Bayesian Semiparametric estimation of densities with unknown support

Johan Van Der Molen Moris

Doctor of Philosophy
University of Edinburgh
2019

Abstract

We address the nonregular semiparametric problem of estimating a boundary point of the support of an unknown density, under local asymptotic exponentiality. The aim is to find the limiting marginal posterior distribution of the nonregular parameter and the rate of concentration for the density. Here we investigate two approaches. The first consists in extending the results found for parametric models to the case where the dimension of the regular nuisance parameter grows to infinity along with the number of observations. We used a Log-Spline prior to obtain the local concentration result for the marginal posterior of the lower support point; a Bernstein - von Mises type theorem with exponential limiting distribution. We also obtained contraction for the density at minimax rate up to a log factor.

In the second approach, we constructed an adaptive mixture prior for a decreasing density with the following properties: a) posterior distribution of the density with known lower support point concentrates at minimax rate, up to log factor, b) the density is estimated consistently, uniformly in a neighbourhood of the lower support point, c) marginal posterior distribution of the lower support point of the density has shifted exponential distribution in the limit. In particular, to ensure that the density is asymptotically consistent pointwise in a neighbourhood of the lower support point, instead of a usual Dirichlet mixture weights, we consider a non-homogeneous Completely Random Measure mixture. This is important since the rate parameter of the limiting Exponential distribution is equal to the value of the density at the lower support point. The general conditions for the BvM type result we have are different from those by Knapik and Kleijn (2013); the latter don't hold for a hierarchical mixture prior we consider. We implement this model using two different representations of the prior process; illustrate performance of this approach on simulated data, and apply it to model distribution of bids in procurement auctions.

Lay Summary

We address the problem of estimating the *true* minimum of a variable, where the data is a random sample of its values. A motivating example is a procurement auction, where companies bid the amount of money they require to perform a certain project, and the aim is to estimate its true cost. Some results in the literature have been obtained under the assumption that the underlying distribution of the observations belongs to a family that can be characterised using a fixed number of parameters. We remove this assumption extending the results to the case where the distribution has a density that belongs to a much more flexible class of functions.

We study the properties of Bayesian estimators in this model when the number of observations grows to infinity. That is, we express prior knowledge about the true minimum in the form of a probability distribution and update it using the observed data and Bayes theorem to obtain the posterior distribution, which is used for estimation. We prove that for a wide class of prior distributions, as we get more and more observations, the posterior distribution tends to concentrate around the true value of the minimum with the shape of a shifted Exponential distribution. Considering the density function to be unknown also requires to model the prior knowledge on the density function. We do this in two different ways; one that uses approximating properties of polynomials (B-Splines) and in the other we approximate any density by a mixture of Gamma densities.

Finally, we illustrate performance of this approach on simulated data, and apply it to model distribution of bids in procurement auctions.

Acknowledgements

I would like to express my gratitude and appreciation to my supervisor Dr Natalia Bochkina for the guidance and valuable assistance on this project, from the moment I decided to choose this topic until the very last correction. I would also like to thank Prof. Judith Rousseau for the guidance and great contributions to this work. Special thanks to my parents, my sister and my partner Daniela for being always so supportive especially when I needed it the most, I could not have finished this without them.

Finally, this PhD programme has been funded by *Becas Chile* programme of scholarships from Chilean government, project CONICYT PAI/INDUSTRIA 79090016.

Contents

Chapter 1

Introduction

1.1 Overview

This work addresses the problem of estimating the lowest endpoint of the support of a probability density function without assuming that it belongs to a particular parametric family of distributions. Consider an i.i.d. sample from a distribution with density function f. We define the lowest endpoint as the infimum of its support as $\theta(f) = \inf\{x \in \mathbb{R} : f(x) > 0\} = \inf(\text{supp}(f))$ and we assume that this variable is well defined and finite, that is, the density f is supported on a semiline $[\theta, \infty)$ for some $\theta \in \mathbb{R}$. Additionally, we assume that the density has a discontinuity located at θ. We are interested in estimating θ when f is unknown. See Figure 1.1 for a simple representation of the problem. Note this is equivalent to estimating the highest endpoint of the support when f is supported on the semiline $(-\infty, \theta]$.

This problem is considered relevant from both theoretical and practical perspectives. First, when the density function is discontinuous at the estimated point, the parameter becomes non-regular and standard results do not hold. For instance, it is easy to find consistent estimates with rate of convergence equal to n, and with an exponential limit distribution, in contrast to the well known rate of convergence $n^{1/2}$ and normal limiting distribution of both maximum likelihood estimator and bayesian estimators in the regular framework [30]. Also, MLEs are often inefficient in nonregular models (See for instance [11]). From a practical point of view, estimation of the location of jumps of a density has many applications mainly in econometrics such as procurement

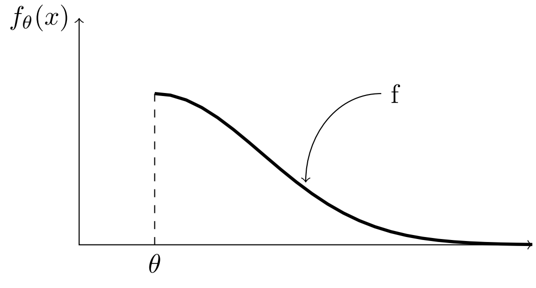

Figure 1.1: Example of the model of interest. The density function f_θ is unknown and supported on the semiline $(\theta, \infty]$. The parameter of interest is θ.

auctions or equilibrium job-search models [14], [15], [16], [11].

General results based on the limit likelihood ratio process when the number of observations tends to infinity have been found in the parametric version of the problem where the shape of the density function is considered to be known and so the only (nonregular) parameter of the model is the location of the lower endpoint of the support [30], and in the extension where the family of densities is indexed by an additional regular euclidean parameter [22]. However, there are still open questions regarding the semiparametric case, where the parametrization of the model has an infinite-dimensional nuisance parameter as mentioned in [32].

The purpose of this work is to study semiparametric estimates of the lower endpoint of the support of an unknown density and its asymptotic behaviour from a Bayesian perspective. In particular it is of interest to prove consistency, find the corresponding rate of convergence for the density and derive the limiting marginal distribution of the nonregular parameter of interest. This is what in literature is known as a Bernstein-von Mises (BvM) type of result.

Bayesian inference and particularly Bayesian nonparametric and semiparametric models have become more popular in recent years. Since they allow us to avoid arbitrary parametric assumptions, they can be applied to more general frameworks and wide variety of data. Additionally, the development of MCMC algorithms have made possible their application and implementation, solving real-world problems in diverse research areas such as finance, geosciences, biology, epidemiology and machine learning among many others. Therefore it is important to keep developing the theory that supports its use. In the parametric case where the density is known, Bayesian estimators are asymptotically efficient whereas the MLE in general can be improved by de-biasing (See Section V.4 in [30]). This illustrates some of the advantages Bayesian estimators in our model may have compared to their frequentist counterparts and motivates the idea of studying them further. Posterior concentration and Bernstein-von Mises theorem are key results in Bayesian analysis. Consistency is considered to be a way of validating Bayesian inference through frequentist properties. For example, in a simulated experiment with a known parameter, a consistent Bayesian estimator will be close to the known truth given enough data. It also ensures robustness with respect to the choice of prior since data eventually overrides them. Bernstein-von Mises theorem specifies the limiting posterior distribution which implies that posterior credible sets are also asymptotically confidence sets, as shown in Section 3.2.1. This justifies the use of credible sets from a frequentist point of view, and in applications where the posterior distribution may be intractable the approximation given by BvM supports the use of the limiting distribution as a good approximation at least when the sample size is large. In regular models BvM is also relevant for studying efficiency of estimators and it would be interesting to investigate this in nonregular models as well.

The discontinuity in the density function results in non standard inference theory, and some difficulties emerge to obtain likelihood approximation results especially for the nonparametric part of the model. Finding a prior model with suitable conditions of hyperparameters is particularly challenging in this model as well. In addition to the usual concentration in Hellinger distance, we require uniform consistency near the point of discontinuity among the sufficient conditions for a BvM theorem to hold. It seems this is related to the fact that the rate parameter of the limiting Exponential distribution is equal to the limit from the right of density function at the discontinuity point. This extra condition is non trivial to satisfy in general, and in fact the nonparametric MLE estimator is not consistent at the discontinuity point.

In this work we investigated two different approaches. The first one corresponds to consider a parametric model with a nuisance regular parameter of increasing dimension along with the number of observations using what is known in literature as a sieve prior (Section 2.1), whereas the second considers a mixture with a kernel supported on the semiline as a prior (Section 3.1). For both models we proved a BvM theorem for the marginal posterior distribution of the nonregular parameter. These results together with the implementation of an MCMC algorithm for the mixture prior correspond to the main contributions of this thesis. A general nonregular BvM result for a sieve prior can be found in Proposition 2 (Section 2.3), and Theorem 2 (Section 2.5.2), shows a BvM result for a log-spline model. Results for a general LAE model and a a Shift LAE model are found in Proposition 3 (Section 3.2.1) and Theorem 7 (Section 3.2.2) respectively. The Bvm theorem for the mixture prior is Theorem 10 in Section 3.3.7. Implementation of the mixture prior and numerical results are in Section 3.4.

Now we summarise other importat results and contributions in this thesis. For the sieve estimator we showed that the likelihood ratio process converges to an exponential distribution for the nonregular parameter and a Gaussian process for the regular nuisance parameter (See Theorem 3). Using widely studied properties of B-splines we investigated a log-spline prior for the density. Since the data is generated from a density supported on the whole semiline and support of B-splines is a finite interval, we did the approximation through a truncation of the true density and we had to adapt and extend the usual consistency results of B-splines to the case of an expanding support and that the densities can go to 0 as n goes to infinity (See Theorem 5 and Corollary 3). Finally we obtained a BvM theorem combining consistency and likelihood approximation.

The second approach is joint work with Judith Rousseau (University of Oxford) and J.B. Salomond (Université Paris-Est Créteil). We studied a mixture prior with a kernel constructed as a convolution of a Gamma distribution and a Uniform distribution. General sufficient conditions were found for a BvM Theorem under LAE assumption and it was shown that they are satisfied by our model with a mixture prior for f and a prior for nonregular parameter with positive and continuous density having polynomial tails. In particular, we proved L^1 consistency at minimax rate (up to a logarithmic factor, see Proposition 5), local uniform consistency near the discontinuity point (see Proposition 6), and the most challenging condition, the interaction term between function and parameter uniformly going to 0 (see Proposition 9). In order to obtain all of these conditions simultaneously we used a non-standard prior for the mixing distribution. We implemented this prior model and obtained numerical results from simulated data that illustrate theoretical results. Finally, we applied our algorithm to real data from procurement auctions (see Section 3.4.1).

The rest of the document is organised as follows. In the next sections of this chapter we present a review of relevant literature and the model of interest. Chapter 2 describes main results of the first approach using B-splines for sieve estimation. Our progress on the second approach with a mixture prior is shown in chapter 3. We finish with our conclusions and plans for future work in the last chapter.

1.2 Literature Review

Early works on estimating the location of a discontinuity of a density in a parametric model include [10], [43] and [40] which show asymptotic properties of MLE, while Polfeldt (1970) in [39] and [38] analysed the order of the minimum variance of unbiased estimators. The problem of estimating a one-dimensional nonregular parameter has been studied in detail by Ibragimov and Hasminskii (1981) in [30]. In particular, Chapter V presents their work on densities with discontinuities in a very general framework, with several points of discontinuity depending on the parameter, weak assumptions on smoothness and covering both one and two-sided kind of jumps. The first refers to jumps from zero to a positive value and the second when jumping from a positive value to another. They reduced the problem to studying the likelihood ratio as a stochastic process, previously used in [43] and [40], this time used to derive properties for both frequentist and Bayesian estimators. They also proved that under some conditions the minimax rate of convergence is n instead of $\sqrt{n}$ as in regular models. Additionally the concept of Local Asymptotic Exponentiality (see Section 1.3) was introduced analogously to Local Asymptotic Normality defined by Le Cam [34] and a convolution theorem similar to the one deduced by Hájek (1972) [29]. Some generalisations are due to Pflug (1982, 1983), regarding the dimensionality of the nonregular parameter [36] and the jump measure of the process [37]. Goria (1982) [26] proposed some estimators for a specific family of densities with a discontinuity and Smith (1985) [48] studied asymptotic properties of MLE for another familiy of densities with one nonregular parameter and some other regular parameters. Later works by Ghosh et al (1994) [25], Ghosal et al (1995) [20] provide sufficient and necessary conditions over convergence of posterior distribution under the setup of Ibragimov and Hasminskii. Furthermore Ghosal and Samanta (1995) extended their results and gave such conditions when a multidimensional regular parameter is added [22].

There is also a connection between our model of interest and Extreme Value Theory (EVT). Indeed, Extremal Types theorem, one of the main results in EVT, claims that the distribution of the maximum $X_{(n)}$ (or minimum $X_{(1)}$ in our case) of an i.i.d. sample $X_1, \ldots, X_n$, always converges to a Generalised Extreme Value distribution (GEV), characterised by three parameters, $\mu \in \mathbb{R}, \sigma > 0$ and $\gamma \in \mathbb{R}$ which correspond to location, scale and shape. Parameter γ is also called Extreme Value Index (EVI), and it is related to the tail of the original density. For more details and formal definitions we refer to [1]. In our context, the density has a left tail with a finite end point which means that the EVI is negative, and in fact, having a discontinuity at this point corresponds to $\gamma = -1$. For this particular value, the GEV becomes a shifted Exponential distribution with scale σ and located at $\mu+\sigma$. Thus, EVT ensures that the minimum is asymptotically distributed this Exponential distribution and that our parameter of interest θ is equal to $\mu + \sigma$. Effectively, this coincides with the result from nonregular models showing that in our context $n(X_{(1)} - \theta_0)$ is distributed as an Exponential with rate equal to $f_0(0)$. This is an important fact since $n(X_{(1)} - \theta_0)$ plays the role of the centering variable in the definition of Local Asymptotic Exponentiality (LAE) introduced in Section 1.3, which in turn is key to obtain a BvM theorem in the same way that Local Asymptotic Normality is needed to have a BvM theorem in the regular case. In conclusion we could say that the relation between Extremal Types Theorem and LAE is analogous to the Central Limit Theorem and LAN in the regular case.

Applications of nonregular models are usually found in Econometrics, for instance auction models are analysed by Donald and Paarsch (1993, 1996) (see [14], [15]) using MLE in nonregular contexts where the support of the density of bids depends on the estimated parameter. They extend the results in [30] by modeling the location of the jump by a regression curve with discrete regressors. In more recent works, Hirano and Porter (2003) [28] use local asymptotic minimax criterion to compare efficiency of MLE and BEs in a more general regression model. Moreover, Chernozhukov and Hong (2004) [11] develop likelihood-based estimation and inference methods addressing both one and two sided jumps primarily motivated by procurement auctions and equilibrium job-search models respectively.

However, most of nonregular theory has been developed only for parametric models. Indeed, all of the literature mentioned above addresses that kind of problems. Similarly, theory of regular semiparametric models has been well developed in the frequentist framework; see for instance [2] or [49]. In the Bayesian framework, most of the work has been very recent, for instance, [3] by Bickel and Kleijn (2012) and [9] by Castillo and Rousseau (2013) among others ([8, 42, 35]) have proved semiparametric versions of Bernstein-Von Mises Theorem.

In its semiparametric version, the problem of estimating the boundary points of the support of an unknown density is far from being completely answered, specially for Bayesian estimators as stated in [32]. In the frequentist framework, Chu and Cheng (1996) [13] use a kernel density estimator to estimate the nonparametric part of the model but their parameter of interest is estimated with a rate slower than n. Gayraud (2002) [18] proposes an estimator which achieves that rate based on differences of histograms, however it is defined only for jumps located in the interior the support of the density. On the other hand, in the Bayesian framework Kleijn and Knapik (2013) [32] have a preprint where they propose a likelihood ratio based Theorem analogue to Bernstein-Von Mises for densities under Local Asymptotic Exponentiality under strong assumptions on the prior distribution. Related works on BvM theorems for nonregular models include [5] by Bochkina and Green (2014) where the parameter of interest lies on the boundary of the parameter space and the work by Resiss and Schmidt-Hieber (2018) [41] where where they study the recovery of the boundary function of the intensity of a Poisson point process.

In our first approach we consider a model with the one-dimensional nonregular parameter and a multidimensional regular nuisance parameter with dimension growing to infinity similar to works done in purely regular frameworks such as [19], [6,] or in an adaptive way with a prior on the dimension [45]. Now we apply this idea to our nonregular context and this way we find conditions over the rate of growth of the dimension such that the rate of convergence of parametric nonregular models is preserved which is a novel result.

Our second approach is closer to [32] but we use a mixture prior similar to that of Bochkina and Rousseau (2017) [4] that is more flexible. Other articles that contain useful results for this approach are [47], [9], [51], [21] and [24].

Regarding numerical results, we implemented a slice sampler algorithm as in Kalli and Griffin (2011) [31], but using different representations for our prior process.

1.3 Nonregular Models and LAE condition

We start defining what a nonregular model is in the i.i.d. case. For this purpose we need to define the concept of regularity.

Let $\mathcal{P}$ be a statistical model, that is a collection of probability measures, in some measurable space $(\mathbb{X}, \mu)$, parametrised by some finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^k$. Suppose that $X_1, \ldots, X_n$ are i.i.d. with common distribution $P \in \mathcal{P}$, dominated by μ, and consider the map $s : P \in \mathcal{P} \mapsto \sqrt{f} \in L^2(\mu)$, where $f = \frac{dP}{d\mu}$ is the density with respect to measure μ. We write $s(\theta)$ to denote $s(P_\theta)$, omitting its dependency on $\mathbb{X}$. When needed we specify it as $s(x, \theta)$ and we do similarly with f.

Definition 1. A parametric i.i.d. model is called **regular** if

(i) The parametrisation is Fréchet differentiable for every $\theta \in \Theta$, that is, there exists a linear operator $\dot{s}_\theta : \Theta \to L^2(\mu)$ such that,

$$\|s(\theta + h) - s(\theta) - \dot{s}_\theta(h)\|_{L^2(\mu)} = o(\|h\|_{\mathbb{R}^k})$$

and the map $\theta \to \dot{s}_\theta$ is continuous. Note that $\dot{s}_\theta$ can be identified with a k-dimensional vector of elements of $L^2(\mu)$, say $(\dot{s}_1(\cdot, \theta), \dot{s}_2(\cdot, \theta), \ldots, \dot{s}_k(\cdot, \theta))$.

(ii) The matrix $\int \dot{s}_\theta \dot{s}_\theta^t d\mu$ is nonsingular.

In literature, the Fréchet differentiability condition stated above is found also as differentiable in quadratic mean or Hellinger differentiable. The following Proposition gives a sufficient condition to establish regularity of a model that in general is easier to check.

Proposition 1. *Suppose Θ is open and for all $\theta \in \Theta$:*

(i) $f(x, \theta)$ is continuously differentiable in θ for almost all x with gradient $\dot{f}(x, \theta)$.

(ii) $\left\|\dot{l}(\theta)\right\| \in L^2(P_\theta)$, where $\dot{l}(\theta) = \frac{\dot{f}(\theta)}{f(\theta)} \mathbb{1}(f(\theta) > 0)$.

(iii) the matrix $\int \dot{l}(\theta)\dot{l}(\theta)^t$ is nonsingular and continuous in θ.

Then the parametrisation $\theta \to P_\theta$ of the model is regular with $\dot{s_\theta} = \frac{1}{2} f(\theta)^{-1/2} \dot{f}(\theta) \mathbb{1}(f(\theta) > 0)$.

A proof can be found in [2]. It is worth mentioning that there exist models that are regular but do not satisfy assumptions in Proposition 1. The same reference provides an example.

We are interested in working on a semiparametric model, therefore we need to define regularity within that framework.

Definition 2. Consider a semiparametric model $\mathcal{P}$ and a fixed $P_0 \in \mathcal{P}$. P_0 is said to be regular if it belongs to a regular parametric submodel $\mathcal{Q} \subset \mathcal{P}$. Normally P_0 represents the measure for the 'true values' of the parameters.

An important class of nonregular models is the set of models that satisfy *Local asymptotic exponentiality* (LAE). Let us formally define the concept of LAE.

Definition 3. Denote by $\mathcal{E}(\gamma)$ an Exponential distribution with parameter $\gamma > 0$. Assume that the data $\mathbf{X}^n = (X_1, \cdots, X_n)$ has probability density $f^n_{\theta,\eta}$, $\theta \in \mathbb{R}$ and $\eta \in H$ where H is possibly infinite dimensional. We say this model satisfies the *Local asymptotic exponentiality* (LAE) condition at θ_0, η_0, if there exists γ_0 and a random variable ζ_n (also called centering variable), such that ζ_n converges in distribution to an $\mathcal{E}(\gamma_0)$ as n goes to infinity under $f^n_{\theta_0,\eta_0}$, and for all $h \in \mathbb{R}$, all h_n converging to h,

$$\frac{f^n_{\theta_0 + \frac{h_n}{n}, \eta_0}(\mathbf{X}^n)}{f^n_{\theta_0,\eta_0}(\mathbf{X}^n)} = \exp\{\gamma_0 h + R_n\} \mathbb{1}(h \leq \zeta_n), \quad P_{\theta_0,\eta_0}(|R_n| > \epsilon) = o(1), \tag{1.1}$$

for all $\epsilon > 0$. Moreover the LAE condition holds uniformly over a subset $\Theta_0 \times H_0$ of $\mathbb{R} \times H$ if

$$\sup_{\theta_0 \in \Theta_0, \eta_0 \in H_0} P_{\theta_0,\eta_0}(|R_n| > \epsilon) = o(1), \quad \forall \epsilon > 0.$$

1.4 Model of interest

Now we proceed to describe our Bayesian model of interest presenting the likelihood and general setup for prior distribution.

1.4.1 Likelihood

Consider the following semi-parametric model. Assume $X_1, \ldots, X_n$ are i.i.d. random variables $X_i \sim f_\theta(\cdot) = f(\cdot - \theta)$ with unknown f and θ, where $\theta \in \Theta \subseteq \mathbb{R}$ and f belongs to $\mathcal{F} \subset L^1(\mathbb{R}^+)$.

Note that $X_{(1)} = \min_{i=1,\ldots,n} X_i$ is a sufficient statistic for θ.

The class of functions $\mathcal{F}$ will vary between chapters, but it always contains functions with the following conditions

(i) $f(x) \geq 0$ for all $x > 0$ and $\int_{\mathbb{R}^+} f = 1$

(ii) f has a discontinuity at 0 but it is continuous from the right, with $\lim_{x \searrow 0} f(x) > 0$

and the fact that the discontinuity of f_θ depends on the parameter θ is the key condition that makes this a LAE model.

1.4.2 Prior

We consider a Bayesian model with a prior of the form $(\theta, f) \sim \Pi_n = \Pi_1 \otimes \Pi_{2,n}$ where Π_1 is a probability distribution on the real line with density function $\pi_1(\cdot)$ that is continuous and positive at θ_0. Different additional conditions will be required in each model and will be specified in each chapter.

Regarding $\Pi_{2,n}$, as stated in [12] a prior distribution defined on an infinite-dimensional space such as our functional class should cover a large section of it in a topological sense in order to have consistent posterior. Such priors may be thought of as a stochastic process taking values in the given function space and are usually constructed by some mechanism depending on hyper-parameters that reflect prior believes. For instance, they may be put by describing a sampling scheme to generate a random function or by describing the finite dimensional laws.

In this work we study two possible prior distributions for the density $\Pi_{2,n}$ which correspond to our two approaches and they are described in the chapters 2 and 3 respectively. In general, our prior distribution will depend on n, the number of observations and thus we use the subscript in the notation.

1.4.3 Objective of the study

Our aim is to study the asymptotic marginal posterior distribution of the nonregular parameter θ, the posterior contraction rate for the density function f with respect to the L_1 norm, and derive a BvM theorem for this setup and each of the two prior models.

Chapter 2

Semiparametric BvM Theorem for a Sieve Prior

In this chapter we cover the details of the approach using a sieve estimator for the density, that is, we work with the space generated by a finite number functions of a given basis of the functional space. As the number of observations goes to infinity we also increase the number of basis functions considered. We study the limiting likelihood ratio process and general conditions of a Semiparametric Bernstein-von Mises Theorem in this context, and show that such conditions are satisfied by a Log-spline model. For this purpose we also prove consistency at nonparametric minimax rate.

2.1 Method of Sieves

Suppose we have a nonparametric model with an infinite-dimensional parameter $f \in \mathcal{F}$. Consider a sequence of approximating spaces $\mathcal{F}_n$ such that the closure of $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ coincides with the functional space $\mathcal{F}$. Usually the spaces F_n are constructed as the linear span of a basis with a finite number of elements, say $\{\phi_j\}_{1 \leq j \leq J_n}$, that is $\mathcal{F}_n = \text{span}\{\phi_1, \ldots, \phi_{J_n}\}$. Therefore, for any function in $\mathcal{F}$ and for all $\epsilon > 0$ there exists $n > 0$ such that $\|f - \sum_{j=1}^{J_n} \eta_j \phi_j\| < \epsilon$, where $\|\cdot\|$ is a suitable norm, typically supremum norm. Thus the method of sieves uses a sequence of approximate function spaces with increasing complexity over which estimation is carried out. These sets are called sieves. It is possible to generalise this idea using a link function Ψ so that $f = \Psi(\sum_{j=1}^{J_n} \eta_j \phi_j)/c(\eta)$, where $c(\eta)$ is the normalising factor and Ψ is smooth, strictly monotonic and with an inverse that is also smooth. Typically Ψ is chosen to be the exponential function.

Using the approximation introduced by sieves generates a bias corresponding to the approximation error that must be controlled allowing the dimension J_n to grow to infinity with the number of observations n, but the rate of growth is also determined by the rate of convergence of the estimator that we want to achieve.

It is also worth mentioning that in this method we map the parameters in the functional space $\mathcal{F}_n$ to $\mathbb{R}^{J_n}$ and then priors can be defined as distributions over $\mathbb{R}^{J_n}$.

2.2 Setup

Consider the following semi-parametric model. Let $X^{(n)} = (X_1, \ldots, X_n)$ where $X_i \overset{i.i.d.}{\sim} f_{0,\theta_0} = f_0(\cdot - \theta_0)$ and $f_0 \in \mathcal{F} \subset L^1(\mathbb{R}^+)$. Here $\mathcal{F}$ is the class of functions such that

(i) $f(x) > 0$ for all $x > 0$ and $\int_{\mathbb{R}^+} f = 1$

(ii) f has a discontinuity at 0 but it is continuous from the right, with $\lim_{x \searrow 0} f(x) > 0$

(iii) f is Hölder continuous on $(0, \infty)$ with parameter $\beta > 2$.

(iv) f has a tail that satisfies,

$$e^{-cx^\tau} \lesssim f_0(x) \lesssim x^{-\kappa} \tag{2.1}$$

$$\left|\tfrac{d^t}{dx^t} \log f_0(x)\right| \lesssim x^\tau \tag{2.2}$$

for some $\kappa > 1$, $c, \tau > 0$, $t = 1, \ldots, \lfloor \beta \rfloor$, and $x > x_0$ for some x_0 large.

Let us define

$$f_n = \frac{f_0 \mathbb{1}([0, a_n])}{F_0(a_n)} \tag{2.3}$$

where $F_0(x) = \int_0^x f_0$ the c.d.f. of f_0, thus f_n corresponds to f_0 truncated on $[0, a_n]$. Throughout this chapter we will denote $\gamma_0 := f_0(0+)$.

Following the method described in the previous section, we consider a sequence of spaces $\mathcal{F}_n$ that only contains functions supported on $[0, a_n]$ that can be expressed as

$$f(\cdot; \eta) = \frac{\Psi(\sum_{j=1}^{J_n} \eta_j \phi_j)}{c(\eta)}$$

with a link function Ψ and a basis $\{\phi_j\}_{1 \le j \le J_n}$ as described in Section 2.1, and therefore it can be parametrised by a regular finite dimensional parameter denoted η with dimension J_n. In other words, the model is $X_i | \theta, \eta \overset{i.i.d.}{\sim} f_\theta(x; \eta) = f(x - \theta; \eta)$ where $\operatorname{supp}(f(\cdot; \eta)) = [0, a_n]$, $\theta \in \Theta \subset \mathbb{R}$ and $\eta \in H_{J_n} \subset \mathbb{R}^{J_n}$ with J_n and a_n being deterministic sequences that go to infinity as n grows. We denote $\eta_0 \in H_{J_n}$ the parameter such that $f(x; \eta_0)$ attains the best approximation to f_n, that is, $\eta_0 = \arg\min_{\eta \in H_{J_n}} \| \sum_{j=1}^{J_n} \eta_j \phi_j - \Psi^{-1}(f_n) \|_\infty$. Note that η_0 depends on n.

We construct a prior for this model simply by defining priors on parameters $\theta \in \mathbb{R}$ and $\eta \in H_{J_n}$ independently. Let us denote this prior $d\Pi(\theta, \eta) = d\Pi_\theta(\theta) d\Pi_\eta(\eta)$ with corresponding densities with respect to Lebesgue measure π, π_θ and π_η respectively. We assume that the density function π_θ is continuous and strictly positive at θ_0. For the sake of simplicity of notation we will drop subindexes θ and η when there is no ambiguity. Note that in general, prior on η will depend on n.

We denote the posterior distribution given a vector of data $\underline{X}^{(n)}$ as

$$\Pi_n(A | \underline{X}^{(n)}) = \frac{\int_A \prod_{i=1}^n f_\theta(\underline{X}_i; \eta) d\Pi(\theta, \eta)}{\int_{\mathbb{R} \times H_{J_n}} \prod_{i=1}^n f_\theta(\underline{X}_i; \eta) d\Pi(\theta, \eta)}$$

with $A \subset \mathbb{R} \times H_{J_n}$. We will use the same notation for the corresponding marginals for θ and η when it is expressed explicitly that $A \subset \mathbb{R}$ or $A \subset H_{J_n}$.

This definition coincides with the usual posterior distribution only when $\underline{X}^{(n)}$ equals $X^{(n)}$ the full vector of observations. In our context $\underline{X}^{(n)}$ represents a random subset of the observations, making it a nonstandard choice for the posterior, however we will define it as a subset that contains a growing proportion of the total number of observations, and containing all of them in the limit (with probability going to 1) as the data size grows.

The motivation to introduce this modified posterior distribution is following. Since we are working with a sample from a density supported on $[\theta_0, \infty]$ and our model uses functions $f_\theta(\cdot;\eta)$ with support $[\theta, \theta + a_n]$, whenever we sample an observation that is greater than $\theta + a_n$, the likelihood will vanish and the log-likelihood will not be finite. We could consider a_n to grow to infinity fast enough to cover all observations with high probability or even take the random sequence $(a_n \vee X_{(n)})$, with $X_{(n)}$ the maximum of the X_is. However this condition will be in conflict with the sufficient conditions found later to obtain consistency, especially when the true density f_0 decreases to 0 too fast. Therefore we will impose a_n to grow slowly and consider only observations that lie in the support of $f_\theta(\cdot;\eta)$ for all $\eta \in H$ and θ in some suitable subset of Θ. If θ_0 is known we can consider observations $\underline{X}^{(n)} = (X_i : X_i \leq \theta_0 + a_n)$, however in our model of interest this is not the case and we will consider $\underline{X}^{(n)} = (X_i : X_i \leq X_{(1)} + a_n - d_n/n)$ where $X_{(1)}$ is the minimum of the X_is and d_n is a deterministic sequence that grows to infinity slowly. More precise statements will be presented in Lemma 1 below, with conditions that ensure both definitions of $\underline{X}^{(n)}$ match and with high probability contains at least $p_n n$ observations, where p_n is a sequence that goes to 1 as n goes to infinity.

Choosing a restricted sample such as $\underline{X}^{(n)}$ aims to obtain convergence of $f(\cdot;\eta)$ to f_n, but this also affects θ since the likelihood vanishes when θ is less than $\underline{X}_{(\underline{n})} - a_n$ where $\underline{X}_{(\underline{n})}$ is the maximum of the observations that are smaller than a_n. This together with the fact that the likelihood always vanishes when $\theta > X_{(1)}$ (the minimum of the observations) determines the support of the posterior distribution for θ. Thus, Lemma 1 also provides a lower bound on the length of the support with high probability which is important for the study of the limiting posterior distribution of θ and in particular the BvM result. This will be revisited with more details in section 2.5.5.

Lemma 1. *Let $X^{(n)} = (X_1, \dots, X_n)$ where $X_i \overset{i.i.d.}{\sim} f_{0,\theta_0} = f_0(\cdot - \theta_0)$. Let a_n, d_n be deterministic sequences such that $a_n, d_n \to \infty$ as $n \to \infty$ and $d_n f_0(a_n) \to 0$. Let us define $\underline{\tilde{X}} = (X_i : \ X_i < \theta_0 + a_n)$, $\underline{X} = (X_i : \ X_i < X_{(1)} + a_n - d_n/n)$ and denote $\underline{\tilde{n}}$, $\underline{n}$ their corresponding number of components. Then*

$$P_0^{(n)}(\underline{\tilde{X}}_{(\underline{\tilde{n}})} \leq \theta_0 + a_n - 2d_n/n) \geq e^{-8d_n f_0(a_n)} \to 1 \tag{2.4}$$

and with probability at least $1 - \exp(-c_0 d_n) - (8 d_n f_0(a_n))^{1/2}$,

$$\underline{\tilde{X}} = \underline{X} \tag{2.5}$$

where $c_0 > 0$ depends only on $f_0(0)$. Additionally, with probability at least $1 - \rho_n$,

$$\underline{n} \geq p_n n \tag{2.6}$$

where

$$\begin{aligned}
p_n =&1, \ \rho_n = n(1 - F_0(a_n)) && \text{if } n(1 - F_0(a_n)) \to 0 \\
p_n =&2F_0(a_n) - 1, \ \rho_n = F(a_n)/n(1 - F_0(a_n)) && \text{if } n(1 - F_0(a_n)) \to \infty \\
p_n =&DF_0(a_n) - (D - 1), \ \rho_n = C_0/(l(D - 1)) && \text{if } n(1 - F_0(a_n)) \to l > 0
\end{aligned}$$

for some constant $C_0 > 0$.

The proof of this Lemma can be found in section 2.6.3.

We finalise this section defining some additional notation. Throughout this chapter we will use the following notation for localised parameters and likelihood ratio; $h := n(\theta - \theta_0)$, $g := \sqrt{n}(\eta - \eta_0)$, $\ell(\theta, \eta) = \sum_{i=1}^{n} \log f_\theta(\underline{X}_i; \eta)$ and $Z_n(h, g) = \exp(\ell(\theta_0 + h/n, \eta_0 + g/\sqrt{n}) - \ell(\theta_0, \eta_0))$. Additionally, P_0 refers to the probability measure associated to f_0 and similarly P_{f_n} the probability measure corresponding to the truncated density f_n.

In this chapter density functions are usually compared using Hellinger distance that is expressed as $d_H(\cdot, \cdot)$.

We also denote by $\mathcal{E}(\gamma)$ the exponential distribution with parameter $\gamma > 0$. We add a subscript to represent a shifted exponential distribution and the minus sign $(-)$ to denote a negative exponential. For instance $\mathcal{E}_t^-(\gamma)$ corresponds to a negative exponential distribution supported on $(-\infty, t]$.

2.3 Bernstein-von Mises with a Sieve Prior

In this section we present sufficient conditions for a nonregular version of a Bernstein-von Mises type of result with a Sieve prior for the density.

Proposition 2. *Consider the model described in section 2.2 and let*

$$A_n = \left\{ (\theta, \eta) : |\theta - \theta_0| \leq \frac{R_n}{n},\ \|\eta - \eta_0\|_2 \leq \frac{\sqrt{J_n} S_n}{\sqrt{n}} \right\} \tag{2.7}$$

where R_n and S_n are sequences that go to infinity as n goes to infinity, such that $R_n/n \to 0$ and $\sqrt{J_n} S_n/\sqrt{n} \to 0$. For all $(\theta, \eta) \in A_n$ denote $h = n(\theta - \theta_0)$ and $g = \sqrt{n}(\eta - \eta_0)$. Suppose that the following assumptions hold.

(H1) Assume that

$$\sup_{(\theta,\eta) \in A_n} \left| \log Z_n(h, g) - \left(\gamma_0 h + g^t \Delta_n - \frac{1}{2} g^t i(\eta_0) g \right) \right| \mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n \right) \xrightarrow[n \to \infty]{P_0} 0$$

and $Z(h, g)$ vanishes for all $h \notin [\tilde{\zeta}_n, \zeta_n]$, where $i(\eta_0) = -\int (\frac{\partial^2}{\partial \eta^2} f(x; \eta_0)) f_0(x) dx$, $\Delta_n = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta} \ell(\theta_0, \eta_0)$, $\zeta_n = n\left(X_{(1)} - \theta_0 \right)$ and $\tilde{\zeta}_n = n(\underline{X}_{(\underline{n})} - a_n - \theta_0)$.

(H2) $\Pi_n(A_n^C | \underline{X}^{(n)}) \to 0$ in P_0-probability.

Then

$$\|\Pi_n - \mathcal{E}_{\zeta_n}^-(\gamma_0)\|_{TV} \to 0 \tag{2.8}$$

in P_0-probability, where Π_n denotes the posterior distribution of h.

Note that from frequentist results we know that $\zeta_n \to \mathcal{E}(\gamma_0)$ and by equation (2.4) from Lemma 1, $\tilde{\zeta}_n \overset{P_0}{\to} -\infty$ since $d_n \to \infty$. Thus, assumption (H1) implies that this model satisfies LAE condition at (θ_0, η_0).

Proof. Let A a measurable subset of $\mathbb{R}$.

$$\left|\int_A \pi(h|\underline{X}^{(n)}) - \gamma_0 e^{\gamma_0(h-\zeta_n)}dh\right|$$
$$\leq \int_{A\cap\{|h|\leq R_n\}} \left|\frac{\int_{\|g\|_2\leq\sqrt{J_n}S_n} Z_n(h,g)\pi_n(h)d\Pi_n(g)}{\int_{A_n} Z_n(h,g)d\Pi_n(h,g)} - \gamma_0 e^{\gamma_0(h-\zeta_n)}\right| dh\Pi_n(A_n|\underline{X}^{(n)})$$
$$+\Pi_n(A_n^C|\underline{X}^{(n)}) + \int_{|h|\geq R_n} \gamma_0 e^{\gamma_0(h-\zeta_n)}dh \quad (2.9)$$

We prove that the right-hand side of the inequality goes to 0 in P_0-probability. Indeed, the second term goes to 0 in probability by assumption (H2). For n sufficiently large $R_n > \zeta_n$ with probability tending to 1, since $R_n \to \infty$ and $\zeta_n \overset{P_0}{\to} \zeta \sim \mathcal{E}(\gamma_0)$ then the third term is equal to

$$\int_{-\infty}^{-R_n} \gamma_0 e^{\gamma_0(h-\zeta_n)}dh = e^{-\gamma_0(R_n+\zeta_n)} \overset{P_0}{\to} 0$$

Finally, we bound the first term. By assumption (H1), for all (h,g) such that the corresponding (θ,η) is in the set A_n, $Z_n(h,g)$ can be expressed as

$$Z_n(h,g) = \exp(\gamma_0 h)\mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n\right)\exp\left(g^t\Delta_n - \frac{1}{2}g^t i(\eta_0)g\right)\exp(o_P(1))$$

where the term $o_P(1)$ goes to zero uniformly over $(\theta,\eta) \in A_n$. Similarly, $\pi_n(h) = \pi_n(0)(1+o(1))$ by continuity of π_n. Therefore,

$$\frac{\int_{\|g\|_2\leq\sqrt{J_n}S_n} Z_n(h,g)\pi_n(h)d\Pi_n(g)}{\int_{A_n} Z_n(h,g)d\Pi_n(h,g)} = \frac{e^{\gamma_0 h}\mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n\right)}{\int e^{\gamma_0 h}\mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n \cap |h| \leq R_n\right)}(1+o_P(1))$$

since the term involving g cancels out. We conclude the proof noting that

$$\left|\int e^{\gamma_0 h}\mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n \cap |h| \leq R_n\right) - \gamma_0 e^{\zeta_n}\right| = o_P(1)$$

since $R_n \to \infty$ and $\tilde{\zeta}_n \overset{P_0}{\to} -\infty$ by Lemma 1. ☐

2.4 Limiting Likelihood ratio process

The following Theorem shows log-likelihood approximation as the number of observations goes to infinity. It shows that in the limiting log-likelihood normalised by corresponding rates of convergence can be expressed as the sum of log-likelihood of a Negative Exponential density in θ and a Gaussian Process in η. This also implies that in the limit there is no interaction between the nonregular parameter θ and the regular parameter η.

Theorem 1. *Let $X^{(n)} = (X_1,\ldots,X_n)$ where $X_i \overset{i.i.d.}{\sim} f_{0,\theta_0} = f_0(\cdot - \theta_0)$. Let a_n, d_n be deterministic sequences such that $a_n, d_n \to \infty$ as $n \to \infty$ and $d_n f_0(a_n) \to 0$. Consider the model $f_\theta(x;\eta) = f(x-\theta;\eta)$ supported on $[0,a_n]$ with $\theta \in \Theta \subset \mathbb{R}$ and $\eta \in H_{J_n} \subset \mathbb{R}^{J_n}$, where $f_{\theta_0}(x;\eta_0)$ the "true" misspecified density. Define $\underline{X} = (X_i : \ X_i < X_{(1)} + a_n - d_n/n)$, and denote $\underline{n}$ its* *number of components. Let us define the set*

$$A_n = \left\{(\theta, \eta) : |\theta - \theta_0| \leq \frac{R_n}{n},\ \|\eta - \eta_0\|_2 \leq \frac{\sqrt{J_n} S_n}{\sqrt{n}}\right\} \tag{2.10}$$

where R_n and S_n are sequences that go to infinity as n goes to infinity, such that $R_n/n \to 0$ and $\sqrt{J_n} S_n/\sqrt{n} \to 0$. Let $(\theta, \eta) \in A_n$ and define $h = n(\theta - \theta_0)$, $g = \sqrt{n}(\eta - \eta_0)$ and $\ell(\theta, \eta) = \sum_{i=1}^{n} \log f_\theta(\underline{X}_i, \eta)$. Suppose the following assumptions hold,

(I) $f_\theta(x; \eta)$ *is continuously differentiable in θ and twice continuously differentiable in η, $\forall x \in (\theta, \theta + a_n)$*

(II) $\int_0^\infty |f_0'(x)|\, dx < \infty$

(III) $\sup_{\|g\|_2 \leq \sqrt{J_n} S_n} |\frac{1}{2n} g^\top \frac{\partial^2}{\partial \eta^2} \ell(\theta_0, \eta_0) g - \frac{1}{2} g^t (\int (\frac{\partial^2}{\partial \eta^2} f(x; \eta_0)) f_0(x) dx) g| = o_P(1)$

(IV) $\sup_{|h| \leq R_n} \sup_{\tilde{\theta} \in (\theta, \theta_0)} \left| \frac{h}{n} \left(\frac{\partial}{\partial \theta} \ell(\tilde{\theta}, \eta_0) - \frac{\partial}{\partial \theta} \ell(\theta_0, \eta_0) \right) \right| = o_P(1).$

(V) $\sup_{\|g\|_2 \leq \sqrt{J_n} S_n} \sup_{\tilde{\eta} \in \langle \eta, \eta_0 \rangle} \frac{1}{n} \left| g^\top \left(\frac{\partial^2}{\partial \eta^2} \ell(\theta_0, \tilde{\eta}) - \frac{\partial^2}{\partial \eta^2} \ell(\theta_0, \eta_0) \right) g \right| = o_P(1)$ *where $\langle \eta, \eta_0 \rangle$ is the line connecting η and η_0.*

(VI) $\sup_{(\theta, \eta) \in A_n} |\ell(\theta, \eta) - \ell(\theta, \eta_0) - \ell(\theta_0, \eta) + \ell(\theta_0, \eta_0)| = o_P(1)$

(VII) $\left| \frac{1}{\underline{n}} \sum_{i=1}^{\underline{n}} \frac{\partial}{\partial \theta} \log f_{\theta_0}(\underline{X}_i; \eta_0) - \frac{1}{\underline{n}} \sum_{i=1}^{\underline{n}} \frac{\partial}{\partial \theta} \log f_{n, \theta_0}(\underline{X}_i) \right| = o_P(1)$ *where f_n is defined in equation* (2.3)*.*

Then the localised likelihood ratio tends up to a constant to the product of the density of a Negative Exponential distribution in h and a Gaussian process in g, i.e.,

$$\sup_{(\theta, \eta) \in A_n} \left| \log Z_n(h, g) - \left(\gamma_0 h + g^t \Delta_n - \frac{1}{2} g^t i(\eta_0) g \right) \right| \mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n \right) \xrightarrow[n \to \infty]{P_0} 0$$

and $Z(h, g)$ vanishes for all $h \notin [\tilde{\zeta}_n, \zeta_n]$, where $i(\eta_0) = -\int (\frac{\partial^2}{\partial \eta^2} f(x; \eta_0)) f_0(x) dx$, $\Delta_n = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta} \ell(\theta_0, \eta_0)$, $\zeta_n = n \left(X_{(1)} - \theta_0 \right)$ and $\tilde{\zeta}_n = n(\underline{X}_{(\underline{n})} - a_n - \theta_0)$.

The proof of this Theorem can be found in section 2.6.1.

2.5 Application to a Log-Spline estimator

Now we proceed to show the application of Theorem 1 to a Log-Spline estimator for the density function. We analyse this checking the conditions of the Theorem in this particular case. We start showing the definition of a B-spline estimator and some basic properties.

2.5.1 B-splines and Log-Spline model

We investigate a Log-Spline model where the log-likelihood is a linear combination of B-Splines, namely

$$
\begin{aligned}
f_\theta(x;\eta) =& f(x-\theta;\eta) \\
f(x;\eta) =& e^{\eta^t B_q(x) - c(\eta)} = e^{\sum_{j=1}^{J_n} \eta_j B_{j,q}(x) - c(\eta)} \\
c(\eta) =& \log \int_0^\infty e^{\eta^t B_q(x)} dx
\end{aligned} \tag{2.11}
$$

where $J_n > 0$ is the total number of B-Splines which we let grow with n, $B_{j,q}(x)$ is the B-Spline of order q with knots $\{t_{j-q}, \dots, t_j\}$ which are defined as

$$
t_i = \begin{cases} 0, & i = -q+1, \dots, 0 \\ i/K_n, & i = 1, \dots, a_n K_n - 1 \\ a_n, & i = a_n K_n, \dots, a_n K_n + q - 1 \end{cases}
$$

and $K_n = (J_n - q + 1)/a_n$.

Now denote $\eta_0 \in \mathbb{R}^{J_n}$, the vector of coefficients corresponding to the best approximation of $\log f_n$, that is $\eta_0 = \arg\min_{\eta \in H_{J_n}} \|\eta^t B_q - \log f_n\|_\infty$, then

$$
\|\log f_n - \eta_0^t B_q\|_\infty \leq C_{q,\beta} K_n^{-\beta} \|\log f_n\|_{C^\beta}. \tag{2.12}
$$

for some constant $C_{q,\beta} > 0$ and $\|\eta_0\|_\infty \leq \|\log f_n\|_{C^\beta}$, as long as $q \geq \beta$. A similar bound holds for the derivative, namely, there exists a constant $C'_{q,\beta} > 0$ such that

$$
\|(\log f_n)' - \eta_0^t B'_q\|_\infty \leq C'_{q,\beta} K_n^{-(\beta-1)} \|\log f_n\|_{C^\beta}. \tag{2.13}
$$

this is a well known result from B-splines approximation theory that can be found for example in [7] or [46].

Moreover, the following equations bound the uniform norm of differences between densities in terms of uniform norm of parameters and will be helpful throughout the chapter. Indeed, from Lemma 9.2 and the proof of Lemma 2.5 in [23],

$$
|c(\eta) - c(\eta_0)| \leq \|(\eta - \eta_0)\top B_q\|_\infty \tag{2.14}
$$

$$
|c(\eta_0)| \leq \|\eta_0^\top B_q - \log f_n\|_\infty \tag{2.15}
$$

therefore

$$
\begin{aligned}
|\log f(x;\eta) - \log f(x;\eta_0)| \leq& 2\|(\eta - \eta_0)^\top B_q\|_\infty \leq 2\|\eta - \eta_0\|_\infty \\
\leq& 2\|\eta - \eta_0\|_2
\end{aligned} \tag{2.16}
$$

$$
|\log f(x;\eta_0) - \log f_n(x)| \leq 2\|\eta_0^\top B_q - \log f_n\|_\infty \tag{2.17}
$$

Now we proceed to study the derivatives of the log of the density. Note that

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log f_\theta(x;\eta) =& -\eta^t B_q'(x-\theta)\\
\frac{\partial}{\partial\eta_j}\log f_\theta(x;\eta) =& B_j(x-\theta)-\frac{\partial}{\partial\eta_j}c(\eta)\\
\frac{\partial^2}{\partial\theta\partial\eta_j}\log f_\theta(x;\eta) =& -B_j'(x-\theta)\\
\frac{\partial^2}{\partial\eta_j\partial\eta_k}\log f_\theta(x;\eta) =& -\frac{\partial^2}{\partial\eta_j\partial\eta_k}c(\eta)
\end{aligned} \tag{2.18}$$

and

$$\begin{aligned}
\frac{\partial}{\partial\eta_j}c(\eta) =& \int_0^\infty B_{j,q}(x)f(x;\eta)dx\\
\frac{\partial^2}{\partial\eta_j\partial\eta_k}c(\eta) =& \int_0^\infty B_{j,q}(x)B_{k,q}(x)f(x;\eta)dx - \frac{\partial}{\partial\eta_j}c(\eta)\frac{\partial}{\partial\eta_k}c(\eta)
\end{aligned} \tag{2.19}$$

The following formula is helpful to control the derivative $B_q'(x)$ expressing it in terms of $B_{q-1}(x)$. Indeed, for all $j=1,\dots,J_n$

$$B_{j,q}'(x) = (q-1)\left(\frac{B_{j-1,q-1}(x)}{t_{j-1}-t_{j-q}} - \frac{B_{j,q-1}(x)}{t_j-t_{j+1-q}}\right) \tag{2.20}$$

Example 1 (a B-Spline and its derivative)**.** We show the first B-Spline and its derivative for $q=3$ and $J_n=5$. The corresponding full set of knots is $(t)_{i=-2}^5 = \{0,0,0,a_n/3,2a_n/3,a_n,a_n,a_n\}$, and the first four of them are used to construct $B_{1,3}$. Effectively, using the Cox-de Boor recursion formula

$$B_{1,3}(x) = \frac{t_1-x}{t_1-t_{-1}}B_{1,2}(x) = (1-\tfrac{3x}{a_n})B_{1,2}(x) = \begin{cases}(1-\frac{3x}{a_n})^2 & x\in[0,a_n/3)\\ 0 & \text{otherwise}\end{cases}$$

and applying formula (2.20) we obtain

$$\begin{aligned}
(q-1)\left(\frac{B_{j-1,q-1}(x)}{t_{j-1}-t_{j-q}} - \frac{B_{j,q-1}(x)}{t_j-t_{j+1-q}}\right) =& -2\frac{B_{1,2}(x)}{t_1-t_{-1}}\\
=& \begin{cases}-\frac{6}{a_n}(1-\frac{3x}{a_n}) & x\in[0,a_n/3)\\ 0 & \text{otherwise}\end{cases}
\end{aligned}$$

which is indeed the derivative.

We finish this section with some comments on the support of the prior defined by the log-spline model (2.11). Due to uniform approximation property when $\theta=\theta_0$ is known given by (2.12), the log-spline model has full support for L_1 convergence and KL divergence on the set of restrictions to $[\theta_0,\theta_0+a_n]$ of elements of our functional parameter class $\mathcal{F}$ as long as $q\geq\beta$. It actually charges all densities on $[\theta_0,\theta_0+a_n]$ with regularity $\beta\leq q$, and this is because under the exponential link model for densities $p_f=e^{f-c(f)}$, KL divergence and L_1 distance between two densities p_f and p_g are bounded by the uniform distance between f and g that share the same support (see Lemma 2.5 in [23]). However when considering the joint prior, the KL support is empty since for any pair (θ_0, f_0) the KL divergence $KL(f_{0,\theta_0}, f_\theta(\cdot;\eta)) = \infty$ for any $\theta \neq \theta_0$ and any $\eta \in H_J$. This is a known issue for parameters that shift densities supported on finite intervals, nonetheless, L_1 support of a prior is always bigger than the KL support and thus having full KL support is not a necessary condition obtain L_1 consistency as we show later in this chapter.

2.5.2 Bernstein-von Mises for a Log-spline model

In this section we present a nonregular Bernstein-von Mises Theorem for the log-spline model applying Proposition 2.

Theorem 2. *Let $\beta > 5/2 + \delta$ with $\delta > 0$ arbitrarily small. Suppose $f_0 \in C^\beta$ is uniformly bounded from above by $\bar{M} > 0$ and has a tail that satisfies* (2.1) *and* (2.2). *Let $X_1, \ldots, X_n \sim f_0$. Consider the Log-Spline model with $q \geq \beta$, $J_n \asymp n^{1/(2\beta+1)}$, $K_n = (J_n - q + 1)/a_n$ and, a_n is such that $a_n \to \infty$ and $a_n^\tau \lesssim \log\log n$ (e.g. $a_n \asymp (\log\log n)^\gamma$ with $0 < \gamma < 1/\tau$ or $a_n \asymp \log\log\log n$).*

Let $\eta \sim \pi_{\eta,n}$ the density of a prior defined on $\{\eta \in [-M_n, M_n]^{J_n} : \eta^\top 1 = 0\}$ such that $\underline{c}^{J_n} \leq \pi_n(\eta) \leq \bar{c}^{J_n}$, for some $0 < \underline{c} < \bar{c} < \infty$.

Let $\theta \sim \pi_{\theta,n}$ the density of a prior defined on $\mathbb{R}$ such that $\sup_\theta \pi(\theta) \leq \bar{k}$, and for all $\theta_0 \in \mathbb{R}$ there exists $t_0 > 0$, such that $\inf_{[\theta_0 - t_0, \theta_0]} \pi(\theta) \geq \underline{k}$, for some $0 < \underline{k} < \bar{k} < \infty$. Then,

$$\|\Pi_n - \mathcal{E}^-_{\zeta_n}(\gamma_0)\|_{TV} \to 0 \tag{2.21}$$

in P_0-probability, where Π_n denotes the posterior distribution of h.

Proof. The proof consists in verifying the two assumptions in Proposition 2. Corollary 3 in section 2.5.5 states that posterior convergence rate for θ is $n^{-2\beta/(2\beta+1)}(\log n)^2(\log\log n)^{2(2+\beta/\tau)}$ and the rate for η is $n^{-\beta/(2\beta+1)}(\log n)^2(\log\log n)^{2+\beta/\tau+1/2}$. Therefore we can consider the definition of set A_n with $S_n = n^s$ and $R_n = n^r$ where $r = 1/(2\beta+1) + \delta/(3(\delta+3))$ and $s = r/2$, which implies condition (H2). Condition (H1) holds due to Theorem 3 in section 2.5.3 with $\rho = 1/(2\beta+1)$. Indeed, let us check that the four assumptions of Theorem 3 hold. Assumption (i) holds by definition of a_n. Replacing $s = r/2$, assumption (iii) becomes $\rho + 3r < 1$ and assumption (iv), $\rho + r < 1/3$. Therefore (iv) implies (ii) and (iii). Now we conclude noting that $\rho + r = 2/(2\beta+1) + \delta/(3(\delta+3))$ and given $\beta > 5/2 + \delta$, then $\rho + r < 1/(3+\delta) + \delta/(3(\delta+3)) = 1/3$. $\square$

In the next sections we will obtain the results used in this proof.

2.5.3 Limiting Likelihood ratio process

We show that the log-spline model satisfies conditions of Theorem 1.

Theorem 3. *Let $\beta > 1$. Suppose $f_0 \in C^\beta$ is uniformly bounded from above, by $\bar{M} > 0$, $\int |f_0'(x)|dx < \infty$ and has a tail that satisfies conditions* (2.1) *and* (2.2). *Assume $R_n \asymp n^r$ and $S_n \asymp n^s$ for some $r, s > 0$, and define set A_n as in* (2.7). *Let $(\theta, \eta) \in A_n$ and define $h = n(\theta - \theta_0)$, $g = \sqrt{n}(\eta - \eta_0)$.*

Consider the B-Spline model with $q \geq 3$, $J_n \asymp n^\rho$ and $K_n = (J_n - q + 1)/a_n$ where and a_n and $\rho > 0$ satisfy

(i) *$a_n \to \infty$ and $a_n = o(n^a)$ for any $a > 0$.*

(ii) *$\rho + r < 1/2$*

(iii) *$\rho + 6s < 1$*

(iv) $3\rho + 2(r+s) < 1$

then the localised likelihood ratio $Z_n(h,g)$ tends up to a constant to the product of the density of a Negative Exponential distribution in h and a Gaussian process in g, i.e.,

$$\sup_{(\theta,\eta)\in A_n} \left|\log Z_n(h,g) - \left(\gamma_0 h + g^t \Delta_n - \frac{1}{2} g^t i(\eta_0) g\right)\right| \mathbb{1}\left(\tilde{\zeta}_n < h < \zeta_n\right) \xrightarrow[n\to\infty]{P_0} 0$$

and $Z(h,g)$ vanishes for all $h \notin [\tilde{\zeta_n}, \zeta_n]$, where $i(\eta_0) = -\int(\frac{\partial^2}{\partial\eta^2} f(x;\eta_0)) f_0(x)dx$, $\Delta_n = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\eta}\ell(\theta_0,\eta_0)$, $\zeta_n = n\left(X_{(1)} - \theta_0\right)$ and $\tilde{\zeta}_n = n(\underline{X}_{(\underline{n})} - a_n - \theta_0)$.

This Theorem is proved in section 2.6.1.

Conditions (ii), (iii) and (iv) set an upper bound on the rate of growth of dimension (or number of components) in the model. They also imply that the bigger the order of neighbourhoods around (θ_0, η_0) considered, the smaller the order of dimension should be. In particular if we imposed $r = s = 0$, for instance with $R_n = S_n = \log n$ (since R_n and S_n have to go to infinity), we obtain the usual critical dimension of $n^{1/3}$. However given that the posterior distribution does not concentrate on balls around η_0 of order $1\sqrt{n}$ but at nonparametric rate which are slightly larger, we state our conditions for likelihood approximation with r and s strictly positive.

Similarly, we can obtain the likelihood approximation relaxing condition (i) to just $a_n \to \infty$, however if we allow this sequence to grow too fast it will interfere in conditions (ii)-(iv) and in order to obtain consistency it is required that a_n goes to infinity rather slowly.

2.5.4 Convergence rate with known θ_0

Now we prove consistency. Firstly, we show consistency and find contraction rate in parameter η only assuming θ_0 known, by extending the results in section 9.1 of [23]. The main difference between the model studied there and our model is the fact that the density that generates the data changes with n. In particular the support expands and the minimum of the function goes to 0 as n grows to infinity. We find suitable conditions for the rates of growth of the density support and decrease of the density minimum, in order to obtain the minimax nonparametric rate in Hellinger distance. Our model assumes the density is uniformly bounded from above, but we believe similar results should hold when allowing the maximum to increase as fast as the minimum decreases.

We include these findings as they are of interest in themselves but also they are a helpful to understand how to prove joint consistency later.

Let us define the following sets that are useful for the rest of this chapter.

$$\begin{aligned}
H_{J,M} &= \{\eta \in [-M,M]^J : \eta^\top 1 = 0\} \\
H(J,\epsilon) &= \{\eta \in H_{J,M} : \|\eta - \eta_0\|_2 \le \sqrt{J}\epsilon\} \\
B_{J,M}(f_n,\epsilon) &= \{\eta \in H_{J,M} : K(f_n; f(\cdot;\eta)) < \epsilon^2, V_2(f_n; f(\cdot;\eta)) < \epsilon^2\} \\
B_2(f_n,\epsilon) &= \{f(\cdot;\eta) : \eta \in B_{J,M}(f_n,\epsilon)\} \\
C_{J,M}(f_n,\epsilon) &= \{\eta \in H_{J,M} : d_H(f_n, f(\cdot;\eta)) < \epsilon\}
\end{aligned} \tag{2.22}$$

Theorem 4. *Let $\beta > 1$. Suppose $f_0 \in C^\beta$ is uniformly bounded from above by $\bar{M} > 0$ and has a tail that satisfies* (2.1), (2.2). *Let $X_1, \ldots, X_n \sim f_n$ with f_n from equation* (2.3). *Consider the Log-Spline model with $q \ge \beta$, $J_n \asymp n^{1/(2\beta+1)}$, $K_n = (J_n - q + 1)/a_n$ and, a_n is such that $a_n \to \infty$ and $a_n^\tau \lesssim \log\log n$ (e.g. $a_n \asymp (\log\log n)^\gamma$ with $0 < \gamma < 1/\tau$ or $a_n \asymp \log\log\log n$).*

Let $\eta \sim \pi_n$ the density of a prior defined on $\{\eta \in [-M_n, M_n]^{J_n} : \eta^\top 1 = 0\}$ such that $\underline{c}^{J_n} \leq \pi_n(\eta) \leq \overline{c}^{J_n}$, for some $0 < \underline{c} < \overline{c} < \infty$ then

$$\Pi_n(\eta : d_H(f(\cdot;\eta), f_n(\cdot)) \geq S_n n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}|X^{(n)}) \to 0 \tag{2.23}$$

in $P^n_{f_n}$-probability, for every $S_n \to \infty$.

The proof of this Theorem can be found in section 2.6.2 and is based on the proof of Theorem 9.1 in [23] adapted to the case where the parameter that here we call η has a bound that can grow with n. The rate of convergence is the minimax nonparametric rate for Hellinger distance up to a logarithmic factor. From the proof we observe that this extra factor is a consequence of allowing the bound on η, namely M_n to grow with n. In turn this is a consequence allowing the function being approximated by B-splines to have support that expands as n grows. It is not clear whether this phenomenon is unavoidable or just an artefact of the proof. More precisely, it is Lemma 9.3 in [23] that brings up this phenomenon. This Lemma compares Hellinger distance between two densities to Euclidean distance between corresponding parameters η. In particular, inequality

$$c_0 e^{-M_n}(\|\eta_1 - \eta_2\|_2 \wedge \sqrt{J_n}) \leq \sqrt{J_n} d_H(f(\cdot;\eta_1), f(\cdot;\eta_2))$$

becomes relevant when $d_H(f(\cdot;\eta_1), f(\cdot;\eta_2)) < c_0 e^{-M_n}$, and thus we require M_n not to grow too fast. Moreover, e^{M_n} becomes the factor that transforms the convergence rate in Hellinger distance to the convergence rate in Euclidean distance for η. This is the reason why we choose M_n of order $\log\log n$ so that η concentrates at nonparametric rate times a $\log n$ factor. Finally, it is worth mentioning that the term e^{-M_n} appears in the proof of Lemma 9.3 as a uniform lower bound for $f(\cdot;\eta)$ and then it could be replaced if a more suitable bound can be found.

We finish this section with Corollaries. In Theorem 4 we assumed that the sample was distributed according to f_n the truncated version of f_0. The first Corollary links the result with a sample from f_0.

Corollary 1. *Consider f_0 and the Log-Spline model as in Theorem 4 and a sample $X^{(n)} = (X_1, \dots, X_n)$ i.i.d. from f_0. Define $\underline{X}^{(n)} := (X_i : X_i \leq a_n)$ and $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$, then*

$$\Pi_n(\eta : d_H(f(\cdot;\eta), f_n(\cdot)) \geq S_n \epsilon_n |\underline{X}^{(n)}) \to 0 \tag{2.24}$$

in $P_0^{(n)}$-probability. Furthermore, if $d_H(f_0, f_n) \lesssim \bar{\epsilon}_n$ for some $\bar{\epsilon}_n \geq \epsilon_n$ then

$$\Pi_n(\eta : d_H(f(\cdot;\eta), f_0(\cdot)) \geq S_n \bar{\epsilon}_n |\underline{X}^{(n)}) \to 0 \tag{2.25}$$

in $P_0^{(n)}$-probability.

Proof. Given that $a_n \lesssim (\log\log n)^{(1/\tau)}$ and $1 - F_0(a_n) \lesssim a_n^{-(\kappa-1)}$ with $\kappa > 1$ by condition (2.1), then applying Lemma 1 we obtain that with probability at least $1 - C(a_n^{\kappa-1} - 1)/n \to 1$ for some constant $C > 0$, the number of elements in $\underline{X}^{(n)}$ is greater than $n(1 - 2Ca_n^{-(\kappa-1)}) \geq n(1 - 2\tilde{C}(\log\log n)^{-(\kappa-1)/\tau})$ for some $\tilde{C} > 0$.

Therefore, conditioned on this event, applying Theorem 4 we conclude (2.24) holds.

To prove (2.25), we use triangular inequality of d_H and the fact that equation (2.24). $\square$

The second Corollary provides the convergence rate for the Euclidean distance in η as discussed above.

Corollary 2. *Under the conditions of Corollary 1,*

$$\Pi_n(\eta : \|\eta - \eta_0\|_2 \leq S_n \sqrt{J_n} e^{M_n} \epsilon_n |\underline{X}^{(n)}) \to 0 \tag{2.26}$$

in $P_0^{(n)}$-probability, for every $S_n \to \infty$. Hence if $J_n \asymp n^{1/(2\beta+1)}$, then $\epsilon_n \asymp n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$ and $e^{M_n} \lesssim \log n$ and the contraction rate for η is $n^{-(\beta-1/2)/(2\beta+1)}(\log n)(\log\log n)^{2+\beta/\tau}$, for $\beta > 1$.

Proof. Using Lemma 9.4(i) in [23] we have $C_{J_n,M_n}(f_n, S_n\epsilon_n) \subset H(J_n, 2e^{M_n}c_o^{-1}S_n\epsilon_n)$ for $2S_n\epsilon_n < c_0 e^{-M_n}$. By Corollary 1 we conclude that for any $S_n \to \infty$

$$\Pi_n(\eta : H^C(J_n, 2e^{M_n}c_o^{-1}S_n\epsilon_n)|\underline{X}^{(n)}) \leq \Pi_n(\eta : C^C_{J_n,M_n}(f_n, S_n\epsilon_n)|\underline{X}^{(n)}) \to 0 \tag{2.27}$$

in $P_0^{(n)}$-probability. □

2.5.5 Joint Consistency and Convergence rate

Now we proceed to prove joint consistency of θ and η.

In this section we will use the sets defined in equations (2.22) and also the following sets

$$\begin{aligned} \Theta(\epsilon) =&\{\theta \in \mathbb{R} : |\theta - \theta_0| \leq \epsilon\} \\ E_{J,M}(f_n, \epsilon) =&\{\theta \in \mathbb{R}, \eta \in H_{J,M} : d_H(f_n, f_\theta(\cdot;\eta)) < \epsilon\} \\ E(f_n, \epsilon) =&\{f_\theta(\cdot;\eta) : (\theta, \eta) \in E_{J,M}(f_n, \epsilon)\} \end{aligned} \tag{2.28}$$

We start with a Lemma that is analogue to Lemma 9.3 in [23] that compares Hellinger distance with Euclidean distances. It is a generalisation that includes the variation in the new parameter θ.

Lemma 2. *Let $\theta_1, \theta_2 \in \mathbb{R}$ and $\eta_1, \eta_2 \in H_{J,M}$. Let $a > 0$ such that* $\operatorname{supp}(f) = [0, a]$ *and define $c_\eta = (\inf_{x\in[0,1]} f(x;\eta_1) \;\wedge\; \inf_{x\in[0,1]} f(x;\eta_2))$,*
$C_{\eta_i} = (\sup_{x\in[0,1]} f(x;\eta_i) + \sup_{x\in[a-1,a]} f(x;\eta_i) + \int |f'(x;\eta_i)|\,dx)$ for $i = 1, 2$ and $C_\eta = (C_{\eta_1} \;\wedge\; C_{\eta_2})$. Then

$$d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2))^2 \leq 2\left(C_\eta|\theta_1 - \theta_2| + C_0^2 e^{2M}\|\eta_1 - \eta_2\|_2^2/J\right) \tag{2.29}$$

with $C_0 > 0$ is a universal constant.

Additionally, if

$$d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2))^2 < e^{-2M} \leq \left(\inf_{x\in[0,a]} f(x;\eta_1) \;\wedge\; \inf_{x\in[0,a]} f(x;\eta_2)\right) \tag{2.30}$$

then $|\theta_1 - \theta_2| < 1$ and

$$\begin{aligned} &\frac{\sqrt{C_\eta}}{1 + 2\sqrt{C_\eta/c_\eta}}|\theta_1 - \theta_2|^{1/2} + \frac{c_0 e^{-M}}{1 + 2\sqrt{C_\eta/c_\eta}}(\|\eta_1 - \eta_2\|_2/J^{1/2} \wedge 1) \\ &\qquad\qquad\qquad\qquad \leq d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) \end{aligned} \tag{2.31}$$

where c_0 is a universal constant.

