# Notes on the theory of Bayesian statistics

Rui Wang

Friday 24th May, 2019

**Abstract**

This document provides notes on the theory of Bayesian statistics.

## 1 Consistency

Nonparametric convergence rate. The work of [6] is seminal. A simlar result was obtained by [8]. The work of [6] focused on iid case. [9] generalized the results to non iid case. This line of research relies on the assumption that there exists a sequence of uniformly consistent test.

## 2 Results in [9]

For each $n \in \mathbb{N}$ and $\theta \in \Theta$, let $P_\theta^{(n)}$ admit densities $p_\theta^{(n)}$ relative to a $\sigma$-finite measure $\mu^{(n)}$. Assume that $(x, \theta) \mapsto p_\theta^{(n)}(x)$ is jointly measurable relative to $\mathscr{A} \otimes \mathscr{B}$, where $\mathscr{B}$ is a $\sigma$-field on $\Theta$. The posterior distribution is given by

$$\Pi_n(B|X^{(n)}) = \frac{\int_B p_\theta^{(n)}(X^{(n)}) \,\mathrm{d}\Theta_n(\theta)}{\int_\Theta p_\theta^{(n)}(X^{(n)}) \,\mathrm{d}\Theta_n(\theta)}, \quad B \in \mathscr{B}.$$

Here $X^{(n)}$ is generated according to $P_{\theta_0}^{(n)}$ for some given $\theta_0 \in \Theta$.

**Assumption 1.** *For each $n$, let $d_n$ and $e_n$ be semimetrics on $\Theta$ with the property that there exist universal constants $\xi > 0$ and $K > 0$ such that for every $\epsilon > 0$ and for each $\theta_1 \in \Theta$ with $d_n(\theta_1, \theta_0) > \epsilon$, there exists a test $\phi_n$ such that*

$$P_{\theta_0}^{(n)} \phi_n \le e^{-Kn\epsilon^2}, \qquad \sup_{\theta \in \Theta : e_n(\theta, \theta_1) < \epsilon\xi} P_\theta^{(n)}(1 - \phi_n) \le e^{-Kn\epsilon^2}.$$

**Lemma 1.** *Suppose Assumption 1 holds. Suppose that for some nonincreasing function $\epsilon \mapsto N(\epsilon)$ and some $\epsilon_n \ge 0$,*

$$N\left(\frac{\epsilon\xi}{2}, \{\theta \in \Theta : d_n(\theta, \theta_0) < \epsilon\}, e_n\right) \le N(\epsilon) \quad \textit{for all } \epsilon > \epsilon_n.$$

*Then for every $\epsilon > \epsilon_n$, there exist tests $\phi_n$, $n \geq 1$, (depending on $\epsilon$) such that $P^{(n)}_{\theta_0}\phi_n \leq N(\epsilon)e^{-Kn\epsilon^2}/(1-e^{-Kn\epsilon^2})$ and $P^{(n)}_{\theta}(1-\phi_n) \leq e^{-Kn\epsilon^2 j^2}$ for all $\theta \in \Theta$ such that $d_n(\theta,\theta_0) > j\epsilon$ and for every $j \in \mathbb{N}$.*

*Proof.* For a given $j \in \mathbb{N}$, let $\Theta_j = \{\theta \in \Theta : j\epsilon < d_n(\theta,\theta_0) \leq (j+1)\epsilon\}$ and choose a set $\Theta'_j \subset \Theta_j$ such that $\{B_{e_n}(\theta_{(j,i)}, j\epsilon\xi), i = 1,\ldots,|\Theta'_j|\}$ is a minimal $j\epsilon\xi$-covering. Since $\Theta_j \subset \{\theta \in \Theta : d_n(\theta,\theta_0) \leq 2j\epsilon\}$, we have

$$|\Theta'_j| \leq N\left(j\epsilon\xi, \{\theta \in \Theta : d_n(\theta,\theta_0) < 2j\epsilon\}, e_n\right) \leq N(2j\epsilon).$$

By assumption, for every point $\theta_{(j,i)} \in \Theta'_j$, there exists a test $\phi_n^{(j,i)}$ with the following properties

$$P^{(n)}_{\theta_0}\phi_n^{(j,i)} \leq e^{-Knj^2\epsilon^2}, \qquad \sup_{\theta \in B_{e_n}(\theta_{(j,i)}, j\epsilon\xi)} P^{(n)}_{\theta}(1-\phi_n^{(j,i)}) \leq e^{-Knj^2\epsilon^2}.$$

