in Section 7. We conclude the article with a brief discussion in Section 8.

## 2 Concentration Ratio and Dissonance

### 2.1 HDR

For a parameter $\boldsymbol{\theta}$ with the probability density function $f(\boldsymbol{\theta})$, the $100(1-\alpha)\%$ highest density region (HDR) is the subset of the parameter space $\Theta$ such that

$$R(\alpha) = \{\boldsymbol{\theta} \in \Theta : f(\boldsymbol{\theta}) \geq f_\alpha\}, \tag{2.1}$$

where $f_\alpha$ is the largest constant such that $\Pr(\boldsymbol{\theta} \in R(\alpha)) \geq 1 - \alpha$.

Recognizing that log-concave distributions are frequently encountered in practice, we provide an important property of a HDR for a log-concave distribution in the following theorem.

**Theorem 2.1.** *For a log-concave density, the $100(1-\alpha)\%$ HDR is a closed convex set.*

*Proof.* For any two points $\boldsymbol{a}, \boldsymbol{b}$ within the $100(1-\alpha)\%$ HDR, we have $f(\boldsymbol{a}) \geq f_\alpha, f(\boldsymbol{b}) \geq f_\alpha$. Then, by the log-concavity of $f$, for any $\lambda \in (0,1)$,

$$f(\lambda\boldsymbol{a} + (1-\lambda)\boldsymbol{b}) \geq f(\boldsymbol{a})^\lambda f(\boldsymbol{b})^{1-\lambda} \geq f_\alpha^\lambda f_\alpha^{1-\lambda} = f_\alpha.$$

Therefore, the point $\lambda\boldsymbol{a} + (1-\lambda)\boldsymbol{b}$ also falls within the $100(1-\alpha)\%$ HDR. $\qquad\square$

### 2.2 Definitions of Concentration Ratio and Dissonance

Consider two probability distributions of $\boldsymbol{\theta}$ with probability densities $f_1(\boldsymbol{\theta})$ and $f_2(\boldsymbol{\theta})$, and $R_1(\alpha)$ and $R_2(\alpha)$ as their corresponding $100(1-\alpha)\%$ HDRs. We have

$$R_i(\alpha) = \{\boldsymbol{\theta} \in \Theta : f_i(\boldsymbol{\theta}) \geq f_{\alpha,i}\}, \quad i = 1, 2,$$

where $f_{\alpha,i}$ is the largest constant such that $\Pr(\boldsymbol{\theta} \in R_i(\alpha)) \geq 1 - \alpha$.

Our measure $LCR(\alpha)$ is based on the comparison of $V(R_1(\alpha))$ and $V(R_2(\alpha))$, where $V(\cdot)$ denotes the volume.

**Definition 1.** *Let the concentration ratio $CR(\alpha) = V(R_1(\alpha))/V(R_2(\alpha))$. The logarithm of the concentration ratio of the first distribution to the second distribution is*

*defined as*

$$LCR(\alpha) = \log CR(\alpha) = \log \frac{V(R_1(\alpha))}{V(R_2(\alpha))}.$$

For a fixed $\alpha$, $LCR(\alpha)$ measures the volume change between two HDRs. When the first distribution is the prior distribution and the second one the posterior distribution, $LCR(\alpha)$ quantifies the amount of information provided by the data beyond our prior belief.



(a) No Concentration Change     (b) Positive $LCR(\alpha)$     (c) Negative $LCR(\alpha)$
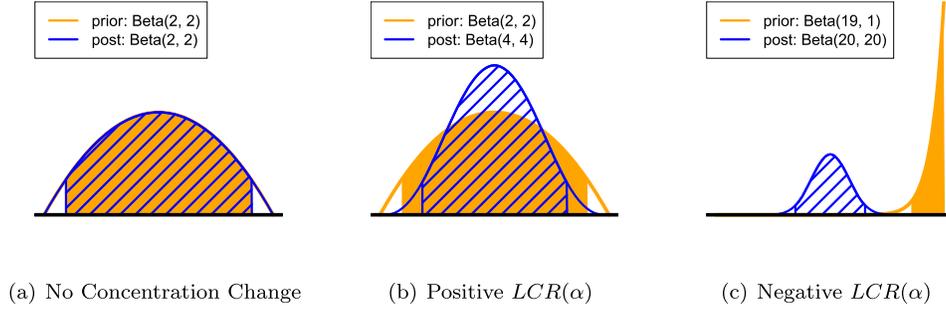
Figure 1: Graphical illustration of $LCR(\alpha)$ in the one-dimensional case.

When the posterior distribution is identical to the prior, the ratio $CR(\alpha) = 1$ and thus $LCR(\alpha) = 0$ (Figure 1(a)). The set of plausible values remains the same after the data are considered, which means no concentration change is provided by the data in addition to the prior knowledge. When the volume of the posterior HDR $V(R_2(\alpha))$ is smaller than $V(R_1(\alpha))$, the ratio $CR(\alpha) > 1$, which results in positive $LCR(\alpha)$ (Figure 1(b)). In this case, the data effectively exclude some outcomes which are considered plausible by the prior. Finally, data may make us less certain about the value of the parameter, increasing the size of the set of plausible values specified by the prior. This represents negative $LCR(\alpha)$ (Figure 1(c)).

**Remark 1.** *For $\boldsymbol{\theta} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the $100(1-\alpha)\%$ HDR is*

$$R(\alpha) = \left\{ \theta \in \Theta : (\boldsymbol{\theta} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \leq \chi^2_{k,\alpha} \right\}, \tag{2.2}$$

*where $\chi^2_{k,\alpha}$ is the upper $100\alpha th$ percentile of the chi-square distribution with $k$ degrees of freedom. From Appendix A in Supplementary Material (Shi et al., 2021), the volume of (2.2) is*

$$V(R(\alpha)) = \frac{\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)} \left(\chi^2_{k,\alpha}\right)^{\frac{k}{2}} (\det(\boldsymbol{\Sigma}))^{\frac{1}{2}}.$$

*Thus, when comparing two normal distributions $N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$,*

$$LCR(\alpha) = \frac{1}{2} \log \frac{\det(\boldsymbol{\Sigma}_1)}{\det(\boldsymbol{\Sigma}_2)}. \tag{2.3}$$

*Note that in this case, $LCR(\alpha)$ has the same value over different choices of $\alpha$. Assuming $\det(\boldsymbol{\Sigma}_1)/\det(\boldsymbol{\Sigma}_2) = r$, then for $r = 2, 4, 8, 16, 32$, $LCR(\alpha) = 0.35, 0.69, 1.04, 1.39, 1.73$. This can be used as a guide for calibrating $LCR(\alpha)$ when comparing approximately normal distributions.*

While $LCR(\alpha)$ compares the dispersion of each distribution, it cannot capture conflict caused by a location shift. Therefore, we introduce a dissonance measure, $\mathcal{D}(\alpha)$, which complements $LCR(\alpha)$ by focusing on how far apart the two distributions are.

**Definition 2.** *For $R_1(\alpha)$ and $R_2(\alpha)$, let $R_{\min}(\alpha)$ denote the region with a smaller volume and $R_{\max}(\alpha)$ denote the larger one. The dissonance $\mathcal{D}(\alpha)$ is measured as the fraction of the volume of the smaller HDR that is not overlapping with the larger HDR, i.e.,*

$$\mathcal{D}(\alpha) = \frac{V(R_{\min}(\alpha) \cap \overline{R_{\max}(\alpha)})}{V(R_{\min}(\alpha))}, \tag{2.4}$$

*where $R_{\min}(\alpha) \cap \overline{R_{\max}(\alpha)} = \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} \in R_{\min}(\alpha), \boldsymbol{\theta} \notin R_{\max}(\alpha)\}$.*



(a) No Dissonance          (b) Partial Dissonance          (c) Complete Dissonance
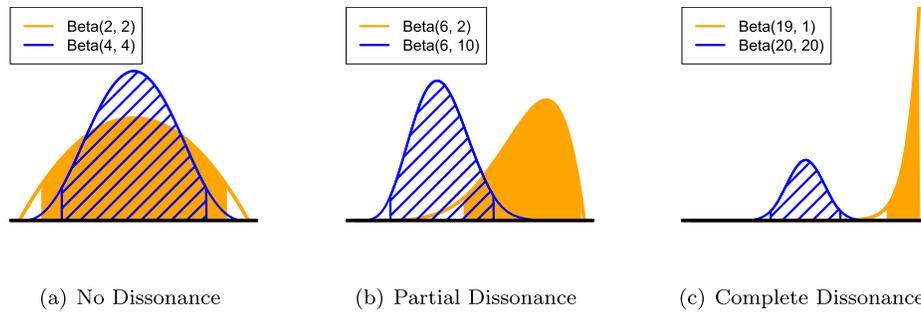
Figure 2: Graphical illustration of $\mathcal{D}(\alpha)$ in the one-dimensional case.

As $\mathcal{D}(\alpha)$ is defined as the fraction of conflict in terms of the volume of the smaller HDR, the range of $\mathcal{D}(\alpha)$ is $[0, 1]$. When the smaller HDR $R_{\min}(\alpha)$ entirely falls within the larger HDR $R_{\max}(\alpha)$, $\mathcal{D}(\alpha) = 0$, referred as 'no dissonance' (Figure 2(a)). When there is partial overlap of two HDRs (Figure 2(b)), referred as 'partial dissonance'; $\mathcal{D}(\alpha)$ is between 0 and 1. When two HDRs are disjoint, $\mathcal{D}(\alpha) = 1$, indicating complete dissonance (Figure 2(c)).

## 2.3   Summarizing Concentration Ratio and Dissonance over Different Content Levels

The measures $LCR(\alpha)$ and $\mathcal{D}(\alpha)$ are related to the content level $\alpha$. $LCR(\alpha)$ is relatively stable over multiple choices of $\alpha$. We show in (2.3) that when comparing two normal distributions, $LCR(\alpha)$ is actually unrelated to $\alpha$. In our first real data analysis (Section 6.1), we present results of $LCR(\alpha)$ for 5 choices of $\alpha$ to illustrate the stability of $LCR(\alpha)$ empirically.

In contrast, $\mathcal{D}(\alpha)$ is very sensitive to the choice of $\alpha$. In the partial dissonance case as in Figure 2(b), it is trivial to find a content level $\alpha$ leading to complete dissonance and thus the measure $\mathcal{D}(\alpha)$ over different $\alpha$ values yields differing conclusions. As $\mathcal{D}(\alpha)$ is a function of $\alpha$ and bounded by 0 and 1, we plot the curve of $\mathcal{D}(\alpha)$ versus $\alpha$ and summarize the extent of conflict using the area under the curve, denoted as $\mathcal{D}$-AUC. The formal definition of $\mathcal{D}$-AUC is given as follows:

**Definition 3.** *$\mathcal{D}$-AUC is the area under the $\mathcal{D}(\alpha)$ curve, i.e.,*

$$\mathcal{D}\text{-}AUC = \int_0^1 \mathcal{D}(\alpha)d\alpha.$$

*It is a summary of $\mathcal{D}(\alpha)$ over all possible choices of $\alpha$ and measures the degree of conflict between two distributions.*

We can revisit the three examples in Figure 2 to better understand $\mathcal{D}$-AUC. In the case of no dissonance, $\mathcal{D}(\alpha)$ is always 0 over different values of $\alpha$ as the smaller HDR is inside the larger HDR for whatever content level. For partial dissonance, $\mathcal{D}(\alpha)$ starts with value 0 when $\alpha = 0$, i.e., the 100% HDRs of two distributions are both the whole parameter space and thus no dissonance. As $\alpha$ increases, that is when we compare $99\%, 98\%, \ldots$ HDRs, we begin to have partial non-overlapping regions and thus $\mathcal{D}(\alpha)$ has a value between 0 and 1. At a certain content level $\alpha$, two HDRs become disjoint and $\mathcal{D}(\alpha)$ equals 1 and stays at 1 thereafter. In the extreme case of complete dissonance, $\mathcal{D}(\alpha) = 1$ for all values of $\alpha > 0$, resulting in $\mathcal{D}$-AUC $= 1$.

In conclusion, the range of $\mathcal{D}$-AUC is also $[0, 1]$, with smaller value suggesting compatibility and larger value indicating contradiction. The interpretation is consistent with that of $\mathcal{D}(\alpha)$ with a fixed $\alpha$. Moreover, $\mathcal{D}$-AUC avoids the necessity for an arbitrary choice of $\alpha$. In our real data applications, we use $\mathcal{D}$-AUC to report the results of dissonance.

