Statistica Sinica **32** (2022), 787-823
doi:https://doi.org/10.5705/ss.202019.0450

GENERAL ROBUST BAYES PSEUDO-POSTERIORS: EXPONENTIAL CONVERGENCE RESULTS WITH APPLICATIONS

Abhik Ghosh[1], Tuhin Majumder[2] and Ayanendranath Basu[2]

[1]*Indian Statistical Institute and* [2]*North Carolina State University*

Abstract: Although Bayesian inference is a popular paradigm among a large segment of scientists, including statisticians, most applications consider objective priors and need critical investigations. While it has several optimal properties, Bayesian inference lacks robustness against data contamination and model misspecification, which becomes a problem when using objective priors. As such, we present a general formulation of a Bayes pseudo-posterior distribution that leads to robust inference. Exponential convergence results related to the new pseudo-posterior and the corresponding Bayes estimators are established under a general parametric setup, and illustrations are provided for independent stationary and nonhomogeneous models. Several additional details and properties of the procedure are described, including estimation under fixed-design regression models.

Key words and phrases: Bayesian linear regression, density power divergence, exponential convergence, logistic regression, robust Bayes pseudo-posterior.

1. Introduction

Bayesian analysis is arguably one of the most popular statistical paradigms, with applications across different scientific disciplines. It is widely preferred by many nonstatisticians owing to its nice interpretability and incorporation of prior knowledge. From a statistical point of view, it is widely accepted, even among many nonBayesians, because of its nice optimal (asymptotic) properties. Bayesian inference is built on the well-known Bayes theorem, described in the celebrated 1763 paper by Thomas Bayes, and combines prior knowledge with experimental evidence to produce posterior conclusions. However, Bayesian inference has also been criticized, with several debates still ongoing (Efron (2013)). In addition to the controversies related to its internal logic (Halpern (1999); Dupre and Tipler (2009)), a major practical drawback of Bayesian inference is its nonrobust nature against misspecification in models (including data contamination and outliers) and priors, as has been extensively observed in the literature;

Corresponding author: Ayanendranath Basu, Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India. Email: ayanbasu@isical.ac.in.

see Berk (1966), Weiss (1996), Millar and Stewart (2007), De Blasi and Walker (2012), and Owhadi, Scovel and Sullivan (2015), and the references therein. An optimal solution to this problem has been developed mainly for prior misspecifications (Berger (1994); Berger and Berliner (1986); Gelfand and Dey (1991); Dey and Birmiwal (1994); Delampady and Dey (1994); Gustafson and Wasserman (1995); Ghosh, Delampady and Samanta (2006)), with Bayesians traditionally viewing the model as perfect for the given data. Thus, the possibility of model misspecification and data contamination has been generally ignored, until some very recent publications, some of which we describe later in this section.

In applying Bayesian inference to complicated data sets, we need to use complex and sophisticated models, which are highly prone to misspecification or data contamination. In reality, when "all models are wrong," the Bayesian philosophy of refining the fixed model adaptively (Gelman, Meng and Stern (1996)) often fails to handle complex scenarios or leads to "a model as complex as the data" (Wang and Blei (2018)). Data contamination can lead to erroneous posterior conclusions. The problem becomes clearer, but pernicious in the case of inferences with objective or reference priors. For example, the Bayes estimate of the mean of a normal model, with any objective prior and symmetric loss function, is the highly nonrobust sample mean. Of greater concern, as noted by Efron (2013), is that most recent Bayesian inference applications hinge on objective priors, and so always need to be scrutinized carefully, sometimes even from a frequentist perspective. The posterior nonrobustness against model misspecification and data contamination makes the process vulnerable, and we clearly need a solution to this problem.

From a true Bayesian perspective, there are a few solutions to the problem of model misspecification (Ritov (1985, 1987); Sivaganesan (1993); Shyamalkumar (2000)). However, most, if not all, assume that the perturbation in the model is known beforehand, such as gross error contaminated models with a known contamination proportion ϵ. For modern complex data sets, this is rarely meaningful. Several recent publications are motivated by the need to safeguard Bayesian inference against model misspecification by relying on a generalized (pseudo) posterior expressed in terms of a loss function and a tuning parameter η (Alquier and Lounici (2011); Catoni (2007); Jiang and Tanner (2008); Walker and Hjort (2001); Kleijn and Van der Vaart (2006); Gruenwald and van Ommen (2017); Holmes and Walker (2017); Ramamoorthi et al. (2015); De Blasi and Walker (2012)). This approach, referred to as the PAC-Bayesian approach generated from Gibb's posterior, has been quite successful in regression and other supervised classification problems with misspecified model assumptions. However, the

resulting inference is not robust against outliers with respect to a specified model that is correct for the majority of the data. This is because every sample observation, including outliers, receives equal weight in the PAC-Bayesian approach and, hence, it closely resembles a robust nonparametric analysis; see Ghosh and Basu (2016a).

To achieve robustness against data contamination (outliers) in Bayesian inference, some attempts have been made to develop alternative solutions by linking Bayesian inference suitably to the frequentist concept of robustness. In the frequentist sense, there are two major approaches to achieve robustness, namely using heavy-tailed distributions (e.g., using a t-distribution instead of a normal distribution), and using new (robust) inference methodologies (Hampel et al. (1986); Basu, Shioya and Park (2011)). The first approach has been adopted by some Bayesian scientists; see Andrade and O'Hagan (2006, 2011) and Desgagne (2013) among others. However, the difficulty with this approach is the availability of appropriate heavy-tailed alternatives in complex scenarios, and it indeed does not solve the nonrobustness of a Bayesian inference for a specified model (which might have a lighter tail). The second approach serves the purpose, but differs in the strictest probabilistic sense from the Bayesian philosophy, because one needs to alter the posterior density appropriately to achieve robustness against data contamination or model misspecification. The resulting modified posteriors are generally referred to as pseudo-posterior densities. Various pseudo-posteriors have been proposed by Greco, Racugno and Ventura (2008), Agostinelli and Greco (2013), Hooker and Vidyashankar (2014), Ghosh and Basu (2016a), Danesi et al. (2016), Atkinson, Corbellini and Riani (2017), and Nakagawa and Hashimoto (2017), but all primarily consider independent stationary models and have different pros and cons. Another recent attempt, between these two approaches, has been proposed by Wang and Blei (2018), who transformed the given model to a localized model involving hyperparameters to be estimated using the empirical Bayes approach.

1.1. Background: $R^{(\alpha)}$-posterior for independent and identically distributed setup

We consider a particular pseudo-posterior originally proposed by Ghosh and Basu (2016a) in the independently and identically distributed (i.i.d.) setup. This choice is motivated by its several nice properties and its potential for extension to more general setups. As a brief description, consider n i.i.d. random variables $X_1, \ldots, X_n$ taking values in a measurable space $(\chi, \mathcal{B})$. Assume that there is an underlying true probability space $(\Omega, \mathcal{B}_\Omega, P)$ such that, for $i = 1, \ldots, n$, X_i is $\mathcal{B}/$

Ω measurable, independent with respect to P, and has an induced distribution $G(x)$ with an absolutely continuous density $g(x)$ with respect to a dominating σ-finite measure $\lambda(dx)$. We model G by a parametric family $\{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, which is assumed to be absolutely continuous with respect to λ having density $f_{\boldsymbol{\theta}}$. Consider a prior density for $\boldsymbol{\theta}$ over the parameter space Θ given by $\pi(\boldsymbol{\theta})$. Ghosh and Basu (2016a) defined a robust pseudo-posterior density, namely the $R^{(\alpha)}$-posterior density of $\boldsymbol{\theta}$, given the sample $\underline{\boldsymbol{x}}_n = (x_1, \ldots, x_n)^T$ on the random variable $\underline{\boldsymbol{X}}_n = (X_1, \ldots, X_n)^T$, as

$$\pi_n^{(\alpha)}(\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n) = \frac{\exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi(\boldsymbol{\theta})}{\int \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}'))\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}'}, \qquad \alpha \geq 0, \tag{1.1}$$

where $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ is the α-likelihood of $\underline{\boldsymbol{x}}_n$ given by

$$q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \frac{1}{\alpha}\sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\alpha}(x_i) - \frac{n}{1+\alpha}\int f_{\boldsymbol{\theta}}^{1+\alpha} - \frac{n}{\alpha} = \sum_{i=1}^{n} q_{\boldsymbol{\theta}}^{(\alpha)}(x_i), \tag{1.2}$$

with G_n being the empirical distribution based on the data and

$$q_{\boldsymbol{\theta}}^{(\alpha)}(y) = \frac{1}{\alpha}\left(f_{\boldsymbol{\theta}}^{\alpha}(y) - 1\right) - \frac{1}{1+\alpha}\int f_{\boldsymbol{\theta}}^{1+\alpha}. \tag{1.3}$$

In a limiting sense, $q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^{n}(\log(f_{\boldsymbol{\theta}}(x_i)) - 1)$, which is the usual log-likelihood (plus a constant); thus, the $R^{(0)}$-posterior is just the ordinary Bayes posterior. The idea came from a frequentist robust estimator, the minimum density power divergence (DPD) estimator (MDPDE) of Basu et al. (1998), which has proven to be a useful robust generalization of the maximum likelihood estimator (MLE); see Ghosh and Basu (2016a) for details. The similarity of this approach (at $\alpha > 0$) to the usual Bayes posterior (at $\alpha = 0$) is that it does not require nonparametric smoothing, as some other pseudo-posteriors do, and it is additive in the data so that the posterior update is easy with new observations. Ghosh and Basu (2016a) demonstrate its robustness and prove a Bernstein–von Mises-type limiting result under the i.i.d. setup.

1.2. The contribution of this study

We provide a generalization of the $R^{(\alpha)}$-posterior density for a general parametric model setup beyond i.i.d. data, using a suitable structural definition of the α-likelihood function, and derive the exponential convergence results associated with the new pseudo-posterior for the general setup. These, in fact, generalize the

corresponding results for the usual Bayes posterior (Barron (1988)) for the $R^{(\alpha)}$-posterior, and their advantages are illustrated by means of several applications. Our major contributions are summarized as follows:

- This study is the first to define a robust pseudo-posterior for the general class of parametric models with a finite set of parameters. Previous works on pseudo-posteriors are confined to the i.i.d. setup or to a particular example of a non-i.i.d. case. Our model setup is extremely general to cover the i.i.d. case and all types of nonhomogeneous and dependent observations, provided the inference is performed based on a finite set of parameters. We define a robust $R^{(\alpha)}$-posterior and the associated estimators for this general class of statistical inference problems, covering a wide range of applications.

- To illustrate the wide applicability of our proposal, we explicitly present the forms of the $R^{(\alpha)}$-posterior and the α-likelihood function for several important cases, such as independent nonhomogeneous data, including linear and logistic regressions, time series and Markov models, diffusion processes, and so on. Our $R^{(\alpha)}$-posteriors also contain the usual Bayes posterior at $\alpha \to 0$ and, hence, provide a direct generalization of the latter at $\alpha > 0$.

- All existing pseudo-posteriors sacrifice the conditional probability interpretation of the usual Bayes theory. This study is the first to discuss a pseudo-posterior, namely the $R^{(\alpha)}$-posterior, that retains this conditional probability interpretation with respect to a suitably modified model and modified prior; indeed the $R^{(\alpha)}$-posterior becomes the ordinary Bayes posterior for such a modified setup (Remark 1). We also introduce the $R^{(\alpha)}$-marginal density of data, a robust generalization of the usual marginal.

- Beyond the methodological proposals, we also establish the theoretical properties of the proposed $R^{(\alpha)}$-posterior under the fully general parametric setup. We study the asymptotic properties of the $R^{(\alpha)}$-marginal and the corresponding joint density of the data and the parameters. We also derive the exponential convergence of the $R^{(\alpha)}$-posterior probabilities and, hence, the exponential consistency of the associated $R^{(\alpha)}$-Bayes estimators under the fully general setup. To the best of our knowledge, such an optimal asymptotic property is not available for any other pseudo-posterior.

- The assumptions needed for our theoretical derivations are extensions of those required for the classical Bayes theory (Barron (1988)). They are based on the usual concepts of information denseness, merging of distributions in probability, (modified) prior negligibility, and the existence of

uniform exponential consistent tests. We further simplify these conditions for the i.i.d. and nonhomogeneous setups, and verify them for common examples, such as the linear regression with a known or an unknown error variance and the logistic regression models. Although the initial set of conditions under the general parametric models look more stringent than those in the current literature, we show that they hold under very mild conditions in common examples. For example, for linear or logistic regressions, they are seen to hold only under the boundedness conditions on the fixed design matrix and the positive definiteness of the associated variance matrix.

- We separately examine the interesting cases of discrete priors under the i.i.d. setup, and the associated maximum $R^{(\alpha)}$-posterior estimator with their exponential consistency.

- Finally, to bridge the gap between the theoretical developments and their practical applicability, we discuss several important practical issues, such as the computation of the $R^{(\alpha)}$-posterior and the associated estimates, and the choice of the tuning parameter α. The usefulness of our proposal is illustrated numerically for linear regressions with known and unknown error variances and a logistic regression, along with the corresponding algorithms and R code.

For brevity, all proofs and the R-code are given in the online Supplementary Material.

