

# Differentially private Bayesian tests

Abhisek Chakraborty, Saptati Datta

**Abstract.** Differential privacy has emerged as a significant cornerstone in the realm of scientific hypothesis testing utilizing confidential data. When data are not confidential, Bayesian tests are widely used in reporting scientific discoveries, as they effectively address the key criticisms of p-values, namely lack of interpretability and inability to quantify evidence in support of competing hypotheses. In this article, we introduce a novel framework for differentially private Bayesian hypothesis testing, thereby expanding the applicability of Bayesian testing to confidential data. This framework naturally arises from a principled data-generative mechanism, ensuring that the resulting inferences retain interpretability while maintaining privacy. Further, by focusing on differentially private Bayes factors based on test statistics, we circumvent the need to model the complete data generative mechanism and ensure substantial computational benefits. We also provide a set of sufficient conditions to establish Bayes factor consistency under the proposed framework. Finally, the utility of the proposed methodology is showcased via several numerical experiments.

**MSC2020 subject classifications:** Primary 60K35, 60K35; secondary 60K35.

**Keywords:** Bayes Factors, Bayes Factor Consistency, Differential Privacy, Interpretable ML, Laplace Mechanism.

## 1 Introduction

Hypothesis testing is an indispensable tool to answer scientific questions in the context of clinical trials, bioinformatics, social sciences, etc. The data within such domains often involves sensitive and private information pertaining to individuals. Researchers often bear legal obligations to safeguard the privacy of such data. In this context, differential privacy (Dwork, 2006) has emerged as a popular framework for ensuring privacy in statistical analyses with confidential data. Consequently, differentially private versions of numerous commonly used hypothesis tests have been developed, although exclusively from a frequentist view point. This encompasses private adaptations of test of binomial proportions (Awan and Slavković, 2018), significance in linear regression (Alabi and Vadhan, 2023), goodness of fit (Kwak et al., 2021), analysis of variance (Swanberg et al., 2019), high-dimensional normal means (Narayanan, 2022), to name a few. Differentially private versions of common non-parametric tests (Couch et al., 2019), permutation tests (Kim and Schrab, 2023), etc., have emerged as well. Another recent line of work proposes to automatically create private versions of the existing frequentist tests in a black-box

---

arXiv: 2010.00000

\*Department of Statistics, Texas A&M University, College Station, TX, USA

†Both the authors contributed equally.

‡There was no external or internal funding for this project.

fashion (Kazan et al., 2023; Peña and Barrientos, 2023), utilizing the subsample-and-aggregate method, with the aggregation done by the uniformly most powerful binomial test (Awan and Slavković, 2018).

In this article, we intend to complement the existing literature via providing a novel generalised framework for differentially private Bayesian testing. This is crucial since p-values obtained from frequentist tests are routinely criticised for its lack of transparency, inability to quantify evidence in support of the null hypothesis, and inability to measure the importance of a result. Readers may refer to the American Statistical Association statement on statistical significance and p-values Wasserstein and Lazar (2016) for further discussions on this. On the other hand, Bayes factors (Kass and Raftery, 1995; Jeffreys, 1961; Morey and Rouder, 2011; Rouder et al., 2009; Wagenmakers et al., 2010; Johnson et al., 2023a) can quantify the relative evidence in favor of competing hypotheses, effectively addressing the persistent criticisms of p-values. This automatically calls for development of differentially private Bayes Factors to test the scientific hypotheses of interest with confidentiality guarantees.

Prior to presenting the proposed methodology, we provide a concise overview of pertinent concepts in Bayesian hypothesis testing and differential privacy, laying the groundwork for subsequent discussions.

## 1.1 Bayesian hypothesis testing

Suppose we intend to test the null hypothesis  $H_0 : \theta \in \Theta_0$  against the alternative hypothesis  $H_1 : \theta \in \Theta_1$ . The Bayes Factor (Kass and Raftery, 1995; Jeffreys, 1961; Morey and Rouder, 2011; Rouder et al., 2009; Wagenmakers et al., 2010; Johnson et al., 2023a) in favour of the alternative hypothesis  $H_1$  against the null hypothesis  $H_0$  is defined as

$$BF_{10} = \frac{P(H_1 | x)}{P(H_0 | x)} = \frac{m_1(x)}{m_0(x)} \times \frac{P(H_1)}{P(H_0)}, \quad (1.1)$$

where  $P(H_i | x)$  is the posterior probability of  $H_i$  given the data  $x$ ,  $P(H_i)$  is the prior probability assigned to  $H_i$ , and  $M_i(x)$  denotes the marginal probability assigned to the data under  $H_i$ ,  $i = 0, 1$ . The marginal densities are computed via

$$m_i(x) = \int_{\Theta_i} \pi(x | \theta) \pi_i(\theta) d\theta,$$

where  $\pi(x | \theta)$  is the data generative model given the parameter  $\theta$ , and  $\pi_i(\theta)$  is the prior density for  $\theta$  under  $H_i$ ,  $i = 0, 1$ . Bayes factor effectively addresses the persistent criticisms regarding lack of interpretability of p-values (Wasserstein and Lazar, 2016) in reporting outcomes of hypothesis tests, since it directly quantifies the relative evidence in favor of the competing hypotheses  $H_0$  and  $H_1$ . Noteworthy, the value of a Bayes factor depends on the choices of  $\pi_i(\theta)$ ,  $i = 0, 1$ , and it is generally difficult to justify any single default choice. Moreover, the expression of  $M_i(x)$ ,  $i = 0, 1$  may not be available in closed form, and require computing a potentially high-dimensional numerical integral. Several modifications to the existing Bayes factor methodology are proposed to improve

the reporting of scientific findings. Two such modifications of traditional Bayes factors, that will be utilized to construct differentially private Bayesian tests, are Bayes factors constructed based on common test statistics (BFBOTS) (Johnson, 2005) and Bayes factors reported as a function of effect size (BFFs) (Johnson et al., 2023b). We briefly review both the ideas next.

## 1.2 Bayes factor based on test statistics

To circumnavigate the issues with traditional Bayes factors, Johnson (2005) proposed to define Bayes factors based on common test statistics, and put priors directly on the non-centrality parameter of the test statistics to specify the various competing hypotheses. Bayes factors are first defined in terms of standard  $z$ ,  $t$ ,  $\chi^2$ , and  $F$  test statistics. The distribution of these test statistics under the null hypothesis is known. Under the alternative hypotheses, their asymptotic distributions are common non-central distributions, characterized solely by scalar-valued non-centrality parameters. As a result, the specification of the prior density under the alternative hypothesis is simplified, and no prior density is required under the null hypothesis. For example, suppose we observe data  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , and want to test  $H_0 : \mu = 0$  against  $H_1 : \mu \neq 0$ . In this context, under  $H_0$ , the  $t$ -statistic  $t = \frac{\sqrt{n}\bar{X}}{\hat{\sigma}}$  follows a student's  $t_{n-1}(0)$  distribution with degrees of freedom  $n - 1$ , where  $\bar{X}$  and  $\hat{\sigma}$  are the mean and standard deviation of the observed data. Under  $H_1$  the  $t$ -statistic  $t = \frac{\sqrt{n}\bar{X}}{\hat{\sigma}}$  follows a (non-central)  $t_{n-1}(\delta)$  distribution with non centrality parameter  $\delta \geq 0$  and degrees of freedom  $n - 1$ . In other words, the non centrality parameter  $\delta = 0$  under  $H_0$  and  $\delta > 0$  under  $H_1$ . So, we can elicit priors directly on the non-centrality parameter  $\delta$  under  $H_1$ , and conduct a Bayesian testing. Moreover, if we utilise non-local priors (Johnson and Rossell, 2010a) on  $\delta$  under  $H_1$  instead of local priors, not only the expression of the marginal likelihoods are often available in closed form, but also it ensures quick accumulation of evidence in favor of both true  $H_0$  and true  $H_1$ . Data-driven approaches for the prior hyper-parameter tuning (Johnson et al., 2023a; Datta et al., 2024) were subsequently suggested, adding to the objectivity of such Bayesian testing methodology.

## 1.3 Bayes factor functions(BFFs)

In Johnson et al. (2023b), the authors defined Bayes factor functions (BFFs) as a mapping from standardized effect sizes of interest to corresponding Bayes factors, or more precisely, as a mapping from prior densities centered on the standardized effect sizes of interest to the corresponding Bayes factors. For a given value of the test statistic, a range of Bayes factors based on the test statistic value is calculated by varying the prior densities elicited on the non-centrality parameter that defines the alternative hypothesis. In other words, the families of prior densities used to define these Bayes factors are indexed by the standardized effect size. Through this mapping, BFFs make the relationship between Bayes factors and prior assumptions more transparent by providing a clear interpretation of Bayes factors as a function of prior densities used in their formulation. Additionally, another crucial feature of BFFs relevant to us is that the BFFs enable the accumulation of evidence across multiple studies examining the same phenomenon.

To elucidate the construction of a BFF, let us reconsider a two-sided  $t$ -test as previously discussed. Let  $X_1, \dots, X_n$  be independent and identically distributed as  $N(\mu, \sigma^2)$ , where  $\sigma^2$  is unknown, and define the test statistic as  $t = \sqrt{n}\bar{x}/\sigma$ . Under this set up,  $t \sim t_{n-1}(\sqrt{n}\pi^*)$ , where  $\pi^* = \mu/\sigma$  represents the standardized effect, and  $\lambda = \sqrt{n}\pi^*$  is the non-centrality parameter associated with the test statistic. Under the null hypothesis, we have  $H_0 : \lambda = 0$ , and no prior specification on  $\lambda$  is required. Under the alternative hypothesis  $H_1 : \lambda \neq 0$ , we may specify a normal-moment prior density (Johnson and Rossell, 2010b), as follows:

$$H_1 : t \mid \lambda \sim t_{n-1}(\lambda), \quad \lambda \sim j(\lambda \mid \tau^2),$$

where

$$j(\lambda \mid \tau^2) = \frac{\lambda^2}{\sqrt{2\pi}\tau^3} \exp\left(-\frac{\lambda^2}{2\tau^2}\right), \quad \lambda \in \mathcal{R},$$

and  $\tau^2$  is a prior hyper-parameter. The mode of the prior density  $j(\lambda \mid \tau^2)$  occurs at  $\pm\sqrt{2}\tau$ . For a given value of standardized effect size  $\pi^*$ , if we set  $\tau^2 = n\pi^{*2}/2$ , then the mode of the prior density equals  $\sqrt{n}\pi^*$ . Consequently, the mode of the prior density aligns with the non-centrality parameter  $\lambda$  of the test statistic. Utilizing this relationship, Bayes Factor Functions (BFFs) are developed to associate standardized effect sizes with corresponding Bayes factors. Specifically, for a given standardized effect size, the Bayes factors are computed using alternative prior densities that are centered on non-centrality parameters aligned with the specified standardized effect size. We next provide a succinct overview of relevant concepts from differential privacy literature.

## 1.4 Differential privacy

Differential privacy (Dwork, 2006; Baraheem and Yao, 2022) presents a principled framework for safeguarding sensitive information in confidential data utilized for statistical analyses. It achieves this by ensuring that the probability of any specific output of the statistical procedure is nearly the same, regardless of the data belonging any one specific individual in the database. In this context, databases that vary only with respect to the data corresponding to a single individual are termed *neighboring* databases.

**Definition 1** (Differential Privacy, (Dwork, 2006)). *A randomized algorithm  $f : \mathcal{D} \rightarrow \mathcal{S}$  is  $\varepsilon$ -differentially private if for all possible outcomes  $S \subseteq \mathcal{S}$  and for any two neighbouring databases  $D, D' \in \mathcal{D}$ , we have  $\mathbb{P}[f(D) \in S] \leq e^\varepsilon \mathbb{P}[f(D') \in S]$ .*

Further, the measure of how much the modification of a single row in a database can influence the output of a query is referred to as *global sensitivity*.

**Definition 2** (Global sensitivity). *The global sensitivity of a function  $f$  is  $\text{GS}_f = \max_{D, D' \in \mathcal{D}} |f(D) - f(D')|$ , where  $D$  and  $D'$  are neighboring databases.*

The *privacy budget*  $\varepsilon$  quantifies the degree of privacy protection, with smaller values of  $\varepsilon$  indicating a stronger privacy guarantee. In particular,  $\varepsilon$ -differential privacy ensures that the privacy loss is bounded by  $\varepsilon$  almost surely, which implies that the

ratio of the output probabilities under neighboring datasets is constrained by  $e^\epsilon$  for all outcomes. As  $\epsilon$  becomes smaller, the privacy of the mechanism  $f$  increases, since the output distributions under neighboring datasets  $f(D)$  and  $f(D')$  are forced to be similar (Barrientos et al., 2019). Every differentially private algorithm achieves this via incorporating an element of randomness. One of the most widely adopted such technique is the Laplace mechanism (Dwork, 2006), that involves introducing a privacy noise sampled from Laplace distribution to the result of the query that requires privacy protection. The scale of the Laplace distribution employed to introduce the privacy noise, is contingent on both the privacy budget  $\epsilon$  and the global sensitivity  $GS_f$ . Given any function  $f$ , the *Laplace mechanism* is defined as

$$\tilde{f}(D) = f(D) + \eta,$$

where  $\eta$  is a privacy noise drawn from  $\text{Laplace}(0, GS_f/\epsilon)$ , and  $GS_f$  is the global sensitivity of  $f$ . We now have reviewed all the necessary tools to introduce our methodology.

