Statistical guarantees for data-driven posterior tempering

Ruchira Ray, Marco Avella Medina, and Cynthia Rush

Abstract

α-posteriors reduce the influence of the likelihood in the calculation of the posterior by raising the likelihood to a constant fractional power, α. This procedure, also known as posterior tempering, has been shown to exhibit appealing properties, including robustness to model misspecification and asymptotic normality (Bernstein-von Mises). However, practical recommendations for selecting α and statistical guarantees for data-driven selection of α remain open questions. We engage with these issues by connecting posterior tempering to penalized estimation. Data-driven approaches to tuning α in these settings suggest a novel asymptotic regime where the distributuon of the selected α is a mixture with a point mass at $\alpha = \infty$ and with remaining mass converging to 0. We formalize the limiting distribution of the α-posterior in this new regime. Furthermore, in the regime where α converges to 0, we provide sufficient conditions for (i) asymptotic normality of the α-posterior, (ii) consistensy of the α-posterior moments, and (iii) $\sqrt{n}$-consistency of the α-posterior mean estimator. Our results hold for α that can depend on the data in an arbitrary way.

1 Introduction

α-posteriors, or power posteriors, are proportional to the prior density multiplied by the likelihood raised to a fractional exponent, α ([21, Chapter 8.6]). This procedure, also known as posterior tempering, distorts standard Bayesian inference by reducing the influence of the likelihood in the calculation of the posterior.

Recent literature on α-posterior inference has studied the robustness properties of α-posteriors ([24, 25, 28, 50, 4]). Consequently, the statistical properties of α-posteriors including concentration ([8, 2, 67]) and asymptotic behavior ([4, 54]) have been studied as well. In particular, [4] showed that α-posteriors of parametric, low-dimensional models obey a Bernstein-von Mises theorem (BvM), even under model-misspecification. The BvM for standard posteriors states that the distance between the posterior and a Gaussian centered around the Maximum Likelihood Estimator (MLE) with covariance related to the "curvature" of the (potentially-misspecified) likelihood converges to 0 with growing sample size ([44, 40]). The BvM for α-posteriors that was established in [4] states that the limiting Gaussian of the α-posterior is that same as that of the standard posterior but with variance divided by α. In the same setting as [4], [54] established the consistency of the α-posterior moments and the $\sqrt{n}$-consistency of the α-posterior mean.

1.1 Our contribution

In this work, we consider the asymptotic behavior of the α-posterior when α is chosen in a data-driven fashion. Throughout this work, we refer to these posteriors as "$\hat{\alpha}_n$-posteriors," where $\hat{\alpha}_n$ denotes an α that is (potentially) data-dependent). The optimal selection of α has remained an open question with several approaches proposed in the literature (see [24, 25, 63, 28, 50, 53, 57, 65, 48, 13]). However, the effects of these selection methods on the resulting α-posterior have not been studied. Our theoretical results are of the same form of the results from [4] and [54]. We outline the main contributions of our work below.

First, in Section 3, we consider the problem of selecting α through an example of misspecified linear regression. We investigate natural tuning selection methods based on cross-validation. Our numerical experiments reveal that when the model is correctly specified, Bayesian-based cross-validation and train-test splitting can be unstable in a way that is consistent with findings from other works ([56, 53, 48]). In particular, these methods select $\alpha = \infty$ with some probability – that does not vanish with n – and select $\alpha = \alpha_n$ otherwise, where $\frac{1}{n} \ll \alpha_n \ll 1$.

We study $\hat{\alpha}_n$-posteriors in the aforementioned asymptotic regimes. Specifically, we establish a BvM (Theorem 1; Corollary 1) and a moment-consistency result (Theorem 2; Corollary 2) for data-dependent $\hat{\alpha}_n$-posteriors where $\frac{1}{n} \ll \hat{\alpha}_n \ll 1$ with high probability. We also establish that $\frac{1}{n} \ll \hat{\alpha}_n$ is a necessary condition for a BvM result to hold (Proposition 1).

We also establish the $\sqrt{n}$-consistency of the $\hat{\alpha}_n$-posterior mean (Theorem 3; Corollary 3) when $\frac{1}{\sqrt{n}} \ll \hat{\alpha}_n \ll 1$. Our proof technique involves establishing a new Laplace approximation result (Lemma 1). We also use this Laplace approximation to formalize the understanding that the "∞-posterior" (i.e., the α-posterior, where $\alpha = \infty$) is equal to a point mass at the MLE.

1.2 Related Work

Our work draws upon several themes, which we outline below.

α-Variational Inference and generalized posteriors: First, we outline how α-posteriors arise naturally from a modified formulation of Variational Inference (VI). VI is an optimization-based approach to approximate an intractable posterior distribution ([33, 10, 62]). The core idea of VI is to approximate the posterior with the distribution $q^* \in \mathcal{Q}$ closest to the posterior in terms of KL-divergence, where $\mathcal{Q}$ is a family of computationally tractable distributions.

$$q^* = \arg\min_{q \in \mathcal{Q}} d_{\mathrm{KL}}\left(q || \pi(\theta | X^n)\right). \tag{1}$$

The KL-minimization, which is infeasible, is equivalent to maximizing the Evidence Lower Bound (ELBO) ([10]).

$$q^* = \arg\max_{q \in \mathcal{Q}} \int \log f_n(X^n | \theta) \pi(\theta) q(\theta) d\theta - d_{\mathrm{KL}}\left(q(\theta) || \pi(\theta)\right). \tag{2}$$

The ELBO comprises a log-likelihood term and a regularization term that forces the variational approximation to be closer to the prior distribution, balancing goodness of fit and complexity. Adding a penalization parameter to the regularization term results in a modified ELBO, which is maximized by the α-posterior when $\mathcal{Q}$ is unrestricted [4].

$$q^* = \arg\max_{q \in \mathcal{Q}} \int \log f_n(X^n | \theta) \pi(\theta) q(\theta) d\theta - \frac{1}{\alpha} d_{\mathrm{KL}}\left(q(\theta) || \pi(\theta)\right). \tag{3}$$

Such regularized ELBOs and related quantities have been referred to as α-VI in the literature ([29, 67]), and this objective has been used in training variational autoencoders ([27, 14]). The α-VI objective can be further generalized by replacing the log-likelihood with another loss function. When $\mathcal{Q}$ is restricted, this problem is known as generalized VI ([41]). When $\mathcal{Q}$ is unrestricted, this objective is maximized by the "generalized posterior" [9]. Such posteriors are also known as Gibbs posteriors [66, 3]. We note an advantage of generalized posteriors over power posteriors is the ability to use more "robust" losses than the log-likelihood. We discuss such posteriors in the next subsection.

We also note that – beyond VI – Gibbs (or generalized) posteriors arise as optimal solutions to several statistical problems such as aggregation of estimators ([61, 18, 55]), stochastic model selection [47], and minimum complexity density estimation/information complexity minimization ([6, 69, 70]).

Robust posteriors: As discussed in Section 1, α-posteriors have been shown to be more robust than traditional posteriors in many settings. The robustness properties of α-posteriors were initially explored in the context of linear regression, where standard posteriors can be inconsistent if the additive errors are assumed to be homoskedastic [25]. α-posteriors can also be modified to be robust to data corruption by applying an individual tempering parameter to each observation to downweight the impact of corrupt observations [63]. Furthermore, α-posteriors and their variational approximations were – asymptotically – shown to be robust to general misspecifications of the likelihood [4].

We also discuss other robust posteriors and their connections to α-posteriors. The coarsened posterior conditions on the data being perturbed – rather than being observed exactly, as it is in the calculation of the standard posterior [50]. This design of the posterior enables robustness to – for example – misspecification of the likelihood. Putting a prior on the magnitude of the perturbation reveals – asymptotically – that the coarsened posterior is approximately a power posterior, where the power depends on the sample size. However, the coarsened posterior does not contract because the sequence α_n decays to 0 too quickly (see [50, 4] and Proposition 1). Robust posteriors are often derived from constructing generalized posteriors with more "robust" losses ([22]) and test statistics ([5]).

Posterior tempering is also useful as a technique for more robust prior elicitation. For example, partial Bayes factors indicate a way to specify priors in Bayes factors by holding out a portion of the data to compute meaningful priors. Fractional Bayes factors are an approximation to partial Bayes factors that involves raising the likelihood to the fraction of held-out samples, and they tend to be more robust to prior specification ([52]). Similarly, power priors construct informative priors from historical data by computing an α-posterior from the historical data to use as a prior for the current study ([31, 32, 15]).

Data-driven posterior tempering: As mentioned in Section 1, our work is motivated by the problems of **(I)** selecting the tempering parameter, α, and **(II)** understanding the statistical properties of "data-driven" α-posteriors. Several approaches to learn the tempering parameter from the data have been proposed in the literature. The SafeBayesian algorithm proposed in [24] and explored in [25] chooses the tempering parameter that minimizes the cumulative posterior expected log loss. That is, the tempering parameter is chosen to maximize the "prequential" log-likelihood, which is similar to Bayesian cross-validation criteria (see [68]). The Bayesian data reweighting approach proposed by [63] assigns a tempering parameter to each observation (as opposed to a single tempering parameter applied to the entire likelihood) and jointly models the weights and observations by putting a prior on the weights. Another approach proposed by [28] chooses the tempering parameter such that the expected gain in information between the prior and tempered posterior matches the expected gain in information between the prior and standard posterior. The tempering parameter of the coarsened posterior proposed by [50] is chosen to balance goodness of fit and model complexity. Another class of methods chooses the tempering parameter (which controls the posterior variance as shown in [4, Theorem 1]) such that the credible intervals obtain the nominal frequentist coverage probability ([57, 64]). Another approach proposes a closed-form expression of the optimal inverse temperature in the Gibbs posterior that minimizes the population risk ([13]). Recent work proposes to choose the tempering parameter to optimize posterior predictive criteria but also shows that this is an ill-posed problem [48].

Motivated by the connections between α-posterior inference and penalized estimation – which we outline in Section 3 – we investigate both traditional and Bayesian-based cross-validation methods as well as the SafeBayesian algorithm. See Section 3.2 for further details.

Asymptotic behavior of posterior distributions: We also provide an overview of previous work related to studying the asymptotic behavior of posterior distributions and how our work relates to this literature. Our results rely heavily on the Bernstein-von Mises theorem for Bayesian posteriors derived from parametric, low-dimensional likelihoods. The classical Bernstein-von Mises theorem states that the total variation distance between the posterior and a Gaussian centered around the MLE with asymptotic variance equal to the inverse Fisher information converges in probability to 0 ([44] and references therein). The required regularity conditions on the likelihood are similar to those ensuring asymptotic normality of the maximum likelihood estimator. When the likelihood is misspecified (i.e., the data-generating distribution is different from the likelihood model), a Bernstein-von Mises result holds, but the asymptotic behavior of the misspecified posterior is determined with respect to the "pseudo-true" model ([40]). In the same setting, α-posteriors have the same limiting distribution as that of standard posteriors but with asymptotic variance divided by α ([4]).

We also note here other forms of results from which we could derive BvM results for α-posteriors. BvMs for quasi-posteriors are discussed in [17] and [22], and in [49] for generalized posteriors. Semiparametric BvMs were derived in [16], and the extensions to the sorts of α_n-posteriors we are considering (in the semiparametric setting) were derived in [42]. A BvM result for Markov Processes was derived in [12].

1.3 Organization

The rest of this paper is organized as follows. We introduce notation and definitions in Section 2. Section 3 demonstrates the connections between α-posterior inference and their corresponding problems in penalized estimation. We present our main results in Section 4. Section 5 presents a discussion and our concluding remarks. The proofs of our main results are in Appendix D.

2 Notation and Preliminaries

Let $\phi(\cdot|\mu, \Sigma)$ denote the multivariate normal density with mean μ and covariance matrix Σ. For a matrix, A, let $|A|$ denote the determinant of A. The indicator function of an event A is denoted $\mathbb{1}\{A\}$. For a sequence of distributions P_n on random variables (or matrices) X_n, if $\lim_n P_n(||X_n||_2 > \epsilon) = 0$ for every $\epsilon > 0$, we say $X_n = o_{P_n}(1)$. We say X_n is 'bounded in P_n-probability', or $O_{P_n}(1)$, if for every $\epsilon > 0$ there exists $M_\epsilon > 0$ such that $P_n(||X_n||_2 < M_\epsilon) \geq 1-\epsilon$. We use $\|X\|_p = (\sum_i |X_i|^p)^{1/p}$ to denote the p-norm. Let $\bar{B}_v(\delta) \coloneqq \{\theta : ||\theta - v|| \leq \delta\}$ denote a closed ball of radius δ around vector $v \in \mathbb{R}^p$ and $B_v(\delta) \coloneqq \{\theta : ||\theta - v|| < \delta\}$ denote the open ball. For any two probability densities p and q, define the total variation distance, $d_{\text{TV}}(p, q)$, between them as

$$d_{\text{TV}}(p, q) = \frac{1}{2} \int_{\mathbb{R}^p} |p(h) - q(h)| dh$$

and the KL-divergence, $d_{\text{KL}}(p, q)$, between them as

$$d_{\text{KL}}(p, q) = \int_{\mathbb{R}^p} p(h) \log\left(\frac{p(h)}{q(h)}\right) dh.$$

Statistical Model: Let $\mathcal{F}_n = \{f_n(\cdot|\theta) : \theta \in \mathbb{R}^p\}$ be a parametric family of densities used as a statistical model for the random sample $X^n = (X_1, \ldots, X_n) \in \mathcal{X}$. This model is allowed to be misspecified in the sense that $f_{0,n}(X^n)$, the true density of the random sample X^n, may not belong to $\mathcal{F}_n$. We will use the notion of the pseudo-true parameter θ^*, which is the true data-generating θ if the model is well-specified, meaning $f_{0,n}(X^n) = f_n(X^n|\theta^*) \in \mathcal{F}_n$. If the model is misspecified, $\theta^* = \arg\min_{\theta\in\mathbb{R}^p} D_{KL}\left(f_{0,n}(X^n)||f_n(X^n|\theta)\right)$. Denote the (pseudo) MLE by $\hat{\theta} = \arg\max_\theta f_n(X^n|\theta)$. Let $\mathbb{E}_{f_{0,n}}$ and $\mathbb{P}_{f_{0,n}}$ denote the expectation and probability taken with respect to $f_{0,n}$.

$\boldsymbol{\alpha_n}$ **and** $\boldsymbol{\hat{\alpha}_n}$ **-posteriors:** Given the statistical model $\mathcal{F}_n$, a prior density $\pi(\theta)$ for θ, and a sequence of scalars α_n, the power posterior, or α_n-posterior, has the density:

$$\pi_{n,\alpha_n}(\theta|X^n) = \frac{f_n(X^n|\theta)^{\alpha_n}\pi(\theta)}{\int_{\mathbb{R}^p} f_n(X^n|\theta)^{\alpha_n}\pi(\theta)d\theta}. \tag{4}$$

In this work, we subscript with α with n to emphasize the dependence of α on the sample size, n. We also differentiate α_n-posteriors – where the sequence, α_n is chosen a priori – and $\hat{\alpha}_n$-posteriors – where $\hat{\alpha}_n$ is learned from the data or depends on the data in any way.

Variational approximations: Consider a family $\mathcal{Q}$ of distributions. Examples include the mean-field family (independent marginals) or the Gaussian mean-field family (independent Gaussian marginals). The variational approximation of a distribution, p, is the distribution $q^* \in \mathcal{Q}$ that is closest to p in terms of KL divergence:

$$q^* = \arg\min_{q\in\mathcal{Q}} D_{\mathrm{KL}}\left(q||p\right). \tag{5}$$

3 A motivating example: misspecified linear regression

In this section, we consider the problem of tuning the regularization parameter in ridge regression. This problem can be equivalently stated as tuning the tempering parameter in Bayesian linear regression. We investigate the limiting distribution of the tuned parameter values selected according to methods in the context of both ridge regression and Bayesian linear regression.

We conduct our numerical experiments in the following regression setting. Consider paired data $\{x_i, y_i\}_{i=1}^n$, where $x_i = [x_{i,1}, x_{i,2}, x_{i,3}]^\top \in \mathbb{R}^3$ and $y_i \in \mathbb{R}$ for all $i = 1, \ldots, n$. We simulate each sample according to the following generative model

$$\begin{aligned} x_i &\sim \mathcal{N}(0, I_3) \\ y_i &= x_i^\top \beta^* + \epsilon_i, \end{aligned} \tag{6}$$

where $\epsilon_i \sim \mathcal{N}(0,1)$ and $\beta^* = [\beta_1^*, -0.5, 0.1]$. We describe how we set $\beta_1^* \in \mathbb{R}$ in Section 3.1.

3.1 Model specification, $\boldsymbol{\alpha_n}$-posterior and connection to ridge regression

In our experiments, we consider the following model specification:

$$y_i = \tilde{x}_i^\top \tilde{\beta} + \epsilon_i, \tag{7}$$

where $\epsilon_i \sim \mathcal{N}(0,1)$ and $\tilde{x}_i = [x_{i,2}, x_{i,3}]^\top \in \mathbb{R}^2$. Notice that if $\beta_1^* = 0$ in (6), then (7) is "well specified." If $\beta_1^* \neq 0$ in (6), then (7) is incorrectly specified. In our experiments, we set $\beta_1^* = 1$ and denote this setting as "misspecified."

We note the equivalence between ridge regression and α_n-posterior inference in the linear regression model with a Gaussian prior. Indeed, the model in (7) gives rise to a likelihood for the

Behavior of $\sqrt{n}(\hat{\beta}^{\text{Ridge}} - b(\lambda) - \beta_0)$

Regime		α_n-posterior	$\hat{\beta}^{\text{Ridge}}$ (α_n-posterior mean)	
λ_n	α_n	Variance	$b(\lambda)$	$\sqrt{n}(\hat{\beta}^{\text{Ridge}} - \hat{\beta})$
$\frac{1}{\sqrt{n}} \ll \lambda_n \ll c$	$\frac{1}{n} \ll \alpha_n \ll \frac{1}{\sqrt{n}}$	$O_p\left(\frac{1}{\alpha_n n}\right)$	$O_p\left(\lambda_n\right)$	Diverges
$\lambda_n \sim \frac{1}{\sqrt{n}}$	$\alpha_n \sim \frac{1}{\sqrt{n}}$	$O_p\left(\frac{1}{\sqrt{n}}\right)$	$O_p\left(\frac{1}{\sqrt{n}}\right)$	$O_p(1)$
$\frac{1}{n} \ll \lambda_n \ll \frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}} \ll \alpha_n \ll \alpha$	$O_p\left(\frac{1}{\alpha_n n}\right)$	$O_p\left(\lambda_n\right)$	$o_p(1)$
$\lambda_n \sim \frac{1}{n}$	$\alpha_n \sim \alpha$	$O_p\left(\frac{1}{\alpha n}\right)$	$O_p\left(\frac{1}{n}\right)$	$o_p(1)$

Table 1: Asymptotic behavior of $\hat{\beta}^{\text{Ridge}}$ and the corresponding α-posterior. We denote the bias of the ridge estimator/α-posterior mean by $b(\lambda)$. The last three columns denote the variance of the α_n-posterior, the bias of the ridge estimator, and the order of $\sqrt{n}(\hat{\beta}^{\text{Ridge}} - \hat{\beta})$.

data $\{x_i, y_i\}_{i=1}^n$, given the coefficient vector β. Combining this likelihood with a $\mathcal{N}(0, \sigma^2 I_2)$ prior for β and using $\alpha_n = \frac{1}{n\lambda_n}$ yields the α-posterior

$$\pi_{n,\alpha_n}(\beta|X,Y) = \phi\left(\beta \,\middle|\, \hat{\beta}^{\text{Ridge}},\ \frac{\sigma^2}{\alpha_n n}\left(\frac{1}{n}X^\top X + \frac{1}{\alpha_n n}I_p\right)^{-1}\right), \tag{8}$$

where $\lambda_n = \frac{1}{n\alpha_n}$,

$$\hat{\beta}^{\text{Ridge}} = \arg\min_{\beta} \frac{1}{n}\|Y - X\beta\|_2^2 + \lambda_n\|\beta\|_2^2, \tag{9}$$

$X = [x_1 \cdots x_n]^\top$, and $Y = [y_1, \ldots, y_n]$. See Appendix A.1 for additional details. We note that standard frequentist theory shows that under some regularity conditions, $\hat{\beta}^{\text{Ridge}}$ is consistent as long as $\lambda_n = o(1)$ and $\hat{\beta}^{\text{Ridge}}$ is $\sqrt{n}$-asymptotically normally distributed if $\lambda_n = o(1/\sqrt{n})$. In Table 1, we compare the asymptotic behavior of $\hat{\beta}^{\text{Ridge}}$ to the to that of the α_n-posterior in (8) in these regimes of λ_n (and the corresponding regimes of α_n). Table 1 suggests that the regimes $1/n \ll \alpha_n \ll 1$ and $1/\sqrt{n} \ll \alpha_n \ll 1$ are particularly interesting. We will see in Section 4 that these regimes are indeed the most relevant ones for establishing Bernstein-von-Mises thereoms and asymptotic normality of the α-posterior mean estimator. Before showing these theoretical results, we will see that natural data-driven choices of α do indeed lead us to these types of regimes as well.

3.2 Data-driven choices of the tempering parameter

In both the well specified and misspecified model settings (Section 3.1), we investigate the distribution of $\hat{\alpha}_n$, where $\hat{\alpha}_n$ is selected according to the following methods.
Bayesian cross-validation (CV): We choose α to maximize the leave-one-out log pointwise predictive density ([19, 60])

$$\overline{\text{lppd}}_{\text{LOO}} = \frac{1}{n}\sum_{i=1}^{n} \log \int p\left(x_i, y_i|\beta\right) \pi_{n,\alpha}\left(\beta|X_{-i}Y_{-i}\right) d\beta, \tag{10}$$

where X_{-i} and Y_{-i} are the design matrix and dependent variable, respectively, with the i^{th} entries removed. The leave-one-out $\overline{\text{lppd}}$ was discussed in the context of tuning hyperparameters in Bayesian inference in [67] (referred to as elpd_{LOO}). See Appendix A.2 for the $\overline{\text{lppd}}_{\text{LOO}}$ calculation for the posterior in (8).

Bayesian CV + VI: We also consider (10) with the α-posterior replaced with its Gaussian mean field variational approximation

$$\overline{\text{lppd}}_{\text{LOO-VI}} = \frac{1}{n}\sum_{i=1}^{n} \log \int p\left(x_i, y_i|\beta\right) \tilde{\pi}_{n,\alpha}\left(\beta|X_{-i}Y_{-i}\right) d\beta, \tag{11}$$

where $\tilde{\pi}_{n,\alpha}(\cdot)$ denotes the density of the Gaussian mean field variational approximation of (8)

$$\tilde{\pi}_{n,\alpha}\left(\beta|X,Y\right) = \phi\left(\beta\middle|\hat{\beta}^{\text{Ridge}},\ \frac{\sigma^2}{\alpha n}\left[\text{diag}\left(\frac{1}{n}X^\top X + \lambda I\right)\right]^{-1}\right). \tag{12}$$

See Appendix A.1 for the derivation of (12) and Appendix A.3 for the $\overline{\text{lppd}}_{\text{LOO-VI}}$ calculation for the posterior in (8).

LOOCV: We choose λ to minimize Allen's PRESS statistic ([1])

$$\text{PRESS} = \frac{1}{n}\|B(I_n - H)Y\|_2^2, \tag{13}$$

where B is diagonal with entries $B_{ii} = (1 - H_{ii})^{-1}$, $H = X(X^\top X + \lambda I_p)^{-1}X^\top$, and $\lambda = \frac{1}{\alpha}$. Allen's PRESS statistic is equivalent to the leave one out cross-validation (LOOCV) squared error of $\hat{\beta}^{\text{Ridge}}$.

Train-test split: Given a data split into a training set and test set, we choose α to minimize the squared test error of $\hat{\beta}^{\text{Ridge}}$ estimated on the training dataset (we use a 70-30 training-test data split).

SafeBayes: The SafeBayesian algorithm proposed in [23] and explored in [25] chooses α in the interval $[0, 1]$ that minimizes the cumulative posterior expected log loss (i.e., it maximizes the "prequential" log-likelihood). Since we assume our noise variance is known, we use the R-square-Safe-Bayesian Ridge Regression algorithm ([26]).

We use grid searches to select the optimal parameter according to each method. For methods selecting λ, we map $\hat{\lambda}_n$ to its equivalent $\hat{\alpha}_n$ value in the context of the $\hat{\alpha}_n$-posterior defined in (28). See Table 4 in Appendix A.4 for these mappings and the grids used.

3.3 Simulation results

Figure 1 displays the distribution of $\hat{\alpha}_n$ – over 1,000 independent replications of the data – selected according to the methods described in 3.2 in the model specification settings described in Section 3.1. To understand the asymptotic behavior of $\hat{\alpha}_n$, we compare the distribution of $\hat{\alpha}_n$ when $n = 100$ and when $n = 1,000$. Among the methods we consider, we note the following classes of asymptotic behavior of $\hat{\alpha}_n$.

Constant limit: In both model specification settings, LOOCV and SafeBayes suggest a $\hat{\alpha}_n$ that approaches a constant. The asymptotics of the resulting constant α-posterior were studied in [4] and [54]. In particular, a Bernstein-von Mises theorem for the α-posterior was established in [4]. The convergence of the moments of the α-posterior and $\sqrt{n}$-consistency of the α-posterior mean were established in [54].

Method	Model Specification	Est. exponent	95% CI
Bayesian CV	Well specified	-0.63	[-0.66, -0.59]
	Misspecified	-1.04	[-1.05, -1.03]
Bayesian CV + VI	Well specified	-0.63	[-0.66, -0.59]
	Misspecified	-1.04	[-1.05, -1.02]
Train-test split	Well specified	-0.63	[-0.66, -0.6]
	Misspecified	-0.61	[-0.64, -0.58]

Table 2: Estimated stochastic order of $\hat{\alpha}_n$. We model $\hat{\alpha}_n$ as $\hat{\alpha}_n = Cn^{\gamma}$ and report the estimates and confience intervals of γ. The corresponding curves of best fit are shown in Figure 5. See Appendix A.5 for details on the estimation of γ.

Quickly vanishing $\hat{\boldsymbol{\alpha}}_{\mathbf{n}}$: In the misspecified model setting, Bayesian CV and Bayesian CV + VI suggest a $\hat{\alpha}_n$ that decreases to 0. According to Table 2, this rate of convergence is estimated to be on the order of $\frac{1}{n}$. For details on the estimation of these convergence rates and plots of the curves of best fit, see Appendix A.5 and Figure 5. We study the resulting $\hat{\alpha}_n$-posterior when $\alpha \sim \frac{1}{n}$ in Section 4.1.2 and show that a Bernstein-von Mises result does not hold for exponential family conjugate models in this regime of $\hat{\alpha}_n$.

Mixed $\hat{\boldsymbol{\alpha}}_{\mathbf{n}}$: The remaining settings (Bayesian CV and Bayesian CV + VI in the well-specified setting and train-test split in both model specification settings) suggest a $\hat{\alpha}_n$ that has a point mass at what is effectively $\alpha = \infty$ – which we call a "corner solution" – with remaining mass appearing to converge to 0. For more details on how we recode $\hat{\alpha}_n$ values to be ∞, see Appendix A.4. According to Table 2, this rate of convergence is estimated to be slower than $\frac{1}{n}$. Furthermore, approximately half of the $\hat{\alpha}_n$ are corner solutions (see Panel C of Figure 5), asymptotically. These corner solutions have been observed in other works as well. When α is selected to optimize the expected log predictive density (a quantity similar to the log point predictive density), [48] also observe a point mass in the at $\alpha = \infty$. Furthermore, when λ is selected according to a train-test split, [56] observe that, asymptotically, around half of the selected values are corner solutions at $\lambda = 0$ (i.e., $\alpha = \infty$) and formalize this observation into a result about the asymptotic distribution of the selected regularization parameter. In this work, we study the asymptotics of the $\hat{\alpha}_n$-posterior that arise from this distribution of $\hat{\alpha}_n$. We separately study the α_n-posterior when $\frac{1}{n} \ll \alpha_n \ll 1$ (Theorem 1 – 3) and when $\alpha_n = \infty$ (Proposition 7). We extend these results to data-dependent sequences, $\hat{\alpha}_n$ (Corollary 1 – Corollary 3). We combine these results to formalize a Bernstein-von Mises for what we call the "mixed $\hat{\alpha}_n$-posterior" (Corollary 5).

3.4 Application: CPS1988

We illustrate the relevance of the asymptotic regimes of $\hat{\alpha}_n$ in real data by studying the distributions of $\hat{\alpha}_n$ and the resulting estimated coefficients on the CPS1988 dataset. This is a cross-sectional dataset containing data from the U.S. Census Bureau's March 1988 Current Population Survey (CPS). It consists of 28,155 observations containing information about the wage, education, work experience, and ethnicity of men between the ages of 18 and 70 who are making more than \$50 yearly and are neither self-employed nor working without pay ([39]).

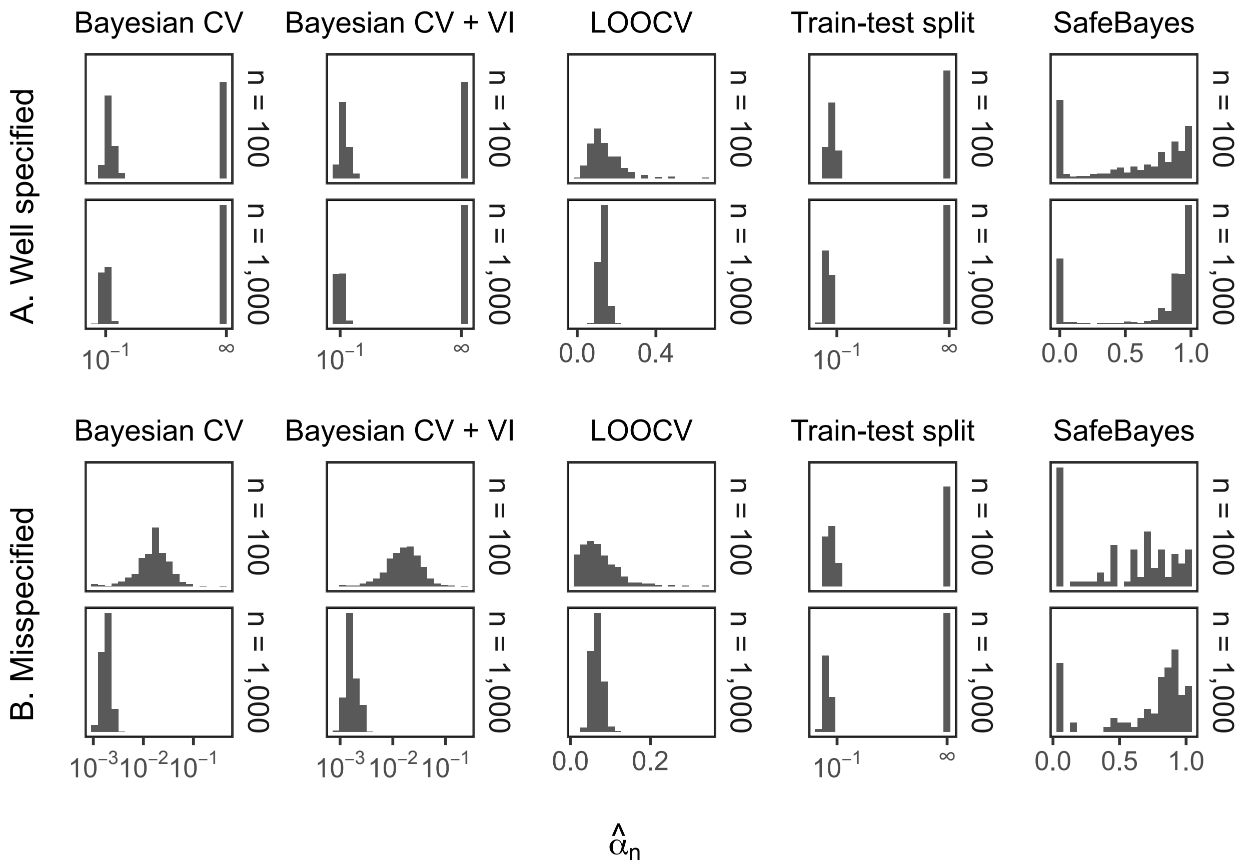


Figure 1: Distribution of $\hat{\alpha}_n$ over 1,000 replications in the well-specified model setting (Panel A) and misspecified model setting (Panel B). Within each panel, columns correspond to a selection method of α (Bayesian cross-validation, Bayesian CV + VI, LOOCV, train-test split, SafeBayes), and rows correspond to sample size per replication ($n = 100$, $n = 1000$).

3.4.1 Model specifications

As a baseline, we start with the model specified in [39], which we refer to as the "full model"

$$\log(\text{wage}) = \beta_0 + \beta_1\text{education} + \beta_2\text{ethnicity} + \beta_3\text{experience} + \beta_4\text{experience}^2 + \epsilon. \tag{14}$$

We consider two additional model specifications that remove covariates from (14). A first model model with the ethnicity variable removed

$$\log(\text{wage}) = \beta_0 + \beta_1\text{education} + \beta_2\text{experience} + \beta_3\text{experience}^2 + \epsilon, \tag{15}$$

and another model with the squared experience term removed

$$\log(\text{wage}) = \beta_0 + \beta_1\text{education} + \beta_2\text{ethnicity} + \beta_3\text{experience} + \epsilon. \tag{16}$$

We note that eliminating the squared experience term will greatly affect the estimation of the coefficient of the linear term. We do not claim that (14) is the correct or true model, but it seems like it is closer to correctly specified than either (15) or (16).

3.4.2 Asymptotic regimes of $\hat{\alpha}_n$

First, we repeat the experiments in Figures 1 and 5 by taking subsamples from the full dataset. We note that the results of this experiment are not equivalent to the results in Figures 1 – 5 because those experiments involve independent replications of the data, whereas subsamples are dependent. Nevertheless, we find that similar asymptotic regimes arise.

First, we examine the suggested distributions of $\hat{\alpha}_n$ in Figure 2. We discuss the appearance of the asymptotic regimes discussed in Section 3.3.

Constant limit: Under all model specifications, SafeBayes suggests a $\hat{\alpha}_n$ that approaches a constant.

Quickly vanishing $\hat{\boldsymbol{\alpha}}_{\mathbf{n}}$: We do not observe the "quickly vanishing" regime under any of the model specifications.

Mixed $\hat{\boldsymbol{\alpha}}_{\mathbf{n}}$: According to Figure 2, the distribution of $\hat{\alpha}_n$ selected according to both Bayesian CV + VI (under the full model and model excluding ethnicity) and train-test split (under all model specifications) exhibits the mixture behavior discussed in our simulation example. For plots of the curves of best fit see Figure 6 in Appendix A.5. According to Table 3, the non-corner solutions selected according to Bayesian CV + VI and train-test split are estimated to be of order slower than $\frac{1}{n}$ under all model specifications. We also note that LOOCV also appears to exhibit the same mixture behavior to some extent under all model specifications. However, we find that the values of the "non-corner" $\hat{\alpha}_n$ appear to remain constant or grow – rather than shrink – with n (see Figure 6 in Appendix A.5 for more details). Finally – while this is not mixture behavior – we note that Bayesian CV selects corner solutions under all model specifications. We formalize the behavior of the resulting α-posterior in Corollary 4.

3.4.3 Stability of $\hat{\alpha}_n$

To investigate the stability of the selection methods discussed in Section 3.2 on real data, we study the distributions of $\hat{\alpha}_n$ and the resulting estimated coefficients on subsamples from the dataset.