2. A general form of the $R^{(\alpha)}$-posterior distribution

In order to extend the $R^{(\alpha)}$-posterior density to a more general setup, let us assume that the random variable $\underline{\boldsymbol{X}}_n$ is defined on a general measurable space $(\boldsymbol{\chi}_n, \mathcal{B}_n)$, for each n (sample size). In addition, we assume there is an underlying true probability space $(\Omega, \mathcal{B}_\Omega, P)$ such that, for each $n \geq 1$, $\underline{\boldsymbol{X}}_n$ is $\mathcal{B}_n/\Omega$ measurable and its induced distribution $G^n(\underline{\boldsymbol{x}}_n)$ is absolutely continuous with respect to some σ-finite measure $\lambda^n(d\underline{\boldsymbol{x}}_n)$ with a "true" probability density $g^n(\underline{\boldsymbol{x}}_n)$. We wish to model it using a parametric family of distributions $\mathcal{F}_n = \{F^n(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n \subseteq \mathbb{R}^p\}$, where the elements of $\mathcal{F}_n$ are assumed to be absolutely continuous with respect to λ^n, with density $f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$, for each n. Note that we have not assumed that the parameter space Θ_n is independent of the sample size n. Similarly, the prior measure $\pi_n(\boldsymbol{\theta})$ on Θ_n may be n-dependent, with $\pi_n(\Theta_n) \leq 1$. Consider a σ-field $\mathcal{B}_{\Theta_n}$ on the parameter space Θ_n. Generalizing from (1.2), we propose defining the α-likelihood function $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ such

that

$$q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) := \lim_{\alpha\downarrow 0} q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \log f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) - n, \text{ for all } \underline{\boldsymbol{x}}_n \in \chi_n. \tag{2.1}$$

Our definition should guarantee that the α-likelihood, as a function of $\boldsymbol{\theta}$, is $\mathcal{B}_{\Theta_n}$ measurable for each $\underline{\boldsymbol{x}}_n$, and jointly $\mathcal{B}_n \times \mathcal{B}_{\Theta_n}$ measurable when both $\underline{\boldsymbol{X}}_n$ and $\boldsymbol{\theta}$ are random. Then, for this general setup, we define the corresponding $R^{(\alpha)}$-posterior probabilities as

$$\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n) = \frac{\int_{A_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\Theta_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}}, \qquad A_n \in \mathcal{B}_{\Theta_n}, \tag{2.2}$$

whenever the denominator is finitely defined and is positive; otherwise we may define it arbitrarily, for example, $\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n) = \pi_n(A_n)$. Definition (2.1) ensures that $\pi_n^{(0)}$ is the usual Bayes posterior.

For a useful alternative representation, we define $Q_n^{(\alpha)}(S_n|\boldsymbol{\theta}) := \int_{S_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))d\underline{\boldsymbol{x}}_n$, $M_n^{(\alpha)}(S_n, A_n) := \int_{A_n} Q_n^{(\alpha)}(S_n|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}$, and $M_n^{(\alpha)}(S_n) := M_n^{(\alpha)}(S_n, \Theta_n)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)$, for $S_n \in \mathcal{B}_n$ and $A_n \in \mathcal{B}_{\Theta_n}$. In the following, we assume that the model and priors are chosen to satisfy $0 < M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n) < \infty$. Then, the last two measures have densities with respect to $\lambda^n(d\underline{\boldsymbol{x}}_n)$ given by $m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n) = \int_{A_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}$ and $m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n) = m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, \Theta_n)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)$, respectively. Clearly, $m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)$ is a proper probability density, which we refer to as the $R^{(\alpha)}$-marginal density of $\underline{\boldsymbol{X}}_n$; the associated $R^{(\alpha)}$-marginal distribution is $M_n^{(\alpha)}(\cdot)$. At $\alpha > 0$, it provides a robust version of the ordinary Bayes marginal $m_n^{(0)}(\underline{\boldsymbol{x}}_n)$. Whenever $0 < m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n) < \infty$, we can re-express the $R^{(\alpha)}$-posterior probabilities (2.2) in terms of this $R^{(\alpha)}$-marginal density as $\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n) = m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n)/m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, \Theta_n) = (m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n))/m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)$, for $A_n \in \mathcal{B}_{\Theta_n}$. Then, the $R^{(\alpha)}$-Bayes joint posterior law of the parameter $\boldsymbol{\theta}$ and the data $\underline{\boldsymbol{X}}_n$ is defined as

$$L_n^{(\alpha)Bayes}(d\boldsymbol{\theta}, d\underline{\boldsymbol{x}}_n) = \pi_n^{(\alpha)}(d\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n)\, M_n^{(\alpha)}(d\underline{\boldsymbol{x}}_n) = \frac{M_n^{(\alpha)}(d\underline{\boldsymbol{x}}_n, d\boldsymbol{\theta})}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)}. \tag{2.3}$$

This provides a nice interpretation of the quantity $M_n^{(\alpha)}(S_n, A_n)$, when properly normalized, as the product measure associated with the $R^{(\alpha)}$-Bayes joint posterior distribution of $\boldsymbol{\theta}$ and $\underline{\boldsymbol{X}}_n$. At $\alpha = 0$, these simplify to the ordinary Bayes measures.

Example 1. (Independent Stationary Data). The simplest possible setup

is that of i.i.d. observations, as described in Section 1. In terms of the general notation presented above, we have $\underline{\boldsymbol{X}}_n = (X_1, \ldots, X_n)$, with its observed value $\underline{\boldsymbol{x}}_n = (x_1, \ldots, x_n)$, and the general measurable space $(\boldsymbol{\chi}_n, \mathcal{B}_n)$ is the n-fold product of $(\chi, \mathcal{B})$. Additionally, we have $G^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n G(x_i)$, $g^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n g(x_i)$, $\lambda^n(d\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n \lambda(dx_i)$, $F^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n F_{\boldsymbol{\theta}}(x_i)$, and $f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n f_{\boldsymbol{\theta}}(x_i)$, and so $\mathcal{F}_n$ is also the n-fold product of the family of individual distributions $F_{\boldsymbol{\theta}}$. Under this notation, the α-likelihood $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$, given by (1.2), satisfies the required measurability assumptions, along with the condition in (2.1).

Then, under suitable assumptions on the prior distribution, as before, the corresponding $R^{(\alpha)}$-posterior distribution is defined by (2.2), which is now equivalent to (1.1) and can be written as a product of the stationary independent terms corresponding to each x_i (additivity). Other related measures can be defined from these quantities; see Section 4.

Example 2. (Independent Nonhomogeneous Data). Suppose $X_1, \ldots, X_n$ are independently but not identically distributed random variables, where each X_i is defined on a measurable space $(\chi^i, \mathcal{B}^i)$, for $i = 1, \ldots, n$. Considering an underlying common probability space $(\Omega, \mathcal{B}_\Omega, P)$, the random variable X_i is assumed to be $\mathcal{B}^i/\Omega$ measurable and independent with respect to P, and its induced distribution $G_i(x)$ has an absolutely continuous density $g_i(x)$ with respect to some common dominating σ-finite measure $\lambda(dx)$, for each $i = 1, \ldots, n$. For each i, the true distribution G_i is modeled by a parametric family $\mathcal{F}^i = \{F_{i,\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, which is absolutely continuous with respect to λ, having density $f_{i,\boldsymbol{\theta}}$. Note that although the densities are potentially different for each i, they are assumed to share the common unknown parameter $\boldsymbol{\theta}$, leaving us with enough degrees of freedom for the estimation of $\boldsymbol{\theta}$.

This setup of independent nonhomogeneous (INH) observations covers many interesting practical problems, the most common being a regression with a fixed design. Let $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ be n fixed, k-variate design points. For each $i = 1, \ldots, n$, given $\boldsymbol{t}_i$, we independently observe x_i, which has the parametric model density $f_{i,\boldsymbol{\theta}}(x_i) = f(x_i; \boldsymbol{t}_i, \boldsymbol{\theta})$, depending on $\boldsymbol{t}_i$ through a regression structure. This can, for example, have the form

$$E(X_i) = \psi(\boldsymbol{t}_i, \boldsymbol{\beta}), \quad i = 1, \ldots, n, \tag{2.4}$$

where $\boldsymbol{\beta} \subseteq \boldsymbol{\theta}$ denotes the unknown regression coefficients, and ψ is a suitable link function. In general, the unknown parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$ may additionally contain some variance parameter σ. For the subclass of generalized linear models, we take $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = \psi(\boldsymbol{t}_i^T \boldsymbol{\beta})$ and f from the exponential family of distributions. For the

normal linear regression, we have $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = \boldsymbol{t}_i^T \boldsymbol{\beta}$, and f is the normal density with mean $\boldsymbol{t}_i^T \boldsymbol{\beta}$ and variance σ^2. Here, the underlying random variables X_i, associated with the observations x_i, have the INH structure with the common parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$ and different densities $f_{i,\boldsymbol{\theta}}$. We can further extend this setup to include heterogeneous variances (by taking different σ_i for different $f_{i,\boldsymbol{\theta}}$, but involving some common unknown parameters) as a part of our INH setup. In terms of the general notation, the random variable $\underline{\boldsymbol{X}}_n = (X_1, \ldots, X_n)$ is defined on the measurable space $(\boldsymbol{\chi}_n, \mathcal{B}_n) = \otimes_{i=1}^n (\chi^i, \mathcal{B}^i)$, and we have $G^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n G_i(x_i)$, $g^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n g_i(x_i)$, $\lambda^n(d\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n \lambda(dx_i)$, $F^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n F_{i,\boldsymbol{\theta}}(x_i)$, and $f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n f_{i,\boldsymbol{\theta}}(x_i)$, such that $\mathcal{F}_n = \otimes_{i=1}^n \mathcal{F}^i$.

Now, under this INH setup, we define the $R^{(\alpha)}$-posterior by suitably extending the definition of the α-likelihood function $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ from its i.i.d. version in (1.2), keeping in mind the general requirement (2.1). Borrowing from Ghosh and Basu (2013), who developed the MDPDE for the INH setup, and following the intuition behind the construction of the α-likelihood (1.2) of Ghosh and Basu (2016a), one possible extended definition for the α-likelihood in the INH case can be given by

$$q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^n \left[\frac{1}{\alpha} f_{i,\boldsymbol{\theta}}^{\alpha}(x_i) - \frac{1}{1+\alpha} \int f_{i,\boldsymbol{\theta}}^{1+\alpha} \right] - \frac{n}{\alpha} = \sum_{i=1}^n q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i), \quad (2.5)$$

with $q_{i,\boldsymbol{\theta}}^{(\alpha)}(y) = (1/\alpha)(f_{i,\boldsymbol{\theta}}^{\alpha}(y) - 1) - (1/(1+\alpha)) \int f_{i,\boldsymbol{\theta}}^{1+\alpha}$. Note that we have $q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^n (\log(f_{i,\boldsymbol{\theta}}(x_i)) - 1)$, satisfying the required condition in (2.1). Thus, assuming a suitable prior for $\boldsymbol{\theta}$, the $R^{(\alpha)}$-posterior for the INH observations is defined using (2.2) with $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ being given by (2.5). Note that the resulting posterior is again a product of independent, but nonhomogeneous terms; see Section 5.

Remark 1. In the first introduction of the $R^{(\alpha)}$-posterior under an i.i.d. setup (Ghosh and Basu (2016a)), it was noted that its only drawback is the loss of the probabilistic interpretation. Thus far, we have defined the $R^{(\alpha)}$-posterior differently to the conditional probability approach of the usual Bayes theory, calling it a pseudo-posterior. However, it can also be interpreted as an ordinary Bayes posterior under a suitably modified model and prior, which becomes prominent in our general setup. To see this, define an α-modified model density $\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))/Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})$ and the α-modified prior density $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) = Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)$. Both are proper densities and satisfy the required measurability assumptions whenever the relevant integrals exist finitely. Furthermore, $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ only (independent of

the data) and, hence, may be used as a prior density in a Bayesian inference; however it depends on α and the model. In particular, at $\alpha = 0$, $\widetilde{\pi}_n^{(0)}(\boldsymbol{\theta}) = \pi_n(\boldsymbol{\theta})$ and $\widetilde{q}_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$, and so represent modifications of the model and the prior, respectively, in order to achieve robustness against data contamination. Now, for any measurable $A_n \in \mathcal{B}_{\Theta_n}$, the standard Bayes (conditional) posterior probability of A_n with respect to the (α-modified) model family $\mathcal{F}_{n,\alpha} = \{\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\}$ and the (α-modified) prior $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})$ is given by $\int_{A_n} \widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})d\boldsymbol{\theta} / \int_{\Theta_n} \widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})d\boldsymbol{\theta}$, which simplifies to $\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n)$, as in (2.2).

In the following, we briefly present the forms of the α-likelihood for some other practically important model setups; however, detailed investigations are left to future research.

Example 3. (Time Series Data). Consider the true probability space $(\Omega, \mathcal{B}_\Omega, P)$ and an index set T. A measurable time series $X_t(\omega)$ is a function defined on $T \times \Omega$, which is a random variable on $(\Omega, \mathcal{B}_\Omega, P)$, for each $t \in T$. Given a time series $\{X_t(\omega) : t \in T\}$, they are assumed to be associated with an increasing sequence of sub σ-fields $\{\mathcal{G}_t\}$ and have absolute continuous densities $g(X_t|\mathcal{G}_t)$, for $t \in T$. For a stationary time series, one might take $\mathcal{G}_t = \mathcal{F}_{t-1}$, the σ-field generated by $\{X_{t-1}, X_{t-2}, \ldots\}$, for each $t \in T$. In a parametric inference, we model $g(X_t|\mathcal{G}_t)$ using a parametric density $f_{\boldsymbol{\theta}}(X_t|\mathcal{F}_{t-1})$, and try to infer the unknown parameter $\boldsymbol{\theta}$ from an observed sample $\underline{\boldsymbol{x}}_n = \{x_t : t \in \{1, 2, \ldots, n\}\}$ of size n. For example, in a Poisson autoregressive model, we assume $f_{\boldsymbol{\theta}}(x_t|\mathcal{F}_{t-1})$ to be a Poisson density with mean $\lambda_t = h_{\boldsymbol{\theta}}(\lambda_{t-1}, X_{t-1})$, for all $t \in T = \mathbb{Z}$ and some known function $h_{\boldsymbol{\theta}}$ involving the unknown parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. In the Bayesian paradigm, we additionally assume a prior density $\pi(\boldsymbol{\theta})$, and update it to obtain an inference based on the posterior density of $\boldsymbol{\theta}$, given the observed sample data. We can develop a robust Bayesian inference for any such time series model using the proposed $R^{(\alpha)}$-posterior density, provided a suitable α-likelihood function can be defined. Following the construction of the MDPDE in such time series models (Kim and Lee (2011, 2013); Kang and Lee (2014), among others), we can define the corresponding α-likelihood function as

$$q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{t=1}^{n} \left[\frac{1}{\alpha} f_{\boldsymbol{\theta}}^{\alpha}(x_t|\mathcal{F}_{t-1}) - \frac{1}{1+\alpha} \int f_{\boldsymbol{\theta}}^{1+\alpha}(x|\mathcal{F}_{t-1})dx \right] - \frac{n}{\alpha}. \quad (2.6)$$

We have $q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^{n} (\log(f_{\boldsymbol{\theta}}(x_t|\mathcal{F}_{t-1})) - 1)$, which satisfies the required

Condition (2.1). A robust $R^{(\alpha)}$-posterior inference about $\boldsymbol{\theta}$ can be developed using this α-likelihood function.