## 2.4   Prior-Data Comparison and Data-Data Comparison

We have defined $LCR(\alpha)$ and $\mathcal{D}(\alpha)$ in very general terms, so they can be applied to comparing prior and data, or two data sets.

In the context of prior-data comparison, in addition to comparing prior versus posterior distribution, we can also compare prior and likelihood by converting the likelihood function to a posterior distribution through a noninformative prior. More details are given in Section 4.

For the data-data comparison, consider two data sets $S_1$ and $S_2$ from similar studies with common parameter $\boldsymbol{\theta}$ defined on the same parameter space $\Theta$. We use the same prior distribution for $\boldsymbol{\theta}$ for each data set and compare the two data sets based on their posterior distributions.

Letting $R_1(\alpha)$ and $R_2(\alpha)$ be the posterior HDRs for $S_1$ and $S_2$, respectively, $LCR(\alpha)$ defined in Definition 1 compares the concentration of $S_1$ to that of $S_2$. Consequently, $LCR(\alpha) = 0$ means that $S_1$ and $S_2$ are equally concentrated; $LCR(\alpha) > 0$ indicates that

$S_2$ is more concentrated than $S_1$; and $LCR(\alpha) < 0$ implies that $S_2$ is less concentrated than $S_1$.

To construct a dissonance measure $\mathcal{D}(\alpha)$ between these two data sets, we can apply Definition 2 with $R_1(\alpha)$ and $R_2(\alpha)$ specified to be the posterior HDRs for $S_1$ and $S_2$, respectively. Definition 3 for $\mathcal{D}$-AUC still applies in this case.

**Remark 2.** *We expect $LCR(\alpha)$ for comparing two data sets to be robust to the prior choice, given that it is evaluated from the "central" region of each posterior distribution. $\mathcal{D}(\alpha)$ is more complicated. Although $\mathcal{D}(\alpha)$ is also evaluated from two posterior distributions, the numerator of (2.4) depends on the shapes of each distribution, so this measure is not expected to be robust to the prior choice. Therefore, we have used diffuse priors in real data applications to focus more on the data.*

## 3 Computational Algorithm

### 3.1 Computation of the Volume of an HDR

A two-stage Monte Carlo algorithm is developed to estimate the volume of the $100(1 - \alpha)\%$ HDR in (2.1). It is often difficult to calculate $f_\alpha$ in (2.1) analytically. Therefore, in the first stage, we obtain an approximation to $f_\alpha$ through a Monte Carlo technique. Let $\psi = f(\boldsymbol{\theta})$, which is obtained by transforming $\boldsymbol{\theta}$ by its own density function. Then $f_\alpha$ is the $100\alpha^{th}$ percentile of $\psi$ since it satisfies $\Pr(f(\boldsymbol{\theta}) \geq f_\alpha) = 1 - \alpha$. Therefore $f_\alpha$ can be estimated by the corresponding sample quantile of $\psi$, denoted as $\hat{f}_\alpha$. The approximated $100(1 - \alpha)\%$ HDR is therefore defined as $\hat{R}(\alpha) = \{\boldsymbol{\theta} \in \Theta : f(\boldsymbol{\theta}) \geq \hat{f}_\alpha\}$. According to Hyndman (1996), when the sample size goes to infinity, $\hat{f}_\alpha \to f_\alpha$ and so $\hat{R}(\alpha) \to R(\alpha)$. In the second stage, we employ a needle-dropping technique to estimate the volume. We first construct a hyper-rectangle with known volume that covers the approximated region $\hat{R}(\alpha)$, generate a large number of random points uniformly from the hyper-rectangle region and count the proportion of points that fall within $\hat{R}(\alpha)$ to estimate the volume. Again, as the number of points increases, the estimated volume $\hat{V}(\hat{R}(\alpha))$ approaches the actual volume of $\hat{R}(\alpha)$. Detailed steps are summarized in Algorithm 1.

When the distribution is approximately normal, we can improve the algorithm in the second stage using a hyper-ellipsoid region. Since an HDR of a multivariate normal distribution is a hyper-ellipsoid (2.2), we construct a hyper-ellipsoid that just covers the approximated region $\hat{R}(\alpha)$. Given the estimated mean ($\hat{\boldsymbol{z}}$) and covariance matrix (**S**), the approximated hyper-ellipsoid is $\{\boldsymbol{z} : (\boldsymbol{z} - \hat{\boldsymbol{z}})^\top \mathbf{S}^{-1}(\boldsymbol{z} - \hat{\boldsymbol{z}}) \leq \chi^2_{k,\alpha}\}$. This hyper-ellipsoid may not be large enough to contain all points known to be in the HDR, so the lengths of its semi-axes are increased by 2% and the constructed hyper-ellipsoid is $\{\boldsymbol{z} : (\boldsymbol{z} - \hat{\boldsymbol{z}})^\top \mathbf{S}^{-1}(\boldsymbol{z} - \hat{\boldsymbol{z}}) \leq 1.02^2 \times \chi^2_{k,\alpha}\}$. An expansion factor of 2% (1.02) is arbitrary, but the expansion factor may be determined in any particular case by finding the inflection point in a plot of estimated HDR volume against expansion factor. The estimated HDR volume increases linearly while the expansion factor is too small, then flattens when the hyper-ellipsoid defined by the expansion factor is large enough to contain the true HDR. Similarly, we generate uniformly distributed random points from this hyper-ellipsoid region using the method in Dezert and Musso (2001) and calculate the proportion falling within $\hat{R}(\alpha)$. Algorithm 2 gives the detailed steps.

---

**Algorithm 1** Computing the volume of a HDR using a hyper-rectangle region.

1. Draw a Markov chain Monte Carlo (MCMC) sample $\{\boldsymbol{\theta}_t, t = 1, \ldots, T\}$ from $f(\boldsymbol{\theta})$.
2. Calculate $f_t = f(\boldsymbol{\theta}_t)$ for $t = 1, \ldots, T$. Choose $\hat{f}_\alpha = f_{(j)}$, where $f_{(j)}$ is the $j$th smallest number of $\{f_t\}$ and $j = [\alpha T]$, the integer part of $\alpha T$. The approximated $100(1 - \alpha)\%$ HDR is given by $\hat{R}(\alpha) = \{\boldsymbol{\theta} \in \Theta : f(\boldsymbol{\theta}) \geq \hat{f}_\alpha\}$. (Hyndman, 1996).
3. Obtain the minimum and maximum of each dimension in the MCMC sample $\{\boldsymbol{\theta}_t\}$, denoted as $\{[\theta_{1(1)}, \theta_{1(T)}], \ldots, [\theta_{k(1)}, \theta_{k(T)}]\}$. Draw a sample $\{\boldsymbol{z}_i, i = 1, \ldots, N\}$ uniformly from the hyper-rectangle region $\{[\theta_{1(1)}, \theta_{1(T)}] \times \cdots \times [\theta_{k(1)}, \theta_{k(T)}]\}$.
4. Calculate the proportion $p$ of $\{\boldsymbol{z}_i\}$ that falls within $\hat{R}(\alpha)$, i.e.,

$$p = \frac{\#\{i : f(\boldsymbol{z}_i) \geq \hat{f}_\alpha\}}{N},$$

   where $\#\{\cdot\}$ denotes the cardinality of $\{\cdot\}$.
5. The estimated HDR volume is given by

$$\hat{V}(\hat{R}(\alpha)) = p \times (\theta_{1(T)} - \theta_{1(1)}) \times \cdots \times (\theta_{k(T)} - \theta_{k(1)}).$$

**Remark 3.** *The density function $f(\cdot)$ in Algorithm 1 is only required to be known up to a normalizing constant.*

---

---

**Algorithm 2** Computing the volume of a HDR using a hyper-ellipsoid region.

Follow the same steps as in Algorithm 1, except steps 3 and 5 are replaced as follows:

$3'$. Obtain the sample mean $(\hat{\boldsymbol{z}})$ and sample covariance matrix $(\mathbf{S})$ from the MCMC sample $\{\boldsymbol{\theta}_t\}$. Generate random points $\{\boldsymbol{z}_i, i = 1, \ldots, N\}$ uniformly distributed in hyper-ellipsoid $\{\boldsymbol{z} : (\boldsymbol{z} - \hat{\boldsymbol{z}})^\top \mathbf{S}^{-1}(\boldsymbol{z} - \hat{\boldsymbol{z}}) \leq 1.02^2 \times \chi^2_{k,\alpha}\}$ (Dezert and Musso, 2001):

   - Generate random points $\{\boldsymbol{u}_i, i = 1, \ldots, N\}$ by drawing k-dimensional standard normal random vectors $\boldsymbol{a}_i$ independently and computing $\boldsymbol{u}_i = \boldsymbol{a}_i/\|\boldsymbol{a}_i\|$, where $\|\cdot\|$ represents the Euclidean norm.
   - Generate $r_i = v_i^{1/k}, i = 1, \ldots, N$ with $v_i \overset{iid}{\sim} U(0, 1)$.
   - Compute $\boldsymbol{y}_i = r_i \boldsymbol{u}_i, i = 1, \ldots, N$.
   - Compute the Cholesky decomposition of $\mathbf{S}$, where $\mathbf{S} = \mathbf{T}\mathbf{T}^\top$.
   - Compute $\boldsymbol{x}_i = \mathbf{T}\boldsymbol{y}_i, i = 1, \ldots, N$.
   - Compute and return $\boldsymbol{z}_i = 1.02 \times \sqrt{\chi^2_{k,\alpha}} \boldsymbol{x}_i + \hat{\boldsymbol{z}}$.

$5'$. The estimated HDR volume is given by

$$\hat{V}(\hat{R}(\alpha)) = p \times \frac{\left(\pi \chi^2_{k,\alpha}\right)^{\frac{k}{2}} \times 1.02^k}{\Gamma(\frac{k}{2} + 1)} (\det(\mathbf{S}))^{\frac{1}{2}}.$$

---

In Appendix B in Supplementary Material (Shi et al., 2021), we explain why we drop needles uniformly to estimate the volume.

We evaluate the performance of both algorithms using simulation studies for a bivariate normal distribution and a 5-dimensional normal distribution and the details are given in Appendix C in Supplementary Material (Shi et al., 2021). Our simulation results show that, when $T$ is large enough, increasing $N$ improves performance, and, when $N$ is large enough, increasing $T$ also improves the performance. For the bivariate normal case, the simulation study suggests we need at least $T = 10^4$ and $N = 10^4$ for the hyper-rectangle method (Algorithm 1) and at least $T = 10^4$ and $N = 10^2$ for the hyper-ellipsoid method (Algorithm 2). For the 5-dimensional normal case, at least $T = 10^4$ and $N = 10^6$ are required for the hyper-rectangle method and at least $T = 10^4$ and $N = 10^3$ are required for the hyper-ellipsoid method. As expected, Algorithm 2 produces more accurate results than Algorithm 1 for both normal distributions.

## 3.2   Calculating Concentration Ratio and Dissonance

We now list the steps required to compute our measures using the algorithms of the previous section.

**Step 1.** Apply Algorithm 1 (Algorithm 2) to the distribution $f_1(\boldsymbol{\theta})$ to obtain $\hat{V}(\hat{R}_1(\alpha))$. Denote the corresponding approximated cut-off value as $\hat{f}_{\alpha,1}$ and the needle-dropping sample as $\{\boldsymbol{z}_{i,1}\}$.

**Step 2.** Apply Algorithm 1 (Algorithm 2) to the distribution $f_2(\boldsymbol{\theta})$ to obtain $\hat{V}(\hat{R}_2(\alpha))$. Denote the corresponding approximated cut-off value as $\hat{f}_{\alpha,2}$ and the needle-dropping sample as $\{\boldsymbol{z}_{i,2}\}$.

