Supplementary Material for “ General Robust Bayes Pseudo-Posteriors: Exponential Convergence results with Applications”

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

Indian Statistical Institute, Kolkata, India.

[†] abhianik@gmail.com

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

1 Proofs of the Results of Section 3 in Main paper

1.1 Proof of Theorem 1

First note that, by help of Equation (2.3) of the main paper, we can rewrite the $R^{(\alpha)}$-Bayes joint distribution as

$$L_n^{(\alpha)Bayes}(d\boldsymbol{\theta}, d\underline{\boldsymbol{x}}_n) = \frac{M_n^{(\alpha)}(d\underline{\boldsymbol{x}}_n, d\boldsymbol{\theta})}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} = \widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})\widetilde{\pi}_n^{(\alpha)}(d\boldsymbol{\theta})\lambda^n(d\underline{\boldsymbol{x}}_n), \tag{1.1}$$

which has a density $\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta})$ with respect to $\widetilde{\pi}_n^{(\alpha)} \times \lambda^n$. On the other hand, the frequnetist approximation $L_n^{*(\alpha)}$ has the density function $e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}g_n\left(\underline{\boldsymbol{x}}_n\right)/c_n$ and hence we get

$$\begin{aligned}
KLD\left(L_n^{*(\alpha)}, L_n^{(\alpha)Bayes}\right) &= E_{L_n^{*(\alpha)}}\left[\log \frac{e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}g_n\left(\underline{\boldsymbol{X}}_n\right)/c_n}{\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta})}\right] \\
&= E_{\pi_n^{*(\alpha)}}E_{G_n}\left[-nD_n^{(\alpha)}(\boldsymbol{\theta}) + \log\frac{g_n\left(\underline{\boldsymbol{X}}_n\right)}{\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta})} - \log c_n\right] \\
&= E_{\pi_n^{*(\alpha)}}\left[-nD_n^{(\alpha)}(\boldsymbol{\theta}) + E_{G_n}\log\frac{g_n\left(\underline{\boldsymbol{X}}_n\right)}{\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta})}\right] - \log c_n \\
&= E_{\pi_n^{*(\alpha)}}\left[-nD_n^{(\alpha)}(\boldsymbol{\theta}) + nD_n^{(\alpha)}(\boldsymbol{\theta})\right] - \log c_n \\
&= -\log c_n = -\log\left[\int e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}\widetilde{\pi}_n^{(\alpha)}(d\boldsymbol{\theta})\right]
\end{aligned}$$

Therefore, for any $\epsilon > 0$, we get

$$\begin{aligned}
\frac{1}{n}KLD\left(L_n^{*(\alpha)}, L_n^{(\alpha)Bayes}\right) &= -\frac{1}{n}\log\left[\int e^{-nD_n^{(\alpha)}(\boldsymbol{\theta})}\widetilde{\pi}_n^{(\alpha)}(d\boldsymbol{\theta})\right] \\
&\leq \frac{\epsilon}{2} - \frac{1}{n}\log\widetilde{\pi}_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right) \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \qquad \text{for all but finitely many } n,
\end{aligned} \tag{1.2}$$

by applying Assumption (M1) with $r = \frac{\epsilon}{2}$. Since $\epsilon > 0$ is arbitrary, this completes the proof of the first equation in Part (a) of the theorem.

The second equation of Part (a) and the first equation in Part (b) of the theorem follows by the relation

$$\frac{1}{n}KLD\left(L_n^{*(\alpha)}, L_n^{(\alpha)Bayes}\right) = \frac{1}{n}E_{G^n}\left[KLD\left(\pi_n^{*(\alpha)}(\cdot), \pi_n^{(\alpha)}(\cdot|\underline{\boldsymbol{X}}_n)\right)\right] + \frac{1}{n}KLD\left(g^n, m_n^{(\alpha)}\right).$$

Finally, to proof the the last part of (b) in the theorem, we note that the Kullback-Leibler divergence satisfies the relation

$$E\left|\log\frac{g_n(\underline{\boldsymbol{X}}_n)}{m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n)}\right| \leq KLD\left(g^n, m_n^{(\alpha)}\right) + \frac{2}{e}.$$

Therefore, by the first part of (b), we get $\lim_{n\to\infty} E\left|\log\frac{g_n(\underline{\boldsymbol{X}}_n)}{m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n)}\right| = 0$ and hence G^n and $M_n^{(\alpha)}$ merge in probability by using Markov inequality. □

1.2 Proof of Theorem 2

To show Assumption (M1), let us fix $\epsilon, r > 0$ and define $\rho_n(\boldsymbol{\theta}) = e^{nr}\frac{d\pi_n}{d\widetilde{\pi}}(\boldsymbol{\theta})$. Then, using Fatou's Lemma, we get

$$\begin{aligned}
\liminf_{n\to\infty} e^{nr}\pi_n\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right) &= \liminf_{n\to\infty} e^{nr}\int I\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right)\pi_n(d\boldsymbol{\theta}) \\
&= \liminf_{n\to\infty}\int I\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right)\rho_n(\boldsymbol{\theta})\widetilde{\pi}(d\boldsymbol{\theta}) \\
&= \int \liminf_{n\to\infty} I\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right)\rho_n(\boldsymbol{\theta})\widetilde{\pi}(d\boldsymbol{\theta}) \\
&= \int I\left(\left\{\boldsymbol{\theta} : \bar{D}^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right)\widetilde{\pi}(d\boldsymbol{\theta}) \\
&= \widetilde{\pi}\left(\left\{\boldsymbol{\theta} : D_n^{(\alpha)}(\boldsymbol{\theta}) < \epsilon\right\}\right), \qquad (1.3)
\end{aligned}$$

which is strictly positive by the information denseness with respect to $\mathcal{F}_{n,\alpha}$ (Definition 2 of the main paper). This implies Assumption (M1) and we are done via Theorem 1. □

1.3 Proof of Theorem 3

We use an argument similar to that used by Barron (1988). Let us consider the following two assumptions in addition to Assumptions (A1)–(A3) and (A3)*.

(A4) The true distribution G^n and the $R^{(\alpha)}$-marginal distribution $M_n^{(\alpha)}$ satisfy

$$\lim_{n\to\infty} P\left(\frac{m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n)}{g^n(\underline{\boldsymbol{X}}_n)} \geq a_n\right) = 1.$$

(A4)* The true distribution G^n and the $R^{(\alpha)}$-marginal distribution $M_n^{(\alpha)}$ satisfy

$$P\left(\frac{m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n)}{g^n(\underline{\boldsymbol{X}}_n)} < a_n \ \text{ i.o.}\right) = 0.$$

Note that, if Conditions (A4) and (A4)* hold with $a_n = e^{-n\epsilon}$ for every $\epsilon > 0$, they indicate that the true distribution G^n and the $R^{(\alpha)}$-marginal distribution $M_n^{(\alpha)}$ merge in probability or with probability one respectively.

Now, we start with two primary results on the convergence of the $R^{(\alpha)}$-posterior probabilities.

