Statistica Sinica Preprint No: SS-2019-0450

Title	General Robust Bayes Pseudo-Posteriors: Exponential Convergence Results with Applications
Manuscript ID	SS-2019-0450
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0450
Complete List of Authors	Abhik Ghosh, Tuhin Majumder and Ayanendranath Basu
Corresponding Author	Ayanendranath Basu
E-mail	ayanbasu@isical.ac.in
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

General Robust Bayes Pseudo-Posteriors: Exponential Convergence Results with Applications

Abhik Ghosh[†], Tuhin Majumder[‡] and Ayanendranath Basu[†*]

[†] Indian Statistical Institute, India. [‡] North Carolina State University, USA.

[*]Corresponding author: ayanbasu@isical.ac.in

Abstract

Although Bayesian inference is an immensely popular paradigm among a large segment of scientists including statisticians, most applications consider objective priors and need critical investigations [20]. While it has several optimal properties, a major drawback of Bayesian inference is the lack of robustness against data contamination and model misspecification, which becomes pernicious in the use of objective priors. This paper presents the general formulation of a Bayes pseudo-posterior distribution yielding robust inference. Exponential convergence results related to the new pseudo-posterior and the corresponding Bayes estimators are established under the general parametric set-up and illustrations are provided for the independent stationary as well as non-homogeneous models. Several additional details and properties of the procedure are described, including the estimation under fixed-design regression models.

Keywords: Robust Bayes Pseudo-Posterior, Density Power Divergence, Exponential Convergence, Bayesian Linear Regression, Logistic Regression.

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 non-statisticians due to its nice interpretability and incorporation of prior knowledge. From a statistical point of view, it is widely accepted even among many non-Bayesians, because of its nice optimal (asymptotic) properties. Bayesian inference is built on the famous 'Bayes theorem', the celebrated 1763 paper of Thomas Bayes, which combines prior knowledge with experimental evidence to produce posterior conclusions. However, over these 250 years, Bayesian inference has also been subject to several criticisms and some of these debates are still ongoing [20]. Other than the controversies about its internal logic [19, 34], a major practical drawback of Bayesian inference is its non-robust nature against misspecification in models (including data contamination and outliers) and priors, as has been extensively observed in the literature; see [12, 13, 45, 47, 61] and the references therein. The optimal solution to this problem has been developed mainly for prior misspecifications [10, 11, 17, 18, 22, 30, 33]; the Bayesians traditionally viewed the model to be perfect for the given data. Thus the possibility of model misspecification and data contamination has been generally ignored for a long time till the appearance of some very recent publications, some of which we describe later in this section.

In applying Bayesian inference to the complicated datasets of the present era, we need to use complex and sophisticated models which are highly prone to misspecification or data contamination. In reality, where "All models are wrong", the Bayesian philosophy of refining the fixed model adaptively [23] often fails to handle complex scenarios or leads to "a model as complex as the data" [59]. Data contamination can lead to erroneous posterior conclusions. The problem becomes more clear but pernicious in case of inference 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 non-robust sample mean. What is a matter of greater concern, as noted by Efron [20], is that most of the recent applications of Bayesian inference hinge on objective priors and so they always need to be scrutinized carefully, sometimes even from a frequentist perspective. The posterior non-robustness 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 only few solutions to the problem of model misspecification [49, 50, 53, 54]. However, most of them, if not all, assume that the perturbation in the model is known beforehand, such as gross error contaminated models with known contamination proportion ϵ. For modern complex datasets, this is rarely meaningful. There has been several recent publications which are motivated by the need to safeguard Bayes inference against model misspecification by relying on a generalized (pseudo) posterior which is expressed in terms of a loss function and a tuning parameter η [1, 13, 14, 32, 36, 38, 42, 48, 57]. 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. But the resulting inference is not robust against outliers with respect to a specified model which is correct for the majority of the data. This is because every sample observation, including outliers, receives equal weight in PAC-Bayesian approach and hence it closely resembles the model robust non-parametric analysis; see [28].

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 the use of heavy tailed distributions (e.g., t-distribution in place of normal), or new (robust) inference methodologies [9, 35]. The first one has been adopted by some Bayesian scientists; see [3, 4] and [16] 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 non-robustness of Bayesian inference for a specified model (which might be of a lighter tail). The second approach of frequentist robustness serves the purpose but differs in the strictest probabilistic sense from the Bayesian philosophy, since 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. Different such pseudo-posteriors have been proposed by [2, 5, 15, 28, 31, 37, 46]; but all of them have primarily considered independent stationary models and have different pros and cons. Another recent attempt, in the borderline of these two approaches, has been proposed by [59], who have transformed the given model to a localized model involving hyperparameters to be estimated through the empirical Bayes approach.

1.1 Background: $R^{(\alpha)}$-posterior for IID set-up

We consider a particular pseudo-posterior originally proposed by [28] in the independently and identically distributed (IID) set-up. This choice has been motivated by its several nice properties and its potential for extension to more general set-ups. As a brief description, consider n IID 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 $\mathcal{B}/\Omega$ measurable, independent with respect to P and it's induced distribution $G(x)$ has 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 [28] 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

$$\pi_n^{(\alpha)}(\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n) = \frac{\exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi(\boldsymbol{\theta})}{\int \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}'))\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}'}, \qquad \alpha \geq 0, \tag{1}$$

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

$$q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) \;=\; \frac{1}{\alpha}\sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\alpha}(x_i) - \frac{n}{1+\alpha}\int f_{\boldsymbol{\theta}}^{1+\alpha} - \frac{n}{\alpha} = \sum_{i=1}^{n} q_{\boldsymbol{\theta}}^{(\alpha)}(x_i), \tag{2}$$

with G_n being the empirical distribution based on the data and

$$q_{\boldsymbol{\theta}}^{(\alpha)}(y) = \frac{1}{\alpha}\left(f_{\boldsymbol{\theta}}^{\alpha}(y) - 1\right) - \frac{1}{1+\alpha}\int f_{\boldsymbol{\theta}}^{1+\alpha}. \tag{3}$$

In a limiting sense, $q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(\log(f_{\boldsymbol{\theta}}(x_i)) - 1\right)$, which is the usual log-likelihood (plus a constant); so 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 [8], which has proven to be a useful robust generalization of the maximum likelihood estimator (MLE); see [28] for details. The similarity of this approach (at $\alpha > 0$) with the usual Bayes posterior (at $\alpha = 0$) is that, it does not require nonparametric smoothing like some other pseudo-posteriors and it is additive in the data so that the posterior update is easy with new observations. In [28], its robustness is demonstrated and a Bernstein-von Mises type limiting result is proved under the IID set-up.

1.2 The Contribution of This Paper

We provide a generalization of the $R^{(\alpha)}$-posterior density for a completely general parametric model set-up beyond IID data, through a suitable structural definition of the α-likelihood function, and derive the exponential convergence results associated with the new pseudo-posterior for the general set-up. These, in fact, generalize the corresponding results for the usual Bayes posterior [6] for the $R^{(\alpha)}$-posterior and their advantages are illustrated through several applications. Our major contribution in the present paper can be summarized as follows.

- This paper is the first to define a robust pseudo-posterior for the general class of parametric models with a finite set of parameters. All the previous literature on pseudo-posterior are confined to the IID set-up or a particular example of a non-IID case. Our model set-up is extremely general to cover the IID case as well as every type of non-homogeneous and dependent observations provided the inference is to be performed based on a finite set of parameters. We have defined a robust $R^{(\alpha)}$-posterior and the associated estimators for such a general class of statistical inference problems covering enormous applications.

- To illustrate the wide applicability of our proposal, we have explicitly presented the forms of the $R^{(\alpha)}$-posterior or the α-likelihood function for several important cases like the independent non-homogeneous data including linear and logistic regressions, time series and Markov models, diffusion processes, etc. Our $R^{(\alpha)}$-posteriors also contain the usual Bayes posterior at $\alpha \to 0$ and hence provides a direct generalization of the latter at $\alpha > 0$.

