Bayesian Analysis (2022) **17**, Number 3, pp. 737–764

Quantifying Observed Prior Impact*

David E. Jones[†], Robert N. Trangucci[‡], and Yang Chen[§,¶]

Abstract. When summarizing a Bayesian analysis, it is important to quantify the contribution of the prior distribution to the final posterior inference because this informs other researchers whether the prior information needs to be carefully scrutinized, and whether alternative priors are likely to substantially alter the conclusions drawn. One appealing and interpretable way to do this is to report an *effective prior sample size* (EPSS), which captures how many observations the information in the prior distribution corresponds to. However, typically the most important aspect of the prior distribution is its *location relative to the data*, and therefore traditional information measures are somewhat deficit for the purpose of quantifying EPSS, because they concentrate on the variance or spread of the prior distribution (in isolation from the data). To partially address this difficulty, Reimherr et al. (2014) introduced a class of EPSS measures based on prior-likelihood discordance. In this paper, we take this idea further by proposing a new measure of EPSS that not only incorporates the general mathematical form of the likelihood (as proposed by Reimherr et al., 2014) but also the specific data at hand. Thus, our measure considers the location of the prior relative to the current observed data, rather than relative to the average of multiple datasets from the working model, the latter being the approach taken by Reimherr et al. (2014). Consequently, our measure can be highly variable, but we demonstrate that this is because the impact of a prior on a Bayesian analysis can intrinsically be highly variable. Our measure is called the (posterior) mean Observed Prior Effective Sample Size (mOPESS), and is a Bayes estimate of a meaningful quantity. The mOPESS well communicates the extent to which inference is determined by the prior, or framed differently, the amount of sampling effort saved due to having relevant prior information. We illustrate our ideas through a number of examples including Gaussian conjugate and non-conjugate models (continuous observations), a Beta-Binomial model (discrete observations), and a linear regression model (two unknown parameters).

Keywords: effective prior sample size, statistical information, Wasserstein distance, Bayes estimate, sensitivity analysis.

1 Introduction

Prior knowledge and assumptions are central to many statistical problems, and in practice it is important to assess their impact on the final inference. When such an assessment is missing, it can be difficult to tell whether the results could be reproduced

*This work is supported by NSF DMS-1811083 (PI: Yang Chen, 2018–2021).

[†]Texas A&M University, College Station, TX, USA, david.jones@tamu.edu

[‡]University of Michigan, Ann Arbor, MI, USA, trangucc@umich.edu

[§]University of Michigan, Ann Arbor, MI, USA, ychenang@umich.edu

[¶]Correspondence to: Yang Chen, Department of Statistics and Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, ychenang@umich.edu

 https://doi.org/10.1214/21-BA1271

with a different prior, or whether similar studies could be made more efficient by incorporating existing information that was neglected. Of course, each researcher chooses whichever prior seems most appropriate to them, but reporting the impact of the prior allows others, and even the original researchers, to better interpret the results. These considerations are important in many scientific studies. For example, Chen et al. (2019) propose a Bayesian analysis of the brightnesses of a large collection of stars based on a multi-telescope astronomical dataset, and highlight that scientific prior distributions provide key information about each of the specific instruments and play a substantial role in the final inference. For both the scientists directly involved in the study, and also others who rely on their work, it is important to understand the role of the prior distributions used, e.g., do the priors associated with one particular instrument have a much greater impact on the inference than those for other instruments?

One appealing and interpretable way to assess prior impact is to provide a measure of the *effective* prior sample size (EPSS), i.e., the approximate number of observations to which the information in the prior is equivalent. Gaussian conjugate models offer a canonical example: with observed data $y_i \overset{iid}{\sim} N(\mu, \sigma^2)$, for $i = 1, \ldots, n$, and conjugate prior distribution $\mu \sim N(\mu_0, \sigma^2/r)$, the posterior distribution of μ is $N(w_n \bar{y}_n + (1 - w_n)\mu_0, \sigma^2/(n+r))$, where $w_n = n/(n+r)$. Based on the denominator $n+r$ in the expression for the posterior variance, the effect of the prior appears to be equivalent to that of r samples, so we say that the EPSS is r. However, this formulation faces two challenges: (a) it is not immediately clear how to generalize beyond conjugate models; and more importantly, (b) when μ_0 is arbitrarily different to $\bar{y}_n$, the prior impact on the posterior mean is arbitrarily large, and is therefore clearly not equivalent to that of r samples.

EPSS measures have gained substantial attention in the literature, and a number of strategies have been proposed in response to the two challenges above, e.g., Clarke (1996), Morita et al. (2008), and Morita et al. (2010). Most of the strategies proposed rely on a comparison between the actual prior π and a default or baseline prior π^b, e.g., the improper prior $\pi^b(\mu) \propto 1$ would be a natural choice for the baseline prior in the Gaussian conjugate model above. This comparative information approach is necessary because there is no universal "non-informative" prior against which to measure prior impact, and Bayesian inference cannot be conducted without a prior. Early generalizations along these lines sought to match the prior π to a hypothetical posterior distribution constructed using the baseline prior π^b and some hypothetical previous samples, that is, they interpreted the prior π as the posterior from a previous analysis. The EPSS is then defined as the number of observations used in the hypothetical posterior distribution, e.g., Clarke (1996) and Morita et al. (2008). These approaches successfully generalize the notion of EPSS, but do not address concern (b) regarding the real impact of the prior when the data mean and prior mean differ substantially. Indeed, these methods do not consider the observed data or the real posterior distribution at all.

Reimherr et al. (2014) instead suggested minimizing the discrepancy between *two* posterior distributions, one using the real prior π and the other using the baseline prior π^b. In this case the EPSS is defined as the difference in the number of samples used by the two posteriors. Similar ideas have also been proposed in slightly different contexts, e.g., see Lin et al. (2007) and Wiesenfarth and Calderazzo (2019). The Reimherr et al. (2014) method offers many improvements over early approaches and goes beyond simply capturing the variance of the prior; it also partially quantifies the impact of the prior *location*. However, it averages over the data using the bootstrap, and therefore does not quantify the impact of the prior for the specific analysis carried out with the observed data at hand, which is of most interest in practice.

Another recent approach introduced by Neuenschwander et al. (2020) defines the EPSS as the expected local-information-ratio (ELIR), i.e., the prior mean of the ratio of the prior information and Fisher information (of a single observation). This approach has the elegant property that a sample of size n from the posterior predictive distribution has an effective sample size of n plus the EPSS. However, similarly to the strategies already mentioned, this method does not take into account the observed data and therefore does not fully capture the impact of the prior on the Bayesian analysis at hand.

In this paper, we follow a similar approach to Reimherr et al. (2014) but propose a new EPSS measure which addresses the above limitations by conditioning on the observed data, and thereby directly quantifies the prior impact for the actual analysis performed. Our measure is called the *mean Observed Prior Effective Sample Size* (mOPESS), where 'Observed' indicates that the observed data is treated as fixed, and 'mean' refers to an average over additional future samples drawn from the posterior predictive distribution (the purpose of which will become clear). This new measure was inspired by the work of Efron and Hinkley (1978) which highlighted that observed Fisher information is sometimes more useful than expected Fisher information. By providing an explicit definition of the mOPESS in terms of future observations, we also identify the real-world estimand of interest, which we call the Observed Prior Effective Sample Size (OPESS), i.e., a quantity that would be realized if future samples were actually collected. The interpretation of the OPESS is essentially the number of additional samples that must be combined with the baseline prior π^b in order to obtain similar inference to that under our actual prior π. In other words, the OPESS communicates how much sampling effort is saved by having access to the information in the prior π, rather than only the default information captured by π^b. Further appealing properties of our mOPESS measure include a Bayes estimate interpretation and no lower limit on the observed data sample size n. The latter property is important because prior impact is often most pronounced, and therefore of most interest, when the sample size is small. In contrast, Reimherr et al. (2014) require n to be large because their method relies on the bootstrap and an accurate estimate of the "true parameter" value, see Section 2.2 for a review. In summary, our approach represents a substantially improved method for quantifying prior impact in practice, and its real-world interpretation makes it a valuable tool for clearly reporting the contribution of priors in Bayesian analyses.

One possible limitation of our approach is the need for a baseline or default prior against which to compare the prior at hand. However, in our opinion, a key purpose of an EPSS measure is to *communicate* the impact of a prior, and for this objective comparing to a standard prior that many researchers are already familiar with is in fact a strength rather than a weakness. Of course, if necessary our mOPESS measure could be computed for several different baseline priors, but we suspect one version will usually provide a sufficient summary of prior impact. As mentioned above, most of the existing literature on EPSS measures has similarly concluded that comparing against a baseline prior is desirable or necessary or both.

This paper is organized as follows. Section 2 briefly overviews some topics from the broader literature connected to EPSS measures and their uses, then summarizes the EPSS methods on which we build, and lastly provides a motivating Gaussian example to illustrate our mOPESS measure. Section 3 defines the mOPESS and discusses its computation, first in general and then in the specific case of the motivating Gaussian example introduced in Section 2.3. Section 4 provides intuition and theory supporting our method. Section 5 provides additional numerical results in the form of non-conjugate Gaussian, Beta-Binomial, and regression model examples. Section 6 provides a summary. Proofs are given in the appendices.

2 Connections with Existing Work and a Motivating Example

2.1 EPSS and the Broader Literature on Prior Distributions

To provide greater context for the importance of EPSS measures, we now briefly discuss several concepts that have been studied in the literature on prior distributions, and highlight their connections to EPSS measures.

Prior-likelihood conflict, e.g., Evans et al. (2006); Bousquet (2008); Walter and Augustin (2009); Nott et al. (2020, 2021). The topic of prior likelihood-data conflict is very much related to EPSS measures in the sense that such conflict can indicate that the prior distribution has a large influence on the final analysis. On the other hand, measures of EPSS are not restricted to quantifying prior-likelihood conflict: in the case of prior-likelihood alignment or weakly informative priors, the mOPESS measure proposed here gives non-zero values which describe the extent to which the prior is facilitating the inference, e.g., by increasing posterior concentration.

Sample size determination, e.g., Wang et al. (2002); Sahu and Smith (2006); Clarke et al. (2006); Gupta et al. (2016). In clinical trials, a classical problem is to determine the sample size needed to achieve a certain power when performing a hypothesis test for the presence of a treatment effect, see Sahu and Smith (2006) for more detailed discussion and examples. In its classical form this line of research does not usually emphasize prior impact and is not closely related to our work. However, the concepts of prior-likelihood conflict and EPSS are important for sample size determination in *adaptive* clinical trials, where priors are typically constructed from an interim analysis, e.g., Hobbs et al. (2013) uses effective historical sample size (EHSS) to determine a randomization procedure for allocating patients. Similarly, Wiesenfarth and Calderazzo (2020) compare EPSS measures that quantify prior information in terms of historical/external samples (e.g., Morita et al., 2008) or current/new samples (e.g., Reimherr et al., 2014); and tailor the method in Reimherr et al. (2014) to the adaptive design setting. Wiesenfarth and Calderazzo (2020) specifically investigates priors that adaptively discard prior information in the case of prior-data conflict, e.g., robust mixture (Berger et al., 1986), power (Ibrahim et al., 2015) and commensurate (Hobbs et al., 2012) priors. Our mOPESS measure could be applied in such studies to identify other similarly "adaptive" priors and to better quantify their impact in Bayesian clinical trials.

