Stability and Convergence of the Posterior in Non-Regular Problems

Jayanta K. Ghosh*, Subhashis Ghosal and Tapas Samanta
Indian Statistical Institute, Calcutta

Abstract. It is shown that the posterior converges in a weak sense in a fairly general set up which includes the exponential, gamma, and Weibull with a location parameter, and a reliability change point problem. A representation is obtained for the limiting posterior in case a stronger convergence, e.g., almost sure convergence, holds. This leads to a necessary condition under the above general set up and settles the question of almost sure convergence in the above examples. Under very general conditions it is also shown that the posterior is asymptotically free of the prior. Though primarily developed for non-regular problems, all the theorems apply to regular cases also.

1 Introduction

In Bayesian analysis, one starts with a prior and the resultant analysis is based on the posterior, given data. Since this involves a prior, naturally we are interested in how to what extent the Bayesian analysis is sensitive to the choice of a prior.

Bhattacharyya (1971), under mild conditions the prior has little or no effect if sample size is large, so that almost same conclusions will follow from almost any reasonable prior, i.e., to have almost complete prior robustness.

The next natural question is whether the posterior stabilizes as the sample size increases indefinitely. In this case the inference stabilizes and we want to know whether the posterior approaches a simple form. If so, the Bayesian analysis is remarkably simple; the approximate computation based on the simplified form is often quite accurate, even for moderately large sample sizes. (For an example, see Berger (1985, p. 231).)

Also, if one is interested in studying the frequentist coverage probabilities of Bayesian confidence intervals, the job is very much simplified.

The problem has been well investigated in the so-called “regular” cases, where the posterior has been observed to be normal, centered at the maximum likelihood estimator (MLE), tends to the normal distribution. This fact was first observed by Laplace (1774) and Rayleigh (1875) and subsequently referred to as the Bernstein-von Mises Theorem or the Bayesian central limit theorem. Le Cam (1947, 1953) gave a rigorous proof of this result for independent and identically distributed (i.i.d.) observations. Various modifications and extensions of this have been made by several authors including Wedderburn and Yahav (1969), Walker and Chao (1967), Dawid (1970), Heide and Johnstone (1979), Kallianpur, Borodin and Prakasa Rao (1971), Clarke and Barron (1980). A detailed discussion on the conditions underlying the Bernstein-von Mises theorem can be found in Le Cam.

* Research was completed while the first author was visiting Purdue University.

183
Convergence of Posterior

2 Main Results

Let \( \{X^{(n)}, A^{(n)}, P^{(n)}, \theta \in \Theta \} \) be a sequence of statistical experiments by observation \( X^{(n)} \in A^{(n)} \) where \( \Theta \subset \mathbb{R}^k \) is a Borel set with non-empty \( \Theta \). We assume that for each \( n \geq 1 \), there exists a \( \sigma \)-finite measure \( \mu^{(n)} \) on \( A^{(n)} \) dominating the family \( \{P^{(n)}, \theta \in \Theta \} \) and let the Radon-Nikodym derivative

\[
p^{(n)}(x^{(n)}, \theta) = \frac{d\mu^{(n)}}{d\nu^{(n)}}(x^{(n)}).
\]

Let \( \pi(.) \) be a prior density (possibly improper) on \( \Theta \). The posterior densities, given \( X^{(n)} \), by Bayes' theorem, is given by

\[
\pi(\theta | X^{(n)}) = \frac{\pi(\theta)p^{(n)}(X^{(n)}, \theta)}{\int_{\Theta} \pi(\theta)p^{(n)}(X^{(n)}, \theta)d\theta}
\]

provided the integral converges. We assume that the posterior is non-trivial (i.e., sufficiently large) \( n \). The posterior probability of a set \( A \subset \Theta \) is denoted by

\[
\pi(\theta \in A | X^{(n)}) = \int_A \pi(\theta | X^{(n)})d\theta .
\]

The following theorem shows that under certain conditions, the posterior is insensitive to the choice of prior for large \( n \).

Theorem 2.1 Let \( \theta_0 \) be an interior point of \( \Theta \) and \( \pi_1, \pi_2 \) be two prior densities which are positive and continuous at \( \theta_0 \). We assume that the posterior \( \pi_i(\theta | X^{(n)}) \) is continuous for \( i = 1, 2 \), i.e., given any \( \eta > 0 \) and a neighborhood \( V \) of \( \theta_0 \), there exists \( n_0 \geq 1 \) such that for \( i = 1, 2 \)

\[
\pi_i(\theta | X^{(n)}) \geq 1 - \eta \text{ for all } n \geq n_0 \text{ a.s. (}\theta_0\).
\]

Thus

\[
\lim_{n \to \infty} \int \pi_1(\theta | X^{(n)}) - \pi_2(\theta | X^{(n)})d\theta = 0 \text{ a.s. (}\theta_0\).
\]

Proof: From consistency of posterior, it follows that for \( n \geq n_0 \)

\[
\int \pi_1(\theta | X^{(n)}) - \pi_2(\theta | X^{(n)})d\theta \leq \int \pi_1(\theta | X^{(n)})[1 - \pi_1(\theta | X^{(n)})]\pi_2(\theta | X^{(n)})d\theta + 2\eta.
\]

