Local Convergence of Nonparametric Density Estimation Problems to Gaussian Shift Experiments on a Hilbert Space

By

Mark G. Low

Technical Report No. 225
October 1989

Department of Statistics
University of California
Berkeley, California

Local Convergence of Nonparametric Density Estimation Problems to Gaussian Shift Experiments on a Hilbert Space

By

Mark G. Low

Section 1. Introduction.

Millar (1979) has illustrated the power of looking at nonparametric problems from the viewpoint of Le Cams theory of experiments, with the parameter space indexed by an infinite dimensional Hilbert space. This highly successful approach led to a simple (at least conceptually!) unified theory for estimating distribution functions under for example sup norm loss for a variety of parameter spaces. The power of Hilbert space parametrizations has not however been exploited in pointwise estimation problems arising for example from density estimation, nonparametric regression or estimation of a variable intensity function from a Poisson process. In fact it is only recently that Le Cams theory of experiments has even been brought to bear on the problem of density estimation, and then only through a sequence of one dimensional parametrizations. Donoho and Liu [1988] however did pick these one dimensional parametrizations in an optimal way and showed the power of Le Cams methodology. Romano (1989) has also applied Le Cams theory to sequences of one dimensional experiments naturally arising from estimation of the mode. In Section 2 we introduce a new sequence of experiments with a Hilbert space parametrization, that arises naturally from density estimation problems. The consideration of this sequence of experiments was inspired by Donoho and Liu [1988]. However it can also be looked upon as a generalization of some sequences of experiments given by Millar [1979]. At the end of section 2 we state a convergence theorem for this new sequence of experiments. The limiting experiment is a Gaussian shift experiment on a Hilbert space. We leave a proof of this theorem to Section 4. Applications of the theorem are given in Section 3.

Section 2. A sequence of experiments.

The main theorem in this section is given in terms of the convergence of a sequence of experiments to a Gaussian shift experiment. In other words the theorem states that the distributions of the likelihood ratios for the sequence of experiments converges weakly to the distributions of the likelihood ratios for the limiting experiment. A more detailed and complete description for the reader unfamiliar with this idea can be found either in Le Cam [1986] or Millar [1979].

We now turn to a description of the limiting Gaussian experiment. Let

$$H = \{h : R \to R, \int h^2 < \infty, \sup_x |h(x)| < \infty, \quad \int |h| < \infty\}$$

The limiting experiment in fact is just the standard Gaussian shift experiment on H. Following Le Cam [1986] the standard Gaussian process on H, where H has inner product $< \ , \ >$ and norm $\| \ \|$, is the process $h \to <Z, h>$ such that

$$E<Z,h> = 0, \quad E|<Z,h>|^2 = \|h\|^2.$$

Take G_0 for the distribution of this process and take G_h to be the measure whose density with respect to G_0 is $\exp\{<Z,h> - \frac{1}{2}\|h\|^2\}$. It then immediately follows that under G_0 the distribution of $\log \frac{dG_h}{dG_0}$ is $N(-\frac{1}{2}\|h\|^2, \|h\|^2)$. The experiment $G = \{G_h : h \in H\}$ is then called the standard Gaussian shift experiment on H. We may give a more concrete representation of this shift experiment as follows. Let $W_1(t)$ and $W_2(t)$ be independent Brownian motions on $[0,\infty)$ and let

$$W(t) = \begin{cases} W_1(t) & t \geq 0 \\ W_2(-t) & t \leq 0. \end{cases}$$

Then the experiment generated by

(0)
$$dX_t = h(t)\,dt + dW_t \qquad -\infty < t < \infty$$
$$h \in H$$

gives a concrete representation of this standard Gaussian shift on H.

