Supplementary material for "Semiparametric Bayesian inference for local extrema of functions in the presence of noise"

In this supplementary material, we present proofs of all results in the main paper, additional technical lemmas, and additional numerical experiments.

A Proofs

A.1 Proof of Proposition 1

By Bayes' theorem, it suffices to show that the likelihood takes the form of Equation (5). Recall that $\boldsymbol{y}|X,t \sim N(0,\Sigma_t)$, where $\Sigma_t = \sigma^2(n\lambda)^{-1}(A+B)$ with $A = K(X,X) + n\lambda \boldsymbol{I}_n$ and $B = -K_{01}(X,t)K_{11}^{-1}(t,t)K_{10}(t,X) = -\boldsymbol{a}\boldsymbol{a}^T$ by letting $\boldsymbol{a} = K_{01}(X,t)K_{11}^{-1/2}(t,t)$. Note that the condition $\widehat{\sigma}^2_{f'}(t) > 0$ for any t ensures $K_{11}(t,t) > 0$ in view of Equation (4).

In view of the Sherman–Morrison formula, we have $\det(A+B) = (1-\boldsymbol{a}^TA^{-1}\boldsymbol{a})\det(A)$ and $(A+B)^{-1} = A^{-1} + \frac{A^{-1}\boldsymbol{a}\boldsymbol{a}^TA^{-1}}{1-\boldsymbol{a}^TA^{-1}\boldsymbol{a}}$, assuming $1-\boldsymbol{a}^TA^{-1}\boldsymbol{a} \neq 0$. Substituting these two identities into the multivariate normal density $\ell(t)$ yields

$$\begin{aligned}
\ell(t) &= \{2\pi\sigma^2(n\lambda)^{-1}\}^{-n/2}\det(A+B)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2(n\lambda)^{-1}}\boldsymbol{y}^T(A+B)^{-1}\boldsymbol{y}\right\} \\
&= \{2\pi\sigma^2(n\lambda)^{-1}\}^{-n/2}(\det(A))^{-1/2}\exp\left\{-\frac{\boldsymbol{y}^TA^{-1}\boldsymbol{y}}{2\sigma^2(n\lambda)^{-1}}\right\} \\
&\quad \cdot (1-\boldsymbol{a}^TA^{-1}\boldsymbol{a})^{-1/2}\exp\left\{-\frac{\boldsymbol{y}^TA^{-1}\boldsymbol{a}\boldsymbol{a}^TA^{-1}\boldsymbol{y}}{2\sigma^2(n\lambda)^{-1}(1-\boldsymbol{a}^TA^{-1}\boldsymbol{a})}\right\} \\
&= C\{\sigma^2(n\lambda)^{-1}(1-\boldsymbol{a}^TA^{-1}\boldsymbol{a})\}^{-1/2}\exp\left\{-\frac{\boldsymbol{y}^TA^{-1}\boldsymbol{a}K_{11}(t,t)\boldsymbol{a}^TA^{-1}\boldsymbol{y}}{2\sigma^2(n\lambda)^{-1}K_{11}(t,t)(1-\boldsymbol{a}^TA^{-1}\boldsymbol{a})}\right\},
\end{aligned}$$

where

$$C = \{2\pi\sigma^2(n\lambda)^{-1}\}^{-n/2}(\det(A))^{-1/2}\exp\left\{-\frac{\boldsymbol{y}^TA^{-1}\boldsymbol{y}}{2\sigma^2(n\lambda)^{-1}}\right\}\cdot\{\sigma^2(n\lambda)^{-1}\}^{1/2}$$

does not depend on t. The proof is completed by noticing that $\widehat{\mu}_{f'}(t) = K_{11}^{1/2}(t,t)\boldsymbol{a}^TA^{-1}\boldsymbol{y}$ and $\widehat{\sigma}^2_{f'}(t) = \sigma^2(n\lambda)^{-1}K_{11}(t,t)(1-\boldsymbol{a}^TA^{-1}\boldsymbol{a})$. This completes the proof.

A.2 Proof of Lemma 1

A one-dimensional version of Theorem 4 in Liu and Li (2023) shows that

$$\begin{aligned}\|\widehat{\mu}_{f'}^{(k)} - f_{\lambda}^{(k+1)}\|_{\infty} \leq & \frac{\sqrt{\kappa\kappa_{k+1,k+1}}\|f_0\|_{\infty}\sqrt{\log(9/\delta)}}{\sqrt{n}\lambda}\left(10 + \frac{4\kappa\sqrt{\log(9/\delta)}}{3\sqrt{n}\lambda}\right) \\ & + \frac{C_2\sqrt{\kappa\kappa_{k+1,k+1}}\sigma\sqrt{\log(3/\delta)}}{\sqrt{n}\lambda}, \quad 0 \leq k \leq 3,\end{aligned} \tag{S1}$$

For any bounded $f \in L^2_{p_X}(\mathcal{X})$, we define a bias of estimators of f by matrix and integral operation as

$$\begin{aligned}E(K, X, f) &= (L_{K,X} + \lambda I)^{-1} L_{K,X} f - (L_K + \lambda I)^{-1} L_K f \\ &= K(\cdot, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-1} f(X) - (L_K + \lambda I)^{-1} L_K f,\end{aligned}$$

which belongs to $\mathbb{H}$. Consider any $j, l \geq 1$ and $j + l \leq 5$, taking $f = K_{0l,x}$ yields

$$\partial^j E(K, X, K_{0l,x}) = K_{j0}(\cdot, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-1} K_{0l}(X, x) - \partial^j (L_K + \lambda I)^{-1} L_K K_{0l,x}.$$

Thus,

$$\begin{aligned}\partial^j E(K, X, K_{0l,x})(x) &= K_{j0}(x, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-1} K_{0l}(X, x) - \partial^j (L_K + \lambda I)^{-1} L_K K_{0l,x}(x) \\ &= K_{j0}(x, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-1} K_{0l}(X, x) - (L_K + \lambda I)^{-1} L_K K_{jl,x}(x)\end{aligned}$$

We write $(L_K + \lambda I)^{-1} L_K K_{jl,x}(x) = K_{jl,x}(x) - \lambda(L_K + \lambda I)^{-1} K_{jl,x}(x) = K_{jl}(x, x) - \lambda\varphi_{jl}(x)$. Then, by Theorem 16 in Liu and Li (2023) we have that for any $\delta \in (0, 1)$, with $\mathbb{P}_0$-probability at least $1-\delta$ it holds

$$\begin{aligned}
&|K_{j0}(x,X)[K(X,X)+n\lambda\boldsymbol{I}_n]^{-1}K_{0l}(X,x)-K_{jl}(x,x)+\lambda\varphi_{jl}(x)| \\
&\leq \|\partial^j E(K,X,K_{0l,x})\|_\infty \\
&\leq \frac{\sqrt{\kappa\kappa_{jj}}\|K_{0l,x}\|_\infty\sqrt{\log(3/\delta)}}{\sqrt{n}\lambda}\left(10+\frac{4\sqrt{\kappa}\sqrt{\log(3/\delta)}}{3\sqrt{n\lambda}}\right) \\
&= \frac{\sqrt{\kappa\kappa_{jj}}\kappa_{0l}\sqrt{\log(3/\delta)}}{\sqrt{n}\lambda}\left(10+\frac{4\sqrt{\kappa}\sqrt{\log(3/\delta)}}{3\sqrt{n\lambda}}\right).
\end{aligned}$$

In view of Equation (7), $\widehat{\sigma}^{2(k)}_{f'}(x)$ is a linear combination of quadratic forms

$$K_{jl}(x,x)-K_{j0}(\cdot,X)[K(X,X)+n\lambda\boldsymbol{I}_n]^{-1}K_{0l}(X,x).$$

Therefore, for any $0\leq k\leq 3$, we have

$$\begin{aligned}
&|\widehat{\sigma}^{2(k)}_{f'}(x)-\sigma^2 n^{-1}\sum_{i=0}^{k}\binom{k}{i}\varphi_{i+1,k+1-i}(x)| \\
&\leq \sum_{i=0}^{k}\binom{k}{i}\left[\frac{\sqrt{\kappa\kappa_{i+1,i+1}}\kappa_{0,k+1-i}\sigma^2\sqrt{\log(3/\delta)}}{n\sqrt{n}\lambda^2}\left(10+\frac{4\sqrt{\kappa}\sqrt{\log(3/\delta)}}{3\sqrt{n\lambda}}\right)\right].
\end{aligned} \tag{S2}$$

The above Equations (S1) and (S2) can hold simultaneously with $\mathbb{P}_0$-probability $1-8\delta$. Let $8\delta=n^{-10}$, and A_n be the corresponding event. We immediately have $\mathbb{P}_0(A_n)\geq 1-n^{-10}$ with $\log(3/\delta)\leq 10\log n+4$ and $\log(9/\delta)\leq 10\log n+5$ in the upper bound. This completes the proof.

A.3 Proof of Lemma 2

First we prove that for any local extremum t_m of f_0, there exists a local extremum $t_{\lambda,m}$ of f_λ such that $t_{\lambda,m}\to t_m$ as $\lambda\to 0$. There exists $\delta>0$ such that for any $0<\epsilon<\delta$, it holds that $f_0'(t_m-\epsilon)<0$, $f_0'(t_m+\epsilon)>0$ and $f_0''(t_m\pm\epsilon)\neq 0$ without loss of generality. By Assumption C, we have

$$|f_\lambda'(t_m-\epsilon)-f_0'(t_m-\epsilon)|\lesssim\lambda^{r_1}.$$

Hence, for sufficiently small λ, it holds $f_\lambda'(t_m - \delta/2) < 0$. Similarly, we have $f_\lambda'(t_m + \delta/2) > 0$. According to the continuity of f_λ, there exists a $t_{\lambda,m} \in (t_m - \delta/2, t_m + \delta/2)$ such that $f_\lambda'(t_{\lambda,m}) = 0$. It can also be shown that $f_\lambda''(t) \neq 0$ for any $t \in (t_m - \delta/2, t_m + \delta/2)$ and sufficiently small λ, which implies $f_\lambda''(t_{\lambda,m}) \neq 0$. Finally, we have $t_{\lambda,m} \to t_m$ as $\delta \to 0$ and $\lambda \to 0$.