In our stability experiments, we take 100 subsamples each of size $n = 5,000$ from the dataset. For each subsample, we select the relevant regularization/tempering parameter according to the methods described in Section 3.2. See Table 5 for the grids used in our grid searches. We also

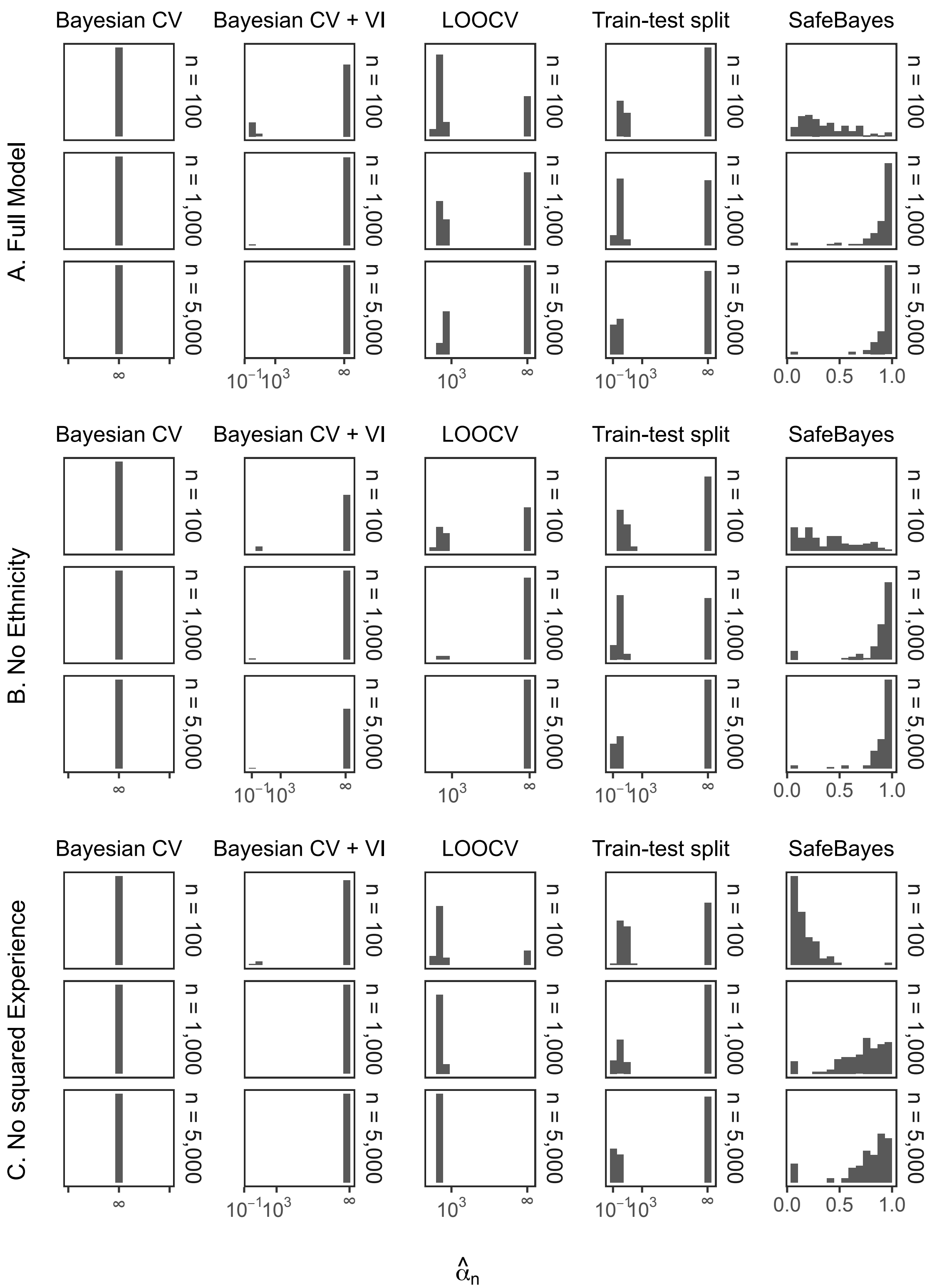


Figure 2: Distribution of $\hat{\alpha}_n$ over 100 subsamples from the CPS1988 dataset under the full model (Panel A), model excluding ethnicity (Panel B), and model excluding squared experience (Panel C). Within each panel, columns correspond to a selection method of α (Bayesian cross-validation, Bayesian CV + VI, LOOCV, train-test split, SafeBayes), and rows correspond to sample size per replication ($n = 100$, $n = 1,000$, $n = 5,000$).

Method	Model Specification	Solution	Est. exponent	95% CI
Bayesian CV + VI	Full Model	non-corner	-0.82	[-0.95, -0.68]
	No Ethnicity	non-corner	-0.85	[-0.98, -0.71]
LOOCV	Full Model	non-corner	0.36	[0.27, 0.44]
	No Ethnicity	non-corner	0.47	[0.3, 0.64]
	No squared Experience	non-corner	0.00	[-0.04, 0.04]
Train-test split	Full Model	non-corner	-0.66	[-0.75, -0.56]
	No Ethnicity	non-corner	-0.70	[-0.77, -0.64]
	No squared Experience	non-corner	-0.59	[-0.69, -0.5]

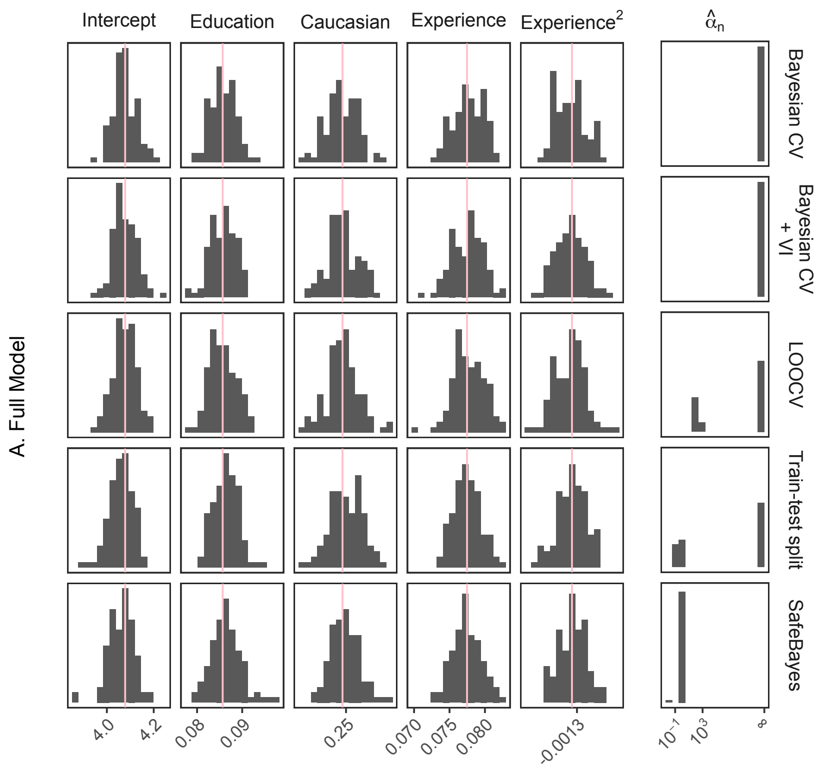
Table 3: Estimated stochastic order of $\hat{\alpha}_n$. We model $\hat{\alpha}_n$ as $\hat{\alpha}_n = Cn^{\gamma}$ and report the estimates and confience intervals of γ. The corresponding curves of best fit are shown in Figure 6. See Appendix A.5 for details on the estimation of γ.

compute the mean of the resulting α-posterior according to (8) (which is the same as the mean of its variational approximations according to (12)). We compare the distribution of α-posterior mean values to the least squares estimates of the full model coefficients (on the full dataset). We call these the "full coefficient values." The results of these experiments are displayed in Figure 3a (full model), Figure 3b (model excluding ethnicity), and Figure 3c (model excluding squared experience). We make the following observations:

Overall stability of $\hat{\alpha}_n$: Given the relatively large subsample size, we would expect for $\hat{\alpha}_n$ values to "agree" across subsamples. Indeed, across all methods and model specifications, the $\hat{\alpha}_n$ values chosen according to Bayesian CV, Bayesian CV+VI, and SafeBayes are not sensitive to the choice of subsample. The volatility of the distribution of $\hat{\alpha}_n$ chosen according to train-test split is likely due to the additional randomness induced by data-splitting. The $\hat{\alpha}_n$ chosen according to LOOCV are somewhat sensitive to choice of subsample under the full model (Figure 3a). This is the only case where the distribution of $\hat{\alpha}_n$ in this data example is different from the distribution of $\hat{\alpha}_n$ in the simulation example (Figure 1).

Agreement between methods: Under all model specifications, all methods tend to favor higher values of $\hat{\alpha}_n$, including $\alpha = \infty$. We note the following exceptions to this observation. First, SafeBayes constrains $\alpha \in [0, 1]$, but Safebayes overwhelmingly selects α to be close to 1 under all model specifications. Second, train-test split exhibits mixture behavior similar to that in Figure 1, where a proportion of values of $\hat{\alpha}_n$ are effectively ∞. However, this mixture behavior is attributable to the additional randomness induced by data-splitting as discussed before.

Effect of model specification: In our discussion of the different model specifications, we conjectured that the models excluding covariates are more misspecified than the full model. Hence, we would expect the selected values of $\hat{\alpha}_n$ to be comparatively lower. However, this is not the case for all of the methods. In our discussions of stability of $\hat{\alpha}_n$, we note that $\hat{\alpha}_n$ is stable to model specification in most settings. The exceptions are LOOCV and SafeBayes, which select smaller $\hat{\alpha}_n$ values in the model excluding squared experience. Since excluding squared experience affects the estimation error of other coefficients more so than excluding ethnicity (Figure 3c vs. Figure 3b), it seems reasonable that some methods choose comparatively lower $\hat{\alpha}_n$. However, it appears to be the case that the selected values of $\hat{\alpha}_n$ are still moderately large, even for methods that are selecting comparatively smaller $\hat{\alpha}_n$. Hence, the distribution of estimated coefficients does not appear to



(a) Full model

change with model specification.

4 Main Results

In this section, we establish results about the $\hat{\alpha}_n$-posterior when $\hat{\alpha}_n$ exhibits the "mixed $\hat{\alpha}_n$" behavior discussed in Section 3. In Sections 4.1 – 4.3, we study the asymptotics of α_n-posteriors, α_n is a sequence fixed a priori satisfying $1/n \ll \alpha_n \ll 1$. We extend all our results to data-dependent $\hat{\alpha}_n$ (i.e., $\hat{\alpha}_n$-posteriors), where $1/n \ll \hat{\alpha}_n \ll 1$ with high probability. In Section 4.4, we study the α_n-posterior when $\alpha_n = \infty$ and the $\hat{\alpha}_n$-posterior when $\hat{\alpha}_n$ exhibits the "mixed $\hat{\alpha}_n$" behavior.

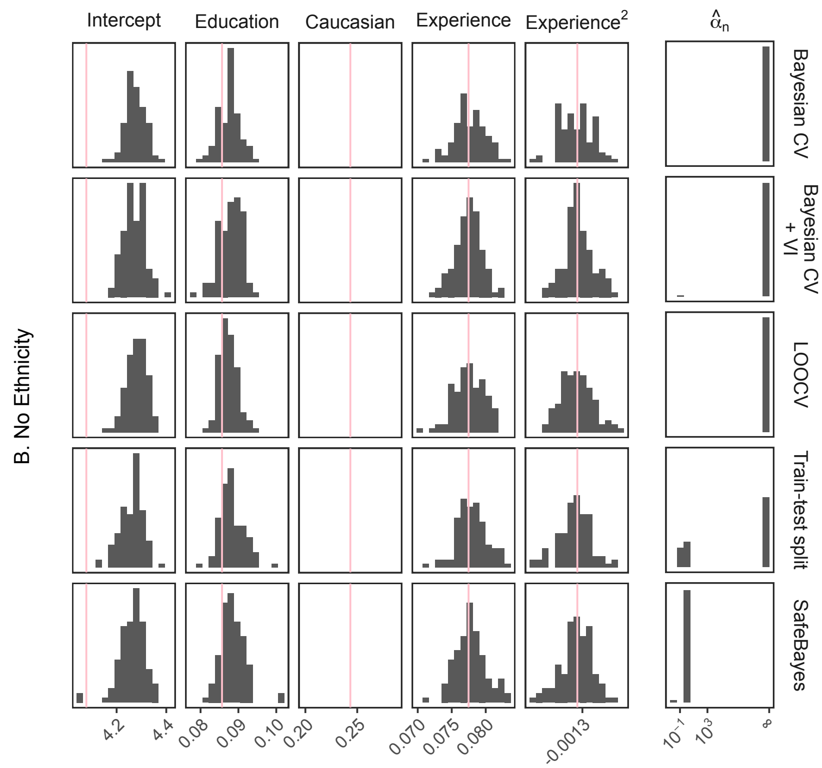
4.1 Bernstein-von Mises Theorem for α_n-posterior

4.1.1 Main result

Theorem 1 establishes that the α_n-posterior is asymptotically normal for any sequence, α_n, such that $\frac{1}{n} \ll \alpha_n \ll 1$. In particular, we show the total variation distance between the α_n-posterior and a multivariate Gaussian with mean $\hat{\theta}$ and variance $\frac{1}{\alpha_n n} V_{\theta^*}$ converges in $f_{0,n}$-probability to 0 as $n \to \infty$. We will require the following conditions.

(A0) The MLE, $\hat{\theta}$, is unique and there exists a θ^* in the interior of $\mathbb{R}^p$ such that $\hat{\theta} \to \theta^*$ in $f_{0,n}$-probability and $\sqrt{n}(\hat{\theta} - \theta^*)$ is asymptotically normal.

(A1) $\pi(\theta)$ is continuous and positive in a neighborhood of θ^*.



(b) Model excluding ethnicity

(A2) Define $\Delta_{n,\theta^*} = \sqrt{n}(\hat{\theta}-\theta^*)$. There exists a positive definite matrix V_{θ^*} such that for all compact sets $K \subseteq \mathbb{R}^p$, we have $\sup_{h \in K} |R_n(h)| \to 0$ in $f_{0,n}$-probability, where $R_n(h)$ is defined as

$$-\alpha_n n \left(-\frac{1}{n} \log f_n \left(X^n \middle| \theta^* + \frac{h}{\sqrt{\alpha_n n}} \right) - \left[-\frac{1}{n} \log f_n(X^n|\theta^*) \right] \right) - \sqrt{\alpha_n} h^\top V_{\theta^*} \Delta_{n,\theta^*} + \frac{1}{2} h^\top V_{\theta^*} h.$$

(A3) The α_n-posterior concentrates at rate $\sqrt{\alpha_n n}$ around θ^*: there exists a sequence $\epsilon_n \to 0$ with $\alpha_n n \epsilon_n^2 \to \infty$ and $C > 0$ sufficiently large, such that

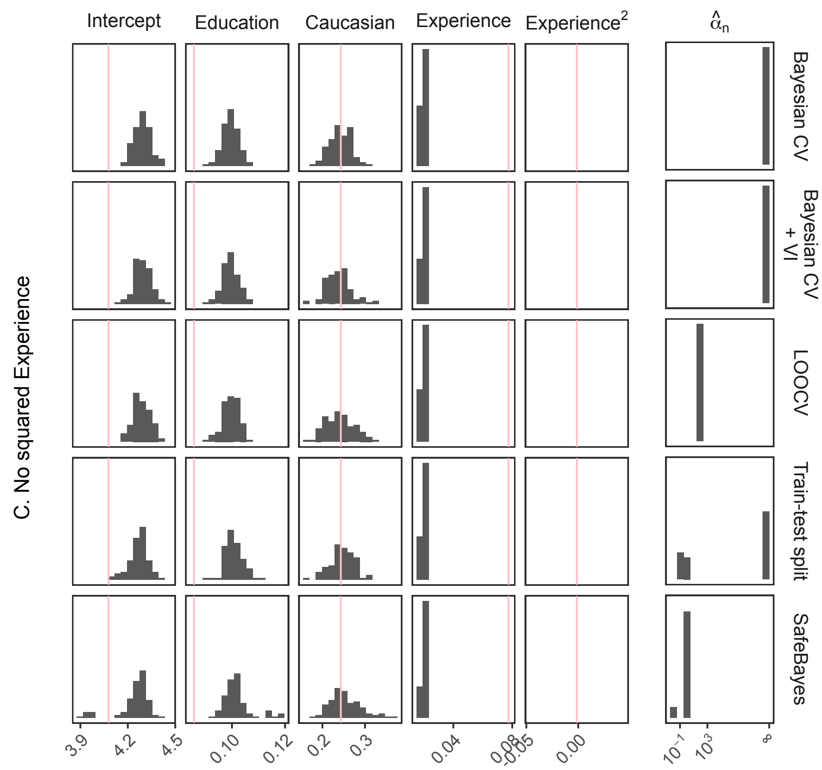
$$\mathbb{E}_{f_{0,n}} \left[\mathbb{P}_{\pi_{n,\alpha_n}} \left(\|\theta - \theta^*\|_2 > C\epsilon_n \right) \right] \to 0.$$

Theorem 1. (Bernstein-von Mises) Let Assumptions **(A0)**, **(A1)**, **(A2)**, and **(A3)** hold. Then for any sequence α_n such that $\frac{1}{n} \ll \alpha_n \ll 1$, in $f_{0,n}$-probability,

$$\mathrm{d_{TV}} \left(\pi_{n,\alpha_n}(\theta|X^n), \phi \left(\theta \middle| \hat{\theta}, \frac{1}{\alpha_n n} V_{\theta^*}^{-1} \right) \right) \to 0. \tag{17}$$

The full proof of this result is in Appendix D.1, but we provide a brief sketch here. By the posterior concentration assumption, **(A3)**, studying the the total variation distance between the α_n-posterior and the limiting Gaussian in (17) amounts to comparing to the log α_n-posterior ratio

$$\log \left(\frac{\pi_{n,\alpha_n}(\theta^* + h/\sqrt{\alpha_n n}|X^n)}{\pi_{n,\alpha_n}(\theta^*|X^n)} \right) \tag{18}$$



(c) Model excluding squared experience

Figure 3: Plots of the distributon of $\hat{\alpha}_n$ (last column) and resulting posterior means of each coefficient over 100 subsamples of size $n = 5,000$. Each row corresponds to a selection method. The pink vertical lines correspond to the least squares estimates of the coefficients of the full model (estimated on the full dataset).

to the ratio of the limiting Gaussians

$$\log\left(\frac{(\alpha_n n)^{-\frac{p}{2}}\phi(\theta^* + h/\sqrt{\alpha_n n}|\hat{\theta}, V_{\theta^*}^{-1}/\alpha_n n)}{(\alpha_n n)^{-\frac{p}{2}}\phi(\theta^*|\hat{\theta}, V_{\theta^*}^{-1}/\alpha_n n)}\right). \tag{19}$$

By the prior mass condition, **(A1)** and regularity conditions on the log-likelihood, **(A0)** and **(A2)**, (18) and (19) are equal up to $o_{f_{0,n}}(1)$ terms.

4.1.2 Discussion

Theorem 1 extends the result in [4], which proves an analogous result in the regime where α_n is a constant. Recall that [4] showed the following,

$$\mathrm{d_{TV}}\left(\pi_{n,\alpha}(\theta|X^n), \phi\left(\theta\middle|\hat{\theta}, \frac{1}{\alpha n}V_{\theta^*}^{-1}\right)\right) \to 0 \tag{20}$$

in $f_{0,n}$ probability, for a fixed $\alpha > 0$. The limiting distribution in (20) is the same as that of the Bernstein-von Mises theorem for standard posteriors in [40] but with asymptotic variance divided by α. Our result says this also holds for $\frac{1}{n} \ll \alpha_n \ll 1$. Since $\sqrt{\alpha_n n} \ll \sqrt{n}$ by assumption, the rate of our Bernstein-von Mises theorem is slower than the $\sqrt{n}$ rate in [44, 40, 4]. This slower concentration is also discussed in [42] and [48].

We note the importance of the requirement that $\frac{1}{n} \ll \alpha_n$ because a Bernstein-von Mises theorem does not hold when $\alpha_n \sim \frac{1}{n}$. We show this explicitly for exponential families. Consider the statistical model

$$f_n(X^n|\eta) = h(X^n)\exp\left(\eta^\top T(X^n) - nA(\eta)\right) \tag{21}$$

and prior density

$$\pi(\eta) = \tilde{h}(\xi, \nu)\exp\left(\eta^\top \xi - \nu A(\eta) - \psi(\xi, \nu)\right), \tag{22}$$

where $\psi(\xi, \nu)$ and $\tilde{h}(\xi, \nu)$ are chosen such that (22) integrates to 1. The α_n-posterior has the form

$$\pi_{n,\alpha_n}(\eta|X^n) = \tilde{h}(\xi', \nu')\exp\left(\eta^\top \tilde{\xi} - \tilde{\nu}A(\eta) - \psi(\tilde{\xi}, \tilde{\nu})\right), \tag{23}$$

where

$$\begin{aligned}\tilde{\xi} &= \alpha_n T(X^n) + \xi\\ \tilde{\nu} &= \alpha_n n + \nu.\end{aligned}$$

Hence, (23) belongs to the same family as (22). Suppose that $\alpha_n n \to \alpha_0$ for some $\alpha_0 > 0$ and $\frac{1}{n}T(X^n)$ converges in $f_{0,n}$-probability to some limit $g(\eta^*)$, where η^* denotes the true (natural) parameter. Then the limit of the characteristic function of π_{n,α_n} is

$$\begin{aligned}\varphi_{\pi_{n,\alpha_n}}(t) &= \int_\Xi \exp\left(\eta^\top(it + \tilde{\xi}) - \tilde{\nu}A(\eta) - \psi(\tilde{\xi}, \tilde{\nu})\right)d\eta\\ &= \exp\left(\psi(it + \tilde{\xi}, \tilde{\nu}) - \psi(\tilde{\xi}, \tilde{\nu})\right)\\ &\to \exp\left(\psi(it + \xi', \nu') - \psi(\xi', \nu')\right) \coloneqq \varphi_\infty(t),\end{aligned} \tag{24}$$

where

$$\begin{aligned}\xi' &= \alpha_0 g(\eta^*) + \xi\\ \nu' &= \alpha_0 + \nu.\end{aligned}$$

We have therefore given a counterexample showing that $\alpha_n = o(1/n)$ is a necessary condition for establishing a Bernstein-von-Mises theorem. We have indeed established the following result.

Proposition 1. Suppose the model (21) is well specified. Then, if $\alpha_n n \to \alpha_0 > 0$, the α_n-posterior (23) converges weakly to a distribution belonging to the same family as (22).

Appendix C.2 complements the result with a discussion and some explicit limiting calculations for conjugate priors. We note that in the case where both the likelihood and prior are Gaussian, Proposition 1 holds, but Theorem 1 does not (see Appendix C.2 for details).

4.1.3 Sufficient conditions and extension to data-driven choice of α_n.

The proof of Theorem 1 closely follows the arguments of [4, Theorem 1], which, in turn, adapts the arguments of [40, Theorem 2.1]. Hence, Theorem 1 inherits the assumptions from [4] and [40] (assumptions **(A0)**–**(A3)** in this work), but we note key differences here.

First, our **(A2)** is a slightly different condition on the log-likelihood ratio than the $\sqrt{n}$-stochastic LAN assumption from [4] and [40]. We provide sufficient conditions that are easier to check in the following result.

Proposition 2. Suppose that in a neighborhood of θ^*, the log-likelihood function $\log f_n(X^n|\theta)$ is three-times continuously differentiable in θ and the entries of $\frac{1}{n}\nabla^3 \log f_n(X^n|\theta)$ satisfy

$$\max_{i_1,i_2,i_3} \left| \left\{ \frac{1}{n} \nabla^3 \log f_n(X^n|\theta) \right\}_{i_1,i_2,i_3=1}^{p} \right| < \frac{1}{n} \tilde{M}(X_n),$$

where $\mathbb{E}_{f_{0,n}} \left| \frac{1}{n} \tilde{M}(X_n) \right| < \infty$. Furthermore, suppose

$$\frac{1}{\sqrt{n}} \nabla \log f_n(X^n|\theta^*) - V_{\theta^*} \Delta_{n,\theta^*} \to 0 \tag{25}$$

and

$$-\frac{1}{n} \nabla^2 \log f_n(X^n|\theta^*) - V_{\theta^*} \to 0 \tag{26}$$

in $f_{0,n}$-probability. Then assumption **(A2)** holds.

The proof of this proposition is in Appendix C. Sufficient conditions to ensure (25) and (26) for i.i.d. data can be found in Lemma 2.1 and Lemma 2.2 in [40]. We note that the conditions in Proposition 2 are the usual regularity conditions that also imply $\sqrt{n}$-stochastic LAN ([59, Chapter 7]) and can be relaxed for i.i.d. data as shown in Lemmas 2.1-2.2 in [40].

We also note that our **(A3)** assumes a posterior concentration rate of $\sqrt{\alpha_n n}$ – which reflects the rate of our BvM result – as opposed to the $\sqrt{n}$ concentration rate assumed in [4] and [40]. This rate is a consequence of tempering the likelihood, which changes the effective number of samples from n samples to $\alpha_n n$ samples. Sufficient conditions for **(A3)** are discussed in Appendix A.3 of [48].

We conclude this section by noting that a Bernstein-von Mises result also holds for random, potentially data-dependent sequences that converge at the appropriate rates, as noted in the following corollary. Its proof can be found in Appendix B.1.

Corollary 1. Consider a random sequence $\hat{\alpha}_n$ satisfying the following

$$\begin{aligned} &\forall \epsilon > 0, \ \mathbb{P}_{f_{0,n}}(\hat{\alpha}_n > \epsilon) \to 0 \\ &\forall \epsilon > 0, \ \mathbb{P}_{f_{0,n}}(r_n \hat{\alpha}_n > \epsilon) \to 1 \end{aligned} \tag{27}$$

for some $r_n \to \infty$. Additionally, define the $\hat{\alpha}_n$-posterior

$$\pi_{n,\hat{\alpha}_n}(\theta|X^n) = \frac{f_n(X^n|\theta)^{\hat{\alpha}_n}\pi(\theta)}{\int_{\mathbb{R}^p} f_n(X^n|\theta)^{\hat{\alpha}_n}\pi(\theta)d\theta}. \tag{28}$$

Assume the conditions of Theorem 1 hold. Then, for $\hat{\alpha}_n$ satisfying (27),

$$\mathrm{d_{TV}}\left(\pi_{n,\hat{\alpha}_n}(\theta|X^n), \phi\left(\theta \,\middle|\, \hat{\theta}, \frac{1}{\hat{\alpha}_n n}V_{\theta^*}^{-1}\right)\right) \to 0 \tag{29}$$

in $f_{0,n}$-probability.

4.2 Convergence of α_n-posterior moments

4.2.1 Main result

Theorem 2 establishes that in the same asymptotic regime as that in Theorem 1, the difference between the k^{th} moment of the α_n-posterior and the k^{th} moment of the limiting Gaussian from Theorem 1 converges to 0 with growing sample size. Theorem 2 inherits assumptions **(A0)**–**(A2)** from Theorem 1 but requires the following assumption in lieu of assumption **(A3)**.

(A3') For all $n \in \mathbb{N}$, there exists a $\gamma > 0$ such that for some $k_0 \in \mathbb{N}$, the following is finite:

$$\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} \|h\|_2^{k_0(1+\gamma)}(\alpha_n n)^{-\frac{p}{2}}\pi_{n,\alpha_n}\left(\theta^* + \frac{h}{\sqrt{\alpha_n n}}\middle| X^n\right)dh\right].$$

Remark 1. Assumption **(A3')** is stronger than assumption **(A3)**. By Markov's inequality,

$$\begin{aligned}
&\mathbb{E}_{f_{0,n}}\left[\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(||\theta-\theta^*||_2 > C\epsilon_n)\right] && (30)\\
&= \mathbb{E}_{f_{0,n}}\left[\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(||\sqrt{\alpha_n n}(\theta-\theta^*)||_2 > C\sqrt{\alpha_n n}\epsilon_n)\right] && (31)\\
&= \mathbb{E}_{f_{0,n}}\left[\mathbb{P}_{(\alpha_n n)^{-p/2}\pi_{n,\alpha_n}(\theta^*+h/\sqrt{\alpha_n n}|X^n)}(||h||_2 > C\sqrt{\alpha_n n}\epsilon_n)\right]\\
&\leq (C\sqrt{\alpha_n n}\epsilon_n)^{-k_0(1+\gamma)}\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^{k_0(1+\gamma)}(\alpha_n n)^{-\frac{p}{2}}\pi_{n,\alpha_n}\left(\theta^*+\frac{h}{\sqrt{\alpha_n n}}\middle|X^n\right)dh\right]
\end{aligned}$$

Since $C\sqrt{\alpha_n n}\epsilon_n \to \infty$, **(A3)** holds (i.e., (30) converges to 0) if **(A3')** holds.

Theorem 2. Assume **(A0)**, **(A1)**, **(A2)** and **(A3')** hold. Then, for any sequence α_n such that $\frac{1}{n} \ll \alpha_n \ll 1$ and any integer $k \in [1, k_0]$,

$$\int_{\mathbb{R}^p}\left\|h^{\otimes k}\right\|_1\left|(\alpha_n n)^{-\frac{p}{2}}\pi_{n,\alpha_n}\left(\theta^*+\frac{h}{\sqrt{\alpha_n n}}|X^n\right) - (\alpha_n n)^{-\frac{p}{2}}\phi\left(h\middle|\sqrt{\alpha_n n}(\hat{\theta}-\theta^*), V_{\theta^*}^{-1}\right)\right|dh \to 0, \tag{32}$$

in $f_{0,n}$-probability, where

$$\left\|h^{\otimes k}\right\|_1 = \sum_{i_1=1}^{p}\cdots\sum_{i_k=1}^{p}|h_{i_1}\times\cdots\times h_{i_k}|, \tag{33}$$

is the 1-norm of a k^{th} order tensor.

This result holds for random sequences of α_n in the following corollary. The proof is given in Appendix B.2

Corollary 2. Assume Theorem 2 hold. Then for $\hat{\alpha}_n$ satisfying (27),

$$\int_{\mathbb{R}^p} \left\| h^{\otimes k} \right\|_1 \left| (\hat{\alpha}_n n)^{-\frac{p}{2}} \pi_{n,\hat{\alpha}_n} \left(\theta^* + \frac{h}{\sqrt{\hat{\alpha}_n n}} | X^n \right) - (\hat{\alpha}_n n)^{-\frac{p}{2}} \phi \left(h \middle| \sqrt{\hat{\alpha}_n n}(\hat{\theta} - \theta^*), V_{\theta^*}^{-1} \right) \right| dh \to 0, \tag{34}$$

in $f_{0,n}$-probability.

Remark 2. Note that as long as $k_0 \geq 2$, Theorem 2 shows that both of the following converge to zero in $f_{0,n}$-probability:

$$\int_{\mathbb{R}^p} \sqrt{\alpha_n n} \|\theta - \theta^*\|_1 \left| \pi_{n,\alpha_n}(\theta|X^n) - \phi\left(\theta \middle| \hat{\theta}, \frac{V_{\theta^*}^{-1}}{n\alpha_n} \right) \right| d\theta,$$

$$\int_{\mathbb{R}^p} \alpha_n n \|[\theta - \theta^*][\theta - \theta^*]^\top\|_1 \left| \pi_{n,\alpha_n}(\theta|X^n) - \phi\left(\theta \middle| \hat{\theta}, \frac{V_{\theta^*}^{-1}}{n\alpha_n} \right) \right| d\theta.$$

The full proof of Theorem 2 is in Appendix D.2 and very closely follows the proof of Theorem 1 in [54], which we simplify in this work. The proof of Theorem 2 follows a similar approach to the proof of Theorem 1 by arguing that computing the object in (32) also amounts to comparing (18) and (19), which follows from the finite moment condition, **(A3')**.

4.2.2 Discussion

The similarity between the proofs of Theorem 2 and Theorem 1 suggests that Theorem 2 should hold under the same conditions as those of Theorem 1. However, we note that neither this result, nor the analogous result in for the constant α case ([54, Theorem 1]) is a trivial consequence of a Bernstein-von Mises result. It seems plausible that Theorem 2 should follow immediately from Theorem 1 without additional conditions. However, this is not necessarily the case. Indeed, consider the following random variable

$$X_n \sim \begin{cases} \mathcal{N}(n, 1) & \text{w.p. } \frac{1}{n} \\ \mathcal{N}(0, 1) & \text{w.p. } 1 - \frac{1}{n}. \end{cases} \tag{35}$$

The distribution of (35) has a limit in total variation distance, but its first moment odoes not converge to the first moment of its total variation limit. Additionally, while (35) has a finite absolute first moment for all n, Assumption **(A3')** does not hold. We collect these results in the proposition below.

Proposition 3. Consider the random variable (35) and denote its density by f_{X_n}, Let Z be a standard normal random variable. Then, **(1)** $\mathrm{d}_{\mathrm{TV}}(f_{X_n}, \phi) \to 0$, **(2)** $\mathbb{E}[X_n] \not\to \mathbb{E}[Z]$, **(3)** $\mathbb{E}[|X_n|] < \infty$, $\forall n \in \mathbb{N}$, and **(4)** $\mathbb{E}\left[|X_n|^{1+\gamma}\right] \to \infty$ for $\gamma > 0$.

The proof of this result is in Appendix C.3.

4.3 Limiting Distribution of the α_n-posterior Mean Estimator

4.3.1 Main result

Theorem 3 shows that the posterior mean has the same limiting distribution as the maximum likelihood estimator any sequence, α_n, such that $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$. We will require the following assumptions.

(A1") The prior mean is finite,

$$\int_{\mathbb{R}^p} \|\theta\|_2 \pi(\theta) d\theta < \infty. \tag{36}$$

Furthermore, $\pi(\theta)$ is twice differentiable in a neighborhood of θ^*.

(A2") For all $\epsilon > 0$ and compact sets $K \subseteq \mathbb{R}^p$, there exist a matrix $H_n \in \mathbb{R}^{p\times p}$ that is positive definite with probability tending to 1, a tensor $S_n \in \mathbb{R}^{p\times p\times p}$, and $M > 0$ such that $\mathbb{P}_{f_{0,n}}\left(\left|\tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right| > \frac{M}{\alpha_n n}\|h\|_2^4\right) < \epsilon$ for all $h \in K$ and n sufficiently large, where

$$\begin{aligned} \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right) \equiv -\alpha_n n\left[-\frac{1}{n}\log f_n\left(X^n\middle|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\right) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] \\ - \left[-\frac{1}{2}h^\top H_n h + \frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle\right]. \end{aligned} \tag{37}$$

(A3") For all $\epsilon > 0$ and all $\delta > 0$, there exists $N \in \mathbb{N}$ such that for $n > N$,

$$\mathbb{P}_{f_{0,n}}\left(\inf_{B_{\hat{\theta}}(\delta)^{\mathsf{c}}}\left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] < 0\right) < \epsilon. \tag{38}$$

Theorem 3 is an analogue of Theorem 2 in [54], which establishes this posterior mean convergence result for the case where α is constant.

Theorem 3. Assume **(A0)**, **(A1") (A2")**, and **(A3")** hold. Then, for any sequence, α_n such that $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$, the α_n-posterior mean estimator $\hat{\theta}^B = \int_{\mathbb{R}^p} \theta \pi_{n,\alpha_n}(\theta|X^n)d\theta$ is asympotically equivalent to $\hat{\theta}$. That is, $\sqrt{n}\left(\hat{\theta}^B - \hat{\theta}\right) \to 0$ in $f_{0,n}$-probability.

This main result can be extended random sequences of α_n as shown in the following corollary. We relegate its proof to Appendix B.3.

Corollary 3. Assume Theorem 3 holds for (28) and a deterministic sequence, α_n, where $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$. Then for $\hat{\alpha}_n$ satisfying (27),

$$\sqrt{n}\left(\hat{\theta}^B - \hat{\theta}\right) \to 0 \tag{39}$$

in $f_{0,n}$-probability.

4.3.2 Discussion

According to the Bernstein-von Mises theorem for α-posteriors ([4, Theorem 1]) and α_n-posteriors (Theorem 1), raising the likelihood to a power does not affect the centering of the resulting posterior in the limit. Hence, intuitively the limiting distribution of the mean of the power posterior should be unaffected by the power. We note however that while Theorems 1 and 2 extend analogous results in the constant α case ([4, Theorem 1] and [54, Theorem 2], respectively) to the regime where $\frac{1}{n} \ll \alpha_n \ll 1$, Theorem 3 extends [54, Theorem 2] only to the restricted asymptotic regime $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$. To understand this discrepancy, we sketch the high-level argument of the Theorem

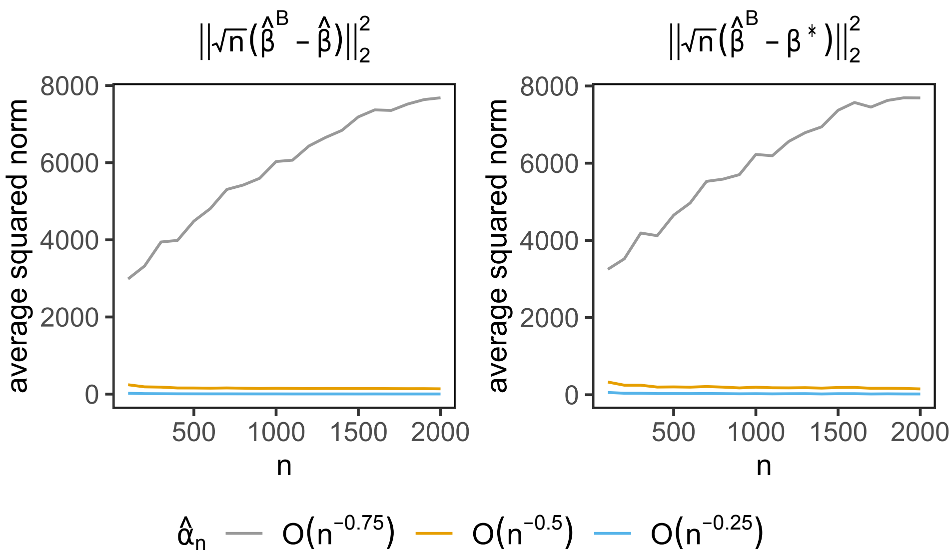


Figure 4: The squared norm of $\sqrt{n}(\hat{\beta}^B - \hat{\beta})$ (left) and $\sqrt{n}(\hat{\beta}^B - \beta^*)$ (right) – averaged over 100 replications – as a function of sample size. The α_n-posterior mean was obtained by samples from the α_n-posterior, where $\alpha_n = 0.5n^{-3/4}$ (gray), $\alpha_n = 0.5n^{-1/2}$ (orange), and $\alpha_n = 0.5n^{-1/4}$ (blue).