Let

$$\phi_n = \sup_{\{(j,i): i \in 1,\ldots,|\Theta'_j|, j \in \mathbb{N}\}} \phi_n^{(j,i)}.$$

Then

$$P^{(n)}_{\theta_0}\phi_n \leq \sum_{j=1}^{\infty}\sum_{i=1}^{|\Theta'_j|} P^{(n)}_{\theta_0}\phi_n^{(j,i)} \leq \sum_{j=1}^{\infty} |\Theta'_j| e^{-Knj^2\epsilon^2} \leq N(\epsilon)\sum_{j=1}^{\infty} e^{-Knj\epsilon^2} = N(\epsilon)\frac{e^{-Kn\epsilon^2}}{1-e^{-Kn\epsilon^2}}.$$

On the other hand, for any $\theta \in \Theta$ such that $d_n(\theta,\theta_0) > j\epsilon$, there exists $(j',i')$ such that $j' \geq j$, $\theta \in B_{e_n}(\theta_{(j,i)}, j\epsilon\xi)$.

$$P^{(n)}_{\theta}(1-\phi_n) \leq P^{(n)}_{\theta}(1-\phi_n^{(j',i')}) \leq \sup_{\theta \in B_{e_n}(\theta_{(j,i)}, j\epsilon\xi)} P^{(n)}_{\theta}(1-\phi_n^{(j',i')}) \leq e^{-Knj'^2\epsilon^2} \leq e^{-Knj^2\epsilon^2}.$$

This completes the proof.

□

**Corollary 1.** *If the conclution of Lemma 1 holds, then for $d_n(\theta,\theta_0) > \epsilon$,*

$$P^{(n)}_{\theta}(1-\phi_n) \leq \exp\left\{-\frac{K}{4}nd_n^2(\theta,\theta_0)\right\}.$$

*Proof.* Lemma 1 asserts that $P^{(n)}_{\theta}(1-\phi_n) \leq e^{-Kn\epsilon^2 j^2}$ for all $\theta \in \Theta$ such that $d_n(\theta,\theta_0) > j\epsilon$ and for every $j \in \mathbb{N}$. Then $P^{(n)}_{\theta}(1-\phi_n) \leq e^{-Kn\epsilon^2 j^2}$ for $\theta \in \Theta$ such that $j\epsilon < d_n(\theta,\theta_0) \leq (j+1)\epsilon$ for every $j \in \mathbb{N}$. Thus if $d_n(\theta,\theta_0) > \epsilon$,

$$\begin{aligned} P^{(n)}_{\theta}(1-\phi_n) &\leq e^{-Kn\epsilon^2 j^2} \\ &= e^{-Kn\epsilon^2(j+1)^2\frac{j^2}{(j+1)^2}} \\ &\leq e^{-\frac{K}{4}nd_n^2(\theta,\theta_0)} \end{aligned}$$

□

For a given $k > 1$, let

$$B_n(\theta_0, \epsilon; k) = \left\{\theta \in \Theta : K\left(p_{\theta_0}^{(n)}, p_\theta^{(n)}\right) \le n\epsilon^2, V_{k,0}\left(p_{\theta_0}^{(n)}, p_\theta^{(n)}\right) \le n^{k/2}\epsilon^k\right\},$$

where $V_{k,0}(f,g) = \int f|\log(f/g) - K(f,g)|^k \, \mathrm{d}\mu$.

**Lemma 2.** *For $k \ge 2$, every $\epsilon > 0$ and every probability measure $\bar{\Pi}_n$ supported on the set $B_n(\theta_0, \epsilon; k)$, we have, for every $C > 0$,*

$$P_{\theta_0}^{(n)}\left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \, \mathrm{d}\bar{\Pi}_n(\theta) \le e^{-(1+C)n\epsilon^2}\right) \le \frac{1}{C^k(n\epsilon^2)^{k/2}}$$