Bayesian prior sensitivity analysis, e.g., Berger (1990); Weiss (1996); Roos et al. (2011, 2015). This line of work is aimed at assessing the impact on the posterior inference when the prior distribution is perturbed, thus possibly identifying parameters for which the posterior inference is highly sensitive to *changes* in the prior. In our work, we only assess the impact of the given prior on the current inference, as compared to a baseline prior, and do not consider whether this impact would be different for other similar priors. On the other hand, if posterior inference substantially varied across a group of priors, then our mOPESS measure would likely be high for some or all of the priors in question, and so our measure could be used to detect this type of sensitivity.

Prior construction for objective Bayes, e.g., Kass and Wasserman (1996); Ghosh et al. (2011b); Berger et al. (2015); Consonni et al. (2018); Leisen et al. (2020). For example, Ghosh et al. (2011a) constructs "objective priors" by maximizing an approximate expression for the distance between the prior and the posterior under a general divergence criterion. In general this line of research focuses on constructing priors in such a way as to avoid the subjective nature of Bayesian inference, e.g., ensuring the prior has little influence in cases where the data contain substantial information. Our approach is about assessing the prior influence and giving an intuitively quantitative measure of the impact of an informative prior or subjective prior. Moreover, as opposed to approximate measures based on asymptotic expansions (e.g., Ghosh et al., 2011a), our method works on the exact posteriors and is particularly designed for finite sample settings, in which prior impact is potentially substantial. In our approach, reference priors (typically, but not necessarily, non-informative) are only used as a baseline to compare our working priors against. On the other hand, our mOPESS measure could likely be used to assess whether a prior is a suitable alternative to the current non-informative prior of choice.

2.2 Some Existing Methods of Measuring EPSS

Suppose that π is our prior distribution for a collection of unknown parameters of interest $\theta \in \Theta$. Let π^b be a baseline prior and $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be unknown *hypothetical* previous data with probability density $f(\boldsymbol{x}|\theta)$. Imagine that our real prior π is the posterior distribution $q^b(\cdot|\boldsymbol{x}) \propto f(\boldsymbol{x}|\cdot)\pi^b(\cdot)$ computed using the unknown hypothetical dataset $\boldsymbol{x}$. Under this formulation, Clarke (1996) considers

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathcal{X}}{\arg\min}\, \mathrm{KL}(q^b(\cdot|\boldsymbol{x}), \pi(\cdot)), \tag{2.1}$$

where $\mathrm{KL}(g, h)$ denotes the Kullback-Leibler divergence (KL divergence) defined as $\int_\Theta \log(g(\theta)/h(\theta))g(\theta)d\theta$, and $\mathcal{X}$ is the support of f (for simplicity we assume $\mathcal{X}$ to be the same for all $\theta \in \Theta$). In words, the approach of Clarke (1996) is to find the hypothetical dataset $\boldsymbol{x}^*$ that, when combined with the baseline prior π^b, produces the posterior distribution $q^b(\cdot|\boldsymbol{x}^*)$ with minimum KL divergence from our true prior π. The EPSS can then be quantified as the number of individual observations contained in $\boldsymbol{x}^*$. Note that the density f is a user specified hypothetical distribution for prior data, and is not necessarily the same as the model for any actual data.

The above approach is distinguished from most other methods (such as those mentioned below) in that it gives a specific dataset $\boldsymbol{x}^*$ which represents the information in the prior. An advantage of this approach is that $\boldsymbol{x}^*$ can potentially capture other aspects of the information contained in π in addition to the EPSS. However, $\boldsymbol{x}^*$ has no concrete relation to the likelihood or data at hand, which we consider to be a drawback, at least when the impact of the prior on a specific analysis is of primary interest.

Morita et al. (2008) adopt a similar approach but measure distance using the difference of the second derivative of the log densities rather than KL divergence. Furthermore, to avoid the peculiarity of reporting a specific dataset $\boldsymbol{x}^*$, and to take account of uncertainty regarding the hypothetical dataset; they take an expectation over $\boldsymbol{x}$, i.e., they compute $E[\frac{\partial^2}{\partial\theta^2}\log q^b(\cdot|X)]$. This treatment of the hypothetical previous data $\boldsymbol{x}$ may be preferable to that of Clarke (1996). But for the purpose of assessing prior impact on a specific analysis, the Morita et al. (2008) method suffers from the same fundamental problem of not taking the likelihood of any actual data into account.

To address this limitation, Reimherr et al. (2014) introduced the notion of prior-likelihood discordance and incorporated it in their measures of EPSS. The key change they proposed was to compare *two* posterior distributions rather than comparing a prior to a (hypothetical) posterior. To make the comparison, under each prior π, they consider the expected mean squared error (MSE) when a draw from the posterior is used to estimate the true parameter θ_T, i.e.,

$$U_{\pi,\theta_T}(I) = E_{\theta_T}[\mathrm{MSE}(\pi, X_I)] = \int_{\mathcal{X}_I} \mathrm{MSE}(\pi, \boldsymbol{x}_I) f(\boldsymbol{x}_I|\theta_T) d\boldsymbol{x}_I,$$

where the "posterior MSE", as defined in Reimherr et al. (2014), is

$$\mathrm{MSE}(\pi, \boldsymbol{x}_I) = \mathrm{Var}_\pi(\theta|\boldsymbol{x}_I) + \left(E_\pi[\theta|\boldsymbol{x}_I] - \theta_T\right)^2.$$

Reimherr et al. (2014) use I to indicate the information contained in the hypothetical data $\boldsymbol{x}_I$ and in their main examples it represents the sample size (because the samples are assumed to be independent and identically distributed). Let n be the sample size of the *real data*, denoted by $\boldsymbol{y}$. For hypothetical sample size $k \ll n$, Reimherr et al. (2014) estimate the EPSS of an informative prior π relative to a baseline prior π^b by the smallest integer r such that

$$\hat{U}_{\pi,\hat{\theta}}(k) \approx \hat{U}_{\pi_b,\hat{\theta}}(k+r),$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ_T based on $\boldsymbol{y}$, and $\hat{U}$ is computed by averaging over datasets of size k drawn from the empirical distribution (hence the constraint that $k \ll n$). By a slight abuse of terminology, we refer to their averaging method as bootstrapping (as they do). One of the novel aspects of this formulation is that their estimate of EPSS, r, is allowed to be negative. This is helpful when, for example, we are trying to assess if π is a low-information prior and therefore might feasibly have less impact than π^b.

The approach of Reimherr et al. (2014) described above has a number of advantages over earlier methods: (i) it focuses on the impact of the prior on posterior inference;

(ii) it incorporates the likelihood, although for reduced data size; and (iii) it proposes a potentially reasonable method for generating datasets to combine with π and π^b (bootstrapping). There are however still some limitations of their approach. Firstly, their method averages over the data and therefore their measure of EPSS does not tell us what the impact of the prior is on the inference using the observed data $\boldsymbol{y}$, which is of most interest in practice. Secondly, their approach relies on bootstrapping the data and estimating θ_T which both require n to be large, but the impact of a prior is usually greatest and of most interest when n is small. Lastly, their use of MSE is not necessarily the best way of quantifying the difference between two posterior distributions and therefore the prior impact. Indeed, there is in fact no reason to introduce the notion of a true parameter value θ_T in order to measure prior impact.

2.3 Motivating Gaussian Example

Suppose that we have observed data $y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, for $i = 1, \ldots, n$. Assume that σ^2 is known, that our prior for μ is a conjugate prior, denoted $\pi(\mu) \equiv \mathcal{N}(\mu_0, \lambda_0^2)$, and that the baseline prior is $\pi^b(\mu) \propto 1$. Suppose that $\mu = \mu_0 = 0$, $\sigma^2 = 1$, and $\lambda_0^2 = 0.1$. We now use this example to illustrate the behavior of our EPSS measure, which is called the *mean Observed Prior Effective Sample Size* (mOPESS). All mathematical and computational details are deferred to Section 3.

The top left panel of Figure 1 shows the mOPESS plotted against the data mean $\bar{y}_n$. It can be seen that the mOPESS increases with the difference $|\bar{y}_n - \mu_0|$ between the prior mean and data mean, which is a key feature of our proposed measure of EPSS. The top right panel of Figure 1 shows the quantiles of the *Observed Prior Effective Sample Size* (OPESS), i.e., it summarizes the distribution of which the mOPESS is the mean for any given $\bar{y}_n$. The OPESS represents a real-world quantity (defined in Section 3.1) which captures the impact of the prior, and can in principle be observed by collecting more samples. Variability in the OPESS for a fixed value of $\bar{y}_n$ indicates genuine uncertainty about the future observations. The mOPESS (i.e., the mean of the OPESS distribution) averages over this uncertainty and is our preferred single number summary of prior impact.

The bottom left panel of Figure 1 corresponds to the point indicated by a "+" symbol in the top left panel, i.e., one of two points with the largest value of $|\bar{y}_n - \mu_0|$. In particular, the bottom left panel shows the priors π and π^b, as well as the corresponding posterior distributions denoted q_n and q_n^b, respectively. Note that the posterior distribution q_n is pulled towards zero by the informative conjugate prior π. The top left panel shows that in this case the mOPESS is larger than for values of $\bar{y}_n$ that are closer to the prior mean $\mu_0 = 0$. In particular, the mOPESS has a value of about 14.5, which has the interpretation that on average an investigator using q_n^b would need to collect 14.5 additional samples to obtain similar inference to an investigator using π, *for the specific value of* $\bar{y}_n$ *currently at hand, i.e., for* $\bar{y}_n \approx -0.6$.

The bottom right panel of Figure 1 illustrates the case where $|\bar{y}_n - \mu_0|$ is smallest across the points plotted in the top left panel, i.e., the point indicated by a cross in the top left panel. From the bottom right panel we can see that in this case both

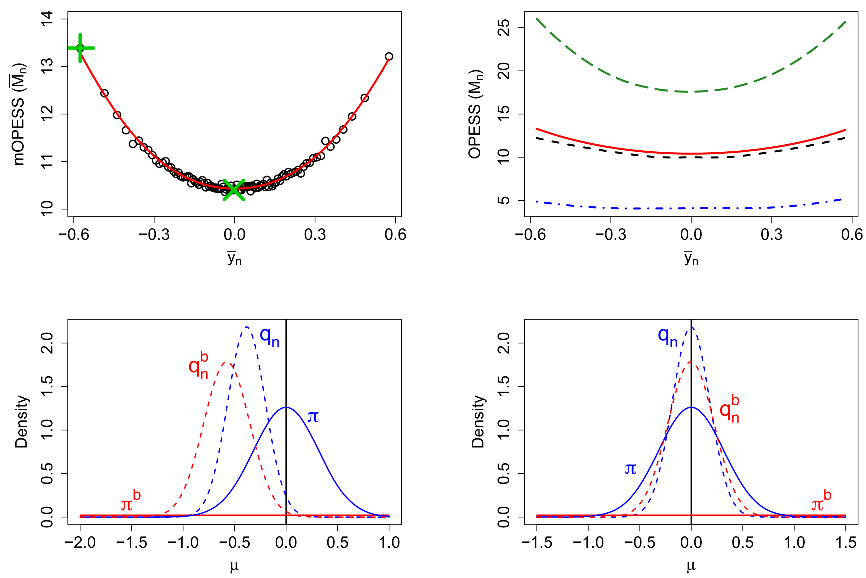

Figure 1: (Left top) mOPESS as a function of $\bar{y}_n$. The solid curve shows a LOESS (LOcal polynomial regrESSion) fit to the points. (Right top) 95% quantile (long dash curve), median (short dash curve), and 5% quantile (dash-dot curve) of the OPESS as a function of $\bar{y}_n$. The solid curve is the same as in the left panel. (Left and right bottom) posteriors q_n and q_n^b (dashed lines) and priors π and π^b (solid lines) for the dataset indicated by a green "+" symbol and a cross, respectively, in the top left plot.

posteriors are centered very close to zero. Specifically, the posteriors have substantial overlap because the conjugate prior π is centered at $\mu = \mu_0 = 0 \approx \bar{y}_n$ and so does not cause q_n to have a substantially different mean to q_n^b, only a smaller variance. Returning to the top left panel we can see that this case corresponds to a mOPESS value of around 10.5, which is one of the smallest among the points plotted.