3 proof. To show that the MLE and the α_n-posterior mean are asymptotically equivalent, we characterize the difference between them and show

$$\hat{\theta}^B - \hat{\theta} = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right) \tag{40}$$

$$\implies \sqrt{n}\left(\hat{\theta}^B - \hat{\theta}\right) = O_{f_{0,n}}\left(\frac{1}{\alpha_n \sqrt{n}}\right) = o_{f_{0,n}}(1), \tag{41}$$

where the last equality requires we take $\alpha_n\sqrt{n} \to \infty$. The result in (40) provides an important insight into the impact of tempering on the limiting distribution of the α_n-posterior. Tempering induces a bias in the resulting posterior mean, which depends on the tempering parameter, α_n. However, this bias disappears asymptotically and does not affect the distribution of α_n-posterior mean in the limit.

For an example that cannot be worked out in closed form but meets the assumptions of Theorem 3, we demonstrate numerically that $\frac{1}{\sqrt{n}} \ll \alpha_n$ is the critical regime for the posterior mean to be asymptotically equivalent to the MLE in Figure 4. In our experiments, we generate data from and specify a logistic regression model (see Appendix A.6). We sample from the α_n-posterior under three settings of α_n: $\alpha_n \sim n^{-3/4}$, $\alpha_n \sim n^{-1/2}$, and $\alpha_n \sim n^{-1/4}$. We note that $\alpha_n \sim n^{-1/4}$ is in the "allowable" regime specified by Theorem 3, while $\alpha_n \sim n^{-3/4}$ falls outside this regime. Furthermore, $\alpha_n \sim n^{-1/2}$ is on the border of the critical regime. Indeed, we see in the left panel of Figure 4 that the $\sqrt{n}$-scaled difference between the posterior mean and the MLE diverges when $\alpha_n \sim n^{-3/4}$, vanishes when $\alpha_n \sim n^{-1/4}$, and is constant order when $\alpha_n \sim n^{-1/2}$. These behaviors are also reflected in the right panel of Figure 4, which shows the $\sqrt{n}$-rescaled difference between the α_n-posterior mean and the true parameter value.

To obtain the result in (40), we note that

$$\hat{\theta}^B = \frac{\int_{\mathbb{R}^p} \theta\pi(\theta) \exp\left(-\alpha_n n \left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right) d\theta}{\int_{\mathbb{R}^p} \pi(\theta) \exp\left(-\alpha_n n \left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right) d\theta} \tag{42}$$

is a ratio of two integrals of the form

$$\int_{\mathbb{R}^p} g(\theta) \exp\left(-\alpha_n n \left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right) d\theta. \tag{43}$$

We control (43) using a new Laplace approximation discussed in the next subsection.

Usually, given a moment convergence result like Theorem 2, Theorem 3 is a simple consequence such a result and requires only slight changes in proof technique (see [44, Theorems 8.2-8.3] for an example for standard posteriors). For the α-posterior with constant α, [54, Theorem 2] follows almost immediately from [54, Theorem 1]. However, this standard proof technique fails in the regime where $\alpha_n \to 0$ so we needed to derive the Laplace approximation stated next to get (40). To see why this is the case, consider, for a moment, the α-posterior, where α is constant. The proof of Theorem 2 in [54] shows that

$$\sqrt{n}(\hat{\theta}^B - \hat{\theta}) \leq \int_{\mathbb{R}^p} \|h\|_1 \left| n^{-\frac{p}{2}} \pi_{n,\alpha}(\theta^* + h/\sqrt{n}|X^n) - n^{-\frac{p}{2}} \phi\left(h \middle| \sqrt{n}(\hat{\theta} - \theta^*), (\alpha V_{\theta^*})^{-1}\right)\right| dh, \tag{44}$$

where (44) is $o_{f_{0,n}}(1)$ by [54, Theorem 1]. Examining the limiting Gaussian in (44) suggests that the density of $\sqrt{n}(\theta - \theta^*)$ has a second moment of order $1/\alpha$, which is constant order when α is constant but diverges when $\alpha \to 0$.

4.3.3 Laplace approximation

We use the following additional assumption to obtain a general Laplace approximation to integrals of the form (43). This is formalized in Lemma 1. This result is the main tool that allow us to establish the approximation (40) which is used in the proof of Theorem 3. Note that the specific form of Assumption $(A4'')$ needed in Theorem 3 is in fact guaranteed by Proposition 4.

(A4") Define $b(\theta) \coloneqq g(\theta)\pi(\theta)$, where b is assumed to be integrable and finite in a neighborhood of θ^*. For all $\epsilon > 0$ and compact sets $K \subseteq \mathbb{R}^p$, there exist a $v \in \mathbb{R}^p$ and $M > 0$ such that

$$\mathbb{P}_{f_{0,n}}\left(R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) > \frac{M}{\alpha_n n}\|h\|_2^2\right) < \epsilon \tag{45}$$

for all $h \in K$ and n sufficiently large, where

$$R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) = \left[b\left(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\right) - b(\hat{\theta})\right] - v^\top \frac{h}{\sqrt{\alpha_n n}}. \tag{46}$$

We are now ready to state our Laplace approximation result.

Lemma 1. Suppose assumptions **(A0)**, **(A1")**, **(A2")**, **(A3")**, and **(A4")** hold. Then, for any sequence, α_n such that $\alpha_n n \to \infty$, the following hold:

1. For each $n \in \mathbb{N}$, the integral

$$I = \int_{\mathbb{R}^p} b(\theta) \exp\left(\alpha_n \log f_n(X^n|\theta)\right) d\theta \tag{47}$$

is finite almost-surely. Furthermore, it can be written as

$$I = \exp\left(-\alpha_n n \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right) |H_n|^{-\frac{1}{2}} \left(\frac{2\pi}{\alpha_n n}\right)^{p/2} \left\{b(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)\right\}, \tag{48}$$

where H_n was defined in assumption **(A2")**.

2. The expectation of $g(\theta)$ with respect to the α-posterior satisfies the approximation

$$\frac{\int_{\mathbb{R}^p} g(\theta)\pi_{n,\alpha_n}(\theta|X^n)\pi(\theta)d\theta}{\int_{\mathbb{R}^p} \pi_{n,\alpha_n}(\theta|X^n)\pi(\theta)d\theta} = g(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right). \tag{49}$$

The proof of this result is in Appendix E.1. We also provide sufficient conditions for the high-probability control of $R_b(\cdot)$ in assumption **(A4")** (proof in Appendix C.4).

Proposition 4. Assume **(A0)** and **(A1")** hold and that in a neighborhood of θ^*, $g(\theta)$ is twice continuously differentiable with probability tending to 1 and finite. Then $b(\theta)$ is finite in a neighborhood of θ^*. Furthermore, the result (45) and the representation (46) in assumption **(A4")** hold.

The proof of Lemma 1 borrows techniques from the Laplace approximation literature. In the context of Theorem 3, we take "Laplace approximation" to mean an approximation of the expectation of a function with respect to a posterior distribution ([46, 45, 35, 34]) and not Bernstein-von Mises results ([38]) or approximations of the normalizing constant of the posterior ([7, 49]), although these are all interrelated concepts.

The proof of Lemma 1 follows most closely the proof of Theorem [34], who provide the most rigorous treatment of Laplace approximations compared to the heuristic arguments of previous works ([46, 45, 35]). However, we modify their results to accommodate α_n-posteriors and weaken the sense of convergence from almost-sure convergence to convergence in $f_{0,n}$-probability. We also adapt their approach to obtain a coarser approximation to the order $(\alpha_n n)^{-1}$ – as opposed to $(\alpha_n n)^{-2}$ – that is suitable for our purposes as shown in (40)–(41). As a result, we require 2 fewer derivatives of the prior (assumption **(A4")** and Proposition 4) and log-likelihood (assumption **(A2")** and Proposition 6). By relaxing the form of convergence to that in probability, we can also weaken the differentiability requirements on the prior to specify differentiability with high probability (see Proposition 4). Doing so allows us to apply our technique to prove the result in Proposition 7.

Finally, while our assumptions **(A1")** and **(A3")** look different from their analogous tail conditions in our previous two results (i.e., assumptions **(A3)** and **(A3')**), similar assumptions have appeared in the literature to show Bernstein-von Mises results and Laplace approximations (e.g., assumptions (B3) and (B5) pp. 488-489 in [44], condition (iii') pg. 479 in [34], assumption (2), pg. 5 in [49]). Furthermore, in the result, below, we show that a stronger version of assumption **(A1)** and a modified version of **(A3")** are sufficient conditions for the posterior concentration assumption **(A3)** (proof in Appendix C.5). We will introduce these assumptions here.

Proposition 5. Assume assumption **(A0)** holds. Furthermore, assume the following hold

(A1"') Define

$$\begin{aligned} K(\theta^*,\theta) &= \lim_{n\to\infty} \mathbb{E}_{f_{0,n}}\left[\frac{1}{n}\log\left(\frac{f_n(X^n|\theta)}{f_n(X^n|\theta^*)}\right)\right] \\ V(\theta^*,\theta) &= \lim_{n\to\infty} \mathrm{Var}_{f_{0,n}}\left(\frac{1}{\sqrt{n}}\log\left(\frac{f_n(X^n|\theta)}{f_n(X^n|\theta^*)}\right)\right) \\ G_n &= \{\theta\in\mathbb{R}^p : \max\{K(\theta^*,\theta), V(\theta^*,\theta)\} \le t_n\} \end{aligned} \tag{50}$$

There exists a positive sequence $t_n \to 0$ such that $\alpha_n n t_n \to \infty$ such that

$$\int_{G_n} \pi(\theta)d\theta \ge e^{-2\alpha_n n t_n}. \tag{51}$$

(A3"') For all $\epsilon > 0$, there exists a positive sequence $s_n \to 0$ such that $\alpha_n n s_n^2 \to \infty$, a constant $c_1 > 0$, and $N \in \mathbb{N}$ such that for $n > N$,

$$\mathbb{P}_{f_{0,n}}\left(\inf_{B_{\hat{\theta}}(s_n)^{\mathsf{c}}}\left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] < c_1 s_n^2\right) < \epsilon. \tag{52}$$

Then, assumption **(A3)** (posterior concentration) holds.

Assumption **(A1"')** is similar to Assumption 4 from [48], where we consider the loss function to be the negative log-likelihood. We note that assumption **(A1"')** is stronger than the prior mass assumption **(A1)** in the sense that it specifies the minimal prior mass required on a specific neighborhood of θ^*.

Assumption **(A3"')** is similar to Assumption 3 from [48], where we consider the loss function to be the negative log-likelihood and center the neighborhood around $\hat{\theta}$ instead of θ^* to match assumption **(A3")**. We note that assumption **(A3"')** is both stronger and weaker than **(A3")** in some sense. Since $s_n \to 0$, for any $\delta > 0$, we may pick N_0 large enough so that $s_n < \delta$. Then for $n > N_0$,

$$\begin{aligned}&\mathbb{P}_{f_{0,n}}\left(\inf_{B_{\hat{\theta}}(s_n)^{\mathsf{c}}}\left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] < c_1 s_n^2\right)\\ &\leq \mathbb{P}_{f_{0,n}}\left(\inf_{B_{\hat{\theta}}(\delta)^{\mathsf{c}}}\left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] < c_1 \delta^2\right).\end{aligned}$$

However, **(A3")** assumes the difference to be less than 0, which is a stronger assumption.

We also specify sufficient conditions for the representation in **(A2")** to hold in the fo(proof in Appendix C.6).

Proposition 6. Suppose that in a neighborhood of θ^*, the log-likelihood function $\log f_n(X^n|\theta)$ is four-times continuously differentiable in θ and the entries of $\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)$ satisfy

$$\max_{i_1,i_2,i_3,i_4}\left|\left\{\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)\right\}_{i_1,i_2,i_3,i_4=1}^{p}\right| < \tilde{M}(X_n),$$

where $\mathbb{E}_{f_{0,n}}\left|\tilde{M}(X_n)\right| < \infty$. Furthermore, suppose $\nabla \log f_n(X^n|\hat{\theta}) = 0$, and $\mathbb{P}_{f_{0,n}}(\lambda_{\min}\left(-\nabla^2 \frac{1}{n}\log f_n(X^n|\theta)\right) > 0) \to 1$ for all θ in a neighborhood of θ^*. Then assumption **(A2")** and the representation of the log-likelihood in (37) hold.

We note that to obtain the more precise approximation in Theorem 1 of [34], the expansion in (133) would include additional terms of order $(\alpha_n n)^{-3/2}$ and $(\alpha_n n)^{-2}$ and require 2 additional derivatives of $\log f_n(X^n|\theta)$.

We analyze the integral in **(A3')** using a similar technique as the proof of Lemma 1 and show that assumption **(A3')** does not hold. See Appendix C.7 for an informal analysis. Hence, Theorem 2 cannot be applied to the integral in (44).

4.4 Bernstein-von Mises Theorem for "mixed" α_n-posteriors

Some of the limiting distributions of $\hat{\alpha}_n$ suggested by Figure 1 are distributions with one point mass at $\alpha = \infty$ and remaining probability mass converging to 0 slower than $1/n$ (Figure 5). We formalize the limiting behavior of the resultant α-posterior in this section. We will first need to establish a key intermediate result formalizing the behavior of the α-posterior when $\alpha \to \infty$. Based on the limiting distribution of the α-posterior suggested by Theorem 1, the limit of the α-posterior as $\alpha \to \infty$ should be a point mass at $\hat{\theta}$. This has been claimed informally in the literature ([48]). Furthermore, a similar result was established for Gibbs measures in terms of weak convergence ([30]). We formalize this claim for α-posteriors in terms of total variation distance in the following result.

Proposition 7. Define the $\pi_{n,\infty}(\theta|X^n)$ posterior as

$$\lim_{\alpha\to\infty} \pi_{n,\alpha}(\theta|X^n). \tag{53}$$

Furthermore, assume **(A0)**, **(A1)**, **(A1")**, **(A2")**, and **(A3")** hold. Then,

$$\mathrm{d_{TV}}\left(\pi_{n,\infty}(\theta|X^n), \delta_{\hat{\theta}}\right) \to 0 \tag{54}$$

in $f_{0,n}$-probability, where $\delta_{\hat{\theta}}$ denotes a point mass at $\hat{\theta}$.

The proof of Proposition 7 is given in Appendix D.4. The main ingredient in the argument is the use of the Laplace approximation technique discussed in Section 4.3.3. The result can be extended to data-dependent $\hat{\alpha}_n$ as stated next. The proof can be found in Appendix B.4.

Corollary 4. Consider a random sequence $\hat{\alpha}_n$ satisfying the following

$$\forall M > 0,\ \mathbb{P}_{f_{0,n}}(\hat{\alpha}_n > M) = 1. \tag{55}$$

Furthermore, assume the conditions of Theorem 7 hold. Then

$$\mathrm{d_{TV}}\left(\pi_{n,\hat{\alpha}_n}(\theta|X^n), \delta_{\hat{\theta}}\right) \to 0 \tag{56}$$

in $f_{0,n}$-probability.

We are now ready to state our main result regarding data-driven α_n-posteriors where α_n follows a mixture distribution with a mass point at infinty. The proof of this result is in Appendix B.5.

Corollary 5. Assume the conditions of Proposition 5, Proposition 6, and Proposition 7 hold. Furthermore, suppose a random sequence, $\hat{\alpha}_n$, satisfies the following:

1. For all $\epsilon > 0$,
$$\mathbb{P}_{f_{0,n}}(\hat{\alpha}_n \le \epsilon,\ n\hat{\alpha}_n > \epsilon) \to p$$
2. For all $M > 0$,
$$\mathbb{P}_{f_{0,n}}(\hat{\alpha}_n > M) \to 1 - p.$$

Then

$$\mathrm{d_{TV}}\left(\pi_{n,\hat{\alpha}_n}(\theta|X^n), p\phi\left(\theta \,\middle|\, \hat{\theta}, \frac{1}{\hat{\eta}_n n} V_{\theta^*}^{-1}\right) + (1-p)\delta_{\hat{\theta}}\right) \to 0 \tag{57}$$

in $f_{0,n}$-probability, where $\mathbb{P}_{f_{0,n}}(\hat{\eta}_n \le \epsilon,\ n\hat{\eta}_n > \epsilon) \to 1$.

5 Discussion

In this work, we have studied the asymptotic behavior of $\hat{\alpha}_n$-posteriors where $\hat{\alpha}_n = \infty$ with some probability and is a sequence that tends to 0 otherwise. This analysis was motivated by numerical results in cross-validation based selection methods of α that suggest this regime. We discuss these points in more detail in Section 3.

In the setting where $\alpha_n \to 0$, we establish asymptotic normality of α_n-posteriors (Theorem 1); consistency of α_n-posterior moments (Theorem 2); and asymptotic normality of the α_n-posterior mean (Theorem 3). We also study the limiting distribution of the $\hat{\alpha}_n$-posterior under the regime suggested by data-driven selection methods for α in Corollary 5.

This are several avenues for future work. First, we only analyzed the mean of the α_n-posterior but not other point estimators derived from posteriors such as the posterior mean and mode. We may be able to use the Laplace approximation we developed (Lemma 1) to do so, but we leave this for future work. It would also be nice to investigate the asymptotic behavior of α-posteriors in parametric, high-dimensional settings (i.e., where the dimension, p, diverges as well as the sample size, n). Asymptotic normality results for high-dimensional Bayesian linear regression were established in [20, 11]. Furthermore, sufficient conditions for and asymptotic regimes for which valid Laplace approximations exist are given in [36, 38, 37]. Another potential technique is to determine the behavior of variational approximations to conclude behavior of the posterior ([51]). These works may give similar insights to establish similar guarantees in high-dimensional asymptotic regimes.

References

[1] David M. Allen. "Mean Square Error of Prediction as a Criterion for Selecting Variables". In: *Technometrics* 13.3 (1971), pp. 469–475. DOI: 10.2307/1267161.

[2] Pierre Alquier and James Ridgway. "Concentration of tempered posteriors and of their variational approximations". In: *The Annals of Statistics* 48.3 (June 2020), pp. 1475–1497. DOI: 10.1214/19-AOS1855.

[3] Gholamali Aminian et al. "An Exact Characterization of the Generalization Error for the Gibbs Algorithm". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 8106–8118.

[4] Marco Avella Medina et al. "On the Robustness to Misspecification of alpha-posteriors and Their Variational Approximations". In: *Journal of Machine Learning Research* 23.147 (2022), pp. 1–51.

[5] Yannick Baraud. *From robust tests to Bayes-like posterior distributions*. Jan. 2024. DOI: 10.48550/arXiv.2107.12011.

[6] A.R. Barron and T.M. Cover. "Minimum complexity density estimation". In: *IEEE Transactions on Information Theory* 37.4 (July 1991), pp. 1034–1054. DOI: 10.1109/18.86996.

[7] Anirban Bhattacharya and Debdeep Pati. *Nonasymptotic Laplace approximation under model misspecification*. June 2020. DOI: 10.48550/arXiv.2005.07844.

[8] Anirban Bhattacharya, Debdeep Pati, and Yun Yang. "Bayesian fractional posteriors". In: *The Annals of Statistics* 47.1 (Feb. 2019), pp. 39–66. DOI: 10.1214/18-AOS1712.

[9] P. G. Bissiri, C. C. Holmes, and S. G. Walker. "A general framework for updating belief distributions". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 78.5 (2016). Publisher: [Royal Statistical Society, Wiley], pp. 1103–1130.

[10] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. “Variational Inference: A Review for Statisticians”. In: *Journal of the American Statistical Association* 112.518 (Apr. 2017), pp. 859–877. ISSN: 0162-1459. DOI: 10.1080/01621459.2017.1285773.

[11] Dominique Bontemps. “Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors”. In: *The Annals of Statistics* 39.5 (Oct. 2011), pp. 2557–2584. DOI: 10.1214/11-AOS912.

[12] J. Borwanker, G. Kallianpur, and B. L. S. Prakasa Rao. “The Bernstein-Von Mises Theorem for Markov Processes”. In: *The Annals of Mathematical Statistics* 42.4 (Aug. 1971), pp. 1241–1253. DOI: 10.1214/aoms/1177693237.

[13] Yuheng Bu. “Towards Optimal Inverse Temperature in the Gibbs Algorithm”. In: *2024 IEEE International Symposium on Information Theory (ISIT)*. July 2024, pp. 2257–2262. DOI: 10.1109/ISIT57864.2024.10619441.

[14] Christopher P. Burgess et al. *Understanding disentangling in \$\beta\$-VAE*. Apr. 2018. DOI: 10.48550/arXiv.1804.03599.

[15] Luiz Max Carvalho and Joseph G. Ibrahim. “On the normalized power prior”. In: *Statistics in Medicine* 40.24 (2021), pp. 5251–5275. ISSN: 1097-0258. DOI: 10.1002/sim.9124.

[16] Ismaël Castillo and Judith Rousseau. “A Bernstein-von Mises theorem for smooth functionals in semiparametric models”. In: *The Annals of Statistics* 43.6 (Dec. 2015). DOI: 10.1214/15-AOS1336.

[17] Victor Chernozhukov and Han Hong. “An MCMC approach to classical estimation”. In: *Journal of Econometrics* 115.2 (Aug. 2003), pp. 293–346. DOI: 10.1016/S0304-4076(03)00100-3.

[18] A. Dalalyan and A. B. Tsybakov. “Aggregation by exponential weighting, sharp PAC-Bayesian bounds and sparsity”. In: *Machine Learning* 72.1 (Aug. 2008), pp. 39–61. DOI: 10.1007/s10994-008-5051-0.

[19] Andrew Gelman, Jessica Hwang, and Aki Vehtari. “Understanding predictive information criteria for Bayesian models”. en. In: *Statistics and Computing* 24.6 (Nov. 2014), pp. 997–1016. DOI: 10.1007/s11222-013-9416-2.

[20] Subhashis Ghosal. “Asymptotic normality of posterior distributions in high-dimensional linear models”. In: *Bernoulli* 5.2 (Apr. 1999). Publisher: Bernoulli Society for Mathematical Statistics and Probability, pp. 315–331.

[21] Subhashis Ghosal and Aad van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2017. ISBN: 978-0-521-87826-5.

[22] Abhik Ghosh and Ayanendranath Basu. “Robust Bayes estimation using the density power divergence”. en. In: *Annals of the Institute of Statistical Mathematics* 68.2 (Apr. 2016), pp. 413–437. DOI: 10.1007/s10463-014-0499-0.

[23] Peter Grünwald. “Safe Learning: bridging the gap between Bayes, MDL and statistical learning theory via empirical convexity”. In: *Proceedings of the 24th Annual Conference on Learning Theory*. JMLR Workshop and Conference Proceedings, Dec. 2011, pp. 397–420.

[24] Peter Grünwald. “The Safe Bayesian”. In: *Algorithmic Learning Theory*. Ed. by Nader H. Bshouty et al. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 169–183. DOI: 10.1007/978-3-642-34106-9_16.

[25] Peter Grünwald and Thijs van Ommen. "Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It". In: *Bayesian Analysis* 12.4 (Dec. 2017), pp. 1069–1103. DOI: 10.1214/17-BA1085.

[26] Rianne de Heide et al. *SafeBayes: Generalized and Safe-Bayesian Ridge and Lasso Regression.* Oct. 2016. URL: https://CRAN.R-project.org/package=SafeBayes (visited on 01/19/2025).

[27] Irina Higgins et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: Nov. 2016.

[28] C. C. Holmes and S. G. Walker. "Assigning a value to a power likelihood in a general Bayesian model". In: *Biometrika* 104.2 (June 2017), pp. 497–503. DOI: 10.1093/biomet/asx010.

[29] Chin-Wei Huang et al. "Improving Explorability in Variational Inference with Annealed Variational Objectives". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018.

[30] Chii-Ruey Hwang. "Laplace's Method Revisited: Weak Convergence of Probability Measures". In: *The Annals of Probability* 8.6 (Dec. 1980), pp. 1177–1182. DOI: 10.1214/aop/1176994579.

[31] Joseph G. Ibrahim and Ming-Hui Chen. "Power Prior Distributions for Regression Models". In: *Statistical Science* 15.1 (2000), pp. 46–60.

[32] Joseph G. Ibrahim et al. "The Power Prior: Theory and Applications". In: *Statistics in medicine* 34.28 (Dec. 2015), pp. 3724–3749. DOI: 10.1002/sim.6728.

[33] Michael I. Jordan et al. "An Introduction to Variational Methods for Graphical Models". In: *Machine Learning* 37.2 (Nov. 1999), pp. 183–233. DOI: 10.1023/A:1007665907178.

[34] R. Kass, L. Tierney, and J. Kadane. "The validity of posterior expansions based on Laplace's method". In: 1990.

[35] Robert E. Kass and Duane Steffey. "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)". In: *Journal of the American Statistical Association* 84.407 (1989), pp. 717–726. DOI: 10.2307/2289653.

[36] Anya Katsevich. *Improved Scaling with Dimension in the Bernstein-von Mises Theorem for Two Statistical Models.* Aug. 2023. DOI: 10.48550/arXiv.2308.06899.

[37] Anya Katsevich. *Tight Bounds on the Laplace Approximation Accuracy in High Dimensions.* Aug. 2023. DOI: 10.48550/arXiv.2305.17604.

[38] Anya Katsevich. *Tight skew adjustment to the Laplace approximation in high dimensions.* Aug. 2023. DOI: 10.48550/arXiv.2306.07262.

[39] Christian Kleiber and Achim Zeileis. *Applied Econometrics with R.* New York, NY: Springer, 2008. ISBN: 978-0-387-77316-2 978-0-387-77318-6.

[40] B. J. K. Kleijn and A. W. van der Vaart. "The Bernstein-Von-Mises theorem under misspecification". In: *Electronic Journal of Statistics* 6.none (Jan. 2012), pp. 354–381. DOI: 10.1214/12-EJS675.

[41] Jeremias Knoblauch, Jack Jewson, and Theodoros Damoulas. *Generalized Variational Inference: Three arguments for deriving new Posteriors.* Dec. 2019. DOI: 10.48550/arXiv.1904.02063.

[42] Alice L'Huillier et al. "Semiparametric Inference Using Fractional Posteriors". In: *Journal of Machine Learning Research* 24.389 (2023), pp. 1–61.

[43] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Berlin, Heidelberg: Springer, 1991. ISBN: 978-3-642-20211-7 978-3-642-20212-4.

[44] "Asymptotic Optimality". In: *Theory of Point Estimation*. Ed. by E. L. Lehmann and George Casella. Springer Texts in Statistics. New York, NY: Springer, 1998, pp. 429–519. DOI: 10.1007/0-387-22728-8_6.

[45] D. V. Lindley. "Approximate Bayesian methods". In: *Trabajos de Estadistica Y de Investigacion Operativa* 31.1 (Feb. 1980), pp. 223–245. DOI: 10.1007/BF02888353.

[46] D. V. Lindley. "The Use of Prior Probability Distributions in Statistical Inference and Decisions". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Vol. 4.1. University of California Press, Jan. 1961, pp. 453–469.

[47] David A Mcallester. "PAC-Bayesian Stochastic Model Selection". In: (2003).

[48] Yann McLatchie et al. *Predictive performance of power posteriors*. Aug. 2024. DOI: 10.48550/arXiv.2408.08806.

[49] Jeffrey W. Miller. "Asymptotic Normality, Concentration, and Coverage of Generalized Posteriors". In: *Journal of Machine Learning Research* 22.168 (2021), pp. 1–53. ISSN: 1533-7928.

[50] Jeffrey W. Miller and David B. Dunson. "Robust Bayesian Inference via Coarsening". In: *Journal of the American Statistical Association* 114.527 (July 2019), pp. 1113–1125. DOI: 10.1080/01621459.2018.1469995.

[51] Sumit Mukherjee and Subhabrata Sen. "Variational inference in high-dimensional linear regression". In: *The Journal of Machine Learning Research* 23.1 (Jan. 2022), 304:13703–304:13758.

[52] Anthony O'Hagan. "Fractional Bayes Factors for Model Comparison". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1 (1995), pp. 99–138.

[53] Lucie Perrotta. *Practical calibration of the temperature parameter in Gibbs posteriors*. Apr. 2020. DOI: 10.48550/arXiv.2004.10522.

[54] Ruchira Ray, Marco Avella Medina, and Cynthia Rush. "Asymptotics for power posterior mean estimation". In: *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Sept. 2023, pp. 1–8. DOI: 10.1109/Allerton58177.2023.10313460.

[55] Philippe Rigollet and Alexandre B. Tsybakov. "Sparse Estimation by Exponential Weighting". In: *Statistical Science* 27.4 (Nov. 2012), pp. 558–575. DOI: 10.1214/12-STS393.

[56] Nandana Sengupta and Fallaw Sowell. "On the Asymptotic Distribution of Ridge Regression Estimators Using Training and Test Samples". In: *Econometrics* 8.4 (Dec. 2020), p. 39. DOI: 10.3390/econometrics8040039.

[57] Nicholas Syring and Ryan Martin. "Calibrating general posterior credible regions". In: *Biometrika* 106.2 (June 2019), pp. 479–486. DOI: 10.1093/biomet/asy054.

[58] Nicholas Syring and Ryan Martin. "Gibbs posterior concentration rates under sub-exponential type losses". In: *Bernoulli* 29.2 (May 2023), pp. 1080–1108. DOI: 10.3150/22-BEJ1491.

[59] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1998. ISBN: 978-0-521-78450-4.

[60] Aki Vehtari, Andrew Gelman, and Jonah Gabry. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and Computing* 27.5 (Sept. 2017), pp. 1413–1432. DOI: 10.1007/s11222-016-9696-4.

[61] Volodimir G. Vovk. "Aggregating strategies". In: *Proceedings of the third annual workshop on Computational learning theory*. COLT '90. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., July 1990, pp. 371–386.

[62] Martin J. Wainwright and Michael I. Jordan. "Graphical Models, Exponential Families, and Variational Inference". English. In: *Foundations and Trends® in Machine Learning* 1.1–2 (Nov. 2008), pp. 1–305. DOI: 10.1561/2200000001.

[63] Yixin Wang, Alp Kucukelbir, and David M. Blei. "Robust probabilistic modeling with Bayesian data reweighting". In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. ICML'17. Sydney, NSW, Australia: JMLR.org, Aug. 2017, pp. 3646–3655.

[64] Pei-Shien Wu and Ryan Martin. "A Comparison of Learning Rate Selection Methods in Generalized Bayesian Inference". In: *Bayesian Analysis* 18.1 (Mar. 2023), pp. 105–132. DOI: 10.1214/21-BA1302.

[65] Pei-Shien Wu and Ryan Martin. *Calibrating generalized predictive distributions*. July 2021.

[66] Aolin Xu and Maxim Raginsky. "Information-theoretic analysis of generalization capability of learning algorithms". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.

[67] Yun Yang, Debdeep Pati, and Anirban Bhattacharya. "$\alpha $-variational inference with statistical guarantees". In: *The Annals of Statistics* 48.2 (Apr. 2020), pp. 886–905. DOI: 10.1214/19-AOS1827.

[68] Yuling Yao and Aki Vehtari. "Make cross-validation Bayes again". In: ().

[69] Tong Zhang. "From epsilon-Entropy to KL-Entropy: Analysis of Minimum Information Complexity Density Estimation". In: *Ann. Stat.* 34.5 (2006), pp. 2180–2210.

[70] Tong Zhang. "Information Theoretical Upper and Lower Bounds for Statistical Estimation". In: *IEEE Trans. Inf. Theory* 52.4 (2006), pp. 1307–1321.

A Additional Experimental Details

A.1 α-posterior and GMF variational approximation calculation

Given design matrix $X \in \mathbb{R}^{n\times p}$ and $\beta \in \mathbb{R}^p$, we consider the linear model

$$Y = X\beta + \epsilon, \tag{58}$$

where $\epsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$. This gives rise to the likelihood $\ell(X, Y|\beta) \sim \prod_{i=1}^n \mathcal{N}(y_i - x_i^\top \beta, \sigma^2)$. Now, putting a prior on β, with $\pi(\beta) \sim \mathcal{N}(\mu_0, \Sigma_0)$, we calculate the α-posterior

$$\pi_{n,\alpha}(\beta|X,Y) \propto \ell(X,Y|\beta)^\alpha \pi(\beta) = \left[\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(y_i - x_i^\top\beta\right)^2}\right]^\alpha \frac{1}{\sqrt{(2\pi)^p|\Sigma_0|}} e^{-\frac{1}{2}(\beta-\mu_0)^\top\Sigma_0^{-1}(\beta-\mu_0)}. \tag{59}$$

Simplifying (59), we find

$$
\begin{aligned}
\pi_{n,\alpha}(\beta|X,Y) &\propto e^{-\frac{\alpha}{2\sigma^2}\sum_{i=1}^n \left(y_i - x_i^\top \beta\right)^2 - \frac{1}{2}(\beta-\mu_0)^\top \Sigma_0^{-1}(\beta-\mu_0)} \\
&\propto e^{\frac{\alpha}{\sigma^2}\sum_{i=1}^n \left[y_i x_i^\top \beta - \frac{1}{2}(x_i^\top \beta)^2\right] - \frac{1}{2}\beta^\top \Sigma_0^{-1}\beta + \beta^\top \Sigma_0^{-1}\mu_0} \\
&\propto e^{-\frac{1}{2}\beta^\top \left(\frac{\alpha}{\sigma^2}\sum_{i=1}^n x_i x_i^\top\right)\beta - \frac{1}{2}\beta^\top \Sigma_0^{-1}\beta + \beta^\top \left(\frac{\alpha}{\sigma^2}\sum_{i=1}^n y_i x_i\right) + \beta^\top \Sigma_0^{-1}\mu_0} \\
&\propto e^{-\frac{1}{2}\left[\beta^\top \left(\frac{\alpha}{\sigma^2}X^\top X + \Sigma_0^{-1}\right)\beta - 2\beta^\top \left(\frac{\alpha}{\sigma^2}X^\top Y + \Sigma_0^{-1}\mu_0\right)\right]} \\
&\propto e^{-\frac{1}{2}(\beta-\mu_n)^\top \Sigma_n^{-1}(\beta-\mu_n)},
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma_n &= \left(\frac{\alpha}{\sigma^2}X^\top X + \Sigma_0^{-1}\right)^{-1} = \frac{\sigma^2}{\alpha n}\left(\frac{1}{n}X^\top X + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1} \\
\mu_n &= \left(\frac{\alpha}{\sigma^2}X^\top X + \Sigma_0^{-1}\right)^{-1}\left(\frac{\alpha}{\sigma^2}X^\top Y + \Sigma_0^{-1}\mu_0\right) = \left(\frac{1}{n}X^\top X + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left(\frac{1}{n}X^\top Y + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right).
\end{aligned}
$$

Suppose $\Sigma_0 = \sigma^2 I_p$ and $\mu_0 = \mathbf{0}_p$. Setting $\alpha_n = \frac{1}{n\lambda_n}$, the ridge estimator is the posterior mean:

$$
\Sigma_n = \lambda_n \sigma^2 \left(\frac{1}{n}X^\top X + \lambda_n I_p\right)^{-1} \quad \text{and} \quad \mu_n = \left(\frac{1}{n}X^\top X + \lambda_n I_p\right)^{-1}\left(\frac{1}{n}X^\top Y\right). \tag{60}
$$

We obtain the Gaussian mean-field variational approximation of (60) as follows. Considering the KL-divergence between two multivariate Gaussians, we can write the Gaussian mean-field variational inference problem in (5) as

$$
\begin{aligned}
&\arg\min_{\mu_0,\Sigma_0} d_{\mathrm{KL}}\left(\phi(\cdot|\mu_0,\Sigma_0)||\phi(\cdot|\mu_1,\Sigma_1)\right) && (61)\\
&= \arg\min_{\mu_0,\sigma_1^2,\dots,\sigma_p^2} d_{\mathrm{KL}}\left(\phi(\cdot|\mu_0,\mathrm{diag}(\sigma_1^2,\dots\sigma_p^2))||\phi(\cdot|\mu_1,\Sigma_1)\right) && (62)\\
&= \arg\min_{\mu_0,\sigma_1^2,\dots,\sigma_p^2} \frac{1}{2}\Bigg[\mathrm{trace}(\Sigma_1^{-1}\mathrm{diag}(\sigma_1^2,\dots\sigma_p^2)) - p + (\mu_1-\mu_0)^\top \Sigma_1^{-1}(\mu_1-\mu_0) && (63)\\
&\qquad + \log\left(\frac{|\Sigma_1|}{|\mathrm{diag}(\sigma_1^2,\dots\sigma_p^2)|}\right)\Bigg]. && (64)
\end{aligned}
$$

First, since Σ_1 is positive definite, (61) is minimized with respect to μ_0 by setting $\mu_0 = \mu_1$. Substituting this into (61) and eliminating terms that do not depend on $\sigma_1^2,\dots,\sigma_p^2$, we can rewrite the optimization problem in (61) as

$$
\begin{aligned}
&\arg\min_{\sigma_1^2,\dots,\sigma_p^2} \mathrm{trace}(\Sigma_1^{-1}\mathrm{diag}(\sigma_1^2,\dots\sigma_p^2)) - \log\left(|\mathrm{diag}(\sigma_1^2,\dots\sigma_p^2)|\right) && (65)\\
&= \arg\min_{\sigma_1^2,\dots,\sigma_p^2} \sum_{j=1}^p \sigma_j^2\left(\Sigma_1^{-1}\right)_{jj} - \log(\sigma_j^2). && (66)
\end{aligned}
$$

Minimizing (65) with respect to $\sigma_1^2,\dots,\sigma_p^2$, we find that

$$
\sigma_j^2 = ((\Sigma_1^{-1})_{jj})^{-1}. \tag{67}
$$

Hence, the density of the Gaussian mean-field variational approximation of (8) is (12).