Lemma 1.1 *Suppose Assumptions (A1)–(A3) and (A4) hold with* $\lim b_n = \lim c_n = 0$ *such that* $r_n := (b_n + c_n)/a_n$ *is finitely defined. Then, for all* $\delta > 0$*, we have*

$$\limsup_{n\to\infty} P\left(\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n) > \frac{r_n}{\delta}\right) \leq \delta. \tag{1.4}$$

Further, if additionally Assumptions (A3)* and (A4)* are satisfied, then for any summable sequence $\delta_n > 0$ *we have*

$$P\left(\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n) > \frac{r_n}{\delta_n} \ i.o.\right) = 0. \tag{1.5}$$

Proof: Note that, with G^∞ probability one, the $R^{(\alpha)}$-posterior probability can be re-expressed as

$$\pi_n^{(\alpha)}\left(A_n^c|\underline{\boldsymbol{X}}_n\right) = \frac{m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n, A_n^c)/M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)g^n(\underline{\boldsymbol{X}}_n)}{m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n)/g^n(\underline{\boldsymbol{X}}_n)}, \tag{1.6}$$

since $g^n(\underline{\boldsymbol{X}}_n)$ is non-zero for each n with G^∞ probability one. Let us first consider the numerator in (1.6) and define E_n to be the event that the numerator is greater than $(b_n + c_n)/\delta$. Note that, $G^n(E_n) \leq G^n(E_n \cap S_n^c) + G^n(S_n)$ for any sequence of measurable sets $S_n \in \mathcal{B}_n$. So, taking S_n to be the critical sets of Assumption (A3), we get

$$\begin{aligned}
& G^n(E_n \cap S_n^c) = \int_{E \cap S_n^c} G^n(d\underline{\boldsymbol{x}}_n) \\
\leq\; & \frac{\delta}{(b_n + c_n)} \int_{S_n^c} \frac{m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n^c)}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)g^n(\underline{\boldsymbol{x}}_n)} G^n(d\underline{\boldsymbol{x}}_n) \quad \text{[by Markov's inequality and definition of } E_n\text{]} \\
=\; & \frac{\delta}{(b_n + c_n)M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \int_{S_n^c} m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n^c) d\underline{\boldsymbol{x}}_n \\
=\; & \frac{\delta}{(b_n + c_n)M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \int_{S_n^c} \int_{A_n^c} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}d\underline{\boldsymbol{x}}_n \\
=\; & \frac{\delta}{(b_n + c_n)M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \int_{A_n^c} Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta} \qquad \text{[by Fubini Theorem]} \\
\leq\; & \frac{\delta}{(b_n + c_n)M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \left[\int_{B_n} Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta} + \int_{C_n} Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}\right] \\
& \qquad \text{[for the sets } B_n \text{ and } C_n \text{ from Assumptions (A1)–(A3)]} \\
\leq\; & \frac{\delta}{(b_n + c_n)M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \left[\int_{B_n} Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta} + \int_{C_n} \frac{Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta}\right] \\
\leq\; & \frac{\delta}{(b_n + c_n)M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \left[M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n) + \sup_{\boldsymbol{\theta} \in C_n} \frac{Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} M_n^{(\alpha)}(\boldsymbol{\chi}_n, C_n)\right] \\
\leq\; & \frac{\delta}{(b_n + c_n)} \left[b_n + c_n \frac{M_n^{(\alpha)}(\boldsymbol{\chi}_n, C_n)}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)}\right] \qquad \text{[by Assumptions (A2) and (A3)]} \\
\leq\; & \delta.
\end{aligned}$$

Hence, $G^n(E_n) \le \delta + G^n(S_n)$ and using Assumption (A3) we get $\limsup_{n\to\infty} G^n(E_n) \le \delta$. Further, by Assumption (A4) the denominator in (1.6) is less than a_n has probability tending to zero. Combining the numerator and denominator probabilities (using the bound by the union of events related to numerator and denominator), we get the desired result (1.4).

To prove the second part (1.5), we proceed as before by noting that $P(\underline{\boldsymbol{X}}_n \in E_n \text{ i.o.}) \le P(\underline{\boldsymbol{X}}_n \in E_n \cap S_n^c \text{ i.o.}) + P(\underline{\boldsymbol{X}}_n \in S_n \text{ i.o.})$. Then, defining E_n with any summable sequence δ_n and proceeding as before, we get $P(\underline{\boldsymbol{X}}_n \in E_n \cap S_n^c \text{ i.o.}) = 0$ by Borel-Cantelli Lemma. Next, by Assumption (A3)*, we have $P(\underline{\boldsymbol{X}}_n \in S_n \text{ i.o.}) = 0$ and hence $P(\underline{\boldsymbol{X}}_n \in E_n \text{ i.o.}) = 0$. Then, the desired result (1.5) follows by noting that the denominator in (1.6) is less than a_n infinitely often with probability zero by Assumption (A4)*. □

Lemma 1.2 *Suppose, for some sequence of constants r_n, we have*

$$\lim_{n\to\infty} P\left(\pi_n^{(\alpha)}\left(A_n^c|\underline{\boldsymbol{X}}_n\right) \le r_n\right) = 1. \tag{1.7}$$

Then, for any sequences b_n and c_n satisfying $b_n c_n \ge r_n$, there exists parameter sets $B_n, C_n \subset \Theta_n$ such that Conditions (A1)–(A3) hold. Moreover, if additionally we have

$$P\left(\pi_n^{(\alpha)}\left(A_n^c|\underline{\boldsymbol{X}}_n\right) > r_n \ i.o.\right) = 0, \tag{1.8}$$

then Conditions (A1), (A2) and (A3)* hold.

Proof: Let us define $S_n = \left\{\underline{\boldsymbol{x}}_n : \pi_n^{(\alpha)}\left(A_n^c|\underline{\boldsymbol{x}}_n\right) > r_n\right\}$ so that $\lim_{n\to\infty} G^n(S_n) = 0$ by Assumption

(1.7). Next, for any sequence c_n, we construct the parameter sets

$$C_n = \left\{ \boldsymbol{\theta} : \frac{Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} \le c_n \right\}, \quad B_n = \left\{ \boldsymbol{\theta} \in A_n^c : \frac{Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} > c_n \right\}.$$

Then, Conditions (A1) and (A3) hold by constructions of C_n and B_n. Finally, to show Condition (A2), note that $m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n^c) \le r_n M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n) m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)$ for all $\underline{\boldsymbol{x}}_n \in S_n^c$ by its definition. Then,

$$\begin{aligned}
\frac{M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n)}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} &= \frac{1}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \int_{B_n} Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&\le \frac{1}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)c_n} \int_{A_n^c} Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})\pi_n(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&\qquad \text{[by Definition of } B_n \text{ and Markov's inequality]} \\
&\le \frac{1}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)c_n} \int_{A_n^c} \int_{S_n} \exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))d\underline{\boldsymbol{x}}_n \pi_n(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&\le \frac{1}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)c_n} \int_{S_n} m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n, A_n^c)d\underline{\boldsymbol{x}}_n \qquad \text{[by Fubini Theorem]} \\
&\le \frac{1}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)c_n} \int_{S_n} r_n M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n) m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)d\underline{\boldsymbol{x}}_n \quad \text{[by the construction of } S_n\text{]} \\
&\le \frac{r_n}{c_n} \int_{S_n} m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)d\underline{\boldsymbol{x}}_n \\
&\le \frac{r_n}{c_n} \qquad \left[\textstyle\int_{S_n} m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)d\underline{\boldsymbol{x}}_n \le \int_{\boldsymbol{\chi}_n} m_n^{(\alpha)}(\underline{\boldsymbol{x}}_n)d\underline{\boldsymbol{x}}_n = 1\right] \\
&\le b_n \qquad \text{[for any sequence } b_n \text{ satisfying } b_n c_n \ge r_n\text{]}
\end{aligned}$$

For the second part of the Lemma , we use the same definitions of sets as above. Then, by Assumption (1.8), we have $P(\underline{\boldsymbol{X}}_n \in S_n \ i.o.) = 0$ and hence Condition (A3)* holds by the construction of C_n. Other two conditions then hold similarly as before. $\square$

Proof of Theorem 3:

Theorem 3 now follows directly from the above two lemmas.

The sufficiency part of the theorem follows from Lemma 1.1 by taking $b_n = e^{-nr_1}$, $c_n = e^{-nr_2}$, $a_n = e^{-n\epsilon}$ and $\delta_n = e^{-n\Delta}$ (for Part 2) with $\epsilon, \Delta > 0$ and $\epsilon + \Delta < \min\{r_1, r_2\}$ Then, r_n and $r_n' = r_n/\delta_n$ tend to zero exponentially fast.