- All the previous pseudo-posteriors currently available in the literature sacrifice the conditional probability interpretation of the usual Bayes theory. In this paper, for the first time, we 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; the $R^{(\alpha)}$-posterior indeed becomes the ordinary Bayes posterior for such a modified set-up (Remark 2.1). 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 set-up. We study the asymptotic properties of the $R^{(\alpha)}$-marginal and the corresponding joint density of data and 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 set-up. As per our knowledge, such an optimal asymptotic property is not available for any other pseudo-posterior.

- The assumptions needed for our theoretical derivations are indeed extensions of those required for the classical Bayes theory [6]; 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 have further simplified these conditions for the IID and the non-homogeneous set-ups. They are verified for common examples like linear regression with known or unknown error variance and logistic regression models. Although the initial set of conditions under the general parametric models look more stringent than the current literature, we have illustrated that they indeed hold under very mild conditions in common examples; e.g., 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 have also separately studied the interesting cases of discrete priors under IID set-up, and the associated maximum $R^{(\alpha)}$-posterior estimator with their exponential consistency.

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

For brevity, all proofs and the R-codes are given in the Online Supplement.

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

In order to extend the $R^{(\alpha)}$-posterior density to a more general set-up, 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). Also assume that 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 $\lambda^n(d\underline{\boldsymbol{x}}_n)$ having "true" probability density $g^n(\underline{\boldsymbol{x}}_n)$. We wish to model it by 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 having density $f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ for each n. Note that, we have not assumed the parameter space Θ_n to be 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 (2), we propose to define the α-likelihood function $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ in such a way that ensures

$$q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) := \lim_{\alpha \downarrow 0} q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \log f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) - n, \text{ for all } \underline{\boldsymbol{x}}_n \in \chi_n. \tag{4}$$

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 set-up, we define the corresponding $R^{(\alpha)}$-posterior probabilities as

$$\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n) = \frac{\int_{A_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\Theta_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}}, \qquad A_n \in \mathcal{B}_{\Theta_n}, \tag{5}$$

whenever the denominator is finitely defined and is positive; otherwise we may define it arbitrarily, e.g., $\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n) = \pi_n(A_n)$. Definition (4) ensures that $\pi_n^{(0)}$ is the usual Bayes posterior.

For an 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 will 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 (5) in terms of this $R^{(\alpha)}$-marginal density as $\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n) = \frac{m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n)}{m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, \Theta_n)} = \frac{m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)}{m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)}$, 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

$$L_n^{(\alpha)Bayes}(d\boldsymbol{\theta}, d\underline{\boldsymbol{x}}_n) = \pi_n^{(\alpha)}(d\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n)\, M_n^{(\alpha)}(d\underline{\boldsymbol{x}}_n) = \frac{M_n^{(\alpha)}(d\underline{\boldsymbol{x}}_n, d\boldsymbol{\theta})}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)}. \tag{6}$$

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 $\alpha = 0$, all these again simplify to the ordinary Bayes measures.

Example 2.1 [Independent Stationary Data]:

The simplest possible set-up is that of IID 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 $G^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n G(x_i)$, $g^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n g(x_i)$, $\lambda^n(d\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n \lambda(dx_i)$, $F^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n F_{\boldsymbol{\theta}}(x_i)$, $f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n f_{\boldsymbol{\theta}}(x_i)$ and so $\mathcal{F}_n$ is also the n-fold product of the family of individual distributions $F_{\boldsymbol{\theta}}$. Under these notations, the α-likelihood $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$, which is given by (2), satisfies the required measurability assumptions along with the condition in (4).

Then, under suitable assumptions on the prior distribution as before, the corresponding $R^{(\alpha)}$-posterior distribution is defined by (5) which is now equivalent to (1) and can be written as a product of stationary independent terms corresponding to each x_i (additivity). Other related measures can be defined from these quantities; we will come back to them again in Section 4. □

Example 2.2 [Independent Non-homogeneous 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, 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 to be 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 estimation of $\boldsymbol{\theta}$.

This set-up of independent non-homogeneous (INH) observations covers many interesting practical problems, the most common one being the regression with fixed design. Suppose $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ 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

$$E(X_i) = \psi(\boldsymbol{t}_i, \boldsymbol{\beta}), \quad i = 1, \ldots, n, \tag{7}$$

where $\boldsymbol{\beta} \subseteq \boldsymbol{\theta}$ is 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 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_is, associated with observations x_is, have the INH structure with the common parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$ and the different densities $f_{i,\boldsymbol{\theta}}$. We can further extend this set-up to include the 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 set-up. 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 $G^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n G_i(x_i)$, $g^n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n g_i(x_i)$, $\lambda^n(d\underline{\boldsymbol{x}}_n) = \prod_{i=1}^n \lambda(x_i)$, $F^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n F_{i,\boldsymbol{\theta}}(x_i)$ and $f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \prod_{i=1}^n f_{i,\boldsymbol{\theta}}(x_i)$ so that $\mathcal{F}_n = \otimes_{i=1}^n \mathcal{F}^i$.

Now, under this INH set-up, we can 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 IID version in (2) keeping in mind the general requirement (4). Borrowing ideas from [26], who have developed the MDPDE for the INH set-up, and following the intuition behind the construction of the α-likelihood (2) of [28], one possible extended definition for the α-likelihood in the INH case can be given by

$$q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^{n}\left[\frac{1}{\alpha}f_{i,\boldsymbol{\theta}}^{\alpha}(x_i) - \frac{1}{1+\alpha}\int f_{i,\boldsymbol{\theta}}^{1+\alpha}\right] - \frac{n}{\alpha} = \sum_{i=1}^{n} q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i), \tag{8}$$

with $q_{i,\boldsymbol{\theta}}^{(\alpha)}(y) = \frac{1}{\alpha}\left(f_{i,\boldsymbol{\theta}}^{\alpha}(y) - 1\right) - \frac{1}{1+\alpha}\int f_{i,\boldsymbol{\theta}}^{1+\alpha}$. Note that, we have $q_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(\log(f_{i,\boldsymbol{\theta}}(x_i)) - 1\right)$, satisfying the required condition in (4). So, assuming a suitable prior for $\boldsymbol{\theta}$, the $R^{(\alpha)}$-posterior for the INH observations is defined through (5) with $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ being given by (8). Note that, the resulting posterior is again a product of independent but non-homogeneous terms. We will discuss their properties in detail in Section 5. □

Remark 2.1. *In the first introduction of the $R^{(\alpha)}$-posterior under IID set-up [28], it was noted that its only drawback is the loss of the probabilistic interpretation. Here also, so far, we have defined the $R^{(\alpha)}$-posterior differently than the conditional probability approach of the usual Bayes theory and called it a pseudo-posterior. But, in fact, it can also be interpreted as an ordinary Bayes posterior under a suitably modified model and prior, which becomes prominent in our general set-up. To see this, define an α-modified model density $\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \frac{\exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})}$ and the α-modified prior density $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) = \frac{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})}{M_n^{(\alpha)}(\boldsymbol{\chi}_n,\Theta_n)}$. Both are proper densities and satisfy the required measurability assumptions whenever the relevant integrals exist finitely. Further, $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ only (independent of the data) and hence may be used as a prior density in Bayesian inference; but it depends on α and the model. In particular, at $\alpha = 0$, $\widetilde{\pi}_n^{(0)}(\boldsymbol{\theta}) = \pi_n(\boldsymbol{\theta})$ and $\widetilde{q}_n^{(0)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = f^n(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ so they indeed 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 $\mathcal{F}_{n,\alpha} = \left\{\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\right\}$ and the (α-modified) prior $\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})$ is given by $\frac{\int_{A_n}\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\Theta_n}\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})d\boldsymbol{\theta}}$, which simplifies to $\pi_n^{(\alpha)}(A_n|\underline{\boldsymbol{x}}_n)$ as in (5).*

In the following we briefly present the forms of the α-likelihood for some other practically important model set-ups, but their detailed investigations are kept for the future.