A.2 $\overline{\text{lppd}}$ calculation

Recall the definition of leave-one-out $\overline{\text{lppd}}$,

$$\overline{\text{lppd}}_{\text{LOO}} = \frac{1}{n}\sum_{i=1}^{n} \log \int p\left(x_i, y_i|\beta\right) \pi_{n,\alpha}\left(\beta|X_{-i}Y_{-i}\right) d\beta, \tag{68}$$

where X_{-i}, Y_{-i} denote the design matrix and dependent variable, respectively, with the ith entry removed. Substituting the calculations from Section A.1, we have

$$\overline{\text{lppd}}_{\text{LOO}} = \frac{1}{n}\sum_{i=1}^{n} \log \int \frac{1}{\sqrt{2\pi\sigma^2}}(2\pi)^{-\frac{p}{2}}|\Sigma_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}A_n\right) d\beta, \tag{69}$$

where

$$A_n = \left(\frac{1}{\sigma^2}(y_i - x_i^\top\beta)^2 + (\beta-\mu_n)^\top\Sigma_n^{-1}(\beta-\mu_n)\right).$$

We compute A_n as follows

$$\begin{aligned}
A_n &= \frac{1}{\sigma^2}(y_i - x_i^\top\beta)^2 + (\beta-\mu_n)^\top\Sigma_n^{-1}(\beta-\mu_n) \\
&= \frac{1}{\sigma^2}y_i^2 - \frac{2}{\sigma^2}y_i x_i^\top\beta + \frac{1}{\sigma^2}(x_i^\top\beta)^2 + \beta^\top\Sigma_n^{-1}\beta - 2\beta^\top\Sigma_n^{-1}\mu_n + \mu_n^\top\Sigma_n^{-1}\mu_n \\
&= \left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n\right] + \frac{1}{\sigma^2}\beta^\top x_i x_i^\top\beta + \beta^\top\Sigma_n^{-1}\beta - \frac{2}{\sigma^2}\beta^\top x_i y_i - 2\beta^\top\Sigma_n^{-1}\mu_n \\
&= \left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n\right] + \beta^\top\left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)\beta - 2\beta^\top\left(\frac{1}{\sigma^2}x_i y_i - \Sigma_n^{-1}\mu_n\right) \\
&= (\beta - m_n)^\top\left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)(\beta - m_n) \\
&\quad - \left(\frac{1}{\sigma^2}x_i y_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)\left(\frac{1}{\sigma^2}x_i y_i + \Sigma_n^{-1}\mu_n\right) + \left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n\right],
\end{aligned} \tag{70}$$

where

$$m_n = \left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_i y_i + \Sigma_n^{-1}\mu_n\right)$$

We set $B_n = \exp\left(-\frac{1}{2}A_n\right)$ and obtain,

$$\begin{aligned}
B_n &= \exp\left(-\frac{1}{2}A_n\right) \\
&= \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n - \left(\frac{1}{\sigma^2}x_i y_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_i y_i + \Sigma_n^{-1}\mu_n\right)\right]\right) \\
&\quad \times \exp\left(-\frac{1}{2}\left(\beta - \tilde{\mu}_n\right)^\top\tilde{\Sigma}_n^{-1}\left(\beta - \tilde{\mu}_n\right)\right),
\end{aligned} \tag{71}$$

where

$$\begin{aligned}
\tilde{\Sigma}_n &= \left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)^{-1} \\
\tilde{\mu}_n &= \left(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_i y_i + \Sigma_n^{-1}\mu_n\right).
\end{aligned} \tag{72}$$

We set $C_n = \frac{1}{\sqrt{2\pi\sigma^2}}(2\pi)^{-\frac{p}{2}}|\Sigma_n|^{-\frac{1}{2}} \int B_n d\beta$ and compute,

$$
\begin{aligned}
&\frac{C_n}{(2\pi)^{-\frac{p+1}{2}}|\sigma^2\Sigma_n|^{-\frac{1}{2}}} \\
&= \int B_n d\beta \\
&= \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n - \left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_ix_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)\right]\right) \\
&\quad\times \int \exp\left(-\frac{1}{2}\left(\beta - \tilde{\mu}_n\right)^\top\tilde{\Sigma}_n^{-1}\left(\beta - \tilde{\mu}_n\right)\right) d\beta \\
&= \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n - \left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_ix_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)\right]\right) \\
&\quad\times (2\pi)^{\frac{p}{2}}|\tilde{\Sigma}_n|^{\frac{1}{2}} \int (2\pi)^{-\frac{p}{2}}|\tilde{\Sigma}_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\beta - \tilde{\mu}_n\right)^\top\tilde{\Sigma}_n^{-1}\left(\beta - \tilde{\mu}_n\right)\right) d\beta.
\end{aligned}
\tag{73}
$$

Hence,

$$
C_n = (2\pi)^{-\frac{1}{2}}|\sigma^2\Sigma_n\tilde{\Sigma}_n^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n - v_n\right]\right), \tag{74}
$$

where

$$
v_n = \left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_ix_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right).
$$

We set $D_n = \log(C_n)$ and obtain

$$
\begin{aligned}
D_n &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log|\sigma^2\Sigma_n\tilde{\Sigma}_n^{-1}| \\
&\quad - \frac{1}{2}\left[\frac{1}{\sigma^2}y_i^2 + \mu_n^\top\Sigma_n^{-1}\mu_n - \left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_ix_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)\right] \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log|\sigma^2\Sigma_n\tilde{\Sigma}_n^{-1}| \\
&\quad - \frac{1}{2\sigma^2}y_i^2 - \frac{1}{2}\mu_n^\top\Sigma_n^{-1}\mu_n + \frac{1}{2}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_ix_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right).
\end{aligned}
\tag{75}
$$

Dropping terms that are constant in α (and, thus, do not matter in the optimization of $\overline{\mathrm{lppd}}_{\mathrm{LOO}}$), we find

$$
\begin{aligned}
D_n &\sim -\frac{1}{2}\log|\sigma^2\Sigma_n\tilde{\Sigma}_n^{-1}| - \frac{1}{2}\mu_n^\top\Sigma_n^{-1}\mu_n \\
&\quad + \frac{1}{2}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right)^\top\left(\frac{1}{\sigma^2}x_ix_i^\top + \Sigma_n^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}x_iy_i + \Sigma_n^{-1}\mu_n\right).
\end{aligned}
\tag{76}
$$

Finally, we substitute

$$\begin{aligned}
\tilde{\Sigma}_n &= \left(\frac{1}{\sigma^2} x_i x_i^\top + \Sigma_n^{-1}\right)^{-1} \\
\Sigma_n &= \frac{\sigma^2}{\alpha n}\left(\frac{1}{n-1} X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1} \\
\mu_n &= \left(\frac{1}{n-1} X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left(\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right).
\end{aligned} \tag{77}$$

Hence,

$$\begin{aligned}
\sigma^2 \Sigma_n \tilde{\Sigma}_n^{-1} &= \sigma^2 \Sigma_n \left(\frac{1}{\sigma^2} x_i x_i^\top + \Sigma_n^{-1}\right) = \Sigma_n x_i x_i^\top + \sigma^2 I \\
\mu_n^\top \Sigma_n^{-1} \mu_n &= \frac{\alpha n}{\sigma^2}\left(\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right)^\top \left(\frac{1}{n-1} X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left(\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right) \\
\Sigma_n^{-1}\mu_n &= \frac{\alpha n}{\sigma^2}\left(\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right).
\end{aligned} \tag{78}$$

Substituting (78) into (75) gives us $\overline{\text{lppd}}_{\text{LOO}}$ up to α-dependent constants

$$\overline{\text{lppd}}_{\text{LOO}} \sim T_1 + T_2 + T_3 \tag{79}$$

where

$$T_1 = -\frac{1}{2n}\sum_{i=1}^n \log\left|\frac{\sigma^2}{\alpha n}\left(\frac{1}{n-1} X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1} x_i x_i^\top + \sigma^2 I\right|, \tag{80}$$

$$\begin{aligned}
T_2 = -\frac{1}{2n}\sum_{i=1}^n \frac{\alpha n}{\sigma^2}&\left[\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right]^\top \\
&\left(\frac{1}{n-1} X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left[\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right],
\end{aligned} \tag{81}$$

and

$$\begin{aligned}
T_3 = \frac{1}{2n}\sum_{i=1}^n &\left[\frac{1}{\sigma^2} x_i y_i + \frac{\alpha n}{\sigma^2}\left(\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right)\right]^\top \\
&\left(\frac{1}{\sigma^2} x_i x_i^\top + \frac{\alpha n}{\sigma^2}\left(\frac{1}{n-1} X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)\right)^{-1} \\
&\left[\frac{1}{\sigma^2} x_i y_i + \frac{\alpha n}{\sigma^2}\left(\frac{1}{n-1} X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right)\right]
\end{aligned} \tag{82}$$

In our numerical experiments, we set $\mu_0 = \mathbf{0}_p$ and $\Sigma_0 = I_p$, and $\alpha = \frac{1}{n\lambda}$.

A.3 $\overline{\text{lppd}}_{\text{VI-LOO}}$ calculation

We also consider the $\overline{\text{elpd}}$, where density of the Gaussian mean field variational approximation of (8) is substituted in the calculation of (10). Using a similar approach to the calculation of $\overline{\text{elpd}}$ (Section A.2), we obtain

$$\overline{\text{lppd}}_{\text{LOO-VI}} \sim T_1 + T_2 + T_3 \tag{83}$$

where

$$T_1 = -\frac{1}{2n}\sum_{i=1}^{n}\log\left|\frac{\sigma^2}{\alpha n}\left(\mathrm{diag}\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)\right)^{-1} x_i x_i^\top + \sigma^2 I\right|, \tag{84}$$

$$\begin{aligned} T_2 = -\frac{1}{2n}\sum_{i=1}^{n}\frac{\alpha n}{\sigma^2}&\left[\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left(\frac{1}{n-1}X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right)\right]^\top \\ &\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1} \\ &\left[\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left(\frac{1}{n-1}X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right)\right]^\top, \end{aligned} \tag{85}$$

and

$$T_3 = \frac{1}{2n}\sum_{i=1}^{n} v^\top\left(\frac{1}{\sigma^2}x_i x_i^\top + \frac{\alpha n}{\sigma^2}\mathrm{diag}\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)\right)^{-1} v, \tag{86}$$

where

$$v = \left[\frac{1}{\sigma^2}x_i y_i + \frac{\alpha n}{\sigma^2}\mathrm{diag}\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)\left(\frac{1}{n-1}X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\right)^{-1}\left(\frac{1}{n-1}X_{-i}^\top Y_{-i} + \frac{\sigma^2}{\alpha n}\Sigma_0^{-1}\mu_0\right)\right] \tag{87}$$

A.4 Grids used in grid searches

We list the grids used in the grid searches in our simulation example (Section 3, Table 4) and our data application example (Section 3.4, Table 5). Each table contains the following columns:

- Method: Selection method/loss function. See Section 3.2 for details on each method.
- Parameter: The parameter with respect to which the loss function is minimized. All methods except for Safebayes – which optimizes for the tempering parameter, α – optimize for λ (the regularization parameter in (9)).
- Spacing: whether the grid is linearly spaced or logarithmically spaced.
- Interval: lower and upper bounds of the grid.
- Density: number of gridpoints.
- Mapping to α: How the selected parameter is mapped back to $\hat{\alpha}_n$ in (28).

After $\hat{\alpha}_n$ is computed from Tables 4 and 5, we recode $\hat{\alpha}_n$ that are greater than 10^6 to be "effectively" infinity, and label such values as ∞ in Figures 1, 2, and 3a – 3c.

Method	Parameter	Spacing	Interval	Density	Mapping to α
Bayesian CV	λ	Logarithmic	$[10^{-12}, 10]$	200	$(n\lambda)^{-1}$
Bayesian CV + VI	λ	Logarithmic	$[10^{-12}, 10]$	200	$(n\lambda)^{-1}$
LOOCV	λ	Linear	$[10^{-12}, 30]$	200	λ^{-1}
Train-test split	λ	Linear	$[10^{-12}, 5]$	200	$(n\lambda)^{-1}$
Safebayes	α	Linear	$[0, 1]$	30	α

Table 4: Grids used in the simulation example described in Section 3.

Method	Parameter	Spacing	Interval	Density	Mapping to α
Bayesian CV	λ	Logarithmic	$[10^{-12}, 10]$	200	$(n\lambda)^{-1}$
Bayesian CV + VI	λ	Logarithmic	$[10^{-12}, 10]$	200	$(n\lambda)^{-1}$
LOOCV	λ	Linear	$[10^{-12}, 0.5]$	200	λ^{-1}
Train-test split	λ	Linear	$[10^{-12}, 0.05]$	200	$(n\lambda)^{-1}$
Safebayes	α	Linear	$[0, 1]$	30	α

Table 5: Grids used in the data application example described in Section 3.4.

A.5 Relationship between $\hat{\alpha}_n$ and n

In this section, we explain how we estimate the rate of convergence of $\hat{\alpha}_n$ to 0 that appear in Tables 2 and 3. To obtain a curve of best fit of $\hat{\alpha}_n$ vs. n, we assume $\alpha = Cn^{\gamma}$, and our goal is to estimate the parameters C and γ. We regress $\log(\alpha)$ on $\log(n)$, since

$$\log(\alpha) = \log(C) + \gamma \log n \coloneqq \beta_0 + \beta_1 \log n, \tag{88}$$

the estimated curve of best fit is $\hat{\alpha}_n = e^{\hat{\beta}_0} n^{\hat{\beta}_1}$.

$$\begin{aligned} \beta_0 &= \log(C) \iff C = e^{\beta_0} \\ \beta_1 &= \gamma. \end{aligned} \tag{89}$$

We report the estimates and 95% confidence intervals for β_1 for our simulation example in Table 2 and for our data application example in Table 3.

Furthermore, we plot the corresponding curves of best fit in Figure 5 (simulation example) and Figure 6 (data application example).

A.6 Model specification for logistic regression example

We conduct the posterior mean experiment (Figure 4) in the following logistic regression setting. Consider paired data $\{x, y\}_{i=1}^n$, where $x_i = [x_{i,1}, x_{i,2}, x_{i,3}]^\top \in \mathbb{R}^3$ and $y_i \in \mathbb{R}$ for all $i = 1, \ldots, n$. We simulate each sample according to the following generative model

$$\begin{aligned} x_i &\sim \mathcal{N}(0, I_3) \\ g(x_i) &= x_i^\top \beta^* \\ p(y_1 = 1|x_i) &= \frac{1}{1 + \exp(-g(x_i))} \end{aligned} \tag{90}$$

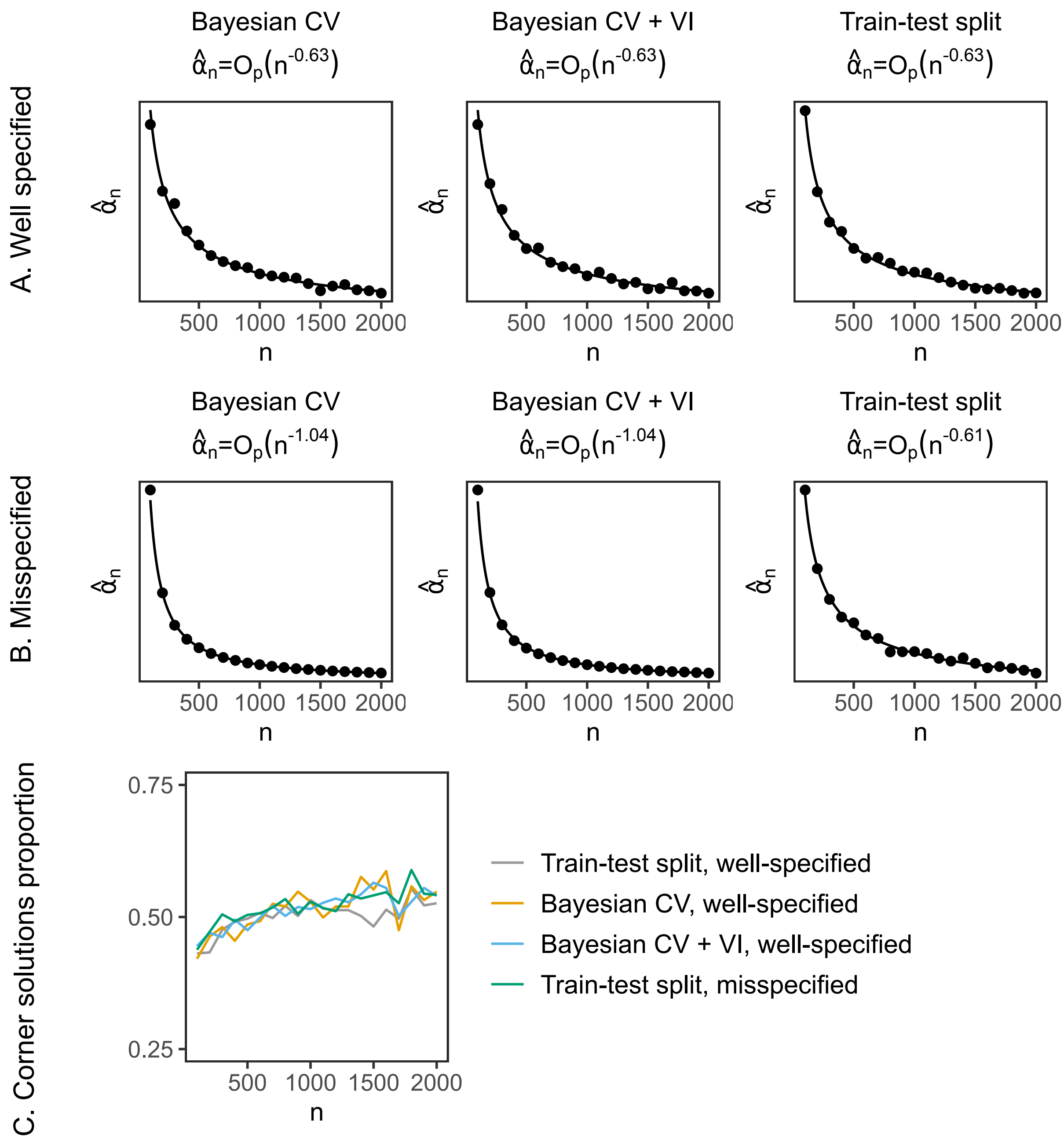


Figure 5: Panel A and B: dependence of $\hat{\alpha}_n$ on n in the well-specified model setting (A) and misspecified model setting (B). Points correspond to $\hat{\alpha}_n$ values averaged over 1,000 replications with overlaid curves of best fit. Plot titles denote the estimated stochastic order of $\hat{\alpha}_n$. Corner solutions at $\hat{\alpha}_n = \infty$ are discarded in the plotting and estimation in the following settings: Bayesian CV/well-specified, Bayesian CV+VI/well-specified, train-test/well-specified and misspecified. Panel C: dependence of the proportion of corner solutions on n.

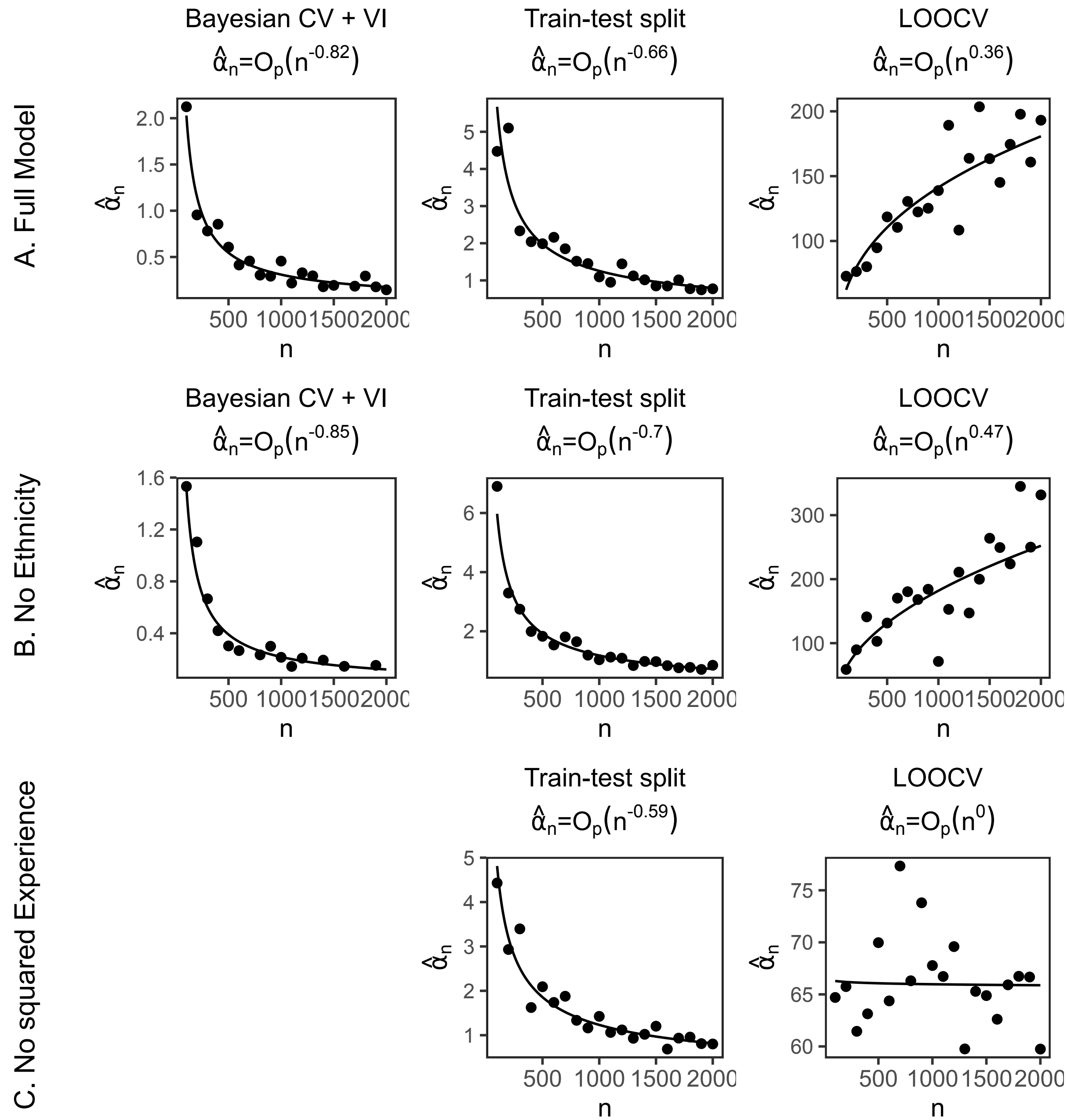


Figure 6: Panel A, B, and C: dependence of $\hat{\alpha}_n$ on n under the full model setting (A), model excluding ethnicity (B), and model excluding squared experience. Points correspond to $\hat{\alpha}_n$ values (with corner solutions discarded) averaged over 100 replications with overlaid curves of best fit. Plot titles denote the estimated stochastic order of $\hat{\alpha}_n$. The plot for Bayesian CV + VI under the model excluding squared experience is not plotted because "non-corner solutions" only appear for n = 100. We also not that the "non-corner" solutions according to LOOCV appear to be growing – rather than shrinking – with n. We include this result for completeness.

where $\beta^* = [1, -0.5, 0.1]$. We note that for this setting, the generative model and specified model are the same (i.e., the model is well-specified).

B Extension to Data-dependent α_n

Before proving Corollary 1, we will establish and prove the following Lemma:

Lemma 2. Suppose for a nonramdom function, H_n, we have that $H_n(\alpha_n) \to 0$ for sequences $\alpha_n > 0$ and $r_n \to \infty$ satisfying $\alpha_n \to 0$ and $r_n\alpha_n \to \infty$. Then $H_n(\hat{\alpha}_n) \to 0$ in $f_{0,n}$-probability for a sequence of positive random variables, $\hat{\alpha}_n$ satsfying (27).

Proof. Towards a contradiction, assume that $H_n(\hat{\alpha}_n)$ does not converge to 0 in $f_{0,n}$-probability. Hence, there exists $\tilde{\epsilon} > 0$ such that $\limsup_{n\to\infty} \mathbb{P}_{f_{0,n}}(H_n(\hat{\alpha}_n) > \tilde{\epsilon}) \to c$, where $c > 0$. Furthermore, by assumption on $\hat{\alpha}_n$, there exists a sequence δ_n such that $\delta_n \to 0$ and $r_n\delta_n \to \infty$ such that $\mathbb{P}_{f_{0,n}}\left(\frac{1}{r_n\delta_n} \leq \hat{\alpha}_n \leq \delta_n\right) \to 1$. Thus, we have

$$\limsup_{r_n\to\infty} \mathbb{P}_{f_{0,n}}\left(\frac{1}{r_n\delta_n} \leq \hat{\alpha}_n \leq \delta_n \cap H_n(\hat{\alpha}_n) > \tilde{\epsilon}\right) = \tilde{c},$$

where $\tilde{c} > 0$. Hence, the set

$$\left\{\frac{1}{r_n\delta_n} \leq \hat{\alpha}_n \leq \delta_n \cap H_n(\hat{\alpha}_n) > \tilde{\epsilon}\right\}$$

is nonempty. Thus, there exists a sequence α_n such that $\frac{1}{r_n\delta_n} \leq \hat{\alpha}_n \leq \delta_n$ and $H_n(\hat{\alpha}_n) \to c'$, where $c' > 0$. This is a contradiction. □

B.1 Proof of Corollary 1

By assumption, Theorem 1 holds for $\frac{1}{n} \ll \alpha_n \ll 1$. Hence, $r_n = n$, and we apply Lemma 2 with

$$H_n(\alpha_n) = \mathbb{P}_{f_{0,n}}\left(\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}(\theta|X^n), \phi\left(\theta\middle|\hat{\theta}, \frac{1}{\alpha_n n}V_{\theta^*}^{-1}\right)\right) > \epsilon\right).$$

B.2 Proof of Corollary 2

By assumption, Theorem 2 holds for $\frac{1}{n} \ll \alpha_n \ll 1$. Hence, $r_n = n$, and we apply Lemma 2 with

$$H_n(\alpha_n) = \mathbb{P}_{f_{0,n}}\left(\int_{\mathbb{R}^p} \left\|h^{\otimes k}\right\|_1 \left|(\alpha_n n)^{-\frac{p}{2}}\pi_{n,\alpha_n}\left(\theta^* + \frac{h}{\sqrt{\alpha_n n}}|X^n\right) - (\alpha_n n)^{-\frac{p}{2}}\phi\left(h\middle|\sqrt{\alpha_n n}(\hat{\theta}-\theta^*), V_{\theta^*}^{-1}\right)\right| dh > \epsilon\right).$$

B.3 Proof of Corollary 3

Recall the proof of Theorem 3 relies on establishing

$$\hat{\theta}^B - \hat{\theta} = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right) \tag{91}$$

since

$$\sqrt{n}(\hat{\theta}^B - \theta^*) = \sqrt{n}(\hat{\theta}^B - \hat{\theta}) + \sqrt{n}(\hat{\theta} - \theta^*) = O_{f_{0,n}}\left(\frac{1}{\alpha_n\sqrt{n}}\right) + \sqrt{n}(\hat{\theta} - \theta^*).$$

By assumption, (91) holds for $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$. Hence, $r_n = \sqrt{n}$, and we apply Lemma 2 with

$$H_{n,j}(\alpha_n) = \mathbb{P}_{f_{0,n}}\left(|\hat{\theta}_j^B - \hat{\theta}_j| > \frac{M}{\alpha_n n}\right),\ 1 \leq j \leq p.$$

B.4 Proof of Corollary 4

We will prove this result with a technique similar to that used in the proof of Lemma 2. Define the function

$$H_n(\alpha) = \mathbb{P}_{f_{0,n}}\left(\mathrm{d_{TV}}\left(\pi_{n,\alpha}(\theta|X^n), \delta_{\hat{\theta}}\right) > \epsilon\right). \tag{92}$$

By Theorem 7, $H(\alpha) \to 0$ as $\alpha \to \infty$ and $n \to \infty$. To establish (54), we would like to show $H_n(\hat{\alpha}_n) \to 0$ in $f_{0,n}$-probability for random sequences, $\hat{\alpha}_n$, satisfying (55). Towards a contradiction, suppose this is not the case. Then there exists $\tilde{\epsilon} > 0$ such that $\limsup_{n\to\infty} \mathbb{P}_{f_{0,n}}(H_n(\alpha_n) > \tilde{\epsilon}) \to c$, where $c > 0$. Furthermore, by assumption on $\hat{\alpha_n}$, we have

$$\limsup_{M\to\infty} \mathbb{P}_{f_{0,n}}(\hat{\alpha}_n > M \cap H_n(\alpha_n) > \tilde{\epsilon}) = c.$$

Hence, the set

$$\{\hat{\alpha}_n > M \ \forall M > 0 \cap H_n(\alpha_n) > \tilde{\epsilon}\} \tag{93}$$

is nonempty, and there exists a sequence satisfying (55) where $H_n(\alpha_n) \not\to 0$. This is a contradiction.

B.5 Proof of Corollary 5

By assumption on $\hat{\alpha}_n$, we can write $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ as

$$\pi_{n,\hat{\alpha}_n}(\theta|X^n) = p_n \pi_{n,\hat{\eta}_n}(\theta|X^n) + (1 - p_n)\pi_{n,\hat{\gamma}_n}(\theta|X^n), \tag{94}$$

where $\hat{\eta}_n$ satisfies (27) and $\hat{\gamma}_n$ satisfies (55), and

$$p_n = \mathbb{P}_{f_{0,n}}(\hat{\eta}_n \le \epsilon,\ n\hat{\eta}_n > \epsilon) = 1 - \mathbb{P}_{f_{0,n}}(\hat{\gamma}_n > M) \to p.$$

By convexity of total variation distance,

$$\begin{aligned}
&\mathrm{d_{TV}}\left(\pi_{n,\hat{\alpha}_n}(\theta|X^n), p\phi\left(\theta \,\middle|\, \hat{\theta}, \frac{1}{\hat{\eta}_n n} V_{\theta^*}^{-1}\right) + (1-p)\delta_{\hat{\theta}}\right) \\
&\le p_n \mathrm{d_{TV}}\left(\pi_{n,\hat{\eta}_n}(\theta|X^n), \phi\left(\theta \,\middle|\, \hat{\theta}, \frac{1}{\hat{\eta}_n n} V_{\theta^*}^{-1}\right)\right) + (1-p_n)\mathrm{d_{TV}}\left(\pi_{n,\hat{\gamma}_n}(\theta|X^n), \delta_{\hat{\theta}}\right)
\end{aligned} \tag{95}$$

Since the conditions of Proposition 6 hold, **(A2)** holds since Proposition 6 is stronger than Proposition 2 (which provides sufficient conditions for **(A2)** to hold). Furthermore, since the conditions of Proposition 5 hold, **(A3)** holds. Hence, the conditions of Theorem 1 hold and

$$\mathrm{d_{TV}}\left(\pi_{n,\hat{\beta}_n}(\theta|X^n), \phi\left(\theta \,\middle|\, \hat{\theta}, \frac{1}{\hat{\beta}_n n} V_{\theta^*}^{-1}\right)\right) \to 0$$

in $f_{0,n}$-probability by Corollary 1. Furthermore, since the conditions of Proposition 7 hold,

$$\mathrm{d_{TV}}\left(\pi_{n,\hat{\gamma}_n}(\theta|X^n), \delta_{\hat{\theta}}\right) \to 0$$

in $f_{0,n}$-probability by Corollary 4. Hence, (95) converges to 0 in $f_{0,n}$-probability, and we conclude (57).