Example 4. (Markov Process). Example 3 can be easily generalized to Markov processes with stationary transitions. Consider the random variables $X_1, \ldots, X_n$ defined on the underlying true probability space $(\Omega, \mathcal{B}_\Omega, P)$, with true transition probabilities $g(X_{k+1}|X_k)$, for $k = 0, 1, 2, \ldots, n-1$, with X_0 being the initial value of the process. We model it using a parametric family of stationary probabilities $f_{\boldsymbol{\theta}}(X_{k+1}|X_k)$ depending on some unknown parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. Then, the α-likelihood function, given the sample $\underline{\boldsymbol{x}}_n = (x_1, \ldots, x_n)$, can be defined as $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{k=1}^n [(1/\alpha) f_{\boldsymbol{\theta}}^{\alpha}(x_{k+1}|x_k) - (1/(1+\alpha)) \int f_{\boldsymbol{\theta}}^{1+\alpha}(x|x_k)dx] - n/\alpha$. Clearly, it satisfies Condition (2.1), and it is possible to perform a robust $R^{(\alpha)}$-Bayes inference about $\boldsymbol{\theta}$ under this setup.

Example 5. (Diffusion Process). Consider again a (true) probability space $(\Omega, \mathcal{B}_\Omega, P)$ and an index set T. A measurable random variable X_t defined on T follows a diffusion process if $dX_t = a(X_t, \boldsymbol{\mu})dt + b(X_t, \sigma)dW_t$, for $t \geq 0$, with $X_0 = x_0$ and two known functions a and b. Here $\{W_t : t \geq 0\}$ is a standard Wiener process and the parameter of interest is $\boldsymbol{\theta} = (\boldsymbol{\mu}, \sigma)^T \in \Theta$, a convex compact subset of $\mathbb{R}^p \times \mathbb{R}^+$. This model has important applications in finance, where some inference about $\boldsymbol{\theta}$ is desired based on discretized observations $X_{t_i^n}$, for $i = 1, \ldots, n$, from the above diffusion process. In general, we assume $t_i^n = ih_n$, with $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Robust (frequentist) MDPDEs of $\boldsymbol{\theta}$ based on such observations are developed for two of its special cases, $a(X_t, \mu) = a(X_t)$ and $b(X_t, \sigma) = \sigma$ by Song et al. (2007) and Lee and Song (2013), respectively. However, whenever we have some prior knowledge about $\boldsymbol{\theta}$, quantified using a prior $\pi(\boldsymbol{\theta})$, one would apply the Bayesian approach. A robust Bayes inference can be done using our $R^{(\alpha)}$-posterior. For this purpose, note that $X_{t_i^n} = X_{t_{i-1}^n} + a(X_{t_{i-1}^n}, \boldsymbol{\mu})h_n + b(X_{t_{i-1}^n}, \sigma)\sqrt{h_n} Z_{n,i} + \Delta_{n,i}$, for $i = 1, \ldots, n$, where $\Delta_{n,i} = \int_{t_{i-1}^n}^{t_i^n} \left[a(X_s, \boldsymbol{\mu}) - a(X_{t_{i-1}^n}, \boldsymbol{\mu})\right] ds + \int_{t_{i-1}^n}^{t_i^n} \left[b(X_s, \sigma) - b(X_{t_{i-1}^n}, \sigma)\right] dW_s$ and $Z_{n,i} = h_n^{-1/2} \left(W_{t_i^n} - W_{t_{i-1}^n}\right)$. Clearly, $Z_{n,i}$ are i.i.d. standard normal variables, for $i = 1, \ldots, n$. Therefore, whenever $\Delta_{n,i}$ can be ignored in P-probability, for large enough n, $X_{t_i^n}|\mathcal{G}_{i-1}^n$, for $i = 1, \ldots, n$, behave as INH variables with densities $f_{i,\boldsymbol{\theta}}(\cdot|\mathcal{G}_{i-1}^n) \equiv N\left(X_{t_{i-1}^n} + a(X_{t_{i-1}^n}, \boldsymbol{\mu})h_n, b(X_{t_{i-1}^n}, \sigma)^2 h_n\right)$, where $\mathcal{G}_{i-1}^n$ is the σ-field generated by $\{W_s : s \leq t_i^n\}$. Then, the corresponding α-likelihood function based on the observed data $\underline{\boldsymbol{x}}_n = (x_{t_1^n}, \ldots, x_{t_n^n})$ can be derived as in Example 3. This satisfies the general requirement (2.1), and has the simplified form $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^n q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_{t_i^n})$, with

$$q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_{t_i^n}) = \begin{cases} \dfrac{1}{(2\pi b(x_{t_{i-1}^n},\sigma)^2 h_n)^{\alpha/2}} \left[\dfrac{1}{\alpha} e^{-\alpha(x_{t_i^n} - x_{t_{i-1}^n} - a(x_{t_{i-1}^n},\boldsymbol{\mu})h_n)^2/(2b(x_{t_{i-1}^n},\sigma)^2 h_n)} \right. \\ \left. - \dfrac{1}{(1+\alpha)^{3/2}} \right] - \dfrac{1}{\alpha}, & \text{if } \alpha > 0, \\ - \dfrac{\alpha \left(x_{t_i^n} - x_{t_{i-1}^n} - a(x_{t_{i-1}^n},\boldsymbol{\mu})h_n \right)^2}{2b(x_{t_{i-1}^n},\sigma)^2 h_n} - \dfrac{1}{2}\log\left(2\pi b(x_{t_{i-1}^n},\sigma)^2 h_n\right) - 1, \\ & \text{if } \alpha = 0. \end{cases}$$

The robust $R^{(\alpha)}$-posterior can be obtained easily using this α-likelihood function.

3. Exponential Convergence Results under the General Setup

Exponential consistency is an important property of a posterior (Bayes) inference. It was first demonstrated in Barron (1988), and later refined by several authors (see Ghosal, Ghosh and van der Vaart (2000); Walker (2004); Ghosal and van der Vaart (2007); Walker, Lijoi and Prunster (2007), among others). We follow the approach of Barron (1988) to show that our new robust $R^{(\alpha)}$-posterior probabilities and the corresponding parameter estimates also enjoy such asymptotic optimality properties.

3.1. Properties of the joint and marginal $R^{(\alpha)}$-Bayes distributions

Recall the general setup of Section 2, along with the α-modified model and prior densities $\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta})$ and $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})$, as defined in Remark 1. Consider the Kullback–Leibler divergence between two absolutely continuous densities f_1 and f_2 with respect to the common σ-finite measure λ, defined as $KLD(f_1, f_2) = \int f_1 \log(f_1/f_2)\, d\lambda$, and put $D_n^{(\alpha)}(\boldsymbol{\theta}) = 1/nKLD(g^n(\cdot), \widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta}))$. We define a joint (frequentist) law of $\boldsymbol{\theta}$ and $\underline{\boldsymbol{X}}_n$ given by $L_n^{*(\alpha)}(d\boldsymbol{\theta}, d\underline{\boldsymbol{x}}_n) = \pi_n^{*(\alpha)}(d\boldsymbol{\theta})\, G_n(d\underline{\boldsymbol{x}}_n)$, where the probability distribution $\pi_n^{*(\alpha)}$ of $\boldsymbol{\theta}$ on Θ_n is defined as $\pi_n^{*(\alpha)}(d\boldsymbol{\theta}) = e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}\widetilde{\pi}_n^{(\alpha)}(d\boldsymbol{\theta})/c_n$, with $c_n = \int e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}\widetilde{\pi}_n^{(\alpha)}(d\boldsymbol{\theta})$. We show that this joint law $L_n^{*(\alpha)}$ provides a frequentist large-deviation approximation to the joint $R^{(\alpha)}$-Bayes distribution (2.3) of $\boldsymbol{\theta}$ and $\underline{\boldsymbol{X}}_n$. To quantify their closeness, we consider the concept of "*merging*" for probability distributions (Barron (1988)).

Definition 1. Consider two probability distributions G_1^n and G_2^n of $\underline{\boldsymbol{X}}_n$, with densities g_1^n and g_2^n, respectively, with respect to λ^n.

- They are said to *merge in probability* if for all $\epsilon > 0$, $\lim_{n\to\infty} P(g_2^n(\underline{\boldsymbol{X}}_n)/$

$g_1^n(\underline{\boldsymbol{X}}_n) > e^{-n\epsilon}) = 1$.

- They are said to *merge with probability one* if for every $\epsilon > 0$, $P(g_2^n(\underline{\boldsymbol{X}}_n)/g_1^n(\underline{\boldsymbol{X}}_n) > e^{-n\epsilon}$ for all large $n) = 1$.

An application of Markov's inequality shows that Definition 1 is equivalent to the conditions $\lim_{n\to\infty}(1/n)\log g_2^n(\underline{\boldsymbol{X}}_n)/g_1^n(\underline{\boldsymbol{X}}_n) = 0$ in probability or with probability one, respectively. See Barron (1988, Sec. 4) for more results on merging. Additionally, we assume the following condition.

Assumption (M1): For any $\epsilon, r > 0$, there exists a positive integer N such that $\widetilde{\pi}_n^{(\alpha)}(\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\}) \geq e^{-nr}$, for all $n \geq N$.

Theorem 1. *Under Assumption* (M1)*, we have the following results:*

a) $\lim_{n\to\infty}(1/n)KLD(L_n^{*(\alpha)}, L_n^{(\alpha)Bayes}) = 0$, $\lim_{n\to\infty}(1/n)E_{G^n}[KLD(\pi_n^{*(\alpha)}(\cdot), \pi_n^{(\alpha)}(\cdot|\underline{\boldsymbol{X}}_n))] = 0$.

b) $\lim_{n\to\infty}(1/n)KLD(g^n, m_n^{(\alpha)}) = 0$, *such that* G^n *and* $M_n^{(\alpha)}$ *merge in probability.*

Although Assumption (M1) might look a bit complicated, it can be simplified using the common notion of *information denseness* of priors π_n with respect to a suitable family of model densities. This notion of information denseness is frequently used in large-sample analyses of the usual Bayesian methods, and is defined precisely below for our context.

Definition 2. Suppose $\Theta_n = \Theta$ is independent of n and we define $\bar{D}^{(\alpha)}(\boldsymbol{\theta}) = \limsup_{n\to\infty} D_n^{(\alpha)}(\boldsymbol{\theta})$. Then, the prior sequence π_n is said to be information dense at G^n with respect to $\mathcal{F}_{n,\alpha} = \{\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\}$ if there exists a finite measure $\widetilde{\pi}$ such that $\widetilde{\pi}\left(\left\{\boldsymbol{\theta} : \bar{D}^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right) > 0$, for all $\epsilon > 0$, and

$$\liminf_{n\to\infty} e^{nr} \frac{d\widetilde{\pi}_n^{(\alpha)}}{d\widetilde{\pi}}(\boldsymbol{\theta}) \geq 1, \quad \text{for all } r > 0, \boldsymbol{\theta} \in \Theta. \tag{3.1}$$

Theorem 2. *If the prior is information dense with respect to* $\mathcal{F}_{n,\alpha}$*, as in Definition* 2*, then Assumption* (M1) *holds, and hence the results of Theorem* 1 *hold.*

3.2. Consistency of the $R^{(\alpha)}$-posterior probabilities

We now prove the exponential convergence results for our robust $R^{(\alpha)}$-posterior probabilities. For measurable sets $A_n, B_n, C_n \subseteq \Theta_n$ and constants b_n, c_n, we assume the following:

(A1) A_n, B_n, and C_n together complete Θ_n; that is, $A_n \cup B_n \cup C_n = \Theta_n$, for each $n \geq 1$.

(A2) B_n satisfies $\widetilde{\pi}_n^{(\alpha)}(B_n) = M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n) \leq b_n$, for each $n \geq 1$.

(A3) $\{C_n\}$ is such that there exists $S_n \in \mathcal{B}_n$ satisfying $\lim_{n\to\infty} G^n(S_n) = 0$, and $\sup_{\boldsymbol{\theta}\in C_n} Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})/Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta}) \leq c_n$.

(A3)* $\{C_n\}$ is such that there exists $S_n \in \mathcal{B}_n$ satisfying $P(\underline{\boldsymbol{X}}_n \in S_n \text{ i.o.}) = 0$ and $\sup_{\boldsymbol{\theta}\in C_n} Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})/Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta}) \leq c_n$, where i.o. denotes "*infinitely often*."

Here, we need either Condition (A3) or Condition (A3)*, which help us to prove the convergence results in probability or with probability one, respectively. Condition (A3)* is stronger and implies (A3), but (A3) is sufficient in most cases yielding a convergence in probability-type result. In addition, if Condition (A3) holds with $c_n = e^{-nr}$ for some $r > 0$, then it ensures the existence of a uniformly exponentially consistent (UEC) test for G^n against the family of α-modified probability distributions $\{Q_n^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) : \boldsymbol{\theta} \in C_n\}$ corresponding to the α-modified model density $\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta})$ defined in Remark 1. Although complex, these conditions are straightforward extensions of those used by Barron (1988) to prove the exponential convergence of ordinary Bayes posterior probabilities; they indeed coincide at $\alpha = 0$. In particular, at $\alpha = 0$, Condition (A2) simplifies to $\pi_n(B_n) \leq b_n$; that is, B_ns have negligible prior probabilities if $b_n \to 0$, and (A3) assumes the existence of a UEC test against the models with $\boldsymbol{\theta} \in C_n$. Under these conditions, along with the concept of merging (Subsection 3.1), we have the following main theorem.

