Some Theory of Likelihood

Monday 15^{th} January, 2018

1 To be done

- Give the theory of posterior Bayes factor under exponential family.
- Beyond exponential family. (Berger et al., 2003).
- Neyman-Scott problems.
- Bartlett correction.
- General integral likelihood ratio test.
- Nonasymptotic. Read Spokoiny (2012)'s paper.
- Consider the sparse case as in Stadler and Mukherjee (2017).

2 Introduction

3 Results for exponential family

The content of this section is adapted from Ghosal (2000).

The following result, known as acute angle principle, is a key tool for the analysis.

Lemma 1 (J. M. Ortega (1987), Theorem 6.3.4.)**.** *Let C be an open, bounded set in $\mathbb{R}^n$ and assume that $F : \bar{C} \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ is countinuous and satisfies $(x - x_0)^T F(x) \geq 0$ for some $x_0 \in C$ and all $x \in \partial C$. Then $F(x) = 0$ has a solution in $\bar{C}$.*

We make the following assumptions.

Assumption 1. *The p dimensional independent random samples $X_1, \ldots, X_n$ are from a standard exponential family with density*

$$f(x; \theta_n) = \exp[x^T \theta_n - \psi_n(\theta_n)]$$

with respect to μ_n. Where $\theta_n \in \Theta_n$, an ***open*** *subset of $\mathbb{R}^n$. Sometimes we suppress the subscript n.*

The true parameter is denoted by θ_0. To prevent θ_0 approaching the boundary as $n \to \infty$, we assume that for a fixed $\epsilon_0 > 0$ independent of n, $B(\theta_0, \epsilon_0) \subset \Theta$.

It's well known that $\operatorname{E} X_1 = \psi'(\theta_0)$ and $\operatorname{Var} X_1 = \psi''(\theta_0)$. $\psi''(\theta_0)$ is also the Fisher information matrix. We assume that $\psi''(\theta_0)$ is positive definite.

Assumption 2. *$p \to \infty$ as $n \to \infty$.*

Let the positive definite matrix $\mathbf{J}$ be the square root of $\psi''(\theta_0)$, that is $\psi''(\theta_0) = \mathbf{J}^2$. The MLE $\hat{\theta}$ of θ is unique and satisfies $\psi'(\hat{\theta}) = \bar{X}$.

For a square matrix $\mathbf{A}$, $\|\mathbf{A}\|$ will stand for its operator norm.

The function $\psi(\theta)$ is in fact the cumulant generating function of X_1. Portnoy (1988) gave the following Taylor series expansions.

Proposition 1. *Suppose Assumption 1 holds. For any θ and θ_0 in Θ, the following expansions hold for some $\tilde{\theta}$ between θ and θ_0:*

$$\begin{aligned}
\psi(\theta) =& \psi(\theta_0) + (\theta-\theta_0)^T\psi'(\theta_0) + \frac{1}{2}(\theta-\theta_0)^T\psi''(\theta_0)(\theta-\theta_0) \\
&+ \frac{1}{6}\operatorname{E}_{\theta_0}\Big((\theta-\theta_0)^T(U-\operatorname{E}_{\theta_0}U)\Big)^3 \\
&+ \frac{1}{24}\Big\{\operatorname{E}_{\tilde{\theta}}\Big((\theta-\theta_0)^T(U-\operatorname{E}_{\tilde{\theta}}U)\Big)^4 - 3\big[\operatorname{E}_{\tilde{\theta}}\big((\theta-\theta_0)^T(U-\operatorname{E}_{\tilde{\theta}}U)\big)^2\big]^2\Big\},
\end{aligned}$$

$$\begin{aligned}
\alpha^T\psi'(\theta) =& \alpha^T\psi'(\theta_0) + \alpha^T\psi''(\theta_0)(\theta-\theta_0) \\
&+ \frac{1}{2}\operatorname{E}_{\theta_0}\big((\theta-\theta_0)^T(U-\operatorname{E}_{\theta_0}U)\big)^2\alpha^T(U-\operatorname{E}_{\theta_0}U) \\
&+ \frac{1}{6}\operatorname{E}_{\tilde{\theta}}\big((\theta-\theta_0)^T(U-\operatorname{E}_{\tilde{\theta}}U)\big)^3\alpha^T(U-\operatorname{E}_{\tilde{\theta}}U) \\
&- \frac{1}{2}\operatorname{E}_{\tilde{\theta}}\big((\theta-\theta_0)^T(U-\operatorname{E}_{\tilde{\theta}}U)\big)^2\operatorname{E}_{\tilde{\theta}}(\theta-\theta_0)^T(U-\operatorname{E}_{\tilde{\theta}}U)\alpha^T(U-\operatorname{E}_{\tilde{\theta}}U),
\end{aligned}$$

$$\alpha^T\psi'(\theta) = \alpha^T\psi'(\theta_0) + \alpha^T\psi''(\theta_0)(\theta-\theta_0) + \frac{1}{2}\operatorname{E}_{\tilde{\theta}}\big((\theta-\theta_0)^T(U-\operatorname{E}_{\tilde{\theta}}U)\big)^2\alpha^T(U-\operatorname{E}_{\tilde{\theta}}U),$$

where U is a random variable with density $f(x;\theta)$, α is a fixed p dimensional vector.

Let, for $c \geq 0$,

$$\begin{aligned}
B_{1n}(c) &= \sup\Big\{\operatorname{E}_\theta|a^T\mathbf{J}^{-1}(U-\operatorname{E}_\theta U)|^3 : \|a\|=1, \|\mathbf{J}(\theta-\theta_0)\| \leq c\Big\}, \\
B_{2n}(c) &= \sup\Big\{\operatorname{E}_\theta|a^T\mathbf{J}^{-1}(U-\operatorname{E}_\theta U)|^4 : \|a\|=1, \|\mathbf{J}(\theta-\theta_0)\| \leq c\Big\},
\end{aligned}$$

Consistency.