Example 2.3 [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 parametric inference, we model $g(X_t|\mathcal{G}_t)$ by a parametric density $f_{\boldsymbol{\theta}}(X_t|\mathcal{F}_{t-1})$ and try to infer about 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 get inference based on the posterior density of $\boldsymbol{\theta}$ given the observed sample data. We can develop the robust Bayesian inference for any such time series model through 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 [39–41, among others], we can define the corresponding α-likelihood function as

$$q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{t=1}^{n} \left[\frac{1}{\alpha} f_{\boldsymbol{\theta}}^{\alpha}(x_t|\mathcal{F}_{t-1}) - \frac{1}{1+\alpha} \int f_{\boldsymbol{\theta}}^{1+\alpha}(x|\mathcal{F}_{t-1}) dx \right] - \frac{n}{\alpha}. \tag{9}$$

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 (4). Robust $R^{(\alpha)}$-posterior inference about $\boldsymbol{\theta}$ can be developed using this α-likelihood function. □

Example 2.4 [Markov Process]:

Example 2.3 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 $(\Omega, \mathcal{B}_\Omega, P)$ having true transition probabilities $g(X_{k+1}|X_k)$, $k = 0, 1, 2, \ldots, n-1$, with X_0 being the initial value of the process. We model it by 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 $q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \sum_{k=1}^{n} \left[\frac{1}{\alpha} f_{\boldsymbol{\theta}}^{\alpha}(x_{k+1}|x_k) - \frac{1}{1+\alpha} \int f_{\boldsymbol{\theta}}^{1+\alpha}(x|x_k) dx \right] - \frac{n}{\alpha}$. Clearly it satisfies Condition (4) and it is possible to perform robust $R^{(\alpha)}$-Bayes inference about $\boldsymbol{\theta}$ under this set-up. □

Example 2.5 [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 $dX_t = a(X_t, \boldsymbol{\mu})dt + b(X_t, \sigma)dW_t$, $t \geq 0$, with $X_0 = x_0$ and two known functions a and b, where $\{W_t : t \geq 0\}$ 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 $X_{t_i^n}$, $i = 1, \ldots, n$, from the above diffusion process. We generally 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 $b(X_t, \sigma) = \sigma$, respectively, by [55] and [43]. However, whenever we have some prior knowledge about $\boldsymbol{\theta}$, quantified through a prior $\pi(\boldsymbol{\theta})$, one would apply the Bayesian approach. A robust Bayes inference can be done by using our $R^{(\alpha)}$-posterior. For this purpose, we note that $X_{t_i^n} = X_{t_{i-1}^n} + a(X_{t_{i-1}^n}, \boldsymbol{\mu})h_n + b(X_{t_{i-1}^n}, \sigma)\sqrt{h_n} Z_{n,i} + \Delta_{n,i}$, $\quad i = 1, \ldots, n$, 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 IID standard normal variables for $i = 1, \ldots, n$. Therefore, whenever $\Delta_{n,i}$ can be ignored in P-probability, for large enough n, $X_{t_i^n}|\mathcal{G}_{i-1}^n$, $i = 1, \ldots, n$, 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 2.3. It satisfies the general requirement (4) 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

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

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

3 Exponential Convergence Results under the General Set-up

Exponential consistency is an important property of posterior (Bayes) inference; it was first demonstrated in [6] and later refined by several authors [see 24, 25, 56, 58, among others]. We follow the approach of [6] to show that of 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

Let us recall the general set-up 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 2.1. 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 $KLD(f_1, f_2) = \int f_1 \log\left(\frac{f_1}{f_2}\right) d\lambda$, and put $D_n^{(\alpha)}(\boldsymbol{\theta}) = \frac{1}{n} KLD\left(g^n(\cdot), \widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta})\right)$. We define a joint (frequentist) law of $\boldsymbol{\theta}$ and $\underline{\boldsymbol{X}}$ 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 $\pi_n^{*(\alpha)}(d\boldsymbol{\theta}) = \frac{e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}\widetilde{\pi}_n^{(\alpha)}(d\boldsymbol{\theta})}{c_n}$, 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 (6) of $\boldsymbol{\theta}$ and $\underline{\boldsymbol{X}}_n$; to quantify their closeness we consider the concept of "*merging*" of probability distributions [6].

Definition 3.1. *Consider two probability distributions G_1^n and G_2^n of $\underline{\boldsymbol{X}}_n$, having densities g_1^n and g_2^n respectively with respect to λ^n.*

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

An application of Markov's inequality shows that Definition 3.1 is equivalent to the conditions $\lim\limits_{n\to\infty} \frac{1}{n} \log \frac{g_2^n(\underline{\boldsymbol{X}}_n)}{g_1^n(\underline{\boldsymbol{X}}_n)} = 0$ in probability or with probability one, respectively. See Barron [6, Section 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 $\widetilde{\pi}_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right) \geq e^{-nr}$, for all $n \geq N$.

Theorem 3.1. *Under Assumption (M1), we have the following results.*

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

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

Although Assumption (M1) might look a bit complicated, it can be further simplified in terms of 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 usual Bayesian methods and is precisely defined below for our context.

Definition 3.2. *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* $\mathcal{F}_{n,\alpha} = \left\{\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\right\}$ *if there exists a finite measure* $\widetilde{\pi}$ *such that* $\widetilde{\pi}\left(\{\boldsymbol{\theta} : \bar{D}^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\}\right) > 0$*, for all* $\epsilon > 0$*, and*

$$\liminf_{n\to\infty} e^{nr} \frac{d\widetilde{\pi}_n^{(\alpha)}}{d\widetilde{\pi}}(\boldsymbol{\theta}) \geq 1, \quad \textit{for all } r > 0, \boldsymbol{\theta} \in \Theta. \tag{10}$$

Theorem 3.2. *If the prior is information dense with respect to* $\mathcal{F}_{n,\alpha}$ *as in Definition 3.2, then Assumption (M1) holds and hence the three results of Theorem 3.1 also 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, i.e., $A_n \cup B_n \cup C_n = \Theta_n$, for each $n \geq 1$.

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

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

(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 $\sup_{\boldsymbol{\theta}\in C_n} \frac{Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} \leq c_n$, where i.o. denotes "*infinitely often*".

Here we need either Condition (A3) or Condition (A3)* which, respectively, help us to prove the convergence results in probability or with probability one. Condition (A3)* is stronger and imply (A3), but (A3) is sufficient in most practices yielding a convergence in probability type result. Also, 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 $\left\{ \frac{Q_n^{(\alpha)}(\cdot|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta})} : \boldsymbol{\theta} \in C_n \right\}$ corresponding to the α-modified model density $\widetilde{q}_n^{(\alpha)}(\cdot|\boldsymbol{\theta})$ defined in Remark 2.1. Although complex looking, these conditions are straightforward extensions of the conditions used by [6] for proving the exponential convergence of ordinary Bayes posterior probabilities; they indeed coincide at $\alpha = 0$. In particular, at $\alpha = 0$, Condition (A2) simplifies to $\pi_n(B_n) \leq b_n$, i.e., B_n 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 3.3. *[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, $\limsup_{n\to\infty} P\left(\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n) < e^{-nr}\right) = 1$, 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, $P\left(\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n) \geq e^{-nr} \text{ i.o.}\right) = 0$, for some $r > 0$, if and only if there exists 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 3.3 coincides with the classical exponential convergence results of ordinary Bayes posterior probabilities as proved in [6]. Our theorem generalizes it for the robust $R^{(\alpha)}$-posterior probabilities under suitable conditions. Hence, the $R^{(\alpha)}$-posterior distribution, besides yielding robust results under data contamination, is asymptotically optimal in exactly 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 (5) under the general set-up of Section 2. In the decision-theoretic framework, we consider the problem of estimation of 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 (like mean) of P. For the given parametric family $F^n(\cdot|\boldsymbol{\theta})$, let us 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 non-negative loss function $L_n(\phi, \widehat{\phi})$ on $\Phi \times \Phi$ denoting the loss in estimating ϕ by $\widehat{\phi}$; let $L_n(\phi_{\boldsymbol{\theta}}, \phi)$ is $\mathcal{B}_{\Theta_n}$ 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

$$\widehat{\phi} = \arg\min_{\phi \in \Phi} \int L_n(\phi_{\boldsymbol{\theta}}, \phi) \pi_n^{(\alpha)}(d\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n), \tag{11}$$

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 squared error loss provided it exists finitely, or a median of the $R^{(\alpha)}$-posterior distribution for absolute error loss.