We now introduce our sequence of experiments. First fix a probability density f_0 on R such that f_0 is continuous at 0, $f_0(0) > 0$, $\sup_x f(x) < \infty$, $\int |f_0| < \infty$ and $\int f_0^2 < \infty$. Corresponding to f_0 and $\{\alpha_n\}_{n=1}^{\infty}$, $\{\beta_n\}_{n=1}^{\infty}$ any nondecreasing sequence of positive numbers satisfying $\lim_{n\to\infty} \alpha_n = \infty$ and

(1)
$$\frac{\alpha_n^2 \beta_n}{f_0(0)\, n} \to 1$$

will be the following sequence of experiments. For $h \in H$ let

(2)
$$h_n = \int \frac{h(\beta_n x)}{\alpha_n} f_0(x).$$

h_n is finite since h and f_0 are square integrable. Furthermore the conditions we imposed on h and f_0 imply that $h_n = O(n^{-1})$. If

(3)
$$1 + \frac{h(\beta_n x)}{\alpha_n} - h_n \geq 0 \text{ for all } x$$

define

(4)
$$f_n(h;x) = \left(1 + \frac{h(\beta_n x)}{\alpha_n} - h_n\right) f_0(x)$$

otherwise let

(5)
$$f_n(h;x) = f_0(x).$$

Defining $f_n(h;x)$ by (5) when (3) is not satisfied is only a technical condition. Its only purpose is to make the sequence of experiments given below to have parameter space H for each n. Note that since $\sup_x h(x) < \infty$ and α_n increases to infinity, for any given h, $f_n(h;x)$ is defined by (4) for all sufficiently large n. Finally, define P_h^n to be the probability on R^n having density

$$(6) \qquad \prod_{i=1}^{n} f_n(h;x_i).$$

The collection $\{P_h^n : h \in H\}$ now defines an experiment for each n.

Theorem: The sequence of experiments $\{P_h^n : h \in H\}$ constructed above converges weakly to the standard Gaussian experiment $\{P_h : h \in H\}$.

The importance of this theorem is contained in the following corollary which is just a statement of Hajek-Le Cam minimax theorem in the present context.

First, we need to establish some notation which we shall use throughout the rest of this paper. Write $\delta_n = \delta_n(X_1, \ldots, X_n)$ to be any decision procedure based on n independent observations from a density $f_n(h)$ where $h \in H$. Also by $E_h l(h, \delta_n)$ we mean the risk of the estimator (in estimating h) when the density is $f_n(h)$. We write δ (no subscript) to be any decision procedure based on one observation from the Gaussian shift problem given in (0) and $E_h^G l(h, \delta)$ for the associated risk in estimating h.

Cor: Let $K \subset H$. If l is any loss function $l : K \times R \to R$, lower semicontinuous in the second argument, then

$$(7) \qquad \overline{\lim_{n\to\infty}} \inf_{\delta_n} \sup_{h\in K} E_h l(h, \delta_n) \geq \inf_{\delta} \sup_{h\in K} E_h^G l(h, \delta).$$

If K is compact then the inequality in (6) can be replaced by an equality.

Section 3. Applications.

In this section we give two simple applications of Theorem 1. Our main purpose is to illustrate by an example how α_n, β_n and K can be appropriately chosen. In our second example we improve on some lower bounds given by Wahba (1975) for estimating a density function at a point under Sobolev constraints. We should also mention that Millar (1979) has exploited Theorem 1, with $\alpha_n = (f_0(0))^{1/2} n^{1/2}$, $\beta_n = 1$ to obtain lower bounds for estimating distribution functions.

Example 1: Suppose we observe $X_1, \ldots, X_n$ i.i.d. with density $f \in \mathbf{F}$ and we want to estimate f either at 0 or in a neighborhood of 0.

