Adaptive posterior rate of convergence for infinite-dimensional exponential families

C. Scricciolo

2003.25

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Dicembre 2003
ADAPTIVE POSTERIOR RATE OF CONVERGENCE
FOR INFINITE-DIMENSIONAL EXPONENTIAL FAMILIES

BY CATIA SCRICCIOLI

University of Padova

PRELIMINARY DRAFT

The problem of estimating probability densities on the unit interval whose
log-functions belong to a periodic Sobolev space is studied adopting a Bayesian
approach. A prior is constructed so that the posterior converges at optimal rate in
the minimax sense under Hellinger loss whichever is the degree of smoothness of
the log-density. Thus, the point-wise posterior expectation of the density function
provides an optimal non-parametric adaptive estimation procedure.

1. Introduction. We consider the problem of estimating densities on a closed in-
terval, taken to be $[0, 1]$ without loss of generality. Log-density functions are supposed
to belong to a periodic Sobolev space. From the Bayesian perspective, the most natural
estimation procedure is the posterior expected density. The goal is to obtain estimates
achieving the optimal rate of convergence in the minimax sense under Hellinger loss,
whichever is the degree of smoothness of the sampling density.

Suppose the generic density function $f_\theta$ with respect to Lebesgue measure is of the
form

$$f_\theta(x) = \frac{\exp \{ \theta(x) \}}{\int_0^1 \exp \{ \theta(t) \} \, dt}, \quad x \in [0, 1],$$

with $\theta$ a square-integrable function. Trigonometric series expansion transforms such prob-
lem into a discrete one. Let $\{\phi_j\}_{j \geq 0}$ be the trigonometric orthonormal system of $L_2[0, 1]$.


Key words and phrases. Adaptive rate of convergence, infinite-dimensional exponential family, posterior
distribution.
i.e. for $x \in [0, 1]$, $k \geq 1$,

$$
\phi_j(x) = \begin{cases} 
1, & j = 0, \\
\sqrt{2}\cos(2\pi k x), & j = 2k - 1, \\
\sqrt{2}\sin(2\pi k x), & j = 2k.
\end{cases}
$$

For any $\theta \in L_2[0, 1]$, $j \geq 0$, let $\theta_j = \int_0^1 \theta(x)\phi_j(x) \, dx$ be the sequence of Fourier coefficients. Then, $\theta$ admits the series expansion $\sum_{j=0}^{\infty} \theta_j \phi_j$. For convenience, let $\theta \cdot \phi = \sum_{j=0}^{\infty} \theta_j \phi_j$ and $\psi(\theta) = \log(\int_0^1 \exp\{\theta \cdot \phi(t)\} \, dt)$. The density in (1) takes the form

$$
f_\theta(x) = \exp\{\theta \cdot \phi(x) - \psi(\theta)\}, \quad x \in [0, 1].
$$

For later use, we introduce the definition of a Sobolev space. For any integer $p \geq 1$, the Sobolev functional space $W_p^p$ comprises all square-integrable functions having degree of smoothness at least $p$, with the $p$th derivative bounded in $L_2$-norm and derivative of order $p - 1$ which is absolutely continuous. The periodic Sobolev space is defined as the subclass of all periodic functions with period 1 satisfying certain boundary conditions:

$$
W_p^{\text{per}} = \{ \theta \in W_p^p : \theta^{(r)}(0) = \theta^{(r)}(1), \ r = 0, \ldots, p - 1 \}.
$$