The Necessity part of the theorem follows from Lemma 1.2 with $r_n = e^{-nr}$ and then letting $b_n = e^{-nr_1}$, $c_n = e^{-nr_2}$ for any $r_1, r_2 > 0$ with $r_1 + r_2 \leq r$. $\square$

2 Proofs of the Results of Section 4 in Main paper

2.1 Proof of Theorem 7

Note that, by the definition of $\widehat{\boldsymbol{\theta}}_\alpha$, it is sufficient to show that

$$\sup_{\boldsymbol{\theta} \in A_n^c} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) \widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}) < \sup_{\boldsymbol{\theta}} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) \widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}) e^{-n\delta_n} \quad a.s.[G], \qquad \text{for all large } n. \tag{2.1}$$

Now, by the information denseness assumption, Theorem 2 of the main paper implies that G^n and $M_n^{(\alpha)}$ merge in probability. Therefore, the exponential convergence of $\pi_n^{(\alpha)}(A_n^c|\underline{\boldsymbol{X}}_n)$ is equivalent to

$$\begin{aligned} \sum_{\boldsymbol{\theta} \in A_n^c} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}) \widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}) &\leq m_n^{(\alpha)}(\underline{\boldsymbol{X}}_n) e^{-nr_1} \\ &< g_n(\underline{\boldsymbol{X}}_n) e^{-nr} \quad a.s.[G], \\ &\qquad \text{for all large } n, \ \text{for some } r_1, r > 0. \end{aligned}$$

Let us now choose a $\boldsymbol{\theta}^* \in \Theta$ such that $KLD(g, \widetilde{q}^{(\alpha)}(\cdot|\boldsymbol{\theta}^*)) < r/4$. Then, using SLLN along with Assumption (4.2) of the main paper, we get

$$g_n(\underline{\boldsymbol{X}}_n) < \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}^*)\widetilde{q}^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}^*)e^{nr/2} \quad a.s.[G], \qquad \text{for all large } n.$$

Therefore, for all large n, we have with $a.s.[G]$,

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in A_n^c} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}) &\leq \sum_{\boldsymbol{\theta}\in A_n^c} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}) \\
&< g_n(\underline{\boldsymbol{X}}_n)e^{-nr} \\
&< \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta}^*)\widetilde{q}^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta}^*)e^{-nr/2} \\
&< \sup_{\boldsymbol{\theta}} \widetilde{\pi}_n^{(\alpha)}(\boldsymbol{\theta})\widetilde{q}_n^{(\alpha)}(\underline{\boldsymbol{X}}_n|\boldsymbol{\theta})e^{-n\delta_n}.
\end{aligned}$$

This completes the proof that $\widehat{\boldsymbol{\theta}}_\alpha \in A_n$ $a.s.[G]$, for all sufficiently large n. □

2.2 Proof of Theorem 8

Using the equivalence of d_1 and d_H (the Hellinger metric), it is enough to show that $\pi_n^{(\alpha)}(A^c|\underline{\boldsymbol{X}}_n)$ is exponentially small with probability one, with $A = \{\boldsymbol{\theta} : d_H(g, f_{\boldsymbol{\theta}}) \geq \epsilon\}$ for each fixed $\epsilon > 0$. Note that, G^n and $M_n^{(\alpha)}$ merge in probability by applying Theorem 2 of the main paper. So, we will use Theorem 3 by constructing suitable parameter sets B_n and C_n with $A \cup B_n \cup C_n = \Theta$.

Put $B_n = \left\{\boldsymbol{\theta} : \pi_n(\boldsymbol{\theta}) < e^{-n\epsilon/4}\right\}$ and $C_n = \left\{\boldsymbol{\theta} \in A^c : \pi_n(\boldsymbol{\theta}) \geq e^{-n\epsilon/4}\right\}$.

Then, clearly $A \cup B_n \cup C_n = \Theta$. Further, for some $\tau \in (0,1)$,

$$\frac{M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n)}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} = \sum_{\boldsymbol{\theta} \in B_n} \frac{Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta})}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \pi_n(\boldsymbol{\theta}) \leq \frac{e^{-\frac{n(1-\tau)\epsilon}{4}}}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \sum_{\boldsymbol{\theta} \in B_n} Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) \pi_n(\boldsymbol{\theta})^\tau$$

But, since the prior sequence π_n satisfies Assumption (4.2) of the main paper, we get, for all sufficiently large n, (assuming all the relevant quantities exists finitely)

$$\begin{aligned} M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n) &= \sum_{\boldsymbol{\theta} \in \Theta} Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) \pi_n(\boldsymbol{\theta}) \\ &\geq e^{-n(1-\tau)\epsilon/8} \sum_{\boldsymbol{\theta} \in \Theta} Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) \pi_n(\boldsymbol{\theta})^\tau \\ &\geq e^{-n(1-\tau)\epsilon/8} \sum_{\boldsymbol{\theta} \in B_n} Q_n^{(\alpha)}(\chi_n|\boldsymbol{\theta}) \pi_n(\boldsymbol{\theta})^\tau, \end{aligned}$$

and hence

$$\frac{M_n^{(\alpha)}(\boldsymbol{\chi}_n, B_n)}{M_n^{(\alpha)}(\boldsymbol{\chi}_n, \Theta_n)} \leq e^{-n(1-\tau)\epsilon/8}.$$

Thus, the first two conditions of Theorem 3 hold. For the third condition related to C_n, note that $\sum_{\boldsymbol{\theta} \in C_n} \pi_n(\boldsymbol{\theta}) \leq 1$ and so the number of points in C_n is less than $e^{n\epsilon/4}$. Then, consider the likelihood ratio test for g_n against $\left\{ \frac{\exp(q_n^{(\alpha)}(\cdot|\boldsymbol{\theta}))}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} : \boldsymbol{\theta} \in C_n \right\}$ having the critical sets

$$S_n = \left\{ \underline{\boldsymbol{x}}_n : \max_{\boldsymbol{\theta} \in C_n} \frac{\exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} > g_n(\underline{\boldsymbol{x}}_n) \right\}.$$

We will show that this S_n serves as the desired set in the required condition (A3) on C_n. For note that, $S_n = \cup_{\boldsymbol{\theta} \in C_n} S_{n,\boldsymbol{\theta}}$, where $S_{n,\boldsymbol{\theta}} = \left\{ \underline{\boldsymbol{x}}_n : \left[\frac{\exp(q_n^{(\alpha)}(\underline{\boldsymbol{x}}_n|\boldsymbol{\theta}))}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} \right]^{1/2} > g_n(\underline{\boldsymbol{x}}_n)^{1/2} \right\}$. But for each of these sets, we get from Markov inequality that,

$$G_n(S_{n,\boldsymbol{\theta}}) \leq \left[1 - \frac{1}{2}d_H(g, f_{\boldsymbol{\theta}})\right]^n < e^{-n\epsilon/2},$$

and hence $G_n(S_n) < e^{-n\epsilon/4}$. Similarly, we can also show that

$$\frac{Q_n^{(\alpha)}(S_n^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} \leq \frac{Q_n^{(\alpha)}(S_{n,\boldsymbol{\theta}}^c|\boldsymbol{\theta})}{Q_n^{(\alpha)}(\boldsymbol{\chi}_n|\boldsymbol{\theta})} < e^{-n\epsilon/2},$$

uniformly over $\boldsymbol{\theta} \in C_n$. Hence, all the required conditions of Theorem 3 hold and we get the first part of the present theorem.