C Sufficient Conditions

C.1 Proof of Proposition 2

A second-order Taylor expansion of $-\frac{1}{n}\log f_n\left(X^n\middle|\theta^*+\frac{h}{\sqrt{\alpha_n n}}\right)$ shows that

$$
\begin{aligned}
R_n(h) &= -\alpha_n n\left[-\frac{1}{n}\log f_n\left(X^n\middle|\theta^*+\frac{h}{\sqrt{\alpha_n n}}\right)-\left(-\frac{1}{n}\log f_n(X^n|\theta^*)\right)\right] \\
&\quad-\left[\sqrt{\alpha_n}V_{\theta^*}\Delta_{n,\theta^*}-\frac{1}{2}h^\top V_{\theta^*}h\right] \\
&= -\alpha_n n\Bigg[-\frac{h}{\sqrt{\alpha_n n}}^\top\left(\frac{1}{n}\nabla\log f_n(X^n|\theta^*)\right)-\frac{h}{\sqrt{2\alpha_n n}}^\top\left(\frac{1}{n}\nabla^2\log f_n(X^n|\theta^*)\right)\frac{h}{\sqrt{\alpha_n n}}-R'_n(h)\Bigg] \\
&\quad-\left[\sqrt{\alpha_n}V_{\theta^*}\Delta_{n,\theta^*}-\frac{1}{2}h^\top V_{\theta^*}h\right] \\
&= \sqrt{\alpha_n}\left(\frac{1}{\sqrt{n}}\log f_n(X^n|\theta^*)-V_{\theta^*}\Delta_{n,\theta^*}\right)-\frac{1}{2}h^\top\left(-\frac{1}{n}\nabla^2\log f_n(X^n|\theta^*)-V_{\theta^*}\right)h+\alpha_n nR'_n(h) \\
&= o_{f_{0,n}}(1)+\alpha_n nR'_n(h),
\end{aligned}
\tag{96}
$$

where the last line of (96) follows from (25) and (26). Fix $K=\bar{B}_0(r)$. It remains to verify that for all $\epsilon>0$, $r>0$, and $M>0$,

$$
\mathbb{P}_{f_{0,n}}\left(\sup_{h\in\bar{B}_0(r)}\left|\alpha_n nR'_n(h)\right|>M\right)<\epsilon \tag{97}
$$

for n sufficiently large. First, we argue that $R'_n(h)$ is well defined on $\bar{B}_0(r)$ as follows. By assumption, there exists $r_0>0$ such that $\log f_n(X^n|\theta)$ has continuous third derivatives on $\bar{B}_{\theta^*}(r_0)$. For any $r>0$, there exists $N_0 \coloneqq N_0(r_0,r)=\lceil 4r^2/r_0^2\rceil$ such that $\theta^*+h/\sqrt{\alpha_n n}\in\bar{B}_{\theta^*}(r_0)$ whenever $h\in\bar{B}_0(r)$ and $\alpha_n n\geq N_0$. To see this, note that if $\|h\|_2\leq r$ and $\alpha_n n\geq 4r^2/r_0^2$, then

$$
\left\|\theta^*+\frac{h}{\sqrt{\alpha_n n}}-\theta^*\right\|_2\leq\frac{r}{\sqrt{\frac{4r^2}{r_0^2}}}<\frac{r_0}{2}<r_0. \tag{98}
$$

Hence, for any $h\in\bar{B}_0(r)$, $R'_n(h)$ is well defined for n sufficiently large. For $\alpha_n n\geq N_0$, we can represent $R'_n(h)$ as

$$
\begin{aligned}
&\sup_{h\in\bar{B}_0(r)}\ \alpha_n nR'_n(h) \\
&= \sup_{h\in\bar{B}_0(r)}\frac{1}{2\sqrt{\alpha_n n}}\int_0^1(1-t)^2\left\langle\frac{1}{n}\nabla^3\log f_n\left(X^n\middle|\theta^*+t\frac{h}{\sqrt{\alpha_n n}}\right),h^{\otimes 3}\right\rangle dt \\
&\leq \frac{1}{2\sqrt{\alpha_n n}}\left[\sup_{\theta\in\bar{B}_{\theta^*}(r_0)}\max_{j,k,l}\left|\left\{\frac{1}{n}\nabla^3\log f_n(X^n|\theta)\right\}_{j,k,l=1}^p\right|\right]\sup_{h\in\bar{B}_0(r)}\sum_{j,k,l}h_jh_kh_l\int_0^1(1-t)^2dt \\
&\leq \frac{p^{3/2}r^3}{6\sqrt{\alpha_n n}}\left[\sup_{\theta\in\bar{B}_{\theta^*}(r_0)}\max_{j,k,l}\left|\left\{\frac{1}{n}\nabla^3\log f_n(X^n|\theta)\right\}_{j,k,l=1}^p\right|\right],
\end{aligned}
\tag{99}
$$

The inequality in (99) follows from noting that when $\alpha_n n \geq N_0$, $h \in \bar{B}_0(r)$ implies $\theta \in \bar{B}_{\theta^*}(r_0)$. Hence,

$$\sup_{t\in[0,1],\ h\in\bar{B}_0(r)} \max_{j,k,l} \left\{ \frac{1}{n}\nabla^3 \log f_n \left(X^n \middle| \theta^* + t\frac{h}{\sqrt{\alpha_n n}} \right) \right\}_{j,k,l=1}^{p}$$
$$\leq \sup_{\theta\in\bar{B}_{\theta^*}(r_0)} \max_{j,k,l} \left| \left\{ \frac{1}{n}\nabla^3 \log f_n(X^n|\theta) \right\}_{j,k,l=1}^{p} \right|.$$

Taking $\alpha_n n > N_0$, we apply (99) to the probability in (97) to obtain

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\left(\left|\alpha_n n R_n'(h)\right| > M\right) \\
&\leq \mathbb{P}_{f_{0,n}}\left(\frac{p^{3/2}r^3}{6\sqrt{\alpha_n n}} \left[\sup_{\theta\in\bar{B}_{\theta^*}(r_0)} \max_{j,k,l} \left| \left\{ \frac{1}{n}\nabla^3 \log f_n(X^n|\theta) \right\}_{j,k,l=1}^{p} \right| \right] > M \right) \\
&\leq \frac{p^{3/2}r^3}{6M\sqrt{\alpha_n n}} \mathbb{E}_{f_{0,n}} \left[\sup_{\theta\in\bar{B}_{\theta^*}(r_0)} \max_{j,k,l} \left| \left\{ \frac{1}{n}\nabla^3 \log f_n(X^n|\theta) \right\}_{j,k,l=1}^{p} \right| \right] \\
&\leq \frac{p^{3/2}r^3}{6M\sqrt{\alpha_n n}} \mathbb{E}_{f_{0,n}} \left[\left| \frac{1}{n}\tilde{M}(X^n) \right| \right] < \frac{\epsilon}{2} < \epsilon
\end{aligned} \tag{100}$$

where we obtain the final inequality by taking

$$\alpha_n n > \left\lceil \frac{p^3 r^6 \mathbb{E}_{f_{0,n}}\left[\left|\frac{1}{n}\tilde{M}(X^n)\right|\right]^2}{9M^2\epsilon^2} \right\rceil := N_1,$$

which is finite (i.e., N_1 exists) by assumption. Hence, choosing $n > \max(N_0, N_1)$ and taking

$$\alpha_n n > \max\left\{ \left\lceil \frac{4r^2}{r_0^2} \right\rceil, \left\lceil \frac{p^3 r^6 \mathbb{E}_{f_{0,n}}\left[\left|\frac{1}{n}\tilde{M}(X^n)\right|\right]^2}{9M^2\epsilon^2} \right\rceil \right\}$$

yields **(A2)**.

C.2 Explicit counterexamples with conjugate priors

In this section, we provide specific examples of likelihood models and priors where Proposition 1 holds as well as the limiting characteristic functions in these cases.

Exponential-Gamma: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Exp}(\lambda)$. Denote the true parameter value by λ^*. Furthermore, let $\lambda \sim \Gamma(a, b)$. We can write $f_n(X^n|\lambda)$ as

$$f(X^n|\lambda) = \exp\left(-\lambda \sum_{i=1}^{n} X_i - n\left(-\log(\lambda)\right) \right).$$

Hence, writing $f_n(X^n|\eta)$ in the form of (21) yields $\eta = -\lambda$, $A(\eta) = -\log(-\eta)$, $T(X^n) = \sum_{i=1}^{n} X_i$, and $h(X^n) = 1$. We can write $\pi(\lambda)$ as

$$\pi(\lambda) = \exp\left(-\lambda b - (1-a)(-\log(\lambda)) - \log\left(\frac{b^a}{\Gamma(a)}\right) \right).$$

Hence, writing $\pi(\eta)$ in the form of (22) yields $\xi = b$, $\nu = 1 - a$, $\psi(\xi, \nu) = \log\left(\frac{\xi^{1-\nu}}{\Gamma(1-\nu)}\right)$, and $\tilde{h}(\xi, \nu) = 1$. We can write the characteristic function of the limit of the α_n-posterior in the form of (24)

$$\varphi_\infty(t) = \exp\left(\log\left(\frac{(\xi' + it)^{1-\nu'}}{\Gamma(1-\nu')}\right) - \log\left(\frac{\xi'^{1-\nu}}{\Gamma(1-\nu')}\right)\right) = \left(\frac{\xi' + it}{\xi'}\right)^{1-\nu'},$$

where

$$\begin{aligned} g(\eta^*) &= -\frac{1}{\eta^*} = \frac{1}{\lambda^*} \\ \xi' &= \alpha_0 g(\eta^*) + \xi = \frac{\alpha_0}{\lambda^*} + b \\ \nu' &= \alpha_0 + \nu = \alpha_0 + 1 - a. \end{aligned}$$

Hence, Proposition 1 holds.

Exponential-Gamma: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Exp}(\lambda)$. Denote the true parameter value by λ^*. Furthermore, let $\lambda \sim \Gamma(a, b)$. We can write $f_n(X^n|\lambda)$ as

$$f(X^n|\lambda) = \exp\left(-\lambda \sum_{i=1}^n X_i - n\left(-\log(\lambda)\right)\right).$$

Hence, writing $f_n(X^n|\eta)$ in the form of (21) yields $\eta = -\lambda$, $A(\eta) = -\log(-\eta)$, $T(X^n) = \sum_{i=1}^n X_i$, and $h(X^n) = 1$. We can write $\pi(\lambda)$ as

$$\pi(\lambda) = \exp\left(-\lambda b - (1-a)(-\log(\lambda)) - \log\left(\frac{b^a}{\Gamma(a)}\right)\right).$$

Hence, writing $\pi(\eta)$ in the form of (22) yields $\xi = b$, $\nu = 1 - a$, $\psi(\xi, \nu) = \log\left(\frac{\xi^{1-\nu}}{\Gamma(1-\nu)}\right)$, and $\tilde{h}(\xi, \nu) = 1$. We can write the characteristic function of the limit of the α_n-posterior in the form of (24)

$$\varphi_\infty(t) = \exp\left(\log\left(\frac{(\xi' + it)^{1-\nu'}}{\Gamma(1-\nu')}\right) - \log\left(\frac{\xi'^{1-\nu}}{\Gamma(1-\nu')}\right)\right) = \left(\frac{\xi' + it}{\xi'}\right)^{1-\nu'},$$

where

$$\begin{aligned} g(\eta^*) &= -\frac{1}{\eta^*} = \frac{1}{\lambda^*} \\ \xi' &= \alpha_0 g(\eta^*) + \xi = \frac{\alpha_0}{\lambda^*} + b \\ \nu' &= \alpha_0 + \nu = \alpha_0 + 1 - a. \end{aligned}$$

Hence, Proposition 1 holds.

Pareto-Gamma: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Pareto}(x_m, k)$, where x_m is known. Denote the true parameter value by k^*. Furthermore, let $k \sim \Gamma(a, b)$. We can write $f_n(X^n|k)$ as

$$f(X^n|\eta) = \exp\left(-(k+1)\sum_{i=1}^n \log\left(\frac{X_i}{x_m}\right) - n\log\left(\frac{x_m}{k}\right)\right).$$

Hence, writing $f_n(X^n|\eta)$ in the form of (21) yields $\eta = -(k+1)$, $A(\eta) = \log\left(\frac{x_m}{-\eta-1}\right)$, $T(X^n) = \sum_{i=1}^n \log\left(\frac{X_i}{x_m}\right)$, and $h(X^n) = 1$. We can write $\pi(k)$ as

$$\pi(k) = \exp\left(-(k+1)b - (1-a)\log\left(\frac{x_m}{k}\right) - \left(-\log\left(\frac{b^a}{\Gamma(a)}\right) - (1-a)\log x_m - b\right)\right).$$

Hence, writing $\pi(\eta)$ in the form of (22) yields $\xi = b$, $\nu = 1-a$, $\psi(\xi,\nu) = -\log\left(\frac{\xi^{1-\nu}}{\Gamma(1-\nu)}\right) - \nu\log x_m - \xi$, and $\tilde{h}(\xi,\nu) = 1$. We can write the characteristic function of the limit of the α_n-posterior in the form of (24)

$$\varphi_\infty(t) = \exp\left(\log\left(\frac{\Gamma(1-\nu')}{(it+\xi')^{1-\nu'}x_m^{\nu'}\exp(it+\xi')}\right) - \log\left(\frac{\Gamma(1-\nu')}{\xi'^{1-\nu'}x_m^{\nu'}\exp(\xi')}\right)\right) = \left(\frac{\xi'}{\xi'+it}\right)^{1-\nu'} e^{-it},$$

where

$$\begin{aligned} g(\eta^*) &= -\frac{1}{\eta^*+1} = \frac{1}{k^*} \\ \xi' &= \alpha_0 g(\eta^*) + \xi = \frac{\alpha_0}{k^*} + b \\ \nu' &= \alpha_0 + \nu = \alpha_0 + 1 - a. \end{aligned}$$

Hence, Proposition 1 holds.

Bernoulli-Beta: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(p). Denote the true parameter value by p^*. Furthermore, let $p \sim \text{beta}(a,b)$. We can write $f_n(X^n|p)$ as

$$f(X^n|\eta) = \exp\left(\log\left(\frac{p}{1-p}\right)\sum_{i=1}^n X_i - n\log\left(\frac{1}{1-p}\right)\right).$$

Hence, writing $f_n(X^n|\eta)$ in the form of (21) yields $\log\left(\frac{p}{1-p}\right)$, $A(\eta) = \log(1+e^\eta)$, $T(X^n) = \sum_{i=1}^n X_i$, and $h(X^n) = 1$. We can write $\pi(p)$ as

$$\pi(p) = \exp\left((a-1)\log\left(\frac{p}{1-p}\right) - (a+b+2)\log\left(\frac{1}{1-p}\right) - \log\left(\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\right)\right).$$

Hence, writing $\pi(\eta)$ in the form of (22) yields $\xi = a-1$, $\nu = a+b+2$, $\psi(\xi,\nu) = \log\left(\frac{\Gamma(1+\xi)\Gamma(\nu-\xi-3)}{\Gamma(\nu-2)}\right)$, and $\tilde{h}(\xi,\nu) = 1$. We can write the characteristic function of the limit of the α_n-posterior in the form of (24)

$$\begin{aligned} \varphi_\infty(t) &= \exp\left(\log\left(\frac{\Gamma(1+\xi'+it)\Gamma(\nu'-\xi'-it-3)}{\Gamma(\nu'-2)}\right) - \log\left(\frac{\Gamma(1+\xi')\Gamma(\nu'-\xi'-3)}{\Gamma(\nu'-2)}\right)\right) \\ &= \frac{\Gamma(1+\xi'+it)\Gamma(\nu'-\xi'-it-3)}{\Gamma(1+\xi')\Gamma(\nu'-\xi'-3)}, \end{aligned}$$

where

$$\begin{aligned} g(\eta^*) &= \frac{e^{\eta^*}}{1+e^{\eta^*}} = p^* \\ \xi' &= \alpha_0 g(\eta^*) + \xi = \alpha_0 p^* + a - 1 \\ \nu' &= \alpha_0 + \nu = \alpha_0 + a + b + 2 \end{aligned}$$

Hence, Proposition 1 holds.

We note that in the case where the likelihood model and prior distributions are Gaussian, Proposition 1 holds, but Theorem 1 does not. We demonstrate this below.

Gaussian-Gaussian: Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Denote the pseudo-true parameter value by μ^*. Furthermore, let $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. We can write $f_n(X^n|\mu)$ as

$$f(X^n|\mu) = \exp\left(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} X_i^2\right)\exp\left(\mu\frac{\sum_{i=1}^{n} X_i}{\sigma^2} - n\left(\frac{\mu^2}{2\sigma^2}\right)\right).$$

Hence, writing $f_n(X^n|\eta)$ in the form of (21) yields $\eta = \mu$, $A(\eta) = \frac{\eta}{2\sigma^2}$, $T(X^n) = \sum_{i=1}^{n} X_i/\sigma^2$, and $h(X^n) = \exp\left(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} X_i^2\right)$. We can write $\pi(\mu)$ as

$$\pi(k) = \exp\left(-\frac{1}{2}\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\mu_0^2\right)\exp\left(\mu\frac{\mu_0}{\sigma_0^2} - \frac{\mu_0^2}{2\sigma_0^2}\right)$$

Hence, writing $\pi(\eta)$ in the form of (22) yields $\xi = \mu_0/\sigma_0^2$, $\nu = \sigma^2/\sigma_0^2$, and $\psi(\xi,\nu) = \frac{1}{2}\left(\frac{\nu}{\sigma^2}\right)\left(\frac{\xi}{\nu/\sigma^2}\right)^2 + \frac{1}{2}\log\left(2\pi(\nu/\sigma^2)^{-1}\right)$, and $\tilde{h}(\xi,\nu) = \exp\left(-\frac{1}{2}\log(2\pi\sigma^2/\nu) - \frac{\xi^2\sigma^2}{2\nu}\right)$ We can write the characteristic function of the limit of the α_n-posterior in the form of (24)

$$\varphi_\infty(t) = \exp\left(\frac{1}{2}\left(\frac{\nu'}{\sigma^2}\right)\left(\frac{\xi' + it}{\nu'/\sigma^2}\right)^2 - \frac{1}{2}\left(\frac{\nu'}{\sigma^2}\right)\left(\frac{\xi'}{\nu'/\sigma^2}\right)^2\right) = \exp\left(\frac{\xi' it}{\nu'/\sigma^2} - \frac{t^2}{2(\nu'/\sigma^2)}\right)$$

where

$$\begin{aligned} g(\eta^*) &= \frac{\eta^*}{\sigma^2} = \frac{\mu^*}{\sigma^2} \\ \xi' &= \alpha_0 g(\eta^*) + \xi = \frac{\alpha_0\mu^*}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \\ \nu' &= \alpha_0 + \nu = \alpha_0 + \frac{\sigma^2}{\sigma_0^2}. \end{aligned}$$

The alleged limiting distribution of the α_n-posterior according to Theorem 1 is

$$\mathcal{N}\left(\hat{\mu}, \frac{\sigma^2}{\alpha_n n}\right). \tag{101}$$

Recognizing that we can write this distribution in the form of (22), we note the parameters of this distribution are yields $\bar{\xi} = \frac{\alpha_n n\hat{\mu}}{\sigma^2}$ and $\bar{\nu} = \alpha_n n$ and the characteristic function is

$$\varphi_{\pi_{n,\alpha_n}}(t) = \exp\left(\frac{\bar{\xi} it}{\bar{\nu}/\sigma^2} - \frac{t^2}{2(\bar{\nu}/\sigma^2)}\right) \to \exp\left(\frac{\xi'' it}{\nu''/\sigma^2} - \frac{t^2}{2(\nu''/\sigma^2)}\right),$$

where

$$\begin{aligned} \bar{\xi} &\to \xi'' = \frac{\alpha_0\mu^*}{\sigma^2} \\ \bar{\nu} &\to \nu'' = \alpha_0 \end{aligned}$$

Hence, although Proposition 1 holds, Theorem 1 does not hold.

C.3 Proof of Proposition 3

Proof. We rewrite the distribution in (35) as

$$
\begin{aligned}
Y &\sim \text{Bernoulli}(1/n) \\
X_n &\sim \begin{cases} \mathcal{N}(n,1) \text{ if Y=1} \\ \mathcal{N}(0,1) \text{ if Y=0.} \end{cases}
\end{aligned} \tag{102}
$$

Proof of (1): Note that

$$
f_{X_n}(x) = \frac{1}{n}\phi(x|n,1) + \left(1 - \frac{1}{n}\right)\phi(x|0,1), \tag{103}
$$

where $\phi(\cdot|\mu,\sigma^2)$ denotes the density function of a normal distribution with mean μ and variance σ^2. Using the density of X_n defined in (103), we have

$$
\begin{aligned}
\mathrm{d}_{\mathrm{TV}}(X_n, Z) &= \frac{1}{2}\int \left| \frac{1}{n}\phi(x|n,1) + \left(1-\frac{1}{n}\right)\phi(x|0,1) - \phi(x|0,1)\right| dx \\
&= \frac{1}{2n}\int |\phi(x|n,1) - \phi(x|0,1)|\, dx \\
&= \frac{1}{n}\mathrm{d}_{\mathrm{TV}}\left(\mathcal{N}(n,1), \mathcal{N}(0,1)\right).
\end{aligned} \tag{104}
$$

To show that the right hand side of (104) converges to 0, we use the fact that for two distributions P and Q, $H^2(P,Q) \le \mathrm{d}_{\mathrm{TV}}(P,Q) \le \sqrt{2}H(P,Q)$, where $H(P,Q)$ is the Hellinger distance between P and Q. Applying this to (104), we find

$$
\frac{1}{n}\left(1 - e^{-\frac{1}{4}n^2}\right) \le \frac{1}{n}\mathrm{d}_{\mathrm{TV}}\left(\mathcal{N}(n,1), \mathcal{N}(0,1)\right) \le \frac{\sqrt{2}}{n}\sqrt{1 - e^{-\frac{1}{4}n^2}} \tag{105}
$$

We arrive at our conclusion by noting that both sides of (105) converge to 0.

Proof of (2): By the law of iterated expectation,

$$
\begin{aligned}
\mathbb{E}[X_n] &= \frac{1}{n}\mathbb{E}[X_n|Y=1] + \left(1-\frac{1}{n}\right)\mathbb{E}[X_n|Y=0] \\
&= \frac{1}{n}\mathbb{E}\left[\mathcal{N}(n,1)\right] + \left(1-\frac{1}{n}\right)\mathbb{E}\left[\mathcal{N}(0,1)\right] \\
&= 1 \neq 0 = \mathbb{E}[Z].
\end{aligned}
$$

Proof of (3): By the law of iterated expectation,

$$
\begin{aligned}
\mathbb{E}[|X_n|] &= \frac{1}{n}\mathbb{E}[|X_n||Y=1] + \left(1-\frac{1}{n}\right)\mathbb{E}[|X_n||Y=0] \\
&= \frac{1}{n}\mathbb{E}[|\mathcal{N}(n,1)|] + \left(1-\frac{1}{n}\right)\mathbb{E}[|\mathcal{N}(0,1)|].
\end{aligned} \tag{106}
$$

The second term of (106) is finite for all $n \in \mathbb{N}$ by the properties of the folded normal distribution. It remains to check the first term.

$$\frac{1}{n}\mathbb{E}\left[|\mathcal{N}(n,1)|\right] \leq \frac{1}{n}\mathbb{E}\left[|\mathcal{N}(n,1)-n|+n\right] \leq \frac{1}{n}\mathbb{E}\left[|\mathcal{N}(n,1)-\mathbb{E}\left[\mathcal{N}(n,1)\right]|\right]+1 = \frac{1}{n}\sqrt{\frac{2}{\pi}}+1,$$

which is finite for all $n \in \mathbb{N}$.

Proof of (4): By the law of iterated expectation,

$$\begin{aligned}\mathbb{E}[|X_n|^{1+\gamma}] &= \frac{1}{n}\mathbb{E}[|X_n|^{1+\gamma}|Y=1] + \left(1-\frac{1}{n}\right)\mathbb{E}[|X_n|^{1+\gamma}|Y=0] \\ &= \frac{1}{n}\mathbb{E}[|\mathcal{N}(n,1)|^{1+\gamma}] + \left(1-\frac{1}{n}\right)\mathbb{E}[|\mathcal{N}(0,1)|^{1+\gamma}].\end{aligned} \tag{107}$$

The second term of (107) is finite for all $n \in \mathbb{N}$ by the properties of the folded normal distribution. We will show that the first term of (107) diverges with n. By Jensen's inequality,

$$\frac{1}{n}\mathbb{E}[|\mathcal{N}(n,1)|^{1+\gamma}] \geq \frac{1}{n}|\mathbb{E}[\mathcal{N}(n,1)]|^{1+\gamma} = n^{\gamma} \to \infty \tag{108}$$

for $\gamma > 0$.

C.4 Proof of Proposition 4

By assumption **(A1")** and the assumption that g is finite in a neighborhood of θ^*, it follows that b is also finite in a neighborhood of θ^*. It remains to establish the condition on R_b. A first-order Taylor expansion of $b\left(\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\right)$ around $b(\hat{\theta})$ shows that

$$\begin{aligned}R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) &= \left[b\left(\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\right)-b(\hat{\theta})\right] - v^{\top}\frac{h}{\sqrt{\alpha_n n}} \\ &= \left(\nabla b(\hat{\theta})-v\right)^{\top}\frac{h}{\sqrt{\alpha_n n}} + R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right) \\ &= R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right),\end{aligned} \tag{109}$$

where (109) follows from the differentiability assumptions on g and π and setting

$$v = \nabla b(\hat{\theta}) = \pi(\hat{\theta})\nabla g(\hat{\theta}) + g(\hat{\theta})\nabla\pi(\hat{\theta}).$$

Fix $K = \bar{B}_0(r)$. It remains to verify that for all $\epsilon > 0$, $r > 0$, and all $h \in \bar{B}_{\theta^*}(r)$, there exists $M > 0$ such that

$$\mathbb{P}_{f_{0,n}}\left(\left|R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right| > \frac{M}{\alpha_n n}\|h\|_2^2\right) < \epsilon \tag{110}$$

for n sufficiently large. First, we argue that $R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right)$ is well defined on $\bar{B}_0(r)$ with high probability for n sufficiently large. By assumption on g and **(A1")**, there exists $r_0 > 0$ and N_0 such that the event

$$\mathcal{E}_0 := \{b \in \mathcal{C}^2 \text{ on } \bar{B}_{\theta^*}(r_0)\} \tag{111}$$

holds with probability at least $1-\epsilon/4$ whenever $n > N_0$. Furthermore, by **(A0)**, there exists N_1 such that the event

$$\mathcal{E}_1 := \{\hat{\theta} \in \bar{B}_{\theta^*}(r_0/4)\} \tag{112}$$

holds with probability at least $1-\epsilon/4$ whenever $n > N_1$. Finally, there exists

$$N_2 \coloneqq N_2(r_0, r) = \max\left(N_1, \left\lceil \frac{16r^2}{r_0^2} \right\rceil\right) \tag{113}$$

such that the event

$$\mathcal{E}_2 \coloneqq \{\hat{\theta} + h/\sqrt{\alpha_n n} \in \bar{B}_{\theta^*}(r_0)\} \tag{114}$$

holds with probability at least $1-\epsilon/2$ whenever $n > N_2$ and $h \in \bar{B}_0(r)$. To see this, note that if $\|h\|_2 \le r$, $\alpha_n n \ge 16r^2/r_0^2$, and $n > \max(N_0, N_2)$, then

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}(\mathcal{E}_2^{\mathsf{c}}) &\le \mathbb{P}_{f_{0,n}}(\mathcal{E}_2^{\mathsf{c}} | \mathcal{E}_0 \cap \mathcal{E}_1) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_0^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}}) \\
&= \mathbb{P}_{f_{0,n}}\left(\left\| \hat{\theta} + \frac{h}{\sqrt{\alpha_n n}} - \theta^* \right\|_2 > r_0 \middle| \mathcal{E}_0 \cap \mathcal{E}_1\right) + \frac{\epsilon}{2} \\
&\le \mathbb{P}_{f_{0,n}}\left(\|\hat{\theta} - \theta^*\|_2 + \frac{r}{\sqrt{\frac{16r^2}{r_0^2}}} > r_0 \middle| \mathcal{E}_0 \cap \mathcal{E}_1\right) + \frac{\epsilon}{2} \\
&\le \mathbb{P}_{f_{0,n}}\left(\frac{r_0}{4} + \frac{r_0}{4} > r_0 \middle| \mathcal{E}_0 \cap \mathcal{E}_1\right) + \frac{\epsilon}{2} = \frac{\epsilon}{2}.
\end{aligned} \tag{115}$$

On $\mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2$, we can represent $R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right)$ as

$$\begin{aligned}
R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right) &= \frac{1}{\alpha_n n}\int_0^1 (1-t) h^\top \left[\nabla^2 b\left(\hat{\theta} + t\frac{h}{\sqrt{\alpha_n n}}\right)\right] h dt \\
&\le \frac{1}{\alpha_n n}\left[\sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{j,k} \left|\left\{\nabla^2 b(\theta)\right\}_{j,k=1}^p\right|\right] \sum_{j,k} h_k h_k \int_0^1 (1-t) dt && (116) \\
&\le \frac{p}{2\alpha_n n}\left[\sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{j,k} \left|\left\{\nabla^2 b(\theta)\right\}_{j,k=1}^p\right|\right] \|h\|_2^2, && (117)
\end{aligned}$$

where $\nabla^2 b(\theta) = \pi(\theta)\nabla^2 g(\theta) + 2\nabla g(\theta)\nabla\pi(\theta) + g(\theta)\nabla^2\pi(\theta)$. The inequality in (116) follows from noting that on the event $\mathcal{E}_2$, $h \in \bar{B}_0(r)$ implies $\theta \in \bar{B}_{\theta^*}(r_0)$.

Taking $\alpha_n n \ge 16r^2/r_0^2$ and $n > \max(N_0, N_2)$, we apply (117) to the probability in (110) to obtain

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\left(\left|R_b'\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right| > \frac{M}{\alpha_n n}\|h\|_2^2\right) \\
&\le \mathbb{P}_{f_{0,n}}\left(\frac{p}{2\alpha_n n}\left[\sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{i_1,i_2} \left|\left\{\nabla^2 b(\theta)\right\}_{i_1,i_2=1}^p\right|\right] \|h\|_2^2 > \frac{M}{\alpha_n n}\|h\|_2^2 \middle| \mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2\right) \\
&\quad + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}_0^{\mathsf{c}}\right) + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}_1^{\mathsf{c}}\right) + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}_2^{\mathsf{c}}\right) \\
&\le \mathbb{P}_{f_{0,n}}\left(\sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{i_1,i_2} \left|\left\{\nabla^2 b(\theta)\right\}_{i_1,i_2=1}^p\right| > \frac{2M}{p} \middle| \mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_2\right) + \epsilon = \epsilon.
\end{aligned} \tag{118}$$

We obtain the final inequality by setting

$$M = p \sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{i_1,i_2} \left|\left\{\nabla^2 b(\theta)\right\}_{i_1,i_2=1}^p\right|,$$

which is finite on the event $\mathcal{E}_0$, since $\nabla^2 b(\theta)$ is continuous on the compact set $\bar{B}_{\theta^*}(r_0)$. Hence, choosing $n > \max(N_0, N_2)$ and taking $\alpha_n n \ge 16r^2/r_0^2$ yields (46).

C.5 Proof of Proposition 5

We will show that the following object

$$\mathbb{E}_{f_{0,n}}[\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\epsilon_n)] \tag{119}$$

converges to 0. Define $\epsilon_n = \max\{s_n, t_n^{1/2}\}$, where t_n and s_n were defined in assumptions **(A1"')** and **(A3"')**, respectively.

Given constants $c > 0$, $c_1 > 0$, $C > 0$ sufficiently large, the set G_n defined in (50), and $\tilde{Z}_n$ to be defined later, we define the following events

$$\begin{aligned}
\mathcal{A} &= \left\{ \tilde{Z}_n \geq e^{-2\alpha_n n t_n} \int_{G_n} \pi(\theta) d\theta \right\} \\
\mathcal{B} &= \left\{ \|\hat{\theta} - \theta^*\| < \frac{C}{2}\tilde{\epsilon}_n \right\} \\
\mathcal{C} &= \left\{ \inf_{\|\theta-\hat{\theta}\| > \frac{C}{2}\epsilon_n} \left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta}) \right) \right] > \frac{c_1 C^2}{4}\epsilon_n^2 \right\}.
\end{aligned} \tag{120}$$

Since

$$\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\epsilon_n) < 1, \tag{121}$$

we can bound (119) by a union bound as follows

$$\begin{aligned}
&\mathbb{E}_{f_{0,n}}[\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\epsilon_n)] \\
&\leq \mathbb{E}_{f_{0,n}}[\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\epsilon_n)\mathbb{1}\{\mathcal{ABC}\}] + \mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{B}^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{C}^{\mathsf{c}})
\end{aligned} \tag{122}$$

We will bound each term of (122).

First term of (122)**:** We will analyze

$$\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\tilde{\epsilon}_n)$$

under the events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. Define

$$\tilde{Z}_n \coloneqq \int_{\mathbb{R}^p} \exp\left(-\alpha_n n \left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta}) \right) \right] \right) \pi(\theta) d\theta. \tag{123}$$

We have that

$$
\begin{aligned}
&\mathbb{E}_{f_{0,n}}\left[\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\epsilon_n)\right] \\
&= \mathbb{E}_{f_{0,n}}\left[\frac{\exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right)\int_{\bar{B}_{\theta^*}(C\epsilon_n)^c} \exp\left(-\alpha_n n\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right)\pi(\theta)d\theta}{\exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right)\int_{\mathbb{R}^p} \exp\left(-\alpha_n n\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right)\pi(\theta)d\theta}\right] \\
&= \mathbb{E}_{f_{0,n}}\left[\frac{\int_{\bar{B}_{\theta^*}(C\epsilon_n)^c} \exp\left(-\alpha_n n\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right)\pi(\theta)d\theta}{\tilde{Z}_n}\right] \\
&\le \mathbb{E}_{f_{0,n}}\left[\frac{\exp\left(-\alpha_n n \inf_{\|\theta-\theta^*\|>C\epsilon_n}\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right)}{\tilde{Z}_n}\right] \\
&\le \mathbb{E}_{f_{0,n}}\left[\frac{\exp\left(-\alpha_n n \inf_{\|\theta-\hat{\theta}\|+\|\hat{\theta}-\theta^*\|>C\epsilon_n}\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right)}{\tilde{Z}_n}\right] \\
&\overset{(a)}{\le} \mathbb{E}_{f_{0,n}}\left[\frac{\exp\left(-\alpha_n n \inf_{\|\theta-\hat{\theta}\|>C\epsilon_n/2}\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right)}{\tilde{Z}_n}\right] \\
&\overset{(b)}{\le} \mathbb{E}_{f_{0,n}}\left[\frac{e^{-\frac{c_1C^2}{4}\alpha_n n\epsilon_n^2}}{e^{-2\alpha_n n t_n}\int_{G_n}\pi(\theta)d\theta}\right] \\
&\overset{(c)}{\le} \mathbb{E}_{f_{0,n}}\left[\frac{e^{-\frac{c_1C^2}{4}\alpha_n n\epsilon_n^2}}{e^{-2\alpha_n n t_n}e^{-2\alpha_n n t_n}}\right] \\
&= \mathbb{E}_{f_{0,n}}\left[\frac{e^{-\frac{c_1C^2}{4}\alpha_n n\epsilon_n^2}}{e^{-4\alpha_n n t_n}}\right] \\
&\overset{(d)}{\le} \mathbb{E}_{f_{0,n}}\left[\frac{e^{-\frac{c_1C^2}{4}\alpha_n n\epsilon_n^2}}{e^{-4\alpha_n n \epsilon_n^2}}\right] \\
&= \mathbb{E}_{f_{0,n}}\left[\exp\left(-\alpha_n n\epsilon_n^2\left(\frac{c_1C^2}{4}-4\right)\right)\right].
\end{aligned} \tag{124}
$$

Since $\alpha_n n\epsilon_n^2 \to \infty$ by assumption, the above can be bounded above by $\epsilon/4$ for

$$
\alpha_n n\epsilon_n^2 \ge \log\left(\frac{c_1C^2-16}{\epsilon}\right) \tag{125}
$$

and $c_1C^2 > 16$. Step (a) of (124) follows from conditioning on $\mathcal{B}$. Step (b) follows from conditioning on $\mathcal{A}$ and $\mathcal{C}$. Step (c) follows from assumption **(A1"')**. Step (d) follows from the definition of $\epsilon_n = \max\{s_n, t_n^{1/2}\}$, which implies that $\epsilon_n^2 \ge t_n$.

Second term of (122)**:** We will show that

$$
\mathbb{P}_{f_{0,n}}\left(\tilde{Z}_n \le e^{-2\alpha_n n t_n}\int_{G_n}\pi(\theta)d\theta\right) < \frac{\epsilon}{4} \tag{126}
$$

for n sufficiently large. We have that

$$\begin{aligned}
\tilde{Z}_n &= \int_{\mathbb{R}^p} \exp\Bigg(- \alpha_n n \Big[- \frac{1}{n} \log f_n(X^n|\theta) - \Big(- \frac{1}{n} \log f_n(X^n|\theta^*) \Big) \\
&\qquad\qquad - \frac{1}{n} \log f_n(X^n|\theta^*) - \Big(- \frac{1}{n} \log f_n(X^n|\hat{\theta}) \Big) \Big] \Bigg) \pi(\theta) d\theta \\
&= \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta^*) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta}) \right) \right] \right) \\
&\quad \times \int_{\mathbb{R}^p} \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\theta^*) \right) \right] \right) \pi(\theta) d\theta \\
&= \exp\left(\alpha_n \log \left(\frac{f_n(X_n|\theta^*)}{f_n(X^n|\hat{\theta})} \right) \right) \\
&\quad \times \int_{\mathbb{R}^p} \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\theta^*) \right) \right] \right) \pi(\theta) d\theta \\
&\leq \int_{\mathbb{R}^p} \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\theta^*) \right) \right] \right) \pi(\theta) d\theta := Z_n,
\end{aligned} \tag{127}$$

since, by assumption **(A0)**, $f_n(X^n|\theta^*) \leq f_n(X^n|\hat{\theta})$, which implies $\alpha_n \log\left(\frac{f_n(X_n|\theta^*)}{f_n(X^n|\hat{\theta})}\right) < 0$, so $\exp\left(\alpha_n \log\left(\frac{f_n(X_n|\theta^*)}{f_n(X^n|\hat{\theta})}\right)\right) < 1$. Since $\tilde{Z}_n \leq Z_n$, by Lemma 4 of [48] (which is similar to Lemma 1 of [58]), we have that

$$\mathbb{P}_{f_{0,n}} \left(\tilde{Z}_n \leq e^{-2\alpha_n n t_n} \int_{G_n} \pi(\theta) d\theta \right) < \frac{2}{\alpha_n n t_n} < \frac{\epsilon}{4} \tag{128}$$

for $\alpha_n n t_n \geq 12/\epsilon$.