However, if the minimum in (11) is not attained, we may define the approximate $R^{(\alpha)}$-Bayes estimator $\widehat{\phi}$ of ϕ through the relation $\int L_n(\phi_{\boldsymbol{\theta}}, \widehat{\phi}) \pi_n^{(\alpha)}(d\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n) \leq \inf_{\phi \in \Phi} \int L_n(\phi_{\boldsymbol{\theta}}, \phi) \pi_n^{(\alpha)}(d\boldsymbol{\theta}|\underline{\boldsymbol{x}}_n) + \delta_n$, with $\lim_{n\to\infty} \delta_n = 0$. An useful example is the approximate mode of the $R^{(\alpha)}$-posterior for discrete parameter space, which is an approximate $R^{(\alpha)}$-Bayes estimator under 0-1 loss. Also, note that, if the $R^{(\alpha)}$-Bayes estimator exists, it is also an approximate $R^{(\alpha)}$-Bayes estimator.

Definition 3.3. *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 3.4. *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 3.4 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. Also, 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 loss. The proof mimics that of Lemma 12 in [6].

Theorem 3.4 (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. Also, for any $\epsilon > 0$, define $A_{\epsilon,n} = \{\boldsymbol{\theta} : d_n(\phi_P, \phi_{\boldsymbol{\theta}}) \leq \epsilon\}$. Then, we have $d_n(\phi_P, \widehat{\phi}_n) \leq \epsilon + h_2 \left(\frac{\epsilon + \bar{L}\pi_n^{(\alpha)}\left(A^c_{h_1^{-1}(\epsilon),n} | \underline{\boldsymbol{x}}_n\right)}{1 - \pi_n^{(\alpha)}(A^c_{\epsilon,n} | \underline{\boldsymbol{x}}_n)} \right)$. 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 3.4 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. But, Theorem 3.3 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 3.4 applies to the $R^{(\alpha)}$-Bayes estimators with respect to the squared Hellinger loss and the L_1-loss. to deduce their L_1 consistency.

4 Application (I): Independent Stationary Models

4.1 $R^{(\alpha)}$-Posterior Convergence

Consider the set-up of the independent stationary model as in Example 2.1. Let us study the conditions required for the exponential convergence of the $R^{(\alpha)}$-posterior for this particular set-up. First, to verify the merging of G^n and $M_n^{(\alpha)}$, we define the individual α-modified density as $\widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta}) = \exp\left(q_{\boldsymbol{\theta}}^{(\alpha)}(\cdot)\right)/Q^{(\alpha)}(\chi|\boldsymbol{\theta})$ and the α-modified prior $\widetilde{\pi}_n^{(\alpha)}$ as in Remark 2.1 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 4.1. *The prior π under the IID model is information dense at G with respect to $\mathcal{F}_\alpha$ if there exists a finite measure $\widetilde{\pi}$ satisfying (10) and $\widetilde{\pi}\left(\left\{\boldsymbol{\theta} : KLD(g, \widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta})) < \epsilon\right\}\right) > 0$ for all $\epsilon > 0$.*

Note that, the above definition is equivalent to the general notion of information denseness given in Definition 3.2. Thus, in view of Theorem 3.2, it implies the merging of G^n and $M_n^{(\alpha)}$ in probability for independent stationary models. Then, Theorem 3.3 may be restated as follows.

Proposition 4.1. *Consider the set-up 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 4.1. 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 exists constants $r_1, r_2 > 0$ and sets $B_n, C_n \in \mathcal{B}_\Theta$ such that such that (A1)–(A3) are satisfies 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 $\left\{\frac{Q^{(\alpha)}(\cdot|\boldsymbol{\theta})}{Q^{(\alpha)}(\chi|\boldsymbol{\theta})} : \boldsymbol{\theta} \in C_n\right\}$. We can further simplify it by using a necessary and sufficient condition for the existence of UEC from [7] 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 $p^T(x) = \frac{1}{\lambda(A)}\int_A p(y)\lambda(dy)$, for $x \in A \in T$, whenever $\lambda(A) \neq 0$, and $p^T = 0$ otherwise. We call a sequence of partitions T_n to be "*rich*" if the corresponding sequence of densities g^{T_n} converges to g in L_1-distance. Also, 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, and consider the following assumption.

Assumption (B): *For $\epsilon > 0$, $\widetilde{\pi}_n^{(\alpha)}(B_\epsilon^{T_n}) = \frac{M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_\epsilon^{T_n})}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta)}$ 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. So, applying it with $B_n = \left\{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) \geq \epsilon, d_{T_n}\left(G, \frac{Q^{(\alpha)}(\cdot|\boldsymbol{\theta})}{Q^{(\alpha)}(\chi|\boldsymbol{\theta})}\right) < \epsilon/2\right\} \subset B_{\epsilon/4}^{T_n}$ and the existence result of UEC tests with $C_n = \left\{\boldsymbol{\theta} : d_{T_n}\left(G, \frac{Q^{(\alpha)}(\cdot|\boldsymbol{\theta})}{Q^{(\alpha)}(\chi|\boldsymbol{\theta})}\right) > \epsilon/2\right\}$, Proposition 4.1 yields the asymptotic exponential concentration of the $R^{(\alpha)}$-posterior probability in the L_1-neighborhood $A_n = \{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) < \epsilon\}$. Note that, clearly $A_n \cup B_n \cup C_n = \Theta_n$ for these choices.

Theorem 4.2. *Consider the set-up of IID 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 4.1. If Assumption (B) holds then, for every $\epsilon > 0$, $\pi_n^{(\alpha)}\left(\{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) \geq \epsilon\} | \underline{\boldsymbol{X}}_n\right)$ is exponentially small with P-probability one.

Note that the final Assumption (B) is easy to verify for model 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 in terms of T_n-variance distance in place of the L_1 distance. The idea goes back to [6] for a similar result in case of the ordinary posterior; an extended version for the $R^{(\alpha)}$-posterior is given in the following.

Theorem 4.3. *Consider the set-up of IID 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 4.1. Then, for any sequence of partitions T_n with effective card. of order n, $\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : d_{T_n}\left(G, \frac{Q_n^{(\alpha)}(\cdot|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta})}\right) \geq \epsilon\right\} \middle| \underline{\boldsymbol{X}}_n\right)$ 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 3.4 along with Proposition 4.1–4.3. Let us now consider, in more detail, the particular case of discrete priors and the maximum $R^{(\alpha)}$-posterior estimator.

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

$$\widetilde{\pi}_n^{(\alpha)}(\widehat{\boldsymbol{\theta}}_\alpha)\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\widehat{\boldsymbol{\theta}}_\alpha) > \sup_{\boldsymbol{\theta}} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})e^{-n\delta_n}, \tag{12}$$

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 2.1). 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 data $\underline{\boldsymbol{x}}_n$, is measurable, and consider such prior sequence that satisfies

$$\liminf_{n\to\infty} e^{nr}\widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) \geq 1, \qquad \text{for all } r > 0,\ \boldsymbol{\theta} \in \Theta. \tag{13}$$

Assumption (13) signifies that the (α-modified) prior probabilities are not exponentially small anywhere in Θ. Then, we have the following theorems.