The second part then follows directly from Theorem 7 of the main paper. □

3 Proofs of the Results of Section 5 in Main paper

3.1 Proof of Theorem 10

By straightforward calculation, it turns out that $d_1(g_i, f_{i,\boldsymbol{\beta}}) = 4\Phi\left(\frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|}{2}\right) - 2$, where Φ is the cumulative distribution function of the standard Normal distribution. Therefore,

$$\begin{aligned}
&\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^n d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\right\}\middle|\underline{\boldsymbol{x}}_n\right) \\
&\quad = \pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^n \Phi\Big(\frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|}{2}\Big) \geq \frac{\epsilon}{4} + \frac{1}{2}\right\}\middle|\underline{\boldsymbol{x}}_n\right) \\
&\quad \leq \sum_{i=1}^n \pi_n^{(\alpha)}\Big(\{\boldsymbol{\beta} : \Phi\Big(\frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|}{2}\Big) \geq \frac{\epsilon}{4} + \frac{1}{2}\}|\underline{\boldsymbol{x}}_n\Big) \\
&\quad = \sum_{i=1}^n \pi_n^{(\alpha)}\Big(\{\boldsymbol{\beta} : \frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|}{2} \geq \Phi^{-1}(\frac{1}{2} + \frac{\epsilon}{4})\}|\underline{\boldsymbol{x}}_n\Big).
\end{aligned}$$

For notational simplicity, let us denote the set $A_{i,n}^c = \left\{\boldsymbol{\beta} : \frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|}{2} \geq \Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})\right\}$, for each $i = 1,\ldots,n$ and let $\widehat{\boldsymbol{\beta}}_n$ be the minimum DPD estimator (MDPDE) of $\boldsymbol{\beta}$ under the same model. After some basic algebra, using the consistency of the MDPDE under (R1)–(R2) (Ghosh and Basu, 2013), it can be shown that the event $A_{i,n}$ implies

$$-2\Phi^{-1}\left(\frac{1}{2}+\frac{\epsilon}{4}\right) - \boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \leq \boldsymbol{t}_i^T(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) \leq 2\Phi^{-1}\left(\frac{1}{2}+\frac{\epsilon}{4}\right) - \boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0).$$

Now, we use Theorem 2.1 of Majumder et al. (2019) to approximate the above probability as follows.

$$\begin{aligned}
\pi_n^{(\alpha)}(A_{i,n}^c|\underline{\boldsymbol{x}}_n) &= 1 - \pi_n^{(\alpha)}(A_{i,n}|\underline{\boldsymbol{x}}_n) \\
&= 1 - \left[\Phi\left(\frac{2\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4}) - \sqrt{n}\boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}\right)\right. \\
&\qquad \left. - \Phi\left(\frac{-2\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4}) - \sqrt{n}\boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}\right)\right] + o_p(1) \\
&= \Phi(u_n) + \Phi(v_n) + o_p(1),
\end{aligned}$$

where $\boldsymbol{\Psi}_n^{-1} = n[\boldsymbol{D}^T\boldsymbol{D}]$, $u_n = \frac{-2\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})+\sqrt{n}\boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}$ and $v_n = \frac{-2\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})-\sqrt{n}\boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}$.

Now, using the consistency of the MDPDE under (R1)–(R2) (Ghosh and Basu, 2013), along with (R1), we get $\boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{p} 0$, for all $i = 1,...,n$. So,

$$\lim_{n\to\infty} P\left(\left|\boldsymbol{t}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\right| < \Phi^{-1}\left(\frac{1}{2}+\frac{\epsilon}{4}\right)\right) = 1.$$

Further, by (R2), $\max\limits_{1\leq i\leq n} \boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i = \max\limits_{1\leq i\leq n} n\boldsymbol{t}_i^T[\boldsymbol{D}^T\boldsymbol{D}]^{-1}\boldsymbol{t}_i = O(1)$ and hence we get

$$P\left(u_n < \frac{-\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}\right) \to 1$$

and

$$P\left(v_n < \frac{-\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}\right) \to 1,$$

as $n \to \infty$. Since Φ is continuous, in turn, we have

$$P\left(\Phi(u_n)+\Phi(v_n) < 2\Phi\left(\frac{-\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}\right)\right) \to 1, \qquad \text{as } n \to \infty.$$

Let us now denote $l_n = \frac{\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})}{\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}$. Note that $l_n > 0$ for all n, and $l_n \to \infty$ as $n \to \infty$. So, $\Phi(-l_n) \sim \dfrac{\phi(l_n)}{l_n}$, where ϕ is the density of the standard normal distribution. Hence, for large n, $\Phi(-l_n) \leq 2\dfrac{\phi(l_n)}{l_n}$ so that we get, with probability tending to one,

$$\Phi(u_n)+\Phi(v_n) < 4\frac{\phi(l_n)}{l_n} = \frac{4\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i}{\sqrt{2\pi}\sqrt{n}\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})} e^{-\dfrac{n(\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4}))^2}{2(\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i)^2}}.$$

Next, by Assumption (R2), there exists a constant $C_0 > 0$ satisfying $\boldsymbol{t}_i^T\boldsymbol{\Psi}_n^{-1}\boldsymbol{t}_i \leq C_0$. Thus, with probability tending to one, we have

$$\Phi(u_n)+\Phi(v_n) < \frac{C_1}{\eta\sqrt{n}}e^{-nC_2}$$

for some constants $C_1, C_2 > 0$ and $\eta = \Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})$. Also, since $C_2 > 0$, we get $\dfrac{C_1\sqrt{n}}{\eta} < e^{\frac{nC_2}{2}}$ for all sufficiently large n. Therefore, with probability tending to one, we have

$$\pi_n^{(\alpha)}\Big(\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\}|\underline{\boldsymbol{x}}_n\Big) < \frac{C_1\sqrt{n}}{\eta}e^{-nC_2} < e^{\frac{-nC_2}{2}},$$

and hence the theorem holds with $r = \frac{C_2}{2}$. $\square$

3.2 Proof of Theorem 11

In this case, the parameter is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ and, after some basic algebra, we find that

$$d_1(g_i, f_{i,\boldsymbol{\theta}}) \leq 2\frac{|\sigma - \sigma_0|}{\sigma} + 4\Phi\left(\frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|}{2\sigma_0}\right) - 2.$$

Hence, for any $\epsilon \in (0, 1)$, it follows that

$$\begin{aligned}
& \pi_n^{(\alpha)}\left(\left\{\boldsymbol{\theta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\right\}\middle|\underline{\boldsymbol{x}}_n\right) \\
\leq \; & \pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{4}{n}\sum_{i=1}^{n} \Phi\Big(\frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|}{2\sigma_0}\Big) - 2 \geq \frac{\epsilon}{2}\right\}\middle|\underline{\boldsymbol{x}}_n\right) + \pi_n^{(\alpha)}\left(\left\{\sigma : 2\frac{|\sigma - \sigma_0|}{\sigma} \geq \frac{\epsilon}{2}\right\} |\underline{\boldsymbol{x}}_n\right) \\
\leq \; & \sum_{i=1}^{n} \pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{|\boldsymbol{t}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|}{2\sigma_0} \geq \Phi^{-1}\left(\frac{1}{2} + \frac{\epsilon}{8}\right)\right\}\middle|\underline{\boldsymbol{x}}_n\right) + \pi_n^{(\alpha)}\left(\left\{\sigma : |\sigma - \sigma_0| \geq \frac{\epsilon/4}{1 + \epsilon/4}\right\} \Big|\underline{\boldsymbol{x}}_n\right).
\end{aligned} \tag{3.1}$$

The first term in above (3.1) is exponentially small in probability under (R1)–(R2), as proved in Theorem 10. However, if $\widehat{\sigma}_n^2$ denotes the MDPDE of σ, using its consistency under (R1)–(R2) (Ghosh and Basu, 2013), we have

$$\pi_n^{(\alpha)}\Big(\{\sigma : |\sigma - \sigma_0| \geq \eta\}|\underline{\boldsymbol{x}}_n\Big) \leq \pi_n^{(\alpha)}\Big(\{\sigma : |\sigma - \widehat{\sigma}_n| \geq \eta/2\}|\underline{\boldsymbol{x}}_n\Big) + o_p(1).$$