Let \( \eta > 0 \) be given. Then by continuity of \( \pi_i \), get a neighborhood \( V \) of \( \theta_0 \) so that on \( V \)

\[
\pi_i(\theta; \theta_0)(1 - \delta) \leq \pi_i(\theta) \leq \pi_i(\theta; \theta_0)(1 + \delta)
\]

and hence

\[
(1 - \delta)\pi_i(\theta; \theta_0)C_n \leq \int \pi_i(\theta)p^{(n)}(X^{(n)}, \theta)d\theta \leq (1 + \delta)\pi_i(\theta; \theta_0)C_n
\]

where

\[
C_n = \int \pi_i(\theta)p^{(n)}(X^{(n)}, \theta)d\theta.
\]

Using (2.6)

\[
(1 + \eta)[(1 - \delta)p^{(n)}(x^{(n)}, \theta) - \pi_i(\theta | X^{(n)})] \leq \pi_i(\theta | X^{(n)}) \leq \frac{(1 + \delta)}{(1 - \delta)C_n}p^{(n)}(x^{(n)}, \theta) .
\]

We have

\[
(1 - \eta)(1 - \delta) \leq \pi_i(\theta | X^{(n)}) \leq (1 - \eta)^{-1}(1 + \delta) .
\]

Adding \( \delta \) and \( \delta \) small enough and putting (2.8) in (2.5), we have the desired
Remark 2.1. If $F$ is a family of priors such that $\pi(\theta) > 0$ for all $\theta \in F$, $F$ is equicontinuous at $\theta_0$ and (2.5) is satisfied uniformly in $\theta \in F$, then
\[
\lim_{n \to \infty} \sup \left\{ \int [\pi_2(\theta)^n(X_0^n) - \pi_2(\theta_0)^n(X_0^n)]d\mu: \theta_1, \theta_2 \in F \right\} = 0 \quad a.s.
\] (2.12)

Remark 2.2. The assumption (2.3) in Theorem 2.1 about posterior consistency generally would appear to be a mild condition on the case of a finite dimensional parameter space. For this reason,Diaconis and Freedman (1980) and references therein. Conditions (2.3) of Theorem 2.1 holds if Conditions (III) below hold and $E[|\pi_n(t)|^s] < \infty$ for some $s > 0$. The latter condition holds in all common examples, where $(\pi_n)$ is a power of $n$.

We now investigate the existence of posterior limit, which, if it exists, is independent of the prior in view of Theorem 2.1. The first result, essentially due to Hira and Hasmukh (1981) (henceforth abbreviated as H.H) states that under general conditions, posterior probabilities of the normalized parameter converge weakly. The set up and assumptions are described below.

Let $(\pi_n)$ be a sequence of $k \times k$ positive definite matrices converging to $\pi_0$ and $\theta_1 \in \Theta$.

Let $\mathcal{U}_n$ be a neighborhood of 0 in $\mathbb{R}^k$ tending to $\mathbb{R}^k$. Define the "likelihood process"
\[
Z_n(u) = Z_{n,\theta_1}(u) = \frac{p_n(\theta_1)^n}{p_n(\theta_1)^n}\rho(\theta_1, \theta_1)
\]
considered as a random function in $u \in \mathcal{U}_n$. For studying various asymptotics, it is necessary to choose $\pi_n$ properly. See H.H (1981) in this context. Essentially unique choice of $\pi_n$ is shown.

Now onwards, expectations and probabilities refer to the 'true' prior $\theta_1$.

Conditions (III).

1. For some $\alpha > 0$, $K > 0$, $n > 0$, $n_0 \geq 1$
\[
\sup \{|u_1 - u_2|^{1/2} E_{u_1/2}(u_2) - E_{u_1/2}(u_1): |u_1| \leq R_1, |u_2| \leq R_2 \} \leq K(1 + |u_1|^{1/2})
\]
for all $n \geq n_0$.

2. For all $u \in \mathcal{U}_n, n \geq n_0$
\[
E_{u_1/2}(u) \leq \exp[-g_n(u)]
\]
where $(g_n)$ is a sequence of real valued functions on $[0, \infty)$ satisfying
(a) For fixed $n \geq 1$, $g_n(u) \to 0$ as $v \to \infty$
(b) For any $N > 0$, $\lim_{u \to \infty} g_N \exp[-g_n(u)] = 0$

3. Let $Z(u): u \in \mathbb{R}^k$ be a stochastic process not identically zero such that of $Z$.

**Notation:** Unless otherwise indicated, integral with respect to $u$ is taken over the whole set where integrand is defined. We also write
\[
\mathcal{L}(u) = \int \pi(\theta_1 + \pi_n u)Z(u)du \quad \text{and} \quad \mathcal{L}(u) = \int \pi(\theta_1 + \pi_n u)Z(u)du
\]

Theorem 2.2. Let Conditions (III) be satisfied and $\pi$ be a prior density continuous at $\theta_0$. Then for any Borel set $A \subset \mathbb{R}^k$,\n\[
\mathcal{L}(u) \in A(X_0^n) \Rightarrow \int_{A}(Z(u))du = \int_{A}(Z(u))du
\]

The proof is implicit in the proof of Theorem 1.1(2) of H.H (1981).