We consider two classes of loss functions

1) l_p loss given by $l_p : \mathbf{F} \times \mathbf{R} \to \mathbf{R}$

$$l_p(f, a) = (f(0) - a)^P$$

2) $\tilde{l}_{n,p}$ loss given by $\tilde{l}_{n,p} : \mathbf{F} \times \mathbf{R} \to \mathbf{R}$

$$\tilde{l}_{n,p}(f, a) = \left[\frac{n^\gamma}{2d} \int_{-dn^{-\gamma}}^{dn^{-\gamma}} f(x)\,dx - a \right]^P$$

In 2) we have suppressed in the notation the dependence on d and γ.

To the best of our knowledge noone has considered sequences of loss functions given in 2). The motivation behind them is that if we are interested in the local behaviour of f at 0 it might make at least as much sense to look at a shrinking neighborhood of zero rather than just at the point zero.

There are two major obstacles to applying Theorem 1 in this context.

1) The loss functions l_p and $\tilde{l}_{n,p}$ are defined on the functions $f_n(h;\ )$ whereas the loss function in the corollary is defined on h.

2) As mentioned above we need to be able to choose α_n, β_n and K appropriately.

To answer these questions we consider particular classes of **F**. Write $f^k(x)$ for the k^{th} derivative of f.

Let $\mathbf{F}(a, k, M) = \{f: \mathbf{R} \to \mathbf{R}^+;\ f(0) \le a,\ \int f = 1,\ \sup_x |f^k(x)| \le M\}$.

First fix some $f_0 \in \mathbf{F}(a, k, M)$ such that

1) $f_0(0) = b < a$

2) $|f^k(x)| < M$ for all x

3) For some $\varepsilon > 0$ $f_0(x) = f_0(0)$ for $|x| \le \varepsilon$.

Conditions 1 and 2 make sure that f_0 is an interior point of the set $\mathbf{F}(a, k, M)$. Condition 3 facilitates the construction of the perturbations given below.

Let

$$K(c) = \{h: |h^k(x)| \le 1,\ h(x) = 0 \text{ if } |x| \ge c\}.$$

We impose the condition that $h(x) = 0$ for $|x| \ge c$ primarily to make K(c) compact and hence insure strong convergence.

Now let $\alpha_n = M^{-\frac{1}{2k+1}} (f_0(0))^{\frac{k+1}{2k+1}} n^{\frac{k}{2k+1}}$, $\beta_n = M^{\frac{2}{2k+1}} (f_0(0))^{\frac{-1}{2k+1}} n^{\frac{1}{2k+1}}$. A few simple calculations which we leave to the reader (partly made easy by requiring $h(x) = 0$ for $|x| \ge 0$), show that for some N, $f_n(h; x) \in \mathbf{F}(a, k, M)$ for all h when $n \ge N$. Note also that $\alpha_n^2 \beta_n = f_0(0) \cdot n$. These same calculations should also give the reader a good idea of why we imposed 1) and 2) on f_0.

Now l_p and $\tilde{l}_{n,p}$, are all defined on $f_n(h;x)$. To apply Corollary 1 we need to have a loss function on h. Of course l_p is also defined on $h \in K(c)$ by

$$l_p(h,a) = (h(0) - a)^P.$$

We may also define $\tilde{l}$ on $h \in K(c)$ by

$$\tilde{l}(h,a) = \left[\frac{1}{2d}\int_{-d}^{d} h(x) - a\right]^P$$

Cor 1 then yields

$$\lim_{n\to\infty} \inf_{\delta_n} \sup_{h\in K(c)} E_h l(h,\delta_n) = \inf_{\delta} \sup_{h\in K(c)} E_h^G l(h,\delta) \tag{8}$$

for l being l_p or $\tilde{l}$.