But $\pi_n^{(\alpha)}\Big(\{\sigma : |\sigma - \widehat{\sigma}_n| \geq \eta/2\}|\underline{\boldsymbol{x}}_n\Big) = \pi_n^{(\alpha)}\Big(\{\sigma : |\sigma^2 - \widehat{\sigma}_n^2| \geq \eta'\}|\underline{\boldsymbol{x}}_n\Big)$ for some $\eta' > 0$. And, it follows from Theorem 2.1 of Majumder et al. (2019) that the posterior distribution of $\sqrt{n}(\sigma^2 - \widehat{\sigma}_n^2)$ is $\mathcal{N}(0, \zeta_\alpha)$, where ζ_α is some function of σ_0 and α. So, combining them, we get the following approximation

$$\pi_n^{(\alpha)}\Big(\{\sigma : |\sigma - \widehat{\sigma}_n| \geq \eta/2\}|\underline{\boldsymbol{x}}_n, \boldsymbol{D}\Big) = 2\Phi\Big(\frac{-\sqrt{n}\eta'}{\zeta_\alpha}\Big) + o_p(1).$$

Note that the right-hand side of the above equation is again exponentially small implying the same for its left-hand side. Thus the last term in (3.1) is also exponentially small, completing the proof of the theorem. □

3.3 Proof of Theorem 12

In the present case of logistic set-up, the L_1 distance between the true density g_i and model density $f_{i,\boldsymbol{\beta}}$ turns out to be

$$d_1(g_i, f_{i,\boldsymbol{\beta}}) = 2|p_i(\boldsymbol{\beta}_0) - p_i(\boldsymbol{\beta})|,$$

where $p_i(\boldsymbol{\beta}) = \dfrac{e^{\boldsymbol{t}_i'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{t}_i'\boldsymbol{\beta}}}$. Recall that $\boldsymbol{\beta}_0$ is the true parameter. Now, applying the mean value theorem on the function $g(t) = \frac{e^t}{1+e^t}$, we get

$$d_1(g_i, f_{i,\boldsymbol{\beta}}) = 2|\boldsymbol{t}_i'\boldsymbol{\beta}_0 - \boldsymbol{t}_i'\boldsymbol{\beta}|\frac{e^{t_i}}{(1 + e^{t_i})^2} \leq 2|\boldsymbol{t}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|.$$

Hence, we get

$$\begin{aligned}
\pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\right\}\Bigg|\underline{\boldsymbol{x}}_n\right) \quad &\leq \quad \pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : \frac{2}{n}\sum_{i=1}^{n}|\boldsymbol{t}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| \geq \epsilon\right\}\Bigg|\underline{\boldsymbol{x}}_n\right) \\
&\leq \quad \sum_{i=1}^{n} \pi_n^{(\alpha)}\left(\left\{\boldsymbol{\beta} : |\boldsymbol{t}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| \geq \frac{\epsilon}{2}\right\}\Big|\underline{\boldsymbol{x}}_n\right).
\end{aligned}$$

Denote the set $\{\boldsymbol{\beta} : |\boldsymbol{t}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| \geq \frac{\epsilon}{2}\} = A_{i,n}^c$ and let $\widehat{\boldsymbol{\beta}}_n$ is the MDPDE of $\boldsymbol{\beta}$. After some basic algebra, it can be shown that the event $A_{i,n}$ implies

$$-\frac{\epsilon}{2} - \boldsymbol{t}_i'(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \leq \boldsymbol{t}_i'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) \leq \frac{\epsilon}{2} - \boldsymbol{t}_i'(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0).$$

Now, by the consistency of the MDPDE under (R1) and (R3) (Majumder et al., 2019), we have $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \xrightarrow{p} 0$, and the supremum of the elements of the vectors $\boldsymbol{t}_i$ are bounded by (R3). Hence, $\max_i \boldsymbol{t}_i'(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{p} 0$, implying $-\epsilon \leq \boldsymbol{t}_i'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) \leq \epsilon$ with probability tending to one. Then, we find that

$$\pi_n^{(\alpha)}\Big(- \epsilon \leq \boldsymbol{t}_i'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) \leq \epsilon | \underline{\boldsymbol{x}}_n \Big) = 2\Phi\left(\frac{\sqrt{n}\epsilon}{\boldsymbol{t}_i'\boldsymbol{\Psi}_n(\boldsymbol{\beta})^{-1}\boldsymbol{t}_i}\right) - 1 + o_p(1).$$

So, with probability tending to one, we get

$$\pi_n^{(\alpha)}\Big(A_{i,n}^c | \underline{\boldsymbol{x}}_n, \boldsymbol{D}\Big) \leq 2\left[1 - \Phi\left(\frac{\sqrt{n}\epsilon}{\boldsymbol{t}_i'\boldsymbol{\Psi}_n(\boldsymbol{\beta})^{-1}\boldsymbol{t}_i}\right)\right] = 2\Phi\left(-\frac{\sqrt{n}\epsilon}{\boldsymbol{t}_i'\boldsymbol{\Psi}_n(\boldsymbol{\beta})^{-1}\boldsymbol{t}_i}\right).$$

Using boundedness of $\boldsymbol{t}_i'\boldsymbol{\Psi}_n(\boldsymbol{\beta})^{-1}\boldsymbol{t}_i$ and the fact that $\Phi(-l_n) \sim \dfrac{\phi(l_n)}{l_n}$ as $n \to \infty$, we can derive finally that

$$\pi_n^{(\alpha)}\Big(A_{i,n}^c | \underline{\boldsymbol{x}}_n, \boldsymbol{D}\Big) \leq \frac{C_0}{\sqrt{n}} e^{-nC_1}$$

for some positive constants C_0 and C_1. Thus

$$\pi_n^{(\alpha)}\Big(\{\boldsymbol{\beta} : \frac{1}{n}\sum_{i=1}^{n} d_1(g_i, f_{i,\boldsymbol{\beta}}) \geq \epsilon\} | \underline{\boldsymbol{x}}_n, \boldsymbol{D}\Big) \leq \sqrt{n} C_0 e^{-nC_1} < e^{-nr}$$

for some $r > 0$, with probability tending to one, proving the theorem.

4 Additional Description of Figures 1, 2 and 3 of the Main Paper

In Figures 1, 2 and 3 of the main paper it may appear that the MSE is practically zero at $\alpha = 1$ (in case of pure data, as well as for each contaminated scenario). If that is so, why should we not set $\alpha = 1$ all the time, rather than going through the exercise of choosing an optimal α? Actually this is a false impression created by the scale of these figures. Under contamination, the inflations in the MSE of the estimators corresponding to low values of α are of such a high magnitude, that in trying to accommodate them within the same frame of the figure, the MSEs of several of the stable estimators corresponding to large values of α (and not just for $\alpha = 1$) appear to be zero, or to be very close to it. The phenomenon can be better explained by looking at blow-ups of these MSE curves by restricting the Y-axis (the MSE axis) to a small range around zero. We provide such a representative figure (Figure 1 below), which corresponds to such a blown up MSE curve for the estimation of σ with a sample size of $n = 100$ in the linear regression model with unknown σ and the conjugate priors. Notice that under pure data, the MSE curve in Figure 1 is steadily increasing, indicating that under the model the performance progressively deteriorates with increasing α. With increasing contamination the optimal value of α (the value which minimizes the MSE) keeps getting shifted upward; the optimal values of α for $\epsilon_C = 0.05, 0.1$ and 0.2 are, approximately, 0.35, 0.5 and 1. In this example, therefore, it is clear that the choice of the optimal α is very much a function of the amount of the anomaly in the data, and the blanket selection of $\alpha = 1$ does not necessarily provide the best solution in all cases. A tuning parameter selection does remain an important component of our methodology as described in Section 7.2 of the main paper.