Theorem 3. *(Exponential Consistency of $R^{(\alpha)}$-Posterior Probabilities).*

(1) Suppose that G^n and $M_n^{(\alpha)}(\cdot)$ merge in probability, and let $A_n \in \mathcal{B}_{\Theta_n}$ be any sequence of sets. Then, $\limsup_{n\to\infty} P(\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n) < e^{-nr}) = 1$, for some $r > 0$, if and only if there exist $r_1, r_2 > 0$ and sets $B_n, C_n \in \mathcal{B}_{\Theta_n}$ such that (A1)–(A3) *are satisfied, with $b_n = e^{-nr_1}$ and $c_n = e^{-nr_2}$.*

(2) Suppose that G^n and $M_n^{(\alpha)}(\cdot)$ merge with probability one and let $A_n \in \mathcal{B}_{\Theta_n}$ be any sequence of sets. Then, $P(\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n) \geq e^{-nr} \text{ i.o.}) = 0$, for some $r > 0$, if and only if there exist constants $r_1, r_2 > 0$ and sets $B_n, C_n \in \mathcal{B}_{\Theta_n}$ such that Assumptions (A1), (A2), *and* (A3)* *are satisfied, with $b_n = e^{-nr_1}$ and $c_n = e^{-nr_2}$.*

Note that, for $\alpha = 0$, Theorem 3 coincides with the classical exponential convergence results of ordinary Bayes posterior probabilities, as proved in Barron

(1988). Our theorem generalizes this for the robust $R^{(\alpha)}$-posterior probabilities under suitable conditions. Hence, the $R^{(\alpha)}$-posterior distribution, in addition to yielding robust results under data contamination, is asymptotically optimal in the same exponential rate as the ordinary posterior for all $\alpha \geq 0$.

3.3. Consistency of the $R^{(\alpha)}$-Bayes estimators

Let us now examine the asymptotic properties of the $R^{(\alpha)}$-Bayes estimators associated with the $R^{(\alpha)}$-posterior distribution (2.2) under the general setup of Section 2. In the decision-theoretic framework, we estimate a functional $\phi_P := \phi(P)$ of the true probability P; for example, ϕ_P could be the probability density of P, or any summary measure (e.g., mean) of P. For the given parametric family $F^n(\cdot|\boldsymbol{\theta})$, denote $\phi_{\boldsymbol{\theta}} := \phi_{F^n(\cdot|\boldsymbol{\theta})}$. Then, our action space is $\Phi = \{\phi_Q : Q \text{ is a probability measure on } (\Omega, \mathcal{B}_\Omega)\}$. Consider a nonnegative loss function $L_n(\phi, \widehat{\phi})$ on $\Phi \times \Phi$ denoting the loss in estimating ϕ by $\widehat{\phi}$; let $L_n(\phi_{\boldsymbol{\theta}}, \phi)$ be $\mathcal{B}_{\Theta_n}$ measurable for each $\phi \in \Phi$. Then, the general $R^{(\alpha)}$-Bayes estimator $\widehat{\phi} = \widehat{\phi}(\cdot; \underline{\boldsymbol{x}}_n)$ of ϕ is defined as

$$\widehat{\phi} = \operatorname*{argmin}_{\phi \in \Phi} \int L_n(\phi_{\boldsymbol{\theta}}, \phi) \pi_n^{(\alpha)} \left(d\boldsymbol{\theta} | \underline{\boldsymbol{x}}_n\right), \tag{3.2}$$

provided the minimum is attained. In particular, the $R^{(\alpha)}$-Bayes estimator of $\phi_{\boldsymbol{\theta}} = \boldsymbol{\theta}$ is the mean of the $R^{(\alpha)}$-posterior distribution for a squared error loss, provided it exists finitely, or a median of the $R^{(\alpha)}$-posterior distribution for an absolute error loss.

However, if the minimum in (3.2) is not attained, we may define the approximate $R^{(\alpha)}$-Bayes estimator $\widehat{\phi}$ of ϕ using the relation

$$\int L_n(\phi_{\boldsymbol{\theta}}, \widehat{\phi}) \pi_n^{(\alpha)} \left(d\boldsymbol{\theta} | \underline{\boldsymbol{x}}_n\right) \leq \inf_{\phi \in \Phi} \int L_n(\phi_{\boldsymbol{\theta}}, \phi) \pi_n^{(\alpha)} \left(d\boldsymbol{\theta} | \underline{\boldsymbol{x}}_n\right) + \delta_n,$$

with $\lim_{n \to \infty} \delta_n = 0$. A useful example is the approximate mode of the $R^{(\alpha)}$-posterior for a discrete parameter space, which is an approximate $R^{(\alpha)}$-Bayes estimator under a 0–1 loss. In addition, note that if the $R^{(\alpha)}$-Bayes estimator exists, it is also an approximate $R^{(\alpha)}$-Bayes estimator.

Definition 3. A loss function L_n on $\Phi \times \Phi$ is said to be bounded if there exists $\bar{L} < \infty$ such that $L_n(\phi_{\boldsymbol{\theta}}, \phi_P) \leq \bar{L}$, for all n and all $\boldsymbol{\theta} \in \Theta_n$.

Definition 4. A loss L_n on $\Phi \times \Phi$ is said to be equivalent to a pseudo-metric d_n on $\Phi \times \Phi$ if there exist two strictly increasing functions h_1 and h_2 on $[0, \infty)$ that are continuous at 0, with $h_1(0) = h_2(0) = 0$, and satisfy $L_n \leq h_1(d_n)$ and

$d_n \leq h_2(L_n)$ on $\Phi \times \Phi$ and for all n.

Note that Definition 4 indicates $\lim_{n\to\infty} L_n(\phi_n, \widehat{\phi}_n) = 0$ if and only if $\lim_{n\to\infty} d_n(\phi_n, \widehat{\phi}_n) = 0$. As an example, the squared Hellinger loss is bounded and equivalent to the L_1-distance. In addition, the absolute error (L_1) loss is equivalent to itself and bounded by twice the Hellinger loss.

We now establish the asymptotic consistency of $R^{(\alpha)}$-Bayes and approximate $R^{(\alpha)}$-Bayes estimators of $\phi_{\boldsymbol{\theta}}$ to the true value ϕ_P for such a loss. The proof mimics that of Lemma 12 in Barron (1988).

Theorem 4. *(Consistency of $R^{(\alpha)}$-Bayes Estimators). Given any sample data $\underline{\boldsymbol{x}}_n$, let $\widehat{\phi}_n = \widehat{\phi}(\cdot; \underline{\boldsymbol{x}}_n)$ be an approximate $R^{(\alpha)}$-Bayes estimator (or the $R^{(\alpha)}$-Bayes estimator) of ϕ_P with respect to a loss function L_n that is bounded and equivalent to a pseudo-metric d_n. In addition, for any $\epsilon > 0$, define $A_{\epsilon,n} = \{\boldsymbol{\theta} : d_n(\phi_P, \phi_{\boldsymbol{\theta}}) \leq \epsilon\}$. Then, we have $d_n(\phi_P, \widehat{\phi}_n) \leq \epsilon + h_2((\epsilon + \bar{L}\pi_n^{(\alpha)}(A^c_{h_1^{-1}(\epsilon),n}|\underline{\boldsymbol{x}}_n))$ $/(1 - \pi_n^{(\alpha)}\left(A^c_{\epsilon,n}|\underline{\boldsymbol{x}}_n\right)))$. Consequently, if $\lim_{n\to\infty} \pi_n^{(\alpha)}\left(A^c_{\epsilon,n}|\underline{\boldsymbol{X}}_n\right) = 0$ in probability or with probability one for all $\epsilon > 0$, then $\lim_{n\to\infty} d_n(\phi_P, \widehat{\phi}_n) = 0$ in probability or with probability one, respectively.*

In simple language, Theorem 4 states that whenever the target ϕ_P is close enough to the model value $\phi_{\boldsymbol{\theta}}$ in the pseudo-metric d_n asymptotically under the $R^{(\alpha)}$-posterior probability, the corresponding $R^{(\alpha)}$-Bayes estimator with respect to L_n is asymptotically consistent for ϕ_P in d_n. However, Theorem 3 yields $\lim_{n\to\infty} \pi_n^{(\alpha)}\left(A^c_{\epsilon,n}|\underline{\boldsymbol{X}}_n\right) = 0$ under appropriate conditions and, hence, the corresponding $R^{(\alpha)}$-Bayes estimators are consistent in suitable d_n. In particular, Theorem 4 applies to the $R^{(\alpha)}$-Bayes estimators with respect to the squared Hellinger loss and the L_1-loss in terms of deducing their L_1 consistency.

4. Application (I): Independent Stationary Models

4.1. $R^{(\alpha)}$-posterior convergence

Consider the setup of the independent stationary model in Example 1. Let us study the conditions required for the exponential convergence of the $R^{(\alpha)}$-posterior for this particular setup. First, to verify the merging of G^n and $M_n^{(\alpha)}$, we define the individual α-modified density as $\widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta}) = \exp(q_{\boldsymbol{\theta}}^{(\alpha)}(\cdot))/Q^{(\alpha)}(\chi|\boldsymbol{\theta})$ and the α-modified prior $\widetilde{\pi}_n^{(\alpha)}$ as in Remark 1, with $\pi_n = \pi$. Then, we consider the information denseness of the prior π under independent stationary models with respect to $\mathcal{F}_\alpha = \left\{\widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\right\}$, defined as follows.

Definition 5. The prior π under the i.i.d. model is information dense at G with

respect to $\mathcal{F}_\alpha$ if there exists a finite measure $\widetilde{\pi}$ satisfying (3.1) and $\widetilde{\pi}(\{\boldsymbol{\theta} : KLD(g, \widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta})) < \epsilon\} > 0$, for all $\epsilon > 0$.

Note that the above definition is equivalent to the general notion of information denseness given in Definition 2. Thus, in view of Theorem 2, it implies the merging of G^n and $M_n^{(\alpha)}$ in probability for independent stationary models. Thus, Theorem 3 may be restated as follows.

Proposition 1. *Consider the setup of independent stationary models, and assume that the prior π is independent of n and is information dense at g with respect to $\mathcal{F}_\alpha$, as per Definition 5. Take any sequence of measurable parameter sets $A_n \subset \Theta$. Then, $\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n)$ is exponentially small with P-probability tending to one if and only if there exist constants $r_1, r_2 > 0$ and sets $B_n, C_n \in \mathcal{B}_\Theta$ such that* (A1)–(A3) *are satisfied with $b_n = e^{-nr_1}$ and $c_n = e^{-nr_2}$.*

Next, note that for the present case, (A3) holds under the assumption of the existence of a UEC test for G against the family $\{Q^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q^{(\alpha)}(\chi|\boldsymbol{\theta}) : \boldsymbol{\theta} \in C_n\}$. We can further simplify it by using a necessary and sufficient condition for the existence of a UEC from Barron (1989), which states that, "*for every $\epsilon > 0$ there exists a sequence of UEC tests for the hypothesized distribution P versus the family of distributions $\{Q : d_{T_n}(P, Q) > \epsilon/2\}$ if and only if the sequence of partitions T_n has effective cardinality (eff. card.) of order n with respect to P*". Here, for any measurable partition T, d_T denotes the T-variation norm $d_T(P, Q) = \sum_{A \in T} |P(A) - Q(A)|$. Using this, we show that the $R^{(\alpha)}$-posterior asymptotically concentrates on the L_1 model neighborhood of the true density g. Define, for any density p and any partition T, the "theoretical histogram" density p^T as $p^T(x) = (1/\lambda(A)) \int_A p(y)\lambda(dy)$, for $x \in A \in T$, whenever $\lambda(A) \neq 0$, and $p^T = 0$ otherwise. We call a sequence of partitions T_n "*rich*" if the corresponding sequence of densities g^{T_n} converges to g in the L_1-distance. Furthermore, define $B_\epsilon^{T_n} = \left\{\boldsymbol{\theta} : d_1\left(f_{\boldsymbol{\theta}}, \widetilde{q}^{(\alpha)T_n}(\cdot|\boldsymbol{\theta})\right) > \epsilon\right\}$, for any $\epsilon > 0$ and sequence of partition T_n, where d_1 denotes the L_1 distance.

Assumption (B): *For $\epsilon > 0$, $\widetilde{\pi}_n^{(\alpha)}(B_\epsilon^{T_n}) = M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_\epsilon^{T_n})/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta)$ is exponentially small for a rich sequence of partitions T_n with eff. card. of order n.*

Note that Assumption (B) implies Assumption (A2) for $B_\epsilon^{T_n}$, or any smaller subset of it. Thus, applying it with $B_n = \{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) \geq \epsilon, d_{T_n}(G, Q^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q^{(\alpha)}(\chi|\boldsymbol{\theta})) < \epsilon/2\} \subset B_{\epsilon/4}^{T_n}$ and the existence result of UEC tests with $C_n = \left\{\boldsymbol{\theta} : d_{T_n}\left(G, Q^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q^{(\alpha)}(\chi|\boldsymbol{\theta})\right) > \epsilon/2\right\}$, Proposition 1 yields the asymptotic exponential concentration of the $R^{(\alpha)}$-posterior probability in the L_1-neighborhood

$A_n = \{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) < \epsilon\}$. Clearly, $A_n \cup B_n \cup C_n = \Theta_n$ for these choices.

Theorem 5. *Consider the setup of i.i.d. models, and assume that the prior π is independent of n and information dense at g with respect to $\mathcal{F}_\alpha$, as per Definition 5. If Assumption (B) holds, then, for every $\epsilon > 0$, $\pi_n^{(\alpha)}(\{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) \geq \epsilon\} | \underline{\boldsymbol{X}}_n)$ is exponentially small with P-probability one.*

Note that the final Assumption (B) is easy to verify for models and priors belonging to the standard exponential family of distributions with exponentially decaying tails. However, if Assumption (B) does not hold, we can deduce a weaker conclusion by using the T_n-variance distance in place of the L_1 distance. This idea was proposed by Barron (1988) for a similar result in the case of the ordinary posterior. An extended version for the $R^{(\alpha)}$-posterior is given in the following theorem.