It is known that any function $\theta$ admitting the series expansion $\sum_{j=0}^{\infty} \theta_j \phi_j$ belongs to $W_p^{\text{per}}$ if and only if the sequence $\{\theta_j\}_{j \geq 0}$ satisfies the condition $\sum_{j=0}^{\infty} a_j^2 \theta_j^2 < \infty$, with $\{a_j\}_{j \geq 0}$ defined as follows. For $k \geq 1$,

$$
a_j = \begin{cases} 
0, & j = 0, \\
(2k)^p, & j = 2k - 1, \\
(2k)^p, & j = 2k,
\end{cases}
$$

equivalently, $a_0 = 0$ and $a_{2k-1} = a_{2k} = (2k)^p$. To ease the notation, let $\Theta = (\theta_0, \theta_1, \ldots)$. A function $\theta = \sum_{j=0}^{\infty} \theta_j \phi_j$ belongs to $W_p^{\text{per}}$ if and only if $\theta$ belongs to the Sobolev space $\Theta_p = \{ \theta \in \ell^2 : \sum_{j=0}^{\infty} a_j^2 \theta_j^2 < \infty \}$. Note that for any pair $\theta, \theta' \in W_p^{\text{per}}$, $||\theta - \theta'||_2 = ||\theta' - \theta||_{\ell^2}$. We shall put a prior on $\mathcal{P} = \{ f_\theta = e^{\theta \cdot \phi - \psi(\theta)}, \ \theta \in \ell^2 \}$, even though assuming that the sampling density $f_\theta \equiv f_{\theta_0}$, with $\theta_0 \in \Theta_{p_0}$ for some $p_0$. Randomness of $f_\theta$ is driven by randomness of $\theta$. Thus, any prior $\pi$ on $\ell^2$ induces a prior $\Pi$ on $\mathcal{P}$ via the map $\theta \mapsto f_\theta$. We shall consider a mixture prior of the form $\sum_{p=1}^{\infty} \lambda_p \Pi_p$ and show that it leads to the optimal posterior rate. A similar problem has been addressed by Belitser and Ghosal (2003).
2. Main results. We assume that conditionally on \( p \), the \( \theta_j \)'s are independent, zero-mean normals with variance \( j^{-2q} \), i.e. \( \theta_j \sim N(0, j^{-2q}) \) for \( j \geq 0 \), with \( q \) to be suitably specified. The joint prior density for \( \theta \) is an infinite product of normal densities

\[
\pi_p(\theta) = \prod_{j=0}^{\infty} \pi_j(\theta_j) = \delta_0(\theta_0) \times \prod_{j=1}^{\infty} \frac{j^q}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} j^{2q} \theta_j^2 \right),
\]

where \( \delta_0(\theta_0) \) denotes point mass at zero for \( \theta_0 \). Let \( \Pi_p \) be the induced prior on \( \mathcal{P} \). The overall prior is then \( \Pi = \sum_{p=1}^{\infty} \lambda_p \Pi_p \). Let \( \mathcal{P} = \{p_0, \ldots, p_m, p_{m+1}, \ldots\} \) be the set of possible values for \( p \). Let \( \lambda_m = \lambda(p = p_m) \). We shall assume that \( \lambda_m > 0 \) for all \( m \). We shall write \( \Pi_m \) for \( \Pi_{p_m} \). For \( m \geq 0 \), let \( \lambda_m = \lambda_m / \sum_{j=0}^{\infty} \lambda_j \) and \( \Pi \sum_{m=0}^{\infty} \lambda_m \Pi_m \).

**Theorem 1.** Let \( q_0 = 2p_0 + \frac{\beta}{2} \), with \( \beta > 1 \) fixed. Let \( \theta_0 \) be such that

\[
\sum_{j=1}^{\infty} j^{2q_0} \theta_j^2 < \infty \text{ and } \sum_{j=1}^{\infty} j^{2q_0} E_0[\phi_j(X_1)^2].
\]