**Step 3.** Calculate $LCR(\alpha)$, which is

$$LCR(\alpha) = \log\left\{\frac{\hat{V}(\hat{R}_1(\alpha))}{\hat{V}(\hat{R}_2(\alpha))}\right\}.$$

**Step 4.** After evaluation, we know which HDR is smaller.

If $\hat{R}_{\min}(\alpha) = \hat{R}_1(\alpha)$, calculate $\mathcal{D}(\alpha)$ as

$$\mathcal{D}(\alpha) = \frac{\#\{i : f_1(\boldsymbol{z}_{i,1}) \geq \hat{f}_{\alpha,1}, f_2(\boldsymbol{z}_{i,1}) < \hat{f}_{\alpha,2}\}}{\#\{i : f(\boldsymbol{z}_{i,1}) \geq \hat{f}_{\alpha,1}\}};$$

If $\hat{R}_{\min}(\alpha) = \hat{R}_2(\alpha)$, calculate $\mathcal{D}(\alpha)$ as

$$\mathcal{D}(\alpha) = \frac{\#\{i : f_2(\boldsymbol{z}_{i,2}) \geq \hat{f}_{\alpha,2}, f_1(\boldsymbol{z}_{i,2}) < \hat{f}_{\alpha,1}\}}{\#\{i : f_2(\boldsymbol{z}_{i,2}) \geq \hat{f}_{\alpha,2}\}}.$$

# 4    Comparing Dissonance and Conflict P-Values

To better understand our dissonance measure $\mathcal{D}$-AUC, we compare it with two conflict $p$-value approaches in detecting prior-data conflict.

Evans and Moshonov (2006) explored prior-data conflict using the $p$-value of a minimal sufficient statistic with its prior-predictive distribution as the reference.

Presanis et al. (2013) discussed another type of conflict $p$-value using a node splitting approach that evaluated a conflict locally at a particular node or group of nodes in a directed acyclic graph. The evidence about $\theta$ is split into two parts, resulting in two independent distributions $p(\theta_a|\boldsymbol{y}_a)$ and $p(\theta_b|\boldsymbol{y}_b)$. They defined $\delta = h(\theta_a) - h(\theta_b)$, where $h(\cdot)$ is a Jeffreys' transformation, and calculated the $p$-value as the probability that the density of $\delta$ is smaller than that at 0, i.e.,

$$p_{\text{Presanis}} = \Pr(p_\delta(\delta|\boldsymbol{y}_a, \boldsymbol{y}_b) < p_\delta(0|\boldsymbol{y}_a, \boldsymbol{y}_b)).$$

In this section, we compare our $p$-values of $\mathcal{D}$-AUC (prior/posterior) and a variation of $\mathcal{D}$-AUC (prior/likelihood), denoted by $\tilde{\mathcal{D}}$-AUC, to the above Bayesian conflict $p$-values with a normal location example first, then a binomial success probability example.

## 4.1    Normal Location Model

Suppose we have a single observation $y$ from a normal distribution with known variance. Then $y \sim N(\mu, \sigma^2)$, and the prior for $\mu$ is assumed to be $N(\mu_0, \sigma_0^2)$. Let $w = \sigma_0^2/(\sigma_0^2+\sigma^2)$, the posterior distribution for $\mu$ is $N((1-w)\mu_0 + wy, w\sigma^2) = N(\hat{\mu}, \hat{\sigma}^2)$, where $\hat{\mu}$ denotes the posterior mean of $\mu$ with posterior standard deviation $\hat{\sigma}$.

The prior predictive distribution of $y$ is $N(\mu_0, \sigma^2 + \sigma_0^2)$ with density denoted by $p(y)$. The Bayesian conflict $p$-value of Evans and Moshonov (2006), denoted by $p_{\text{EM}}$, is

$$p_{\text{EM}} = \Pr(p(Y) \leq p(y)) = 2\left(1 - \Phi\left(\frac{|y - \mu_0|}{\sqrt{\sigma^2 + \sigma_0^2}}\right)\right).$$

Instead of evaluating the conflict between prior and posterior, we can compare prior and likelihood by converting the likelihood to a posterior distribution through a noninformative prior, using the setup by Presanis et al. (2013):

$$\text{data-translated likelihood: } \mu_a \sim N(y, \sigma^2),$$
$$\text{prior: } \mu_b \sim N(\mu_0, \sigma_0^2),$$
$$\delta = \mu_a - \mu_b \sim N(y - \mu_0, \sigma^2 + \sigma_0^2),$$
$$p_{\text{Presanis}} = \Pr(p_\delta(\delta) < p_\delta(0)) = 2\left(1 - \Phi\left(\frac{|y - \mu_0|}{\sqrt{\sigma^2 + \sigma_0^2}}\right)\right).$$

In this case, it is the same as $p_{\text{EM}}$.

When comparing prior $N(\mu_0, \sigma_0^2)$ versus posterior $N(\hat{\mu}, \hat{\sigma}^2)$, our dissonance measure

$$\mathcal{D}(\alpha) = \begin{cases} 0 & \text{if } z_{\alpha/2} \geq \frac{|\hat{\mu}-\mu_0|}{\sigma_0-\hat{\sigma}}, \\ \frac{|\hat{\mu}-\mu_0|}{2z_{\alpha/2}\times\hat{\sigma}} - \frac{\sigma_0-\hat{\sigma}}{2\hat{\sigma}} & \text{if } \frac{|\hat{\mu}-\mu_0|}{\sigma_0+\hat{\sigma}} < z_{\alpha/2} < \frac{|\hat{\mu}-\mu_0|}{\sigma_0-\hat{\sigma}}, \\ 1 & \text{if } z_{\alpha/2} \leq \frac{|\hat{\mu}-\mu_0|}{\sigma_0+\hat{\sigma}}, \end{cases}$$

where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. Let $a = \frac{|\hat{\mu}-\mu_0|}{\sigma_0+\hat{\sigma}}, b = \frac{|\hat{\mu}-\mu_0|}{\sigma_0-\hat{\sigma}}$, then

$$\begin{aligned}\mathcal{D}\text{-AUC} &= \int_0^1 \mathcal{D}(\alpha)d\alpha \\ &= \frac{|\hat{\mu} - \mu_0|}{2\sqrt{2\pi}\hat{\sigma}} \left( E_1 \left( \frac{a^2}{2} \right) - E_1 \left( \frac{b^2}{2} \right) \right) - \frac{\sigma_0 - \hat{\sigma}}{\hat{\sigma}} (\Phi(b) - \Phi(a)) + 2\Phi(a) - 1,\end{aligned}$$

where $E_1$ is the exponential integral defined as $E_1(x) = \int_x^\infty \exp(-t)/t\,dt, x > 0$ (Abramowitz and Stegun, 1972).

The $p$-value of $\mathcal{D}$-AUC is

$$p_{\mathcal{D}\text{-AUC}} = \Pr(\mathcal{D}\text{-AUC}(Y) \geq \mathcal{D}\text{-AUC}(y)),$$

where $Y \sim N(\mu_0, \sigma^2 + \sigma_0^2)$ and $y$ is the observed data.

When comparing prior $N(\mu_0, \sigma_0^2)$ and likelihood $N(y, \sigma^2)$, denote the dissonance measures by $\tilde{\mathcal{D}}(\alpha)$. When $\sigma_0 > \sigma$, we can show:

$$\tilde{\mathcal{D}}(\alpha) = \begin{cases} 0 & \text{if } z_{\alpha/2} \geq \frac{|y-\mu_0|}{\sigma_0-\sigma}, \\ \frac{|y-\mu_0|}{2z_{\alpha/2}\times\sigma} - \frac{\sigma_0-\sigma}{2\sigma} & \text{if } \frac{|y-\mu_0|}{\sigma_0+\sigma} < z_{\alpha/2} < \frac{|y-\mu_0|}{\sigma_0-\sigma}, \\ 1 & \text{if } z_{\alpha/2} \leq \frac{|y-\mu_0|}{\sigma_0+\sigma}. \end{cases}$$

Let $c = \frac{|y-\mu_0|}{\sigma_0+\sigma}, d = \frac{|y-\mu_0|}{\sigma_0-\sigma}$,

$$\begin{aligned}\tilde{\mathcal{D}}\text{-AUC} &= \int_0^1 \tilde{\mathcal{D}}(\alpha)d\alpha \\ &= \frac{|y - \mu_0|}{2\sqrt{2\pi}\sigma} \left( E_1 \left( \frac{c^2}{2} \right) - E_1 \left( \frac{d^2}{2} \right) \right) - \frac{\sigma_0 - \sigma}{\sigma} (\Phi(d) - \Phi(c)) + 2\Phi(c) - 1.\end{aligned}$$

When $\sigma_0 < \sigma$, we can carry out the derivation similarly. Combining both cases, we have

$$\begin{aligned}\tilde{\mathcal{D}}\text{-AUC} &= \frac{|y - \mu_0|}{2\sqrt{2\pi}\min\{\sigma_0, \sigma\}} \left( E_1 \left( \frac{c^2}{2} \right) - E_1 \left( \frac{e^2}{2} \right) \right) \\ &\quad - \frac{|\sigma_0 - \sigma|}{\min\{\sigma_0, \sigma\}} (\Phi(e) - \Phi(c)) + 2\Phi(c) - 1,\end{aligned}$$

where $e = \frac{|y-\mu_0|}{|\sigma_0-\sigma|}$ (set $e = \infty$ when $\sigma_0 = \sigma$). Derivations of these expressions are given in Appendix D in Supplementary Material (Shi et al., 2021). Numerical comparison

among $\mathcal{D}$-AUC, $\tilde{\mathcal{D}}$-AUC, Kullback-Leibler divergence, and Hellinger distance is given in Appendix E in Supplementary Material (Shi et al., 2021).

Similarly, the $p$-value of $\tilde{\mathcal{D}}$-AUC is

$$p_{\tilde{\mathcal{D}}\text{-AUC}} = \Pr(\tilde{\mathcal{D}}\text{-AUC}(Y) \geq \tilde{\mathcal{D}}\text{-AUC}(y)),$$

with $Y \sim N(\mu_0, \sigma^2 + \sigma_0^2)$ and $y$ as the observed data.

The two $p$-value methods, $p_{\text{EM}}$ and $p_{\text{Presanis}}$, are identical in this example, so they are plotted using one $p$-value curve. Plotting $p$-value (EM, Presanis et al.), $p$-value of $\mathcal{D}$-AUC, and $p$-value of $\tilde{\mathcal{D}}$-AUC as a function of $y$ shows that three curves are identical for several choices of $\sigma$ and $\sigma_0$ (Figure 3 shows the result of $\sigma_0 = 1, \sigma = 1$). This suggests that our dissonance check is equivalent to that of Evans and Moshonov (2006) and Presanis et al. (2013) in this example.

As mentioned in Nott et al. (2020), if two discrepancy statistics $A(y)$ and $B(y)$ are related as a function of $y$ by a monotone transformation, denoted by $A(y) \doteq B(y)$, they will result in the same predictive $p$-values. Here

$$\mathcal{D}\text{-AUC}(y) \doteq |\hat{\mu} - \mu_0| = w|y - \mu_0| \doteq |y - \mu_0|,$$
$$\tilde{\mathcal{D}}\text{-AUC}(y) \doteq |y - \mu_0|.$$

For the conflict check of Evans and Moshonov (2006), $p(y) \doteq (y - \mu_0)^2 \doteq |y - \mu_0|$. In this case, all these checks compare $|y - \mu_0|$ to the distribution of $|Y - \mu_0|$ for $Y \sim N(\mu_0, \sigma^2 + \sigma_0^2)$ and therefore yield the same $p$-values.

## 4.2   Binomial Model

A binomial model has been applied to events consisting of two possible outcomes, for example, cured or not cured in clinical trial, success or failure in a designed experiment. Binomial models are common in disease monitoring, drug development, opinion poll, quality control, etc.

For $n$ independent trials, assume the probability of a success in each trial remains constant at $\theta$. Let $y$ denote the number of successes in $n$ independent Bernoulli trials. Assuming $n$ is known, consider $y \sim Bin(n, \theta)$, a binomial distribution with parameter $\theta$. For a Bayesian analysis, further assume the $\text{Beta}(a, b)$ distribution as the prior for $\theta$. Then the posterior for $\theta|y \sim \text{Beta}(a + y, b + n - y)$.