**Proof:** Let $A$ be any measurable subcollection of $\mathcal{B}_N$, the Borel $\sigma$-field in $\mathbb{R}^k$. Then by a slight modification of the proof, it can be shown that the $\mathbb{R}^m$-valued process $\int_{A}(Z(u))du: A \subset A$ converges in distribution to the $\mathbb{R}^m$-valued process $\int_{A}(Z(u))du: A \subset A$ under Conditions (III).

Let $P$ denote the space of all absolutely continuous probability on $\mathbb{R}^k$ equipped with the total variation norm. $P$ is isometrically identified with the space of all probability densities on $\mathbb{R}^k$ with $L^1$-norm.

The next result is somewhat technical and is a strengthened version of Theorem 2.2. This is used in Theorem 2.3 and is also of independent interest. Note that by Lemma A.4, one can consider $\mathcal{L}_n$ and $\mathcal{L}$ to be $P$-valued or $L^1(\mathbb{R}^k)$-valued random variables.

Theorem 2.3. Let $\mathcal{L}_n$ converge to the process $\mathcal{L}$ in $L^1(\mathbb{R}^k)$.

**Proof:** A view of Remark 2.3, it is enough to verify that $\{\mathcal{L}_n\}$ forms a tight family.

We use Lemma A.1. The first condition is trivial and it is enough to verify that
\[
\sup_{|u| \leq M} \left| \int_{|u| \leq M} (\mathcal{L}_n(u) - \mathcal{L}_n) \right| du = n \geq n_0, |u| < \delta
\]
and
\[
\sup_{|u| > M} \left| \int_{|u| > M} (\mathcal{L}_n(u) - \mathcal{L}_n) \right| du \geq n \geq n_0
\]
unfortunately small with probability arbitrarily close to one if one chooses $\delta$ small enough.

The above has been verified for (a) in H.H (1981). To verify this for (a) we note that
\[
\int_{|u| \leq M} (\mathcal{L}_n(u) - \mathcal{L}_n) \right| du = \int_{|u| \leq M} (Z_n(u) - Z_n(u)) \right| du = \int_{|u| \leq M} (Z_n(u) - Z_n(u)) \right| du
\]
and
\[
\int_{|u| > M} (Z_n(u) - Z_n(u)) \right| du = \int_{|u| > M} (Z_n(u) - Z_n(u)) \right| du = \int_{|u| > M} (Z_n(u) - Z_n(u)) \right| du
\]

(2.14)
by Cauchy-Schwartz inequality and the inequality \((a + b)^2 \leq 2(a^2 + b^2)\). Expression (2.14) is in turn less than or equal to
\[
4 \int Z_n(u)du^{-1/3} \left\{ \int_{\|u\| \leq \Delta} |Z_n(u + x) - Z_n(u)|^2/2 du \right\}^{1/2}.
\]
Now by Condition (II) and by Chebyshev's inequality we have
\[
P(\{|Z_n(u + x) - Z_n(u)| > \varepsilon\} \leq \frac{\varepsilon^2}{2}\int Z_n(u)du)^{-1/3}
\]
which can be made arbitrarily small whatever be \(\varepsilon\) and \(M\). Thus with high probability the inequality
\[
\int_{\|u\| \leq \Delta} |\varepsilon_n(u + x) - \varepsilon_n(u)|du \leq 4\varepsilon \sqrt{M} \int Z_n(u)du^{1/3}
\]
is satisfied. Clearly the last term can be made arbitrarily small uniformly in \(x\) if \(\varepsilon\) is small enough. Thus the result follows.

Remark 2.4. A result similar to Theorem 2.2 for the posterior with a random centering \(\delta\) also holds if \(\delta\) is such that
\[
(G_1^{-1}(\delta - \delta_0), \varepsilon_n(\delta_0)) \in (W_n, \xi_n(\delta_0)) \in \mathbb{R}^3 \times L^1(\mathbb{R}^3).
\]
In this case the posterior probability of a set is given by (2.33).

The above technical results give only the weak limit of posterior probability. However, we have familiar examples where the suitably centered posterior is a limit almost surely. In regular cases, the posterior centered at the MLE is a normal distribution almost surely under certain conditions. Therefore we investigate whether the posterior, with suitable centering, goes to a limit almost surely, or at least in probability. In the remaining portion of this section, we provide necessary conditions for the existence of such a centering under a general setup.

We confine ourselves only to the iid case where observation \(X^{(n)}\) is an \((X_1, \ldots, X_n)\) and \(p_n(X_1, \ldots, X_n, \theta) = p_n(X_1, \theta, \ldots, X_n, \theta)\) is a probability density with respect to a \(\sigma\)-finite measure \(\nu\) on a standard Borel space \((X, A)\).

The next result shows that a limit, if it exists, must be free of the sample.

Proposition 2.1. Let \(\theta = \theta(X_1, \ldots, X_n)\) and \(T = T(X_1, \ldots, X_n)\) be functions of \(X_1, \ldots, X_n\), which may or may not involve \(\theta\). Put \(\psi = \psi(T)\) let \(A \in \mathbb{B}^t\).

Suppose for each sample sequence \(X\), there exists \(c(X)\) such that
\[
\pi(\psi \in A|X^{(n)}|) \overset{P}{=} c(X).
\]
Then \(c(X)\) does not depend on the sample sequence \(X\).