Then, for any sequence \( M_n \to \infty \),

\[
\Pi \left( \left\{ P_\theta \in \mathcal{P} : d_{\Pi}(P_0, P_\theta) \geq M_n n^{-p_0/(2p_0+1)} \right\} \right| X_1, \ldots, X_n \to 0
\]

in probability.
PROOF. We denote by \( r_n = r_n(p_0) = n^{p_0/(2p_0+1)} \). Let \( L > 0 \) be a constant. Clearly, 
\[
\Pi \left( \{ P_\theta : r_n d_{\Pi}(P_\theta, P_0) \geq M_n | X^n \} \right) = \Pi \left( \left\{ P_\theta : r_n d_{\Pi}(P_\theta, P_0) \geq M_n, \quad p \geq p_0, \quad \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 > L | X^n \right\} \right)
+ \Pi \left( \left\{ P_\theta : r_n d_{\Pi}(P_\theta, P_0) \geq M_n, \quad p \geq p_0, \quad \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 \leq L | X^n \right\} \right)
+ \Pi \left( \left\{ P_\theta : r_n d_{\Pi}(P_\theta, P_0) \geq M_n, \quad p < p_0 | X^n \right\} \right)
\]
\[ \leq \Pi \left( \left\{ P_\theta : \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 > L | X^n \right\} \right)
+ \Pi \left( \left\{ P_\theta : r_n d_{\Pi}(P_\theta, P_0) \geq M_n, \quad \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 \leq L | X^n \right\} \right)
+ \Pi \left( p < p_0 | X^n \right). \]

For all \( n \) large, given \( \epsilon > 0 \), \( L \) can be chosen sufficiently large so that the first addendum is less than \( \epsilon \) by Lemma 1. For this \( L \), Theorem 4 in Scricciolo (2002) can be applied to the second term. The last term goes to zero in probability by Lemma 2. □

We can provide an approximation for \( n \) large of the posterior expected density. By laying out the same arguments as in the proof of Proposition 1 in Scricciolo (2002), we get the following expression
\[
\tilde{f}_n(x) = E[f_\theta(x)|x^n] = \frac{\sum_{p=1}^{\infty} \lambda_p \int f_\theta(x) \prod_{i=1}^{n} f_\theta(x_i) \pi_p(\theta) \, d\theta}{\sum_{q=1}^{\infty} \lambda_q \int \prod_{i=1}^{n} f_\theta(x_i) \pi_q(\theta) \, d\theta} \times \frac{\sum_{p=1}^{\infty} \lambda_p \prod_{j \geq 1} (1 + (n + 1)j^{-2p+\beta})^{-1/2} \exp \left\{ \frac{1}{2} \sum_{j \geq 1} \frac{(n \phi_j + \phi_j(x))^2}{(n+1)^{j-2p+\beta}} \right\}}{\sum_{q=1}^{\infty} \lambda_q \prod_{j \geq 1} (1 + n^{-1}j^{-2q+\beta})^{-1/2} \exp \left\{ \frac{1}{2} \sum_{j \geq 1} \frac{(n \phi_j)^2}{(n+1)^{j-2q+\beta}} \right\}}.
\]

The Bayes's estimator provides an adaptive Bayesian density estimation procedure, yet, its approximate expression involves infinitely many terms which renders impossible to carry out computations. Even though a finite number of values for the smoothness parameter can be considered, there are still infinite terms due to the orthonormal series expansion. The usual way for getting rid of infinite series is to truncate the series to a number of terms varying with sample size. But, at this stage of knowledge, we are not
able to say whether it is possible to suitably choose the maximum number of terms as a function on \( n \) in such a way that the optimal rate of convergence be preserved. Research on this issue is in course and we hope to report on it in a future work.

**APPENDIX**

**Lemma 1.** If \( \sum_{j=1}^{\infty} j^{2p_0} (E_0[\phi_j(X_1)])^2 < \infty \), then there exists \( N \) not depending on \( L \) such that

\[
\lim_{L \to \infty} \sup_{n \geq N} E_0^n \left[ \prod \left( \left\{ \mathcal{P}_0 : \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 > L \right\} \right| X^n \right] = 0.
\]