Theorem 4.4. *Consider the set-up of IID models with fixed countable $\Theta_n = \Theta$ and discrete prior sequence π_n satisfying Assumption (13). Suppose π_n is information dense at the true probability mass function g with respect to $\mathcal{F}_\alpha$ as in Definition 4.1 and $\pi_n^{(\alpha)}\left(A_n^c|\underline{\boldsymbol{X}}_n\right)$ 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 4.5. *Consider the set-up of stationary independent models with fixed countable $\Theta_n = \Theta$ and a discrete prior sequence π_n satisfying Assumption (13). Then, for any true density g which is an information limit of the (countable) family $\left\{\widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\right\}$ and for any $\epsilon > 0$, we have $\pi_n^{(\alpha)}\left(\{\boldsymbol{\theta} : d_1(g, f_{\boldsymbol{\theta}}) \geq \epsilon\}\,|\underline{\boldsymbol{X}}_n\right)$ is exponentially small with probability one. So $\lim_{n\to\infty} d_1(g, f_{\widehat{\boldsymbol{\theta}}_\alpha}) = 0$, with probability one for any AMRPE $\widehat{\boldsymbol{\theta}}_\alpha$.*

Remark 4.1. *Theorem 4.5, in a special case $\alpha = 0$, yields a stronger version of Theorem 15 of [6]. Our result requires fewer assumptions than required by Barron's result.*

5 Application (II): Independent Non-homogeneous Models

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

Let us now consider the set-up of independent but non-homogeneous (INH) models as described in Example 2.2 of Section 2, and simplify the exponential convergence results for the $R^{(\alpha)}$-posterior probabilities under this INH set-up. 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 $\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}) = \frac{\prod_{i=1}^n \exp\left(q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i)\right)}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} = \prod_{i=1}^n \widetilde{q}^{(i,\alpha)}(x_i|\boldsymbol{\theta})$ with $\widetilde{q}^{(i,\alpha)}(x_i|\boldsymbol{\theta}) = \frac{\exp\left(q_{i,\boldsymbol{\theta}}^{(\alpha)}(x_i)\right)}{Q^{(i,\alpha)}(\chi^i|\boldsymbol{\theta})}$. Thus, in the notation of Section 3.1, we have $D_n^{(\alpha)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n KLD\left(g_i, \widetilde{q}^{(i,\alpha)}(\cdot|\boldsymbol{\theta})\right)$, and hence the definition of information denseness can be simplified for the INH models as follows.

Definition 5.1. *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 (10) such that $\widetilde{\pi}\left(\left\{\boldsymbol{\theta} : \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n KLD\left(g_i, \widetilde{q}^{(i,\alpha)}(\cdot|\boldsymbol{\theta})\right) < \epsilon\right\}\right) > 0$, for all $\epsilon > 0$.*

When $f_{i,\boldsymbol{\theta}} = f_{\boldsymbol{\theta}}$ is independent of i, then the INH set-up coincides with the IID set-up and the information denseness in Definition 5.1 coincides with that in Definition 4.1. Further, Definition 5.1 is also equivalent to the general Definition 3.2 and hence implies that G^n and $M_n^{(\alpha)}$ merge in probability. Then, we have the following simplified results for the INH set-up.

Proposition 5.1. *Consider the set-up 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 5.1. Then, for any sequence of measurable parameter sets $A_n \subset \Theta$, $\pi_n^{(\alpha)}\left(A_n^c|\underline{\boldsymbol{X}}_n\right)$ is exponentially small with P-probability one, if and only if there exists sequences of measurable parameter sets $B_n, C_n \subset \Theta$ such that $A_n \cup B_n \cup C_n = \Theta$, $\frac{M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n)}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \leq e^{-nr}$ for $r > 0$ and a UEC test for G^n against $\left\{Q_n^{(\alpha)}(\cdot|\boldsymbol{\theta})/Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) : \boldsymbol{\theta} \in C_n\right\}$ exists.*

However, the existence of the required UEC in Proposition 5.1 is equivalent to the existence of a UEC test for G_i against $\left\{\frac{Q^{(i,\alpha)}(\cdot|\boldsymbol{\theta})}{Q^{(i,\alpha)}(\chi^i|\boldsymbol{\theta})} : \boldsymbol{\theta} \in C_n\right\}$ uniformly over $i = 1, \ldots, n$. Following the discussions of Section 4.1, this holds if Assumption (B) is satisfied for $\widetilde{B}_\epsilon^{T_n} = \left\{\boldsymbol{\theta} : \frac{1}{n}\sum_{i=1}^n d_1(f_{i,\boldsymbol{\theta}}, \widetilde{q}^{(i,\alpha)}(\cdot|\boldsymbol{\theta})^{T_n}) > \epsilon\right\}$ in place of $B_\epsilon^{T_n}$. This leads to following simplification.

Theorem 5.2. *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 5.1. 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 $\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : \frac{1}{n}\sum_{i=1}^n d_1(g_i, f_{i,\boldsymbol{\theta}}) \geq \epsilon\right\}|\underline{\boldsymbol{x}}_n\right)$ is exponentially small with P-probability one for $\epsilon > 0$.*

We note that the Bernstein-von Mises type asymptotic results for the $R^{(\alpha)}$-posterior distribution under the INH set-up would be extremely important to provide contraction rates for our new robust pseudo-posterior; similar results for IID models were discussed in [28]. However, considering the length of the present paper and to keep its focus clear on the exponential convergence results, we propose to present the results on contraction rates for INH models in a sequel paper; for the time being, they are made available in the ArXiv version [44].

5.2 Robust Bayes Estimation under Fixed Design Regression Models

As noted in Example 2.2, the most common example of the general INH set-up is the fixed design regression models. We consider the important example of model (7) with n fixed k-variate design points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ and $f_{i,\boldsymbol{\theta}}(x) = \frac{1}{\sigma} f\left(\frac{x-\psi(\boldsymbol{t}_i,\boldsymbol{\beta})}{\sigma}\right)$ 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 $q_{i,(\boldsymbol{\beta},\sigma)}^{(\alpha)}(x_i) = \frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i-\psi(\boldsymbol{t}_i,\boldsymbol{\beta})}{\sigma}\right)^\alpha - \frac{M_{f,\alpha}}{(1+\alpha)\sigma^\alpha} - \frac{1}{\alpha}$, 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 the Jeffrey's prior given by $\pi(\boldsymbol{\beta},\sigma) = \sigma^{-1}$. Then, the $R^{(\alpha)}$-posterior density of $(\boldsymbol{\beta},\sigma)$ is given by (5) which now simplifies as

$$\pi_n^{(\alpha)}\left((\boldsymbol{\beta},\sigma)|\underline{\boldsymbol{x}}_n\right) = \frac{\prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i-\psi(\boldsymbol{t}_i,\boldsymbol{\beta})}{\sigma}\right)^\alpha - \frac{M_{f,\alpha}}{(1+\alpha)\sigma^\alpha}\right]\pi(\boldsymbol{\beta},\sigma)}{\int\int\prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i-\psi(\boldsymbol{t}_i,\boldsymbol{\beta})}{\sigma}\right)^\alpha - \frac{M_{f,\alpha}}{(1+\alpha)\sigma^\alpha}\right]\pi(\boldsymbol{\beta},\sigma)d\boldsymbol{\beta}d\sigma}. \tag{14}$$

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 can get the simplified form for the $R^{(\alpha)}$-posterior density of $\boldsymbol{\beta}$ as given by

$$\pi_n^{(\alpha)}\left(\boldsymbol{\beta}|\underline{\boldsymbol{x}}_n\right) = \frac{\prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i-\psi(\boldsymbol{t}_i,\boldsymbol{\beta})}{\sigma}\right)^\alpha\right]\pi(\boldsymbol{\beta})}{\int\prod_{i=1}^n \exp\left[\frac{1}{\alpha\sigma^\alpha} f\left(\frac{x_i-\psi(\boldsymbol{t}_i,\boldsymbol{\beta})}{\sigma}\right)^\alpha\right]\pi(\boldsymbol{\beta})d\boldsymbol{\beta}}. \tag{15}$$

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, here we assume that the error variance σ is known; the unknown σ case is considered later. In this case, we can simplify the $R^{(\alpha)}$-posterior from (15) 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).