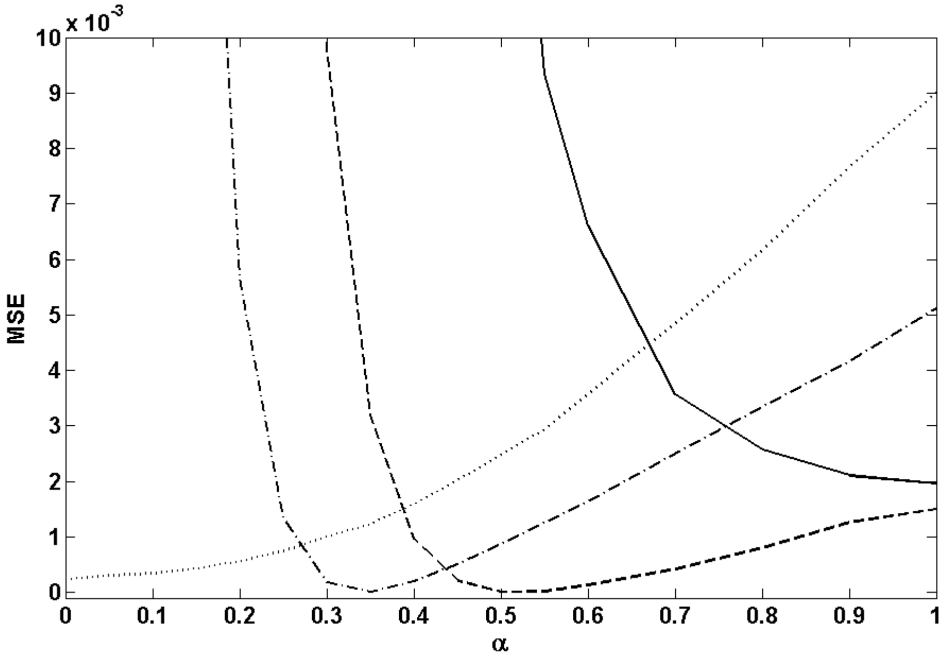
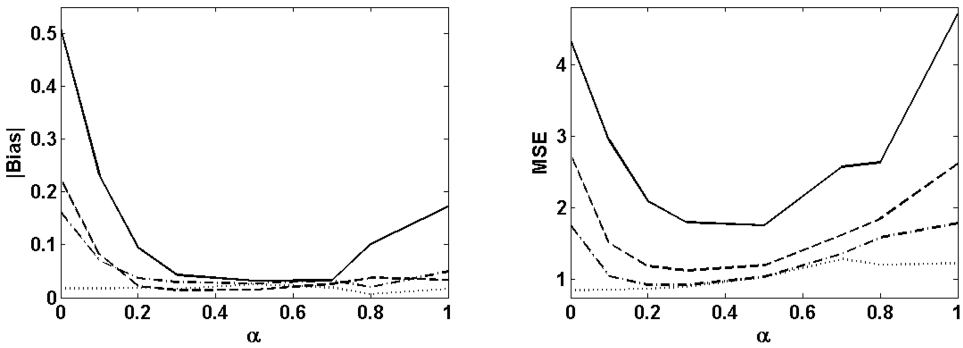


Figure 1: Empirical MSE of the ERPE of σ for $n = 100$ in the linear regression model with unknown σ and the conjugate priors. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$] (rescaled version of Fig. 2(c), lower panel, of the main paper)

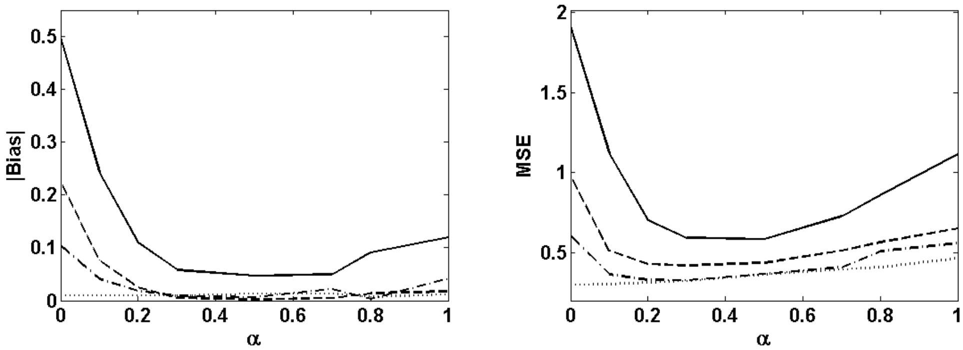
Similar phenomena occur for the graphs of MSEs under all the cases reported in Figures 1, 2 and 3 of the main paper; so we do not repeat them for brevity.

5 Additional Simulation results for Normal Linear Regression Model with fixed σ

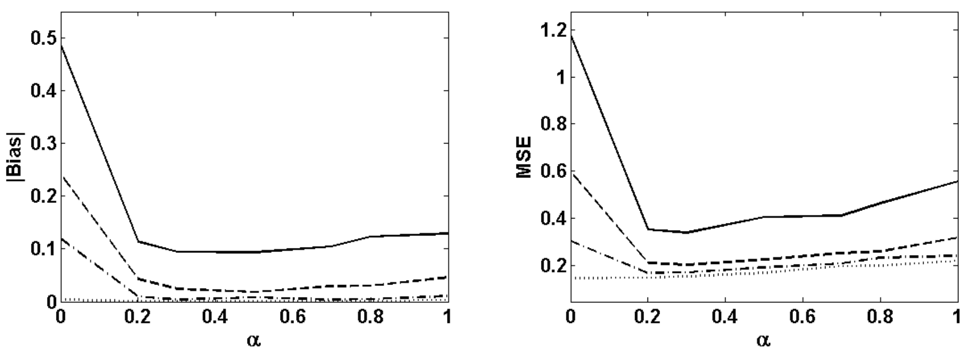
Recall the simulation set-up and notation as in Section 6.1 of the main paper to illustrate the performance of the ERPE for fixed-design linear regression model with known error variance σ and two different choices for the prior. As the first choice of the prior $\pi(\boldsymbol{\beta})$, we consider the non-informative uniform prior $\pi(\boldsymbol{\beta}) \equiv 1$. Secondly, we consider the conjugate normal prior $\pi(\boldsymbol{\beta}) \equiv N_k(\boldsymbol{\beta}_0, \tau^2 I_k)$ which signifies that the prior belief about our true parameter value is quantified by a symmetric structure with uncertainty quantified by τ. The resulting values of the total absolute bias and the total MSE (over the two components of $\boldsymbol{\beta}$) are shown in Figure 2 and 3, respectively. As noted in the main paper, these results clearly demonstrate the significant improvement for the ERPE over the usual Bayes estimators under data contamination with only a slight loss in case of pure data.