The proof of this Lemma is provided in section 2.6.3. Note that the term involving η is of the same order as in the original Lemma, and we recover it taking $\theta_1 = \theta_2$. In terms of θ we observe that due to the nonregular nature of the parameter, the variation in this case if of order Hellinger distance squared whereas the variation in η is of the same order of the Hellinger distance. Ibragimov and Hansminskii [30] in Chapter V also showed this when the density is known and [22] in equation (3.5) showed a lower bound similar to this when there is a nuisance regular parameter but with fixed dimension. For this reason they did not require to be specific with the constants; in our case we show explicit expressions for them since they might depend on n.

This is a key fact for determining convergence rate of parameters. For instance in a fully parametric model when the dimension of η is fixed, the inequality in [22] together with posterior contraction rate of $1/\sqrt{n}$ in Hellinger distance implies the classic $1/\sqrt{n}$ rate for regular parameters and $1/n$ for the nonregular paramter.

The following Lemma provides inclusions of neighbourhood for Hellinger distance around f_n and Euclidean distance around (θ_0, η_0) that are derived from the previous Lemma.

Lemma 3. *Suppose* $\text{supp}(f) = \text{supp}(f_n) = [0, a]$ *and there are constants* $M_0, D_0, K_0, K_0' > 0$ *such that*

$$\begin{aligned}
\|f_n(x)\|_\infty \leq & M_0 && (2.32)\\
\|\log f_n\|_{C^\beta} \leq & K_0 K^\beta / a && (2.33)\\
\|(\log f_n(\cdot))'\|_{C^{\beta-1}} \leq & K_0' K^{\beta-1} && (2.34)\\
\int_0^a |f_n'(x)|\, dx \leq & D_0 && (2.35)
\end{aligned}$$

Let

$$\begin{aligned}
\tilde{C}_0 = & \left(2(M_0 + \tilde{K}_0/a) + 2\tilde{C}_{q,\beta} K_0' + \tilde{K}_0 + D_0\right)^{1/2}\\
\tilde{K}_0 = & 2C_{q,\beta} K_0 e^{2C_{q,\beta} K_0/a}
\end{aligned}$$

where $C_{q,\beta}, \tilde{C}_{q,\beta}$ *are constants that only depend on* q *and* β, *the order of B-splines and smoothness of function* f_n. *Then*

- *For all* $\epsilon \geq d_H(f_{\theta_0}(\cdot; \eta_0), f_n)$,
$$E(f_n, 2\epsilon) \supset \Theta((\epsilon/(2\tilde{C}_0))^2) \times H_M(J, \epsilon/(2C_0 e^M)). \tag{2.36}$$
- *For all* $\epsilon < c_0 e^{-M}/2(1 + 2e^M \tilde{C}_0)$,
$$E_{J,M}(f_n, \epsilon) \subset \Theta\left(4\epsilon^2(\tilde{C}_0^{-1} + 2e^M)^2\right) \times H_M\left(J,\ 2c_0^{-1} e^M (1 + 2e^M \tilde{C}_0)\epsilon\right). \tag{2.37}$$
- *Additionally, if there exists* $0 < E_0 \leq aKJ^{1/2}/4$ *such that*
$$\epsilon \leq E_0 (aK)^{-1} 2J^{-1/2} c_0 e^{-M} (1 + 2e^M \tilde{C}_0)^{-1}$$
then for all $\eta \in H_M\left(J,\ 2c_0^{-1} e^M (1 + 2e^M \tilde{C}_0)\epsilon\right)$
$$\|f(\cdot; \eta)\|_\infty \leq M_0 + \tilde{K}_0/a + 2e^{2E_0/(aK)} E_0/(aK) \tag{2.38}$$
and
$$\int_0^a |f'(x; \eta)|\, dx \leq E_0/a + 2e^{2E_0/(aK)} E_0/K + 2\tilde{C}_{q,\beta} K_0' + \tilde{K}_0 + D_0 \tag{2.39}$$

The proof this Lemma is also found in section 2.6.3.

The next Lemma is analogue to Lemma 9.5 in [23] regarding the bound on the entropy number that corresponds to one of the sufficient conditions to prove consistency.

Lemma 4. *Suppose* $\mathrm{supp}(f) = \mathrm{supp}(f_n) = [0, a]$ *and there are constants* $M_0, D_0, K_0, K_0' > 0$ *as in assumptions* (2.32) - (2.35) *in Lemma 3, that is,*

$$\begin{aligned}\|f_n(x)\|_\infty \leq& M_0\\ \|\log f_n\|_{C^\beta} \leq& K_0 K^\beta / a\\ \|(\log f_n(\cdot))'\|_{C^{\beta-1}} \leq& K_0' K^{\beta-1}\\ \int_0^a |f_n'(x)|\, dx \leq& D_0\end{aligned}$$

then

$$\log N(\epsilon/5, E(f_n, \epsilon), d_H) \leq J \log(60 C_0 e^{2M}(1 + 2e^M \tilde{C}_0)/c_0) + \log(12 \cdot 10^2 (\tilde{C}_0^{-1} + 2e^M)^2 (2\tilde{M}_0 + \tilde{M}_0')) \quad (2.40)$$

where

$$\begin{aligned}\tilde{C}_0 =& \Big(2(M_0 + \tilde{K}_0/a) + 2\tilde{C}_{q,\beta} K_0' + \tilde{K}_0 + D_0\Big)^{1/2}\\ \tilde{K}_0 =& 2C_{q,\beta} K_0 e^{2C_{q,\beta} K_0 / a}\\ \tilde{M}_0 =& M_0 + \tilde{K}_0/a + 2e^{2E_0/(aK)} E_0/(aK)\\ \tilde{M}_0' =& E_0/a + 2e^{2E_0/(aK)} E_0/K + 2\tilde{C}_{q,\beta} K_0' + \tilde{K}_0 + D_0\end{aligned}$$

$E_0 > 0$ *is any constant such that* $E_0 \leq aKJ^{1/2}/4$ *and* $C_{q,\beta}, \tilde{C}_{q,\beta}$ *are constants that only depend on* q *and* β*, the order of B-splines and smoothness of function* f_n*.*

The proof is in section 2.6.2.

Before we prove joint contraction rate, we have to slightly modify Theorem 8.11 in [23] for us to use it in this model. Note that $K(f_{n,\theta_0}; f_\theta(\cdot;\eta)) < \epsilon^2$ implies $\theta = \theta_0$. Indeed, when $\theta > \theta_0$, for any $x \in [\theta_0, \theta]$, $f_\theta(x;\eta) = 0$ but $f_{n,\theta_0}(x) > 0$, then $K(f_{n,\theta_0}; f_\theta(\cdot;\eta)) = \infty$. Likewise, the same result holds when $\theta < \theta_0$ since $f_\theta(x;\eta)$ vanishes on $[\theta + a_n, \theta_0 + a_n]$, where f_{n,θ_0} is strictly positive.

Condition (i) in Theorem 8.11 [23] requires prior mass on the set $\{\theta \in \mathbb{R}, \eta \in H_{J,M} : K(f_{n,\theta_0}; f_\theta(\cdot;\eta)) < \epsilon^2, V_2(f_{n,\theta_0}; f_\theta(\cdot;\eta)) < \epsilon^2\}$ to be strictly positive which in this case implies positive prior mass assigned to a singleton, that is, $\Pi(\{\theta_0\}) > 0$. This rules out all priors on θ that are absolutely continuous with respect to the Lebesgue measure.

To solve this problem, we analyse condition (i) in Theorem 8.11 in [23] which is the only one that involves prior mass on Kullback-Leibler neighbourhoods. More specifically, this prior mass is used to bound from below what is called the *evidence*, that refers to the normalising term in the posterior distribution that does not depend on the parameters.

This is done through the results shown in Lemmas 6.26 and 8.10 of [23] which we include here with the notation of our model.

Lemma 5 (Evidence lower bound as in Lemma 6.26 of [23])**.** *For any probability measure* Π *on*

$\mathbb{R} \times H_{J,M}$*, and any constants* $D > 1$ *and* $\epsilon \geq n^{-1}$*, with* $P_0^{(n)}$*-probability at least* $1 - D^{-1}$*,*

$$\int \int \prod_{i=1}^{n} \frac{f_\theta(X_i;\eta)}{f_{n,\theta_0}(X_i)} d\Pi(\theta)d\Pi(\eta) \geq \Pi(f_\theta(\cdot;\eta) : K(f_{n,\theta_0}; f_\theta(\cdot;\eta)) < \epsilon)e^{-2Dn\epsilon} \tag{2.41}$$

Lemma 6 (Evidence lower bound as in Lemma 8.10 of [23])**.** *For every* $k \geq 2$ *there exists a constant* $d_k > 0$ *(with* $d_2 = 1$*) such that for any probability measure* Π *on* $\mathbb{R} \times H_{J,M}$ *and any positive constants* ϵ, D*, with* $P_0^{(n)}$*-probability at least* $1 - d_k(D\sqrt{n}\epsilon)^{-k}$*,*

$$\int \int \prod_{i=1}^{n} \frac{f_\theta(X_i;\eta)}{f_{n,\theta_0}(X_i)} d\Pi(\theta)d\Pi(\eta) \geq \Pi(D_k(f_{n,\theta_0}, \epsilon))e^{-(1+D)n\epsilon^2} \tag{2.42}$$

where $D_k(f_{n,\theta_0}, \epsilon) := \{f_\theta(\cdot;\eta) : K(f_{n,\theta_0}; f_\theta(\cdot;\eta)) < \epsilon^2, V_k(f_{n,\theta_0}; f_\theta(\cdot;\eta)) < \epsilon^k\}$*.*

We prove a result similar to Lemma 6 replacing $\Pi(D_k(f_{n,\theta_0}, \epsilon))$ by

$$\Pi(B_k(f_n, \epsilon/\sqrt{2}))\Pi([\theta_0 - \epsilon_n^2(2M_nK_n)^{-1}), \theta_0]).$$

where $B_k(f_n, \epsilon) := \{f(\cdot;\eta) : K(f_n; f(\cdot;\eta)) < \epsilon^2, V_k(f_n; f(\cdot;\eta)) < \epsilon^k\}$ is a Kullback-Leibler neighbourhood with respect to η only where $\theta_0 = 0$ is fixed.

Note also that in Lemmas 5 and 6 it is assumed that observations $X = (X_i,\ i = 1\ldots,n)$ are i.i.d. from density f_{n,θ_0}, however our sample comes from density f_{0,θ_0} with observations potentially greater than $\theta_0 + a_n$ where f_{n,θ_0} vanishes. In Corollary 1 we considered only the observations that lie in the support of f_{n,θ_0}, that is the interval $[\theta_0, \theta_0 + a_n]$. This way the posterior distribution does not vanish and the corresponding sample is distributed according to f_{n,θ_0}.

In that case, $f(\cdot;\eta)$ and f_n shared the same support, however after the introduction of parameter θ we also need to consider the support of $f_\theta(\cdot,\eta)$ as we did in Theorem 1. Given a sample X, and considering the restriction to $[\theta_0, \theta_0 + a_n]$, say $\underline{X}$, the likelihood $\prod_i f_\theta(\underline{X}_i;\eta)$ is strictly positive only for $\theta \in [\underline{X}_{(\underline{n})} - a_n, X_{(1)}]$, where $X_{(1)} = \min\{X_i,\ i = 1,\ldots,n\}$, $\underline{n}$ is the number of components in $\underline{X}$ and $\underline{X}_{(\underline{n})} = \max\{\underline{X}_i,\ i = 1,\ldots,n\}$. It is crucial then to measure the length of that interval, to find the desired lower bound on the evidence. We have shown in Lemma 1 that with high probability this length is of order d_n/n, with d_n a sequence that goes to infinity more slowly than $1/f_0(a_n) \lesssim \exp(a_n^\tau)$. Finally, Lemma 1 also allows us to define $\underline{X}$ not involving θ_0 which in practice it is unknown. Specifically, we define $\underline{X} = (X_i : X_i < X_{(1)} + a_n - 2d_n/n)$ which by Lemma 1 contains at least $p_n n$ elements with $p_n \to 1$. We use this to prove the desired lower bound on the evidence, shown in the following Lemma. Its proof follows the same ideas in the proof of Lemma 8.10 [23] adapted to this context.

Lemma 7. *Let* $X = (X_i,\ i = 1\ldots,n) \overset{i.i.d.}{\sim} f_{0,\theta_0}$*. Define* $\tilde{\underline{X}} = (X_i : X_i < \theta_0 + a_n)$ *and* $\underline{X} = (X_i : X_i < X_{(1)} + a_n - d_n/n)$*, with* $d_n \to \infty$ *such that* $d_n f_0(a_n) \to 0$*. Denote* $\underline{n}$ *the number of components in* $\underline{X}$*.*

There is a positive constant c_0 *depending only on* $f_0(0)$ *and for every* $k \geq 2$ *there exists a constant* $d_k > 0$ *(with* $d_2 = 1$*) such that for any probability measure* Π *on* $\mathbb{R} \times H_{J_n,M_n}$ *and any positive constants* ϵ, D*, such that* $\epsilon^2/(2(q-1)M_nK_n) \leq 2d_n/n$*, the following bound holds with*

$P_0^{(n)}$*-probability at least* $1 - d_k(D\sqrt{2n}\epsilon)^{-k} - e^{-c_0 d_n} - (8d_n f_0(a_n))^{1/2}$,

$$\begin{aligned}\int\int\prod_{i=1}^{\tilde{n}}\frac{f_\theta(\tilde{\underline{X}}_i;\eta)}{f_{n,\theta_0}(\tilde{\underline{X}}_i)}d\Pi(\theta)d\Pi(\eta) &= \int\int\prod_{i=1}^{n}\frac{f_\theta(\underline{X}_i;\eta)}{f_{n,\theta_0}(\underline{X}_i)}d\Pi(\theta)d\Pi(\eta)\\ &\geq \Pi(B_k(f_n,\epsilon/\sqrt{2}))\Pi([\theta_0-\epsilon^2(2C_qM_nK_n)^{-1},\ \theta_0])e^{-(1+D)n\epsilon^2}\end{aligned} \tag{2.43}$$

where

$$B_k(f_n,\epsilon) := \{f(\cdot;\eta) : K(f_n;f(\cdot;\eta)) < \epsilon^2, V_k(f_n;f(\cdot;\eta)) < \epsilon^k\}.$$

Next we prove nonparametric convergence rate up to a logarithmic factor of the joint posterior distribution with respect to Hellinger distance.

Theorem 5. *Consider* f_0, f_n *and the Log-Spline model as in Theorem 4 and a sample* $X^{(n)} = (X_1,\ldots,X_n)$ *i.i.d. from* f_0*. Define* $\underline{X}^{(n)} := (X_i : X_i - X_{(1)} \leq a_n - d_n/n)$ *where* $d_n \to \infty$ *such that* $d_n = o(a_n^{-\kappa})$ *with* $\kappa > 0$ *from tail condition* (2.1) *(e.g.* $d_n = \log a_n$*), and* $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$.

Let $\eta \sim \pi_{\eta,n}$ *the density of a prior defined on* $\{\eta \in [-M_n,M_n]^{J_n} : \eta^\top 1 = 0\}$ *such that* $\underline{c}^{J_n} \leq \pi_n(\eta) \leq \overline{c}^{J_n}$*, for some* $0 < \underline{c} < \overline{c} < \infty$.

Let $\theta \sim \pi_{\theta,n}$ *the density of a prior defined on* $\mathbb{R}$ *such that* $\sup_\theta \pi(\theta) \leq \overline{k}$*, and for all* $\theta_0 \in \mathbb{R}$ *there exists* $t_0 > 0$*, such that* $\inf_{[\theta_0-t_0,\theta_0]}\pi(\theta) \geq \underline{k}$*, for some* $0 < \underline{k} < \overline{k} < \infty$*. Then for* $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$,

$$\Pi_n((\theta,\eta) : d_H(f_\theta(\cdot;\eta), f_{n,\theta_0}(\cdot)) \geq S_n\epsilon_n|\underline{X}^{(n)}) \to 0 \tag{2.44}$$

in $P_0^{(n)}$*-probability. Furthermore, if* $d_H(f_0,f_n) \lesssim \bar{\epsilon}_n$ *for some* $\bar{\epsilon}_n \geq \epsilon_n$ *then*

$$\Pi_n((\theta,\eta) : d_H(f_\theta(\cdot;\eta), f_{0,\theta_0}(\cdot)) \geq S_n\bar{\epsilon}_n|\underline{X}^{(n)}) \to 0 \tag{2.45}$$

in $P_0^{(n)}$*-probability.*

The proof of this Theorem is in section 2.6.2.

We conclude this section with a Corollary showing convergence rate for parameters θ and η.

Corollary 3. *Under the conditions of Theorem 5,*

$$\Pi_n((\theta,\eta) : |\theta-\theta_0| \geq S_n^2 e^{2M_n}\epsilon_n^2,\ \|\eta-\eta_0\|_2 \geq S_n\sqrt{J_n}e^{2M_n}(\log\log n)^{1/2}\epsilon_n|\underline{X}^{(n)}) \to 0 \tag{2.46}$$

in $P_0^{(n)}$*-probability, for every* $S_n \to \infty$*. with* $\epsilon_n \asymp n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$ *and* $e^{M_n} \lesssim \log n$.

Proof. Applying Lemma 3, since

$$S_n\epsilon_n = S_n n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau} < (\log n)^{-3} < c_0 e^{-M_n}/2(1+2e^{M_n}\tilde{C}_0)$$

we obtain,

$$\begin{aligned}E_{J,M}(f_n,\epsilon_n) \subset \Theta&\left(4\epsilon_n^2(\tilde{C}_0^{-1}+2e^{M_n})^2\right)\\ &\times H_{M_n}\left(J_n,\ 2c_0^{-1}e^{M_n}(1+2e^{M_n}\tilde{C}_0)\epsilon_n\right)\\ &\subset \Theta\left(A_0\epsilon_n^2 e^{2M_n}\right)\\ &\times H_{M_n}\left(J_n,\ B_0 e^{2M_n}(\log\log n)^{1/2}\epsilon_n\right)\end{aligned}$$

for some universal constants $A_0, B_0 > 0$. $\square$

We conclude this section with a remark.

Remark 1. An important implication of this Corollary as well as Corollary 2 is that we achieve not only consistency in Hellinger distance for the densities but also consistency in uniform norm in any bounded interval of the real line. Indeed, this is a consequence of convergence of η together with uniform approximation given by equations (2.16) and (2.17).

2.5.6 Adaptation via undersmoothing

It is important to note that the results presented so far in this work are not adaptive since parameters q and J_n depend on smoothness parameter β. However, this can be easily solved at least for an interval of possible values of β if we are willing to obtain consistency at a suboptimal rate. This is shown in the following Theorem.

Theorem 6. *Let $X_1, \ldots, X_n \sim f_0$. Let $\rho = 1/(6+\delta)$ for some $\delta > 0$ arbitrarily small, and consider the Log-Spline model with $q \geq \beta_1$, for some $\beta_1 \geq 5/2+\delta$, $J_n \asymp n^\rho$, $K_n = (J_n - q + 1)/a_n$ and, a_n is such that $a_n \to \infty$ and $a_n^\tau \lesssim \log\log n$ (e.g. $a_n \asymp (\log\log n)^\gamma$ with $0 < \gamma < 1/\tau$ or $a_n \asymp \log\log\log n$).*

Let $\eta \sim \pi_{\eta,n}$ the density of a prior defined on $\{\eta \in [-M_n, M_n]^{J_n} : \eta^\top 1 = 0\}$ such that $\underline{c}^{J_n} \leq \pi_n(\eta) \leq \overline{c}^{J_n}$, for some $0 < \underline{c} < \overline{c} < \infty$.

Let $\theta \sim \pi_{\theta,n}$ the density of a prior defined on $\mathbb{R}$ such that $\sup_\theta \pi(\theta) \leq \overline{k}$, and for all $\theta_0 \in \mathbb{R}$ there exists $t_0 > 0$, such that $\inf_{[\theta_0 - t_0, \theta_0]} \pi(\theta) \geq \underline{k}$, for some $0 < \underline{k} < \overline{k} < \infty$. Then, for any $f_0 \in \mathcal{F}(5/2+\delta, \beta_1, M)$,

$$\|\Pi_n - \mathcal{E}^-_{\zeta_n}(\gamma_0)\|_{TV} \to 0 \tag{2.47}$$

in P_0-probability, where Π_n denotes the posterior distribution of h, and $\mathcal{F}(\beta_0, \beta_1, M) = \{f \in C^\beta,\ \beta \in [\beta_0, \beta_1],\ ||f||_\infty \leq M\}$.

Proof. The proof is identical to the proof of Theorem 2 except that now joint consistency in Hellinger distance will occur at rate $n^{-e}(\log\log n)^{2+\beta/\tau}$, where $e = ((1-\rho)/2 \wedge \beta\rho)$. Therefore we can choose $S_n = n^s$ and $R_n = n^r$ where $r = 1 - 2e + \delta/(6(6+\delta))$ and $s = r/2$ and then condition (H2) is satisfied. Again, assumption (i) of Theorem 3 is satisfied by definition of a_n and assumption (iv) implies (ii) and (iii). Now we consider to cases, whether $\rho \leq 1/(2\beta+1)$ or not.

If $\rho \leq 1/(2\beta+1)$ then $e = \beta\rho$ and $r = 1 - 2\beta\rho + \delta/(6(6+\delta))$, then $\rho + r = \rho(1-2\beta) + 1 + \delta/(6(6+\delta))$ and given $\rho = 1/(6+\delta)$, $\beta > 5/2+\delta$ then $\rho + r < (-4-2\delta)/(6+\delta) + 1 + \delta/(6(6+\delta)) = (6(2-\delta)+\delta)/6(6+\delta) < 1/3$.

If $\rho > 1/(2\beta+1)$ then $e = (1-\rho)/2$, therefore $r = 1 - 2e + \delta/(6(6+\delta)) = \rho + \delta/(6(6+\delta))$. Then $\rho + r = 2/(6+\delta) + \delta/(6(6+\delta)) = (12+\delta)/(6(6+\delta)) < 1/3$. $\square$

This Theorem implies that if we choose q big enough, $\delta > 0$ small and take $J_n \asymp n^{1/(6+\delta)}$ then we obtain the BvM result for all $\beta \in (5/2+\delta, q]$.

We finish this chapter mentioning that another possible way to obtain adaptive results keeping optimal concentration rate is to put a prior on J_n. Several constructions of such priors have been investigated in the literature. See for instance, Section 10.3.4 in [23], however this will be left as future work.

2.6 Proofs

2.6.1 Limiting Likelihood ratio process

Here we provide the proofs of the two main theorems regarding approximation of the localised log-likelihood ration process. We start with the general results of Theorem 1.

Proof of Theorem1. We separate the analysis in two cases; when $\underline{X}_i > \theta_0 + {}^h\!/\underline{n} > \underline{X}_i - a_n$, $\forall i = 1, \ldots, \underline{n}$, that is $X_{(1)} > \theta_0 + {}^h\!/\underline{n} > \underline{X}_{(\underline{n})} - a_n$ holds, and otherwise. Note that $\underline{X}_{(\underline{n})} - a_n < X_{(1)}$ with probability greater than $1 - \exp(-c_0 d_n) - (8 d_n f_0(a_n))^{1/2}$ by Lemma 1. In other words, we express the likelihood ratio as

$$Z_n(h,g) = Z_n(h,g)\mathbb{1}(B_n) + Z_n(h,g)\mathbb{1}\left(B_n^C\right) \tag{2.48}$$

where $B_n = \{\underline{X}_{(1)} > \theta_0 + {}^h\!/\underline{n} > \underline{X}_{(\underline{n})} - a_n\}$. It is easy to see that on the set B_n^C, for any $X_i \in [\theta_0, \theta_0 + {}^h\!/\underline{n}] \cup [\theta_0 + {}^h\!/\underline{n} + a_n, \infty)$.

$$\frac{f(X_i - (\theta_0 + {}^h\!/n); \eta_0 + {}^g\!/\sqrt{n})}{f(X_i - \theta_0; \eta_0)} = 0$$

because in the numerator X_i lies outside the support of the density which implies $Z_n(h,g) = 0$ for all $n > 0$. Therefore, equation (2.48) is simplified to

$$Z_n(h,g) = Z_n(h,g)\mathbb{1}(B_n) \tag{2.49}$$

Note that on B_n, $\tilde{\zeta}_n = n(\underline{X}_{(\underline{n})} - a_n - \theta_0) < h < \zeta_n = n\left(X_{(1)} - \theta_0\right)$. Recall that $\theta_0 + {}^h\!/n = \theta$ and $\eta_0 + {}^g\!/n = \eta$. We expand the log-likelihood ratio as the variation in θ, variation in η and the interaction term

$$\begin{aligned}
\log f_\theta(x;\eta) - \log f_{\theta_0}(x;\eta_0) = &\log f_\theta(x;\eta_0) - \log f_{\theta_0}(x;\eta_0) \\
&+ \log f_{\theta_0}(x;\eta) - \log f_{\theta_0}(x;\eta_0) \\
&+ \log f_\theta(x;\eta) - \log f_\theta(x;\eta_0) \\
&- (\log f_{\theta_0}(x;\eta) - \log f_{\theta_0}(x;\eta_0))
\end{aligned} \tag{2.50}$$

and by assumption (I) we can express the difference with respect to θ using a first order Taylor expansion and the difference with respect to η using a second order Taylor expansion. Indeed,

$$\log f_\theta(x;\eta_0) - \log f_{\theta_0}(x;\eta_0) = (\theta - \theta_0)\frac{\partial}{\partial\theta}\log f(x - \tilde{\theta};\eta_0) \tag{2.51}$$

for some $\tilde{\theta}$ between θ and θ_0, and

$$\begin{aligned}
\log f_{\theta_0}(x;\eta) - \log f_{\theta_0}(x;\eta_0) = &(\eta - \eta_0)^\top \frac{\partial}{\partial\eta}\log f_{\theta_0}(x;\eta_0) \\
&+ \frac{1}{2}(\eta - \eta_0)^\top \frac{\partial^2}{\partial\eta^2}\log f_{\theta_0}(x;\tilde{\eta})(\eta - \eta_0)
\end{aligned} \tag{2.52}$$

for some $\tilde{\eta} \in \langle \eta, \eta_0 \rangle$, the line connecting the two points. Combining equations (2.50), (2.51) and (2.52) and expressing $\theta = \theta_0 + h/n$, $\eta = \eta_0 + g/\sqrt{n}$ we obtain

$$\sup_{(\theta,\eta)\in A_n}\left|\ell\left(\theta_0+\tfrac{h}{n},\eta_0+\tfrac{g}{\sqrt{n}}\right)-\ell(\theta_0,\eta_0)-\tfrac{h}{n}\tfrac{\partial}{\partial\theta}\ell(\theta_0,\eta_0)-\tfrac{1}{\sqrt{n}}g^\top\tfrac{\partial}{\partial\eta}\ell(\theta_0,\eta_0)\right.$$
$$\left.-\tfrac{1}{2n}g^\top\tfrac{\partial^2}{\partial\eta^2}\ell(\theta_0,\eta_0)g\right|$$
$$\le\sup_{|h|\le R_n}\sup_{\tilde\theta\in(\theta,\theta_0)}\left|\tfrac{h}{n}\left(\tfrac{\partial}{\partial\theta}\ell(\tilde\theta,\eta_0)-\tfrac{\partial}{\partial\theta}\ell(\theta_0,\eta_0)\right)\right|$$
$$+\sup_{\|g\|_2\le\sqrt{J_n}S_n}\sup_{\tilde\eta\in\langle\eta,\eta_0\rangle}\tfrac{1}{n}\left|g^\top\left(\tfrac{\partial^2}{\partial\eta^2}\ell(\theta_0,\tilde\eta)-\tfrac{\partial^2}{\partial\eta^2}\ell(\theta_0,\eta_0)\right)g\right|$$
$$+\sup_{(\theta,\eta)\in A_n}|\ell(\theta,\eta)-\ell(\theta,\eta_0)-\ell(\theta_0,\eta)-\ell(\theta_0,\eta_0)| \quad (2.53)$$

which goes to 0 by assumptions (IV), (V) and (VI).

Now we study the limit of $\frac{h}{n}\frac{\partial}{\partial\theta}\ell(\theta_0,\eta_0)$.

$$\left|\frac{1}{\underline{n}}\frac{\partial}{\partial\theta}\ell(\theta_0,\eta_0)-\int[\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(x)]f_{0,\theta_0}(x)dx\right|$$
$$\le\left|\frac{1}{\underline{n}}\frac{\partial}{\partial\theta}\ell(\theta_0,\eta_0)-\frac{1}{\underline{n}}\sum_{i=1}^{\underline{n}}\tfrac{\partial}{\partial\theta}\log f_{n,\theta_0}(\underline{X}_i)\right|$$
$$+\left|\frac{1}{\underline{n}}\sum_{i=1}^{\underline{n}}\tfrac{\partial}{\partial\theta}\log f_{n,\theta_0}(\underline{X}_i)-\frac{1}{n}\sum_{i=1}^{n}\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right|$$
$$+\left|\frac{1}{n}\sum_{i=1}^{n}\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)-\int[\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(x)]f_{0,\theta_0}(x)dx\right| \quad (2.54)$$

The first term on the right-hand side goes to 0 in probability by assumption (VII) and the third term goes to 0 in probability due to assumption (II) and Law of Large Numbers. Now we analyse the second term. First note that $\frac{\partial}{\partial\theta}\log f_{n,\theta_0}=\frac{\partial}{\partial\theta}\log f_{0,\theta_0}$ on $[0,a_n]$, then

$$\left|\frac{1}{\underline{n}}\sum_{i=1}^{\underline{n}}\tfrac{\partial}{\partial\theta}\log f_{n,\theta_0}(\underline{X}_i)-\frac{1}{n}\sum_{i=1}^{n}\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right|$$
$$=\left|\frac{n-\underline{n}}{\underline{n}n}\sum_{i:X_i<a_n}\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)+\frac{1}{n}\sum_{i:X_i>a_n}\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right|$$
$$\le\frac{n-\underline{n}}{\underline{n}n}\sum_{i:X_i<a_n}\left|\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right|+\frac{1}{n}\left|\sum_{i:X_i>a_n}\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right| \quad (2.55)$$

We bound these two terms separately. Once again, by Assumption (II) and Law of Large Numbers we obtain

$$\frac{n-\underline{n}}{\underline{n}n}\sum_{i:X_i<a_n}\left|\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right|\le\frac{n-\underline{n}}{\underline{n}n}\sum_{i=1}^{n}\left|\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right|$$
$$\le\frac{n-\underline{n}}{\underline{n}}\left[\int\left|\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(x)\right|f_{0,\theta_0}(x)dx+o_P(1)\right] \quad (2.56)$$

which goes to 0 in probability since by Lemma 1, there exists a sequence p_n such that $n/\underline{n} \leq 1/p_n \to 1$ in probability. Finally we bound the term $\frac{1}{n}\left|\sum_{i:X_i>a_n} \frac{\partial}{\partial\theta} \log f_{0,\theta_0}(X_i)\right|$. Let $\epsilon > 0$ and $I_n = \{\underline{n} \geq p_n n\}$ with p_n as defined in Lemma 1, and $P_0(I_n^C) \to 0$. Note that on I_n, there exists $D > 0$ such that

$$\#\{i : X_i > a_n\} = n - \underline{n} \leq (1 - p_n)n = Dn(1 - F_n(a_n))$$

and by Markov's inequality

$$\begin{aligned} P_0\left(\frac{1}{n}\sum_{i:X_i>a_n}\left|\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_i)\right| > \epsilon \middle| I_n\right) \\ \leq \frac{Dn(1-F_n(a_n))E(|\frac{\partial}{\partial\theta}\log f_{0,\theta_0}(X_1)|\mathbb{1}([\theta_0 + a_n, \infty)))}{n\epsilon} \\ \leq \frac{D(1-F_n(a_n))\int_{\theta_0+a_n}^{\infty}|\frac{\partial}{\partial\theta} f_{0,\theta_0}(x)|dx}{\epsilon(1-F_n(a_n))} \end{aligned} \tag{2.57}$$

which goes to 0 since the integral is over the interval $[\theta_0 + a_n, \infty)$ with $a_n \to \infty$. We conclude that

$$\begin{aligned} \left|\frac{1}{\underline{n}}\frac{\partial}{\partial\theta}\ell(\theta_0,\eta_0) - \int[\tfrac{\partial}{\partial\theta}\log f_{0,\theta_0}(x)]f_{0,\theta_0}(x)dx\right| \\ = \left|\frac{1}{\underline{n}}\frac{\partial}{\partial\theta}\ell(\theta_0,\eta_0) - \int\tfrac{\partial}{\partial\theta} f_{0,\theta_0}(x)dx\right| \\ = \left|\frac{1}{\underline{n}}\frac{\partial}{\partial\theta}\ell(\theta_0,\eta_0) - \gamma_0\right| = o_P(1) \end{aligned} \tag{2.58}$$

where $\gamma_0 = \int \frac{\partial}{\partial\theta} f_{0,\theta_0}(x)dx = -\int_0^\infty f_0'(x)dx = f_0(0+)$, since $\lim_{x\to\infty} f_0(x) = 0$ by the tail assumption (2.1).

We finalise the proof combining this result with equation (2.53), denoting $\Delta_n = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\eta}\ell(\theta_0,\eta_0)$ and applying assumption (III),

$$\sup_{\|g\|_2\leq\sqrt{J_n}S_n}\left|\tfrac{1}{2n}g^\top\tfrac{\partial^2}{\partial\eta^2}\ell(\theta_0,\eta_0)g + \frac{1}{2}g^t i(\eta_0)g\right| = o_P(1)$$

with $i(\eta_0) = -\int(\frac{\partial^2}{\partial\eta^2}f(x;\eta_0))f_0(x)dx$. $\square$

Now we continue with the proof of Theorem 3.

Proof of Theorem 3. Given that f_0 is uniformly bounded from above by $\bar{M} > 0$ and has a tail that satisfies $e^{-cx^\tau} \lesssim f_0(x) \searrow 0$ for all x large, then for all $x \leq a_n$

$$\begin{aligned} |\log f_n(x)| &= |\log f_0(x) - \log F_0(a_n)| \lesssim a_n^\tau \\ |\tfrac{d^t}{dx^t}\log f_n(x)| &= |\tfrac{d^t}{dx^t}\log f_0(x)| \lesssim a_n^\tau \end{aligned} \tag{2.59}$$

for $t = 1, \ldots, \lfloor\beta\rfloor$, therefore $\|\log f_n\|_{C^\beta} \lesssim a_n^\tau$.

Let $\theta = \theta_0 + h/n$ and $\eta = \eta_0 + g/\sqrt{n}$ with $|h| \leq R_n$ and $\|g\|_2 \leq \sqrt{J_n}S_n$. We will show that the conditions in Theorem 1 hold. Note that assumption (I) holds since $\exp(\eta^\top B_q(x) - \int \eta^\top B_q(x)dx)$ is smooth in η and $q \geq 3$ for the B-splines. Condition (II) corresponds to assumption $\int |f_0'(x)|dx < \infty$. Condition (III) is satisfied since, as shown equations (2.18), $\frac{\partial^2}{\partial\eta^2}\ell(\theta_0,\eta_0) = \frac{\partial^2}{\partial\eta^2}c(\eta_0)$ that does not depend on the observations $X^{(n)}$. Now we check assumption (IV),

$$\sup_{|h|\leq R_n}\sup_{\tilde\theta\in(\theta,\theta_0)}\frac{|h|}{n}\sum_{i=1}^n\left|\frac{\partial}{\partial\theta}\log f(X_i-\tilde\theta;\eta_0)-\frac{\partial}{\partial\theta}\log f(X_i-\theta_0;\eta_0)\right| = o_P(1) \tag{2.60}$$

The left hand side is equal to

$$\sup_{|h|\leq R_n}\sup_{\tilde\theta\in(\theta,\theta_0)}\frac{|h|}{n}\sum_{i=1}^n\left|\eta_0^t(B_q'(X_i-\theta_0)-B_q'(X_i-\tilde\theta))\right| \tag{2.61}$$

and by equation (2.20)

$$\begin{aligned}B_{j,q}'(X_i-\theta_0)-B_{j,q}'(X_i-\tilde\theta) =&\\ &(q-1)\left(\frac{B_{j-1,q-1}(X_i-\theta_0)-B_{j-1,q-1}(X_i-\tilde\theta)}{t_{j-1}-t_{j-q}}\right.\\ &\qquad\left.-\frac{B_{j,q-1}(X_i-\theta_0)-B_{j,q-1}(X_i-\tilde\theta)}{t_j-t_{j+1-q}}\right)\end{aligned} \tag{2.62}$$

Additionally $t_j - t_{j+1-q}\geq 1/K_n$ for all j, and

$$B_{j,q-1}(x+y)-B_{j,q-1}(x) = yB_{j,q-1}'(\tilde x) \tag{2.63}$$

for some $\tilde x\in[x,x+y]$ and using equation (2.20) again,

$$\begin{aligned}\sum_{j=1}^{J_n}|B_{j,q-1}'(\tilde x)| =&(q-2)\sum_{j=1}^{J_n}\left|\frac{B_{j-1,q-2}(\tilde x)}{t_{j-1}-t_{j-q+1}}-\frac{B_{j,q-2}(\tilde x)}{t_j-t_{j+1-q+1}}\right|\\ \leq&2(q-2)K_n\end{aligned} \tag{2.64}$$

since $\sum_{j=1}^{J_n}|B_{j,q}(x)|=1$ for all $x\in[0,a_n]$, and $q\geq 1$, therefore

$$\sum_{j=1}^{J_n}|B_{j,q}'(X_i-\theta_0)-B_{j,q}'(X_i-\tilde\theta)|\leq C_qK_n^2|\tilde\theta-\theta_0| \tag{2.65}$$

with $C_q = 4(q-1)(q-2)$, and

$$\begin{aligned}\left|\eta_0^t(B_q'(X_i-\theta_0)-B_q'(X_i-\tilde\theta))\right| \leq&\|\eta_0\|_\infty\sum_{j=1}^{J_n}|B_q'(X_i-\theta_0)-B_q'(X_i-\tilde\theta)|\\ \lesssim&\|\log f_n\|_{C^\beta}K_n^2|\tilde\theta-\theta_0|\\ \lesssim&a_n^{(\tau-2)}n^{2\rho-1}|\tilde h|\end{aligned} \tag{2.66}$$

Hence, combining display (2.61) and equation (2.66) we obtain the bound

$$\begin{aligned} \sup_{|h|\le R_n} \sup_{\tilde\theta\in(\theta,\theta_0)} \frac{|h|}{n} \sum_{i=1}^n \left|\eta_0^t(B_q'(X_i-\theta_0)-B_q'(X_i-\tilde\theta))\right| \lesssim & \sup_{|h|\le R_n} \sup_{\tilde h\le R_n} a_n^{(\tau-2)} n^{2\rho-1}|h||\tilde h| \\ =& a_n^{(\tau-2)} n^{2\rho-1+2r} \end{aligned} \tag{2.67}$$

which goes to 0 by conditions (i) and (ii).

Now we proceed to check assumption (V) of Theorem 1,

$$\sup_{\sqrt{n}\|\eta-\eta_0\|_2\le\sqrt{J_n}S_n} \sup_{\tilde\eta\in\langle\eta,\eta_0\rangle} \sum_{i=1}^n \left|(\eta-\eta_0)^t\left(H(X_i-\theta_0;\tilde\eta)-H(X_i-\theta_0;\eta_0)\right)(\eta-\eta_0)\right| = o_P(1) \tag{2.68}$$

where $H(x;\eta) := \frac{\partial^2}{\partial\eta^2}\log f(x;\eta)$. In this case $H(x;\eta) = H(\eta) = -\frac{d^2}{d\eta^2}c(\eta)$. By the proof of Lemma 9.3 in [23] (page 235),

$$\begin{aligned} \left|t^\top(H(\tilde\eta)-H(\eta_0))t\right| =& \left|\min_\mu \int_0^\infty [(t-\mu 1)^\top B_q(x)]^2(f(x;\tilde\eta)-f(x;\eta_0))dx\right| \\ \le& \min_\mu \|(t-\mu 1)^\top B_q\|_2^2\|f(x;\tilde\eta)-f(x;\eta_0)\|_\infty \end{aligned} \tag{2.69}$$

By Lemma E.6 in the same book (adapted to our knots), $\|(t-\mu 1)^\top B_q\|_2^2 \le \|t-\mu 1\|_2^2/K_n$. Minimising with respect to μ this reduces to $\|t\|_2^2/K_n$ when $1^\top t = 0$. Thus,

$$-t^\top(H(\tilde\eta)-H(\eta))t \le K_n^{-1}\|t\|_2^2\|f(\cdot;\tilde\eta)-f(\cdot;\eta_0)\|_\infty \tag{2.70}$$

Additionally using equations (2.17) and (2.12) we can apply Lemma 8 with $f_{n,1} = f(x;\eta_0)$ and $f_{n,2} = f_n$ which is bounded from above by a multiple of $\bar M$, and since $K_n^{-\beta}\|\log f_n\|_{C^\beta} \lesssim a_n^{\beta+\tau} J_n^{-\beta} \asymp a_n^{\beta+\tau} n^{-\beta\rho}$ which goes to 0 by condition (i) and $\rho > 0$. Therefore for all $x > 0$,

$$f(x;\eta_0) \le 2\bar M \tag{2.71}$$

Similarly, using equation (2.16) for any η such that $\|\eta-\eta_0\|_2 \le \sqrt{J_n}S_n/\sqrt{n}$, we apply Lemma 8 with $f_{n,1} = f(x;\eta)$ and $f_{n,2} = f(x;\eta_0)$ which we now know is bounded from above by $2\bar M$, and since condition (iii) implies $\sqrt{J_n}S_n/\sqrt{n} = n^{1/2(\rho-1)+s} \to 0$. Thus, for all $x > 0$

$$f(x;\eta) \le 3\bar M \tag{2.72}$$

and

$$|f(x;\eta)-f(x;\eta_0)| \lesssim \|\eta-\eta_0\|_2 \tag{2.73}$$

In particular this is true for $\eta = \tilde\eta$, then

$$\|f(\cdot;\tilde\eta)-f(\cdot;\eta_0)\|_\infty \lesssim \|\tilde\eta-\eta_0\|_2 \tag{2.74}$$

Hence, using equations (2.74), (2.71) and (2.72) in equation (2.70) replacing $t = \eta - \eta_0$,

$$\begin{aligned}\sup_{\sqrt{n}\|\eta-\eta_0\|_2 \le \sqrt{J_n}S_n} \sup_{\tilde{\eta}\in\langle\eta,\eta_0\rangle} \sum_{i=1}^{n} |(\eta-\eta_0)^\top (H(\tilde{\eta}) - H(\eta))(\eta-\eta_0)| \\ \lesssim \sup_{\sqrt{n}\|\eta-\eta_0\|_2 \le \sqrt{J_n}S_n} \sup_{\tilde{\eta}\in\langle\eta,\eta_0\rangle} nK_n^{-1}\|\eta-\eta_0\|_2^2 \|\tilde{\eta}-\eta_0\|_2 \\ \lesssim K_n^{-1} J_n^{3/2} \frac{S_n^3}{n^{1/2}} \asymp a_n J_n^{1/2} \frac{S_n^3}{n^{1/2}} \asymp a_n n^{1/2(\rho-1)+3s}\end{aligned} \tag{2.75}$$

which goes to 0 due to conditions (i) and (iii).

Now we check assumption (VI)

$$\sup_{(\theta,\eta)\in A_n} \sum_{i=1}^{n} |\log f_\theta(X_i;\eta) - \log f_\theta(X_i;\eta_0) - \log f_{\theta_0}(X_i;\eta) - \log f_{\theta_0}(X_i;\eta_0)| = o_P(1) \tag{2.76}$$

The expression $\log f_\theta(X_i;\eta) - \log f_\theta(X_i;\eta_0)$ is equal to $(\eta-\eta_0)^\top B_q(X_i-\theta)$, therefore condition (2.76) simplifies to

$$\sup_{(\theta,\eta)\in A_n} \sum_{i=1}^{n} |(\eta-\eta_0)^\top (B_q(X_i-\theta) - B_q(X_i-\theta_0))| \tag{2.77}$$

and following the calculations used for the first condition, specifically equations (2.63) and (2.64)

$$\begin{aligned}\sup_{(\theta,\eta)\in A_n} \sum_{i=1}^{n} |(\eta-\eta_0)^\top (B_q(X_i-\theta) - B_q(X_i-\theta_0))| &\lesssim \sup_{(\theta,\eta)\in A_n} n\|\eta-\eta_0\|_\infty K_n|\theta-\theta_0| \\ &\lesssim n^{-1/2} J^{1/2} K_n R_n S_n \\ &\asymp n^{-1/2} J_n^{3/2} a_n^{-1} R_n S_n \\ &\asymp a_n^{-1} n^{3\rho/2-1/2+r+s}\end{aligned} \tag{2.78}$$

which goes to 0 by conditions (i) and (iv).

Finally, we check assumption (VII). Given equation (2.13)

$$\begin{aligned}|(\log f_n(x))' - (\log f(x;\eta_0))'| &\lesssim K_n^{-(\beta-1)} \|\log f_n\|_{C^\beta} \\ &\asymp a_n^{(\beta-1)+\tau} n^{-(\beta-1)\rho} \to 0\end{aligned} \tag{2.79}$$

by condition (i), $\rho > 0$ and $\beta > 1$.

□

2.6.2 Convergence rate results

We start this section with the proof of Theorem 4 that shows minimax nonparametric rate up to a logarithmic factor in Hellinger distance.

Proof of Theorem 4. We will show that conditions (i) and (ii) in Theorem 8.11 in [23] are fulfilled.

Lemma 9.5 in [23] states that

$$\log N(\epsilon/5, C_{J,M}(f_n,\epsilon), d_H) \le (2M + \log(30C_0/c_0))J \tag{2.80}$$

In our case $J = J_n \asymp n^{1/(2\beta+1)}$ and $M = M_n \asymp \|\log f_n\|_{C^\beta} \lesssim a_n^\tau \lesssim (\log\log n)$, therefore condition (ii) in Theorem 8.11 is satisfied for $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$ which implies $J_n\|\log f_n\|_{C^\beta} \lesssim n\epsilon_n^2$.

By Lemma 9.4(iii) in [23], there is a constant $D_0 > 0$ such that

$$C_{J,M}(f_n, \epsilon) \subset B_{J,M}(f_n, D_0 M\epsilon) \tag{2.81}$$

given that $\|\log f_n\|_\infty \leq M$. By Lemma 9.6(ii) and (i) in the same reference we have that

$$d_H(f(\cdot;\eta_0), f_n) \leq d_1 K_n^{-\beta}\|\log f_n\|_{C^\beta} e^{d_1 K_n^{-\beta}\|\log f_n\|_{C^\beta}} \tag{2.82}$$

$$d_0\|\eta_0\|_\infty \lesssim \|\log f_n\|_\infty + d_1 K_n^{-\beta}\|\log f_n\|_{C^\beta} \tag{2.83}$$

for some universal constants $d_0, d_1 > 0$. Thus, $d_H(f(\cdot;\eta_0), f_n) \lesssim K_n^{-\beta}\|\log f_n\|_{C^\beta}$ and $\|\eta_0\|_\infty \lesssim \|\log f_n\|_\infty$.

Additionally by Lemma 9.4(ii), $C_{J,M}(f_n, 2\epsilon) \supset \{\eta \in H_{J,M} : \|\eta - \eta_0\|_2 \leq e^{-M}C_0^{-1}\sqrt{J}\epsilon\} = H(J, e^{-M}C_0^{-1}\epsilon)$, for $\epsilon \geq d_H(f_n, f(\cdot;\eta_0))$. Combining this with equation (2.81) we obtain

$$B_{J,M}(f_n, \epsilon) \supset C_{J,M}(f_n, D_0^{-1}M^{-1}\epsilon) \supset H(J, e^{-M}C_0^{-1}D_0^{-1}M^{-1}\epsilon/2) \tag{2.84}$$

Therefore,

$$2D_0 M_n d_H(f_n, f(\cdot;\eta_0)) \lesssim K_n^{-\beta}\|\log f_n\|_{C^\beta}^2 \asymp n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$$

which implies that

$$D_0^{-1}M_n^{-1}\epsilon_n/2 \gtrsim d_H(f_n, f(\cdot;\eta_0))$$

for $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$ times a big constant if necessary. Then

$$B_{J_n,M_n}(f_n, \epsilon_n) \supset H(J_n, e^{-M_n}C_0^{-1}D_0^{-1}M_n^{-1}\epsilon_n/2) \tag{2.85}$$

By Lemma 9.4(i) if $2\epsilon < c_0 e^{-M}$ then

$$C_{J,M}(f_n, \epsilon) \subset \{\eta \in H_{J,M} : \|\eta - \eta_0\|_2 \leq 2e^M c_0^{-1}\sqrt{J}\epsilon\} = H(J, 2e^M c_0^{-1}\epsilon) \tag{2.86}$$

For n sufficiently large we obtain

$$4j\epsilon_n = 4jn^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau} < c_0/\log n \asymp c_0 e^{-\|\log f_n\|_\infty}.$$

Therefore,

$$C_{J_n,M_n}(f_n, 2j\epsilon_n) \subset H(J_n, 4e^{M_n}c_0^{-1}j\epsilon_n) \tag{2.87}$$

Hence, using equations (2.85) and (2.87)

$$\begin{aligned}
\frac{\Pi_n(f(\cdot;\eta) : d_H(f(\cdot;\eta), f_n) \leq 2j\epsilon_n)}{\Pi_n(B_2(f_n, \epsilon_n))} &\leq \frac{\Pi_n(H(J_n, 4e^{M_n}c_0^{-1}j\epsilon_n))}{\Pi_n(H(J_n, e^{-M_n}C_0^{-1}D_0^{-1}M_n^{-1}\epsilon_n/2))} \\
&\leq \frac{\sup_\eta \pi_n(\eta)(\sqrt{J_n}4e^{M_n}c_0^{-1}j\epsilon_n)^{J_n-1}\mathrm{vol}\{x \in \mathbb{R}^{J_n} : \|x\| \leq 1\}}{\inf_\eta \pi_n(\eta)(\sqrt{J_n}e^{-M_n}C_0^{-1}D_0^{-1}M_n^{-1}\epsilon_n/2)^{J_n-1}\mathrm{vol}\{x \in \mathbb{R}^{J_n} : \|x\| \leq 1\}} \\
&\leq \left(\frac{\bar{c}}{\underline{c}}\right)^{J_n}\left(\tilde{C}jM_n e^{2M_n}\right)^{J_n-1} \lesssim e^{J_n(\log(jM_n)+2M_n)}
\end{aligned} \tag{2.88}$$

which satisfies condition (i) of Theorem 8.11 since $J_n M_n \lesssim n\epsilon_n^2$. ☐

Now we include the proof of Lemma 4 on the entropy condition for proving joint consistency.

Proof of Lemma 4. Let $\epsilon < E_0(aK)^{-1}2J^{-1/2}c_0e^{-M}(1+2e^M\tilde{C}_0)^{-1}$ for some $0 < E_0 \le aKJ^{1/2}/4$ so that we also have $\epsilon < c_0e^{-M}/2(1+2e^M\tilde{C}_0)$. By Lemma 3

$$E_{J,M}(f_n,\epsilon) \subset \Theta\left(4\epsilon^2(\tilde{C}_0^{-1}+2e^M)^2\right) \times H_M\left(J,\ 2c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon\right)$$

and for all $\eta \in H_M\left(J,\ 2c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon\right)$

$$\|f(\cdot;\eta)\|_\infty \le \tilde{M}_0 \tag{2.89}$$

and

$$\int_0^a |f'(x;\eta)|\,dx \le \tilde{M}_0' \tag{2.90}$$

where $\tilde{M}_0$ and $\tilde{M}_0'$ depend on $M_0, K_0, K_0', D_0, E_0, K, a, q$ and β as stated in the Lemma.

Now let $\tilde{\theta} \in \Theta(4\epsilon^2(\tilde{C}_0^{-1}+2e^M)^2)$ and $\tilde{\eta} \in H_M(J,\ 2c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon)$. Denote $\Theta(\tilde{\theta},\tilde{\epsilon}) := \{\theta \in \mathbb{R} : |\theta-\tilde{\theta}| \le \tilde{\epsilon}\}$ the interval of length $2\tilde{\epsilon}$ around $\tilde{\theta}$, and $H(\tilde{\eta},\tilde{\epsilon}) := \{\eta \in H_{J,M} : \|\eta-\tilde{\eta}\|_2 \le \tilde{\epsilon}\}$ the ball of radius $\tilde{\epsilon}$ with respect to Euclidean distance centred at $\tilde{\eta}$.

Let $\tilde{B}(\tilde{\theta},\tilde{\eta},(\epsilon/10)^2(2\tilde{M}_0+\tilde{M}_0')^{-1},(\epsilon/10)\sqrt{J}e^{-M}(C_0)^{-1})$ be the set of elements $f_\theta(\cdot;\eta)$ where $\theta \in \Theta(\tilde{\theta},(\epsilon/10)^2(2\tilde{M}_0+\tilde{M}_0')^{-1})$ and $\eta \in H(\tilde{\eta},(\epsilon/10)\sqrt{J}e^{-M}(C_0)^{-1})$. Applying inequality (2.29) in Lemma 2 with $\theta_1=\theta$, $\theta_2=\tilde{\theta}$, $\eta_1=\eta$ and $\eta_2=\tilde{\eta}$, we obtain that $d_H(f_\theta(\cdot;\eta), f_{\tilde{\theta}}(\cdot;\tilde{\eta}))^2$ bounded by

$$2\left(C_\eta|\theta-\tilde{\theta}| + C_0^2e^{2M}\|\eta-\tilde{\eta}\|_2^2/J\right) \le (\epsilon/5)^2$$

where $C_\eta = 2\tilde{M}_0+\tilde{M}_0'$. Therefore, the following inclusion holds

$$\begin{aligned}&\tilde{B}(\tilde{\theta},\tilde{\eta},(\epsilon/10)^2(2\tilde{M}_0+\tilde{M}_0')^{-1},(\epsilon/10)\sqrt{J}e^{-M}(C_0)^{-1}) \subset \\ &\qquad\qquad \{\theta \in \mathbb{R}, \eta \in H_{J,M} : d_H(f_\theta(\cdot;\eta), f_{\tilde{\theta}}(\cdot;\tilde{\eta})) \le \epsilon/5\}.\end{aligned} \tag{2.91}$$

This implies

$$\begin{aligned}&N(\epsilon/5, E(f_n,\epsilon), d_H) \le \\ &\qquad N((\epsilon/10)^2(2\tilde{M}_0+\tilde{M}_0')^{-1}, \{|\theta-\theta_0| \le 4\epsilon^2(\tilde{C}_0^{-1}+2e^M)^2\}, |\cdot|) \\ &\qquad \times N((\epsilon/10)\sqrt{J}e^{-M}(C_0)^{-1}, \{\|\eta-\eta_0\|_2 \le 2\sqrt{J}c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon\}, \|\cdot\|_2)\end{aligned} \tag{2.92}$$

Finally, using Proposition C.2 in [23]

$$\begin{aligned}&N((\epsilon/10)^2(2\tilde{M}_0+\tilde{M}_0')^{-1}, \{|\theta-\theta_0| \le 4\epsilon^2(\tilde{C}_0^{-1}+2e^M)^2\}, |\cdot|) \\ &\qquad \le \frac{12\epsilon^2(\tilde{C}_0^{-1}+2e^M)^2}{(\epsilon/10)^2(2\tilde{M}_0+\tilde{M}_0')^{-1}} = 12\cdot 10^2(\tilde{C}_0^{-1}+2e^M)^2(2\tilde{M}_0+\tilde{M}_0')\end{aligned}$$

and

$$N((\epsilon/10)\sqrt{J}e^{-M}(C_0)^{-1}, \{\|\eta-\eta_0\|_2 \le 2\sqrt{J}c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon\}, \|\cdot\|_2)$$
$$\le \left(\frac{6\sqrt{J}c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon}{(\epsilon/10)\sqrt{J}e^{-M}(C_0)^{-1}}\right)^J = (60C_0e^{2M}(1+2e^M\tilde{C}_0)/c_0)^J$$

The entropy for $\epsilon \ge E_0(aK)^{-1}2J^{-1/2}c_0e^{-M}(1+2e^M\tilde{C}_0)^{-1}$ is bounded by the entropy just obtained. □

The following proof corresponds to Lemma 7 on lower bound for the evidence.

Proof of Lemma 7. By Lemma 1, with probability at least $1-e^{-c_0d_n}-(8d_nf_0(a_n))^{1/2}$

$$\{\theta : \textstyle\prod_i f_\theta(\underline{\tilde{X}}_i;\eta)>0\} = [\underline{\tilde{X}}_{(\tilde{n})}-a_n,\ X_{(1)}] \supset [\theta_0-2d_n/n,\, \theta_0]. \tag{2.93}$$

and $\underline{\tilde{X}} = \underline{X}$. Our first step is to bound the variation of $f_\theta(\cdot;\eta)$ with respect to θ. For any $\theta \in [\theta_0-\epsilon^2(2M_nK_n)^{-1},\ \theta_0]$ and any $x \in [\theta, \theta+a_n]$,

$$\begin{aligned}\left|\log\frac{f_{\theta_0}(x;\eta)}{f_\theta(x;\eta)}\right| =& |\eta^\top(B_q(x-\theta)-B_q(x-\theta_0))| \\ =& |\eta^\top B_q'(x-\bar{\theta})||\theta-\theta_0| \\ =& (q-1)\left|\sum_j \eta_j\left(\frac{B_{j-1,q-1}(x)}{t_{j-1}-t_{j-q}}-\frac{B_{j,q-1}(x)}{t_j-t_{j+1-q}}\right)\right||\theta-\theta_0| \\ \le& 2(q-1)M_nK_n|\theta-\theta_0|\end{aligned} \tag{2.94}$$

for some $\bar{\theta} \in [\theta, \theta_0]$. Here we have used equation (2.20), $\|\eta\|_\infty \le M_n$ and $\sum_j B_{j,q-1}(x) = 1$ for all x. Let us denote $C_q = 2(q-1)$, therefore if $\theta \in [\theta_0-\epsilon^2/(2C_qM_nK_n),\ \theta_0]$ then $|\log f_{\theta_0}(x;\eta)-\log f_\theta(x;\eta)| \le \epsilon^2/2$.

Now we continue following the ideas of the proof of Lemma 8.10 in [23]. For simplicity of notation let us denote $B := B_k(f_n, \epsilon/\sqrt{2})$. Given that $\epsilon^2/(2C_qM_nK_n) \le 2d_n/n$, the integral that we want to bound on the left-hand side of equation (2.43) is bigger than the same integral restricted to the set $B \times [\theta_0-\epsilon^2/(2C_qM_nK_n),\ \theta_0]$. Then we can multiply and divide by $\Pi(B_k(f_n,\epsilon/\sqrt{2}))\Pi([\theta_0-\epsilon^2(2C_qMK)^{-1},\ \theta_0])$ and then we can consider Π to be the restriction over $B \times [\theta_0-\epsilon^2/(2C_qM_nK_n),\ \theta_0]$ and all that is left to prove is

$$\int_B\int_{\theta_0-\epsilon^2/(2C_qMK)}^{\theta_0}\prod_{i=1}^{\tilde{n}}\frac{f_\theta(\underline{\tilde{X}}_i;\eta)}{f_{n,\theta_0}(\underline{\tilde{X}}_i)}d\Pi(\theta)d\Pi(\eta) \ge e^{-(1+D)n\epsilon^2} \tag{2.95}$$

Now by Jensen's inequality

$$\log\int_B\int_{\theta_0-\epsilon^2/(2C_qMK)}^{\theta_0}\prod_{i=1}^{\tilde{n}}\frac{f_\theta(\underline{\tilde{X}}_i;\eta)}{f_{n,\theta_0}(\underline{\tilde{X}}_i)}d\Pi(\theta)d\Pi(\eta)$$
$$\ge \int_B\int_{\theta_0-\epsilon^2/(2C_qMK)}^{\theta_0}\sum_{i=1}^{\tilde{n}}\log\frac{f_\theta(\underline{\tilde{X}}_i;\eta)}{f_{n,\theta_0}(\underline{\tilde{X}}_i)}d\Pi(\theta)d\Pi(\eta) \tag{2.96}$$

Next, using the fact that $\|\log f_{\theta_0}(\cdot;\eta) - \log f_\theta(\cdot;\eta)\|_\infty \le \epsilon^2/2$ for all $\theta \in [\theta_0 - \epsilon^2/(2C_q MK),\ \theta_0]$ and $\eta \in B$ we obtain

$$\int_B \int_{\theta_0-\epsilon^2/(2C_qMK)}^{\theta_0} \sum_{i=1}^{\underline{\tilde{n}}} \log \frac{f_\theta(\underline{\tilde{X}}_i;\eta)}{f_{n,\theta_0}(\underline{\tilde{X}}_i)} d\Pi(\theta) d\Pi(\eta)$$
$$\ge \sum_{i=1}^{\underline{\tilde{n}}} \int_B \log \frac{f_{\theta_0}(\underline{\tilde{X}}_i;\eta)}{f_{n,\theta_0}(\underline{\tilde{X}}_i)} d\Pi(\eta) - \underline{\tilde{n}}\epsilon^2/2 := Z \tag{2.97}$$

From the proof of lemma 1 we know that the joint density of variables $\underline{\tilde{X}}_i$ is given by

$$f_{\underline{\tilde{X}}_1,\dots,\underline{\tilde{X}}_{\underline{n}}|\underline{\tilde{n}}=k}(x_1,\dots,x_{\underline{n}}) = \prod_{i=1}^{k} f_{n,\theta_0}(x_i).$$

then

$$E(Z) = E(E\,(Z\,|\,\underline{\tilde{n}})) = -E(\underline{\tilde{n}}) \int_B K(f_n; f(\cdot;\eta)) d\Pi(\eta) - E(\underline{\tilde{n}})\epsilon^2/2 \ge -n\epsilon^2$$

by definition of B. Due to Marcinkiewicz-Zygmund inequality, there exists a constant $d_k > 0$ that only depends on k (and $d_2 = 1$), such that

$$E\left|\frac{Z - E(Z)}{\sqrt{n}}\right|^k \le d_k \int_B V_k(f_n; f(\cdot;\eta)) d\Pi(\eta) \le d_k \epsilon^k/2^{k/2} \tag{2.98}$$

We conclude using Markov's inequality

$$P_0^{(n)}(\{Z < -(1+D)n\epsilon^2\}) \le P_0^{(n)}(\{Z - E(Z) < -Dn\epsilon^2\}) \le \frac{E|Z-E(Z)|^k}{(Dn\epsilon^2)^k} \tag{2.99}$$

which is bounded by $d_k/(D\sqrt{2n}\epsilon)^k$.

Thus,

$$\int_B \int_{\theta_0-\epsilon^2/(2C_qMK)}^{\theta_0} \prod_{i=1}^{\underline{\tilde{n}}} \frac{f_\theta(\underline{\tilde{X}}_i;\eta)}{f_{n,\theta_0}(\underline{\tilde{X}}_i)} d\Pi(\theta) d\Pi(\eta) \ge e^Z$$

and $\underline{\tilde{X}} = \underline{X}$ with probability greater than $1 - e^{-c_0 d_n} - (8 d_n f_0(a_n))^{1/2}$ and

$$e^Z \ge e^{-(1+D)n\epsilon^2}$$

with probability greater than $1 - d_k/(D\sqrt{2n}\epsilon)^k$. □

This section concludes with the proof of Theorem 5 on joint posterior concentration rate.

Proof of Theorem 5. First we prove the contraction rate for $\underline{\tilde{X}}$. We start showing that conditions on f_n in Lemmas 3 and 4 are satisfied. For n is large enough $F_0(a_n) > 1/2$ then given that $\|f_0(x)\|_\infty \le \bar{M}$ we obtain $\|f_n(x)\|_\infty \le 2\bar{M}$. Since $a_n \lesssim (\log\log n)^{1/\tau}$, $J_n \asymp n^{1/(2\beta+1)}$, and $K_n \asymp J_n/a_n$ then

$$a_n \|\log f_n\|_{C^\beta} \lesssim a_n M_n \lesssim (\log\log n)^{1+1/\tau} \le K_0 K_n^\beta$$

for some universal constant $K_0 > 0$. Note also that $(\log f_n(x))' = (\log f_0(x))'$ for all $x \in [0, a_n]$.

By assumption (2.2) on the derivatives of $\log f_0$

$$|\tfrac{d^t}{dx^t}(\log f_n(a_n))| = |\tfrac{d^t}{dx^t}(\log f_0(a_n))| \leq a_n^\tau \lesssim \log\log n$$

for $t = 1, \ldots, \lfloor\beta\rfloor$. Thus,

$$\|(\log f_n(\cdot))'\|_{C^{\beta-1}} = \|(\log f_0(\cdot))'\|_{C^{\beta-1}} \lesssim \log\log n \leq K_0' K_n^{\beta-1}$$

for some $K_0' > 0$ and $\int |f_n'(x)|\, dx = \int f_n(x)\, |(\log f_n(x))'|\, dx \leq \log\log n$.

Conditions of Lemma 4 are satisfied with $M_0 = 2\bar{M}$, $K_0 > 0$ and K_0' as defined above, $D_0 = \log\log n$, $E_0 = 1/4$, obtaining

$$\begin{aligned}\log N(\epsilon_n/5, E(f_n, \epsilon_n), d_H) \leq J_n \log(60 C_0 e^{2M_n}(1 + 2e^{M_n}\tilde{C}_0)/c_0) \\ + \log(12 \cdot 10^2(\tilde{C}_0^{-1} + 2e^{M_n})^2(2\tilde{M}_0 + \tilde{M}_0'))\end{aligned} \tag{2.100}$$

where

$$\begin{aligned}\tilde{C}_0 =& \left(2(M_0 + \tilde{K}_0/a_n) + 2\tilde{C}_{q,\beta}K_0' + \tilde{K}_0 + D_0\right)^{1/2} \\ \lesssim& (\log\log n)^{1/2} \\ \tilde{K}_0 =& 2C_{q,\beta}K_0 e^{2C_{q,\beta}K_0/a_n} \\ \leq& T_0 \\ \tilde{M}_0 =& M_0 + \tilde{K}_0/a_n + 2e^{2E_0/(a_n K_n)}E_0/(a_n K_n) \\ \leq& T_0 \\ \tilde{M}_0' =& E_0/a_n + 2e^{2E_0/(a_n K_n)}E_0/K_n + 2\tilde{C}_{q,\beta}K_0' + \tilde{K}_0 + D_0 \\ \lesssim& (\log\log n)^{1/2}\end{aligned}$$

for some universal constant T_0. Therefore

$$\begin{aligned}\log N(\epsilon_n/5, E(f_n, \epsilon_n), d_H) \lesssim& J_n \log(e^{3M_n}(\log\log n)^{1/2}) \\ \lesssim& n^{1/(2\beta+1)}\log((\log n)^3(\log\log n)^{1/2}) \\ \lesssim& n^{1/(2\beta+1)}(\log\log n)\end{aligned} \tag{2.101}$$

thus condition (ii) of Theorem 8.11 in [23] is satisfied with $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$. Applying Lemma 3, since

$$2j\epsilon_n = 2jn^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau} < (\log n)^{-3} < c_0 e^{-M_n}/2(1 + 2e^{M_n}\tilde{C}_0)$$

we obtain,

$$\begin{aligned}E_{J,M}(f_n, 2j\epsilon_n) \subset& \Theta\left(16j^2\epsilon_n^2(\tilde{C}_0^{-1} + 2e^{M_n})^2\right) \\ &\times H_{M_n}\left(J_n,\ 4jc_0^{-1}e^{M_n}(1 + 2e^{M_n}\tilde{C}_0)\epsilon_n\right) \\ \subset& \Theta\left(A_0 j^2\epsilon_n^2 e^{2M_n}\right) \\ &\times H_{M_n}\left(J_n,\ B_0 j e^{2M_n}(\log\log n)^{1/2}\epsilon_n\right)\end{aligned}$$

for some universal constants $A_0, B_0 > 0$. Additionally, equation (2.85) from the proof of Theorem 4 states that for $\epsilon_n = n^{-\beta/(2\beta+1)}(\log\log n)^{2+\beta/\tau}$, we have $D_0^{-1}M_n^{-1}\epsilon_n/2 \gtrsim d_H(f_n, f(\cdot;\eta_0))$ and then there exists a constant $\tilde{B_0} > 0$ such that

$$B_{J_n,M_n}(f_n, \epsilon_n/\sqrt{2}) \supset H(J_n, \tilde{B}_0 e^{-M_n} M_n^{-1}\epsilon_n)$$

with $B_{J,M}(f_n, \epsilon)$ and $H(J, \epsilon)$ defined in equations (2.22). Hence,

$$\begin{aligned}
&\frac{\Pi_n(f_\theta(\cdot;\eta) : d_H(f_\theta(\cdot;\eta), f_{n,\theta_0}) \leq 2j\epsilon_n)}{\Pi_n(B_k(f_n, \epsilon_n/\sqrt{2}))\Pi_n([\theta_0 - \epsilon_n^2(2C_qM_nK_n)^{-1},\ \theta_0])} \\
&\qquad \leq \frac{\Pi_n(H_{M_n}(J_n,\ B_0 j e^{2M_n}(\log\log n)^{1/2}\epsilon_n))\Pi_n(\Theta(A_0 j^2\epsilon_n^2 e^{2M_n}))}{\Pi_n(H_{M_n}(J_n, \tilde{B}_0 e^{-M_n}M_n^{-1}\epsilon_n))\Pi_n([\theta_0 - \epsilon_n^2(2C_qM_nK_n)^{-1},\ \theta_0])} \\
&\qquad \leq \frac{\sup_\eta \pi_n(\eta)(\sqrt{J_n}B_0 j e^{2M_n}(\log\log n)^{1/2}\epsilon_n)^{J_n-1}\mathrm{vol}\{x \in \mathbb{R}^{J_n} : \|x\| \leq 1\}}{\inf_\eta \pi_n(\eta)(\sqrt{J_n}\tilde{B}_0 e^{-M_n}M_n^{-1}\epsilon_n)^{J_n-1}\mathrm{vol}\{x \in \mathbb{R}^{J_n} : \|x\| \leq 1\}} \\
&\qquad\qquad \times \frac{\Pi_n(|\theta - \theta_0| \leq A_0 j\epsilon_n^2 e^{2M_n})}{\Pi_n([\theta_0 - \epsilon_n^2(2C_qM_nK_n)^{-1},\ \theta_0]))} \\
&\qquad \leq \left(\frac{\bar{c}}{\underline{c}}\right)^{J_n}\left(\frac{B_0}{\tilde{B}_0} jM_n e^{3M_n}(\log\log n)^{1/2}\right)^{J_n-1}\frac{4A_0C_q\overline{k}}{\underline{k}} je^{2M_n}M_nK_n \\
&\qquad \lesssim \exp\{J_n(\log(jM_n^{3/2}) + 3M_n) + \log(jM_nK_n) + 2M_n\}
\end{aligned}$$

and this is bounded by $\exp\{Cn\epsilon_n^2 j^2/2\}$ for some constant $C > 0$, since $J_nM_n \leq n\epsilon_n^2$. Here we have used that there exists $t_0 > 0$ such that $\inf_{[\theta_0 - t_0, \theta_0]} \pi(\theta) \geq \underline{k}$ and $\sup_\theta \pi(\theta) \leq \overline{k}$ then

$$\Pi_n(|\theta - \theta_0| \leq A_0 j\epsilon_n^2 e^{2M_n}) \leq 2A_0 j\epsilon_n^2 e^{2M_n} \sup_\theta \pi(\theta) \leq 2A_0\overline{k} j\epsilon_n^2 e^{2M_n}$$

and

$$\begin{aligned}
\Pi_n([\theta_0 - \epsilon_n^2(2C_qM_nK_n)^{-1},\ \theta_0])) &\geq \inf_{[\theta_0 - \epsilon_n^2(2C_qM_nK_n)^{-1},\ \theta_0])} \pi(\theta)\frac{\epsilon_n^2}{2C_qM_nK_n} \\
&\geq \frac{\underline{k}\epsilon_n^2}{2C_qM_nK_n}
\end{aligned}$$

then the ratio is less than

$$\frac{4A_0C_q\overline{k}}{\underline{k}} je^{2M_n}M_nK_n$$

taking n sufficiently large such that $t_0 \geq \epsilon_n^2(2C_qM_nK_n)^{-1}$. Note also that Lemma 7 holds since $n\epsilon^2/(2(q-1)M_nK_n)$ is bounded from above and $d_n \to \infty$. Thus condition (i) of the Theorem is satisfied and this concludes the proof of contraction rate for the posterior given $\tilde{\underline{X}}$. Finally, we conclude the same for the posterior given $\underline{X}$ since both vectors are equal with probability going to 1 due to Lemma 1. $\square$

2.6.3 Auxiliary results

We start this section with the proof of Lemma 1.