Proof. The posterior density of \(\psi\) is
\[
\pi(\psi \in A|X^{(n)}) = \frac{\pi(\psi \in A|X^{(n)}\pi(\nu_0)})}{\pi(\nu_0)} f(x, \psi | X^{(n)}).
\]

This is a symmetric function of \(X_1, \ldots, X_n\) and so is
\[
\pi(\psi \in A|X^{(n)}) = \int_X \pi(\psi \in A|\psi^{(n)})d\psi.
\]
By going through a subsequence, if necessary, (2.16) can be assumed to hold a.s. By an application of the Hewitt-Savage zero-one law (Chow-Teicher, 1988), it follows that \(c(X)\) does not depend on \(X\).

Often it is true that
\[
\psi, T \overset{P}{=} \Sigma,
\]
where \(\Sigma\) is a p.d. matrix. In such cases, one may assume that \(T = \psi^{-1}\).

The following definition will be useful.

Definition 2.1. An \(\mathbb{R}^t\)-valued random variable \(\theta = \theta(X_1, \ldots, X_n)\), symmetric in its arguments, is called a proper centering if for each \(A \in \mathbb{B}^t\), there exists a non-random \(Q(A)\) such that
\[
\sup\{|\pi(\psi^{-1}(\theta - \theta) \in A|X^{(n)}) - Q(A)|: A \in \mathbb{B}^t| \overset{P}{=} 0.
\]

\(\psi^{-1}\) is called a semi-proper centering or wide sense proper centering if for each \(A \in \mathbb{B}^t\),
\[
\pi(\psi^{-1}(\theta - \theta) \in A|X^{(n)}) \overset{P}{=} Q(A).
\]

Remark 2.5. If (2.20) is satisfied, it automatically follows that \(Q(\cdot)\) is an absolutely continuous countably additive probability, and if (2.21) is satisfied, then \(Q(\cdot)\) is a finitely additive probability. Proposition 2.1 makes clear why \(Q(\cdot)\) must be non-null.

Proposition 2.2. Let \(\theta\) be a proper centering such that \(W_n = \psi^{-1}(\theta - \theta_0)\) converges weakly in a random variable \(W\). Then for any countable subcollection \(A \in \mathbb{B}^t\), the \(\mathbb{R}^t\) valued process \(\pi(\psi^{-1}(\theta - \theta) \in A|X^{(n)}): A \in \mathbb{A}\) converges weakly to the process \(\{\xi \in \mathbb{A}: A \in \mathbb{A}\}\).

Proof. By weak convergence theory in \(\mathbb{R}^t\), it is enough to prove the result for a finite collection \((A_1, \ldots, A_r)\). Since \(Q\) is absolutely continuous, the mapping
\[
x \mapsto (Q(A_1 - x), \ldots, Q(A_r - x))
\]
continuous by Lemma A.2 and hence
\[
\sup\{|Q(A_1 - W_n), \ldots, Q(A_r - W_n)| \leq (Q(A_1 - W), \ldots, Q(A_r - W))
\]
\[
\sup\{|\pi(\psi^{-1}(\theta - \theta) \in A|X^{(n)}) - Q(A - W)|: A \in \mathbb{B}^t| \overset{P}{=} 0.
\]

By Slutsky's theorem, the result now follows.
Proposition 2.3 Assume Conditions (III) and let \( \hat{\theta} \) be a proper centering. Then \( \psi_n^{-1}(\hat{\theta} - \theta_0) \) is weakly convergent.

**Proof.** We first show that \( W_n = \psi_n^{-1}(\hat{\theta} - \theta_0) \) is tight. If not, there exists \( \epsilon > 0 \) such that for any \( \lambda > 0 \), there is a subsequence \( \{n'\} \) of \( \{n\} \) for which

\[
P(||W_{n'}|| > \lambda) > \epsilon \quad \text{for all } n'
\]  

Put

\[
\begin{align*}
  u &= \psi_n^{-1}(\hat{\theta} - \theta_0) \\
  v &= \psi_n^{-1}(\hat{\theta} - \theta_0) 
\end{align*}
\]

Then

\[
\pi(v \in A|X^{(n)}) = \int_{A \times W_n} \xi_n(u)du.
\]  

Fix a bounded set \( A \) and using arguments of III (1981), find \( M \) large enough that for \( ||u|| > M \), \( \xi_n(u)du \) can be made as small as desired with probability close to 1 uniformly in \( n \geq n_0 \). Choose \( \lambda > 0 \) large enough such that for all \( \lambda \)

\[
A + \varepsilon \subset \{||u|| > M\}
\]

Combining (2.22) to (2.24), it follows that one must have \( Q(A) = 0 \). Clearly cannot be true for every bounded set. So \( W_n \) is tight.

If \( W \) and \( W' \) are two subsequential limits, then by Proposition 2.2,

\[
Q(A - W) \overset{\text{d}}{=} Q(A - W') \quad \text{for all } A \in B^b.
\]

An application of Lemma A.3 completes the proof.

Remark 2.6 The conclusion of Proposition 2.3 is still valid even if Conditions are not satisfied, instead there exists a proper centering \( \hat{\theta} \) such that \( \psi_n^{-1}(\hat{\theta} - \theta_0) \) is tight.