Theorem 1. *Suppose Assumption 1 holds. Assume that for every $M > 0$, we have $\{\theta : \|\mathbf{J}(\theta-\theta_0)\| \leq M\sqrt{p/n}\} \subset \Theta$ for large n. Assume that for all $M > 0$, $M\sqrt{p/n}B_{1n}(M\sqrt{p/n}) \to 0$. Then*

$$\|\mathbf{J}(\hat{\theta}-\theta_0)\| = O_P(\sqrt{p/n}).$$

Proof. The MLE $\hat{\theta}$ is unique and satisfies $\bar{X} - \psi'(\hat{\theta}) = 0$. By Lemma 1, the inequality

$$\sup_{\|\mathbf{J}(\theta-\theta_0)\|=c}(\theta-\theta_0)^T(\bar{X}-\psi'(\theta)) \leq 0$$

implies $\|\mathbf{J}(\hat\theta-\theta_0)\|\le c$. By proposition 1, for θ satisfying $\|\mathbf{J}(\theta-\theta_0)\|=c$, we have

$$\begin{aligned}
(\theta-\theta_0)^T(\bar X-\psi'(\theta)) &= (\theta-\theta_0)^T(\bar X-\psi'(\theta_0))-(\theta-\theta_0)^T\psi''(\theta_0)(\theta-\theta_0)-\frac{1}{2}\,\mathrm{E}_{\tilde\theta}\left((\theta-\theta_0)^T(U-\mathrm{E}_{\tilde\theta}U)\right)^3\\
&=(\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))-c^2-\frac{1}{2}\,\mathrm{E}_{\tilde\theta}\left((\theta-\theta_0)^T(U-\mathrm{E}_{\tilde\theta}U)\right)^3\\
&\le c\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|-c^2-\frac{1}{2}\,\mathrm{E}_{\tilde\theta}\left((\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right)^3\\
&\le c\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|-c^2+\frac{1}{2}c^3B_{1n}(c).
\end{aligned}$$

Since $\mathrm{E}\,\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|^2=\mathrm{tr}\,\mathrm{Var}(\mathbf{J}^{-1}\bar X)=p/n$, we have $\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|=O_P(\sqrt{p/n})$. Hence for every $\delta>0$, there is an M such that $\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|\le M\sqrt{p/n}$ with probability at least $1-\delta$. Taking $c=2M\sqrt{p/n}$ yields that with probability at least $1-\delta$,

$$\sup_{\|\mathbf{J}(\theta-\theta_0)\|=M\sqrt{p/n}}(\theta-\theta_0)^T(\bar X-\psi'(\theta))\le -2M^2\frac{p}{n}+2M^2\frac{p}{n}\Big(2M\sqrt{p/n}B_{1n}(2M\sqrt{p/n})\Big),$$

which is less than 0 eventually. Hence for large n, with probability at least $1-\delta$, $\|\mathbf{J}(\theta-\theta_0)\|\le M\sqrt{p/n}$. This proves $\|\mathbf{J}(\theta-\theta_0)\|=O_P(\sqrt{p/n})$.

□

Proposition 2. *Suppose that Assumption 1 holds. Let $\hat\theta$ be the MLE. Then on the event $\{\hat\theta\in\Theta\}$, for any $\alpha\in\mathbb{R}^p$, we have*

$$\begin{aligned}
\alpha^T\mathbf{J}(\hat\theta-\theta_0)=&\alpha^T\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))-\frac{1}{2}\,\mathrm{E}_{U,\theta_0}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^2\alpha^T\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\\
&+O(1)\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\alpha\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|),
\end{aligned}$$

where $|O(1)|\le 2/3$.

Proof. The MLE $\hat\theta$ satisfies $\alpha^T\mathbf{J}^{-1}\bar X=\alpha^T\mathbf{J}^{-1}\psi'(\hat\theta))$. Applying Proposition 1 yields

$$\begin{aligned}
\alpha^T\mathbf{J}^{-1}\bar X=&\alpha^T\mathbf{J}^{-1}\psi'(\hat\theta))\\
=&\alpha^T\mathbf{J}^{-1}\psi'(\theta_0)+\alpha^T\mathbf{J}(\hat\theta-\theta_0)\\
&+\frac{1}{2}\,\mathrm{E}_{U,\theta_0}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^2\alpha^T\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\\
&+\frac{1}{6}\,\mathrm{E}_{U,\tilde\theta}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right)^3\alpha^T\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\\
&-\frac{1}{2}\,\mathrm{E}_{U,\tilde\theta}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right)^2\mathrm{E}_{U,\tilde\theta}(\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\alpha^T\mathbf{J}^{-1}(U-\mathrm{E}_{\hat\theta}U).
\end{aligned}$$

For the second last term, we have

$$\begin{aligned}
&\mathrm{E}_{U,\tilde\theta}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right)^3\alpha^T\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\\
\le&\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\alpha\|\sup_{\|a\|=1,\|b\|=1}\left|a^T\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right|^3\left|b^T\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right|\\
\le&\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\alpha\|\sup_{\|a\|=1}\left|a^T\mathbf{J}^{-1}(U-\mathrm{E}_{\tilde\theta}U)\right|^4\\
\le&\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\alpha\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|).
\end{aligned}$$

The last term satisfies the same bound. This proves the proposition.