*Proof.* By Jenson's inequality applied to the logarithm, with $\ell_{n,\theta} = \log\left(p_\theta^{(n)}/p_{\theta_0}^{(n)}\right)$, we have $\log\int(p_\theta^{(n)}/p_{\theta_0}^{(n)}) \, \mathrm{d}\bar{\Pi}_n(\theta) \ge \int \ell_{n,\theta} \, \mathrm{d}\bar{\Pi}_n(\theta)$. Thus,

$$\begin{aligned}
&P_{\theta_0}^{(n)}\left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \, \mathrm{d}\bar{\Pi}_n(\theta) \le e^{-(1+C)n\epsilon^2}\right)\\
\le& P_{\theta_0}^{(n)}\left(\int \ell_{n,\theta} \, \mathrm{d}\bar{\Pi}_n(\theta) \le -(1+C)n\epsilon^2\right)\\
=& P_{\theta_0}^{(n)}\left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right) \mathrm{d}\bar{\Pi}_n(\theta) \le -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)}\ell_{n,\theta} \, \mathrm{d}\bar{\Pi}_n(\theta)\right)\\
=& P_{\theta_0}^{(n)}\left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right) \mathrm{d}\bar{\Pi}_n(\theta) \le -(1+C)n\epsilon^2 + \int K\left(p_{\theta_0}^{(n)}, p_\theta^{(n)}\right) \mathrm{d}\bar{\Pi}_n(\theta)\right)\\
\le& P_{\theta_0}^{(n)}\left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right) \mathrm{d}\bar{\Pi}_n(\theta) \le -Cn\epsilon^2\right)\\
\le& \frac{P_{\theta_0}^{(n)}\int |\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}|^k \, \mathrm{d}\bar{\Pi}_n(\theta)}{(Cn\epsilon^2)^k}\\
=& \frac{\int V_{k,0}(f,g) \, \mathrm{d}\bar{\Pi}_n(\theta)}{(Cn\epsilon^2)^k}\\
\le& \frac{1}{C^k(n\epsilon^2)^{k/2}}.
\end{aligned}$$

□

**Theorem 1.** *Suppose Assumption 1 holds. Let $\epsilon_n > 0$, $\epsilon_n \to 0$, $(n\epsilon_n^2)^{-1} = O(1)$, $k > 1$, and $\Theta_n \subset \Theta$ be such that,*

$$\sup_{\epsilon > \epsilon_n} \log N\left(\frac{1}{2}\epsilon\xi, \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}, e_n\right) \le n\epsilon_n^2, \tag{1}$$

*for some $C > 0$,*

$$\frac{e^{(1+C)n\epsilon_n^2}}{\Pi_n(B_n(\theta_0, \epsilon_n; k))} \int_{\{\theta \in \Theta_n : d_n(\theta,\theta_0) \ge M_n\epsilon_n\}} \exp\left\{-\frac{K}{4}nd_n^2(\theta, \theta_0)\right\} \mathrm{d}\Pi_n(\theta) \to 0. \tag{2}$$

*Then for every $M_n \to \infty$, we have that*

$$P_{\theta_0}^{(n)} \Pi_n \left( \theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n | X^{(n)} \right) \to 0.$$

*Proof.* From Corollary 1, applied with $N(\epsilon) = \exp\left(n\epsilon_n^2\right)$ and $\epsilon = M_n \epsilon_n$ (W.L.O.G $M_n \geq 1$) in its assertion, there exist tests $\phi_n$ that satisfy

$$P_{\theta_0}^{(n)} \phi_n \leq e^{n\epsilon_n^2} \frac{e^{-KnM_n^2\epsilon_n^2}}{1 - e^{-KnM_n^2\epsilon_n^2}},$$

$$P_\theta^{(n)}(1 - \phi_n) \leq \exp\left\{-\frac{K}{4} n d_n^2(\theta, \theta_0)\right\} \text{ for all } \theta \in \Theta_n \text{ s.t. } d_n(\theta, \theta_0) > M_n \epsilon_n.$$

The first assertion implies that if $M_n$ is sufficiently large to ensure that $KM_n^2 - 1 > KM_n^2/2$, then as $n \to \infty$, we have

$$P_{\theta_0}^{(n)} \left[ \Pi_n \left( \theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n | X^{(n)} \right) \phi_n \right] \leq P_{\theta_0}^{(n)} \phi_n \lesssim \exp\left\{-KM_n^2 n\epsilon_n^2/2\right\}.$$