Proof of Lemma 1. Note that the joint density of variables $\underline{\tilde{X}}_i$ given $\underline{\tilde{n}}$ is given by

$$f_{\underline{\tilde{X}}_1,\dots,\underline{\tilde{X}}_{\underline{n}}|\underline{\tilde{n}}=k}(x_1,\dots,x_k) = \prod_{i=1}^{k} f_{n,\theta_0}(x_i).$$

Indeed, let us denote I_j for $j \in \{1,\dots,m := \binom{n}{k}\}$ all the subsets of $\{1,\dots,n\}$ that have k elements. This represents all possible ways of obtaining a vector $\underline{\tilde{X}}$ of size $\underline{n} = k$ considering elements in I_j as indices of variables X_i that are smaller than $\theta_0 + a_n$. Denote the event $S_j := (\forall i \in I_j,\, X_i \leq \theta_0 + a_n) \wedge (\forall i \in I_j^C,\, X_i > \theta_0 + a_n)$, thus for all $x_1,\dots,x_k \leq \theta_0 + a_n$

$$\begin{aligned}
P_0\left(\underline{\tilde{X}}_1 \leq x_1 \wedge \ldots \wedge \underline{\tilde{X}}_k \leq x_k \,\middle|\, \underline{\tilde{n}} = k\right) =& P_0\left(\underline{\tilde{X}}_1 \leq x_1 \wedge \ldots \wedge \underline{\tilde{X}}_k \leq x_k \,\middle|\, \bigcup_{j=1}^{m} S_j\right)\\
=& \sum_{j=1}^{m} P_0\left(\underline{\tilde{X}}_1 \leq x_1 \wedge \ldots \wedge \underline{\tilde{X}}_k \leq x_k \,\middle|\, S_j\right) \frac{P_0(S_j)}{\sum_{j=1}^{m} P_0(S_j)}\\
=& \sum_{j=1}^{m} P_0\left(\underline{\tilde{X}}_1 \leq x_1 \wedge \ldots \wedge \underline{\tilde{X}}_k \leq x_k \,\middle|\, S_j\right) \frac{1}{m}\\
=& \frac{1}{m}\sum_{j=1}^{m} P_0\left(\forall i \in I_j,\, X_i \leq x_{l_{i,j}} \,\middle|\, \forall i \in I_j,\, X_i \leq \theta_0 + a_n\right)\\
=& \prod_{i \in I_j} P_0\left(X_i \leq x_{l_{i,j}} \,\middle|\, X_i \leq \theta_0 + a_n\right)\\
=& \prod_{i=1}^{k} F_{n,\theta_0}(x_i)
\end{aligned}$$

where the index $l_{i,j}$ takes values in $\{1,\dots,k\}$ and is defined such that $X_i = \underline{\tilde{X}}_{l_{i,j}}$ for all $i \in I_j$. Also $F_n(\cdot) = F_0(\cdot)/F_0(a_n)$ is the c.d.f. corresponding to f_n. Here we have used that variables X_i are independent, and the events S_j are disjoint and have equal probability

$$P_0(S_j) = F_0(a_n)^k (1 - F_0(a_n))^{n-k},\ \forall j = 1,\dots,m$$

We bound the probability of the event $X_{(1)} \leq \theta_0 + d_n/n$. The c.d.f of $X_{(1)} - \theta_0$ is given by $1 - (1 - F_0(\cdot))^n$, where F_0 corresponds to the c.d.f. of $X_1 - \theta_0$. Additionally, $F_0(d_n/n) \geq cf_0(0)d_n/n$ for some positive constant $0 < c \leq 1$. Denote $c_0 = cf_0(0)$, thus

$$P_0^{(n)}(X_{(1)} > \theta_0 + d_n/n) = (1 - F_0(d_n/n))^n \leq (1 - c_0 d_n/n)^n \leq e^{-c_0 d_n} \to 0 \tag{2.102}$$

since $d_n \to \infty$. Now we bound the probability of the event $\underline{\tilde{X}}_{(\underline{\tilde{n}})} = \max\{\underline{\tilde{X}}_i,\ i = 1,\dots \underline{\tilde{n}}\} < \theta_0 + a_n - 2d_n/n$. The c.d.f. of $(\underline{\tilde{X}}_{(\underline{\tilde{n}})} - \theta_0)|(\underline{\tilde{n}} = k)$ is $F_n(\cdot)^k$ then

$$P_0^{(n)}(\underline{\tilde{X}}_{(\underline{\tilde{n}})} \leq \theta_0 + a_n - 2d_n/n \mid \underline{\tilde{n}} = k) = F_n(a_n - 2d_n/n)^k \tag{2.103}$$

and

$$F_n(a_n - 2d_n/n) = 1 - \frac{F_0(a_n) - F_0(a_n - 2d_n/n)}{F_0(a_n)} = 1 - \frac{2f_0(\bar{a}_n)d_n/n}{F_0(a_n)} \tag{2.104}$$

for some $\bar{a}_n \in [a_n - 2d_n/n, a_n]$. Then for n sufficiently large, $f_0(\bar{a}_n) \geq 2f_0(a_n)$ and $F_0(a_n) > 2/3$. Therefore

$$\begin{aligned}
P_0^{(n)}(\tilde{\underline{X}}_{(\tilde{\underline{n}})} \leq \theta_0 + a_n - 2d_n/n) =& \sum_{k=0}^{n} P_0^{(n)}(\tilde{\underline{X}}_{(\tilde{\underline{n}})} \leq \theta_0 + a_n - 2d_n/n \mid \tilde{\underline{n}} = k)P(\tilde{\underline{n}} = k) \\
\geq& \sum_{k=0}^{n} (1 - 6d_n f_0(a_n)/n)^k P(\tilde{\underline{n}} = k) \\
\geq& e^{-8d_n f_0(a_n)} \to 1
\end{aligned}$$

since $d_n f_0(a_n) \to 0$ and $k \leq n$. Then the probability of the complement of this event is bounded from above by $1 - e^{-8d_n f_0(a_n)} \leq (8d_n f_0(a_n))^{1/2}$. Hence, with probability at least $1 - e^{-c_0 d_n} - (8d_n f_0(a_n))^{1/2}$

$$\tilde{\underline{X}}_{(\tilde{\underline{n}})} \leq \theta_0 + a_n - 2d_n/n \leq X_{(1)} + a_n - 2d_n/n \tag{2.105}$$

$$X_{(1)} \leq \theta_0 + d_n/n \tag{2.106}$$

which imply $\underline{X} = (X_i : X_i < X_{(1)} + a_n - d_n/n) = \tilde{\underline{X}}$.

Let $0 < p \leq 1$. Now we proceed to show that $\underline{n} > pn$ with high probability. For simplicity of notation we will denote $P_0^{(n)}$ just as P_0. We will analyse the cases $n(1 - F_0(a_n)) \to 0$ and $n(1 - F_0(a_n)) \nrightarrow 0$ separately.

Firstly consider $n(1 - F_0(a_n)) \to 0$. We can take $p = 1$ and then the statement $\underline{n} < n$ is equivalent to the maximum of the observations $X_{(n)}$ being greater than a_n, thus

$$P_0(\underline{n} = n) = P_0(X_{(n)} > a_n) \leq n(1 - F_0(a_n)) \to 0. \tag{2.107}$$

hence, $\underline{n} = n$ with probability $1 - n(1 - F_0(a_n))$.

Now consider the case $n(1-F_0(a_n)) \nrightarrow 0$. This means that either $n(1-F_0(a_n)) \to l$ for some $l > 0$ or $n(1 - F_0(a_n)) \to \infty$. The probability of interest is given by the c.d.f. of a Binomial distribution. Indeed, the probability of an observation X_i being less than a_n can be seen as a Bernoulli trial with success probability equal to $F_0(a_n)$. Therefore $\underline{n} \sim Bin(n, F_0(a_n))$ and

$$P_0(\underline{n} \leq pn) = \sum_{j=1}^{\lfloor pn \rfloor} F_0(a_n)^j (1 - F_0(a_n))^{n-j} \tag{2.108}$$

We can bound this probability using Chebyshev's inequality,

$$\begin{aligned}
P_0(\underline{n} \leq pn) \leq& P\left(|\underline{n} - \mu_n| > \sigma_n \frac{(F_0(a_n) - p)\sqrt{n}}{\sqrt{F_0(a_n)(1 - F_0(a_n))}} \right) \\
\leq& \frac{F_0(a_n)(1 - F_0(a_n))}{(F_0(a_n) - p)^2 n}
\end{aligned} \tag{2.109}$$

with $\mu_n = nF_0(a_n)$ and $\sigma_n^2 = nF_0(a_n)(1 - F_0(a_n))$. Assume first that $n(1 - F_0(a_n)) \to \infty$, then for n big enough such that $F_0(a_n) > 1/2$, replacing $p = 2F_0(a_n) - 1 < F_0(a_n)$ we obtain

$$P_0(\underline{n} \leq (2F_0(a_n) - 1)n) \leq \frac{F_0(a_n)}{(1 - F_0(a_n))n} \to 0 \tag{2.110}$$

and then $\underline{n} > (2F_0(a_n) - 1)n$ with probability at least $1 - F_0(a_n)/((1 - F_0(a_n))n)$.

Finally, in the case where $n(1 - F_0(a_n)) \to l > 0$, we choose $D > 1$ arbitrarily large, n such that $F_0(a_n) > (D-1)/D$ and $p = DF_0(a_n) - (D-1) < F_0(a_n)$. Similarly to the last case we find

$$P_0(\underline{n} \leq (DF_0(a_n) - (D-1))n) \leq \frac{F_0(a_n)}{(D-1)(1-F_0(a_n))n} \to \frac{1}{(D-1)l} \tag{2.111}$$

hence we conclude that $\underline{n} > (DF_0(a_n) - (D-1))n$ with probability at least $1 - C_0/(l(D-1))$ for some constant $C_0 > 0$ and D arbitrarily large. $\square$

We continue with a Lemma that compares the differences of densities, log of densities and ratio between them.

Lemma 8. *Assume for all $x \in [a,b]$, $0 \leq a < b$, $sup_{x\in[a,b]}|\log f_{n,1}(x) - \log f_{n,2}(x)| = e_n = o(1)$. Let $C > 1$, then there exists N_C, such that for $n \geq N_C$*

$$\left|\frac{f_{n,2}(x)}{f_{n,1}(x)} - 1\right| \lesssim e_n. \tag{2.112}$$

If additionally there exists $\bar{M} > 0$ such that $\sup_{x\in[a,b]} f_{n,2}(x) \leq \bar{M}$ then

$$sup_{x\in[a,b]} f_{n,1}(x) \leq C\bar{M} \tag{2.113}$$

Moreover,

$$|f_{n,1}(x) - f_{n,2}(x)| \lesssim e_n \tag{2.114}$$

Proof. Take N_C as the smallest n such that $sup_{x\in[a,b]}|\log f_{n,1}(x) - \log f_{n,2}(x)| \leq \log C$ then, for all $x \in [a,b]$,

$$\begin{aligned} \left|\frac{f_{n,2}(x)}{f_{n,1}(x)} - 1\right| \leq& \frac{f_{n,1}(x) \vee f_{n,2}(x)}{f_{n,1}(x)} \left|1 - e^{-|\log f_{n,1}(x) - \log f_{n,2}(x)|}\right| \\ \leq& (e^{e_n} \vee 1)e_n \leq Ce_n \end{aligned} \tag{2.115}$$

Now we prove that $f_{n,1}$ is bounded

$$f_{n,1}(x) = e^{\log f_{n,1}(x)} \leq e^{\log f_{n,2}(x) + |\log f_{n,1}(x) - \log f_{n,2}(x)|} \leq Cf_{n,2}(x) \leq C\bar{M} \tag{2.116}$$

Finally,

$$\begin{aligned} |f_{n,1}(x) - f_{n,2}(x)| =& (f_{n,1}(x) \vee f_{n,2}(x)) \left|1 - e^{-|\log f_{n,1}(x) - \log f_{n,2}(x)|}\right| \\ \leq& (f_{n,1}(x) \vee f_{n,2}(x))|\log f_{n,1}(x) - \log f_{n,2}(x)| \\ \leq& C\bar{M}e_n \end{aligned} \tag{2.117}$$

$\square$

Next, we include the proofs of Lemmas used in Theorem 5 that shows joint convergence rate.

Proof of Lemma 2. Without loss of generality consider $\theta_1 \leq \theta_2$. We start proving the upper bound in equation (2.29). Using the inequality $(a+b)^2 \leq 2(a^2+b^2)$ with $a = f_{\theta_1}^{1/2}(x;\eta_1) -$

$f_{\theta_1}^{1/2}(x;\eta_2)$ and $b = f_{\theta_1}^{1/2}(x;\eta_2) - f_{\theta_2}^{1/2}(x;\eta_2)$,

$$\begin{aligned} d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2))^2 \leq & 2(d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_1}(\cdot;\eta_2))^2 \\ & + d_H(f_{\theta_1}(\cdot;\eta_2), f_{\theta_2}(\cdot;\eta_2))^2) \end{aligned} \tag{2.118}$$

The first term on the right-hand side is equal to $d_H(f(\cdot;\eta_1), f(\cdot;\eta_2))^2$ which is bounded by $C_0^2 e^{2M}\|\eta_1 - \eta_2\|_2^2/J$ due to Lemma 9.3 in [23]. Now we bound the second term, for $i = 1, 2$

$$\begin{aligned} & d_H(f_{\theta_1}(\cdot;\eta_i), f_{\theta_2}(\cdot;\eta_i))^2 \\ & \qquad \leq \|f_{\theta_1}(\cdot;\eta_i) - f_{\theta_2}(\cdot;\eta_i)\|_1 = \int |f_{\theta_1}(x;\eta_i) - f_{\theta_2}(x;\eta_i)| dx \\ & = \int_{\theta_1}^{\theta_2} f_{\theta_1}(\cdot;\eta_i) + \int_{\theta_1+a}^{\theta_2+a} f_{\theta_2}(\cdot;\eta_i) + |\theta_1 - \theta_2| \int_{\theta_2}^{\theta_1+a} \left| \frac{\partial}{\partial\theta} f_t(x;\eta_i) \right| dx \\ & \leq |\theta_1 - \theta_2| \left(\sup_{x\in[0,|\theta_1-\theta_2|]} f(x;\eta_i) + \sup_{x\in[a-|\theta_1-\theta_2|,a]} f(x;\eta_i) + \int_0^a |f'(x;\eta_i)|\, dx \right) \\ & \leq C_{\eta_i}|\theta_1 - \theta_2| \end{aligned} \tag{2.119}$$

for some t between θ_1 and θ_2. Hence we have the upper bound as in (2.29) with C_{η_1} instead of C_η. Following the same steps with $a = f_{\theta_1}^{1/2}(x;\eta_1) - f_{\theta_2}^{1/2}(x;\eta_1)$ and $b = f_{\theta_2}^{1/2}(x;\eta_1) - f_{\theta_2}^{1/2}(x;\eta_2)$ we also obtain the upper bound with C_{η_2} and conclude.

Now we prove the lower bound (2.31). First we note that

$$\begin{aligned} d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2))^2 & \geq \int_{\theta_1}^{\theta_2} f_{\theta_1}(x;\eta_1)dx \\ & \geq \left(\inf_{x\in[0,|\theta_1-\theta_2|]} f(x;\eta_1)|\theta_1 - \theta_2| \wedge 1 \right) \end{aligned} \tag{2.120}$$

thus assumption (2.30) implies $\inf_{x\in[0,|\theta_1-\theta_2|]} f(x;\eta_1)|\theta_1 - \theta_2| < e^{-2M}$ and then $|\theta_1 - \theta_2| < 1$. Equation (2.120) becomes

$$d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) \geq \left(\inf_{x\in[0,1]} f(x;\eta_1)|\theta_1 - \theta_2| \right)^{1/2} \geq c_\eta^{1/2}|\theta_1 - \theta_2|^{1/2} \tag{2.121}$$

By Lemma 9.3 in [23], for $i = 1, 2$,

$$d_H(f_{\theta_i}(\cdot;\eta_1), f_{\theta_i}(\cdot;\eta_2))^2 = d_H(f(\cdot;\eta_1), f(\cdot;\eta_2))^2 \geq c_0^2 e^{-2M}(\|\eta_1 - \eta_2\|_2^2/J \wedge 1) \tag{2.122}$$

with $c_0 > 0$ a universal constant. Now using triangular inequality and equations (2.119), (2.122)

$$\begin{aligned} d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) & \geq d_H(f_{\theta_2}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) - d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_1)) \\ & \geq c_0 e^{-M}(\|\eta_1 - \eta_2\|_2/J^{1/2} \wedge 1) - C_{\eta_1}^{1/2}|\theta_1 - \theta_2|^{1/2} \end{aligned} \tag{2.123}$$

and

$$\begin{aligned} d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) & \geq d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_1}(\cdot;\eta_2)) - d_H(f_{\theta_1}(\cdot;\eta_2), f_{\theta_2}(\cdot;\eta_2)) \\ & \geq c_0 e^{-M}(\|\eta_1 - \eta_2\|_2/J^{1/2} \wedge 1) - C_{\eta_2}^{1/2}|\theta_1 - \theta_2|^{1/2} \end{aligned} \tag{2.124}$$

therefore

$$d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) \geq c_0 e^{-M}(\|\eta_1-\eta_2\|_2/J^{1/2}\wedge 1) - C_\eta^{1/2}|\theta_1-\theta_2|^{1/2} \tag{2.125}$$

Now multiplying both sides of equation (2.121) by $2\sqrt{C_\eta/c_\eta}$ and combining it with equation (2.125) we obtain

$$\begin{aligned}\left(1+2\sqrt{C_\eta/c_\eta}\right) d_H(f_{\theta_1}(\cdot;\eta_1), f_{\theta_2}(\cdot;\eta_2)) \geq \sqrt{C_\eta}|\theta_1-\theta_2|^{1/2} \\ + c_0 e^{-M}(\|\eta_1-\eta_2\|_2/J^{1/2}\wedge 1)\end{aligned} \tag{2.126}$$

□

The following proof corresponds to Lemma 3.

Proof of Lemma 3. Note that from equation (2.17),

$$\|\log f_n - \log f(\cdot;\eta_0)\|_\infty \leq 2C_{q,\beta}K^{-\beta}\|\log f_n\|_{C^\beta}$$

and using the proof of Lemma 8 and assumption (2.33) we have

$$\begin{aligned}\|f_n - f(\cdot;\eta_0)\|_\infty \leq & 2C_{q,\beta}K^{-\beta}\|\log f_n\|_{C^\beta} e^{2C_{q,\beta}K^{-\beta}\|\log f_n\|_{C^\beta}} \\ \leq & 2C_{q,\beta}K_0 e^{2C_{q,\beta}K_0/a}/a \\ = & \tilde{K}_0/a\end{aligned} \tag{2.127}$$

Moreover, $\|(\log f(\cdot;\eta_0))' - (\log f_n(\cdot))'\|_\infty \leq 2\tilde{C}_{q,\beta}K^{-(\beta-1)}\|(\log f_n(\cdot))'\|_{C^{\beta-1}}$, then

$$f(x;\eta_0) \leq f_n(x) + \tilde{K}_0/a \tag{2.128}$$

$$(\log f(x;\eta_0))' \leq (\log f_n(x)' + 2\tilde{C}_{q,\beta}K^{-(\beta-1)}\|(\log f_n(\cdot))'\|_\infty \tag{2.129}$$

Therefore

$$\|f(\cdot;\eta_0)\|_\infty \leq M_0 + \tilde{K}_0/a \tag{2.130}$$

and

$$\begin{aligned}\int_0^a |f'(x;\eta_0)|\,dx = & \int_0^a f(x;\eta_0)\,|(\log f(x;\eta_0))'|\,dx \\ \leq & 2\tilde{C}_{q,\beta}K_0' + \int_0^a f(x;\eta_0)\,|(\log f_n(x))'|\,dx \\ \leq & 2\tilde{C}_{q,\beta}K_0' + \int_0^a \tilde{K}_0/a + \int_0^a |f_n'(x)|\,dx \\ \leq & 2\tilde{C}_{q,\beta}K_0' + \tilde{K}_0 + D_0\end{aligned} \tag{2.131}$$

First we prove inclusion (2.36). Let $\theta \in \Theta((\epsilon/(2\tilde{C}_0))^2)$ and $\eta \in H_M(J, \epsilon/(2C_0e^M))$ Using upper bound (2.29) in Lemma 2 with $\theta_1 = \theta$, $\theta_2 = \theta_0$, $\eta_1 = \eta$ and $\eta_2 = \eta_0$ we take $C_\eta = 2(M_0 + \tilde{K}_0/a) + 2\tilde{C}_{q,\beta}K_0' + \tilde{K}_0 + D_0 = \tilde{C}_0^2$ and obtain

$$d_H(f_\theta(\cdot;\eta), f_{\theta_0}(\cdot;\eta_0))^2 \leq 2\left(\tilde{C}_0^2|\theta-\theta_0| + C_0^2e^{2M}\|\eta-\eta_0\|_2^2/J\right) \leq \epsilon^2 \tag{2.132}$$

hence equation (2.36) holds due to $d_H(f_{\theta_0}(\cdot;\eta_0), f_n) \le \epsilon$ and triangular inequality.

Now we prove inclusion (2.37). Let $(\theta,\eta) \in E_{J,M}(f_n,\epsilon)$. Applying inequality (2.31) of Lemma 2 with $\theta_1 = \theta$, $\theta_2 = \theta_0$, $\eta_1 = \eta$ and $\eta_2 = \eta_0$, we can take $c_\eta = e^{-2M}$ and $C_\eta = \tilde{C}_0^2$.

If $d_H(f_\theta(\cdot;\eta), f_n(\cdot)) \le \epsilon$ then $d_H(f_\theta(\cdot;\eta), f_{\theta_0}(\cdot;\eta_0)) \le 2\epsilon$ since we can assume that $d_H(f_{\theta_0}(\cdot;\eta_0), f_n(\cdot)) \le \epsilon$, otherwise the set $E_{J,M}(f_n,\epsilon)$ is empty.

Condition $2\epsilon < c_0 e^{-M}/(1+2e^M\tilde{C}_0) < e^{-M}$ together with Lemma 2 imply

$$|\theta_1-\theta_2|^{1/2} \le 2\epsilon\left(\frac{1}{\sqrt{C_\eta}}+\frac{2}{\sqrt{c_\eta}}\right) = 2\left(\tilde{C}_0^{-1}+2e^M\right)\epsilon \tag{2.133}$$

and

$$\begin{aligned}\|\eta_1-\eta_2\|_2 \le& J^{1/2}2c_0^{-1}e^M\left(1+2\sqrt{C_\eta/c_\eta}\right)\epsilon\\ =& J^{1/2}2c_0^{-1}e^M\left(1+2e^M\tilde{C}_0\right)\epsilon\end{aligned} \tag{2.134}$$

which proves (2.37).

Now let $\eta \in H_M\left(J,\ 2c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon\right)$. Assuming $J^{1/2}2c_0^{-1}e^M(1+2e^M\tilde{C}_0)\epsilon \le E_0/(aK)$ and following similar steps as in the beginning of the proof

$$\|\log f(\cdot;\eta) - \log f(\cdot;\eta_0)\|_\infty \le 2\|\eta-\eta_0\|_\infty$$

and

$$\begin{aligned}\|f(\cdot;\eta)-f(\cdot;\eta_0)\|_\infty \le& 2\|\eta-\eta_0\|_\infty e^{2\|\eta-\eta_0\|_\infty}\\ \le& 2\|\eta-\eta_0\|_2 e^{2\|\eta-\eta_0\|_2}\\ =& 2e^{2E_0/(aK)}E_0/(aK)\end{aligned} \tag{2.135}$$

Moreover,

$$\begin{aligned}\|(\log f(\cdot;\eta))' - (\log f(\cdot;\eta_0))'\|_\infty =&\\ \le& K\|\eta-\eta_0\|_\infty\\ \le& E_0/a\end{aligned} \tag{2.136}$$

then

$$f(x;\eta) \le f(x;\eta_0) + 2e^{2E_0/(aK)}E_0/(aK) \tag{2.137}$$

$$(\log f(x;\eta))' \le (\log f_n(x;\eta_0))' + E_0/a \tag{2.138}$$

Hence

$$\|f(\cdot;\eta)\|_\infty \le M_0 + \tilde{K}_0/a + 2e^{2E_0/(aK)}E_0/(aK) \tag{2.139}$$

and

$$
\begin{aligned}
\int_0^a |f'(x;\eta)|\,dx =& \int_0^a f(x;\eta)\,|(\log f(x;\eta))'|\,dx \\
\leq& E_0/a + \int_0^a f(x;\eta)\,|(\log f(x;\eta))'|\,dx \\
\leq& E_0/a + \int_0^a 2e^{2E_0/(aK)}E_0/(aK) + \int_0^a |f'(x;\eta_0)|\,dx \\
\leq& E_0/a + 2e^{2E_0/(aK)}E_0/K + 2\tilde{C}_{q,\beta}K_0' + \tilde{K}_0 + D_0
\end{aligned}
\tag{2.140}
$$

□

Chapter 3

Semiparametric BvM Theorem for a Mixture Prior

In this chapter we show our second approach which corresponds to using a mixture prior model for the density. Our general result states that if we have an i.i.d. sample from a distribution with density $f_\theta = f(\cdot - \theta)$ where f and θ do not depend on each other, and f is supported on $(0, \infty)$, with a discontinuity at 0 but continuous from the right and monotonic non-increasing, then the nonregular version of BvM for θ holds under L^1-consistency for f, local uniform consistency for f around 0 and the prior for θ being proper and having polynomial tails. We also require that the interaction term between f and θ in the likelihood approximation goes to 0 as n goes to infinity. This condition is potentially the most problematic to prove for a particular model. Nonetheless, we showed that the conditions for the BvM Theorem to hold are satisfied by a mixture prior where the kernel is a convolution of gamma and uniform densities (so it only generates decreasing densities) and the mixing distribution follows a non-homogeneous completely random measure around 0 – a Dirichlet Process with modified jumps depending on the locations – which helps us to prove pointwise consistency, and a Dirichlet Process elsewhere.

Additionally, we have developed and implemented the algorithms that help us to illustrate how our models work in practice. Indeed, we implemented in programming language R a slice sampler for f with our mixture prior. We derived two versions of it, in the first we used the stick-breaking representation of the DP and it is simpler but we needed to introduce an approximation in the conditional posterior of one of the variables. In the second we expressed one of the DP as a normalised Gamma Process, it is slightly more complex, but uses only exact quantities. We have applied our models to simulated data obtaining the limiting type of distribution we expected from our theoretical results. Furthermore we have run our algorithm with real data from Procurement Auctions. It is often assumed for this kind of data that the bids follow a distribution from a parametric family such as Pareto. Our results allow us to remove this assumption and work with a wider set of densities.

This is work in collaboration with my supervisor Dr. Natalia Bochkina (University of Edinburgh), Prof. Judith Rousseau (University of Oxford & Université Paris Dauphine) and Dr. J.B. Salomond (Université Paris-Est Créteil). This chapter mostly contains text that has been written collaboratively by the authors in the draft of an article that is being prepared to submit for publication. Likewise, all the work presented in this chapter has been a product of the collaboration led by Prof. Judith Rousseau and my supervisor Dr. Natalia Bochkina. To be more precise with my contributions, I participated in most of the discussion including the process of definition of the prior model that is being used. I also contributed with some results involved in

$L1$ consistency and bound on Interaction term including Lemmas 17 and 29, and participated in Lemmas 15 and 16. Finally, I developed the implementation, from the derivation of the algorithm to coding it in R and running it with data.

Now we start this chapter introducing the model and some notation.

3.1 Setup and Notation

We start this section defining our model and some notation. Consider $\mathbf{X}^n = (X_1, \cdots, X_n)$, where

$$X_i \mid \theta, f \overset{i.i.d.}{\sim} f_\theta = f(\cdot - \theta), \quad i = 1, \dots, n, \tag{3.1}$$

where $\theta \in \Theta \subset \mathbb{R}$ and the probability density $f(\cdot)$ has support on $[0, +\infty)$ and is positive at 0. In literature this is called a location (or shift) LAE model. Definition of a LAE model can be found in section 1.3. Additionally we will consider that f belongs to $\mathcal{F}$, the set of monotone non increasing probability densities on $\mathbb{R}_+ = [0, +\infty)$ however we also provide a high level theorem which is valid for all LAE models, see Definition 1.1.

We denote by P_0 the true distribution of the observations, and write $E_0(f(X))$ for the expectation of $f(X))$ under E_0. For the sake of conciseness and if there is no ambiguity, use E_0 both for expectation under P_0 and under $P_0^{\otimes n}$.

Let Π be the prior distribution on (θ, f), we assume implicitly a sigma-field associated to $\mathbb{R} \times \mathcal{F}$. We consider that $\Pi(d\theta, df) = \Pi_\theta(d\theta)\Pi_f(df)$ and we denote by $\Pi(\cdot|\mathbf{X}^n)$ the posterior distribution, either marginal or not, as defined by (3.2) and (3.3).

$$\Pi(A|\mathbf{X}^n) = \frac{\int_A \prod_{i=1}^n f_\theta(X_i) d\Pi(\theta, f)}{\int_{\mathbb{R}\times\mathcal{F}} \prod_{i=1}^n f_\theta(X_i) d\Pi(\theta, f)}, \quad A \subset \mathbb{R} \times \mathcal{F}, \tag{3.2}$$

$$\Pi(B|\mathbf{X}^n) = \frac{\int_{\mathcal{F}} \mathbf{1}_{\theta\in B} \prod_{i=1}^n f_\theta(X_i) d\Pi(\theta, f)}{\int_{\mathbb{R}\times\mathcal{F}} \prod_{i=1}^n f_\theta(X_i) d\Pi(\theta, f)}, \quad B \subset \mathbb{R} \tag{3.3}$$

We study the frequentist properties of the posterior distribution, hence throughout the paper we denote by P_0 the true distribution of the observations, and write $E_0(f(X))$ for the expectation of $f(X))$ under E_0. For the sake of conciseness and if there is no ambiguity, use E_0 both for expectation under P_0 and under $P_0^{\otimes n}$.

We will also use the following notation: η represents in general an infinite-dimensional parameter and we denote by $\ell_n(\eta, \theta)$ the log -likelihood at (η, θ). For any function f, define $\mathbb{P}_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_i)$, $\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - \mathbb{P}_0(f))$. The Total variation distance between two probability measures P_1 and P_2 is denoted by $\|P_1 - P_2\|_{TV}$. Recall that the total variation distance is equivalent to the L_1 norm when the distributions are absolutely continuous with respect to a fixed measure, which we denote by $\|f_1 - f_2\|_1$ for integrable functions f_1, f_2. We denote by $h(f_1, f_2)$ the Hellinger distance between densities f_1, f_2, by $\mathcal{KL}(f_0, f_1) = \int f_0 \log(f_0/f_1)(x)dx$ the Kullback-Leibler divergence and we set $\mathcal{V}(f_0, f_1) = \int f_0(x)(\log(f_0/f_1)(x))^2 dx$.

Finally, we use the following notation for the exponential distribution with parameter $\gamma > 0$: $\mathcal{E}(\gamma)$, i.e. this represents the distribution whose density is given by $\gamma e^{-\gamma x}$, $x \geq 0$.

3.2 General Results

Our main focus is on shift location models (3.1), and particularly using a mixture prior for the density, nonetheless we first provide a high level theorem for general LAE models to highlight where the non standard limit of the posterior distribution comes from.

3.2.1 General LAE model

We consider a Bayesian model with independent priors for θ, Π_θ, and for η, Π_η defined on some functional class H, and we denote by $\Pi(\cdot|\mathbf{X}^n)$ the posterior distribution.

Note that in the case of shift LAE models defined by (3.1) , $\eta = f$, $\mathbf{X}^n$ is a n-sample of random variables independently distributed from $f_{\theta,\eta}(x) = f(x-\theta)$ and if $\eta_0 = f_0$, $\gamma_0 = f_0(0)$, see Chapter V in [30]. Note also that scale LAE models defined by $f_\tau(x) = \tau^{-1} f(x/\tau)$ with $x, \tau > 0$ can be viewed as equivalent to shift LAE models using the transformation $Z_i = \log(X_i)$, $\theta = \log \tau$, see Section 3.2.3 for details.

Define Δ_n as

$$\Delta_n(\eta, \theta) = \ell_n(\eta, \theta) - \ell_n(\eta_0, \theta) - [\ell_n(\eta, \theta_0) - \ell_n(\eta_0, \theta_0)], \tag{3.4}$$

the following proposition gives sufficient conditions for a LAE Bernstein-von Mises Theorem, in total variation, to hold.

Proposition 3. *Consider the model where data $\mathbf{X}^n$ has distribution $P^n_{\theta,\eta}$ with density $f^n_{\theta,\eta}$, $\theta \in \Theta \subset \mathbb{R}$, $\eta \in H$, with a prior $\Pi(d\theta, d\eta) = \pi_\theta(\theta) d\theta \Pi_\eta(d\eta)$ and assume that the true generating process is $P^n_0 = P^n_{\theta_0,\eta_0}$.*

(i) Assume that there exist $\epsilon_n, \varepsilon_n = o(1)$ and $A_n \subset \{d(\eta, \eta_0) \le \epsilon_n,\ |\theta - \theta_0| \le \varepsilon_n\}$, such that

$$E_{P_0} \Pi_n(A_n^c \mid \mathbf{X}^n) \to 0, \quad \textit{as} \quad n \to \infty,$$

where d is some metric on H.

(ii) Assume that the model satisfies the LAE condition (1.1) at θ_0, η_0 with centering ζ_n.

(iii) Assume that the prior density $\pi_\theta(\theta)$ is continuous and positive at θ_0.

(iv) Assume that

$$\sup_{(\theta,\eta) \in A_n} \frac{|\Delta_n(\eta, \theta)|}{1 + n|\theta - \theta_0|} = o_p(1). \tag{3.5}$$

Then

$$||\Pi_n - \mathcal{E}_{\gamma_0}||_{TV} = o(1), \tag{3.6}$$

where Π_n is the posterior distribution of $n(\zeta_n - \theta)$.

An important application of the result (3.6) is that posterior credible sets for θ will typically be also confidence sets asymptotically. Indeed consider a one sided credible interval: $(\theta_1(\mathbf{X}^n), \zeta_n)$ with $\theta_1(\mathbf{X}^n)$ defined by

$$\Pi(\theta \le \theta_1(\mathbf{X}^n) | \mathbf{X}^n) = \alpha, \tag{3.7}$$

then (3.6) implies that

$$\theta_1(\mathbf{X}^n) = \zeta_n - \frac{\log(1/\alpha)}{n\gamma_0}(1 + o(1)),$$

so that

$$P_0\left(\theta \in (\theta_1(\mathbf{X}^n), \zeta_n)\right) = 1 - \alpha + o(1)$$

due to the LAE condition that distribution of $n(\zeta_n - \theta_0)$ converges to $\mathcal{E}(\gamma_0)$. Note that (3.6) can be also made uniform over some sets so that the frequentist coverage is also uniformly approximately equal to $1-\alpha$.

In Proposition 3, conditions (i)-(iii) are standard conditions. Condition (i) is merely a posterior concentration rate condition, at the, typically, non parametric rate. There is now a large number of tools to study such conditions in the literature, thanks to the seminal work of [21]. Here the main difference from the usual results is that the posterior concentration is required separately on θ and η. We will see in Section 3.2.2 that in the case of shift and scale LAE models, an ϵ_n posterior concentration rate on $d(\eta,\eta_0)+|\theta-\theta_0|$ is easily deduced from an ϵ_n posterior concentration rate on $h(f_{\eta_0,\theta_0}, f_{\eta,\theta})$. Condition (ii) corresponds to the parametric LAE condition and that has been well studied for a number of models, in particular for the shift and scale LAE models, discussed below. The continuity condition on the prior π_θ at θ_0 is also very mild since θ is univariate. The most demanding condition is (iv), since it requires a uniform control on Δ_n. In the case of shift and scale LAE models, we propose a set of sufficient conditions for (iv) to hold and we verify these conditions for the non trivial prior models of nonparametric mixture priors on f.

In [32] the authors propose another set of sufficient conditions, based on a different type of proof, for a weaker version of (3.6) to hold. Their set of conditions is neither stronger nor weaker than ours, but does not hold for instance for the family of priors considered in Section 3.3.1. The main differences between their conditions and ours are the following. Instead of directly assuming the posterior concentration rate in (θ,η), the authors consider a stronger version of the usual Kullback-Leibler neighbourhood, together with a bounded entropy condition on H. The Kullback-Leibler condition, together with the bound on the entropy are known to be sufficient conditions to derive posterior concentration rates on $f_{\theta,\eta}$, see [21]. In addition the authors assume a posterior concentration rate of order $1/n$ for the parameter θ, which is a non trivial assumption to verify. However they do not assume a uniform bound on Δ_n, but instead require that, uniformly over η in the ϵ_n Hellinger ball around η_0 ($h(f_{\eta_0,\theta_0}, f_{\eta,\theta_0}) \le \epsilon_n$), $\int f_{\eta,\theta_0}(\mathbf{X}^n) e^{\ell_n(\eta,\theta_0+h_n/n)-\ell_n(\eta,\theta_0)} d\mathbf{X}^n < +\infty$ for any bounded random variables h_n. Finally the authors also require that uniformly over η such that $h(f_{\eta_0,\theta_0}, f_{\eta,\theta_0}) > \epsilon_n$, $h(f_{\eta,\theta_0+h/n}, f_{\eta,\theta_0}) = o(h(f_{\eta,\theta_0}, f_{\eta_0,\theta_0}))$. In the case of nonparametric mixture prior models, the latter is difficult to obtain since the set $h(f_{\eta_0,\theta_0}, f_{\eta,\theta_0}) > \epsilon_n$ is typically complex and often contains non regular functions. In order to study such complex prior models, we consider the case of shift LAE models in the following section.

To put this result in perspective with corresponding results for regular models, the conditions in Proposition 3 share many similarities with the ones in Theorem 1 of [8], which provides a Semiparametric Bernstein-Von Mises Theorem for regular models in the case of no information loss. Indeed, in general terms the result is obtained under assumptions of posterior concentration, positivity and continuity of the prior for θ, LAN expansion of the log-likelihood and uniform control of the remainder term. Our conditions are analogous to these, however there are some differences that are worth mentioning. The concentration condition (C) in [8] requires concentration of the joint posterior, and the posterior for f with known $\theta=\theta_0$, whereas we only require the first one and then we derive concentration for the marginals as discussed above. Regarding likelihood approximation, condition (N) in [8] assumes LAN expansion in θ and f, whereas we only need LAE expansion in θ for f_0 fixed. Additionally, the remainder terms are similar but not exactly the same. Note that using our notation for the log-likelihood, the numerator of the term in [8] can be expressed as

$$\ell_n(\eta,\theta) - \ell_n(\eta,\theta_0) + nI_{\eta_0}(\theta-\theta_0)^2/2 - \sqrt{n}(\theta-\theta_0)W_n(1,0).$$

Here we see the connection with our term $\Delta_n(\eta,\theta)$, since this expression is equal to

$$\Delta_n(\eta,\theta) + \ell_n(\eta_0,\theta) - \ell_n(\eta_0,\theta_0) + n\|\theta-\theta_0, 0\|_L^2/2 - \sqrt{n}W_n(\theta-\theta_0, 0).$$

Note that the log-likelihood difference here is for η_0 instead of η, and therefore bounding this term implies bounding the term $\Delta_n(\eta,\theta)$ and showing that the LAN expansion holds for the parametric model with fixed η_0. In Proposition 3 we state those conditions separately, evidently replacing LAN expansion by LAE expansion, and we only require a uniform bound on $\Delta_n(\eta,\theta)$. Finally, the denominators are also different because of the different normalisation rates; in the regular case it corresponds to $n(\theta-\theta_0)^2$ and in our case we have $n|\theta-\theta_0|$.

In the same reference, there is an example where Theorem 1 is applied to the estimation of a translation parameter in Gaussian white noise. The Theorem conditions are satisfied using a Gaussian process prior under some smoothness assumption on the nonparametric part of the model. Indeed, it is required that the unknoun funciton f belongs to a Hölder space with parameter $\beta > 3/2$. This coincides with the minimum smoothness required for the density in the BvM Theorem for a Shift LAE model under a mixture prior we show in Theorem 10. In fact, in both cases this smoothness condition is needed to prove the uniform bound on the likelihood approximation discussed above, that is condition (N) in [8] and condition (iv) in Proposition 3, however, it remains to be investigated if in our case this condition is necessary or if it can be relaxed. In fact, $\beta > 3/2$ is only needed to prove the last inequality in Theorem 9.

3.2.2 Shift LAE model

In this section we apply Proposition 3 to the case of the shift LAE model defined by (3.1), where both θ and $\eta = f$ are unknown. Recall that in this case the model satisfies the LAE condition. Similarly to [32] we assume that f belongs to $\mathcal{F}$, the class of monotone non increasing densities on $(0,+\infty)$ and we define $\mathcal{F}_0 \subset \mathcal{F}$ as the set of functions satisfying:

1. $f \in \mathcal{F}$, $f(x) > 0$ for all $x > 0$.
2. There exist constants $0 < a, \gamma, M < +\infty$ such that f is absolutely continuous on $[0,a]$, $f(0) > 0$ and $\sup_{x\in(0,a)} |f'(x)| \le M < \infty$.

We write the prior on (η,θ) as a product prior on (f,θ): $d\Pi(f,\theta) = d\Pi_f(f)\pi_\theta(\theta)d\theta$. As discussed in Section 3.2.1, to apply Proposition 3 the hardest condition to verify is (3.5) and to some extend (i). Condition (i) is non standard in that one needs to derive a posterior concentration rate separately on θ and on f, however in the context of the shift model with monotone density functions f, this is easily deduced from the Hellinger posterior concentration rates on $f_{\theta,\eta}$.

Theorem 7. *Consider the model* (3.1) *with a prior* $d\Pi(f,\theta) = d\Pi_f(f)\pi_\theta(\theta)d\theta$ *on* $\mathcal{F}\times\mathbb{R}$*. We assume that the following conditions hold.*

(H1) Hellinger concentration: There exists $\mathcal{F}_1 \subset \mathcal{F}_0$ *where for all* $f_0 \in \mathcal{F}_1$*, there exists* $\epsilon_n \to 0$ *with* $n\epsilon_n^2 \to +\infty$ *such that for all compact subset* Θ_0 *of* Θ*,*

$$\sup_{\theta_0\in\Theta_0}\sup_{f_0\in\mathcal{F}_1} \mathbb{E}_{\theta_0,f_0}\Pi(h(f_\theta, f_{0,\theta_0}) \ge \epsilon_n \mid X^{(n)}) = o(1).$$

Assume also that there exists $a, M, \gamma_0, e > 0$ *with* $aM < \gamma_0$ *such that for all* $f \in \mathcal{F}_1$*,* $f(0) \ge \gamma_0$ *and* $\sup_{x\in(0,a)} |f'(x)| \le M$ *and* $f(a) \le f(0) - \iota$ *for a small* $\iota > 0$*.*

(H2) Uniform concentration near 0*: there exists* $u_n \to 0$ *as* $n \to \infty$ *such that*

$$\sup_{\theta_0\in\Theta_0}\sup_{f_0\in\mathcal{F}_1} \mathbb{E}_{\theta_0,f_0}\Pi_n(\sup_{x\in[0,a]} |f(x) - f_0(x)| > u_n \mid X^{(n)}) \to 0.$$

(H3) The prior density π_θ satisfies assumption (iii) of Proposition 3.

(H4) For all $f_0 \in \mathcal{F}_1$, there exist sets of functions $\mathcal{F}_n \subset \{f \in \mathcal{F}; h(f_0, f) \leq \epsilon_n\}$ such that any $f \in \mathcal{F}_n$ is differentiable and for all $\epsilon > 0$

$$\sup_{\theta_0 \in \Theta_0} \sup_{f_0 \in \mathcal{F}_1} \mathbb{E}_{\theta_0, f_0} \Pi_n(f \notin \mathcal{F}_n \mid X) = o(1),$$

$$\sup_{f_0 \in \mathcal{F}_1} \mathbb{P}_{0, f_0} \left(\sup_{f \in \mathcal{F}_n} |\mathbb{G}_n(\frac{f'}{f})| > \epsilon\sqrt{n} \right) = o(1),$$

$$\sup_{f \in \mathcal{F}_n} \left(\frac{M_n}{n} \left\| \frac{f'}{f} \right\|_\infty \right) = o(1)$$

and there exist $C_0 > 0$, such that for all $M_0 > 0$ and $M_n = M_0 \log n$ and for all $f_0 \in \mathcal{F}_1$, there exists $M(\cdot)$ on $\mathbb{R}_+$ satisfying $\int M(y) f_0(y) dy \leq C_0$ and for all $\epsilon > 0$, when n is large enough for all $\forall f \in \mathcal{F}_n$,

$$\left| \frac{f'(y+u)}{f(y+u)} - \frac{f'(y)}{f(y)} - \left(\frac{f_0'(y+u)}{f_0(y+u)} - \frac{f_0'(y)}{f_0(y)} \right) \right| \leq M(y)\epsilon,$$

for all $(y, u) \in (0, \infty) \times (-M_n/n, M_n/n)$.

Then with $\zeta_n = X_{(1)} = \min \mathbf{X}^n$ and

$$\sup_{\theta_0 \in \Theta_0} \sup_{f_0 \in \mathcal{F}_1} E_{\theta_0, f_0} \left(||\Pi_n - \mathcal{E}_{f_0(0)}||_{TV} \right) = o(1).$$

Note that the rate ϵ_n and the sets $\mathcal{F}_n$ are allowed to depend on f_0, θ_0.

Condition (H1) can be proved using the technique of [21] and large families of prior models on densities on have been investigated. Condition (H3) has been discussed and is very mild. Condition (H2) is slightly stronger than the pointwise consistency at 0. In the context of monotone non increasing densities, this condition is non trivial if the prior on f consists on a mixture of Uniform distributions, see for instance [44] and in particular the nonparametric maximum likelihood estimator is not consistent at 0. Condition (H4) implies condition (3.5) in Proposition 3.

For instance, the prior considered in [32] trivially satisfy most of these assumptions. Indeed the prior is defined by modelling

$$f(x) = \frac{e^{-\alpha x + \int_0^x \ell(y) dy}}{\int_0^\infty e^{-\alpha x + \int_0^x \ell(y) dy}},$$

and the prior on f is defined as the transformation of a prior on ℓ. In [32] , the authors assume that with probability 1 under the prior, ℓ belongs to the set of continuous functions, bounded by a given constant S and converging at infinity. Since $f'/f = \ell - \alpha$ is uniformly bounded by $S + \alpha$ over H. Moreover since H is included in the set of monotone non increasing functions, bounded by $S + \alpha$ and under the assumption that $\Pi(\|\ell - \ell_0\|_\infty \leq \epsilon) > 0$, then ϵ_n can be defined as a sequence going to 0 arbitrarily slowly. From that pointwise and uniform (locally around 0) consistency can be deduce from the bound on f'/f together with L_1 consistency. Finally, the first part of (H4) is proved with $\mathcal{F}_n = \{f \in H; d_H(f_0, f) \leq \epsilon_n\}$, which a Donsker class. The only condition which would need to be proved is the last part of (H4) and more precisely

$$\sup_{f\in\mathcal{F}_n}\left(\sup_{|u|\le M_n/n}\sup_y e^{-sy}\,|\ell(y+u)-\ell(y)|\right)=o(1),$$

for some $s<\alpha$, where H can be replaced by some set $H_n\subset H$ which has posterior probability going to 1. Note that the above condition can be verified either by proving sup-norm consistency on ℓ or equicontinuity.

3.2.3 Scale LAE model

Consider a scale LAE model $Z_i\sim\tau_0^{-1}g_0(x/\tau_0)$, $i=1,\dots,n$ independently, where g_0 is supported on $(1,\infty)$ and is independent of scale $\tau_0>0$. Example of such distributions is Pareto family with pdf $p(x\mid\tau,a)=ax^{-a-1}\tau^a I(x\ge\tau)$, $a,\tau>0$.

We can turn a scale model into a location model, by transforming the data to $X_i=\log(Z_i)$ and setting $f_0(x)=e^x g_0(e^x)$ and $\theta_0=\log\tau_0$, we get $X_i\sim f_0(\cdot-\theta_0)$.

We show that if the BvM is satisfied for the location parameter of a shift model, it is satisfied for the corresponding scale parameter of the corresponding scale model.

Proposition 4. *Let $X_i\sim f_0(\cdot-\theta_0)$ iid $i=1,\dots,n$. Suppose that under a Bayesian model $X_i\mid f,\theta\sim f(\cdot-\theta)$, iid $i=1,\dots,n$ and prior $\Pi(f,\theta)$,*

$$||\Pi_n(n(X_{(1)}-\theta)\mid X^{(n)})-\mathcal{E}(f_0(0))||_{TV}\le\alpha_n\overset{P_{f_0,\theta_0}}{\to}0$$

as $n\to\infty$ and $n(X_{(1)}-\theta_0)$ weakly converges to $\mathcal{E}(f_0(0))$.

Then, for the corresponding scale model with $Y_i=\exp(X_i)$, $\tau=\exp(\theta)$, $g(y)=f(\log y)/y$, and $g_\tau(y)=g(y/\tau)/\tau$, and prior on (g,τ) induced by the prior $\Pi(f,\theta)$,

$$||\Pi_n(n(Y_{(1)}-\tau_0)\mid Y^{(n)})-\mathcal{E}(g_{0,\tau_0}(\tau_0))||_{TV}=O_{P_{g_{0,\tau_0}}}(\max(\alpha_n,[\log(1/\alpha_n)]^2/n))\overset{P_{g_0,\tau_0}}{\to}0$$

as $n\to\infty$ for an appropriate choice of α_n, and $n(Y_{(1)}-\tau_0)$ weakly converges to $\mathcal{E}(g_{0,\tau_0}(\tau_0))$.

3.3 BvM Theorem for Mixture Prior

3.3.1 Mixture Prior

Recall that $\mathcal{F}$ is a set of monotone non increasing densities on $\mathbb{R}_+$, using the mixture representation of such functions, see [50]. We consider a prior on f by modelling a prior on probabilities G on $\mathbb{R}_+$ where $f(x)=\int_x^\infty\theta^{-1}dG(\theta)$. Moreover, since we restrict ourselves to continuously differentiable densities, we do not consider the usual Dirichlet process prior on G, but instead model it as a nonparametric mixture of Gamma densities following the recent work of [4], so that f is modelled as $f_{P,z}$: let $z>2$,

$$f_{P,z}(x)=\int_x^\infty\frac{1}{\theta}g_{P,z}(\theta)d\theta=\int_0^\infty\left(\frac{z}{\epsilon}\right)^z\frac{1}{\Gamma(z)}\left(\int_x^\infty\theta^{z-2}e^{\frac{-\theta z}{\epsilon}}d\theta\right)dP(\epsilon). \tag{3.8}$$

To ensure that $f_{P,z}$ is continuously differentiable on $\mathbb{R}$ we consider the following model for P: let $\delta>0,p\in(0,1)$, and $Q^{(0)},Q^{(1)}$ be probability distribution on $(0,\delta)$ and $(\delta,+\infty)$ respectively,

then

$$dP(\epsilon) = \frac{p\epsilon^2 dQ^{(0)}(\epsilon) + (1-p)dQ^{(1)}(\epsilon)}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p} \tag{3.9}$$

The prior on $f_{P,z}$ is defined as

$$Q^{(0)} \sim DP(M^{(0)}, G^{(0)}), \quad Q^{(1)} \sim DP(M^{(1)}, G^{(1)}), \quad (z,p) \sim \pi_z \otimes \pi_p$$

where $M^{(0)}, M^{(1)} > 0$, $G^{(0)}$ is a probability measure on $[0,\delta)$ with density $g^{(0)}$ and $G^{(1)}$ is a probability measure on $[\delta,\infty)$ with density $g^{(1)}$.

We consider the following assumptions on $g^{(0)}$ and $g^{(1)}$: there exist $-1 < a_0' \leq a_0 < \infty$ and $1 < a_1 \leq a_1' < \infty$ such that

$$x^{a_0} \lesssim g^{(0)}(x) \lesssim x^{a_0'} \text{ as } x \to 0, \text{ and } \lim_{x \nearrow \delta} g^{(0)}(x) > 0, \tag{3.10}$$

$\lim_{x\searrow\delta} g^{(1)}(x) > 0$ and

$$x^{-a_1} \lesssim g^{(1)}(x) \lesssim x^{-a_1'} \text{ as } x \to \infty. \tag{3.11}$$

The prior on z, Π_z satisfies : for some constants $c \geq c' > 0, c_0 > 0$ and $\rho_z \geq 0$,

$$\begin{aligned} &\Pi_z([x,2x]) \gtrsim e^{-c\sqrt{x}(\log x)^{\rho_z}}, \quad \Pi_z([x,+\infty)] \lesssim e^{-c'\sqrt{x}(\log x)^{\rho_z}} \quad \text{as } x \to +\infty, \\ &\Pi_z([0,x]) \lesssim x^{c_0} \quad \text{for } x \to 0. \end{aligned} \tag{3.12}$$

Consider a continuous prior measure on p, Π_p with density π_p, supported on $(0,p_n)$ such that for some fixed $0 < p_0 < p_1 < 1$, $p_1 \leq p_n$,

$$\begin{aligned} &\Pi_p([0,x]) \leq C/\log(1/x) \text{ for } x \in (0,p_0), \\ &\pi_p(x) \gtrsim \exp\{-d[\log(1/x)]^{5/2}\} \text{ for } x \in (0,p_0), \\ &\pi_p(x) \geq c > 0 \text{ for } x \in (p_0,p_1), \\ &\pi_p(x) \gtrsim \exp\{-d[\log(1/(1-x))]^{5/2}\} \text{ for } x \in (p_1,p_n), \end{aligned} \tag{3.13}$$

for some constants $c, d, d' > 0$ independent of n. Assume that $1/(1-p_n) \leq [\log n]^s$ for some $s > 0$.

We assume that the prior distribution on θ satisfies the following conditions.

Assumption 1. We assume that the prior Π_θ on $\Theta = \mathbb{R}$ has positive and continuous density with respect to Lebesgue measure on $\mathbb{R}$ such that there exist $\kappa > 1$, $C_\kappa > 0$: $\Pi_\theta(|\theta| \geq x) \leq C_\kappa x^{-\kappa}$ for large enough x.

Note that under this mixture prior, $f(0) \leq p_n\delta/(1-p_n) + 1/\delta$, and $f'(0) \leq p_n/(1-p_n) + 1/\delta^2$. Also note that by construction and the approximation properties proved in section sec:proofs-cont-disc-approx, this prior model has full KL and L_1 support on $\mathcal{F} \times \mathbb{R}$ where $\mathcal{F}$ is the functional class defined in the next section. Indeed, since the prior on f is supported on densities supported on the whole semiline $[0,\infty)$, for any given pair (θ_0, f_0), the KL divergence $KL(f_{0,\theta_0}, f_\theta)$ is finite for all $\theta \leq \theta_0$ and any $f\mathcal{F}$, avoiding the problem with KL neighbourhoods we faced in Chapter 2.

3.3.2 Functional Class

Let the true density of X_is be $f_{0,\theta_0}(x) = f_0(x - \theta_0)$ where f_0 is in the set $\mathcal{F}$ of functions f. We assume the following smoothness condition on f_0.

Definition 4. Let $\mathcal{P}(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu)$, with $\nu \in (0, 1/3)$, contain the set of functions $f : \mathbb{R}^+ \to [0, \infty)$ which are r times continuously differentiable with $r = \lceil \beta \rceil - 1$ and which satisfy for all $x \in \mathbb{R}_+$ and y: $y > -x$ and $|y| \leq \Delta$,

$$\left| f^{(r)}(x+y) - f^{(r)}(x) \right| \leq L(x)|y|^{\beta - r}(1 + |y|^\gamma), \quad \sum_{\ell=0}^{r} |f^{(\ell)}(0)| + L(0) \leq C_0, \tag{3.14}$$

defining $r_0 = \lceil \beta/2 \rceil - 1$ for $\beta > 2$ and $r_0 = 0$ if $\beta \leq 2$,

$$\begin{aligned}
&\sum_{j=2}^{r} \int_0^\infty \left(\frac{x^j |f^{(j)}(x)|}{g(x)} \right)^{\frac{\beta/\nu + e}{j}} g(x)dx + \int_0^\infty \left(\frac{x^\beta L(x)}{g(x)} \right)^{(\beta/\nu + e)/\beta} g(x)dx \leq C_1 \\
&\int_0^\infty \left(\frac{(1 + x^{\gamma + 2r_0}) x^\beta L(x)}{g(x)} \right)^2 g(x)dx + \int_0^\infty f(x)^{1-2\nu} dx \leq C_1 \\
&\sum_{j=1}^{r} \int_0^\infty f(x) \left(\frac{(x^j + 1)|f^{(j)}(x)|}{f(x)} \right)^{\frac{\beta/\nu + e}{j}} dx \\
&+ \int_0^\infty f(x) \left(\frac{(x^\beta + 1) L(x) (1 + x^\gamma)}{f(x)} \right)^{(\beta/\nu + e)/\beta} dx \leq C_1
\end{aligned} \tag{3.15}$$

for some $e > 0$ where $g(x) = -xf'(x)$. Here $L(\cdot)$ is a fixed positive function from $\mathbb{R}^+$ to $\mathbb{R}^+$. If $r < 2$, the first condition in (3.15) becomes

$$\int_0^\infty \left(\frac{x^\beta L(x)}{g(x)} \right)^{3+e} g(x)dx \leq C_0.$$

Note that if f is a density in $\mathcal{F}$ then it is easy to show that $g(x) = -xf'(x)$ is also a density on $\mathbb{R}_+$.

3.3.3 Posterior Concentration in L^1 norm with known θ

First we consider a nonparametric density estimation problem where the density is from the class described in Section 3.3.2 with known $\theta = \theta_0 = 0$. To add an additional moment condition, we introduce

$$\mathcal{P}'(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu) = \left\{ \begin{array}{l} f_0 \in \mathcal{P}(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu) : \\ \int x^{2\beta/((1-\nu)\beta - 2\nu)} f_0(x)dx < \infty. \end{array} \right\}. \tag{3.16}$$

$$\exists \rho_1 > 0 \,\&\, C_2 > 0 : \int_x^\infty y^2 f_0(y)dy \leq C_2(1+x)^{-\rho_1}. \tag{3.17}$$

We denote by $\mathcal{T}(\rho_1, C_2)$ the set of densities satisfying (3.17). Note that his condition is mild, it is satisfied by a Student t distribution with $\nu > 2$ degrees of freedom.

Remark 2. If f_0 satisfies condition (3.17) and is bounded, then condition $\int_0^\infty f_0^{1-2\nu}(y)dy < \infty$ holds with any $\nu \leq 1/3$ such that $\rho_1 > (3\nu - 1)/[\nu(1 - 2\nu)]$. This is due to Hölder inequality with $p = 1/(1-2\nu)$ and $q = 1/(1-1/p) = 1/(2\nu)$,

$$\begin{aligned} \int_x^\infty f_0^{1-2\nu}(y)dy &\leq \left[\int_x^\infty y^2 f_0(y)dy\right]^{1-2\nu} \left[\int_x^\infty y^{2-1/\nu}dy\right]^{2\nu} \\ &\leq C[1+x]^{-\rho_1(1-2\nu)}[x^{3-1/\nu}[1/\nu-3]^{-1}\mathbb{1}(\nu<1/3) + \log x \mathbb{1}(\nu=1/3)] \end{aligned}$$

which is finite for large x under the above conditions.

We also need the following assumption on p_n and δ:

$$\delta \leq a, \quad \log\log\left(\frac{\delta^2 p_n}{(1-p_n)}\right) \leq C\log\log n, \quad \log\log\left(\frac{p_n}{(1-p_n)}\right) \leq C\log n, \tag{3.18}$$

where a comes from the definition of $\mathcal{F}_0(a, M, u)$.

Theorem 8. *Consider the prior on f defined in Section 3.3.1 under assumption (3.18) and assume that $\mathbf{X}^n = (X_1, \cdots, X_n)$ is a sample of independent observations identically distributed according to a probability P_0 on $\mathbb{R}^+$ having density f_0 with respect to Lebesgue measure.*

Then, for any $\beta_1 \geq \beta_0 > 1$, $L(\cdot)$, γ, $C_0, C_1, e, \Delta, C_2, \rho_1, a > 0$, $0 < \nu \leq 1/3$, M_0, $u > 0$, there exists $M > 0$ such that

$$\sup_{\beta\in[\beta_0,\beta_1]} \sup_{f_0\in\mathcal{Q}_\beta(\cdots)} E_{P_0}\left[\Pi\left(d_H(f, f_0) > M\frac{(\log n)^q}{n^{\beta/(2\beta+1)}}|\mathbf{X}^n\right)\right] = o(1) \tag{3.19}$$

with $q = 5\beta/(4\beta+2)$ if $\rho_z \leq 5/2$ and $q = 2\rho_z\beta/(4\beta+2)$ if $\rho_z > 5/2$ and

$$\mathcal{Q}_\beta(\cdots) = \mathcal{P}'(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu) \cap \mathcal{T}(\rho_1, C_2) \cap \mathcal{F}_0(a, M_0, u).$$

Theorem 8 is proved in Section 3.6.

Remark 3. If $f_0(x) \lesssim x^{-\rho-1}$ for large x for some $\rho > 0$, then condition (3.17) holds for $\rho > 2$, $\int f_0^{1-2\nu}(x)dx < \infty$ holds for any $\nu \in (0, 1/3]$ and $\rho > 2$, and condition (3.16) holds for $\nu \in (0, 1/3]$ such that $\nu < (1-2/\rho)\beta/(\beta+2)$.

3.3.4 Joint Posterior Concentration in L^1 norm

For this proposition, we need the following assumption on δ and p_n: there exists a sequence $0 < m_0 \leq \tilde{M}_n \leq m_1[\log n]^A$ for some $A \geq 0$ such that

$$\log\left[\frac{\delta p_n}{1-p_n} + \frac{1}{\delta}\right] \leq \tilde{M}_n. \tag{3.20}$$

Proposition 5. *Consider the shift model 3.1 with prior on (f, θ) defined in Section 3.3.1 under assumption (3.18) and (3.20). Assume that $\mathbf{X}^n = (X_1, \cdots, X_n)$ is a sample of independent observations identically distributed according to a probability P_0 on $\mathbb{R}^+$ having density f_0 with respect to Lebesgue measure.*

Then, for any $\beta_1 \geq \beta_0 > 1$, $L(\cdot)$, γ, $C_0, C_1, e, \Delta, C_2, \rho_1, a > 0$, $0 < \nu \leq 1/3$, M_0, $u > 0$, and any compact $\Theta_0 \subset \mathbb{R}$, there exists $M > 0$ such that

$$\sup_{\theta_0\in\Theta_0} \sup_{\beta\in[\beta_0,\beta_1]} \sup_{f_0\in\mathcal{Q}_\beta(\cdots)} E_{P_0}\left[\Pi\left(d_H(f(\cdot-\theta), f_0(\cdot-\theta_0)) > M\frac{(\log n)^q}{n^{\beta/(2\beta+1)}}|\mathbf{X}^n\right)\right] = o(1) \tag{3.21}$$

with $q = 5\beta/(4\beta+2)$ *if* $\rho_z \le 5/2$ *and* $q = 2\rho_z\beta/(4\beta+2)$ *if* $\rho_z > 5/2$ *and*

$$\mathcal{Q}_\beta(\cdots) = \mathcal{P}'(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu) \cap \mathcal{T}(\rho_1, C_2) \cap \mathcal{F}_0(a, M_0, u).$$

This proposition is proved in Section 3.5.4.

3.3.5 Local Consistency of density with known θ

Proposition 6. *Under assumptions of Theorem 8 and* $(1-p_n)\delta^{-2}\epsilon_n^{1/3} = o(1)$*,*

$$\Pi_n\Big(\sup_{x\in[0,a]} |f(x) - f_0(x)| \le u_n \mid X^{(n)}\Big) \to 1$$

as $n \to \infty$ *where* $u_n = C_0\epsilon_n^{1/3}\max(1,\delta)(1-p_n)^{-1}$ *with* C_0 *large enough.*

This proposition is proved in Section 3.10.1

3.3.6 Interaction Term

In the theorem below we verify the sufficient conditions for the interaction term to vanish stated in Theorem 7.

We need the following additional assumptions. First, for uniform bounds, we need additional conditions for f_0 in a neighbourhood of 0.

Definition 5. Let a class of functions $\tilde{\mathcal{F}}_0(a, M, u, v)$ with $0 < a, M < \infty$, $0 < aM < u < \infty$, small $v > 0$, be a class of monotone non-increasing positive densities on $(0, +\infty)$ such that for all $f_0 \in \tilde{\mathcal{F}}_0(a, M, u, v)$,

$$f_0(0) \ge u, \qquad \sup_{x\in(0,a)} |f_0'(x)| \le M, \qquad f_0(a) \le f_0(0) - v. \tag{3.22}$$

Note that condition $aM < u$ implies that $f_0(a)$ is uniformly bounded from below by $u - aM > 0$. We need the following additional assumptions. For $f_0 \in \tilde{\mathcal{F}}_0(a, M, u, v)$,

$$\delta \le M^{-1}(f_0(0) - v), \quad \delta^2 \le (1-p_n)/p_n \wedge a/4, \quad 1 - p_n \le C[\log n]^{-1}, \tag{3.23}$$

with C denoting any finite positive constant.