$\square$

Theorem 2. *Suppose that Assumption 1 holds. Let $\hat{\theta}$ be the MLE. Then on the event $\{\hat{\theta} \in \Theta\}$, for any $\alpha \in \mathbb{R}^p$, we have*

$$\left|\sqrt{n}\alpha^T \mathbf{J}(\hat{\theta}-\theta_0)-\sqrt{n}\alpha^T \mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\right| \leq \frac{\|\alpha\|}{2}\sqrt{n}\|\mathbf{J}(\hat{\theta}-\theta_0)\|^2 B_{1n}(0)+\frac{2\|\alpha\|}{3}\sqrt{n}\|\mathbf{J}(\hat{\theta}-\theta_0)\|^3 B_{2n}(\|\mathbf{J}(\hat{\theta}-\theta_0)\|).$$

Furthermore, we have

$$\begin{aligned}\left\|\sqrt{n}\mathbf{J}(\hat{\theta}-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\right\|^2 \leq & 2n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\| B_{2n}(\|\mathbf{J}(\hat{\theta}-\theta_0)\|) \\ & + 2n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^4 B_{2n}(\|\mathbf{J}(\hat{\theta}-\theta_0)\|).\end{aligned}$$

Proof. The first assertion follows directly from Proposition 2. Next we prove the second assertion. Write

$$\begin{aligned}&\left\|\sqrt{n}\mathbf{J}(\hat{\theta}-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\right\|^2 \\ =& n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^2 - 2n(\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0)) + n\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\|^2.\end{aligned}$$

For the first term, using Proposition 2 with $\alpha = n\mathbf{J}(\hat{\theta}-\theta_0)$, we have

$$\begin{aligned}n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^2 =& n(\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0)) - \frac{n}{2}\operatorname{E}_{U,\theta_0}\left((\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\operatorname{E}_{\theta_0}U)\right)^3 \\ & + O(1)n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^4 B_{2n}(\|\mathbf{J}(\hat{\theta}-\theta_0)\|).\end{aligned}$$

Therefore,

$$\begin{aligned}&\left\|\sqrt{n}\mathbf{J}(\hat{\theta}-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\right\|^2 \\ =& -n(\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0)) + n\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\|^2 \\ & - \frac{n}{2}\operatorname{E}_{U,\theta_0}\left((\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\operatorname{E}_{\theta_0}U)\right)^3 + O(1)n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^4 B_{2n}(\|\mathbf{J}(\hat{\theta}-\theta_0)\|).\end{aligned}$$

For the first term, using Proposition 2 with $\alpha = n\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))$, we have

$$\begin{aligned}& n(\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0)) \\ =& n\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\|^2 \\ & - \frac{n}{2}\operatorname{E}_{U,\theta_0}\left((\hat{\theta}-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\operatorname{E}_{\theta_0}U)\right)^2(\bar{X}-\psi'(\theta_0))^T\mathbf{J}^{-2}(U-\operatorname{E}_{\theta_0}U) \\ & + O(1)n\|\mathbf{J}(\hat{\theta}-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\| B_{2n}(\|\mathbf{J}(\hat{\theta}-\theta_0)\|).\end{aligned}$$

Thus,

$$
\begin{aligned}
&\left\|\sqrt{n}\mathbf{J}(\hat\theta-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\right\|^2\\
=&\frac{n}{2}\,\mathrm{E}_{U,\theta_0}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^2(\bar X-\psi'(\theta_0))^T\mathbf{J}^{-2}(U-\mathrm{E}_{\theta_0}U)\\
&+O(1)n\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
&-\frac{n}{2}\,\mathrm{E}_{U,\theta_0}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^3+O(1)n\|\mathbf{J}(\hat\theta-\theta_0)\|^4B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
=&\frac{n}{2}\,\mathrm{E}_{U,\theta_0}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^2\left(\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))-\mathbf{J}(\hat\theta-\theta_0)\right)^T\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\\
&+O(1)n\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
&+O(1)n\|\mathbf{J}(\hat\theta-\theta_0)\|^4B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
\le&\frac{n}{2}\left(\mathrm{E}_{U,\theta_0}\left((\hat\theta-\theta_0)^T\mathbf{J}\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^4\right)^{1/2}\left(\mathrm{E}_{U,\theta_0}\left(\left(\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))-\mathbf{J}(\hat\theta-\theta_0)\right)^T\mathbf{J}^{-1}(U-\mathrm{E}_{\theta_0}U)\right)^2\right)^{1/2}\\
&+\frac{2}{3}n\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
&+\frac{2}{3}n\|\mathbf{J}(\hat\theta-\theta_0)\|^4B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
\le&\frac{1}{2}\left(n\|\mathbf{J}(\hat\theta-\theta_0)\|^4B_{2n}(0)\right)^{1/2}\left(n\left\|\left(\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))-\mathbf{J}(\hat\theta-\theta_0)\right\|^2\right)^{1/2}\\
&+\frac{2}{3}n\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)\\
&+\frac{2}{3}n\|\mathbf{J}(\hat\theta-\theta_0)\|^4B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|).
\end{aligned}
$$

Let $\epsilon=n\|\mathbf{J}(\hat\theta-\theta_0)\|^3\|\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\|B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)+n\|\mathbf{J}(\hat\theta-\theta_0)\|^4B_{2n}(\|\mathbf{J}(\hat\theta-\theta_0)\|)$. Then

$$\left\|\sqrt{n}\mathbf{J}(\hat\theta-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\right\|^2\le\frac{1}{2}\sqrt{\epsilon}\left\|\sqrt{n}\mathbf{J}(\hat\theta-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\right\|+\frac{2}{3}\epsilon.$$

Thus $\left\|\sqrt{n}\mathbf{J}(\hat\theta-\theta_0)-\sqrt{n}\mathbf{J}^{-1}(\bar X-\psi'(\theta_0))\right\|^2\le 2\epsilon$. □