However, being a particular case of the INH set-up, the exponential consistency of the $R^{(\alpha)}$-posterior of $\boldsymbol{\beta}$ directly holds under the assumptions of Proposition 5.1. We now verify the required conditions for this present case 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 square 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 the 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 $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{ik})^T$, $i = 1, \ldots, n$, are such that, for all $j, l, s = 1, \ldots, k$,

$$\sup_{n>1} \max_{1\leq i\leq n} |t_{ij}| = O(1), \qquad \max_{1\leq i\leq n} |t_{ij}||t_{il}| = O(1), \qquad \frac{1}{n}\sum_{i=1}^{n} |t_{ij}t_{il}t_{is}| = O(1). \tag{16}$$

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

Note that these 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 [26]. They are easy to verify for any given design matrix; in particular they hold if $\boldsymbol{t}_i$'s are generated from some non-singular k-variate distributions. It is shown in [44] that these two conditions indeed ensure a Bernstein-von Mises type result for the associated $R^{(\alpha)}$-posterior.

It is really fascinating to see that, despite the complex natures 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 $\alpha \geq 0$ for the example of linear regression (along with some mild conditions on the prior). The result is presented in the following theorem.

Theorem 5.3. *Consider the normal linear regression set up with known error variance. Assume that the true parameter value is $\boldsymbol{\beta}_0$, i.e., $g_i = f_{i,\boldsymbol{\beta}_0}$ 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*

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

or equivalently,

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

i.e., the $R^{(\alpha)}$-posterior probabilities asymptotically concentrates on the neighborhoods of the true regression line at a 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 normal linear regression with unknown error variance. Consider the set-up 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 0 and variance σ; but now we consider σ^2 to be also 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 (14) with $M_f = (2\pi)^{-\alpha/2}(1+\alpha)^{-1/2}$.

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

Theorem 5.4. *Consider the normal linear regression set up with unknown error variance. Assume that the true parameter value is $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \sigma_0^2)$, i.e., $g_i = f_{i,\boldsymbol{\theta}_0}$ 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*

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

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 notation of Example 2.2, given fixed-design points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$, the logistic regression model considers binary response variables x_i, respectively, having Bernoulli distribution with expectation $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = \dfrac{e^{\boldsymbol{t}_i^T \beta}}{1 + e^{\boldsymbol{t}_i^T \beta}}$, for $i = 1, \ldots, n$. As in Example 2.2, this model clearly belongs to the INH set-up 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 (8) with $f_{i,\boldsymbol{\theta}}$ being the probability mass function of 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 (5) given any prior $\pi(\boldsymbol{\beta})$, which does not have a closed form and needs to be computed numerically; see Section 6 for illustrations.

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 the Assumption (R1) on the fixed design points and consider the new condition (R3) in terms of the matrix $\boldsymbol{\Psi}_n(\boldsymbol{\beta}) = n^{-1} E_{g_i}\left[\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta})\right]$.

(R3) $\inf_n$ [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 [29], as well as in the Bernstein-von Mises type results for the corresponding $R^{(\alpha)}$-posterior distribution [44]. 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 set-up, as presented in the following theorem.

Theorem 5.5. *Consider the fixed-design regression set up as above. Assume that the true parameter value is $\boldsymbol{\beta}_0$, i.e., $g_i = f_{i,\boldsymbol{\beta}_0}$ 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*

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

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 equals one. The corresponding $R^{(\alpha)}$-posterior is given by (15), as discussed in Section 5.3, and has no closed form solution. So, we have computed the ERPE through 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 through 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 have 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 20000 steps in the importance sampling Monte-Carlo with the proposal density $N_k\left(\widehat{\boldsymbol{\beta}}, n^{-1}(\boldsymbol{D}^T\boldsymbol{D})^{-1}\right)$. We replicate the above procedure 1000 times to compute the empirical bias and MSE of the ERPE for two different priors, namely the non-informative uniform prior and the conjugate normal prior, which are presented in the Online Supplement (Figures 1 and 2) due to page restriction. 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 very 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 unknown error variance σ in the above linear regression model, as discussed in Section 5.4. We repeat the above simulation exercise for the unknown σ case as well, by taking the true value of $\sigma_0 = 1$ and the conjugate prior on $(\boldsymbol{\beta}, \sigma)$ given by $\pi(\boldsymbol{\beta}, \sigma) = \pi(\boldsymbol{\beta}|\sigma)\pi(\sigma)$, where $\pi(\beta|\sigma)$ is taken to be $N_2(\boldsymbol{\beta}_0, \sigma^2 I_2)$ density and $\pi(\sigma)$ is the density of the square root of Inverse chi-square distribution with 5 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; alternatively we have used the Metropolis-Hastings algorithm.

Algorithm 1: Computation of ERPE in LRM with unknown variance:

We generate 20000 sample observation from $R^{(\alpha)}$ 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 k^{th} 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 $\gamma = \frac{\exp[q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta}^*, \sigma^*)g_1(\boldsymbol{\beta}^*)g_2(\sigma^*)}{\exp[q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\beta}^{(k-1)}, \sigma^{(k-1)})g_1(\boldsymbol{\beta}^{(k-1)})g_2(\sigma^{(k-1)})}$.

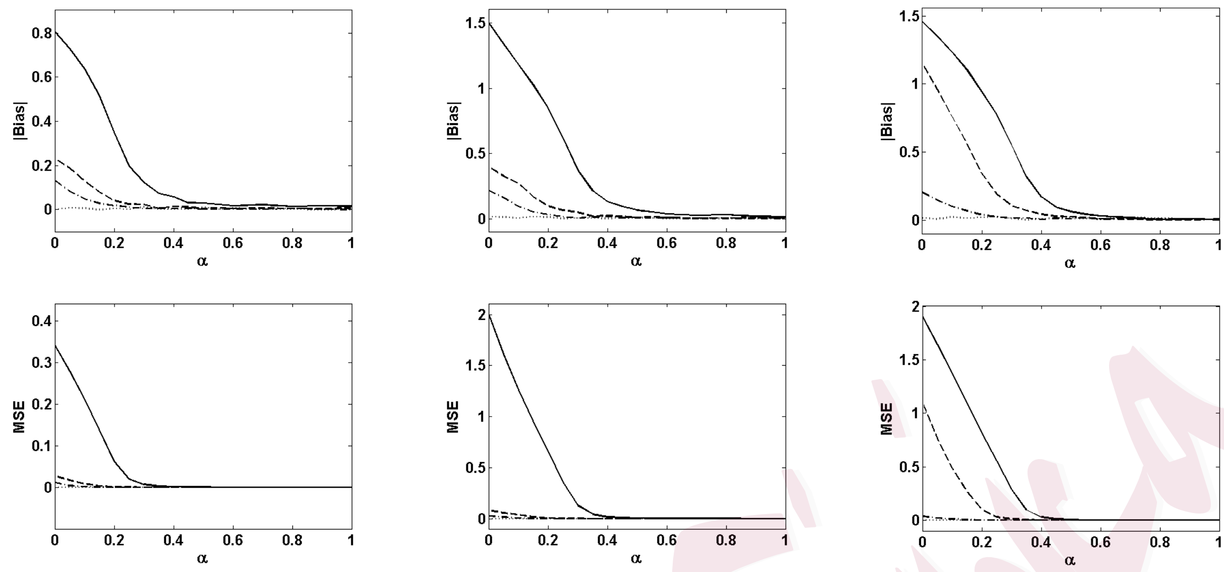
Step 4. If $U < \gamma$, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$ and $\sigma^{(k)} = \sigma^*$. Otherwise, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)}$ and $\sigma^{(k)} = \sigma^{(k-1)}$.

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

In each cases, the first 5000 values generated are rejected as burn-in and the remaining 15000 parameter values are averaged to get a good approximation of the ERPE of $(\boldsymbol{\beta}, \sigma)$. □

The process is replicated 1000 times to compute the empirical biases and MSEs of the ERPEs of $\boldsymbol{\beta}$ and σ at different α for the previous simulation set-up. The resulting values of total absolute bias and the total MSE over the two components of $\boldsymbol{\beta}$ as well as the absolute bias and MSE of the ERPE of σ are presented in Figures 1 and 2, respectively.