Theorem 2.4 Assume Conditions (III) and let \( Z(u) = \exp[Y(u)] \). If a centering \( \hat{\theta} \) exists, then there exist a random variable \( W \) such that

\[
\psi_n^{-1}(\hat{\theta} - \theta_0) \overset{\text{d}}{=} W
\]

and for almost all \( u \in \mathbb{R}^d \), \( \xi(u - W) \) is non-random, equivalently for \( u_0, u_1 \in \mathbb{R}^d \),

\[
Y(u_1 - W) - Y(u_0 - W) \text{ is non-random.}
\]

**Proof.** By Proposition 2.3, such a \( W \) exists.

By Lemma A.3, \( \xi = \{Q(A - W): A \in B^b\} \) is an \( M_{Q_1} \)-valued random field. Fix a countable field \( \mathcal{A} \) which generates \( B^b \). By Remark 2.6 and Proposition 2.3,

\[
(\int_A \xi(u)du: A \in \mathcal{A}) \overset{\text{d}}{=} (Q(A - W): A \in \mathcal{A})
\]

and hence \( \xi \overset{\text{d}}{=} \xi \).

Since \( P(\xi \in M_{Q_1}) = 1 \), we also have \( P(\xi \in M_{Q_2}) = 1 \). Define \( \psi_1 \) as in Lemma A.3 and hence by (2.26)

\[
(\xi, \psi^{-1}(\xi)) \overset{\text{d}}{=} (\zeta, \psi^{-1}(\zeta)) = (\zeta, W).
\]  

Define a map \( \Lambda: \mathcal{P} \times \mathbb{R}^d \to \mathcal{P} \) by \( \Lambda(Q, x) = Q^* \) where \( Q^*(A) = Q(A + x) \) for all \( A \in B^b \).

By Lemma A.3, for a fixed \( Q \), \( \Lambda \) is continuous in \( x \) whereas for fixed \( x \), \( \Lambda \) is an isometry in \( Q \). Thus (2.27) implies that

\[
\Lambda(\xi, \psi^{-1}(\xi)) \overset{\text{d}}{=} \Lambda(\zeta, W).
\]

Putting \( \psi^{-1}(\xi) = W^* \), we have \( W^* \overset{\text{d}}{=} W \) and

\[
\int_{A \times W^*} \xi(u)du = Q(A) \quad \text{for all } A \in B^b,
\]

i.e.,

\[
\int_A \xi(u - W^*)du = Q(A) \quad \text{for all } A \in B^b.
\]  

The conclusion is now immediate.

The next result shows that a proper centering, if it exists, is essentially unique.

Proposition 2.4 Assume Conditions (III) and let \( \hat{\theta} \) and \( \tilde{\theta} \) be two proper centerings. Then the associated probabilities and weak limits are shifts of each other.

**Proof.** Let \( Q_1, Q_2 \) denote the associated measures and \( W_1, W_2 \) denote the weak limits of \( \psi_n^{-1}(\hat{\theta} - \theta_0) \) and \( \psi_n^{-1}(\tilde{\theta} - \theta_0) \) respectively (which exist by Proposition 2.3). By Proposition 2.2, it follows that \( \mathcal{P} \)-valued random process \( \{Q_1(A - W_1): A \in B^b\} \) has the same distribution as \( \{Q_2(A - W_2): A \in B^b\} \). Hence it follows that \( Q_1 \), \( Q_2 \) are \( M_{Q_1} \)-valued. Using arguments similar to Theorem 2.4, it follows that \( \psi_1^{-1}(\hat{\theta} - \theta_0) \equiv \psi_1^{-1}(\tilde{\theta} - \theta_0) \).

Remark 2.7 Conditions (III) are used only to guarantee the existence of the stated limits. So one can use Remark 2.6 instead of Proposition 2.3.

In the remaining part of this section, we give a partial answer to the question of whether a nonproper centering exists.

**Theorem 2.8** A semiproper centering \( \hat{\theta} \) is called regular if there exists a continuous \( A \) from \( \mathcal{P} \) to \( \mathbb{R}^d \) such that

\[
\psi_n^{-1}(\hat{\theta} - \theta_0) = \Lambda(\xi).
\]  

(2.30)
Theorem 2.5 Assume Conditions (II) and let \( \hat{\theta} \) be a regular semiparametric centering with associated measure \( Q \). Then \( Q \) is countably additive and there exists a random variable \( W \) satisfying
\[
\begin{align*}
(\alpha) \quad & \psi_n(\hat{\theta} - \theta) \overset{d}{\rightarrow} W \\
(\beta) \quad & \xi(u - W) \text{ is non-random for almost every } u .
\end{align*}
\]
Further, if \( \hat{\theta} \) is another regular semiparametric centering with associated measure \( W' \), then \( W' \) is a shift of \( W \).

Proof. By Theorem 2.3, it follows that (\(a\)) is satisfied with \( W = \Lambda(\xi) \) and
\[
(\zeta, \psi_n^{-1}(\hat{\theta} - \theta)) \overset{d}{\rightarrow} \Lambda(\xi, W) .
\]

For any Borel set \( A \) in \( \mathbb{R}^t \), the map \( (f, x) \rightarrow \int_{A+x} f(u)du \) from \( L^1(\mathbb{R}^t) \times \mathbb{R}^t \) to \( \mathbb{R} \) is continuous by lemma A.2. So (2.31) now implies that
\[
\pi(\psi_n^{-1}(\hat{\theta} - \theta) \in A) \overset{d}{\rightarrow} \int_A \xi(\xi)du
\]
combining (2.21) with (2.32)
\[
\int_A \xi(u - W)du = Q(A) .
\]
Since the left hand side of (2.33) is countably additive, so is \( Q \) and this proves the first part.