In conclusion, we can see that the mOPESS is larger the further $\bar{y}_n$ is from $\mu_0 = 0$ and that this is because the prior has more impact on the posterior in these cases. Thus, at least in this simple example, our mOPESS measure of EPSS seems to have an intuitive interpretation that well captures the way the prior impact changes with the observed data.

3 mOPESS Definition and Computation

3.1 mOPESS Definition

Our guiding intuition is that we want to know how many extra samples are needed to obtain similar inference under the baseline prior as that under our informative prior. To that end, for each $m = n + r$, where $r \in \mathbb{Z}_{\geq 0}$, we introduce a hypothetical expanded dataset $\boldsymbol{x}^{(m)} = \{x_1^{(m)}, \ldots, x_m^{(m)}\} \equiv \boldsymbol{y} \cup \{x_{n+1}^{(m)}, \ldots, x_m^{(m)}\}$, i.e., $x_i^{(m)} = y_i$, for $i = 1 \ldots, n$. The superscripts '(m)' are necessary because we do not assume that $\boldsymbol{x}^{(m+1)} = \boldsymbol{x}^{(m)} \cup \{x_{m+1}\}$, for reasons to be explained in Section 3.2. If $\boldsymbol{x}^{(m)}$ was known for all m then intuitively we would choose the EPSS to be $r = m - n$ for the m that minimizes the distance between the two posterior distributions $q_n \equiv q(\cdot|\boldsymbol{y}) \propto f(\boldsymbol{y}|\cdot)\pi(\cdot)$ and $q_m^b \equiv q^b(\cdot|\boldsymbol{x}^{(m)}) \propto f(\boldsymbol{x}^{(m)}|\cdot)\pi^b(\cdot)$, where π is our real prior whose EPSS is to be measured, π^b is the baseline prior, and f is the model. We denote the distance for a given m by $D(q^b(\cdot|\boldsymbol{x}^{(m)}), q(\cdot|\boldsymbol{y}))$. The collection of expanded datasets is denoted by $\boldsymbol{x}_{\text{L}} = \{\boldsymbol{x}^{(n)}, \boldsymbol{x}^{(n+1)}, \ldots, \boldsymbol{x}^{(L)}\}$, where L is the maximum feasible value of m, or in other words $L - n$ is the maximum feasible magnitude of the EPSS associated with π.

We must account for the possibility that our prior π is in fact *less* impactful than the baseline prior π^b. This happens when the prior π is more diffuse than the baseline π^b or is similarly diffuse but has greater location agreement with the data than π^b. Thus, we also consider the alternative distance $D(q^b(\cdot|\boldsymbol{y}), q(\cdot|\tilde{\boldsymbol{x}}^{(m)}))$, where the extra hypothetical samples are combined with our real prior π rather than with the baseline π_b. Again we do not assume that $\tilde{\boldsymbol{x}}^{(m)} = \boldsymbol{x}^{(m)}$, for reasons to be explained in Section 3.2. We write $\boldsymbol{x}_{\text{L}}^{\text{all}} = \boldsymbol{x}_{\text{L}} \cup \tilde{\boldsymbol{x}}_{\text{L}} = \{\boldsymbol{x}^{(n)}, \tilde{\boldsymbol{x}}^{(n)}, \boldsymbol{x}^{(n+1)}, \tilde{\boldsymbol{x}}^{(n+1)}, \ldots, \boldsymbol{x}^{(L)}, \tilde{\boldsymbol{x}}^{(L)}\}$ to denote all the future samples combined, and for conciseness introduce the notation $D(m)$ and $\widetilde{D}(m)$ as shorthand for the distances $D\left(q^b(\cdot|\boldsymbol{x}^{(m)}), q(\cdot|\boldsymbol{y})\right)$ and $D\left(q^b(\cdot|\boldsymbol{y}), q(\cdot|\tilde{\boldsymbol{x}}^{(m)})\right)$, respectively. We can now define the underlying quantity of interest.

Definition 3.1. *For a given realization of $\boldsymbol{x}_{\text{L}}^{all}$, the Observed Prior Effective Sample Size (OPESS) is*

$$M_n(\boldsymbol{x}_{\text{L}}^{all}) = \begin{cases} \operatorname{argmin}_{n \le m \le L} \{D(m)\} - n & \text{if} \quad S_n(\boldsymbol{x}_{\text{L}}^{all}) = 1, \\ n - \operatorname{argmin}_{n \le m \le L} \left\{\widetilde{D}(m)\right\} & \text{if} \quad S_n(\boldsymbol{x}_{\text{L}}^{all}) = -1, \end{cases} \tag{3.1}$$

where

$$S_n(\boldsymbol{x}_{\text{L}}^{all}) = \begin{cases} 1 & \text{if} \quad \min_{n \le m \le L} \{D(m)\} \le \min_{n \le m \le L} \left\{\widetilde{D}(m)\right\}, \\ -1 & \text{if} \quad \min_{n \le m \le L} \{D(m)\} > \min_{n \le m \le L} \left\{\widetilde{D}(m)\right\}. \end{cases}$$

The OPESS is negative when $S_n(\boldsymbol{x}_{\text{L}}^{\text{all}}) = -1$ because this suggests that π is less informative than the baseline prior π^b. In practice, the future samples $\boldsymbol{x}_{\text{L}}^{\text{all}}$ are unknown and therefore the OPESS must be estimated. The mOPESS defined in Definition 3.2 below is simply the posterior mean of the OPESS, and as such provides a convenient estimate of the OPESS. Posterior quantiles of the OPESS distribution and other summaries could also be reported to provide a measure of uncertainty.

Definition 3.2. *The theoretical (posterior) mean Observed Prior Effective Sample Size (mOPESS) is*

$$\overline{M}_n^T = \int_{\mathcal{X}} M_n(\boldsymbol{x}_{\text{L}}^{all}) p(\boldsymbol{x}_{\text{L}}^{all}|\boldsymbol{y}, \pi) d\boldsymbol{x}_{\text{L}}^{all}, \tag{3.2}$$

where $\mathcal{X}$ is the domain of $\boldsymbol{x}_{\text{L}}^{all}$ and $p(\cdot|\boldsymbol{y}, \pi)$ is the corresponding posterior predictive distribution.

3.2 Discussion of the General mOPESS Definition and Computation

In practice, it makes sense for the posterior predictive distribution in Definition 3.2 to factorise, i.e.,

$$p(\boldsymbol{x}_{\mathrm{L}}^{\mathrm{all}}|\boldsymbol{y},\pi) = \int_{\Theta} p(\boldsymbol{x}_{\mathrm{L}}|\boldsymbol{y},\theta)p(\tilde{\boldsymbol{x}}_{\mathrm{L}}|\boldsymbol{y},\theta)q(\theta|\boldsymbol{y})d\theta \tag{3.3}$$

$$= \int_{\Theta} \prod_{m=n}^{L} p(\boldsymbol{x}^{(m)}|\boldsymbol{y},\theta) \prod_{m=n}^{L} p(\tilde{\boldsymbol{x}}^{(m)}|\boldsymbol{y},\theta)q(\theta|\boldsymbol{y})d\theta, \tag{3.4}$$

as we now explain. A researcher who does not want to use π would not collect many additional samples and then attempt to find the m to minimize the distance between their posterior and $q(\cdot|\boldsymbol{y})$. Instead, the researcher would simply collect a fixed number of additional samples $r = m - n$ (fixed in the sense that no minimization of the distance to $q(\cdot|\boldsymbol{y})$ is performed). Therefore the correct question to ask when trying to estimate the OPESS of π is as follows: if there are multiple independent researchers each of whom chooses a different value of r, then whose inference will most closely agree with our inference? This is why, for the mOPESS to correspond to normal scientific procedure, the hypothesized future samples $\boldsymbol{x}^{(m)}$ (and $\tilde{\boldsymbol{x}}^{(m)}$) need to be conditionally independent across values of m, given θ. We avoid assuming that $\tilde{\boldsymbol{x}}_{\mathrm{L}} = \boldsymbol{x}_{\mathrm{L}}$ for essentially the same reason: for the interpretation of the mOPESS to correspond to normal scientific procedure, we cannot assume that each individual researcher computes both $q^b(\cdot|\boldsymbol{x}^{(m)})$ and $q(\cdot|\boldsymbol{x}^{(m)})$ and then decides which to use depending on whether $D(m)$ or $\widetilde{D}(m)$ is smaller. We instead assume there are two researchers in the population for each value of m, one who computes $q^b(\cdot|\boldsymbol{x}^{(m)})$ and one who computes $q(\cdot|\tilde{\boldsymbol{x}}^{(m)})$, and since the researchers will likely have different laboratories it is natural to assume that $\boldsymbol{x}^{(m)}$ and $\tilde{\boldsymbol{x}}^{(m)}$ are independent. Unconditionally, all the future samples are dependent, which corresponds to the real-world in that all additional samples collected would be generated using the same underlying, but unknown, value of θ.

In this paper we set the discrepancy measure D to be the 2-Wasserstein distance, and from hereon replace "D" by "W_2" in our notation. For $p \geq 1$, let u and v be probability measures defined on $\mathcal{M}$ with finite p^{th} moment. The p-Wasserstein distance between u and v is defined as

$$W_p(u,v) = \left(\inf_{\gamma \in \Gamma(u,v)} \int_{\mathcal{M}\times\mathcal{M}} d(x,y)^p \, \mathrm{d}\gamma(x,y) \right)^{1/p}, \tag{3.5}$$

where $\Gamma(u,v)$ denotes the set of measures on $\mathcal{M}\times\mathcal{M}$ with marginals u and v respectively, and d is a metric on $\mathcal{M}$. Conveniently, in the case of multivariate Gaussian distributions the 2-Wasserstein distance can be computed in closed form, and more generally there are efficient software packages for approximating it given samples from the two distributions at hand, e.g., Schuhmacher et al. (2019). On the other hand, our framework is general and other measures of posterior discrepancy could also be used, e.g., Kullback-Leibler

Algorithm 1: General procedure for computing the mOPESS.

Step 1: Compute $q_n^b \equiv q^b(\cdot|\boldsymbol{y})$ and $q_n \equiv q(\cdot|\boldsymbol{y})$.

Step 2: For $j = 1, \ldots S$:

- **Part a:** Generate extra samples $\boldsymbol{x}_L^{\text{all}} = \boldsymbol{x}_L \cup \tilde{\boldsymbol{x}}_L$ from (3.3)–(3.4).
- **Part b:** Compute $q_m^b \equiv q^b(\cdot|\boldsymbol{x}^{(m)})$ and $q_m \equiv q(\cdot|\tilde{\boldsymbol{x}}^{(m)})$, for $m = n+1, \ldots, L$.
- **Part c:** Compute the distances $W_2(m), \widetilde{W}_2(m)$, for $m = n, \ldots, L$.
- **Part d:** Calculate the OPESS $M_n^{(j)}$ given by (3.1).