The prior predictive density for $y$ is beta-binomial,

$$p(y) = \binom{n}{y} \frac{B(a + y, b + n - y)}{B(a, b)}, y = 0, \ldots, n, \text{where } B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

The Bayesian conflict $p$-value of Evans and Moshonov (2006) is

$$p_{\text{EM}} = \Pr(p(Y) \leq p(y)) = \sum_{Y:p(Y) \leq p(y)} \binom{n}{Y} \frac{B(a + Y, b + n - Y)}{B(a, b)}.$$
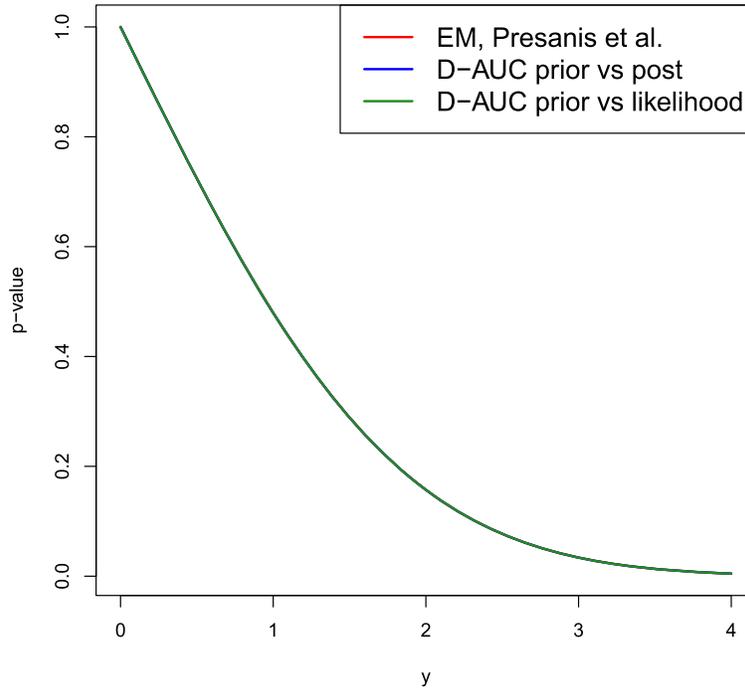
Figure 3: Comparison among Bayesian conflict $p$-values, $p$-value of $\mathcal{D}$-AUC, $p$-value of $\tilde{\mathcal{D}}$-AUC for the Normal Location Model, $\sigma_0 = 1, \sigma = 1$.

For Presanis et al. (2013), we consider:

$$\text{data-translated likelihood: } \theta_a \sim \text{Beta}(1 + y, 1 + n - y),$$
$$\text{prior: } \theta_b \sim \text{Beta}(a, b),$$
$$\text{let } \delta = \theta_a - \theta_b, \text{ then } p_{\text{Presanis}} = \Pr(p_\delta(\delta) < p_\delta(0)) = 2 \times \min\{\Pr(\delta > 0), 1 - \Pr(\delta > 0)\},$$

assuming that the distribution of $\delta$ is roughly symmetric and unimodal. In this case, $p_{\text{Presanis}}$ is calculated using MCMC samples by counting the number of times $\theta_a > \theta_b$.

When comparing prior $\text{Beta}(a, b)$ versus posterior $\text{Beta}(a + y, b + n - y)$, the $p$-value of $\mathcal{D}$-AUC in this case is $p_{\mathcal{D}\text{-AUC}} = \Pr(\mathcal{D}\text{-AUC}(Y) \geq \mathcal{D}\text{-AUC}(y))$, where $Y$ follows the prior predictive distribution with pmf $p(y)$ and $y$ is the observed data.

Similarly, when comparing prior $\text{Beta}(a, b)$ and likelihood $\text{Beta}(1 + y, 1 + n - y)$, the $p$-value of $\tilde{\mathcal{D}}$-AUC is $p_{\tilde{\mathcal{D}}\text{-AUC}} = \Pr(\tilde{\mathcal{D}}\text{-AUC}(Y) \geq \tilde{\mathcal{D}}\text{-AUC}(y))$, with $Y$ from the prior predictive distribution and $y$ as the observed data.

Assuming $n = 10$, we have plotted the above 4 $p$-values ($p_{\text{EM}}, p_{\text{Presanis}}, p_{\mathcal{D}\text{-AUC}}$ and $p_{\tilde{\mathcal{D}}\text{-AUC}}$) for each $y$ and 2 prior choices in Figure 4. For most $y$ values, the $p$-value of $\mathcal{D}$-AUC is the same as the $p$-value of $\tilde{\mathcal{D}}$-AUC and is close to either $p_{\text{EM}}$ or $p_{\text{Presanis}}$.

(a) $n = 10, a = 5, b = 15$                         (b) $n = 10, a = 10, b = 10$
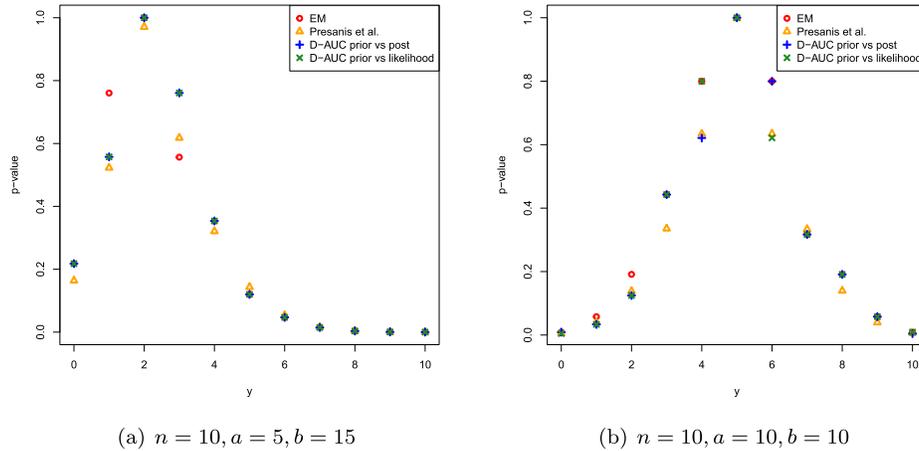
Figure 4: Comparison among Bayesian conflict $p$-values, $p$-value of $\mathcal{D}$-AUC, $p$-value of $\tilde{\mathcal{D}}$-AUC for the Binomial Model.

# 5    Numerical Results from Comparing Two Distributions

To better understand how the two measures behave, we design the following experiments to explore various scenarios between the two distributions, which are inspired by Section 3.3 of Holmes et al. (2015). Let $y_i$ denote a sample from each distribution, $i = 1, 2$. Then we compare the two distributions in each of the following scenarios.

1. Mean shift: $y_1 \sim N(0,1), y_2 \sim N(\theta,1), \theta = 0, 0.1, 0.2, \ldots, 4$.

2. Variance shift: $y_1 \sim N(0,1), y_2 \sim N(0,\theta^2), \theta = 1, 1.1, 1.2, \ldots, 3$.

3. Tails: $y_1 \sim N(0,1), y_2 \sim t_\theta, \theta = 10^3, 10^2, 10, 5, 2, 1$.

4. Mixture: $y_1 \sim N(0,1), y_2 \sim \frac{1}{2}N(\theta,1) + \frac{1}{2}N(-\theta,1), \theta = 0, 0.1, 0.2, \ldots, 3$.

5. Skewed: $y_1 \sim Gamma(1,1), y_2 \sim Gamma(\theta,1), \theta = 1, 1.1, 1.2, \ldots, 4$.

The results of the 5 scenarios are shown in Figure 5, with $LCR$ plotted in the left column and $\mathcal{D}$-AUC plotted in the right column.

With scenario 1 of mean shift, there is no scale change, so $LCR(\alpha) = 0, \forall \alpha$ and we focus instead on the dissonance measure, plotting $\mathcal{D}$-AUC as a function of $\theta$. We observe that $\mathcal{D}$-AUC increases as $\theta$ increases.

For scenario 2, there is no location change, so $\mathcal{D}$-AUC $= 0$ and thus we only plot the $LCR(\alpha)$ as a function of $\theta$ for $\alpha = 0.05$. We see that $LCR$ values are negative and become more negative with increased $\theta$.
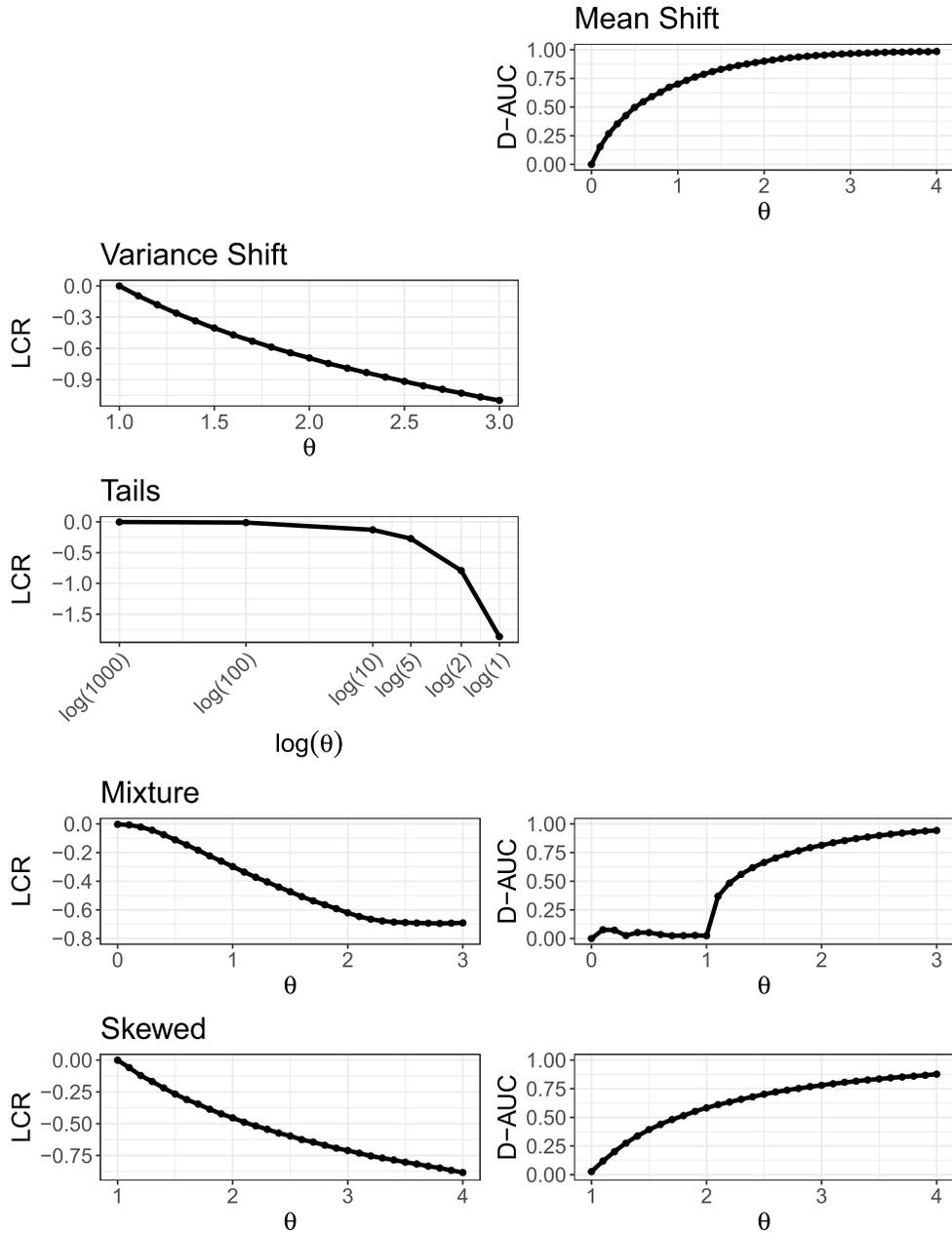
Figure 5: Results on $LCR(\alpha)$ with $\alpha = 0.05$ and $\mathcal{D}$-AUC with 5 scenarios for comparing two distributions.

For scenario 3, $\mathcal{D}$-AUC $= 0$, so we report only $LCR(\alpha)$ with $\alpha = 0.05$ as a function of $\theta$. As expected, values become more negative because the distribution for $y_2$ becomes more dispersed than that for $y_1$ when the degree of freedom $\theta$ decreases.

For scenarios 4 and 5, we plot both $LCR(\alpha)$ with $\alpha = 0.05$, and $\mathcal{D}$-AUC as a function of $\theta$. All $LCR$ values are negative as expected because the distribution for $y_2$ is more diffuse than $y_1$. For the mixture scenario, $\mathcal{D}$-AUC remains low when $\theta \leq 1$ and increases as a function of $\theta$ afterwards. This is because, when $\theta$ is small, the mixture distribution is still unimodal. Once $\theta > 1$, the distribution for $y_2$ becomes bimodal in this case. For the skewed scenario, $\mathcal{D}$-AUC increases as a function of $\theta$ as expected.