Third term of (122)**:** We will show that

$$\mathbb{P}_{f_{0,n}} \left(\|\hat{\theta} - \theta^*\| > \frac{C}{2} \epsilon_n \right) < \frac{\epsilon}{4} \tag{129}$$

for n sufficiently large. We rewrite this probability as

$$\mathbb{P}_{f_{0,n}} \left(\|\hat{\theta} - \theta^*\| > \frac{C}{2} \epsilon_n \right) = \mathbb{P}_{f_{0,n}} \left(\sqrt{\alpha_n n} \|\hat{\theta} - \theta^*\| > \frac{C}{2} \sqrt{\alpha_n n} \epsilon_n \right). \tag{130}$$

By assumption **(A0)**, $\sqrt{\alpha_n n}\|\hat{\theta} - \theta^*\| = o_{f_{0,n}}(1)$ since $\alpha_n \to 0$ by assumption. Furthermore, $\sqrt{\alpha_n n}\epsilon_n \to \infty$ by assumption. Hence, there exists $N_0 := N_0(\epsilon)$ such that for $n > N_0$, (129) holds.

Fourth term of (122)**:** We will show that

$$\mathbb{P}_{f_{0,n}} \left(\inf_{\|\theta - \hat{\theta}\| > \frac{C}{2}\epsilon_n} \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta}) \right) \right] < \frac{c_1 C^2}{4} \epsilon_n^2 \right) < \frac{\epsilon}{4} \tag{131}$$

for n sufficiently large. This follows from assumption **(A3"')** and noting that, by definition of ϵ_n, $s_n \leq \epsilon_n$. Thus,

$$\inf_{\|\theta-\hat{\theta}\|>\frac{C}{2}\epsilon_n}\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]$$
$$\geq \inf_{\|\theta-\hat{\theta}\|>\frac{C}{2}s_n}\left[-\frac{1}{n}\log f_n(X^n|\theta)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right].$$

Furthermore, $\frac{c_1C^2}{4}s_n^2 < \frac{c_1C^2}{4}\epsilon_n^2$. Hence, there exists $N_1 := N_1(\epsilon)$ such that for $n > N_1$, (131) holds.

Final bound: Taking $n > \max(N_0, N_1)$, $\alpha_n n\epsilon_n^2 \geq \log\left(\frac{c_1C^2-16}{\epsilon}\right)$, and $\alpha_n n t_n \geq 12/\epsilon$, we have

$$\mathbb{E}_{f_{0,n}}\left[\mathbb{P}_{\pi_{n,\alpha_n}}(||\theta-\theta^*||_2 > C\epsilon_n)\right] < \frac{\epsilon}{4}+\frac{\epsilon}{4}+\frac{\epsilon}{4}+\frac{\epsilon}{4} = \epsilon. \tag{132}$$

C.6 Proof of Proposition 6

A third-order Taylor expansion of $-\frac{1}{n}\log f_n\left(X^n\middle|\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\right)$ around $-\frac{1}{n}\log f_n(X^n|\hat{\theta})$ shows that

$$\begin{aligned}
\tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right) &= -\alpha_n n\left[-\frac{1}{n}\log f_n\left(X^n\middle|\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\right)-\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] \\
&\quad -\left[-\frac{1}{2}h^\top H_n h+\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle\right] \\
&= -\alpha_n n\Bigg[-\frac{h}{\sqrt{\alpha_n n}}^\top\left(\frac{1}{n}\nabla\log f_n(X^n|\hat{\theta})\right)-\frac{h}{\sqrt{2\alpha_n n}}^\top\left(\frac{1}{n}\nabla^2\log f_n(X^n|\hat{\theta})\right)\frac{h}{\sqrt{\alpha_n n}} \\
&\qquad -\frac{1}{6(\alpha_n n)^{\frac{3}{2}}}\left\langle\frac{1}{n}\nabla^3\log f_n(X^n|\hat{\theta}), h^{\otimes 3}\right\rangle-\tilde{R}'_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)\Bigg] \\
&\quad -\left[-\frac{1}{2}h^\top H_n h+\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle\right] \\
&= -\frac{1}{2}h^\top\left(-\frac{1}{n}\nabla^2\log f_n(X^n|\hat{\theta})-H_n\right)h+\frac{1}{6\sqrt{\alpha_n n}}\left\langle\frac{1}{n}\nabla^3\log f_n(X^n|\hat{\theta})-S_n, h^{\otimes 3}\right\rangle \\
&\quad +\alpha_n n\tilde{R}'_n\left(\frac{h}{\sqrt{\alpha_n n}}\right) \\
&= \alpha_n n\tilde{R}'_n\left(\frac{h}{\sqrt{\alpha_n n}}\right),
\end{aligned} \tag{133}$$

where the last line of (133) follows from the condition that $\nabla\log f_n(X^n|\hat{\theta}) = 0$ and setting $-\frac{1}{n}\nabla^2\log f_n(X^n|\hat{\theta}) = H_n$ and $\frac{1}{n}\nabla^3\log f_n(X^n|\hat{\theta}) = S_n$. Fix $K = \bar{B}_0(r)$. It remains to verify that for all $\epsilon > 0$, $r > 0$, and all $h \in \bar{B}_0(r)$, there exists $M > 0$ such that

$$\mathbb{P}_{f_{0,n}}\left(\left|\alpha_n n\tilde{R}'_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right| > \frac{M}{\alpha_n n}\|h\|_2^4\right) < \epsilon \tag{134}$$

for n sufficiently large. First, we argue that $\tilde{R}'_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)$ is well defined on $\bar{B}_0(r)$ with high probability for n sufficiently large. By assumption, there exists $r_0 > 0$ such that $\log f_n(X^n|\theta)$ has continuous fourth derivatives on $\bar{B}_{\theta^*}(r_0)$. Furthermore, by **(A0)**, there exists N_1 such that the event

$$\mathcal{E}_1 \coloneqq \{\hat{\theta} \in \bar{B}_{\theta^*}(r_0/4)\} \tag{135}$$

holds with probability at least $1 - \epsilon/4$ whenever $n > N_1$. Finally, there exists

$$N_2 \coloneqq N_2(r_0, r) = \max\left(N_1, \left\lceil \frac{16r^2}{r_0^2} \right\rceil\right) \tag{136}$$

such that the event

$$\mathcal{E}_2 \coloneqq \{\hat{\theta} + h/\sqrt{\alpha_n n} \in \bar{B}_{\theta^*}(r_0)\} \tag{137}$$

holds with probability at least $1 - \epsilon/4$ whenever $n > N_2$ and $h \in \bar{B}_0(r)$. To see this, note that if $\|h\|_2 \leq r$, $\alpha_n n \geq 16r^2/r_0^2$, and $n > N_2$, then

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}(\mathcal{E}_2^{\mathsf{c}}) &\leq \mathbb{P}_{f_{0,n}}(\mathcal{E}_2^{\mathsf{c}} | \mathcal{E}_1) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}}) \\
&= \mathbb{P}_{f_{0,n}}\left(\left\|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}} - \theta^*\right\|_2 > r_0 \middle| \mathcal{E}_1\right) + \frac{\epsilon}{4} \\
&\leq \mathbb{P}_{f_{0,n}}\left(\|\hat{\theta} - \theta^*\|_2 + \frac{r}{\sqrt{\frac{16r^2}{r_0^2}}} > r_0 \middle| \mathcal{E}_1\right) + \frac{\epsilon}{4} \\
&\leq \mathbb{P}_{f_{0,n}}\left(\frac{r_0}{4} + \frac{r_0}{4} > r_0 \middle| \mathcal{E}_1\right) + \frac{\epsilon}{4} = \frac{\epsilon}{4}.
\end{aligned} \tag{138}$$

On $\mathcal{E}_1 \cap \mathcal{E}_2$, we can represent $\alpha_n n \tilde{R}_n'\left(\frac{h}{\sqrt{\alpha_n n}}\right)$ as

$$\begin{aligned}
&\alpha_n n \tilde{R}_n'\left(\frac{h}{\sqrt{\alpha_n n}}\right) \\
&= \frac{1}{6\alpha_n n}\int_0^1 (1-t)^3 \left\langle \frac{1}{n}\nabla^4 \log f_n\left(X^n \middle| \hat{\theta} + t\frac{h}{\sqrt{\alpha_n n}}\right), h^{\otimes 4}\right\rangle dt \\
&\leq \frac{1}{6\alpha_n n}\left[\sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{j,k,l,m} \left|\left\{\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)\right\}_{j,k,l,m=1}^p\right|\right] \sum_{j,k,l,m} h_j h_k h_l h_m \int_0^1 (1-t)^3 dt \qquad (139) \\
&\leq \frac{p^2}{24\alpha_n n}\left[\sup_{\theta \in \bar{B}_{\theta^*}(r_0)} \max_{j,k,l,m} \left|\left\{\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)\right\}_{j,k,l,m=1}^p\right|\right] \|h\|_2^4. \qquad (140)
\end{aligned}$$

The inequality in (139) follows from noting that on the event $\mathcal{E}_2$, $h \in \bar{B}_0(r)$ implies $\theta \in \bar{B}_{\theta^*}(r_0)$.

Taking $\alpha_n n \geq 16r^2/r_0^2$ and $n > N_2$, we apply (140) to the probability in (134) to obtain

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\left(\left|\alpha_n n \tilde{R}_n'\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right| > \frac{M}{\alpha_n n}\|h\|_2^4\right) \\
&\leq \mathbb{P}_{f_{0,n}}\left(\frac{p^2}{24\alpha_n n}\left[\sup_{\theta\in\bar{B}_{\theta^*}(r_0)}\max_{j,k,l,m}\left|\left\{\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)\right\}_{j,k,l,m=1}^{p}\right|\right]\|h\|_2^4 > \frac{M}{\alpha_n n}\|h\|_2^4\middle|\mathcal{E}_1\cap\mathcal{E}_2\right) \\
&\quad + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}_1^{\mathsf{c}}\right) + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}_2^{\mathsf{c}}\right) \qquad (141) \\
&\leq \mathbb{P}_{f_{0,n}}\left(\sup_{\theta\in\bar{B}_{\theta^*}(r_0)}\max_{j,k,l,m}\left|\left\{\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)\right\}_{j,k,l,m=1}^{p}\right| > \frac{24M}{p^2}\middle|\mathcal{E}_1\cap\mathcal{E}_2\right) + \frac{\epsilon}{2} \\
&\leq \frac{p^2}{24M}\mathbb{E}_{f_{0,n}}\left[\sup_{\theta\in\bar{B}_{\theta^*}(r_0)}\max_{j,k,l,m}\left|\left\{\frac{1}{n}\nabla^4 \log f_n(X^n|\theta)\right\}_{j,k,l,m=1}^{p}\right|\mathbb{1}\{\mathcal{E}_1\cap\mathcal{E}_2\}\right] + \frac{\epsilon}{2} \\
&\leq \frac{p^2}{24M}\mathbb{E}_{f_{0,n}}\left[\left|\frac{1}{n}\tilde{M}\left(X^n\right)\right|\right] + \frac{\epsilon}{2} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \qquad (142)
\end{aligned}$$

where we obtain the final inequality by setting

$$M = \frac{p^2\mathbb{E}_{f_{0,n}}\left[\left|\frac{1}{n}\tilde{M}\left(X^n\right)\right|\right]}{12\epsilon},$$

which is finite by assumption. Hence, choosing $n > N_2$ and taking $\alpha_n n \geq 16r^2/r_0^2$ yields (37).

C.7 Checking (A3') for the $\sqrt{n}$-rescaled α_n-posterior

We provide a heuristic analysis of the expectation

$$n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\|h\|_2^{1+\gamma}\pi_{n,\alpha_n}\left(\theta^* + \frac{h}{\sqrt{n}}\middle|X^n\right)dh = \frac{n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\|h\|_2^{1+\gamma}\pi\left(\theta^*+\frac{h}{\sqrt{n}}\right)f_n\left(X^n\middle|\theta^*+\frac{h}{\sqrt{n}}\right)dh}{n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\pi\left(\theta^*+\frac{h}{\sqrt{n}}\right)f_n\left(X^n\middle|\theta^*+\frac{h}{\sqrt{n}}\right)dh} := \frac{\tilde{Z}}{Z}$$

We use a similar proof technique as that of Lemma 1 but with the change of variable $h = \sqrt{n}(\theta-\theta^*)$ to analyze Z and $\tilde{Z}$.

Starting with Z, we expand the log-likelihood and prior around θ^* and appeal to (25) and (26) from Proposition 2 to obtain

$$\begin{aligned}
&\frac{Z}{\exp\left(\alpha_n \log f_n(X^n|\theta^*)\right)} \\
&= n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\pi\left(\theta^*+\frac{h}{\sqrt{n}}\right)\exp\left(-\alpha_n n\left[-\frac{1}{n}\log f_n\left(X^n\middle|\theta^*+\frac{h}{\sqrt{n}}\right) - \left(-\frac{1}{n}\log f_n\left(X^n|\theta^*\right)\right)\right]\right)dh \\
&\approx n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\pi\left(\theta^*+\frac{h}{\sqrt{n}}\right)\exp\left(\alpha_n h^\top\frac{1}{\sqrt{n}}\nabla\log f_n(X^n|\theta^*)\right)\exp\left(-\frac{1}{2}h^\top\left(-\frac{\alpha_n}{n}\nabla^2\log f_n(X^n|\theta^*)\right)h\right)dh \\
&\approx n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\pi\left(\theta^*+\frac{h}{\sqrt{n}}\right)\exp\left(\alpha_n h^\top V_{\theta^*}\Delta_{n,\theta^*}\right)\exp\left(-\frac{1}{2}h^\top\left(\alpha_n V_{\theta^*}\right)h\right)dh \\
&\approx n^{\frac{p}{2}}\int_{\mathbb{R}^p}\left[\pi(\theta^*)+\frac{1}{\sqrt{n}}h^\top\nabla\pi(\theta^*)\right]\left[1+\alpha_n h^\top V_{\theta^*}\Delta_{n,\theta^*}\right]\exp\left(-\frac{1}{2}h^\top\left(\alpha_n V_{\theta^*}\right)h\right)dh.
\end{aligned} \qquad (143)$$

From (143), we see that Z (up to remainder terms) is approximately the expectation of a polynomial in h. We will examine the leading order term

$$n^{-\frac{p}{2}}\pi(\theta^*)\int_{\mathbb{R}^p}\exp\left(-\frac{1}{2}h^\top\left(\alpha_n V_{\theta^*}\right)h\right)dh=\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}}|V_{\theta^*}|^{-1}\pi(\theta^*).$$

Hence,

$$Z\sim\exp\left(\alpha_n\log f_n(X^n|\theta^*)\right)\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}}|V_{\theta^*}|^{-1}\pi(\theta^*).$$

We now perform a similar analysis for $\tilde{Z}$. We substitute $\|h\|_2^{1+\gamma}\pi(\theta^*+h/\sqrt{n})$ for $\pi(\theta^*+h/\sqrt{n})$ in (143) to obtain

$$\begin{aligned}
&\frac{\tilde{Z}}{\exp\left(\alpha_n\log f_n(X^n|\theta^*)\right)}\\
&=n^{-\frac{p}{2}}\int_{\mathbb{R}^p}\|h\|_2^{1+\gamma}\pi\left(\theta^*+\frac{h}{\sqrt{n}}\right)\exp\left(-\alpha_n n\left[-\frac{1}{n}\log f_n\left(X^n\middle|\theta^*+\frac{h}{\sqrt{n}}\right)-\left(-\frac{1}{n}\log f_n\left(X^n|\theta^*\right)\right)\right]\right)dh\\
&\approx n^{\frac{p}{2}}\int_{\mathbb{R}^p}\left[\|h\|_2^{1+\gamma}\pi(\theta^*)+\frac{1}{\sqrt{n}}h^\top\nabla(\|\theta\|_2^{1+\gamma}\pi(\theta))\big|_{\theta=\theta^*}\right]\left[1+\alpha_n h^\top V_{\theta^*}\Delta_{n,\theta^*}\right]\exp\left(-\frac{1}{2}h^\top\left(\alpha_n V_{\theta^*}\right)h\right)dh.
\end{aligned}\tag{144}$$

We examine the leading order term

$$\begin{aligned}
&n^{-\frac{p}{2}}\pi(\theta^*)\int_{\mathbb{R}^p}\|h\|_2^{1+\gamma}\exp\left(-\frac{1}{2}h^\top\left(\alpha_n V_{\theta^*}\right)h\right)dh\\
&=\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}}|V_{\theta^*}|^{-1}\pi(\theta^*)\int_{\mathbb{R}^p}(2\pi)^{-\frac{p}{2}}|V_{\theta^*}|\|h\|_2^{1+\gamma}\exp\left(-\frac{1}{2}h^\top\left(\alpha_n V_{\theta^*}\right)h\right)dh.
\end{aligned}\tag{145}$$

Consider the simple case where $p=1$, so $\|h\|_2^{1+\gamma}=|h|^{\frac{1+\gamma}{2}}$. Then the expectation in (145) becomes

$$\frac{2^{\frac{1+3\gamma}{2}-1}p^{\frac{1+\gamma}{2}-1}}{\sqrt{\pi}}\Gamma\left(\frac{2+\gamma}{2}\right)\left[\frac{V_{\theta^*}^{-1}}{\alpha_n}\right]^{\frac{1+\gamma}{2}}.\tag{146}$$

Hence,

$$\tilde{Z}\sim\exp\left(\alpha_n\log f_n(X^n|\theta^*)\right)\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}}|V_{\theta^*}|^{-1}\pi(\theta^*)\frac{2^{\frac{1+3\gamma}{2}-1}p^{\frac{1+\gamma}{2}-1}}{\sqrt{\pi}}\Gamma\left(\frac{2+\gamma}{2}\right)\left[\frac{V_{\theta^*}^{-1}}{\alpha_n}\right]^{\frac{1+\gamma}{2}}$$

and

$$\begin{aligned}
\frac{\tilde{Z}}{Z}&\sim\frac{\exp\left(\alpha_n\log f_n(X^n|\theta^*)\right)\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}}|V_{\theta^*}|^{-1}\pi(\theta^*)\frac{2^{\frac{1+3\gamma}{2}-1}p^{\frac{1+\gamma}{2}-1}}{\sqrt{\pi}}\Gamma\left(\frac{2+\gamma}{2}\right)\left[\frac{V_{\theta^*}^{-1}}{\alpha_n}\right]^{\frac{1+\gamma}{2}}}{\exp\left(\alpha_n\log f_n(X^n|\theta^*)\right)\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}}|V_{\theta^*}|^{-1}\pi(\theta^*)}\\
&=\frac{2^{\frac{1+3\gamma}{2}-1}p^{\frac{1+\gamma}{2}-1}}{\sqrt{\pi}}\Gamma\left(\frac{2+\gamma}{2}\right)\left[\frac{V_{\theta^*}^{-1}}{\alpha_n}\right]^{\frac{1+\gamma}{2}}\to\infty
\end{aligned}$$

D Proofs of Main Results

D.1 Proof of Theorem 1

Proof. We will show that the total variation distance between the α_n-posterior, $\pi_{n,\alpha_n}(\theta|X^n)$, and a Gaussian centered around $\hat{\theta}$ with covariance $\frac{1}{\alpha_n n}V_{\theta^*}^{-1}$ converges to 0 in $f_{0,n}$-probability. Our approach adapts the proof of [4, Theorem 1], which proves asymptotic normality of the α-posterior and follows closely to that of [40, Theorem 2.1]. As in [4], by Lemma 5, the total variation distance between the α_n-posterior and its limiting Gaussian can be expressed in terms of their ratios ((18)–(19)) and their tail probabilities. As discussed in Section 4.1, we appeal to Lemma 3 to conclude that (18) and (19) are equal up to $o_{f_{0,n}}(1)$ terms. Finally, the tail probabilities of the α_n-posterior and its limiting Gaussian vanish by **(A3)** and [40, Lemma 5.2], respectively.

By invariance of the total variation distance to simultaneous centering and scaling, it suffices to study the total variation distance between the two distributions after centering them around θ^* and scaling them by $\sqrt{\alpha_n n}$. Denote these transformed densities by

$$\pi_{n,\alpha_n}^{LAN}(h) \equiv \frac{1}{(\alpha_n n)^{\frac{p}{2}}}\pi_{n,\alpha_n}\left(\theta^* + \frac{h}{\sqrt{\alpha_n n}}\,\Big|\, X^n\right), \quad \text{and} \quad \phi_n(h) \equiv \frac{1}{(\alpha_n n)^{\frac{p}{2}}}\phi\left(h \,\big|\, \sqrt{\alpha_n n}(\hat{\theta}-\theta^*), V_{\theta^*}^{-1}\right). \tag{147}$$

We note that the multivariate densities in (147) are proper densities. Indeed, using the change of variable $h = \sqrt{\alpha_n n}(\theta - \theta^*)$, it is straightforward to show that they integrate to one.

To control the total variation distance between the α_n-posterior and its limiting Gaussian, we note that by Markov's inequality we have

$$\mathbb{P}_{f_{0,n}}\left(\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}^{LAN}(h), \phi_n(h)\right) > t\right) \leq \frac{1}{t}\mathbb{E}_{f_{0,n}}\left[\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}^{LAN}(h), \phi_n(h)\right)\right]. \tag{148}$$

Hence, it suffices to show that the expectation on the right hand side of (148) converges to zero. We control this expectation by (i) controlling it conditioned on a well chosen event $\mathcal{A}$ (defined in (150)) and (ii) showing its complement $\mathcal{A}^{\mathsf{c}}$ occurs with vanishing probability.

For vectors $g, h \in K$, where K is a compact subset of $\mathbb{R}^p$, define the random variable

$$f_n^+(g,h) = \left\{1 - \frac{\phi_n(h)\,\pi_{n,\alpha_n}^{LAN}(g|X^n)}{\pi_{n,\alpha_n}^{LAN}(h|X^n)\,\phi_n(g)}\right\}^+, \tag{149}$$

where $\{x\}^+ = \max\{0,x\}$. For a constant $C > 0$ sufficiently large, define the sequence $\tilde{r}_n := \max\{r_n, C\sqrt{\alpha_n n}\epsilon_n\}$, where ϵ_n is the sequence from **(A3)** and r_n is the sequence from Lemma 3. For some $\eta > 0$ (that will be specified later), define the event

$$\mathcal{A} = \left\{\sup_{g,h \in \bar{B}_0(\tilde{r}_n)} f_n^+(g,h) \leq \eta\right\}. \tag{150}$$

We notice that for each n, the set $\bar{B}_0(\tilde{r}_n)$ is compact. To guarantee that $f_n^+(g,h)$ is well defined on K with high probability under event $\mathcal{A}$, we need to ensure we are not dividing by zero. This is shown in Lemma 3.

Because $\mathrm{d_{TV}}(\cdot,\cdot) \leq 1$, the following holds:

$$\begin{aligned}\mathbb{E}_{f_{0,n}}\left[\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}^{LAN}(h), \phi_n(h)\right)\right] &= \mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\}\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}^{LAN}(h), \phi_n(h)\right)\right] \\ &\quad + \mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}^{\mathsf{c}}\}\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}^{LAN}(h), \phi_n(h)\right)\right] \\ &\leq \mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\}\mathrm{d_{TV}}\left(\pi_{n,\alpha_n}^{LAN}(h), \phi_n(h)\right)\right] + \mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}).\end{aligned} \tag{151}$$

We aim to show that each term on the right side of (151) converges to zero.

Bounding (151)**.** For all $\eta > 0$ and all $\epsilon > 0$, there exists $N_0 := N_0(\eta, \epsilon)$ such that for all $n > N_0$,

$$\mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}) = \mathbb{P}_{f_{0,n}}\left(\sup_{g,h \in \bar{B}_0(\tilde{r}_n)} f_n^+(g,h) > \eta \right) < \epsilon. \tag{152}$$

This follows from Lemma 3, which shows there exists a sequence $r_n \to \infty$ such that

$$\mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}) = \mathbb{P}_{f_{0,n}}\left(\sup_{g,h \in \bar{B}_0(r_n)} f_n^+(g,h) > \eta \right) < \epsilon \tag{153}$$

and the fact that $r_n \leq \tilde{r}_n$ by definition.

Now we upper bound the first term on the right side of (151). By Lemma 5 using $s(h) = 1/2$ and $K = \bar{B}_0(\tilde{r}_n)$, for n large enough to ensure positivity of π^{LAN}_{n,α_n},

$$\begin{aligned} \mathrm{d_{TV}}\left(\pi^{LAN}_{n,\alpha_n}(h), \phi_n(h)\right) &= \frac{1}{2}\int_{\mathbb{R}^p} \left|\pi^{LAN}_{n,\alpha_n}(h|X^n) - \phi_n(h)\right| dh \\ &\leq \frac{1}{2} \sup_{g,h \in \bar{B}_0(\tilde{r}_n)} f_n^+(g,h) + \frac{1}{2}\int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \pi^{LAN}_{n,\alpha_n}(h|X^n) dh + \frac{1}{2}\int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \phi_n(h) dh. \end{aligned} \tag{154}$$

Therefore,

$$\begin{aligned} &\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\}\mathrm{d_{TV}}\left(\pi^{LAN}_{n,\alpha_n}(h), \phi_n(h)\right)\right] \\ &\quad\leq \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\} \sup_{g,h \in \bar{B}_0(\tilde{r}_n)} f_n^+(g,h)\right] + \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\} \int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \pi^{LAN}_{n,\alpha_n}(h|X^n) dh\right] \\ &\qquad + \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\} \int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \phi_n(h) dh\right]. \end{aligned} \tag{155}$$

First, we note that under event $\mathcal{A}$, the first term of (155) is upper bounded by

$$\frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\} \sup_{g,h \in \bar{B}_0(\tilde{r}_n)} f_n^+(g,h)\right] \leq \frac{\eta}{2}.$$

Since $0 \leq \mathbb{1}\{\mathcal{A}\} \leq 1$, we can upper bound the sum of the second and third terms of (155) by

$$\begin{aligned} &\frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\} \int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \pi^{LAN}_{n,\alpha_n}(h|X^n) dh\right] + \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\} \int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \phi_n(h) dh\right] \\ &\quad\leq \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \pi^{LAN}_{n,\alpha_n}(h|X^n) dh\right] + \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\int_{h \in \bar{B}_0(\tilde{r}_n)^{\mathsf{c}}} \phi_n(h) dh\right]. \end{aligned} \tag{156}$$

Label the terms on the right hand side of (156) as T_1 and T_2, respectively. We argue that, for arbitrary $\epsilon > 0$, each can be upper bounded by $\epsilon/2$.

Consider term T_1 in (156), which we claim can be made arbitrarily small by the posterior concentration assumption, **(A3)**. Using the change of variable $h = \sqrt{\alpha_n n}(\theta - \theta^*)$, we have

$$
\begin{aligned}
T_1 &= \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\int_{h \in \bar{B}_0(\tilde{r}_n)^c} \pi^{LAN}_{n,\alpha_n}(h|X^n)dh\right] \\
&= \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\int \mathbb{1}\{||\sqrt{\alpha_n n}(\theta-\theta^*)||_2 > \tilde{r}_n\}\pi_{n,\alpha_n}(\theta|X^n)d\theta\right] \\
&\le \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\int \mathbb{1}\{||\sqrt{\alpha_n n}(\theta-\theta^*)||_2 > C\sqrt{\alpha_n n}\epsilon_n\}\pi_{n,\alpha_n}(\theta|X^n)d\theta\right] \\
&= \frac{1}{2}\mathbb{E}_{f_{0,n}}\left[\int \mathbb{1}\{||(\theta-\theta^*)||_2 > C\epsilon_n\}\pi_{n,\alpha_n}(\theta|X^n)d\theta\right]
\end{aligned}
\tag{157}
$$

By **(A3)**, there exists $N_1 := N_1(\epsilon)$ such that for $n > N_1$, we have $T_1 < \epsilon/2$.

Now we bound term T_2 in (156). By **(A0)** and the fact that $\alpha_n = o_{f_{0,n}}(1)$, we have $\sqrt{\alpha_n n}(\hat{\theta} - \theta^*) = o_{f_{0,n}}(1)$. Thus, the sequence of means of $\phi_n(h)$ is uniformly tight and by [40, Lemma 5.2], there exists an $N_2 := N_2(\eta, \epsilon)$ such that $T_2 < \epsilon/2$ when $n > N_2(\eta, \epsilon)$.

Therefore, we have shown that whenever $n \ge \max(N_1, N_2)$,

$$
\mathbb{E}_{f_{0,n}}\left[\mathbb{1}\{\mathcal{A}\}\mathrm{d_{TV}}\left(\pi^{LAN}_{n,\alpha_n}(h), \phi_n(h)\right)\right] < \frac{\eta}{2} + \frac{\epsilon}{2} + \frac{\epsilon}{2} = \frac{\eta}{2} + \epsilon. \tag{158}
$$

Final bound. By using (158) and (152) in (151), for $\epsilon > 0$, $\eta > 0$, and $n > \max(N_0, N_1, N_2)$, we have

$$
\mathbb{E}\left[\int_{\mathbb{R}^p}\left|\pi^{LAN}_{n,\alpha_n}(h|X^n) - \phi_n(h)\right|dh\right] < \frac{\eta}{2} + 2\epsilon,
$$

which gives us the desired result by (148).

□

D.2 Proof of Theorem 2

The proof closely follows the proof of [54, Theorem 1], which adapts arguments from [4, Theorem 1] and [40, Theorem 2.1] to show convergence of the posterior moments rather than the total variation distance. Let Z_0 denote the integral we are trying to control, namely

$$
Z_0 := \int_{\mathbb{R}^p}\left\|h^{\otimes k}\right\|_1\left|\pi^{LAN}_{n,\alpha_n}(h) - \phi_n(h)\right|dh, \tag{159}
$$

where the transformed densities are defined in (147). Recalling the definition of $\left\|h^{\otimes k}\right\|_1$ in (33) and using that $\|h\|_1 \le \sqrt{p}\|h\|_2$, we have for fixed p that

$$
\left\|h^{\otimes k}\right\|_1 = \sum_{i_1=1}^{p}\cdots\sum_{i_k=1}^{p}|h_{i_1}\times\cdots\times h_{i_k}| = \sum_{i_1=1}^{p}|h_{i_1}|\cdots\sum_{i_k=1}^{p}|h_{i_k}| = \|h\|_1^k \le p^{\frac{k}{2}}\|h\|_2^k. \tag{160}
$$

By the bound in 160 and monotonicity of the integral, we have that

$$
Z_0 \le p^{\frac{k}{2}}\int_{\mathbb{R}^p}\|h\|_2^k\left|\pi^{LAN}_{n,\alpha_n}(h) - \phi_n(h)\right|dh := Z. \tag{161}
$$

By Markov's inequality and the bound established in (161), we have

$$
\mathbb{P}_{f_{0,n}}(|Z_0| > t) \le \frac{1}{t}\mathbb{E}_{f_{0,n}}[Z_0] \le \frac{1}{t}\mathbb{E}_{f_{0,n}}[Z]. \tag{162}
$$

Hence, it suffices to show that the expectation on the right hand side of (162) converges to 0. We control this expectation by (i) controlling it conditioned on the event $\mathcal{A}$ (defined in (150)) and (ii) showing the complement of the event $\mathcal{A}^{\mathsf{c}}$ occurs with vanishing probability, which we do in Lemma 3. Next, we upper bound $\mathbb{E}_{f_{0,n}}[Z]$ as follows using Hölder's Inequality for some $\gamma > 0$:

$$\begin{aligned} \mathbb{E}_{f_{0,n}}[Z] &= \mathbb{E}_{f_{0,n}}[Z\mathbb{1}\{\mathcal{A}\}] + \mathbb{E}_{f_{0,n}}[Z\mathbb{1}\{\mathcal{A}^{\mathsf{c}}\}] \\ &\leq \mathbb{E}_{f_{0,n}}[Z\mathbb{1}\{\mathcal{A}\}] + \mathbb{E}_{f_{0,n}}[Z^{1+\gamma}]^{\frac{1}{1+\gamma}} \left(\mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}})\right)^{\frac{\gamma}{1+\gamma}}. \end{aligned} \tag{163}$$

We aim to show that each term in (163) converges to zero.

First term in (163)**.** We upper bound the first term of (163), which is $\mathbb{E}_{f_{0,n}}[Z\mathbb{1}\{\mathcal{A}\cap\mathcal{B}\}]$. Using (161) and appealing to Lemma 5 with $s(h) = p^{\frac{k}{2}}\|h\|_2^k$, we have that

$$\begin{aligned} Z &= p^{\frac{k}{2}} \int_{\mathbb{R}^p} \|h\|_2^k \left|\pi_{n,\alpha_n}^{LAN}(h|X^n) - \phi_n(h)\right| dh \\ &\leq p^{\frac{k}{2}} \left[\sup_{g,h\in\bar{B}_0(r_n)} f_n^+(g,h)\right] \int_{\mathbb{R}^p} \|h\|_2^k \left(\pi_{n,\alpha_n}^{LAN}(h|X^n) + \phi_n(h)\right) dh \\ &\quad + p^{\frac{k}{2}} \int_{\bar{B}_0(r_n)^{\mathsf{c}}} \|h\|_2^k \left(\pi_{n,\alpha_n}^{LAN}(h|X^n) + \phi_n(h)\right) dh. \end{aligned} \tag{164}$$

Therefore, to study $\mathbb{E}_{f_{0,n}}[Z\mathbb{1}\{\mathcal{A}\}]$, we notice that the suprema in (164) is bounded. This gives the upper bound

$$\begin{aligned} \mathbb{E}_{f_{0,n}}[Z\mathbb{1}\{\mathcal{A}\}] &\leq p^{\frac{k}{2}}\eta\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right] + p^{\frac{k}{2}}\eta\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh\right] \\ &\quad + p^{\frac{k}{2}}\mathbb{E}_{f_{0,n}}\left[\int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\phi_n(h)dh\right] + p^{\frac{k}{2}}\mathbb{E}_{f_{0,n}}\left[\int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right]. \end{aligned} \tag{165}$$

Label the four terms in (165) as T_1, T_2, T_3, and T_4. We argue that, for arbitrary $\epsilon > 0$, each can be upper bounded by $\epsilon/4$ (possibly for large enough n).

Consider term T_1 of (165). Recalling that $k \in [1, k_0]$, by **(A3')**, there exists $M_1 < \infty$ such that

$$\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^k\,\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right] < M_1.$$

Thus, we can pick $\eta < \frac{\epsilon}{4M_1p^{k/2}}$ such that

$$T_1 = p^{k/2}\eta\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^k\,\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right] < \frac{\epsilon}{4}.$$

For term T_2, we notice that by the change of variable $h = \sqrt{\alpha_n n}(\theta - \hat{\theta} + \theta^*)$,

$$\begin{aligned} \int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh &= \int_{\mathbb{R}^p}(\alpha_n n)^{-\frac{p}{2}}\|h\|_2^k\phi\left(h\,\middle|\,\sqrt{\alpha_n n}(\hat{\theta}-\theta^*), V_{\theta^*}^{-1}\right)dh \\ &= \int_{\mathbb{R}^p}(\alpha_n n)^{\frac{k}{2}}\|\theta-\hat{\theta}+\theta^*\|_2^k\phi\left(\theta\,\middle|\,0, V_{\theta^*}^{-1}/(n\alpha_n)\right)d\theta \\ &= (\alpha_n n)^{\frac{k}{2}}\mathbb{E}_{\mathcal{N}(0,V_{\theta^*}^{-1}/(n\alpha_n))}\left[\|\theta-\hat{\theta}+\theta^*\|_2^k\right]. \end{aligned} \tag{166}$$

We upper bound the above using Hölder's Inequality[1].

$$
\begin{aligned}
(\alpha_n n)^{\frac{k}{2}}\mathbb{E}\left[\|\theta - \hat{\theta} + \theta^*\|_2^k\right] &= (\alpha_n n)^{\frac{k}{2}}\mathbb{E}\left[\left(\sum_{i=1}^p (\theta_i - \hat{\theta}_i + \theta_i^*)^2\right)^{\frac{k}{2}}\right] \\
&\leq (\alpha_n n)^{\frac{k}{2}} p^{\frac{k}{2}-1}\sum_{i=1}^p \mathbb{E}\left[|\theta_i - \hat{\theta}_i + \theta_i^*|^k\right] \\
&\leq (\alpha_n n)^{\frac{k}{2}} 2^{k-1} p^{\frac{k}{2}-1}\sum_{i=1}^p \mathbb{E}\left[|\theta_i|^k\right] + (\alpha_n n)^{\frac{k}{2}} 2^{k-1} p^{\frac{k}{2}-1}\sum_{i=1}^p |\hat{\theta}_i - \theta_i^*|^k.
\end{aligned}
\tag{167}
$$

Using centered absolute Gaussian moments, the first term on the right side of (167) equals

$$
\begin{aligned}
(\alpha_n n)^{\frac{k}{2}} 2^{k-1} p^{\frac{k}{2}-1}\sum_{i=1}^p \mathbb{E}[|\theta_i|^k] &= (\alpha_n n)^{\frac{k}{2}} \frac{2^{\frac{3k}{2}-1} p^{\frac{k}{2}-1}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right)\sum_{i=1}^p \left[\frac{[V_{\theta^*}^{-1}]_{ii}}{\alpha_n n}\right]^{\frac{k}{2}} \\
&= \frac{2^{\frac{3k}{2}-1} p^{\frac{k}{2}-1}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right)\sum_{i=1}^p \left[[V_{\theta^*}^{-1}]_{ii}\right]^{\frac{k}{2}},
\end{aligned}
\tag{168}
$$

and the second term on the right side of (167) equals $2^{k-1}p^{\frac{k}{2}-1}\sum_{i=1}^p(\sqrt{\alpha_n n}|\hat{\theta}_i - \theta_i^*|)^k$. Because $\sqrt{n}(\hat{\theta}_i - \theta_i^*)$ is asymptotically normal by Assumption **(A0)**, and $\alpha_n = o(1)$ by assumption, there exists an $N_1 := N_1(p, k, 1)$ such that for all $n > N_1$, the second term on the right side of (167) can be made arbitrarily small, say, less than one, with high probability. From (166)–(168) we have shown for all $n > \max(N_0, N_1)$,

$$
\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h)dh\right] \leq M_2,
\tag{169}
$$

where

$$
M_2 := \frac{2^{\frac{3k}{2}-1} p^{\frac{k}{2}-1}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right)\sum_{i=1}^p \left[[V_{\theta^*}^{-1}]_{ii}\right]^{\frac{k}{2}} + 1.
$$

Putting this together, we have

$$
T_2 = p^{\frac{k}{2}}\eta\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h)dh\right] \leq p^{\frac{k}{2}}\eta M_2.
$$

Hence, by choosing $\eta < \epsilon/(4M_2 p^{\frac{k}{2}})$ we have $T_2 \leq \epsilon/4$.