Next we consider the asymptotic normality of posterior distribution. Let $\pi(\theta)$ be the prior density with respect to Lebesgue measure. Then the posterior density of θ is given by

$$\frac{f(x;\theta)\pi(\theta)}{\int f(x;\theta)\pi(\theta)\,d\theta}$$

Put $u=\sqrt{n}\mathbf{J}(\theta-\theta_0)$. The likelihood ratio, as a function of u, is given by

$$Z_n(u)=\frac{f(x;\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{f(x;\theta_0)}.$$

And the posterior density of u is given by

$$\pi^*(u)=\frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)\,du}.$$

Let $\Delta_n = \sqrt{n}\mathbf{J}^{-1}(\bar{X}-\mu)$. We have

$$
\begin{aligned}
\log Z(u) =& \sqrt{n}\bar{X}^T\mathbf{J}^{-1}u - n(\psi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)-\psi(\theta_0)) \\
=& \Delta_n^T u - n\big(\psi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)-\psi(\theta_0)-n^{-1/2}\mu^T\mathbf{J}^{-1}u\big) \\
=& \Delta_n^T u - \frac{1}{2}\|u\|^2 - n\big(\psi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)-\psi(\theta_0)-n^{-1/2}\mu^T\mathbf{J}^{-1}u - n^{-1}\frac{1}{2}\|u\|^2\big).
\end{aligned}
$$

If the smaller order terms can be omitted, then the posterior density of u is approximately $\phi_p(u;\Delta_n,\mathbf{I}_p)$, where $\phi_p(\cdot;\mu,\Sigma)$ stands for the density of $N_p(\mu,\Sigma)$. The following theorem makes this assertion rigorous.

Theorem 3 (Asymptotic normality of posterior distribution)**.** *Suppose Assumption 1 holds. Let C be a quantity satisfying $C \gg \sqrt{p}$. Suppose that for large n, $\{\theta : \|\mathbf{J}(\theta-\theta_0)\| \le n^{-1/2}C\} \subset \Theta$. Suppose that $\frac{1}{3}\big(\frac{1}{n^{1/2}}CB_{1n}(0)+\frac{1}{n}C^2B_{2n}(n^{-1/2}C)\big) \le 1/2$ for sufficiently large n. Then for any $\epsilon > 0$, for sufficiently large n, with probability larger than $1-\epsilon$,*

$$
\begin{aligned}
&\int |\pi^*(u) - \phi_p(u;\Delta_n,\mathbf{I}_p)|\,du \\
\le& \Big|\exp\Big\{\frac{1}{6}\big(\frac{1}{n^{1/2}}C^3B_{1n}(0)+\frac{1}{n}C^4B_{2n}(n^{-1/2}C)\big)\Big\}-1\Big| \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C}\frac{\pi(\theta)}{\pi(\theta_0)} \\
&+ \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C}\Big|\frac{\pi(\theta)}{\pi(\theta_0)}-1\Big| \\
&+\exp\Big\{\frac{p}{2}\log\frac{n}{2\pi}+\frac{1}{2}\log|\psi''(\theta_0)|\Big\}\int_{\|\mathbf{J}(\theta-\theta_0)\|>n^{-1/2}C}\exp\Big\{-\frac{\sqrt{n}}{4}C\|\mathbf{J}(\theta-\theta_0)\|\Big\}\frac{\pi(\theta)}{\pi(\theta_0)}\,d\theta \\
&+\exp\big(-\frac{1}{4}\big(C-(1/\sqrt{\epsilon}+1)\sqrt{p}\big)^2\big).
\end{aligned}
$$

Proof. Let $\tilde{Z}_n(u) = \exp[\Delta_n^T u - \frac{1}{2}\|u\|^2]$. Note that $\phi_p(u;\Delta_n,\mathbf{I}_p) = \tilde{Z}_n(u)\pi(\theta_0)/\int\tilde{Z}_n(u)\pi(\theta_0)\,du$. We have

$$
\begin{aligned}
&\int |\pi^*(u)-\phi_p(u;\Delta_n,\mathbf{I}_p)|\,du = \int\Big|\frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int Z_n(w)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}w)\,dw}-\frac{\tilde{Z}_n(u)\pi(\theta_0)}{\int\tilde{Z}_n(w)\pi(\theta_0)\,dw}\Big|\,du \\
\le& \int\Big|\frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int Z_n(w)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}w)\,dw}-\frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int\tilde{Z}_n(w)\pi(\theta_0)\,dw}\Big|\,du \\
&+\int\Big|\frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int\tilde{Z}_n(w)\pi(\theta_0)\,dw}-\frac{\tilde{Z}_n(u)\pi(\theta_0)}{\int\tilde{Z}_n(w)\pi(\theta_0)\,dw}\Big|\,du \\
=&\Big|1-\frac{\int Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)\,du}{\int\tilde{Z}_n(u)\pi(\theta_0)\,du}\Big|+\frac{\int\big|Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)-\tilde{Z}_n(u)\pi(\theta_0)\big|\,du}{\int\tilde{Z}_n(u)\pi(\theta_0)\,du} \\
\le&\Big|1-\frac{\int Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)\,du}{\int\tilde{Z}_n(u)\pi(\theta_0)\,du}\Big|+\frac{\int\big|Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)-\tilde{Z}_n(u)\pi(\theta_0)\big|\,du}{\int\tilde{Z}_n(u)\pi(\theta_0)\,du} \\
\le&2\frac{\int\big|Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)-\tilde{Z}_n(u)\pi(\theta_0)\big|\,du}{\int\tilde{Z}_n(u)\pi(\theta_0)\,du} \\
=&2\int\Big|\exp\Big\{\log Z_n(u)-\frac{p}{2}\log(2\pi)-\frac{1}{2}\|\Delta_n\|^2\Big\}\frac{\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}-\phi_p(u;\Delta_n,\mathbf{I}_p)\Big|\,du
\end{aligned}
$$