## 1.5 Relevant literature

There has been growing interest in developing Bayesian inferential methods under the differential privacy (DP) framework. A notable contribution in this area is by Peña and Barrientos (2023), who proposed a differentially private hypothesis test based on a subsample-and-aggregate scheme. While their method primarily aligns with frequentist principles—focusing on assessing null hypothesis fit—they also extend it to the Bayesian paradigm by computing posterior odds based on the distribution of the test statistics they introduce. However, due to the nature of their test statistic, the resulting Bayes factor lacks consistency under the alternative hypothesis. Additionally, the method relies on a subjective prior for the test’s power, rendering the Bayes factor sensitive to this prior specification. To address these limitations, Peña and Barrientos (2024) introduced an alternative differentially private test that uses truncated likelihood ratio statistics or truncated Bayes factor, effectively bounding the global sensitivity of the Bayes factor. While this approach improves privacy guarantees, it raises interpretability concerns, as the truncation is not derived from the underlying generative model.

In the context of linear regression, Amitai and Reiter (2018) proposed two approaches for Bayesian inference based on posterior probabilities of regression coefficients. The first approach assumes a normal approximation to the posterior and requires unbiased posterior estimates of the coefficients, making it suitable primarily for large sample sizes. The second approach aggregates posterior probabilities using Fisher’s method and assumes that the resulting statistic follows a  $\chi^2$  distribution. This assumption fails under the alternative hypothesis. Moreover, their method involves clipping posterior probabilities at a fixed threshold, which may suppress evidence against the null. Importantly, their work does not focus specifically on Bayesian hypothesis testing.

Other notable contributions include Bernstein and Sheldon (2019), who proposed a differentially private method for posterior sampling of regression parameters, although their work does not directly address hypothesis testing. In parallel, Dimitrakakis et al. (2017); Heikkilä et al. (2019); Hu et al. (2023) developed various techniques for drawing

posterior samples under differential privacy constraints. These methods typically rely on the assumption of a bounded likelihood and require a privacy budget that scales with the number of posterior samples. Consequently, when the privacy budget is small, such approaches may yield unreliable Monte Carlo approximations.

Drawing inspiration from these works, we propose an objective Bayesian testing procedure that is consistent under the true hypothesis.

## 1.6 Our contributions

The key contributions in the article are three-fold. First, to the best of our knowledge, we introduce the first-ever differentially private objective Bayesian testing framework and ensure the consistency of the Bayes factor under the true model. This development is important in light of the American Statistical Association’s statement on statistical significance and p-values (Wasserstein and Lazar, 2016), which warns against making scientific decisions based solely on p-values, due to their lack of transparency, inability to quantify evidence supporting the null hypothesis, and failure to measure the practical significance of a result. While Bayes factor-based tests offer a solution to many of the limitations associated with p-values, extending these tests to a differentially private setting—while preserving their inherent interpretability—presents substantial challenges. To address this, we integrate widely-used strategies in differentially private testing, such as sub-sample and aggregate methods and truncated test functions, within the proposed data-generative mechanism. This careful integration ensures that the resulting Bayesian testing procedures emerge naturally from the data generation models, maintaining the interpretability of the resulting inferences.

Second, we introduced differentially private Bayes factors that rely on widely-used test statistics, extending the work of Johnson (2005) to a privacy-preserving framework. In this approach, we place priors directly on the non-centrality parameter of the test statistic and employ data-driven techniques for tuning prior hyperparameters. This strategy eliminates the need to model the entire data-generating process and avoids placing priors on potentially high-dimensional model parameters. As a result, the method offers significant computational advantages when calculating marginal likelihoods, which typically involve intractable integrals in fully parametric Bayesian testing frameworks. Furthermore, we establish a set of sufficient conditions to ensure the consistency of the Bayes factor (Chib and Kuffner, 2016; Chatterjee et al., 2020) within the proposed framework.

Finally, we critically leverage the proposed Bayes factor based on test statistics framework, to offer a general template for constructing size- $\alpha$  differentially private Bayesian tests that adhere to a predefined privacy budget, provided the investigator has pre-specified the effect size of interest. Additionally, we propose numerical schemes for data-driven hyperparameter tuning. Specific examples under particular hypothesis testing scenarios are presented. Nevertheless, the applicability of the proposed schemes extends beyond the Bayes factors based on test statistics discussed in this article.

In a nutshell, the key differences between our contribution and the existing works of Peña and Barrientos (2023); Peña and Barrientos (2024) and Amitai and Reiter (2018)

are as follows:

1. We introduce Bayes factors for hypothesis testing within an objective Bayesian framework, which eliminates the dependence on subjective prior elicitation. This ensures that the resulting Bayes factors are interpretable and mitigates the sensitivity to prior specifications that is present in the approaches of [Peña and Barrientos \(2023\)](#).
2. Unlike [Peña and Barrientos \(2024\)](#), our method does not require any direct truncation scheme to guarantee the boundedness of the General Sensitivity Parameter (GSS). This property is inherently satisfied by our model specification and prior choice. As a result, we retain full interpretability of the Bayes factor, which may be compromised in [Peña and Barrientos \(2024\)](#) due to the non-generative truncation required to bound the GSS. Consequently, the Bayes factor in their work can be hard to interpret.
3. We evaluate the Bayes factor over a sequence of alternative hypotheses indexed by the effect size of interest, which introduces additional flexibility.
4. As outlined in our general framework, presented in Section 2, the proposed method extends well beyond test statistic-based Bayes factors. We emphasize Bayes factors based on test statistics primarily to avoid subjective priors on nuisance parameters. We use sufficient statistics to define Bayes factors, thereby ensuring there is no potential information loss. However, our general framework can encompass a broader class of models.
5. In contrast to [Amitai and Reiter \(2018\)](#), our primary objective is hypothesis testing for the purpose of model selection. We aim to achieve both interpretability and consistency in model selection, both under the null and alternative hypotheses.

Rest of the article is organized as follows. Section 2 introduces the proposed framework for differentially private Bayesian testing in complete generality, and presents the key features of the methodology. In Section 3, we introduce differentially private Bayesian tests based on common test statistics, discuss their asymptotic properties under various common hypothesis testing problems and discuss schemes for the hyperparameter tuning under the proposed technology. In Section 4, we study the numerical efficacy of the proposed technology in different hypothesis testing problems with practical utility. Finally, we conclude with a discussion.

## 2 General framework

In this section, the objective is to lay down a hierarchical model equipped with carefully chosen priors on the parameter of interest to specify competing hypotheses, which naturally gives rise to the proposed differentially private Bayesian tests. Notably, this attribute remains elusive in the existing literature and that hinders the interpretability of the resulting inference. In particular, many existing differentially private frequentist

testing procedures (Peña and Barrientos, 2023; Kazan et al., 2023) rely on an ad-hoc sampling and aggregation scheme coupled with a test statistic truncation step, to formulate the desired tests. Despite the empirical success of such methods, a significant deficiency stems from the lack of a principled probabilistic interpretation, i.e, such methods does not enable us to quantify evidence in favor of the null and alternative hypotheses. We specifically tackle this issue in the subsequent discussions.

## 2.1 Hierarchical model and prior specification

For a positive integer  $t$ , denote  $[t] := \{1, \dots, t\}$ . Suppose we observe data  $\mathbf{x} = (x_1, \dots, x_n)'$  from a probability distribution having density  $\pi(\cdot | \theta)$ ,  $\theta \in \Theta$ . We then randomly divide the data into  $M_n$  partitions  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$ . We can assume that the data in each of the  $M_n$  partitions arise from

$$\mathbf{x}^{(i)} \sim \prod_{j=1}^{n_i} \pi(\mathbf{x}_j^{(i)} | \theta^{(i)}), \quad \text{independently for } i \in [M_n], \quad (2.1)$$

such that  $\sum_{i=1}^{M_n} n_i = n$ . The above data generating scheme comes in handy to formalize the *sample and aggregate* scheme under the proposed setup. In privacy-preserving inference, this subsample-and-aggregate structure plays a crucial role in controlling the global sensitivity of the resulting quantity of interest, e.g, test statistic, Bayes factor. By splitting the dataset into multiple partitions, computing intermediate quantities within each subset, and then aggregating these results, one can approximate the full-sample quantity of interest. This construction provides the foundation for the proposed differentially private Bayes factor, as it allows the privacy noise to scale with the sensitivity of the partition-level statistics rather than the entire dataset.

Next, we posit the priors on the parameter of interest  $\theta$  under the competing hypotheses. To that end, we assume that

$$\begin{aligned} \theta^{(i)} | \tau_{0,i}, \tau_{1,i} &\sim (1 - \omega_n) \pi_0(\theta^{(i)} | \tau_{0,i}^2) + \omega_n \pi_1(\theta^{(i)} | \tau_{1,i}^2) \quad \text{under } H_0, \\ \theta^{(i)} | \tau_{0,i}, \tau_{1,i} &\sim \omega_n \pi_0(\theta^{(i)} | \tau_{0,i}^2) + (1 - \omega_n) \pi_1(\theta^{(i)} | \tau_{1,i}^2) \quad \text{under } H_1, \end{aligned} \quad (2.2)$$

independently for  $i \in [M_n]$ , where  $\{\pi_i(\cdot), i = 0, 1\}$  are the prior probability distributions,  $(\tau_{0,i}, \tau_{1,i})$  are prior hyper-parameters, and  $\omega_n \in (0, \frac{1}{2})$  is a hyper-parameter that ensures that the Bayes factor for testing  $H_0$  against  $H_1$  is bounded. We will later on show that  $\omega_n$  depends on the sample size  $n$  and hence find it necessary to index it with  $n$ . If the Bayes factor for testing  $H_0$  against  $H_1$  were to be unbounded, the global sensitivity parameter of the Bayes factor becomes unbounded too. An immediate fix against it would involve truncating the log Bayes factor beyond an interval of the form  $[-a_n, a_n]$ ,  $a_n > 0$  or more generally  $[L_n, U_n]$  with  $L_n < U_n$ , akin to the approach in Peña and Barrientos (2024). If one were to adopt the framework of Peña and Barrientos (2024), impose non-local priors on the parameters of interest, and implement an objective Bayesian hyper-parameter tuning strategy similar to that used in the present work, one would obtain a Bayes factor with operating characteristics comparable to the differentially private Bayes factor developed here. However, such a construction would

lack the interpretability afforded by our formulation, which arises from a hierarchical model with carefully specified priors that directly encode the competing hypotheses and thereby yield the proposed differentially private Bayesian tests in a principled manner. Upon closer look, truncating the log Bayes factor in principle means that the maximum evidence that we can obtain in favor of  $H_0$  (or  $H_1$ ) against  $H_1$  (or  $H_0$ ) is bounded from above – we precisely impose this constraint a-priori via the hierarchical specification (2.2), thereby eliminating the need for truncation/post-processing. For each partition  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$ , we record the notations for the integrals

$$m_k(\mathbf{x}^{(i)} | \tau_{k,i}^2) = \int_{\Theta} \left[ \prod_{j=1}^{n_j} \pi(\mathbf{x}_j^{(i)} | \theta^{(i)}) \right] \pi_k(\theta^{(i)} | \tau_{k,i}^2) d\theta^{(i)}, \quad k = 0, 1.$$

Given  $\omega_n, \tau_{0,i}, \tau_{1,i}$ , and assuming independence, the marginal  $m_k(\mathbf{x} | \tau_{k,i}^2)$  is the product of the marginals across each partition,  $m_k(\mathbf{x}^{(i)} | \tau_{k,i}^2, \omega_n)$ . Consequently, since the Bayes factor is the ratio of these marginal densities with equal prior probabilities on the hypotheses, it becomes the product of the individual Bayes factors for each partition. Aggregation of evidence across partition is further detailed in Section 2.2. In summary, the mixture priors on  $\{\theta^{(i)}, i \in [M_n]\}$  under the competing hypotheses are instrumental in formalizing the *truncation* scheme necessary to ensure that the global sensitivity parameter of the privacy mechanism is bounded, while preserving the interpretability of the resulting Bayes factors. One may criticize the appearance of the parameter  $\omega_n$  in our formulation. To that end, we carefully introduce an asymptotic regime to analyze the proposed methodology, so that the effect of  $\omega_n$  washes away as the sample size diverges to  $\infty$ .