**Remark 4.** *From Figure 5 the mean shift scenario, we see $\mathcal{D}\text{-}AUC = 0.5, 0.7, 0.9, 0.97$ for $\theta = 0.5, 1, 2, 3$. Therefore, we can categorize $\mathcal{D}\text{-}AUC < 0.5$ as small dissonance, $0.5 \leq \mathcal{D}\text{-}AUC < 0.7$ as moderate dissonance, $0.7 \leq \mathcal{D}\text{-}AUC < 0.9$ as large dissonance, $0.9 \leq \mathcal{D}\text{-}AUC < 0.97$ as extra large dissonance, and $0.97 \leq \mathcal{D}\text{-}AUC$ as extremely large dissonance.*

# 6 Real Data Applications

## 6.1 Analysis of Benchmark Dose Data in Toxicology

In this study, we select two data sets from similar studies and would like to answer the questions: (1) are the two data sets compatible? and (2) which of these two data sets is more concentrated?

The benchmark approach is a useful tool in toxicology. The benchmark dose is the dose of an environmental toxicant that produces a predetermined change in response compared with the background response level. The toxicological data contain $n$ binomial responses $\boldsymbol{y} = (y_1, \ldots, y_n)$ with $y_i \sim Bin(n_i, p_i)$, where $n_i$ is the number of animals tested at dose level $x_i$ and $p_i$ is the probability that an animal gives an adverse response at dose level $x_i$. Using the logistic regression model, we have

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, i = 1, \ldots, n.$$

The Kociba study (Kociba et al., 1978) is a lifetime feeding study of both female and male Sprague Dawley rats, with 50 rats tested in each group at doses of 0, 1, 10, and 100 ng/kg/day. Inferences derived from the Kociba study have been widely used as the basis for risk assessments for 2,3,7,8-tetrachlorodibenzodioxin (TCDD). The National Toxicology Program (NTP) study (National Toxicology Program, 1982) is a study in which groups of 50 male rats, 50 female rats, and 50 male mice received TCDD as a suspension in 9:1 corn oil: acetone by gavage twice each week to achieve doses of 0, 10, 50, or 500 ng/kg/week for two years. These exposures correspond to daily averaged doses of 0, 1.4, 7.1, or 71 ng/kg/day for rats. Liver tumor (neoplastic nodule) incidences of female rats from both studies, shown in Table 1, are chosen as the data for this analysis. Posterior mean and posterior standard deviation (SD) for $\beta_0$ and $\beta_1$ for each study are also given.

| Kociba | | | NTP | | |
|---|---|---|---|---|---|
| $y$ | $n$ | $x$ | $y$ | $n$ | $x$ |
| 9 | 86 | 0 | 5 | 75 | 0 |
| 3 | 50 | 1 | 1 | 49 | 1.4 |
| 18 | 50 | 10 | 3 | 50 | 7.1 |
| 34 | 48 | 100 | 12 | 49 | 71 |
| $\beta_0$ Mean(SD) $= -1.785(0.210)$ | | | $\beta_0$ Mean(SD) $= -3.030(0.366)$ | | |
| $\beta_1$ Mean(SD) $= 0.028(0.004)$ | | | $\beta_1$ Mean(SD) $= 0.026(0.007)$ | | |

Table 1: Benchmark dose data summary.

Denote the Kociba study data by $S_1 = (\boldsymbol{n_1}, \boldsymbol{y_1}, \boldsymbol{x_1})$ and the NTP study data by $S_2 = (\boldsymbol{n_2}, \boldsymbol{y_2}, \boldsymbol{x_2})$. Shao and Small (2011) showed that while $S_1$ and $S_2$ are not compatible in terms of all parameters $\boldsymbol{\theta} = (\beta_0, \beta_1)^\top$, they are compatible in terms of one common parameter $\beta_1$. Our measure $\mathcal{D}(\alpha)$ confirms their conclusion. The settings for prior distributions, likelihood functions, and corresponding posterior distributions are given as follows:

$$\pi(\beta_0) \sim N(0, 10000),$$
$$\pi(\beta_1) \sim N(0, 10000),$$
$$L(\beta_0, \beta_1 \mid S_1) = \prod_{i=1}^n \left[ \frac{\exp(\beta_0 + \beta_1 x_{1i})}{1 + \exp(\beta_0 + \beta_1 x_{1i})} \right]^{y_{1i}} \left[ \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{1i})} \right]^{n_{1i} - y_{1i}},$$
$$L(\beta_0, \beta_1 \mid S_2) = \prod_{i=1}^n \left[ \frac{\exp(\beta_0 + \beta_1 x_{2i})}{1 + \exp(\beta_0 + \beta_1 x_{2i})} \right]^{y_{2i}} \left[ \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{2i})} \right]^{n_{2i} - y_{2i}},$$
$$\pi(\beta_0, \beta_1 \mid S_1) \propto \pi(\beta_0)\pi(\beta_1)L(\beta_0, \beta_1 \mid S_1),$$
$$\pi(\beta_0, \beta_1 \mid S_2) \propto \pi(\beta_0)\pi(\beta_1)L(\beta_0, \beta_1 \mid S_2).$$

The estimated marginal posterior densities of $\beta_0$ based on the Kociba data and the NTP data, respectively, have different locations, inducing almost complete dissonance in $\beta_0$ and thus are not compatible (Figure 6(a)). The estimated marginal posterior densities of $\beta_1$ for each data set share a similar location leading to small $\mathcal{D}$-AUC values for $\beta_1$ and thus are compatible with each other (Figure 6(b)). Because the posterior distributions for the two parts are largely non-overlapping, there is almost complete dissonance for $(\beta_0, \beta_1)$ jointly (Figure 6(c)).

**Remark 5.** *The priors $\pi(\beta_0)$ and $\pi(\beta_1)$ are specified as normal distributions and thus are log-concave. The likelihood functions of the logistic regression $L(\beta_0, \beta_1 \mid S_1)$ and $L(\beta_0, \beta_1 \mid S_2)$ are log-concave. Log-concavity is preserved under multiplication, i.e., the product of log-concave functions is also log-concave. Therefore, the posterior distributions $\pi(\beta_0, \beta_1 \mid S_1)$ and $\pi(\beta_0, \beta_1 \mid S_2)$ are both log-concave and HDRs corresponding to each distribution are both closed convex sets according to Theorem 2.1.*

To compare these two data sets, we estimated the HDR $R_1(\alpha)$ based on the posterior distribution $\pi(\beta_0, \beta_1 \mid S_1)$ and the HDR $R_2(\alpha)$ based on the posterior distribution
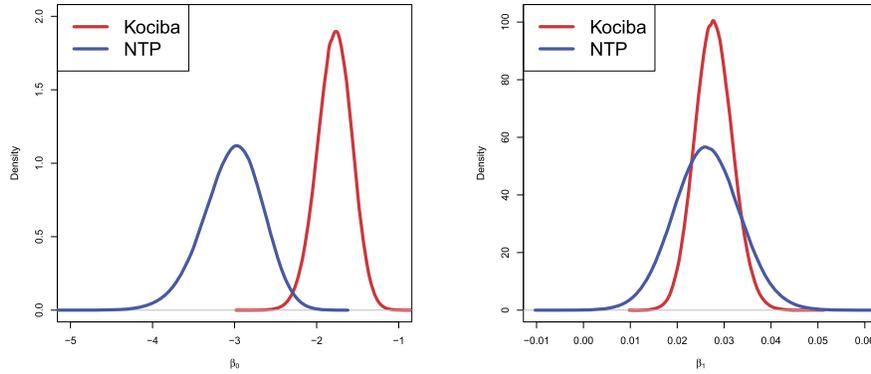
(a) Marginal Density Plot of $\beta_0$

(b) Marginal Density Plot of $\beta_1$
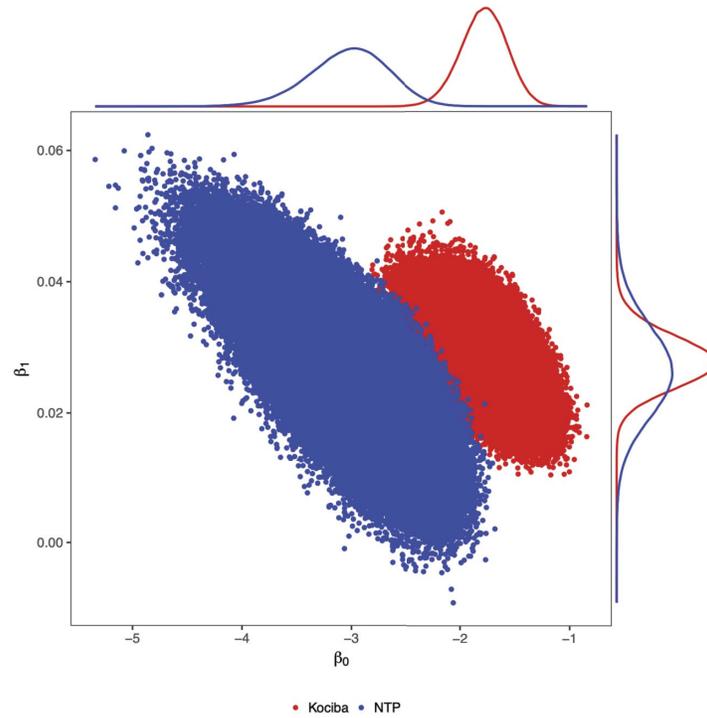
(c) MCMC samples

Figure 6: Posterior distributions of $(\beta_0, \beta_1)$ for Kociba and NTP data.

$\pi(\beta_0, \beta_1 \mid S_2)$. We calculated $LCR(\alpha)$ with respect to $\beta_0$, $\beta_1$ and $(\beta_0, \beta_1)$ for several different level $\alpha$ values (Table 2). Over five choices of $\alpha$ values, $LCR(\alpha)$ is quite stable, indicating that $LCR(\alpha)$ is not sensitive to the choice of $\alpha$. The results show that the

| $\alpha$ | $\beta_0$ | $\beta_1$ | $(\beta_0, \beta_1)$ |
|------|---------|---------|------------------|
| 0.05 | $-0.544$ | $-0.580$ | $-0.920$ |
| 0.25 | $-0.537$ | $-0.576$ | $-0.906$ |
| 0.50 | $-0.532$ | $-0.573$ | $-0.903$ |
| 0.75 | $-0.530$ | $-0.572$ | $-0.910$ |
| 0.95 | $-0.530$ | $-0.575$ | $-0.957$ |

Table 2: $LCR$ of Kociba data to NTP data.



(a) $\beta_0$        (b) $\beta_1$        (c) $(\beta_0, \beta_1)$
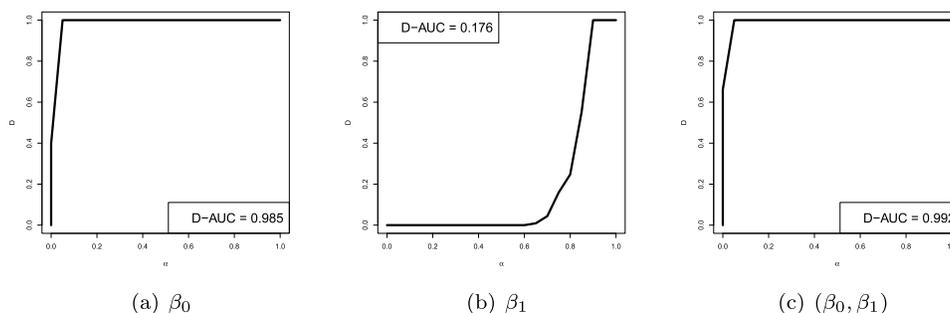
Figure 7: Plot of $\mathcal{D}(\alpha)$ versus $\alpha$ for Kociba to NTP comparison.

NTP data is less concentrated than the Kociba data marginally for each parameter and also jointly for both parameters given that $LCR(\alpha)$ is negative in all three cases. Plotting $\mathcal{D}(\alpha)$ versus $\alpha$ and summarizing the results using $\mathcal{D}$-AUC (Figure 7) shows almost complete dissonance for $\beta_0$ and jointly for $(\beta_0, \beta_1)$, as 0.985 and 0.992 are quite close to 1, and little dissonance for $\beta_1$, as expected.

## 6.2 Analysis of Six Cities Data

In this subsection, we illustrate how our dissonance measure can be useful in the missing data problem and would like to know if the data with no missing values are adequate and how much information is contributed by handling the missing data.