Theorem 6. *Consider the setup of i.i.d. models, and assume that the prior π is independent of n and information dense at g with respect to $\mathcal{F}_\alpha$, as per Definition 5. Then, for any sequence of partitions T_n with eff. card. of order n, $\pi_n^{(\alpha)}(\{\boldsymbol{\theta} : d_{T_n}(G, Q_n^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta})) \geq \epsilon\}|\underline{\boldsymbol{X}}_n)$ is exponentially small with P-probability one.*

4.2. The cases of discrete priors: maximum $R^{(\alpha)}$-posterior estimator

We can derive the exponential consistency of the $R^{(\alpha)}$-Bayes estimators with respect to the bounded loss functions from Theorem 4, along with Proposition 1 to Theorem 6. Let us now consider, in more detail, the particular case of discrete priors and the maximum $R^{(\alpha)}$-posterior estimator.

Consider the setup of i.i.d. models, but now with a countable Θ. On this countable parameter space, we consider a sequence of discrete priors $\pi_n(\boldsymbol{\theta})$ that are sub-probability mass functions; that is, $\sum_{\boldsymbol{\theta}} \pi_n(\boldsymbol{\theta}) \leq 1$. The most common loss-function to consider under this setup is the 0–1 loss function, for which the resulting $R^{(\alpha)}$-Bayes estimator is the (global) mode of the $R^{(\alpha)}$-posterior density; we call this estimator of $\boldsymbol{\theta}$ the "maximum $R^{(\alpha)}$-posterior estimator (MRPE)." When this mode is not attained, we consider an approximate version $\widehat{\boldsymbol{\theta}}_\alpha$, referred to as an "approximate maximum $R^{(\alpha)}$-posterior estimator (AMRPE)," defined by the relation

$$\widetilde{\pi}_n^{(\alpha)}(\widehat{\boldsymbol{\theta}}_\alpha)\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\widehat{\boldsymbol{\theta}}_\alpha) > \sup_{\boldsymbol{\theta}} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})e^{-n\delta_n}, \tag{4.1}$$

with $\lim_{n\to\infty} \delta_n = 0$, where $\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta})$ and $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})$ are the α-modified model and prior densities (see Remark 1), respectively. This definition follows from the

fact that the $R^{(\alpha)}$-posterior density is proportional to $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$. Note that if the MRPE exists, then it is also an AMRPE. Assume that this estimator $\widehat{\boldsymbol{\theta}}_\alpha = \widehat{\boldsymbol{\theta}}_\alpha(\underline{\boldsymbol{x}}_n)$, as a function of the data $\underline{\boldsymbol{x}}_n$, is measurable, and consider a prior sequence that satisfies

$$\liminf_{n\to\infty} e^{nr}\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) \geq 1, \quad \text{for all } r > 0,\ \boldsymbol{\theta} \in \Theta. \tag{4.2}$$

Assumption (4.2) signifies that the (α-modified) prior probabilities are not exponentially small anywhere in Θ. Then, we have the following theorems.

Theorem 7. *Consider the setup of i.i.d. models with fixed countable $\Theta_n = \Theta$ and discrete prior sequence π_n satisfying Assumption* (4.2)*. Suppose π_n is information dense at the true probability mass function g with respect to $\mathcal{F}_\alpha$, as in Definition* 5*, and $\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n)$ is exponentially small with probability one for a sequence of measurable subsets $A_n \subseteq \Theta$. Then, any AMRPE $\widehat{\boldsymbol{\theta}}_\alpha \in A_n$, for all sufficiently large n, with probability one.*

Theorem 8. *Consider the setup of stationary independent models with fixed countable $\Theta_n = \Theta$ and a discrete prior sequence π_n satisfying Assumption* (4.2)*. Then, for any true density g that is an information limit of the (countable) family $\{\widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\}$ and for any $\epsilon > 0$, we have that $\pi_n^{(\alpha)}(\{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) \geq \epsilon\}|\underline{\boldsymbol{X}}_n)$ is exponentially small with probability one. Therefore, $\lim_{n\to\infty} d_1(g, f_{\widehat{\boldsymbol{\theta}}_\alpha}) = 0$, with probability* 1 *for any AMRPE $\widehat{\boldsymbol{\theta}}_\alpha$.*

Remark 2. Theorem 8, in the special case $\alpha = 0$, yields a stronger version of Theorem 15 of Barron (1988). Our result requires fewer assumptions than required by Barron's result.

5. Application (II): Independent Nonhomogeneous Models

5.1. Convergences of $R^{(\alpha)}$-posterior and $R^{(\alpha)}$-Bayes estimators

Let us now consider the setup of independent but nonhomogeneous (INH) models, as described in Example 2 of Section 2, and simplify the exponential convergence results for the $R^{(\alpha)}$-posterior probabilities under this INH setup. Note that in this case, $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^n q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i)$ for any observed data $\underline{\boldsymbol{x}}_n = (x_1, \ldots, x_n)$, and hence $Q_n^{(\alpha)}(S_n|\boldsymbol{\theta}) = \prod_{i=1}^n Q^{(i,\alpha)}(S^i|\boldsymbol{\theta})$, for any $S_n = S^1 \times S^2 \times \cdots \times S^n \in \mathcal{B}_n$, with $S^i \in \mathcal{B}^i$, for all i and $Q^{(i,\alpha)}(S^i|\boldsymbol{\theta}) = \int_{S^i} \exp(q_{i,\boldsymbol{\theta}}^{(\alpha)}(y))dy$. Assume that $\Theta_n = \Theta$ and $\pi_n = \pi$ are independent of n. Then, we have $\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n \exp(q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i))/Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta}) = \prod_{i=1}^n \widetilde{q}^{(i,\alpha)}(x_i|\boldsymbol{\theta})$, with $\widetilde{q}^{(i,\alpha)}(x_i|\boldsymbol{\theta}) = \exp(q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i))/Q^{(i,\alpha)}(\chi^i|\boldsymbol{\theta})$. Thus, in the notation of Section 3.1, we have $D_n^{(\alpha)}(\boldsymbol{\theta}) =$

$(1/n)\sum_{i=1}^{n} KLD\left(g_i, \widetilde{q}^{(i,\alpha)}(\cdot|\boldsymbol{\theta})\right)$, and hence the definition of information denseness can be simplified for the INH models as follows.

Definition 6. The prior π under the INH model is said to be information dense at $\boldsymbol{G}_n = (G_1, \ldots, G_n)$ with respect to $\mathcal{F}_{n,\alpha} = \otimes_{i=1}^{n} \mathcal{F}_{\alpha}^{i}$ if there exists a finite measure $\widetilde{\pi}$ satisfying (3.1), such that $\widetilde{\pi}(\{\boldsymbol{\theta} : \limsup_{n\to\infty}(1/n)\sum_{i=1}^{n} KLD\left(g_i, \widetilde{q}^{(i,\alpha)}(\cdot|\boldsymbol{\theta})\right) < \epsilon\}) > 0$, for all $\epsilon > 0$.

When $f_{i,\boldsymbol{\theta}} = f_{\boldsymbol{\theta}}$ is independent of i, then the INH setup coincides with the i.i.d. setup and the information denseness in Definition 6 coincides with that in Definition 5. Furthermore, Definition 6 is equivalent to the general Definition 2, and hence implies that G^n and $M_n^{(\alpha)}$ merge in probability. Then, we have the following simplified results for the INH setup.

Proposition 2. *Consider the setup of INH models with $\Theta_n = \Theta$, and assume that the prior π is independent of n and information dense at $\boldsymbol{G}_n$ with respect to $\mathcal{F}_{n,\alpha}$, as per Definition* 6. *Then, for any sequence of measurable parameter sets $A_n \subset \Theta$, $\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n)$ is exponentially small with P-probability one if and only if there exist sequences of measurable parameter sets $B_n, C_n \subset \Theta$ such that $A_n \cup B_n \cup C_n = \Theta$, $M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n) \leq e^{-nr}$ for $r > 0$ and a UEC test for G^n against $\{Q_n^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) : \boldsymbol{\theta} \in C_n\}$ exists.*

However, the existence of the required UEC in Proposition 2 is equivalent to the existence of a UEC test for G_i against $\{Q^{(i,\alpha)}(\cdot|\boldsymbol{\theta})/Q^{(i,\alpha)}(\chi^i|\boldsymbol{\theta}) : \boldsymbol{\theta} \in C_n\}$ uniformly over $i = 1, \ldots, n$. Following Section 4.1, this holds if Assumption (B) is satisfied for $\widetilde{B}_{\epsilon}^{T_n} = \left\{\boldsymbol{\theta} : (1/n)\sum_{i=1}^{n} d_1(f_{i,\boldsymbol{\theta}}, \widetilde{q}^{(i,\alpha)}(\cdot|\boldsymbol{\theta})^{T_n}) > \epsilon\right\}$ in place of $B_{\epsilon}^{T_n}$. This leads to the following simplification.

Theorem 9. *Consider the INH models with $\Theta_n = \Theta$, and assume that the prior π is independent of n and information dense at G_n with respect to $\mathcal{F}_{n,\alpha}$, as per Definition* 6. *If Assumption* (B) *holds for $\widetilde{B}_{\epsilon}^{T_n}$ in place of $B_{\epsilon}^{T_n}$, for every $\epsilon > 0$, the $R^{(\alpha)}$-posterior probability $\pi_n^{(\alpha)}(\{\boldsymbol{\theta} : (1/n)\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\theta}}) \geq \epsilon\}|\underline{\boldsymbol{x}}_n)$ is exponentially small with P-probability one, for $\epsilon > 0$.*

Note that the Bernstein–von Mises-type asymptotic results for the $R^{(\alpha)}$-posterior distribution under the INH setup are extremely important to providing contraction rates for our new robust pseudo-posterior; similar results for i.i.d. models are discussed in Ghosh and Basu (2016a). However, considering the length of the present paper and its focus on the exponential convergence results, we propose presenting the results on contraction rates for INH models in a subsequent paper; for the time being, they are available in the ArXiv version (Majumder, Basu and Ghosh (2019)).

5.2. Robust Bayes estimation under fixed design regression models

As noted in Example 2, the most common example of the general INH setup is that of fixed-design regression models. We consider the important example of model (2.4) with n fixed k-variate design points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ and $f_{i,\boldsymbol{\theta}}(x) = (1/\sigma) f\left(x - \psi(\boldsymbol{t}_i, \boldsymbol{\beta})/\sigma\right)$, for some univariate density f. The corresponding α-likelihood is given by $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|(\boldsymbol{\beta}, \sigma)) = \sum_{i=1}^n q_{i,(\boldsymbol{\beta},\sigma)}^{(\alpha)}(x_i)$, with $q_{i,(\boldsymbol{\beta},\sigma)}^{(\alpha)}(x_i) = 1/(\alpha\sigma^\alpha) f\left((x_i - \psi(\boldsymbol{t}_i, \boldsymbol{\beta}))/\sigma\right)^\alpha - M_{f,\alpha}/((1+\alpha)\sigma^\alpha) - 1/\alpha$, where $M_{f,\alpha} = \int f^{1+\alpha}$. Consider a prior density $\pi(\boldsymbol{\beta}, \sigma)$ for the parameters $(\boldsymbol{\beta}, \sigma)$ over the space $\Theta = \mathbb{R}^k \times (0, \infty)$ $[p = k+1]$. This prior can be chosen to be the conjugate prior or any subjective or objective prior; a common objective prior is Jeffrey's prior, given by $\pi(\boldsymbol{\beta}, \sigma) = \sigma^{-1}$. Then, the $R^{(\alpha)}$-posterior density of $(\boldsymbol{\beta}, \sigma)$ is given by (2.2), which now simplifies to

$$\pi_n^{(\alpha)}((\boldsymbol{\beta}, \sigma)|\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i - \psi(\boldsymbol{t}_i, \boldsymbol{\beta})}{\sigma}\right)^\alpha - \frac{M_{f,\alpha}}{(1+\alpha)\sigma^\alpha}\right] \pi(\boldsymbol{\beta}, \sigma) \Bigg/ \int\int \prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i - \psi(\boldsymbol{t}_i, \boldsymbol{\beta})}{\sigma}\right)^\alpha - \frac{M_{f,\alpha}}{(1+\alpha)\sigma^\alpha}\right] \pi(\boldsymbol{\beta}, \sigma) d\boldsymbol{\beta} d\sigma. \tag{5.1}$$

If σ is known, as in the Poisson or logistic regression models (or can be assumed to be known with properly scaled variables), we consider a prior only on $\boldsymbol{\beta}$ given by, say, $\pi(\boldsymbol{\beta})$, which is either the objective uniform prior or the conjugate prior, or some other proper prior. In such cases, we obtain the following simplified form for the $R^{(\alpha)}$-posterior density of $\boldsymbol{\beta}$:

$$\pi_n^{(\alpha)}(\boldsymbol{\beta}|\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i - \psi(\boldsymbol{t}_i, \boldsymbol{\beta})}{\sigma}\right)^\alpha\right] \pi(\boldsymbol{\beta}) \Bigg/ \int \prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i - \psi(\boldsymbol{t}_i, \boldsymbol{\beta})}{\sigma}\right)^\alpha\right] \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}. \tag{5.2}$$

One can obtain the $R^{(\alpha)}$-Bayes estimators of $\boldsymbol{\beta}$, σ under any suitable loss. We now study the exponential convergence for some regression examples, providing simplifications for the required assumptions.

5.3. Example: normal linear regression model with known variance

We consider the normal regression model, a particular member of the class of regression models considered in Section 5.2, where $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = \boldsymbol{t}_i^T \boldsymbol{\beta}$, with f being

a standard normal density. For simplicity, we assume that the error variance σ is known; the case of unknown σ is considered later. In this case, we can simplify the $R^{(\alpha)}$-posterior from (5.2), and compute the expected $R^{(\alpha)}$-posterior estimator (ERPE) of $\boldsymbol{\beta}$; however, the resulting $R^{(\alpha)}$-posterior has no explicit form and, hence, the corresponding ERPE needs to be computed numerically (see Sections 6, 7).