The model and prior specified in equations (2.1)-(2.2), together with the assumption that

$$P(H_0) = P(H_1) = 0.5, \quad (2.3)$$

a-priori, completely describe our generative model of interest.

## 2.2 Properties

We note that for each partition  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$ , the quantity  $m_k(\mathbf{x}^{(i)} | \tau_{k,i}^2)$ ,  $k = 0, 1$ , can be interpreted as the marginal likelihood of the data partition  $\mathbf{x}^{(i)}$  under the hypothesis  $H_k$  when  $\omega_n = 0$ . Under this set up, as we eluded to earlier, the Bayes factor for testing  $H_0$  against  $H_1$  can be unbounded, and consequently the global sensitivity parameter of the privacy mechanism is unbounded too. This renders the development of differentially private Bayesian tests non-trivial, and throughout the article we shall assume that the hyper-parameter  $\omega_n > 0$ . Under the alternative hypothesis  $H_1$  in (2.1)-(2.3), the marginal distribution of the data partition  $\mathbf{x}^{(i)}$  is expressed as

$$m_1^t(\mathbf{x}^{(i)} | \tau_{0,i}^2, \tau_{1,i}^2, \omega_n) = \omega_n m_0(\mathbf{x}^{(i)} | \tau_{0,i}^2) + (1 - \omega_n) m_1(\mathbf{x}^{(i)} | \tau_{1,i}^2), \quad i \in [M_n];$$

and under the null hypothesis  $H_0$ , it is given by:

$$m_0^t(\mathbf{x}^{(i)} | \tau_{0,i}^2, \tau_{1,i}^2, \omega_n) = (1 - \omega_n) m_0(\mathbf{x}^{(i)} | \tau_{0,i}^2) + \omega_n m_1(\mathbf{x}^{(i)} | \tau_{1,i}^2), \quad i \in [M_n].$$

Consequently, in light of the data partition  $\mathbf{x}^{(i)}$ , the Bayes factor against the null hypothesis  $H_0$  is expressed as

$$\begin{aligned} & \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n) \\ &= \frac{\omega_n + (1 - \omega_n)[m_1(\mathbf{x}^{(i)} \mid \tau_{1,i}^2) / m_0(\mathbf{x}^{(i)} \mid \tau_{0,i}^2)]}{(1 - \omega_n) + \omega_n[m_1(\mathbf{x}^{(i)} \mid \tau_{1,i}^2) / m_0(\mathbf{x}^{(i)} \mid \tau_{0,i}^2)]}, \quad i \in [M_n]. \end{aligned} \quad (2.4)$$

We proceed by listing out the key properties of the Bayes factor against the null hypothesis  $\text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  for the data partition  $\mathbf{x}^{(i)}$ .

**Lemma 2.1.** For  $\omega_n \in (0, 1/2)$ ,  $\text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  is bounded between  $[\frac{\omega_n}{1-\omega_n}, \frac{1-\omega_n}{\omega_n}]$ .

**Lemma 2.2.** Under (2.1)-(2.3), the Bayes Factor against the null hypothesis  $H_0$  combined across all the partitions  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$  is expressed as

$$\text{BF}_{10}^t(\mathbf{x} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n) = \prod_{i=1}^{M_n} \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n).$$

**Corollary 2.2.1.** Lemma 2.1 together with Lemma 2.2 implies that, the combined Bayes Factor against the null hypothesis  $H_0$  across all the partitions  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$ ,  $\text{BF}_{10}^t(\mathbf{x} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  is bounded between  $(\frac{\omega_n}{1-\omega_n})^{M_n}$  and  $(\frac{1-\omega_n}{\omega_n})^{M_n}$  for  $\omega_n \in (0, 1/2)$ .

In the sequel, we construct differential private Bayesian tests based on the randomised mechanism, denoted by  $f$ , that map a data set  $\mathbf{x}$  to the average log Bayes factor

$$f_{10}(\mathbf{x}) = \frac{1}{M_n} \log \text{BF}_{10}^t(\mathbf{x} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n) = \frac{1}{M_n} \sum_{i=1}^{M_n} \log \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n). \quad (2.5)$$

We take refuge to common differential privacy mechanism of adding calibrated privacy noise to the randomised mechanism  $f$  to ensure differential privacy of the inferential procedure. For every data set  $\mathbf{x}$ , since the log Bayes factor in each of the partitions  $\log \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  is bounded between  $[-a_n, a_n]$  with  $a_n = -\log(\omega_n/1 - \omega_n)$ , the global sensitivity of  $\log \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  is  $2a_n$ . Consequently, the global sensitivity of  $\text{GS}_f$  of the log Bayes factor averaged over the partitions, i.e of  $f$ , is  $(2a_n/M_n)$ . We now have all the necessary ingredients to introduce our privacy preserving mechanism.

To that end, let  $\mathcal{D}$  denote the collection of all possible data sets. We consider a random mechanism  $H : \mathcal{D} \rightarrow \mathbf{R}$  expressed as

$$H_{10}(\mathbf{x}) = f_{10}(\mathbf{x}) + \eta, \quad \mathbf{x} \in \mathcal{D}, \quad (2.6)$$

where  $\eta$  is a Laplace(0,  $\text{GS}_f/\varepsilon$ ) distributed privacy noise. A test based on  $H_{10}(\mathbf{x})$  is then obtained as

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } H_{10}(\mathbf{x}) \geq \gamma_{\alpha,\varepsilon}(n), \\ 0 & \text{otherwise,} \end{cases} \quad (2.7)$$

for some cut-off  $\gamma_{\alpha,\varepsilon}(n)$ , depending on the sample size  $n$  and size  $\alpha$  of the test. Here,  $\phi(\mathbf{x}) = 1$  signifies rejection of  $H_0$ / acceptance of  $H_1$ , and if  $\phi(\mathbf{s}) = 0$ , accept  $H_0$ .

**Theorem 2.3.** *The function  $\phi(\mathbf{x})$  for testing  $H_0$  against  $H_1$ , as described in (2.5)-(2.7), is  $\varepsilon$ -differentially private.*

*Proof.* The global sensitivity  $GS_f$  of  $f(\mathbf{x})$  is  $\frac{2a_n}{M_n}$ . Then, defining  $H(\mathbf{x}) = f(\mathbf{x}) + \eta$ , where  $\eta$  follows  $\text{Laplace}(0, GS_f/\varepsilon)$  completes the proof, by definition.  $\square$

Next, we develop differentially private Bayes factors based on widely-used test statistics, and crucially leverage the proposed framework to provide a systematic approach for selecting  $\gamma_{\alpha,\varepsilon}(n)$  that maintains a predetermined privacy budget, assuming the investigator specifies the effect size of interest in advance.

### 3 Differentially private Bayes factors based on test statistics

The general framework presented in Section 2 can be employed to devise fully parametric Bayesian tests that preserves differential privacy. However, given the limitations of Bayes factors derived from fully parametric models, as outlined in the introduction, and in order to ensure straightforward preservation of  $(\varepsilon, 0)$ -differential privacy in the proposed mechanism, we instead develop differentially private Bayes factors based on test statistics. On top of computational simplicity and objectivity, a major benefit of representing Bayes factors in terms of test statistics that is crucial for us is that they can be formulated as a function of the standardized effect size (Johnson et al., 2023b; Datta et al., 2024), if we specify the competing hypotheses in (2.2) using a mixture of a spike at 0 and an appropriately chosen first order non-local prior (Johnson and Rossell, 2010b; Johnson et al., 2023b) on the non-centrality parameter of the appropriate test statistic. Moreover, such non-local priors contain a single scale parameter that control the mode of the prior densities that in turn can be utilized to define the null and alternative hypotheses. By determining the prior mode using a function of the standardized effect size (analogous to BFFs.), we can derive Bayes factors for a sequence of null and alternative hypothesis. In particular, the hyper parameter  $\tau_{1,i}^2$  in (2.2) can be objectively determined based on the standardized effect size of interest to express our Bayes factors as functions of the sequence of null and alternatives. This enables us to determine a privatized cut-off  $\gamma_{\alpha,\varepsilon}(n)$  that varies for each sequence of null and alternatives, while adhering to pre-specified privacy budget.

Beyond aiding the computation of the cut-off value  $\gamma_{\alpha,\varepsilon}(n)$ , such that  $(\varepsilon, 0)$ -differential privacy is ensured, the choice of priors throughout the rest of the article is further guided by two other considerations. First, the carefully chosen non-local priors ensure that both the marginal likelihoods under the two competing hypotheses of interest are available in closed form. This eliminates the computational expenses associated with the numerical evaluation of the otherwise intractable multiple integrals. Secondly, and perhaps more

important, the prior specification ensures rapid accumulation of evidence in favor of the true hypothesis.

We shall now formally introduce the differentially private Bayes factor based on test statistics. We note that, the partition specific log Bayes factor based on test statistic  $\log \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$ , as shown later in Section 3.3, depends on  $\mathbf{x}^{(i)}$  only through the partition specific test statistic, denoted by  $s^{(i)}$ ,  $i \in [M_n]$ . To make it explicit, we use the notation  $\log \text{BF}_{\text{stat},10}^t(s^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  instead of  $\log \text{BF}_{10}^t(\mathbf{x}^{(i)} \mid \tau_{0,i}^2, \tau_{1,i}^2, \omega_n)$  in the rest of the section. Similarly, we shall use  $f_{\text{stat},10}(\mathbf{s})$  and  $H_{\text{stat},10}(\mathbf{s})$  to denote the non-privatized and privatized average log Bayes factors, instead of  $f_{10}(\mathbf{x})$  and  $H_{10}(\mathbf{x})$ , respectively, where  $\mathbf{s} = (s^{(1)}, \dots, s^{(M_n)})^\top$ . With that, the Bayesian test function based on test statistic  $H_{\text{stat},10}(\mathbf{s})$  is expressed as

$$H_{\text{stat},10}(\mathbf{s}) = f_{\text{stat},10}(\mathbf{s}) + \eta, \quad \mathbf{s} \in \mathcal{D}, \quad (3.1)$$

where  $\eta$  is a  $\text{Laplace}(0, \text{GS}_f/\varepsilon)$  distributed privacy noise. Similar to the general framework, a test based on  $H_{\text{stat},10}(\mathbf{s})$  is then obtained as

$$\phi(\mathbf{s}) = \begin{cases} 1 & \text{if } H_{\text{stat},10}(\mathbf{s}) \geq \gamma_{\alpha,\varepsilon}(n), \\ 0 & \text{otherwise,} \end{cases} \quad (3.2)$$

for some cut-off  $\gamma_{\alpha,\varepsilon}(n)$ , depending on the sample size  $n$  and size  $\alpha$  of the test. Here,  $\phi(\mathbf{s}) = 1$  signifies rejection of  $H_0$ / acceptance of  $H_1$ , and if  $\phi(\mathbf{s}) = 0$ , accept  $H_0$ .

Next, given the sample size  $n$ , privacy mechanism, the user defined parameters  $(\varepsilon, a_n, M_n)$ , and the size  $\alpha$  of the test, we determine the cut off  $\gamma_{\alpha,\varepsilon}(n)$  by satisfying the *size constraint*

$$\Xi_0(M_n \mid \varepsilon, \alpha, n) = \mathbb{E}[\phi(\mathbf{s}) \mid H_0] = \mathbb{P}_{H_0}[H(\mathbf{s}) \geq \gamma_{\alpha,\varepsilon}(n)] = \alpha. \quad (3.3)$$

Exact calculation of the cut off  $\gamma_{\alpha,\varepsilon}(n)$  may not be amenable under certain cases. However, we can always obtain  $\gamma_{\alpha,\varepsilon}(n)$  via a Monte Carlo simulation. We present the formalized algorithm to determine the cutoff value  $\gamma_{\alpha,\varepsilon}(n)$  in Algorithm 1. Determination of the cut off is further demonstrated in the experiment section.

### 3.1 Selection of size $\alpha$ cut-off $\gamma_{\alpha,\varepsilon}(n)$

Given the privatized average logarithm of Bayes factor based on test statistic  $H_{\text{stat},10}$  as defined in (3.2), our goal in Algorithm 1 is to compute the associated size  $\alpha$  cut-off  $\gamma_{\alpha,\varepsilon}$  via a Monte Carlo scheme. Such Monte Carlo approaches are ubiquitous in literature in finding cut-offs of test functions; see Barrientos et al. (2019) for an example in context of differentially private frequentist tests.