The six cities longitudinal study of the health effects of respiratory function in children (Ware et al., 1984) is a well known environmental data set that has been analyzed extensively in the literature. We only select two cities from this data set as an example. The binary response is the wheezing status of a child at age 11 with $y = 0$ representing no wheezing and $y = 1$ for wheezing. The wheezing status is modeled as a function of the city of residence $(x_1)$ and the smoking status of the mother $(x_2)$. The city of residence $x_1$ is a binary covariate that equals 1 if the child lived in Kingston–Harriman, Tennessee, the more polluted city, and 0 if the child lived in Portage, Wisconsin. The covariate $x_2$ is maternal cigarette smoking measured by number of cigarettes per day. There are 2315 subjects in the data set. The covariate $x_1$ is missing for 32.8% of the cases. We present a brief summary of the data in Table 3. Details of the data set can be found in the original paper.

| $y$ | $x_1$ | $x_2$ | $n$ | $y$ | $x_1$ | $x_2$ | $n$ |
|-----|-------|-------|-----|-----|-------|-------|-----|
| 0 | 0 | 0 | 418 | 1 | 0 | 0 | 127 |
| 0 | 1 | 0 | 323 | 1 | 1 | 0 | 106 |
| 0 | 0 | $\geq 1$ | 226 | 1 | 0 | $\geq 1$ | 72 |
| 0 | 1 | $\geq 1$ | 201 | 1 | 1 | $\geq 1$ | 83 |
| 0 | NA | 0 | 369 | 1 | NA | 0 | 86 |
| 0 | NA | $\geq 1$ | 229 | 1 | NA | $\geq 1$ | 75 |

$y$: wheezing status. 0: no wheezing; 1: wheezing.

$x_1$: city of residence. 0: clean city; 1: polluted city; NA: missing.

$x_2$: maternal smoking (number of cigarettes per day).

Table 3: Summary of Two Cities data.

We use a logistic regression model for $y$ given $x_1$ and $x_2$. i.e.,

$$P(y_i = 1 \mid x_{1i}, x_{2i}, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$. We further model $x_1$ given $x_2$ by a logistic regression to handle the missing data problem. Specifically, we assume $x_{1i}$ given $x_{2i}$ are independent Bernoulli variables with probability

$$P(x_{1i} = 1 \mid x_{2i}, \boldsymbol{\alpha}) = \frac{\exp(\alpha_0 + \alpha_1 x_{2i})}{1 + \exp(\alpha_0 + \alpha_1 x_{2i})}, i = 1, \ldots, n,$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$. The proposed joint prior distribution for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \pi(\boldsymbol{\alpha} \mid c_0)\pi(\boldsymbol{\beta} \mid d_0),$$

where $\boldsymbol{\alpha} \mid c_0 \sim N_2(0, c_0 I_2)$ and $\boldsymbol{\beta} \mid d_0 \sim N_3(0, d_0 I_3)$.

Let $r_i$ be the missing indicator with $r_i = 1$ if $x_{1i}$ is missing and $r_i = 0$ otherwise. Let $D_{cc}$ (complete case) denote the subset of the data where both $x_1$ and $x_2$ are observed with no missing values ($r_i = 0$). Then the joint likelihood function based on the complete case data subset $D_{cc}$ is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{cc}) = \prod_{i:r_i=0} f(y_i \mid x_{1i}, x_{2i}, \boldsymbol{\beta})f(x_{1i} \mid x_{2i}, \boldsymbol{\alpha}),$$

where

$$f(y_i \mid x_{1i}, x_{2i}, \boldsymbol{\beta}) = \frac{\exp[y_i(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})]}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})},$$

$$f(x_{1i} \mid x_{2i}, \boldsymbol{\alpha}) = \frac{\exp[x_{1i}(\alpha_0 + \alpha_1 x_{2i})]}{1 + \exp(\alpha_0 + \alpha_1 x_{2i})}.$$

Let $D_{ac}$ denote the whole data set with the subscript ac meaning all case. The joint likelihood function based on all data $D_{ac}$ is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{ac}) = L(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{cc})L_1(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{ac} \backslash D_{cc}),$$

where

$$L_1(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{ac} \backslash D_{cc}) = \prod_{i:r_i=1} [f(y_i \mid x_{1i} = 1, x_{2i}, \boldsymbol{\beta}) f(x_{1i} = 1 \mid x_{2i}, \boldsymbol{\alpha})$$
$$+ f(y_i \mid x_{1i} = 0, x_{2i}, \boldsymbol{\beta}) f(x_{1i} = 0 \mid x_{2i}, \boldsymbol{\alpha})].$$

That is, for the missing data part, we marginalize over $x_{1i}$.

The posterior distributions of the complete case $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{cc})$ and all case $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{ac})$ are defined as:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{cc}) \propto L(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{cc})\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}),$$
$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{ac}) \propto L(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid D_{ac})\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

| | $D_{cc}$ | | | $D_{ac}$ | |
|---|---|---|---|---|---|
| | Mean | SD | | Mean | SD |
| $\beta_0$ | $-1.252$ | 0.089 | $\beta_0$ | $-1.337$ | 0.081 |
| $\beta_1$ | 0.145 | 0.118 | $\beta_1$ | 0.145 | 0.118 |
| $\beta_2$ | 0.011 | 0.005 | $\beta_2$ | 0.013 | 0.004 |

Table 4: Posterior Mean and Standard Deviation of Parameters $\boldsymbol{\beta}$ for Two Cities Data.

Using $c_0 = d_0 = 100$, a MCMC sample of size 1,000,000 was generated from each posterior distribution. As expected, for $\alpha = 0.05$, comparing $D_{cc}$ versus $D_{ac}$, $LCR(\alpha)$ for $\beta_0$ and $\beta_2$ are 0.099 and 0.195 indicating that the marginal posterior distributions for $\beta_0$ and $\beta_2$ become more concentrated with the addition of data, while $LCR(\alpha)$ for $\beta_1$ is $-0.001$ (nearly zero), which is expected given that the additional observations are completely missing for the covariate $x_1$. Marginal posterior densities of $\beta_0$ and $\beta_2$ show some dissonance between the complete case and the all case, whereas marginal posteriors for $\beta_1$ are nearly identical (Figure 8). Posterior mean and standard deviation for each parameter are reported for the $D_{cc}$ and $D_{ac}$ respectively in Table 4. Because the modes of $\beta_1$ from the complete case and the all case are almost identical, $\mathcal{D}$-AUC is very close to 0 (0.06) (Figure 9). The modes of $\beta_2$ from the two cases are still quite similar, with $\mathcal{D}$-AUC = 0.383. For $\beta_0$, there is a location shift between the complete case and the all case. The $\mathcal{D}$-AUC for $\beta_0$ is 0.686. The $\mathcal{D}$-AUC for $(\beta_0, \beta_1, \beta_2)$ jointly is 0.566. The dissonance measure shows that adding 32.8% missing data to the complete case does not affect $\beta_1$, affects $\beta_2$ slightly, affects $\beta_0$ moderately, and has moderate effect on $(\beta_0, \beta_1, \beta_2)$ jointly.

## 6.3 Analysis of Pediatric Cancer Data

Conducting pediatric clinical trials as part of drug development for children and adolescents can be difficult due to the often low incidence of the disease, making accrual slow or infeasible. In addition, low morbidity in this population makes it impractical to achieve adequate power. In this case, pediatric care providers are accustomed to relying on evidence from adult studies, so it is natural to consider leveraging information from adult trials. Our measures $LCR(\alpha)$ and $\mathcal{D}(\alpha)$ can be very useful in deciding
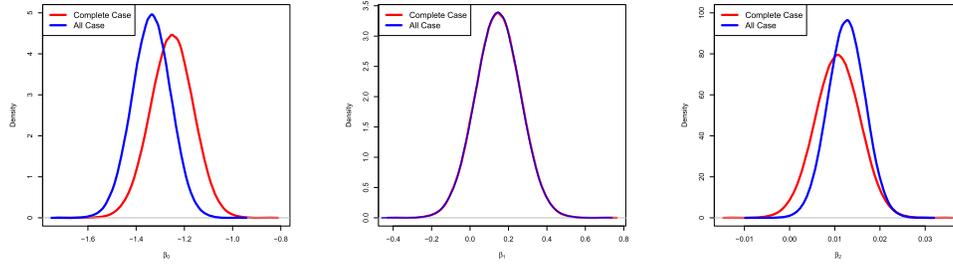
(a) Marginal Density Plot of $\beta_0$   (b) Marginal Density Plot of $\beta_1$   (c) Marginal Density Plot of $\beta_2$

Figure 8: Posterior distributions of $\beta_0$, $\beta_1$ and $\beta_2$ from the complete case and the all case for Two Cities Data.



(a) $\beta_0$                              (b) $\beta_1$



(c) $\beta_2$                          (d) $(\beta_0, \beta_1, \beta_2)$
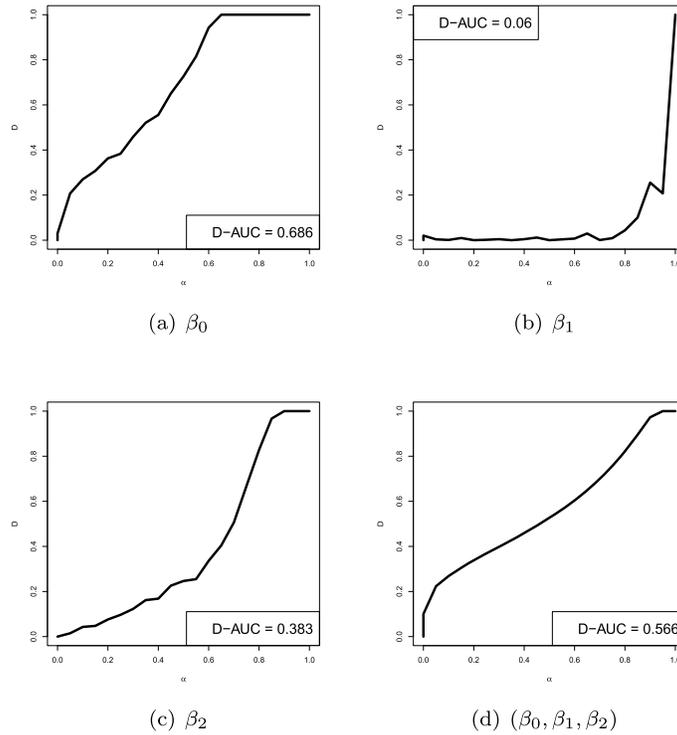
Figure 9: Plot of $\mathcal{D}(\alpha)$ versus $\alpha$ for the complete case compared to the all case for the Two Cities Data.

the extent we should borrow. Here we illustrate this using the pediatric cancer study of Philadelphia chromosome positive chronic myeloid leukemia (Ph+CML). We have adult and pediatric clinical trials for Tasigma (nilotinib) for newly diagnosed Ph+CML in

chronic phase. The primary efficacy endpoint was major molecular response (MMR) at 12 months after the start of study medication. MMR was defined as less than or equal to 0.1% BCR-ABL/ABL % (BCR-ABL: breakpoint cluster region - Abelson murine leukemia gene, ABL: Abelson murine leukemia gene) by international scale measured by real-time quantitative polymerase chain reaction (RQ-PCR), which corresponds to a greater than or equal to 3 log reduction of BCR-ABL transcript from standardized baseline. A summary of two trials is provided in Table 5. Other detailed information can be found in https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/022068s027lbl.pdf.

|             | Adult              | Pediatric          |
|-------------|--------------------|--------------------|
| Endpoint    | MMR at 12 month    | MMR at 12 month    |
| Design      | Randomized         | Single-arm         |
| Sample Size | 282                | 25                 |
| Age         | 18+                | 2-18               |
| Results     | $125/282 = 44\%$   | $15/25 = 60\%$     |

Table 5: Summary of Adult and Pediatric Clinical Trials.

The adult data are denoted $D_A$, where the sample size $n_A$ is 282, and the number of subjects having MMR $y_A$ is 125. Similarly, the pediatric data are denoted $D_P$, with $n_P = 25$ and $y_P = 15$. We use a binomial distribution to model $y_A$ and $y_P$: $y_A \sim Bin(n_A, p)$ and $y_P \sim Bin(n_P, p)$, where $p$ is the efficacy for each population. To determine the extent of information we could borrow from the adult data, we used the power prior method (Ibrahim et al., 2015). Letting $\pi(p)$ denote the initial prior distribution for $p$, the power prior distribution of $p$ with the weight $a_0$ for the pediatric cancer study is defined as

$$\pi(p \mid D_A, a_0) \propto \pi(p)[L(p \mid D_A)]^{a_0},$$

where $a_0$ controls the weight of the adult data relative to the pediatric data. Our settings can be summarized as follows:

$$\pi(p) \propto p^{-1}(1-p)^{-1},$$
$$\pi(p \mid D_A, a_0) \propto \pi(p)[L(p \mid D_A)]^{a_0},$$
$$\pi(p \mid D_A, D_P, a_0) \propto \pi(p \mid D_A, a_0)L(p \mid D_P)$$
$$\propto \pi(p)[L(p \mid D_A)]^{a_0}L(p \mid D_P)$$
$$\sim \text{Beta}(a_0 y_A + y_P, a_0(n_A - y_A) + (n_P - y_P)).$$

The range of $a_0$ is restricted between 0 and 1, with $a_0 = 0$ meaning no incorporation of the adult data. We compare the posterior distributions $\pi(p \mid D_A, D_P, a_0)$ to $\pi(p \mid D_A, D_P, a_0 = 0)$, i.e., borrowing versus no borrowing (pediatric data only), through various choices of value $a_0$. Ideally, if two distributions are nearly identical, we gain information from incorporating the adult data with certain weight $a_0$. In this case, our measure $\mathcal{D}(\alpha)$ helps to quantify the extent of discrepancy between the two distributions and $LCR(\alpha)$ can measure how much information to gain by borrowing from the adult data versus not borrowing at all.