Note that, being a particular case of the INH setup, the exponential consistency of the $R^{(\alpha)}$-posterior of $\boldsymbol{\beta}$ holds directly under the assumptions of Proposition 2. We now verify the required conditions for normal linear regression models with known σ. For this purpose, let us denote $\boldsymbol{D} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n]^T$, the fixed-design matrix, and $\boldsymbol{x} = (x_1, \ldots, x_n)^T$. Recall that, provided $\boldsymbol{D}$ has full column rank, the ordinary least squares estimate of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{D}^T\boldsymbol{x}$, which is also the ordinary Bayes estimator under the uniform prior and has variance $n^{-1}(\boldsymbol{D}^T\boldsymbol{D})^{-1}$. We assume the following intuitive assumptions on the fixed-design matrix $\boldsymbol{D}$ of the linear regression models:

(R1) The design points $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{ik})^T$, for $i = 1, \ldots, n$, are such that, for all $j, l, s = 1, \ldots, k$, we have

$$\sup_{n>1} \max_{1\le i\le n} |t_{ij}| = O(1), \quad \max_{1\le i\le n} |t_{ij}||t_{il}| = O(1), \quad \frac{1}{n}\sum_{i=1}^{n} |t_{ij}t_{il}t_{is}| = O(1). \tag{5.3}$$

(R2) The matrix $\boldsymbol{D}$ satisfies $\inf_n[\text{min eigenvalue of } n^{-1}(\boldsymbol{D}^T\boldsymbol{D})] > 0$, which also implies the matrix $\boldsymbol{D}$ has full column rank, and $\max_{1\le i\le n}[\boldsymbol{t}_i^T(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{t}_i] = O(n^{-1})$.

Note that Assumptions (R1)–(R2) imply the (weak) consistency of the corresponding (frequentist) MDPDE of $\boldsymbol{\beta}$ obtained by minimizing the negative of the associated α-likelihood function (Ghosh and Basu (2013)). They are easy to verify for any given design matrix; in particular, they hold if $\boldsymbol{t}_i$ are generated from some nonsingular k-variate distributions. It is shown in Majumder, Basu and Ghosh (2019) that these two conditions indeed ensure a Bernstein–von Mises-type result for the associated $R^{(\alpha)}$-posterior.

It is fascinating to see that, despite the complexity of our earlier assumptions for general INH models, these two simple Assumptions (R1)–(R2) imply the exponential consistency of the $R^{(\alpha)}$-posterior probability at any $\alpha \ge 0$ for a linear regression (along with some mild conditions on the prior). The result is presented in the following theorem.

Theorem 10. *Consider the normal linear regression setup with known error*

variance. Assume that the true parameter value is $\boldsymbol{\beta}_0$*, that is,* $g_i = f_{i,\boldsymbol{\beta}_0}$ *for all* i*, and the prior on* $\boldsymbol{\beta}$ *is continuous and positive at* $\boldsymbol{\beta}_0$*. Take any* $\alpha \geq 0$*. Then, under Assumptions* (R1)–(R2)*, given any* $\epsilon > 0$*, there exists* $r > 0$ *such that*

$$\lim_{n\to\infty} P\left[\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\right\} \middle| \underline{\boldsymbol{x}}_n\right) < e^{-nr}\right] = 1,$$

or equivalently,

$$\lim_{n\to\infty} P\left[\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{t}_i^T|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \geq \epsilon\right\} \middle| \underline{\boldsymbol{x}}_n\right) < e^{-nr}\right] = 1;$$

that is, the $R^{(\alpha)}$*-posterior probabilities asymptotically concentrate on the neighborhoods of the true regression line at an exponential rate of convergence.*

5.4. Example: normal linear regression model with unknown variance

We now consider an extended version of the previous example of a normal linear regression with an unknown error variance. Consider the setup and notation of the previous subsection, with $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = \boldsymbol{t}_i^T\boldsymbol{\beta}$ and f being a normal density with mean zero and variance σ. However, we now consider σ^2 to also be an unknown parameter, along with the regression coefficient $\boldsymbol{\beta}$. Given a prior $\pi(\boldsymbol{\beta}, \sigma)$, in this case, the $R^{(\alpha)}$-posterior distribution is given by (5.1), with $M_f = (2\pi)^{-\alpha/2}(1+\alpha)^{-1/2}$.

Furthermore, we have simplified the required conditions for the exponential convergence of the $R^{(\alpha)}$-posterior probabilities. The result is presented in the following theorem; interestingly, the same sets of conditions as in the case of known σ suffice.

Theorem 11. *Consider the normal linear regression setup with an unknown error variance. Assume that the true parameter value is* $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \sigma_0^2)$*; that is,* $g_i = f_{i,\boldsymbol{\theta}_0}$*, for all* i*, and the prior on* $\boldsymbol{\theta}$ *is continuous and positive at* $\boldsymbol{\theta}_0$*. Take any* $\alpha \geq 0$*. Then, under Assumptions* (R1)–(R2)*, given any* $\epsilon > 0$*, there exists* $r > 0$ *such that*

$$\lim_{n\to\infty} P\left[\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\theta}}) \geq \epsilon\right\} \middle| \underline{\boldsymbol{x}}_n\right) < e^{-nr}\right] = 1.$$

5.5. Example: logistic regression model

We now consider the important logistic regression model, which does not belong to the class of location-scale-type regressions in Section 5.2. In the nota-

tion of Example 2, given the fixed-design points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$, the logistic regression model considers binary response variables x_i that follow a Bernoulli distribution with expectation $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = e^{\boldsymbol{t}_i^T \beta}/(1 + e^{\boldsymbol{t}_i^T \beta})$, for $i = 1, \ldots, n$. As in Example 2, this model clearly belongs to the INH setup with the only parameter being the regression coefficient $\boldsymbol{\theta} = \boldsymbol{\beta}$; there is no scale parameter here. Thus, the α-likelihood $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta})$ of $\boldsymbol{\beta}$ is given by (2.5), with $f_{i,\boldsymbol{\theta}}$ being the probability mass function of the Bernoulli$(\psi(\boldsymbol{t}_i, \boldsymbol{\beta}))$ distribution, and the integral being the sum over its support $\chi^i = \{0, 1\}$; the underlying measure is the counting measure. The $R^{(\alpha)}$ is obtained by using (2.2), given any prior $\pi(\boldsymbol{\beta})$, which does not have a closed form and needs to be computed numerically; see Section 6.

Let us now simplify the conditions required for the exponential consistency of the $R^{(\alpha)}$-posterior for the logistic regression model. For this purpose, we recall Assumption (R1) on the fixed-design points, and consider the new condition (R3) in terms of the matrix $\boldsymbol{\Psi}_n(\boldsymbol{\beta}) = n^{-1} E_{g_i}[\partial^2 q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T]$.

(R3) $\inf_n[\text{min eigenvalue of } \boldsymbol{\Psi}_n(\boldsymbol{\beta})] > 0$, for all $\boldsymbol{\beta}$.

The matrix $\boldsymbol{\Psi}_n(\boldsymbol{\beta})$ appears in the asymptotic variance of the (frequentist) MDPDE of $\boldsymbol{\beta}$ under the fixed-design logistic regression model (Ghosh and Basu (2016b)), as well as in the Bernstein–von Mises-type results for the corresponding $R^{(\alpha)}$-posterior distribution (Majumder, Basu and Ghosh (2019)). Thus, in view of those results, Assumption (R3) is extremely intuitive and easy to verify for any given design matrix. We have shown that Assumptions (R1) and (R3) also imply the exponential convergence of our generalized $R^{(\alpha)}$-posterior probability in this logistic regression setup, as presented in the following theorem.

Theorem 12. *Consider the fixed-design regression setup given above. Assume that the true parameter value is $\boldsymbol{\beta}_0$; that is, $g_i = f_{i,\boldsymbol{\beta}_0}$, for all i, and the prior $\pi(\boldsymbol{\beta})$ is continuous and positive at $\boldsymbol{\beta}_0$. Take any $\alpha \geq 0$. Then, under Assumptions* (R1) *and* (R3)*, given any $\epsilon > 0$, there exists $r > 0$ such that*

$$\lim_{n\to\infty} P\left[\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\right\} \middle| \underline{\boldsymbol{x}}_n\right) < e^{-nr}\right] = 1.$$

6. Numerical Illustrations: Simulations

6.1. Performance of ERPE in normal linear regression model

Let us now reconsider the regression model described in Sections 5.3–5.4, and examine the finite-sample performance of the expected $R^{(\alpha)}$-posterior estimator (ERPE) of the parameters.

We first assume that the error variance σ is known and is equal to one. The corresponding $R^{(\alpha)}$-posterior is given by (5.2), as discussed in Section 5.3, and has no closed-form solution. Thus, we compute the ERPE using an importance sampling Monte-Carlo. We first simulate n observations $t_{11}, \ldots, t_{1n}$ independently from $N(5,1)$ to fix the predictor values $\boldsymbol{t}_i = (1, t_{1i})^T$. Then, n independent error values $\epsilon_1, \ldots, \epsilon_n$ are generated from $N(0,1)$ (note $\sigma = 1$), and the responses are obtained from the linear regression structure $x_i = \boldsymbol{t}_i^T \boldsymbol{\beta} + \epsilon_i$, for $i = 1, \ldots, n$, with the true value of $\boldsymbol{\beta}$ being $\boldsymbol{\beta}_0 = (5,2)^T$. We considered different sample sizes $n = 20, 50, 100$, and different contamination proportions ϵ_C =0% (pure data), 5%, 10%, 20% to examine the finite-sample robustness properties of our proposal. For contaminated samples, $[n\epsilon_C]$ error values are contaminated by generating them from $N(5,1)$ instead of $N(0,1)$. In each case, given a prior, the ERPE at different $\alpha \geq 0$ are computed using 20,000 steps in the importance sampling Monte Carlo, with the proposal density $N_k(\widehat{\boldsymbol{\beta}}, n^{-1}(\boldsymbol{D}^T\boldsymbol{D})^{-1})$. We replicate the above procedure 1,000 times to compute the empirical bias and MSE of the ERPE for two priors, namely the non-informative uniform prior and the conjugate normal prior, which are presented in the Supplementary Material (Figures 1 and 2) to conserve space. The figures show that, under pure data, the bias and the MSE are the least for the usual Bayes estimator of $\boldsymbol{\beta}$ at $\alpha = 0$, but their inflations are not significant for the ERPEs with moderate $\alpha > 0$. Under contamination, the usual Bayes estimator (at $\alpha = 0$) has severely inflated bias and MSE, and becomes highly unstable. Our ERPEs with $\alpha > 0$ are much more stable under contamination in terms of both bias and MSE; the maximum stability is observed for tuning parameters $\alpha \in [0.4, 0.6]$, yielding significantly improved robust Bayes estimators.

Next, we consider the case of an unknown error variance σ in the above linear regression model, as discussed in Section 5.4. We repeat the above simulation exercise for the case of unknown σ as well by taking the true value of $\sigma_0 = 1$ and the conjugate prior on $(\boldsymbol{\beta}, \sigma)$ given by $\pi(\boldsymbol{\beta}, \sigma) = \pi(\boldsymbol{\beta}|\sigma)\pi(\sigma)$. Here $\pi(\beta|\sigma)$ is taken as the $N_2(\boldsymbol{\beta}_0, \sigma^2 I_2)$ density, and $\pi(\sigma)$ is the density of the square root of the inverse chi-square distribution with five degrees of freedom (i.e., prior for σ^2 is Inverse-χ^2_5). However, in this case, the computation of the ERPE could not be done efficiently using the simple importance sampling method, as in the case of known σ; as such we used the Metropolis–Hastings algorithm.

The process is replicated 1,000 times to compute the empirical biases and MSEs of the ERPEs of $\boldsymbol{\beta}$ and σ at different α for the previous simulation setup. The resulting values of the total absolute bias and total MSE over the two components of $\boldsymbol{\beta}$ and the absolute bias and MSE of the ERPE of σ are presented in

Algorithm 1 Computation of ERPE in LRM with unknown variance.

We generate 20,000 sample observations from the $R^{(\alpha)}$ posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$, as follows:

Step 1. Start with $\boldsymbol{\theta}^{(0)} = (0, 0, 2)^T$. Set $k = 1$.

Step 2. After generating $\boldsymbol{\theta}^{(k-1)} = (\boldsymbol{\beta}^{(k-1)}, \sigma^{(k-1)})$ in the $(k-1)$th step, at the kth step, generate $\boldsymbol{\beta}^*$ and σ^* from the proposal densities $g_1 \equiv \mathcal{N}_2(\boldsymbol{\beta}^{(k-1)}, I_2)$ and $g_2 \equiv$Exponential$(\sigma^{(k-1)})$, respectively.

Step 3. Generate $U \sim U(0,1)$ and compute $\gamma = \exp[q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta}^*, \sigma^*)g_1(\boldsymbol{\beta}^*)g_2(\sigma^*)]/\exp[q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta}^{(k-1)}, \sigma^{(k-1)})g_1(\boldsymbol{\beta}^{(k-1)})g_2(\sigma^{(k-1)})]$.

Step 4. If $U < \gamma$, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$ and $\sigma^{(k)} = \sigma^*$. Otherwise, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)}$ and $\sigma^{(k)} = \sigma^{(k-1)}$.

Step 5. Set $k = k + 1$, and go to Step 2.

In each case, the first 5,000 values generated are rejected as burn-in, and the remaining 15,000 parameter values are averaged to obtain a good approximation of the ERPE of $(\boldsymbol{\beta}, \sigma)$.

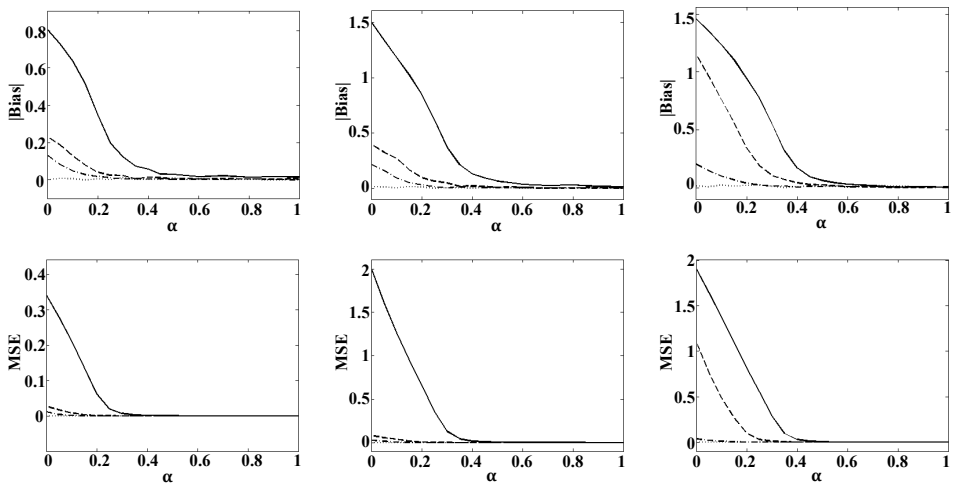
Figures 1 and 2, respectively.