Again by Assumption C we can see that

$$|f_\lambda'(t_{\lambda,m}) - f_0'(t_{\lambda,m})| \lesssim \lambda^{r_1}.$$

Since $f_\lambda'(t_{\lambda,m}) = 0$, in view of the mean value theorem, we have

$$|f_0'(t_m) + f_0''(\xi_1)(t_{\lambda,m} - t_m)| \lesssim \lambda^{r_1},$$

where ξ_1 lies between and $t_{\lambda,m}$ and t_m. Since $\xi_1 \to t_m$ and $f_0''(t_m) \neq 0$ by Assumption A, we obtain

$$|t_{\lambda,m} - t_m| \lesssim \lambda^{r_1}.$$

Under A_n, the existence and convergence rate of $\hat{t}_m$ can be shown similarly by applying Equation (8). This completes the proof.

A.4 Proof of Theorem 1

The proof is based on the high probability event A_n defined in Lemma 1. Conditions of Theorem 1 imply $n^{\frac{1}{2}-2\beta}\varphi_{11,m} = o(1)$, yielding $\widehat{\sigma}_{f'}^2(t_{\lambda,m}) > 0$ in view of Equation (11). Invoking the likelihood function $\ell(t)$ in Equation (5), which holds at $t_{\lambda,m}$ and in its small neighborhood, we have

$$\Lambda(t, \Delta t) = \log \frac{\ell(t + \Delta t)}{\ell(t)} = \frac{\ell'(t)}{\ell(t)}\Delta t + \frac{\ell''(t)\ell(t) - \ell'(t)^2}{2\ell(t)^2}(\Delta t)^2 + R_3(\xi)(\Delta t)^3,$$

where $R_3(\xi)=\{2\ell'(t)^3+\ell'''(\xi)\ell(\xi)^2-3\ell'(\xi)\ell''(\xi)\ell(\xi)\}/\{6\ell(\xi)^3\}$ and ξ is bewteen t and $t+\Delta t$. Thus,

$$
\begin{aligned}
\Lambda(t_{\lambda,m},\frac{u}{n^\beta}) = &\left[-\frac{\widehat{\sigma}^{2\prime}_{f'}(t_{\lambda,m})}{2\widehat{\sigma}^2_{f'}(t_{\lambda,m})}-\frac{\widehat{\mu}_{f'}(t_{\lambda,m})\widehat{\mu}'_{f'}(t_{\lambda,m})}{\widehat{\sigma}^2_{f'}(t_{\lambda,m})}+\frac{\widehat{\mu}_{f'}(t_{\lambda,m})^2\widehat{\sigma}^{2\prime}_{f'}(t_{\lambda,m})}{2\widehat{\sigma}^2_{f'}(t_{\lambda,m})^2}\right]\frac{u}{n^\beta}\\
&+\frac{1}{2}\left[\frac{\widehat{\sigma}^{2\prime}_{f'}(t_{\lambda,m})^2}{2\widehat{\sigma}^2_{f'}(t_{\lambda,m})^2}-\frac{\widehat{\sigma}^{2\prime\prime}_{f'}(t_{\lambda,m})}{2\widehat{\sigma}^2_{f'}(t_{\lambda,m})}+\frac{2\widehat{\mu}_{f'}(t_{\lambda,m})\widehat{\mu}'_{f'}(t_{\lambda,m})\widehat{\sigma}^{2\prime}_{f'}(t_{\lambda,m})}{\widehat{\sigma}^2_{f'}(t_{\lambda,m})^2}\right.\\
&\qquad-\frac{\widehat{\mu}'_{f'}(t_{\lambda,m})^2+\widehat{\mu}_{f'}(t_{\lambda,m})\widehat{\mu}''_{f'}(t_{\lambda,m})}{\widehat{\sigma}^2_{f'}(t_{\lambda,m})}\\
&\qquad\left.-\frac{1}{2}\widehat{\mu}_{f'}(t_{\lambda,m})^2\left(\frac{2\widehat{\sigma}^{2\prime}_{f'}(t_{\lambda,m})^2}{\widehat{\sigma}^2_{f'}(t_{\lambda,m})^3}-\frac{\widehat{\sigma}^{2\prime\prime}_{f'}(t_{\lambda,m})}{\widehat{\sigma}^2_{f'}(t_{\lambda,m})^2}\right)\right]\frac{u^2}{n^{2\beta}}\\
&+\frac{1}{6}R_3(\xi)\frac{u^3}{n^{3\beta}}.
\end{aligned}
$$

Based on the rates given by Equations (9), (10), (11) and (12), we obtain

$$
\begin{aligned}
|\widehat{\mu}_{f'}(t_{\lambda,m})|\lesssim n^{-\beta}(\log n)^{-a},\quad |\widehat{\mu}^{(k)}_{f'}(t_{\lambda,m})|\lesssim 1,\\
|\widehat{\sigma}^2_{f'}(t_{\lambda,m})|\lesssim n^{-1}\varphi_{11,m},\quad |\widehat{\sigma}^{2(k)}_{f'}(t_{\lambda,m})|\lesssim n^{-\frac{1}{2}-\beta}(\log n)^{-\frac{1}{2}-a},
\end{aligned}
$$

for $1\le k\le 3$. Further calculation gives $R_3(\xi)\lesssim\frac{\widehat{\sigma}^{2\prime\prime}_{f'}(\xi)}{\widehat{\sigma}^2_{f'}(\xi)^2}=O\left(n^{\frac{3}{2}-\beta}(\log n)^{-\frac{1}{2}-a}\varphi^{-2}_{11,m}\right)$. Substituting these into the above $\Lambda(t_{\lambda,m},\frac{u}{n^\beta})$ yields

$$
\begin{aligned}
\Lambda(t_{\lambda,m},\frac{u}{n^\beta}) &= -\frac{\widehat{\mu}_{f'}(t_{\lambda,m})\widehat{\mu}'_{f'}(t_{\lambda,m})}{n^\beta\widehat{\sigma}^2_{f'}(t_{\lambda,m})}u-\frac{\widehat{\mu}'_{f'}(t_{\lambda,m})^2}{2n^{2\beta}\widehat{\sigma}^2_{f'}(t_{\lambda,m})}u^2+o(n^{\frac{3}{2}-4\beta}\varphi^{-2}_{11,m})\\
&= -\frac{n^{1-\beta}\widehat{\mu}_{f'}(t_{\lambda,m})\widehat{\mu}'_{f'}(t_{\lambda,m})}{n\widehat{\sigma}^2_{f'}(t_{\lambda,m})}u-\frac{n^{1-2\beta}\widehat{\mu}'_{f'}(t_{\lambda,m})^2}{2n\widehat{\sigma}^2_{f'}(t_{\lambda,m})}u^2+o(n^{\frac{3}{2}-4\beta}\varphi^{-2}_{11,m})\\
&= n^{1-2\beta}\varphi^{-1}_{11,m}\left\{-\frac{\widehat{\mu}'_{f'}(t_{\lambda,m})^2u^2}{2n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m})}-\frac{n^\beta\widehat{\mu}_{f'}(t_{\lambda,m})^2u}{n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m})}\right\}+o(1),
\end{aligned}
$$

when $n^{\frac{3}{2}-4\beta}\varphi^{-2}_{11,m}=O(1)$.

Then we study the convergence of $\mu_{n,m}$ and $\sigma^2_{n,m}$. According to Equations (9) and (10), we have

$$
|n^\beta\widehat{\mu}_{f'}(t_{\lambda,m})|\lesssim(\log n)^{-a},
$$

$$|\widehat{\mu}'_{f'}(t_{\lambda,m}) - f''_\lambda(t_{\lambda,m})| \lesssim n^{-\beta}(\log n)^{-a}.$$

In view of Lemma 2, Assumption A2, and Assumption C, we obtain that $\widehat{\mu}'_{f'}(t_{\lambda,m})$ converges to $f''_0(t_m)$, and thus is bounded away from zero and infinity for sufficiently large n. Therefore,

$$|\mu_{n,m}| = \left|\frac{n^\beta \widehat{\mu}_{f'}(t_{\lambda,m})}{\widehat{\mu}'_{f'}(t_{\lambda,m})}\right| \asymp |n^\beta \widehat{\mu}_{f'}(t_{\lambda,m})| \lesssim (\log n)^{-a}.$$

From Equation (11) we have

$$\left| n\varphi_{11,m}^{-1}\widehat{\sigma}^2_{f'}(t_{\lambda,m}) - \sigma^2 \right| \lesssim n^{\frac{1}{2}-2\beta}(\log n)^{-1-a}\varphi_{11,m}^{-1}.$$

Therefore,

$$\begin{aligned}
&\left|\sigma^2_{n,m} - \frac{\sigma^2}{f''_\lambda(t_{\lambda,m})^2}\right| = \left|\frac{n\varphi_{11,m}^{-1}\widehat{\sigma}^2_{f'}(t_{\lambda,m})}{\widehat{\mu}'_{f'}(t_{\lambda,m})^2} - \frac{\sigma^2}{f''_\lambda(t_{\lambda,m})^2}\right| \\
&\asymp \left| f''_\lambda(t_{\lambda,m})^2 n\varphi_{11,m}^{-1}\widehat{\sigma}^2_{f'}(t_{\lambda,m}) - \widehat{\mu}'_{f'}(t_{\lambda,m})^2\sigma^2 \right| \\
&\lesssim \left| f''_\lambda(t_{\lambda,m})^2 \left[n\varphi_{11,m}^{-1}\widehat{\sigma}^2_{f'}(t_{\lambda,m}) - \sigma^2\right] \right| + \left| \left[\widehat{\mu}'_{f'}(t_{\lambda,m})^2 - f''_\lambda(t_{\lambda,m})^2\right]\sigma^2 \right| \\
&\lesssim n^{\frac{1}{2}-2\beta}(\log n)^{-1-a}\varphi_{11,m}^{-1} + n^{-\beta}(\log n)^{-a} \\
&\lesssim n^{\frac{1}{2}-2\beta}(\log n)^{-\frac{3}{2}}\varphi_{11,m}^{-1}.
\end{aligned} \tag{S3}$$

On the other hand,

$$\left|\frac{\sigma^2}{f''_\lambda(t_{\lambda,m})^2} - \frac{\sigma^2}{f''_0(t_m)^2}\right| \asymp |f''_0(t_m)^2 - f''_\lambda(t_{\lambda,m})^2| \lesssim \left[f''_0(t_m) - f''_\lambda(t_{\lambda,m})\right]. \tag{S4}$$

Since $K \in C^8(\mathcal{X}, \mathcal{X})$, we have $f_\lambda \in C^4(\mathcal{X})$. Then the mean value theorem gives

$$f''_0(t_m) - f''_\lambda(t_{\lambda,m}) = f''_0(t_m) - f''_\lambda(t_m) - f'''_\lambda(\xi)(t_{\lambda,m} - t_m).$$

By Assumption C and Lemma 2 we have

$$|f_0''(t_m) - f_\lambda''(t_{\lambda,m})| \lesssim \lambda^{r_2} + \lambda^{r_1} \le 2\lambda^r. \tag{S5}$$

Combining Equations (S3), (S4) and (S5), we obtain

$$\left|\sigma_{n,m}^2 - \frac{\sigma^2}{f_0''(t_m)^2}\right| \lesssim \lambda^r.$$

This completes the proof.