Also (2.33) implies that
\[
Q(A + W) = \int_A \xi(\xi)du
\]
from which the second part follows as in Proposition 2.4.

The condition on \( \hat{\theta} \) is likely to be satisfied if it is taken as a quasi-posterior.

3 Examples

We now apply the results established in Sect. 2 to several important examples.

Example 3.1. Regular case. It is well known that in the usual regular case the posterior distribution centered at the MLE converges to a normal distribution with variance norm a.s. In those cases, the limiting likelihood ratio process is
\[
\mathcal{Z}(u) = \exp(u'\Delta - \frac{1}{2}u'\Sigma)
\]
where \( I \) is Fisher's information matrix and \( \Delta \) is a random vector having \( N_p(0, I) \). Indeed, if one assumes conditions of Sect. III.3.1 of III (1981), Conditions (II) of Sect. 2 are satisfied. One can easily see that the necessary and sufficient conditions stated in Theorem 2.4 is satisfied with \( W = I^{-1}\Delta \).

Example 3.2. Non-Regular Case — Densities with Jumps. We consider the set up and assumptions of III (1981, Chap. V, p. 242). We have a sequence of i.i.d. observations with values in \( \mathbb{R} \) and common density \( f(x, \theta) \) with respect to the Lebesgue measure where the parameter set \( \Theta \) is an open interval (finite or infinite) in \( \mathbb{R} \). Let \( f(x, \theta) \) possess \( r \) jumps at \( x_1(\theta), \ldots, x_r(\theta) \) and let \( p_i(\theta) = \lim_{x \to x_i(\theta)} f(x, \theta), q_i(\theta) = \lim_{x \to x_i(\theta)} f(x, \theta) \), \( i = 1, 2, \ldots, r \).

We fix \( \theta \in \Theta \) and write \( p_i, q_i, x_i, x_i' \) in place of \( p_i(\theta), q_i(\theta), x_i(\theta), x_i'(\theta) \) respectively. It is shown in III (1981) that Conditions (II) of Sect. 2 are satisfied in the case with \( \psi(\theta) = n^{-1} \).

Whether a limit of the posterior exists or not depends on the nature of the jumps. Below we consider several important special cases.

Case 1. Assume that for each \( r \) one of the numbers \( p_i \) and \( q_i \) is zero.

Let
\[
\begin{align*}
I^+ &= \{(i; p_i = 0 \text{ and } x_i' > 0) \cup (i; p_i = 0 \text{ and } x_i' < 0) \} \\
I^- &= \{(i; q_i = 0 \text{ and } x_i' > 0) \cup (i; q_i = 0 \text{ and } x_i' < 0) \}
\end{align*}
\]
and
\[
c = \sum_{i=1}^r (p_i - q_i)x_i'.
\]

Assume 1A. Suppose that both \( I^+ \) and \( I^- \) are nonempty. In this case the limiting likelihood ratio process is given by
\[
\mathcal{Z}(u) = \begin{cases}
\exp(u'\Delta - \frac{1}{2}u'\Delta), & \text{if } -r^- < u < r^+ \\
0, & \text{otherwise}
\end{cases}
\]
where \( r^- \) and \( r^+ \) are independent exponentially distributed random variables with parameters \( a = \sum_{i \in I^+} (p_i - q_i) \) and \( b = \sum_{i \in I^-} (p_i - q_i) \) respectively.

If \( c = 0 \), we have
\[
\xi(u) = \begin{cases}
(t^+ + r^-)^{-1}, & \text{if } -r^- < u < r^+ \\
0, & \text{otherwise}
\end{cases}
\]
and in case \( c \neq 0 \), we have
\[
\xi(u) = \begin{cases}
\exp(t^+ + r^-), & \text{if } -r^- < u < r^+ \\
0, & \text{otherwise}
\end{cases}
\]
In both cases, it is clear that the necessary condition of Theorem 2.4 is not satisfied (except when \( \psi = 0 \)) and \( U(\theta, \Delta) \) (with \( c \neq 0 \)) and \( U(\theta, 2\Delta) \) (with \( c \neq 0 \)).

Case 1B. Suppose that one of \( I^- \) and \( I^+ \) is empty. In case \( I^- \) is empty (the case is similar), we have
\[
\xi(u) = \begin{cases}
\exp(c(u - r^-)), & \text{if } u < r^+ \\
0, & \text{otherwise}
\end{cases}
\]
and the necessary condition of Theorem 2.4 holds with \( W = -r^+ \).
Indeed a limit has been obtained in Samanta (1988) for a special situation of Subcase 1B, where the support of the density is an interval which is either increasing or decreasing in \( \theta \). Samanta (1988) assumed conditions similar to those of Weiss and Wolfowitz (1974, Chap. 5) and a uniform integrability type condition on \( f \) and obtained an exponential limit. In these situations there exists a statistic \( Z_0 \) such that the set \( \{(X_1, \ldots, X_n), f(X, \theta) > 0 \text{ for all } i \} \) can be expressed as \( \{Z_0(X) > \theta\} \) or \( \{Z_0(X) < \theta\} \) according as the support is decreasing or increasing in \( \theta \). The \( Z_0 \) acts as a proper centering.