(a) Density Plot with $a_0 = 0.01$     (b) Density Plot with $a_0 = 0.04$
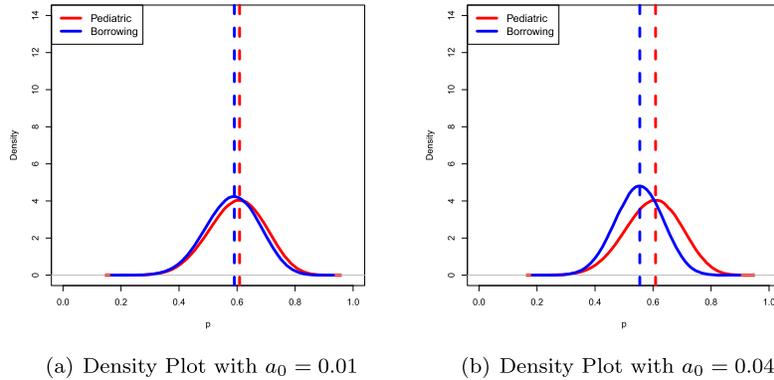
Figure 10: Posterior distributions of Pediatric Data Only versus Borrowing.



(a) Plot of $\mathcal{D}(\alpha)$ versus $\alpha$ with $a_0 = 0.01$ (b) Plot of $\mathcal{D}(\alpha)$ versus $\alpha$ with $a_0 = 0.04$
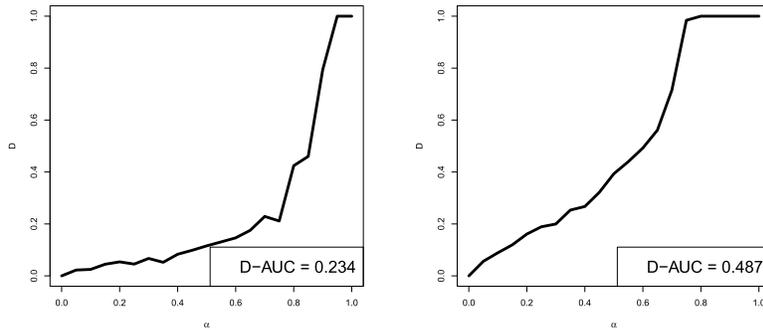
Figure 11: Plot of $\mathcal{D}(\alpha)$ versus $\alpha$ for borrowing versus pediatric data only.

We plot the posterior densities of $p$ with $a_0 = 0.01$ or $a_0 = 0.04$ (borrowing from the adult data) versus $a_0 = 0$ (no borrowing) (Figure 10). For each value $a_0$, we calculate $\mathcal{D}(\alpha)$ and summarize the result using $\mathcal{D}$-AUC. With weight $a_0$ increasing, we gain more information as the variance becomes smaller; however, the location shift also becomes more severe (Figure 11). As the borrowing power $a_0$ increases, the difference of two modes increases as well since the borrowing distribution is pulled more towards the adult side (Figure 12(a)). As a result, $\mathcal{D}$-AUC also increases.

As the weight $a_0$ increases, both effective sample size and $LCR$ increase due to more borrowing from the adult study, which has a larger sample size and smaller variance (Figure 12(b)). The relationship between $LCR$ and the effective sample size is logarithmic. Both distributions are roughly symmetric (Figure 10), which leads to $LCR$ being almost equivalent to the log ratio of the standard deviations of these two distributions. The effective sample size is proportional to the ratio of variances of these two distributions, thus we see a logarithmic trend in the plot. These two plots work together to
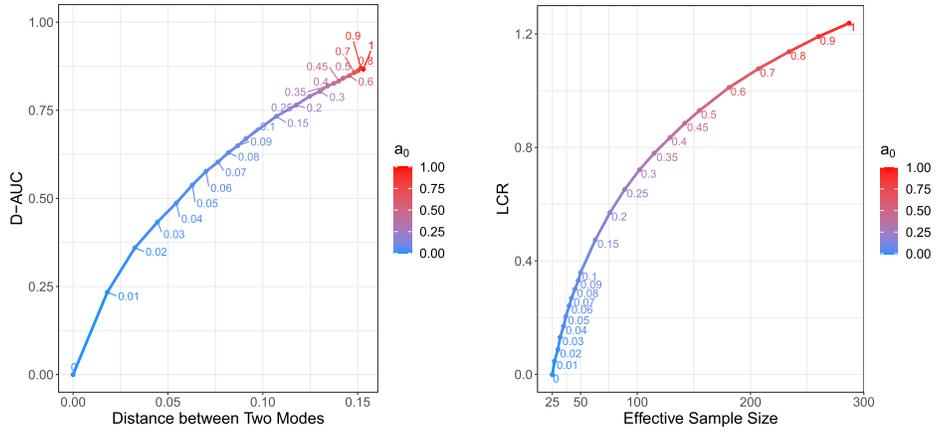
(a) Plot of $\mathcal{D}$-AUC versus distance between two modes

(b) Plot of $LCR$ versus effective sample size

Figure 12: Plots of $\mathcal{D}$-AUC and $LCR$ over different choices of $a_0$.

determine the weight $a_0$ for borrowing from the adult study such that the borrowing distribution is similar to the pediatric study and we can gain information from incorporating the adult data. For instance, we can choose weight $a_0$ to be 0.04, with $\mathcal{D}$-AUC $= 0.487$ slightly less than 0.5, $LCR = 0.170$ and effective sample size is 35. In this way, the extent of difference is acceptable, and we gain $100 \times (\exp(LCR) - 1) = 19\%$ more information by borrowing the amount of information equivalent to 10 subjects from the adult study compared to using pediatric data only.

# 7 Extension of Concentration Ratio and Dissonance Using Invariant HDR

## 7.1 Invariant HDR

We note that the use of HDRs makes our proposed measures not invariant to reparametrization. Consider for example $y \sim Bin(1, \theta)$ and $\theta \sim U(0, 1)$, the posterior HDR for $\theta$ is $\text{HDR}_\theta(\alpha) = \{\theta : \theta^y(1-\theta)^{1-y} \geq f_\alpha\}$. For a new parameterization $\gamma = \text{logit}(\theta)$, the posterior HDR for $\gamma$ is $\text{HDR}_\gamma(\alpha) = \{\gamma : e^{\gamma(y+1)}/(1 + e^\gamma)^3 \geq f_\alpha\}$. When $y = 1$, the posterior HDR for $\theta$ is one-sided while the posterior HDR for $\gamma$ is two-sided and thus, $\text{HDR}_\gamma(\alpha) \neq \text{logit}(\text{HDR}_\theta(\alpha))$. It means that with data unchanged, our measures may lead to different conclusions under different parameterizations, which is undesirable. So we explore an invariant version of the HDR proposed in Druilhet and Marin (2007).

They argued that the lack of invariance for HDRs was mainly due to the choice of the Lebesgue measure as dominating measure and thus they chose the Jeffreys measure instead, which makes HDRs invariant under 1-1 smooth reparametrization.

Let $\lambda$ denote the Lebesgue measure. Let $\pi_\lambda(\theta)$ and $\pi_\lambda(\theta|x) \propto f(x|\theta)\pi_\lambda(\theta)$ denote the prior and posterior density of $\theta$ with respect to $\lambda$. The Jeffreys measure for $\theta$, denoted by $J_\theta$, is the measure with density $|I(\theta)|^{1/2}$ with respect to $\lambda$, where $I(\theta)$ represents the Fisher information. Assuming that the Fisher information $I(\theta)$ is well defined and positive definite for every $\theta \in \Theta$, the $100(1-\alpha)\%$ HDR for $\theta$ with the Jeffreys measure as dominating measure (JHDR) is defined by

$$\mathrm{JHDR}_\theta(\alpha) = \{\theta : f(x|\theta)|I(\theta)|^{-\frac{1}{2}}\pi_\lambda(\theta) \geq k_\alpha\},$$

where $k_\alpha$ is the largest constant such that $\Pr(\theta \in \mathrm{JHDR}_\theta(\alpha)) \geq 1 - \alpha$.

Note that, if $I(\theta) \propto 1$, the JHDR is the same as the posterior HDR. Recall the normal location example in Section 4.1 where $y \sim N(\mu, \sigma^2)$ with known variance and $\mu \sim N(\mu_0, \sigma_0^2)$, we have $I(\mu) \propto 1$ and thus the posterior HDR for $\mu$ is the JHDR.

As our measures involve calculating the volume of a JHDR, we provide an important property of the volume of a JHDR in the following theorem.

**Theorem 7.1.** *The volume of a JHDR is invariant under differentiable reparametrization.*

*Proof.* Consider a differentiable reparametrization $\gamma = h(\theta)$ and denote the Jeffreys measure for $\gamma$ by $J_\gamma$,

$$V_{J_\theta}(\mathrm{JHDR}_\theta(\alpha)) = \int_{\mathrm{JHDR}_\theta(\alpha)} dJ_\theta = \int \mathbf{1}(\theta \in \mathrm{JHDR}_\theta(\alpha))|I(\theta)|^{\frac{1}{2}}d\theta,$$

$$V_{J_\gamma}(\mathrm{JHDR}_\gamma(\alpha)) = \int_{\mathrm{JHDR}_\gamma(\alpha)} dJ_\gamma = \int \mathbf{1}(\gamma \in \mathrm{JHDR}_\gamma(\alpha))|I(\gamma)|^{\frac{1}{2}}d\gamma$$

$$= \int \mathbf{1}(h(\theta) \in h(\mathrm{JHDR}_\theta(\alpha)))|I_\gamma(h(\theta))|^{\frac{1}{2}}h'(\theta)d\theta$$

$$= \int \mathbf{1}(\theta \in \mathrm{JHDR}_\theta(\alpha))|I(\theta) \times (h'(\theta))^{-2}|^{\frac{1}{2}}h'(\theta)d\theta$$

$$= \int \mathbf{1}(\theta \in \mathrm{JHDR}_\theta(\alpha))|I(\theta)|^{\frac{1}{2}}d\theta$$

$$= V_{J_\theta}(\mathrm{JHDR}_\theta(\alpha)). \qquad \square$$

## 7.2 Calculating JHDR Volume

We can tailor Algorithm 1 to calculate the volume of a JHDR, with detailed steps given in Algorithm 3.

When it is easy to generate a sample from the Jeffreys prior $\pi(\theta) \propto |I(\theta)|^{1/2}$ and we know the normalizing constant $m = \int_\Theta |I(\theta)|^{1/2}d\theta$, Step 3 in Algorithm 3 can be replaced by drawing a sample $\{z_i, i = 1, \ldots, N\}$ from the Jeffreys prior distribution and the estimated JHDR volume is $\hat{V}_{J_\theta} = \#\{i : f(x|z_i)|I(z_i)|^{-\frac{1}{2}}\pi_\lambda(z_i) \geq \hat{k}_\alpha\} \times m/N$.