The performance of the ERPE of regression coefficient and the error variance are the same as before in that the proposed ERPE with a larger α provides extremely stable estimates, even under contamination up to 20%. Under pure data, the usual Bayes estimators give the minimum absolute bias and MSEs, but the ERPEs with $\alpha > 0$ are close to these values. However, under data contamination, the usual Bayes estimates (at $\alpha = 0$) become extremely nonrobust yielding a significantly higher bias and MSEs, even though we are using a strong conjugate prior. As the contamination proportion increases, we need larger values of α in the proposed ERPE to produce smaller biases and MSEs close to the pure data scenarios; in particular, $\alpha \geq 0.5$ always has excellent robust performance.

6.2. Performance of ERPE in logistic regression model

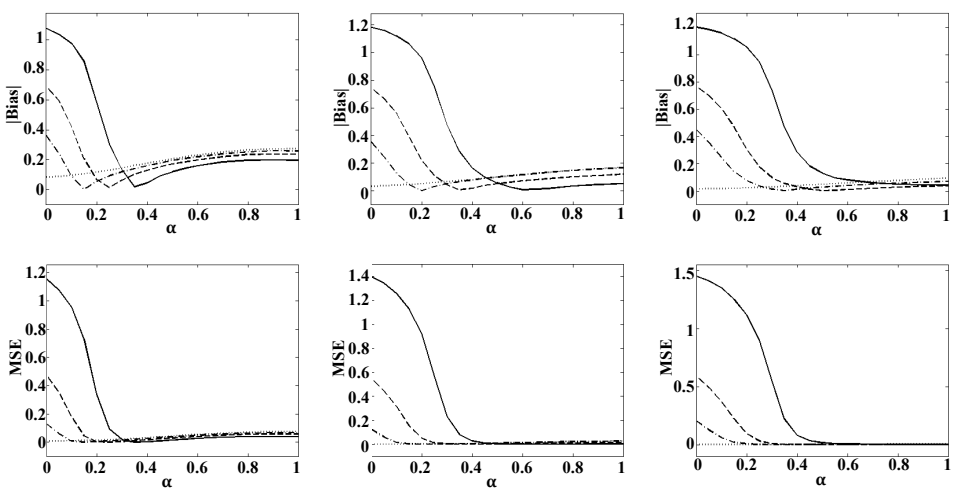
We now consider the fixed-design logistic regression model in Section 5.5, and study the finite-sample properties of the ERPE, the expectation of the regression coefficient $\boldsymbol{\beta}$ under the proposed $R^{(\alpha)}$-posterior distribution. Because the corresponding $R^{(\alpha)}$-posterior has no closed-form solution, we computed the ERPE numerically in our simulation exercise.

We first simulate n values $t_{11}, \ldots, t_{1n}$ independently from $U(-5, 5)$ and fix the design points as $\boldsymbol{t}_i = (1, t_{1i})^T$. Then, the n response values $x_1, \ldots, x_n$ are obtained from the logistic regression structure, with x_i generated from a Bernoulli distri-



(a) $n = 20$ (b) $n = 50$ (c) $n = 100$

Figure 1. Empirical total absolute bias and total MSE of the ERPE of $\boldsymbol{\beta}$ in the linear regression model with unknown σ and the conjugate priors. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$] (see the Supplementary Material for an additional discussion.)



(a) $n = 20$ (b) $n = 50$ (c) $n = 100$

Figure 2. Empirical absolute bias and MSE of the ERPE of σ in the linear regression model with unknown σ and the conjugate priors. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$] (see the Supplementary Material for an additional discussion.)

bution with mean parameter $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = e^{\boldsymbol{t}_i^T \beta}/(1 + e^{\boldsymbol{t}_i^T \beta})$, for each $i = 1, \ldots, n$; the true parameter value is taken as $\boldsymbol{\beta}_0 = (0, 5)^T$. Again, we have considered different sample sizes $n = 20, 50, 100$ and different contamination proportions ϵ_C =0% (pure data), 5%, 10%, 20%. The contaminated observations, $[n\epsilon_C]$ many in a sample of size n, are forced by misspecifying the response values, that is, by changing x_i to $(1 - x_i)$, and the prior is taken as the (bivariate) normal distribution, $\pi(\boldsymbol{\beta}) \equiv N_2(\boldsymbol{\beta}_0, I_2)$. However, in this case, the importance sampling fails to provide a good approximation to the ERPE; thus, we use the Metropolis–Hastings method. Note that the target density, that is, $R^{(\alpha)}$ posterior density, is proportional to $g(\boldsymbol{\beta}) = \exp[q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta})]\pi(\boldsymbol{\beta})d\boldsymbol{\beta}$.

Algorithm 2 Computation of ERPE in Logistic Regression.

We generate 20,000 sample observations from the $R^{(\alpha)}$ posterior distribution of $\boldsymbol{\beta}$, as follows:

Step 1. Start with $\boldsymbol{\beta}^{(0)} = (0, 0)^T$.

Step 2. After generating $\boldsymbol{\beta}^{(k-1)}$ in the $(k-1)$th step, at the kth step, generate $\boldsymbol{\beta}^*$ from $\mathcal{N}_2(\boldsymbol{\beta}^{(k-1)}, I_2)$.

Step 3. Generate $U \sim U(0, 1)$, and compute $\gamma = g(\boldsymbol{\beta}^*)/g(\boldsymbol{\beta}^{(k-1)})$.

Step 4. If $U < \gamma$, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$. Otherwise, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)}$.

Step 5. Set $k = k + 1$, and go to Step 2.

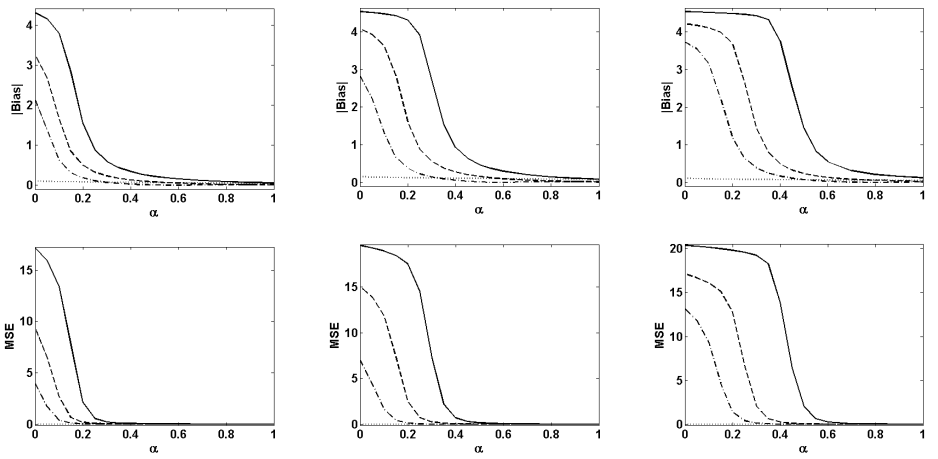
In each case, the first 5,000 values generated are rejected as burn-in, and the remaining 15,000 parameter values are averaged to obtain a good approximation of the ERPE.

The simulation exercise is replicated 1,000 times to compute 1,000 ERPEs of $\boldsymbol{\beta}$. Their empirical biases and MSEs are presented in Figure 3. Clearly, the moderately larger values of α produce highly robust estimates under contaminations, with only a slight loss in efficiency under pure data. Under contamination, the MSEs of the ERPEs remain stable for $\alpha \geq 0.5$; however, we need slightly larger $\alpha \geq 0.7$ to get smaller biases under heavy contamination of 20%.

7. Practical Aspects

7.1. On the computation of the $R^{(\alpha)}$-Bayes estimators

A complex and challenging aspect of the proposed $R^{(\alpha)}$-Bayes estimators is their computation. This is, in fact, a common problem with all pseudo-posteriors that replace the likelihood with some robust loss function. In a frequentist sense, using a suitable optimization algorithm to derive a point estimator from some



(a) $n = 20$ (b) $n = 50$ (c) $n = 100$

Figure 3. Empirical total absolute bias and total MSE of the ERPE of $\boldsymbol{\beta}$ in the logistic regression model with a normal prior. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$] (see the Supplementary Material for an additional discussion.)

robust loss function results in scalable computation for many applications. In contrast, the computation of the whole pseudo-posterior is challenging for complicated models, and needs careful attention (even for the usual Bayes methods).

For our $R^{(\alpha)}$-posterior, no closed-form expressions exist in most applications and, hence, we need to compute the corresponding $R^{(\alpha)}$-Bayes estimators numerically. One such approach could be the importance sampling technique, which is seen to work well in our illustrations for normal means (Ghosh and Basu (2016a)) or linear models with known σ (Section 6.1). However, this simple approach is useful only when it is possible to use some conjugacy structure; here, the standard posterior distribution is used as the proposal distribution, owing to its conjugacy. However, when the model is more complicated and we do not have a good proposal distribution, importance sampling fails to provide good approximations to the proposed $R^{(\alpha)}$-Bayes estimators. This is because the α-likelihood parts do not enjoy some conjugacy when the model is little bit more complicated, for example, in the case of the linear regression with unknown variance or the logistic regression models. In such cases, we propose using a suitable Metropolis-Hastings algorithm, which works very well for the computations of the proposed ERPE under the above-mentioned two cases; the corresponding algorithms are given in Sections 6.1 and 6.2, respectively. We also supply the relevant R code for the computations of the ERPEs for our examples in the Supplementary Material.

We hope that, with advances in modern computers, it will be possible to develop similar algorithms for the computation of the $R^{(\alpha)}$-posterior and the $R^{(\alpha)}$-Bayes estimators for other useful models. However, if the model becomes too complex, the usual Bayes computation also becomes challenging, and we have to develop appropriate computation algorithms more carefully. An alternative approach can be to approximate the $R^{(\alpha)}$-Bayes estimators for larger sample sizes using asymptotic expansions, such as Laplace's one. Such approximations for our $R^{(\alpha)}$-posterior and its expectations are provided in Majumder, Basu and Ghosh (2019) for general nonhomogeneous (but independent) observations. These computational aspects of our robust pseudo-posterior would surely form a sequence of interesting future works.

7.2. On the choice of the tuning parameter α

We have proposed a class of robust pseudo-posteriors, indexed by the tuning parameter $\alpha > 0$, which coincides with the nonrobust but (asymptotically) most efficient ordinary Bayes posterior as $\alpha \to 0$. In all our illustrations in Section 6, it is observed that, with increasing values of $\alpha > 0$, the asymptotic performance of the proposed $R^{(\alpha)}$-Bayes estimators deteriorates slightly under pure data, but their robustness under data contamination improves significantly compared with that of the usual Bayes estimates (at $\alpha = 0$). Thus, a natural and practical question arises: which α should one use for a given data set? As we have observed numerically, with a conjugate prior, any $\alpha \geq 0.5$ provides an extremely robust inference under contamination, whereas the empirically suggested range for cases with a uniform prior is $\alpha \in (0.4, 0.7)$. Thus, from our simulations presented here (along with numerous others not presented for brevity), $\alpha \approx 0.5$ seems to be a good choice in most cases.

However, a more systematic procedure for selecting this tuning parameter depending on the given data would surely be useful for reliable applications of our proposal. In this regard, note that the asymptotic distribution of the proposed ERPE at any $\alpha \geq 0$ is the same as that of the corresponding frequentist MDPDE for both the i.i.d. and INH cases (Ghosh and Basu (2016a); Majumder, Basu and Ghosh (2019)). Therefore, finding the optimal tuning parameter for the ERPE becomes an asymptotically equivalent problem of choosing an α for the optimal control between the robustness and efficiency of the MDPDE. The second one has received some attention in the literature; one such approach chooses α by minimizing an asymptotic MSE of the MDPDE, with respect to

$\alpha \in [0,1]$, given by

$$\widehat{\mathrm{AMSE}}(\alpha) = (\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^P)^T(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^P) + \frac{1}{n}\mathrm{Trace}(\Sigma_\alpha(\widehat{\boldsymbol{\theta}}_\alpha)), \tag{7.1}$$

where $\widehat{\boldsymbol{\theta}}_\alpha$ is the MDPDE at α, Σ_α is the asymptotic variance of $\sqrt{n}\widehat{\boldsymbol{\theta}}_\alpha$, and $\boldsymbol{\theta}^P$ is some suitable pilot estimator. The details can be found in Warwick and Jones (2005) and Ghosh and Basu (2015) for i.i.d. and INH setups, respectively, where some suggestions for the choice of the pilot $\boldsymbol{\theta}^P$ are also provided.

Because the asymptotic MSE of the MDPDE is indeed the same as the frequentist MSE of our ERPE, the same process can be used to choose an optimum α for the ERPE when using improper non-informative priors, with $\widehat{\boldsymbol{\theta}}_\alpha$ being replaced by the corresponding ERPE, say $\widehat{\boldsymbol{\theta}}^*_\alpha$, at any given α. However, if we have a proper subjective prior, say $\pi(\boldsymbol{\theta})$, then we can improve this approach appropriately by taking the pilot $\boldsymbol{\theta}^P$ as a random variable following $\pi(\boldsymbol{\theta})$, and then taking the expected bias in (7.1); the modified criterion is then given by

$$\widehat{\mathrm{AMSE}}^*(\alpha) = \int (\widehat{\boldsymbol{\theta}}^*_\alpha - \boldsymbol{\theta})^T(\widehat{\boldsymbol{\theta}}^*_\alpha - \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} + \frac{1}{n}\mathrm{Trace}(\Sigma_\alpha(\widehat{\boldsymbol{\theta}}^*_\alpha)), \tag{7.2}$$

which we can minimize with respect to α, possibly using a grid search over $[0,1]$, to choose an appropriate tuning parameter value. However, this proposal clearly needs further detailed investigation, which is left to future work.