In particular, conditions (3.20) and assumption of Proposition 6 that $(1-p_n)\delta^{-2}\epsilon_n^{1/3} = o(1)$ hold for any $\beta > 1$ under (3.23).

Theorem 9. *Assume that* $\mathbf{X}^n = (X_1, \cdots, X_n)$ *is a sample of independent observations identically distributed according to a probability* P_0 *on* $\mathbb{R}^+$ *having density* f_0 *with respect to Lebesgue measure. Consider the prior defined in Section 3.3.1 under assumption* (3.23)*.*

Then, there exist sets of functions $\mathcal{F}_n \subset \mathcal{F} \cap \mathcal{P}'(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu) \cap \tilde{\mathcal{F}}_0(a, M, u, v)$ *for*

$\nu \in (0, 1/3]$ *and* $\beta > 3/2$ *such that any* $f \in \mathcal{F}_n$ *is differentiable and, for all* $\epsilon > 0$,

$$\sup_{\theta_0 \in \Theta_0} \sup_{f_0 \in \mathcal{F}_1} \mathbb{E}_{P_0} \Pi_n(f \notin \mathcal{F}_n \mid X) = o(1),$$
$$\sup_{f_0 \in \mathcal{F}_1} \mathbb{P}_{0,f_0} \left(\sup_{f \in \mathcal{F}_n} |\mathbb{G}_n(\frac{f'}{f})| > \epsilon\sqrt{n} \right) = o(1),$$
$$\sup_{f \in \mathcal{F}_n} \left(\frac{M_n}{n} \left\| \frac{f'}{f} \right\|_\infty \right) = o(1)$$

and there exist $C_0 > 0$, *such that for all* $M_0 > 0$ *and* $M_n = M_0 \log n$ *and for all* $f_0 \in \mathcal{F}_1$, *there exists* $M(\cdot)$ *on* $\mathbb{R}_+$ *satisfying* $\int M(y) f_0(y) dy \le C_0$ *and for all* $\epsilon > 0$, *when* n *is large enough for all* $\forall f \in \mathcal{F}_n$,

$$\left| \frac{f'(y+u)}{f(y+u)} - \frac{f'(y)}{f(y)} - \left(\frac{f_0'(y+u)}{f_0(y+u)} - \frac{f_0'(y)}{f_0(y)} \right) \right| \le M(y)\epsilon,$$

for all $(y, u) \in (0, \infty) \times (-M_n/n, M_n/n)$.

This theorem is proved in Section 3.10.2.

3.3.7 Bernstein-von Mises theorem for the mixture model

Combining previous results we have verified the sufficient conditions for the BvM theorem for the marginal posterior distribution of θ stated in Theorem 7 which are summarised below.

Under the mixture prior on f defined in Section 3.3.1, we make assumptions (3.18) and (3.23). The latter assumptions hold for large enough n if

$$1/(1 - p_n) = C[\log n]^s, \quad \delta = c[\log n]^{-s/2} \tag{3.24}$$

with $0 < s \le 1$, $C \in (0, \infty)$ and $0 < c < C^{-1/2}$.

Theorem 10. *Assume that* $\mathbf{X}^n = (X_1, \cdots, X_n)$ *is a sample of independent observations identically distributed according to a probability* P_0 *on* $\mathbb{R}^+$ *having density* $f_0(\cdot - \theta_0)$ *with respect to Lebesgue measure that satisfies the LAE assumption with* $\gamma_{\theta_0} = f_0(0)$.

Consider the model (3.1) *with a prior* $d\Pi(f, \theta) = d\Pi_f(f)\pi_\theta(\theta)d\theta$ *on* $\mathcal{F} \times \mathbb{R}$ *where the prior on* f *is defined in Section 3.3.1 under additional conditions* (3.24), *and that the prior on* θ *satisfies Assumption 1.*

Then, for some $\beta_1 \ge \beta_0 \ge 3/2$, $L(\cdot)$, γ, C_0, C_1, e, Δ, C_2, ρ_1, a, M, M_0, $u, v > 0$, $0 < \nu \le 1/3$, *and any compact set* $\Theta_0 \subset \mathbb{R}$,

$$\sup_{\theta_0 \in \Theta_0} \sup_{f_0 \in \mathcal{Q}_\beta(\cdots)} E_{\theta_0, f_0} \left(||\Pi_n(n(X_{(1)} - \theta) \mid \mathbf{X}^n) - \mathcal{E}_{f_0(0)}||_{TV} \right) = o(1)$$

where

$$\mathcal{Q}_\beta(\cdots) = \mathcal{P}'(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu) \cap \mathcal{T}(\rho_1, C_2) \cap \tilde{\mathcal{F}}_0(a, M, u, v).$$

3.4 Numerical Results

3.4.1 Implementation

We developed two versions of a Slice sampler algorithm for the following prior model

$$
\begin{gathered}
f_\theta(x) = f(x-\theta); \quad f(x) = \int_x^\infty \frac{1}{y}\int_0^\infty g_z(y;\epsilon)dP(\epsilon)dy \\
dP(\epsilon) = \frac{p\epsilon^2 dQ^{(0)}(\epsilon) + (1-p)dQ^{(1)}(\epsilon)}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p}; \quad Q^{(i)} \sim DP(m, H^{(i)}),\ i = 0, 1. \\
H^{(0)}(x) \propto x^a,\ x \in [0,1]; \quad H^{(1)}(x) \propto x^{-a},\ x \in (1,\infty); \quad a > 1. \\
p = qp_n,\ q \sim Beta(\alpha,\beta),\ \alpha,\beta > 0,\ p_n \in (0,1) \\
\Pi_z : \sqrt{z} \sim \Gamma(b,c),\ b,c > 0; \quad \theta \sim t_d
\end{gathered}
$$

First note that if $X \sim f(x)$ and $Y \sim f_Y(y) := \int_0^\infty g_z(y;\epsilon)dP(\epsilon)$ then $X|Y \sim Unif(0,Y)$, therefore X can be expressed as $X = \xi Y$ where $\xi \sim Unif(0,1)$. Thus, that equality and ξ being independent from Y implies $X|\xi \sim f_Y(x/\xi)/\xi$. Thus the joint density of X and ξ is

$$
f(x,\xi) = \mathbb{1}(0 \le \xi \le 1)\int_0^\infty g_z(x/\xi;\epsilon)/\xi dP(\epsilon) \tag{3.25}
$$

First Method

Using the stick-breaking representation of the Dirichlet Process

$$
f(x,\xi) = \frac{p\sum_{j=1}^\infty (\epsilon_j^{(0)})^2 q_j^{(0)} g_z(x/\xi;\epsilon_j^{(0)})/\xi + (1-p)\sum_{j=1}^\infty q_j^{(1)} g_z(x/\xi;\epsilon_j^{(1)})/\xi}{p\sum_{j=1}^\infty (\epsilon_j^{(0)})^2 q_j^{(0)} + 1 - p} \tag{3.26}
$$

with

$$
q_j^{(i)} = V_j^{(i)} \prod_{k=1}^{j-1}(1 - V_k^{(i)}) \tag{3.27}
$$

and $V_j^{(i)} \sim Beta(1,m)$.

Let us denote $T = p\sum_{j=0}^{N^{(0)}} (\epsilon_j^{(0)})^2 q_j^{(0)} + 1 - p$. We augment the model with other auxiliary variables; $(v|p, V^{(0)}, V^{(1)}, \epsilon) \sim \Gamma(n, T)$ to handle the normalising constant, $u = (u_1, \ldots, u_n)$ for truncation, and $c = (c_1, \ldots, c_n)$ for allocation.

Therefore, the full likelihood becomes

$$
\begin{aligned}
L_n(X^n, u, c, v, \xi) = & \frac{v^{n-1} e^{-v\left[p\sum_{j=0}^\infty (\epsilon_j^{(0)})^2 q_j^{(0)} + 1 - p\right]}}{\Gamma(n)} \\
& \times \prod_{i=1}^n \mathbb{1}(u_i \le p_{c_i})\, \mathbb{1}(0 \le \xi_i \le 1)\, g_z((X_i - \theta)/\xi_i; \epsilon_{c_i})/\xi_i
\end{aligned} \tag{3.28}
$$

where

$$
p_j = \begin{cases} p\epsilon_j^2 q_k^{(0)} & \text{if } \epsilon_j \le \delta \\ (1-p)q_k^{(1)} & \text{otherwise.} \end{cases} \tag{3.29}
$$

Note that T contains a sum of an infinite number of elements. As a first approach we use the approximation $T' = p\sum_{j:p(\epsilon_j^{(0)})^2 q_j^{(0)} > L} (\epsilon_j^{(0)})^2 q_j^{(0)} + 1 - p$ and $L = \min_i u_i$. Note that the error in the approximation is $T - T' = p\sum_{j:p(\epsilon_j^{(0)})^2 q_j^{(0)} \le L} (\epsilon_j^{(0)})^2 q_j^{(0)}$ and the distribution of L given the weights p_{c_i} is given by

$$f(L|p_{c_1}, \ldots, p_{c_n}) = \sum_{i=1}^{n} \left[\frac{1}{p_{c_i}} \prod_{j \neq i} \left(1 - \frac{L}{p_{c_j}} \right) \right] \tag{3.30}$$

then

$$\frac{p_{min}}{n+1} \le E(L|p_{c_1}, \ldots, p_{c_n}) \le \frac{p_{max}}{n+1} \tag{3.31}$$

$$Var(L|p_{c_1}, \ldots, p_{c_n}) \le \frac{np_{\max}}{(n+1)^2(n+2)} \tag{3.32}$$

with $p_{min} := \min\{p_{c_i}\}_{i=1,\ldots,n}$ and $p_{max} := \max\{p_{c_i}\}_{i=1,\ldots,n}$

This allows us to use a Gibbs sampler algorithm based on the following conditional distributions

- $[v|\ldots] \sim \Gamma(n, T')$, $\quad T' = p\sum_{j:p(\epsilon_j^{(0)})^2 q_j^{(0)} > L} (\epsilon_j^{(0)})^2 q_j^{(0)} + 1 - p$, $\quad L = \min_i u_i$.
- $[\xi_i|\ldots] \propto \mathbb{1}(0 \le \xi_i \le 1)\, g_z((X_i - \theta)/\xi_i; \epsilon_{c_i})/\xi_i$
- $[z|\ldots] \propto \frac{\pi_z(z) z^{nz}}{\Gamma(z)^n} e^{-z\sum_j S_j/\epsilon_j + z\sum_i \log((X_i-\theta)/\xi_i) - z\sum_j n_j \log \epsilon_j}$, where $S_j = \sum_{c_i=j}(X_i - \theta)/\xi_i$, and $n_j = \sum_{i=0}^{n} \mathbb{1}(c_i = j)$
- $[u_i|\ldots] \sim Unif(0, p_{c_i})$
- $Pr[c_i = j|\ldots] \propto \mathbb{1}(u_i \le p_j)\, g_z(X_i/\xi_i; \epsilon_j)$
- $[\epsilon_j^{(0)}|\ldots$ excluding $u] \sim e^{-zS_j/\epsilon_j - vpq_j\epsilon_j^2} \epsilon_j^{-(zn_j - 2 - a)} \mathbb{1}(0 \le \epsilon_j \le \delta)$
- $[\epsilon_j^{(1)}|\ldots$ excluding $u] \sim e^{-zS_j/\epsilon_j} \epsilon_j^{-(zn_j + a)} \mathbb{1}(\epsilon_j > \delta)$
- $[V_j^{(0)}|\ldots$ excluding $u] \sim e^{-BV_j} V_j^{n_j} (1 - V_j)^{m-1+\sum_{l>j} n_l} \mathbb{1}(0 \le V_j \le 1)$, $\quad B = vp[\epsilon_j^2 \prod_{k<j}(1 - V_k) - \sum_{k>j} \epsilon_k^2 V_k \prod_{l<k, l \neq j}(1 - V_l)]$
- $[V_j^{(1)}|\ldots$ excluding $u] \sim Beta(n_j + 1,\ m + \sum_{k>j} n_k)$
- $[p|\ldots$ excluding $u] \sim e^{pvR} p^{n_L + \alpha - 1}(1-p)^{n_U}(p_n - p)^{\beta - 1}$, $\quad R = 1 - \sum_j \left(\epsilon_j^{(0)}\right)^2 q_j^{(0)}$, $\quad n_L = \sum_{i=0}^{n} \mathbb{1}(\epsilon_{c_i} \le \delta)$, $\quad n_U = n - n_L$
- $[\theta|\ldots] \sim \prod_{i=1}^{n}(X_i - \theta)^{z-1} e^{\frac{z\theta}{\epsilon_{c_i}\xi_i}} \mathbb{1}\left(\theta \le X_{(1)}\right) \pi_\theta(\theta)$

Second Method

We can follow the approach in Griffin & Walker (2011) [27] and marginalise over the weights that are not allocated but their idea cannot be applied directly as the weights from the Dirichlet

Process are not independent. Therefore we need to find a way to express our process with independent jumps. For this purpose use the stick-breaking representation of the Dirichlet Process only for $Q^{(1)}$ and express $Q^{(0)}$ as a normalised Gamma Process.

$$f(x,\xi) = \frac{p\sum_{j=1}^{\infty}(\epsilon_j^{(0)})^2 \frac{J_j}{\sum_k J_k} g_z(x/\xi;\epsilon_j^{(0)})/\xi + (1-p)\sum_{j=1}^{\infty} q_j^{(1)} g_z(x/\xi;\epsilon_j^{(1)})/\xi}{p\sum_{j=1}^{\infty}(\epsilon_j^{(0)})^2 \frac{J_j}{\sum_k J_k} + 1 - p} \tag{3.33}$$

with

$$\sum_{j=1}^{\infty} J_j \delta_{\epsilon_j^{(0)}} \sim GP(m, H^{(0)})$$
$$q_j^{(1)} = V_j^{(1)} \prod_{k=1}^{j-1}(1 - V_k^{(1)})$$

and $V_j^{(1)} \sim Beta(1,m)$.

Let us define $w := \sum_k J_k$ and add it to the Gibbs sampler. We know that the prior for w is $w \sim \Gamma(m,1)$.

Therefore, the full likelihood becomes

$$\begin{aligned} L_n(X^n,u,c,v,\xi) =& \frac{v^{n-1} e^{-v\left[p\sum_{j=0}^{\infty}(\epsilon_j^{(0)})^2 J_j/w + 1 - p\right]}}{\Gamma(n)} \\ &\times \prod_{i=1}^{n} \mathbb{1}(u_i \le p_{c_i})\, \mathbb{1}(0 \le \xi_i \le 1)\, g_z((X_i - \theta)/\xi_i;\epsilon_{c_i})/\xi_i \end{aligned} \tag{3.34}$$

where

$$p_j = \begin{cases} p\epsilon_j^2 J_k/w & \text{if } \epsilon_j \le \delta \\ (1-p)q_k^{(1)} & \text{otherwise.} \end{cases} \tag{3.35}$$

Following Griffin & Walker (2011) [27] we introduce a new variable $0 < L \le \min\{p_{c_i} : \epsilon_{c_i} \le \delta\}$ and marginalise out the weights that are smaller than L. Let S_L be the set of weights greater than $L := \min\{u_i\}_{i=1,\dots,n} = u_{(1)}$, $S_L = \{j : p(\epsilon_j^{(0)})^2 J_j > L\}$. Denote $J_{S_L} = \sum_{j\in S_L} \frac{p}{w}(\epsilon_j^{(0)})^2 J_j$ and $J_{S_L^C}$ similarly. Thus,

$$\begin{aligned} L_n(X^n,u,L,c,v,\xi) =& \frac{v^{n-1} e^{-vJ_{S_L}} e^{-v(1-p)}}{\Gamma(n)} E(e^{-vJ_{S_L^C}})\delta_{u_{(1)}}(L) \\ &\times \prod_{i=1}^{n} \mathbb{1}(u_i \le p_{c_i})\, \mathbb{1}(0 \le \xi_i \le 1)\, g_z((X_i - \theta)/\xi_i;\epsilon_{c_i})/\xi_i \end{aligned} \tag{3.36}$$

where $\delta_a(\cdot)$ denotes a point mass at a. From the Levy-Khintchine representation of the process we obtain

$$
\begin{aligned}
E(e^{-vJ_{S_L^C}}) =& m\int_0^\delta \int_0^{L/p} \left[1-e^{-svp/w}\right] e^{-s/x^2} s^{-1} ds dH^{(0)}(x) \\
=& m\int_0^\delta \int_0^{L/p} \left[1-e^{-svp/w}\right] e^{-s/x^2} s^{-1} \frac{(a+1)}{\delta} x^a ds dx
\end{aligned}
\tag{3.37}
$$

Note that

$$
\int_0^\delta e^{-s/x^2} x^a dx = \frac{s^{(a+1)/2}}{2}\Gamma\left(-\frac{a+1}{2}, \frac{s}{\delta^2}\right) = \frac{\delta^{a+1}}{2} E_{\frac{a+3}{2}}\left(\frac{s}{\delta^2}\right) \tag{3.38}
$$

where $\Gamma(a,x)$ is the incomplete Gamma function and $E_a(x)$ is the Exponential integral function. Therefore

$$
E(e^{-vJ_{S_L^C}}) = m\frac{(a+1)}{2\delta}\int_0^{L/p} \left[1-e^{-svp/w}\right] s^{(a-1)/2}\Gamma\left(-\frac{a+1}{2}, \frac{s}{\delta^2}\right) ds \tag{3.39}
$$

which can be calculated via numerical integration.

If variables u_i are integrated out.

$$
\begin{aligned}
L_n(X^n, L, c, v, \xi) =& \frac{v^{n-1} e^{-vJ_{S_L}} e^{-v(1-p)}}{\Gamma(n)} E(e^{-vJ_{S_L^C}}) \\
& \times \sum_{i=1}^n \Big[\mathbb{1}(0 \le \xi_i \le 1)\, g_z((X_i-\theta)/\xi_i; \epsilon_{c_i})/\xi_i \\
& \qquad \prod_{j\neq i} (p_{c_j} - L)\mathbb{1}(0\le \xi_j \le 1)\, g_z((X_j-\theta)/\xi_j; \epsilon_{c_j})/\xi_j\Big]
\end{aligned}
\tag{3.40}
$$

which is not very useful for a Gibbs sampler. Note that this differs from the expression in the work by Griffin and Walker, which did not consider the randomness of L.

As an alternative to this we just obtain all conditional posterior from the full likelihood given by equation (3.36) without marginalising out any variable. Hence the corresponding Gibbs sampler algorithm is based on the following conditional distributions

- $[v|\ldots] \sim v^{n-1} e^{-v(J_{S_L}+1-p)} \int_0^\delta \int_0^{L/p} \left[1-e^{-svp/w}\right] e^{-s/x^2} s^{-1} x^a ds dx.$
- $[\xi_i|\ldots] \propto \mathbb{1}(0\le \xi_i \le 1)\, g_z((X_i-\theta)/\xi_i; \epsilon_{c_i})/\xi_i$
- $[z|\ldots] \propto \frac{\pi_z(z) z^{nz}}{\Gamma(z)^n} e^{-z\sum_j S_j/\epsilon_j + z\sum_i \log((X_i-\theta)/\xi_i) - z\sum_j n_j \log \epsilon_j}$, where $S_j = \sum_{c_i=j}(X_i-\theta)/\xi_i$ and $n_j = \sum_{i=0}^n \mathbb{1}(c_i = j)$
- $[u_i|\ldots] \sim Unif(0, p_{c_i})$
- $Pr[c_i = j|\ldots] \propto \mathbb{1}(u_i \le p_j)\, g_z(X_i/\xi_i; \epsilon_j)$
- $[\epsilon_j^{(0)}|\ldots] \sim e^{-zS_j/\epsilon_j - vpJ_j\epsilon_j^2/w} \epsilon_j^{-(zn_j-a)} \mathbb{1}(0 \le \epsilon_j \le \delta)\, \mathbb{1}\left(\epsilon_j \ge \sqrt{\frac{w\bar{u}_j}{pJ_j}}\right),\ \bar{u}_j = \max\{u_i\}_{i:\ c_i=j}$
- $[\epsilon_j^{(1)}|\ldots] \sim e^{-zS_j/\epsilon_j} \epsilon_j^{-(zn_j+a)} \mathbb{1}(\epsilon_j > \delta)$
- $[J_j|\ldots] \sim \mathbb{1}\left(J_j \ge \frac{w\bar{u}_j}{p(\epsilon_j^{(0)})^2}\right) e^{-vp(\epsilon_j^{(0)})^2 J_j/w} e^{-J_j} J_j^{-1}$

- $[V_j^{(1)}|\ldots] \sim (1-V_j)^m \mathbb{1}(u_{l,j} \le V_j \le u_{u,j})$, where $u_{l,j} = \frac{\bar{u}_j}{(1-p)\prod_{k<j}(1-V_k)}$, $u_{u,j} = 1 - \frac{\underline{u}_j}{(1-p)V_k \prod_{l<k, l\neq j}(1-V_l)}$, $\underline{u}_j = \min\{u_i\}_{i:\ c_i=k>j}$

- $[w|\ldots] \sim e^{-w^{-1}v\sum_{j\in S_L} p(\epsilon_j^{(0)})^2 J_j} w^{m-1} e^{-w} \mathbb{1}(w \le \bar{w})$
$\times \int_0^\delta \int_0^{L/p} \left[1-e^{-svp/w}\right] e^{-s/x^2} s^{-1} x^a ds dx,\ \bar{w} = \min\left\{\frac{p\epsilon_{c_i}^2 J_{c_i}}{u_i}\right\}_{i=1,\ldots,n}$

- $[p|\ldots] \sim e^{pvR} p^{\alpha-1}(p_n - p)^{\beta-1} \mathbb{1}\left(\underline{p} \le p \le \bar{p}\right) \int_0^\delta \int_0^{L/p} \left[1-e^{-svp/w}\right] e^{-s/x^2} s^{-1} x^a ds dx$,
$R = 1 - \sum_{j\in S_L} \left(\epsilon_j^{(0)}\right)^2 J_j/w,\ \underline{p} = \max\left\{\frac{wu_i}{\epsilon_{c_i}^2 J_{c_i}}\right\}_{i:\epsilon_{c_i}\le\delta},\ \bar{p} = \min\left\{\frac{q_{c_i}^{(1)} - u_i}{q_{c_i}^{(1)}}\right\}_{i:\epsilon_{c_i}>\delta}$

- $[\theta|\ldots] \sim \prod_{i=1}^n (X_i - \theta)^{z-1} e^{\frac{z\theta}{\epsilon_{c_i}\xi_i}} \mathbb{1}\left(\theta \le X_{(1)}\right) \pi_\theta(\theta)$

We implemented in programming language R the first method, that worked very well in practice. In order to sample from non-standard conditional distributions we used rejection sampling. To find suitable envelope functions, for each of the expressions determining the conditionals, it was studied the number of modes and support depending on different values of the parameters. In all cases we have supports that are finite intervals or semi-lines and most of them define unimodal distributions. Thus, we used Gamma and truncated Gamma distributions as envelope functions. To define the value of the paramters of the envelope we imposed two conditions.

The first is that the modes coincide and that the envelope is equal to target density at that point. This can be done by finding the mode of the target conditional through the zeros of the derivative of the log of the density, and using the expression of the mode of a Gamma distribution. In most of the cases the zeros had analytical expressions, and only a couple had to be found numerically. This slows down each iteration of the sampler but in practice it was not critical. In some cases the mode was located at one of the boundary points of the support, but this is not a problem since it is also possible to construct (truncated) Gamma distributions with that characteristic.

The second condition to determine the parameters of the envelope Gamma distribution was to impose the definition of being an envelope, that is, being greater or equal to the target density. This gave us a condition for the parameters for each point in the support. More specifically, we used the condition on the modes to express the scale parameter in terms of the shape parameter, and plugging in that expression on the second condition we obtained a critical value for the shape parameter for each point of the support. Increasing the value of the shape parameter results in tighter envelopes and this critical value is the threshold for which at a certain point, target and envelope are equal and thus the shape parameter was chosen as the biggest possible that was below the threshold for all points in the support, or in other words, the minimum of the critical values throughout the support. Again, in order to find this minimum sometimes it was possible analytically and sometimes numerically.

We show an example. Recall the conditional for parameter z in the first method.

$$[z|\ldots] \propto \frac{\pi_z(z) z^{nz}}{\Gamma(z)^n} e^{-z\sum_j S_j/\epsilon_j + z\sum_i \log((X_i-\theta)/\xi_i) - z\sum_j n_j \log\epsilon_j},$$

where $S_j = \sum_{c_i=j}(X_i-\theta)/\xi_i$,and $n_j = \sum_{i=0}^n \mathbb{1}(c_i = j)$. Recall that in rejection sampling the normalising constant is irrelevant. Therefore the logarithm of the density can be expressed as

$$\ell(z) := -zV - c\sqrt{z} + (b/2-1)\log(z) + n(z\log(z) - \log\Gamma(z))$$

for some variable V that depends on the observations and other parameters. This function goes to $-\infty$ as z goes to 0 or infinity and has one maximum in between. Thus the mode can be found solving the equation

$$b - c\sqrt{z} + 2nz(1 + \log(z) - \Gamma'(z)/\Gamma(z)) - 2Vz - 2 = 0$$

since there is no closed expression for the solution we find it numerically. Let us denote such solution as μ, and define as α and β the shape and scale parameters of the envelope Gamma distribution. Since we impose equal modes we obtain the expression $\beta = (\alpha - 1)/\mu$ and the envelope function is $K_\mu + (\alpha - 1)\log z - \beta z$ with $K_\mu = \ell(\mu) - (\alpha - 1)\log\mu - \beta\mu$ so that the functions coincide at μ. Now imposing the envelope condition at a point z

$$(\alpha - 1)\log(z/\mu) - \beta(z - \mu) \geq \ell(z) - \ell(\mu)$$

replacing β and reorganising we obtain that the critical value for α at z, say $\tilde{\alpha}(z)$ is

$$\tilde{\alpha}(z) = 1 + (\ell(\mu) - \ell(z))/(\log(\mu/z) + (z/\mu - 1))$$

and exploring this function we realised that it is decreasing with a strictly positive limit at infinity equal to $1 + \mu(V - n)$. Hence this is the value of α that gives us the tightest envelope.

When there are more than one mode, the strategy is to use a mixture of Gammas and follow similar steps. In the following sections we show the outcomes of using this algorithm with simulated and real data.

3.4.2 Simulated Data

We simulate data from two different true distributions, Exp(2) and a mixture of $\Gamma(2.5, 1)$ with the uniform distribution (as in the prior), number of observations is $n = 100$ in both cases, and prior parameters are $a = b = 2$, $c = 1$, $d = 2$, $\delta = 1$, $m = 1$, $\alpha = \beta = 0.5$, $p_n = 0.99$, using the first sampler described in Section 3.4.1. We ran a total of 40000 iterations with a burn-in of 30000 and thinning of 5 to reduce autocorrelation. Figures 3.1 and 3.2 show posterior samples for f and normalised location parameter $n(\theta - X_{(1)})$ corresponding to data from $Exp(2)$ and Gamma-Uniform mixture respectively.

In the plots for f, red solid line represents the posterior mean and green solid line represents the true density. In the histograms, solid black line in plots for θ shows the (Negative) Exponential density that fit best the sample. We observe that in both cases the parameter of the Exponential distribution for (normalised) θ is close to $f_0(0)$ as it was expected from theoretical results.

For $Exp(2)$: the 95% credible interval for $f(0)$ is $(1.142, 2.925)$ (with true value being 2) and the fitted exponential for $n(\theta - X_{(1)})$ had rate 1.64 (true value is 2). For $\Gamma(2.5, 1)$-Uniform mixutre: the 95% credible interval for $f(0)$ is $(0.520, 0.878)$ (with true value being 2/3) and the fitted exponential for $n(\theta - X_{(1)})$ had rate 0.675 (true value is 2/3).

3.4.3 Auctions Data

3.4.4 Procurement auctions

In this section we apply the model we propose to the private procurement auctions. For a definition and discussion, see e.g. [28, 33]. Typical examples of private procurement auctions are where bids are invited by government for building roads, buildings etc. Consider observed bids from a procurement auction $b_1, \ldots, b_n$ from n bidders. For each bidder, given their value

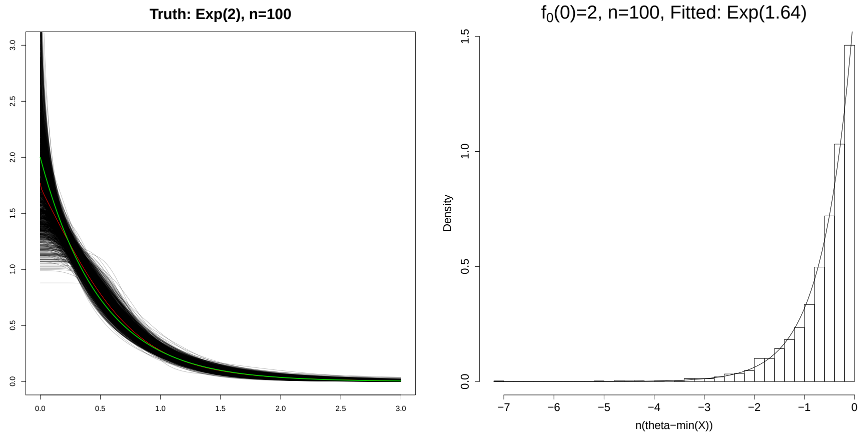

Figure 3.1: Samples from posterior distribution of f and $n(\theta - X_{(1)})$, for simulated data from $Exp(2)$ Distribution. $f_0(0) = 2$. Fitted Exponential distribution is shown in histogram for $n(\theta - X_{(1)})$. Red: Posterior mean; Green: Truth. Total iterations: 40000, Burn in: 30000, Thinning: 5, $m = 1$, $a = b = 2$, $c = 1$, $d = 2$.

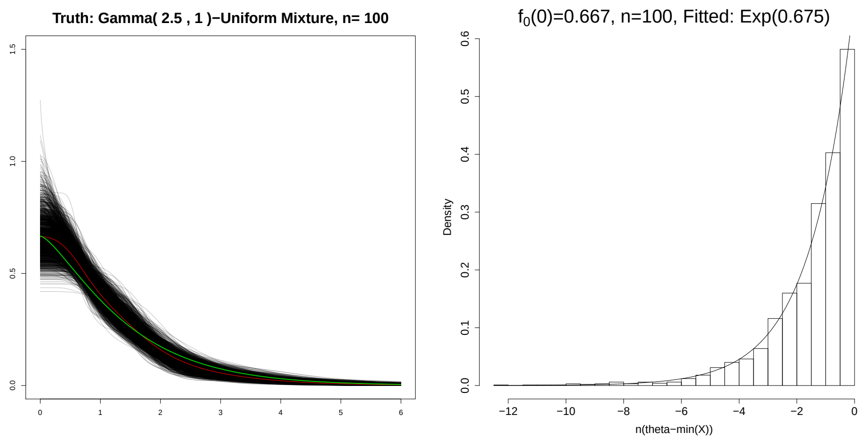

Figure 3.2: Samples from posterior distribution of f and $n(\theta - X_{(1)})$, for simulated data from $GammaUnif(2.5, 1)$-mixture. $f_0(0) = 2/3$. Fitted Exponential distribution is shown in histogram for $n(\theta - X_{(1)})$. Red: Posterior mean; Green: Truth. Total iterations: 40000, Burn in: 30000, Thinning: 5, $m = 1$, $a = b = 2$, $c = 1$, $d = 2$.

of the cost c, the optimal bidding strategy is the winning bid b (under the assumption of the Bayes-Nash equilibrium) given by

$$b(c) = c + \frac{\int_c^\infty (1 - G(x))^{n-1} dx}{(1 - G(c))^{n-1}}$$

where $C \sim G$ is the distribution of cost [28], and the costs of individual bidders are assumed to be independent. Examples of such distributions are shifted exponential $c + Exp(\lambda)$ with $b(c) = c + 1/[\lambda(n-1)]$ and Pareto distribution with $G(x;c) = (1 - (x/c)^{-a})I(x > c)$ with $b(c) = c(1 + \frac{1}{a(n-1)-1})$.

In practice, it is of interest to determine the actual cost c of given observed bids which is the lower bound of the support of G. Typically G is assumed known (see eg [11]). We address the case where both G and c are unknown.

In our setting, for the location LAE model, the denominator is 1, so the winning bid function becomes

$$b(c) = c + \int_c^\infty (1 - G(x;c))^{n-1} dx = c + \int_0^\infty (1 - G(x;0))^{n-1} dx. \tag{3.41}$$

Note that for the location model, (3.41) is easy to invert as the bid is a shifted cost, with the shift independent of c:

$$c(b) = b - \int_0^\infty (1 - G(x;0))^{n-1} dx. \tag{3.42}$$

In the model considered in the paper, for observed b_is, we estimate θ - the lower support point of the distribution of b, and the distribution of the bids

$$F_\theta(b) = F(b - \theta) = G(c^{-1}(b)) = G(b - \theta - s_{G,n}), \quad s_{G,n} = \int_0^\infty (1 - G(x;0))^{n-1} dx.$$

We estimate the density of F and θ, and we need to estimate $s_{G,n}$ to obtain the distribution of the cost.

Note that given F, shift $s = s_{G,n}$ satisfies the following equation:

$$s = \int_0^\infty (1 - F(x + s))^{n-1} dx. \tag{3.43}$$

3.4.5 Checking the identification assumption

Now we discuss the identification assumptions of Theorem 1 stated in [33] (supplement) that the probability of winning the bid $\pi_0(k)$ decreases in k hodls where

$$\pi_0(k) = \int (b_k(c) - c)(1 - G_0(c))^{k-1} g_0(c) dc$$

and $b_k(c)$ is the bid corresponding to cost c based on k bidders.

For a location model, the bid function is defined by (3.41) which implies that $b(c) - c$ is independent of c, and hence the probability of winning is

$$\pi_0(k) = k^{-1} \int_0^\infty (1 - G(x;0))^{k-1} dx$$

Differentiating the log for a continuous range of k, we have

$$\begin{aligned}\log \pi_0(k)' &= [-\log k + \log \int_0^\infty (1-G(x;0))^{k-1}dx]' \\ &= -1/k + [\int_0^\infty \log(1-G(x;0))(1-G(x;0))^{k-1}dx][\int_0^\infty (1-G(x;0))^{k-1}dx]^{-1} < 0\end{aligned}$$

i.e. it decreases.

For a scale LAE model,

$$b(c) = c + \frac{\int_c^\infty (1-G(x))^{n-1}dx}{(1-G(c))^{n-1}} = c + c\frac{\int_1^\infty (1-G(cy))^{n-1}dy}{(1-G(c))^{n-1}}$$

and the probability of winning the bid $\pi_0(k)$ is

$$\pi_0(k) = \int (1-G(y))^{k-1}dy \int I(y>c)g(c)dc = \int G(y)(1-G(y))^{k-1}dy.$$

Differentiating under the integral for a continuous range of k, we get

$$\frac{d}{dk}\pi_0(k) = -\int g(y)G(y)(1-G(y))^{k-2}dy$$

i.e. $\pi_0(k)$ decreases in k.

3.4.6 Distribution of the smallest bid given observed costs

Suppose we observe costs $c_1, \ldots, c_n$ and estimate their distribution $f_\theta(x)$.

Note that for all $M > 0$ large enough

$$\int_0^\infty (1-G(x;0))^{n-1}dx = \int_0^{M\log n/n} (1-G(x;0))^{n-1}dx + \int_{M\log n/n}^\infty (1-G(x))^{n-1}dx$$

with

$$\begin{aligned}\int_{M\log n/n}^\infty (1-G(x))^{n-1}dx &\le E_G(X)(1-G(M\log n/n))^{n-2} \\ &\le E_G(X)\left(1-g(M\log n/n)\frac{M\log n}{n}\right)^{n-2} \\ &\le E_G(X)e^{-Mg(M\log n/n)\log n} \\ &\le E_G(X)n^{-Mg_0(0)/2}\end{aligned}$$

for all G such that $\sup_{x\in[0,\delta]}|g(x)-g_0(x)| \le g_0(0)/2$. So if we can control $E_G(X)$ in A_{ϵ_n} so that

$$E_G(X) \le e^{[Mg_0(0)/2-2]\log n} \tag{3.44}$$

$$\int_0^\infty (1-G(x;0))^{n-1}dx = \int_0^{M\log n/n} (1-G(x;0))^{n-1}dx + O(n^{-2})$$

Now let $u = G(x;0)$, we have $du = g(x;0)dx = g_0(0)(1+o(1))dx$ uniformly on $[0, M\log n/n]$ so long as $M \log n/n = o(1)$ and

$$\int_0^{M \log n/n} (1 - G(x;0))^{n-1} dx$$
$$= \frac{1}{g_0(0)(1+o(1))} \int_0^{G(M \log n/n)} e^{-(n-1)u} du (1 + O(M^2 \log n^2/n))$$
$$= \frac{1}{ng_0(0)(1+o(1))}$$

and under (3.44)

$$b(c) = c + \frac{1}{ng_0(0)} + o(1/n)$$

uniformly on A_{ϵ_n} so that studying BvM for b is equivalent to studying BvM for c. By choosing $M_n = o(\sqrt{n}/\log n)$ all the conditions above are satisfied and as soon as

$$E^\Pi \left(\int_0^\infty x g(x) dx \right) < +\infty$$

and for all $\delta > 0$

$$\Pi(E_G(X) > e^{M_n \delta \log n}) \leq e^{-\delta M_n \log n} E^\Pi [\int_0^\infty x g(x) dx] = o(e^{-Cn\epsilon_n^2})$$

as soon as $\epsilon_n = o(n^{-1/4-\delta_0})$ for some $\delta_0 > 0$.

3.4.7 Asymptotic expression for the shift

We can see in Section 3.4.6, that asymptotically for large n, the equation for the shift $s = s_{G,n}$ (3.43) can be written as

$$s = 1/(nf_0(s)) + o(1/n).$$

Now we show that asymptotically, $s = s_{G,n} = \frac{1}{nf_0(0)}(1 + o(1))$.

As $f_0(x) \leq f_0(0)$ for all x, $s \geq 1/(nf_0(0)) + o(1/n)$. If $s > a$ (here $a > 0$ is a constant used in the definition of class $\mathcal{F}$) then the equation implies $f_0(s) < a/(n + o(1))$ for all natural n which contradicts the assumption that $f_0(0) = \gamma > 0$, i.e. is separated away from 0. Therefore, $s \in [1/(nf_0(0)), a]$. As $\sup_{x\in(0,a)} |f_0'(x)| \leq M$, this implies that

$$f_0(s) = f_0(0) - s(-f_0'(s)) \geq f_0(0) - sM$$

which implies that $s \leq 1/(f_0(0)n)(1 + o(1))$ which together with the first inequality implies that $s = 1/(f_0(0)n)(1 + o(1))$.

If $\sup_{x\in(0,a)} |f_0'(x)|/f_0(x) \leq M$, this implies that for some $s_0 \in (0, s)$,

$$f_0(s) = f_0(0) - s(-f_0'(s_0)) \geq f_0(0) - sMf_0(s_0) \geq f_0(0)(1 - sM)$$

which implies that $s \leq 1/(f_0(0)n)(1 + o(1))$ which together with the first inequality implies that $s = 1/(f_0(0)n)(1 + o(1))$.

3.4.8 Credible interval for the unknown cost

When $f = f_0$ is known, under a location LAE model, asymptotically, $(1-\alpha)100\%$ credible interval for its lower support point is $[X_{(1)} + \frac{\log(1-\alpha)}{nf_0(0)}, X_{(1)}]$.

Under the procurement auctions model, for known g_0, the credible interval for the cost is asymptotically given by

$$[X_{(1)} + \frac{\log(1-\alpha)}{nf_0(0)} - s_{G,n}, X_{(1)} - s_{G,n}],$$

as $f_0(0) = g_0(s_{G,n})$.

Under the conditions of marginal BvM for θ, $s_{G,n} \approx 1/(nf_0(0))$, and hence $f_0(0) = g_0(s_{G,n}) \approx g_0(0)$ as on A_{ϵ_n}, $\sup_{g,\,0\le h\le nu_n} |g(h) - g_0(0)| \le M\epsilon_n^{1/3}$, where $nu_n \to \infty$ as $n \to \infty$.

Therefore, the credible interval for $b(c)$ with unknown G asymptotically becomes

$$[X_{(1)} - \frac{\log(1-\alpha)+1}{ng_0(0)} + o(1/n), X_{(1)} - \frac{1}{ng_0(0)} + o(1/n)]$$

i.e. the price to pay for unknown G is of the order of the estimation error of θ_0.

3.4.9 Application to California Transport procurement auctions

We apply the model to the bids observed in Californian state procurement auctions of highway and street maintenance projects carried out by the California Department of Transportation (Caltrans) between January 2002 and December 2005. The data are freely available on the Caltrans website, and were studied by [33]. Here θ is the bid corresponding to the true cost of the project. In these auctions, small companies had a preferential treatment and hence a potentially different distribution of bids from the large companies participating in the same auction. Therefore, here small and large companies are analysed separately.

We analysed bids by small companies in Auction 128 ($n = 7$), and bids by large companies in auction 438 ($n = 6$), with and without log transform, i.e. using scale and location models, keeping same parameters used with simulated data. These are two of the auctions with highest number of bids that have histograms suggesting a decreasing density function (Figure 3.3).

For Auction 128, the draws from the posterior density of f_θ and the marginal posterior distribution of θ are given in Figure 3.4 for the location model, and under the log transform in Figure 3.5. The 95% credible intervals for θ_0 and for $f_0(\theta_0)$ under both models are given in Table 3.1. We can see that both densities have high uncertainty around the lower support point which is reflected in fairly wide credible intervals for $f_{0,\theta_0}(\theta_0)$ however the marginal distribution of θ_0 is approximated very well by and exponential distribution, despite a relative small number of observations. This may be due to a faster rate of contraction of the posterior.

Uncertainty about θ_0 is higher under the scale model than under the location model. This may be due to a simpler estimation process for $f_{0,\theta_0}(\theta_0)$ under the location model (as $f(0)$) compared to the one under the log transform and then back to the original scale ($f_{0,\theta_0}(\theta_0) = \theta_0^{-1} f_0(\cdot/\theta_0)$ is estimated as $e^{-\tau} g(0)$ for the corresponding location model $g(\cdot - \tau)$).

Draws from the posterior densities of the bids with and without log transform for Auction 438 are given in Figures 3.7 and 3.6, respectively, and the credible intervals for θ_0 and $f_{0,\theta_0}(\theta_0)$ are given in Table 3.2. The uncertainty of $f_{0,\theta_0}(\theta_0)$ is larger for both models compared to that for Auction 128, otherwise the conclusions are similar.

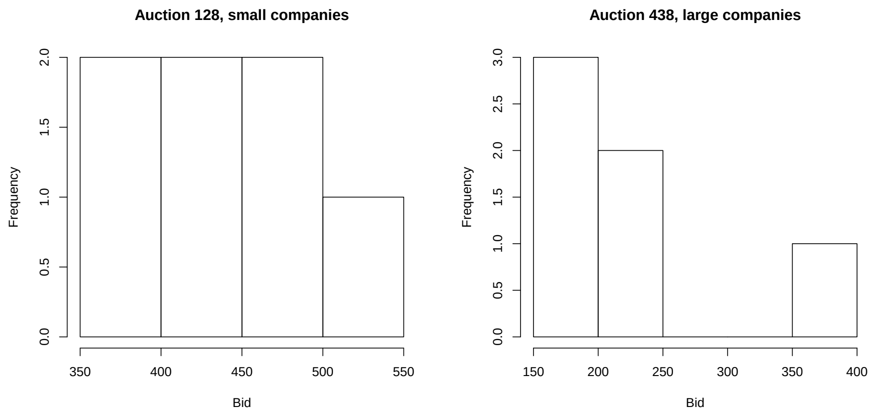

Figure 3.3: Histograms for Auction 128 (small companies) and Auction 438 (large companies), CalTrans.

	Scale model	Location model
95% CI for θ_0	$[2.809, 3.697]$	$[3.275, 3.697]$
95% CI for $f_{0,\theta_0}(\theta_0)$	$[0.262, 5.658]$	$[0.495, 6.596]$

Table 3.1: Auction 128, small companies, bids divided by 10^5.

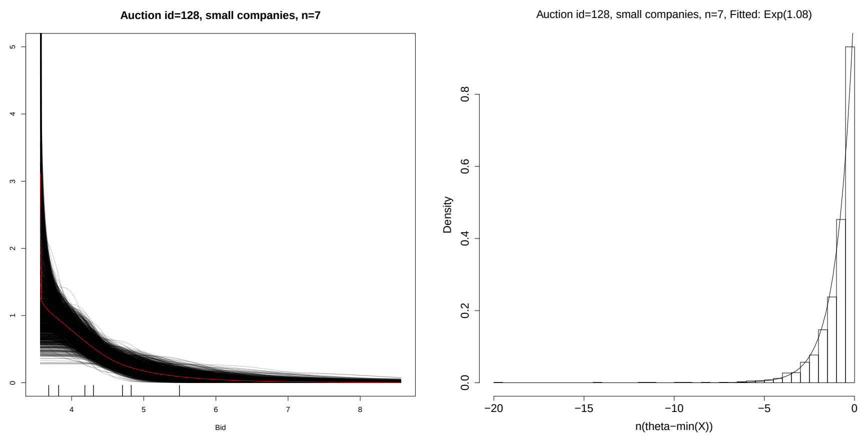

Figure 3.4: Auction 128, small companies, location model. Left: Posterior distribution of bids, black lines represent the draws from the posterior distribution of f, red line is posterior mean; lower support point is the posterior mean of θ. Right: Posterior distribution of the smallest bid (recentred by the smallest bid and rescaled by sample size).

	Scale model	Location model
95% CI for θ_0	$[1.288, 1.812]$	$[1.392, 1.812]$
95% CI for $f_0(\theta_0)$	$[0.299, 20.887]$	$[1.030, 8.712]$

Table 3.2: Auction 438, large companies, bids divided by 10^5.

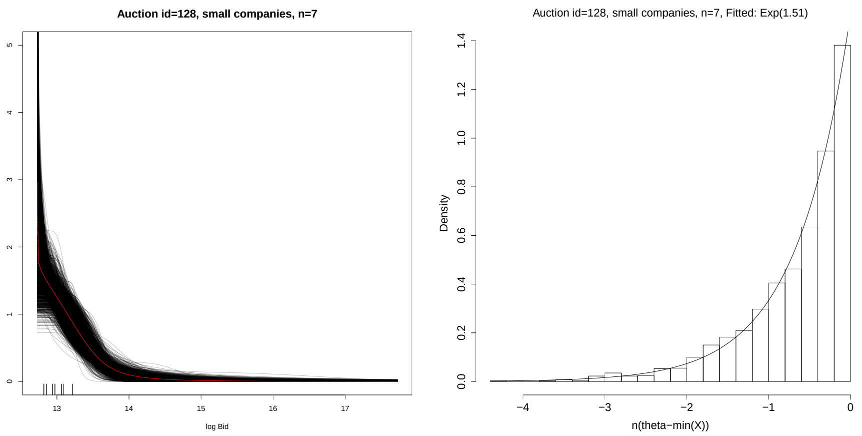

Figure 3.5: Auction 128, small companies, scale model. Left: Draws from the posterior distribution of $g(x) = f(\log x)/x$, red line is posterior mean; lower support point is the posterior mean of $\log \theta$. Right: Posterior distribution of the log of the smallest bid (recentred by the smallest bid and rescaled by sample size).

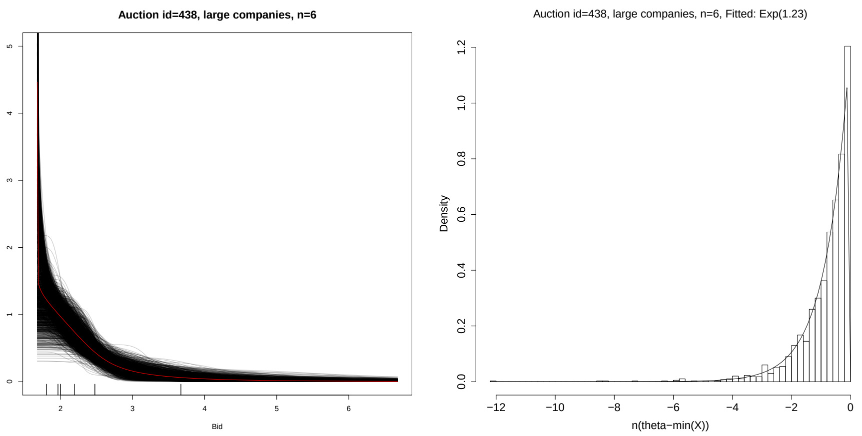

Figure 3.6: Auction 438, large companies, location model. Left: Posterior distribution of bids, black lines represent the draws from the posterior distribution of f, red line is posterior mean; lower support point is the posterior mean of θ. Right: Posterior distribution of the smallest bid (recentred by the smallest bid and rescaled by sample size).

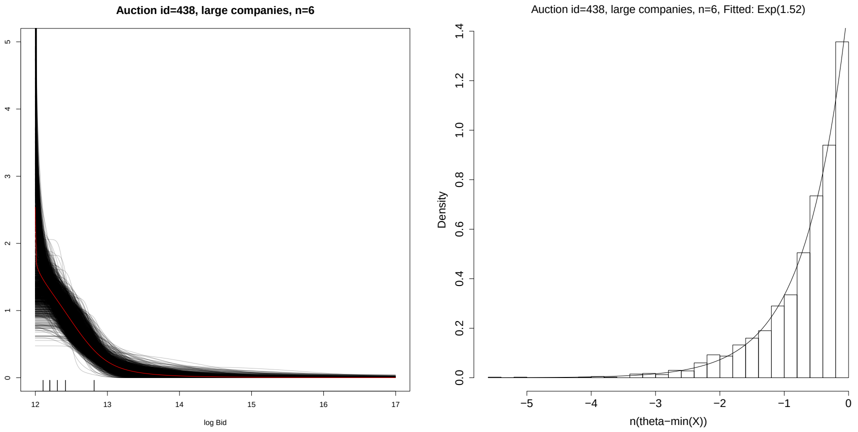

Figure 3.7: Auction 438, large companies, scale model. Left: Draws from the posterior distribution of f, red line is posterior mean; lower support point is the posterior mean of θ. Right: Posterior distribution of the smallest bid (recentred by the smallest bid and rescaled by sample size).

3.5 Proofs

3.5.1 Proof of Proposition 3

Condition (i) implies that if $h = n(\theta - \theta_0)$ and π_n is the posterior density of h, then

$$\begin{aligned}\int |\pi_n(h) - \frac{\gamma_0 e^{\gamma_0 h}\mathbf{1}_{h\le\zeta_n}}{e^{\gamma_0\zeta_n}-1}|dh &\le \int_{|h|\le n\varepsilon_n} |\pi_n(h) - \gamma_0 e^{-\gamma_0 h}\mathbf{1}_{h\ge 0}|dh \\ &\quad + \Pi(|\theta-\theta_0| > \varepsilon_n|\mathbf{X}^n) + e^{-n\gamma_0\varepsilon_n} \\ &= \int_{|h|\le n\varepsilon_n} |\pi_n(h) - \gamma_0 e^{-\gamma_0 h}\mathbf{1}_{h\ge 0}|dh + o_{P_0}(1).\end{aligned}$$

so that we can restrict ourselves to $|h| \le n\varepsilon_n$. Moreover for all $|h| \le n\varepsilon_n$

$$\begin{aligned}\pi_n(h) = &\frac{n\int \mathbf{1}_{\eta\in A_{\epsilon_n}} e^{\ell_n(\theta_0+h/n,\eta)-\ell_n(\theta_0,\eta_0)}\Pi_\eta(d\eta)}{\int \mathbf{1}_{(\eta,\theta')\in A_{\epsilon_n}} e^{\ell_n(\theta',\eta)-\ell_n(\theta_0,\eta_0)}\Pi_\eta(d\eta)\pi(\theta')d\theta'}\Pi(A_{\epsilon_n}|\mathbf{X}^n) \\ &+ \frac{n\int \mathbf{1}_{\eta\in A^c_{\epsilon_n}} e^{\ell_n(\theta_0+h/n,\eta)-\ell_n(\theta_0,\eta_0)}\Pi_\eta(d\eta)}{\int e^{\ell_n(\theta',\eta)-\ell_n(\theta_0,\eta_0)}\Pi_\eta(d\eta)\pi(\theta')d\theta'}\end{aligned}$$

where $A_{\epsilon_n} = \{\eta : d(\eta, \eta_0) \le \epsilon_n\}$ so that

$$
\begin{aligned}
&\int |\pi_n(h) - \gamma_0 e^{-\gamma_0 h} \mathbf{1}_{h\ge 0}| dh \\
&\le \int_{|h|\le n\varepsilon_n} \left| \frac{n \int \mathbf{1}_{\eta\in A_{\epsilon_n}} e^{\ell_n(\theta_0+h/n,\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)}{\int \mathbf{1}_{(\eta,\theta')\in A_{\epsilon_n}} e^{\ell_n(\theta',\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)\pi(\theta')d\theta'} - \gamma_0 e^{-\gamma_0 h} \mathbf{1}_{h\ge 0} \right| dh \\
&\quad + \Pi(|\theta - \theta_0| > \varepsilon_n | \mathbf{X}^n) + 2\Pi(A^c_{\epsilon_n} | \mathbf{X}^n) + o(1) \\
&= \int_{|h|\le n\varepsilon_n} \left| \frac{n \int \mathbf{1}_{\eta\in A_{\epsilon_n}} e^{\ell_n(\theta_0+h/n,\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)}{\int \mathbf{1}_{(\eta,\theta')\in A_{\epsilon_n}} e^{\ell_n(\theta',\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)\pi(\theta')d\theta'} - \gamma_0 e^{-\gamma_0 h} \mathbf{1}_{h\ge 0} \right| dh + o(1).
\end{aligned}
$$

Denote

$$
\bar{\pi}_n(h) = \frac{n \int \mathbf{1}_{\eta\in A_{\epsilon_n}} e^{\ell_n(\theta_0+h/n,\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)}{\int \mathbf{1}_{(\eta,\theta')\in A_{\epsilon_n}} e^{\ell_n(\theta',\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)\pi(\theta')d\theta'}.
$$

Note that by continuity and positivity of the prior density π_θ at θ_0, $\pi(\theta_0 + h/n) = \pi(\theta_0)(1 + o(1))$ where the term $o(1)$ is uniform over $|h| \le n\epsilon_n$ and

$$
\bar{\pi}_n(h) = \frac{(1 + o(1)) \int e^{\ell_n(\theta_0+h/n,\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)}{\int \mathbf{1}_{|h'|\le n\varepsilon_n} \int \mathbf{1}_{\eta\in A_{\epsilon_n}} e^{\ell_n(\theta_0+h'/n,\eta)-\ell_n(\theta_0,\eta_0)} \Pi_\eta(d\eta)dh'}.
$$

We consider the following expansion of the log-likelihood ratio:

$$
\begin{aligned}
\ell_n(\eta, \theta_0 + h/n) - \ell_n(\eta_0, \theta_0) &= \ell_n(\eta_0, \theta_0 + h/n) - \ell_n(\eta_0, \theta_0) + J_n(\theta), \qquad (3.45) \\
e^{J_n(\theta)} &= \int_H e^{\ell_n(\eta,\theta_0+h/n)-\ell_n(\eta_0,\theta_0+h/n)} d\Pi_\eta(\eta).
\end{aligned}
$$

The first term, $\ell_n(\eta_0, \theta) - \ell_n(\eta_0, \theta_0)$, represents log likelihood ratio for a parametric model with unknown θ and known η_0, and we can write

$$
e^{J_n(\theta)} = \int_H e^{\ell_n(\eta,\theta_0)-\ell_n(\eta_0,\theta_0)+\Delta_n(\eta,\theta)} d\Pi_\eta(\eta).
$$

Under assumption (3.5) , we have that

$$
J_n(\theta_0 + h/n) = J_n(\theta_0)(1 + o_{P_0}(1 + h))
$$

uniformly on $|h| \le n\varepsilon_n$ so that

$$
\begin{aligned}
\bar{\pi}_n(h) &= \frac{(1 + o_{P_0}(1)) e^{\ell_n(\theta_0+h/n,\eta_0)-\ell_n(\theta_0,\eta_0)} J_n(\theta_0)}{\int \mathbf{1}_{|h'|\le n\epsilon_n} e^{\ell_n(\theta_0+h'/n,\eta_0)-\ell_n(\theta_0,\eta_0)} J_n(\theta_0) dh'} \\
&= \frac{(1 + o_{P_0}(1)) e^{\ell_n(\theta_0+h/n,\eta_0)-\ell_n(\theta_0,\eta_0)}}{\int \mathbf{1}_{|h'|\le n\varepsilon_n} e^{\ell_n(\theta_0+h'/n,\eta_0)-\ell_n(\theta_0,\eta_0)} dh'},
\end{aligned}
$$

where we can use the parametric LAE Bernstein von-Mises result from [30], Theorem V.5.1, so that

$$
\bar{\pi}_n(h) = \frac{(1 + o_{P_0}(1)) \gamma_{\theta_0} e^{\gamma_{\theta_0} h(1+o_{P_0}(1))} \mathbf{1}_{h\le\zeta_n}}{e^{\gamma_{\theta_0}\zeta_n} - 1}
$$

which terminates the proof.

3.5.2 Proof of Theorem 7

To prove Theorem 7, we verify assumptions (i)-(iv) of Proposition 3. We first prove (i) holds for $A_n = \{d_H(f_0, f) \le \epsilon_n, |\theta - \theta_0| \le M_n/n\}$ where ϵ_n is given in condition (H1) and $M_n = M_0 \log n$ where $M_0 > 0$ is a constant large enough.

To do that we first note that $\Pi(B_n|\mathbf{X}^{(n)}) = 1 + o_P(1)$ uniformly over $\mathcal{F}_1 \times \Theta_0$, where $B_n = \{d_H(f_0, f) \le \epsilon_n, |\theta - \theta_0| \le 8\epsilon_n^2/f_0(0)\}$, due to Lemma 9. Moreover note that

$$\Pi(\theta > X_{(1)}|\mathbf{X}^{(n)}) = 0, \quad P_0(X_{(1)} > \theta_0 + M_n/n) \le f_0(0)/M_n = o(1)$$

We now prove that $\Pi(\theta \le \theta_0 - M_n/n|\mathbf{X}^{(n)}) = o_P(1)$, using an adaptation of Lemma 4.3 of [32]. We have, writing $\tilde{B}_n = \{f \in \mathcal{F}_n; d_H(f_0, f) \le \epsilon_n\}$ and

$$S_n(\theta) = \int_{d_H(f,f_0)\le\epsilon_n} e^{\ell_n(\theta,f)-\ell_n(\theta_0,f_0)} d\Pi_f(f),$$

and uniformly over $\mathcal{F}_1 \times \Theta_0$,

$$\begin{aligned}\Pi(\theta \le \theta_0 - M_n/n|\mathbf{X}^{(n)}) &= \Pi(\{\theta \le \theta_0 - M_n/n\} \cap B_n|\mathbf{X}^{(n)}) + o_{P_0}(1)\\ &\le \Pi(\theta \le \theta_0 - M_n/n|B_n, \mathbf{X}^{(n)}) + o_{P_0}(1)\\ &= \frac{\int_{\theta_0 - C_0\epsilon_n^2 \le \theta \le \theta_0 - M_n/n} S_n(\theta)d\theta}{\int_{\theta_0 - C_0\epsilon_n^2 \le \theta \le \theta_0 + M_n/n} S_n(\theta)d\theta}(1 + o(1)) + o_{P_0}(1).\end{aligned}$$

Let

$$\Omega_n(\theta_0) = \{\sup_{\theta \le \theta_0 - M_n/n} \sup_{f \in \tilde{B}_n} \ell_n(\theta, f) - \ell_n(\theta_0, f) \le -C_1 M_n\}$$

and

$$\Omega_n^{'}(\theta_0) = \{\inf_{\theta_0 \le \theta \le \theta_0 + \epsilon^2/n} \inf_{f \in \tilde{B}_n} \ell_n(\theta, f) - \ell_n(\theta_0, f) \ge 0 \},$$

with $\epsilon < 1/f_0(0)$. Note that $P_{f_0,\theta_0}(\Omega_n(\theta_0)) = P_{f_0,0}(\Omega_n(0))$ and similarly with $\Omega_n^{'}$ and we show in Lemma 18 that

$$P_{f_0,0}(\Omega_n(0)^c) = o(1), \quad P_{f_0,0}((\Omega_n^{'})(0)^c) = o(1)$$

uniformly over $\mathcal{F}_1$, where $C_1 > 0$ is a fixed constant depending only on $\mathcal{F}_1$, if ϵ is small enough. Then on $\Omega_n(\theta_0)$, for all $\theta_0 - C_0\epsilon_n^2 \le \theta \le \theta_0 - M_n/n$,

$$S_n(\theta) \le S_n(\theta_0)e^{-C_1 M_n}$$

and on $\Omega_n^{'}(\theta_0)$, for all $\theta_0 \le \theta \le \theta_0 + \epsilon^2/n$,

$$S_n(\theta) \ge S_n(\theta_0)$$

so that

$$\int_{\theta_0 - C_0\epsilon_n^2 \le \theta \le \theta_0 + M_n/n} S_n(\theta)d\theta \ge \frac{\epsilon^2}{n}$$

and on $\Omega_n^{'} \cap \Omega_n(\theta_0)$

$$\Pi(\theta \leq \theta_0 - M_n/n|\mathbf{X}^{(n)}) \leq \frac{e^{-C_1 M_n} n}{\epsilon^2} + o_p(1) = o_p(1)$$

as soon as $M_n \geq 2\log n/C_1 := M_0 \log n$. This proves that (i) holds with $A_n = \{(f,\theta), \in \mathcal{F}_n, d_H(f_0, f) \leq \epsilon_n, |\theta - \theta_0| \leq M_n/n\}$. Note that (ii) is proved in [30] and that (iii) is a consequence of (H3). We therefore need only prove that (iv) holds.

We now study $\Delta_n(f,\theta)$ defined in (3.4), using the change of variable $Y_i = X_i - \theta_0 \sim f_0$ under f_{0,θ_0} and defining $\mathbb{P}_n^y$ the empirical measure associated to $Y_i, i \leq n$ and $\mathbb{G}_n^y$ its centred version scaled by $\sqrt{n}$. Let $h = n(\theta_0 - \theta)$, then $|h| \leq M_n$.

$$\begin{aligned}
|\Delta_n(f,\theta)| &= |n\mathbb{P}_n^y\left[\log f(\cdot + h/n) - \log f(\cdot) - \log f_0(\cdot + h/n) + \log f_0(\cdot)\right]| \\
&\leq |h| \left|\mathbb{P}_n^y\left[\frac{f'}{f} - \frac{f_0'}{f_0}\right]\right| \\
&+ |h| \sup_{|u|\leq M_n} \sup_{y\geq 0} \left|\frac{f'(y+u)}{f(y+u)} - \frac{f_0'(y+u)}{f_0(y+u)} - \frac{f'(y)}{f(y)} + \frac{f_0'(y)}{f_0(y)}\right|.
\end{aligned}$$

From the second part of condition (H4), the second part of the right hand side is bounded by $|h|M(y)\epsilon$ uniformly over $f \in A_n$. Moreover

$$\mathbb{P}_n^y\left[\frac{f'(\cdot)}{f(\cdot)} - \frac{f_0'(\cdot)}{f_0(\cdot)}\right] = \int_0^\infty \left[\frac{f'(y)}{f(y)} - \frac{f_0'(y)}{f_0(y)}\right] f_0(y)dy + n^{-1/2}\mathbb{G}_n^y\left[\frac{f'(\cdot)}{f(\cdot)} - \frac{f_0'(\cdot)}{f_0(\cdot)}\right].$$

Note that

$$n^{-1/2}\mathbb{G}_n^y\left[\frac{f_0'(\cdot)}{f_0(\cdot)}\right] = \frac{1}{n}\sum_{i=1}^n \frac{f_0'(Y_i)}{f_0(Y_i)} - f_0(0)$$

converges almost surely to 0 due to $\mathbb{P}_0(|f_0'(\cdot)|/f_0(\cdot)) = f_0(0) < \infty$ [17]. This convergence is uniform over the class of functions where $f_0(0)$ is uniformly bounded above (condition (H1)), following the proof given in [17]. The first part of (H4) implies that

$$n^{-1/2}\mathbb{G}_n^y\left[\frac{f'(\cdot)}{f(\cdot)}\right] = o_{P_0}(1)$$

uniformly. Moreover

$$\begin{aligned}
\left|\int_0^\infty \left[\frac{f'(y)}{f(y)} - \frac{f_0'(y)}{f_0(y)}\right] f_0(y)dy\right| &\leq |f_0(0) - f(0)| + \left|\int_0^\infty \frac{f'(y)}{f(y)}(f_0(y) - f(y))dy\right| \\
&\leq |f_0(0) - f(0)| + \left\|\frac{f'(\cdot)}{f(\cdot)}\right\|_\infty d_H^2(f_0,f) + 2\left\|\frac{f'(\cdot)}{f(\cdot)}\right\|_\infty^{1/2} d_H(f_0,f)\sqrt{f(0)} = o(1)
\end{aligned}$$

on A_n. We thus have that $|\Delta_n(f,\theta)| = o_{P_0}(|h|)$ uniformly over A_n and Theorem 3.5.2 is proved.

3.5.3 Marginal Concentration in L^1 norm

Lemma 9. *Consider the shift LAE model with $f_{\eta,\theta}(x) = f(x-\theta) := f_\theta(x)$ where $f \in \mathcal{F}$. Assume that $f_0 \in \mathcal{F}_0(a, M, u)$ and that $\theta_0 \in \mathbb{R}$. If $d_H(f_\theta, f_{0,\theta_0}) \leq \epsilon_n$, for any $\epsilon_n \leq f_0(0)/(2M) \wedge a$, then,*

$$|\theta - \theta_0| \leq 8\epsilon_n^2/f_0(0) \leq 8\epsilon_n^2/u, \quad \textit{and} \quad d_H(f, f_0) \leq \epsilon_n. \tag{3.46}$$

Proof of Lemma 9. First let $\theta > \theta_0$. Using the change of variables $y = x - \theta_0$ and noting that $f(y - \theta + \theta_0) = 0$ for $y \in (0, \theta - \theta_0)$ we write

$$\begin{aligned}
d_H^2(f_{0,\theta_0}, f_\theta) &= \int_0^{\theta-\theta_0} f_0(y)dy + \int_{\theta-\theta_0}^{\infty} \left(\sqrt{f_0}(y) - \sqrt{f}(y - \theta + \theta_0)\right)^2 dy \\
&\geq 2 - 2\int_{\theta-\theta_0}^{\infty} \sqrt{f_0}(y)\sqrt{f}(y - \theta + \theta_0)dy \\
&\geq 2 - 2\int_{\theta-\theta_0}^{\infty} \sqrt{f_0}(y - \theta + \theta_0)\sqrt{f}(y - \theta + \theta_0)dy \\
&= d_H^2(f_0, f).
\end{aligned}$$

The same argument holds if $\theta \leq \theta_0$ so that

$$d_H^2(f_0, f) \leq d_H^2(f_{0,\theta_0}, f_\theta).$$

Moreover

$$d_H^2(f_{0,\theta}, f_{0,\theta_0}) \leq 2d_H^2(f_0, f) + 2d_H^2(f_{0,\theta_0}, f_\theta) \leq 4\epsilon_n^2.$$

Without loss of generality we can assume that $\theta > \theta_0$ and

$$d_H^2(f_{0,\theta}, f_{0,\theta_0}) \geq \int_0^{\theta-\theta_0} f_0(x)dx = F_0(\theta - \theta_0)$$

Since f_0' exists and is bounded by $M > 0$ on $(0, a)$ for some $a, M > 0$, by choosing $\epsilon_0 < a$ satisfying $M\epsilon_0 \leq f_0(0)/2$, $f_0(\epsilon_0) \geq f_0(0)/2$ and if $\epsilon_n^2 \leq \epsilon_0$, $\theta - \theta_0 \leq F_0^{-1}(4\epsilon_n^2) \leq \frac{8\epsilon_n^2}{f_0(0)}$ which concludes the proof. ☐

3.5.4 Joint posterior concentration rate

Proof of Proposition 5. We prove the theorem by verifying conditions of Theorem 2.1 of [21]. Note that we are under the conditions of Theorem 8 which is proved by verifying the conditions of this theorem for $d_H(f, f_0)$ stated in Lemmas 13 and 17. Without loss of generality, consider the case $\theta_0 = 0$.

Define the following neighbourhoods of f and θ:

$$\begin{aligned}
S_n &:= \{f : KL(f_0, f) \leq \epsilon_n^2; \quad V(f_0, f) \leq M\epsilon_n^2\}, \\
\Omega_n &:= \{\theta : -\delta_n \leq \theta \leq 0\}
\end{aligned}$$

with $\delta_n \to 0$ and $M_n \to \infty$ as $n \to \infty$. By Lemma 17,

$$\Pi(f \in S_n) \geq C_p e^{-\varkappa n^{1/(2\beta+1)}(\log n)^{2q}}, \quad \text{and } \Pi(\theta \in \Omega_n) \geq \min(\pi_\theta(0), \pi_\theta(-\delta_n))\delta_n$$

if the prior on θ satisfies the assumptions of the proposition where $\epsilon_n = \epsilon_0 n^{-\beta/(2\beta+1)}(\log n)^q$, with q defined in Theorem 8.