A.5 Proof of Theorem 2

We first present a technical lemma and leave its proof to Section A.10.

Lemma S1. *Suppose Assumption B1 holds and let* $\lambda = n^{-\frac{1}{2}+\beta}(\log n)^{\frac{1}{2}+a}$ *for some* $\frac{1}{4} < \beta < \frac{1}{2}$ *and* $a > 0$. *Under event* A_n, *there exists* $C > 0$ *such that*

$$\left| n^{-\frac{1}{2}} \varphi_{11,m}^{\frac{1}{2}} \ell(t_{\lambda,m}) \Big/ \exp\left(-n^{1-2\beta} \varphi_{11,m}^{-1} \frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right) - \frac{C}{|f_0''(t_m)|\sigma_m^*} \right| \lesssim n^{\frac{1}{2}-2\beta}(\log n)^{-1-a}.$$

For any $x \ge 0$, define the error function as

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

By changing of variable, we have

$$\int_{-A}^{B} a \exp\left(-a^2 \frac{(u+b)^2}{c}\right) du = \sqrt{\frac{\pi c}{2}} \left[\mathrm{Erf}\left(\frac{a(A-b)}{\sqrt{c}}\right) + \mathrm{Erf}\left(\frac{a(B+b)}{\sqrt{c}}\right) \right], \tag{S6}$$

where $a, c, A, B > 0$ and $b \in \mathbb{R}$.

A.5.1 Proof of (i)

The proof will follow three steps.

Step 1: According to Theorem 1.3 in Devroye et al. (2018) and Lemma 2, we have that for any $z \in \mathbb{R}$,

$$\begin{aligned}
&\left| \sum_{m=1}^{M} \pi_m \Phi(z \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2}) - \sum_{m=1}^{M} \pi_m \Phi(z \mid t_m, n^{-1}\varphi_{11,m}\sigma_m^{*2}) \right| \\
&\le d_{TV}\left(\sum_{m=1}^{M} \pi_m \phi(\cdot \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2}),\ \sum_{m=1}^{M} \pi_m \phi(\cdot \mid t_m, n^{-1}\varphi_{11,m}\sigma_m^{*2}) \right) \\
&\lesssim \sum_{m=1}^{M} \pi_m |t_{\lambda,m} - t_m| \lesssim \lambda^{r_1} = o(1),
\end{aligned}$$

where d_{TV} is the total variation distance between two distributions. Thus, we only need to show

$$\left| \Pi_n(t \le z \mid X, \boldsymbol{y}) - \sum_{m=1}^{M} \pi_m \Phi(z \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2}) \right| \to 0$$

for any $z \in \mathbb{R}$ in $\mathbb{P}_0$-probability.

Step 2: We work under the high probability event A_n henceforth in this proof, that is, all convergence rates and bounding integrals only hold under A_n.

Define a sequence of functions

$$\tilde{h}_n(t) = \sum_{m=1}^{M} \ell(t_{\lambda,m}) \tilde{\phi}_{n,m}(t) \pi(t_m), \tag{S7}$$

where

$$\tilde{\phi}_{n,m}(t) = \exp\left(-\frac{(t - t_{\lambda,m})^2}{2n^{-1}\varphi_{11,m}\sigma_m^{*2}} + n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2} \right).$$

In this step, we will prove that

$$\left| \int_{-\infty}^{z} \ell(t)\pi(t)dt - \int_{-\infty}^{z} \tilde{h}_n(t)dt \right| \to 0 \tag{S8}$$

for any $z \in \mathbb{R}$. That is, $\tilde{h}_n(t)$ approximates the unnormalized limit density where each mixture component is properly rescaled. In line with the LAN condition (14), we expand

$\tilde{h}_n(t)$ at $t = t_{\lambda,m} + u/n^\beta$ for $m = 1, \ldots, M$, transforming $\tilde{\phi}_{n,m}(t)$ to

$$\nu_{n,m}(u) = \exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\left(-\frac{u^2}{2\sigma_m^{*2}} + \frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\right).$$

We consider three cases for z: (1) $z \leq 0$, (2) $0 < z \leq 1$, and (3) $z > 1$.

Case (1) ($z \leq 0$). Since $\ell(t) = 0$ in $\mathbb{R}\backslash[0,1]$, the left hand side of Equation (S8) becomes

$$\begin{aligned}\int_{-\infty}^{z} \tilde{h}_n(t)dt &\leq \sum_{m=1}^{M} \int_{-\infty}^{z} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt \\ &= \sum_{m=1}^{M} \int_{L_{n,m}} n^{-\beta}\ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m)du\end{aligned}$$

where we let $t = t_{\lambda,m} + u/n^\beta$ and $L_{n,m} = (-\infty, (z - t_{\lambda,m})n^\beta]$. By Lemma S1 and Equation (S6) , we have

$$\begin{aligned}\int_{L_{n,m}} n^{-\beta}\ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m)du &\lesssim \int_{L_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_m^{*2}}\right) du \\ &= \sqrt{\frac{\pi\sigma_m^{*2}}{2}}\left[1 - \text{ Erf}\left(\frac{\sqrt{n}(t_{\lambda,m} - z)}{\sqrt{2\sigma_m^{*2}\varphi_{11,m}}}\right)\right] \\ &\lesssim \exp\left(-\frac{n(t_{\lambda,m} - z)^2}{2\sigma_m^{*2}\varphi_{11,m}}\right),\end{aligned}$$

where we use the well known inequality that $1 - \text{ Erf}(x) \leq \frac{2}{\sqrt{\pi}}e^{-\frac{x^2}{2}}$. Therefore, there holds that for $z \leq 0$,

$$\int_{-\infty}^{z} \tilde{h}_n(t)dt \lesssim M \exp\left(-\frac{n(t_{\lambda,1} - z)^2}{4\sigma_m^{*2}\varphi_{11,m}}\right) \to 0. \tag{S9}$$

Case (2) ($0 < z \leq 1$). Now the left hand side of Equation (S8) becomes

$$\left|\int_{-\infty}^{z} \ell(t)\pi(t)dt - \int_{-\infty}^{z} \tilde{h}_n(t)dt\right| \leq \int_{-\infty}^{0} \tilde{h}_n(t)dt + \left|\int_{0}^{z} \ell(t)\pi(t)dt - \int_{0}^{z} \tilde{h}_n(t)dt\right|.$$

Taking $z = 0$ in Equation (S9) gives that $\int_{-\infty}^{0} \tilde{h}_n(t)dt \to 0$. We next bound the second term. We divide $[0, 1]$ into M disjoint intervals (up to overlapping endpoints that do not affect estimates of integrals), each of which centering around $t_{\lambda,m}$:

$$[0,1] = \bigcup_{m=1}^{M} I_{n,m}, \quad I_{n,m} = [t_{\lambda,m} - \xi_{n,m-1}, t_{\lambda,m} + \xi_{n,m}], \quad m = 1, \ldots, M,$$

where $\xi_{n,0} = t_{\lambda,1}$, $\xi_{n,m} = (t_{\lambda,m+1} - t_{\lambda,m})/2$ for $m = 1, \ldots, M-1$, and $\xi_{n,M} = 1 - t_{\lambda,M}$. Suppose $z \in I_{n,m_0}$ for some $1 \le m_0 \le M$ and let $I'_{n,m_0} = [t_{\lambda,m_0} - \xi_{n,m_0-1}, z]$. By the triangle inequality, we have

$$\begin{aligned}
\left| \int_0^z \ell(t)\pi(t)dt - \int_0^z \tilde{h}_n(t)dt \right| &= \left| \left(\sum_{m=1}^{m_0-1} \int_{I_{n,m}} + \int_{I'_{n,m_0}} \right) \left[\ell(t)\pi(t) - \tilde{h}_n(t) \right] dt \right| \\
&\le \sum_{m=1}^{m_0-1} \left| \int_{I_{n,m}} \left[\ell(t)\pi(t) - \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m) \right] dt \right| && \text{(S10)} \\
&\quad + \sum_{m=1}^{m_0-1} \int_{[0,z]\setminus I_{n,m}} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt && \text{(S11)} \\
&\quad + \left| \int_{I'_{n,m_0}} \left[\ell(t)\pi(t) - \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m) \right] dt \right| && \text{(S12)} \\
&\quad + \int_{[0,z]\setminus I'_{n,m_0}} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt. && \text{(S13)}
\end{aligned}$$

Again, after changing of variable with $t = t_{\lambda,m} + u/n^{\beta}$, each term in Equation (S10) becomes

$$\begin{aligned}
&\left| \int_{I_{n,m}} \left[\ell(t)\pi(t) - \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m) \right] dt \right| \\
&= \left| \int_{J_{n,m}} \left[\ell(t_{\lambda,m} + u/n^{\beta})\pi(t_{\lambda,m} + u/n^{\beta}) - \ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m) \right] n^{-\beta} du \right| \\
&= \left| \int_{J_{n,m}} n^{-\beta}\ell(t_{\lambda,m}) \left[Z_{n,m}(u)\pi(t_{\lambda,m} + u/n^{\beta}) - \nu_{n,m}(u)\pi(t_m) \right] du \right|,
\end{aligned}$$

where $Z_{n,m}(u) = \ell(t_{\lambda,m} + u/n^{\beta})/\ell(t_{\lambda,m})$ and $J_{n,m} = [-n^{\beta}\xi_{n,m-1}, n^{\beta}\xi_{n,m}]$. Applying the triangle inequality yields an upper bound of the preceding display:

$$\begin{aligned}
&\left|\int_{J_{n,m}} n^{-\beta}\ell(t_{\lambda,m})Z_{n,m}(u)[\pi(t_{\lambda,m}+u/n^{\beta})-\pi(t_m)]du\right| \\
&+\left|\int_{J_{n,m}} n^{-\beta}\ell(t_{\lambda,m})[Z_{n,m}(u)-\nu_{n,m}(u)]\pi(t_m)du\right| \\
=\ & I_1+I_2.
\end{aligned}$$