We split the integral into the region $\|u\| \le C$ and $\|u\| > C$, where C will be specified latter. Then

$$
\begin{aligned}
&\int \Big| \exp\Big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u; \Delta_n, \mathbf{I}_p) \Big| \, du \\
\le &\int_{\|u\|\le C} \Big| \exp\Big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u; \Delta_n, \mathbf{I}_p) \Big| \, du \\
&+ \int_{\|u\|>C} \exp\Big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} \, du + \int_{\|u\|>C} \phi_p(u; \Delta_n, \mathbf{I}_p) \, du
\end{aligned}
\tag{1}
$$

We deal the three terms of (1) separately. Consider the first term. For $\|u\| \le C$, we have

$$
\begin{aligned}
&\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\|u - \Delta_n\|^2 \\
&- n\Big(\frac{1}{6n^{3/2}} \operatorname{E}_{\theta_0} \left(u^T \mathbf{J}^{-1}(U - \operatorname{E}_{\theta_0} U) \right)^3 + O(1)\frac{1}{n^2}\|u\|^4 B_{2n}(n^{-1/2}\|u\|) \Big).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\Big| \Big(\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big) - \Big(-\frac{p}{2}\log(2\pi) - \frac{1}{2}\|u - \Delta_n\|^2 \Big) \Big| \\
\le &\frac{1}{6}\Big(\frac{1}{n^{1/2}}\|u\|^3 B_{1n}(0) + \frac{1}{n}\|u\|^4 B_{2n}(n^{-1/2}\|u\|) \Big) \\
\le &\frac{1}{6}\Big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C) \Big).
\end{aligned}
\tag{2}
$$

It follows that

$$
\begin{aligned}
&\int_{\|u\|\le C} \Big| \exp\Big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u; \Delta_n, \mathbf{I}_p) \Big| \, du \\
\le &\int_{\|u\|\le C} \Big| \exp\Big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big\} - \phi_p(u; \Delta_n, \mathbf{I}_p) \Big| \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} \, du \\
&+ \int_{\|u\|\le C} \Big| \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - 1 \Big| \phi_p(u; \Delta_n, \mathbf{I}_p) \, du \\
\le &\Big| \exp\Big\{ \frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C) \big) \Big\} - 1 \Big| \int_{\|u\|\le C} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} \phi_p(u; \Delta_n, \mathbf{I}_p) \, du \\
&+ \int_{\|u\|\le C} \Big| \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - 1 \Big| \phi_p(u; \Delta_n, \mathbf{I}_p) \, du \\
\le &\Big| \exp\Big\{ \frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C) \big) \Big\} - 1 \Big| \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C} \frac{\pi(\theta)}{\pi(\theta_0)} \\
&+ \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C} \Big| \frac{\pi(\theta)}{\pi(\theta_0)} - 1 \Big|.
\end{aligned}
$$

Next we deal with the last term of (1). Note that $\operatorname{E}\Delta_n = \mathbf{0}_p$ and $\operatorname{Var}\Delta_n = \mathbf{I}_p$. By Chebyshev's inequality, for $\epsilon > 0$, there is an $M = 1/\sqrt{\epsilon}$ such that

$$
\sup_n \Pr(\|\Delta_n\| \ge M\sqrt{p}) < \epsilon.
$$

Denote $\mathcal{A} = \{\|\Delta_n\| \le M\sqrt{p}\}$. On the event $\mathcal{A}$, for $M_1 > 0$,

$$\int_{\|u\|>(M+1)\sqrt{p}+M_1} \phi_p(u;\Delta_n,\mathbf{I}_p)\,du \le \int_{\|u\|>M_1+\sqrt{p}} \phi_p(u;\mathbf{0}_p,\mathbf{I}_p)\,du \le \exp\big(-\frac{1}{4}M_1^2\big).$$

Hence for large n such that $C > (M+1)\sqrt{p}$, we have

$$\int_{\|u\|>C} \phi_p(u;\Delta_n,\mathbf{I}_p)\,du \le \exp\big(-\frac{1}{4}\big(C-(M+1)\sqrt{p}\big)^2\big).$$

Now we deal with the second term of (1). For $\|u\| \ge C$, by the concavity of $\log Z_n(u)$, we have that

$$(1-\frac{C}{\|u\|})\log Z_n(0) + \frac{C}{\|u\|}\log Z_n(u) \le \log Z_n\big(\frac{C}{\|u\|}u\big).$$

Hence

$$\log Z_n(u) \le \frac{\|u\|}{C}\log Z_n(\frac{C}{\|u\|}u).$$

This, combined with (2), yields

$$\begin{aligned}
\log Z_n(u) \le & \frac{\|u\|}{C}\Big(-\frac{1}{2}\big\|\frac{C}{\|u\|}u-\Delta_n\big\|^2 + \frac{1}{2}\|\Delta_n\|^2 + \frac{1}{6}\big(\frac{1}{n^{1/2}}C^3B_{1n}(0) + \frac{1}{n}C^4B_{2n}(n^{-1/2}C)\big)\Big) \\
= & -\frac{1}{2}C\|u\| + \Delta_n^T u + \frac{\|u\|}{C}\frac{1}{6}\big(\frac{1}{n^{1/2}}C^3B_{1n}(0) + \frac{1}{n}C^4B_{2n}(n^{-1/2}C)\big).
\end{aligned}$$

Hence on the event $\mathcal{A}$, for sufficiently large n, we have

$$\begin{aligned}
& \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \\
\le & -\frac{p}{2}\log(2\pi) - \frac{1}{2}C\|u\| + M\sqrt{p}\|u\| + \frac{\|u\|}{C}\frac{1}{6}\big(\frac{1}{n^{1/2}}C^3B_{1n}(0) + \frac{1}{n}C^4B_{2n}(n^{-1/2}C)\big) \\
= & -\frac{p}{2}\log(2\pi) - \frac{1}{2}C\|u\|\Big(1-\frac{2M}{C}\sqrt{p} - \frac{1}{3}\big(\frac{1}{n^{1/2}}CB_{1n}(0) + \frac{1}{n}C^2B_{2n}(n^{-1/2}C)\big)\Big) \\
\le & -\frac{p}{2}\log(2\pi) - \frac{1}{4}C\|u\|.
\end{aligned}$$