The following upper bound on $\log f(0)$ holds under the considered prior:

$$\log[\delta p_n/(1-p_n) + 1/\delta] \leq \tilde{M}_n$$

under the assumptions of the proposition.

By Lemma 10, under stated assumptions,

$$\mathcal{KL}(f_0, f_\theta) = O(\epsilon_n^2 + \tilde{M}_n \delta_n), \quad \mathcal{V}(f_0, f_\theta) = O(\epsilon_n^2 + \tilde{M}_n^2 \delta_n).$$

Hence, it is sufficient to take $\delta_n = O(\epsilon_n^2/\tilde{M}_n^2)$ which tends to 0 by the assumption of the proposition. This implies that the prior probability of the event

$$\mathcal{KL}(f_0, f_\theta) = O(\epsilon_n^2), \quad \mathcal{V}(f_0, f_\theta) = O(\epsilon_n^2)$$

is bounded from below by

$$\Pi(f \in S_n \,\&\, \theta \in \Omega_n) \geq C e^{-\varkappa n \epsilon_n^2} \delta_n \geq C e^{-0.5 \varkappa n \epsilon_n^2}$$

for large enough n, for any δ_n such that $\delta_n = O(\epsilon_n^2/\tilde{M}_n^2)$ and $\delta_n \geq e^{-0.5\varkappa n\epsilon_n^2}$. Since $\tilde{M}_n \leq [\log n]^A$ for some $A \geq 0$, we can take e.g. $\delta_n = Cn^{-1}$ which satisfies both conditions.

The entropy condition for the prior for f is verified in Lemma 13 due to assumptions (3.18), and the lemma, together with Proposition 7, imply the entropy condition for the prior for f_θ. The assumption of the proposition that

$$\tilde{M}_n = \log M_n \leq \log(4) + n\epsilon_n^2(\kappa - 1)/\kappa + 2\log \epsilon_n$$

holds since $\tilde{M}_n \leq m_1 [\log n]^A$.

This completes the proof.

□

Proposition 7. *Suppose that probability density $f(\cdot - \theta)$ is such that*

1. $f \sim \Pi$ *satisfying* $f(x) = \int_x^\infty y^{-1} g(y) dy$ *and* $f(0) \leq M_n$,
2. $\theta \sim \pi(\theta)$ *- proper prior, independent of f, satisfying for some $\kappa > 1$ and for large $\theta > 0$:*

$$\pi((-\infty, -\theta)) \lesssim \theta^{-\kappa}, \quad \pi((\theta, \infty)) \lesssim \theta^{-\kappa}.$$

If $M_n \leq 4e^{n\epsilon_n^2(\kappa-1)/\kappa}\epsilon_n^2$ and there exists a sieve Q for prior Π of f such that

$$\Pi(Q^c) \leq ce^{-n\epsilon_n^2} \quad \& \quad \log N(\zeta\epsilon_n, Q, d_H) \leq Cn\epsilon_n^2$$

with $\epsilon_n = n^{-\gamma}[\log n]^q$, then

$$\Pi(\tilde{Q}^c) \leq (c+1)e^{-n\epsilon_n^2} \quad \& \quad \log N((\zeta+1)\epsilon_n, \tilde{Q}, d_H) \leq (C+1)n\epsilon_n^2,$$

where $\tilde{Q} = \{f(\cdot - \theta) : f \in Q \,\&\, \theta \in [-B, B]\}$ with $B = 2\exp\{n\epsilon_n^2/\kappa\}\epsilon_n^2$.

Proof of Proposition 7. We need A, B such that $\pi((-\infty, A)) \leq \exp\{-n\epsilon_n^2\}/2$, $\pi((B,\infty)) \leq \exp\{-n\epsilon_n^2\}/2$. Consider a finite sieve $\hat{\Theta} = \{A + hs,\, s = 1, \ldots, S = \lceil (B-A)/h \rceil\}$ with h such that $\log\lceil (B-A)/h \rceil \leq n\epsilon_n^2$.

Then, for a $\hat{f} \in \hat{Q}$, a $\zeta\epsilon_n$ net of Q, $f \in Q$, $\theta \in [A, B]$ and $\hat{\theta} \in \hat{\Theta}$ (assuming $\hat{\theta} \geq \theta$ without loss of generality),

$$\begin{aligned} ||f(\cdot - \theta) - \hat{f}(\cdot - \hat{\theta})||_1 &\leq ||f(\cdot - \theta) - \hat{f}(\cdot - \theta)||_1 + ||\hat{f}(\cdot - \theta) - \hat{f}(\cdot - \hat{\theta})||_1 \\ &\leq \zeta\epsilon_n + \int_{\hat{\theta}-\theta}^{\infty} |\hat{f}(x) - \hat{f}(x + \theta - \hat{\theta})|dx \\ &\leq \zeta\epsilon_n + \int_0^{\hat{\theta}-\theta} \hat{f}(x)dx - (\hat{\theta} - \theta)\int_{\hat{\theta}-\theta}^{\infty} dx \int_{x-(\hat{\theta}-\theta)}^{x} g(y)/ydy \\ &\leq \zeta\epsilon_n + 2\hat{f}(0)h. \end{aligned}$$

For the Hellinger distance,

$$\begin{aligned} d_H(f(\cdot - \theta), \hat{f}(\cdot - \hat{\theta})) &\leq d_H(f(\cdot - \theta), \hat{f}(\cdot - \theta)) + d_H(\hat{f}(\cdot - \theta), \hat{f}(\cdot - \hat{\theta})) \\ &\leq \zeta\epsilon_n + \sqrt{||\hat{f}(\cdot - \theta) - \hat{f}(\cdot - \hat{\theta})||_1} \\ &\leq \zeta\epsilon_n + \sqrt{2\hat{f}(0)h}, \end{aligned}$$

i.e. we need $h = [0.5\epsilon_n/M]^2$, since for the considered prior $\hat{f}(0) \leq M$.

Hence, it is sufficient to take A, B such that $\pi((-\infty, A)) \leq \exp\{-n\epsilon_n^2\}/2$, $\pi((B, \infty)) \leq \exp\{-n\epsilon_n^2\}/2$, $B \leq 2e^{n\epsilon_n^2}(\epsilon_n/M_n)^2$, $A \geq -2e^{n\epsilon_n^2}(\epsilon_n/M_n)^2$. As the conditions are symmetric in B and $-A$, we take $A = -B$.

As $\pi((-\infty, -\theta)) \leq C\theta^{-\kappa}$ and $\pi((\theta, \infty)) \leq C\theta^{-\kappa}$ for some $\kappa > 1$ for large $\theta > 0$, the conditions above can be written as

$$\pi((-\infty, A)) \leq C(-A)^{-\kappa} \leq \exp\{-n\epsilon_n^2\}/2, \quad \pi((B, \infty)) \leq CB^{-\kappa} \leq \exp\{-n\epsilon_n^2\}/2,$$

i.e. we need $-A = B \geq \exp\{n\epsilon_n^2/\kappa\}/2$ and $B \leq 2e^{n\epsilon_n^2}(\epsilon_n/M_n)^2$. These conditions are compatible if

$$M_n \leq 4e^{n\epsilon_n^2(\kappa-1)/\kappa}\epsilon_n^2 \tag{3.47}$$

hence the proposition is proved.

□

Lemma 10. *Assume that f, f_0 are decreasing density functions on $[0, \infty)$, $f_0 \in \mathcal{P}(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu)$, and $f(0) < \infty$.*

Then, $\mathcal{KL}(f_0, f(\cdot - \theta)) = \mathcal{V}(f_0, f(\cdot - \theta)) = \infty$ if $\theta > 0$, and if $-\Delta \leq \theta \leq 0$, then

$$\mathcal{KL}(f_0, f_\theta) \leq \mathcal{KL}(f_0, f) + (\log f(0))_+|\theta|[C_0 + 2C_1],$$
$$\mathcal{V}(f_0, f_\theta) \leq \mathcal{V}(f_0, f) + 2|\theta|[(\log C_0)_+ + (\log f(0))_+](\log f(0))_+[C_0 + 2C_1].$$

Proof of Lemma 10. First we study $\mathcal{KL}(f_0, f_\theta)$:

$$\mathcal{KL}(f_0, f_\theta) - \mathcal{KL}(f_0, f) = \int_0^\infty \log f(x) f_0(x) dx - \int_0^\infty \log f(x-\theta) f_0(x) dx$$
$$= \int_0^{-\theta} \log f(x) f_0(x) dx + \int_{-\theta}^\infty \log f(x)[f_0(x) - f_0(x+\theta)] dx$$
$$\leq (\log f(0))_+ |\theta| f_0(0) + (\log f(0))_+ \int_{-\theta}^\infty [\sum_{j=1}^r |f_0^{(j)}(x)|\, |\theta|^j / j! + 2L(x)|\theta|^\beta] dx$$
$$\leq (\log f(0))_+ |\theta| f_0(0)$$
$$+ |\theta| (\log f(0))_+ \left[\sum_{j=1}^r 1/j! \int_0^\infty \frac{|f_0^{(j)}(x)|^{\rho_j}}{f_0^{\rho_j}(x)} f_0(x) dx + 2 \int_0^\infty \frac{|L(x)|^{\rho_\beta}}{f_0^{\rho_\beta}(x)} f_0(x) dx \right]$$
$$\leq (\log f(0))_+ |\theta| [f_0(0) + 2C_1]$$

by the definition of the class $\mathcal{P}(\beta, L(\cdot), \gamma, \ldots)$ with the appropriate ρ_j and $\beta > 1$, since $|\theta| \leq \Delta$.

Now we consider $\mathcal{V}(f_0, f_\theta)$. Recall that $f(x) \geq f(x-\theta)$ since f decreases and $\theta \leq 0$. Hence,

$$\mathcal{V}(f_0, f_\theta) = \int_0^\infty [\log f_0(x) - \log f(x-\theta)]^2 f_0(x) dx$$
$$= \int_0^\infty [\log f_0(x)]^2 f_0(x) dx - 2 \int_0^\infty \log f_0(x) \log f(x-\theta) f_0(x) dx + \int_0^\infty [\log f(x-\theta)]^2 f_0(x) dx$$
$$\leq \mathcal{V}(f_0, f) + 2(\log f_0(0))_+ [\mathcal{KL}(f_0, f_\theta) - \mathcal{KL}(f_0, f)] + \int_0^\infty [[\log f(x-\theta)]^2 - [\log f(x)]^2] f_0(x) dx.$$

Consider the last integral. As $f(x)$ decreases, for $\theta \leq 0$ it is bounded above by 0. For $\theta > 0$,

$$\begin{aligned} & \int_0^\infty [[\log f(x-\theta)]^2 - [\log f(x)]^2] f_0(x) dx \\ \leq \quad & 2(\log f(0))_+ \int_0^\infty [\log f(x-\theta) - \log f(x)] f_0(x) dx \\ \leq \quad & 2(\log f(0))_+ [\mathcal{KL}(f_0, f) - \mathcal{KL}(f_0, f_\theta)]. \end{aligned}$$

Hence,

$$\mathcal{V}(f_0, f_\theta) \leq \mathcal{V}(f_0, f) + 2[(\log f_0(0))_+ + (\log f(0))_+](\log f(0))_+ |\theta| [f_0(0) + 2C_1].$$

□

3.6 Proofs: posterior concentration rate of f for known θ_0

Proof of Theorem 8. The theorem is proved by verifying the assumptions of Theorem 2.1 of [21]. The first assumption, on the prior mass of Kullback - Leibler neighbourhood of the true density, is verified in Lemma 17. In Lemma 11 we control the Hellinger entropy of the sieves defined below. Fix an arbitrary $\zeta > 0$ to be defined later, and take a sieve $Q_n = Q(\zeta\varepsilon_n, J_n, a_n, b_n, \underline{z}, \bar{z})$

as defined by (3.49) in Lemma 11 with

$$\begin{aligned} \varepsilon_n &= n^{-\beta/(2\beta+1)}[\log n]^q, \quad \bar{z} = n^d, \quad \underline{z} = 2, \quad d > 2/(2\beta+1) \\ J_{1,n} &= j_1 n^{1/(2\beta+1)}(\log n)^{2q-1}, \quad J_{0,n} = j_0 n^{1/(2\beta+1)}(\log n)^{2q-1}, \quad , \\ a &= \exp\{-Cn^{1/(2\beta+1)}[\log n]^{2q}\}, \quad b = \exp\{Cn^{1/(2\beta+1)}[\log n]^{2q}\}, \\ p_{\min} &= \exp\{-Cn^{1/(2\beta+1)}[\log n]^{2q}\}, \quad p_n \leq 1 - \exp(-n^s) \end{aligned} \tag{3.48}$$

for some constants $C, s > 0$, and j_0, j_1 large enough and q as defined in the theorem.

Then, by Lemma 13,

$$\Pi(Q^c) \leq e^{-n\epsilon_n^2} \text{ and } \log N(\zeta\epsilon_n, Q, d_H) \leq Cn\epsilon_n^2.$$

Choosing ϵ_0 large enough in the definition of $\epsilon_n = \epsilon_0\varepsilon_n$ completes the proof of Theorem 8.

□

3.6.1 Entropy condition

Denote the density of $\Gamma(z, z/\epsilon)$ by

$$g_{z,\epsilon}(x) = \frac{1}{\Gamma(z)}\left(\frac{z}{\epsilon}\right)^z x^{z-1}e^{\frac{-zx}{\epsilon}}, \quad x > 0.$$

Lemma 11. *Fix $\varepsilon > 0$, $J_0, J_1 \in \mathbb{N}$, $0 < a < \delta < b < \infty$, $2 \leq \underline{z} < \bar{z} < \infty$, $0 < p_{\min} < p_n$, $p_n \geq 1/2$, and introduce the following class of densities:*

$$\mathcal{Q} = \left\{ \begin{array}{l} f(x) = \int_x^\infty \theta^{-1}d\theta \int g_{z,\epsilon}(\theta)\left[\frac{p\epsilon^2 dQ^{(0)}(\epsilon) + (1-p)dQ^{(1)}(\epsilon)}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p}\right], \text{ with} \\ Q^{(k)} = \sum_{j=1}^\infty \pi_j^{(k)}\delta_{(\epsilon_j^{(k)})} \text{ where } \sum_{j>J_0}\pi_j^{(0)} < \varepsilon/(6\delta^2 r_{\max}), \quad \sum_{j>J_1}\pi_j^{(1)} < \varepsilon/2, \\ \epsilon_j^{(0)} \in [a,\delta] \text{ for } j = 1,\ldots,J_0, \quad \epsilon_j^{(1)} \in [\delta, a+b] \text{ for } j = 1,\ldots,J_1; \\ \quad z \in [\underline{z}, \bar{z}], \quad p \in [p_{\min}, p_n] \end{array} \right\} \tag{3.49}$$

where $r_{\max} = p_n/(1-p_n) \geq 1$.

Then, for $\varepsilon \leq \min(1, (\bar{z})^{1/4})$,

$$\begin{aligned} \log N(5\varepsilon, Q, d_H) &\leq C + J\left[\log\log(b/a) - 2\log(2\varepsilon/5) + 0.5\log(\bar{z})\right] \\ &\quad + \log\log(\bar{z}/\underline{z}) + \log\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right), \\ \Pi(Q^c) &\leq J_0 G^{(0)}([0,a]) + J_1 G^{(1)}((a+b,\infty)) \\ &\quad + 1 - \Pi_z([\underline{z},\bar{z}]) + 1 - \Pi([p_{\min}, p_n]) \\ &\quad + \left(\frac{em}{J_0}\log(6\sqrt{5/2}\delta^2 r_{\max}/\sqrt{\varepsilon})\right)^{J_0} + \left(\frac{em}{2J_1}\log(2.5/\varepsilon)\right)^{J_1}. \end{aligned}$$

In particular if $\mathcal{Q}$ is defined by (3.48)*, then*

$$\Pi(Q_n^c) \leq e^{-n\epsilon_n^2} \textit{ and } \log N(\zeta\epsilon_n, Q_n, d_H) \leq Cn\epsilon_n^2$$

Proof. Proof of Lemma 11

This lemma follows from Lemma 12 due to inequality $d_H^2(f_1,f_2)\le 2||f_1-f_2||_1$ which implies $N(\varepsilon^2/2,Q,d_H)\le N(\varepsilon,Q,\|\cdot\|_1)$.

Denoting $v=5\varepsilon^2/2$, and hence $\varepsilon=\sqrt{2v/5}$, Lemma 12 implies

$$
\begin{aligned}
\log N(5v,Q,d_H) &\le \log N(5\sqrt{v/5},Q,\|\cdot\|_1)\\
&\le C+J\left[\log\log(b/a)-2\log(2v/5)+0.5\log(\bar z)\right]\\
&\quad +\log\log(\bar z/\underline{z})+\log\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right),\\
\Pi(Q^c) &\le J_0G^{(0)}([0,a])+J_1G^{(1)}((a+b,\infty))\\
&\quad +1-\Pi_z([\underline{z},\bar z])+1-\Pi([p_{\min},p_n])\\
&\quad +\left(\frac{em}{J_0}\log(6\sqrt{5/2}\delta^2 r_{\max}/\sqrt{v})\right)^{J_0}+\left(\frac{em}{2J_1}\log(2.5/v)\right)^{J_1}.
\end{aligned}
$$

□

Lemma 12. *Fix $\varepsilon>0$, $J_0,J_1\in\mathbb{N}$, $0<a<\delta<b<\infty$, $2\le\underline{z}<\bar z<\infty$, $0<p_{\min}\le p_n$, $p_n\ge 1/2$, $\delta\le 1$ and introduce the following class of densities:*

$$
\mathcal{Q}=\left\{\begin{array}{l}
f(x)=\int_x^\infty\theta^{-1}d\theta\int g_{z,\epsilon}(\theta)\left[\frac{p\epsilon^2 dQ^{(0)}(\epsilon)+(1-p)dQ^{(1)}(\epsilon)}{p\int_0^\delta\epsilon^2 dQ^{(0)}(\epsilon)+1-p}\right],\ \textit{with}\\
Q^{(k)}=\sum_{j=1}^\infty\pi_j^{(k)}\delta_{(\epsilon_j^{(k)})}\ \textit{where}\ \sum_{j>J_0}\pi_j^{(0)}<\frac{\varepsilon}{(6\delta^2 r_{\max})},\quad \sum_{j>J_1}\pi_j^{(1)}<\varepsilon/2,\\
\epsilon_j^{(0)}\in[a,\delta]\ \textit{for}\ j=1,\dots,J_0,\quad \epsilon_j^{(1)}\in[\delta,a+b]\ \textit{for}\ j=1,\dots,J_1;\\
\quad z\in[\underline{z},\bar z],\quad p\in[p_{\min},p_n]
\end{array}\right\}
$$

where $r_{\max}=p_n/(1-p_n)\ge 1$.

Then, for $\varepsilon\le\min(1,\sqrt{\bar z})$,

$$
\begin{aligned}
\log N(5\varepsilon,Q,\|\cdot\|_1) &\le C+J\left[\log\log(b/a)-4\log\varepsilon+0.5\log(\bar z)\right]\\
&\quad +\log\log(\bar z/\underline{z})+\log\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right),\\
\Pi(Q^c) &\le J_0G^{(0)}([0,a])+J_1G^{(1)}((a+b,\infty))\\
&\quad +1-\Pi_z([\underline{z},\bar z])+\Pi([0,p_{\min}])\\
&\quad +\left(\frac{em}{J_0}\log(6\delta^2 r_{\max}/\varepsilon)\right)^{J_0}+\left(\frac{em}{J_1}\log(1/\varepsilon)\right)^{J_1}.
\end{aligned}
$$

Proof of Lemma 12. Take any $f\in\mathcal{Q}$, that is, $f(x)=\sum_{j=1}^\infty p_j\int_x^\infty\theta^{-1}d\theta g_{z,\epsilon_j}(\theta)$ where

$$
p_j=\pi_j\frac{p\epsilon_j^2 I(\epsilon_j\le\delta)+(1-p)I(\epsilon_j>\delta)}{p\sum_{j=1}^\infty\pi_j\epsilon_j^2 I(\epsilon_j\le\delta)+(1-p)},
$$

with superscripts k for $p_j^{(k)}$ corresponding to $\epsilon_j^{(k)}$, with a slight abuse of notation. Note that $\sum_{j=1}^{J_0}p_j^{(0)}+\sum_{j=1}^{J_1}p_j^{(1)}\le 1$.

Fix $\delta_2=\varepsilon/C$ for the constant C defined in proof of Lemma 4.2 in [4] (seethe statement of this Lemma in section 3.11), $\delta_1=\varepsilon/(6\sqrt{2z})$ and $\delta_p=\varepsilon/2$. Note that since $\varepsilon\le 1$ and $z>2$, $\delta_1<1$, $\delta_2<1$. We have also assumed that $\delta\le 1$.

Let $\hat A$ be the following set $\{a(1+\delta_1)^k\}_{k=0}^K$ with $K=K_z=\lceil\log(1+b/a)/\log(1+\delta_1)\rceil$, with the interval corresponding to $k = k_\delta$: $a(1+\delta_1)^k < \delta < a(1+\delta_1)^{k+1}$ split into two intervals: $[a(1+\delta_1)^{k_\delta}, \delta]$ and $(\delta, a(1+\delta_1)^{k_\delta+1}]$ (if such k_δ exists).

Define also $\hat{Z} = \{\underline{z}(1+\delta_2)^\ell\}_{\ell=0}^L$ with $L = \lceil \log(\bar{z}/\underline{z})/\log(1+\delta_2)\rceil$ and $\hat{P} = \{p = x/(x+1),\, x \in \hat{R}\}$ where $\hat{R} = \{p_{\min}/(1-p_{\min})(1+\delta_p)^r\}_{r=0}^R$ with $R = \lceil \log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right)/\log(1+\delta_p)\rceil$. In particular, for any $z \in [\underline{z}(1+\delta_2)^\ell, \underline{z}(1+\delta_2)^{\ell+1})$ for some $\ell \in \{0,1,\ldots,L\}$, $\inf_{\hat{z}\in\hat{Z}} |\hat{z}/z - 1| \le \delta_2$ and $\inf_{\hat{p}\in\hat{P}} \frac{|p-\hat{p}|}{p(1-\hat{p})} \le \delta_p$. Let $\hat{S}$ be an ε-net for $S = \{(\tilde{\pi}_1^{(0)}, \ldots, \tilde{\pi}_{J_0}^{(0)}, \tilde{\pi}_1^{(1)}, \ldots, \tilde{\pi}_{J_1}^{(1)}) : \tilde{\pi}_j^{(k)} = \pi_j^{(k)}/(\sum_{j=1}^{J_k} \pi_j^{(k)})\, \forall j, k = 0, 1\}$.

Given $f \in \mathcal{Q}$, define

$$\widehat{\mathcal{Q}} = \left\{\begin{array}{l} \hat{f}(x) = \sum_{j\in J_{01}} \hat{\pi}_j \int_x^\infty \theta^{-1} g_{\hat{z},\hat{\epsilon}_j}(\theta)d\theta \text{ where } \hat{z} \in \hat{Z}, \quad |\hat{z}/z - 1| < \delta_2, \\ \quad \hat{\epsilon}_j \in \hat{A},\, j \in J_{01},\ \max_{j\in J_{01}} |\hat{\epsilon}_j/\epsilon_j - 1| < \delta_1, \\ \quad \hat{\pi} = (\hat{\pi}_j^{(k)}) \in \hat{S} \text{ and } \sum_{j\in J_0} |\hat{\pi}_j^{(0)} - \tilde{\pi}_j^{(0)}| < \varepsilon/(6\delta^2 r_{\max}), \\ \sum_{j\in J_1} |\hat{\pi}_j^{(1)} - \tilde{\pi}_j^{(1)}| < \varepsilon/2 \\ \quad \text{with } \tilde{\pi}_j^{(k)} = \hat{\pi}_j^{(k)}/[\sum_{j\in J_k} \hat{\pi}_j^{(k)}],\ k = 0, 1, \\ \quad \hat{p} \in \hat{P} : \frac{|p-\hat{p}|}{p(1-\hat{p})} \le \delta_p \end{array}\right\}.$$

First we show that for all $f \in \mathcal{Q}$ there exists $\hat{f} \in \widehat{\mathcal{Q}}$ such that

$$||f - \hat{f}||_1 = \|\tilde{K}_z P_Q(x) - \tilde{K}_z P_{\hat{Q}}(x)\|_1 \le \|K_z P_Q(x) - K_z P_{\hat{Q}}(x)\|_1 \le 5\varepsilon. \tag{3.50}$$

Note that for any $G(\theta)$ independent of x,

$$\begin{aligned} ||\int_x^\infty \theta^{-1}G(\theta)d\theta||_1 &\le \int_0^\infty \left|\int_x^\infty \theta^{-1}G(\theta)d\theta\right| dx \le \int_0^\infty \int_0^\infty \theta^{-1}|G(\theta)|I(\theta > x)d\theta dx \\ &= \int_0^\infty \theta^{-1}|G(\theta)| \left[\int_0^\infty I(\theta > x)dx\right] d\theta = \int_0^\infty |G(\theta)|d\theta \\ &= ||G(\cdot)||_1. \end{aligned}$$

If $G \ge 0$ then there is equality. Therefore, is it sufficient to prove that

$$||f - \hat{f}||_1 = \|K_z P_Q(x) - K_z P_{\hat{Q}}(x)\|_1 \le 5\varepsilon,$$

where $K_z P$ is a mixture of gammas defined by (3.51).

Denote $J_{01} = \{j : \epsilon_j \in [a, a+b]\}$. We have, using the inequalities on the L_1 norms shown above,

$$\begin{aligned} || \sum_{j\in J_{01}} \hat{p}_j \int_x^\infty \theta^{-1} g_{\hat{z},\hat{\epsilon}_j}(\theta)d\theta - \sum_{j=1}^\infty p_j \int_x^\infty \theta^{-1} g_{z,\epsilon_j}(\theta)||_1 &= || \sum_{j\in J_{01}} \hat{p}_j g_{\hat{z},\hat{\epsilon}_j} - \sum_{j=1}^\infty p_j g_{z,\epsilon_j}||_1 \\ \le \sum_{j\in J_{01}} \hat{p}_j ||(g_{z,\hat{\epsilon}_j} - g_{\hat{z},\hat{\epsilon}_j})||_1 &+ \sum_{j\in J_{01}^c} p_j ||g_{z,\epsilon_j}||_1 \\ + \sum_{j\in J_{01}} p_j ||(g_{z,\hat{\epsilon}_j} - g_{z,\epsilon_j})||_1 &+ \sum_{j\in J_{01}} |\hat{p}_j - p_j| ||g_{z,\hat{\epsilon}_j}||_1 \\ \le \sum_{j\in J_{01}} \hat{p}_j ||g_{z,\hat{\epsilon}_j} - g_{\hat{z},\hat{\epsilon}_j}||_1 + \sum_{j\in J_{01}^c} p_j &+ \sum_{j\in J_{01}} p_j ||g_{z,\hat{\epsilon}_j} - g_{z,\epsilon_j}||_1 + \sum_{j\in J_{01}} |\hat{p}_j - p_j|. \end{aligned}$$

To bound the first and third terms, we use Lemma C.1 of [4] (See Lemma 33 in Section 3.11):

$$||g_{z,\hat{\epsilon}_j} - g_{z,\epsilon_j}||_1 \quad \leq \quad \sqrt{2KL(g_{z,\hat{\epsilon}_j}, g_{z,\epsilon_j})} \leq \sqrt{2z}\delta_1 = \varepsilon/6$$

by the definition of δ_1 and

$$||g_{z,\hat{\epsilon}_j} - g_{\hat{z},\hat{\epsilon}_j}||_1 \leq \sqrt{2C}\delta_2 = \varepsilon$$

so that

$$\sum_{j \in J_{01}} \hat{p}_j ||g_{z,\hat{\epsilon}_j} - g_{\hat{z},\hat{\epsilon}_j}||_1 + \sum_{j \in J_{01}} p_j ||g_{z,\hat{\epsilon}_j} - g_{z,\epsilon_j}||_1 \leq 7\varepsilon/6.$$

Denote $c = \sum_{j=1}^{\infty} \pi_j^{(0)} [\epsilon_j^{(0)}]^2$.

$$\sum_{j>J_0} p_j^{(0)} \quad = \quad \frac{p}{pc+1-p} \sum_{j>J_0} \pi_j^{(0)} [\epsilon_j^{(0)}]^2 \leq \frac{p\delta^2}{pc+1-p} \varepsilon/(6\delta^2 r_{\max}) \leq \varepsilon/6.$$

$$\sum_{j>J_1} p_j^{(1)} \quad = \quad \frac{1-p}{pc+1-p} \sum_{j>J_1} \pi_j^{(1)} \leq \frac{1-p}{pc+1-p} \varepsilon \leq \varepsilon.$$

Also,

$$\sum_{j=1}^{J_k} |\pi_j^{(k)} - \hat{\pi}_j^{(k)}| \leq \sum_{j=1}^{J_k} |\tilde{\pi}_j^{(k)} - \hat{\pi}_j^{(k)}| + \sum_{>J_k} \pi_j^{(k)}.$$

For $k = 0$, the upper bound is $\varepsilon/(3\delta^2 r_{\max})$, and for $k = 1$, the upper bound is ε.

Then, for $j \leq J_1$,

$$\begin{aligned} |p_j^{()} 1 - \hat{p}_j^{(1)}| \quad &= \quad \left| \frac{(1-p)\pi_j^{(1)}}{pc + (1-p)} - \frac{(1-\hat{p})\hat{\pi}_j^{(1)}}{\hat{p}\hat{c} + (1-\hat{p})} \right| \\ &\leq \quad \pi_j^{(1)} \left| \frac{(1-p)}{pc + (1-p)} - \frac{(1-\hat{p})}{\hat{p}\hat{c} + (1-\hat{p})} \right| + \frac{(1-\hat{p})|\hat{\pi}_j^{(1)} - \pi_j^{(1)}|}{\hat{p}\hat{c} + (1-\hat{p})}. \end{aligned}$$

The first term, up to the factor of $\pi_j^{(1)}$, equals to

$$\begin{aligned} & \frac{|(1-p)(\hat{p}\hat{c} + (1-\hat{p})) - (1-\hat{p})(pc + (1-p))|}{(\hat{p}\hat{c} + (1-\hat{p}))(pc + (1-p))} \\ = \quad & p\hat{p} \frac{\left| \frac{(1-p)}{p}\hat{c} - \frac{(1-\hat{p})}{\hat{p}} c \right|}{(\hat{p}\hat{c} + (1-\hat{p}))(pc + (1-p))} \\ \leq \quad & \frac{c\,|p - \hat{p}| + (1-p)\hat{p}|c - \hat{c}|}{(\hat{p}\hat{c} + (1-\hat{p}))(pc + (1-p))} \end{aligned}$$

Also, we have

$$
\begin{aligned}
|\hat{c}-c| &= \sum_{j=1}^{\infty} |\pi_j^{(0)}[\epsilon_j^{(0)}]^2 - \hat{\pi}_j^{(0)}[\hat{\epsilon}_j^{(0)}]^2| \\
&\leq \sum_{j\in J_0} \pi_j^{(0)}[\epsilon_j^{(0)}]^2|1-[\hat{\epsilon}_j^{(0)}]^2/[\epsilon_j^{(0)}]^2| + \sum_{j\in J_0}[\hat{\epsilon}_j^{(0)}]^2|\pi_j^{(0)}-\hat{\pi}_j^{(0)}| + \sum_{j\in J_0^c}[\epsilon_j^{(0)}]^2\pi_j^{(0)} \\
&\leq 2\delta_1 c + \varepsilon/(3r_{\max}) \leq c\varepsilon/6 + \varepsilon/(3r_{\max})
\end{aligned}
$$

since $1/\sqrt{z} \leq 1/\sqrt{2}$.

Therefore,

$$
\begin{aligned}
\sum_{j\leq J_1} |p_j^{(1)} - \hat{p}_j^{(1)}| &\leq \frac{\hat{c}\,|p-\hat{p}| + (1-\hat{p})p|c-\hat{c}|}{(\hat{p}\hat{c}+(1-\hat{p}))(pc+(1-p))}\sum_{j\leq J_1}\pi_j^{(1)} + \frac{(1-\hat{p})}{\hat{p}\hat{c}+(1-\hat{p})}\sum_{j\leq J_1}|\hat{\pi}_j^{(1)}-\pi_j^{(1)}| \\
&\leq \frac{|p-\hat{p}|}{(1-\hat{p})p} + \frac{p[c\varepsilon/6+\varepsilon/(3r_{\max})]}{(pc+(1-p))} \\
&\leq \delta_p + 3\varepsilon/2.
\end{aligned}
$$

Now consider $k=0$:

$$
\begin{aligned}
|p_j^{(0)} - \hat{p}_j^{(0)}| &= |\frac{p\pi_j^{(0)}[\epsilon_j^{(0)}]^2}{pc+(1-p)} - \frac{\hat{p}\hat{\pi}_j^{(0)}[\hat{\epsilon}_j^{(0)}]^2}{\hat{p}\hat{c}+(1-\hat{p})}| \\
&\leq \pi_j^{(0)}[\epsilon_j^{(0)}]^2\left|\frac{p}{pc+(1-p)} - \frac{\hat{p}}{\hat{p}\hat{c}+(1-\hat{p})}\right| + \frac{\hat{p}\pi_j^{(0)}[\epsilon_j^{(0)}]^2|1-[\hat{\epsilon}_j^{(0)}]^2/[\epsilon_j^{(0)}]^2|}{\hat{p}\hat{c}+(1-\hat{p})} \\
&\quad + \frac{\hat{p}|\pi_j^{(0)}-\hat{\pi}_j^{(0)}||[\hat{\epsilon}_j^{(0)}]^2}{\hat{p}\hat{c}+(1-\hat{p})}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\sum_{j\leq J_0}|p_j^{(0)}-\hat{p}_j^{(0)}| &\leq \frac{|p-\hat{p}|+\hat{p}p|c-\hat{c}|}{(\hat{p}\hat{c}+(1-\hat{p}))(pc+(1-p))}\sum_{j\leq J_0}\pi_j^{(0)}[\epsilon_j^{(0)}]^2 \\
&\quad + \sum_{j\leq J_0}\frac{\hat{p}\pi_j^{(0)}[\epsilon_j^{(0)}]^2|1-[\hat{\epsilon}_j^{(0)}]^2/[\epsilon_j^{(0)}]^2|}{\hat{p}\hat{c}+(1-\hat{p})} + \frac{\hat{p}\delta^2}{\hat{p}\hat{c}+(1-\hat{p})}\sum_{j\leq J_0}|\pi_j^{(0)}-\hat{\pi}_j^{(0)}| \\
&\leq \delta_p + \varepsilon\frac{c}{3\hat{c}} + \varepsilon/3.
\end{aligned}
$$

Since $\hat{c}\leq 1$, $\varepsilon \leq 1$ and $r_{\max}\geq 1$, using the upper bound on $\hat{c}-c$, we have

$$
\hat{c}/c \geq 1 - \varepsilon(1/6+1/(3r_{\max})) \geq 1/2.
$$

Hence, using $\delta_p = \varepsilon/2$,

$$
\sum_{j\in J_{01}}|p_j - \hat{p}_j| \leq \varepsilon[1/2+2/3+1/2+1] = 8\varepsilon/3.
$$

Therefore, combining all the inequalities together, we obtain that

$$||\sum_{j\in J_{01}}\hat{\pi}_j\int_x^\infty \theta^{-1}g_{\hat{z},\hat{\epsilon}_j}(\theta)d\theta - \sum_{j=1}^\infty \pi_j\int_x^\infty \theta^{-1}g_{z,\epsilon_j}(\theta)d\theta||_1 \quad \le \quad 5\varepsilon,$$

and hence $\hat{Q}$ is a 5ε-net of Q.

Now we study cardinality of $\hat{\mathcal{Q}}$. Denote $J = |J_{01}| = J_0 + J_1$. Following [4], $|\hat{S}|$ is $\lesssim \varepsilon^{-J}$, $K_z \lesssim \frac{\sqrt{\bar{z}}\log(b/a)}{\varepsilon}$ for large b/a, assuming that $\varepsilon \le \sqrt{\bar{z}}$.

Then, for $\varepsilon \le \sqrt{\bar{z}}$, the cardinality of $\hat{Q}$ is bounded by

$$\begin{aligned}
|\hat{Q}| \quad &\le \quad |\hat{S}|\cdot|\hat{P}|\sum_{\ell=1}^L |K_z|^J \lesssim R\varepsilon^{-J}L\left[\frac{\sqrt{\bar{z}}\log(b/a)}{\varepsilon}\right]^J\\
&\lesssim \quad \left[\frac{\log(b/a)\sqrt{\bar{z}}}{\varepsilon^2}\right]^J \frac{\log(\bar{z}/\underline{z})}{\varepsilon}\frac{\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right)}{\varepsilon}
\end{aligned}$$

due to $\delta_2 = C\varepsilon$ and by the definition of L.

$$\begin{aligned}
\log N(5\varepsilon, Q, ||\cdot||_1) \quad &\le \quad C + J\left[\log\log(b/a) - 4\log\varepsilon + 0.5\log(\bar{z})\right] + \log\log(\bar{z}/\underline{z})\\
&\quad + \log\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right).
\end{aligned}$$

The lower bound on the prior mass of $\mathcal{Q}^c$ is proved following the same route as in the proof of Proposition 2 in [47]. For the independent Dirichlet process priors for $(Q^{(0)})$ and $(Q^{(1)})$ with masses m_0 and m_1 respectively,

$$\begin{aligned}
\Pi(Q^c) \quad &\le \quad \Pi_z([\underline{z},\bar{z}]^c) + J_0G_0([0,a]) + J_1G_1((a+b,\infty)) + \Pi([p_{\min},p_n]^c)\\
&\quad +\Pi(\sum_{j>J_0}\pi_j^{(0)} > \varepsilon/(6\delta^2 r_{\max})) + \Pi(\sum_{j>J_1}\pi_j^{(1)} > \varepsilon/2)\\
&\le \quad J_0G_0([0,a]) + J_1G_1((a+b,\infty)) + 1 - \Pi_z([\underline{z},\bar{z}]) + 1 - \Pi([p_{\min},p_n])\\
&\quad + \left(\frac{em_0}{J_0}\log(6\delta^2 r_{\max}/\varepsilon)\right)^{J_0} + \left(\frac{em_1}{J_1}\log(2/\varepsilon)\right)^{J_1}.
\end{aligned}$$

□

Lemma 13. *Consider the prior* Π *on* f *defined in Section 3.3.1 with* $\log\log(\delta^2/(1-p_n)) \le C\log\log n$ *for some* $C > 0$.

Then, in the notation of Lemma 11,

$$\Pi(Q^c) \le e^{-n\epsilon_n^2} \text{ and } \log N(\zeta\epsilon_n, Q, d_H) \le Cn\epsilon_n^2$$

for $\epsilon_n = n^{-(1-\gamma)/2}(\log n)^t$ *with*

$$\begin{aligned}
&J_1 = n^\gamma[\log n]^{2t-1}, \quad J_0 = n^\gamma[\log n]^{2t-1},\\
\bar{z} = n, \quad \underline{z} = 2, \quad &a = \exp\{-Cn^\gamma[\log n]^{2t}\}, \quad b = \exp\{Cn^\gamma[\log n]^{2t}\},\\
&p_{\min} = \exp\{-Cn^\gamma[\log n]^{2t}\}, \quad 1 - p_{\max} = p_n.
\end{aligned}$$

For instance, $n^{-\beta/(2\beta+1)} = n^{-(1-\gamma)/2}$ with $\gamma = 1/(2\beta+1)$.

Proof of Lemma 13. Need $\Pi(Q^c) \le e^{-n\epsilon_n^2}$ and $\log N(\zeta\epsilon_n, Q, d_H) \le Cn\epsilon_n^2$. According to Lemma 11, these conditions are satisfied if

$$\begin{aligned}
&J_1 G^{(1)}((a+b,\infty)) \le e^{-n\epsilon_n^2},\\
&1-\Pi_z([\underline{z},\bar{z}]) \le e^{-n\epsilon_n^2}, \quad 1-\Pi([p_{\min},p_n]) \le e^{-n\epsilon_n^2},\\
&J_0\left[\log J_0 - \log(em_0) - \log\log(6\sqrt{5/2}\delta^2 r_{\max}/(\zeta\sqrt{\epsilon_n}))\right] \ge n\epsilon_n^2,\\
&J_1 \log\left(\frac{em_1}{J_1}\log(2.5/(\zeta\epsilon_n))\right) \le -n\epsilon_n^2,
\end{aligned}$$

and

$$\begin{aligned}
J &\le Cn\epsilon_n^2/\log(\epsilon_n^{-1}), \quad \log\log(b/a) \le \log(\epsilon_n^{-1}),\\
\log(\bar{z}) &\le C\log(\epsilon_n^{-1}), \log\log(\bar{z}/\underline{z}) \le n\epsilon_n^2,\\
&\log\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right) \le n\epsilon_n^2.
\end{aligned}$$

In our case, $\epsilon_n = n^{-(1-\gamma)/2}(\log n)^t$ and hence $n\epsilon_n^2 = n^\gamma(\log n)^{2t}$.
By assumptions on the prior of Lemma 17,

$$\begin{gathered}
\log G^{(0)}([0,a]) \le C\log a, \quad \log G^{(1)}([a+b,\infty)) \le -C\log(a+b),\\
\log(\Pi_z([\bar{z},\infty))) \le C - C\sqrt{\bar{z}}[\log\bar{z}]^{\rho_z}.
\end{gathered}$$

Hence, we need

1. for J_0, J_1:

$$\begin{aligned}
&J_0\left[\log J_0 - \log\log(\delta^2 r_{\max}) - \log\log n\right] \ge n^\gamma[\log n]^{2t},\\
&J_1\left[\log J_1 - \log\log n\right] \ge n^\gamma[\log n]^{2t},\\
&J_0 + J_1 \le Cn^\gamma[\log n]^{2t-1},
\end{aligned}$$

   which hold if $J_1 = n^\gamma[\log n]^{2t-1}$, $J_0 = n^\gamma[\log n]^{2t-1}$ and $\log\log(\delta^2 r_{\max}) \le C\log\log n$;

2. for a, b:

$$a \le \exp\{-Cn^\gamma[\log n]^{2t}\}, \quad b \ge \exp\{Cn^\gamma[\log n]^{2t}\}, \quad \log\log(b/a) \le \log n$$

   e.g. we can take

$$a = \exp\{-Cn^\gamma[\log n]^{2t}\}, \quad b = \exp\{Cn^\gamma[\log n]^{2t}\};$$

3. for $\underline{z}$, $\bar{z}$:

$$\begin{aligned}
\Pi_z([0,\underline{z}]) \le e^{-n^\gamma[\log n]^{2t}}, \bar{z}[\log\bar{z}]^{2\rho_z} &\ge Cn^{2\gamma}[\log n]^{4t}, \quad \log(\bar{z}) \le C\log n,\\
\log\log(\bar{z}/\underline{z}) &\le n^\gamma[\log n]^{2t}
\end{aligned}$$

e.g. we can take $\bar{z} = n^d$ with $d > 2\gamma$. If $\rho_z > 2t$, we can take $\bar{z} = n^{2\gamma}$. Note that $\underline{z} \geq 2$, hence we can take $\underline{z} = 2$.

4. for $p_{\min}$, p_n:

$$\begin{aligned} 1 - p_n &\leq p_n \exp\{-Cn^\gamma[\log n]^{2t}\}, \quad \log(p_{\min}) \leq -n^\gamma[\log n]^{2t}, \\ \log\log\left(\frac{p_n}{\delta^2(1-p_n)}\right) &\leq C\log\log n, \quad \log\log\left(\frac{p_n(1-p_{\min})}{(1-p_n)p_{\min}}\right) \leq n^\gamma[\log n]^{2t}. \end{aligned}$$

Taking $1 - p_n = C[\log n]^{-s}$ implies $\Pi_p(p > p_n) = 0$ and

$$\log\log\left(\frac{p_n}{(1-p_n)}\right) \leq \log\log\left(\frac{p_n}{(1-p_n)}\right) \leq C\log n \leq n\epsilon_n^2$$

for the given ϵ_n. For $p_{\min}$ defined by

$$\log\log(1/p_{\min}) = 0.5n\epsilon_n^2$$

we have

$$\log\log\left(\frac{(1-p_{\min})}{p_{\min}}\right) \leq 0.5n\epsilon_n^2$$

and

$$\log(\Pi_p([0, p_{\min}])) \leq C - \log\log(1/p_{\min}) = C - 0.5n\epsilon_n^2 \leq -0.4n\epsilon_n^2$$

for large enough n.

□

3.6.2 Continuous and discrete approximation

Continuous approximation

In this section we study how to approximate $f_0(x)$ by a continuous mixture (3.8). Define for any measurable and integrable function h

$$\tilde{K}_z h(x) = \int_x^\infty \frac{1}{\theta} K_z h(\theta) d\theta = \int_x^\infty \int_{\mathbb{R}^+} \frac{1}{\theta} h(\epsilon) g_{\Gamma(z,z/\epsilon)}(\theta) d\epsilon d\theta, \; x \geq 0,$$

where operator $K_z f(x)$ was considered in [4]

$$K_z f(x) = \int f(\epsilon) g_{\Gamma(z,z/\epsilon)}(x) d\epsilon, \tag{3.51}$$

$$g_{\Gamma(z,z/\epsilon)}(x) = \frac{(z/\epsilon)^z}{\Gamma(z)} x^{z-1} e^{-xz/\epsilon}. \tag{3.52}$$

To simplify the notation, we use $g_{z,\epsilon}(x) = g_{\Gamma(z,z/\epsilon)}(x)$. Note that if

$$f(x) = \int_x^\infty g(\theta)\frac{d\theta}{\theta} \quad \text{then} \quad g(x) = -xf'(x).$$

The idea behind the approximation is that as z goes to infinity, when $g_0(x) = -xf_0'(x)$ is continuous, $\tilde{K}_z g_0$ approximates $f_0(x)$ to the order $z^{-\beta/2}$. This is made more precise in Lemma

14, where we make the form of $\tilde{K}_z g_0$ explicit. Using this, we can construct $g_\beta \geq 0$ such that $\tilde{K}_z g_\beta$ is approximately equal to f_0 to the same order. This is presented in the following proposition.

Proposition 8. *Assume that $f_0 \in \mathcal{P}(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu)$ with $\nu \leq 1/3$, $\beta > (2-e)\nu \vee 1$ and set $g_0(x) = -xf_0'(x)$. Assume also that*

$$\int_0^\infty x^m f_0(x)dx \leq C_1, \quad m = \frac{2\beta}{(1-\nu)\beta - 2\nu} \vee 2. \tag{3.53}$$

Then there exist function $h_j(x)$, $j < \beta/2$, linear combinations of functions in the form $-x^\ell f_0^{(\ell)}(x)$, with fixed coefficients with $\ell \leq 2j$ and polynomial functions of $1/\sqrt{z}$, $c_j(z)$, such that $c_j(z) = 1 + O(z^{-1/2})$ and if

$$\begin{aligned} g_1(x) &= \frac{(-xf_0'(x))}{1+c_0(z)} + \sum_{j=1}^{\lceil \beta/2 \rceil - 1} \frac{(-xh_j(x))'}{z^j} c_j(z), \\ g_\beta(x) &= c_{0,g}^{-1} \left(g_1(x) I_{\mathcal{A}_z(a)} + \frac{g_0(x)}{2} I_{\mathcal{A}_z(a)^c} \right), \quad \int_0^\infty g_\beta(x)dx = 1 \end{aligned} \tag{3.54}$$

with

$$\mathcal{A}_z(a) = \left\{ x;\ |x^j f_0^{(j)}(x)| \leq a \frac{z^{j/2}}{(\log z)^{j/2}} g_0(x);\ \forall 2 \leq j \leq r;\ x^\beta L(x) \leq a \frac{z^{\beta/2}}{(\log z)^{\beta/2}} g_0(x) \right\},$$

then for all $H \geq \beta$, there exists $a > 0$,

$$\begin{aligned} \tilde{K}_z g_\beta(x) &= f_0(x) + O\left(z^{-(\beta+e\nu)/2} (\log z)^{(\beta+e\nu)} \right) x^{-\nu} [f_0(x)(1 + H_0(x, z^{-1}\log z))]^{1-\nu} \\ &\quad + O\left(\frac{f_0(x)(1 + H_0(x,1))}{z^{\beta/2}} + z^{-H} \right) \ \text{as } z \to \infty, \end{aligned} \tag{3.55}$$

where

$$H_0(x, \alpha) = \frac{\sum_{j=1}^r \alpha^j |x^j f_0^{(j)}(x)| + L(x) x^\beta \alpha^\beta (1 + x^\gamma)}{f_0(x)}.$$

Note that condition $\beta > (2-e)\nu$ holds for any $\nu \in (0, 1/3]$ if $\beta > 1$. Condition $\beta \geq 2[(1-\nu)\beta - 2\nu]$ holds if $\beta \leq \frac{4\nu}{(1-2\nu)}$. For $\nu = 1/3$, $m = \frac{3\beta}{2(\beta-1)}$.

Proof of Proposition 8. We first prove (3.55).

Using Lemma 14, we have that on $\mathcal{P}(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu)$, for any large enough $H > 0$,

$$\tilde{K}_z(-xf_0'(x)) = f_0(x) + \sum_{j=1}^{\lfloor \beta/2 \rfloor} z^{-j} h_j(x) + O\left(\frac{L(x)x^\beta + \sum_{j=0}^k x^j |f_0^{(j)}(x)|}{z^{\beta/2}} \right) + O(z^{-H})$$

with $h_j(x)$ a linear combination of $x^l f_0(x)^{(l)}$ for $l \leq 2j$. Then, using a recursive construction, we can define coefficients $c_j(z) = 1 + O(z^{-1/2})$ which are polynomial functions of $1/\sqrt{z}$ such that, if

$$g_1(x) = \frac{(-xf_0'(x))}{1+c_0(z)} + \sum_{j=1}^{r_0} \frac{(-xh_j(x))'}{z^j} c_j(z), \quad r_0 = \lceil \beta/2 \rceil - 1$$

then

$$\tilde{K}_z g_1(x) = f_0(x) + O\left(\frac{f_0(x) + \sum_{j=1}^r |x^j f_0^{(j)}(x)| + L(x)x^\beta}{z^{\beta/2}} + z^{-H} \right). \tag{3.56}$$

We now show that $\int \tilde{K}_z g_1(x)dx = 1+O(z^{-\beta/2})$. Let $H \geq \beta$, then note that $\int_{\mathbb{R}_+} |\tilde{K}_z g_1(x)|dx \leq \int_{\mathbb{R}_+} |g_1(x)|dx < \infty$ and

$$\begin{aligned}\int_{\mathbb{R}_+} \tilde{K}_z g_1(x)dx &= \int_0^{z^{H/2}} [\tilde{K}_z g_1(x) - f_0(x)]dx + \int_0^{z^{H/2}} f_0(x)dx + \int_{z^{H/2}}^\infty \tilde{K}_z g_1(x)dx \\ &= 1 + \int_{z^{H/2}}^\infty \tilde{[}K_z g_1(x) - f_0(x)]dx + O\left(z^{-H/2} + z^{-\beta/2} \int_0^{z^{H/2}} H_0(x,1)dx \right).\end{aligned}$$

We have, by the moment condition,

$$\int_{z^{H/2}}^\infty f_0(x)dx \leq z^{-H} \int_0^\infty x^2 f_0(x)dx,$$

and, since z is large,

$$\begin{aligned}\int_{z^{H/2}}^\infty \tilde{K}_z g_1(x)dx &= \int_0^\infty \int_{z^{H/2}}^\infty \int_x^\infty \frac{1}{\theta} g_{z,\epsilon}(\theta) g_1(\epsilon) d\epsilon dx \\ &\leq \frac{z}{z-1} \int_0^{z^{H/2}/2} \frac{g_1(\epsilon)}{\epsilon} \int_{z^{H/2}}^\infty Pr[\Gamma(z-1,z) \geq x/\epsilon] dx d\epsilon \\ &+ \int_{z^{H/2}/2}^\infty |g_1|(\epsilon)d\epsilon \\ &\leq \int_0^{z^{H/2}/2} \frac{|g_1(\epsilon)|}{\epsilon} \int_{z^{H/2}}^\infty e^{-czx/\epsilon}dx + \int_{z^{H/2}/2}^\infty |g_1|(\epsilon)d\epsilon \\ &\leq z^{-1}e^{-2zc} \int_{\mathbb{R}_+} |g_1(x)|dx + \int_{z^{H/2}/2}^\infty |g_1|(\epsilon)d\epsilon \\ &\lesssim e^{-2zc} \int_{\mathbb{R}_+} [g_0(x) + \sum_{j=1}^r \frac{x^j |f_0^{(j)}|(x)}{z^{j/2}}]dx \\ &+ \int_{z^{H/2}/2}^\infty [g_0(x) + \sum_{j=2}^r \frac{x^j |f_0^{(j)}|(x)}{z^{j/2}}]dx.\end{aligned}$$

Moreover

$$\begin{aligned}\int_{z^{H/2}/2}^\infty g_0(x)dx &\leq 2z^{H/2} f_0(z^{H/2}/2) + \int_{z^{H/2}/2}^\infty f_0(x)dx \\ &\leq 2z^{H/2} f_0(z^{H/2}/2) + 4z^{-H} \int_0^\infty x^2 f_0(x)dx\end{aligned}$$

and

$$x^{-2} \int_x^\infty u^2 f_0(u)du \geq \int_{x/2}^\infty f_0(u)du \geq \frac{x}{2} f_0(x/2)$$

so that

$$\int_{z^{H/2}/2}^{\infty} g_0(x)dx \lesssim z^{H/2} f_0(z^{H/2}/2) + z^{-H} \lesssim z^{-H}.$$

Also, with $q_j = (\beta/\nu + e)/j \geq 1$ and $H \geq \beta$, and using Hölder inequality,

$$\begin{aligned}
\int_{z^{H/2}/2}^{\infty} \frac{x^j |f_0^{(j)}|(x)}{z^{j/2}} dx &\leq z^{-j/2} \left(\int_{z^{H/2}/2}^{\infty} \left(\frac{x^j |f_0^{(j)}|(x)}{z^{j/2}} \right)^{q_j} dx \right)^{1/q_j} \left(\int_{z^{H/2}/2}^{\infty} g_0(x)dx \right)^{1-1/q_j} \\
&\lesssim z^{-j/2} \left(\int_{z^{H/2}/2}^{\infty} g_0(x)dx \right)^{1-1/q_j} \lesssim z^{-\beta/2}
\end{aligned}$$

for all $j \leq \beta$. We thus obtain that if $H \geq \beta$,

$$\int_{\mathbb{R}_+} \tilde{K}_z g_1(x)dx = 1 + O(z^{-\beta/2}).$$

Obviously g_1 may be negative, so that we replace it by g_β defined by (3.54). We now show that $\tilde{K}_z g_\beta$ remains a good approximation of f_0. By definition of $\mathcal{A}_z(a)$, if a is small enough, $\mathcal{A}_z(a) \subset \{x;\, g_0(x)/2 \leq g_1(x) \leq 2g_0(x)\}$.

For simplicity we write the normalisation constant as $c_0 = c_{0,g}$. Now we show that c_0 equals to 1, up to the error terms of order $z^{-\beta/2}$. We have that

$$\begin{aligned}
c_0 &= \int_{\mathcal{A}_z(a)} g_1(x)dx + 0.5 \int_{\mathcal{A}_z(a)^c} g_0(x)dx \\
&\geq \int_{\mathcal{A}_z(a)} g_1(x)dx = \int_{\mathbb{R}_+} \tilde{K}_z g_1(x)dx - \int_{\mathcal{A}_z(a)^c} g_1(x)dx
\end{aligned}$$

and

$$c_0 \leq \int_{\mathcal{A}_z(a)^c} \left(\frac{g_0(x)}{2} - g_1(x) \right) dx + \int_0^\infty g_1(x)dx.$$

Note that for large z,

$$\int_{\mathcal{A}_z(a)^c} g_1(x)dx \lesssim \int_{\mathcal{A}_z(a)^c} \left[g_0(x) + \sum_{j=2}^{r} \frac{x^j |f_0^{(j)}(x)|}{z^{j/2}} \right] dx.$$

Denote $G_0(A) = \int_A g_0(x)dx$. Using Markov inequalities, with $q_j = (\beta/\nu + e)/j \geq 1$ we have

$$\begin{aligned}
G_0[\mathcal{A}_z(a)^c] &= \int_{\mathcal{A}_z(a)^c} g_0(x)dx \\
&\lesssim z^{-(\beta/\nu+e)/2} (\log z)^{(\beta/\nu+e)/2} \sum_{j=2}^{r} \int g_0(x) \left(\frac{x^j |f_0^{(j)}(x)|}{g_0(x)} \right)^{q_j} dx \\
&\quad + z^{-(\beta/\nu+e)/2} (\log z)^{(\beta/\nu+e)/2} \int g_0(x) \left(\frac{x^\beta L(x)}{g_0(x)} \right)^{(\beta/\nu+e)/\beta} dx \\
&\lesssim z^{-(\beta/\nu+e)/2} (\log z)^{(\beta/\nu+e)/2}.
\end{aligned}$$

Similarly,

$$F_0[\mathcal{A}_z(a)^c] = \int_{\mathcal{A}_z(a)^c} f_0(x)dx \lesssim z^{-(\beta/\nu+e)/2}(\log z)^{(\beta/\nu+e)/2}.$$

Also for all $2 \le j \le r$,

$$\begin{aligned}
&\int_{\mathcal{A}_z(a)^c} x^j |f_0^{(j)}(x)| dx \\
&\lesssim [G_0(\mathcal{A}_z(a)^c)]^{\frac{(\beta/\nu+e-j)}{(\beta/\nu+e)}} \left[\int g_0(x) \left(\frac{x^j|f_0^{(j)}(x)|}{g_0(x)}\right)^{(\beta/\nu+e)/j} dx\right]^{j/(\beta/\nu+e)} \\
&\lesssim z^{-(\beta/\nu+e-j)/2}(\log z)^{(\beta/\nu+e-1)/2}.
\end{aligned}$$

This implies that

$$\int_{\mathcal{A}_z(a)^c} g_1(x)dx \lesssim z^{-\beta/\nu}, \quad \int_{\mathcal{A}_z(a)^c} \left[\frac{g_0(x)}{2} - g_1(x)\right] dx \lesssim z^{-\beta/\nu}, \tag{3.57}$$

and hence

$$1 + O(z^{-\beta/2}) \le c_0 \le 1 + O(z^{-\beta/2}). \tag{3.58}$$

Therefore, we can write

$$\tilde{K}_z g_\beta = c_0^{-1}\tilde{K}_z g_1 + c_0^{-1}\tilde{K}_z[(g_0/2 - g_1)I_{\mathcal{A}_z(a)^c}] =: c_0^{-1}(\tilde{K}_z g_1 + \Delta_z). \tag{3.59}$$

Using Lemma C.2 of [4] (see Lemma 34 in Section 3.11), we have if $\delta_z = \sqrt{2H\log z/z}$,

$$\begin{aligned}
|\Delta_z(x)| &= |\int_{\mathcal{A}_z(a)^c} [g_0(\epsilon)/2 - g_1(\epsilon)] \int_x^\infty \theta^{-1} g_{z,\epsilon}(\theta) d\theta d\epsilon| \\
&= (1 + O(z^{-\beta/2}))|\int_{\mathcal{A}_z(a)^c} \frac{[g_0(\epsilon)/2 - g_1(\epsilon)]}{\epsilon} \int_{x/\epsilon}^\infty \frac{z^z u^{z-2} e^{-zu}}{\Gamma(z)} du d\epsilon| \\
&\le 2\int_{x/(1+\delta_z^2)}^\infty I_{\mathcal{A}_z(a)^c} \frac{|g_0(\epsilon)/2 - g_1(\epsilon)|}{\epsilon} d\epsilon \\
&+ z^{-H} \int_0^{x/(1+\delta_z^2)} I_{\mathcal{A}_z(a)^c}(\epsilon) \frac{|g_0(\epsilon)/2 - g_1(\epsilon)|}{\epsilon} d\epsilon.
\end{aligned}$$

The first integral is bounded by, up to a constant,

$$\begin{aligned}
&\int_{x/(1+\delta_z^2)}^\infty I_{\mathcal{A}_z(a)^c}(\epsilon) \left| -|f_0'(\epsilon)| + \sum_{j=2}^r \frac{|\epsilon^{j-1} f_0^{(j)}(\epsilon)|}{z^{\lceil j/2 \rceil}} c_j(z) \right| d\epsilon \\
&\lesssim \int_{x/(1+\delta_z^2)}^\infty I_{\mathcal{A}_z(a)^c}(\epsilon)[-f_0'(\epsilon)] d\epsilon + \sum_{j=2}^r z^{-\lceil j/2 \rceil} \int_{x/(1+\delta_z^2)}^\infty I_{\mathcal{A}_z(a)^c}(\epsilon) \epsilon^{j-1} |f_0^{(j)}(\epsilon)| d\epsilon.
\end{aligned}$$

Using that f_0 decreases and applying Hölder inequality with exponent $1/\nu$ for $\nu \in (0,1)$,

$$\begin{aligned}
&\int_{x/(1+\delta_z^2)}^{\infty} I_{\mathcal{A}_z(a)^c}(\epsilon)|f_0'(\epsilon)|d\epsilon \\
&\lesssim \left[\int_{x/(1+\delta_z^2)}^{\infty} I_{\mathcal{A}_z(a)^c}(\epsilon)[-f_0'(\epsilon)]d\epsilon\right]^{\nu}\left[\int_{x/(1+\delta_z^2)}^{\infty} [-f_0'(\epsilon)]d\epsilon\right]^{1-\nu} \\
&\lesssim [x^{-1}G_0(\mathcal{A}_z(a)^c)]^{\nu}\left[f_0(x/(1+\delta_z^2))\right]^{1-\nu} \\
&\lesssim z^{-(\beta+e\nu)/2}(\log z)^{(\beta+e\nu)/2}x^{-\nu}\left[f_0(x/(1+\delta_z^2))\right]^{1-\nu}.
\end{aligned}$$

Similarly, applying Hölder inequality twice, with $q_j = (\beta/\nu + e)/j > 1/\nu$ and with $1/\nu_j$ with $\nu_j = \frac{\nu - 1/q_j}{1-1/q_j} \in (0,1)$,

$$\begin{aligned}
&\int_{x/(1+\delta_z^2)}^{\infty} I_{\mathcal{A}_z(a)^c}(\epsilon)\frac{\epsilon^{j-1}|f_0^{(j)}(\epsilon)|}{-f_0'(\epsilon)}[-f_0'(\epsilon)]d\epsilon \\
\lesssim&\left[\int_{x/(1+\delta_z^2)}^{\infty} I_{\mathcal{A}_z(a)^c}[-f_0'(\epsilon)]d\epsilon\right]^{1-1/q_j} x^{-1/q_j}\left[\int_0^{\infty}\left[\frac{\epsilon^j|f_0^{(j)}(\epsilon)|}{g_0(\epsilon)}\right]^{q_j} g_0(\epsilon)d\epsilon\right]^{1/q_j} \\
\lesssim& x^{-\nu_j(1-1/q_j)}[z/\log z]^{-0.5(\beta+e\nu)\nu_j(1-1/q_j)/\nu}\left[f_0(x/(1+\delta_z^2))\right]^{(1-\nu_j)(1-1/q_j)} x^{-1/q_j} \\
\lesssim& x^{-\nu}z^{-0.5(\beta+e\nu)+j/2}(\log z)^{0.5(\beta+e\nu)-j/2}\left[f_0(x/(1+\delta_z^2))\right]^{(1-\nu)}.
\end{aligned}$$

Therefore, the first integral is bounded by

$$\begin{aligned}
&\lesssim z^{-(\beta+e\nu)/2}(\log z)^{(\beta+e\nu)/2}x^{-\nu}\left[f_0(x/(1+\delta_z^2))\right]^{1-\nu}[1+\sum_{j=2}^{r} z^{-\lceil j/2\rceil}z^{j/2}(\log z)^{-j/2}] \\
&\lesssim z^{-(\beta+e\nu)/2}(\log z)^{(\beta+e\nu)/2}x^{-\nu}\left[f_0(x/(1+\delta_z^2))\right]^{1-\nu}.
\end{aligned}$$

Moreover,

$$f_0(x/(1+\delta_z^2)) = \sum_{\ell=0}^{r}\frac{\delta_z^{2\ell}}{\ell!}f_0^{(\ell)}(x) + O(L(x)\delta_z^{2\beta}) \le f_0(x)(1+H_0(x,\delta_z^2)).$$

The second integral is bounded by z^{-H} multiplied by

$$\begin{aligned}
&\int_0^{x/(1+\delta_z^2)} I_{\mathcal{A}_z(a)^c}(\epsilon)\left[\frac{-|f_0'(\epsilon)|(2-c_0(z))}{0.5(1+c_0(z))}+\sum_{j=2}^{r}\frac{|\epsilon^{j-1}f_0^{(j)}(\epsilon)|}{z^{\lceil j/2\rceil}}c_j(z)\right]d\epsilon \\
&\lesssim f_0(0)+\sum_{j=2}^{r} z^{-\lceil j/2\rceil}\left[\int_0^{\infty}\left[\frac{1+\epsilon^j|f_0^{(j)}(\epsilon)|}{f_0(\epsilon)}\right]^{q_j} f_0(\epsilon)d\epsilon\right]^{1/q_j} \\
&= O(1)
\end{aligned}$$

by the moment conditions of $\mathcal{P}(\beta, L(\cdot), \gamma, C_0, C_1, e, \Delta, \nu)$ and decreasing f_0. Therefore we obtain, writing $e' = e\nu$,

$$|\Delta_z(x)| \lesssim z^{-H} + z^{-(\beta+e')/2}(\log z)^{(\beta+e')/2} x^{-\nu} \left[f_0(x)(1 + H_0(x, \delta_z^2))\right]^{1-\nu}. \tag{3.60}$$

Finally combining (3.56) with (3.58), (3.59) and (3.60) we prove (3.55).

□

Lemma 14. *Assume that f_0 satisfies the following local Hölder property*

$$f_0(y) - f_0(x) = \sum_{j=1}^{\lfloor\beta\rfloor} \frac{(y-x)^j}{j!} f_0^{(j)}(x) + R(x,y), \quad |R(x,y)| \leq L(x)|y-x|^\beta(1 + |x-y|^\gamma) \tag{3.61}$$

where $\beta > 1$ and $\gamma \geq 0$.

Then, for any $H > 0$,

$$\tilde{K}_z(-xf_0'(x)) = f_0(x) + \sum_{j=1}^{\lceil\beta/2\rceil-1} z^{-j} h_j(x) + O\left(\frac{L(x)x^\beta + \sum_{j=0}^{k} x^j |f_0^{(j)}(x)|}{z^{\beta/2}}\right) + O(z^{-H})$$

with $h_j(x)$ a linear combination of $x^l f_0(x)^{(l)}$ for $l \leq 2j$.

Note that for $\beta \leq 2$ and for any $H > 0$,

$$\tilde{K}_z(-xf_0'(x)) = f_0(x) + O\left(\frac{L(x)x^\beta + \sum_{j=0}^{1} x^j |f_0^{(j)}(x)|}{z^{\beta/2}}\right) + O(z^{-H}) \quad \text{as } z \to \infty.$$

Proof of Lemma 14. We apply Lemma 22 to $-xf_0'(x)$, this leads to

$$\begin{aligned}
&\tilde{K}_z(-xf_0'(x)) \\
&= \int_{\mathbb{R}_+} \frac{zf_0(\epsilon)}{\epsilon} \left(Pr(X_1 > x | X_1 \sim \Gamma(z, z/\epsilon)) - Pr(X_2 > x | X_2 \sim \Gamma(z-1, z/\epsilon))\right) d\epsilon.
\end{aligned}$$

Note that X_1 has the same distribution as $\epsilon(Y + Y_1)$ with $Y \sim \Gamma(z-1, z)$ and $Y_1 \sim \Gamma(1, z)$, $Y \perp Y_1$ and that X_2 has the same distribution as ϵY. Therefore

$$\begin{aligned}
&Pr(X_1 > x | X_1 \sim \Gamma(z, z/\epsilon)) - Pr\left(X_2 > x | X_2 \sim \Gamma(z-1, z/\epsilon)\right)) \\
&= P\left(\frac{x}{\epsilon} - Y_1 < Y < \frac{x}{\epsilon}\right) \\
&= P_Z\left(\sqrt{z}(x/\epsilon - 1) - \sqrt{z}Y_1 < Z < \sqrt{z}(x/\epsilon - 1)\right)
\end{aligned}$$

with $Z \sim \sqrt{z}(\Gamma(z,z) - 1)$. Let $u = \sqrt{z}(x/\epsilon - 1)$, then $du = \sqrt{z}(1 + u/\sqrt{z})d\epsilon/\epsilon$ and

$$\tilde{K}_z(-xf_0'(x)) = z^{1/2} \int_{-\sqrt{z}}^{\infty} (1 + u/\sqrt{z})^{-1} f_0(x(1 + u/\sqrt{z})^{-1}) P_Z(u - \sqrt{z}Y_1 < Z < u) du$$

Note that $Z \Rightarrow \mathcal{N}(0,1)$ and that $\sqrt{z}Y_1 = O_p(1/\sqrt{z})$ with $\sqrt{z}E[Y_1] = z^{-1/2}$ and $zE[Y_1^2] = 2/z$.