---

**Algorithm 3** Computing the volume of a JHDR.

---

1. Draw a MCMC sample $\{\theta_t, t = 1, \ldots, T\}$ from the posterior distribution $\pi_\lambda(\theta|x)$.
2. Calculate $k_t = f(x|\theta_t)|I(\theta_t)|^{-\frac{1}{2}}\pi_\lambda(\theta_t)$ for $t = 1, \ldots, T$. Choose $\hat{k}_\alpha = k_{(j)}$, where $k_{(j)}$ is the $j$th smallest number of $\{k_t\}$ and $j = [\alpha T]$, the integer part of $\alpha T$.
3. Obtain the minimum and maximum in the MCMC sample $\{\theta_t\}$, denoted by $\theta_{(1)}$ and $\theta_{(T)}$. Draw a sample $\{z_i, i = 1, \ldots, N\}$ with $z_i \overset{iid}{\sim} U(\theta_{(1)}, \theta_{(T)})$.
4. Calculate the weighted proportion $w$ of $\{z_i\}$ that falls within the approximated JHDR, that is,

$$w = \frac{1}{N}\sum_i |I(z_i)|^{\frac{1}{2}}, \text{ where } i \in \{i : f(x|z_i)|I(z_i)|^{-\frac{1}{2}}\pi_\lambda(z_i) \geq \hat{k}_\alpha\}.$$

5. The estimated JHDR volume is given by

$$\hat{V}_{J_\theta} = w \times (\theta_{(T)} - \theta_{(1)}).$$

---

## 7.3 Analysis of Pediatric Cancer Data using JHDR

We apply our measures using JHDR to the pediatric cancer study in Section 6.3. The Fisher information of $p$ is $I(p) = n_P p^{-1}(1-p)^{-1}$, and the posterior density of $p$ with respect to $J_p$ is

$$
\begin{aligned}
\pi_{J_p}(p \mid D_A, D_P, a_0) &\propto \pi(p \mid D_A, a_0)L(p \mid D_P)|I(p)|^{-\frac{1}{2}} \\
&\propto \pi(p)[L(p \mid D_A)]^{a_0}L(p \mid D_P)|I(p)|^{-\frac{1}{2}} \\
&\propto \text{Beta}\left(a_0 y_A + y_P + \frac{1}{2}, a_0(n_A - y_A) + (n_P - y_P) + \frac{1}{2}\right).
\end{aligned}
$$

The JHDR for $p$ and the corresponding volume are given by

$$\text{JHDR}_p(\alpha) = \{p : \pi_{J_p}(p \mid D_A, D_P, a_0) \geq k_\alpha\},$$

$$V_{J_p}(\text{JHDR}_p(\alpha)) = \int \mathbf{1}(p \in \text{JHDR}_p(\alpha))|I(p)|^{\frac{1}{2}} dp.$$

The results of $\mathcal{D}$-AUC and $LCR$ using JHDR are given in Figure 13. They are quite similar to those in Figure 12, with slightly changes that $\mathcal{D}$-AUC $= 0.473$ when $a_0 = 0.04$, and $LCR = 0.193$.

## 8 Discussion

We propose two new Bayesian measures, $LCR(\alpha)$ and $\mathcal{D}(\alpha)$, for quantifying change in data scale and location, respectively. We provide algorithms for computing both measures based on volumes of HDRs and show that both measures can be calibrated.
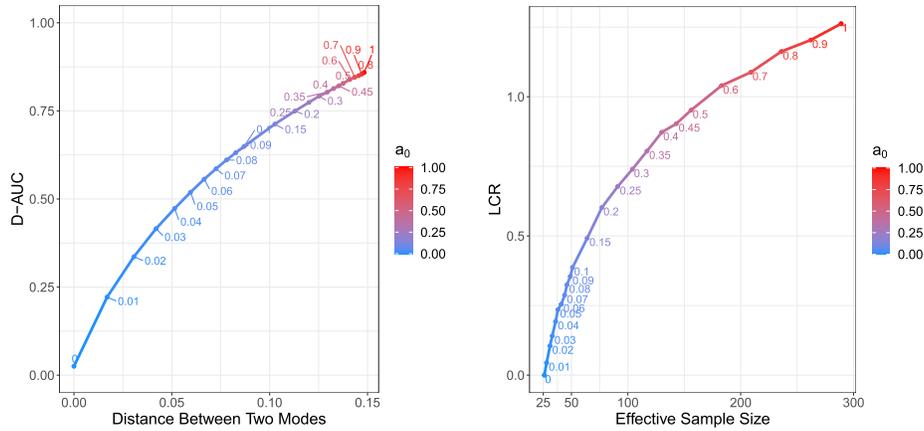
(a) Plot of $\mathcal{D}$-AUC versus distance between two modes

(b) Plot of $LCR$ versus effective sample size

Figure 13: Plots of $\mathcal{D}$-AUC and $LCR$ using JHDR over different choices of $a_0$.

We provide three examples to illustrate how $LCR(\alpha)$ and $\mathcal{D}(\alpha)$ are useful in different areas: comparing two similar studies, measuring the information in missing data, and leveraging adult data in a pediatric cancer study.

We demonstrate that our measures can be computed and calibrated for approximately normal distributions with moderate dimension. For high dimensional non-normal distributions, finding efficient algorithms for evaluating the HDR volume remains challenging.

The use of HDRs means that our proposed measures are not invariant to reparametrization. If the Fisher information is available in closed form, an invariant version JHDR may be used and we provide an algorithm to compute the volume. Calculating our measures using JHDR marginally requires further investigation.

## Supplementary Material

Supplementary Material for "Bayesian Concentration Ratio and Dissonance" (DOI: 10.1214/21-BA1277SUPP; .pdf).

## References

Abramowitz, M. and Stegun, I. (1972). "Handbook of Mathematical Functions." Dover, New York. MR0208797. 827

Ali, S. M. and Silvey, S. D. (1966). "A general class of coefficients of divergence of one distribution from another." *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1): 131–142. MR0196777. 818

Bregman, L. M. (1967). "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming." *USSR Computational Mathematics and Mathematical Physics*, 7(3): 200–217. MR0215617. 818

Chen, M.-H. and Shao, Q.-M. (1999). "Monte Carlo estimation of Bayesian credible and HPD intervals." *Journal of Computational and Graphical Statistics*, 8(1): 69–92. MR1705909. doi: https://doi.org/10.2307/1390921. 818

Conway, K. P., Vullo, G. C., Kennedy, A. P., Finger, M. S., Agrawal, A., Bjork, J. M., Farrer, L. A., Hancock, D. B., Hussong, A., Wakim, P., et al. (2014). "Data compatibility in the addiction sciences: an examination of measure commonality." *Drug and Alcohol Dependence*, 141: 153–158. 817

Csiszár, I. and Shields, P. C. (2004). "Information theory and statistics: A tutorial." *Foundations and Trends® in Communications and Information Theory*, 1(4): 417–528. 818

Dezert, J. and Musso, C. (2001). "An efficient method for generating points uniformly distributed in hyperellipsoids." In *Proceedings of the Workshop on Estimation, Tracking and Fusion: A Tribute to Yaakov Bar-Shalom*. 823, 824

Druilhet, P. and Marin, J.-M. (2007). "Invariant {HPD} credible sets and {MAP} estimators." *Bayesian Analysis*, 2(4): 681–691. MR2361970. doi: https://doi.org/10.1214/07-BA227. 841

Evans, M. and Moshonov, H. (2006). "Checking for prior-data conflict." *Bayesian Analysis*, 1(4): 893–914. MR2282210. doi: https://doi.org/10.1016/j.spl.2011.02.025. 818, 819, 826, 828

Gamalo-Siebers, M., Savic, J., Basu, C., Zhao, X., Gopalakrishnan, M., Gao, A., Song, G., Baygani, S., Thompson, L., Xia, H. A., et al. (2017). "Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation." *Pharmaceutical Statistics*, 16(4): 232–249. 817

Gåsemyr, J. and Natvig, B. (2009). "Extensions of a conflict measure of inconsistencies in Bayesian hierarchical models." *Scandinavian Journal of Statistics*, 36(4): 822–838. MR2573310. doi: https://doi.org/10.1111/j.1467-9469.2009.00659.x. 818

Hellinger, E. (1909). "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen." *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1909(136): 210–271. MR1580780. doi: https://doi.org/10.1515/crll.1909.136.210. 818

Higgins, J. P. and Thompson, S. G. (2002). "Quantifying heterogeneity in a meta-analysis." *Statistics in Medicine*, 21(11): 1539–1558. 817

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). "Measuring inconsistency in meta-analyses." *BMJ*, 327(7414): 557–560. 817

Holmes, C. C., Caron, F., Griffin, J. E., Stephens, D. A., et al. (2015). "Two-sample

Bayesian nonparametric hypothesis testing." *Bayesian Analysis*, 10(2): 297–320. MR3420884. doi: https://doi.org/10.1214/14-BA914.   830

Hyndman, R. J. (1996). "Computing and graphing highest density regions." *The American Statistician*, 50(2): 120–126.   823, 824

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015). "The power prior: theory and applications." *Statistics in Medicine*, 34(28): 3724–3749. MR3422144. doi: https://doi.org/10.1002/sim.6728.   839

Kociba, R., Keyes, D., Beyer, J., Carreon, R., Wade, C., Dittenber, D., Kalnins, R., Frauson, L., Park, C., Barnard, S., et al. (1978). "Results of a two-year chronic toxicity and oncogenicity study of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin in rats." *Toxicology and Applied Pharmacology*, 46(2): 279–303.   832

Kullback, S. and Leibler, R. A. (1951). "On information and sufficiency." *The Annals of Mathematical Statistics*, 22(1): 79–86. MR0039968. doi: https://doi.org/10.1214/aoms/1177729694.   818

Lakshminarayanan, M. and Natanegara, F. (2019). *Bayesian Applications in Pharmaceutical Development*. CRC Press.   817

Lewis, P. O., Chen, M.-H., Kuo, L., Lewis, L. A., Fučíková, K., Neupane, S., Wang, Y.-B., and Shi, D. (2016). "Estimating Bayesian phylogenetic information content." *Systematic Biology*, 65(6): 1009–1023.   817

Lindley, D. V. (1956). "On a measure of the information provided by an experiment." *The Annals of Mathematical Statistics*, 27(4): 986–1005. MR0083936. doi: https://doi.org/10.1214/aoms/1177728069.   818

Marshall, E. and Spiegelhalter, D. (2007). "Identifying outliers in Bayesian hierarchical models: a simulation-based approach." *Bayesian Analysis*, 2(2): 409–444. MR2312289. doi: https://doi.org/10.1214/07-BA218.   818

Morimoto, T. (1963). "Markov processes and the H-theorem." *Journal of the Physical Society of Japan*, 18(3): 328–331. MR0167200. doi: https://doi.org/10.1143/JPSJ.18.328.   818

National Toxicology Program (1982). "Carcinogenesis Bioassay of 2, 3, 7, 8-Tetrachlorodibenzo-p-dioxin (CAS No. 1746-01-6) in Osborne-Mendel Rats and B6C3F1 Mice (Gavage Study)." *National Toxicology Program Technical Report Series*, 209: 1.   832

Nott, D. J., Wang, X., Evans, M., Englert, B.-G., et al. (2020). "Checking for prior-data conflict using prior-to-posterior divergences." *Statistical Science*, 35(2): 234–253. MR4106603. doi: https://doi.org/10.1214/19-STS731.   818, 828

Presanis, A. M., Ohlssen, D., Spiegelhalter, D. J., and De Angelis, D. (2013). "Conflict diagnostics in directed acyclic graphs, with applications in Bayesian evidence synthesis." *Statistical Science*, 28(3): 376–397. MR3135538. doi: https://doi.org/10.1214/13-STS426.   818, 819, 826, 828, 829

Rényi, A. (1961). "On measures of entropy and information." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 547–561. The Regents of the University of California. MR0132570. 818

Schoenfeld, D. A., Zheng, H., and Finkelstein, D. M. (2009). "Bayesian design using adult data to augment pediatric trials." *Clinical Trials*, 6(4): 297–304. 817

Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27(3): 379–423. MR0026286. doi: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x. 818

Shao, K. and Small, M. J. (2011). "Potential uncertainty reduction in model-averaged benchmark dose estimates informed by an additional dose study." *Risk Analysis: An International Journal*, 31(10): 1561–1575. 833

Shi, D. (2017). "Statistical Methods for Information Assessment and Data Compatibility with Applications." Ph.D. thesis, University of Connecticut. 818

Shi, W., Chen, M.-H., Kuo, L., and Lewis, P. O. (2021). "Supplementary Material for "Bayesian Concentration Ratio and Dissonance"." *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1277SUPP. 818, 820, 825, 827, 828

Tanner, M. A. (1996). "Normal Approximations to Likelihoods and to Posteriors." In *Tools for Statistical Inference*, 14–39. Springer. MR1396311. doi: https://doi.org/10.1007/978-1-4612-4024-2. 818

Ware, J., Dockery, D., Spiro Iii, A., Speizer, F., and Ferris Jr, B. (1984). "Passive smoking, gas cooking, and respiratory health of children living in six cities." *American Review of Respiratory Disease*, 129(3): 366–374. 835