Hence the second term of (1) can be bounded by

$$\begin{aligned}
& \int_{\|u\|>C} \exp\big\{\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2\big\}\frac{\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}\,du \\
\le & \int_{\|u\|>C} \exp\Big\{-\frac{p}{2}\log(2\pi) - \frac{1}{4}C\|u\|\Big\}\frac{\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}\,du \\
= & \int_{\|\mathbf{J}(\theta-\theta_0)\|>n^{-1/2}C} \exp\Big\{-\frac{p}{2}\log(2\pi) - \frac{\sqrt{n}}{4}C\|\mathbf{J}(\theta-\theta_0)\|\Big\}\frac{\pi(\theta)}{\pi(\theta_0)}n^{p/2}|\mathbf{J}|\,d\theta \\
= & \exp\Big\{\frac{p}{2}\log\frac{n}{2\pi} + \frac{1}{2}\log|\psi''(\theta_0)|\Big\}\int_{\|\mathbf{J}(\theta-\theta_0)\|>n^{-1/2}C} \exp\Big\{-\frac{\sqrt{n}}{4}C\|\mathbf{J}(\theta-\theta_0)\|\Big\}\frac{\pi(\theta)}{\pi(\theta_0)}\,d\theta.
\end{aligned}$$

This proves the theorem. $\square$

4 Bayes consistency

For the models more general than the exponential families, the tail behavior of the likelihood is hard to control. As a result, Bayes consistency is not trivial. We consider the general case. Suppose that we observe a random sample $X_1, \ldots, X_n$ from a distribution P_0 with densitu p relative to some reference measure μ on the sample space $(\mathbb{X}, \mathbb{A})$. Let P_0^n denote the expectation with respect to $X_1, \ldots, X_n$. Let μ^n denote the n-fold product measure of μ. Let p^n denote the density of P^n with respect to μ^n. Let $\mathbf{X}^n = (X_1, \ldots, X_n)$ be the pooled data. Suppose the model space is $\mathcal{P}$. Given some prior distribution Π on the set $\mathcal{P}$, the posterior distribution is the random measure given by

$$\Pi(B|X_1, \ldots, X_n) = \frac{\int_B \prod_{i=1}^n p(X_i)\, d\Pi_n(P)}{\int \prod_{i=1}^n p(X_i)\, d\Pi_n(P)}. \tag{3}$$

To prove the consistency result, i.e., the posterior probability of $\{P : d(P_0, P) > \epsilon\}$ ($d(\cdot, \cdot)$ is certain distance) tends to 0, we need to lower bound the denominator of (3) and upper bound the numerator of (3). There is a commonly used method for lower bounding the denominator. The following lemma is adapted from Ghosal et al. (2000) and Shen and Wasserman (2001). Let

$$D_{KL}(P||Q) = P \log(dP/dQ), \quad V(P||Q) = \mathrm{Var}_P\big(\log(dP/dQ)\big).$$

Lemma 2. *Let $\alpha > 0$ and $\epsilon > 0$. Let*

$$A_\epsilon = \{P : D_{KL}(P_0, P) \le \epsilon,\, V(P_0||P) \le \epsilon\}.$$

Then for every prior probability measure Π and every $C > 0$, we have

$$P_0^n\Big(\int_{\mathscr{P}} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(P) < \Pi(A_\epsilon)\exp\big(-(1+C)n\epsilon\big)\Big) \le \frac{\alpha^2}{C^2 n\epsilon}$$

Proof. Without loss of generality, we assume $\Pi(A_\epsilon) > 0$. Let Π_ϵ be the restriction of Π on A_ϵ.

Then

$$
\begin{aligned}
&P_0^n\Big(\int_{\mathscr{P}}\Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^{\alpha}\,d\Pi(P) < \Pi(A_\epsilon)\exp\big(-(1+C)n\epsilon\big)\Big)\\
\le& P_0^n\Big(\log\int_{\mathscr{P}}\Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^{\alpha}\,d\Pi_\epsilon(P) < -(1+C)n\epsilon\Big)\\
\le& P_0^n\Big(\int\alpha\log\frac{p^n}{p_0^n}(\mathbf{X}^n)\,d\Pi_\epsilon(P) < -(1+C)n\epsilon\Big)\\
=& P_0^n\Big(\sum_{i=1}^n\int\log\frac{p}{p_0}(X_i)\,d\Pi_\epsilon(P) < -(1+C)n\epsilon/\alpha\Big)\\
\le& P_0^n\Big(\sum_{i=1}^n\int\log\frac{p}{p_0}(X_i)+D_{KL}(P_0||P)\,d\Pi_\epsilon(P) < -Cn\epsilon/\alpha\Big)\\
\le&\frac{\alpha^2}{C^2n^2\epsilon^2}nP_0\Big(\int\log\frac{p}{p_0}+D_{KL}(P_0||P)\,d\Pi_\epsilon(P)\Big)^2\\
\le&\frac{\alpha^2}{C^2n\epsilon^2}P_0\int\Big(\log\frac{p}{p_0}+D_{KL}(P_0||P)\Big)^2\,d\Pi_\epsilon(P)\\
=&\frac{\alpha^2}{C^2n\epsilon^2}\int P_0\Big(\log\frac{p}{p_0}+D_{KL}(P_0||P)\Big)^2\,d\Pi_\epsilon(P)\\
=&\frac{\alpha^2}{C^2n\epsilon^2}\int V(P_0||P)\,d\Pi_\epsilon(P)\le\frac{\alpha^2}{C^2n\epsilon}.
\end{aligned}
$$

□

The hard part is the numerator. Shen and Wasserman (2001) directly upper bound $p^n/p_0^n(X)$ to upper bound the numerator. Ghosal (2000) imposed a test condition to upper bound the numerator. If no additional assumption is adopted, the numerator can not be bounded. In fact, there are counterexamples, see Diaconis and Freedman (1986).