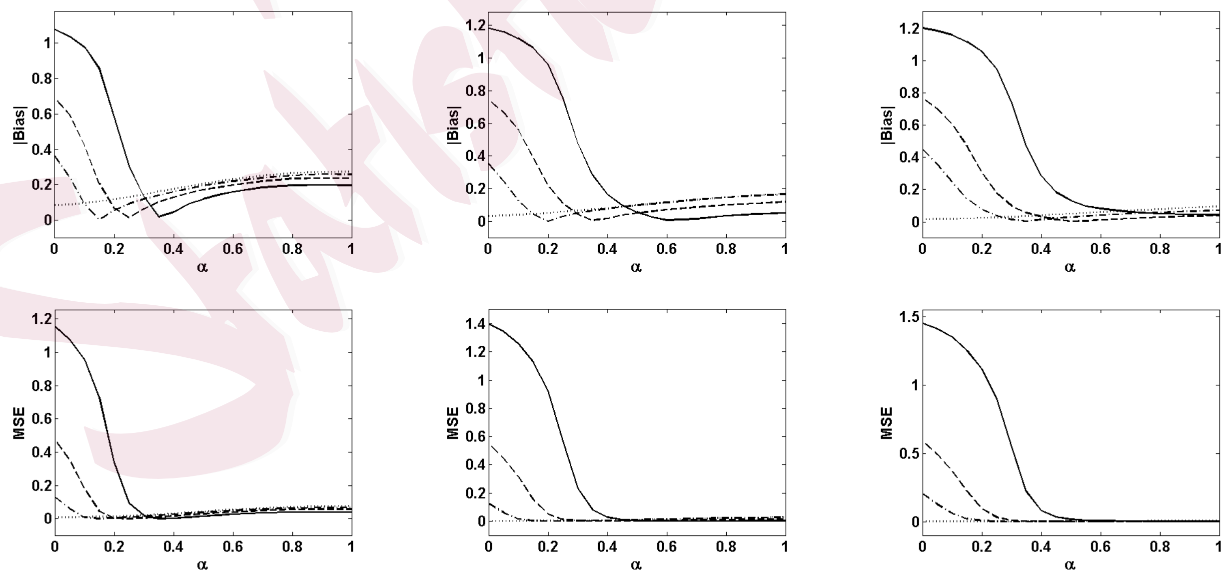
The performances of the ERPE of regression coefficient and error variance are again the same as before in that the proposed ERPE with larger α provide extremely stable estimates even under contamination up to 20%. Under pure data the usual Bayes estimators give minimum absolute bias and MSEs, but the ERPEs with $\alpha > 0$ are also not very far away. However, under data contamination, the usual Bayes estimates (at $\alpha = 0$) become extremely non-robust yielding significantly higher



(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 additional discussions in the Online Supplement)

bias and MSEs even though we are using 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.



(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 additional discussions in the Online Supplement)

6.2 Performance of ERPE in Logistic Regression Model

We now consider the fixed-design logistic regression model as 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. Since the corresponding $R^{(\alpha)}$-posterior has no closed form solution, we have 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 through the logistic regression structure with x_i generated from Bernoulli distribution with mean parameter $\psi(\boldsymbol{t}_i, \boldsymbol{\beta}) = \dfrac{e^{\boldsymbol{t}_i^T\beta}}{1+e^{\boldsymbol{t}_i^T\beta}}$ 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 through misspecification of the response values, i.e., by changing x_i to $(1 - x_i)$, and the prior is taken as the (bivariate) normal distribution as $\pi(\boldsymbol{\beta}) \equiv N_2(\boldsymbol{\beta}_0, I_2)$. However, in this case also, the importance sampling is seen to fail to provide a good approximation to the ERPE and we have alternatively used the Metropolis-Hastings method. Note that, the target density, i.e $R^{(\alpha)}$ posterior density here 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 20000 sample observation from $R^{(\alpha)}$ posterior distribution of $\boldsymbol{\beta}$ as follows.

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

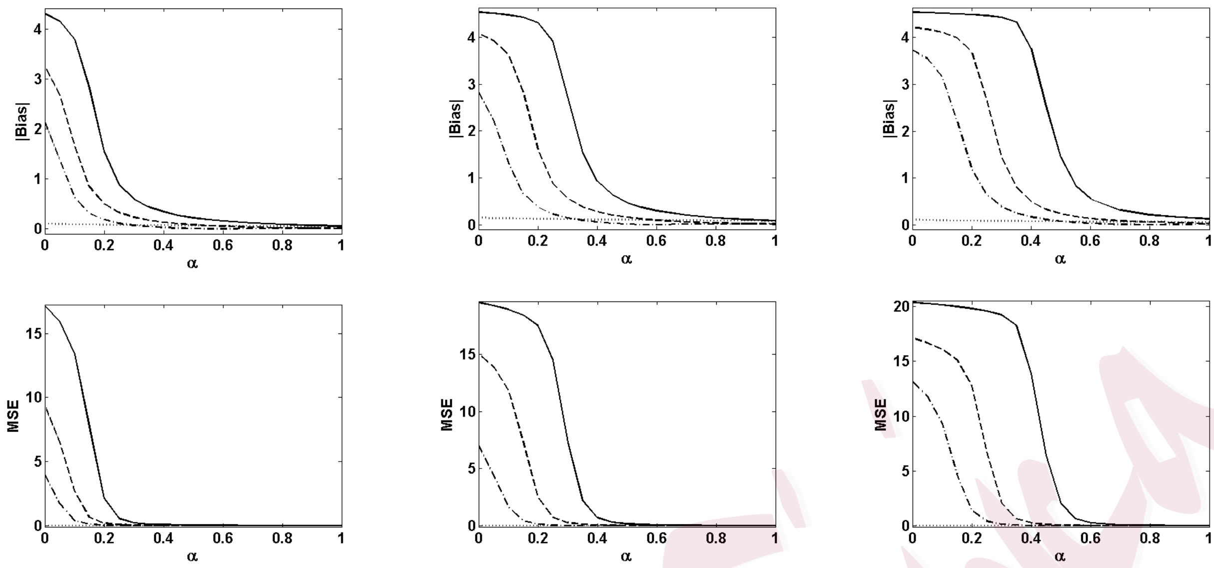
Step 2. After generating $\boldsymbol{\beta}^{(k-1)}$ in the $(k-1)$-th step, at the k^{th} step, generate $\boldsymbol{\beta}^*$ from $\mathcal{N}_2(\boldsymbol{\beta}^{(k-1)}, I_2)$.

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 5000 values generated are rejected as burn-in and the remaining 15000 parameter values are averaged to get a good approximation of the ERPE. $\square$



(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 normal prior. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$] (See additional discussions in the Online Supplement)

The simulation exercise is replicated 1000 times to compute 1000 ERPEs of $\boldsymbol{\beta}$. Their empirical biases and MSEs are presented in Figure 3. Here also, it is clearly observed that 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 MSE 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 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 also, no closed form expressions exist in most applications and hence we need to compute the corresponding $R^{(\alpha)}$-Bayes estimators numerically. One such possible approach could be the use of the importance sampling technique, which is seen to work well in our illustrations for normal means ([28]) or linear models with known σ (Section 6.1). But, this simple approach can be useful only when it is possible to utilize some conjugacy structure; in our cases, the standard posterior distribution is used as the proposal distribution due to their 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, the linear regression with unknown variance or the logistic regression models. In such cases, we propose to use a suitable Metropolis-Hastings algorithm that is seen to work 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 have also supplied the relevant R codes for the computations of the ERPEs for our examples in the Online Supplement.

We hope that, with the advance in modern computers, it would 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 like Laplace's one; such an approximation for our $R^{(\alpha)}$-posterior and its expectations are provided in [44] for general non-homogeneous (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 non-robust 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 deteriorate slightly under pure data, but their robustness under data contamination improves significantly compared to 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 that, with conjugate prior, any $\alpha \geq 0.5$ provides extremely robust inference under contamination, whereas the empirically suggested range for the cases with uniform prior is $\alpha \in (0.4, 0.07)$; 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 selection of this tuning parameter depending on the given data at hand would surely be useful for reliable applications of our proposal. In this regard, we 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 IID and INH cases [28, 44]. Therefore, finding the optimal tuning parameter for the ERPE becomes an asymptotically equivalent problem of choosing an α for the optimal control between 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

$$\widehat{\text{AMSE}}(\alpha) = (\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^P)^T(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^P) + \frac{1}{n}\text{Trace}(\Sigma_\alpha(\widehat{\boldsymbol{\theta}}_\alpha)), \tag{17}$$

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 [60] and [27] for IID and INH set-ups, respectively, where some suggestions regarding the choice of pilot $\boldsymbol{\theta}^P$ are also provided.