8. Real-Data Applications

8.1. Hertzsprung–russell star cluster data

As our first application, let us consider the famous star cluster (CYG OB1) data from the Hertzsprung–Russell diagram containing the logarithms of the light intensity (L/L_0) and the effective temperature (T_e) at the surface of 47 stars in the direction of Cygnus (Table 3, Chapter 2, Rousseeuw and Leroy (1987)). These data have been studied by several authors (e.g., Rousseeuw and Leroy (1987); Ghosh and Basu (2013)) to demonstrate robust methods using a simple linear regression, with (L/L_0) being the response and T_e as the covariate. It has been observed there that four stars in the data (with indices 11, 20, 30, and 34) are significantly different from the remaining stars, and produce nonrobust outlier effects when using classical estimation methods.

Here, we perform Bayesian analyses of the simple linear regression model with different conjugate and improper priors. As is common practice, we assume the error variance σ^2 to be unknown. For brevity, we present only the results

Table 1. The ERPEs of the coefficients and the error variance σ^2 in the simple linear regression models for the Hertzsprung–Russell data with a uniform prior.

	Original Data			Without Four Outliers		
α	Intercept	Slope	σ	Intercept	Slope	σ
0	7.33	−0.54	0.55	−3.38	1.89	0.41
0.1	6.83	−0.42	0.58	−4.90	2.24	0.42
0.25	−8.91	3.14	0.41	−5.78	2.43	0.41
0.4	−6.13	2.51	0.42	−8.73	3.10	0.39
0.5	−6.60	2.62	0.43	−7.75	2.88	0.38
0.6	−7.19	2.75	0.41	−9.68	3.31	0.39
0.8	−7.22	2.76	0.42	−7.76	2.88	0.42

for the extreme case of uniform priors $\pi(\boldsymbol{\beta}, \sigma) = \sigma^{-1}$; the resulting values of the ERPE (with and without the outliers) are presented in Table 1. It can be clearly observed that the usual Bayes estimates (at $\alpha = 0$) are extremely nonrobust producing regression coefficients of opposite signs, owing to the presence of outliers. However, our proposed $R^{(\alpha)}$-Bayes approach and the corresponding ERPEs remain extremely stable for moderately large values of α and successfully counter the effect of outliers.

8.2. Skin data

Let us now consider another popular example of logistic regression model that has problems with outliers, namely a controlled study on the occurrence of "vaso constrictions" in the skin of digits due to air inspiration after a single deep breath (Finney (1947)). This Skin data set has been analyzed by several authors, including the recent work by Ghosh and Basu (2016b), where the logistic regression parameters are robustly estimated using the MDPDEs. Here, the important covariates to model the vaso constriction occurrences are the logarithms of the volume of inspired air ("log.Vol") and the rate of inspiration ("log.Rate"). One can observe by plotting these data (see, e.g., Ghosh and Basu (2016b)) that the fourth and 18th observations are the outliers making it difficult to separate the responses. The MLE of the corresponding regression coefficients in the logistic regression model also changes significantly, having the values (−2.88, 4.56, 5.18) in the presence of outliers, and (−24.58, 31.94, 39.55) after removing the outliers.

Here, we have considered the Bayesian modeling of the same regression model with different types of priors. Again, for brevity, we present only the case of a uniform prior over the cube $[-50, 50]^3$, having the most extreme effect of outliers. The resulting ERPEs for different values of α under the full data (including

Table 2. The ERPEs of the coefficients in a logistic regression for the Skin data with a uniform prior.

	Original Data			Without Outliers (4^{th} and 18^{th} obs.)		
α	Intercept	log(Rate)	log(Vol)	Intercept	log(Rate)	log(Vol)
0	−4.68	7.26	7.23	−22.35	35.17	29.58
0.1	−5.73	9.02	8.46	−22.32	34.96	29.62
0.25	−19.45	30.21	26.03	−22.53	34.91	30.02
0.4	−22.38	34.15	29.94	−22.91	34.92	30.61
0.5	−22.94	34.54	30.72	−23.18	34.88	31.02
0.6	−23.29	34.61	31.20	−23.41	34.79	31.37
0.8	−23.63	34.45	31.72	−23.71	34.54	31.80

outliers) and under the data without outliers are given in Table 2. Note that the values corresponding to $\alpha = 0$ give the usual Bayes estimator (posterior mean). Clearly, the usual Bayes estimates are highly affected by the presence of only two outliers, whereas our $R^{(\alpha)}$-Bayes estimators, the ERPEs, with α around 0.5 provide extremely stable results, even in the presence of outliers.

9. Conclusion

This paper presents a general Bayes pseudo-posterior under a general parametric setup and the corresponding pseudo-Bayes estimators that incorporate prior belief, in the general spirit of the Bayesian philosophy, but are also robust against data contamination. The exponential consistency of the proposed pseudo-posterior probabilities and the corresponding estimators are proved and illustrated for the cases of independent stationary and nonhomogeneous models; separate attention is given to the case of discrete priors with stationary models. Further applications of the proposed pseudo-Bayes estimators are described in the context of linear and logistic regression models. All results of Barron (1988) turn out to be special cases of our results when the tuning parameter α is set to zero.

On the whole, we trust that this study will open a new and interesting area of research on robust hybrid inference that has the flexibility to incorporate prior beliefs and inherits optimal properties from the Bayesian paradigm, along with the frequentists' robustness against data contamination. Hence, it could be very helpful in complex practical problems. In this sense, all Bayesian inference methodologies can be extended using this new pseudo-posterior. In particular, a detailed study of the examples discussed in Section 2 should be an interesting future work for different applications. Extended versions of the Bayes testing and

model selection criteria based on this new pseudo-posterior can also be developed to achieve greater robustness for inference under data contamination.

Supplementary Material

The online Supplementary Material contains proofs of all the theoretical results, additional descriptions of Figures 1–3, additional simulation results for the normal linear regression model with fixed σ, and the R code used to compute the ERPEs under different regression setups.

Acknowledgments

The authors wish to thank the editor, associate editor, and two anonymous referees for their careful reading of the manuscript and several constructive suggestions.

References

Alquier, P. and Lounici, K. (2011). PAC-Bayesian bounds for sparse regression estimation with exponential weights. *Electron. J. Stat.* **5**, 127–145.

Agostinelli, C. and Greco, L. (2013) A weighted strategy to handle likelihood uncertainty in Bayesian inference *Comput. Stat.* **28**, 319–239.

Andrade, J. A. A. and O'Hagan, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Anal.* **1**, 169–188

Andrade, J. A. A. and O'Hagan, A. (2011). Bayesian robustness modelling of location and scale parameters. *Scand. J. Stat.* **38**, 691–711.

Atkinson, A. C., Corbellini, A. and Riani, M. (2017). Robust Bayesian regression with the forward search: theory and data analysis. *TEST*, 1–18.

Barron, A. R. (1988). *The Exponential Convergence of Posterior Probabilities with Implications for Bayes Estimators of Density Functions.* Tech-Report. University of Illinois.

Barron, A. R. (1989). Uniformly powerful goodness of fit tests. *Ann. Stat.* **17**, 107–124.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85**, 549–559.

Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach.* Chapman & Hall/CRC, Boca Raton, FL.

Berger, J. O. (1994). An overview of robust Bayesian analysis. *TEST* **3**, 5–124.

Berger, J. and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *Ann. Statist.* **14**, 461–486.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Stat.* **37**, 51–58.

De Blasi, P. and Walker, S. G. (2012). Bayesian asymptotics with misspecified models. *Statist. Sinica* **23**, 169–187.

Catoni, O. (2007). PAC-Bayesian supervised classification. *Thermodyn. Stat. Learn.* **37**, IMS.

Danesi, I. L., Piacenza, F., Ruli, E. and Ventura, L. (2016). Optimal B-robust posterior distri-

butions for operational risk. *J. Operat. Risk* **11**, 35–54.

Desgagne, A. (2013). Full robustness in Bayesian modelling of a scale parameter. *Bayesian Anal.* **8**, 187–220

Delampady, M. and Dey, D. K. (1994) Bayesian robustness for multiparameter problems *J. Stat. Plann. Inf.*, 375–382.

Dey, D. K. and Birmiwal, L. (1994). Robust Bayesian analysis using divergence measures. *Stat. Prob. Lett.* **20**, 287–294.

Dupre, M. J. and Tipler, F. J. (2009). New axioms for rigorous Bayesian probability. *Bayesian Anal.* **4**, 599–606.

Efron, B. (2013). Bayes' theorem in the 21st century. *Science* **340**, 1177–1178.

Finney, D. J. (1947). The estimation from individual records of relationship between dose and quantal response. *Biometrika* **34**, 320–334.

Gelfand, A. E. and Dey, D. K. (1991). On Bayesian robustness of contaminated classes of priors. *Statist. Decisions* **9**, 63–80.

Gelman, A., Meng, X. L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6**, 733–807.

Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Stat.* **28**, 500–531.

Ghosal, S. and van der Vaart, A. W. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Stat.* **35**, 192–223.

Ghosh, A. and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.* **7**, 2420–2456.

Ghosh, A. and Basu, A. (2015). Robust estimation for non-homogeneous data and the selection of the optimal tuning parameter: The DPD approach. *J. App. Stat.* **42**, 2056–2072.

Ghosh, A. and Basu, A. (2016a). Robust Bayes estimation using the density power divergence. *Ann. Inst. Stat. Math.* **68**, 413–437.

Ghosh, A. and Basu, A. (2016b). Robust estimation in generalized linear models : The density power divergence approach. *TEST* **25**, 269–290.

Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. Springer.

Greco, L., Racugno, W. and Ventura, L. (2008). Robust likelihood functions in Bayesian analysis. *J. Stat. Plann. Inf.* **138**, 1258–1270

Gruenwald, P. and van Ommen, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal.* **12**, 1069–1103

Gustafson, P. and Wasserman, L. (1995). Local sensitivity diagnostics for Bayesian inference. *Ann. Stat.* **23**, 2153–2167.

Halpern, J.Y. (1999). A counterexample to theorems of Cox and fine. *J. Art. Int. Res.* **10**, 67–85.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J. and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons.

Holmes, C. C. and Walker, S. G. (2017). Assigning a value to a power likelihood in a general Bayesian model. *Biometrika* **104**, 497–503.

Hooker, G. and Vidyashankar, A. N. (2014). Bayesian model robustness via disparities. *TEST* **23**, 556–584.

Jiang, W. and Tanner, M. A. (2008). Gibbs posterior for variable selection in high dimensional

classification and data mining. *Ann. Statist.* **36**, 2207–2231.

Kang, J. and Lee, S. (2014). Minimum density power divergence estimator for Poisson autoregressive models. *Comput. Stat. Data Anal.* **80**, 44–56.

Kim, B. and Lee, S. (2011). Robust estimation for the covariance matrix of multi-variate time series. *J. Time Ser. Anal.* **32**, 469–481.

Kim, B. and Lee, S. (2013). Robust estimation for the covariance matrix of multivariate time series based on normal mixtures. *Comput. Stat. Data Anal.* **57**, 125–140.

Kleijn, B. and Van der Vaart, A. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Stat.* **34**, 837–877.

Lee, S. and Song, J. (2013). Minimum density power divergence estimator for diffusion processes. *Ann. Inst. Stat. Math.* **65**, 213–236.

Majumder, T., Basu, A. and Ghosh, A. (2019). On robust pseudo-Bayes estimation for the independent non-homogeneous set-up. *ArXiv preprint, arXiv:1911.12160.*

Millar, R. B. and Stewart, W. S. (2007). Assessment of locally influential observations in Bayesian models. *Bayesian Anal.* **2**, 365–384.

Nakagawa, T. and Hashimoto, S. (2017). Robust Bayesian inference based on quasi-posterior under heavy contamination. *Hiroshima Research Group Technical Report*, TR17-5.

Owhadi, H., Scovel, C. and Sullivan, T. J. (2015). Brittleness of Bayesian inference under finite information in a continuous world. *Electron. J. Stat.* **9**, 1–79.

Ramamoorthi, R.V., Sriram, K. and Martin, R. (2015). On posterior concentration in misspecified models. *Bayesian Anal.* **10**, 759–789.

Ritov, Y. A. (1985). Robust Bayes decision procedures - gross error in the data distribution. *Ann. Stat.* **13**, 626–637.

Ritov, Y. A. (1987). Asymptotic results in robust quasi-Bayesian estimation. *J. Mult. Anal.* **23**, 290–302.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons.

Shalizi, C. R. (2009). Dynamics of Bayesian updating with dependent data and misspecified models, *Electron. J. Stat.* **3**, 1039–1074.

Shyamalkumar, N. D. (2000). Likelihood Robustness. In *Robust Bayesian Analysis*, 127–143. Springer.

Sivaganesan, S. (1993). Robust Bayesian diagnostic. *J. Stat. Plann. Inf.* **35**, 171–188.

Song, J., Lee, S., Na, O. and Kim, H. (2007). Minimum density power divergence estimator for diffusion parameter in discretely observed diffusion processes. *Korean Comm. Stat.* **14**, 267–280.

Walker, S. G. (2004). New approaches to Bayesian consistency. *Ann. Stat.* **32**, 2028–2043.

Walker, S. G. and Hjort, N. L. (2001). On Bayesian consistency. *J. Royal Stat. Soc. B* **63**, 811–821.

Walker, S. G., Lijoi, A. and Prunster, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Stat.* **35**, 738–746.

Wang, C. and Blei, D. M. (2018). A general method for robust Bayesian modeling. *Bayesian Analysis* **13**, 1163–1191.

Warwick, J. and Jones, M. C. (2005). Choosing a robustness tuning parameter. *J. Stat. Comput. Simul.* **75**, 581–588.

Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *J. Royal Stat. Soc. B* **58**, 739–

750.

Abhik Ghosh

Interdisciplinary Statistical Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, West Bengal, India.

E-mail: abhik.ghosh@isical.ac.in

Tuhin Majumder

Department of Statistics, North Carolina State University, 2800 Brigadoon Dr., Apt 22, Raleigh, NC 27606, USA.

E-mail: tmajumd@ncsu.edu

Ayanendranath Basu

Interdisciplinary Statistical Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, West Bengal, India.

E-mail: ayanbasu@isical.ac.in

(Received January 2019; accepted August 2020)