Let $|u| \geq M_0 \log z$, then there exists $c > 0$ such that if $u > 0$

$$\begin{aligned}
P_Z(u - \sqrt{z}Y_1 < Z < u) &\leq P_Z(u/2 < Z < u) + P_Z(\sqrt{z}Y_1 > u/2) \\
&\leq \frac{z^z}{\Gamma(z)} \int_{u/2}^{u} e^{-z(t/\sqrt{z}+1)}(t/\sqrt{z}+1)^{z-1}dt + e^{-\sqrt{z}u/2} \\
&\leq \mathbf{1}_{|u|\leq\sqrt{z}/4}\varphi(u/4) + \mathbf{1}_{|u|>\sqrt{z}/4}e^{-cu\sqrt{z}}.
\end{aligned} \tag{3.62}$$

and if $u < 0$

$$\begin{aligned}
P_Z(u - \sqrt{z}Y_1 < Z < u) &\leq P_Z(2u < Z < u) + P_Z(\sqrt{z}Y_1 > |u|) \\
&\leq \frac{z^z}{\Gamma(z)} \int_{2u}^{u} \mathbf{1}_{t>-\sqrt{z}}e^{-z(t/\sqrt{z}+1)}(t/\sqrt{z}+1)^{z-1}dt + e^{-\sqrt{z}|u|/2} \\
&\leq \mathbf{1}_{|u|\leq\sqrt{z}/4}\varphi(u/4) + \mathbf{1}_{u>-\sqrt{z}/4}(3e^{1/4}/4)^z\sqrt{z}
\end{aligned} \tag{3.63}$$

If $|u| \leq M_0 \log z$, using a Taylor expansion of f_0

$$f_0(x(1+u/\sqrt{z})^{-1}) = f_0(x) + \sum_{j=1}^{r} \frac{(-1)^j x^j f_0^{(j)}(x)u^j}{z^{j/2}j!(1+u/\sqrt{z})^j} + R_z(x,u)$$

with

$$|R_z(x,u)| \leq L(x)|u|^\beta z^{-\beta/2}(1+|u|^\gamma z^{-\gamma/2}), r = \lceil\beta\rceil - 1.$$

Note also that, for any $k > 3$ and if $b_z = H_0 \log z$, with $H_0 \geq H$,

$$\begin{aligned}
P_Z(u - \sqrt{z}Y_1 < Z < u) &= z^{1/2} \times \frac{z^z}{\Gamma(z)} \int_{\mathbb{R}+} e^{-zy_1} \int_{u-\sqrt{z}y_1}^{u} e^{-z(t/\sqrt{z}+1)}(t/\sqrt{z}+1)^{z-1}dt \\
&= z^{1/2} \times \frac{z^z}{\Gamma(z)} \int_{\mathbb{R}+} e^{-zy_1} \int_{u-\sqrt{z}y_1}^{u} e^{-z[(t/\sqrt{z}+1)-(1-1/z)\log(t/\sqrt{z}+1)]}dt \\
&= z^{1/2} \times \frac{z^z e^{-z}}{\Gamma(z)} \int_{0}^{(1+b_z)/z} e^{-zy_1} \int_{u-\sqrt{z}y_1}^{u} e^{-t/\sqrt{z}-(1-1/z)[t^2/2+P_k(t;z)]}dt + J_2
\end{aligned}$$

with

$$P_k(t;z) = \sum_{j=1}^{k} \frac{(-1)^j t^{j+2}}{(j+2)z^{j/2}}$$

and

$$\begin{aligned}
J_2 &= z^{1/2} \times \frac{z^z}{\Gamma(z)} \int_{(1+b_z)/z}^{+\infty} e^{-zy_1} \int_{u-\sqrt{z}y_1}^{u} e^{-z[(t/\sqrt{z}+1)-(1-1/z)\log(t/\sqrt{z}+1)]}dt \\
&\leq P[zY_1 - 1 > b_z] = e^{-(1+b_z)} \leq z^{-H}.
\end{aligned}$$

if $b_z = H_0 \log z$ with $H_0 \geq H$. This implies that

$$
\begin{aligned}
& P_Z(u - \sqrt{z}Y_1 < Z < u) \\
& = (1 + d(z)) \int_0^{1+b_z} e^{-y_1} \int_{u-y_1/\sqrt{z}}^{u} \varphi(t) e^{-t/\sqrt{z}+(1-1/z)P_k(t;z)} dt + O(z^{-H}) \\
& = (1 + d(z))\varphi(u) \left(\frac{1}{\sqrt{z}} + Q_k(u; z) + O(u^{k+1}/z^{k+1}) \right) + O(z^{-H})
\end{aligned}
$$

where $d(z) = O(1/z)$ is a polynomial function of $1/z$ and $Q_k(u; z)$ has the form

$$
Q_k(u; z) = \sum_{j=1}^{k} \frac{Q_j(u)}{z^{(j+1)/2}}, \quad Q_j \text{ is a polynomial function with degree less than } j + 2
$$

Combining the above computations with (3.62) and (3.63), we finally obtain that for z large enough and $k \geq 2H + 1$,

$$
\begin{aligned}
\tilde{K}_z(-xf_0'(x)) &= z^{1/2} \int_{-\sqrt{z}}^{+\infty} (1 + u/\sqrt{z})^{-1} f_0(x(1 + u/\sqrt{z})^{-1}) P_Z(u - \sqrt{z}Y_1 < Z < u) du \\
&= z^{1/2} \int_{|u| \leq M_0 \log z} (1 + u/\sqrt{z})^{-1} f_0(x(1 + u/\sqrt{z})^{-1}) P_Z(u - \sqrt{z}Y_1 < Z < u) du + z^{-H} \\
&= f_0(x) \left(\int_{|u| \leq M_0 \log z} (1 + u/\sqrt{z})^{-1} (1 + d(z))\varphi(u) \left(1 + \sqrt{z}Q_k(u; z)\right) du \right) \\
&+ (1 + c(z))(1 + d(z)) \sum_{j=1}^{r} \int_{|u| \leq M_0 \log z} \frac{f_0^{(j)}(x)(-1)^j x^j u^j}{z^{j/2}(1 + u/\sqrt{z})^{j+1}} \varphi(u) \left(1 + \sqrt{z}Q_k(u; z)\right) \\
&+ O\left(\frac{L(x)x^\beta}{z^{\beta/2}}\right) + O(z^{-H}) \\
&= f_0(x) + \sum_{j=1}^{\lceil \beta/2 \rceil - 1} z^{-j} h_j(x) + O\left(\frac{L(x)x^\beta + \sum_{j=0}^{k} x^j |f_0^{(j)}(x)|}{z^{\beta/2}}\right) + O(z^{-H})
\end{aligned}
$$

with $h_j(x)$ a linear combination of $x^l f_0^{(l)}(x)$ for $l \leq 2j$.

$\square$

3.6.3 Discrete pointwise approximation

We now construct the discrete pointwise approximation.

Lemma 15. *Assume $f \in \mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$. Let $e_z = e_0 z^{-\beta/4}$. with e_0 a small constant and E_z satisfying $f_0(E_z) = z^{-\beta/\nu}$.*

Then for any $H > 0$ there exists $P_N = \sum_{i=1}^{N} p_i \delta_{u_i}$ with $u_i \in [e_z, E_z]$, $N \leq N_0 \sqrt{z} (\log z)^{3/2}$ such that

$$
|\tilde{K}_z P_N(x) - \tilde{K}_z g_\beta(x)| \lesssim z^{-H}, \quad \forall x \geq 0, \tag{3.64}
$$

where g_β is defined in Proposition 8.

Proof of Lemma 15. Throughout the proof $c > 0$ denotes a generic constant. Denote

$$G_0(A) = \int_A g_0(x)dx, \quad G_\beta(A) = \int_A g_\beta(x)dx. \tag{3.65}$$

First note that, using $\int f_0(x)^{1-2\nu}dx < \infty$, with $\nu < 1/2$, we have

$$\begin{aligned} 1 - F_0(E_z) &= \int_{E_z}^\infty f_0(x)dx = o(f_0(E_z)^{2\nu}) = o(z^{-2\beta}), \\ E_z &= o(z^{\beta/\nu}) \end{aligned} \tag{3.66}$$

where the latter comes from the fact that $xf_0(x) \to 0$ at infinity. Note also that

$$\begin{aligned} 1 - G_0(E_z) &\leq \left(\int \left(\frac{-xf_0'(x)}{f_0(x)} \right)^{\beta/\nu} f_0(x)dx \right)^{\nu/\beta} (1 - F_0(E_z))^{1-\nu/\beta} \\ &= o(z^{-2\beta+2\nu}) = o(z^{-\beta/2}). \end{aligned} \tag{3.67}$$

Define

$$\bar{g_\beta} = \frac{g_\beta \mathbb{1}([e_z, E_z])}{G_\beta([e_z, E_z])} := \frac{g_\beta \mathbb{1}([e_z, E_z])}{c_\beta}, \quad G_\beta(A) = \int_A g_\beta(x)dx, \tag{3.68}$$

then

$$\begin{aligned} \int_0^{e_z} g_\beta(\epsilon)d\epsilon &\lesssim \int_0^{e_z} [g_0(u) + \sum_{j=1}^r z^{-j/2}u^j |f_0^{(j)}(u)|]du \\ &\leq f_0(0)e_z^2 + \sum_{j=1}^r z^{-j/2} \int_0^{e_z} u^j |f_0^{(j)}(u)|du. \end{aligned} \tag{3.69}$$

Since for all $u < 1$

$$f_0^{(j)}(u) = f_0^{(j)}(0) + \sum_{\ell=j+1}^r f_0^{(j+\ell)}(0)\frac{u^\ell}{\ell!} + O(L(0)u^{\beta-j}) \tag{3.70}$$

we have

$$\int_0^{e_z} g_\beta(\epsilon)d\epsilon \lesssim e_z^2 + \sum_{j=2}^r z^{-j/2}e_z^{j+1} \lesssim e_z^2.$$

Moreover, recall that $q_j = (\beta/\nu + e)/j$, and we have, using (3.67),

$$\begin{aligned} \int_{E_z}^\infty g_\beta(\epsilon)d\epsilon &\leq 1 - G_0(E_z) + O\left(\sum_{j=2}^\infty z^{-j/2} \int_{E_z} \epsilon^j |f_0^{(j)}(\epsilon)|d\epsilon \right) \\ &= o(z^{-\beta/2}) + O\left(\sum_{j=2}^r [1 - G_0(E_z)]^{1-1/q_j} z^{-j/2} \right) = o(z^{-\beta/2}), \end{aligned}$$

since since $\beta/\nu + e > \beta$ for $\nu \in [0,1)$ and $e > 0$. Therefore,

$$c_\beta = \int_{e_z}^{E_z} g_\beta(x)dx \geq 1 - z^{-\beta/2}.$$

We can then bound

$$\begin{aligned}
\left|\tilde{K}_z g_\beta - \tilde{K}_z \bar{g_\beta}\right|(x) &\leq \int_{e_z}^{E_z} g_\beta(\epsilon) \int_x^\infty \frac{1}{\theta} g_{z,z/\epsilon}(\theta)d\theta \frac{|c_\beta - 1|}{c_\beta} \\
&+ \int_0^{e_z} g_\beta(\epsilon) \int_x^\infty \frac{1}{\theta} g_{z,\epsilon}(\theta)d\theta + \int_{E_z}^\infty g_\beta(\epsilon) \int_x^\infty \frac{1}{\theta} g_{z,\epsilon}(\theta)d\theta \\
&\leq 2z^{-\beta/2}\tilde{K}_z g_\beta(x) + \frac{z}{z-1}\int_{E_z}^\infty \frac{g_\beta(\epsilon)}{\epsilon}\int_{x/\epsilon}^\infty g_{\Gamma(z-1,z)}(u)du \\
&+ \frac{z}{z-1}\int_0^{e_z} \frac{g_\beta(\epsilon)}{\epsilon}\int_{x/\epsilon}^\infty g_{\Gamma(z-1,z)}(u)du \\
&\leq 2z^{-\beta/2}\tilde{K}_z g_\beta(x) + \frac{z}{z-1} f_0(E_z) + \frac{(f_0(0) - f_0(e_z))z}{z-1} I_{x \leq 2e_z} + I_{x > 2e_z} e^{-czx/e_z} \\
&\lesssim z^{-\beta/2}\tilde{K}_z g_\beta(x) + f_0(E_z) + e_z I_{x \leq 2e_z} + I_{x > 2e_z} e^{-czx/e_z}
\end{aligned} \tag{3.71}$$

for some $c > 0$.

Using Lemma B.1 of [4] (see Lemma 32 in Section 3.11), for all $H > 0$, there exists P_N with at most $N_0\sqrt{z}(\log z)^{3/2}$ supporting points such that for all $\theta \in [e_z/2, 2E_z]$, $|K_z * g_\beta - K_z * P_N|(\theta) \leq z^{-2H}$. This implies that for all $x \geq e_z/2$,

$$\int_x^{2E_z} \frac{1}{\theta}|K_z * g_\beta - K_z * P_N|(\theta)d\theta \leq z^{-2H}(\log E_z - \log x) \leq z^{-H}$$

where the latter inequality comes from (3.66). We also have

$$\begin{aligned}
\int_{2E_z}^\infty \frac{1}{\theta}[K_z * g_\beta + K_z * P_N](\theta)d\theta &\leq \int_{e_z}^{E_z}[g_\beta(\epsilon)d\epsilon + P(d\epsilon)]\left(\int_{2E_z/\epsilon} u^{z-2}e^{-zu}\frac{z^z}{\Gamma(z)}du\right) \\
&\leq e^{-c' E_z/e_z} \leq e^{-cz^{\beta/4}}
\end{aligned}$$

for some $c > 0$, when z is large enough. Hence when z is large enough, for all $x \geq e_z/2$,

$$\left|\tilde{K}_z g_\beta(x) - \tilde{K}_z P_N(x)\right| \leq z^{-H}.$$

Now let $x < e_z/2$, then there exists $c > 0$ such that

$$\begin{aligned}
\left|\tilde{K}_z g_\beta(x) - \tilde{K}_z P_N(x)\right| &\leq \frac{z^z\Gamma(z-1)}{z^{z-1}\Gamma(z)}\left|\int_{e_z}^{E_z}\frac{1}{\epsilon}[g_\beta(\epsilon)d\epsilon - dP_N(\epsilon)]\right| \\
&+ \frac{z}{z-1}\int_{[e_z,E_z]^c}\frac{1}{\epsilon}[g_\beta(\epsilon)d\epsilon + dP_N(\epsilon)](x/\epsilon)^{cz}d\epsilon.
\end{aligned}$$

The first term of the right hand side is zero by construction of P_N and the second is bounded by

$$2^{-cz}e_z^{-1} \leq 2^{-cz/2}, \quad \text{since} \quad e_z \geq z^{-\beta/2}.$$

We thus obtain that for all $H > 0$, there exists P_N with at most $\sqrt{z}(\log z)^{3/2}$ supporting points in $[e_z, E_z]$ such that for all $x \in \mathbf{R}_+$

$$\left|\tilde{K}_z g_\beta(x) - \tilde{K}_z P_N(x)\right| \leq z^{-H}, \tag{3.72}$$

which in turn implies Lemma 15.

□

Lemma 16. *Assume $f_0 \in \mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$ and with $\nu \leq 1/3$ and that*

$$\int x^m f_0(x)dx < \infty, \quad m = 2\beta/((1-\nu)\beta - 2\nu).$$

If P satisfies $P[e_z; E_z] = 1$, where e_z, E_z are defined as in Lemma 15 and

$$\sup_x \left|\tilde{K}_z g_\beta(x) - \tilde{K}_z P(x)\right| \leq z^{-H},$$

for $H \geq 2\beta/\nu$, with g_β as defined in Proposition 8, then,

$$d_H^2(f_0, \tilde{K}_z P) \lesssim z^{-\beta}, \quad KL(f_0, \tilde{K}_z P) \lesssim z^{-\beta}, \quad V_0(f_0, \tilde{K}_z P) \lesssim z^{-\beta}.$$

Proof of Lemma 16.

$$KL(f_0,\tilde{K}_z P) \leq KL(f_0, \tilde{K}_z g_\beta) + \left|\int f_0[\log \tilde{K}_z P - \log \tilde{K}_z g_\beta]\right|.$$

Recall that $g_\beta \geq g_0/2$ and that, due to Lemma 14,

$$|\tilde{K}_z g_0 - f_0(x)| \lesssim \sum_{j=1}^{r_0} z^{-j}\left(jx^{2j}|f_0^{(2j)}(x)| + jx^{2j+1}|f_0^{(2+1)}(x)|\right) + z^{-\beta}L(x)x^\beta.$$

Therefore on the set

$$\tilde{A}_z(a) = \{x^{2j}j|f_0^{(2j)}(x)| + x^{2j+1}j|f_0^{(2j+1)}(x)| \leq az^j f_0(x),\, L(x)x^\beta \leq az^\beta f_0(x)\}$$

for $1 \leq j \leq r_0$, and if a is small enough,$\tilde{K}_z g_0 \geq f_0/2$. Moreover using Lemma B2 of [47]

$$\int_{\tilde{A}_z(a)} f_0(\log f_0/\tilde{K}_z g_\beta) \lesssim d_H^2(\bar{f}_0, \bar{f}_\beta) + F_0(\tilde{A}_z(a))[\log(F_0(\tilde{A}_z(a)) - \log \tilde{K}_z g_\beta(\tilde{A}_z(a))],$$

where

$$\bar{f}_0 = f_0 \mathbf{1}_{\tilde{A}_z(a)}/F_0(\tilde{A}_z(a)), \quad \bar{f}_\beta = \tilde{K}_z g_\beta \mathbf{1}_{\tilde{A}_z(a)}/\tilde{K}_z g_\beta(\tilde{A}_z(a)).$$

Markov's inequality implies that

$$F_0(\tilde{A}_z(a)^c) \lesssim z^{-\beta/\nu - e}.$$

Moreover

$$\begin{aligned}\tilde{K}_z\bar{g}_\beta(\tilde{A}_z(a)^c) &\lesssim \tilde{K}_z g_\beta(\tilde{A}_z(a)^c)\\ &\lesssim \int_0^{z^{H/2}} \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x)(1+H_0(x,1)z^{-\beta/2})dx\\ &+ \int_0^{z^{H/2}} \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x)^{1-\nu} z^{-\beta/(2\nu)}(1+H_0(x,1))dx + z^{-H/2}\end{aligned}$$

and using Markov's inequality

$$\tilde{K}_z\bar{g}_\beta(\tilde{A}_z(a)^c) \lesssim z^{-\beta/\nu}.$$

Now

$$d_H^2(\bar{f}_0,\bar{f}_\beta) \le \int_0^{E_z} \frac{(\bar{f}_0-\bar{f}_\beta)^2}{\bar{f}_0(x)+\bar{f}_\beta(x)}dx + \int_{E_z}^{\infty}\left(\sqrt{\bar{f}_0}-\sqrt{\bar{f}_\beta}\right)^2(x)dx$$

Using a similar argument as in the proof of Proposition 8, $\bar{f}_\beta \le 2f_0$ and

$$d_H^2(\bar{f}_0,\bar{f}_\beta) \lesssim \int_0^{E_z} \frac{(\bar{f}_0-\bar{f}_\beta)^2}{\bar{f}_0(x)+\bar{f}_\beta(x)}dx + z^{-\beta}.$$

Moreover, from Proposition 8,

$$\begin{aligned}&\int_0^{E_z} \frac{(f_0(x)-\tilde{K}_z\bar{g}_\beta(x))^2}{f_0(x)}dx\\ &\le 2\int_0^{E_z} \frac{(f_0(x)-\tilde{K}_z g_1(x))^2}{f_0(x)}dx + 2(1-c_0)^2 + \int_0^{E_z}\frac{\Delta_z(x)^2}{f_0(x)}dx\\ &\le 2\frac{z^{-2H}E_z}{f_0(E_z)} + 2z^{-\beta}\int(1+H_0^2(x,1))f_0(x)dx + \int_0^{E_z}\frac{\Delta_z(x)^2}{f_0(x)}dx.\end{aligned}$$

Also

$$\begin{aligned}&\int_0^{E_z}\frac{\Delta_z(x)^2}{f_0(x)}dx \le z^{-2H}\int_0^{E_z} f_0^{-1}(x)dx\\ &+ z^{-(\beta+e\nu)}(\log z)^{(\beta+e\nu)}\int_0^{E_z} f_0^{-1}(x)x^{-2\nu}\left[f_0(x)(1+H_0(x,\log z/z))\right]^{2(1-\nu)}dx\\ &\lesssim z^{-2H}E_z f_0^{-1}(E_z) + z^{-(\beta+e\nu/2)}\int_0^{E_z} f_0^{1-2\nu}(x)x^{-2\nu}dx + o(z^{-(\beta+e\nu)})\\ &\lesssim o(z^{-2H+2\beta/\nu}) + z^{-(\beta+e\nu/2)} = o(z^{-\beta})\end{aligned}$$

as long as $2H - 2\beta/\nu \geq \beta$ and due to

$$\begin{aligned}
&\int_0^{E_z} f_0^{1-2\nu}(x) x^{-2\nu} [H_0(x, z^{-1}\log z)]^{2(1-\nu)} dx \\
&\lesssim \sum_{j=1}^{r} [z^{-1}\log z]^{2(1-\nu)j} \int_0^{E_z} f_0^{1-2\nu}(x) \left[\frac{(1+x^j)|f_0^{(j)}(x)|}{f_0(x)}\right]^{2(1-\nu)} dx \\
&+ [z^{-1}\log z]^{2(1-\nu)\beta} \int_0^{E_z} f_0^{1-2\nu}(x) \left[\frac{L(x)(1+x^{\beta})(1+x^{\gamma})}{f_0(x)}\right]^{2(1-\nu)} dx
\end{aligned}$$

and

$$\begin{aligned}
&[z^{-1}\log z]^{2(1-\nu)j} \int_0^{E_z} f_0^{-1}(x) \left[(1+x^j)|f_0^{(j)}(x)|\right]^{2(1-\nu)} dx \\
&\leq [z^{-1}\log z]^{2(1-\nu)j} \left[\int_0^{E_z} f_0^{\frac{[-1+(q_j-1)/\tilde{q}_j]}{(1-1/\tilde{q}_j)}}(x) dx\right]^{1-1/\tilde{q}_j} \left[\int_0^{E_z} \left[\frac{(1+x^j)|f_0^{(j)}(x)|}{f_0(x)}\right]^{q_j} f_0(x) dx\right]^{1/\tilde{q}_j} \\
&\lesssim [z^{-1}\log z]^{2(1-\nu)j} [f_0(E_z)]^{-2\nu/\tilde{q}_j} \left[\int_0^{E_z} f_0^{1-2\nu}(x) dx\right]^{1-1/\tilde{q}_j} \\
&[\log z]^{2(1-\nu)j} z^{-2(1-\nu)j[1-2\beta/(\beta/\nu+e)]} = o(1)
\end{aligned}$$

for $\nu \in (0, 1/2]$ where $\tilde{q}_j = q_j/[2(1-\nu)] > 1$. Therefore

$$KL(f_0, \tilde{K}_z \bar{g}_\beta) \lesssim z^{-\beta} + \int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x)[|\log f_0(x)| + |\log \tilde{K}_z g_\beta(x)|] dx \tag{3.73}$$

Since for any $t > 0$ on $\tilde{A}_z(a)^c$, $f_0 \log f_0 = o(f_0^{1-t})$, we have for any $t > 0$, using Hölder inequality and $\int f_0(x)^{1-2\nu} dx < \infty$,

$$\int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) |\log f_0(x)| dx = o(F_0(\tilde{A}_z(a)^c)^{\nu - t}) = o(z^{-\beta}).$$

Finally, for any small constant $b > e_z$ there exists a constant $c > 0$ such that

$$\begin{aligned}
K_z \bar{g}_\beta(x) &= \frac{z}{z-1} \int_{e_z}^{E_z} \frac{1}{\epsilon} \int_{x/\epsilon}^{\infty} g_{\Gamma(z-1,z)}(u) du g_\beta d(\epsilon) \\
&\gtrsim Pr(\Gamma(z-1, z) \geq x/b) \int_b^{E_z} \frac{g_\beta(\epsilon)}{\epsilon} d\epsilon \\
&\gtrsim e^{-czx} E_z{-1} [G_\beta(E_z) - G_\beta(b)] \gtrsim e^{-czx} E_z{-1},
\end{aligned} \tag{3.74}$$

and and hence for some $m > 1$

$$\begin{aligned}
\int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x)|\log \tilde{K}_z g_\beta(x)|dx &\lesssim z \int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) x dx + \log z \int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) dx \\
&\leq o(z^{-\beta/\nu}) + z[F_0(\tilde{A}_z(a)^c)]^{1-1/m} \left(\int x^m f_0(x) dx \right)^{1/m} \\
&\leq o(z^{-\beta/\nu}) + z^{1-(\beta/\nu+e)(1-1/m)} \\
&= o(z^{-\beta})
\end{aligned}$$

as long as $\int x^m f_0(x)dx < \infty$ and $1 - (\beta/\nu + e)(1 - 1/m) < -\beta$, i.e. we can take any $m \geq \beta/[\beta(1-\nu) - \nu]$. We take $m = 2\beta/[\beta(1-\nu) - 2\nu]$, as assumed in the lemma. We also have that

$$KL(f_0, \tilde{K}_z \bar{g}_\beta) \lesssim z^{-\beta}.$$

Since (3.74) can also be used for $\log K_z P$ and since $|K_z g_\beta - K_z \bar{P}| \leq z^{-H}$ for H arbitrarily large, on $\tilde{A}_z(a) \cap \{f_0 \geq z^{-H/2}\}$, $\tilde{K}_z g_\beta \gtrsim f_0$ and $\tilde{K}_z P \gtrsim f_0$ therefore

$$\begin{aligned}
\int f_0 |\log \tilde{K}_z P - \log \tilde{K}_z g_\beta| &\lesssim z^{-H/2} + z^{-\beta} \\
&\quad + \int \mathbf{1}_{\{f_0(x) \geq z^{-H/2}\}} f_0(x) |\log \tilde{K}_z P - \log \tilde{K}_z g_\beta|(x) dx \\
&\lesssim z \int \mathbf{1}_{\{f_0(x) \geq z^{-H/2}\}} x f_0(x) dx + z^{-\beta} \lesssim z^{-H\nu/2} + z^{-\beta}
\end{aligned}$$

and

$$KL(f_0, \tilde{K}_z P) \lesssim z^{-\beta}.$$

Similar computations can be obtained for $V_0(f_0, \tilde{K}_z P)$, where the major difference is that (3.73) is replaced with

$$V_0(f_0, \tilde{K}_z g_\beta) \lesssim z^{-\beta} + \int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) [|\log f_0(x)|^2 + |\log \tilde{K}_z g_\beta(x)|^2] dx$$

where, for some $k > 1$,

$$\begin{aligned}
&\int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) |\log \tilde{K}_z g_\beta(x)|^2 dx \\
&\quad \lesssim z^2 \int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) x^2 dx + [\log z]^2 \int \mathbf{1}_{\tilde{A}_z(a)^c} f_0(x) dx \\
&\quad \leq o(z^{-\beta}) + z^2 [F_0(\tilde{A}_z(a)^c)]^{1-1/k} \left(\int x^{2k} f_0(x) dx \right)^{1/k} = o(z^{-\beta})
\end{aligned}$$

as soon as $\int x^{2k} f_0(x)dx < \infty$ and $2 - (\beta/\nu + e)(1 - 1/k) < -\beta$, i.e. $(1 - 1/k) \geq \nu(\beta + 2)/\beta$. The latter inequality holds for $k \geq \beta/[\beta(1-\nu) - 2\nu]$, as soon as $\nu \leq 1/3$. We take $k = \beta/[\beta(1-\nu) - 2\nu]$ note that in the above, $2k = m \geq \beta/[\beta(1-\nu) - \nu]$.

Similarly, we obtain the bound on the Hellinger distance:

$$d_H(f_0, \tilde{K}_z P) \leq d_H(f_0, \tilde{K}_z g_\beta) + d_H(\tilde{K}_z g_\beta, \tilde{K}_z P)$$

Both $\tilde{K}_z g_\beta$ and $\tilde{K}_z P$ are supported on $[e_z, E_z]$ hence, due to the pointwise bound,

$$d_H^2(\tilde{K}_z g_\beta, \tilde{K}_z P) \le E_z z^{-H} = o(z^{-\beta})$$

for $E_z = o(z^{\beta/\nu})$ due to equation (3.66). □

Similarly to the derivation of the upper bound on $d_H(\bar{f}_0, \bar{f}_\beta)$, the first term satisfies

$$\begin{aligned} d_H^2(f_0, \tilde{K}_z g_\beta) &\lesssim z^{-\beta} + \int_{e_z}^{E_z} \frac{[f_0(x) - (\tilde{K}_z g_\beta)(x)]^2}{f_0(x)} dx \\ &\lesssim z^{-\beta} \end{aligned}$$

due to Lemma 14 and the moment conditions of class $\mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$.

3.6.4 Prior mass of KL neighbourhood

Lemma 17. *Consider $X_i \sim f_0$, $i = 1, \ldots, n$, independently. Assume $f_0 \in \mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$ for $\beta > 1$*

Consider prior $\Pi(f)$ defined in Section 3.3.1, and assume that $\delta \le a$ and

$$p_n \ge 1 - c[\log n]^{-1}, \quad \textit{for some } c > 0 \tag{3.75}$$

then, there exists $c_1 > 0$ such that

$$\Pi(\mathcal{K}_n) \gtrsim e^{-c_1 n \epsilon_n^2}$$

where $\mathcal{K}_n := \{f : KL(f, \tilde{K}_z * P) \le \epsilon_n^2; V(f, \tilde{K}_z * P) \le \epsilon_n^2 \log n\}$ and $\epsilon_n = n^{-\beta/(1+2\beta)}[\log n]^q$ with $q = \beta \max[\rho_z, 5/2]/(1 + 2\beta)$.

This lemma is proved in Section 3.9.

Remark 4. In Lemma 17, a possibly weaker condition on p_n in terms of δ is given by

$$1 - p_n \le \frac{1 - F_0(\delta) + \delta f_0(\delta)}{1 - F_0(\delta) + \delta f_0(\delta) + \int_{e_z}^{\delta}(-f_0'(x))x^{-1}dx}, \tag{3.76}$$

where e_z is defined in Lemma 15.

3.7 Technical lemmas on likelihood ratio

Lemma 18. *Under the conditions of Theorem 7, assume that $f_0 \in \mathcal{F}_0$ and $f_0 \in \mathcal{P}(\beta, L(\cdot), \ldots)$ with $\beta > 1$. Define $\tilde{B}_n = \{f \in \mathcal{F}_n; d_H(f_0, f) \le \epsilon_n\}$.*

Then, there exists a constant C_1 independent of n which may depend on f_0 and θ_0, such that for $M_n \to \infty$ and $M_n^4/n \to 0$ as $n \to \infty$,

$$P_0\left(\sup_{\theta \le \theta_0 - M_n/n} \sup_{f \in \tilde{B}_n} [\ell_n(\theta, f) - \ell_n(\theta_0, f)] > -C_1 M_n\right) = o(1)$$

and for all $0 < \epsilon < 1/f_0(0)$,

$$P_0\left(\inf_{\theta_0 + \epsilon^2/n \ge \theta \ge \theta_0} \inf_{f \in \tilde{B}_n} [\ell_n(\theta, f) - \ell_n(\theta_0, f)] < 0\right) \le \epsilon.$$

Proof of Lemma 18. We have that for $\theta = \theta_0 - h/n$ with $h > 0$, since f is monotone non increasing

$$\ell_n(\theta, f) - \ell_n(\theta_0, f) = \sum_{i=1}^{n} [\log f(y_i + h/n) - \log f(y_i)] \leq \sum_{i=1}^{n} \mathbf{1}_{y_i \in (t_n, a)} [\log f(y_i + h/n) - \log f(y_i)].$$

We therefore have if $h \geq M_n$,

$$\begin{aligned}
\ell_n(\theta, f) - \ell_n(\theta_0, f) &\leq \sum_{i=1}^{n} \mathbf{1}_{y_i \in (t_n, a)} [\log f(y_i + M_n/n) - \log f(y_i)] \\
&= \sum_{i=1}^{n} \mathbf{1}_{y_i \in (t_n, a)} [\log f(y_i + M_n/n) - \log f(y_i) - \log f_0(y_i + M_n/n) - \log f_0(y_i)] \\
&\quad + \sum_{i=1}^{n} \mathbf{1}_{y_i \in (t_n, a)} [\log f_0(y_i + M_n/n) - \log f_0(y_i)].
\end{aligned}$$

Moreover

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbf{1}_{y_i \in (t_n, a)} [\log f_0(y_i + M_n/n) - \log f_0(y_i)] \\
&\qquad \leq n \int_{t_n}^{a} f_0(y) [\log f_0(y + M_n/n) - \log f_0(y)] dy \\
&\qquad\quad + \sqrt{n} \mathbb{G}_n([\log f_0(\cdot + M_n/n) - \log f_0(\cdot)]).
\end{aligned}$$

Since

$$\int_{t_n}^{a} f_0(y) [\log f_0(y + M_n/n) - \log f_0(y)]^2 dy \leq \frac{\sup_{(0,a)} |f_0^{'}(x)|^2 M_n^2}{n^2 f_0(a)^2} \leq \frac{C_2 M_n^2}{n^2}$$

with probability greater than $1 - M_n^{-1}$

$$\sqrt{n} |\mathbb{G}_n([\log f_0(\cdot + M_n/n) - \log f_0(\cdot)])| \leq C_2 M_n^2 / \sqrt{n} = o(1)$$

and

$$\begin{aligned}
\sum_{i=1}^{n} \mathbf{1}_{y_i \in (t_n, a)} [\log f_0(y_i + M_n/n) - \log f_0(y_i)] &\leq -M_n [f_0(t_n) - f_0(a)] \\
&\quad + \frac{2M_n}{f_0(a)} \sup_{|u| \leq M_n/n} \int_{t_n}^{a} \left| f_0^{'}(y + u) - f_0^{'}(y) \right| dy + o(1) \\
&\leq -M_n [f_0(t_n) - f_0(a)] + o(M_n) \leq -M_n (f_0(0) - f_0(a))/2.
\end{aligned}$$

Therefore, using (H4)

$$
\begin{aligned}
\ell_n(\theta, f) - \ell_n(\theta_0, f) &\leq -M_n(f_0(0) - f_0(a))/2 + M_n \int_{t_n}^{a} f_0(y) \left(\frac{f'(y)}{f(y)} - \frac{f_0'(y)}{f_0(y)} \right) dy \\
&+ \frac{M_n}{\sqrt{n}} \mathbb{G}_n^y \left(\frac{f'}{f} - \frac{f_0'}{f_0} \right) \\
&+ M_n \mathbb{P}_n^y \left(\mathbf{1}_{y \in (t_n, a)} \sup_{0 \leq u \leq M_n/n} \left| \frac{f'(y+u)}{f(y+u)} - \frac{f'(y)}{f(y)} - \frac{f_0'(y+u)}{f_0(y+u)} + \frac{f_0'(y)}{f_0(y)} \right| \right) \\
&\leq [f_0(0) - f_0(a)] \left(-\frac{M_n}{2} + 3M_n \frac{\|f_0 - f\|_{\infty, a}}{f(a)} \right) + \frac{M_n}{\sqrt{n}} \sup_{f \in \tilde{B}_n} \mathbb{G}_n^y \left(\mathbf{1}_{y \in (t_n, a)} \frac{f'}{f} \right) \\
&+ M_n \epsilon \mathbb{P}_n^y \left(\mathbf{1}_{y \in (t_n, a)} M(y) \right) + o_{P_0}(M_n) \leq -M_n [f_0(0) - f_0(a) - 2\epsilon],
\end{aligned}
$$

with probability going to 1, and the first relation of Lemma 18 is proved. We now prove the second relation. We have for all $\epsilon > 0$,

$$
P_0 \left(X_{(1)} \leq \theta_0 + \frac{\epsilon}{n f_0(0)} \right) \leq n \int_0^{\epsilon / n f_0(0)} f_0(x) dx \leq \epsilon.
$$

On the event $X_{(1)} > \theta_0 + \frac{\epsilon}{n f_0(0)}$, $y_i = X_i - \theta_0 \geq \frac{\epsilon}{n f_0(0)}$. For $\theta = \theta_0 + h/n$ with $h \in (0, \epsilon^2)$, on the event $X_{(1)} > \theta_0 + \frac{\epsilon}{n f_0(0)}$,

$$
y_i - h/n \geq \frac{\epsilon}{n f_0(0)} - \frac{\epsilon^2}{n} = \frac{\epsilon(1 - \epsilon f_0(0))}{n f_0(0)} > 0
$$

for $0 < \epsilon < 1/f_0(0)$. Since f is monotone non increasing,

$$
\ell_n(\theta, f) - \ell_n(\theta_0, f) = \sum_{i=1}^{n} [\log f(y_i - h/n) - \log f(y_i)] \leq 0.
$$

□

Lemma 19. *Assume that a probability density f_0 decreases, $f_0(0) > 0$ and is local Hölder: for $x > 0$, $|y| \leq \Delta$, $y > -x$, $|f_0(x) - f_0(x + y)| \leq |y| L(x)(1 + |y|^\gamma)$, for some $\gamma, \Delta > 0$, and $\int L^2(x)/f_0(x) dx \leq C_L$.*

Then,

$$
d_H^2(f_{0,\theta_0}, f_{0,\theta}) \leq |\theta - \theta_0| f_0(0) + 0.25 C_L |\theta - \theta_0|^2 (1 + |\theta - \theta_0|^\gamma).
$$

Proof of Lemma 19. Without loss of generality, assume that $\theta \geq \theta_0$. Denote $v = \theta - \theta_0 \geq 0$.

Then,

$$\begin{aligned}
d_H^2(f_{0,\theta_0}, f_{0,\theta}) &= \int_{\theta_0}^{\theta} f_0(x-\theta_0)dx + \int_{\theta}^{\infty} (\sqrt{f_0(x-\theta)} - \sqrt{f_0}(x-\theta_0))^2 dx \\
&\leq v f_0(0) + \int_0^{\infty} (f_0(y) - f_0(y+v))^2/(f_0(y+v))^2 dx \\
&\leq v f_0(0) + 0.25 \int_0^{\infty} (vL(y+v)(1+v^{\gamma}))^2/f_0(y+v)dx \\
&\leq v f_0(0) + 0.25[v(1+v^{\gamma})]^2 C_L.
\end{aligned}$$

□

3.8 Proof of BvM for scale LAE model

Proof of Proposition 4. Consider $\Omega_n = \{n(X_{(1)} - \theta_0) \leq \tilde{R}_n\}$ for $\tilde{R}_n \to \infty$ and $\tilde{R}_n/n \to 0$ so that $P_{f_0,\theta_0}(\Omega_n) \to 1$. Then, on Ω_n, for large enough n,

$$n(Y_{(1)} - \tau_0) = ne^{X_{(1)}}(1 - e^{-(X_{(1)}-\theta_0)}) \leq ne^{\theta_0 + \tilde{R}_n/n}(X_{(1)} - \theta_0)) \leq 2e^{\theta_0}\tilde{R}_n.$$

Denote $\gamma_0 = f_0(0)$. Condition $||\pi_n(n(X_{(1)} - \theta \mid X^{(n)})) - Exp(f_0(0))||_{TV} \leq \alpha_n$ implies

$$\int_{0 \geq u \leq R_n} |\pi_n(u \mid X^{(n)}) - \gamma_0 e^{-\gamma_0 u}|du \leq 2\alpha_n \int_{u > R_n} \pi_n(u \mid X^{(n)})du \leq 2\alpha_n$$

for any $R_n \geq \log(0.5/\alpha_n)/\gamma_0$ and $R_n/n \to 0$, where $u = n(X_{(1)} - \theta)$ since

$$\begin{aligned}
2||\Pi_n(u \mid X^{(n)}) - Exp(\gamma_0)||_{TV} &= \int_0^{\infty} |\pi_n(u \mid X^{(n)}) - \gamma_0 e^{-\gamma_0 u}|d\theta \\
&= \int_{0 \geq u \leq R_n} |\pi_n(u \mid X^{(n)}) - \gamma_0 e^{-\gamma_0 u}|d\theta \\
&+ \int_{u > R_n} |\pi_n(u \mid X^{(n)}) - \gamma_0 e^{-\gamma_0 u}|du \leq 2\alpha_n.
\end{aligned}$$

Now, consider $S_n \geq \tau_0 \log(0.5/\alpha_n)/\gamma_0$ and $S_n/n \to 0$. For $v = n(Y_{(1)} - \tau)$, condition $v \leq S_n$

$$S_n \geq v = n(Y_{(1)} - \tau) = ne^{X_{(1)}}(1 - e^{-(X_{(1)}-\theta)}) \geq ne^{\theta_0}(X_{(1)} - \theta) \geq e^{\theta_0}u$$

implies that $u \leq e^{-\theta_0}S_n$.

Similarly, condition $v \geq S_n$ on Ω_n

$$\begin{aligned}
S_n &\leq v = n(Y_{(1)} - \tau) = ne^{X_{(1)}}(1 - e^{-(X_{(1)}-\theta)}) \leq n(X_{(1)} - \theta)e^{\theta_0 + \tilde{R}_n/n} = ue^{\theta_0 + \tilde{R}_n/n} \\
&\leq 2ue^{\theta_0}
\end{aligned}$$

implies $u \geq 0.5S_n e^{-\theta_0}$.

Now consider

$$2||\Pi_n(n(Y_{(1)}-\tau)\mid Y^{(n)})-Exp(\tau_0^{-1}\gamma_0)||_{TV}\le\int_{v\ge S_n}\pi_n(v\mid Y^{(n)})dv+2\alpha_n$$
$$+\int_{0\le v\le S_n}|\pi_n(v\mid Y^{(n)})-\tau_0^{-1}\gamma_0 e^{-\tau_0^{-1}\gamma_0 v}|dv$$
$$\le\int_{0\le u\le S_n e^{-\theta_0}}|\pi_n(u\mid X^{(n)})-\tau_0^{-1}\gamma_0 e^{-\tau_0^{-1}\gamma_0 n(e^{X_{(1)}}-e^{X_{(1)}-u/n})}e^{X_{(1)}-u/n}|du$$
$$+\int_{u\ge 0.5S_n e^{-\theta_0}}\pi_n(u\mid X^{(n)})du+2\alpha_n$$
$$\le\int_{0\le u\le S_n e^{-\theta_0}}\gamma_0|e^{-\gamma_0 u}-e^{-\gamma_0 n e^{X_{(1)}-\theta_0}(1-e^{-u/n})}e^{X_{(1)}-\theta_0-u/n}|du+6\alpha_n$$

making the change of variables $v=n(Y_{(1)}-\tau)=n(e^{X_{(1)}}-e^{\theta})=n(e^{X_{(1)}}-e^{X_{(1)}-u/n})$ ($dv=e^{X_{(1)}-u/n}du$) and taking $S_n=0.5e^{\theta_0}\log(0.5/\alpha_n)/\gamma_0$.

Now consider the remaining integral. Note that on Ω_n and for $0\le u\le S_n e^{-\theta_0}$,

$$e^{-\gamma_0 n e^{X_{(1)}-\theta_0}(1-e^{-u/n})}e^{X_{(1)}-\theta_0-u/n}-e^{-\gamma_0 n}\le e^{-\gamma_0 u(1-0.5u/n)}e^{\tilde{R}_n/n}-e^{-\gamma_0 n}$$
$$\le e^{-\gamma_0 u}[e^{\tilde{R}_n/n+0.5\gamma_0(S_n e^{-\theta_0})^2/n}-1]\le 2e^{-\gamma_0 u}[\tilde{R}_n+0.5\gamma_0(S_n e^{-\theta_0})^2]n^{-1}$$

for large enough n under assumption $S_n^2/n\to 0$, and

$$e^{-\gamma_0 n e^{X_{(1)}-\theta_0}(1-e^{-u/n})}e^{X_{(1)}-\theta_0-u/n}-e^{-\gamma_0 u}\ge e^{-\gamma_0 u e^{\tilde{R}_n/n}-S_n e^{-\theta_0}/n}-e^{-\gamma_0 u}$$
$$\ge e^{-\gamma_0 u e^{\tilde{R}_n/n}}[1-S_n e^{-\theta_0}/n]-e^{-\gamma_0 u}\ge -e^{-\gamma_0 u}[e^{\tilde{R}_n/n}-1+S_n e^{-\theta_0}/n]$$

using inequalities $1-x\le e^{-x}\ge 1-x+x^2/2$.

Hence,

$$|e^{-\gamma_0 n e^{X_{(1)}-\theta_0}(1-e^{-u/n})}e^{X_{(1)}-\theta_0-u/n}-e^{-\gamma_0 u}|$$
$$\le e^{-\gamma_0 u}\max([2\tilde{R}_n+\gamma_0(S_n e^{-\theta_0})^2]/n, e^{\tilde{R}_n/n}-1+S_n e^{-\theta_0}/n)$$
$$\le a_n e^{-\gamma_0 u}$$

where $a_n\to 0$. Hence, the integral is bounded by a_n and hence it tends to 0.

The weak convergence of $n(Y_{(1)}-\tau_0)$ follows from Lemma 20. ☐

Lemma 20. *Assume that $X_i\sim f(x)$, $x\ge\theta$, iid, $i=1,\dots,n$, and $f(\theta-)=0$ and $f(\theta+)=\gamma_0>0$, f - continuous on (θ,∞).*

Then, the distribution of $n(X_{(1)}-\theta)$ weakly converges to Exp(γ_0).

This lemma has been proved in [30] in a more complex case, we give a simple proof in the case of the density with a one-sided jump below.

Proof of Lemma 20. To show weak convergence, it is sufficient to show that as $n\to\infty$, for all $x>0$,

$$|f_{(n(X_{(1)}-\theta)}(x)-\gamma_0 e^{-\gamma_0 x}|\to 0.$$

For $x > 0$,

$$P(n(X_{(1)} - \theta) > x) = P(X_i \geq x/n + \theta,\, i = 1, \ldots, n) = (1 - F(x/n + \theta))^n$$

hence

$$f_{n(X_{(1)}-\theta)}(x) = f(x/n + \theta)(1 - F(x/n + \theta))^{n-1}, \quad x > 0.$$

For any fixed $x > 0$, $f(x/n+\theta) = f(\theta+)(1+o(1))$, and $F(\theta+x/n) = f(\theta+)x/n(1+o(1))$. Hence,

$$\begin{aligned}\log[(1 - F(x/n + \theta))^{n-1}] &= (n-1)\log(1 - F(\theta + x/n)) = (n-1)\log(1 - f(\theta+)x/n(1 + o(1))) \\ &= -f(\theta+)x(1 + o(1))),\end{aligned}$$

hence, since $f(\theta+) = \gamma_0$,

$$|f_{(n(X_{(1)}-\theta)}(x) - \gamma_0 e^{-\gamma_0 x}| \to 0.$$

□

3.9 Proofs of lemmas, nonparametric concentration rate

Proof of Lemma 17. Let $P_N = \sum_{i=1}^N p_i \delta_{u_i}$ and $A > 0$ from Lemma 15. Split the vector $u = (u_i)_{i=1}^N$ in two parts, $u^{(0)}$ and $u^{(1)}$, the first containing the elements $u_i < \delta$ and the other containing the elements $u_i \geq \delta$. Similarly we define vectors $p^{(0)}$ and $p^{(1)}$ with the corresponding weights p_i. We denote $N^{(0)}$ and $N^{(1)}$ the number of elements in $u^{(0)}$ and $u^{(1)}$ respectively. Define

$$Q_N = \sum_{i=1}^N q_i \delta_{u_i} = \bar{p} \sum_{i=1}^{N^{(0)}} q_i^{(0)} \delta_{u_i^{(0)}} + (1 - \bar{p}) \sum_{i=1}^{N^{(1)}} q_i^{(1)} \delta_{u_i^{(1)}} \tag{3.77}$$

where

$$q_i^{(0)} = \frac{p_i^{(0)}}{(u_i^{(0)})^2 c_N^{(0)}}, \qquad q_i^{(1)} = \frac{p_i^{(1)}}{c_N^{(1)}}, \tag{3.78}$$

$$c_N^{(0)} = \sum_{i=1}^{N^{(0)}} \frac{p_i^{(0)}}{(u_i^{(0)})^2}, \quad c_N^{(1)} = \sum_{i=1}^{N^{(1)}} p_i^{(1)} \tag{3.79}$$

and

$$\bar{p} = \frac{c_N^{(0)}}{c_N^{(0)} + c_N^{(1)}} \tag{3.80}$$

then $dP_N(\epsilon) = \frac{s(\epsilon) dQ_N(\epsilon)}{\int_0^\delta \epsilon^2 dQ_N(\epsilon) + Q_N(\delta, \infty)}$ and $q_i = p_i/(s(u_i)c_N)$ where

$$s(\epsilon) = \epsilon^2 \mathbb{1}(\epsilon \leq \delta) + \mathbb{1}(\epsilon > \delta) \tag{3.81}$$

$$c_N = c_N^{(0)} + c_N^{(1)}. \tag{3.82}$$

Additionally define $I_z = (1 - 2z^{-B}, 1 - z^{-B})$, $B \geq A + 2a$ and

$$\begin{aligned}\tilde{\mathcal{Q}}_z = \Big\{ &(Q^{(0)}, Q^{(1)}, p) : Q^{(0)}(U'_i)/q_i^{(0)} \in I_z,\ i = 1, \dots, N^{(0)}, \\ &Q^{(1)}(U'_i)/q_i^{(1)} \in I_z,\ i = 1, \dots, N^{(1)}, \\ &|p - \bar{p}| \leq \bar{p}_{min} z^{-B} \Big\}\end{aligned} \tag{3.83}$$

where $U'_i = \left[u_i\left(1 - z^{-B}\right), u_i\left(1 + z^{-B}\right)\right]$ and $\bar{p}_{min} = \min\{\bar{p}, 1 - \bar{p}\}$.

Let $(Q^{(0)}, Q^{(1)}, p) \in \tilde{\mathcal{Q}}_z$, and $Q = pQ^{(0)} + (1-p)Q^{(1)}$. Then by definition of $\tilde{\mathcal{Q}}_z$

$$\begin{aligned}Q(U'_i) =& pQ^{(0)}(U'_i) + (1-p)Q^{(1)}(U'_i) \\ &\in \left(q_i(1 - 2z^{-B})(1 - z^{-B}), q_i(1 - z^{-B})(1 + z^{-B})\right)\end{aligned} \tag{3.84}$$

and defining $U'_0 = [0, \infty) \setminus \cup_{i=1}^N U'_i$

$$\begin{aligned}Q(U'_0) \leq& 1 - (1 - 2z^{-B})(1 - z^{-B}) \leq 2z^{-B} \\ \geq& 1 - (1 - z^{-B})(1 + z^{-B}) = z^{-2B}\end{aligned} \tag{3.85}$$

Now define

$$P_Q = s(\epsilon)Q/c_Q, \quad c_Q = \int s(\epsilon) dQ(\epsilon)$$

then

$$\begin{aligned}P_Q(U'_i) - p_i =& \frac{\int_{U'_i} s(\epsilon) dQ(\epsilon)}{c_Q} - p_i \\ \leq& \frac{s(u_i)(1 + z^{-B})^3 q_i (1 - z^{-B})}{c_Q} - p_i \\ =& p_i \left(\frac{(1 + z^{-B})^3 (1 - z^{-B})}{c_Q c_N} - 1 \right)\end{aligned} \tag{3.86}$$

also

$$\begin{aligned}P_Q(U'_i) - p_i \geq& \frac{s(u_i)(1 - z^{-B})^3 q_i (1 - 2z^{-B})}{c_Q} - p_i \\ =& p_i \left(\frac{(1 - z^{-B})^3 (1 - 2z^{-B})}{c_Q c_N} - 1 \right).\end{aligned} \tag{3.87}$$

We bound the normalising constants

$$
\begin{aligned}
c_Q = \int_0^\delta s(\epsilon)dQ(\epsilon) =& \sum_{i=1}^N \int_{U_i'} s(\epsilon)dQ(\epsilon) + \int_{U_0'} s(\epsilon)dQ(\epsilon) \\
\leq& \sum_{i=1}^N s(u_i)(1+z^{-B})^3 q_i(1-z^{-B}) + 2z^{-B} \\
\leq& c_N^{-1}(1+z^{-B})^3(1-z^{-B}) + 2z^{-B}
\end{aligned}
\tag{3.88}
$$

$$
\begin{aligned}
\geq& \sum_{i=1}^N s(u_i)(1-z^{-B})^3(1-2z^{-B}) \\
\geq& c_N^{-1}(1-z^{-B})^3(1-2z^{-B})
\end{aligned}
$$

and since $u_i \geq e_z = z^{-a}$ due to Lemma 15,

$$
c_N \leq e_z^{-2} = z^{2a}, \quad a > 0. \tag{3.89}
$$

Thus, for z sufficiently large

$$
\begin{aligned}
P_Q(U_i') - p_i \leq& p_i \left(\frac{(1+z^{-B})^3}{(1-z^{-B})^2(1-2z^{-B})} - 1 \right) \\
=& p_i \frac{7z^{-B} - 2z^{-2B} + 3z^{-3B}}{(1-z^{-B})^2(1-2z^{-B})} \\
\leq& 8p_i z^{-B}
\end{aligned}
$$

and

$$
\begin{aligned}
\geq& p_i \left(\frac{(1-z^{-B})^3(1-2z^{-B})}{(1+z^{-B})^3(1-z^{-B}) + 2z^{-B}c_N} - 1 \right) \\
=& p_i \left(\frac{-7z^{-B} + 9z^{-2B} - 5z^{-3B} + 3z^{-4B} - 2z^{-B}c_N}{(1+z^{-B})^3(1-z^{-B}) + 2z^{-B}c_N} \right) \\
\geq& p_i \frac{-(7+2c_N)z^{-B}}{2 + 2z^{-B}c_N} \\
\geq& (-4z^{-B} - z^{-B}c_N)p_i \\
\geq& -5p_i c_N z^{-B} \\
\geq& -5p_i z^{-(B-2a)}
\end{aligned}
\tag{3.90}
$$

Therefore taking $B > A + 2a + 1$, if $Q \in \tilde{\mathcal{Q}}_z$ then $P_Q \in \mathcal{P}_z = \mathcal{P}_z = \{P : P(U_i)/p_i \in (1-z^{-A}, 1+z^{-A}),\ \forall i = 1, \ldots, N\}$, for z large enough.

Moreover,

$$
\begin{aligned}
|\tilde{K}_z * P_Q(x) - \tilde{K}_z * P_N(x)| \leq & \int_x^\infty \frac{1}{\theta} \left| \int g_{z,u}(\theta) dP_Q(u) - \sum_{i=1}^N p_i g_{z,u_i}(\theta) \right| d\theta \\
\leq & \int_x^\infty \frac{1}{\theta} \sum_{i=1}^N \left| \int_{U_i} g_{z,u_i}(\theta) dP_Q(u) - p_i g_{z,u_i}(\theta) \right| d\theta \\
& + \int_x^\infty \frac{1}{\theta} \sum_{i=1}^N \int_{U_i} |g_{z,u}(\theta) - g_{z,u_i}(\theta)| dP_Q(u) d\theta \\
\leq & \int_x^\infty \frac{1}{\theta} \sum_{i=1}^N p_i g_{z,u_i}(\theta) \left| dP_Q(U_i) - p_i \right| d\theta \\
& + \frac{1}{x} \sum_{i=1}^N \int_{U_i} \|g_{z,u} - g_{z,u_i}\|_1 dP_Q(u) \\
\leq & 2z^{-2A} \int_x^\infty \frac{1}{\theta} \sum_{i=1}^N p_i g_{z,u_i}(\theta) d\theta \\
& + \frac{1}{x} \sqrt{2z} z^{-2A} \sum_{i=1}^N \int_{U_i} dP_Q(u) \\
\leq & 2z^{-2A} \tilde{K}_z * P_N(x) + \sqrt{2} z^{a+1/2-2A}
\end{aligned}
\tag{3.91}
$$

Combining this with Lemma 15 we obtain

$$
sup_x |\tilde{K}_z * P_Q(x) - \tilde{K}_z * g_\beta(x)| \leq z^{-H} + 2z^{-2A} \tilde{K}_z * P_N(0) + \sqrt{2} z^{a+1/2-2A} \lesssim z^{-H} \tag{3.92}
$$

for any $A > (H + a + 1/2)/2$, since $\tilde{K}_z * P_N(0)$ can be bounded by a constant times $f_0(0)$, using Lemma 15 and Proposition 8.

The assumption on the tail of f_0 implies that we can choose $E_z = z^b$ for some $b \geq \beta/(2 + \rho_1)$ since

$$
\int_{E_z}^\infty f_0(y) dy \leq E_z^{-2} \int_{E_z}^\infty y^2 f_0(y) dy \leq E_z^{-2} (1 + E_z)^{-\rho_1} \leq E_z^{-2-\rho_1} \leq z^{-\beta}
$$

holds if $E_z \geq z^{\beta/(2+\rho_1)}$.

Following Lemma 4.1 in [4] (see Lemma 30 in Section 3.11) we define $z_n = n^{2/(2\beta+1)} (\log n)^t$. We also set $I_n = (z_n, 2z_n)$. For all $z \in I_n$ and all $P \in \mathcal{P}_z$, Lemma 16 implies

$$
\begin{aligned}
KL(f, \tilde{K}_z * P) \lesssim & n^{-2\beta/(2\beta+1)} (\log n)^{-\beta t} \lesssim \epsilon_n^2 \\
V(f, \tilde{K}_z * P) \lesssim & n^{-2\beta/(2\beta+1)} (\log n)^{-\beta t} \lesssim \epsilon_n^2
\end{aligned}
\tag{3.93}
$$

for A large enough (depending on $\beta, L, \gamma, C_0, C_1, e, \Delta$) as long as $t \geq -2q/\beta$. Thus we need to bound from below the prior mass $\Pi(\tilde{\mathcal{Q}}_z,\, z \in I_n)$. Note that

$$
\Pi\left(\tilde{\mathcal{Q}}_z\right) = \Pi\left(\tilde{\mathcal{Q}}_z^{(0)}\right) \Pi\left(\tilde{\mathcal{Q}}_z^{(1)}\right) \Pi\left(\tilde{\mathcal{Q}}^{(p)}\right) \tag{3.94}
$$

where $\tilde{\mathcal{Q}}_z^{(j)} = \{Q^{(j)} : Q^{(j)}(U_i')/q_i^{(j)} \in I_z,\, i = 1, \ldots, N^{(j)}\}$, $j = 0, 1$, and $\tilde{\mathcal{Q}}^{(p)} = \{|p - \bar{p}| \leq \min\{\bar{p}, 1 - \bar{p}\} z^{-B}\}$. Also note that for all $i = 1, \ldots, N$, such that $\delta \notin U_i'$, and corresponding

$j = 0$ or 1,

$$\alpha_i := mG^{(j)}(U_i') = m\int_{u_i(1-z^{-B})}^{u_i(1+z^{-B})} g^{(j)}(u)du \geq 2u_i z^{-B} \min_{x\in U_i'} g^{(j)}(x). \tag{3.95}$$

and $u_i \geq e_z = z^{-a}$. If $U_i' \ni \delta$ then we can replace the corresponding u_i by δ and the new P_N measure satisfies the same properties in Lemma 16 since $|\delta/u_i - 1| \leq z^{-B}$. Therefore we split the integral α_i in two parts $\alpha_i^{(0)}$ and $\alpha_i^{(1)}$,

$$\alpha_i^{(0)} := m\int_{\delta(1-z^{-B})}^{\delta} g^{(0)}(u)du \geq \delta z^{-B} \min_{x\in[\delta(1-z^{-B}),\delta]} g(x) \tag{3.96}$$

and

$$\alpha_i^{(1)} := m\int_{\delta}^{\delta(1+z^{-B})} g^{(1)}(u)du \geq \delta z^{-B} \min_{x\in[\delta,\delta(1+z^{-B})]} g(x). \tag{3.97}$$

Additionally, we split α_0,

$$\alpha_0^{(0)} \geq m\int_0^{u_1(1-z^{-B})} g^{(0)}(u)du \gtrsim \int_0^{e_z(1-z^{-B})} u^{a_0}du \gtrsim e_z^{a_0+1} = z^{-a(a_0+1)} \tag{3.98}$$

and

$$\begin{aligned}\alpha_0^{(1)} \geq m\int_{u_N(1+z^{-B})}^{\infty} g^{(1)}(u)du \gtrsim \int_{u_N(1+z^{-B})}^{\infty} u^{-a_1}du \gtrsim & E_z^{-a_1+1} \\ =& z^{-(a_1-1)b}.\end{aligned} \tag{3.99}$$

Similarly, $\alpha_0^{(0)} \lesssim z^{-a(a_0'+1)}$ and $\alpha_0^{(1)} \lesssim z^{-(a_1'-1)b}$. Thus

$$\sum_{j=1}^{N}(-\log\alpha_j) \leq NB'\log z \tag{3.100}$$

with $B' > 0$ depending on $a, b, r, a_0, a_0', a_1, a_1'$.

Now we can continue with similar computations as in Lemma 4.1 in [4] (see Lemma 34 in Section 3.11) and obtain

$$\begin{aligned}\Pi\left(\tilde{\mathcal{Q}}_{z_n}^{(j)}\right) \geq & \frac{\Gamma(m)}{\prod_i\Gamma(\alpha_i)} z_n^{-A(\alpha_0-1)}2^{-(\alpha_0-1)_-} \\ & \times \int\prod_{i=1}^{N}\mathbb{1}\left(x\in(p_i(1-z_n^{-A}),p_i(1+z_n^{-A}))\right)x_i^{\alpha_i-1}dx_i \\ \gtrsim & z_n^{-A(\alpha_0-1+N)}\frac{\Gamma(m)}{\prod_i\Gamma(\alpha_i)}\prod_{i=1}^{N}p_i^{\alpha_i} \\ \gtrsim & e^{-A(N+m+B'N\log N)} \gtrsim e^{-(B'+1)AN\log N}\end{aligned} \tag{3.101}$$

hence

$$\Pi\left(\tilde{\mathcal{Q}}_{z_n}^{(j)}\right) \gtrsim e^{-(B'+2)AN_0\sqrt{z_n}(\log z_n)^{5/2}} \tag{3.102}$$

for $j = 0, 1$.

Now we bound $\Pi\left(\tilde{\mathcal{Q}}^{(p)}\right)$. First we show that $\bar{p}_{min} \gtrsim z^{-A-2a}$. Indeed, $p_i \gtrsim z^{-A}$, $u_i \in [z^{-a}, z^b]$ with $A, a, b > 0$, given by Lemma 15, and $c_N = c_N^{(0)} + c_N^{(1)} \le z^{2a}$. Furthermore, note that $c_N^{(0)} = 0$ or $c_N^{(1)} = 0$ only when the true distribution P_0 has all the mass only on one side of δ, which is not possible. Hence,

$$\bar{p}_{min} = \frac{\min\{c_N^{(0)}, c_N^{(1)}\}}{c_N^{(0)} + c_N^{(1)}} \gtrsim \frac{\min\{N^{(0)} z^{-A}\delta^{-2}, N^{(1)} z^{-A}\}}{z^{2a}} \gtrsim z^{-A-2a}. \tag{3.103}$$

Condition (3.13) with $x_1 = \bar{p} - \bar{p} z_n^{-B}$ and $x_2 = \bar{p} + \bar{p} z_n^{-B}$ implies that for any $\bar{p}$ such that $\bar{p} + \bar{p}_{\min} z_n^{-B} < p_0$,

$$\Pi\left(\tilde{\mathcal{Q}}^{(p)}\right) \ge \bar{p} z_n^{-B} e^{-d(\log 1/\bar{p} - z_n^{-B})^{5/2}} \gtrsim e^{-D(\log z_n)^{5/2} - (A+2a+B)\log z_n} \tag{3.104}$$

for some $D > 0$. A bound for $\bar{p}$ around 1, i.e. such that $\bar{p} - \bar{p}_{\min} z_n^{-B} \ge p_1$, is obtained similarly.

Finally for any $\bar{p}$ such that $(\bar{p} - \bar{p}_{\min} z_n^{-B}, \bar{p} + \bar{p}_{\min} z_n^{-B}) \subset (p_0, p_1)$,

$$\Pi\left(\tilde{\mathcal{Q}}^{(p)}\right) \ge \int_{\bar{p} - \bar{p}_{min} z_n^{-B}}^{\bar{p} + \bar{p}_{min} z_n^{-B}} \pi_p(u) du \gtrsim \bar{p}_{min} z_n^{-B} \min_{x \in [\rho_0, \rho_1]} \pi_p(x) \gtrsim e^{-B' \log z_n} \tag{3.105}$$

therefore, for any $\bar{p} \in (0, p_n)$ we obtain

$$\Pi\left(\tilde{\mathcal{Q}}^{(p)}\right) \gtrsim e^{-B'(\log z_n)^{5/2}} \tag{3.106}$$

for some $B' > 0$.

Condition $\bar{p} \le p_n$ is equivalent to

$$\bar{p} \;=\; \frac{\sum_{i=1}^{N^{(0)}} \frac{p_i^{(0)}}{(u_i^{(0)})^2}}{\sum_{i=1}^{N^{(0)}} \frac{p_i^{(0)}}{(u_i^{(0)})^2} + \sum_{i=1}^{N^{(1)}} p_i^{(1)}} \le p_n.$$

In Lemma 21, we show that $P_N(\delta, \infty)$ approximates $1 - G(\delta) = 1 - F(\delta) + \delta f(\delta)$, and that $\sum_{i=1}^{N^{(0)}} \frac{p_i^{(0)}}{(u_i^{(0)})^2}$ is approximately $\int_{z^{-a}}^{\delta} (-f_0'(x))/x dx \lesssim \log z$ where C_p is a constant. Hence,

$$1 - \bar{p} \;\ge\; \frac{1 - G(\delta)}{1 - G(\delta) + C_p \log z}(1 + o(1)) \ge 1 - p_n.$$

Therefore, since here $z \in I_n$, for δ bounded above by a constant independent of n, $p_n \ge 1 - \tilde{C}_p / \log z_n$.

Due to condition (3.12) on the prior for z,

$$\Pi_z(I_n) \gtrsim e^{-a'\sqrt{z_n}(\log z_n)^{\rho_z}} \tag{3.107}$$

for $z_n = n^{2/(2\beta+1)}(\log n)^t$.

Substituting bounds (3.107), (3.102) and (3.106) into equation (3.94) implies

$$\Pi\left(\tilde{\mathcal{Q}}_z\right) \ge e^{-c' n \epsilon_n^2}$$

for $\epsilon_n = n^{-\beta/(1+2\beta)}[\log n]^q$ such that

$$n\epsilon_n^2 \geq \sqrt{z_n}[\log z_n]^{\max[\rho_z, 5/2]}$$

i.e. $q \geq t/4 + \max[\rho_z, 5/2]/2$. Condition $t \geq -2q/\beta$ implies constraint on q:

$$q \geq \max[\rho_z, 5/2]\beta/(2\beta+1).$$

Taking the smallest $q = \max[\rho_z, 5/2]\beta/(2\beta+1)$ and $t = -2q/\beta = -2\max[\rho_z, 5/2]/(2\beta+1)$,, we conclude the proof.

Note that this choice implies

$$z_n = n^{2/(2\beta+1)}(\log n)^t \tag{3.108}$$

□

Lemma 21. *Under assumptions of Lemma 16, $f \in \mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$ with $\beta > 1$,*

- *for any measurable subset $A \subseteq (0, \infty)$,*

$$|P_N(A) - G(A)| = o(1)$$

  where $G(A) = \int_A g_0(x)dx$ where $g_0(x) = -xf_0'(x)$ where $g(x) = -xf'(x)$;

- *for any $\delta > 0$, large enough z, $q > 0$, $\beta > 1$,*

$$\begin{aligned}\sum_i p_i/u_i^2 I(u_i \leq \delta) = &\int_{z^{-q}}^{\delta} (-f_0'(x))x^{-1}dx(1+o(1)) + o(1)\\ &\leq 3q\log z \sup_{x\in(z^{-q},\delta)} |f_0'(x)|.\end{aligned}$$

The bounds are uniform for $f_0 \in \mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$ if $\beta \geq \beta_0 > 1$.

Proof of Lemma 21. We use notation from the proof of Lemma 15.