4.1 The work of Walker and Hjort (2001)

While the numerator of the posterior distribution is hard to control, a variation of posterior distribution is easier to control. This work is done by Walker and Hjort (2001).

For density f_1 and f_2, let

$$H(f_1,f_2)=\Big(\int(\sqrt{f_1}-\sqrt{f_2})^2\,d\mu\Big)^{1/2}=\Big(2-2\int\sqrt{f_1f_2}\,d\mu\Big)^{1/2},$$

the Hellinger distance of f_1 and f_2. For $0<\alpha<1$, Hellinger integral is defined as

$$\rho_\alpha(f_1,f_2)=\int_{\mathcal{X}}f_1^{\alpha}f_2^{1-\alpha}\,d\mu.$$

For $0<\alpha<1$, define the pseudoposterior distribution Q based on Π as

$$Q^n(A)=\frac{\int_A\big[p^n(\mathbf{X}^n)\big]^{\alpha}\,d\Pi_n(P)}{\int_{\mathscr{P}}\big[p^n(\mathbf{X}^n)\big]^{\alpha}\,d\Pi_n(P)}.$$

For $0 < t < 1$, define $\rho_t(\theta_1, \theta_2) = \int_{\mathcal{X}} p(X|\theta_1)^t p(X|\theta_2)^{1-t}\, d\mu$. Define

$$Z_t(A) = \int_A \Big[p_n(\mathbf{X}^{(n)}|\theta)\Big]^t \pi(\theta)\, d\theta.$$

Theorem 4. *Suppose $\Pi(A_\epsilon) > 0$ for every $\epsilon > 0$. Then for every $\epsilon > 0$ and $C > 0$,*

$$P_0^n \left\{ \frac{Z_t(\{\theta : \rho_t(\theta, \theta_0) \le 1-\epsilon\})}{Z_t(\Theta)} \right\} \le \Pi(A_{\frac{\epsilon}{2(1+C)}})^{-1} \exp(-\frac{1}{2}\epsilon n) + \frac{2(1+C)\alpha^2}{C^2 n\epsilon}.$$

Proof. Note that

$$\frac{Z_t(\{\theta : \rho_t(\theta, \theta_0) \le 1-\epsilon\})}{Z_t(\Theta)} = \frac{\int_{\{\theta:\rho_t(\theta,\theta_0)\le 1-\epsilon\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta}{\int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta}.$$

Consider the expactation of the numerator. It follows from Fubini's theorem that

$$\begin{aligned}
&P_0^n \int_{\{\theta:\rho_t(\theta,\theta_0)\le 1-\epsilon\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta \\
=&\int_{\{\theta:\rho_t(\theta,\theta_0)\le 1-\epsilon\}} \left\{ \int_{\mathcal{X}^n} \Big[p_n(\mathbf{X}^{(n)}|\theta)\Big]^t \Big[p_n(\mathbf{X}^{(n)}|\theta_0)\Big]^{1-t} d\mu^n \right\} \pi(\theta)\, d\theta \\
=&\int_{\{\theta:\rho_t(\theta,\theta_0)\le 1-\epsilon\}} \big[\rho_t(\theta, \theta_0)\big]^n \pi(\theta)\, d\theta \\
\le& \exp(-\epsilon n).
\end{aligned}$$

Consider the denominator. From Lemma 2, for every $\epsilon' > 0$, there exists a set $B_{\epsilon'}$ with $P_0^n(B_{\epsilon'}) > 1 - \alpha^2/(C^2 n\epsilon')$ on which

$$\int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta \ge \Pi(A_{\epsilon'}) \exp\big(-(1+C)\epsilon' n\big).$$

Thus,

$$\begin{aligned}
&P_0^n \Big\{ \frac{Z_t(\{\theta : \rho_t(\theta, \theta_0) \le 1-\epsilon\})}{Z_t(\Theta)} \Big\} \\
\le& P_0^n \Big\{ \mathbf{1}_{B_{\epsilon'}} \frac{Z_t(\{\theta : \rho_t(\theta, \theta_0) \le 1-\epsilon\})}{Z_t(\Theta)} \Big\} + P_0^n(B_{\epsilon'}^C) \\
\le& \Pi(A_{\epsilon'})^{-1} \exp\big[-\epsilon n + (1+C)\epsilon' n\big] + \frac{\alpha^2}{C^2 n\epsilon'}.
\end{aligned}$$

Now taking $\epsilon' = \frac{\epsilon}{2(1+C)}$ yields

$$P_0^n \Big\{ \frac{Z_t(\{\theta : \rho_t(\theta, \theta_0) \le 1-\epsilon\})}{Z_t(\Theta)} \Big\} \le \Pi(A_{\frac{\epsilon}{2(1+C)}})^{-1} \exp\big[-\frac{1}{2}\epsilon n\big] + \frac{2(1+C)\alpha^2}{C^2 n\epsilon}.$$

∎

One deficit of the theorem is that it does not satisfactorily cover finite-dimensional models. When applied to such models, it would yield the rate $1/\sqrt{n}$ times a logarithmic factor rather than $1/\sqrt{n}$ itself.