Important examples of this case are shifts of exponential density, \( U(0, \theta) \) etc.

**Case 2.** We now consider the case when both \( p_0 \) and \( q_0 \) are positive. We only consider the case with \( r = 1 \) and \( x' > 0 \).

In this case we have

\[
Y(u) = \left\{ \begin{array}{ll}
(p_1 - q_1)u + \nu^+(u) \log \frac{p_1}{q_1}, & \text{if } u \geq 0 \\
(p_1 - q_1)u - \nu^-(u) \log \frac{p_1}{q_1}, & \text{if } u < 0
\end{array} \right.
\]

where \( \nu^+(u) \) and \( \nu^-(-u) \) are independent homogeneous Poisson process with rates \( p_1 \alpha' \) and \( q_1 \alpha' \), respectively. One can show that the necessary condition of Theorem 2.4 is not satisfied.

An important example of this kind is the change point problem with

\[
f(x, \theta) = \left\{ \begin{array}{ll}
\alpha \exp(-ax), & \text{if } 0 < x < \theta \\
b \exp(-a\theta - b(x - \theta)), & \text{if } x > \theta,
\end{array} \right.
\]

where \( a > 0, b > 0 \) \( (a > 4) \) are known constants and \( \theta > 0 \) is the parameter of interest. See in this connection Baus, Ghoe, and Joshi (1986) and Ghose and Mukhopadhyay (1972a,b).

**Example 3.3. Non-Regular Case — Densities with Singularities.** We consider a sequence of i.i.d. observations with density \( f(x - \theta) \) where \( \theta \) is a real parameter. \( f(x) \) admits the representation

\[
f(x) = \left\{ \begin{array}{ll}
n(x)x^\alpha & \text{if } x > 0, \\
0 & \text{if } x < 0,
\end{array} \right.
\]

in a neighborhood of zero, where \( p(x) \) is a continuous function with \( p(0) = 1 \).

In this case we say that \( f(x) \) has a singularity of order \( \alpha \) at the point zero.\( -1 < \alpha < 1 \). This example is a special case of singularity treated in [2], where it is shown that in presence higher order singularities (smaller value of \( \alpha \)) of singularities of lower orders do not affect the asymptotic analysis. In this case we assume that there is only one singularity of the highest order \( \alpha \) and the conditions hold:

1. There exists a number \( \lambda > 1 + \alpha \) such that for any neighborhood \( N \) of \( 0 \)

\[
\int_0^{\lambda} [f'^{1/\lambda}(x - h) - f'^{1/\lambda}(x)]^\lambda dx = 0 \text{ for } h > 0.
\]

2. \( \int |x|^\alpha f(x) dx < \infty \) for some \( \delta > 0 \).

Under these assumptions, Conditions (III) of Sect. 2 are satisfied with \( \varphi_n = \theta^{-1/\alpha} \) and \( Y(\cdot) = \log Z(\cdot) \) is given by

\[
Y(u) = \left\{ \begin{array}{ll}
\alpha \int_0^{\infty} \log |1 - \frac{1}{z}(\nu(x) - Ev(x)) dz & \text{if } u \geq 0 \\
-\nu_0 \int_0^{\infty} \log \left(1 - \frac{1}{z^{1/\alpha}} - 1 - \alpha \log |1 - \frac{1}{z^{1/\alpha}}|\right) \nu(x) dz & \text{if } u = 0 \\
-\nu_0 \frac{1}{z^{1/\alpha}} & \text{if } u < r
\end{array} \right.
\]

where \( \nu \) is a non-homogeneous Poisson process with rate function \( \lambda(z) = \varphi^{1/\alpha} \) and \( r \) is the first jump of the process \( \nu \).

We consider the case \( \alpha \neq 0 \). The case \( \alpha = 0 \) is treated in Example 3.2. Let \( W \) be a random variable and \( u_1 > u_0 \) be real numbers to be chosen later. In order that \( Y(u_1 - W) - Y(u_0 - W) \) be non-random it must be true that the set where it is positive, i.e., the set \( \{W + r < u_0\} \) is trivial. Using this fact for different \( u_0 \), one can show that \( W + r \) must be constant, say \( W = c + r \). We choose \( u_1 > u_0 > c \).

Putting \( u_0 = u_0 - r, u'_1 = u_1 - c \) we have

\[
Y(u_1 - W) - Y(u_0 - W) = Y(u'_1 + r) - Y(u'_0 + r)
\]

\[
= -\nu_0 \int_0^{\infty} \log \left(1 - \frac{u'_1 + r}{z} - 1 - \alpha \log |1 - \frac{1}{z^{1/\alpha}}|\right) \nu(x) dz
\]

\[
= \nu_0 \frac{1}{z^{1/\alpha}} - \nu_0 \frac{1}{z^{1/\alpha}} - \nu_0 \frac{1}{z^{1/\alpha}} = \nu_0 \frac{1}{z^{1/\alpha}}
\]

\[
g(u, z) = 1 - \frac{u}{z^{1/\alpha}} - 1 - \log |1 - \frac{u}{z^{1/\alpha}}|
\]

\[
\dot{\nu}(dz) = \nu_0 \cdot dz - Ev_0 dz.
\]