**Proof.** By Markov’s inequality,

\[
\prod \left( \left\{ \mathcal{P}_0 : \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 > L \right\} \right| X^n \right) < L^{-1} \sum_{j=1}^{\infty} j^{2p_0} E_0 \left[ \theta_j^2 \right| X^n \right] = L^{-1} \sum_{j=1}^{\infty} j^{2p_0} \sum_{m=0}^{\infty} \Pi(p = p_m | X^n) E_0 \left[ \theta_j^2 \right| X^n, \ p = p_m \right] \leq L^{-1} \sum_{j=1}^{\infty} j^{2p_0} \sup_{m \geq 0} E_0 \left[ \theta_j^2 \right| X^n, \ p = p_m \right] \right)
\]

then

\[
E_0^n \left[ \prod \left( \left\{ \theta : \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 > L \right\} \right| X^n \right) \right) < L^{-1} \sum_{j=1}^{\infty} j^{2p_0} E_0^n \left[ E_0 \left[ \theta_j^2 \right| X^n, \ p = p_0 \right] \right].
\]

For \( n \) large,

\[
\sum_{j=1}^{\infty} j^{2p_0} E_0^n \left[ E_0 \left[ \theta_j^2 \right| X^n, \ p = p_0 \right] \right] \right) \approx \sum_{j=1}^{\infty} \frac{j^{2p_0}}{n + j^{2p_0 + \beta}} + \sum_{j=1}^{\infty} j^{2p_0} \frac{n^{2} E_0^n [\theta_j^2]}{(n + j^{2p_0 + \beta})^2} \equiv S(n),
\]
where the LHS differs from the RHS by a quantity which is \( o(1) \). Noting that

\[
n^2E_0 \left[ \phi_j^2 \right] = E_0 \left[ \sum_{i=1}^{n} \phi_j(X_i)^2 + \sum_{i \neq l} \phi_j(X_i)\phi_j(X_l) \right]
\]

\[
= \sum_{i=1}^{n} E_0 \left[ \phi_j(X_i)^2 \right] + \sum_{i \neq l} E_0 \left[ \phi_j(X_i) \phi_j(X_l) \right] 
\]

\[
= \sum_{i=1}^{n} E_0 \left[ \phi_j(X_i)^2 \right] + \sum_{i \neq l} E_0 \left[ \phi_j(X_i) \phi_j(X_l) \right] 
\]

\[
= nE_0 \left[ \phi_j(X_1)^2 \right] + n(n-1)(E_0[\phi_j(X_1)])^2 
\]

\[
= n^2E_0[\phi_j(X_1)^2] + n \{ E_0[\phi_j(X_1)^2] - (E_0[\phi_j(X_1)])^2 \} 
\]

\[
= n^2E_0[\phi_j(X_1)^2] + n \text{Var}_0[\phi_j(X_1)].
\]

it can be seen that both sums

\[
S_1^{(n)} = \sum_{j=1}^{\infty} \frac{j^{2p_0}}{n + j^{2p_0 + \beta}} 
\]

\[
S_2^{(n)} = \sum_{j=1}^{\infty} \frac{j^{2p_0} n^2 (E_0[\phi_j(X_1)])^2 + n \text{Var}_0[\phi_j(X_1)]}{(n + j^{2p_0 + \beta})^2}
\]

are finite and bounded in \( n \). For \( \beta > 1 \), clearly \( S_1^{(n)} < \sum_{j=1}^{\infty} j^{-\beta} < \infty \). Now,

\[
S_2^{(n)} = \sum_{j=1}^{\infty} \frac{n^2 j^{2p_0} (E_0[\phi_j(X_1)])^2}{(n + j^{2p_0 + \beta})^2} + \sum_{j=1}^{\infty} \frac{n j^{2p_0} \text{Var}_0[\phi_j(X_1)]}{(n + j^{2p_0 + \beta})^2} \equiv S_{21}^{(n)} + S_{22}^{(n)}.
\]