Setting $\Theta_n^\dagger = \{\theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n\}$, we obtain, by Fubini's theorem,

$$\begin{aligned}
&P_{\theta_0}^{(n)} \left[ \int_{\{\theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n\}} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \, \mathrm{d}\Pi_n(\theta) \, (1 - \phi_n) \right] \\
&= \int_{\mathcal{X}^{(n)}} \int_{\{\theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n\}} \frac{p_\theta^{(n)}(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \, \mathrm{d}\Pi_n(\theta) \left(1 - \phi_n(X^{(n)})\right) p_{\theta_0}^{(n)}(X^{(n)}) \, \mathrm{d}\mu^{(n)} \\
&= \int_{\{\theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n\}} P_\theta^{(n)}(1 - \phi_n) \, \mathrm{d}\Pi_n(\theta) \\
&\leq \int_{\{\theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n \epsilon_n\}} \exp\left\{-\frac{K}{4} n d_n^2(\theta, \theta_0)\right\} \mathrm{d}\Pi_n(\theta)
\end{aligned}$$

Fix some $C > 0$. By Lemma 2, we have, on an event $A_n$ with probability at least $1 - C^{-k}(n\epsilon_n^2)^{-k/2}$,

$$\int_\Theta \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \, \mathrm{d}\Pi_n(\theta) \geq \int_{B_n(\theta_0, \epsilon_n; k)} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \, \mathrm{d}\Pi_n(\theta) \geq e^{-(1+C)n\epsilon_n^2} \Pi_n\left(B_n\left(\theta_0, \epsilon_n; k\right)\right).$$

Thus,

$$
\begin{aligned}
&P_{\theta_0}^{(n)}\left[\Pi_n\left(\theta\in\Theta_n: d_n(\theta,\theta_0)>\epsilon_n M_n|X^{(n)}\right)(1-\phi_n)\mathbf{1}_{A_n}\right]\\
=&P_{\theta_0}^{(n)}\left[\frac{\int_{\{\theta\in\Theta_n:d_n(\theta,\theta_0)\geq M_n\epsilon_n\}}\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\,\mathrm{d}\Pi_n(\theta)}{\int_\Theta\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\,\mathrm{d}\Pi_n(\theta)}(1-\phi_n)\mathbf{1}_{A_n}\right]\\
\leq&\frac{e^{(1+C)n\epsilon_n^2}}{\Pi_n\left(B_n\left(\theta_0,\epsilon_n;k\right)\right)}P_{\theta_0}^{(n)}\left[\int_{\{\theta\in\Theta_n:d_n(\theta,\theta_0)\geq M_n\epsilon_n\}}\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\,\mathrm{d}\Pi_n(\theta)(1-\phi_n)\mathbf{1}_{A_n}\right]\\
\leq&\frac{e^{(1+C)n\epsilon_n^2}}{\Pi_n\left(B_n\left(\theta_0,\epsilon_n;k\right)\right)}P_{\theta_0}^{(n)}\left[\int_{\{\theta\in\Theta_n:d_n(\theta,\theta_0)\geq M_n\epsilon_n\}}\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\,\mathrm{d}\Pi_n(\theta)(1-\phi_n)\right]\\
\leq&\frac{e^{(1+C)n\epsilon_n^2}}{\Pi_n\left(B_n\left(\theta_0,\epsilon_n;k\right)\right)}\int_{\{\theta\in\Theta_n:d_n(\theta,\theta_0)\geq M_n\epsilon_n\}}\exp\left\{-\frac{K}{4}nd_n^2(\theta,\theta_0)\right\}\,\mathrm{d}\Pi_n(\theta).
\end{aligned}
$$

This completes the proof.

□

## 3 Fractional posteriors

**Consistency:**

[11], [2], [Alquier and Ridgway] (variational fractional posterior, also contains an example of mixture model).

nonregular models: [5], [4], [3].

Martingale methods: [10], [12].