By Lemma S1 and Theorem 1, we have

$$\begin{aligned}
I_1 &\lesssim \left|\int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}}\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\left(-\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}+\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\right)\right. \\
&\qquad\qquad \left.\cdot\,[\pi(t_{\lambda,m}+u/n^{\beta})-\pi(t_m)]du\right| \\
&\lesssim \left|\int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}}\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}\right)[\pi(t_{\lambda,m}+u/n^{\beta})-\pi(t_m)]du\right| \\
&\lesssim |t_{\lambda,m}-t_m|\int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}}\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}\right)du \\
&\quad +\left|\int_{J_{n,m}} n^{\frac{1}{2}-2\beta}\varphi_{11,m}^{-\frac{1}{2}}u\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}\right)du\right| \\
&= I_{11}+I_{12}.
\end{aligned}$$

In view of Equations (15), (16), (S6), and Lemma 2, it follows that

$$I_{11}\lesssim|t_{\lambda,m}-t_m|\cdot\sqrt{2\pi\sigma_{n,m}^2}\lesssim|t_{\lambda,m}-t_m|\lesssim\lambda^{r_1},$$

and

$$
\begin{aligned}
I_{12} &= n^{-\beta}\bigg| \int_{J'_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} u \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_{n,m}^2}\right) du \\
&\qquad\qquad - \mu_{n,m}\int_{J'_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_{n,m}^2}\right) du \bigg| \\
&\lesssim n^{-\beta}\left[2\int_0^{n^{\beta}(\xi_{n,m-1}\vee\xi_{n,m})} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} u \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_{n,m}^2}\right) du \right.\\
&\qquad\qquad \left. +|\mu_{n,m}|\int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}\right) du\right] \\
&\lesssim n^{-\beta}\left[\sigma_{n,m}^2\varphi_{11,m}^{\frac{1}{2}}\left(1-\exp\left(-\frac{n(\xi_{n,m-1}\vee\xi_{n,m})^2}{2\sigma_{n,m}^2\varphi_{11,m}}\right)\right) n^{-\frac{1}{2}+\beta} + \mu_{n,m}\sqrt{2\pi\sigma_{n,m}^2}\right] \\
&\lesssim \sqrt{\frac{\varphi_{11,m}}{n}} \wedge n^{-\beta},
\end{aligned}
$$

where $J'_{n,m} = [-n^{\beta}\xi_{n,m-1} + \mu_{n,m}, n^{\beta}\xi_{n,m} + \mu_{n,m}]$. Hence, $I_1 \lesssim \lambda^{r_1} \wedge \sqrt{\frac{\varphi_{11,m}}{n}} \wedge n^{-\beta} \to 0$ under Assumption B2. By Lemma S1 and Theorem 1, we have

$$
\begin{aligned}
I_2 \lesssim &\bigg| \int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\left[\exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\left(-\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}+\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\right)\right.\\
&\qquad \left. - \exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\left(-\frac{u^2}{2\sigma_m^{*2}}+\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\right)\pi(t_m)\right] du\bigg| \\
= &\bigg| \int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}}\left[\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}\right) - \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_m^{*2}}\right)\right] du\bigg| \\
\le &\bigg| \int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}}\left[\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{(u+\mu_{n,m})^2}{2\sigma_{n,m}^2}\right) - \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_{n,m}^2}\right)\right] du\bigg| \\
&+ \bigg| \int_{J_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}}\left[\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_{n,m}^2}\right) - \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_m^{*2}}\right)\right] du\bigg| \\
= &\, I_{21} + I_{22}.
\end{aligned}
$$

Without loss of generality we assume $\mu_{n,m} \ge 0$. Then, combining Equations (15), (16) and

(S6) gives that

$$
\begin{aligned}
I_{21} &= \left| \sqrt{\frac{\pi\sigma_{n,m}^2}{2}} \left[\mathrm{Erf}\left(\frac{\xi_{n,m-1}n^{\frac{1}{2}} - \mu_{n,m}n^{\frac{1}{2}-\beta}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) + \mathrm{Erf}\left(\frac{\xi_{n,m}n^{\frac{1}{2}} + \mu_{n,m}n^{\frac{1}{2}-\beta}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) \right] \right. \\
&\qquad \left. - \sqrt{\frac{\pi\sigma_{n,m}^2}{2}} \left[\mathrm{Erf}\left(\frac{\xi_{n,m-1}n^{\frac{1}{2}}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) + \mathrm{Erf}\left(\frac{\xi_{n,m}n^{\frac{1}{2}}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) \right] \right| \\
&\lesssim \left| \mathrm{Erf}\left(\frac{\xi_{n,m-1}n^{\frac{1}{2}} - \mu_{n,m}n^{\frac{1}{2}-\beta}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) - \mathrm{Erf}\left(\frac{\xi_{n,m-1}n^{\frac{1}{2}}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) \right| \\
&\qquad + \left| \mathrm{Erf}\left(\frac{\xi_{n,m}n^{\frac{1}{2}} + \mu_{n,m}n^{\frac{1}{2}-\beta}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) - \mathrm{Erf}\left(\frac{\xi_{n,m}n^{\frac{1}{2}}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) \right| \\
&\le \mu_{n,m}n^{\frac{1}{2}-\beta} \cdot \exp\left(-\frac{\xi_{n,m-1}^2 n}{2\sigma_{n,m}^2\varphi_{11,m}} \right) + \mu_{n,m}n^{\frac{1}{2}-\beta} \cdot \exp\left(-\frac{\xi_{n,m}^2 n}{2\sigma_{n,m}^2\varphi_{11,m}} \right) \\
&\lesssim e^{-cn\varphi_{11,m}^{-1}}
\end{aligned}
$$

for some $c > 0$. In view of Equations (16), (S6), and Theorem 1, we obtain

$$
\begin{aligned}
I_{22} &= \left| \sqrt{\frac{\pi\sigma_{n,m}^2}{2}} \left[\mathrm{Erf}\left(\frac{\xi_{n,m-1}n^{\frac{1}{2}}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) + \mathrm{Erf}\left(\frac{\xi_{n,m}n^{\frac{1}{2}}}{\sqrt{2\sigma_{n,m}^2\varphi_{11,m}}} \right) \right] \right. \\
&\qquad \left. - \sqrt{\frac{\pi\sigma_m^{*2}}{2}} \left[\mathrm{Erf}\left(\frac{\xi_{n,m-1}n^{\frac{1}{2}}}{\sqrt{2\sigma_m^{*2}\varphi_{11,m}}} \right) + \mathrm{Erf}\left(\frac{\xi_{n,m}n^{\frac{1}{2}}}{\sqrt{2\sigma_m^{*2}\varphi_{11,m}}} \right) \right] \right| \lesssim |\sigma_m^{*2} - \sigma_{n,m}^2| \lesssim \lambda^r.
\end{aligned}
$$

Therefore, $I_2 \to 0$.

Similarly, by changing of variable and Lemma S1, each term in Equation (S11) becomes

$$
\begin{aligned}
I_3 &= \int_{[0,z]\setminus I_{n,m}} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt \\
&= \int_{K_{n,m}} n^{-\beta}\ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m)du \\
&\lesssim \int_{K_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2} \right) \exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\left(-\frac{u^2}{2\sigma_m^{*2}} + \frac{\mu_{n,m}^2}{2\sigma_{n,m}^2} \right) \right) du \\
&= \int_{K_{n,m}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_m^{*2}} \right) du,
\end{aligned}
$$

where $K_{n,m} = [-n^\beta t_{\lambda,m}, -n^\beta \xi_{n,m-1}] \cup [n^\beta \xi_{n,m}, n^\beta(z - t_{\lambda,m})]$. It then follows that

$$I_3 \lesssim n^\beta \cdot n^{\frac{1}{2}-\beta} \varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1} \frac{n^{2\beta}(z - t_{\lambda,m})^2}{2\sigma_m^{*2}}\right) \to 0.$$

Following the same arguments, we can show that Equations (S12) and (S13) converge to zero. This proves Equation (S8) for $0 < z \leq 1$.

Case (3) $(z > 1)$. From Case (2) we can see that

$$\left| \int_{-\infty}^1 \ell(t)\pi(t)dt - \int_{-\infty}^1 \tilde{h}_n(t)dt \right| \to 0.$$

Note that $\ell(t) = 0$ in $\mathbb{R}\backslash[0, 1]$. Using similar arguments as in Case (1), it holds that $\int_1^z \tilde{h}_n(t)dt \to 0$, proving Equation (S8) for $z > 1$.

Step 3: We normalize $\tilde{h}_n(t)$ to a density

$$h_n(t) = \frac{\tilde{h}_n(t)}{\int_{\mathbb{R}} \tilde{h}_n(t)dt}.$$

Note that Equation (S8) implies

$$\left| \left(\int_{\mathbb{R}} \ell(t)\pi(t)dt\right)^{-1} - \left(\int_{\mathbb{R}} \tilde{h}_n(t)dt\right)^{-1} \right| \to 0. \tag{S14}$$

Hence, for any $z \in \mathbb{R}$, we have

$$\begin{aligned}\left| \int_{-\infty}^z \pi_n(t \mid X, \boldsymbol{y})dt - \int_{-\infty}^z h_n(t)du \right| &\leq \left(\int_{\mathbb{R}} \ell(t)\pi(t)dt\right)^{-1} \left| \int_{-\infty}^z \ell(t)\pi(t)dt - \int_{-\infty}^z \tilde{h}_n(t)dt \right| \\ &\quad + \int_{-\infty}^z \left| \left(\int_{\mathbb{R}} \ell(t)\pi(t)dt\right)^{-1} - \left(\int_{\mathbb{R}} \tilde{h}_n(t)dt\right)^{-1} \right| \tilde{h}_n(t)dt \\ &\to 0, \end{aligned} \tag{S15}$$

where the last line follows from Equations (S8) and (S14).