(a) $n = 20$

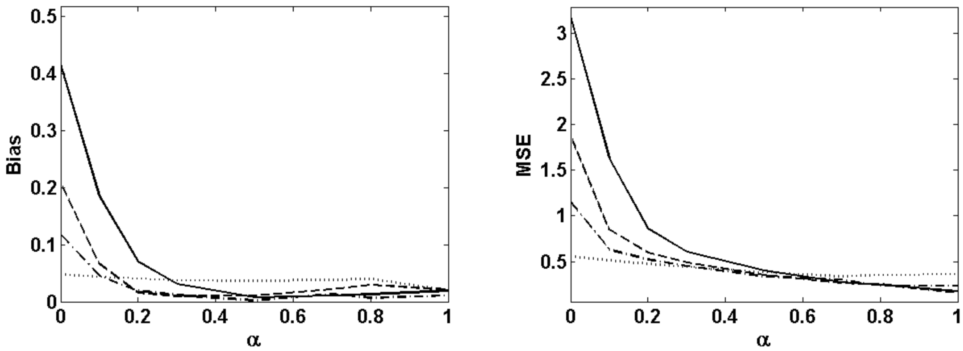


(b) $n = 50$

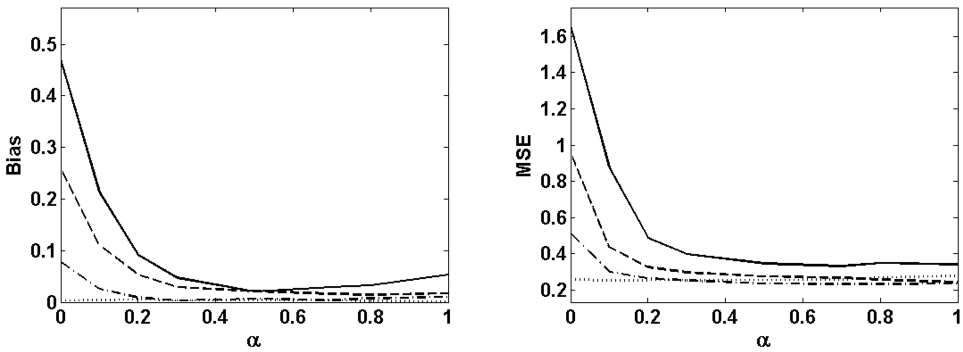


(c) $n = 100$

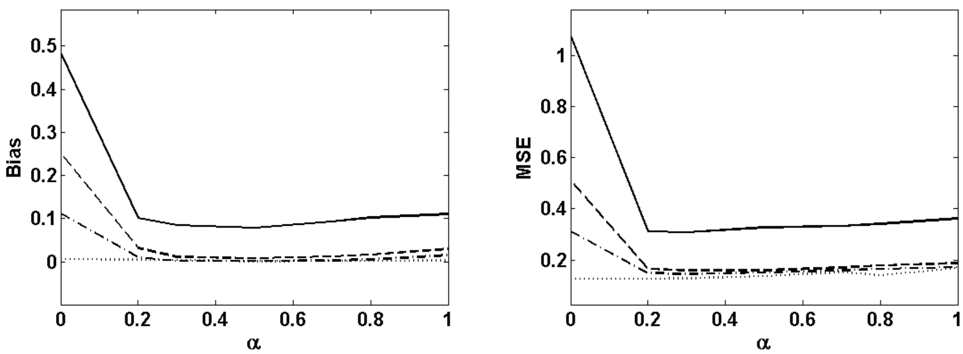
Figure 2: Empirical total absolute bias and total MSE of the ERPE of $\boldsymbol{\beta}$ in the linear regression model with known $\sigma = 1$ and the uniform prior. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$]



(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 linear regression model with known $\sigma = 1$ and the Conjugate normal prior. [Dotted line: $\epsilon_C = 0\%$, Dash-Dotted line: $\epsilon_C = 5\%$, Dashed line: $\epsilon_C = 10\%$, Solid line: $\epsilon_C = 20\%$]

6 R Codes for Computation of the ERPEs

6.1 Fixed-Design Linear Regression with Unknown Error Variance and Jefferey's Prior

```
library("MASS")
library("parallel")

## Computation of ERPE for Fixed-design Linear Regression Model
## with unknonw error variance and
## Non informative Jefferey's Prior pi(\beta, \sigma) = 1/sigma^2


####### Auxiliary Functions  --------------
## Defining the function Q ##
Q=function(beta,sigma,X,y,alpha)
{
  n=nrow(X)
  p=ncol(X)
  M=X%*%beta
  L=matrix(0,n,1,byrow=TRUE)
  for(i in 1:n)
  {
    L[i]=(1/alpha)*(dnorm(y[i],M[i],sigma))^(alpha)
             -1/(((1+alpha)^(3/2))*((sqrt(2*pi)*sigma)^(alpha)))- (1/alpha)
  }
  L_0=matrix(0,n,1,byrow=TRUE)
  for(i in 1:n)
  {
    L_0[i]=log(dnorm(y[i],M[i],sigma))
  }
  if(alpha>0)
  {
    return(sum(L))
  }
  else
  {
    return(sum(L_0))
  }
```

```
}

## Kernel of the prior density ##
prior=function(beta,sigma,true_beta)
{
  return (1/sigma^2)
}

## Kernel of the log of robust posterior density ##
robust_posterior=function(beta,sigma,X,y,alpha,true_beta)
{
  z1=Q(beta,sigma,X,y,alpha)
  z2=log(prior(beta,sigma,true_beta))
  z=z1+z2
  return (z)
}

## Proposal Density ##
proposalfunction=function(beta,sigma)
{
  p=length(beta)
  return (c(rnorm(p,mean=beta,sd=rep(1,p)),rexp(1,1/sigma)))
}

## Metropolis Hastings ##
run_metropolis_MCMC <- function(startvalue_beta,startvalue_sigma,
                                  iterations,X,y,alpha,true_beta)
{
  p=length(startvalue_beta)
  chain_beta = array(dim = c(iterations+1,p))
  chain_sigma = array(dim = c(iterations+1,1))
  chain_beta[1,] = startvalue_beta
  chain_sigma[1]=startvalue_sigma
  for (i in 1:iterations){
    proposal = proposalfunction(chain_beta[i,],chain_sigma[i])
    proposal_beta=proposal[1:p]
    proposal_sigma=proposal[p+1]
    probab1
     = exp(robust_posterior(proposal_beta,proposal_sigma,X,y,alpha,true_beta)
```

```
        -robust_posterior(chain_beta[i,],chain_sigma[i],X,y,alpha,true_beta))
    logprobab2=log(dexp(chain_sigma[i],(1/proposal_sigma)))
                -log(dexp(proposal_sigma,(1/chain_sigma[i])))
    probab2=exp(logprobab2)
    probab=probab1*probab2
    if (runif(1) < probab){
      chain_beta[i+1,] = proposal_beta
      chain_sigma[i+1] = proposal_sigma
    }else{
      chain_beta[i+1,] = chain_beta[i,]
      chain_sigma[i+1,] = chain_sigma[i,]
    }
  }
  chain=cbind(chain_beta,chain_sigma)
}


####### Computation of the ERPE --------------
data=read.table(file="data.txt")   ## Call the data-file  ##
y=data[,2]       ## Response ##
x=data[,-2]       ## Covariate ##
X=cbind(1,x)     ## Design matrix ##
n=nrow(X)        ## Number of observations ##
p=ncol(X)        ## Dimension of regression parameters excluding sigma ##

## Fit linear regression ##
res=lm(y~x)
summary(res)
beta_0=res$coefficients    ## MLE ##

burn=25000       ## burn in ##
max_iter=50000   ## Tortal number of iterations including burn in ##
alpha = 0.5      ## tuning parameter alpha in alpha-likelihood
#set.seed(12345)
output=run_metropolis_MCMC(startvalue_beta=c(-7,3),
                           startvalue_sigma=1,iterations = max_iter,
                           X=X,y=y,alpha=alpha,true_beta=beta_0)
output2=output^2
```

```
#Estimated ERPE, the means of the the R^\alpha-posterior distribution
est_mean=colMeans(output[(burn+1):max_iter+1,])
#Estimated variance of the R^\alpha-posterior distribution
est_var=colMeans(output2[(burn+1):max_iter+1,])-(est_mean)^2
```