Step 3: Report the (estimated) mOPESS: $\overline{M}_n = \frac{1}{S}\sum_{j=1}^{S} M_n^{(j)}$.

(KL) divergence or mean squared error as adopted by Clarke (1996) and Reimherr et al. (2014), respectively. General f-divergences (Ali and Silvey, 1966; Sason and Verdú, 2016), of which KL divergence is a special case, provide further options.

Algorithm 1 summarizes how to estimate the mOPESS in practice. In Step 2, S denotes the number of realizations of the OPESS simulated. The procedure is widely applicable and can be implemented for a large family of models beyond the specific cases considered in this paper. Naturally, we use analytical forms of the posterior distributions and the Wasserstein distances when available; otherwise, we use approximation strategies such as importance sampling or Markov chain Monte Carlo (MCMC) methods (e.g., Marin and Robert, 2007; Liu, 2008; Brooks et al., 2011). For convenience, in Algorithm 1 and the remainder of this paper, we use mOPESS to refer to the estimate $\overline{M}_n = \frac{1}{S}\sum_{j=1}^{S} M_n^{(j)}$, as opposed to the theoretical posterior mean of the OPESS, denoted $\overline{M}_n^T$ in (3.2).

3.3 Implementation and Discussion for the Gaussian Example

Algorithm 2 provides a detailed version of Algorithm 1 for the case of the Gaussian example introduced in Section 2.3. We applied Algorithm 2 to 300 simulations of $\bar{y}_n$, and set $S = 10{,}000$ in Step 2, i.e., for each value of $\bar{y}_n$, the value of $\overline{M}_n$ was computed via 10,000 Monte Carlo samples of $\boldsymbol{x}_{\text{L}}^{\text{all}}$.

The left panel of Figure 2 shows the posteriors q_n and q_n^b (dashed lines) for a single example observed value of $\bar{y}_n$, with $n = 20$. The priors π and π^b are also plotted (solid lines). The right panel of Figure 2 shows the distribution of the mOPESS, i.e., of $\overline{M}_n$, across 300 datasets. Interestingly, in the current context $\overline{M}_n$ is quite variable and is always higher than the nominal EPSS of 10 (vertical line). The nominal EPSS is 10 because of the following three information based analogies between the prior and data: (i) if $n = 10$ then the Fisher information is $n/\sigma^2 = 1/\lambda_0^2 = 10$, (ii) if $n = 10$ then $\bar{y}_n \sim \pi$, and (iii) for any n, the posterior distribution is $\mathcal{N}(0, \sigma^2/(n+10))$.

In Section 4.1 we illustrate that there is a good explanation for this disagreement with the nominal EPSS: classical information measures consider the prior in isolation

Algorithm 2: mOPESS computation for Gaussian example.

Step 1: Compute the initial posterior distributions

$$q_n^b \equiv q^b(\cdot|\boldsymbol{y}) = \mathcal{N}\left(\overline{y}_n, \frac{\sigma^2}{n}\right), \tag{3.6}$$

$$q_n \equiv q(\cdot|\boldsymbol{y}) = \mathcal{N}\left(\mu_n = (1-w_n)\mu_0 + w_n\overline{y}_n, \frac{\sigma^2}{n+z}\right), \tag{3.7}$$

where $w_u = u/(u+z)$, $z = \sigma^2/\lambda_0^2$.

Step 2: For $j = 1, \ldots S$:

Part a: Generate extra samples $\boldsymbol{x}_L^{\text{all}} = \boldsymbol{x}_L \cup \tilde{\boldsymbol{x}}_L$ by drawing $\mu^* \sim q_n$ and then drawing $\boldsymbol{x}_{\text{L}}^{\text{all}}$ from

$$p(\boldsymbol{x}_{\text{L}}^{\text{all}}|\boldsymbol{y}, \mu^*) = \prod_{m=n+1}^{L} \prod_{i=n+1}^{m} \mathcal{N}(x_i^{(m)}|\mu^*, \sigma^2) \prod_{i=n+1}^{m} \mathcal{N}(\tilde{x}_i^{(m)}|\mu^*, \sigma^2).$$

Part b: For $m = n+1, \ldots, L$, compute

$$q_m^b \equiv q^b(\cdot|\boldsymbol{x}^{(m)}) = \mathcal{N}\left(\overline{x}_m, \frac{\sigma^2}{m}\right), \tag{3.8}$$

$$q_m \equiv q(\cdot|\tilde{\boldsymbol{x}}^{(m)}) = \mathcal{N}\left(\mu_m = (1-w_m)\mu_0 + w_m\overline{\tilde{x}}_m, \frac{\sigma^2}{m+z}\right). \tag{3.9}$$

Part c: For $m = n, \ldots, L$, compute the distances

$$W_2(m) \equiv W_2\left(q_m^b, q_n\right) = \mathbb{D}_{m,n} + \left(\frac{\sigma}{\sqrt{m}} - \frac{\sigma}{\sqrt{n+z}}\right)^2, \tag{3.10}$$

$$\widetilde{W}_2(m) \equiv W_2\left(q_m, q_n^b\right) = \widetilde{\mathbb{D}}_{n,m} + \left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma}{\sqrt{m+z}}\right)^2, \tag{3.11}$$

where $\mathbb{D}_{u,v} = (\bar{x}_u - \mu_v)^2$, for $u, v \in \{m, n\}$, and $\bar{x}_u = u^{-1}\sum_i^u x_i^{(u)}$, $\mu_v = (1-w_u)\mu_0 + w_u\bar{x}_u$, $w_u = u/(u+z)$ (and $\widetilde{\mathbb{D}}_{u,v}$ is analogously defined).

Part d: Calculate the OPESS $M_n^{(j)}$ given by (3.1).

Step 3: Report the (estimated) mOPESS: $\overline{M}_n = \frac{1}{S}\sum_{j=1}^{S} M_n^{(j)}$.

and can typically only correspond to the prior impact if there is no data. As soon as some data are collected there is always some disagreement between the prior and the data and therefore $\overline{M}_n$ is usually greater than the nominal EPSS, at least in the current Gaussian conjugate model example. On the other hand, in the right panel of Figure 1, the 5% quantile of $M_n(\boldsymbol{x}_{\text{L}}^{\text{all}})$ (dash-dot curve) shows that there often exist

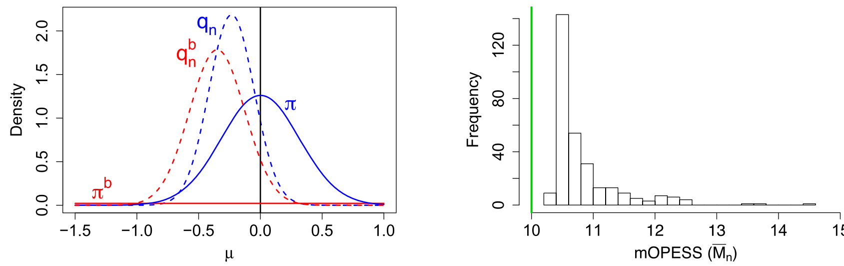

Figure 2: (Left) posteriors q_n and q_n^b (dashed lines) and priors π and π^b (solid lines) for a single simulated dataset $\boldsymbol{y}$. (Right) distribution of the mOPESS ($\overline{M}_n$) across 300 simulated datasets. The vertical line shows the nominal EPSS of 10.

some realizations of $\boldsymbol{x}_{\mathrm{L}}^{\mathrm{all}}$ such that the value of the OPESS M_n is less than the nominal EPSS. Indeed, q_m^b may by chance be closest to q_n after $m - n < 10$ additional samples.

4 Method Justification and Theory for the Gaussian Example

In this section we focus on the Gaussian conjugate model introduced in Sections 2.3 and 3.3, for which we derive theoretical results to justify our proposed method. These results can be seen as general in the sense that the Bernstein von Mises Theorem ensures that the posterior distribution is asymptotically Gaussian under mild conditions, see Van der Vaart (2000) and references therein. On the other hand, the main role of these results is to provide a foundation for understanding and developing the mOPESS, and we emphasize that our method does not rely on asymptotic posterior normality or consistency of the MLE (Maximum Likelihood Estimate). Indeed, our approach is distinguished from many existing methods in that it is designed for small or moderate sample size settings, in which the prior can substantially impact the inference.

4.1 Justification of Sampling Distribution for Extra Observations

Let $r = m - n$ and denote the additional samples collected by $s_1^{(m)}, \ldots, s_r^{(m)}$, i.e., $\{x_1^{(m)}, \ldots, x_m^{(m)}\} = \{y_1, \ldots, y_n, s_1^{(m)}, \ldots, s_r^{(m)}\}$. We write $\bar{s}_r$ to denote $\frac{1}{r}\sum_{i=1}^{r} s_i^{(n+r)}$. Returning to the Gaussian conjugate model introduced earlier, we have

$$W_2(m) = \left(\mu_n - \frac{n}{m}\bar{y}_n - \frac{r}{m}\bar{s}_r\right)^2 + \left(\frac{\sigma}{\sqrt{m}} - \frac{\sigma}{\sqrt{n+z}}\right)^2, \tag{4.1}$$

$$\widetilde{W}_2(m) = \left(\bar{y}_n - \frac{n+z}{m+z}\mu_n - \frac{r}{m+z}\bar{s}_r\right)^2 + \left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma}{\sqrt{m+z}}\right)^2. \tag{4.2}$$

Recall that in our approach described in Section 3.1 the future samples are drawn from the posterior predictive distribution (3.3)–(3.4) under π, meaning that

$$\bar{s}_r|\bar{y}_n, \pi \sim N\left(\mu_n, \left(\frac{1}{r} + \frac{1}{n+z}\right)\sigma^2\right). \tag{4.3}$$

In contrast to our approach, Morita et al. (2008) sample from the distribution of hypothetical previous data and Reimherr et al. (2014) bootstrap the observed data. Our proposed sampling method is therefore not the only option and in order to provide justification for our choice it is instructive to consider the behavior of M_n under several sampling methods. To investigate this, Proposition 4.1 below considers the case where the empirical mean of the future samples $\bar{s}_r$ is exactly equal to its theoretical mean, denoted γ, e.g., $\gamma = \mu_n$ (the posterior mean) under our approach of sampling the future observations from the posterior predictive distribution. If the behavior of M_n for $\bar{s}_r = \gamma$ does not make sense then there is little hope that the corresponding sampling method is useful, and if it does make sense then the investigation may offer valuable insights. The proof of Proposition 4.1 is given in Appendix A by Jones et al. (2021).

Proposition 4.1. *Consider the conjugate Gaussian example, and suppose that $\bar{s}_r = \gamma$. Let z denote the nominal EPSS. Under these settings we have the following results:*

(a) (Posterior predictive sampling) If $\gamma = \mu_n$, then $M_n \geq z$.

(b) (Bootstrap sampling) If $\gamma = \bar{y}_n$, then there exists $\epsilon_l > \epsilon_s > 0$, such that $M_n = z$ whenever $|\bar{y}_n - \mu_n| < \epsilon_s$, and $M_n < 0$ whenever $|\bar{y}_n - \mu_n| > \epsilon_l$.

