Some Theory of Likelihood

Wednesday 20th December, 2017

1 To be done

- Understand existing theory in exponential family. Some paper to be read: Portnoy (1988); Ghosal (2000).
- Give the theory of posterior Bayes factor under exponential family.
- Beyond exponential family. (**?**).
- 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 $\mathrm{E}\, X_1 = \psi'(\theta_0)$ and $\mathrm{Var}\, X_1 = \psi''(\theta_0)$. $\psi''(\theta_0)$ 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 $\|\mathbf{J}(\hat{\theta}-\theta_0)\| \leq c$. By proposition 1, for θ satisfying $\|\mathbf{J}(\theta-\theta_0)\| = c$, we have

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

Since $\operatorname{E}\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\|^2 = \operatorname{tr}\operatorname{Var}(\mathbf{J}^{-1}\bar{X}) = p/n$, 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 $\|\mathbf{J}^{-1}(\bar{X}-\psi'(\theta_0))\| \leq M\sqrt{p/n}$ with probability at least $1-\delta$. Taking $c = 2M\sqrt{p/n}$ yields that with probability at least $1-\delta$,

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

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

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

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

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

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

For the second last term, we have

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

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*

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

Furthermore, we have

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

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

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, $\{\theta : \|\mathbf{J}(\theta - \theta_0)\| \leq n^{-1/2}C\} \subset \Theta$. Suppose that $\frac{1}{3}\big(\frac{1}{n^{1/2}}CB_{1n}(0) + \frac{1}{n}C^2B_{2n}(n^{-1/2}C)\big) \leq 1/2$ for sufficiently large n. Then for any $\epsilon > 0$, for sufficiently large n, with probability larger than $1 - \epsilon$,*

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

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

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

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

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

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

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

Hence

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

It follows that

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

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

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

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

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

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

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

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

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

Hence

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

This, combined with (2), yields

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

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

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

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

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

This proves the theorem. □

Appendices

Appendix A haha1

Appendix B haha2

References

References

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.

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.

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*.