Fractional posterior with power $t$ is defined as

$$
\Pi^{(t)}(\theta\in B|\mathbf{X}^{(n)})=\frac{\int_B\left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t\Pi_n(\mathrm{d}\theta)}{\int_\Theta\left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t\Pi_n(\mathrm{d}\theta)}.
$$

For two parameters $\theta_1$ and $\theta_2$, the $\alpha$ order Rényi divergence ($0<\alpha<1$) of $P_{\theta_1}$ from $P_{\theta_2}$ (suppose $P_{\theta_1}\ll P_{\theta_2}$) is defined to be

$$
D_\alpha(\theta_1\|\theta_2)=-\frac{1}{1-\alpha}\log\rho_\alpha(\theta_1,\theta_2),
$$

where $\rho_\alpha(\theta_1,\theta_2)=\int_{\mathcal{X}}p(X|\theta_1)^\alpha p(X|\theta_2)^{1-\alpha}\mathrm{d}\mu$ is the so-called Hellinger integral.

The Kullback-Leibler between $P$ and $Q$ is

$$
D_1(P\|Q)=\begin{cases}\int_{\mathcal{X}}\log\frac{\mathrm{d}P}{\mathrm{d}Q}\mathrm{d}P, & \text{if } P\ll Q,\\ +\infty & \text{otherwise.}\end{cases}
$$

It is known that $\lim_{\alpha\uparrow1}D_\alpha(\theta_1\|\theta_2)=D_1(\theta_1\|\theta_2)$.

### 3.1 Results in [7], Section 8.6

**Lemma 3.** *Suppose $v$ are nonnegative measurable function from $\Theta$ to $\mathbb{R}$ such that*

$$0 < \int_\Theta v(\theta)\Pi(\mathrm{d}\theta) < \infty.$$

*Let*

$$\mathcal{W} = \left\{ w : w \text{ is a nonnegative measurable function from } \Theta \text{ to } \mathbb{R} \text{ s.t. } \int_\Theta w(\theta)\Pi(\mathrm{d}\theta) = 1 \right\}.$$

*Then*

$$\log \int_\Theta v(\theta)\Pi(\mathrm{d}\theta) = \sup_{w\in\mathcal{W}} \int_\Theta \left[-\log w(\theta) + \log v(\theta)\right] w(\theta)\Pi(\mathrm{d}\theta).$$

*The equality holds when*

$$w(\theta) = \frac{v(\theta)}{\int_\Theta v(\theta)\,\Pi(\mathrm{d}\theta)}.$$

**Remark 1.** This lemma seems well known. One can find it in [**?** ], Lemma 1.1.3.

*Proof.* The conclusion follows from

$$D_1\left( w\Pi(\mathrm{d}\theta) \middle\| \frac{v\Pi(\mathrm{d}\theta)}{\int_\Theta v\,\Pi(\mathrm{d}\theta)} \right) \geq 0.$$

$\square$

**Theorem 2.** *For any numbers $\epsilon > 0$, $\beta \in (0,1)$, $\gamma \geq 0$, for $\alpha = (\gamma+\beta)/(\gamma+1)$,*

$$\frac{1-\beta}{\gamma+1} P_{\theta_0}^{(n)} \int_\Theta D_\beta(\theta\|\theta_0)\Pi^{(\alpha)}(\mathrm{d}\theta|\mathbf{X}^{(n)}) \leq -\log \int_\Theta e^{-\alpha D_1(\theta_0\|\theta)}\Pi(\mathrm{d}\theta).$$

*Proof.* Apply Lemma 3 twice, first with $v(\theta) = (p_\theta/p_{\theta_0})^\delta(X)$, and second with $v(\theta) = (p_\theta/p_{\theta_0})^\beta(X)/\rho_\beta(\theta,\theta_0)$. We have

$$\int_\Theta \left( -\log w(\theta) + \delta \log \frac{p_\theta}{p_{\theta_0}}(X) \right) w(\theta)\Pi(\mathrm{d}\theta) \leq \log \int_\Theta \left(\frac{p_\theta}{p_{\theta_0}}\right)^\delta (X)\Pi(\mathrm{d}\theta),$$

$$\int_\Theta \left( -\log w(\theta) + \beta \log \frac{p_\theta}{p_{\theta_0}}(X) - \log \rho_\beta(\theta,\theta_0) \right) w(\theta)\Pi(\mathrm{d}\theta) \leq \log \int_\Theta \left(\frac{p_\theta}{p_{\theta_0}}\right)^\beta (X)\frac{1}{\rho_\beta(\theta,\theta_0)}\Pi(\mathrm{d}\theta).$$