The inputs of the Algorithm 1 are (a) the privacy budget  $\varepsilon$ , (b) the number of partitions  $M_n$ , and truncation level of the partition specific log Bayes factors  $a_n$ , (c) the standardised effect size of interest  $\pi^*$ , and finally (d) the number of iterations in the Monte-Carlo simulation  $N$ . To find the cut-off  $\gamma_{\alpha,\varepsilon}(n)$ , we note that the partition specific

common test statistic  $s^{(i)}$  follows a non-central distribution, denoted by  $S_{\nu_i}(\lambda^{(i)})$ , under the null hypothesis in (2.1)-(2.2), where

$$\lambda^{(i)} \sim (1 - \omega_n)\delta_0 + \omega_n\pi_1(\lambda^{(i)} \mid \tau_i^2)$$

and  $\nu_i$  is a constant completely determined by the size of the  $i$ -th partition of the data. With this insight, at the iteration  $k \in \{1, \dots, N\}$  of the Monte Carlo scheme, we first independently draw the non-centrality parameter  $\lambda^{(i)}$  from a prior distribution with density  $(1 - \omega_n)\delta_0 + \omega_n\pi_1(\lambda^{(i)} \mid \tau_i^2)$  and pseudo test statistics  $s_k^{(i)}$  from a non-central distribution  $S_{\nu_i}(\lambda^{(i)})$ ,  $i \in [M_n]$ . Crucially, the prior hyper-parameter  $\tau_i^2$  is completely determined by the effect size of interest  $\pi^*$  and size of the  $i$ -th partition  $n_i$ . For example, for a  $t$ -test, we draw the pseudo test statistic  $s_k^{(i)} \equiv t_k^{(i)}$  from non-central  $t_{|n_i|-1}(\lambda^{(i)})$  distributions with non-centrality parameter  $\lambda^{(i)}$ , where

$$\lambda^{(i)} \sim (1 - \omega_n)\delta_0 + \omega_n J(\tau_i^2), \quad \tau_i = \sqrt{n_i}\pi^*,$$

where  $J(\tau_i^2)$  denotes a normal moment density of order 1, the form of which is given by,

$$J(x \mid \tau^2) = \frac{x^2}{\sqrt{2\pi}\tau^3} \exp\left(-\frac{x^2}{2\tau^2}\right), \quad x \in \mathbf{R}. \quad (3.4)$$

The normal moment prior Johnson and Rossell (2010b); Johnson et al. (2023b) has two key features that make it particularly well-suited for Bayesian hypothesis testing. First, it is a nonlocal prior, meaning it assigns zero probability density at the null value (typically zero). This property enforces a clear separation between the null and alternative hypotheses by penalizing parameter values close to the null, thereby reducing false positives and increasing the evidence in favor of the null when it is true. Second, the normal moment prior improves model selection consistency, especially in high-dimensional or sparse settings. Its structure ensures that, as sample size increases, the Bayes factor reliably favors the true model—whether it corresponds to the null or the alternative—leading to more robust and interpretable inference.

For a  $\chi^2$ -test, we draw the test statistic  $s_k^{(i)} \equiv h^{(i)}$  from a non-central  $\chi_{\nu}^2(\lambda^{(i)})$ , where

$$\lambda^{(i)} \sim (1 - \omega_n)\delta_0 + \omega_n \text{G}((k/2) + 1, 1/2\tau_i^2), \quad \tau_i^2 = \frac{n_i\pi^{*\prime}\pi^*}{2},$$

where  $\text{G}(a, b)$  denotes a gamma density with shape parameter  $a_n$  and rate parameter  $b$ . For an  $F$ -test, we draw the test statistic  $f^{(i)}$  from a non-central  $F_{p, n_i-p}(\lambda^{(i)})$  for a  $p \geq 1$ . Then, based on  $\mathbf{s}_k = (s_k^{(1)}, \dots, s_k^{(M_n)})^\top$ , we calculate the pseudo non-private truncated log Bayes factor  $f_{\text{stat},10}(\mathbf{s}_k)$  via Equation (2.5). Then, a privacy noise  $\eta$  generated from  $\text{Laplace}(0, 2a_n/\varepsilon M_n)$  is added to  $f_{\text{stat},10}(\mathbf{s}_k)$  to calculate the pseudo private truncated log Bayes factor  $\text{H}_{\text{stat},10}(\mathbf{s}_k)$  via Equation (3.1). In step (3), we set the size  $\alpha$  cut-off  $\gamma_{\alpha,\varepsilon}(n)$  to the  $100(1 - \alpha)\%$  quantile the  $\{\text{H}_{\text{stat},10}(\mathbf{s}_k), k \in [N]\}$  values. It is important to note that we are sampling the test statistics from their conditional distributions,  $S_{\nu}(\lambda)$ , under the null hypothesis, where the mode of the slab part of the prior is set by the standardized effect size  $\pi^*$ . In this process, we are not sampling pseudo data, but

solely the test statistics which only depend on the partition specific sample size and the standardized effect size of interest. This ensures the preservation of differential privacy of the proposed mechanism.

We note that for a broad class of models, a sufficient test statistic with a tractable null distribution may not exist, even though the full data-generating distribution under both the null and alternative hypotheses is known. In such settings, the general framework described in Section 2 can be applied directly to the full dataset, without reduction to a test statistic. An explicit algorithm for computing the corresponding Bayes factor cutoff in this full-data setting is provided in Section 2 of the supplementary material.

---

**Algorithm 1** (Bayes factor cut-off)

---

- 1: **Input.** (i) The number of partitions  $M_n$ , (ii) the truncation level  $a_n$  of the partition specific log Bayes factors, (iii) privacy parameter  $\varepsilon$ , (iv)  $N$ : The total number of the Monte-Carlo simulations  $N$  to estimate the size  $\alpha$  cut-off  $\gamma_{\alpha,\varepsilon}(n)$ , (v) Standardized effect size of interest  $\pi^* \in (0, 1)$ .
- 2: **for** Standardized effect size of interest  $\pi^*$  **do**
- 3:   **for**  $k = 1$  to  $N$  **do**
  - (i) Compute of  $\tau_i^2, i = 1, 2, \dots, M_n$  in terms of  $\pi^*$  for each of the  $M_n$  partitions of the data. For example, for a  $z$  or  $t$  test  $\tau_i^2 = \frac{n_i \pi^{*t} \pi^*}{2}$  for the  $i$ -th partition. <sup>1</sup>
  - (ii) Draw independent samples of the non-centrality parameter,

$$\lambda^{(i)} \sim (1 - \omega_n)\delta_0 + \omega_n\pi_1(\lambda^{(i)} \mid \tau_i^2),$$

where  $\pi_1(\lambda^{(i)} \mid \tau_i^2)$  is a non-local moment prior distribution,  $i \in [M_n]$ . The prior hyper-parameter  $\tau_i^2$  is completely determined by standardized effect size of interest  $\pi^*$  and the size of the  $i$ -th partition  $n_i$ .

- (ii) Draw test statistics  $s_k^{(i)}$  from  $S_{\nu_i}(\lambda^{(i)})$  independently for  $i \in [M_n]$ .
- (iii) Compute  $\text{BF}_{\text{stat},10}^t(s_k^{(i)} \mid \tau_i^2, \omega_n)$  for  $i \in [M_n]$ .
- (iv) Compute

$$f_{\text{stat},10}(\mathbf{s}_k) = \frac{1}{M_n} \sum_{i=1}^{M_n} \log \text{BF}_{\text{stat},10}^t(s_k^{(i)} \mid \tau_i^2, \omega_n).$$

- (v) Generate  $\eta$  from Laplace  $(0, \frac{2a_n}{\varepsilon M_n})$ .
  - (vi) Compute  $\mathbf{H}_{\text{stat},10}(\mathbf{s}_k) = f_{\text{stat},10}(\mathbf{s}_k) + \eta$ .
  - 4:   **end for**
  - 5:   **Output.** Compute the size  $\alpha$  cut-off  $\gamma_{\alpha,\varepsilon}(n)$ , such that  $100(1 - \alpha)\%$  of the  $\{\mathbf{H}_{\text{stat},10}(\mathbf{s}_1), \dots, \mathbf{H}_{\text{stat},10}(\mathbf{s}_N)\}$  values are greater than  $\gamma_{\alpha,\varepsilon}(n)$ .
  - 6: **end for**
- 

<sup>1</sup>Choice of  $\tau_i^2$  for various tests ( $z, t, \chi^2$  and  $F$ ) can be found in Table 1 of [Johnson et al. \(2023b\)](#).

### 3.2 Hyperparameter tuning

For the use of differentially private Bayesian tests, the concerned parties initially establish the desired privacy level  $\varepsilon$ . Following this determination, we need to select appropriate values for the parameters  $M_n$ , given  $a_n = -\log(\omega_n/1 - \omega_n) = kn^\beta$  with fixed  $0 < \beta < 1, 0 < k \leq 1$ . This iterative process will allow us to experiment with diverse values of  $M_n$ , thereby striking a nuanced balance between attaining the desired privacy level and asymptotic behavior of the Bayes factors. In this article, we consider the case where we do not have the autonomy to select specific value of  $\varepsilon$ , but possess the discretion to choose  $M_n$ . Analysts, endowed with the autonomy of choosing the privacy budget  $\varepsilon$ , may iteratively employ the described methodology, for varying the values of  $\varepsilon$ .

Given the sample size  $n$ , a privacy mechanism, a privacy budget  $\varepsilon$ , standardized effect size of interest  $\pi^*$  and the size  $\alpha$ , we determine the Bayes factor cut off  $\gamma_{\alpha,\varepsilon}(n)$  using (3.3) via a Monte Carlo simulation, for varying values of  $M_n$  given  $a_n$ . Now, let  $\mathcal{M}$  denote the space all possible pairs  $M_n$  under consideration. We operate under the regime that  $M_n = \zeta n$ ,  $0 < \zeta \leq 1$  given  $a_n = kn^\beta$ . The constraints  $0 < \beta < 1$  and  $0 < k \leq 1$  ensure that the  $\text{GS}_f$  decreases as the sample size increases. The order assumption on  $a_n$ , or equivalently  $\omega_n$ , enables us to dictate the rate at which the hypotheses specified in Theorems 3.1-3.2 (Theorems 3.3 and 3.4 of the Supplementary Material) approach the usual/[non-privatized null hypotheses significance testing framework](#). This, in turn, enables us to dictate the convergence rate of the privatized Bayes factor. Specifically, slower the order of  $a_n$ , lower would be the rate of convergence. Refer to Corollary 3.2.2 and Theorem 3.3 for details. The optimal choices  $M_n$  is obtained via *maximising power* of the test

$$\arg \max_{M_n \in \mathcal{M}} \Xi_1(M_n \mid \varepsilon, \alpha, n)$$

where

$$\Xi_1(M_n \mid \varepsilon, \alpha, n) = \mathbb{E}[\phi(\mathbf{s}) \mid H_1] = \mathbb{P}_{H_1}[\mathbb{H}(\mathbf{s}) \geq \gamma_{\alpha,\varepsilon}(n)].$$

The numerical strategy for the hyper parameter tuning in further demonstrated in Section 4 via examples.

### 3.3 Bayes factors based on test statistics and convergence rates of Bayes factors

In this section, we present the closed form expressions for differentially private Bayes factors based on common  $t, \chi^2$  and  $F$  test statistics. Further, we study the asymptotic properties of the proposed Bayes factors, adopting the popular notion of *Conventional Bayes factor consistency* (Chib and Kuffner, 2016; Chatterjee et al., 2020). Such asymptotic properties must be satisfied to ensure that, as the sample size increases, the correct model has a faster rate of posterior contraction as a function of sample size in comparison to an incorrect model/complex model. Results on consistency of Bayes factors based on test statistics (Johnson, 2008; Datta et al., 2024) is not particularly new. However, the introduction of differential privacy to this framework introduces unique challenges.

To ensure the convergence of the privatized Bayes factors, it is essential to incorporate additional meaningful assumptions. This includes considerations about the distribution of test statistics under the competing hypotheses (2.2), order of the truncation parameter  $a_n$  (or equivalently  $\omega_n$ ) and number of data partitions  $M_n$  relative to the sample size, and the characteristics of the global sensitivity parameter.

**Definition 3** (Conventional Bayes factor consistency (Chib and Kuffner, 2016)). *For two competing models  $M_l$  and  $M_k$ , the Bayes factor for comparing models  $M_k$  and  $M_l$ , denoted as  $\text{BF}_{kl} = \frac{m(\mathbf{x}|M_k)}{m(\mathbf{x}|M_l)}$ , is consistent if*

- (i)  $\text{BF}_{kl} \xrightarrow{p} 0$  when  $M_l$  contains the true model, and
- (ii)  $\text{BF}_{kl} \xrightarrow{p} \infty$  when  $M_k$  contains the true model, where  $\mathbf{x}$  is the data.