Rewrite $\tilde{h}_n(t)$ defined in Equation (S7) to

$$\tilde{h}_n(t) = \sum_{m=1}^{M} \pi(t_m)\ell(t_{\lambda,m}) \exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\sqrt{2\pi n^{-1}\varphi_{11,m}\sigma_m^{*2}}\phi(t \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2}),$$

which is a linear combination of $\phi(t \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2})$. Hence, the density function after normalization is

$$h_n(t) = \sum_{m=1}^{M} \tilde{\pi}_{n,m}\phi(t \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2}),$$

with weights

$$\begin{aligned}
\tilde{\pi}_{n,m} &= \frac{\pi(t_m)\ell(t_{\lambda,m}) \exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\sqrt{2\pi n^{-1}\varphi_{11,m}\sigma_m^{*2}}}{\sum_{m=1}^{M}\pi(t_m)\ell(t_{\lambda,m}) \exp\left(n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)\sqrt{2\pi n^{-1}\varphi_{11,m}\sigma_m^{*2}}} \\
&= \frac{\pi(t_m)\sigma_m^* n^{-\frac{1}{2}}\varphi_{11,m}^{\frac{1}{2}}\ell(t_{\lambda,m})/\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)}{\sum_{m=1}^{M}\pi(t_m)\sigma_m^* n^{-\frac{1}{2}}\varphi_{11,m}^{\frac{1}{2}}\ell(t_{\lambda,m})/\exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right)} \\
&= \frac{\pi(t_m)C|f_0''(t_m)|^{-1} + c_{n,m}}{\sum_{m=1}^{M}\pi(t_m)C|f_0''(t_m)|^{-1} + c_{n,m}}
\end{aligned}$$

where the existence of sequences $c_{n,m} = O(n^{\frac{1}{2}-2\beta}(\log n)^{-1-a})$ is guaranteed by Lemma S1. Hence, we arrive at

$$\tilde{\pi}_{n,m} = \frac{\pi(t_m)|f_0''(t_m)|^{-1}}{\sum_{m=1}^{M}\pi(t_m)|f_0''(t_m)|^{-1}} + c'_{n,m} =: \pi_m + c'_{n,m},$$

for some $c'_{n,m} = O(n^{\frac{1}{2}-2\beta}(\log n)^{-1-a})$. It then holds that

$$\begin{aligned}
&\left|\int_{-\infty}^{z} h_n(t)dt - \int_{-\infty}^{z}\sum_{m=1}^{M}\pi_m\phi(t \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2})dt\right| \\
&\leq \sum_{m=1}^{M}\int_{-\infty}^{z}|\tilde{\pi}_{n,m} - \pi_m|\phi(t \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2})dt \to 0. \qquad \text{(S16)}
\end{aligned}$$

Combining Equations (S15) and (S16), we obtain that for any z,

$$\mathbb{E}_{\mathbb{P}_0}\left|\int_{-\infty}^{z}\pi_n(t \mid X, \boldsymbol{y})dt - \int_{-\infty}^{z}\sum_{m=1}^{M}\pi_m\phi(t \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2})dt\right|\mathbf{1}_{A_n} \to 0.$$

This together with $\mathbb{E}_{\mathbb{P}_0}(\mathbf{1}_{A_n^c}) = \mathbb{P}_0(A_n^c) \leq n^{-10}$ gives that

$$\left| \Pi_n(t \leq z \mid X, \boldsymbol{y}) - \sum_{m=1}^{M} \pi_m \Phi(z \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2}) \right| \to 0$$

for any $z \in \mathbb{R}$ in $\mathbb{P}_0$-probability. This completes the proof.

A.5.2 Proof of (ii)

Denote $\zeta_0 = t_1$, $\zeta_m = \frac{1}{2}(t_{m+1} - t_m)$, $m = 1, \ldots, M-1$, $\zeta_M = 1 - t_M$. Then,

$$[0,1] = \bigcup_{m=1}^{M} I_m, \quad I_m = [t_m - \zeta_{m-1}, t_m + \zeta_m], \quad m = 1, \ldots, M.$$

We first bound the unnormalized difference

$$\left| \int_{-\infty}^{z} \ell(t)\mathbf{1}_{I_m}(t)\pi(t)dt - \int_{-\infty}^{z} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt \right| \tag{S17}$$

under A_n by considering three cases for z: (1) $z \leq t_m - \zeta_{m-1}$, (2) $t_m - \zeta_{m-1} < z < t_m + \zeta_m$, and (3) $z \geq t_m + \zeta_m$.

Case 1 ($z \leq t_m - \zeta_{m-1}$). Since $z \notin I_m$, Equation (S17) becomes

$$\begin{aligned}
\int_{-\infty}^{z} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt &= \int_{-\infty}^{(z-t_{\lambda,m})n^{\beta}} \ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m)du \\
&\lesssim \int_{-\infty}^{(z-t_{\lambda,m})n^{\beta}} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_m^{*2}}\right) du \\
&= \sqrt{\frac{\pi\sigma_m^{*2}}{2}} \left[1 - \operatorname{Erf}\left(\frac{\sqrt{n}(t_{\lambda,m} - z)}{\sqrt{2\sigma_m^{*2}\varphi_{11,m}}}\right)\right] \\
&\lesssim \exp\left(-\frac{n(t_{\lambda,m} - t_m - \zeta_{m-1})^2}{2\sigma_m^{*2}\varphi_{11,m}}\right).
\end{aligned}$$

Case 2 ($t_m - \zeta_{m-1} < z < t_m + \zeta_m$). In this case, we consider

$$
\begin{aligned}
&\left|\int_{t_m-\zeta_{m-1}}^{z} \left[\ell(t)\pi(t) - \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)\right] dt\right| \\
= &\left|\int_{H_{n,m}} \left[\ell(t_{\lambda,m}+u/n^\beta)\pi(t_{\lambda,m}+u/n^\beta) - \ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m)\right] n^{-\beta}du\right| \\
= &\left|\int_{H_{n,m}} n^{-\beta}\ell(t_{\lambda,m}) \left[Z_{n,m}(u)\pi(t_{\lambda,m}+u/n^\beta) - \nu_{n,m}(u)\pi(t_m)\right] du\right| \\
\leq &\left|\int_{H_{n,m}} n^{-\beta}\ell(t_{\lambda,m})Z_{n,m}(u)[\pi(t_{\lambda,m}+u/n^\beta) - \pi(t_m)]du\right| \\
&+ \left|\int_{H_{n,m}} n^{-\beta}\ell(t_{\lambda,m})[Z_{n,m}(u) - \nu_{n,m}(u)]\pi(t_m)du\right| \\
&= I_1' + I_2',
\end{aligned}
$$

where $H_{n,m} = [(t_m - t_{\lambda,m} - \zeta_{m-1})n^\beta, (z - t_{\lambda,m})n^\beta]$. Following similar arguments as used in the proof of Part (i), it can be shown that $I_1' \lesssim \lambda^{r_1}$ and

$$
I_2' \lesssim \mu_{n,m} n^{\frac{1}{2}-\beta} \cdot \exp\left(-\frac{(z-t_{\lambda,m})^2 n}{2\sigma_{n,m}^2 \varphi_{11,m}}\right).
$$

Case 3 ($z \geq t_m + \zeta_m$). Again, $z \notin I_m$ and Equation (S17) becomes

$$
\begin{aligned}
\int_{t_m+\zeta_m}^{z} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt &= \int_{(t_m-t_{\lambda,m}+\zeta_m)n^\beta}^{(z-t_{\lambda,m})n^\beta} \ell(t_{\lambda,m})\nu_{n,m}(u)\pi(t_m)du \\
&\lesssim \int_{(t_m-t_{\lambda,m}+\zeta_m)n^\beta}^{\infty} n^{\frac{1}{2}-\beta}\varphi_{11,m}^{-\frac{1}{2}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{u^2}{2\sigma_m^{*2}}\right) du \\
&= \sqrt{\frac{\pi\sigma_m^{*2}}{2}}\left[1 - \operatorname{Erf}\left(\frac{\sqrt{n}(t_m - t_{\lambda,m}+\zeta_m)}{\sqrt{2\sigma_m^{*2}\varphi_{11,m}}}\right)\right] \\
&\lesssim \exp\left(-\frac{n(t_m - t_{\lambda,m}+\zeta_m)^2}{2\sigma_m^{*2}\varphi_{11,m}}\right).
\end{aligned}
$$

Combining the three cases, we obtain that under A_n,

$$
\left|\int_{-\infty}^{z} \ell(t)\mathbf{1}_{I_m}(t)\pi(t)dt - \int_{-\infty}^{z} \ell(t_{\lambda,m})\tilde{\phi}_{n,m}(t)\pi(t_m)dt\right| \lesssim \lambda^{r_1} \vee n^{\frac{1}{2}-\beta} \cdot \exp\left(-\frac{(z-t_{\lambda,m})^2 n}{2\sigma_{n,m}^2\varphi_{11,m}}\right).
$$

Let $\Pi_{n,m}(\cdot \mid X, \boldsymbol{y})$ be the posterior of $t\mathbf{1}_{I_m}$. Following the same arguments as in part (i) again, we can show that

$$\left|\Pi_{n,m}(t' \le z \mid X, \boldsymbol{y}) - \Phi(z \mid t_{\lambda,m}, n^{-1}\varphi_{11,m}\sigma_m^{*2})\right| \lesssim \lambda^{r_1} \vee n^{\frac{1}{2}-\beta} \cdot \exp\left(-\frac{(z-t_{\lambda,m})^2 n}{2\sigma_{n,m}^2\varphi_{11,m}}\right)$$

in $\mathbb{P}_0$-probability.

By Lemma 2, we have $|\hat{t}_m - b_n - t_{\lambda,m}| \lesssim n^{-\beta}\sqrt{\log n} \vee n^{-\beta}\log n = n^{-\beta}\log n$. Thus,

$$\left|\Pi_{n,m}(t' \le z \mid X, \boldsymbol{y}) - \Phi(z \mid \hat{t}_m - b_n, n^{-1}\varphi_{11,m}\sigma_m^{*2})\right| \lesssim \lambda^{r_1} \vee n^{\frac{1}{2}-\beta} \cdot \exp\left(-\frac{(z-t_{\lambda,m})^2 n}{2\sigma_{n,m}^2\varphi_{11,m}}\right) \vee n^{-\beta}\log n.$$

Now we consider the posterior of $\sqrt{\frac{n}{\varphi_{11,m}}}(t\mathbf{1}_{I_m}(t) - \hat{t}_m + b_n)$. By changing of variable, it follows that

$$\begin{aligned}
&\left|\Pi'_{n,m}(t' \le z \mid X, \boldsymbol{y}) - \Phi(z \mid 0, \sigma_m^{*2})\right| \\
=& \left|\Pi_{n,m}\left(t' \le \sqrt{\frac{\varphi_{11,m}}{n}}z + \hat{t}_m - b_n \mid X, \boldsymbol{y}\right) - \Phi\left(\sqrt{\frac{\varphi_{11,m}}{n}}z + \hat{t}_m - b_n | \hat{t}_m - b_n, n^{-1}\varphi_{11,m}\sigma_m^{*2}\right)\right| \\
\lesssim& n^{\frac{1}{2}-\beta} \cdot \exp\left\{-\frac{(\sqrt{\frac{\varphi_{11,m}}{n}}z + \hat{t}_m - t_{\lambda,m} - b_n)^2 n}{2\sigma_{n,1}^2\varphi_{11,m}}\right\} \\
\lesssim& n^{\frac{1}{2}-\beta} \cdot \exp\left\{-\left(\sqrt{\frac{\varphi_{11,m}}{n}}z \vee (\hat{t}_m - t_{\lambda,m}) \vee b_n\right)^2 \frac{n}{2\sigma_{n,m}^2\varphi_{11,m}}\right\} \\
\lesssim& n^{\frac{1}{2}-\beta} \cdot \exp\left\{-\frac{b_n^2 n}{2\sigma_{n,m}^2\varphi_{11,m}}\right\} \\
\lesssim& n^{\frac{1}{2}-\beta} \cdot \exp\left\{-\frac{n^{1-2\beta}(\log n)^2}{2\sigma_{n,m}^2\varphi_{11,m}}\right\} \to 0.
\end{aligned}$$

This completes the proof.