Since the first three terms are functions of \( r \) only, if \( Y(u_1 - W) - Y(u_0 - W) \) is random, then the conditional distribution of \( Y(u_1 - W) - Y(u_0 - W) \) given \( r \) is degenerate. Therefore

\[
\int_0^{\infty} \log |1 - \frac{u'_1 + r}{z} - 1 - \alpha \log |1 - \frac{1}{z^{1/\alpha}}| \nu(x) dz
\]

\[
= \nu_0 \frac{1}{z^{1/\alpha}} \mu(dx)
\]

where \( \mu \) is a degenerate conditional distribution given \( r \). Here

\[
\dot{\nu}(dz) = \nu_0 \cdot dz - Ev_0 dz \text{ and } \mu(x) = \nu(r + x).
\]
However, given \( \tau, \mu \) is again a non-homogeneous Poisson process with rate function
\[ p(x) = \left\{ \begin{array}{ll}
\frac{1}{2} \frac{e^{-x/2}}{x-1} & \text{if } x > 1
\
0 & \text{if } x < 0
\end{array} \right. \]
This contradiction implies the non-existence of a limit of posterior.

Important examples of this kind are the Lognormal density
\[ f(x) = \frac{1}{\sqrt{2\pi x^2}} e^{-\frac{(\ln x - \mu)^2}{2x^2}}, \quad \text{if } x > 0 
= \frac{1}{\sqrt{2\pi x^2}} e^{-\frac{(\ln x - \mu)^2}{2x^2}}, \quad \text{if } x = 0
= 0, \quad \text{if } x < 0
\]
and the Weibull density
\[ f(x) = \frac{\alpha x^{\alpha-1} \exp(-x^\alpha)}{\Gamma(\alpha)}, \quad \text{if } x > 0 
= 0, \quad \text{if } x < 0
\]
where \( 0 < \alpha < 2 \).

Another example is provided by
\[ f(x) = \frac{1}{\Theta x} \exp(-\frac{1}{x^\delta}), \quad \text{if } 0 < x < 1 
= 0, \quad \text{otherwise}
\]
where \( 0 < \alpha < 2, \delta > 0 \) and \( \alpha < \delta \).

The result of this paper holds a representation
\[ f(x) = \frac{1}{\Theta x} \exp(-\frac{1}{x^\delta}), \quad \text{if } 0 < x < 1 
= 0, \quad \text{otherwise}
\]
in a neighborhood of zero is exactly similar and hence omitted.

**Example 3.4. A Two Parameter Case.** Suppose that the observations are a common density
\[ f(x, \theta_1, \theta_2) = \left\{ \begin{array}{ll}
\frac{e^{-\theta_2 x}}{x-1}, & \text{if } x > \theta_1 
= 0, & \text{otherwise}
\end{array} \right. \]
where \(-\infty < \theta_1, \theta_2 < \infty\) and \( \theta_1 > 0 \) are two unknown parameters. As in Examples 3.1 and 3.2 (Case 1), if \( \theta_1 \) is known, the posterior goes to \( \theta_2 \) distribution and on the other hand if \( \theta_2 \) is known it has an exponential limit \( \theta_1 \) and \( \theta_2 \) are unknown, the limiting likelihood ratio process is given by
\[ Z(u_1, u_2) = \left\{ \begin{array}{ll}
\exp\left( (u_1 + u_2)^2 \right), & \text{if } u_1 < r, 
= 0, & \text{otherwise}
\end{array} \right. \]
where \( c > 0 \) and \( f > 0 \) are constants depending on \( \theta_1, \theta_2 \) and \( \alpha \) and \( \alpha \) and \( \alpha \) are independent random variables following \( N(0, 1) \) and exponential distribution with \( \alpha \) respectively. It is interesting to note that (3.2) is the product of the exponential processes obtained in the non-regular case when \( \theta_1 \) is known and in the regular case when \( \theta_1 \) is known. It is easy to see that the necessary condition (3.1) is satisfied in this case. We also expect, but don’t have a proof yet, that (3.1) holds for this example. Indeed, proceeding in a manner similar to that in (199, Chap. 3 and 4) one can show that the posterior has an a.s. limit product of exponential and normal.
Lemma A.4. Let \((\Omega, \mathcal{E})\) be a measurable space and \(\chi: \mathcal{Q} \to \mathcal{P}\) be a map. Then \(\chi\) is measurable if for all \(A\) in \(\mathcal{B}\), the map \(\chi_A\) defined by
\[
\chi_A(w) = \chi(w)(A)
\]
is measurable.

Proof. Since \(\mathcal{Q} \to \mathcal{Q}(A)\) from \(\mathcal{P}\) to \(\mathcal{R}\) is continuous only if \(A\) is trivial.

For if \(A\) let \(\mathcal{F}\) be a countable field generating \(\mathcal{B}\). Then for any \(Q, Q_0 \in \mathcal{P}\)
\[
||Q - Q_0|| = \sup\{||Q(A) - Q_0(A)||: A \in \mathcal{F}\}
\]
by a well known fact in measure theory, (see Halmos (1974, Theorem 13.D) for example).

Now \(\{w: ||\chi(w) - Q_0|| \leq x\} = \bigcap\{w: ||\chi_A(w) - Q_0(A)|| \leq x\} \in \mathcal{F}\) by hypothesis.

References


---

\[
\text{This text was processed using the \LaTeX\ macro package with LNCs style}
\]