Before we proceed, a few notations are in order. The confluent hyper-geometric function of order  $(m, n)$  (Abramowitz, 1974) is denoted by  ${}_mF_n(a, b; z)$ . The Dirac’s delta measure at 0 is denoted by  $\delta_0$ . Further,  $N(a, b)$  denotes normal distribution with a mean of  $a_n$  and variance  $b$ ,  $T_n(\theta)$  denotes the t-distribution with degrees of freedom  $\nu$  and non-centrality parameter  $\theta$ ,  $\chi_n^2(\theta)$  denotes the chi-squared distribution with degrees of freedom  $\nu$  and non-centrality parameter  $\theta$ ,  $F_{k,m}(\theta)$  denotes the F-distribution with degrees of freedom  $(k, m)$  and non-centrality parameter  $\theta$ ,  $G(\alpha, \theta)$  denotes a gamma distribution with a shape parameter  $\alpha$  and a rate parameter  $\theta$ , and  $J(\tau^2)$  represents a normal-moment density of order 1 (Johnson and Rossell, 2010b).

To guarantee consistency of the Bayes factors, we operate under the asymptotic regime where the number of partitions satisfies  $M_n = \zeta n$ , for some constant  $0 < \zeta \leq 1$ , and the truncation parameter is chosen as  $a_n = kn^\beta$ , with  $0 < \beta < 1$  and  $0 < k \leq 1$ . These assumptions apply uniformly across all tests considered. They arise from the selection of the truncation parameter  $a_n = -\log(\omega_n/1 - \omega_n)$ . Recall that, we divide  $n$  observations into  $M_n$  partitions. This asymptotic scaling ensures that the generalized sensitivity  $\text{GS}_f$  vanishes as  $n \rightarrow \infty$ , while preserving the rapid convergence behavior of the privatized Bayes factor. We shall put this statement in concrete terms in Theorem 3.3.

The generalized sensitivity  $\text{GS}_f$  measures the maximum change in the privatized statistic when a single observation is altered. In the proposed construction, the test statistic is computed over  $M_n = \zeta n$  equally sized partitions, each containing approximately  $n/M_n = 1/\zeta$  observations. As the total sample size  $n$  increases, the contribution of any single observation to the aggregated (sample–aggregate) statistic is therefore diluted at rate  $1/n$ . Consequently, the influence of an individual data point on the privatized Bayes factor naturally diminishes as  $n$  grows. More formally, if the truncation parameter is chosen as  $a_n = kn^\beta$  with  $0 < \beta < 1$ , then the sensitivity of the truncated log Bayes factor satisfies

$$\text{GS}_f = O\left(\frac{a_n}{n}\right) = O(n^{\beta-1}),$$

which converges to zero for any  $\beta < 1$ . This falls under the standard “vanishing-sensitivity” asymptotic regime in differential privacy, where the privacy cost per individual decreases as the sample size grows while the statistical power of the test continues

to improve. This assumption is not only mathematically convenient but also conceptually aligned with population-level inference, in which the privacy burden per subject should shrink as more data accumulate. Similar vanishing-sensitivity regimes have been employed in recent work on differentially private hypothesis testing.

With these notations, the Bayes Factors based on common test statistics for each partition  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$  of the observed dataset are presented in Theorems 3.1, 3.2, and Theorems 3.3, 3.4 of the Supplementary Material, in a manner consistent with the developments of Johnson (2005); Johnson and Rossell (2010b); Johnson et al. (2023b).

**Theorem 3.1. (Two-sided  $z$ -test)** *Suppose the generative model for the test statistic  $z$  under  $H_0$  and  $H_1$  are*

$$\begin{aligned} H_0 : z \mid \lambda &\sim N(\lambda, 1), & \lambda \mid \tau^2 &\sim (1 - \omega_n) \delta_0 + \omega_n J(\tau^2) \\ H_1 : z \mid \lambda &\sim N(\lambda, 1), & \lambda \mid \tau^2 &\sim \omega_n \delta_0 + (1 - \omega_n) J(\tau^2), \quad \tau > 0, \end{aligned}$$

respectively. Then, the Bayes factor in favor of the  $H_1$  is of the form  $\text{BF}_{10}^t(z \mid \tau^2, \omega_n) = \frac{\omega_n + (1 - \omega_n)\mathbf{R}}{(1 - \omega_n) + \omega_n \mathbf{R}}$ , where

$$\mathbf{R} = m_1(z \mid \tau^2) / m_0(z) = (1 + \tau^2)^{-\frac{3}{2}} {}_1F_1(3/2, 1/2; \tau^2 z^2 / 2(1 + \tau^2)).$$

*Proof is deferred to the supplementary material.*

**Corollary 3.1.1.** *Under the set up in Theorem 3.1, we further assume that (i)  $n \bmod M_n = 0$  and the partition specific sample sizes satisfy  $n_i = n/M_n \forall i \in [M_n]$ . (ii) the test statistic  $z^{(i)} \sim N(\eta\sqrt{n_i}, 1)$ , for some  $\eta$ ; (iii) the hyper-parameter  $\tau_i^2 = \kappa n_i$ , for some  $\kappa > 0$ ; (iv)  $\log((1 - \omega_n)/\omega_n) = kn^\beta$ ,  $0 < \beta < 1, 0 < k \leq 1$ . (v)  $M_n = \zeta n$ ,  $0 < \zeta \leq 1$ . Then, under  $H_0$ ,  $\prod_{i=1}^{M_n} \text{BF}_{10}^t(z^{(i)} \mid \tau_i^2, \omega_n) = O_p(c^{-n})$ , where  $c$  is a positive constant. Similarly, under  $H_1$ ,  $\prod_{i=1}^{M_n} \text{BF}_{01}^t(z^{(i)} \mid \tau_i^2, \omega_n) = O_p(c^{-n})$ . This demonstrates that the combined Bayes factor, obtained from the partition specific Bayes factors in Theorem 3.1, is consistent under both  $H_0$  and  $H_1$ . Proof is deferred to the supplementary material.*

Assumption (i) could be relaxed by permitting  $n_i$  to increase linearly with  $n$ . This assumption is made primarily for clarity and ease of explanation. The justification for assumption (ii) is provided by the order of commonly utilized  $z$  statistics under  $H_1$ . For instance, consider the test means of an univariate normal population  $N(\mu, 1)$ . The test statistic of interest  $z = \sqrt{n}\bar{X}$  satisfies the assumption, with  $\bar{X}$  and  $n$  being respectively the mean and number of observations collected from the population. Assumption (iii) stems from the selection of  $\tau_i = \sqrt{n_i}\pi^*$ , where  $\pi^*$  represents the standardized effect size for one-sample  $z$  tests. The choice of the scale parameter  $\tau_i^2$  in terms of the sample size and  $\pi^*$  for various linear tests are listed in Table 1 of Johnson et al. (2023b). Assumption (iv) and (v) is shared across all tests, and its justification was provided at the start of this section. This selection ensures that the  $\text{GS}_f$  decreases as the sample size increases, while maintaining the rapid convergence of the privatized Bayes factor.

Now, we carry on with the discussion on privatized Bayes factors based on  $t$ ,  $\chi^2$  and  $F$  test statistics for each partition  $\mathbf{x}^{(i)}$ ,  $i \in [M_n]$  of the observed dataset. The corresponding results for the  $t$ -tests are included in the main manuscript, while those for the  $\chi^2$  and  $F$  tests are provided in Section 3 of the Supplementary Material.

**Theorem 3.2. (Two-sided  $t$ -test)** Suppose the generative model for the test statistic  $t$  under  $H_0$  and  $H_1$  are

$$\begin{aligned} H_0 : t \mid \lambda &\sim T_\nu(\lambda), & \lambda \mid \tau^2 &\sim (1 - \omega_n) \delta_0 + \omega_n J(\tau^2) \\ H_1 : t \mid \lambda &\sim T_\nu(\lambda), & \lambda \mid \tau^2 &\sim \omega_n \delta_0 + (1 - \omega_n) J(\tau^2), \quad \tau > 0, \end{aligned}$$

respectively. Then, the Bayes factor in favor of the  $H_1$  is of the form  $\text{BF}_{10}^t(t \mid \tau^2, \omega_n) = \frac{\omega_n + (1 - \omega_n)\mathbf{R}}{(1 - \omega_n) + \omega_n\mathbf{R}}$ , where

$$\mathbf{R} = \frac{1}{(1 + \tau^2)^{\frac{3}{2}}} {}_2F_1\left(\frac{3}{2}, \frac{\nu + 1}{2}, \frac{1}{2}, \frac{t^2 \tau^2}{(t^2 + \nu)(1 + \tau^2)}\right).$$

*Proof is deferred to the supplementary material.*

**Corollary 3.2.1.** Under the set up in Theorem 3.2, we further assume that (i)  $n \bmod M_n = 0$  and the partition specific sample sizes satisfy  $n_i = n/M_n \forall i \in [M_n]$ . (ii) the test statistic  $t^{(i)} \sim T_\nu(\eta\sqrt{n_i})$ , for some  $\eta$ ; (iii) the hyper-parameters  $\nu_i = \delta n_i - u$  and (iii)  $\tau_i^2 = \kappa n_i$ , for some  $\kappa > 0$ , (iv)  $\log((1 - \omega_n)/\omega_n) = \kappa n^\beta$ ,  $0 < \beta < 1, 0 < \kappa \leq 1$ . (v)  $M_n = \zeta n$ ,  $0 < \zeta \leq 1$ . Then, under  $H_0$ ,  $\prod_{i=1}^{M_n} \text{BF}_{10}^t(t^{(i)} \mid \tau_i^2, \omega_n) = O_p(c^{-n})$ , where  $c$  is a positive constant. Similarly, under  $H_1$ ,  $\prod_{i=1}^{M_n} \text{BF}_{01}^t(t^{(i)} \mid \tau_i^2, \omega_n) = O_p(c^{-n})$ . This demonstrates that the combined Bayes factor, obtained from the partition specific Bayes factors in Theorem 3.2, is consistent under both  $H_0$  and  $H_1$ . *Proof is deferred to the supplementary material.*

Assumption (ii) follows from the degrees of freedom of frequently used  $t$  statistic. For instance, consider the test means of an univariate normal population  $N(\mu, \sigma^2)$ , where  $\sigma^2$  is unknown. Then, the  $t$ -statistic follows a  $t$  distribution with degrees of freedom  $n - 1$ . Rest of assumptions follow from considerations described after Corollary 3.1.1.

Closed-form Bayes factor expressions for the  $\chi^2$  and  $F$  test statistics, together with their corresponding consistency properties, are provided in Section 3 of the Supplementary Material. Theorem 3.3 and Corollary 3.3.1 of the Supplementary present the explicit Bayes factor expression for the  $\chi^2$  test and establish its convergence rate respectively. Likewise, Theorem 3.4 and Corollary 3.4.1 give the corresponding Bayes factor expression and convergence results for the  $F$  test, respectively. We note that when  $z^2 = h$ , the numerical value of the Bayes factor computed from Theorem 3.1 is identical to the value obtained using the  $\chi^2$  test with  $k = 1$  degree of freedom (see Theorem 3.3 of the Supplement). Similarly, when  $t^2 = f$ , the value of the Bayes factor based on Theorem 3.2 coincides with that obtained from the  $F$  test with  $(a = 1, b)$  degrees of freedom (see Theorem 3.4 of the Supplement). The full derivations of these equivalences appear in Johnson et al. (2023b) and are beyond the scope of the current manuscript.

**Corollary 3.2.2.** Under the conditions in corollaries 3.1.1-3.2.1 and noting that the privacy noise  $\eta = o_p(1)$ , it follows that the privatized combined logarithm of the Bayes factor,  $M_n \times \text{H}_{\text{stat},10}(\mathbf{s}) = O_p(n)$  under  $H_0$ . Similarly,  $M_n \times \text{H}_{\text{stat},01}(\mathbf{s}) = O_p(n)$  under  $H_1$ .

Besides Bayes factor consistency, another desirable characteristic of our privatized Bayes factors is that both the privatized and non-privatized Bayes factors will yield similar inference as the sample size diverges to  $\infty$ . Comparable results are presented in [Barrientos et al. \(2019\)](#), in the context of differentially private frequentist tests for assessing significance of linear regression. Denoting the true non-privatized Bayes factor against the null hypothesis for the  $i$ -th data partition by  $R^{(i)}$ ,  $i \in [M_n]$ , we characterize the discrepancy between the average privatized weight of evidence  $H_{\text{stat},10}(\mathbf{s})$  and average non-privatized weights of evidence  $\frac{1}{M_n} \sum_{i=1}^{M_n} \log(R^{(i)})$  in terms of the probability  $P\left(\left|H_{\text{stat},10}(\mathbf{s}) - \frac{1}{M_n} \sum_{i=1}^{M_n} \log(R^{(i)})\right| > c\right)$  for some  $c > 0$  ([Barrientos et al., 2019](#)). Under appropriate conditions, [Theorem 3.3](#) states that, the discrepancy between the average privatized weight of evidence and the average true non-privatized weight of evidence converges in probability to 0.