1. In the proof of Lemma 15, P_N is chosen such that for any $H > 0$

$$|K_z * \bar{g}_\beta(x) - K_z * P_N(x)| \leq z^{-2H} \quad forx \in [e_z/2, 2E_z]. \tag{3.109}$$

From the proof of Lemma B.2 in [**?**], we know that for any measure P with support $[e_z, E_z]$, there exist constants $c_1, c_{0.5} > 0$ such that

$$\int_0^{e_z/2} K_z * P(x)dx \lesssim 2^{-c_1 z}z^{-1}$$

and

$$\int_{2E_z}^{\infty} K_z * P(x)dx \lesssim 2^{-2c_{0.5}z}z^{-1}.$$

Since P_N and $\bar{g}_\beta$ are supported on $[e_z, E_z]$ then, due to equation (3.66),

$$\|K_zP_N - K_z\bar{g}_\beta\|_1 \lesssim z^{-2H}(2E_z - e_z/2) + 2^{-c_2z}z^{-1} \lesssim z^{-2H+\beta/\nu}$$

for some $c_2 > 0$. By definition of $\bar{g}_\beta$

$$K_z g_\beta(x) = c_\beta K_z \bar{g}_\beta(x) + \int_0^{e_z} g_\beta(\epsilon) g_{z,\epsilon}(x) d\epsilon + \int_{E_z}^{\infty} g_\beta(\epsilon) g_{z,\epsilon}(x) d\epsilon$$
$$\geq c_\beta K_z \bar{g}_\beta(x)$$

where $c_\beta = G_\beta(e_z, E_z) \geq 1 - O(z^{-\beta/2})$. Thus

$$d_H^2(K_z\bar{g}_\beta, K_z g_\beta) \leq 2 - 2\int_0^\infty K_z\bar{g}_\beta(x) dx \sqrt{c_\beta} \leq 2 - 2\sqrt{1 - O(z^{-\beta/2})} \lesssim O(z^{-\beta/2})$$

therefore

$$\|K_z\bar{g}_\beta - K_z g_\beta\|_1 \lesssim z^{-\beta/4}.$$

Similarly, due to the Hellinger bound in Lemma 16, we also have

$$\|K_z g_\beta - K_z g_0\|_1 \leq z^{-\beta/2}$$

thus

$$\|K_z P_N - K_z g_0\|_1 \lesssim z^{-2H+\beta/\nu} + z^{-\beta/4}$$

which goes to 0 as z goes to infinity due to $H > \beta/\nu$.

By Lemma C.3 in [4] (see Lemma 35 in Section 3.11) applied to $g_0 \in \mathcal{P}(\beta-1, L, \gamma, C_0, C_1, e, \Delta, \nu)$,

$$K_z g_0(x) = \sum_{j=0}^{r-1} \frac{g_0^{(j)}(x) x^j}{j! z^{j/2}} \mu_j(z) + z^{(\beta-1)/2} R_z(x)$$

where $r = \lceil \beta \rceil - 1$, $\mu_j(z) = \mu_j + O(z^{-H})$, μ_j is the j-th moment of a Gaussian distribution and $|R_z(x)| \leq C_{\beta,z} L_g(x) x^{\beta-1}(1 + x^{\gamma_g}/z^{\gamma/2})$. Therefore,

$$K_z g_0(x) = g_0(x) + O(z^{-(1\wedge(\beta-1))/2})$$

Additionally, for any discrete measure P, and any measurable subset $A \subseteq (0, \infty)$

$$\int_A K_z P(x) = \sum_i p_i \int_A g_{\epsilon_i,z}(x) dx$$

By Lemma C.2 in [4] (see Lemma 34 in Section 3.11), for all $d \in (0, 1)$, and z large enough, there exists $c_d > 0$ such that

$$\int_0^{\epsilon/(1+d)} g_{\epsilon,z}(x) dx \leq (c_d z)^{-1}(1 + d)^{-c_d z}$$

and

$$\int_{\epsilon/(1-d)}^\infty g_{\epsilon,z}(x) dx \leq (z(1 - d))^{-1} e^{-c_d z/(1-d)}$$

For a fixed set A, let us choose $d > 0$ small enough so that for all ϵ_i, A either contains $(\epsilon_i/(1 + d),\ \epsilon_i/(1-d))$ or their intersection is empty. Therefore, for all ϵ_i such that $(\epsilon_i/(1+d),\ \epsilon_i/(1-d)) \subset A$

$$\int_A g_{\epsilon_i,z}(x) dx = 1 + O(z^{-1} e^{-2c_d z})$$

and
$$\int_A K_z P(x) = \sum_{i:\ \epsilon_i \in A} p_i + O(z^{-1}e^{-2c_d z}) = P(A) + O(z^{-1}e^{-2c_d z}).$$

Thus for any measurable subset $A \subseteq (0, \infty)$,

$$|P_N(A) - G(A)| \leq \left| \int_A K_z P_N(x) - K_z g_0(x) dx \right| + O(z^{-1}e^{-2c_d z}) + O(z^{-(1\wedge(\beta-1))/2})$$

and
$$\left| \int_A K_z P_N(x) - K_z g_0(x) dx \right| \leq \|K_z P_N - K_z g_0\|_1 \lesssim z^{-2H+\beta/\nu} + z^{-\beta/4} \to 0$$

For $A = (\delta, E_z)$ this implies $|\sum_i p_i^{(1)} - (1 - G(\delta))| \to 0$ as z goes to infinity.

2. Inequality (3.109) can be written as

$$|K_z * \bar{g}_\beta(x) - K_z * P_N(x)| = |\int_{e_z}^{E_z} s(\epsilon) g_{\epsilon,z}(x)(\bar{g}_\beta(\epsilon)/s(\epsilon) d\epsilon - dP_N(\epsilon)/s(\epsilon))| \leq z^{-2H}$$

where $s(\epsilon) = \epsilon^2 I(\epsilon < \delta) + I(\epsilon \geq \delta)$. For a given $\epsilon \in (e_z, \delta)$, consider x such that $|x/\epsilon - 1| \leq \delta_z = \sqrt{z^{-1}\log z}$, and note that

$$\begin{aligned}
& s(\epsilon) g_{\epsilon,z}(x) = z^z \epsilon^{2-z} \frac{1}{\Gamma(z)} x^{z-1} e^{\frac{-xz}{\epsilon}} \\
= \quad & x^2 e^{-2(x/\epsilon-1)} \frac{e^{-2} z^z}{(z-1)(z-2)^{z-1}} (z-2)^{z-2} \epsilon^{2-z} \frac{1}{\Gamma(z-2)} x^{z-3} e^{\frac{-x(z-2)}{\epsilon}} \\
\geq \quad & x^2 g_{\epsilon,z-2}(x) e^{-2\delta_z}(1 + R(z))
\end{aligned}$$

where $R(z) = O(1/z)$ for large z. Similarly,

$$s(\epsilon) g_{\epsilon,z}(x) \leq x^2 g_{\epsilon,z-2}(x) e^{2\delta_z}(1 + R(z)).$$

Using the approximation of $g_{\epsilon,z}(x)$ for $|x/\epsilon - 1| \leq \delta_z$ as in the proof of Lemma C.2 in [4] (see Lemma 34 in Section 3.11), we have

$$\begin{aligned}
& \int_{c_3 e_z}^{c_4 \delta} g_{\epsilon,z}(x) dx \\
= \quad & [v = (x/\epsilon - 1)\sqrt{z}] = \int_{\sqrt{z}\max(c_3 e_z/\epsilon - 1, -\delta_z)}^{\sqrt{z}\min(c_4\delta/\epsilon - 1, \delta_z)} \phi(v)(1 + R_z(1)) dv \\
= \quad & (1 + O(1/z))
\end{aligned}$$

for $c_3 e_z/\epsilon - 1 \leq -\delta_z$ and $c_4\delta/\epsilon - 1 \geq \delta_z$, i.e. for $c_3 e_z/(1-\delta_z) \leq \epsilon$ and $c_4\delta/(1+\delta_z) \geq \epsilon$; it converges to zero otherwise. Take $c_3 = (1 - \delta_z)$ and $c_4 = 1 + \delta_z$.

Integrating inequality $K_z * P_N(x) \leq z^{-2H} + K_z * \bar{g}_\beta(x)$ multiplied by x^{-2} over $(c_3 e_z, c_4\delta)$, and using approximation of $g_{\epsilon,z}(x)s(\epsilon)$ above, we have

$$\int_{e_z}^{\delta} dP_N(\epsilon)/\epsilon^2 \quad \leq \quad 2\int_{(c_3 e_z, c_4\delta)} x^{-2} \int g_{\epsilon,z}(x) \bar{g}_\beta(\epsilon) d\epsilon + 2z^{-2H} e_z^{-1}.$$

The second term goes to 0 for H large enough. Consider $\beta \in (1, 2]$, then $g_\beta = (g/(1 +$

$c_0(z))I_{\mathcal{A}_z(a)} + (g/2)I_{\mathcal{A}_z(a)^c})/c_0$, with $c_0(z) = 1 + O(z^{-1/2})$ and c_0 the normalising constant. Then the first term is approximated by

$$(1+o(1))\int_{(e_z,\delta)}(-f_0'(x)/x)dx + (1+o(1))z^{-(\beta-1)/2}\int_{(e_z,\delta)} L_g(x)x^{\beta-2}(1+x^{\gamma_g})dx$$

since if f_0 is local Hölder with β, γ, $L(x)$, $\Delta = 1$, then $-xf_0'(x)$ is local Hölder with $\beta_g = \beta - 1$, $\gamma_g = \max(1,\gamma)$, $L_g(x) = \max((x+1)L(x), 2|f_0'(x)|)$ due to

$$\begin{aligned}
|xf_0'(x) - yf_0'(y)| &\leq |x-y||f_0'(x)| + yL(x)|y-x|^{\beta-1}(1+|y-x|^{\gamma}) \\
&\leq |y-x|^{\beta-1}[(x+|y-x|)L(x)(1+|y-x|^{\gamma}) + |f_0'(x)||y-x|^{2-\beta}] \\
&\leq |y-x|^{\beta-1}2\max((x+1)L(x), |f_0'(x)|)[1+|y-x|^{\max(1,\gamma)}].
\end{aligned}$$

For $f_0 \in \mathcal{P}(\beta, L, \gamma, C_0, C_1, e, \Delta, \nu)$ with $\beta > 1$, the second integral is bounded by

$$\begin{aligned}
&(1+\delta^{\gamma_g})(1+\delta)\int_{e_z}^{\delta} L(x)x^{\beta-2}dx \leq [\int_{e_z}^{\delta} x^{2(\beta-2)}g_0(x)dx]^{1/2}[\int_{e_z}^{\delta}[L(x)/g(x)]^2 g(x)dx]^{1/2} \\
\leq\ & (1+\delta^{\gamma_g})(1+\delta)[\int_{e_z}^{\delta} x^{2\beta-3}(-f_0'(x))dx\int_{e_z}^{\delta}[\delta^{-\beta}L(x)x^{\beta}/g_0(x)]^2 g_0(x)dx]^{1/2} \\
\leq\ & (1+\delta^{\gamma_g})(1+\delta)[\sup_{x\in(e_z,\delta)}|f_0'(x)|\delta^{-2}[2(\beta-1)]^{-1}\int_0^{\infty}[L(x)x^{\beta}/g_0(x)]^2 g_0(x)dx]^{1/2} \\
\leq\ & (1+\delta^{\max(\gamma,1)})(1+\delta)[\delta^{-2}[2(\beta-1)]^{-1}C_0C_1]^{1/2}.
\end{aligned}$$

For a uniform bound, need $\beta \geq \beta_0 > 1$.

Note that

$$\int_{x\in(z^{-q},\delta)} g_0(x)/x^2dx = -\int_{x\in(z^{-q},\delta)} f_0'(x)/xdx \leq \sup_{x\in(z^{-q},\delta)}|f_0'(x)|\log(z^q\delta).$$

and

$$-\int_{x\in(z^{-q},\delta)} f_0'(x)/xdx \geq \inf_{x\in(z^{-q},\delta)}|f_0'(x)|\log(z^q\delta).$$

Therefore, if $0 < \inf_{x\in(z^{-q},\delta)}|f_0'(x)| \leq \sup_{x\in(z^{-q},\delta)}|f_0'(x)| < \infty$, then $-\int_{z^{-q}}^{\delta} f_0'(x)/xdx \asymp \log z$.

Therefore, for large enough z, we have

$$\sum_i p_i/u_i^2 I(u_i \leq \delta) \leq 3q\sup_{x\in(z^{-q},\delta)}|f_0'(x)|\log z$$

which completes the proof. □

Lemma 22. $\tilde{K}_z(-xf'(x)) = -(z-1)\tilde{K}_z f(x) + \frac{z^2}{z+1}\tilde{K}_{z+1}f(xz/(z+1))$.

Proof. Consider $g(\theta) = -\theta f'(\theta)$. By definition of K_z we obtain

$$
\begin{aligned}
K_z g(\theta) &= \int_0^\infty \frac{z^z}{\Gamma(z)} \theta^{z-1} \frac{e^{-z\theta/\epsilon}}{\epsilon^z} g(\epsilon) d\epsilon \\
&= \frac{-\theta^{z-1} z^z}{\Gamma(z)} \int_0^\infty \frac{e^{-z\theta/\epsilon}}{\epsilon^z} \epsilon f'(\epsilon) d\epsilon \\
&= \frac{\theta^{z-1} z^z}{\Gamma(z)} \left\{ \frac{e^{-z\theta/\epsilon}}{\epsilon^{z-1}} f(\epsilon) \Big|_0^\infty + \int_0^\infty f(\epsilon) e^{-z\theta/\epsilon} \left[\frac{z\theta}{\epsilon^{z+1}} - \frac{z-1}{\epsilon^z} \right] d\epsilon \right\} \\
&= -(z-1) K_z f(\theta) + \frac{\theta^z z^{z+1}}{\Gamma(z)} \int_0^\infty f(\epsilon) \frac{e^{-z\theta/\epsilon}}{\epsilon^{z+1}} d\epsilon \qquad (3.110)
\end{aligned}
$$

Then

$$
\tilde{K}_z g(x) = \int_x^\infty \frac{K_z g(\theta)}{\theta} d\theta = -(z-1)\tilde{K}_z f(x) + \frac{z^{z+1}}{\Gamma(z)} \int_x^\infty \theta^{z-1} \int_0^\infty f(\epsilon) \frac{e^{-z\theta/\epsilon}}{\epsilon^{z+1}} d\epsilon d\theta \qquad (3.111)
$$

and using the change of variables $d\tilde{\epsilon} = \frac{z+1}{z} d\epsilon$,

$$
\begin{aligned}
\tilde{K}_z g(x) = &- (z-1)\tilde{K}_z f(x) \\
&+ \frac{\Gamma(z+1)}{\Gamma(z)} \frac{z}{z+1} \int_x^\infty \frac{1}{\theta} \int_0^\infty \frac{(z+1)^{z+1}}{\Gamma(z+1)} \theta^z \frac{e^{-(z+1)\theta/\tilde{\epsilon}}}{\tilde{\epsilon}^{z+1}} f\left(\tilde{\epsilon} \frac{z}{z+1} \right) d\tilde{\epsilon} d\theta \qquad (3.112)
\end{aligned}
$$

We conclude noting that $\Gamma(z+1) = z\Gamma(z)$. □

3.10 Proofs: semi-parametric posterior concentration rate

3.10.1 Uniform convergence near zero

Proof of Proposition 6. The proof follows the steps of the proof of Theorem 6 of [44].

For $x \in (0, a]$, define $A_\epsilon^{x,+} = \{f : f(x) - f_0(x) \geq \epsilon\}$, following the proof of Theorem 5 in [44], with ϵ_n as in Theorem 8, we have that

$$
\Pi(A^x_{e_0 \epsilon_n^{2/3}} | X^n) = o_p(1),
$$

for $e_0 > 0$ a large enough constant.

Similarly, define $A_\epsilon^{x,+} = \{f : f(x) - f_0(x) \geq \epsilon\}$. We have that for $h_n = C_1 \epsilon_n^{2/3}$

$$
\begin{aligned}
f(0) - f_0(0) &\leq f(0) - f(h_n) + h_n^{-1} \int_0^{h_n} |f(t) - f_0(t)| dt \\
&\leq f(0) - f(h_n) + h_n^{-1} \epsilon_n \\
&\leq f(0) - f(h_n) + \epsilon_n^{1/3}.
\end{aligned}
$$

Furthermore we have

$$\begin{aligned} f(0) - f(h_n) &= \int_0^{h_n} \frac{1}{\theta} g_{P,z}(\theta) d\theta \\ &= \int_0^{\infty} \frac{z}{z-1} \epsilon^{-1} \Pr\left(\Gamma(z-1,z) \leq \frac{h_n}{\epsilon}\right) dP(\epsilon). \end{aligned}$$

For $k, b > 0$ and x small enough we have

$$\Pr\left(\Gamma(k,b) \leq x\right) \leq 2\frac{b^k x^k}{\Gamma(k)}.$$

Note also that there exists an absolute constant C_1 such that for all $z \geq 2$ and for all $0 < x < 1$

$$\frac{z^{z-1} x^{z-1}}{\Gamma(z)} \leq C_1 x.$$

We thus have, for $\nu_n = C_2 \epsilon_n^{1/3}$,

$$\begin{aligned} \int_0^{\delta} \epsilon \Pr\left(\Gamma(z-1,z) \leq \frac{h_n}{\epsilon}\right) dQ^{(0)}(\epsilon) &\leq \int_0^{\nu_n} \epsilon dQ^{(0)}(\epsilon) + \delta \int_{\nu_n}^{\delta} \Pr\left(\Gamma(z-1,z) \leq C\nu_n\right) dQ^{(0)}(\epsilon) \\ &\leq \nu_n + \delta C_1 C \nu_n. \end{aligned}$$

Furthermore

$$\begin{aligned} \int_{\delta}^{\infty} \epsilon^{-1} \Pr\left(\Gamma(z-1,z) \leq \frac{h_n}{\epsilon}\right) dQ^{(1)}(\epsilon) &\leq \delta^{-1} \Pr\left(\Gamma(z-1,z) \leq \frac{h_n}{\delta}\right) \\ &\leq \delta^{-2} C_1 h_n \end{aligned}$$

Thus, for $z \geq 2$

$$\begin{aligned} f(0) - f(h_n) &\leq \frac{z}{z-1}(1-p_n)^{-1}\left(p\nu_n(1+\delta C_1 C) + (1-p)\delta^{-2} C_1 h_n\right) \\ &\leq 2\left(\nu_n(1+\delta C_1 C)(1-p_n)^{-1} + \delta^{-2} C_1 h_n\right). \end{aligned}$$

Thus there exists $M > 0$ large enough such that

$$f(0) - f_0(0) \leq M r_n$$

where

$$r_n = \epsilon_n^{1/3} \max(1,\delta) \delta_n^{-1}$$

since $\delta^{-2} C_1 \epsilon_n^{2/3} = o(\epsilon_n^{1/3} \max(1,\delta)\delta_n^{-1})$.

We thus have that $\Pi(A_{Mr_n}^{0,+}|X^n) = o_{P_0}(1)$. We conclude the proof using the same arguments as the proof of Theorem 6 in [44]. □

3.10.2 Interaction term

Proof of Theorem 9. Define

$$\mathcal{F}_n = \{f_{P,z}, P \in \mathcal{P}_n, \quad z \leq \bar{z}_n; d_H(f_{P,z}, f_0) \leq \epsilon_n; \sup_{x \in [0,a]} |f_{P,z}(x) - f_0(x)| \leq u_n\},$$

with $\bar{z}_n = n^{2/(2\beta+1)}(\log n)^q$, and $y = x - \theta_0$ and $Y_i = X_i - \theta_0$. Then, by Theorem 8 and Proposition 6, for any compact $\Theta_0 \subset \Theta$,

$$\sup_{\theta_0 \in \Theta_0} \sup_{f_0 \in \mathcal{P}'(\beta,\dots)} \mathbb{E}_{P_0} \mathbb{P}_n(f \notin \mathcal{F}_n \mid \mathbf{X}^n) = o(1).$$

1. By Lemma 26,

$$\sup_{f \in \mathcal{F}_n} \left(\frac{M_n}{n} \left\| \frac{f'}{f} \right\|_\infty \right) \leq C \frac{M_n z_n}{na} \lesssim n^{[1-2\beta]/(2\beta+1)} [\log n]^{1+q} = o(1)$$

   for $\beta > 1/2$.

2. Due to Lemma 27 and Corollary 4

$$\sup_{f_0 \in \mathcal{F}_1} \mathbb{P}_{0,f_0} \left(\sup_{f \in \mathcal{F}_n} |\mathbb{G}_n(\frac{f'}{f})| > \epsilon \sqrt{n} \right) = o(1).$$

3. Now we study the last condition for $|h| \leq M_n$, $f \in \mathcal{F}_n$, $y > 0$:

$$I_1 := \left| \frac{f'(y+h/n)}{f(y+h/n)} - \frac{f_0'(y+h/n)}{f_0(y+h/n)} - \frac{f'(y)}{f(y)} + \frac{f_0'(y)}{f_0(y)} \right|.$$

   We have that

$$\left| \frac{f'(y+t/n)}{f(y+t/n)} - \frac{f'(y)}{f(y)} \right| \leq \frac{|f'(y+t/n) - f'(y)|}{f(y+t/n)} + \frac{|f'(y)|}{f(y)f(y+t/n)}[f(y) - f(y+t/n)].$$

   On $\mathcal{F}_n$, $z \leq \bar{z}_n = n^{2/(2\beta+1)}(\log n)^q$, and hence $z^2/n \leq n^{(3-2\beta)/(2\beta+1)}(\log n)^q \to 0$ for $\beta > 3/2$. Using Lemma 26, for $|t| \leq nM_n$, $z \leq \bar{z}_n < a/M_n$, $t_n < a$,

$$\begin{aligned} I_1 &\leq \sup_{|h| \leq M_n} \sup_{f \in \mathcal{F}_n} \sup_{y} \frac{|f'(y+t/n) - f'(y)|}{f(y+t/n)} + \frac{|f'(y)|}{f(y)f(y+t/n)}[f(y) - f(y+t/n)] \\ &\lesssim \bar{z}_n^2 M_n + \sqrt{\bar{z}_n}[\delta^{-1} + \frac{p_n}{1-p_n}] I(y \leq t_n) + \bar{z}_n^2 M_n. \end{aligned}$$

   Taking $t_n = \frac{\delta+1-p_n}{\log n \sqrt{\bar{z}_n}}$ so that for $\beta > 1$

$$\bar{z}_n T_n / t_n = O(1) \frac{[\log n]^2 \bar{z}_n^{3/2}}{n(\delta + 1 - p_n)} = n^{-2(\beta-1)/(2\beta+1)} [\log n]^{\tilde{q}} = o(1),$$

   we have

$$I_1 \leq n^{[3-2\beta]/(2\beta+1)} (\log n)^{2q+1} + C[\delta^{-1} + \frac{p_n}{1-p_n}] n^{1/(2\beta+1)} (\log n)^{q/2} I(y \leq t_n)$$

   which implies

$$\int [\delta^{-1} + \frac{p_n}{1-p_n}] \sqrt{\bar{z}_n} I(y \leq t_n) f_0(y) dy \leq f_0(0) [\delta^{-1} + \frac{p_n}{1-p_n}] \sqrt{\bar{z}_n} t_n = o(1)$$

   uniformly over f_0 with uniformly bounded $f_0(0)$ and the first term goes to 0 for $\beta > 3/2$.

□

Now we state the lemmas used in the proof. These lemmas are proved in Section 3.10.3.

Lemma 23. *For all $f_{P;z} \in \mathcal{F}_n$, with $u_n \le f_0(0)/2$ and $f_0(0) \le 1/(5\delta)$,*

(i)

$$p \int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) \le (1-p), \quad \int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon \le 4 f_0(0).$$

(ii) If $\delta^2 \le (1-p)/p$, then

$$\int_\delta^\infty \frac{dQ^{(1)}(\epsilon)}{\epsilon} \ge \frac{f_0(0)}{5}.$$

(iii) For all $2\delta \le y \le a$

$$\int_{y/2}^\infty \frac{dQ^{(1)}(\epsilon)}{\epsilon} \ge \frac{z-1}{z}(f_0(y) - u_n) - \frac{2Pr(\Gamma(z-1,z) \ge y/(2\delta))}{\delta}$$

In particular, if δ is small enough

$$\int_\delta^\infty \frac{dQ^{(1)}(\epsilon)}{\epsilon} \ge f_0(0)(1 - P[\Gamma(z-1,z) \ge 2] - 2M\delta)$$

where M is defined by $f_0(0) - f_0(x) \le M|x|$; and since $z \ge 2$,

$$\int_\delta^\infty \frac{dQ^{(1)}(\epsilon)}{\epsilon} \ge f_0(0)(0.98 - 2M\delta) \ge f_0(0)/2.$$

Lemma 24. *Define $t_n \le t_0/\log n$ for some $t_0 > 0$. Let sequence T_n satisfy $T_n \to 0$, $nT_n \to \infty$, $T_n\sqrt{z \log z} \to 0$ and $zT_n/t_n \to 0$ as $n \to \infty$. Assume also that $\delta^2 \le (1-p)/p$.*

If $f \in \mathcal{F}_n$, with $\sup_{y\in(0,a)} |f_{P;z}(y) - f_0(y)| \le u_n = o(1)$ and $f_0(0) \le 1/(5\delta)$

$$\begin{aligned} \left\| \frac{f'_{P,z}(y)}{f_{P,z}(y)} \right\|_{\infty,a} = \sup_{y\in(0,a)} \left| \frac{f'_{P,z}(y)}{f_{P,z}(y)} \right| \le \frac{C\sqrt{z}}{1-p} \left(p + \frac{1-p}{\delta} \right), \\ \sup_{|u|\le T_n} \sup_{y\in(t_n,a)} |f'_{P;z}(y+u) - f'_{P;z}(y)| \le \frac{T_n}{t_n\delta^2(1-p)} \left[z\sqrt{\log z} + t_n^{-1}\sqrt{z} \right]. \end{aligned} \tag{3.113}$$

Lemma 25. *Assume that $f_0 \in \mathcal{F}_1$ and that $f_0(0) \le 1/(5\delta)$. Assume also that $z^2/n = o(1)$ and $z > 2$, and there exists $B > 0$ such that*

$$[\log n]^{-B} \le \delta \wedge (1-p) \wedge f_0(a), \quad \delta^2 \le (1-p)/p \wedge a/4.$$

Then, there exists C depending on $f_0(0), f_0(a), M, a$ such that

$$\left\| \frac{f'_{P,z}}{f_{P,z}} \right\|_\infty \le \frac{Cz}{a}. \tag{3.114}$$

Lemma 26. *Assume that $f_0 \in \mathcal{F}_1$ and that $f_0(0) \le 1/(5\delta)$. Assume also that $z^2/n = o(1)$ and $z > 2$, and there exists $B > 0$ such that*

$$[\log n]^{-B} \le \delta \wedge (1-p) \wedge f_0(a), \quad \delta^2 \le (1-p)/p \wedge a/4.$$

Then, there exists C depending on $f_0(0), f_0(a), M, a$ such that for all $|u| \leq M_n/n$ with M_n going to infinity arbitrarily slowly,

$$
\begin{aligned}
\sup_y |\frac{f_{P;z}(y+u) - f_{P;z}(y)}{f_{P;z}(y+u)}| \leq \frac{CzM_n}{n}, \\
\sup_{y \leq t_n} \left| f'_{P,z}(y) - f'_{P,z}(y+u) \right| \leq C, \\
\sup_{y \geq a} \left| \frac{f'_{P,z}(y) - f'_{P,z}(y+u)}{f_{P,z}(y+u)} \right| \leq C \frac{z^2 M_n}{n},
\end{aligned}
\tag{3.115}
$$

for some small $t_n \leq a$.

Lemma 27. *Let assumptions of Theorem 9 hold. Fix $\epsilon > 0$ arbitrarily small. Define $P_n = P\mathbb{1}([a_n, b_n])$ where $a_n \leq r_n/2 = \epsilon/(3nf_0(0))$ and $b_n = n^{B_1} y_n$ where y_n: $F_0(y_n) \geq 1 - \epsilon/(3n)$ and $B_1 > 0$.*

Then, on $\Omega_n = \{a_n \leq Y_i \leq y_n, \, i = 1, \ldots, n\}$ with $P_0(\Omega_n) \geq 1 - \epsilon$, for B_1 large enough, the following two conditions hold:

for $\beta > 1$,

$$
\sup_{f_{P;z} \in \mathcal{F}_n} \left(\frac{f'_{P;z}(\cdot)}{f_{P;z}(\cdot)} - \frac{f'_{P_n;z}(\cdot)}{f_{P_n;z}(\cdot)} \right) = o(1) \tag{3.116}
$$

and for $\beta > 3/2$,

$$
\sup_{f_{P;z} \in \mathcal{F}_n} \left| \mathbb{G}_n \left(\frac{f'_{P_n;z}(\cdot)}{f_{P_n;z}(\cdot)} \right) \right| = o(\sqrt{n}). \tag{3.117}
$$

Corollary 4. *Under assumptions of Lemma 27, combining the two statements of the lemma, we have*

$$
\sup_{f_0 \in \mathcal{F}_1} \mathbb{P}_{0,f_0} \left(\sup_{f \in \mathcal{F}_n} |\mathbb{G}_n(\frac{f'}{f})| > \epsilon\sqrt{n} \right) = o(1).
$$

3.10.3 Proofs of lemmas, interaction term for mixtures

In this section we prove lemmas used in the proof of Theorem 9.

Proof of Lemma 23 . We first prove (i).

$$
f_{P;z}(0) = \frac{z-1}{z} \frac{p\int_0^\delta \epsilon dQ^{(0)}(\epsilon) + (1-p)\int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + (1-p)} \leq f_0(0) + u_n \tag{3.118}
$$

so that

$$
\int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon \leq \frac{z}{z-1}(f_0(0) + u_n) \left(\frac{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)}{1-p} + 1 \right)
$$

Moreover

$$
p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) \leq 1 - p
$$

since otherwise we would have

$$4(f_0(0)+u_n) \geq \frac{p\int_0^\delta \epsilon dQ^{(0)}(\epsilon)}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)} \geq \frac{1}{\delta},$$

which is not possible. Hence

$$\int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon \leq 4f_0(0)$$

Now $\delta^2 \leq (1-p)/p$ implies that

$$p\int_0^\delta \epsilon dQ^{(0)}(\epsilon) + (1-p)\int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon \leq 2(1-p)\int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon$$

so that, using (3.118) , we obtain

$$\int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon \geq \frac{z}{2(z-1)}(f_0(0)+u_n) \geq \frac{f_0(0)}{5}.$$

Now let $a \geq y \geq 4\delta$, so that $|f(y)-f_0(y)| \leq u_n$,

$$\begin{aligned}
f(y) &\leq \frac{z^z}{\Gamma(z)(1-p)}\int_0^\infty \int_{y/\epsilon}^\infty u^{z-2}e^{-zu}du\frac{dP(\epsilon)}{\epsilon} \\
&\leq \frac{zp\int_0^\delta \epsilon dQ^{(0)}(\epsilon)Pr(\Gamma(z-1,z) \geq y/\delta)}{(z-1)(1-p)} \\
&+ \frac{zPr(\Gamma(z-1,z) \geq y/(2\delta))}{\delta(z-1)} + \frac{z}{z-1}\int_{2\delta}^{+\infty}\frac{dQ^{(1)}(\epsilon)}{\epsilon} \\
&\leq \frac{2zPr(\Gamma(z-1,z) \geq y/(2\delta))}{\delta(z-1)} + \frac{z}{z-1}\int_{y/2}^{+\infty}\frac{dQ^{(1)}(\epsilon)}{\epsilon}
\end{aligned}$$

so that

$$\int_{y/2}^{+\infty}\frac{dQ^{(1)}(\epsilon)}{\epsilon} \geq \frac{z-1}{z}(f_0(y)-u_n) - \frac{2Pr(\Gamma(z-1,z) \geq y/(2\delta))}{\delta}.$$

□

Proof of Lemma 24. For the sake of simplicity, throughout the proof we write $f_{P,z} = f$. Let $y \in (0,a)$ then $f(y) \geq f(a) \geq f_0(a) - u_n \geq f_0(a)/2$ for n large enough and

$$\frac{|f'(y)|}{f(y)} \leq \frac{2|f'(y)|}{f_0(a)} = 2\frac{z^z\int_0^\infty (y/\epsilon)^{z-2}e^{-zy/\epsilon}\epsilon^{-2}dP(\epsilon)}{\Gamma(z)f_0(a)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]}$$

Note that there exists $C > 0$ such that

$$z^z/\Gamma(z)x^{z-2}e^{-zx} \leq C\sqrt{z}e^{-(z-2)h(x)}, \quad h(x) = x-1-\log x, \quad x > 0. \tag{3.119}$$

So that using Lemma 23, for $y \in (0,a)$,

$$\frac{|f'(y)|}{f(y)} \leq C\sqrt{z}\left(\frac{1}{1-p}+\frac{1}{\delta}\right).$$

Now let $t_n \le y \le a$, we have

$$
\begin{aligned}
|f'(y) - f'(y+u)| &= \frac{z^z \int_0^\infty [(y/\epsilon)^{z-2} e^{-zy/\epsilon} - ((y+u)/\epsilon)^{z-2} e^{-z(y+u)/\epsilon}] dP(\epsilon)/\epsilon^2}{\Gamma(z)(p \int \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p)} \\
&\le \frac{C\sqrt{z}|I_2|}{1-p} \le \frac{T_n z \sqrt{\log z}}{t_n(1-p)\delta^2} \\
&+ \frac{\sqrt{z} p [\log n]^{-0.5(z-2)/t_0}}{(1-p)^2} + \frac{T_n \sqrt{z}}{t_n^2 (1-p)\delta^2}
\end{aligned} \tag{3.120}
$$

using Lemma 28. Choosing t_0 small enough, we have

$$
|f'(y) - f'(y+u)| \lesssim \frac{T_n t_n^{-1}}{\delta^2(1-p)} \left[z\sqrt{\log z} + t_n^{-1}\sqrt{z} \right]. \tag{3.121}
$$

□

Lemma 28. *For the mixture prior defined in Section* **??** *with $\delta^2 \le (1-p)/p$, denote*

$$
D_k = e^{z-2} \int_0^\infty [(y/\epsilon)^{z-2} e^{-zy/\epsilon} - ((y+u)/\epsilon)^{z-2} e^{-z(y+u)/\epsilon}] \epsilon^{-k} dP(\epsilon) \tag{3.122}
$$

for $k \in [0,2]$, $y \in (t_n, a)$, $0 \le u \le T_n$ and $z \in [2, \bar{z}_n]$.

Assume that $\bar{z}_n \to \infty$, $T_n, t_n \to 0$, $T_n/t_n \to 0$ and $T_n/t_n = o\left([\bar{z}_n \log(\bar{z}_n)]^{-1/2}\right)$ as $n \to \infty$.

Then, for any arbitrarily large constant $M > 0$,

$$
|D_k| \lesssim \begin{cases} \frac{T_n}{t_n \delta^k} \sqrt{z \log z} & \text{if } z \in [M, \bar{z}_n], \\ p e^{-0.5(z-2)/t_n} t_n^{2(2-k)}/(1-p) + \frac{T_n}{t_n^2 \delta^k} & \text{if } z \in [2, M]. \end{cases} \tag{3.123}
$$

Proof of Lemma 28. Note that there exists $C > 0$ such that

$$
z^z/\Gamma(z) x^{z-2} e^{-zx} \le C\sqrt{z} e^{-(z-2)h(x)}, \quad h(x) = x - 1 - \log x, \quad x > 0. \tag{3.124}
$$

This is a consequence of $z^z/\Gamma(z) \le C\sqrt{z}$ and $e^{z-2}x^{z-2}e^{-zx} \le e^{-(z-2)h(x)}$. Note that $h(x) \ge 0$ for all $x > 0$, and $h'(x) > 0$ for $x > 1$ and $h'(x) < 0$ for $x < 1$. For large enough x ($x > 6$), $h(x) \ge 0.5x$.

1. First we consider the case of large z, namely $z \ge M$ for an arbitrarily large M. For all $|x-1| \ge \delta_z$, we have $h(x) \ge \delta_z^2/3$ where $\delta_z = a_0\sqrt{\log z/z}$ ([4]). By the assumptions of the lemma, $u/y \le T_n/t_n \to 0$. Hence, if $|y/\epsilon - 1| \le \delta_z$ then, for n large enough,

$$
|(y+u)/\epsilon - 1| = |(1+u/y)(y/\epsilon - 1) + u/y| \ge (1 + T_n/t_n)\delta_z - T_n/t_n \ge \frac{\sqrt{3}}{2}\delta_z.
$$

Therefore if $z \geq M$ for an arbitrarily large M, using the lower bound on $h(x)$

$$\begin{aligned}
|D_k| &\leq \int_{y/(1+\delta_z)}^{y(1+\delta_z)} e^{-(z-2)h(y/\epsilon)} \left|1 - e^{z[\log(1+u/y)-u/\epsilon]}(1+O(u/y))\right| dP(\epsilon)/\epsilon^k \\
&\quad + \int_{[y/(1+\delta_z),y(1+\delta_z)]^c} [e^{-(z-2)h(y/\epsilon)} + e^{-(z-2)h((y+u)/\epsilon)}]dP(\epsilon)/\epsilon^k \\
&\leq \int_{y/(1+\delta_z)}^{y(1+\delta_z)} \left|1 - e^{\frac{zu}{y}[(1-y/\epsilon)+O(u/y)]}(1+O(u/y))\right| dP(\epsilon)/\epsilon^k \\
&\quad + 4e^{-(z-2)\delta_z^2/4}\max(p\delta^{2-k}, (1-p)/\delta^k) \\
&\leq \frac{zu}{y}[\delta_z + O(T_n/t_n)]\max(p\delta^2/(1-p),1)/\delta^k + 4e^{-(z-2)\delta_z^2/4}\max(p\delta^2/(1-p),1)/\delta^k \\
&\lesssim [\frac{T_n\sqrt{z\log z}}{t_n} + z^{-a_0/5}]\delta^{-k} \\
&\lesssim \frac{T_n\sqrt{z\log z}}{t_n\delta^k}
\end{aligned}$$

for large enough z and a_0, since $z|1-y/\epsilon||u|/y \leq \sqrt{z\log z}T_n/t_n \to 0$ and $zu^2/y^2 \leq zT_n^2/t_n^2 = o([\log \bar{z}_n]^{-1}) \to 0$. Recall also that $\delta^2 \leq (1-p)/p$.

2. Now consider the case $2 \leq z \leq M$. Note that for small zu/y (e.g. when $zT_n/t_n = o(1)$),

$$\begin{aligned}
|1-(1+u/y)^{z-2}e^{-zu/\epsilon}| = |1-e^{-zu/\epsilon+(z-2)\log(1+u/y)}| &= |1-e^{-zu/\epsilon+(z-2)u/y(1+O(u/y))}| \\
&\leq z|u||1/\epsilon - 1/y|(1+O(zu/y+zu/\epsilon)),
\end{aligned} \tag{3.125}$$

with the last inequality holding for $zu/\epsilon = o(1)$.

For $\epsilon \leq t_n^2$, $h(y/\epsilon) \geq h(1/t_n) \geq 0.5/t_n$ for large enough n, and

$$\begin{aligned}
&e^{z-2}\int_0^{t_n^2} |(y/\epsilon)^{z-2}e^{-zy/\epsilon} - ((y+u)/\epsilon)^{z-2}e^{-z(y+u)/\epsilon}|dP(\epsilon)/\epsilon^k \\
&\leq 2pe^{-(z-2)h(1/t_n)}t_n^{2(2-k)}/(1-p) \leq 2e^{-0.5(z-2)/t_n}t_n^{2(2-k)}/(1-p_n).
\end{aligned}$$

Moreover, using (3.125) again,

$$\begin{aligned}
&e^{z-2}\int_{t_n^2}^{\infty} |(y/\epsilon)^{z-2}e^{-zy/\epsilon} - ((y+u)/\epsilon)^{z-2}e^{-z(y+u)/\epsilon}|dP(\epsilon)/\epsilon^k \\
&\leq \int_{t_n^2}^{\infty} e^{-(z-2)h(y/\epsilon)}|1-(1+u/y)^{z-2}e^{-zu/\epsilon}|dP(\epsilon)/\epsilon^k \\
&\lesssim zut_n^{-2}\int_{t_n^2}^{\infty} dP(\epsilon)/\epsilon^k \lesssim \frac{zT_n}{t_n^2\delta^k}.
\end{aligned}$$

So, when z is small, for $0 \leq u \leq T_n$,

$$|D_k| \lesssim pe^{-0.5(z-2)/t_n}t_n^{2(2-k)}/(1-p) + \frac{zT_n}{t_n^2\delta^k}.$$

For $z = 2$, simple algebra shows that $D_k \leq 2\delta^{-k}u/y \leq 2\delta^{-k}T_n/t_n$.

□

Proof of Lemma 25. We need to prove an upper bound on f'/f for $y > a$.

$$\frac{|f'(y)|}{f(y)} = \frac{\int_0^\infty (y/\epsilon)^{z-2} e^{-zy/\epsilon} dP(\epsilon)/\epsilon^2}{\int_0^\infty \int_{y/\epsilon}^\infty u^{z-2} e^{-zu} du dP(\epsilon)/\epsilon}.$$

Since the function $\epsilon \to \int_{y/\epsilon}^\infty u^{z-2} e^{-zu} du$ is increasing,

$$\frac{\int_0^\delta (y/\epsilon)^{z-2} e^{-zy/\epsilon} dP(\epsilon)/\epsilon^2}{\int_0^\infty \int_{y/\epsilon}^\infty u^{z-2} e^{-zu} du dP(\epsilon)/\epsilon} \leq \frac{4z \int_0^\delta (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(0)}(\epsilon)}{(1-p)(y/(2\delta))^{z-2} e^{-zy/2\delta} \int_{2\delta}^\infty dQ^{(1)}(\epsilon)/\epsilon}$$

and if $z \geq z_0$ so that $f_0(0) \geq 4M\delta \frac{2Pr(\Gamma(z-1,z)\geq 2)}{\delta}$, then

$$\frac{\int_0^\delta (y/\epsilon)^{z-2} e^{-zy/\epsilon} dP(\epsilon)/\epsilon^2}{\int_0^\infty \int_{y/\epsilon}^\infty u^{z-2} e^{-zu} du dP(\epsilon)/\epsilon} \lesssim \frac{z e^{-(z-2)[h(y/\delta)-h(y/(2\delta)]}}{(1-p) f_0(0)}.$$

Otherwise, under the constraint that $\delta^2 \leq (1-p)/p$,

$$\frac{\int_0^\delta (y/\epsilon)^{z-2} e^{-zy/\epsilon} dP(\epsilon)/\epsilon^2}{\int_0^\infty \int_{y/\epsilon}^\infty u^{z-2} e^{-zu} du dP(\epsilon)/\epsilon} \leq \frac{4z \int_0^\delta (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(0)}(\epsilon)}{(1-p)(y/\delta)^{z-2} e^{-zy/\delta} \int_\delta^\infty dQ^{(1)}(\epsilon)/\epsilon} \leq \frac{4z}{(1-p) f_0(0)}$$
$$\lesssim \frac{1}{(1-p) f_0(0)}.$$

Moreover, using (iii) of Lemma 23 we have that if z is large enough so that $P(\Gamma(z-1,z) \geq a/(2\delta)) \leq 5\delta f_0(a)$, i.e. for small δ, if $e^{-za/(4\delta)} \leq 5\delta f_0(a)$,

$$\int_{a/2}^\infty \int_{y/\epsilon} u^{z-2} e^{-zu} du dQ^{(1)}(\epsilon)/\epsilon \geq \int_{2y/a} u^{z-2} e^{-zu} du \int_{a/2}^\infty Q^{(1)}(\epsilon)/\epsilon$$
$$\geq z^{-1} (2y/a)^{z-2} e^{-2zy/a} \int_{a/2}^\infty Q^{(1)}(\epsilon)/\epsilon \geq \frac{f_0(a)}{5} z^{-1} (2y/a)^{z-2} e^{-2zy/a}$$

also

$$\int_\delta^{a/4} (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2 \leq \frac{4(4y/a)^{z-2} e^{-4zy/a} f_0(0)}{\delta}$$

so that

$$\frac{\int_\delta^{a/4} (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_{a/2}^\infty \int_{y/\epsilon} u^{z-2} e^{-zu} du dQ^{(1)}(\epsilon)/\epsilon} \lesssim \frac{z f_0(0) 2^{z-2} e^{-2zy/a}}{f_0(a)\delta} \lesssim \frac{z f_0(0) 2^{z-2} e^{-2z}}{f_0(a)\delta}.$$

We also have

$$\frac{\int_{a/4}^\infty (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_\delta^\infty \int_{y/\epsilon} u^{z-2} e^{-zu} du dQ^{(1)}(\epsilon)/\epsilon} \leq \frac{z \int_{a/4}^\infty (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_\delta^\infty (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon} \leq \frac{4z}{a}$$

and using the same computations if z is small

$$\frac{\int_\delta^\infty (y/\epsilon)^{z-2}e^{-zy/\epsilon}dP(\epsilon)/\epsilon^2}{\int_0^\infty \int_{y/\epsilon}^\infty u^{z-2}e^{-zu}dudP(\epsilon)/\epsilon} \leq \frac{z}{\delta}.$$

So that we finally obtain for all $y > a$, if z is small (i.e. $P(\Gamma(z-1,z) \geq a/(2\delta)) \geq 5\delta f_0(a)$)

$$\left|\frac{f'(y)}{f(y)}\right| \leq \frac{z}{\delta}$$

and if z is such that $e^{-za/(4\delta)} \leq 5\delta f_0(a)$,

$$\left|\frac{f'(y)}{f(y)}\right| \lesssim \frac{4z}{a} + \frac{zf_0(0)2^{z-2}e^{-2z}}{f_0(a)\delta} \lesssim \frac{4z}{a}.$$

□

Proof of Lemma 26. First we study $(f(y) - f(y + t/n))/f(y + t/n)$. It is sufficient to study it for $t > 0$, as for $t < 0$ and $\tilde{y} = y + t/n \geq 0$,

$$\begin{aligned}|f(y) - f(y + t/n)|/f(y + t/n) &= (f(\tilde{y}) - f(\tilde{y} + |t|/n))|/f(\tilde{y}) \\ &\leq (f(\tilde{y}) - f(\tilde{y} + |t|/n))|/f(\tilde{y} + |t|/n)\end{aligned}$$

so the same bound will apply. As f decreases, we can study

$$\begin{aligned}f(y) - f(y + t/n) \leq& \frac{z^z}{\Gamma(z)(1-p)}\int_0^\delta \epsilon \int_{y/\epsilon}^{(y+t/n)/\epsilon} u^{z-2}e^{-zu}dudQ^{(0)}(\epsilon) \\ &+ \frac{z^z zt}{n\Gamma(z)}\int_\delta^\infty [(y+t/n)/\epsilon]^{z-2}e^{-zy/\epsilon}dQ^{(1)}(\epsilon)/\epsilon.\end{aligned}$$

Denote $u = t/n$. Let $y \leq \delta/(1+\delta_z)$ then $y/\epsilon \leq 1/(1+\delta_z)$ and

$$\frac{z^z zu}{\delta\Gamma(z)}\int_\delta^\infty [(y+u)/\epsilon]^{z-2}e^{-zy/\epsilon}dQ^{(1)}(\epsilon)/\epsilon \leq \frac{8f_0(0)z^{3/2}ue^{-\delta_0^2\log z/2}}{\delta}$$

using (i) of Lemma 23 and Lemma C.2 of [4] (see Lemma 34 in Section 3.11), and hence

$$\begin{aligned}\frac{z^z}{\Gamma(z)(1-p)}\int_0^\delta \epsilon \int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2}e^{-zv}dvdQ^{(0)}(\epsilon) &\leq \frac{\sqrt{z}u}{(1-p)}\int_0^\delta e^{-(z-2)h(y/\epsilon)}dQ^{(0)}(\epsilon) \\ &\leq \frac{\sqrt{z}u}{(1-p)}.\end{aligned}$$

Thus, for all $y \leqslant \delta/(1+\delta_z)$ and $\delta_0 > \sqrt{6}$,

$$\frac{|f(y) - f(y+u)|}{f(y+u)} \lesssim \frac{\sqrt{z}u}{f_0(0)(1-p)}.$$

If $\delta/(1+\delta_z) < y \leq a - T_n$, then $y/\epsilon > 1/(1+\delta_z)$ and $y + u \leq a$. Hence, again using Lemma C.2

of [4] (see Lemma 34 in Section 3.11),

$$\begin{aligned}\frac{z^z p\int_0^\delta \epsilon \int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2}e^{-zv}dvdQ^{(0)}(\epsilon)}{\Gamma(z)f(y+u)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]} &\le \frac{z^z p\int_0^\delta \epsilon(u/\epsilon)(y/\epsilon)^{z-2}e^{-zy/\epsilon}dQ^{(0)}(\epsilon)}{\Gamma(z)f(y+u)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]}\\ &\lesssim \frac{\sqrt{z}up\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)}{y^2 f(y+u)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]}\\ &\lesssim \frac{f_0(0)\sqrt{z}u}{\delta^2 f_0(a)}\end{aligned}$$

while

$$\begin{aligned}&\frac{z^z(1-p)\int_\delta^\infty \epsilon^{-1}\int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2}e^{-zv}dvdQ^{(1)}(\epsilon)}{\Gamma(z)f(y+u)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]}\\ &\le \frac{z^z(1-p)u\int_\delta^\infty y^{-2}t(y/\epsilon)^z e^{-zy/\epsilon}dQ^{(0)}(\epsilon)}{n\Gamma(z)f(y+u)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]} \lesssim \frac{\sqrt{z}uf_0(0)}{\delta^2 f_0(a)}.\end{aligned} \tag{3.126}$$

Thus, for $\delta/(1+\delta_z) < y \le a - T_n$,

$$\frac{|f(y)-f(y+u)|}{f(y+u)} \lesssim \frac{\sqrt{z}uf_0(0)}{\delta^2 f_0(a)}.$$

Finally if $y > a/2$, using (iii) of Lemma 23 with $y = 2\delta$ when z is large enough so that $P(\Gamma(z-1,z) \ge 2) \le 4f_0(0)\delta$, we have $\int_{2\delta}^\infty dQ^{(1)}(\epsilon)/\epsilon \ge f_0(a)/4$ and hence

$$\begin{aligned}f(y+u) &\ge \frac{z^z}{\Gamma(z)}\int_0^\infty\int_{(y+u)/\epsilon}^\infty v^{z-2}e^{-zv}dv\frac{[p\epsilon dQ^{(0)}(\epsilon)+(1-p)\epsilon^{-1}dQ^{(1)}(\epsilon)]}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p}\\ &\ge \frac{z^z(1-p)[(y+u)/(2\delta)]^{z-2}e^{-z(y+u)/(2\delta)}}{\Gamma(z)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]}\int_{2\delta}^\infty dQ^{(1)}(\epsilon)/\epsilon\\ &\ge \frac{z^z(1-p)0.25f_0(a)[(y+u)/(2\delta)]^{z-2}e^{-z(y+u)/(2\delta)}}{\Gamma(z)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]}.\end{aligned} \tag{3.127}$$

Therefore,

$$\begin{aligned}&\frac{pz^z\int_0^\delta \epsilon\int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2}e^{-zv}dvdQ^{(0)}(\epsilon)}{\Gamma(z)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon)+1-p]f(y+u)}\\ &\le \frac{4p\int_0^\delta \epsilon[(y+u)/\epsilon]^{z-2}e^{-zy/\epsilon}(1-e^{-zu/\epsilon})dQ^{(0)}(\epsilon)}{(1-p)[(y+u)/(2\delta)]^{z-2}e^{-z(y+u)/(2\delta)}f_0(0)}\\ &= \frac{4p(2\delta)^{(z-2)}e^{z(y+u)/(2\delta)}}{(1-p)f_0(0)}\int_0^\delta \epsilon^{-(z-1)}e^{-zy/\epsilon}(1-e^{-zu/\epsilon})dQ^{(0)}(\epsilon).\end{aligned}$$

Function $H(u) = u^{z-1}e^{-zyu}(1-e^{-ztu/n})$ is monotone non increasing for $u \ge 1/\delta$ (more generally, for $u \geq 1/y$) so that

$$
\begin{aligned}
&\frac{pz^z \int_0^\delta \epsilon \int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2} e^{-zv} dv dQ^{(0)}(\epsilon)}{\Gamma(z)[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p] f(y+u)} \leq \frac{4p(2\delta)^{(z-2)} e^{z(y+u)/(2\delta)} \delta^{1-z} e^{-zy/\delta}(1 - e^{-zu/\delta})}{(1-p) f_0(0)} \\
&= \frac{p2^z \delta^{-1} e^{-z(y-u)/(2\delta)}}{(1-p) f_0(0)} (1 - e^{-zu/\delta}) \leq \frac{p\delta^{-1} \min(1, zu\delta^{-1}) e^{-z(y-u)/(2\delta) + z\log 2}}{(1-p) f_0(0)} \\
&\lesssim \frac{zu}{(1-p)\delta^2} e^{-za/(5\delta)}.
\end{aligned}
\tag{3.128}
$$

Similarly

$$
\begin{aligned}
&\frac{z^z(1-p)\int_\delta^{a/4} \epsilon^{-1} \int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2} e^{-zv} dv dQ^{(1)}(\epsilon)}{\Gamma(z) f(y+u)} \\
&\leq \frac{z\int_\delta^{a/4} \epsilon^{-1} \int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2} e^{-zv} dv dQ^{(1)}(\epsilon)}{([2(y+u)/a]^{z-2} e^{-2z(y+u)/a} \int_{a/2}^\infty dQ^{(1)}(\epsilon)/\epsilon} \\
&\leq \frac{4[4(y+u)/a]^{z-2} e^{-4zy/a}(1 - e^{-4zu/a}) \int_\delta^{a/4} dQ^{(1)}(\epsilon)/\epsilon}{[2(y+u)/a]^{z-2} e^{-2z(y+u)/a} f_0(0)} \\
&\leq 0.25 \min(1, 4zu/a) e^{-2z(y-u)/a + z\log 2} \\
&\lesssim \frac{zu}{a} e^{-z}.
\end{aligned}
$$

Moreover

$$
\frac{z^z \int_{a/4}^\infty \epsilon^{-1} \int_{y/\epsilon}^{(y+u)/\epsilon} v^{z-2} e^{-zv} dv dQ^{(1)}(\epsilon)}{\Gamma(z) f(y+u)} \leq \frac{4tz \int_{a/4}^\infty \epsilon^{-z+1} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)}{na \int_\delta^\infty \epsilon^{-z+1} e^{-zy/\epsilon} Q^{(1)}(\epsilon)} \leq \frac{4uz}{a}.
$$

Therefore, for $|t/n| \leq T_n$,

$$
\frac{|f(y) - f(y+t/n)|}{f(y+t/n)} \lesssim \begin{cases} T_n\sqrt{z}/(1-p), & 0 < y \leq \delta/(1+\delta_z), \\ T_n\sqrt{z}/\delta^2, & \delta/(1+\delta_z) < y \leq a - T_n, \\ T_n z e^{-z}, & y > a - T_n. \end{cases}
$$

Now we bound

$$
\frac{f'(y) - f'(y+u)}{f(y+u)} = -\frac{z^z \int_0^\infty [(y/\epsilon)^{z-2} e^{-zy/\epsilon} - ((y+u)/\epsilon)^{z-2} e^{-z(y+u)/\epsilon}] dP(\epsilon)/\epsilon^2}{\Gamma(z) f(y+u)(p\int \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p)}.
$$

If $y > a/2$, similarly to equation (3.128), using the lower bound (3.127),

$$\begin{aligned}
&\frac{|f'(y) - f'(y+u)|}{f(y+u)} \\
&= \frac{\left|\int_0^\infty [(y/\epsilon)^{z-2} e^{-zy/\epsilon} - ((y+u)/\epsilon)^{z-2} e^{-z(y+u)/\epsilon}][p dQ^{(0)}(\epsilon) + \epsilon^{-2}(1-p) dQ^{(1)}]\right|}{\int_0^\infty \epsilon^{-1} \int_{(y+u)/\epsilon}^\infty v^{z-2} e^{-zv} dv [p\epsilon^2 dQ^{(0)}(\epsilon) + (1-p) dQ^{(1)}(\epsilon)]} \\
&\lesssim \frac{p y^{z-2} \int_0^\delta |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z+2} e^{-zy/\epsilon} dQ^{(0)}(\epsilon)}{0.25(1-p) f_0(a) [(y+u)/(2\delta)]^{z-2} e^{-z(y+u)/(2\delta)}} \\
&\quad + \frac{(1-p) y^{z-2} \int_\delta^\infty |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)}{0.25(1-p) f_0(a) [(y+u)/(2\delta)]^{z-2} e^{-z(y+u)/(2\delta)}} \\
&\leq \frac{4p(2\delta)^{z-2} e^{z(y+u)/(2\delta)}}{(1-p) f_0(a)} \int_0^\delta |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z+2} e^{-zy/\epsilon} dQ^{(0)}(\epsilon) \\
&\quad + \frac{4(2\delta)^{z-2} e^{z(y+u)/(2\delta)}}{f_0(a)} \int_\delta^\infty |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z} e^{-zy/\epsilon} dQ^{(1)}(\epsilon).
\end{aligned}$$

For $\epsilon \leq a/4$ and $y \geq a/2$, there exists $c \in (0,1)$ such that

$$\begin{aligned}
\log[e^{-zu/\epsilon}(1+u/y)^{z-2}] &= -\frac{zt}{n\epsilon} + (z-2)\log(1+u/y) \\
&= -zu\left[\frac{1}{\epsilon} - \frac{1-2/z}{y}\left(1 - \frac{cu}{2y}\right)\right] \\
&\leq -zu\left[\frac{4}{a} - \frac{2(1-2/z)}{a}\right] \leq -\frac{2zu}{a} < 0
\end{aligned}$$

i.e. $e^{-zu/\epsilon}(1+u/y)^{z-2} < 1$. In particular,

$$\begin{aligned}
1 - (1+u/y)^{z-2} e^{-zu/\epsilon} &\leq 1 - e^{-zu/\epsilon} - (z-2)u/y e^{-zu/\epsilon} \\
&\leq I(\epsilon \leq q) + (zu/\epsilon - (z-2)u/y e^{-zu/\epsilon}) I(\epsilon > q)
\end{aligned}$$

for some small $q \leq \delta$.

Around 0,

$$\int_0^q |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z+2} e^{-zy/\epsilon} dQ^{(0)}(\epsilon) \leq q^{-z+2} e^{-zy/q}.$$

Next,

$$\int_q^\delta |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z+2} e^{-zy/\epsilon} dQ^{(0)}(\epsilon) \leq \tfrac{zu}{q} \delta^{-z+2} e^{-zy/\delta}$$

and

$$\int_\delta^{a/2} |1 - (1+u/y)^{z-2} e^{-zu/\epsilon}| \epsilon^{-z} e^{-zy/\epsilon} dQ^{(1)}(\epsilon) \leq \tfrac{zu}{q} \delta^{-z} e^{-zy/\delta}.$$

Therefore,

$$\begin{aligned}
&\frac{4p(2\delta)^{z-2}e^{z(y+u)/(2\delta)}}{(1-p)f_0(a)}\int_0^\delta |1-(1+u/y)^{z-2}e^{-zu/\epsilon}|\epsilon^{-z+2}e^{-zy/\epsilon}dQ^{(0)}(\epsilon)\\
&\quad\le \frac{4p(2\delta)^{z-2}e^{z(y+u)/(2\delta)}}{(1-p)f_0(a)}\left[q^{-z+2}e^{-zy/q}+\frac{zu}{q}\delta^{-z+2}e^{-zy/\delta}\right]\\
&\quad\le \frac{pz|u|e^{-z[(y-u)/(2\delta)-\log 2]}}{(1-p)qaf_0(a)}\left[\frac{nq}{zt}(q/\delta)^{-z+2}e^{-zy(1/q-1/\delta)}+1\right]
\end{aligned}$$

and

$$\begin{aligned}
&\frac{4(2\delta)^{z-2}e^{z(y+u)/(2\delta)}}{f_0(a)}\int_\delta^{a/2}|1-(1+u/y)^{z-2}e^{-zu/\epsilon}|\epsilon^{-z}e^{-zy/\epsilon}dQ^{(1)}(\epsilon)\\
&\le \frac{4(2\delta)^{z-2}e^{z(y+u)/(2\delta)}}{f_0(a)}\frac{z|u|}{q}\delta^{-z}e^{-zy/\delta}\le\frac{z|u|(\delta)^{-2}e^{-z[(y-u)/(2\delta)-\log 2]}}{qf_0(a)}.
\end{aligned}$$

Finally, for $\epsilon > a/2$,

$$\begin{aligned}
&\frac{\int_{a/2}^\infty |(y/\epsilon)^{z-2}e^{-zy/\epsilon}-((y+u)/\epsilon)^{z-2}e^{-z(y+u)/\epsilon}|dQ^{(1)}(\epsilon)}{\int_\delta^\infty \epsilon^{-1}\int_{(y+u)/\epsilon}^\infty v^{z-2}e^{-zv}dvdQ^{(1)}(\epsilon)}\\
&\le\quad 16z^2ue^{2zu/a}\frac{\int_{a/2}^\infty((y+u)/\epsilon)^{z-1}e^{-z(y+u)/\epsilon}dQ^{(1)}(\epsilon)}{\int_\delta^\infty[(y+u)/\epsilon]^{z-1}e^{-z(y+u)/\epsilon}dQ^{(1)}(\epsilon)}\le\frac{16z^2te^{2zu/a}}{n}
\end{aligned}$$

since

$$\begin{aligned}
-(z-2)t/(ny) &\le (1+t/(ny))^{2-z}-1\le(1+t/(ny))^{2-z}e^{zu/\epsilon}-1\le e^{zu/\epsilon}-1\\
&\le zu/\epsilon e^{zu/\epsilon}
\end{aligned}$$

and

$$\begin{aligned}
\max((z-2)u/y, zu/\epsilon e^{zu/\epsilon}) &\le \max(2zu/a, zu/\epsilon e^{zu/\epsilon})\\
&\le 2zu/ae^{2zu/a}.
\end{aligned}$$

Combining these bounds, we obtain for $y \ge a$:

$$\begin{aligned}
\frac{|f'(y)-f'(y+u)|}{f(y+u)} &\lesssim \frac{z^2|t|}{n}+\frac{z^2|t|}{n}\frac{z^{-1}e^{-z[(a-u)/(2\delta)-\log 2]}}{qf_0(a)}\\
&\quad\times\left[\delta^{-2}+\frac{p}{(1-p)a}\left[\frac{nq}{zt}(q/\delta)^{-z+2}e^{-za(1/q-1/\delta)}+1\right]\right].
\end{aligned}$$

Taking $q = \delta/2$ and assuming $|u| \le a/2$ and $a \ge 8\delta\log 2$ (the constant here can be reduced from $8\log 2$ down to 1 if necessary), the bound becomes

$$\begin{aligned}&\frac{|f'(y)-f'(y+u)|}{f(y+u)}\\ &\lesssim z^2|u|\left[1+\frac{z^{-1}e^{-z[(a-u)/(2\delta)-\log 2]}}{0.5\delta f_0(a)}\left[\delta^{-2}+\frac{p}{(1-p)a}\left[\frac{2^{-4}}{zu}e^{-z[a/\delta-\log 2]}+1\right]\right]\right]\\ &\lesssim z^2|u|+z|u|\frac{e^{-za/(8\delta)}}{\delta}\left[\frac{1}{\delta^2}+\frac{p}{(1-p)a}\right]+\frac{p}{(1-p)a\delta}e^{-3z}.\end{aligned}$$

Combining this with the bound for $y\le a/2$, we have

$$\begin{aligned}\frac{|f'(y)-f'(y+u)|}{f(y+u)} &\lesssim z^2|u|+z|u|\tfrac{e^{-z/5}}{\delta}\left[\tfrac{1}{\delta^2}+\tfrac{p}{(1-p)a}\right]+\tfrac{p}{(1-p)a\delta}e^{-3z}+\delta^{-1}+\tfrac{p\sqrt{z}}{1-p}\\ &\lesssim z^2|u|+\max[e^{-2z},z|u|]\tfrac{e^{-z/5}p_n}{(1-p_n)a\delta}+\delta^{-1}+\tfrac{p_n\sqrt{z}}{1-p_n}\end{aligned}$$

since $\delta^{-2}\le p_n/(1-p_n)$.

□

Proof of Lemma 27. Let

$$I_2=\frac{f'(y)}{f(y)}=\frac{f'_{P_n;z}(\cdot)}{f_{P_n;z}(\cdot)}+\left(\frac{f'_{P;z}(\cdot)}{f_{P;z}(\cdot)}-\frac{f'_{P_n;z}(\cdot)}{f_{P_n;z}(\cdot)}\right)$$

and for all $y\in[r_n,y_n]$, writing $d\bar{P}=p\epsilon^2dQ^{(0)}+(1-p)dQ^{(1)}$ we have

$$\begin{aligned}&\frac{f'_{P;z}(\cdot)}{f_{P;z}(\cdot)}-\frac{f'_{P_n;z}(\cdot)}{f_{P_n;z}(\cdot)}=\frac{p\int_0^{a_n}(y/\epsilon)^{z-1}e^{-zy/\epsilon}dQ^{(0)}(\epsilon)+(1-p)\int_{b_n}^\infty(y/\epsilon)^{z-1}e^{-zy/\epsilon}\epsilon^{-2}dQ^{(1)}(\epsilon)}{\int_{a_n}^{b_n}\int_{y/\epsilon}v^{z-2}e^{-zv}dv\epsilon^{-1}d\bar{P}(\epsilon)}\\ &+\frac{|f'_P(y)|}{f_P(y)}\frac{p\int_0^{a_n}\int_{y/\epsilon}v^{z-2}e^{-zv}dv\epsilon dQ^{(0)}(\epsilon)+(1-p)\int_{b_n}^\infty\int_{y/\epsilon}v^{z-2}e^{-zv}dv\epsilon^{-1}dQ^{(1)}(\epsilon)}{\int_{a_n}^{b_n}\int_{y/\epsilon}v^{z-2}e^{-zv}dv\epsilon^{-1}d\bar{P}(\epsilon)}\\ &\le\frac{z(y/a_n)^{z-1}e^{-zy/a_n}}{(y/(2a_n))^{z-2}e^{-zy/(2a_n)}\int_{2a_n}^\infty\epsilon^{-1}d\bar{P}(\epsilon)}+\frac{\Gamma(z-1)(y/b_n)^{z-2}e^{-zy/b_n}\int_{b_n}^\infty\epsilon^{-1}d\bar{P}(\epsilon)}{z^{z-1}\int_y^\infty\epsilon^{-1}d\bar{P}(\epsilon)}\\ &+\frac{|f'_P(y)|}{f_P(y)}\left(\frac{pa_n\int_{y/a_n}v^{z-2}e^{-zv}dv}{\int_{r_n/2}^{b_n}\int_{y/\epsilon}v^{z-2}e^{-zv}dv\epsilon^{-1}d\bar{P}(\epsilon)}+\frac{2(1-p)\int_{b_n}^\infty\epsilon^{-1}dQ^{(1)}(\epsilon)}{\int_{y/2}^{b_n}\epsilon^{-1}d\bar{P}(\epsilon)}\right)\\ &\le\frac{4z2^{z-1}e^{-zy/(2a_n)}}{f_0(a)(1-p)}+4\sqrt{z}e^{-(z-2)h(y/b_n)}+\frac{|f'_P(y)|}{f_P(y)}\frac{4pa_ne^{-z(1-\epsilon)y/a_n}}{(1-p)(2y/r_n)^{z-2}e^{-2zy/r_n}f_0(a)}\\ &+\frac{|f'_P(y)|}{f_P(y)}\mathbb{1}(y\le a)\,\frac{2(1-p)}{b_nf_0(a)}+\frac{|f'_P(y)|}{f_P(y)}\mathbb{1}(y>a)\,\frac{2e_nQ^{(1)}(b_n,+\infty)}{Q^{(1)}(y_n,e_nb_n)}\\ &\lesssim\frac{e^{-zr_n/(4a_n)}}{1-p}+e^{-zn^{B_1}/2}+\left(\sqrt{z}(1-p)\wedge\delta^2+\frac{z}{\delta}\right)n^{-B_1/2}=o(1)\end{aligned}$$

by choosing B_1 large enough. This proves the first statement of the lemma.