We control T_3 by leveraging (169). Indeed, by Cauchy-Schwarz and Jensen's Inequality,

$$
\begin{aligned}
T_3 &= p^{\frac{k}{2}}\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} \|h\|_2^k (1 - \mathbb{1}\{\|h\|_2 \leq r_n\})\phi_n(h)dh\right] \\
&\leq p^{\frac{k}{2}}\mathbb{E}_{f_{0,n}}\left[\sqrt{\int_{\mathbb{R}^p} \|h\|_2^{2k}\phi_n(h)dh} \times \sqrt{\int_{\mathbb{R}^p} (1 - \mathbb{1}\{\|h\|_2 \leq r_n\})\phi_n(h)dh}\right] \\
&\leq p^{\frac{k}{2}}\sqrt{\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} \|h\|_2^{2k}\phi_n(h)dh\right]} \times \sqrt{\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} (1 - \mathbb{1}\{\|h\|_2 \leq r_n\})\phi_n(h)dh\right]}.
\end{aligned}
\tag{170}
$$

[1]Note that for $m > 0$ and $v = (v_1, \ldots, v_p) \in \mathbb{R}^p$, Hölder's inequality implies $(\sum_{j=1}^p |v_j|)^m \leq p^{m-1}\sum_{j=1}^p |v_j|^m$

By appealing to (169) with the power $2k$ instead of k, we know that $\mathbb{E}_{f_{0,n}}[\int_{\mathbb{R}^p} \|h\|_2^{2k}\phi_n(h)dh] \leq \tilde{M}_2$ for $n > N_1$. Because $r_n \to \infty$ and $\phi_n(h) \leq (2\pi)^{-\frac{p}{2}}|V_{\theta^*}^{-1}/\alpha|^{-1/2}$ for all $h \in \mathbb{R}^p$, we see that $(1 - \mathbb{1}\{\|h\|_2 \leq r_n\})\phi_n(h) \to 0$. Thus, by the dominated convergence theorem, there exists $N_2 \coloneqq N_2(\epsilon, \tilde{M}_2, p, k)$ such that for all $n > N_2$,

$$\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}(1 - \mathbb{1}\{\|h\|_2 \leq r_n\})\phi_n(h)dh\right] < \frac{\epsilon^2}{16\tilde{M}_2p^k}.$$

Thus, for $n > \max(N_1, N_2)$, we have that $T_3 < \frac{\epsilon}{4}$.

Finally, to bound T_4, by appealing to Lemma 6 (assumptions are met by appealing to **(A3')**) with $f_Z(z)$ being $\pi_{n,\alpha_n}^{LAN}(z|X^n)$, we pick $N_3 \coloneqq N_3(p, k, \epsilon, \gamma)$ such that for all $n > N_3$,

$$\mathbb{E}_{f_{0,n}}\left[\int_{\bar{B}_0(r_n)^c}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right] < \frac{\epsilon}{4p^{k/2}}.$$

Thus, for $n > N_3$,

$$T_4 = p^{\frac{k}{2}}\mathbb{E}_{f_{0,n}}\left[\int_{\|h\|_2 > r_n}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right] < \frac{\epsilon}{4}.$$

Therefore, we have shown that the first term of (163) is upperbounded by ϵ whenever $\eta < \frac{\epsilon}{4\min\{M_1,M_2\}p^{\frac{k}{2}}}$ and $n \geq \max\{N_1, N_2, N_3\}$.

Second term in (163)**.** First, we consider upper bounding the term $\mathbb{E}_{f_{0,n}}[Z^{1+\gamma}]$. Notice that by the fact that $|\pi_{n,\alpha_n}^{LAN}(h|X^n) - \phi_n(h)| \leq \pi_{n,\alpha_n}^{LAN}(h|X^n) + \phi_n(h)$,

$$\begin{aligned}
Z^{1+\gamma} &= \left[p^{\frac{k}{2}}\int_{\mathbb{R}^p}\|h\|_2^k\left|\pi_{n,\alpha_n}^{LAN}(h|X^n) - \phi_n(h)\right|dh\right]^{1+\gamma} \\
&\leq \left[2p^{\frac{k}{2}}\max\left\{\int_{\mathbb{R}^p}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh, \quad \int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh\right\}\right]^{1+\gamma},
\end{aligned}$$

Next, applying Jensen's Inequality, we have that

$$\begin{aligned}
&\left[\max\left\{\int_{\mathbb{R}^p}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh, \int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh\right\}\right]^{1+\gamma} \\
&= \max\left\{\left(\int_{\mathbb{R}^p}\|h\|_2^k\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right)^{1+\gamma}, \quad \left(\int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh\right)^{1+\gamma}\right\} \\
&\leq \max\left\{\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\pi_{n,\alpha_n}^{LAN}(h|X^n)dh, \quad \int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\phi_n(h)dh\right\} \\
&\leq \int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}(\pi_{n,\alpha_n}^{LAN}(h|X^n) + \phi_n(h))dh.
\end{aligned}$$

Therefore, we have the bound

$$\mathbb{E}_{f_{0,n}}[Z^{1+\gamma}] \leq \left(2p^{\frac{k}{2}}\right)^{1+\gamma}\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\phi_n(h)dh\right] + \left(2p^{\frac{k}{2}}\right)^{1+\gamma}\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\pi_{n,\alpha_n}^{LAN}(h|X^n)dh\right].$$

We argue that the two terms on the right side of the above are upper bounded by constants. By Assumption **(A3)**, there exists an $M_3 < \infty$ such that $\mathbb{E}_{f_{0,n}}[\int_{\mathbb{R}^p} \|h\|_2^{k(1+\gamma)} \pi_{n,\alpha_n}^{LAN}(h|X^n)dh] < M_3$. By appealing to (169) with $k(1+\gamma)$ instead of k, we know that for $n > N_1$,

$$\mathbb{E}_{f_{0,n}}\left[\int_{\mathbb{R}^p} \|h\|_2^{k(1+\gamma)} \phi_n(h)dh\right] < M_4.$$

Thus, we have shown that

$$\begin{aligned}\mathbb{E}_{f_{0,n}}[Z^{1+\gamma}] &\leq (2p^{k/2})^{1+\gamma}(M_3 + M_4) \\ &\leq 2(2p^{k/2})^{1+\gamma} \max\{M_3, M_4\}.\end{aligned}$$

We finally upper bound $\mathbb{P}(\mathcal{A}^{\mathsf{c}})$ and $\mathbb{P}(\mathcal{B}^{\mathsf{c}})$. By Lemma 3, there exists $N_4 := N_4(\eta, \epsilon, M_3, M_4, \gamma, p, k)$ such that $\mathbb{P}(\mathcal{A}^{\mathsf{c}}) = \mathbb{P}(\mathcal{B}^{\mathsf{c}})$ is sufficiently small for all $n > N_4$:

$$\mathbb{P}(\mathcal{A}^{\mathsf{c}}) = \mathbb{P}(\mathcal{B}^{\mathsf{c}}) < \frac{\epsilon^{\frac{1+\gamma}{\gamma}}}{2^{1+\frac{2+\gamma}{\gamma}} p^{\left(\frac{k}{2}\right)\left(\frac{1+\gamma}{\gamma}\right)} \max\{M_3, M_4\}^{\frac{1}{\gamma}}}.$$

Putting this together, we find that for $n > N_4$,

$$\begin{aligned}&\mathbb{E}_{f_{0,n}}[Z^{1+\gamma}]^{\frac{1}{1+\gamma}} \left(\mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{B}^{\mathsf{c}})\right)^{\frac{\gamma}{1+\gamma}} \\ &\leq [2(2p^{k/2})^{1+\gamma} \max\{M_3, M_4\}]^{\frac{1}{1+\gamma}} \times \left[\frac{\epsilon^{\frac{1+\gamma}{\gamma}}}{2^{\frac{2+\gamma}{\gamma}} p^{\left(\frac{k}{2}\right)\left(\frac{1+\gamma}{\gamma}\right)} \max\{M_3, M_4\}^{\frac{1}{\gamma}}}\right]^{\frac{\gamma}{1+\gamma}} = \epsilon.\end{aligned}$$

Final bound. We have shown that whenever η is chosen small enough, for all $n > \max(N_0, N_1, N_2, N_3, N_4)$, for arbitrary $\epsilon > 0$,

$$\mathbb{E}[Z] = p^{k/2}\mathbb{E}\left[\int_{\mathbb{R}^p} \|h\|_2^k \left|\pi_{n,\alpha_n}^{LAN}(h|X^n) - \phi_n(h)\right| dh\right] < 2\epsilon.$$

This gives us the desired result.

D.3 Proof of Theorem 3

Recall the definition of $\hat{\theta}^B$:

$$\hat{\theta}^B = \int_{\mathbb{R}^p} \theta \pi_{n,\alpha_n}(\theta|X^n)d\theta = \frac{\int_{\mathbb{R}^p} \theta\pi(\theta) \exp\left(-\alpha_n n \left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right) d\theta}{\int_{\mathbb{R}^p} \pi(\theta) \exp\left(-\alpha_n n \left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right) d\theta}. \tag{171}$$

We have written $\hat{\theta}^B$ in a form where we can apply Lemma 1 to (171), where $g(\theta) = \theta_j$ for $1 \leq j \leq p$ in the numerator of (171) and $g(\theta) = 1$ in the denominator of (171). Therefore, if the conditions of Lemma 1 hold, we have that

$$\hat{\theta}_j^B - \hat{\theta}_j = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right) \implies \sqrt{n}(\hat{\theta}_j^B - \hat{\theta}_j) = O_{f_{0,n}}\left(\frac{1}{\alpha_n\sqrt{n}}\right) = o_{f_{0,n}}(1). \tag{172}$$

The implication in (172) follows from assumption that $\alpha_n\sqrt{n} \to \infty$. It remains to verify the conditions of Lemma 1, which amounts to verifying that assumption **(A4")** holds for the above specified choices of $b(\theta) = \pi(\theta)$ and $b(\theta) = \theta\pi(\theta)$.

Clearly $\pi(\theta)$ is integrable since π is a density and $g(\theta) = 1$ is trivially finite and twice continuously differentiable everywhere. Hence, assumption **(A4")** holds for $b(\theta) = \pi(\theta)$ by Proposition 4. Furthermore, $g(\theta) = \theta$ is finite in a neighborhood of θ^* and twice continuously differentiable everywhere and by assumption **(A1")**, $b(\theta) = \theta\pi(\theta)$ is integrable. Hence, assumption **(A4")** holds for $b(\theta) = \theta\pi(\theta)$ by Proposition 4.

D.4 Proof of Proposition 7

By definition of the ∞-posterior in (53),

$$\begin{aligned} \mathrm{d_{TV}}\left(\pi_{n,\infty}(\theta|X^n), \delta_{\hat{\theta}}\right) &= \mathrm{d_{TV}}\left(\lim_{\alpha\to\infty} \pi_{n,\alpha}(\theta|X^n), \delta_{\hat{\theta}}\right) \\ &= \lim_{\alpha\to\infty} \mathrm{d_{TV}}\left(\pi_{n,\alpha}(\theta|X^n), \delta_{\hat{\theta}}\right). \end{aligned} \tag{173}$$

Since $\mathrm{d_{TV}}(\cdot,\cdot) \leq 1$, the last equality in (173) follows from the dominated convergence theorem. By definition of total variation distance,

$$\begin{aligned} &\mathrm{d_{TV}}\left(\pi_{n,\alpha}(\theta|X^n), \delta_{\hat{\theta}}\right) \\ &= \sup_A \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A\}\pi_{n,\alpha}(\theta|X^n)d\theta - \delta_{\hat{\theta}}(A) \right| \\ &\leq \sup_{\{A:\hat{\theta}\in A\}} \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A\}\pi_{n,\alpha}(\theta|X^n)d\theta - \delta_{\hat{\theta}}(A) \right| + \sup_{\{A:\hat{\theta}\notin A\}} \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \notin A\}\pi_{n,\alpha}(\theta|X^n)d\theta - \delta_{\hat{\theta}}(A) \right| \\ &= \sup_{\{A:\hat{\theta}\in A\}} \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A\}\pi_{n,\alpha}(\theta|X^n)d\theta - 1 \right| + \sup_{\{A:\hat{\theta}\notin A\}} \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \notin A\}\pi_{n,\alpha}(\theta|X^n)d\theta \right| \\ &= \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A^*\}\pi_{n,\alpha}(\theta|X^n)d\theta - 1 \right| + \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \notin A^{**}\}\pi_{n,\alpha}(\theta|X^n)d\theta \right|, \end{aligned} \tag{174}$$

for some A^* such that $\hat{\theta} \in A^*$ and A^{**} such that $\hat{\theta} \notin A^{**}$. For some set A, we have

$$\int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A\}\pi_{n,\alpha}(\theta|X^n)d\theta = \frac{\int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A\}\pi(\theta)\exp\left(-\alpha n\left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right)d\theta}{\int_{\mathbb{R}^p} \pi(\theta)\exp\left(-\alpha n\left(-\frac{1}{n}\log f_n(X^n|\theta)\right)\right)d\theta}. \tag{175}$$

Since $\alpha \cdot n \to \infty$ as $\alpha \to \infty$, we have written (175) in a form where we can apply Lemma 1 to $\mathbb{1}\{\theta \in B_{\hat{\theta}}(\epsilon)\}$ in the numerator of (175) and $g(\theta) = 1$ in the denominator of (175). Therefore, if the conditions of Lemma 1 hold, we have that

$$\int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A^*\}\pi_{n,\alpha}(\theta|X^n)d\theta = \mathbb{1}\{\hat{\theta} \in A^*\} + O_{f_{0,n}}\left(\frac{1}{\alpha n}\right). \tag{176}$$

Similarly,

$$\int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A^{**}\}\pi_{n,\alpha}(\theta|X^n)d\theta = \mathbb{1}\{\hat{\theta} \in A^{**}\} + O_{f_{0,n}}\left(\frac{1}{\alpha n}\right). \tag{177}$$

Applying (176) and (177) to (174) yields

$$\begin{aligned} &\mathrm{d_{TV}}\left(\pi_{n,\alpha}(\theta|X^n), \delta_{\hat{\theta}}\right) \\ &\leq \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \in A^*\}\pi_{n,\alpha}(\theta|X^n)d\theta - 1 \right| + \left| \int_{\mathbb{R}^p} \mathbb{1}\{\theta \notin A^{**}\}\pi_{n,\alpha}(\theta|X^n)d\theta \right| \\ &= \left| \mathbb{1}\{\hat{\theta} \in A^*\} + O_{f_{0,n}}\left(\frac{1}{\alpha n}\right) - 1 \right| + \left| \mathbb{1}\{\hat{\theta} \in A^{**}\} + O_{f_{0,n}}\left(\frac{1}{\alpha n}\right) \right| \\ &= O_{f_{0,n}}\left(\frac{1}{\alpha n}\right). \end{aligned} \tag{178}$$

By taking $\alpha \to \infty$, we conclude that

$$\mathrm{d_{TV}}\left(\pi_{n,\alpha}(\theta|X^n), \delta_{\hat{\theta}}\right) \to 0 \tag{179}$$

in $f_{0,n}$-probability. By (173), we have proved that for each fixed n, $\pi_{n,\alpha}(\theta|X^n) \to \delta_{\hat{\theta}}$ in total variation distance. By further taking $n \to \infty$, we conclude (56).

It remains to verify the conditions of Lemma 1, which amounts to verifying that assumption **(A4")** holds for the above specified choices of $b(\theta) = \pi(\theta)$ and $b(\theta) = \mathbb{1}\{\theta \in A^*\}\pi(\theta)$, and $b(\theta) = \mathbb{1}\{\theta \in A^{**}\}\pi(\theta)$. Note that we have we have already done so for $b(\theta) = \pi(\theta)$ in the proof of Theorem 3.

We have that $g(\theta) = \mathbb{1}\{\theta \in A^*\}$ and $g(\theta) = \mathbb{1}\{\theta \in A^{**}\}$ are upper bounded by 1 and, hence, are trivially finite everywhere. Furthermore, $\int g(\theta)\pi(\theta)d\theta$ is a probability measure for both choices of g. Hence, $b(\theta)$ is clearly integrable.

Finally, note that both choices of $g(\theta)$ are twice continuously differentiable in a neighborhood of θ^*. Starting with $g(\theta) = \mathbb{1}\{\theta \in A^*\}$, we have, by assumption on A^* that there exists $\epsilon > 0$ such that for all θ in a neighborhood of θ^*, $\mathbb{1}\{\theta \in A^*\} = \mathbb{1}\{\theta \in B_{\hat{\theta}}(\epsilon)\}$. As long as $\|\hat{\theta} - \theta^*\| < \epsilon - \gamma$ for some $\gamma > 0$, there exists a neighborhood around θ^* where $\mathbb{1}\{\theta \in B_{\hat{\theta}}(\epsilon)\}$ is twice continuously differentiable. Indeed, by assumption **(A0)**, we can choose $n(\epsilon, \gamma, \delta)$ large enough such that that this holds with probability at least $1 - \delta$. We can argue similarly for $g(\theta) = \mathbb{1}\{\theta \in A^{**}\}$. Hence, assumption **(A4")** holds for $g(\theta) = \mathbb{1}\{\theta \in A^*\}$ and $g(\theta) = \mathbb{1}\{\theta \in A^{**}\}$ by Proposition 4.

E Technical Lemmas

Lemma 3. Assume that **(A0)**–**(A2)** hold. Then there exists an integer $N(\eta, \epsilon) > 0$ such that for all $\alpha_n n > N(\eta, \epsilon)$, we have that $f_n^+(g, h)$ given in (149) is well defined. Moreover, for any $\eta > 0$ and $\epsilon > 0$, there exists a sequence $r_n \to \infty$, such that

$$\mathbb{P}_{f_{0,n}}\left(\sup_{g,h \in \bar{B}_0(r_n)} f_n^+(g, h) > \eta\right) < \epsilon. \tag{180}$$

Proof. We show (180) in three steps, adapting the arguments of [4, Lemma 5]. In Step 1, we check that $f_n^+(g_n, h_n)$ is well defined for balls of fixed (in n) radius $r > 0$. In Step 2, we prove (180) for balls of fixed radius $r > 0$. In Step 3, we construct a sequence $r_n \to \infty$ where the results of Step 1 and Step 2 hold to give the final result.

Step 1: Check that $f_n^+(g_n, h_n)$ is well defined. We show that there exists an N_0 such that for $\alpha_n n \geq N_0$, the function $f_n^+(g_n, h_n)$ is well defined, meaning we are not dividing by zero, when $\{h_n\}$ and $\{g_n\}$ are sequences in $\bar{B}_0(r)$. Recall the definition of the α_n-posterior in (4) and the scaled densities in (147). Then we see,

$$\begin{aligned}
\frac{\pi^{LAN}_{n,\alpha_n}(g_n|X^n)}{\pi^{LAN}_{n,\alpha_n}(h_n|X^n)} &= \left(\frac{f_n\left(X^n|\theta^* + \frac{g_n}{\sqrt{\alpha_n n}}\right)}{f_n\left(X^n|\theta^* + \frac{h_n}{\sqrt{\alpha_n n}}\right)}\right)^{\alpha_n} \frac{\pi_n(g_n)}{\pi_n(h_n)} \\
&= \frac{\exp\left(-\alpha_n n \times -\frac{1}{n}\log f_n\left(X^n|\theta^* + \frac{g_n}{\sqrt{\alpha_n n}}\right)\right)}{\exp\left(-\alpha_n n \times -\frac{1}{n}\log f_n\left(X^n|\theta^* + \frac{h_n}{\sqrt{\alpha_n n}}\right)\right)} \frac{\pi_n(g_n)}{\pi_n(h_n)}
\end{aligned}$$

where

$$\pi_n(h_n) \equiv (\alpha_n n)^{-p/2}\pi(\theta^* + h_n/\sqrt{\alpha_n n}), \tag{181}$$

is the density of the prior distribution of the transformation $\sqrt{\alpha_n n}(\theta - \theta^*)$. By the definition in (149), with

$$s_n(h_n) = \frac{\exp\left(-\alpha_n n \times -\frac{1}{n}\log f_n\left(X^n|\theta^* + \frac{g_n}{\sqrt{\alpha_n n}}\right)\right)}{\exp\left(-\alpha_n n \times -\frac{1}{n}\log f_n\left(X^n|\theta^*\right)\right)},$$

we have

$$f_n^+(g_n, h_n) = \left\{1 - \frac{\phi_n(h_n)\pi_{n,\alpha}^{LAN}(g_n|X^n)}{\pi_{n,\alpha}^{LAN}(h_n|X^n)\phi_n(g_n)}\right\}^+ = \left\{1 - \frac{\phi_n(h_n)s_n(g_n)\pi_n(g_n)}{\phi_n(g_n)s_n(h_n)\pi_n(h_n)}\right\}^+, \tag{182}$$

To show that (182) is well defined, we justify that $\phi_n(\cdot)$, $s_n(\cdot)$, and $\pi_n(\cdot)$ are nonzero when they take input h_n or g_n. This is obvious for $\phi_n(\cdot)$, as it is the Gaussian density. The positivity of $s_n(\cdot)$ follows from **(A2)**, where we take K to be $\bar{B}_0(r)$ and note that V_{θ^*} is positive definite. It remains to justify the positivity of $\pi_n(\cdot)$, which follows by assumption **(A1)**, provided we verify that the inputs considered, namely $\theta^* + h_n/\sqrt{\alpha_n n}$ and $\theta^* + g_n/\sqrt{\alpha_n n}$, belong to the ball $B_{\theta^*}(\delta)$. Indeed, notice that for any $r > 0$, there exists an integer $N_0 := N_0(r, \delta) := \lceil \frac{4r^2}{\delta^2} \rceil > 0$ such that $\theta^* + h_n/\sqrt{\alpha_n n} \in B_{\theta^*}(\delta)$ whenever $h_n \in \bar{B}_0(r)$ and $\alpha_n n \geq N_0(r, \delta)$. To see this, note that if $||h||_2 \leq r$ and $\alpha_n n \geq \frac{4r^2}{\delta^2}$, then

$$\frac{||h||_2}{\sqrt{\alpha_n n}} \leq \frac{r}{\sqrt{\alpha_n n}} \leq \frac{r}{\sqrt{\frac{4r^2}{\delta^2}}} = \frac{\delta}{2} < \delta.$$

Then as $||h_n||_2/\sqrt{\alpha_n n} < \delta$ and $||g_n||_2/\sqrt{\alpha_n n} < \delta$, we have that $\theta^* + h_n/\sqrt{\alpha_n n}$ and $\theta^* + g_n/\sqrt{\alpha_n n}$ belong to the ball $B_{\theta^*}(\delta)$, so (182) is well defined.

Step 2: Show the statement in (180) **for fixed** $r > 0$**.** For any sequence $h_n \in \bar{B}_0(r)$, notice that **(A2)** implies

$$\begin{aligned}
\log(s_n(h_n)) &= -\alpha_n n\left[-\frac{1}{n}\log f_n\left(X^n|\theta^* + \frac{h_n}{\sqrt{\alpha_n n}}\right) - \left[-\frac{1}{n}\log f_n(X^n|\theta^*)\right]\right] \\
&= \sqrt{\alpha_n} h_n^\top V_{\theta^*}\Delta_{n,\theta^*} - \frac{1}{2}h_n^\top V_{\theta^*} h_n + o_{f_{0,n}}(1) \\
&= -\frac{1}{2}h_n^\top V_{\theta^*} h_n + o_{f_{0,n}}(1),
\end{aligned} \tag{183}$$

where in the last equality, we have used the fact that $\alpha_n \to 0$ and $V_{\theta^*}\Delta_{n,\theta^*} = O_{f_{0,n}}(1)$ by **(A0)** as V_{θ^*} is positive definite by **(A2)**. By definition of the density $\phi_n(h_n)$,

$$\log(\phi_n(h_n)) = -\frac{p}{2}\log(2\pi) + \frac{1}{2}\log(\det(V_{\theta^*})) - \frac{1}{2}h_n^\top V_{\theta^*} h_n.$$

Thus,

$$\log\left[\frac{s_n(h_n)}{\phi_n(h_n)}\right] = o_{f_{0,n}}(1) + \frac{p}{2}\log(2\pi) - \frac{1}{2}\log(\det(V_{\theta^*})). \tag{184}$$

By (184), we notice that

$$\log\left[\frac{\phi_n(h_n)s_n(g_n)}{\phi_n(g_n)s_n(h_n)}\right] = o_{f_{0,n}}(1).$$

By (181) and continuity of π, from which it follows $\pi(\theta^* + g_n/\sqrt{\alpha_n n}), \pi(\theta^* + h_n/\sqrt{\alpha_n n}) \to \pi(\theta^*)$, whenever $\alpha_n n \geq N_0$,

$$\log\left[\frac{\pi_n(g_n)}{\pi_n(h_n)}\right] = \log\left[\frac{\pi(\theta^* + g_n/\sqrt{\alpha_n n})}{\pi(\theta^* + h_n/\sqrt{\alpha_n n})}\right] = o_{f_{0,n}}(1).$$

Putting this all together, we get

$$b_n^+(g_n, h_n) \equiv \log\left[\frac{\phi_n(h_n)s_n(g_n)\pi_n(g_n)}{\phi_n(g_n)s_n(h_n)\pi_n(h_n)}\right] = o_{f_{0,n}}(1). \tag{185}$$

As h_n and g_n are arbitrary sequences in $B_0(r)$, the result in (185) is equivalent to saying that for any fixed r, there exists an integer $\tilde{N}_0 := \tilde{N}_0(r, \epsilon, \eta)$, such that $\mathbb{P}_{f_{0,n}}(|b_n^+(g_n, h_n)| > \eta) < \epsilon$ for $\alpha_n n > \tilde{N}_0$. Because $f_n^+(g_n, h_n)$ is a continuous function of $|b_n^+(g_n, h_n)|$, by the continuous mapping theorem, there exists an integer $\tilde{N}_0^+ := \tilde{N}_0^+(r, \epsilon, \eta)$ such that for $\alpha_n n > \tilde{N}_0^+$,

$$\mathbb{P}_{f_{0,n}}\left(\sup_{g_n, h_n \in \bar{B}_0(r)} f_n^+(g_n, h_n) > \eta\right) < \epsilon. \tag{186}$$

Step 3: Show the statement for $r_n \to \infty$. Let $N^*(\epsilon, \eta) = \max\{\tilde{N}_0^+(1, \epsilon, \eta), \lceil \frac{4}{\delta^2} \rceil\}$ for $\tilde{N}_0^+(r, \epsilon, \eta)$ defined above and for all $\alpha_n n > N^*(\epsilon, \eta)$ let

$$r_n = \max\left\{r \in \mathbb{R} \,\middle|\, r \leq \delta\sqrt{\alpha_n n}/2 \quad \text{and} \quad \alpha_n n > \tilde{N}_0^+(r, \epsilon, \eta)\right\}.$$

For $\alpha_n n \leq N^*(\epsilon, \eta)$, we can define r_n arbitrarily, e.g, $r_n = 1$. We need to check the following:
(A) Check that r_n is well defined (i.e., there exists an r that satisfies the definition of r_n): Indeed, $r = 1$ satisfies the r_n definition for $n > N^*(\epsilon, \eta)$:

$$\frac{\delta\sqrt{\alpha_n n}}{2} > \frac{\delta\sqrt{N^*(\epsilon, \eta)}}{2} \geq \frac{\delta\sqrt{(\frac{4}{\delta^2})}}{2} = \frac{\delta\sqrt{4/\delta^2}}{2} = 1 = r,$$

and $\alpha_n n > N^*(\epsilon, \eta) \geq \tilde{N}_0^+(1, \epsilon, \eta) = \tilde{N}_0^+(r, \epsilon, \eta)$ by the definition of $N^*(\epsilon, \eta)$.
(B) Check that in fact $r_n \to \infty$: Indeed

$$\begin{aligned} r_n &= \max\left\{r \in \mathbb{R} \,\middle|\, r \leq \frac{\delta\sqrt{\alpha_n n}}{2} \text{ and } \alpha_n n > \tilde{N}_0^+(r, \epsilon, \eta)\right\} \\ &= \min\left\{\max\left\{r \in \mathbb{R} \,\middle|\, r \leq \frac{\delta\sqrt{\alpha_n n}}{2}\right\},\ \max\left\{r \in \mathbb{R} \,\middle|\, \alpha_n n > \tilde{N}_0^+(r, \epsilon, \eta)\right\}\right\}, \end{aligned}$$

Clearly $\delta\sqrt{\alpha_n n}/2 \to \infty$, so $r_n \to \infty$ if $\max\{r \in \mathbb{R}|\, \alpha_n n > \tilde{N}_0^+(r, \epsilon, \eta)\} \to \infty$. Notice that for $\max\{r \in \mathbb{R}|\, \alpha_n n > \tilde{N}_0^+(r, \epsilon, \eta)\} \to \infty$ as n grows, we need that $\tilde{N}_0^+(r, \epsilon, \eta)$ is finite as r grows. Otherwise, there would be no such r such that $\alpha_n n > \infty$. Indeed, $\tilde{N}_0^+(r, \epsilon, \eta)$ is finite as r grows. This is true, because if it were not, there would exist $r_{\max}$ such that $\tilde{N}_0^+(r, \epsilon, \eta) = \infty$ for $r > r_{\max}$. However, (185) holds for $g_n, h_n \in B_0(r)$ for any $r < \infty$, which implies the existence of a finite $\tilde{N}_0^+(r, \epsilon, \eta)$. Hence, $\tilde{N}_0^+(r, \epsilon, \eta) < \infty$ for any $r < \infty$.
(C) Show that for all $\alpha_n n > N^*(\epsilon, \eta)$:

$$\mathbb{P}_{f_{0,n}}\left(\sup_{g_n, h_n \in \bar{B}_0(r_n)} f_n^+(g_n, h_n) > \eta\right) < \epsilon.$$

First, $\theta^* + h_n/\sqrt{\alpha_n n} \in B_{\theta^*}(\delta)$ whenever $h_n \in \bar{B}_0(r_n)$ since $||h_n||_2/\sqrt{\alpha_n n} \leq r_n/\sqrt{\alpha_n n} \leq \delta/2$. This guarantees $f_n^+(g_n, h_n)$ is well defined as discussed in Step 1. Then, for all $\alpha_n n > N^*(\epsilon, \eta)$, by the r_n definition, $\alpha_n n > \tilde{N}_0^+(r_n, \epsilon, \eta)$. Hence, the bound in (186) holds and this completes the proof.

Lemma 4. If $X \sim \mathcal{N}(0, \Sigma)$, then

$$\mathbb{P}(\|X\| \geq t) \leq 4 \exp\left(-\frac{t^2}{8\mathrm{tr}(\Sigma)}\right) \tag{187}$$

Proof. By equation (3.5) in [43],

$$\mathbb{P}(\|X\| \geq t) \leq 4 \exp\left(-\frac{t^2}{8\mathbb{E}[\|X\|^2]}\right). \tag{188}$$

Letting $Q\Lambda Q^\top$ denote the spectral decomposition of Σ, we can write X as

$$X \stackrel{d}{=} Q\Lambda^{1/2} Z, \tag{189}$$

where $Z \sim \mathcal{N}(0, I)$. We use this to obtain

$$\begin{aligned}
\mathbb{E}[\|X\|^2] &= \mathbb{E}[(Q\Lambda^{1/2}Z)^\top (Q\Lambda^{1/2}Z)] \\
&= \mathbb{E}[\mathrm{tr}(Z^\top \Lambda Z)] \\
&= \mathrm{tr}(\Lambda \mathbb{E}[ZZ^\top]) \\
&= \mathrm{tr}(\Lambda) \\
&= \mathrm{tr}(\Sigma)
\end{aligned}$$

□

E.1 Proof of Lemma 1

Proof. We start by establishing that (47) is finite almost-surely for each $n \in \mathbb{N}$. We then establish the approximation in (48). Suppose that (48) holds. We will show how this leads to the approximation in (49) as follows. Notice that the expectation in (49) is a ratio of integrals of the form (47) since,

$$\frac{\int_{\mathbb{R}^p} g(\theta)\pi_{n,\alpha_n}(\theta|X^n)\pi(\theta)d\theta}{\int_{\mathbb{R}^p} \pi_{n,\alpha_n}(\theta|X^n)\pi(\theta)d\theta} = \frac{\int_{\mathbb{R}^p} g(\theta)\pi(\theta)\exp\left(\alpha_n \log f_n(X^n|\theta)\right) d\theta}{\int_{\mathbb{R}^p} \pi(\theta)\exp\left(\alpha_n \log f_n(X^n|\theta)\right) d\theta}. \tag{190}$$

We can apply the approximation (48) to the numerator of (190) taking $b(\theta) = g(\theta)\pi(\theta)$, which meets the conditions of assumption **(A4)** by assumption. We can also apply the approximation (48) to the denominator of (190) taking $b(\theta) = \pi(\theta)$ (i.e., setting $g(\theta) = 1$), which meets the conditions of assumption **(A4)** by assumption. Doing so yields

$$\begin{aligned}
&\frac{\int_{\mathbb{R}^p} g(\theta)\pi_{n,\alpha_n}(\theta|X^n)\pi(\theta)d\theta}{\int_{\mathbb{R}^p} \pi_{n,\alpha_n}(\theta|X^n)\pi(\theta)d\theta} \\
&= \frac{\exp\left(-\alpha_n n\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right)|H_n|^{-\frac{1}{2}}\left(\frac{2\pi}{\alpha_n n}\right)^{p/2}\left\{g(\hat{\theta})\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)\right\}}{\exp\left(-\alpha_n n\left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right)|H_n|^{-\frac{1}{2}}\left(\frac{2\pi}{\alpha_n n}\right)^{p/2}\left\{\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)\right\}} \qquad (191) \\
&= \frac{g(\hat{\theta})\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)}{\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)} \\
&= \frac{g(\hat{\theta})\pi(\hat{\theta})}{\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)} + \frac{O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)}{\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)} \\
&= \frac{g(\hat{\theta})\pi(\hat{\theta})}{\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)} + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right) \\
&= g(\hat{\theta}) - \frac{g(\hat{\theta})\pi(\hat{\theta})O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)}{\pi(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)} + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right) \\
&= g(\hat{\theta}) + O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right).
\end{aligned}$$

We note that (191) is well defined with high probability by choosing n large enough such that $\hat{\theta}$ belongs to the neighborhood of θ^* where, by assumption **(A1)**, π is positive and continuous (hence, finite) and g is finite with probability tending to 1. It remains to establish the results in (47) (Part I) and (48) (Part II).