Note that \( S_{21}^{(n)} < \sum_{j=1}^{\infty} j^{2p_0} (E_0[\phi_j(X_1)])^2 < \infty \). Consider \( S_{22}^{(n)} \). Taking \( k = \left[ n^{1/(2p_0+1)} \right] \),

\[
S_{22}^{(n)} \leq \sum_{j=1}^{\infty} \frac{n j^{2p_0} E_0[\phi_j(X_1)^2]}{(n + j^{2p_0 + \beta})^2} \leq 2 \sum_{j=1}^{\infty} \frac{n j^{2p_0}}{(n + j^{2p_0 + \beta})^2} \leq 2n^{-1} \sum_{j=1}^{k} j^{2p_0} + 2n \sum_{j=k+1}^{\infty} j^{-2(\beta_0 + \beta) + 1} 
\]

\[
\leq 2n^{-1} k^{2p_0+1} + (k+1)^{-2(\beta_0 + \beta)+1} \leq 2 \frac{1 + n^{-1}}{k} \leq 4.
\]

It follows that there exists \( N \) such that for all \( n \geq N \), \( S^{(n)} \) is bounded in \( n \). Consequently, for any \( \epsilon > 0 \), by choosing \( L > 0 \) large enough, \( P_0^n \left[ \Pi \left\{ P_0 \mid \{ \sum_{j=1}^{\infty} j^{2p_0} \theta_j^2 > L \} \mid X^n \right\} \right] < \epsilon \), which concludes the proof. \( \square \)

**Lemma 2.** \( \Pi(p < p_0 | X^n) \rightarrow 0 \) in probability.
PROOF. For any $m < 0$,
\[
\Pi(p < p_0 | X^n) = \sum_{m < 0} \Pi(p = p_m | X^n) = \sum_{m < 0} \frac{\lambda_m \int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_m)}{\sum_{i < 0} \lambda_i \int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_i)} \leq \frac{1}{\lambda_0} \sum_{m < 0} \lambda_m \int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_m) \int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_0).
\]
For $n$ large,
\[
\frac{\int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_m)}{\int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_0)} \asymp \prod_{j=1}^{\infty} \left( \frac{n^{-1} + j^{-(2p_0 + \beta)}}{n^{-1} + j^{-(2p_m + \beta)}} \right)^{1/2} \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} (n \phi_j)^2 \left[ (n + j^{2p_m + \beta})^{-1} - (n + j^{2p_0 + \beta})^{-1} \right] \right\} \equiv B_m^{(n)}
\]
and
\[
\sum_{m < 0} \lambda_m \int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_m) \int \prod_{i=1}^n f_\theta(X_i) \Pi(d\theta | p = p_0) \asymp \sum_{m < 0} \lambda_m B_m^{(n)}.
\]
It is simple to see that
\[
a_{jm}^{(n)} = n^2 \frac{[(n + j^{2p_m + \beta})^{-1} - (n + j^{2p_0 + \beta})^{-1}]}{j^{-(2p_m + \beta)} - j^{-(2p_0 + \beta)}} = \frac{n^{-1} + j^{-(2p_m + \beta)}}{(n^{-1} + j^{-(2p_m + \beta)})(n^{-1} + j^{-(2p_0 + \beta)})},
\]
and
\[
0 \leq a_{jm}^{(n)} \leq j^{2p_0 + \beta}.
\]
Thus,
\[
\prod_{j=1}^{\infty} \left( \frac{n^{-1} + j^{-(2p_0 + \beta)}}{n^{-1} + j^{-(2p_m + \beta)}} \right)^{1/2} \leq \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0 + \beta} a_{jm}^{(n)} \right\} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{\infty} j^{-(2p_m + \beta)} - j^{-(2p_0 + \beta)} \right\},
\]
so that
\[
B_m^{(n)} \leq \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0 + \beta} a_{jm}^{(n)} \right\} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{\infty} a_{jm}^{(n)} j^{-(2p_0 + \beta)} \right\}.
\]
Now, we study $\sum_{j=1}^{\infty} (-a_{jm}^{(n)}) j^{-(2p_0 + \beta)}$. Note that
\[
j^{-(2p_0 + \beta)} - j^{-(2p_m + \beta)} \leq \begin{cases} 0, & j \geq 1, \\ \frac{1}{2} j^{-(2p_m + \beta)}, & j > J = 2^{1/(2(p_0 - p_1))}, \end{cases}
\]
If \( j \leq n^{1/(2p_0+\beta)} \), then
\[
(n^{-1} + j^{-1/(2p_0+\beta)})(n^{-1} + j^{-1/(2p_0+\beta)}) \leq 4j^{-2(p_0+p_n+\beta)}.
\]