(c) (Prior sampling) If $\gamma = \mu_0$, then $M_n = z$.

Result (a) of Proposition 4.1 corresponds to our proposed method of sampling the future samples from the posterior predictive distribution (3.3)–(3.4). The result is consistent with the top right panel of Figure 1 in Section 2.3, which shows that the median (dashed curve) value of M_n is always equal to or greater than z. To gain further intuition consider the distance $W_2(m)$ under the conditions of Proposition 4.1:

$$W_2(m) = \left(\frac{n}{m}\right)^2 (\mu_n - \bar{y}_n)^2 + \left(\frac{\sigma}{\sqrt{m}} - \frac{\sigma}{\sqrt{n+z}}\right)^2. \tag{4.4}$$

Inspecting (4.4) reveals that the second term captures the nominal EPSS: setting $m = n + z$ makes the standard deviation of the baseline posterior π_m^b match that of the conjugate posterior q_n, so the second term of (4.4) equals zero, and thus giving z extra samples to the baseline prior minimizes the second term. However, the first term of (4.4) reveals that there is an intuitive reason for the value of M_n to often be larger than z: disagreements between the prior and the data as captured by $(\mu_n - \bar{y}_n)^2 = (z/(z+n))^2 (\mu_0 - \bar{y}_n)^2$ mean that the two posteriors will not be centered in the same location, and the $(n/m)^2$ term in (4.4) suggests that greater agreement is expected to be obtained by adding further samples to the baseline prior, i.e., by increasing m. Thus, our definition of M_n correctly identifies that simply reporting the classical information content of the prior as determined by its standard deviation is not sufficient: we must also take into account the impact of the prior location *relative to the data*. Of course, the results in Proposition 4.1 also take account of $\widetilde{W}_2(m)$, see Appendix A by Jones et al. (2021) for details.

Bayesian methodology stipulates that the extra samples must be drawn from the posterior predictive distribution, as above, but results (b) and (c) of Proposition 4.1 additionally reveal that several other natural approaches do not work well. Result (b) supposes that $\gamma = \bar{y}_n$ which corresponds to the case where the future observations are sampled from the empirical distribution (bootstrap) or from the baseline posterior predictive distribution, i.e., the posterior predictive distribution under π^b and conditional on only $\boldsymbol{y}$. The first part of result (b), $M_n = z$ for $\bar{y}_n \approx \mu_n$, is similar to what is seen in Figure 1. However, the second part of result (b), $M_n < 0$ for large $|\bar{y}_n - \mu_n|$, does not make sense in the current scenario firstly because $z > 0$ and secondly because intuitively the prior impact is large when $\bar{y}_n$ is far from μ_n. Reimherr et al. (2014) avoided this problem by defining the EPSS so that negative values convey a disagreement between the prior and the data, but there are limitations of their approach as discussed in Section 2.2. Furthermore, there is always *some* disagreement between the data and the prior so we find it conceptually more appealing to always have positive prior impact (unless our prior is less informative than the baseline).

Result (c) of Proposition 4.1 corresponds to the case where the additional samples are drawn from the conjugate prior distribution π: just as some might argue that the future data sampling method should not be "contaminated" by the prior, others may argue that it should not be "contaminated" by the data! Under the scenario of the proposition, this sampling scheme yields $M_n = z$, which is at least never negative. However, simply recovering the nominal EPSS regardless of the magnitude of $|\mu_0 - \bar{y}_n|$ does not convey differences in the impact of the prior, which is the purpose of having a measure of prior impact. In summary, drawing the extra samples from the posterior predictive distribution (3.3)–(3.4) seems to yield the most reasonable behavior.

4.2 Theoretical Posterior Distribution of the OPESS

To study the variation in the OPESS for a given observed dataset, we now derive the theoretical distribution of the OPESS conditional on $\boldsymbol{y}$ for the Gaussian conjugate posterior example discussed in Sections 2.3 and 3.3. More generally, the distribution of the OPESS will typically be hard to derive, but it can be empirically approximated, see Algorithm 1 Step 2.

Lemma 4.1 below gives the distribution of the distances $W_2(m)$ and $\widetilde{W}_2(m)$ conditional on $\bar{y}_n$ and μ (drawn from q_n, given by (3.7) in Algorithm 2). The proof is given in Appendix B by Jones et al. (2021). We condition on both $\bar{y}_n$ and μ because then the two distances are independent which facilitates derivation of the OPESS distribution. The distance distributions conditional on only $\bar{y}_n$ are given in Appendix D by Jones et al. (2021). Lemma 4.1 states that both distances follow shifted non-central χ^2 distributions, whose non-centrality parameters depend on μ through λ_m and δ_m (given in the lemma statement).

Lemma 4.1. *Conditional Distribution of Distances. Using the same notation as in Algorithm 2, assume that* $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$*, for* $i = n+1, \dots, m$*. Then we have*

$$\left[W_2\left(q_m^b, q_n\right) \middle| \bar{y}_n, \mu\right] \sim \tau_m \chi_1^2\left(\frac{\lambda_m}{\tau_m}\right) + c_m^2,$$

where

$$c_m^2 = \left(\frac{\sigma}{\sqrt{n+z}} - \frac{\sigma}{\sqrt{m}}\right)^2, \quad \tau_m = \frac{r}{m^2}\sigma^2,$$
$$\lambda_m = \left(\left(\frac{z}{n+z} - \frac{r}{m}\right)(\bar{y}_n - \mu) + (1 - w_n)(\mu - \mu_0)\right)^2;$$

and

$$\left[W_2\left(q_m, q_n^b\right) \middle| \bar{y}_n, \mu\right] \sim \kappa_m \chi_1^2\left(\frac{\delta_m}{\kappa_m}\right) + \tilde{c}_m^2,$$

where

$$\tilde{c}_m^2 = \left(\frac{\sigma}{\sqrt{m+z}} - \frac{\sigma}{\sqrt{n}}\right)^2, \quad \kappa_m = w_m^2 \tau_m,$$
$$\delta_m = \left(\frac{r+z}{m+z}(\bar{y}_n - \mu) + (1 - w_m)(\mu - \mu_0)\right)^2.$$

Furthermore, conditional on $\bar{y}_n$ *and* μ*,* $W_2\left(q_m^b, q_n\right)$ *and* $W_2\left(q_m, q_n^b\right)$ *are independent.*

Theorem 4.1 below gives the posterior distribution of the OPESS conditional on $\bar{y}_n$. The proof is given in Appendix C by Jones et al. (2021). In the theorem statement, v denotes a possible value of M_n (e.g., in the notation $P(M_n = v|\bar{y}_n)$), and t is a dummy variable for the distance corresponding to $M_n = v$, i.e., the distance $W_2(n+v)$, if $v \geq 0$, and the distance $\widetilde{W}_2(n+|v|)$, otherwise. The result gives a separate expression for the case $M_n = 0$ because when $m = n$ the distance between the posteriors (i.e., $W_2(n)$) is not random, meaning that an integral over the distance dummy variable t is not required. In the case $v \in \mathbb{Z}/\{0\}$, the integrands specified are tractable because the products are truncated at $M(t)$ and $\widetilde{M}(t)$ (defined in Appendix C by Jones et al. (2021)), which are finite for all values of $t \leq \sigma^2/(n+z)$.

Theorem 4.1. *OPESS distribution. Let* $F_{m,\mu}$ *and* $\widetilde{F}_{m,\mu}$ *denote the cumulative distribution function of* $\chi_1^2\left(\frac{\lambda_m}{\tau_m}\right)$ *and* $\chi_1^2\left(\frac{\delta_m}{\kappa_m}\right)$*, respectively. Furthermore, let* $h_{m,\mu}$ *and* $\tilde{h}_{m,\mu}$ *denote the conditional probability density function of* $W_2(m) = W_2\left(q_m^b, q_n\right)$ *and* $\widetilde{W}_2(m) = W_2\left(q_m, q_n^b\right)$*, respectively, as given in Lemma 4.1. Lastly, let* $g(t, \mu, v, M, \widetilde{M})$ *denote the function that gives* $P(\min_m(W_2(m), \widetilde{W}_2(m)) > t|\bar{y}_n, \mu)$ *multiplied by the appropriate den-*

sity for t, i.e.,

$$\begin{cases} \prod_{\substack{m=n+1 \\ m\neq v+n}}^{M(t)} (1 - F_{m,\mu}(t_m)) \prod_{m=n+1}^{\widetilde{M}(t)} (1 - \widetilde{F}_{m,\mu}(\tilde{t}_m)) h_{v+n,\mu}(t), & \text{if } v \in \mathbb{Z}_{>0}; \\ \prod_{m=n+1}^{M(t)} (1 - F_{m,\mu}(t_m)) \prod_{\substack{m=n+1 \\ m\neq |v|+n}}^{\widetilde{M}(t)} (1 - \widetilde{F}_{m,\mu}(\tilde{t}_m)) \tilde{h}_{|v|+n,\mu}(t), & \text{if } v \in \mathbb{Z}_{<0}; \\ \prod_{m=n+1}^{M(t)} (1 - F_{m,\mu}(t_m)) \prod_{m=n+1}^{\widetilde{M}(t)} (1 - \widetilde{F}_{m,\mu}(\tilde{t}_m)), & \text{if } v = 0, \end{cases}$$

where $t_m = (t - c_m^2)/\tau_m$ and $\tilde{t}_m = (t - \tilde{c}_m^2)/\kappa_m$, and M and $\widetilde{M}$ are known functions (see Appendix C by Jones et al. (2021)). Then $P(M_n = v|\bar{y}_n)$ is given by

$$\begin{cases} \int_{\mathbb{R}} \int_{T_v} 1_{\{W_2(n), \widetilde{W}_2(n), \sigma^2/(n+z) \geq t\}} g(t, \mu, v, M, \widetilde{M}) dt \, q(\mu|\bar{y}_n) d\mu, & \text{if } v \in \mathbb{Z}/\{0\}, \\ 1_{\{W_2(n) \leq \sigma^2/(n+z)\}} \int_{\mathbb{R}} g(W_2(n), \mu, 0, M, \widetilde{M}) q(\mu|\bar{y}_n) d\mu, & \text{if } v = 0, \end{cases}$$

where T_v is $\mathbb{R}_{\geq c_m^2}$ if $v > 0$ and $\mathbb{R}_{\geq \tilde{c}_m^2}$ otherwise, and $M(t)$ and $\widetilde{M}(t)$ are finite integers for all values of $t \leq \sigma^2/(n+z)$.

Figure 3 shows two examples of the *conditional* posterior distribution of the OPESS given $\bar{y}_n$ *and* μ, where $\bar{y}_n = \mu = 0$ in the left panel and $\bar{y}_n = \mu = 0.45 = 2\sigma/\sqrt{n}$ in the right panel. We plot the conditional posterior distribution to gain intuition about how the particular draw of μ from π impacts the conditional distribution of the OPESS. This is important because in reality the value of μ is fixed but unknown, and it is therefore valuable to understand how the distribution of the OPESS changes when we simulate the future samples based on different fixed choices of μ. In Figure 3, the line-dot density is a close Monte Carlo approximation to the theoretical conditional density of the OPESS given $\bar{y}_n$ and μ, and was obtained by simulating from the theoretical conditional distributions of $W_2(m)$ and $\widetilde{W}_2(m)$ and averaging the resulting values of the integrand given in Theorem 4.1 (except in the case $P(M_n = 0)$ for which no Monte Carlo approximation is needed). The red crosses show the empirical distribution of the OPESS obtained by directly applying the first two steps of Algorithm 1, except with the modification that μ is fixed in Step 2(a). Figure 3 illustrates that for $\bar{y}_n$ (and μ) farther from the prior mean $\mu_0 = 0$ (right panel) the conditional OPESS distribution has larger mean and is more right-skewed. This corroborates the numerical results seen in Figure 1. For some values of $\bar{y}_n$ and μ the conditional posterior distribution of the OPESS is bi-modal, with one mode at positive values and one at negative values (not shown). For other $\bar{y}_n$ and μ, there is a mode at $M_n = 0$, which is prominent in examples where the mOPESS is less than the nominal EPSS. Such scenarios are discussed further in Section 5.2.