A.6 Proof of Theorem 3

A.6.1 Proof of (i)

Let $F_n(z) = \Pi_n(t \le z \mid X, \boldsymbol{y})$ and $G_n(z) = \sum_{m=1}^M \pi_m \Phi(z \mid t_m, n^{-1}\varphi_{11,m}\sigma_m^{*2})$. Note that $G_n(\cdot)$ is a deterministic function, and its derivative $G'_n(\cdot)$ is the density function of a Gaussian mixture. The variance of each component distribution in G_n goes to zero in view of

Assumption B2 and conditions in Theorem 2. For sufficiently large n, using the analytical expression of $G_n'(\cdot)$ and elementary calculus, we can show that $G_n'(\cdot)$ has at least M local modes, denoted by $t_{m,G}$, such that $t_{m,G} \to t_m$. On the other hand, $G_n'(\cdot)$ cannot have more than M local modes in view of Corollary 2.4 in Carreira-Perpinán and Williams (2003); hence, $\{t_{m,1}, \ldots, t_{M,G}\}$ are the only local modes of $G_n'(\cdot)$. For large enough n and each m, we consider an interval $(t_{m,G} - \delta_m, t_{m,G} + \delta_m)$ for some $\delta_m > 0$ such that $G_n''(z) > 0$ when $z \in (t_{m,G} - \delta_m, t_{m,G})$ and $G_n''(z) < 0$ when $z \in (t_{m,G}, t_{m,G} + \delta_m)$.

By Theorem 2 (i), we have $|F_n(z) - G_n(z)| \to 0$ for any $z \in \mathbb{R}$ in $\mathbb{P}_0$-probability. The following arguments and conclusions in Step 1–4 hold with $\mathbb{P}_0$-probability tending to 1 because of this convergence in $\mathbb{P}_0$-probability.

Step 1: We first show that there exists a $t_{m,F}$ in the neighborhood of $t_{m,G}$ such that $F_n''(t_{m,F}) = 0$, for $m = 1, \ldots, M$. Suppose $F_n''(z) \neq 0$ for any $z \in (t_{m,G} - \delta_m, t_{m,G} + \delta_m)$, Without loss of generality we assume $F_n''(z) > 0$ when $z \in (t_{m,G} - \delta_m, t_{m,G} + \delta_m)$. Since $G_n(z)$ is concave on $(t_{m,G}, t_{m,G} + \delta_m)$,

$$G_n(t_{m,G} + \delta_m/2) > (G_n(t_{m,G}) + G_n(t_{m,G} + \delta_m))/2 + \epsilon, \tag{S18}$$

for some $\epsilon > 0$. Since $F_n(z)$ is convex on $(t_{m,G}, t_{m,G} + \delta_m)$,

$$F_n(t_{m,G} + \delta_m/2) < (F_n(t_{m,G}) + F_n(t_{m,G} + \delta_m))/2.$$

For sufficiently large n, it holds that with $\mathbb{P}_0$-probability tending to 1 $|F_n(z) - G_n(z)| < \epsilon/2$ for $z = t_{m,G}, t_{m,G} + \delta_m/2, t_{m,G} + \delta_m$. Therefore,

$$G_n(t_{m,G} + \delta_m/2) > (F_n(t_{m,G}) + F_n(t_{m,G} + \delta_m))/2 + \epsilon/2 > F_n(t_{m,G} + \delta_m/2) + \epsilon/2,$$

which is a contradiction. This proves that there exists $t_{m,F} \in (t_{m,G} - \delta_m, t_{m,G} + \delta_m)$ such that $F_n''(t_{m,F}) = 0$.

Step 2: We show that $t_{m,F} \to t_m$ in $\mathbb{P}_0$-probability. Suppose there exists $\delta > 0$ such that

$|t_{m,G} - t_{m,F}| > \delta$ for any sufficiently large n. Without loss of generality we assume $t_{m,G} < t_{m,F}$ and $F_n''(z) < 0$ when $z \in (t_{m,F}, t_{m,G} + \delta_m)$ and $F_n''(z) > 0$ when $z \in (t_{m,G} - \delta_m, t_{m,F})$. Thus, G_n is concave on $(t_{m,G}, t_{m,F})$ while F_n is convex on $(t_{m,G}, t_{m,F})$. This is a contradiction using the same argument in Step 1. Combining this with $t_{m,G} \to t_m$ shows the convergence of $t_{m,F}$.

Step 3: In this step, we show that $t_{m,F}$ must be a local mode of $F_n'(z)$. Suppose that $F_n''(z) > 0$ when $z \in (t_{m,F}, t_{m,G} + \delta_m)$ and $F_n''(z) < 0$ when $z \in (t_{m,G} - \delta_m, t_{m,F})$, yielding

$$F_n(t_{m,F} + \delta_m/2) < (F_n(t_{m,F}) + F_n(t_{m,F} + \delta_m))/2.$$

For sufficiently large n, it holds with $\mathbb{P}_0$-probability tending to 1 that $|F_n(z) - G_n(z)| < \epsilon/4$ for $x = t_{m,G}, t_{m,G} + \delta_m/2, t_{m,G} + \delta_m$. Invoking Equation (S18),

$$G_n(t_{m,G} + \delta_m/2) > (F_n(t_{m,G}) + F_n(t_{m,G} + \delta_m))/2 + 3\epsilon/4.$$

For sufficiently large n, it holds with $\mathbb{P}_0$-probability tending to 1 that $|F_n(z_1) - F_n(z_2)| < \epsilon/4$ for $z_1 = t_{m,G}$, $z_2 = t_{m_F}$, $z_1 = t_{m,G} + \delta_m/2$, $z_2 = t_{m,F} + \delta_m/2$ and $z_1 = t_{m,G} + \delta_m$, $z_2 = t_{m,F} + \delta_m$. Therefore,

$$G_n(t_{m,G} + \delta_m/2) > (F_n(t_{m,F}) + F_n(t_{m,F} + \delta_m))/2 + \epsilon/2 > F_n(t_{m,F} + \delta_m/2) + \epsilon/2.$$

However,

$$G_n(t_{m,G} + \delta_m/2) < F_n(t_{m,G} + \delta_m/2) + \epsilon/4 < F_n(t_{m,F} + \delta_m/2) + \epsilon/2,$$

which is a contradiction. This completes Step 3.

Step 4: In the last step, we show that the number of local modes of $F_n'(z)$ is exactly M. We have proven that $F_n'(\cdot)$ has at least M local modes. Suppose that there exists $t_{m',F} \in (0, 1)$ and $\delta_{m'} > 0$ such that $t_{m',F}$ is a local mode of $F_n'(z)$ and $G_n''(z) \neq 0$ for $z \in (t_{m',F} - \delta_{m'}, t_{m',F} + \delta_{m'})$ for any sufficiently large n. Without loss of generality assume

$G_n''(z) > 0$ for $z \in (t_{m',F} - \delta_{m'}, t_{m',F} + \delta_{m'})$. Thus, on $(t_{m',F} - \delta_{m'}, t_{m',F} + \delta_{m'})$, $G_n(z)$ is convex while $F_n(z)$ is concave. By similar arguments used in Step 1, we can obtain a contradiction. Hence, the number of local modes of $F_n'(\cdot)$ is exactly M.

This completes the proof.

A.6.2 Proof of (ii)

By Taylor expansion of $\widehat{\mu}_{f'}$, we obtain

$$\widehat{\mu}_{f'}(t) = \widehat{\mu}_{f'}(t_m) + (t - t_m)\widehat{\mu}_{f'}'(\xi)$$

for some ξ between t and t_m. Since $\hat{t}_m$ is a local extremum of $\hat{\mu}_f$, there holds $\widehat{\mu}_{f'}(\hat{t}_m) = 0$. Substituting $t = \hat{t}_m$ into the expansion above yields

$$\widehat{\mu}_{f'}(t_m) + (\hat{t}_m - t_m)\widehat{\mu}_{f'}'(\xi) = 0.$$

Lemma 1 and Assumption C ensure that $\widehat{\mu}_{f'}'(x) \xrightarrow{p} f_0''(x)$, and Lemma 2 implies that $\hat{t}_m \xrightarrow{p} t_m$. Therefore, $\widehat{\mu}_{f'}'(\xi) \xrightarrow{p} f_0''(t_m)$, and thus $\widehat{\mu}_{f'}'(\xi)$ is bounded away from zero and infinity in view of Assumption A3. It thus follows that

$$\hat{t}_m - t_m = -\frac{\widehat{\mu}_{f'}(t_m)}{\widehat{\mu}_{f'}'(\xi)}.$$

Let $\Delta_n(\cdot) = K_{10}(\cdot, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-1} f_0(X)$. Conditioning on X, it holds that

$$\begin{aligned}\widehat{\mu}_{f'}(t_m) &= K_{10}(t_m, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-1}\boldsymbol{y} \\ &\sim N\left(\Delta_n(t_m), \sigma^2 K_{10}(t_m, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T\right).\end{aligned}$$

Hence,

$$\left.\frac{\widehat{\mu}_{f'}(t_m) - \Delta_n(t_m)}{\sigma\sqrt{K_{10}(t_m, X)[K(X, X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T}}\right| X \sim N(0, 1),$$

which implies that

$$\frac{\widehat{\mu}_{f'}(t_m) - \Delta_n(t_m)}{\sigma\sqrt{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T}} \sim N(0,1).$$