For \( n > N_1 = [2(J+1)]^{2p_0+\beta} \), then \( J + 1 < \frac{1}{2} n^{1/(2p_0+\beta)} \) and \( \left( n^{1/(2p_0+\beta)} - J \right) \geq n^{1/(2p_0+\beta)} - (J + 1) > \frac{1}{2} n^{1/(2p_0+\beta)} \) so that
\[
\sum_{j=1}^{\infty} (-a(n))_j j^{-1/(2p_0+\beta)} \leq \sum_{j > J} \frac{(j^{-1/(2p_0+\beta)} - j^{-1/(2p_0+\beta)})j^{-1/(2p_0+\beta)}}{4j^{-2(p_0+p_n+\beta)}}
\]
\[
\leq -\frac{1}{8} \left( n^{1/(2p_0+\beta)} - J \right)
\]
\[
\leq -\frac{1}{16} n^{1/(2p_0+\beta)}.
\]

We have that
\[
\sum_{m<n} \lambda_m B_m^{(n)} \leq \exp \left\{ -\frac{1}{32} n^{1/(2p_0+\beta)} \right\} \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} \frac{\sigma_j^2}{\varphi_j^2} \right\}
\]

Let us now study \( \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} \frac{\sigma_j^2}{\varphi_j^2} \right\} \). By the central limit theorem, \( \sqrt{n}(\overline{\phi}_j - \tau_{0j}) \rightarrow \mathcal{N}(0, \sigma_{0j}^2) \), where \( \sigma_{0j}^2 = \text{Var}_0(\phi_j(X_1)) \). Thus, \( \overline{\phi}_j = \tau_{0j} + O_p(n^{-1/2}) \). By the delta method,
\[
\sqrt{n} \left( \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} \frac{\sigma_j^2}{\varphi_j^2} \right\} - \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} \frac{\sigma_{0j}^2}{\varphi_{0j}^2} \right\} \right) \rightarrow \mathcal{N}(0, \tau_{0j}^2),
\]
with \( \tau_{0j}^2 = \sigma_{0j}^2 \left( \sum_{j=1}^{\infty} j^{2p_0+\beta} (E_0[\phi_j(X_1)])^2 \right) \exp \left\{ \sum_{j=1}^{\infty} j^{2p_0+\beta} (E_0[\phi_j(X_1)])^2 \right\} \). Hence,
\[
\exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} \frac{\sigma_j^2}{\varphi_j^2} \right\} = \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} (E_0[\phi_j(X_1)])^2 \right\} + O_p(n^{-1/2}),
\]
Finally,
\[
\sum_{m<n} \lambda_m B_m^{(n)} \leq \exp \left\{ -\frac{1}{32} n^{1/(2p_0+\beta)} \right\} \exp \left\{ \frac{1}{2} \sum_{j=1}^{\infty} j^{2p_0+\beta} \frac{\sigma_{0j}^2}{\varphi_{0j}^2} \right\} + o_p(n^{-1/2})
\]
and \( \Pi(p < p_0|X^n) \rightarrow 0 \) in probability.

REFERENCES


C. Scricciolo  
Dipartimento di Scienze Statistiche  
Università degli Studi di Padova  
Via C. Battisti 241/243  
I-35121 Padova  
Italy  
E-mail: catia@stat.unipd.it