Now we prove (3.117). The proof is based on Lemma 19.33 of [49]. We construct a bracketing of $\mathcal{G}_n$ in the form: $(g_L^i;g_U^i)$ $i\le N_{[]}$ such that $g_U^i\ge g_L^i$ and

- $\forall g \in \mathcal{G}_n$ there exists $g_L^i \leq g \leq g_U^i$ and
$$\mathbb{P}_{0,0}(g_U^i - g_L^i) = o(1)$$
uniformly where $P_{0,0}$ is the distribution of $Y \sim f_0(y)$, $y \geq 0$,
-
$$\frac{\max_i(\|g_L^i\|_\infty \vee \|g_U^i\|_\infty)}{\sqrt{n}} \log(1 + N_{[]}) + \max_i(\|g_L^i\|_2 \vee \|g_U^i\|_2)\sqrt{\log(1 + N_{[]})} = o(\sqrt{n}).$$

Note that the first part $\mathbb{P}_{0,0}(g_U^i - g_L^i) = o(1)$ uniformly implies that the centred difference is also small, uniformly, which is the first condition of Lemma 19.33 of [49]. For $g_L^i(y) \leq g(y) \leq g_U^i(y)$, the upper bound on the centered g is

$$\begin{aligned} g(y) - \mathbb{P}_{0,0}g &= g(y) - g_U^i(y) + g_U^i(y) - \mathbb{P}_{0,0}g_U^i - \mathbb{P}_{0,0}(g - g_U^i) \\ &\leq g_U^i(y) - \mathbb{P}_{0,0}g_U^i + \mathbb{P}_{0,0}(g_U^i - g_L^i) \end{aligned}$$

and the lower bound is

$$\begin{aligned} g(y) - \mathbb{P}_{0,0}g &= g(y) - g_L^i(y) + g_L^i(y) - \mathbb{P}_{0,0}g_L^i - \mathbb{P}_{0,0}(g - g_L^i) \\ &\geq g_L^i(y) - \mathbb{P}_{0,0}g_L^i - \mathbb{P}_{0,0}(g_U^i - g_L^i). \end{aligned}$$

Then, the difference between the upper and the lower bounds is

$$g_U^i(y) - \mathbb{P}_{0,0}g_U^i - [g_L^i(y) - \mathbb{P}_{0,0}g_L^i + 2\mathbb{P}_{0,0}(g_U^i - g_L^i)] = g_U^i(y) - g_L^i(y) + \mathbb{P}_{0,0}(g_U^i - g_L^i),$$

hence it is sufficient to show $\mathbb{P}_{0,0}(g_U^i - g_L^i) = o(1)$.

Let $\omega_n = o(1)$ arbitrarily slowly, and construct a bracketing net of $\mathcal{G}_n$ where $\mathbb{P}_{0,0}(g_U^i - g_L^i) \leq \omega_n$ for all i. Recall that the elements of $\mathcal{G}_n$ are of the form $g = \frac{f'_{P_n;z}}{f_{P_n;z}}$.

Lemma 29. *Consider the net on $[2, \bar{z}_n]$, $S_z = \{2(1 + r_n)^k, k \leq K_z\}$ with $\bar{z}_n = n^{2/(2\beta+1)}(\log n)^q$ and $K_z = \log(\bar{z}_n/2)(\log(1 + r_n))^{-1}$ and the net on $(0, p_n)$, $S_p = \{d_n k, k \leq p_n/d_n\}$. Let $u > 0$ arbitrarily small and define y'_n such that $F_0(y'_n, +\infty) = u\delta/\bar{z}_n$, $k_{1,n} = \log(y'_n/\delta)[\log(1 + c_n)]^{-1}$. We define the nets on (a_n, δ) and (δ, b_n) respectively*

$$E_0 = \{\epsilon_{n,k} = a_n(1 + c_n)^k, k \leq k_{0,n}\}$$

$$E_1 = \{\epsilon'_{n,k} = \delta(1 + c_n)^k, k \leq k_{1,n},\ \epsilon'_{n,k_{2,n}+1} = b_n\},$$

the net of $[0, 1]$ defined by $S_\pi = \{\pi_n(1 + \tau_n)^k, k \leq K_\pi := \log(1/\pi_n)[\log(1 + \tau_n)]^{-1}\}$. Set $k_n = k_{0,n} + k_{1,n} + 1$. We assume that $\tau_n, d_n, c_n, r_n, \pi_n$ go to 0 and p_n goes to 1.

Define $Q_{0,k} \in S_\pi$ the smallest value larger than $Q^{(0)}(\epsilon_{n,k}, \epsilon_{n,k+1})$ and similarly for $Q_{1,k}$. Let

$$\begin{aligned} \bar{\mathcal{P}}_n = \{(P, z) \in \mathcal{P}_n;\ z \in (2; \bar{z}_n);\ &\int_{y'_n}^{b_n} \epsilon^{-1} dQ^{(1)}(\epsilon) \geq \pi_n, \\ &\min(Q_{0,k}, Q_{1,k'}, k \leq k_{0,n}, k' \leq k_{1,n}) \geq \pi_n\}. \end{aligned}$$

Then if $\pi_n = \exp[-n^B]$ for some $B > 0$, $a_n \geq n^{-B_a}$, $c_n \geq n^{-B_c}$, $C \geq \delta \geq n^{-B_\delta}$ for some $B_a, B_c, B_\delta, C > 0$,

$$\Pi(\bar{\mathcal{P}}_n^c) \leq e^{-C_1 n\epsilon_n^2}, \tag{3.129}$$

and

$$\begin{aligned}\log \mathcal{N} &\lesssim 1 + \frac{\log(y_n'/a_n)}{c_n}\left(\log\log(y_n'/(\pi_n a_n)) - \log(c_n\tau_n)\right) + \log\log(\bar{z}_n) - \log(d_n r_n) \\ &\lesssim n^{B_c}\log n\,(\log n + \log(1/\tau_n)) + \log(1/(d_n r_n)).\end{aligned} \tag{3.130}$$

Remark 5. Under condition (3.17), $1 - F_0(x) = \epsilon$ implies $x \le [\epsilon/C_2]^{-1/(\rho_1+2)}$ since

$$\epsilon = 1 - F_0(x) = \int_x^\infty u^{-2}u^2 f_0(u)du \le C_2 x^{-2}(1+x)^{-\rho_1} \le C_2 x^{-\rho_1-2}.$$

Now we construct upper and lower bounds on $g(y) = \frac{f'_{P_n;z}(y)}{f_{P_n;z}(y)}$, so that $P_{0,0}(g_U - g_L) = o(1)$. We will do it separately for $y < a$ and $y > a$.

Denote

$$\bar{h}(y;\epsilon_{n,k}) = \sup_{\epsilon\in(\epsilon_{n,k},\epsilon_{n,k+1})} (y/\epsilon)^{z-2}e^{-zy/\epsilon}, \quad \underline{h}(y;\epsilon_{n,k}) = \inf_{\epsilon\in(\epsilon_{n,k},\epsilon_{n,k+1})} (y/\epsilon)^{z-2}e^{-zy/\epsilon}$$

or

$$\bar{h}(y;\epsilon'_{n,k}) = \sup_{\epsilon\in(\epsilon'_{n,k},\epsilon'_{n,k+1})} (y/\epsilon)^{z-2}e^{-zy/\epsilon}, \quad \underline{h}(y;\epsilon'_{n,k}) = \inf_{\epsilon\in(\epsilon'_{n,k},\epsilon'_{n,k+1})} (y/\epsilon)^{z-2}e^{-zy/\epsilon}.$$

1. $y \in (0,a)$.

We need to construct a bracketing net of $\mathcal{G}_n$ where $\mathbb{P}_{0,0}(g_U^i - g_L^i) \le \omega_n$ for some $\omega_n = o(1)$. If $g \in \mathcal{G}_n$ then $g = \frac{f'_{P_n;z}}{f_{P_n;z}}$ and we have that $\|\mathbb{1}([0,a])\,(f - f_0)\|_\infty = o(\omega_n)$ by choosing ω_n accordingly. Therefore on $[0,a]$, recalling that $f'_{P_n;z} < 0$ and hence $g(y) < 0$, we have

$$\frac{-f'_{P_n;z}(y)}{f_0(y) + o(\omega_n)} \le -g(y) \le \frac{-f'_{P_n;z}(y)}{f_0(y) - o(\omega_n)}.$$

Hence, for $y \in (0,a)$ it is sufficient to construct a bracketing net of $f'_{P_n;z}$. Recall that

$$-f'_{P_n;z}(y) = \frac{z^z}{\Gamma(z)}\frac{p\int_{a_n}^\delta (y/\epsilon)^{z-2}e^{-zy/\epsilon}dQ^{(0)}(\epsilon) + (1-p)\int_\delta^{b_n}(y/\epsilon)^{z-2}e^{-zy/\epsilon}\epsilon^{-2}dQ^{(1)}(\epsilon)}{p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p}.$$

a) First consider $z \in (z_j, z_{j+1})$ such that $z_{j+1} \le w_n$ where w_n is going to infinity arbitrarily slowly.

Assume that d_n is such that $2d_n \le p \le p_n$ and if $1 - p_n > 2d_n$. Then,

$$\begin{aligned}-f'_{P_n;z}(y) &\le \frac{z_{j+1}^{z+1}}{\Gamma(z_{j+1})}\frac{(p+d_n)\sum_{k=1}^{k_{0,n}}\bar{h}(y;\epsilon_{n,k})Q_{0,k} + (1-p+d_n)\sum_{k=1}^{k_{1,n}}\bar{h}(y;\epsilon'_{n,k})Q_{1,k}/(\epsilon'_{n,k})^2}{(p-d_n)[\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n]_+ + (1-p-d_n)} \\ &\ge \frac{z_j^{z_j}}{\Gamma(z_j)}\frac{(p-d_n)\sum_{k=1}^{k_{0,n}}\underline{h}(y;\epsilon_{n,k})Q_{0,k}(1-\tau_n)}{(p+d_n)[\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n]_+ + (1-p+d_n)} \\ &+ \frac{z_j^{z_j}}{\Gamma(z_j)}\frac{(1-p-d_n)\sum_{k=1}^{k_{1,n}}\underline{h}(y;\epsilon'_{n,k})Q_{1,k}(1-\tau_n)/(\epsilon'_{n,k+1})^2}{(p+d_n)[\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n]_+ + (1-p+d_n)}.\end{aligned}$$

If $\epsilon_{n,k+1} \le y(z_j + r_n)/(z_j + r_n - 2)$ then

$$\bar{h}(y; \epsilon_{n,k}; z) \le (y/\epsilon_{n,k+1})^{z_j - 2} e^{-z_j y/\epsilon_{n,k+1}}, \underline{h}(y; \epsilon_{n,k}; z) \ge (y/\epsilon_{n,k})^{z_j + r_n - 2} e^{-(z_j + r_n) y/\epsilon_{n,k}}$$

and vice versa if $\epsilon_{n,k} \ge y z_j/(z_j - 2)$. If $\epsilon_{n,k} \le y z_j/(z_j - 2) \le \epsilon_{n,k+1}$ then we set

$$\bar{h}_k(y) = e^{-z_j + 2}(1 - 2/z_j)^{z_j - 2}, \quad \underline{h}_k(y) \ge e^{-z_j - r_n + 2}(1 - 2/z_j)^{z_j - 2}(1 - 3c_n)$$

if $w_n c_n = o(1)$.
If $y \le \epsilon_{n,k+1}(z_j + r_n - 2)/(z_j + r_n)$,

$$\underline{h}(y; \epsilon_{n,k}) \ge \bar{h}(y; \epsilon_{n,k}) e^{-z_j y c_n/\epsilon_{n,k+1}} e^{-r_n y/\epsilon_{n,k}}$$

and if $y > \epsilon_{n,k+1}(z_j + r_n - 2)/(z_j + r_n)$,

$$\underline{h}(y; \epsilon_{n,k}) \ge \bar{h}(y; \epsilon_{n,k})(1 + c_n)^{-z_j + 1} e^{-r_n y/\epsilon_{n,k}} \ge \underline{h}(y; \epsilon_{n,k})(1 + c_n)^{-z_j}(1 - \frac{2r_n}{a_n}).$$

Then the difference between both bounds is bounded by (if n is large enough), as $d_n \le p_n/2$,

$$\begin{aligned}
\Delta_{P_n;z}(y) &\le 2(1 + z_j c_n + \frac{2r_n}{a_n} + \tau_n) \\
&\times \frac{z_j^{z_j}}{\Gamma(z_j)} \frac{p \sum_{k=1}^{k_{0,n}} \bar{h}(y; \epsilon_{n,k}) Q_{0,k} + (1 - p) \sum_{k=1}^{k_{1,n}} \bar{h}(y; \epsilon'_{n,k}) Q_{1,k}/(\epsilon'_{n,k})^2}{(p - d_n)[\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n]_+ + (1 - p - d_n)} \\
&+ 2d_n(1 + \delta^2) \frac{z_j^{z_j}}{\Gamma(z_j)} \frac{p \sum_{k=1}^{k_{0,n}} \bar{h}(y; \epsilon_{n,k}) Q_{0,k} + (1 - p) \sum_{k=1}^{k_{1,n}} \bar{h}(y; \epsilon'_{n,k}) Q_{1,k}/(\epsilon'_{n,k})^2}{(p - d_n)[\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n]_+ + (1 - p - d_n)} \\
&+ \frac{2d_n(2 + p) z_j^{z_j}}{\Gamma(z_j)(1 - p)^2} \left(p \sum_{k=1}^{k_{0,n}} \bar{h}(y; \epsilon_{n,k}) Q_{0,k} + (1 - p) \sum_{k=1}^{k_{1,n}} \bar{h}(y; \epsilon'_{n,k}) Q_{1,k}/(\epsilon'_{n,k})^2 \right) \\
&\le C \left(\frac{d_n}{1 - p_n} + p d_n + z_j c_n + \tau_n + \frac{r_n}{a_n} \right) \times \\
&\frac{p \sum_{k=1}^{k_{0,n}} \bar{h}(y; \epsilon_{n,k}) Q_{0,k} + (1 - p) \sum_{k=1}^{k_{1,n}} \bar{h}(y; \epsilon'_{n,k}) Q_{1,k}/(\epsilon'_{n,k})^2}{(1 - p)}
\end{aligned}$$

for some constant C. If also $d_n \le \tau_n p_n$, $r_n \le \tau_n a_n$ there exists $C' > 0$

$$\Delta_{P_n;z}(y) \le C'(\tau_n + w_n c_n) \frac{p \sum_{k=1}^{k_{0,n}} \bar{h}(y; \epsilon_{n,k}) Q_{0,k} + (1 - p) \sum_{k=1}^{k_{1,n}} \bar{h}(y; \epsilon'_{n,k}) Q_{1,k}/(\epsilon'_{n,k})^2}{(1 - p)}.$$

Then

$$\begin{aligned}
\int_{r_n}^{a} f_0(y) \Delta_{P_n;z}(y) dy &\lesssim (\tau_n + w_n c_n) \left(\frac{\sum_{k=1}^{k_{0,n}} Q_{0,k} \epsilon_{n,k+1}}{(1 - p_n)} + \sum_{k=1}^{k_{1,n}} \frac{Q_{1,k}}{\epsilon'_{n,k}} \right) \\
&\lesssim (\tau_n + w_n c_n) \left(\frac{\delta}{1 - p_n} + \frac{1}{\delta} \right) = o(1)
\end{aligned}$$

as soon as $w_n c_n = o((1 - p_n)/\delta + \delta)$ and $\tau_n = o((1 - p_n)/\delta + \delta)$.

b) Now assume that $z_j > w_n$, then

$$\frac{z^z}{\Gamma(z)} u^{z-2} e^{-zu} = \frac{\sqrt{z}}{2\pi} u^{-2} e^{-zh(u)}(1+o(1)), \quad h(u) = u - \log u - 1 = \frac{(u-1)^2}{2}(1+O(u-1)).$$

In particular for all $|y/\epsilon - 1| > \delta_0 \sqrt{\log z/z} := \delta_z$, $zh(y/\epsilon) > c\delta_0 \log z$ if z is large enough and and if $y/\epsilon < 1 - \delta_z$ then

$$\frac{z^z}{\Gamma(z)}(y/\epsilon)^{z-2} e^{-zy/\epsilon} \lesssim e^{-(z-2)h(y/\epsilon)}\sqrt{z} \lesssim e^{-c\delta_0 \log z}\sqrt{z}$$

which implies that

$$\begin{aligned}
-f'_{P_n;z}(y) &\leq \frac{z_{j+1}^{z_{j+1}} e^{-z_{j+1}}}{\Gamma(z_{j+1})} \times \\
&\quad \frac{\left(\int \mathbb{1}(|y/\epsilon - 1| \leq \delta_z)\,(y/\epsilon)^{-2} e^{-z_j h(y/\epsilon)}[pQ^{(0)}(d\epsilon) + (1-p)Q^{(1)}(d\epsilon)/\epsilon^2] + O(z^{-c\delta_0 + 1/2})\right)}{[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p]} \\
&\geq \frac{z_j^{z_j} e^{-z_j}}{\Gamma(z_{j+1})} \frac{\left(\int \mathbb{1}(|y/\epsilon - 1| \leq \delta_z)\,(y/\epsilon)^{-2} e^{-z_j h(y/\epsilon)}[pQ^{(0)}(d\epsilon) + (1-p)Q^{(1)}(d\epsilon)/\epsilon^2]\right)}{[p\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p]}.
\end{aligned}$$

Then we use the same decomposition as before, with $c_n = o(1/\sqrt{z_j \log z_j})$ and $\bar{\epsilon}_k$ defined by $\bar{\epsilon}_k = \epsilon_{n,k}$ if $y > \epsilon_{n,k+1}$, $\bar{\epsilon}_k = \epsilon_{n,k+1}$ if $y < \epsilon_{n,k}$ and $\bar{\epsilon}_k = y$ if $y \in (\epsilon_{n,k}, \epsilon_{n,k+1})$;

$$\begin{aligned}
-f'_{P_n;z}(y) \leq \frac{z_{j+1}^{z_{j+1}} e^{-z_{j+1}}}{\Gamma(z_{j+1})} &\left(\frac{(p+d_n)\sum_k \mathbb{1}(|y/\epsilon_k - 1| \leq \delta_z)\,(y/\epsilon_{n,k+1})^{-2} e^{-z_j h(y/\bar{\epsilon}_k)} Q_{0,k}}{(p+d_n)\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + d_n + 1 - p} \right. \\
&\left. + \frac{(1-p+d_n)\sum_k \mathbb{1}(|y/\epsilon'_k - 1| \leq \delta_z)\,(y/\epsilon'_{n,k+1})^{-2} e^{-z_j h(y/\bar{\epsilon}'_k)} Q_{1,k}/(\epsilon'_{n,k})^2}{(p+d_n)\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + d_n + 1 - p} \right).
\end{aligned}$$

Similarly set $\underline{\epsilon}_k = \epsilon_{n,k}$ if $y < \epsilon_{n,k}$, $\underline{\epsilon}_k = \epsilon_{n,k+1}$ if $y > \epsilon_{n,k+1}$ and $\bar{\epsilon}_k = \epsilon_{n,k}$ if $y \in (\epsilon_{n,k}, \epsilon_{n,k+1})$ so that on $|y - \epsilon_{n,k}| \leq \delta_z$

$$\begin{aligned}
e^{-z_j h(y/\bar{\epsilon}_k)} - e^{-z_{j+1} h(y/\underline{\epsilon}_k)} &\leq e^{-z_j h(y/\bar{\epsilon}_k)}\left(z_{j+1} h(y/\underline{\epsilon}_k) - z_j h(y/\bar{\epsilon}_k)\right) \\
&\leq e^{-z_j h(y/\bar{\epsilon}_k)}\left(z_j r_n \delta_z^2 + 2 z_j c_n \delta_z\right)
\end{aligned}$$

This implies that

$$\begin{aligned}
-f'_{P_n;z}(y) \geq \frac{z_j^{z_j} e^{-z_j}}{\Gamma(z_j)} &\left(\frac{(p-d_n)_+ \sum_k \mathbb{1}(|y/\epsilon_k - 1| \leq \delta_z)\,(y/\epsilon_{n,k})^{-2} e^{-z_j h(y/\underline{\epsilon}_k)} Q_{0,k}(1-\tau_n)}{(p-d_n)_+ \left(\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n\right)_+ + 1 - p - d_n} \right. \\
&\left. + \frac{(1-p-d_n)\sum_k \mathbb{1}(|y/\epsilon'_k - 1| \leq \delta_z)\,(y/\epsilon'_{n,k})^{-2} e^{-z_{j+1} h(y/\underline{\epsilon}'_k)} Q_{1,k}(1-\tau_n)/(\epsilon'_{n,k+1})^2}{(p-d_n)(\int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) - d_n)_+ + 1 - p - d_n} \right)
\end{aligned}$$

and

$$\int_0^a \Delta_{P_n;z}(y) f_0(y)dy \leq \frac{2(r_n + d_n + \tau_n + z_j r_n \delta_z^2 + 2z_j c_n \delta_z)}{(1-p)} \left(1 + \delta^{-2}(d_n + 1 - p)\right)$$
$$+ \frac{d_n}{(1-p)} \frac{\log(1+\delta_z)}{c_n} \left(1 + (1 - p + d_n)\delta^{-2}\right) = o(1)$$

if

$$\begin{aligned} &r_n = o((1-p)(\log z_n)^{-1}); \quad d_n = o(\delta^2(1-p)); \quad \tau_n = o(1-p); \\ &c_n = o((1-p)\sqrt{z_j}/\sqrt{\log z_j}). \end{aligned} \tag{3.131}$$

2. We now study the behaviour for $y \in (a, +\infty)$. First note that since

$$\sup_{y \in (a,\infty)} \left| \frac{f'_{P_n;z}}{f_{P_n;z}} \right| \leq C\frac{z}{a},$$

if $z \in (z_j, z_{j+1})$ and $y'_n(z_j)$ satisfies

$$\int_{y'_u(z_j)}^{\infty} f_0(y)dy \leq u\delta/z_j \quad \text{for a small } u,$$

we can define an upper bound of $-f'_{P_n;z}/f_{P_n;z}$ on $(y'_u(z_j), +\infty)$ by Cz_j/a and a lower bound by 0.

We now study the behaviour for $y \in (a, y'_u(z_j))$. First assume that $z_j \leq Z_0$ fixed but arbitrarily large. We have that for all finite $y \in (0,\infty)$, there exists $\tilde{w}_n = o(1)$ such that

$$\Pi(|f_0(y) - f(y)| > \tilde{w}_n | Y^{(n)}) = o_{P_0}(1),$$

with the proof following the argument of Theorem 3 in [44]. Let $t > 0$ be small and L_t be such that $\int_{L_t}^{\infty} f_0(y)dy \leq t$ and set $t \leq 1/Z_0$, then for all $z \leq Z_0$ $y'_u(z) \leq L_t$ and the bracketing is valid. We can therefore assume that $z \geq Z_0$ where Z_0 is arbitrarily large.

Let $\delta \geq \epsilon_0 > a_n$, then

$$\begin{aligned} \frac{-f'_{P_n;z}}{f_{P_n;z}}(y) &= \frac{p\int_{a_n}^{\delta}(y/\epsilon)^{z-2}e^{-zy/\epsilon}dQ^{(0)}(\epsilon) + (1-p)\int_{\delta}^{b_n}(y/\epsilon)^{z-2}e^{-zy/\epsilon}dQ^{(1)}(\epsilon)/\epsilon^2}{p\int_{a_n}^{\delta}\int_{y/\epsilon}u^{z-2}e^{-zu}\epsilon dQ^{(0)}(\epsilon) + (1-p)\int_{\delta}^{b_n}\int_{y/\epsilon}u^{z-2}e^{-zu}dQ^{(1)}(\epsilon)/\epsilon} \\ &\leq \frac{p\int_{\epsilon_0}^{\delta}(y/\epsilon)^{z-2}e^{-zy/\epsilon}dQ^{(0)}(\epsilon) + (1-p)\int_{\delta}^{b_n}(y/\epsilon)^{z-2}e^{-zy/\epsilon}dQ^{(1)}(\epsilon)/\epsilon^2}{p\int_{a_n}^{\delta}\int_{y/\epsilon}u^{z-2}e^{-zu}\epsilon dQ^{(0)}(\epsilon) + (1-p)\int_{\delta}^{b_n}\int_{y/\epsilon}u^{z-2}e^{-zu}dQ^{(1)}(\epsilon)/\epsilon} \\ &\quad + \frac{4p(y/\epsilon_0)^{z-2}e^{-zy/\epsilon_0}}{(1-p)\int_{y/\delta}v^{z-2}e^{-zv}dv f_0(a)}. \end{aligned}$$

Note that we can take $\epsilon_0 = \delta$ in which case we obtain,

$$\begin{aligned}
\frac{-f'_{P_n;z}}{f_{P_n;z}}(y) &\leq \frac{\int_\delta^{b_n} (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_\delta^{b_n} \int_{y/\epsilon} u^{z-2} e^{-zu} dQ^{(1)}(\epsilon)/\epsilon} + \frac{4pz2^{z-2}e^{-zy/(2\delta)}}{(1-p)f_0(a)} \\
&\leq \frac{\int_\delta^{b_n} (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_\delta^{b_n} \int_{y/\epsilon} u^{z-2} e^{-zu} dQ^{(1)}(\epsilon)/\epsilon} + o(1)
\end{aligned}$$

and

$$\begin{aligned}
\frac{-f'_{P_n;z}}{f_{P_n;z}}(y) &\geq \frac{\int_\delta^{b_n} (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_\delta^{b_n} \int_{y/\epsilon} u^{z-2} e^{-zu} dQ^{(1)}(\epsilon)/\epsilon} - \frac{4pz2^{z-2}e^{-zy/(2\delta)}}{(1-p)f_0(a)} \\
&\geq \frac{\int_\delta^{b_n} (y/\epsilon)^{z-2} e^{-zy/\epsilon} dQ^{(1)}(\epsilon)/\epsilon^2}{\int_\delta^{b_n} \int_{y/\epsilon} u^{z-2} e^{-zu} dQ^{(1)}(\epsilon)/\epsilon} - o(1).
\end{aligned}$$

Denote $h_y(\epsilon) = (y/\epsilon)^{z-2} e^{-zy/\epsilon}$ and similarly it upper and lower bounds on $(\epsilon_{n,k}, \epsilon_{n,k+1})$. Set $G_y(\epsilon) = \int_{y/\epsilon} u^{z-2} e^{-zu} du$, the function is increasing and note that for all $\epsilon > y(1+\delta_z)$ and large z,

$$G_y(\epsilon) = \frac{\Gamma(z-1)}{z^{z-1}}(1+o(1))$$

and $h_y(\epsilon) \leq z G_y(\epsilon)$.

We have, for large z,

$$\frac{-f'_{P_n;z}}{f_{P_n;z}}(y) \leq \frac{1+\tau_n}{1-\tau_n} \times \frac{\sum_k \bar{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon^2_{n,k}}{\sum_k \mathbb{1}(\epsilon_k(u) > \epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k}) Q_{1,k}/\epsilon_{n,k+1} + (1 - z^{-\delta_0^2/2})(1-\tau_n) \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_k(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon}$$

and

$$\frac{-f'_{P_n;z}}{f_{P_n;z}}(y) \geq \frac{1-\tau_n}{1+\tau_n} \frac{\sum_k \underline{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon^2_{n,k+1}}{\sum_k \mathbb{1}(\epsilon_k(u) \geq \epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_k(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon}$$

where $\epsilon_k(u)$ is the smallest $\epsilon_{n,k}$ larger than $y'_u(z_j)(1+\delta_z)$. This leads to

$$
\begin{aligned}
\Delta_n(y) \le 4(\tau_n + c_n) &\frac{\sum_k \bar{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_k(u) \ge \epsilon_{n,k} \ge \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon} \\
&+ \frac{\sum_k [\bar{h}_y(\epsilon'_{n,k}) - \underline{h}_y(\epsilon'_{n,k})] Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_k(u) \ge \epsilon_{n,k} \ge \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon} \\
&+ \frac{2\sum_k \bar{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_k(u) \ge \epsilon_{n,k} \ge \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon} \times \\
&\left(\frac{\sum_k \mathbb{1}(\epsilon_{n,k} \ge \delta)\, Q_{1,k}[G_y(\epsilon_{n,k+1})/\epsilon_{n,k} - G_y(\epsilon_{n,k})/\epsilon_{n,k}]}{\sum_k \mathbb{1}(\epsilon_k(u) > \epsilon_{n,k} \ge \delta)\, G_y(\epsilon_{n,k}) Q_{1,k}/\epsilon_{n,k+1} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon} \right) \\
&+ \frac{z^{-\delta_0^2/2} \frac{\Gamma(z-1)}{z^{z-1}} \sum_k \bar{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_k(u) \ge \epsilon_{n,k} \ge \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon}.
\end{aligned}
$$

In particular, we have, similarly to previous bounds, that if $|y/\epsilon_{n,k} - 1| \le \delta_z$ then

$$
\bar{h}_y(\epsilon'_{n,k}) e^{-z_j} - \underline{h}_y(\epsilon'_{n,k}) e^{-z_{j+1}} \le \bar{h}_y(\epsilon'_{n,k}) e^{-z_j} \left(z_j r_n \delta_z^2 + 2\delta_z c_n z_j \right), \quad G_y(\epsilon_{n,k+1}) \ge \frac{e^{z_j}}{\sqrt{z_j}}
$$

while if $y/\epsilon_{n,k} < 1 - \delta_z$

$$
\bar{h}_y(\epsilon'_{n,k}) \le e^{z_j} z_j^{-\delta_0^2/2}, \quad G_y(\epsilon_{n,k+1}) \ge \frac{e^{z_j}}{\sqrt{z_j}}
$$

and if $y/\epsilon_{n,k} > 1 + \delta_z$

$$
\bar{h}_y(\epsilon'_{n,k}) - \underline{h}_y(\epsilon'_{n,k}) \le \bar{h}_y(\epsilon'_{n,k}) \left(z_j r_n \frac{y}{\epsilon_{n,k'}} + 2c_n z_j \frac{y}{\epsilon'_{n,k}} \right).
$$

In particular, using the same computations as before,

$$
\int_{a/2}^{b_n} dQ^{(1)}(\epsilon)/\epsilon \ge f_0(a) - o(1) - \frac{P(\Gamma(z-1, z) \ge a/\delta)}{\delta} \ge f_0(a)/2
$$

either if $z \ge n^t$ for some $t > 0$ as soon as $1 - p \ge n^{-H}$ and $\delta \ge n^{-H}$ for some $H > 0$, or for fixed z and $\delta \to 0$.

This implies that

$$
\begin{aligned}
\sum_k \mathbb{1}(\epsilon_{n,k} \ge \delta)\, G_y(\epsilon_{n,k}) Q_{1,k}(1+\tau_n)/\epsilon_{n,k} &\ge \frac{\Gamma(z)}{z^z} \int_\delta^{b_n} P(\Gamma(z-1, z) \ge y/\epsilon) dy \\
&\ge \frac{\Gamma(z)}{z^z} P(\Gamma(z-1, z) \ge 2y/a) f_0(a)/2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_a^{y_n} \frac{\sum_k [\bar{h}_y(\epsilon'_{n,k}) - \underline{h}_y(\epsilon'_{n,k})] Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon} f_0(y)dy \\
&\leq \frac{2z_j\sqrt{z_j} e^{-z_j h(4y/a)}\,(r_n + 2c_n)}{\delta^3 P(\Gamma(z-1,z) \geq 2y/a)\sqrt{2\pi} f_0(a)} \int_a^{y_n} y f_0(y)dy \\
&+ \int_a^{y_n} \frac{\sum_k \mathbb{1}\left(\epsilon'_{n,k} \geq a/4\right) \bar{h}_y(\epsilon'_{n,k}) \left(z_j r_n \frac{y}{\epsilon_{n,k'}} + 2c_n z_j \frac{y}{\epsilon'_{n,k}}\right) dQ^{(1)}(\epsilon'_{n,k})/(\epsilon'_{n,k})^2}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k}) Q_{1,k}/\epsilon_{n,k} + \frac{\Gamma(z-1)}{z^{z-1}} \int_{\epsilon_{k+1}(u)}^{b_n} dQ^{(1)}(\epsilon)/\epsilon} (f_0 - \sqrt{f_0 f})(y)dy \\
&\leq \frac{16z_j^2\,(r_n + 2c_n)}{a^2} \int_a^{y_n} y|\sqrt{f_0}(\sqrt{f_0} - \sqrt{f})(y)|dy \\
&+ \frac{2z_j\sqrt{z_j} e^{-z_j h(4y/a)}\,(r_n + 2c_n)}{\delta^3 P(\Gamma(z-1,z) \geq 2y/a)\sqrt{2\pi} f_0(a)} \int_a^{y_n} y f_0(y)dy \\
&\lesssim z_j^2 (r_n + c_n)\epsilon_n \sqrt{\int_a^{y_n} y^2 f_0(y)dy} = o(1)
\end{aligned}
$$

if $r_n + c_n = o(1/(z_n^2 \epsilon_n))$.

We now study

$$
\begin{aligned}
\Delta_2 &= \frac{\sum_k \underline{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}(1+\tau_n)/\epsilon_{n,k}} \times \\
&\quad \left(\frac{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, Q_{1,k}[G_y(\epsilon_{n,k+1})(1+\tau_n)/\epsilon_{n,k} - G_y(\epsilon_{n,k})(1-\tau_n)/\epsilon_{n,k+1}]}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k}) Q_{1,k}(1-\tau_n)/\epsilon_{n,k+1}}\right) \\
&\leq \frac{\sum_k \underline{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}(1+\tau_n)/\epsilon_{n,k}} \times \\
&\quad \left(\frac{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, Q_{1,k}[G_y(\epsilon_{n,k+1})(1+\tau_n)/\epsilon_{n,k} - G_y(\epsilon_{n,k})(1-\tau_n)/\epsilon_{n,k+1}]}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k}) Q_{1,k}/\epsilon_{n,k+1}}\right) \\
&\leq \frac{\sum_k \underline{h}_y(\epsilon'_{n,k}) Q_{1,k}/\epsilon_{n,k}^2}{\sum_k \mathbb{1}(\epsilon_{n,k} \geq \delta)\, G_y(\epsilon_{n,k+1}) Q_{1,k}/\epsilon_{n,k}} (2\tau_n + c_n) \\
&\lesssim \frac{z_j(\tau_n + c_n)}{\delta} \to 0
\end{aligned}
$$

if $z_j(\tau_n + c_n)/\delta \to 0$.

This leads to

$$\int_a^{y_n} f_0(y)\Delta_n(y)dy = o(1).$$

Now we collect the conditions:

$$
\begin{gathered}
r_n + c_n = o(1/(z_n^2 \epsilon_n)) = o(n^{-(4-\beta)/(2\beta+1)} \log n^{-(2t+q)}), \quad z_n(\tau_n + c_n)/\delta \to 0, \\
\delta^2 \leq (1-p_n)/p_n, \quad r_n \leq a_n \tau_n, \quad \tau_n = o((1-p_n)/\delta + \delta), \\
r_n = o((1-p)(\log z_n)^{-1}); \quad d_n = o(\delta^2(1-p)); \quad \tau_n = o(1-p),
\end{gathered}
$$

and for $w_n \le z_j \le \bar{z}_n$,

$$z_j^2 \epsilon_n (c_n + r_n) = o(1), \quad z_j(c_n + r_n)/\delta = o(1), \quad c_n = o(\min((1-p_n)z_j, 1)/\sqrt{z_j \log z_j}).$$

Taking $w_n \ge 1/(1-p_n)$, the above conditions hold for

$$c_n = o(\delta \bar{z}_n^{-1}), \quad r_n = o(\delta \bar{z}_n^{-1}). \tag{3.132}$$

Hence, we have shown that if (3.132) hold, then the constructed net satisfies the first condition.

Now we verify the second condition. By Lemma 29, the entropy of the set is bounded by

$$\mathcal{N} \le \left(\frac{\log(1/\pi_n) \log(k_n)}{\tau_n c_n} \right)^{k_n} K_z / d_n$$

which implies

$$\log \mathcal{N} \lesssim k_n (\log n)^2 = O((\log n)^{2-D_c} \bar{z}_n / \delta)$$

taking $c_n = \delta \bar{z}_n^{-1} (\log n)^{-D_c}$ for some $D_c > 0$. Denote $r = 2 - D_c$.

Now we verify the condition

$$\frac{\max_i (\|g_L^i\|_\infty \vee \|g_U^i\|_\infty)}{\sqrt{n}} \log(1 + N_{[]}) + \max_i (\|g_L^i\|_2 \vee \|g_U^i\|_2) \sqrt{\log(1 + N_{[]})} = o(\sqrt{n}).$$

First note that by Lemma 26,

$$\left\| \frac{f'_{P_n;z}(\cdot)}{f_{P_n;z}(\cdot)} \right\|_\infty \lesssim \frac{z}{\delta} + o(1).$$

Note also that

$$\begin{aligned}
\int_0^{y_n} \left(\frac{f'_{P_n;z}(y)}{f_{P_n;z}(y)} \right)^2 f_0(y) dy &\le \int_0^{y_n} \frac{(f'_{P_n;z}(y))^2}{f_{P_n;z}(y)^2} (f_0(y) - f_{P_n;z}(y)) dy \\
&\quad + \left\| \frac{f'_{P_n;z}(y)}{f_{P_n;z}(y)} \right\|_\infty (f_0(0) + o(1)) \\
&\le \left\| \frac{f'_{P_n;z}(y)}{f_{P_n;z}(y)} \right\|_\infty (f_0(0) + o(1)) + \left\| \frac{f'_{P_n;z}(y)}{f_{P_n;z}(y)} \right\|_\infty^2 d_H^2(f_0, f_{P_n;z}) \\
&\quad + 2 \left\| \frac{f'_{P_n;z}(y)}{f_{P_n;z}} \right\|_\infty^{3/2} [f_0(0) + o(1)]^{1/2} d_H(f_0, f_{P_n;z}) \\
&\lesssim \frac{z}{\delta} \left(1 + \frac{z}{\delta} \epsilon_n^2 + \frac{z^{1/2}}{\delta^{1/2}} \epsilon_n \right) = \frac{z}{\delta} (1 + o(1))
\end{aligned}$$

for $\beta > 1$ and $\delta^{-1} \le [\log n]^s$ for some $s \ge 0$. The expression above is bounded from above, up to a factor, by

$$\begin{aligned}
& \frac{z_n \delta^{-1}}{\sqrt{n}} (\log n)^r \bar{z}_n / \delta + \sqrt{z_n \delta^{-1} (\log n)^r \bar{z}_n / \delta} \\
= \;& n^{1/2} \left[n^{3-2\beta/(2\beta+1)} (\log n)^r \delta^{-2} + n^{0.5(3-2\beta)/(2\beta+1)} (\log n)^{r/2} \delta^{-1} \right] \\
= \;& o(n^{1/2})
\end{aligned}$$

for $\beta > 3/2$ and $\delta^{-1} \le [\log n]^s$ for some $s \ge 0$. $\square$

Proof of Lemma 29. Indeed,

$$\Pi(\bar{\mathcal{P}}_n^c) \le \Pi(\mathcal{P}_n^c) + \Pi_z((2, \bar{z}_n)^c) + \Pi(\min(Q_{0,k}, Q_{1,k'}, k \le k_{0,n}, k' \le k_{1,n}) < \pi_n)$$
$$+\Pi(\int_{y_n'}^{b_n} \epsilon^{-1} dQ^{(1)}(\epsilon) < \pi_n).$$

Now we bound each of the terms. We have shown that $\Pi(\mathcal{P}_n^c) \lesssim e^{-Cn\epsilon_n^2}$.

Note that $y_n' \le [\frac{u}{C_2 \bar{z}_n}]^{-1/(\rho_1+2)}$ due to $F_0(y_n') = 1 - u/(3\bar{z}_n)$ and Remark 5. Under the assumptions of the lemma,

$$k_{0,n} = \log(\delta/a_n)/\log(1+c_n) \lesssim n^{B_c} \log n,$$
$$k_{1,n} = \log(2y_n'/\delta)/\log(1+c_n) \lesssim n^{B_c} \log n$$

and both are bounded by n^B for B large enough.

By assumption (3.12) on prior for z,

$$\Pi_z(\bar{z}_n, \infty) \lesssim e^{-c'\sqrt{\bar{z}_n}(\log \bar{z}_n)^{\rho_z}} = e^{-c'' n^{1/(2\beta+1)}(\log n)^{\rho_z + t/2}} \lesssim e^{-C_2 n\epsilon_n^2}$$

as $2q = t/2 + \max[\rho_z, 5/2] \ge \rho_z + t/2$ by equation (3.108) where $\bar{z}_n = Cn^{2/(2\beta+1)}(\log n)^t$ and $\epsilon_n = n^{-\beta/(1+2\beta)}[\log n]^q$.

Additionally,

$$\begin{aligned}
\Pi(\min(Q_{0,k}, Q_{1,k'},&k \le k_{0,n}, k' \le k_{1,n}) < \pi_n) \\
&= 1 - \prod_{k=1}^{k_{0,n}} (1 - \Pi(Q_{0,k} < \pi_n)) \prod_{k'=1}^{k_{1,n}} (1 - \Pi(Q_{1,k'} < \pi_n)) \\
&\le \sum_{k=1}^{k_{0,n}} \Pi(Q_{0,k} < \pi_n) + \sum_{k'=1}^{k_{1,n}} \Pi(Q_{1,k'} < \pi_n) \\
&\le \sum_{k=1}^{k_{0,n}} \Pi(Q^{(0)}(\epsilon_{n,k}, \epsilon_{n,k+1}) < \pi_n) + \sum_{k'=1}^{k_{1,n}} \Pi(Q^{(1)}(\epsilon'_{n,k'}, \epsilon'_{n,k'+1}) < \pi_n)
\end{aligned}$$

and $Q^{(0)}(\epsilon_{n,k}, \epsilon_{n,k+1}) \sim Beta(\alpha_k, M - \alpha_k)$, with $\alpha_k := MG_0(\epsilon_{n,k}, \epsilon_{n,k+1})$. Then

$$\Pi(Q^{(0)}(\epsilon_{n,k}, \epsilon_{n,k+1}) < \pi_n) = \frac{\Gamma(M)}{\Gamma(\alpha_k)\Gamma(M - \alpha_k)} \int_0^{\pi_n} x^{\alpha_k - 1}(1-x)^{M-\alpha_k-1} dx$$

If $M - \alpha_k \le 1$ then $(1-x)^{M-\alpha_k-1}$ is increasing in x and $(1-x)^{M-\alpha_k-1} \le (1-\pi_n)^{M-\alpha_k-1} \le 2$ for all $x \le \pi_n$ as long as $\pi_n \le 1 - (1/2)^{1/(1-M+\alpha_k)}$. Otherwise, if $M - \alpha_k > 1$ then $(1-x)^{M-\alpha_k-1} \le 1$ for all $x \in [0,1]$. Additionally, $\Gamma(x) \ge 1/2$ for all $x > 0$. Thus,

$$\begin{aligned}
\frac{\Gamma(M)}{\Gamma(\alpha_k)\Gamma(M-\alpha_k)} \int_0^{\pi_n} x^{\alpha_k-1}(1-x)^{M-\alpha_k-1} dx &\le \frac{2\Gamma(M)}{\Gamma(\alpha_k)\Gamma(M-\alpha_k)} \alpha_k^{-1} \pi_n^{\alpha_k} \\
&\le \frac{4\Gamma(M)}{\Gamma(\alpha_k + 1)} \pi_n^{\alpha_k} \\
&\lesssim e^{-n^B \alpha_k}.
\end{aligned}$$

So we need $\alpha_k \gtrsim n^{-B}(n\epsilon_n^2 + \log k_{0,n})$, with $\epsilon_n = n^{-\beta/(2\beta+1)}(\log n)^q$. Similarly for $Q^{(1)}$,

$$\Pi(Q^{(1)}(\epsilon'_{n,k'}, \epsilon'_{n,k'+1}) < \pi_n) \lesssim \pi_n^{\alpha'_{k'}} \le e^{-n^B \alpha'_{k'}}.$$

Recall that $\epsilon_{n,k} = a_n(1+c_n)^k$. If $\epsilon_{n,k}$ is small then

$$\begin{aligned}\alpha_k = \int_{\epsilon_{n,k}}^{\epsilon_{n,k+1}} g_0(x)dx \gtrsim \epsilon_{n,k+1}^{a_0+1} - \epsilon_{n,k}^{a_0+1} =& a_n^{a_0+1}(1+c_n)^{(a_0+1)k}[(1+c_n)^{a_0+1} - 1] \\ \ge& a_n^{a_0+1} c_n \gtrsim n^{-B_c - B_a(a_0+1)}.\end{aligned}$$

For any other $\epsilon_{n,k} \le \delta$

$$\begin{aligned}\alpha_k = \int_{\epsilon_{n,k}}^{\epsilon_{n,k+1}} g_0(x)dx \gtrsim \epsilon_{n,k+1} - \epsilon_{n,k} =& a_n(1+c_n)^k c_n \\ \ge& a_n c_n \gtrsim n^{-B_c - B_a}\end{aligned}$$

and obtain the same type of bound.

Therefore,

$$\begin{aligned}\Pi(\min(Q_{0,k},\ k \le k_{0,n}) < \pi_n) \le \pi_n^{\sum_{k=1}^{k_{0,n}} \alpha_k} \\ \le \exp[-n^B[k_{0,n} \max(n^{-a_c - a_1/[(\rho_1+2)(2\beta+1)]}, n^{-B_\delta - B_c})]] \le \exp\{-Cn\epsilon_n^2\}\end{aligned}$$

for B large enough.

For $\epsilon_{n,k'}$ large

$$\begin{aligned}\alpha'_{k'} = \int_{\epsilon_{n,k'}}^{\epsilon_{n,k'+1}} g_0(x)dx \gtrsim \int_{\epsilon_{n,k'}}^{\epsilon_{n,k'+1}} x^{-a_1}dx \gtrsim& (\epsilon_{n,k'+1} - \epsilon_{n,k'})\epsilon_{n,k'+1}^{-a_1} \\ =& \epsilon_{n,k'} c_n \epsilon_{n,k'+1}^{-a_1} \\ \ge& \delta[y'_n]^{-a_1} c_n \ge C\delta[\delta/\bar{z}_n]^{a_1/(\rho_1+2)} c_n \\ \ge& C n^{-a_c - 2a_1/[(\rho_1+2)(2\beta+1)]}[\log n]^{-qa_1/(\rho_1+2)} \\ &\gtrsim n^{-a_c - a_1/[(\rho_1+2)(2\beta+1)]}\end{aligned}$$

for n large enough, using $y'_n \le [\frac{u\delta}{C_2 \bar{z}_n}]^{-1/(\rho_1+2)}$ due to Remark 5.

For any other $\epsilon_{n,k'} \ge \delta$,

$$\begin{aligned}\alpha'_{k'} = \int_{\epsilon_{n,k'}}^{\epsilon_{n,k'+1}} g_0(x)dx \gtrsim \epsilon_{n,k'+1} - \epsilon_{n,k'} =& \delta(1+c_n)^{k'} c_n \\ \ge& \delta c_n \gtrsim n^{-B_\delta - B_c}.\end{aligned}$$

Therefore,

$$\Pi(\min(Q_{1,k'},\ k' \le k_{1,n}) < \pi_n) \le \pi_n^{\sum_{k'=1}^{k_{1,n}} \alpha'_{k'}} \le \exp[-n^B[n^{B_\delta + B_c(a_0+1)} k_{1,n}]] \le \exp\{-Cn\epsilon_n^2\}$$

for B large enough. Now we bound $\Pi(\int_{y'_n}^{b_n} \epsilon^{-1} dQ^{(1)}(\epsilon) \le \pi_n)$. The probability of this event is bounded by the probability of $Q^{(1)}(y'_n, b_n) \le b_n \pi_n$, since $b_n^{-1} Q^{(1)}(y'_n, b_n) \le \int_{y'_n}^{b_n} \epsilon^{-1} dQ^{(1)}(\epsilon) \le \pi_n$,

then

$$\Pi(Q^{(1)}(y_n', b_n) < b_n\pi_n) \lesssim (b_n\pi_n)^{\alpha_n'} \leq (n^{B_1} y_n e^{-n^B})^{\alpha_n'} \leq (C\epsilon^{-1/(\rho_1+2)} n^{B_1+1/(\rho_1+2)} e^{-n^B})^{\alpha_n'}$$
$$\leq e^{-\alpha_n' n^{B/2}}$$

for large enough n, with $\alpha_n' := MG_0(y_n', b_n)$ and $y_n \leq [\frac{\epsilon}{3C_2 n}]^{-1/(\rho_1+2)}$ due to $F_0(y_n) \geq 1-\epsilon/(3n)$ and Remark 5.

Since $a_1 > 1$,

$$\begin{aligned}\alpha_n' &= \int_{y_n'}^{b_n} g_0(x)dx \gtrsim \int_{y_n'}^{b_n} x^{-a_1} dx = (a_1-1)^{-1}[y_n'^{\,-a_1+1} - b_n^{-a_1+1}] \\ &\gtrsim [n^{-2(a_1-1)/(2\beta+1)}[\log n]^{-q(a_1-1)} - [n^{-(a_1-1)(B_1+1/(2+\rho_1))}\epsilon^{(a_1-1)/(2+\rho_1)}]] \\ &\gtrsim n^{-(a_1-1)/(2\beta+1)}\end{aligned}$$

for B' large enough and $\epsilon \geq n^{-B_\epsilon}$ for some $B_\epsilon > 0$. Hence,

$$\begin{aligned}\Pi(\int_{y_n'}^{b_n} \epsilon^{-1} dQ^{(1)}(\epsilon) \leq \pi_n) &\leq \Pi(Q^{(1)}(y_n', b_n) < b_n\pi_n) \lesssim e^{-\alpha_n' n^{B/2}} \lesssim e^{-n^{-(a_1-1)/(2\beta+1)} n^{B/2}} \\ &\leq \exp\{-Cn\epsilon_n^2\}\end{aligned}$$

for B large enough. This proves the first part of the lemma.

The entropy of the set is bounded by

$$\mathcal{N} \leq (K_\pi k_n)^{k_n} K_z/d_n \lesssim \left(\frac{\log(1/\pi_n)k_n}{\tau_n c_n}\right)^{k_n} \log(\bar{z}_n)/(d_n r_n)$$

with

$$k_n = C[1+\log(y_n'/a_n)]/c_n \lesssim n^{B_c}\log n.$$

Hence,

$$\begin{aligned}\log\mathcal{N} &\leq C + \frac{C\log(y_n'/a_n)}{c_n}(\log\log(1/\pi_n) + \log\log(y_n'/a_n) - \log(c_n\tau_n)) \\ &\quad + \log\log(\bar{z}_n) - \log(d_n r_n) \\ &\lesssim 1 + n^{B_c}\log n\,(\log n + \log\log n + \log(1/\tau_n)) + \log(1/(d_n r_n)).\end{aligned}$$

This completes the proof of Lemma 29. □

3.11 Some technical results from [4]

In this section we include Lemmas from the paper by Bochkina & Rousseau (2017) [4], in which we have based some of our work and have been referenced in this chapter.

Lemma 30 (Lemma 4.1 from [4])**.** *Assume that the probability density $f_0 \in \mathcal{P}_\alpha(\beta, L, \gamma, C_0, C_1, e, \Delta)$ and that there exist $C > 0$ and $\rho_1 > 0$ such that*

$$\int_x^\infty y^2 f_0(y)dy \leq C(1+x)^{-\rho_1}.$$

Then, for any $\epsilon_0 > 0$, there exist $\kappa, C_p > 0$ such that

$$\Pi(KL(f, K_z * P) \leq \epsilon_n^2; V(f, K_z * P) \leq \epsilon_n^2 \log n) \geq C_p e^{-\kappa n^{1/(2\beta+1)} (\log n)^{2q}},$$

for any prior satisfying condition $(\mathcal{P})$ and $n \geq 1$ where $\epsilon_n = \epsilon_0 n^{-\beta/(2\beta+1)} (\log n)^q$, with $q = (5\beta+1)/(4\beta+2)$ if $\rho_z \leq 5/2$ and $q = (2\rho_z\beta+1)/(4\beta+2)$ if $\rho_z > 5/2$, where ρ_z is defined in condition $(\mathcal{P})$. The constants κ and C_p depend on Π, ϵ_0 and on the constants defining the functional class.

Note that condition $(\mathcal{P})$ in [4] corresponds to the same condition (3.12) we have on the prior for z and a Dirichlet Process prior for P with a base measure having polynomial tails, analogous to our conditions (3.10) and (3.11).

Lemma 31 (Lemma 4.2 from [4])**.** *Fix $\varepsilon > 0$, $J \in \mathbb{N}$, $a, b > 0$, $0 < \underline{z} < \bar{z} < \infty$ and introduce the following class of densities:*

$$Q = Q(\varepsilon, J, a, b, \underline{z}, \bar{z}) = \left\{ \begin{array}{c} f = \sum_{j=1}^{\infty} \pi_j g_{z,\epsilon_j} : \sum_{j>J} \pi_j < \epsilon, \ z \in [\underline{z}, \bar{z}], \\ \epsilon_j \in [a, a+b] \text{ for } j = 1, \dots, J \end{array} \right\}$$

Then, for $\varepsilon \leq \sqrt{\bar{z}}$,

$$\begin{aligned} \log N(5\varepsilon, Q, \|\cdot\|_1) \leq & C + J\left[\log\log\left(\frac{b}{a}\right) - 3\log\varepsilon + 0.5\log(\bar{z})\right] + \log\log\left(\frac{\bar{z}}{\underline{z}}\right) \\ \Pi(Q^c) \leq & \left(\frac{em}{J}\log(1/\varepsilon)\right)^J + J(1 - G([a, a+b])) + 1 - \Pi_z([\underline{z}, \bar{z}]) \end{aligned}$$

where Π is a prior satisfying condition $(\mathcal{P})$.

Lemma 32 (Lemma B.1 from [4])**.** *Let $e_z = z^{-a}$, $E_z = z^b$ and H be a probability distribution on $[e_z, E_z]$. Then for all $\kappa > 0$, there exists $N_0 > 0$ and a probability distribution P with at most $\bar{N} = N_0\sqrt{z}(\log z)^{3/2}$ supporting points such that: for all $x \in [\tau_0 e_z, \tau_1 E_z]$ with $0 < \tau_0 < 1 < \tau_1 < +\infty$*

$$|K_z * (H - P)(x)| \leq z^{-\kappa}, \text{ when } z \text{ is large enough.}$$

Lemma 33 (Lemma C.1 from [4])**.** *For all $\delta > 0$, there exists $C > 0$ such that for all ϵ_1, ϵ_2 satisfying $|\epsilon_1/\epsilon_2 - 1| < \delta$*

$$\|g_{z,\epsilon_1} - g_{z,\epsilon_2}\|_1 \leq \sqrt{2KL(g_{z,\epsilon_1}, g_{z,\epsilon_2})} \leq \sqrt{2z}\delta, \quad g_{z,\epsilon_2}(x) \leq g_{z,\epsilon_1}(x) e^{z\delta(1+x/\epsilon_1)}.$$

Lemma 34 (Lemma C.2 from [4])**.** *Let $z > 0$ and $x > 0$, then*

$$I_0(z, x) := \int_0^\infty g_{z,\epsilon}(x) d\epsilon = 1 + \frac{1}{z-1}$$

and for all $k \geq 0$

$$I_k(z, x) := \int_0^\infty (\epsilon - x)^k g_{z,\epsilon}(x) d\epsilon = \frac{x^k z^z}{\Gamma(z)} \int_0^\infty \frac{(u-1)^k e^{-z/u}}{u^z} du$$

Moreover for all $\delta \in (0, 1)$ there exists $c(\delta) > 0$ such that for all z large enough and $u < 1 - \delta$,

$$\frac{z^z e^{-z/u}}{\Gamma(z) u^z} \leq e^{-c(\delta) z/u}$$

and for all $u > 1+\delta$

$$\frac{z^z e^{-z/u}}{\Gamma(z) u^z} \leq u^{-c(\delta) z}.$$

Lemma 35 (Lemma C.3 from [4])**.** *For all $k \geq 0$ and $x > 0$,*

$$I_k(z,x) = \frac{x^k}{z^{k/2}}(1+R(z))^{-1}\left(\mu_k + O(z^{-H})\right) := \frac{x^k}{z^{k/2}}\mu_k(z),\ \forall H > 0,$$

where $\mu_k = \int_{\mathbb{R}} x^k \varphi(x) dx$ with φ the density of a standard Gaussian random variable. We also have

$$K_z f(x) = \sum_{j=0}^{r} \frac{f^{(j)}(x) x^j}{j! z^{j/2}} \mu_j(z) + z^{-\beta/2} R_z(x)$$

where

$$|R_z(x)| \leq C_{\beta,z} L(x) x^{\beta} \left[1 + \frac{x^{\gamma}}{z^{\gamma/2}}\right].$$

For all $g(x) \leq C_1 + C_2 x^a$ for some $a > 0$, then

$$K_z g(x) \leq 2C_1 + 2C_2 x^a,$$

for z large enough and a fixed.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

We begin this section pointing out the main contributions of this work. We have been able to tackle the problem of Semiparametric estimation of densities with unknown support from a Bayesian perspective proving a nonregular version of Bernstein-von Mises Theorem using two prior models. Additionally we implemented an MCMC algorithm for a nonstandard mixture model.

For the sieve prior model we have extended the results of log-spline models to approximate a density defined on a semiline, proving consistency and convergence rate to the truncated density at nonparametric minimax rate. For this purpose we had to prove an alternative lower bound for the denominator in the posterior distribution (also called evidence) which is related to the condition on prior mass of Kullback-Leibler neigbourhoods. Additionally, we extended the limiting log-likelihood ratio results to the case where apart from the nonregular paramter of interest there is a regular nuisance parameter of growing dimension. Finally we proved a nonregular Bernstein-von Mises Theorem for the marginal posterior distribution of the nonregular parameter under a sieve prior and applied it to the log-spline model..

Investigating the second prior model, we were able to find more general Bernstein-von Mises Theorem for LAE models, and from the orignial shift LAE model that motivated our project we found that scale LAE models are equivalent through a log transformation of the data. It is important to note that the results in [32] do not cover our mixture prior model. We also proved posterior concentration for the density in L^1 and local supremum norm near 0. The latter is usually more difficult to obtain, and in this framework seems to be necessary given that the parameter of limiting distribution for θ is precisely the value $f(0)$. Note that in our log-spline prior we also obtain consistency in uniform norm through the approximation properties of B-splines in that norm. It was precisely this type of consistency that motivated the search for a nonstandard prior process for the mixing distribution. Another critical condition required is the uniform control of the interaction term, this was not particularly easy for a flexible prior such as a mixture. Although we used the monotonicity assumption in most of the proofs throughout Chapter 3, it was critical in the proof for uniform consistency motivated by the work of [44]. It is sill not clear whether this can be extended to a more general class of densities. On the other hand, results such as adaptive L^1 consistency can be obtained without monotonicity assumption as shown in [4].

Although the prior process for the mixing distribution is not very standard we were able to implement it through a slice sampling algorithm and applied it on simulated data and real data from procurement auctions proving its applicability and helping to illustrate consistency of f and θ together with the corresponding limiting marginal distribution for θ predicted by theoretical results.

It remains to be seen how general our results are, and find other examples that satisfy the conditions. Finding such a prior model with reasonable assumptions on hyperparameters can be challenging, in particular uniform consistency is not simple to prove in general, although it is only required near the discontinuity point. The usual Dirichlet Process mixture prior seems to be too flexible to obtain it and thus we introduced some modifications to the mixing process and worked only with monotonic nonincreasing densities. It is also important to note that the frequentist counterpart, that is the nonparametric MLE estimator, is not consistent at the discontinuity point. Additionally, another potentially problematic condition is the bound on the interaction term. This implies no loss of information which might not hold in some contexts or be difficult to prove.

We finish this section specifying the reasons for choosing the two particular prior models to which we applied our theorems from Chapters 2 and 3. A B-spline basis for the sieve prior seemed to be a natural choice given their excellent approximation properties, which are key to obtain their known results on minimax contraction rates for densities supported on finite intervals (see Sections 9.1 and E.2 in [23] and the references therein). Additionally, it is easy to control their smoothness and although they are not orthogonal, only a small number of elements of the basis affects the approximation at each point, which simplifies many calculations. Certainly, we could study and apply our theorem to other bases with similar properties, such as Bernstein polynomials and wavelets. However, Bernstein polynomials present suboptimal approximation results leading to suboptimal nonparametric contraction rates and wavelets with arbitrary smoothness level are not easy to construct and many times they do not have closed form expressions. Similarly, we used a gamma density function as the kernel for the mixture prior because of the positive results from [4] on adaptive density estimation on a semiline with minimax contraction rates. There were several ideas from this work that we could borrow. In this case, Gaussian approximation Lemmas 33, 34 and 35 are important to obtain good approximation properties and therefore, other kernels defined on semilines with similar properties are good candidates to obtain the same results as with the Gamma.

4.2 Future Work

There are various ways to keep investigating nonregular models and particularly LAE models from a Bayesian perspective. It would be interesting to keep refining the general sufficient conditions for a BvM Theorem, and also find necessary conditions. It is also of interest to understand whether the prior models presented here are two of a few ad-hoc models that allow us to obtain a BvM Theorem, or if there are many more that can satisfy our general conditions.

More specific steps regarding the two prior studied here are to investigate further adaptation for the log-spline model and whether we can remove the assumption of monotonicity in the mixture prior. Additionally, we could investigate other bases for the sieve prior and other kernels for the mixture prior.

Once we obtain these results, it would be interesting to study other types of non-regularities such as densities with singularities, or multiple jumps, for instance when the support of the density is a finite interval. Also efficiency of estimators in these frameworks can be studied.

Further extensions can be explored such as considering the case when the parameter of interest θ_0 may also be not finite and thus the unknown density could be supported on the whole real line. We could investigate a prior for θ that incorporates this option, e.g. a mixture that includes a point mass at $-\infty$. Additionally, we can investigate inference on multidimensional nonregular parameters, and in particular to study the support of the density of multidimensional data. Finally, we have covered only density estimation so far and therefore we can continue studying nonregular parameters in regression and time series settings.

Bibliography

[1] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Hoboken, NJ, 1 edition edition, Oct. 2004.

[2] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York, 1998 edition edition, May 1998.

[3] P. J. Bickel and B. J. K. Kleijn. The semiparametric Bernstein–von Mises theorem. *The Annals of Statistics*, 40(1):206–237, Feb. 2012.

[4] N. Bochkina and J. Rousseau. Adaptive density estimation based on a mixture of Gammas. *Electronic Journal of Statistics*, 11(1):916–962, 2017.

[5] N. A. Bochkina and P. J. Green. The Bernstein–von Mises theorem and nonregular models. *The Annals of Statistics*, 42(5):1850–1878, Oct. 2014.

[6] D. Bontemps. Bernstein von Mises Theorems for Gaussian Regression with increasing number of regressors. *The Annals of Statistics*, 39(5):2557–2584, Oct. 2011.

[7] C. d. Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences. Springer-Verlag, New York, 1978.

[8] I. Castillo. A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 152(1-2):53–99, Feb. 2012.

[9] I. Castillo and J. Rousseau. A Bernstein-von Mises theorem for smooth functionals in semiparametric models. *arXiv:1305.4482 [math, stat]*, May 2013.

[10] H. Chernoff and H. Rubin. The Estimation of the Location of a Discontinuity in Density. The Regents of the University of California, 1956.

[11] V. Chernozhukov and H. Hong. Likelihood Estimation and Inference in a Class of Nonregular Econometric Models. *Econometrica*, 72(5):1445–1480, Sept. 2004.

[12] N. Choudhuri, S. Ghosal, and A. Roy. Bayesian Methods for Function Estimation. In D. K. D. a. C. R. Rao, editor, *Handbook of Statistics*, volume 25 of *Bayesian ThinkingModeling and Computation*, pages 373–414. Elsevier, 2005.

[13] C. K. Chu and P. E. Cheng. Estimation of jump points and jump values of a density function. *Statistica Sinica*, 6(1), 1996.

[14] S. G. Donald and H. J. Paarsch. Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions. *International Economic Review*, 34(1):121–148, Feb. 1993.

[15] S. G. Donald and H. J. Paarsch. Identification, Estimation, and Testing in Parametric Empirical Models of Auctions within the Independent Private Values Paradigm. *Econometric Theory*, 12(3):517–567, Aug. 1996.

[16] S. G. Donald and H. J. Paarsch. Superconsistent estimation and inference in structural econometric models using extreme order statistics. *Journal of Econometrics*, 109(2):305–340, Aug. 2002.

[17] N. Etemadi. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(1):119–122, Feb. 1981.

[18] G. Gayraud. Minimax Estimation of a Discontinuity for the Density. *Journal of Nonparametric Statistics*, 14(1-2):59–66, Jan. 2002.

[19] S. Ghosal. Asymptotic Normality of Posterior Distributions for Exponential Families when the Number of Parameters Tends to Infinity. *Journal of Multivariate Analysis*, 74(1):49–68, July 2000.

[20] S. Ghosal, J. K. Ghosh, and T. Samanta. On convergence of posterior distributions. *The Annals of Statistics*, 23(6):2145–2152, Dec. 1995.

[21] S. Ghosal, J. K. Ghosh, and A. W. v. d. Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, Apr. 2000.

[22] S. Ghosal and T. Samanta. Asymptotic behaviour of bayes estimates and posterior distributions in multiparameter non regular cases. *Mathematical Methods of Statistics*, 4:P361–388, 1995.

[23] S. Ghosal and A. v. d. Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, June 2017.

[24] S. Ghosal and A. W. v. d. Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *The Annals of Statistics*, 29(5):1233–1263, 2001.

[25] J. K. Ghosh, S. Ghosal, and T. Samanta. Stability and Convergence of the Posterior in Non-Regular Problems. In S. S. Gupta and J. O. Berger, editors, *Statistical Decision Theory and Related Topics V*, pages 183–199. Springer New York, 1994.

[26] M. N. Goria. Estimation of location of discontinuity in a density. *Rivista di matematica per le scienze economiche e sociali*, 5(2):123–141, Sept. 1982.

[27] J. E. Griffin and S. G. Walker. Posterior Simulation of Normalized Random Measure Mixtures. *Journal of Computational and Graphical Statistics*, 20(1):241–259, Jan. 2011.

[28] K. Hirano and J. R. Porter. Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support. *Econometrica*, 71(5):1307–1338, Sept. 2003.

[29] J. Hájek. Local asymptotic minimax and admissibility in estimation. The Regents of the University of California, 1972.

[30] I. A. Ibragimov and R. Z. Has minskii. *Statistical estimation–asymptotic theory*. Springer-Verlag, New York, 1981.

[31] M. Kalli, J. E. Griffin, and S. G. Walker. Slice sampling mixture models. *Statistics and Computing*, 21(1):93–105, Jan. 2011.

[32] B. Kleijn and B. Knapik. Semiparametric posterior limits under local asymptotic exponentiality. *arXiv:1210.6204 [math, stat]*, pages 1–31, Dec. 2013.

[33] E. Krasnokutskaya and K. Seim. Bid Preference Programs and Participation in Highway Procurement Auctions. *The American Economic Review*, 101(6):2653–2686, 2011.

[34] L. Le Cam and G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Science & Business Media, 2nd edition edition, July 2000.

[35] M. Panov and V. Spokoiny. Finite Sample Bernstein – von Mises Theorem for Semiparametric Problems. *Bayesian Analysis*, 10(3):665–710, Sept. 2015.

[36] G. C. Pflug. The Limiting Log-Likelihood Process for Discontinuous Multiparameter Density Families. In W. Grossmann, G. C. Pflug, and W. Wertz, editors, *Probability and Statistical Inference*, pages 287–295. Springer Netherlands, 1982.

[37] G. C. Pflug. The limiting log-likelihood process for discontinuous density families. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 64(1):15–35, Mar. 1983.

[38] T. Polfeldt. Minimum Variance Order when Estimating the Location of an Irregularity in the Density. *The Annals of Mathematical Statistics*, 41(2):673–679, Apr. 1970.

[39] T. Polfeldt. The Order of the Minimum Variance in a Non-Regular Case. *The Annals of Mathematical Statistics*, 41(2):667–672, Apr. 1970.

[40] B. L. S. Prakasa Rao. Estimation of the Location of the Cusp of a Continuous Density. *The Annals of Mathematical Statistics*, 39(1):76–87, Feb. 1968.

[41] M. Reiss and J. Schmidt-Hieber. *Nonparametric Bayesian analysis of the compound Poisson prior for support boundary recovery*. Sept. 2018.

[42] V. Rivoirard and J. Rousseau. Bernstein–von Mises theorem for linear functionals of the density. *The Annals of Statistics*, 40(3):1489–1523, June 2012.

[43] H. Rubin. The Estimation of Discontinuities in Multivariate Densities, and Related Problems in Stochastic Processes. The Regents of the University of California, 1961.

[44] J.-B. Salomond. Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electronic Journal of Statistics*, 8(1):1380–1404, 2014.

[45] W. Shen and S. Ghosal. Adaptive Bayesian Procedures Using Random Series Priors. *Scandinavian Journal of Statistics*, 42(4):1194–1213, 2015.

[46] W. Shen and S. Ghosal. Posterior Contraction Rates of Density Derivative Estimation. *Sankhya A*, 79(2):336–354, Aug. 2017.

[47] W. Shen, S. T. Tokdar, and S. Ghosal. Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika*, 100(3):623–640, Sept. 2013.

[48] R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, Apr. 1985.

[49] A. W. van der Vaart. *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge Univ. Press, Cambridge, 1. paperback ed., 8. printing edition, 2000.

[50] R. E. Williamson. Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal*, 23(2):189–207, June 1956.

[51] W. H. Wong and X. Shen. Probability Inequalities for Likelihood Ratios and Convergence Rates of Sieve MLES. *The Annals of Statistics*, 23(2):339–362, Apr. 1995.