By Slutsky's theorem, we obtain

$$\sqrt{\frac{1}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T}} \left[\hat{t}_m - t_m + \frac{\Delta_n(t_m)}{\widehat{\mu}'_{f'}(t_m)}\right] \xrightarrow{d} N\left(0, \sigma^2 f_0''(t_m)^{-2}\right).$$

Note that

$$\begin{aligned} &\frac{K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(\hat{t}_m, X)^T}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T} - 1 \\ = &\frac{K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2}(K_{10}(\hat{t}_m, X)^T - K_{10}(t_m, X))}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T} \\ &+ \frac{(K_{10}(\hat{t}_m, X)^T - K_{10}(t_m, X))[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X))}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T}. \end{aligned}$$

Consider the eigendecomposition of $K(X,X) = Q_n \Lambda_n Q_n^T$, where $\Lambda_n = \mathrm{diag}(u_1, \ldots, u_n)$ and $Q_n^T = Q_n^{-1}$. Denote $(p_1, \ldots, p_n) = K_{10}(t_m, X)Q_n$, likewise $(q_1, \ldots, q_n) = K_{10}(\hat{t}_m, X)Q_n$. Then

$$\begin{aligned} K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T &= K_{10}(t_m, X) Q_n \Lambda_n^{-2} Q_n^T K_{10}(t_m, X) \\ &= \sum_{i=1}^{\infty} \frac{p_i^2}{(u_i + n\lambda)^2}. \end{aligned}$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned} &K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2}(K_{10}(\hat{t}_m, X)^T - K_{10}(t_m, X)) \\ = &\sum_{i=1}^{\infty} \frac{q_i(q_i - p_i)}{(u_i + n\lambda)^2} \leq \sqrt{\sum_{i=1}^{\infty} \frac{q_i^2}{(u_i + n\lambda)^2} \sum_{i=1}^{\infty} \frac{(q_i - p_i)^2}{(u_i + n\lambda)^2}}. \end{aligned}$$

Since $K_{10}(\hat{t}_m, X_i) - K_{10}(t_m, X_i) \xrightarrow{p} 0$ uniformly for $1 \le i \le n$, we have

$$\sum_{i=1}^{\infty} \frac{q_i^2}{(u_i + n\lambda)^2} \Big/ \sum_{i=1}^{\infty} \frac{p_i^2}{(u_i + n\lambda)^2} \xrightarrow{p} 1$$

and

$$\sum_{i=1}^{\infty} \frac{(q_i - p_i)^2}{(u_i + n\lambda)^2} \Big/ \sum_{i=1}^{\infty} \frac{p_i^2}{(u_i + n\lambda)^2} \xrightarrow{p} 0.$$

Hence, it follows that

$$\frac{K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2}(K_{10}(\hat{t}_m, X)^T - K_{10}(t_m, X))}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T} \xrightarrow{p} 0.$$

Similarly, it can be shown that

$$\frac{(K_{10}(\hat{t}_m, X)^T - K_{10}(t_m, X))[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X))}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T} \xrightarrow{p} 0.$$

Therefore,

$$\frac{K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(\hat{t}_m, X)^T}{K_{10}(t_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(t_m, X)^T} \xrightarrow{p} 1.$$

Recall that $\widehat{\mu}'_{f'}(\hat{t}_m) \xrightarrow{p} f''_0(t_m)$. Therefore, by Slutsky's theorem again, we arrive at

$$\frac{\sigma |\widehat{\mu}'_{f'}(\hat{t}_m)|}{\sqrt{K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(\hat{t}_m, X)^T}} \left[\hat{t}_m - t_m + \frac{\Delta_n(t_m)}{\widehat{\mu}'_{f'}(t_m)} \right] \xrightarrow{d} N(0,1).$$

Hence, an asymptotic $1-\alpha$ confidence interval of $t_m + \Delta_n(t_m)/f''_0(t_m)$ is

$$\hat{t}_m \pm z_{\alpha/2} \frac{\sigma \sqrt{K_{10}(\hat{t}_m, X)[K(X,X) + n\lambda \boldsymbol{I}_n]^{-2} K_{10}(\hat{t}_m, X)^T}}{|\widehat{\mu}'_{f'}(\hat{t}_m)|}.$$

This completes the proof.

A.7 Proof of Theorem 4

For any $j, l \le s$, we have

$$|K_{jl}^{\alpha,s}(x,x')| = \left|\sum_{i=1}^{\infty} \mu_i \psi_i^{(j)}(x)\psi_i^{(l)}(x')\right| \lesssim \sum_{i=1}^{\infty} i^{-2\alpha+j+l},$$

which is finite when $\alpha > \frac{j+l+1}{2}$. Thus, Assumption B1 holds when $s \ge 4$ and $\alpha > 9/2$. According to Lemma 11 in Liu and Li (2023), when $\alpha > \frac{j+l+1}{2}$, we have

$$\sup_{x\in\mathcal{X}} |\varphi_{jl}(x)| = \sup_{x\in\mathcal{X}} \left|\sum_{i=1}^{\infty} \frac{\mu_i}{\lambda+\mu_i}\psi^{(j)}(x)\psi^{(l)}(x)\right| \lesssim \sum_{i=1}^{\infty} \frac{\mu_i i^{j+l}}{\mu_i+\lambda} \asymp \lambda^{-\frac{j+l+1}{2\alpha}}.$$

Hence, Assumption B2 is satisfied when $\alpha > 3$. In view of Lemma 1, Lemma 11 and Lemma 13 in Liu and Li (2023), when $\alpha > k + 1/2$, we have

$$\|f_\lambda^{(k)} - f_0^{(k)}\|_\infty \lesssim \lambda^{\frac{1}{2}-\frac{k}{2\alpha}}.$$

This verifies Assumption C with $r_1 = \frac{\alpha-1}{2\alpha}$, $r_2 = \frac{\alpha-2}{2\alpha}$ when $\alpha > 5/2$. Finally, by Assumption E we have $\varphi_{11}(x) = \sum_{i=1}^{\infty} \frac{\mu_i}{\lambda+\mu_i}\psi_i'(x)^2 \to \infty$ as $\lambda \to 0$. Thus, a sufficient condition for the boundedness of $n^{\frac{3}{2}-4\beta}\varphi_{11,m}^{-2}$ in Theorem 1 and 2 is $n^{\frac{3}{2}-4\beta} = O(1)$, which implies that $\beta \ge \frac{3}{8}$. This completes the proof.

A.8 Proof of Lemma 3

Let $f_0 = \sum_{i=1}^{\infty} f_i\psi_i$. Then, for any $k \le s$,

$$f_\lambda^{(k)} - f_0^{(k)} = -\sum_{i=1}^{\infty} \frac{\lambda}{\lambda+\mu_i} f_i \psi_i^{(k)}.$$

Hence,

$$\|f_\lambda^{(k)} - f_0^{(k)}\|_\infty \le \sum_{i=1}^{\infty} \frac{\lambda}{\lambda+\mu_i}|f_i| \cdot i^k \lesssim \lambda^{\frac{1}{2}-\frac{k}{2e\gamma}} \sum_{i=1}^{\infty} \frac{\lambda^{\frac{1}{2}+\frac{k}{2e\gamma}} \cdot e^{-\gamma i} i^k}{\lambda + e^{-2\gamma i}} e^{\gamma i}|f_i|.$$

Note that $i^k \leq e^{\frac{ki}{e}}$, then by Young's inequality for products, we have

$$\lambda^{\frac{1}{2}+\frac{k}{2e\gamma}} \cdot e^{-\gamma i} i^k \leq \lambda^{\frac{1}{2}+\frac{k}{2e\gamma}} \cdot e^{(\frac{k}{e}-\gamma)i} \leq \left(\frac{1}{2}+\frac{k}{2e\gamma}\right)\lambda + \left(\frac{1}{2}-\frac{k}{2e\gamma}\right) e^{-2\gamma i} \leq \lambda + e^{-2\gamma i}.$$

Therefore, $\|f_\lambda^{(k)} - f_0^{(k)}\|_\infty \lesssim \lambda^{\frac{1}{2}-\frac{k}{2e\gamma}} \sum_{i=1}^\infty e^{\gamma i}|f_i| \lesssim \lambda^{\frac{1}{2}-\frac{k}{2e\gamma}}$. This completes the proof.

A.9 Proof of Theorem 5

It is easy to see that $K_{\gamma,s} \in C^8(\mathcal{X},\mathcal{X})$ for any $\gamma > 0$ and $s \geq 4$; thus Assumption B1 is satisfied.

Note that

$$\sup_{x\in\mathcal{X}} |\varphi_{jl}(x)| = \left| \sum_{i=1}^\infty \frac{\mu_i}{\lambda+\mu_i} \psi_i^{(j)}(x)\psi_i^{(l)}(x) \right| \lesssim \sum_{i=1}^\infty \frac{e^{-2\gamma i} i^{j+l}}{\lambda + e^{-2\gamma i}} \leq \frac{e^{-2\gamma i} e^{\frac{(j+l)i}{e}}}{\lambda + e^{-2\gamma i}}.$$

By Young's inequality for products, when $2e\gamma > j+l$ we have

$$\lambda^{\frac{j+l}{2e\gamma}} \cdot e^{-2\gamma i + \frac{(j+l)i}{e}} \leq \left(1 - \frac{j+l}{2e\gamma}\right)\lambda + \left(\frac{j+l}{2e\gamma}\right) e^{-2\gamma i} \leq \lambda + e^{-2\gamma i}.$$

Hence, $\varphi_{jl} \lesssim \lambda^{-\frac{j+l}{2e\gamma}}$ for $2e\gamma > j+l$ and Assumption B2 holds for $\gamma > \frac{5}{2e}$. In view of Lemma 3, we have Assumption C satisfied with $r_1 = r_2 = \frac{e\gamma-2}{2e\gamma}$ and $\gamma > \frac{2}{e}$.

Finally, since $\lambda = o(1)$, we have $\varphi_{11,m}^{-2} = o(1)$ under Assumption E. Thus, a sufficient condition for the boundedness of $n^{\frac{3}{2}-4\beta}\varphi_{11,m}^{-2}$ is $\beta \geq \frac{3}{8}$. This completes the proof.