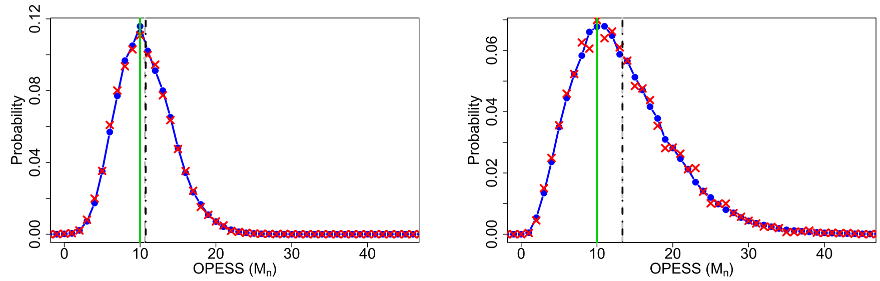

Figure 3: Conditional OPESS distribution given $\bar{y}_n = \mu = 0$ (left panel) and $\bar{y}_n = \mu = 0.45 = 2\sigma/\sqrt{n}$ (right panel). The line-dot line shows the theoretical and empirical distributions, respectively (see the main text for full details). The solid vertical line shows the nominal EPSS (10), and the dash-dot and dashed vertical lines show the mean conditional OPESS based on the theoretical and empirical distributions plotted, respectively.

5 Examples Beyond the Gaussian Conjugate Case

5.1 Non-Conjugate Gaussian Model

We now investigate the properties of Algorithm 1 when using a non-conjugate prior distribution for μ. The setting is identical to that outlined in Section 2.3, except that we set the informative prior to be a t-distribution, $\pi^t(\mu) \equiv \mathcal{T}(\mu|\nu, \mu_0, \lambda_0^2)$. The degrees-of-freedom parameter, ν, controls the heaviness of the tails, and μ_0 and λ_0 are location and scale parameters, respectively. We set $\mu_0 = 0$ and $\lambda_0^2 = 0.1$, and investigate two choices of ν, namely, $\nu = 4, 100$. As $\nu \to \infty$ the t-distribution density converges to a Gaussian density, so for large ν we expect the relationship between the mOPESS and $\bar{y}_n$ to be similar to that seen in the top left-hand panel of Figure 2. However, for relatively small values of ν, we expect the relationship between the mOPESS and $\bar{y}_n$ to be different, because in such cases both the posterior variance and the posterior skewness have non-negligible dependence on $\bar{y}_n - \mu_0$ (in contrast to the conjugate Gaussian example of Section 2.3).

The top-left panel of Figure 4 shows the relationship between the mOPESS and $\bar{y}$, for the t-distribution prior π^t. As in Section 2.3, the values of $\bar{y}$ were chosen to be the quantiles $(k - 0.5)/100$, for $k = 1, \ldots, 100$, of its distribution, namely a zero mean Gaussian with variance $\frac{1}{n}$. The relationship for $\nu = 100$, indicated by a solid red line, mimics the quadratic curve shown in the top-left panel of Figure 1, but we see a concave relationship for $\nu = 4$, represented by a dashed black line. The latter pattern shows that, for $\nu = 4$, the mOPESS is smaller when $|\bar{y}_n - \mu_0|$ is larger. This result can be explained by the fact that the posterior variance of μ increases with $|\bar{y}_n - \mu_0|$, see the upper right-hand panel of Figure 4 (dashed black line). We computed the posterior variance for each $(\nu, \bar{y}_n)$ pair with Monte Carlo integration using 5×10^5 posterior draws generated by `RStan` (Stan Development Team, 2020b,a). In the case of $\nu = 100$ (solid red line), the posterior variance of μ is nearly constant for all values of $|\bar{y}_n - \mu_0|$, and

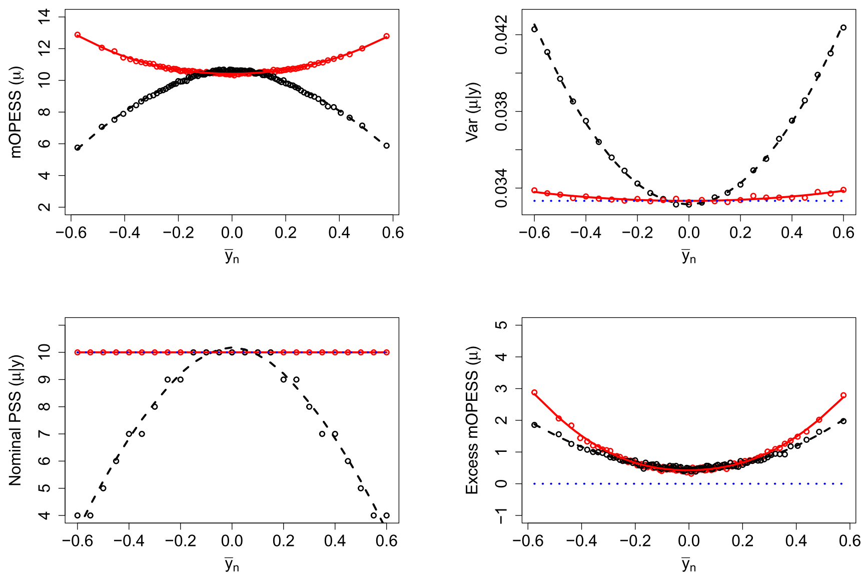

Figure 4: (Top left) mOPESS for $\pi^t \propto \mathcal{T}(\mu|\nu, 0, 0.1)$ with $y_i \sim \mathcal{N}(\mu, 1), i \in [1, \ldots, 20]$. (Top right) Posterior variance for μ under π^t. (Bottom left) Nominal EPSS of the equivalent conjugate Gaussian prior for μ. (Bottom right) Excess mOPESS under π^t. In all plots the black points and dashed line correspond to $\nu = 4$, while the red points and solid line corresponds to $\nu = 100$. The blue dotted line corresponds to a conjugate normal prior with nominal prior sample size of 10.

tracks the constant posterior variance under the prior $\pi(\mu) \equiv \mathcal{N}(\mu|0, 0.1)$, shown by the blue dotted line.

To gain further insight, we define the *equivalent* nominal EPSS of the t-distribution prior by matching posterior variances, i.e., we define it to be the nominal EPSS of the conjugate Gaussian prior π which yields the same posterior variance as under the t-distribution prior π^t. The bottom left panel of Figure 4 shows that when $\nu = 4$ and $\bar{y}_n = -0.6$, the equivalent nominal EPSS is 4 (dashed black line), whereas for $\bar{y}_n$ near 0 it is 10. In contrast, the equivalent nominal EPSS of a t-distribution prior with $\nu = 100$ (solid red line) is 10 for all $\bar{y}_n \in [-0.6, 0.6]$, as in the case of a conjugate Gaussian prior with $z = 10$ (shown as a dotted blue line overlapping with the solid red line).

The pattern described above suggests that subtracting the equivalent nominal EPSS from the mOPESS might shed more light on the comparison between the cases $\nu = 4$ and $\nu = 100$. We call the quantity resulting from this subtraction the *excess mOPESS*. The bottom right-hand panel in Figure 4 shows the excess mOPESS, and reveals that the excess increases with $|\bar{y}_n - \mu_0|$. Indeed, the relationship is now seen to be what we may have initially expected: a t-distribution prior with $\nu = 4$ (dashed black line) has a qualitatively similar, but smaller, impact than a t-distribution prior with $\nu = 100$ (solid red line).

A third factor that is unexplored in the tetraptych above is the posterior skewness of μ under the two values of ν. As $|\bar{y}_n - \mu_0|$ increases, so too does the skewness of the posterior for μ. Furthermore, the sensitivity of the posterior skewness to the quantity $\bar{y}_n - \mu_0$ is a function of the degrees of freedom parameter ν. Skewness of the posterior also impacts the OPESS. If we generate a μ from the tail of π^t_n, then M_n is nearly certain to be negative. For example, suppose that $\bar{y}_n < \mu_0$. Then our posterior π^t_n will be right-skewed. Consequently, Algorithm 1 will tend to generate more posterior draws from the region $\mu \ll \mathbb{E}[\mu|\bar{y}_n, \pi^t]$ than from the region $\mu \gg \mathbb{E}[\mu|\bar{y}_n, \pi^t]$. In the former region, $P(W_2(m) < \widetilde{W}_2(m)|\mu, \bar{y}_n, \pi^t)$ is small because the extra samples generated from the predictive distribution will decrease $W_2(q_m, q^b_n)$, but will increase $W_2(q^b_m, q_n)$, leading to a negative value of M_n. Because the magnitude of the skewness decreases with $|\bar{y}_n - \mu_0|$, smaller values of $|\bar{y}_n - \mu_0|$ yield fewer negative M_n values.

5.2 Beta-Binomial Model

Suppose $\{y_i, 1 \leq i \leq n\}$ are independent observations taking values in the set $\{0, 1\}$. The unknown parameter θ is the probability that $y_i = 1$. We set the informative prior to be $\pi(\theta) \equiv \text{Beta}(\alpha, \beta)$, where α, β are known hyperparameters, and the baseline prior to be $\pi^b(\theta) \equiv \text{Beta}(1, 1)$. As before, $x^{(m)}_i = y_i$ for $i = 1, \ldots, n$, and $x^{(m)}_i | \, \theta \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$ for $i = n+1, \ldots, m$ (but since θ is unknown it is drawn from its posterior distribution when computing the mOPESS, see Algorithm 1 Step 2(a)). Let q^A_n and q^B_m denote the posterior distribution using the original data $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ and the expanded dataset $\boldsymbol{x}^{(m)}$, respectively, under priors A and B. Also define $(F^A_n)^{-1}$ and $(F^B_m)^{-1}$ to be the quantile functions associated with these posterior densities. Then it can be shown that the 2-Wasserstein distance between q^A_n and q^B_m is $\left(\int_0^1 ((F^A_n)^{-1}(u) - (F^B_m)^{-1}(u))^2 du\right)^{1/2}$, see Theorem 2 in Cambanis et al. (1976). Unfortunately, this distance cannot be expressed in closed form in the case of Beta distributions, but it can be approximated to high precision using numerical integration, which is the approach we take.

In our simulations we set $\alpha = \beta = 5$, which corresponds to a nominal EPSS of $\alpha + \beta - 2 = 8$. The subtraction of 2 highlights that the standard nominal EPSS is relative to the prior sample size of the flat prior $\text{Beta}(1, 1)$, which is also our baseline prior π^b. To investigate the prior impact across different datasets, we sample 1,000 datasets of size $n = 20$ with replacement from the sex ratio dataset presented in Section 2.4 of Gelman et al. (2013). The sex ratio dataset consists of the biological sexes of 980 babies born to mothers with placenta previa: 437 of the babies are female, a proportion of 0.446 of the total.