Adding the second inequality to $\gamma$ times the first inequality yields

$$\begin{aligned}
&\int_\Theta \left( -(\gamma+1)\log w(\theta) + (\gamma\delta+\beta) \log \frac{p_\theta}{p_{\theta_0}}(X) - \log \rho_\beta(\theta,\theta_0) \right) w(\theta)\Pi(\mathrm{d}\theta) \\
\leq &\log \int_\Theta \left(\frac{p_\theta}{p_{\theta_0}}\right)^\beta (X)\frac{1}{\rho_\beta(\theta,\theta_0)}\Pi(\mathrm{d}\theta) + \gamma \log \int_\Theta \left(\frac{p_\theta}{p_{\theta_0}}\right)^\delta (X)\Pi(\mathrm{d}\theta).
\end{aligned}$$

Note that the right hand side does not depend on $w(\theta)$, and

$$\begin{aligned}
&P_{\theta_0}\log\int_\Theta\left(\frac{p_\theta}{p_{\theta_0}}\right)^\delta(X)\Pi(\mathrm{d}\theta)\\
\le&\log P_{\theta_0}\int_\Theta\left(\frac{p_\theta}{p_{\theta_0}}\right)^\delta(X)\Pi(\mathrm{d}\theta)=\log\int_\Theta P_{\theta_0}\left(\frac{p_\theta}{p_{\theta_0}}\right)^\delta(X)\Pi(\mathrm{d}\theta)\\
=&\log\int_\Theta P_{\theta_0}\left(\frac{p_\theta}{p_{\theta_0}}\right)^\delta(X)\Pi(\mathrm{d}\theta)=\log\int_\Theta\rho_\delta(\theta,\theta_0)\Pi(\mathrm{d}\theta)\\
\le&\log\int_\Theta\Pi(\mathrm{d}\theta)\le 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&P_{\theta_0}\log\int_\Theta\left(\frac{p_\theta}{p_{\theta_0}}\right)^\beta(X)\frac{1}{\rho_\beta(\theta,\theta_0)}\Pi(\mathrm{d}\theta)\\
\le&\log P_{\theta_0}\int_\Theta\left(\frac{p_\theta}{p_{\theta_0}}\right)^\beta(X)\frac{1}{\rho_\beta(\theta,\theta_0)}\Pi(\mathrm{d}\theta)=\log\int_\Theta P_{\theta_0}\left(\frac{p_\theta}{p_{\theta_0}}\right)^\beta(X)\frac{1}{\rho_\beta(\theta,\theta_0)}\Pi(\mathrm{d}\theta)\\
=&\log\int_\Theta P_{\theta_0}\left(\frac{p_\theta}{p_{\theta_0}}\right)^\beta(X)\frac{1}{\rho_\beta(\theta,\theta_0)}\Pi(\mathrm{d}\theta)=\log\int_\Theta\Pi(\mathrm{d}\theta)\le 0.
\end{aligned}$$

It follows that for every choice of $w(\theta)$ ($w(\theta)$ may depende on $X$), we have

$$P_{\theta_0}\int_\Theta\left(-(\gamma+1)\log w(\theta)+(\gamma\delta+\beta)\log\frac{p_\theta}{p_{\theta_0}}(X)-\log\rho_\beta(\theta,\theta_0)\right)w(\theta)\Pi(\mathrm{d}\theta)\le 0.$$

Note that

$$\int_\Theta-\log\rho_\beta(\theta,\theta_0)w(\theta)\Pi(\mathrm{d}\theta)\ge 0.$$

Then for any $w(\theta)$,

$$P_{\theta_0}\int_\Theta-\log\rho_\beta(\theta,\theta_0)w(\theta)\Pi(\mathrm{d}\theta)\le P_{\theta_0}\int_\Theta\left((\gamma+1)\log w(\theta)-(\gamma\delta+\beta)\log\frac{p_\theta}{p_{\theta_0}}(X)\right)w(\theta)\Pi(\mathrm{d}\theta).$$