**Theorem 3.3.** *Assume that the conditions in corollaries [3.1.1-3.2.1](#) hold. Further, assume  $m_1(s^{(i)})$  and  $m_0(s^{(i)})$  are strictly positive for each  $i \in [M_n]$ . Then,*

$$\left|H_{\text{stat},10}(\mathbf{s}) - \frac{1}{M_n} \sum_{i=1}^{M_n} \log(R^{(i)})\right| \xrightarrow{\mathbb{P}} 0, \text{ as } n \rightarrow \infty, \quad \text{where } R^{(i)} = \frac{m_1(s^{(i)} \mid \tau_{1,i}^2)}{m_0(s^{(i)} \mid \tau_{0,i}^2)}, \quad i \in [M_n].$$

*Proof is deferred to the supplementary material.*

In summary, [Theorem 3.3](#) formalises that both the privatized and non-privatized Bayes factors will favor the same hypothesis with probability converging to 1 when the sample size diverges to  $\infty$ . [Corollary 3.2.2](#) and [Theorem 3.3](#) remain valid for the  $\chi^2$  and  $F$  test settings, provided the respective conditions specified in [Corollary 3.3.1](#) and [Corollary 3.4.1](#) of the Supplement are satisfied.

We conclude the section with a remark. While the proposed prior specification in [Equation 2.2](#) assumes a symmetric mixing weight  $\omega_{1,n} = \omega_{2,n} = \omega_n$ , an alternative and more general formulation is given by

$$\begin{aligned} \theta^{(i)} \mid \tau_{0,i}, \tau_{1,i} &\sim (1 - \omega_{1,n}) \pi_0(\theta^{(i)} \mid \tau_{0,i}^2) + \omega_{1,n} \pi_1(\theta^{(i)} \mid \tau_{1,i}^2) \quad \text{under } H_0, \\ \theta^{(i)} \mid \tau_{0,i}, \tau_{1,i} &\sim \omega_{2,n} \pi_0(\theta^{(i)} \mid \tau_{0,i}^2) + (1 - \omega_{2,n}) \pi_1(\theta^{(i)} \mid \tau_{1,i}^2) \quad \text{under } H_1, \end{aligned} \quad (3.5)$$

In this case, the non-private Bayes factor is bounded by

$$\left[ \frac{\omega_{1,n}}{1 - \omega_{2,n}}, \frac{1 - \omega_{2,n}}{\omega_{1,n}} \right],$$

and one may define

$$a_n = \log\left(\frac{1 - \omega_{2,n}}{\omega_{1,n}}\right), \quad b_n = \log\left(\frac{\omega_{1,n}}{1 - \omega_{2,n}}\right), \quad \text{GS}_f = a_n - b_n,$$

thereby allowing separate tuning of  $(a_n)$  and  $(b_n)$  rather than a single sensitivity sequence. Although this extension offers additional flexibility, the theoretical development

becomes algebraically more involved. The symmetric specification used in this paper, equivalently  $b_n = -a_n$  and  $\text{GS}_f = 2a_n$ , leads to the closed-form expression

$$\omega_n = \frac{1}{1 + e^{kn^\beta}}, \quad 0 < \beta < 1, 0 < k \leq 1,$$

which enables a more transparent presentation of the Bayes factor consistency results. For this reason, we adopt the symmetric model in the main exposition, while noting that the general case can be treated analogously with no change to the final asymptotic conclusions.

## 4 Experiment: Test for Normal means ( $z$ or $t$ -test)

Suppose a confidential database contains samples  $x_1, \dots, x_n$  generated from a univariate  $\text{N}(\mu, \sigma^2)$  with unknown mean  $\mu$  and unknown variance  $\sigma^2$ . The database releases response to an analyst's queries, upon privatization via Laplace mechanism with a fixed privacy budget  $\varepsilon$ . Without the loss of generality, we assume  $\varepsilon \in \{1, 1.5, 2\}$ . In this context, we intend to utilize differentially private Bayes factors based on  $t$ -statistics, to test the null hypothesis against the alternative as described in Theorem 3.2, for an assumed effect size of interest  $\pi^* = 0.01$ . We set  $a_n = \log((1 - \omega_n)/\omega_n) = kn^\beta$  with  $\beta = 1 - 10^{-2}$  and  $k = 1$ , in line with the assumptions in Theorem 3.2. For the purposes of producing the power curves via numerical experiments, suppose the data  $x_1, \dots, x_n$  in the confidential database is generated from  $\text{N}(\mu, 1)$ , where  $\mu \in \{\pm 0.01, \pm 0.02, \dots, \pm 1\}$ . The sample size  $n$  varies in an increasing grid  $\{25, 50, 200, 300, 500, 1000\}$ .

Recall that, in real life application, we shall only have access to privatized log Bayes factor upon privatization via Laplace mechanism with privacy budget  $\varepsilon$ , the sample size  $n$ , and the effect size of interest  $\pi^* = 0.01$ . For the sake of demonstration, for a fixed value of the hyper-parameter  $M_n$  given  $a_n$ , we first showcase the numerical schemes to determine the size  $\alpha$  cut-off for the proposed differentially private Bayes factor. Given the sample size  $n$ , a privacy mechanism, privacy budget  $\varepsilon$ , effect size of interest  $\pi^* = 0.01$ , and the size  $\alpha$  of the test, we determine the cut off  $\gamma_{\alpha, \varepsilon}(n)$  by solving the equation (3.3), utilizing Algorithm 1 based on Monte Carlo simulations.

In Figure 1, we present the distribution of log-Bayes factor, under the null hypothesis, in both non-private and private set up with privacy budget  $\varepsilon = 1$ . In particular, we overlay the non-private and private Bayes factor cut-offs, given size  $\alpha = 0.05$ , sample size  $n = 100$  and privacy budget  $\varepsilon = 1$ , with fixed hyper-parameter  $M_n = 5$  given  $a = 3$ . Due to the insertion of the privacy noise, the null distribution of the differentially private log Bayes factor differ from the null distribution of the non-private log Bayes factor, and the size  $\alpha$  cut-off for differentially private log Bayes factor shifts to the right compared to the size  $\alpha$  cut-off for the non-private log Bayes factor, as expected.

Next, we discuss the numerical scheme to tune the hyper parameter  $M_n \in \mathcal{M}$  given  $a_n$ . For each value of the  $M_n$ , given  $(n, a_n)$ , privacy mechanism,  $\varepsilon$ , and size  $\alpha$ , we can determine the Bayes factor cut off  $\gamma_{\alpha, \varepsilon}(n)$ , as described earlier. Then, the optimal choices of the hyper parameters  $M_n$  is obtained via maximizing power of the privatized

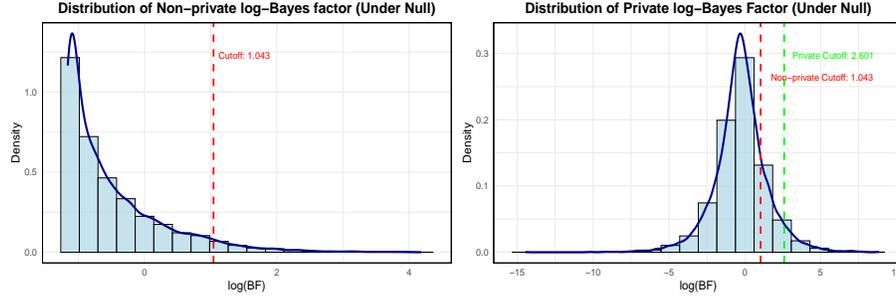


Figure 1: **Determining size  $\alpha$  Bayes factor cut-off ( $t$ -test).** We present the distribution of log-Bayes factor in non-private and private, under  $H_0$ . Non-private and private Bayes factor cut-offs, given size of the test  $\alpha = 0.05$ , sample size  $n = 100$  and privacy budget  $\varepsilon = 1$ , with fixed hyper-parameters  $M_n = 5$  given  $a = 3$ .

Bayes factor based test over  $\mathcal{M} = \{2, \dots, 10\}$ . We again accomplish this via Monte Carlo simulations, similar to Algorithm 1. To that end, for each  $M_n \in \mathcal{M}$  and assumed effect size  $\pi^*$  of interest, we draw the pseudo test statistic  $s_k^{(i)} \equiv t_k^{(i)}$  from non-central  $t_{|n_i|-1}(\lambda^{(i)})$  distributions with non-centrality parameter  $\lambda^{(i)}$ , where

$$\lambda^{(i)} \sim \omega_n \delta_0 + (1 - \omega_n) J(\tau_i^2), \quad \tau_i = \sqrt{n_i} \pi^*, \quad i \in [M_n]$$

and  $\omega_n$  satisfies the condition (iv) in (3.1.1) with  $k = 1$  and  $\beta = 1 - 10^{-2}$ . We then choose the  $\hat{M}_n$  that maximizes power of the privatised Bayes factor with size  $\alpha$  cut-off  $\gamma_{\alpha, \varepsilon}(n)$ . For tuning  $M_n$ , it is important to note that we are only sampling the pseudo test statistics from their conditional distributions under the alternative hypothesis, where the slab part of prior mode is set by the standardized effect size  $\pi^*$  of interest. In this process, we are not sampling pseudo data, but solely the test statistics which only depend on the partition specific sample size and the standardized effect size of interest. This ensures the preservation of differential privacy budget of the proposed mechanism.

We also note that the hyper-parameter tuning scheme remains applicable under a *local* prior specification. The proposed tuning scheme does not exploit any of the separation unique to the non-local prior  $J(\tau_i^2)$ ; therefore, replacing  $J(\tau_i^2)$  with a local prior leaves the tuning procedure unchanged.

In this simulation,  $\pi^* \in \{\pm 0.01, \pm 0.02, \dots, \pm 1\}$  describes the values of  $\mu$  under the alternative hypothesis. Consequently, we choose the  $\hat{M}_n$  that maximizes power based on  $N = 1000$  Monte Carlo simulations where the effect size of interest in each Monte Carlo simulation is sampled uniformly from  $\{\pm 0.01, \pm 0.02, \dots, \pm 1\}$ . In specific applications, the analysts usually have a pretty good idea about the effect size of interest, and the set of values of  $\pi^*$  that describes the alternative hypothesis should be updated accordingly. Refer to Figure 2a for a demonstration of the hyper-parameter tuning scheme.

Finally, we compare the size  $\alpha$  private Bayesian test with the size  $\alpha$  non-private Bayes factor based on  $t$ -statistics (Johnson et al., 2023a), with respect to power. We

consider an increasing grid of sample sizes  $n \in \{25, 50, 100, 200, 500\}$ . For each sample size, we proceed as described previously for determining the size  $\alpha$  cut-offs and tuning the hyper-parameters  $M_n$ . Figure 2b presents the comparison of power of the proposed differentially private Bayesian test based on  $t$ -statistic with varying values of the privacy budget  $\varepsilon$  and hyper-parameter tuning scheme described previous three paragraphs, with the non-private Bayesian test based on  $t$ -statistic proposed in (Johnson et al., 2023b). As expected, we sacrifice on power to ensure privacy, but the difference in the power of the private and non-private Bayesian tests diminishes as we increase sample size. Further, for a fixed sample size, as the value of the privacy budget  $\varepsilon$  increases, the loss in power to ensure privacy decreases, as expected.

As a sensitivity analysis, we next present simulation results under a data generating mechanism that assumes the mixture prior on  $\lambda$ . In this context, we intend to utilize differentially private Bayes factors based on  $t$ -statistics, to test the null hypothesis against the alternative as described in Theorem 3.2, for an assumed effect size of interest  $\pi^* = 0.01$ :

$$\begin{aligned} H_0 : z \mid \lambda &\sim N(\lambda, 1), & \lambda \mid \tau^2 &\sim (1 - \omega_n) \delta_0 + \omega_n J(\tau^2) \\ H_1 : z \mid \lambda &\sim N(\lambda, 1), & \lambda \mid \tau^2 &\sim \omega_n \delta_0 + (1 - \omega_n) J(\tau^2), \quad \tau > 0, \end{aligned}$$

For the purposes of producing the power curves via numerical experiments, suppose the data  $x_1, \dots, x_n$  in the confidential database is generated from  $N(\mu, 1)$ , where

$$\mu \in \{\pm 0.01, \pm 0.02, \dots, \pm 1\},$$

with probability  $(1 - \omega_n)$  and from  $N(0, 1)$  with probability  $\omega_n$ . Like earlier, We set  $\log((1 - \omega_n)/\omega_n) = kn^\beta$  with  $\beta = 1 - 10^{-2}$  and  $k = 1$ , in line with the assumptions in Theorem 3.2. The sample size  $n$  still varies in an increasing grid  $\{25, 50, 200, 300, 500, 1000\}$ . The power comparison results are presented in Figure 2c. As earlier, we sacrifice on power to ensure privacy, but the difference in the power of the private and non-private Bayesian tests diminishes as we increase sample size. Further, for a fixed sample size, as the value of the privacy budget  $\varepsilon$  increases, the loss in power to ensure privacy decreases, as expected.