The top left panel of Figure 5 shows the mOPESS values obtained across the 1,000 simulated datasets. The $\overline{M}_n$ estimates have a similar pattern as in the Gaussian conjugate model of Section 2.3, except that $\overline{M}_n$ is less than the nominal EPSS for datasets with $\bar{y}_n = 0.5$. The top right panel of Figure 5 shows that the median of the posterior distribution of M_n is similar to the mean. It also shows that the posterior distribution is much wider for datasets with means near 0.1. This is due to the fact that the posterior distribution of θ is strongly right skewed when $\bar{y}_n$ is much less than the prior mean of

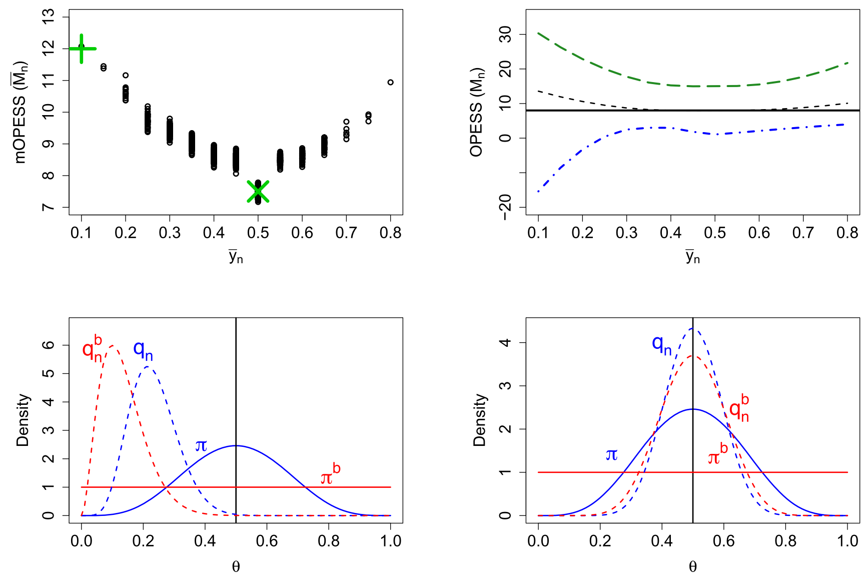

Figure 5: (Top left) mOPESS as a function of $\bar{y}_n$. Each point corresponds to one of the 1,000 simulated datasets. (Top right) Quantiles of the posterior distribution of M_n as a function of $\bar{y}_n$ including the median (dashed black curve), 95% quantile (long dash green curve), and 5% quantile (dash-dot blue curve). The horizontal solid line shows the nominal EPSS of 8. (Bottom left and right) Comparison of the posteriors q_n and q_n^b and their respective priors in the case where $\bar{y}_n = 0.1$ and $\bar{y}_n = 0.5$, indicated in the top left panel by a green "+" and cross, respectively.

0.5. The skewness is in turn reflected in the future observations $\boldsymbol{x}_{\mathrm{L}}^{\mathrm{all}}$, which are simulated conditional on a draw of θ, and this results in large posterior uncertainty for M_n. In particular, the right-skewness of q_n^b means that a large number of future samples can be needed to move the distribution to the right, resulting in some large OPESS values, and the right-skewness of q_n means that future samples to the left can substantially shift it, which results in some negative OPESS values. That posterior skewness is reflected in the resulting OPESS distribution is a strength of our approach. Indeed, the large spread in the posterior distribution of M_n corresponds to genuine uncertainty about the number of extra samples that need to be collected in order to minimize the distance between q_n and q_n (or q_m and q_n^b).

Next, we examine the relationship between the mOPESS and the nominal EPSS in two cases, namely, those where the prior mean and $\bar{y}_n$ are highly discrepant and perfectly aligned, respectively. The top left panel of Figure 5 indicates a simulated dataset for which $\bar{y}_n = 0.1$ (green "+"), and the bottom left panel shows the corresponding initial posterior distributions q_n and q_n^b (as well as the prior distributions). In this case, the mOPESS value is high (around 12) because the initial posteriors are very different. The bottom right panel of Figure 5 shows an analogous plot for a dataset with $\bar{y}_n = 0.5$, i.e., that labeled with green cross in the top left panel. In this case, the initial posteriors are very similar, which is why the mOPESS value is low (approximately 7.5, the variation across datasets is due to Monte Carlo error). In particular, the mOPESS value is *less* than the nominal EPSS of 8, a phenomenon that did not occur in the Gaussian conjugate model example of Section 2.3 for any of the datasets considered.

The low mOPESS value occurs here due to circumstances that arise, in this case, due to the discreteness of the data and the future data, as we now explain. The data mean $\bar{y}_n$ exactly matches the mean of the prior π, which in turn makes the means of q_n and q_n^b exactly equal. Thus, q_n and q_n^b are very similar to begin with, and it is unclear whether adding more samples to one of these posteriors will further reduce the distance between them. Adding more samples to the baseline posterior q_n^b could reduce the width of the distribution, and therefore may lead to greater agreement with q_n. However, the discrete nature of the data means that one additional sample with value one or zero will necessarily move the posterior mean away from 0.5, therefore potentially increasing the 2-Wasserstein distance between the two posteriors. Of course, if we draw an even number of extra samples then their average may be close to 0.5, so a reduction in the width of the baseline posterior q_n^b may be achieved without any substantial change in the mean. However, based on the nominal EPSS value, the approximate number of extra samples needed for matching the posterior widths is 8, but the probability of achieving an average of 0.5 (or very close to this) when drawing around 8 samples is not sufficiently high, and consequently the distance $W(n)$ is often smaller than $W_2(m)$ and $\widetilde{W}_2(m)$ for all $m > n$. Thus, for many simulations of $\boldsymbol{x}_{\mathrm{L}}^{\mathrm{all}}$, we have $M_n = 0$, meaning that the mOPESS $\overline{M}_n$ is shrunk towards zero.

In summary, the mOPESS may be less than the nominal EPSS when the means of the initial posteriors q_n and q_n^b are very similar relative to the size of z. In particular, adding extra samples to one posterior may not reduce the 2-Wasserstein distance between the two posteriors because: (i) if few extra samples are added then the variability in their mean can introduce discrepancies between the posterior means, and (ii) adding many extra samples will introduce discrepancies in the spreads since the initial discrepancy will be over-corrected. Thus, the smallest distance between the posteriors may often be achieved when $M_n = 0$, in which case $\overline{M}_n$ will be small in magnitude. Note that this phenomenon *can* occur in the Gaussian conjugate model example if $n \gg z$ (whereas in Section 2.3 we set $n = 2z$).

5.3 Simple Linear Regression Model

We now consider the setting of a simple linear regression model:

$$Y_i|\boldsymbol{\beta}, \omega_i \sim \mathcal{N}(\beta_1 + \beta_2\omega_i, \sigma^2), \quad \omega_i \sim \mathcal{N}(0, 1), \tag{5.1}$$

for $i = 1, \dots, n$, where σ^2 is known, and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ are the unknown model parameters. We note that in simple linear regression models, distributional assumptions on covariates are typically not made. We assume that the distribution of the covariates ω_i is known in order to simplify the algorithm for generating hypothetical samples. Let

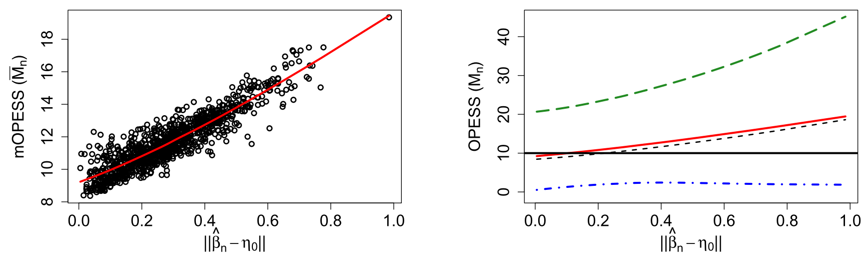

Figure 6: (Left) mOPESS as a function of $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\eta}_0\|_2$, where $\boldsymbol{\eta}_0 = \mathbf{0}$. Each point corresponds to one of the 1,000 simulated datasets. (Right) Quantiles of the posterior distribution of M_n as a function of the L_2-norm including the median (dashed curve), 95% quantile (long-dashed curve), and 5% quantile (dash-dot curve). The solid red line is the same as in the left panel and the horizontal line shows the nominal EPSS (10).

our informative prior $\pi(\boldsymbol{\beta})$ be:

$$\pi(\boldsymbol{\beta}) = \mathcal{N}\left(\boldsymbol{\eta}_0, \Sigma_0\right), \quad \text{where} \quad \Sigma_0 = \begin{bmatrix} \tau_1^2 & 0 \\ 0 & \tau_2^2 \end{bmatrix},$$

and $\boldsymbol{\eta}_0 = (\mu_0, \gamma_0)'$ and τ_1, τ_2 are known hyperparameters. Thus, the nominal EPSS for β_i is given by $\sigma^2/\tau_i^2 = z_i$, for $i \in [1, 2]$. We set the baseline prior to be $\pi^b(\boldsymbol{\beta}) \propto 1$. Define the m^{th} set of hypothetical samples as $\{(y_i^{(m)}, \omega_i^{(m)}), i \in \{1, \ldots, m\}\}$ with $\{(y_i^{(m)}, \omega_i^{(m)}) = (y_i, \omega_i), i \in \{1, \ldots, n\}\}$ for all m. For $i > n$, the hypothetical samples $(y_i^{(m)}, \omega_i^{(m)})$ are generated from (5.1) conditional on a draw of $\boldsymbol{\beta}$ from the posterior distribution q_n. Given that the model is composed of Gaussian conjugates, the posteriors $\{\pi_u(\boldsymbol{\beta}), \pi_u^b(\boldsymbol{\beta}), n \leq u \leq L\}$ are also Gaussian. Closed expressions for the posterior distributions and corresponding Wasserstein distances are given in Appendix E by Jones et al. (2021). Thus, it is straightforward to apply Algorithm 1 to compute the mOPESS.

Our linear regression model simulation study is similar in design to that for the Beta-Binomial model in Section 5.2. We observe $n = 20$ samples from the model (5.1), with $\sigma^2 = 1$, and $\beta_1 = \beta_2 = 0$. We set $z_1 = z_2 = 10$, so the nominal EPSS of π is 10.

As can be seen in Figure 6, the mOPESS increases with $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\eta}_0\|_2$, i.e., the L_2-norm of the ordinary least squares estimator less the prior mean. We use a one-dimensional summary of the two-dimensional measure $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\eta}_0$ to ease visualization, and to account for the fact that the joint prior can be influential even if only one element of $\hat{\boldsymbol{\beta}}_n$ disagrees with the corresponding marginal prior.