Divided both sides by $\gamma+1$, we have

$$\frac{1}{\gamma+1}P_{\theta_0}\int_\Theta-\log\rho_\beta(\theta,\theta_0)w(\theta)\Pi(\mathrm{d}\theta)\le P_{\theta_0}\int_\Theta\left(\log w(\theta)-\log\left(\frac{p_\theta}{p_{\theta_0}}(X)\right)^\alpha\right)w(\theta)\Pi(\mathrm{d}\theta),$$

where $\alpha=\frac{\gamma\delta+\beta}{\gamma+1}$. From Lemma 3, the right hand side is minimized (for $w\in\mathcal{W}$ and $w$ can depends on $X$) when

$$w(\theta)=\frac{p_\theta^\alpha(X)}{\int_\Theta p_\theta^\alpha(X)\Pi(\mathrm{d}\theta)}.$$

And the minimum value of the right hand side is $-P_{\theta_0}\log\int_\Theta\left(\frac{p_\theta}{p_{\theta_0}}(X)\right)^\alpha\Pi(\mathrm{d}\theta)$. In this case,

$w(\theta)\Pi(\mathrm{d}\theta) = \Pi^\alpha(\mathrm{d}\theta|\mathbf{X}^{(n)})$. Thus,

$$
\begin{aligned}
&\frac{1}{\gamma+1} P_{\theta_0} \int_\Theta -\log \rho_\beta(\theta,\theta_0)\Pi^\alpha(\mathrm{d}\theta|\mathbf{X}^{(n)}) \\
\leq & - P_{\theta_0} \log \int_\Theta \left(\frac{p_\theta}{p_{\theta_0}}(X)\right)^\alpha \Pi(\mathrm{d}\theta) \\
= & P_{\theta_0} \inf_{w\in\mathcal{W}} \int_\Theta \left(\log w(\theta) - \log\left(\frac{p_\theta}{p_{\theta_0}}(X)\right)^\alpha\right) w(\theta)\Pi(\mathrm{d}\theta) \\
\leq & \inf_{w\in\mathcal{W}} P_{\theta_0} \int_\Theta \left(\log w(\theta) - \log\left(\frac{p_\theta}{p_{\theta_0}}(X)\right)^\alpha\right) w(\theta)\Pi(\mathrm{d}\theta) \\
= & \inf_{w\in\mathcal{W}} \int_\Theta \left(\log w(\theta) - \alpha P_{\theta_0} \log \frac{p_\theta}{p_{\theta_0}}(X)\right) w(\theta)\Pi(\mathrm{d}\theta) \\
= & \inf_{w\in\mathcal{W}} \int_\Theta \left(\log w(\theta) - \log e^{-\alpha D_1(\theta_0\|\theta)}\right) w(\theta)\Pi(\mathrm{d}\theta) \\
= & -\log \int_\Theta e^{-\alpha D_1(\theta_0\|\theta)}\Pi(\mathrm{d}\theta).
\end{aligned}
$$

This completes the proof.

□

## 4 Hausdorff entropy

[? ], [? ], [? ], [? ],

Their results can be generalized to non iid case.

## 5 $\rho$-estimator

Lucien Birgé and Yannick Baraud's work.

## References


[Alquier and Ridgway] Alquier, P. and Ridgway, J. Concentration of tempered posteriors and of their variational approximations.

[2] Bhattacharya, A., Pati, D., and Yang, Y. (2019). Bayesian fractional posteriors. *The Annals of Statistics*, 47(1):39–66.

[3] Bochkina, N. A. and Green, P. J. (2014). The Bernstein-von Mises theorem and nonregular models. *The Annals of Statistics*, 42(5):1850–1878.

[4] Dudley, R. M. and Haughton, D. (2002). Asymptotic normality with small relative errors of posterior probabilities of half-spaces. *The Annals of Statistics*, 30(5):1311–1344.

[5] Ghosal, S., Ghosh, J. K., and Samanta, T. (1995). On convergence of posterior distributions. *The Annals of Statistics*, 23(6):2145–2152.

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

[7] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference.* Cambridge University Press, 1st edition.

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

[9] van der Vaart, A. and Ghosal, S. (2007). Convergence rates of posterior distributions for noniid observations. *Annals of Statistics*, 35(1):192–223.

[10] Walker, S. G. (2004). New approaches to bayesian consistency. *Annals of Statistics*, 32(5):2028–2043.

[11] 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.

[12] Walker, S. G., Lijoi, A., and Prünster, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *The Annals of Statistics*, 35(2):738–746.