Finally, we conduct a series of simulation studies to evaluate the utility of employing non-local priors in comparison to local priors. This investigation is motivated by the suggested use of the  $g$ -prior in the context of differentially private hypothesis testing for linear regression, since the  $g$ -prior is a local prior Peña and Barrientos (2024). Figure 2d presents the comparison of power of the proposed differentially private Bayesian test based on  $t$ -statistic with varying values of the privacy budget  $\varepsilon$  and hyper-parameter tuning scheme described previous tow paragraphs, with the differentially private Bayesian test with local prior (Peña and Barrientos, 2024). When contrasted with the power curves under the non-local prior shown in Figure 2b, it is evident that employing a non-local prior facilitates more rapid accumulation of evidence in favor of the true alternative hypothesis  $H_1$ . Consequently, this leads to substantially higher statistical power relative to the use of a local prior.

Similarly, a confidential database may have access to samples  $x_1, \dots, x_n$  from a univariate normal distribution  $N(\mu, 1)$  with a known variance. In this context, we would

utilize differentially private Bayes factors based on  $z$ -statistics to test the null hypothesis against the alternative. Due to the similarity of this setup with the test of univariate normal means with unknown variance, we do not provide further numerical results in this context.

Numerical experiments for the  $\chi^2$  and  $F$  tests are presented in Section 4 of the Supplementary Material.

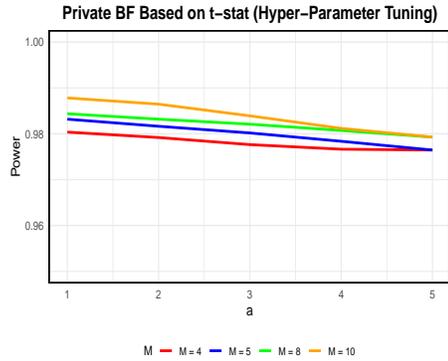
## 5 Gender differences in PHQ-8 scores from Distress Analysis Interview Corpus-Wizard of Oz

The DAIC-WOZ (Distress Analysis Interview Corpus – Wizard of Oz) dataset is a targeted subset of the larger DAIC (Distress Analysis Interview Corpus), designed to support research in the identification of mental health conditions such as depression, anxiety, and post-traumatic stress disorder (PTSD). The dataset is accessible upon request via the [DAIC-WOZ](#) platform. This data collection effort was part of a broader research project aimed at developing automated systems capable of detecting both verbal and non-verbal cues associated with psychological distress. What makes DAIC-WOZ distinctive is its use of “Ellie,” a virtual interviewer. Although Ellie appears autonomous, her interactions were actually controlled in real time by a human operator in a separate location using a methodology commonly referred to as the “Wizard of Oz” technique. Each session in the dataset includes synchronized audio and video recordings along with detailed questionnaire responses.

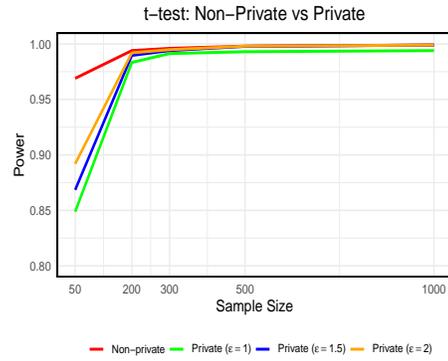
To assess the severity of depressive symptoms, each participant’s data is labeled using the PHQ-8 (Patient Health Questionnaire-8) score. The PHQ-8 is a validated self-report instrument similar to the PHQ-9, but it excludes the item on suicidal ideation due to ethical considerations. PHQ-8 scores range from 0 to 24 and are categorized as follows: 0–4 indicates none or minimal depression, 5–9 corresponds to mild depression, 10–14 represents moderate depression, 15–19 suggests moderately severe depression, and 20–24 reflects severe depression. For more details on the PHQ-8 questionnaire, refer to Figure 1 of the Supplement.

The dataset comprises  $n = 107$  participants, including 44 females and 63 males. Figure 3 presents the gender-specific density of PHQ-8 scores, revealing potential differences in the distributional patterns between males and females. These differences may reflect underlying variations in the experience, expression, or reporting of depressive symptoms across genders. It is plausible that biological gender contributes to these disparities to an extent, potentially through interactions with hormonal, neuro-biological, or psychosocial factors known to influence mental health.

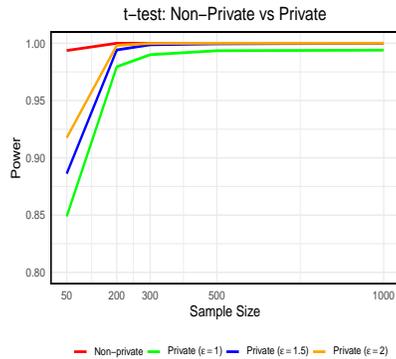
Given the observed skewness in gender-specific PHQ-8 scores, we apply a logarithmic transformation to the variable. Let the confidential DAIC dataset contain log-transformed PHQ-8 scores for males and females, denoted by  $\{x_{g,1}, \dots, x_{g,n_g}\}$ , which are assumed to be independently drawn from a univariate normal distribution  $\mathcal{N}(\mu_g, \sigma^2)$ , where  $g \in \{\text{male}, \text{female}\}$ . Our objective is to conduct hypothesis testing using differentially private Bayes factors based on two-sample  $t$ -statistics, to evaluate the null



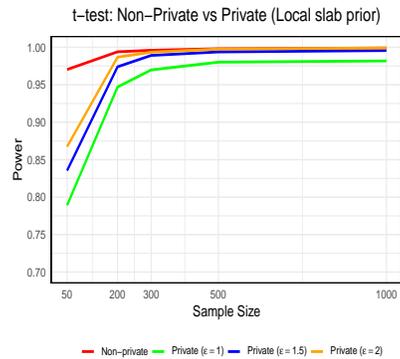
(a) **Hyper-parameter tuning in size  $\alpha$  Bayesian test ( $t$ -test).** The power of the size  $\alpha$  privatized Bayesian test for sample size  $n = 100$  and privacy budget  $\varepsilon = 1$ , with varying values of hyper-parameters  $M_n$  at different fixed values of  $a_n$ .



(b) **Power analysis of size  $\alpha$  non-private and private Bayesian  $t$  test under non-local slab prior.** Comparison of the size  $\alpha$  non-private Bayes factor based on  $t$ -statistic, and size  $\alpha$  private Bayes factor based on  $t$ -statistic with hyper-parameters set at  $\hat{M}_n$ , for varying privacy budget  $\varepsilon \in \{1, 1.5, 2\}$ .



(c) **Power analysis of size  $\alpha$  non-private and private Bayesian  $t$  test under true data generating mechanism with mixture prior on  $\lambda$ .** Comparison of the size  $\alpha$  non-private Bayes factor based on  $t$ -statistic, and size  $\alpha$  private Bayes factor based on  $t$ -statistic with hyper-parameters set at  $\hat{M}_n$ , for varying privacy budget  $\varepsilon \in \{1, 1.5, 2\}$ .



(d) **Power analysis of size  $\alpha$  non-private and private Bayesian  $t$  test under local slab prior.** Comparison of the size  $\alpha$  non-private Bayes factor based on  $t$ -statistic, and size  $\alpha$  private Bayes factor based on  $t$ -statistic with hyper-parameters set at  $\hat{M}_n$ , for varying privacy budget  $\varepsilon \in \{1, 1.5, 2\}$ .

Figure 2: Comparison of non-private and private Bayesian  $t$ -tests under different prior specifications and privacy budgets.

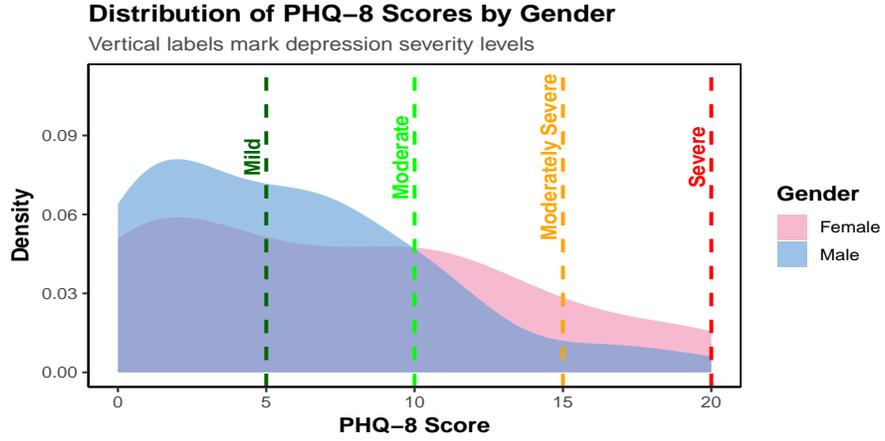


Figure 3: *DAIC-WOZ database*. The plot illustrates the gender-specific densities of the PHQ-8 scores, revealing potential differences between genders.

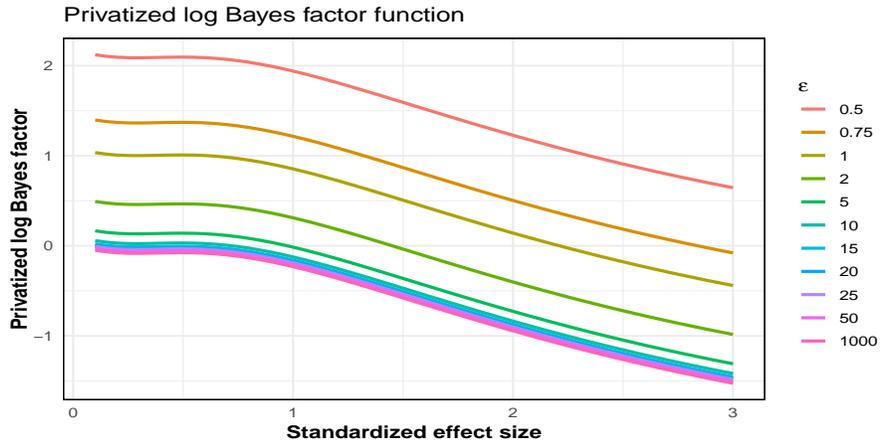


Figure 4: *DAIC-WOZ database*. Privatized log Bayes factor against the null hypotheses, as a function of standardized effect size, for varying values of privacy parameter  $\epsilon$ .

hypothesis  $H_0 : \mu_{\text{male}} = \mu_{\text{female}}$  against  $H_1 : \mu_{\text{male}} \neq \mu_{\text{female}}$ . We assume an effect size of interest defined as

$$\pi^* = \frac{\mu_{\text{male}} - \mu_{\text{female}}}{\sigma} \in \{0.1, 0.2, \dots, 3\}.$$

In this framework, responses to analysts' queries are released through a differential privacy mechanism. Specifically, the database employs the Laplace mechanism with a fixed privacy budget  $\varepsilon$ . Without loss of generality, we consider  $\varepsilon \in \{0.5, 1, 2, 5, 10, 15, 20, 25, 1000\}$ . We recall that, given the sample size  $n = 107$ , a privacy mechanism, privacy budget  $\varepsilon$ , effect size of interest  $\pi^*$ , and the size  $\alpha = 0.05$  of the test, one can determine the cut-off  $\gamma_{\alpha, \varepsilon}(n)$  of the size  $\alpha$  Bayesian test, by solving the equation (3.3), utilizing Algorithm 1 based on Monte Carlo simulations.

Next, we discuss the numerical scheme to tune the hyper parameter  $M_n \in \mathcal{M}$ . For each value of the  $M_n$ , given  $(n, a_n)$  as in Section 4, privacy mechanism,  $\varepsilon$ , and size  $\alpha$ , we can determine the Bayes factor cut off  $\gamma_{\alpha, \varepsilon}(n)$ , as described earlier. Then, the optimal choices of the hyper parameters  $M_n$  is obtained via maximizing power of the privatized Bayes factor based test over  $\mathcal{M} = \{1, \dots, 5\}$ . We again accomplish this via Monte Carlo simulations, similar to Algorithm 1. To that end, for each  $M_n \in \mathcal{M}$  and assumed effect size  $\pi^*$  of interest, we draw the pseudo test statistic  $s_k^{(i)} \equiv t_k^{(i)}$  from non-central  $t_{|n_{1i}|+|n_{2i}|-1}(\lambda^{(i)})$  distributions with non-centrality parameter  $\lambda^{(i)}$ , where

$$\lambda^{(i)} \sim \omega_n \delta_0 + (1 - \omega_n) J(\tau_i^2), \quad \tau_i = \sqrt{|n_{1i}| + |n_{2i}|} \pi^*, \quad i \in [M_n]$$

and  $\omega_n$  satisfies condition in lieu with the condition (iv) in (3.1.1) with  $k = 1$  and  $\beta = 1 - 10^{-2}$ . We then choose the  $\hat{M}_n$  that maximizes power of the privatized Bayes factor with size  $\alpha$  cut-off  $\gamma_{\alpha, \varepsilon}(n)$ . For tuning  $M_n$ , it is important to note that we are only sampling the pseudo test statistics from their conditional distributions under the alternative hypothesis, where the slab part of prior mode is set by the standardized effect size  $\pi^*$  of interest. In this process, we are not sampling pseudo data, but solely the test statistics which only depend on the partition specific sample size and the standardized effect size of interest. This ensures the preservation of differential privacy budget of the proposed mechanism.