A.10 Proof of Lemma S1

The likelihood function (5) gives

$$\begin{aligned} n^{-\frac{1}{2}}\varphi_{11,m}^{\frac{1}{2}}\ell(t_{\lambda,m}) &= \frac{C}{\sqrt{n\varphi_{11,m}^{-1}\widehat{\sigma}_{f'}^2(t_{\lambda,m})}} \exp\left(-\frac{\widehat{\mu}_{f'}(t_{\lambda,m})^2}{2\widehat{\sigma}_{f'}^2(t_{\lambda,m})}\right) \\ &= \frac{C}{\sqrt{n\varphi_{11,m}^{-1}\widehat{\sigma}_{f'}^2(t_{\lambda,m})}} \exp\left(-n^{1-2\beta}\varphi_{11,m}^{-1}\frac{\mu_{n,m}^2}{2\sigma_{n,m}^2}\right), \end{aligned}$$

where $\mu_{n,m}$ and $\sigma^2_{n,m}$ are defined in Theorem 1. Note that

$$\begin{aligned}\left|\frac{1}{\sqrt{n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m})}} - \frac{1}{|f_0''(t_m)|\sigma^*_m}\right| &\asymp ||f_0''(t_m)|\sigma^*_m - \sqrt{n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m})}| \\ &\asymp |f_0''(t_m)^2\sigma^{*2}_m - n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m})|.\end{aligned}$$

Substituting $f_0''(t_m)^2\sigma^{*2}_m = \sigma^2$ into the right side yields

$$\left|f_0''(t_m)^2\sigma^{*2}_m - n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m})\right| = \left|n\varphi^{-1}_{11,m}\widehat{\sigma}^2_{f'}(t_{\lambda,m}) - \sigma^2\right| \lesssim n^{\frac{1}{2}-2\beta}(\log n)^{-1-a}.$$

This completes the proof.

B Additional simulation results

B.1 Effect of noise standard derivation and credible level

We carried out additional experiments to investigate the effect of noise standard derivation and credible levels $1-\alpha$. We used the same regression function shown in the paper and generated more noisy data by increasing the noise standard deviation σ from 0.1 to 0.2. As expected, results worsen, particularly for smaller sample sizes. This is because the GP tends to produce more wiggly curves. For example, looking at the percentages of correctly estimating M for $\alpha = .05$, calculated over 100 replicated datasets, we observed the following results: for $n = 100$ we obtained 19% and 85% for Beta (1,1) and Beta(2,3), respectively, versus 47% and 86% of Figure 3 in the paper; for $n = 500$ we obtained 52% and 94% for Beta (1,1) and Beta(2,3), respectively, versus 95% and 99% for $\sigma = 0.1$. We notice that, as already shown in the main simulation, the $Beta(2,3)$ prior and larger sample sizes help identifying the correct number of local extrema.

Next, we used this additional simulation study to investigate the performance of HPDR for different values of α. Results for sample sizes $n = 100$, $n = 500$ and $n = 1000$ and the two Beta priors are reported in the two tables below. For each combination of prior and sample size, we generated 100 simulated datasets.

Table S1: Beta$(1,1)$. Percentages of correctly estimated number of t's. The results are calculated on 100 simulated data.

Beta$(1,1)$	$\alpha = 0.001$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.03$	$\alpha = 0.05$	$\alpha = 0.1$
$n = 100$	56%	53%	51%	18%	19%	35%
$n = 500$	25%	32%	34%	43%	52%	60%
$n = 1000$	27%	35%	44%	57%	76%	84%

Table S2: Beta$(2,3)$. Percentages of correctly estimated number of t's. The results are calculated on 100 simulated data.

Beta$(2,3)$	$\alpha = 0.001$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.03$	$\alpha = 0.05$	$\alpha = 0.1$
$n = 100$	87%	88%	88%	86%	85%	77%
$n = 500$	95%	96%	95%	95%	94%	89%
$n = 1000$	95%	95%	96%	94%	94%	95%

In this additional study, we observed that increasing values of α did not necessarily correspond to larger estimated numbers of local extrema. This is because situations like the one shown in Figure S1 can occur. Therefore, larger or smaller α values do not necessarily imply more or fewer separated HPDR segments. Overall, results confirm the fairly robust estimation performance of the Beta(2,3) prior in estimating M.

B.2 Highly fluctuated regression function with large M

Upon suggestion from one of the reviewers, we performed a new simulation using the regression function $\sin(k\pi x)$ for $x \in [0,1]$, and assessed how the estimated number of local extrema converges to the true M. We considered $k = 10$ and $k = 100$ with varying n; with this regression function, the true number of local extrema is $M = k$. Other simulation configurations mirrored the main paper's setup, including the noise standard deviation, observed x values, and the number of replications. The proposed method is implemented using the same settings as in the simulation study in the main paper, unless otherwise stated.

We observe that when $k = 10$, our method is able to correctly estimate M 77% of the time even with sample size as small as 30. This percentage increases steadily to (93%, 99%, 100%) as n increases to $(200, 300, 500)$, respectively.

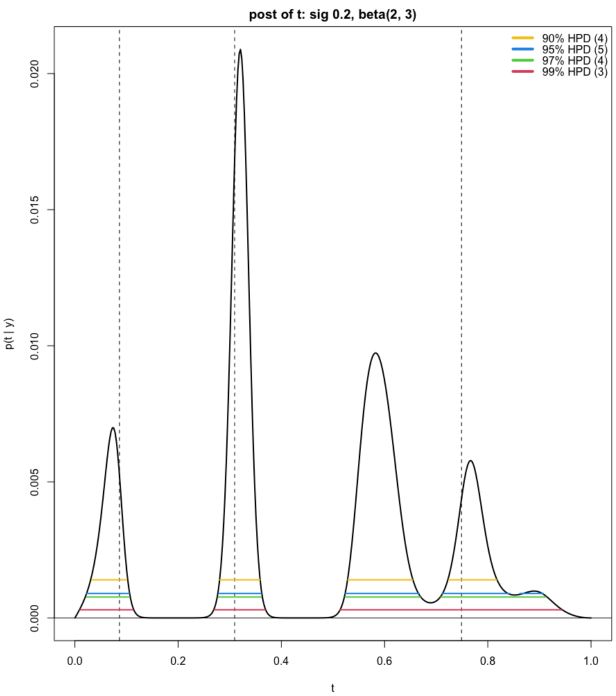

Figure S1: Effect of α on the estimated number of local extrema. The posterior density function is based on one simulated dataset with $n = 100$.

When $k = 100$, M is correctly estimated only 4% of the time when $n = 300$ (compared to 99% when $k = 10$), indicating the challenge of large $k = 100$. We have looked into this challenging scenario and found that for this highly fluctuated function, even simpler tasks such as function estimation become challenging. For example, the model struggles to distinguish between a highly fluctuated function and a flat function when $n = 300$, which is not surprising as indicated in the top plot of Figure S2. This has prompted us to find an effective strategy for this challenging function in which we incorporate the shape of the function into guided hyperparameter tuning. If we have prior knowledge that there are many local extrema, we can confine the hyperparameter searching space, ruling out some basins of the marginal likelihood that do not result in the regression shape being interested. For example, setting the upper bound when searching for (h, λ) to (0.1, 0.0001) as opposed to (10000, 10000) used in our default implementation, leads to the results reported in Table S3, which show a substantially improved estimation of M. For example, the proposed method can estimate the correct value of M with $n = 300$ in all 100 simulations. The posterior distribution of t in one simulation when $k = 100$ is shown in Figure S2. In this simulation, which is typical across 100 replications, our method correctly identifies the number and location of 100 local extrema. We acknowledge that prior information on the shape of the unknown function might not always be available.

	70-79	80-89	90-99	100	> 100
$n = 200$	18	76	6	0	0
$n = 225$	0	1	14	85	0
$n = 250$	0	0	0	99	1
$n = 300$	0	0	0	100	0

Table S3: Frequency of $\hat{M}$ falling in each interval when $k = 100$. Results are based on 100 repeated simulations.

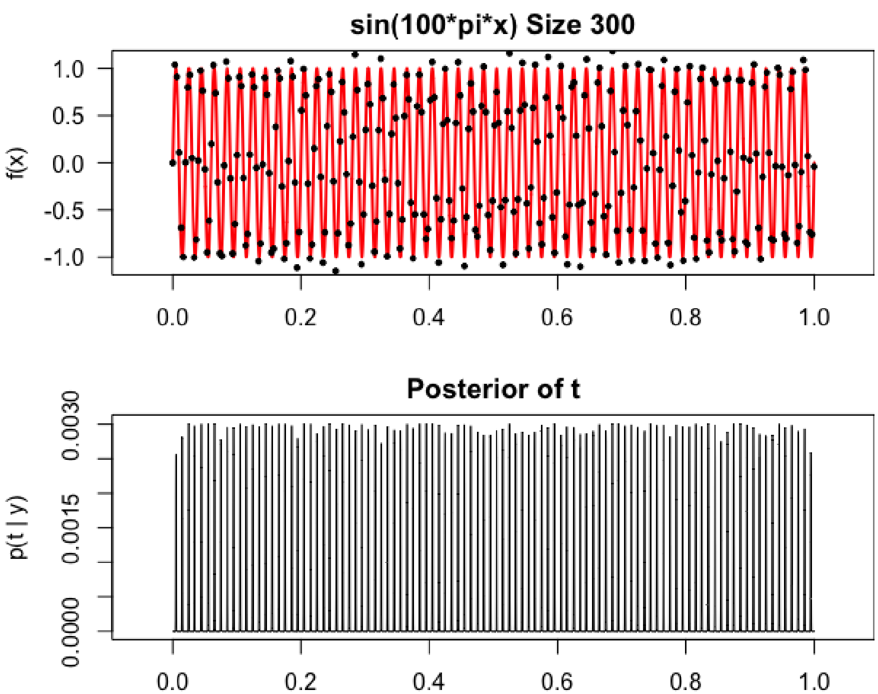

Figure S2: Data (top) and the posterior density of t (bottom) when $f(x) = \sin(100\pi x)$ (red curve in the top plot). Results are based on one simulated dataset with sample size $n = 300$.

References

Carreira-Perpinán, M. A. and Williams, C. K. (2003). On the number of modes of a Gaussian mixture. In *International Conference on Scale-Space Theories in Computer Vision*, pages

625–640. Springer.

Devroye, L., Mehrabian, A., and Reddad, T. (2018). The total variation distance between high-dimensional Gaussians with the same mean. *arXiv preprint arXiv:1810.08693*.

Liu, Z. and Li, M. (2023). On the estimation of derivatives using plug-in kernel ridge regression estimators. *Journal of Machine Learning Research*, 24(266):1–37.