Part I: showing the integral (47) **is bounded:** Recall the definition of I:

$$\begin{aligned}
I = \int_{\mathbb{R}^p} b(\theta)\exp\left(\alpha_n \log f_n(X^n|\theta)\right)d\theta &= \int_{\mathbb{R}^p} g(\theta)\pi(\theta)f_n(X^n|\theta)^{\alpha_n}d\theta \\
&\leq \int_{\mathbb{R}^p} |g(\theta)|\pi(\theta)f_n(X^n|\theta)^{\alpha_n}d\theta.
\end{aligned}$$

Define

$$C = \int_{\mathbb{R}^p} |g(\theta)|\pi(\theta)d\theta < \infty, \qquad (192)$$

which follows from the assumed integrability of $b(\theta)$. Furthermore, define

$$\tilde{\pi}(\theta) = C^{-1}|g(\theta)|\pi(\theta).$$

We note that $\tilde{\pi}(\theta)$ is a valid density since $\tilde{\pi}(\theta) \geq 0$ for all $\theta \in \mathbb{R}^p$ (which follows from nonnegativity of both $|g(\theta)|$ and $\pi(\theta)$) and $\tilde{\pi}(\theta)$ integrates to 1 (which follows from (192)). Hence,

$$I \leq C\int_{\mathbb{R}^p} \tilde{\pi}(\theta)f_n(X^n|\theta)^{\alpha_n}d\theta.$$

For each fixed n, we apply Theorem 7 to conclude that

$$\int_{\mathbb{R}^p} \tilde{\pi}(\theta) f_n(X^n|\theta)^{\alpha_n} d\theta < \infty,$$

almost surely. By this conclusion and (192), I is finite almost surely for each fixed n.

Part II: Obtaining the approximation in (48)**.** Our proof consists of two parts. First, we identify the leading order term

$$I_0 := \exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right) |H_n|^{-\frac{1}{2}} \left(\frac{2\pi}{\alpha_n n}\right)^{p/2} b(\hat{\theta}) \tag{193}$$

in the approximation of (48). Next, we analyze the remainder terms and show that they are of order $\frac{1}{\alpha_n n}$ with high probability for n sufficiently large. To do so, we use an approach similar to our arguments in Theorems 1 and 2, where we analyze the remainder terms on a (growing) neighborhood around $\hat{\theta}$ and its complement. However, we note that unlike the arguments in those proofs, the growing neighborhood in this proof is centered at $\hat{\theta}$ instead of θ^*. We change the centering in order to use the representation of the log-likelihood in Assumption **(A2")** and obtain the desired order of the approximation error.
Step 1: Identifying the leading term. We rewrite the integral in (47) as

$$\begin{aligned}
I &= \int_{\mathbb{R}^p} b(\theta) \exp\left(-\alpha_n n \times -\frac{1}{n} \log f_n(X^n|\theta)\right) d\theta \\
&= \exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right) \int_{\mathbb{R}^p} b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta})\right)\right]\right) d\theta \\
&= \exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right) \int_{\mathbb{R}^p} b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top (\alpha_n n H_n)(\theta - \hat{\theta})\right) d\theta \\
&\quad + \exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right) \int_{\mathbb{R}^p} \left[b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta})\right)\right]\right)\right. \\
&\qquad\qquad\qquad\qquad \left. - b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top (\alpha_n n H_n)(\theta - \hat{\theta})\right)\right] d\theta \\
&=: \exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right) (T_1 + T_2).
\end{aligned} \tag{194}$$

The leading term of the approximation in (48) will come from T_1 because

$$\begin{aligned}
T_1 &= \int_{\mathbb{R}^p} b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top (\alpha_n n H_n)(\theta - \hat{\theta})\right) d\theta \\
&= (2\pi)^{\frac{p}{2}} |\alpha_n n H_n|^{-\frac{1}{2}} b(\hat{\theta}) \int_{\mathbb{R}^p} (2\pi)^{-\frac{p}{2}} |\alpha_n n H_n|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top (\alpha_n n H_n)(\theta - \hat{\theta})\right) d\theta \\
&= |H_n|^{-\frac{1}{2}} \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} b(\hat{\theta}).
\end{aligned}$$

In the last equality we use that the integral integrates to one and $|\alpha_n n H_n|^{-\frac{1}{2}} = (\alpha_n n)^{-\frac{p}{2}} |H_n|^{\frac{1}{2}}$. It follows that

$$I = I_0 + \exp\left(\alpha_n \log f_n(X^n|\hat{\theta})\right) T_2$$

Step 2: Analyzing the remainder term $\mathbf{T_2}$. We will show that the remainder term, T_2 in (194), is comprised of terms of order $\frac{1}{\alpha_n n}$ or lower with high probability for n sufficiently large. To do so, we analyze the integral T_2 on a compact set, K and its complement. We will detail our choice of K below. We begin by noting that

$$\begin{aligned}
T_2 = \int_{K+K^{\mathsf{c}}} \Bigg\{ & b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right) \\
& \qquad\qquad - b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta-\hat{\theta})^\top(\alpha_n n H_n)(\theta-\hat{\theta})\right)\Bigg\} d\theta \\
= \int_{K^{\mathsf{c}}} & b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right) d\theta \\
& - \int_{K^{\mathsf{c}}} b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta-\hat{\theta})^\top(\alpha_n n H_n)(\theta-\hat{\theta})\right) d\theta \\
& + \int_K \Bigg\{ b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right]\right) \\
& \qquad\qquad - b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta-\hat{\theta})^\top(\alpha_n n H_n)(\theta-\hat{\theta})\right)\Bigg\} d\theta \\
=: & \; T_{21} + T_{22} + T_{23}.
\end{aligned} \tag{195}$$

Before we analyze these last three terms, we present and discuss our choice of K.
Choosing K: We will show that the choice $K = \bar{B}_{\theta^*}(\delta_0) \cap \bar{B}_{\hat{\theta}}(\delta_0)$ in (195) leads to the desired approximation (48). To do so, we define the following events

$$\begin{aligned}
\mathcal{A}_\delta &= \left\{\|\hat{\theta} - \theta^*\| < \delta\right\} \\
\mathcal{B}_K &= \left\{\inf_{K^{\mathsf{c}}} \left[-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right] > 0\right\} \\
\mathcal{C}_N &= \left\{\lambda_{\min}(H_n) > 0 \text{ for } n > N\right\}.
\end{aligned} \tag{196}$$

We will show that for appropriate choices of δ and N, K is nonempty. Furthermore, given such δ, N, and K, each of the events in (196) hold with high probability for n sufficiently large.

To see this, note that for any $\theta \in K$, we have by the triangle inequality,

$$\|\hat{\theta} - \theta^*\| \le \|\theta - \hat{\theta}\| + \|\theta - \theta^*\| \le 2\delta_0.$$

Hence, a stronger sufficient condition for K to be nonempty is that $\mathcal{A}_{\delta_0}$ holds.

Given our choice of K, we will establish that the events in (196) hold with high probability for n sufficiently large. First, by Assumption **(A0)**, for all $\epsilon > 0$ there exists $N_0 := N_0(\epsilon, \delta_0)$ such that for $n > N_0$

$$\mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}_{\delta_0}) < \frac{\epsilon}{6}. \tag{197}$$

Next, $K \subseteq B_{\hat{\theta}}(\delta_0)$ and we have by assumption **(A3")**, for all $\epsilon > 0$ and all $\delta_0 > 0$, there exists $\delta_0 > 0$ and $N_1 := N_1(\epsilon, \delta_0)$ such that for $n > N_1$,

$$\mathbb{P}_{f_{0,n}}\left(\mathcal{B}^{\mathsf{c}}_K\right) \le \mathbb{P}_{f_{0,n}}\left(\mathcal{B}^{\mathsf{c}}_{\bar{B}_{\hat{\theta}}(\delta_0)}\right) < \frac{\epsilon}{6} \tag{198}$$

Additionally, by assumption **(A2")**, for all $\epsilon > 0$, there exists $N_2 := N_2(\epsilon)$ such that for $n > N_2$,

$$\mathbb{P}_{f_{0,n}}(\mathcal{C}_{N_2}^{\mathsf{c}}) < \frac{\epsilon}{6}. \tag{199}$$

We conclude that the events in (196) hold with high probability. Furthermore, it will be useful to note that given any event, $\mathcal{D}$, by the law of total probability and a union bound, for $n > \max(N_0, N_1, N_2)$,

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}(\mathcal{D}) &\leq \mathbb{P}_{f_{0,n}}(\mathcal{D} \cap (\mathcal{A}_{\delta_0} \cap \mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2})) + \mathbb{P}_{f_{0,n}}((\mathcal{A}_{\delta_0} \cap \mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2})^{\mathsf{c}}) \\
&\leq \mathbb{P}_{f_{0,n}}(\mathcal{D} \cap (\mathcal{A}_{\delta_0} \cap \mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2})) + \mathbb{P}_{f_{0,n}}(\mathcal{A}_{\delta_0}^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)}^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{C}_{N_2}^{\mathsf{c}}) \\
&< \mathbb{P}_{f_{0,n}}(\mathcal{D} \cap (\mathcal{A}_{\delta_0} \cap \mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2})) + \frac{\epsilon}{2}.
\end{aligned} \tag{200}$$

We are now ready to turn our attention to the three terms in (195).

Analysis of $\mathbf{T_{21}}$: Note that

$$|T_{21}| \leq \exp\left(-\alpha_n n \inf_{\bar{B}_{\hat{\theta}}(\delta_0)^{\mathsf{c}}} \left\{-\frac{1}{n}\log f_n(X^n|\theta) - \left(-\frac{1}{n}\log f_n(X^n|\hat{\theta})\right)\right\}\right)\left(\int_{\mathbb{R}^p} |b(\theta)| d\theta\right).$$

Given $M_5 > 0$, by (200), we have that for $n > \max(N_0, N_1, N_2)$

$$\mathbb{P}_{f_{0,n}}\left(|T_{21}| > \frac{M_5}{\alpha_n n}\right) \leq \mathbb{P}_{f_{0,n}}\left(|T_{21}| > \frac{M_5}{\alpha_n n}\Big| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) + \frac{\epsilon}{2}. \tag{201}$$

Furthermore, $\mathbb{P}_{f_{0,n}}\left(|T_{21}| > \frac{M_5}{\alpha_n n}\Big| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) = 0$ for n large enough since T_{21} decreases exponentially to 0 on the event $\mathcal{B}_{\bar{B}_{\hat{\theta}}(\delta_0)}$ by assumption **(A3")** and the assumed integrability of $b(\theta)$. We conclude that

$$T_{21} = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right).$$

Analysis of $\mathbf{T_{22}}$: consider the the change of variable $h = \sqrt{\alpha_n n}(\theta - \hat{\theta})$, we note that

$$\theta \in K = \bar{B}_{\theta^*}(\delta_0) \cap \bar{B}_{\hat{\theta}}(\delta_0) \iff h \in \tilde{K} = \bar{B}_{\sqrt{\alpha_n n}(\theta^* - \hat{\theta})}(\sqrt{\alpha_n n}\delta_0) \cap \bar{B}_0(\sqrt{\alpha_n n}\delta_0). \tag{202}$$

In particular,

$$\tilde{K}^{\mathsf{c}} = \bar{B}_{\sqrt{\alpha_n n}(\theta^* - \hat{\theta})}(\sqrt{\alpha_n n}\delta_0)^{\mathsf{c}} \cup \bar{B}_0(\sqrt{\alpha_n n}\delta_0)^{\mathsf{c}}. \tag{203}$$

It follows from (202), (203), a union bound and the reverse triangle inequality that

$$
\begin{aligned}
|T_{22}| &= b(\hat{\theta}) \int_{\tilde{K}^{\mathsf{c}}} (\alpha_n n)^{-\frac{p}{2}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \\
&= b(\hat{\theta}) \left(\frac{2\pi}{\alpha_n n}\right)^{p/2} |H_n|^{-1/2} \int_{\tilde{K}^{\mathsf{c}}} \frac{1}{(2\pi)^{p/2}|H_n|^{-1/2}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \\
&\leq b(\hat{\theta}) \left(\frac{2\pi}{\alpha_n n}\right)^{p/2} |H_n|^{-1/2} \Bigg\{ \int_{\bar{B}_{\sqrt{\alpha_n n}(\theta^* - \hat{\theta})}(\sqrt{\alpha_n n}\delta_0)^{\mathsf{c}}} \frac{1}{(2\pi)^{p/2}|H_n|^{-1/2}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \\
&\qquad + \int_{\bar{B}_0(\sqrt{\alpha_n n}\delta_0)^{\mathsf{c}}} \frac{1}{(2\pi)^{p/2}|H_n|^{-1/2}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \Bigg\} \\
&\leq b(\hat{\theta}) \left(\frac{2\pi}{\alpha_n n}\right)^{p/2} |H_n|^{-1/2} \Bigg\{ \int_{\bar{B}_0(\sqrt{\alpha_n n}(\delta_0 + \|\theta^* - \hat{\theta}\|)^{\mathsf{c}}} \frac{1}{(2\pi)^{p/2}|H_n|^{-1/2}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \\
&\qquad + \int_{\bar{B}_0(\sqrt{\alpha_n n}\delta_0)^{\mathsf{c}}} \frac{1}{(2\pi)^{p/2}|H_n|^{-1/2}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \Bigg\}
\end{aligned}
\tag{204}
$$

Conditioning on $\mathcal{C}_{N_2}$ and applying Lemma 4 to this last inequality gives

$$
|T_{22}| \leq 4b(\hat{\theta}) \left(\frac{2\pi}{\alpha_n n}\right)^{p/2} |H_n|^{-1/2} \left\{ \exp\left(-\frac{\alpha_n n(\delta_0 + \|\theta^* - \hat{\theta}\|)^2}{8\mathrm{tr}(H_n)}\right) + \exp\left(-\frac{\alpha_n n \delta_0^2}{8\mathrm{tr}(H_n)}\right) \right\}. \tag{205}
$$

Given $M_6 > 0$, by (200), we have that for $n > \max(N_0, N_1, N_2)$,

$$
\mathbb{P}_{f_{0,n}}\left(|T_{22}| > \frac{M_6}{\alpha_n n}\right) \leq \mathbb{P}_{f_{0,n}}\left(|T_{22}| > \frac{M_6}{\alpha_n n} \Big| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) + \frac{\epsilon}{2}.
$$

Furthermore, there exists $N_7 := N_6(\epsilon, M_6)$ such that $\mathbb{P}_{f_{0,n}}\left(|T_{22}| > \frac{M_6}{\alpha_n n} \Big| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) = 0$ for n large enough since (205) decreases exponentially in $\alpha_n n$ on the event $\mathcal{C}_{N_2} \cap \mathcal{A}_{\delta_0}$. Hence,

$$
T_{22} = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right). \tag{206}
$$

Analysis of $\mathbf{T_{23}}$: Recall that

$$
\begin{aligned}
T_{23} = \int_K \Bigg\{ & b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta})\right)\right]\right) \\
& - b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top (\alpha_n n H_n)(\theta - \hat{\theta})\right) \Bigg\} d\theta
\end{aligned}
$$

We start by analyzing the integrand of T_{23}. Let $T'_{23}(\theta)$ denote the first intergrand of T_{23}, namely

$$
T'_{23}(\theta) = b(\theta) \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n(X^n|\theta) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta})\right)\right]\right) \tag{207}
$$

We reparametrize the exponent in (207) applying the the change of variable $h = \sqrt{\alpha_n n}(\theta - \hat{\theta})$ and Assumption **(A2")** to obtain,

$$
\begin{aligned}
\tilde{T'}_{23}(h) &= b\left(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\right) \exp\left(-\alpha_n n \left[-\frac{1}{n} \log f_n\left(X^n \Big| \hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\right) - \left(-\frac{1}{n} \log f_n(X^n|\hat{\theta})\right)\right]\right) \\
&= b\left(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\right) \exp\left(-\frac{1}{2} h^\top H_n h\right) \exp\left(-\left(-\frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle - \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right)\right).
\end{aligned}
\tag{208}
$$

Furthermore, defining $A(h) = -\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle - \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)$, we see that a Taylor expansion gives

$$\exp(-A(h)) = 1 - A(h) + \frac{1}{2}A(h)^2 - \frac{\exp(-\tilde{A})}{6}A(h)^3, \tag{209}$$

where $\tilde{A} \in (0, A(h))$. By the condition assumed on $b(\cdot)$, we have

$$b\left(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\right) = b(\hat{\theta}) + v^\top \frac{h}{\sqrt{\alpha_n n}} + R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) \tag{210}$$

Combining (208)–(210) yields

$$\begin{aligned}\tilde{T'}_{23}(h) &= \exp\left(-\frac{1}{2}h^\top H_n h\right)\left[1 - A(h) + \frac{1}{2}A(h)^2 + \frac{\exp(-\tilde{A})}{6}A(h)^3\right]\left[b(\hat{\theta}) + \frac{1}{\sqrt{\alpha_n n}}v^\top h + R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right] \\ &= \exp\left(-\frac{1}{2}h^\top H_n h\right)B_n(h), \end{aligned}\tag{211}$$

where

$$B_n(h) = \left[1 - A(h) + \frac{1}{2}A(h)^2 + \frac{\exp(-\tilde{A})}{6}A(h)^3\right]\left[b(\hat{\theta}) + \frac{1}{\sqrt{\alpha_n n}}v^\top h + R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right].$$

Letting

$$\begin{aligned}R_{1,n}(h) &:= \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right) + \frac{1}{2}A(h)^2 - \frac{\exp(-\tilde{A})}{6}A(h)^3 \\ &= \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right) + \frac{1}{2}\left(-\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle - \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right)^2 \\ &\quad - \frac{\exp(-\tilde{A})}{6}\left(-\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle - \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right)^3, \end{aligned}\tag{212}$$

we see that

$$\begin{aligned}B_n(h) &= \left[1 + \frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle + R_{1,n}(h)\right]\left[b(\hat{\theta}) + \frac{1}{\sqrt{\alpha_n n}}v^\top h + R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)\right] \\ &= I_{n,n}(h) + R_{n,n}(h), \end{aligned}\tag{213}$$

where $R_{n,n}(h)$ includes terms depending on $R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)$, $R_{1,n}(h)$, and h^s with $s \geq 4$. More precisely,

$$\begin{aligned}R_{n,n}(h) &= R_{1,n}(h)b(\hat{\theta}) + \frac{1}{6\alpha_n n}\langle S_n, h^{\otimes 3}\rangle \cdot v^\top h + \frac{1}{\sqrt{\alpha_n n}}R_{1,n}(h) \cdot v^\top h \\ &\quad + R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) + \frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle \cdot R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) + R_{1,n}(h)R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) \\ &= R_{1,n}(h)\left[b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}}\right] + R_{1,n}(h)R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) + R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)\left[1 + \frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle\right] \\ &\quad + \frac{1}{6\alpha_n n}\langle S_n, h^{\otimes 3}\rangle \cdot v^\top h. \end{aligned}\tag{214}$$

and

$$I_{n,n}(h) = b(\hat{\theta}) + \frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta}) + \frac{1}{\sqrt{\alpha_n n}} v^\top h.$$

We can therefore rewrite T_{23} using applying the change of variable $h = \sqrt{\alpha_n n}(\theta - \hat{\theta})$, (211) and (213), yielding

$$\begin{aligned}
T_{23} &= \int_K \left\{ T_{23}'(\theta) - b(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top (\alpha_n n H_n)(\theta - \hat{\theta}) \right) \right\} d\theta \\
&= (\alpha_n n)^{-\frac{p}{2}} \int_{\tilde{K}} \tilde{T}_{23}'(h) dh - b(\hat{\theta}) \int_{\tilde{K}} \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \\
&= (\alpha_n n)^{-\frac{p}{2}} \int_{\tilde{K}} \exp\left(-\frac{1}{2} h^\top H_n h \right) R_{n,n}(h) dh \\
&\quad + (\alpha_n n)^{-\frac{p}{2}} \int_{\tilde{K}} \exp\left(-\frac{1}{2} h^\top H_n h \right) I_{n,n}(h) dh - b(\hat{\theta}) \int_{\tilde{K}} \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \\
&= (\alpha_n n)^{-\frac{p}{2}} \int_{\tilde{K}} \exp\left(-\frac{1}{2} h^\top H_n h \right) R_{n,n}(h) dh \\
&\quad + (\alpha_n n)^{-\frac{p}{2}} \int_{\tilde{K}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta}) \right] \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \\
&=: \tilde{V}_1 + \tilde{V}_2.
\end{aligned} \tag{215}$$

We will show that both terms $\tilde{V}_1$ and $\tilde{V}_2$ are of order $\frac{1}{\alpha_n n}$ or lower, which implies $T_{23} = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$.

Analysis of $\tilde{\mathbf{V}}_1$: Define $\tilde{B} := \bar{B}_{\sqrt{\alpha_n n}(\theta^* - \hat{\theta})}(\sqrt{\alpha_n n}\delta_0)$. It follows from $\tilde{K} \subset \tilde{B}$ and (214) that

$$\begin{aligned}
|\tilde{V}_1| &\le (\alpha_n n)^{-\frac{p}{2}} \int_{\tilde{B}} \exp\left(-\frac{1}{2} h^\top H_n h \right) |R_{n,n}(h)| dh \\
&\le \left| \frac{2\pi H_n}{\alpha_n n} \right|^{\frac{1}{2}} \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_{1,n}(h)^2 \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \right)^{\frac{1}{2}} \\
&\qquad \times \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \left(b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \right)^{\frac{1}{2}} \\
&\quad + \left| \frac{2\pi H_n}{\alpha_n n} \right|^{\frac{1}{2}} \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_{1,n}(h)^2 \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \right)^{\frac{1}{2}} \\
&\qquad \times \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_b\left(\frac{h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \right)^{\frac{1}{2}} \\
&\quad + \left| \frac{2\pi H_n}{\alpha_n n} \right|^{\frac{1}{2}} \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_b\left(\frac{h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \right)^{\frac{1}{2}} \\
&\qquad \times \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \left(1 + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \right)^2 \exp\left(-\frac{1}{2} h^\top H_n h \right) dh \right)^{\frac{1}{2}} \\
&\quad + \left| \frac{2\pi H_n}{\alpha_n n} \right|^{\frac{1}{2}} |2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \left(\frac{1}{\alpha_n n} \right) \left| \frac{1}{6} \langle S_n, h^{\otimes 3}\rangle \cdot v^\top h \right| \exp\left(-\frac{1}{2} h^\top H_n h \right) dh.
\end{aligned} \tag{216}$$

The terms of (216) are multiples of

$$\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}} \tag{217}$$

and

$$\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_{1,n}(h)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}}. \tag{218}$$

Given $\mathcal{C}_{N_2}$, the remaining terms are expectations of polynomials with respect to a mean 0 Gaussian with covariance H_n^{-1}. Furthermore, the last term of (216) is $O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$. Hence, we will conclude that $\tilde{V}_1 = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$ by showing that (217) and (218) are $O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$. Before we do so, we note by the reverse triangle inequality that,

$$\begin{aligned} \mathbb{1}\{h \in \tilde{B}\} &= \mathbb{1}\{\|h - \sqrt{\alpha_n n}(\theta^* - \hat{\theta})\| < \sqrt{\alpha_n n}\delta_0\} \\ &\leq \mathbb{1}\{\left|\|h\| - \sqrt{\alpha_n n}\|\theta^* - \hat{\theta}\|\right| < \sqrt{\alpha_n n}\delta_0\} \\ &\leq \mathbb{1}\left\{\frac{\|h\|}{\sqrt{\alpha_n n}} < \delta_0 + \|\theta^* - \hat{\theta}\|\right\}. \end{aligned} \tag{219}$$

We now analyze (217). Define the event

$$\mathcal{E} = \left\{R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right) < \frac{M_1}{\alpha_n n}\|h\|_2^2;\ \frac{\|h\|}{\sqrt{\alpha_n n}} < 2\delta_0\right\}.$$

If we choose N_3 large enough such that $\frac{\|h\|}{\sqrt{\alpha_n n}} < 2\delta_0$ for $n > N_3$, then by assumption **(A4)**, there exists $M_1 > 0$ and N_4 such that for $n > \max(N_3, N_4)$,

$$\mathbb{P}_{f_{0,n}}\left(\mathcal{E}^{\mathsf{c}}\right) < \frac{\epsilon}{2}. \tag{220}$$

Hence, for, $n > \max(N_0, N_1, N_2, N_3, N_4)$,

$$\begin{aligned} &\mathbb{P}_{f_{0,n}}\left(\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}} > \frac{\tilde{M}_1}{\alpha_n n}\right) \\ &\overset{(a)}{\leq} \mathbb{P}_{f_{0,n}}\left(\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \mathbb{1}\left\{\frac{\|h\|}{\sqrt{\alpha_n n}} < \delta_0 + \|\theta^* - \hat{\theta}\|\right\} R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}} > \frac{\tilde{M}_1}{\alpha_n n}\right) \\ &\overset{(b)}{\leq} \mathbb{P}_{f_{0,n}}\left(\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\frac{\|h\|}{\sqrt{\alpha_n n}} < 2\delta_0} R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}} > \frac{\tilde{M}_1}{\alpha_n n}\middle| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_1)} \cap \mathcal{C}_{N_2}\right) + \frac{\epsilon}{2} \\ &\leq \mathbb{P}_{f_{0,n}}\left(\frac{M_1}{\alpha_n n}\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \|h\|_2^4 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}} > \frac{\tilde{M}_1}{\alpha_n n}\middle| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_1)} \cap \mathcal{C}_{N_2} \cap \mathcal{E}\right) \\ &\quad + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}^{\mathsf{c}}\right) + \frac{\epsilon}{2} \overset{(c)}{<} \epsilon \end{aligned} \tag{221}$$

Step (a) of (221) follows from (219). Step (b) follows from (200). Step (c) follows from (220) and choosing $\tilde{M}_1$ so that

$$\mathbb{P}_{f_{0,n}}\left(\frac{M_1}{\alpha_n n}\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \|h\|_2^4 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh\right)^{\frac{1}{2}} > \frac{\tilde{M}_1}{\alpha_n n}\middle| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) = 0.$$

For example, we may let

$$\tilde{M}_1 = 2M_1 \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \|h\|_2^4 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}}.$$

Hence, there exists $\tilde{M}_1 > 0$ such that

$$\mathbb{P}_{f_{0,n}} \left(\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_b \left(\frac{h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}} > \frac{\tilde{M}_1}{\alpha_n n} \right) < \epsilon$$

for n sufficiently large, and we conclude this term is $O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$. We use a similar argument for (218). Inspecting the definition of $R_{1,n}(h)$ in (212), we see that it is a polynomial of terms of order $\frac{1}{\alpha_n n}$ and of $\tilde{R}_n\left(\frac{h}{\alpha_n n}\right)$. Then, the highest order term of

$$\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} R_{1,n}(h)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}},$$

which depends on $\tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right)$ is,

$$\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} \tilde{R}_n \left(\frac{h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}}, \tag{222}$$

which we will show is $O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$. Using a similar argument as that for $R_b\left(\frac{h}{\sqrt{\alpha_n n}}\right)$, we define the event

$$\mathcal{E} = \left\{ \tilde{R}_n\left(\frac{h}{\sqrt{\alpha_n n}}\right) < \frac{M_2}{\alpha_n n}\|h\|_2^4;\ \frac{\|h\|}{\sqrt{\alpha_n n}} < 2\delta_0 \right\}.$$

By assumption **(A2")**, there exists $M_2 > 0$ and N_5 such that for $n > \max(N_3, N_5)$,

$$\mathbb{P}_{f_{0,n}}\left(\mathcal{E}^{\mathsf{c}}\right) < \frac{\epsilon}{2}. \tag{223}$$

We can now apply a similar argument as (220) to conclude for $n > \max(N_0, N_1, N_2, N_3, N_5)$,

$$\begin{aligned} &\mathbb{P}_{f_{0,n}} \left(\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} \tilde{R}_n \left(\frac{h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \right) \\ &\le \mathbb{P}_{f_{0,n}} \left(\frac{M_2}{\alpha_n n} \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \|h\|_2^8 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \middle| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2} \cap \mathcal{E} \right) \\ &\quad + \mathbb{P}_{f_{0,n}}\left(\mathcal{E}^{\mathsf{c}}\right) + \frac{\epsilon}{2} < \epsilon, \end{aligned} \tag{224}$$

where we choose

$$\tilde{M}_2 = 2M_2 \left(|2\pi H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \|h\|_2^8 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}}.$$

Hence, there exists $\tilde{M}_2 > 0$ such that

$$\mathbb{P}_{f_{0,n}} \left(\left(|2\pi H_n|^{-\frac{1}{2}} \int_{\tilde{B}} \tilde{R}_n \left(\frac{h}{\sqrt{\alpha_n n}} \right)^2 \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \right)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \right) < \epsilon$$

for n sufficiently large, and we conclude this term is $O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$.

Given $M_3 > 0$, by (200), we have that for $n > \max(N_0, N_1, N_2, N_3, N_4, N_5)$,

$$\mathbb{P}_{f_{0,n}}\left(|\tilde{V}_1| > \frac{M_3}{\alpha_n n}\right) \leq \mathbb{P}_{f_{0,n}}\left(|\tilde{V}_1| > \frac{M_3}{\alpha_n n}\Big|\mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_1)} \cap \mathcal{C}_{N_2}\right) + \frac{\epsilon}{2}. \tag{225}$$

Furthermore, on the event $\mathcal{C}_{N_2} \cap \mathcal{A}_{\delta_0}$, we have, by the analyses of $\tilde{R}_n(\cdot)$ and $R_b(\cdot)$, there exists $N_6 := N_6(\epsilon, M_3) > \max(N_3, N_4, N_5)$ such that $\mathbb{P}_{f_{0,n}}\left(|\tilde{V}_1| > \frac{M_3}{\alpha_n n}\Big|\mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) < \frac{\epsilon}{2}$. Hence,

$$\tilde{V_1} = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right).$$

Analysis of $\tilde{\mathbf{V}}_\mathbf{2}$: We write $\tilde{V}_2$ as,

$$\begin{aligned}
\tilde{V}_2 &= \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \\
&= \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \Bigg\{ \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \\
&\qquad - \int_{\tilde{K}^c} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \Bigg\} \\
&= -\left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \int_{\tilde{K}^c} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh,
\end{aligned} \tag{226}$$

where the second line of (226) follows from the fact that conditional on $\mathcal{C}_{N_2}$, the first term of (226) is the expectation of a polynomial in h with respect to a Gaussian density with covariance H_n^{-1} and that the odd moments of a mean 0 Gaussian are 0. We use the same union bound and reverse triangle inequality from (204) to write $\tilde{V}_2$ as

$$\begin{aligned}
|\tilde{V}_2| &= \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \int_{\tilde{K}^c} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \\
&\leq \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \Bigg\{ \int_{\bar{B}_0(\sqrt{\alpha_n n}(\delta_0 + \|\theta^* - \hat{\theta}\|))^c} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \\
&\qquad + \int_{\bar{B}_0(\sqrt{\alpha_n n}\delta_0)^c} \frac{1}{(2\pi)^{\frac{p}{2}}|H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2}h^\top H_n h\right) dh \Bigg\}
\end{aligned} \tag{227}$$

Conditioning on $\mathcal{C}_{N_2}$, we appeal to Cauchy-Schwarz and Lemma 4 to this last inequality to yield

$$\begin{aligned}
|\tilde{V}_2| \leq & \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \left(\int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2} h^\top H_n h\right) dh\right)^{\frac{1}{2}} \\
& \times \left\{ \int_{\bar{B}_0(\sqrt{\alpha_n n}(\delta_0 + \|\theta^* - \hat{\theta}\|))^c} \frac{1}{(2\pi)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \right. \\
& \qquad \left. + \int_{\bar{B}_0(\sqrt{\alpha_n n}\delta_0)^c} \frac{1}{(2\pi)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2} h^\top H_n h\right) dh \right\}^{\frac{1}{2}} \\
\leq & \left(\frac{2\pi}{\alpha_n n}\right)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}} \left(\int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}} |H_n|^{-\frac{1}{2}}} \left[\frac{1}{\sqrt{\alpha_n n}} v^\top h + \frac{1}{6\sqrt{\alpha_n n}} \langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})\right] \exp\left(-\frac{1}{2} h^\top H_n h\right) dh\right)^{\frac{1}{2}} \\
& \times \left\{ \exp\left(-\frac{\alpha_n n(\delta_0 + \|\theta^* - \hat{\theta}\|)^2}{8\mathrm{tr}(H_n)}\right) + \exp\left(-\frac{\alpha_n n \delta_1^2}{8\mathrm{tr}(H_n)}\right) \right\}^{\frac{1}{2}}
\end{aligned} \tag{228}$$

Given $M_4 > 0$, by (200), we have that for $n > \max(N_0, N_1, N_2)$,

$$\mathbb{P}_{f_{0,n}}\left(|\tilde{V}_2| > \frac{M_4}{\alpha_n n}\right) \leq \mathbb{P}_{f_{0,n}}\left(|\tilde{V}_2| > \frac{M_4}{\alpha_n n} \middle| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_0)} \cap \mathcal{C}_{N_2}\right) + \frac{\epsilon}{2}. \tag{229}$$

Furthermore, there exists $N_6 := N_6(\epsilon, M_4)$ such that $\mathbb{P}_{f_{0,n}}\left(|\tilde{V}_2| > \frac{M_4}{\alpha_n n} \middle| \mathcal{A}_{\delta_0} \cap \mathcal{B}_{B_{\hat{\theta}}(\delta_1)} \cap \mathcal{C}_{N_2}\right) = 0$ for n large enough since (228) decreases exponentially in $\alpha_n n$ on the event $\mathcal{C}_{N_2} \cap \mathcal{A}_{\delta_0}$. Hence,

$$\tilde{V}_2 = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right)$$

and

$$T_{2,1} = \tilde{V}_1 + \tilde{V}_2 = O_{f_{0,n}}\left(\frac{1}{\alpha_n n}\right).$$

□

F Auxiliary Results

In this section, we collect and restate results from other works that we use in our proofs.

Lemma 5. [4, Lemmas 1 and 4] Consider densities φ and ψ and, for a given compact set $K \subset \mathbb{R}^p$, suppose that the densities are positive on K. Then, for any function, $s(h)$, that is nonnegative on all of $\mathbb{R}^p$,

$$\int_{\mathbb{R}^p} s(h)|\varphi(h) - \psi(h)|dh \leq [\sup_{g,g' \in K} \tilde{f}^+(g, g')] \int_{\mathbb{R}^p} s(h)(\psi(h) + \varphi(h))dh + \int_{\mathbb{R}^p \setminus K} s(h)(\psi(h) + \varphi(h))dh,$$

where

$$\tilde{f}^+(g, g') := \left\{1 - \frac{\varphi(g')\psi(g)}{\psi(g')\varphi(g)}\right\}^+.$$

Remark 3. Unlike [4, Lemma 4], Lemma 5 does not use the quantity $\tilde{f}^-(g, h)$ since, by [54, Eqs. (21)-(22)], we see that $\sup_{g,h \in K} \tilde{f}^+(g, h) = \sup_{g,h \in K} \tilde{f}^-(g, h)$.

Lemma 6. [54, Lemma 2] Suppose a random variable $Z \stackrel{d}{:=} Y|X^n$ has a density $f_Z(\cdot)$ on $\mathbb{R}^p$. Assume there exists a $\gamma > 0$ such that $\mathbb{E}_{f_{0,n}}[\mathbb{E}[\|Z\|_2^{k(1+\gamma)}]] < \infty$. Then, for any $\epsilon > 0$ and $r_n \to \infty$, there exists an integer $N(\epsilon, \gamma, k) > 0$ such that for all $n > N(\epsilon, \gamma, k)$,

$$\mathbb{E}_{f_{0,n}}\left[\int_{\|z\|_2 > r_n} \|z\|_2^k f_Z(z) dz\right] < \epsilon.$$

Lemma 7. [15, Theorem 1] Assume $\int_\Theta \pi(\theta)d\theta = 1$. Then $\int_\Theta f_n(X^n|\theta)^\alpha \pi(\theta)d\theta < \infty$ almost surely for any $\alpha \geq 0$.

Remark 4. There is a slight discrepancy between [15, Theorem 1] and our formulation in Lemma 7: because in our setting $f_n(X^n|\theta)$ is a valid density for each $\theta \in \mathbb{R}^p$, we do not state this as a condition. Furthermore, we state the result holds almost surely: for any proper $\pi(\theta)$, by Fubini's theorem,

$$\int_{\mathcal{X}} \int_{\mathbb{R}^p} \pi(\theta) f_n(X^n|\theta) d\theta dX^n = \int_{\mathbb{R}^p} \pi(\theta) \int_{\mathcal{X}} f_n(X^n|\theta) dX^n d\theta = 1 < \infty.$$

Hence, $\int_{\mathbb{R}^p} \pi(\theta) f_n(X^n|\theta) d\theta$ is finite except on a set of Lebesgue measure 0.