Next we consider finite-dimensional models. Let $\{p_\theta : \theta \in \Theta\}$ be a family of densities parametrized by a Euclidean parameter θ running through an open set $\Theta \subset \mathbb{R}^p$. Assume that for every $\theta, \theta_1, \theta_2 \in \Theta$ and some $\alpha > 0$, there exists positive constants C_1, C_2, C_3, C_4, such that

$$\begin{aligned}
&D_{KL}(p_{\theta_0}||p_\theta) \leq C_1\|\theta - \theta_0\|^{2\alpha}\\
&V(p_{\theta_0}||p_\theta) \leq C_1\|\theta - \theta_0\|^{2\alpha}\\
&C_3\|\theta_1 - \theta_2\|^{2\alpha} \leq 1 - \rho_\alpha(p_\theta, p_{\theta_0}) \leq C_4\|\theta_1 - \theta_2\|^{2\alpha}
\end{aligned}$$

The C_4 seems useless, and we can assume that the third inequality only holds locally. The proof of the following theorem is similar to the corresponding nonparametric one.

Theorem 5. *Under the conditions listed previously and θ_0 interior to Θ, then for any $M_n \to \infty$,*

$$P_0^n\Big\{Q^n\big(\|\theta - \theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}\big)\Big\} \to 0.$$

Proof. Without loss of generality, we assume $\frac{M_n}{n^{\frac{1}{2\alpha}}} \to 0$, otherwise we replace M_n by a smaller one. Consider the expactation of the numerator,

$$\begin{aligned}
&P_0^n \int_{\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}} \Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi_n(\theta)\\
&= \int_{\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}} \int_{\mathcal{X}^n} \Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha p_0^n(\mathbf{X}^n)\, d\mu^n\, d\Pi_n(\theta)\\
&= \int_{\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}} \int_{\mathcal{X}^n} \Big[p_\theta^n(\mathbf{X}^n)\Big]^\alpha \Big[p_0^n(\mathbf{X}^n)\Big]^{1-\alpha} d\mu^n\, d\Pi_n(\theta)\\
&= \int_{\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}} \big(\rho_\alpha(p_\theta, p_{\theta_0})\big)^n d\Pi_n(\theta)\\
&= \sum_{j=1}^{+\infty} \int_{\frac{jM_n}{n^{\frac{1}{2\alpha}}} \leq \|\theta-\theta_0\| \leq \frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}} \big(\rho_\alpha(p_\theta, p_{\theta_0})\big)^n d\Pi_n(\theta)\\
&\leq \sum_{j=1}^{+\infty} \int_{\frac{jM_n}{n^{\frac{1}{2\alpha}}} \leq \|\theta-\theta_0\| \leq \frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}} \big(1 - C_3(\frac{jM_n}{n^{\frac{1}{2\alpha}}})^{2\alpha}\big)^n d\Pi_n(\theta)\\
&\lesssim \sum_{j=1}^{+\infty} \Big[\frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}\Big]^p \exp[-C_3\Big(\frac{jM_n}{n^{\frac{1}{2\alpha}}}\Big)^{2\alpha} n]\\
&= \sum_{j=1}^{+\infty} \Big[\frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}\Big]^p \exp[-C_3\Big(jM_n\Big)^{2\alpha}]
\end{aligned}$$

Consider the denominator. From Lemma 2, on a set B with $P_0^n(B) > 1 - \alpha^2/(C^2 n\epsilon')$, we have

$$\int_\Theta \Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(\theta) \geq \Pi(A_{\epsilon'}) \exp\big(-(1+C)\epsilon' n\big) \geq \Pi\Big(\Big\{\theta : \|\theta-\theta_0\| \leq (\epsilon'/C_1)^{\frac{1}{2\alpha}}\Big\}\Big) \exp\big(-(1+C)\epsilon' n\big).$$

Let $\epsilon' = \frac{C_3 M_n^{2\alpha}}{2(1+C)n}$, we have

$$\int_\Theta \Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(\theta) \gtrsim \Big[\frac{M_n}{n^{\frac{1}{2\alpha}}}\Big]^p \exp\big(-\frac{1}{2}C_3 M_n^{2\alpha}\big).$$

Hence

$$\begin{aligned}
&P_0^n\Big\{Q^n\big(\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}\big)\Big\} \\
\leq &P_0^n\Big\{\mathbf{1}_B Q^n\big(\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}\big)\Big\} + P_0^n(B^C) \\
\leq &\sum_{j=1}^{+\infty}(j+1)^p \exp[-\frac{1}{2}C_3\Big(jM_n\Big)^{2\alpha}] + \frac{2(1+C)\alpha^2}{C^2 C_3 M_n^{2\alpha}} \to 0.
\end{aligned}$$

□

Appendices

Appendix A haha1

Appendix B haha2

References

References

Berger, J. O., Ghosh, J. K., and Mukhopadhyay, N. (2003). Approximations and consistency of bayes factors as model dimension grows. *Journal of Statistical Planning and Inference*, 112:241–258.

Diaconis, P. and Freedman, D. (1986). On the consistency of bayes estimates. *Annals of Statistics*, 14(1):1–26.

Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *Journal of Multivariate Analysis*, 74(1):49–68.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531.

J. M. Ortega, W. C. R. (1987). *Iterative Solution of Nonlinear Equations in Several Variables.* Classics in Applied Mathematics, 30. Society for Industrial Mathematics.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Annals of Statistics*, 16(1):356–366.

Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Annals of Statistics*, 29(3):687–714.

Spokoiny, V. (2012). Parametric estimation. finite sample theory. *Annals of Statistics*, 40(6):2877–2909.

Stadler, N. and Mukherjee, S. (2017). Two-sample testing in high dimensions. *Journal of The Royal Statistical Society Series B-statistical Methodology*.

Walker, S. G. and Hjort, N. L. (2001). On bayesian consistency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):811–821.