Since the asymptotic MSE of the MDPDE is indeed the same as the frequentist MSE of our ERPE, the same process can be used to chose optimum α for the ERPE here 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 expected bias in (17); the modified criterion is then given by

$$\widehat{\mathrm{AMSE}}^*(\alpha) = \int (\widehat{\boldsymbol{\theta}}^*_\alpha - \boldsymbol{\theta})^T (\widehat{\boldsymbol{\theta}}^*_\alpha - \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} + \frac{1}{n} \mathrm{Trace}(\Sigma_\alpha(\widehat{\boldsymbol{\theta}}^*_\alpha)), \tag{18}$$

which we can minimize with respect to α, possibly through a grid search over $[0, 1]$, to chose an appropriate tuning parameter value. However, this proposal clearly needs further detailed investigation which, considering the length of the current paper, we hope to do in a 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 about 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, [51]). These data are studied by several authors (e.g., [26, 51]) for demonstration of robust methods through 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 indeed significantly different from the remaining stars and produce the non-robust outlier effects while using classical estimation methods.

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

Table 1: The ERPEs of the coefficients and error variance σ^2 in the simple linear regression models for the Hertzsprung-Russell data with 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

be clearly observed that the usual Bayes estimates (at $\alpha = 0$) are extremely non-robust producing regression coefficients of opposite sign due 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 models having outlier issue, namely a controlled study on the occurrence of "vaso constrictions" in the skin of digits due to air inspiration after a single deep breath [21]. This Skin dataset was analyzed by several authors including the recent work by [29] where the logistic regression parameters are robustly estimated by 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, for example, [29]) that the 4-th and 18-th observations are indeed the outliers making it difficult to separate the responses; the MLE of the corresponding regression coefficients in the logistic regression model also changes significantly to have the values $(-2.88, 4.56, 5.18)$ in the presence of outliers and $(-24.58, 31.94, 39.55)$ after removal of 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 uniform prior over the cube $[-50, 50]^3$ having the most extreme effect of outliers. The resulting ERPE for different values of α under the full data (including outliers) as well as under the outlier deleted data are given in Table 2; note that the values

Table 2: The ERPEs of the coefficients in a logistic regression for the Skin data with 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

corresponding to $\alpha = 0$ gives the usual Bayes estimator (posterior mean). Clearly, the usual Bayes estimates get highly affected by the presence of only two outliers whereas our $R^{(\alpha)}$-Bayes estimators, the ERPEs, with α around 0.5 provides extremely stable results even in the presence of outliers.

9 Concluding Remark

This paper presents a general Bayes pseudo-posterior under general parametric set-up that produces pseudo-Bayes estimators which incorporate prior belief in the general spirit of 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 non-homogeneous 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 [6] turn out to be special cases of our results when the tuning parameter α is set to 0.

On the whole, we trust that this paper opens up a new and interesting area of research on robust hybrid inference that has the flexibility to incorporate prior belief and inherits optimal properties from the Bayesian paradigm along with the frequentists' robustness against data contamination and hence could be very helpful in different complex practical problems. In this sense, all Bayesian inference methodologies can be extended with 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.

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

References

[1] Alquier, P. and Lounici, K. (2011). PAC-Bayesian bounds for sparse regression estimation with exponential weights. *Electronic Journal of Statistics*, 5, 127–145.

[2] Agostinelli, C. and Greco, L. (2013) A weighted strategy to handle likelihood uncertainty in Bayesian inference *Comput Stat* , 28(1), 319-239.

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

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

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

[6] Barron, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. *Tech-Report*, University of Illinois.

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

[8] 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.

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

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

[11] Berger, J. and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *Ann. Statist.*, 14(2), 461–486.

[12] Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Stat.*, 37, 51–58. [Corrigendum 37, 745–746].

[13] De Blasi, P., Walker, S. G. (2012). Bayesian asymptotics with misspecified models. *Statistica Sinica*, 23, 169–187.

[14] Catoni, O. (2007). PAC-Bayesian Supervised Classification. *Thermodyn. Stat. Learn.*, 37, IMS.

[15] Danesi, I. L., Piacenza, F., Ruli, E., and Ventura, L. (2016). Optimal B-robust posterior distributions for operational risk. *pre-print*.

[16] Desgagne, A. (2013). Full Robustness in Bayesian Modelling of a Scale Parameter. *Bayesian Anal.*, 8(1), 187–220

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

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

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

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

[21] Finney, D. J. (1947), The estimation from individual records of relationship between dose and quantal response. *Biometrika*, 34(3/4), 320-334.

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

[23] Gelman, A., Meng, X. L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 733-760.

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

[25] Ghosal, S., and van der Vaart, A. W. (2007). Convergence rates of posterior distributions for non-IID observations. *Ann. Stat.*, 35(1), 192–223.

[26] Ghosh, A., and Basu, A. (2013). Robust estimation for independent non-homogeneous observa-

tions using density power divergence with applications to linear regression. *Electron. J. Stat.*, 7, 2420—2456.

[27] 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(9), 2056—2072.

[28] Ghosh, A., and Basu, A. (2016a). Robust Bayes Estimation using the Density Power Divergence. *Ann. Inst. Stat. Math.*, 68(2), 413–437.

[29] Ghosh, A., and Basu, A. (2016b). Robust Estimation in Generalized Linear Models : The Density Power Divergence Approach. *TEST, 25(2)*, 269—290.

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

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

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

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

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

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

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

[37] Hooker, G., and Vidyashankar, A. N. (2014). Bayesian Model Robustness via Disparities. *TEST 23(3)*, 556–584.

[38] Jiang, W. and Tanner, M. A. (2008). Gibbs posterior for variable selection in high dimensional classification and data mining. *Ann. Statist.*, 36, 2207–2231.

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

[40] Kim, B., and Lee, S. (2011). Robust estimation for the covariance matrix of multi-variate time series. *J. Time Ser. Anal.*, 32(5), 469–481.

[41] 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.

[42] Kleijn, B., and Van der Vaart, A. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Annals of Statistics*, 34,

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

[44] 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 [math.ST].

[45] Millar, R. B., and Stewart, W. S. (2007). Assessment of Locally Influential Observations in Bayesian Models. *Bayesian Anal.*, 2(2), 365–384.

[46] Nakagawa, T., and Hashimoto, S. (2017). Robust Bayesian inference based on quasi-posterior under heavy contamination. *Technical Report.*

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

[48] Ramamoorthi, R.V., Sriram, K., and Martin, R. (2015). On posterior concentration in misspecified models. Bayesian Analysis, vol. 10, 759-789.

[49] Ritov, Y. A. (1985). Robust Bayes Decision Procedures - Gross Error in the Data Distribution *Ann. Stat.*, 13(2) 626–637

[50] Ritov, Y. A. (1987). Asymptotic results in robust quasi-bayesian estimation. *J. Mult. Anal.*, 23(2), 290-302.

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

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

[53] Shyamalkumar, N. D. (2000). Likelihood Robustness. In: *Robust Bayesian Analysis.* Springer.

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

[55] 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(2), 267–280.

[56] Walker, S. G. (2004). New approaches to Bayesian consistency. *Ann. Stat.*, 32(5), 2028–2043.

[57] Walker, S. G., and Hjort, N.L. (2001). On Bayesian Consistency. *J. Royal Stat. Soc. B* 63(4), 811–821.

[58] 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.

[59] Wang, C., and Blei, D. M. (2016). A General Method for Robust Bayesian Modeling. *ArXiv Preprint*, arXiv:1510.05078.

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

[61] Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *J. Royal Stat. Soc. B*, 58, 739–750.