In this case study,  $\pi^* \in \{\pm 0.1, \pm 0.2, \dots, \pm 3\}$  describes the values of parameter of interest under the alternative hypothesis. In specific applications, the analysts usually have a pretty good idea about the effect size of interest, and the set of values of  $\pi^*$  that describes the alternative hypothesis should be updated accordingly. Finally, we compare the performance of size- $\alpha$  private Bayesian tests, based on  $t$ -statistics, across varying values of the privacy budget  $\varepsilon$ , in terms of the strength of evidence against the null hypothesis. For each value of  $\varepsilon$ , we follow the procedure described previously to determine the size- $\alpha$  decision thresholds and to tune the hyperparameters  $M_n$ . The comparison results are presented in Figure 4. For small effect sizes of interest, the test yields moderate evidence against the null hypothesis. However, as the assumed effect size increases, the evidence against the null diminishes sharply. This underscores that, for small effect sizes of interest, there are notable gender-specific differences in log PHQ-8 scores. However, these differences are not discernible when considering only relatively

larger effect sizes. Furthermore, increasing the privacy budget  $\varepsilon$  leads to a consistent decrease in the strength of evidence against the null hypothesis, which is expected given the additional noise introduced by stronger privacy constraints.

## 6 Discussion

In scientific applications where data are not confidential, Bayesian hypothesis tests are commonly employed in reporting outcomes. The widespread adoption of Bayesian tests stems from their ability to address critical limitations of p-values, particularly lack of interpretability and failure to provide a quantitative measure of evidence supporting competing hypotheses. In this article, we presented a novel differentially private Bayesian hypothesis testing methodology, coherently embedding popular tools from existing literature (e.g., subsample and aggregate, truncation of the test function, etc.) within a generative model framework. In contrast to existing frequentist approaches, the proposed method enables us to systematically accumulate evidence in support of the true hypothesis—a feature desirable in reporting scientific discoveries utilizing confidential data. To mitigate the computational complexities associated with computing marginal likelihoods of the competing hypotheses arising from fully parametric models, we devised our differentially private Bayes factors based on commonly used test statistics. Under appropriate asymptotic regime, we derive the consistency rates of the proposed differentially private Bayes factors.

An important direction for future work concerns the development of privacy-preserving, data-driven procedures for selecting hyperparameters in frequentist or Bayesian hypothesis testing. In the present article, the choice of the hyperparameter  $M_n$  is intentionally not data-dependent, since such dependence would consume additional privacy budget. This design choice is consistent with current practice in the differential privacy literature (e.g., Barrientos et al. (2019)). For instance, in the context of the  $t$ -test setting Section 4, we adopt a simulation-based calibration strategy: for each candidate value  $M_n \in \mathcal{M}$  given  $a_n$ , and for fixed sample size  $n$ , privacy parameter  $\varepsilon$ , and nominal size  $\alpha$ , we compute the Bayes factor rejection threshold  $\gamma_{\alpha, \varepsilon}(n)$  and select  $\hat{M}_n$  as the maximizer of the empirical power of the privatized Bayes factor test over the grid  $\mathcal{M}$ . Importantly, the simulation draws only from the conditional distribution of the test statistics under the alternative, and does not require sampling pseudo data, thereby preserving the differential privacy budget of the mechanism.

Although the finite sample frequentist operating characteristics of the proposed tests are explored via several numerical experiments, concrete theoretical study of the finite sample properties of the proposed tests is beyond the scope of the current article and presents an interesting avenue for future inquiry. Moreover, developing differentially private Bayesian tests tailored to scenarios beyond those described in this work—such as sequential testing (Schönbrodt et al., 2016) and multiple comparison (Berry and Hochberg, 1999)—holds significant practical utility.

## Description of the supplementary material

Supplementary material to “Differentially private Bayesian tests” (Chakraborty and Datta, 2025) contains proof of theoretical results and remarks from Sections 2.2 and 3 in the main document. It also includes the closed-form Bayes factor expressions for the  $\chi^2$  and  $F$  test statistics, together with the corresponding convergence rates. In addition, Section 4 of the Supplementary Material presents numerical experiments for the  $\chi^2$  and  $F$  tests, along with log-scale visualizations for the  $z$ ,  $t$ ,  $\chi^2$ , and  $F$ -based Bayes factor evaluations in Section 5.

### Acknowledgments

There was no external or internal funding for this work.

## References

- Abramowitz, M. (1974). *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*,. USA: Dover Publications, Inc. 16
- Alabi, D. G. and Vadhan, S. P. (2023). “Differentially Private Hypothesis Testing for Linear Regression.” *Journal of Machine Learning Research*, 24(361): 1–50.  
URL <http://jmlr.org/papers/v24/23-0045.html> 1
- Amitai, G. and Reiter, J. (2018). “Differentially private posterior summaries for linear regression coefficients.” *Journal of Privacy and Confidentiality*, 8(1).  
URL <https://journalprivacyconfidentiality.org/index.php/jpc/article/view/683> 5, 6, 7
- Awan, J. and Slavković, A. (2018). “Differentially Private Uniformly Most Powerful Tests for Binomial Data.” In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS’18, 4212–4222. Red Hook, NY, USA: Curran Associates Inc. 1, 2
- Baraheem, S. and Yao, Z. (2022). “A Survey on Differential Privacy with Machine Learning and Future Outlook.” 4
- Barrientos, A. F., Reiter, J. P., Machanavajjhala, A., and Chen, Y. (2019). “Differentially Private Significance Tests for Regression Coefficients.” *Journal of Computational and Graphical Statistics*, 28(2): 440–453.  
URL <https://doi.org/10.1080/10618600.2018.1538881> 5, 12, 19, 27
- Bernstein, G. and Sheldon, D. (2019). *Differentially private Bayesian linear regression*. Red Hook, NY, USA: Curran Associates Inc. 5
- Berry, D. A. and Hochberg, Y. (1999). “Bayesian perspectives on multiple comparisons.” *Journal of Statistical Planning and Inference*, 82(1): 215–227.  
URL <https://www.sciencedirect.com/science/article/pii/S0378375899000440> 27
- Chakraborty, A. and Datta, S. (2025). “Supplementary Material to "Differentially private Bayesian tests".” Both the authors contributed equally. 28

- Chatterjee, D., Maitra, T., and Bhattacharya, S. (2020). “A Short Note on Almost Sure Convergence of Bayes Factors in the General Set-Up.” *The American Statistician*, 74(1): 17–20.  
URL <https://doi.org/10.1080/00031305.2017.1397548> 6, 15
- Chib, S. and Kuffner, T. A. (2016). “Bayes factor consistency.” 6, 15, 16
- Couch, S., Kazan, Z., Shi, K., Bray, A., and Groce, A. (2019). “Differentially Private Nonparametric Hypothesis Testing.” In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19*, 737–751. New York, NY, USA: Association for Computing Machinery.  
URL <https://doi.org/10.1145/3319535.3339821> 1
- Datta, S., Shudde, R., and Johnson, V. E. (2024). “Bayes Factors Based on Test Statistics and Non-Local Moment Prior Densities.”  
URL <https://arxiv.org/abs/2310.16213> 3, 11, 15
- Dimitrakakis, C., Nelson, B., Zhang, Z., Mitrokotsa, A., and Rubinstein, B. I. P. (2017). “Differential Privacy for Bayesian Inference through Posterior Sampling.” *Journal of Machine Learning Research*, 18(11): 1–39.  
URL <http://jmlr.org/papers/v18/15-257.html> 5
- Dwork, C. (2006). “Differential Privacy.” *Proceedings of the 33rd International Conference on Automata, Languages and Programming (ICALP)*, 1–12. 1, 4, 5
- Heikkilä, M., Jälkö, J., Dikmen, O., and Honkela, A. (2019). “Differentially Private Markov Chain Monte Carlo.” In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.  
URL [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/074177d3eb6371e32c16c55a3b8f706b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/074177d3eb6371e32c16c55a3b8f706b-Paper.pdf) 5
- Hu, J., Williams, M. R., and Savitsky, T. D. (2023). “Mechanisms for Global Differential Privacy under Bayesian Data Synthesis.”  
URL <https://arxiv.org/abs/2205.05003> 5
- Jeffreys, H. (1961). *Theory of Probability*. Oxford, UK: Oxford University Press, 3rd edition. 2
- Johnson, V. E. (2005). “Bayes Factors Based on Test Statistics.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(5): 689–701.  
URL <http://www.jstor.org/stable/3647614> 3, 6, 17
- (2008). “Properties of Bayes Factors Based on Test Statistics.” *Scandinavian Journal of Statistics*, 35(2): 354–368.  
URL <http://www.jstor.org/stable/41548598> 15
- Johnson, V. E., Pramanik, S., and Shudde, R. (2023a). “Bayes factor functions for reporting outcomes of hypothesis tests.” *Journal of the American Statistical Association*. Advance online publication.  
URL <https://doi.org/10.1080/01621459.2023.2176534> 2, 3, 21

- (2023b). “Bayes factor functions for reporting outcomes of hypothesis tests.” *Proceedings of the National Academy of Sciences*, 120(8): e2217331120.  
URL <https://www.pnas.org/doi/abs/10.1073/pnas.2217331120> 3, 11, 13, 14, 17, 18, 22
- Johnson, V. E. and Rossell, D. (2010a). “On the Use of Non-Local Prior Densities in Bayesian Hypothesis Tests.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 72(2): 143–170.  
URL <http://www.jstor.org/stable/40541581> 3
- (2010b). “On the Use of Non-Local Prior Densities in Bayesian Hypothesis Tests.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 72(2): 143–170.  
URL <http://www.jstor.org/stable/40541581> 4, 11, 13, 16, 17
- Kass, R. E. and Raftery, A. E. (1995). “Bayes Factors.” *Journal of the American Statistical Association*, 90(430): 773–795. 2
- Kazan, Z., Shi, K., Groce, A., and Bray, A. (2023). “The Test of Tests: A Framework For Differentially Private Hypothesis Testing.” 2, 8
- Kim, I. and Schrab, A. (2023). “Differentially Private Permutation Tests: Applications to Kernel Methods.” Hal-04276141.  
URL <https://hal.archives-ouvertes.fr/hal-04276141> 1
- Kwak, S. W., Ahn, J., Lee, J., and Park, C. (2021). “Differentially Private Goodness-of-Fit Tests for Continuous Variables.” *Econometrics and Statistics*.  
URL <https://www.sciencedirect.com/science/article/pii/S2452306221001143> 1
- Morey, R. D. and Rouder, J. N. (2011). “Bayes factor approaches for testing interval null hypotheses.” *Psychological Methods*, 16(4): 406–419. 2
- Narayanan, S. (2022). “Private High-Dimensional Hypothesis Testing.” *Proceedings of Machine Learning Research*, 178: 1–49. 35th Annual Conference on Learning Theory. 1
- Peña, V. and Barrientos, A. F. (2024). “Differentially private methods for managing model uncertainty in linear regression models.” *J. Mach. Learn. Res.*, 25(1). 5, 6, 7, 8, 22
- Peña, V. and Barrientos, A. F. (2023). “Differentially Private Hypothesis Testing with the Subsampled and Aggregated Randomized Response Mechanism.” 2, 5, 6, 7, 8
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). “Bayesian t-tests for accepting and rejecting the null hypothesis.” *Psychonomic Bulletin & Review*, 16(2): 225–237. 2
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., and Perugini, M. (2016). “Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences.” *Psychological Methods*, 21(2): 206–226. Manuscript accepted for publication. 27

- Swanberg, M., Globus-Harris, I., Griffith, I., Ritz, A., Groce, A., and Bray, A. (2019). “Improved Differentially Private Analysis of Variance.” *Proceedings on Privacy Enhancing Technologies*, 2019(3): 310–330. [1](#)
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). “Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method.” *Cognitive Psychology*, 60(3): 158–189. [2](#)
- Wasserstein, R. L. and Lazar, N. A. (2016). “The ASA’s statement on p-values: Context, process, and purpose.” *The American Statistician*, 70(2): 129–133. [2](#), [6](#)