Indeed, Figure 7 shows how the mOPESS can vary even when conditioning on a small interval of $(\hat{\boldsymbol{\beta}}_n)_{[1]} - \mu_0$, or $(\hat{\boldsymbol{\beta}}_n)_{[2]} - \gamma_0$. In the left panel, we see that for values of $(\hat{\boldsymbol{\beta}}_n)_{[1]} - \mu_0$ that are near zero, there is still quite a range of mOPESS values. Thus the mOPESS is influenced not only by $(\hat{\boldsymbol{\beta}}_n)_{[1]} - \mu_0$, but also the disagreement between γ_0 and $(\hat{\boldsymbol{\beta}}_n)_{[2]}$. For example, the maximum mOPESS value in the left panel, indicated by the cross, occurs at a value of $(\hat{\boldsymbol{\beta}}_n)_{[1]} - \mu_0$ which is not extreme. Turning to the

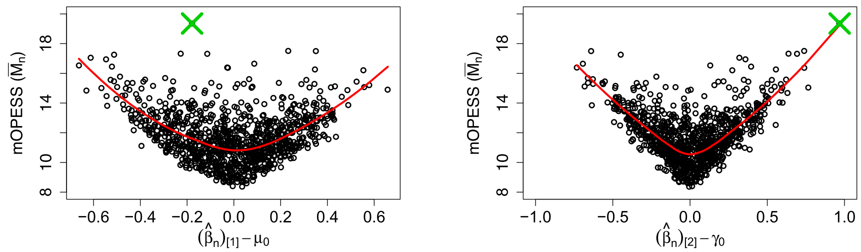

Figure 7: (Left) mOPESS as a function of the of $(\hat{\boldsymbol{\beta}}_n)_{[1]} - \mu_0$ (in this case $\mu_0 = 0$). (Right) mOPESS ($\overline{M}_n$) as a function of $(\hat{\boldsymbol{\beta}}_n)_{[2]} - \gamma_0$ (in this case $\gamma_0 = 0$). Each point corresponds to one of the 1,000 simulated datasets. The cross symbol indicates the maximum observed mOPESS across all 1,000 simulation studies, ≈ 19.

right panel in Figure 7, which shows the mOPESS versus $(\hat{\boldsymbol{\beta}}_n)_{[2]} - \gamma_0$, we see that the maximum mOPESS occurs at the maximum observed value of $(\hat{\boldsymbol{\beta}}_n)_{[2]} - \gamma_0$ (because the maximum value of this discrepancy is large compared to that for $(\hat{\boldsymbol{\beta}}_n)_{[1]} - \mu_0$). In summary, the mOPESS generalizes to two dimensions as we would expect it to, with joint dependence on $\hat{\boldsymbol{\beta}}_n - \eta_0$.

6 Summary

In this paper, we have proposed the *mean Observed Prior Effective Sample Size* (mOPESS) as a measure of the impact of the prior distribution on the Bayesian analysis at hand. Our measure is different from other methods proposed in literature in that we condition on the observed data, instead of averaging over the data. Furthermore, we do not rely on asymptotic results, meaning that our method can be applied to small or moderate sample size settings, where the prior impact is largest and of most interest in practice.

Supplementary Material

Appendix for "Quantifying Observed Prior Impact" (DOI: 10.1214/21-BA1271SUPP; .pdf).

References

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. 747

Berger, J., Berliner, L. M., et al. (1986). "Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors." *The Annals of Statistics*, 14(2): 461–486. MR0840509. doi: https://doi.org/10.1214/aos/1176349933. 740

Berger, J. O. (1990). "Robust Bayesian analysis: sensitivity to the prior." *Journal of statistical planning and inference*, 25(3): 303–328. MR1064429. doi: https://doi.org/10.1016/0378-3758(90)90079-A. 741

Berger, J. O., Bernardo, J. M., Sun, D., et al. (2015). "Overall objective priors." *Bayesian Analysis*, 10(1): 189–221. MR3420902. doi: https://doi.org/10.1214/14-BA915. 741

Bousquet, N. (2008). "Diagnostics of prior-data agreement in applied Bayesian analysis." *Journal of Applied Statistics*, 35(9): 1011–1029. MR2522125. doi: https://doi.org/10.1080/02664760802192981. 740

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of markov chain monte carlo*. CRC press. MR2742422. doi: https://doi.org/10.1201/b10905. 747

Cambanis, S., Simons, G., and Stout, W. (1976). "Inequalities for $Ek(X, Y)$ when the marginals are fixed." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4): 285–294. MR0420778. doi: https://doi.org/10.1007/BF00532695. 756

Chen, Y., Meng, X.-L., Wang, X., van Dyk, D. A., Marshall, H. L., and Kashyap, V. L. (2019). "Calibration concordance for astronomical instruments via multiplicative shrinkage." *Journal of the American Statistical Association*, 114(527): 1018–1037. MR4011755. doi: https://doi.org/10.1080/01621459.2018.1528978. 738

Clarke, B. (1996). "Implications of reference priors for prior information and for sample size." *Journal of the American Statistical Association*, 91(433): 173–184. MR1394071. doi: https://doi.org/10.2307/2291393. 738, 741, 742, 747

Clarke, B., Yuan, A., et al. (2006). "Closed form expressions for Bayesian sample size." *Annals of statistics*, 34(3): 1293–1330. MR2278359. doi: https://doi.org/10.1214/009053606000000308. 740

Consonni, G., Fouskakis, D., Liseo, B., Ntzoufras, I., et al. (2018). "Prior distributions for objective Bayesian analysis." *Bayesian Analysis*, 13(2): 627–679. MR3807861. doi: https://doi.org/10.1214/18-BA1103. 741

Efron, B. and Hinkley, D. V. (1978). "Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information." *Biometrika*, 65(3): 457–483. MR0521817. doi: https://doi.org/10.1093/biomet/65.3.457. 739

Evans, M., Moshonov, H., et al. (2006). "Checking for prior-data conflict." *Bayesian Analysis*, 1(4): 893–914. MR2282210. doi: https://doi.org/10.1016/j.spl.2011.02.025. 740

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC. MR3235677. 756

Ghosh, M., Mergel, V., and Liu, R. (2011a). "A general divergence criterion for prior selection." *Annals of the Institute of Statistical Mathematics*, 63(1): 43–58. MR2748933. doi: https://doi.org/10.1007/s10463-009-0226-4. 741

Ghosh, M. et al. (2011b). "Objective priors: An introduction for frequentists." *Statistical*

Science, 26(2): 187–202. MR2858380. doi: https://doi.org/10.1214/10-STS338. 741

Gupta, K., Attri, J., Singh, A., Kaur, H., and Kaur, G. (2016). "Basic concepts for sample size calculation: critical step for any clinical trials!" *Saudi journal of anaesthesia*, 10(3): 328. 740

Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013). "Adaptive adjustment of the randomization ratio using historical control data." *Clinical Trials*, 10(3): 430–440. 740

Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). "Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models." *Bayesian Analysis (Online)*, 7(3): 639. MR2981631. doi: https://doi.org/10.1214/12-BA722. 740

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. 740

Jones, D. E., Trangucci, R, N., and Chen, Y. (2021). "Appendix for "Quantifying Observed Prior Impact"." *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1271SUPP. 750, 751, 752, 753, 759

Kass, R. E. and Wasserman, L. (1996). "The selection of prior distributions by formal rules." *Journal of the American Statistical Association*, 91(435): 1343–1370. 741

Leisen, F., Villa, C., Walker, S. G., et al. (2020). "On a Class of Objective Priors from Scoring Rules (with Discussion)." *Bayesian Analysis*, 15(4): 1345–1423. MR4194270. doi: https://doi.org/10.1214/19-BA1187. 741

Lin, X., Pittman, J., and Clarke, B. (2007). "Information conversion, effective samples, and parameter size." *IEEE transactions on information theory*, 53(12): 4438–4456. MR2446916. doi: https://doi.org/10.1109/TIT.2007.909168. 738

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media. MR2401592. 747

Marin, J.-M. and Robert, C. (2007). *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Science & Business Media. MR2289769. 747

Morita, S., Thall, P. F., and Müller, P. (2008). "Determining the effective sample size of a parametric prior." *Biometrics*, 64(2): 595–602. MR2432433. doi: https://doi.org/10.1111/j.1541-0420.2007.00888.x. 738, 740, 742, 750

Morita, S., Thall, P. F., and Müller, P. (2010). "Evaluating the impact of prior assumptions in Bayesian biostatistics." *Statistics in biosciences*, 2(1): 1–17. 738

Neuenschwander, B., Weber, S., Schmidli, H., and O'Hagan, A. (2020). "Predictively consistent prior effective sample sizes." *Biometrics*, 76(2): 578–587. MR4125280. doi: https://doi.org/10.1111/biom.13252. 739

Nott, D. J., Seah, M., Al-Labadi, L., Evans, M., Ng, H. K., Englert, B.-G., et al. (2021).

"Using prior expansions for prior-data conflict checking." *Bayesian Analysis*, 16(1): 203–231. MR4194279. doi: https://doi.org/10.1214/20-BA1204. 740

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. 740

Reimherr, M., Meng, X.-L., and Nicolae, D. L. (2014). "Prior sample size extensions for assessing prior impact and prior-likelihood discordance." *arXiv preprint* arXiv:1406.5958. 737, 738, 739, 740, 742, 747, 750, 751

Roos, M., Held, L., et al. (2011). "Sensitivity analysis in Bayesian generalized linear mixed models for binary data." *Bayesian Analysis*, 6(2): 259–278. MR2806244. doi: https://doi.org/10.1214/11-BA609. 741

Roos, M., Martins, T. G., Held, L., Rue, H., et al. (2015). "Sensitivity analysis for Bayesian hierarchical models." *Bayesian Analysis*, 10(2): 321–349. MR3420885. doi: https://doi.org/10.1214/14-BA909. 741

Sahu, S. and Smith, T. (2006). "A Bayesian method of sample size determination with practical applications." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(2): 235–253. MR2225541. doi: https://doi.org/10.1111/j.1467-985X.2006.00408.x. 740

Sason, I. and Verdú, S. (2016). "f-divergence Inequalities." *IEEE Transactions on Information Theory*, 62(11): 5973–6006. MR3565096. doi: https://doi.org/10.1109/TIT.2016.2603151. 747

Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F., and Schmitzer, B. (2019). *transport: Computation of Optimal Transport Plans and Wasserstein Distances*. R package version 0.12-1. 746

Stan Development Team (2020a). "RStan: the R interface to Stan." R package version 2.21.2. URL http://mc-stan.org/. 754

Stan Development Team (2020b). *Stan Modeling Language Users Guide and Reference Manual*. 754

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press. MR1652247. doi: https://doi.org/10.1017/CBO9780511802256. 749

Walter, G. and Augustin, T. (2009). "Imprecision and prior-data conflict in generalized Bayesian inference." *Journal of Statistical Theory and Practice*, 3(1): 255–271. MR2667666. doi: https://doi.org/10.1080/15598608.2009.10411924. 740

Wang, F., Gelfand, A. E., et al. (2002). "A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models." *Statistical Science*, 17(2): 193–208. MR1925941. doi: https://doi.org/10.1214/ss/1030550861. 740

Weiss, R. (1996). "An approach to Bayesian sensitivity analysis." *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4): 739–750. MR1410188. 741

Wiesenfarth, M. and Calderazzo, S. (2019). "Quantification of prior impact in terms of effective current sample size." *Biometrics*. MR4098565. doi: https://doi.org/10.1111/biom.13124. 738

Wiesenfarth, M. and Calderazzo, S. (2020). "Quantification of prior impact in terms of effective current sample size." *Biometrics*, 76(1): 326–336. MR4098565. doi: https://doi.org/10.1111/biom.13124. 740

Acknowledgments

The authors thank Prof. Xiao-Li Meng from Harvard University for helpful discussions and Dr. Vinay Kashyap from the Harvard-Smithsonian Center for Astrophysics (CfA) for collaborating on the astronomical instrument calibration problem.