6.2 Fixed-Design Linear Regression with Unknown Error Variance and Conjugate Prior

```
library("MASS")
library("parallel")

## Computation of ERPE for Fixed-design Linear Regression Model
## with unknonw error variance and
## Conjugate normal-Inverse Gamma prior with prior mean for beta is given ##


####### Auxiliary Functions  --------------
## Defining the function Q ##
Q=function(beta,sigma,X,y,alpha)
{
  n=nrow(X)
  p=ncol(X)
  M=X%*%beta
  L=matrix(0,n,1,byrow=TRUE)
  for(i in 1:n)
  {
    L[i]=(1/alpha)*(dnorm(y[i],M[i],sigma))^(alpha)
             -1/(((1+alpha)^(3/2))*((sqrt(2*pi)*sigma)^(alpha)))- (1/alpha)
  }
  L_0=matrix(0,n,1,byrow=TRUE)
  for(i in 1:n)
  {
    L_0[i]=log(dnorm(y[i],M[i],sigma))
  }
  if(alpha>0)
  {
    return(sum(L))
  }
```

```
  else
  {
    return(sum(L_0))
  }
}

## Kernel of the prior density ##
prior=function(beta,sigma,true_beta)
{
  z1=prod(dnorm(beta,true_beta,sigma))
  z2=sigma^(-5)*exp(-0.5/(sigma^2))
  return (z1*z2)
}

## Kernel of the log of robust posterior density ##
robust_posterior=function(beta,sigma,X,y,alpha,true_beta)
{
  z1=Q(beta,sigma,X,y,alpha)
  z2=log(prior(beta,sigma,true_beta))
  z=z1+z2
  return (z)
}

## Proposal Density ##
proposalfunction=function(beta,sigma)
{
  p=length(beta)
  return (c(rnorm(p,mean=beta,sd=rep(1,p)),rexp(1,1/sigma)))
}

## Metropolis Hastings ##
run_metropolis_MCMC <- function(startvalue_beta,startvalue_sigma,
                                iterations,X,y,alpha,true_beta)
{
  p=length(startvalue_beta)
  chain_beta = array(dim = c(iterations+1,p))
  chain_sigma = array(dim = c(iterations+1,1))
  chain_beta[1,] = startvalue_beta
  chain_sigma[1]=startvalue_sigma
```

```
  for (i in 1:iterations){
    proposal = proposalfunction(chain_beta[i,],chain_sigma[i])
    proposal_beta=proposal[1:p]
    proposal_sigma=proposal[p+1]
    probab1
     = exp(robust_posterior(proposal_beta,proposal_sigma,X,y,alpha,true_beta)
       -robust_posterior(chain_beta[i,],chain_sigma[i],X,y,alpha,true_beta))
    logprobab2=log(dexp(chain_sigma[i],(1/proposal_sigma)))
                 -log(dexp(proposal_sigma,(1/chain_sigma[i])))
    probab2=exp(logprobab2)
    probab=probab1*probab2
    if (runif(1) < probab){
      chain_beta[i+1,] = proposal_beta
      chain_sigma[i+1] = proposal_sigma
    }else{
      chain_beta[i+1,] = chain_beta[i,]
      chain_sigma[i+1,] = chain_sigma[i,]
    }
  }
  chain=cbind(chain_beta,chain_sigma)
}


####### Computation of the ERPE --------------
data=read.table(file="data.txt")   ## Call the data-file  ##
y=data[,2]       ## Response ##
x=data[,-2]       ## Covariate ##
X=cbind(1,x)     ## Design matrix ##
n=nrow(X)        ## Number of observations ##
p=ncol(X)        ## Dimension of regression parameters excluding sigma ##

beta_0=c(-8.03,2.95)  ## Given value for the mean of the normal prior ##
burn=25000
max_iter=50000
alpha = 0.5      ## tuning parameter alpha in alpha-likelihood
#set.seed(12345)
output=run_metropolis_MCMC(startvalue_beta=beta_0,
                           startvalue_sigma=1,iterations = max_iter,
                           X=X,y=y,alpha=alpha,true_beta=beta_0)
```

```
output2=output^2
#Estimated ERPE, the means of the the R^\alpha-posterior distribution
est_mean=colMeans(output[(burn+1):max_iter+1,])
#Estimated variance of the R^\alpha-posterior distribution
est_var=colMeans(output2[(burn+1):max_iter+1,])-(est_mean)^2
```

6.3 Fixed-Design Logistic Regression with Normal Prior for Regression Coefficient

```
library("MASS")
library("mvtnorm")
library("PACBO")

## Computation of ERPE for Fixed-design Logistic Regression Model
## with normal prior for regression coefficient beta
## with prior mean  given ##


####### Auxiliary Functions  ---------------
## Logistic mass function ##
dlogistic=function(t,y)
{
  z1=exp(t*y)/(1+exp(t))
  return (z1)
}

## Defining the function Q ##
Q=function(beta,X,y,alpha)
{
  n=nrow(X)
  p=ncol(X)
  M=X%*%beta
  L=matrix(0,n,1,byrow=TRUE)
  for(i in 1:n)
  {
    L[i]=(1/alpha)*(dlogistic(M[i],y[i]))^(alpha)
              -(1/(1+alpha))*((dlogistic(M[i],0))^(1+alpha)
```

```
                                  +(dlogistic(M[i],1))^(1+alpha))-(1/alpha)
  }
  L_0=matrix(0,n,1,byrow=TRUE)
  for(i in 1:n)
  {
    L_0[i]=log(dlogistic(M[i],y[i]))
  }
  if(alpha>0)
  {
    return(sum(L))
  }
  else
  {
    return(sum(L_0))
  }
}

## Normal Prior with mean true_beta ##
prior=function(beta,true_beta)
{
  z=prod(dnorm(beta,true_beta,1))
  return (z)
}

## Kernel of the log of robust posterior density ##
robust_posterior=function(beta,X,y,alpha,true_beta)
{
  z1=Q(beta,X,y,alpha)
  z2=log(prior(beta,true_beta))
  z=z1+z2
  return (z)
}

## Proposal Density ##
proposalfunction=function(beta)
{
  p=length(beta)
  return (rnorm(p,mean=beta,sd=rep(1,p)))
}
```

```
## Metropolis Hastings ##
run_metropolis_MCMC <- function(startvalue,iterations,
                                X,y,alpha,true_beta)
{
  p=length(startvalue)
  chain = array(dim = c(iterations+1,p))
  chain[1,] = startvalue
  for (i in 1:iterations){
    proposal = proposalfunction(chain[i,])

    probab = exp(robust_posterior(proposal,X,y,alpha,true_beta)
                 -robust_posterior(chain[i,],X,y,alpha,true_beta))
    if (runif(1) < probab){
      chain[i+1,] = proposal
    }else{
      chain[i+1,] = chain[i,]
    }

  }
  return(chain)
}


####### Computation of the ERPE --------------
data=read.table(file="data.txt")   ## Call the data-file  ##
y=data[,2]       ## Response ##
x=data[,-2]       ## Covariate ##
X=cbind(1,x)     ## Design matrix ##
n=nrow(X)        ## Number of observations ##
p=ncol(X)        ## Dimension of regression parameters excluding sigma ##

beta_0=c(-23,39.4,31.8)    ## Given value for the mean of the normal prior ##

burn=25000
max_iter=50000
alpha = 0.5      ## tuning parameter alpha in alpha-likelihood
#set.seed(12345)
output=run_metropolis_MCMC(startvalue=beta_0,iterations = max_iter,
```

```
                                                X=X,y=y,alpha=alpha,true_beta=beta_0)
output2=output^2
#Estimated ERPE, the means of the the R^\alpha-posterior distribution
est_mean=colMeans(output[(burn+1):max_iter+1,])
#Estimated variance of the R^\alpha-posterior distribution
est_var=colMeans(output2[(burn+1):max_iter+1,])-(est_mean)^2
```

References

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

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

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].