We shall now connect equation (8) to the problem of estimating $f_n(h;\ )$ instead of h. We will examine our two loss functions, one at a time.

a) l_p **loss:**

Note that $l_p(f_0(x)(1 + \frac{h(\beta_n x)}{\alpha_n}), f_0(0)(1 + \frac{\delta}{\alpha_n}))$

$$= \left[\frac{f_0(0)}{\alpha_n}\right]^P l_p(h,\delta).$$

Hence $\overline{\lim_{n\to\infty}} \left[\frac{\alpha_n}{f_0(0)}\right]^P \inf_{\delta_n} \sup_{h\in K(c)} E_h l(f_0(x)(1 + \frac{h(\beta_n x)}{\alpha_n}), \delta_n)$

$$= \overline{\lim_{n\to\infty}} \inf_{\delta_n} \sup_{h\in K(c)} E_h l(h,\delta_n)$$

$$= \inf_{\delta} \sup_{h\in K(c)} E_h^G l(h,\delta).$$

Furthermore since $h_n = O(n^{-1})$ it follows that

$$\overline{\lim_{n\to\infty}} \left[\frac{\alpha_n}{f_0(0)}\right]^P \inf_{\delta_n} \sup_{h\in K(c)} E_h l(f_0(x)(1 + \frac{h(\beta_n x)}{\alpha_n} - h_n), \delta_n)$$

$$= \inf_{\delta} \sup_{h\in K(c)} E_h^G l(h,\delta).$$

Now let $c \to \infty$ and note that $\alpha_n = M^{-\frac{1}{2k+1}} (f_0(0))^{\frac{k+1}{2k+1}} n^{\frac{k}{2k+1}}$ and we get

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} M^{-\frac{P}{2k+1}} (f_0(0))^{\frac{-Pk}{2k+1}} n^{\frac{Pk}{2k+1}} \inf_{\delta_n} \sup_{h\in K(c)} E_h l(f_n(h), \delta_n)$$

$$= \lim_{c\to\infty} \inf_{\delta} \sup_{h\in K(c)} E_h^G l\,(h,\delta).$$

Note that since for $h \in K(c)$, $f_n(h) \in F(a,k,M)$ this last equation immediately yields on taking $\sup_{f\in F} f(0)$

(9)
$$\overline{\lim_{n\to\infty}}\, M^{\frac{-P}{2k+1}}\, a^{\frac{-Pk}{2k+1}}\, n^{\frac{Pk}{2k+1}} \inf_{\delta_n} \sup_{f\in F(a,k,M)} E_f l\,(f,\delta_n)$$
$$\geq \lim_{c\to\infty} \inf_{\delta} \sup_{h\in K(c)} E_h^G l\,(h,\delta).$$

b) $\tilde{l}_{n,p}$.

For the class $F(a,k,M)$ we shall take γ in the loss $\tilde{l}_{n,p}$ to be such that $\gamma \geq \frac{1}{2k+1}$. The reader can explore the case $\gamma < \frac{1}{2k+1}$. We then note that

(10)
$$\frac{n^\gamma}{2d} \int_{-dn^{-\gamma}}^{dn^{-\gamma}} f_n(h;x)\,dx = \frac{n^\gamma}{2d} \int_{-dn^{-\gamma}}^{dn^{-\gamma}} f_0(x) + f_0(x)\frac{h(\beta_n x)}{\alpha_n} - f_0(x)\,h_n\,dx$$

where α_n and β_n were defined earlier. Note that since $f_0(x) = f_0(0)$ for $|x| \leq \varepsilon$ we may replace the R.H.S. of (10) for sufficiently large n by

$$f_0(0)\,[\,1 + \frac{1}{\alpha_n}\frac{n^\gamma}{2d} \int_{-dn^{-\gamma}}^{dn^{-\gamma}} h(\beta_n x)\,dx - h_n\,].$$

Now $\beta_n = M^{\frac{2}{2k+1}}\,(f_0(0))^{\frac{-1}{2k+1}}\,n^{\frac{1}{2k+1}}$. Let $y = \beta_n x$. Then

$$\frac{n^\gamma}{2d} \int_{-dn^{-\gamma}}^{dn^{-\gamma}} h(\beta_n x)\,dx = \frac{n^\gamma}{2d} \int_{-dn^{-\gamma+\frac{1}{2k+1}}}^{dn^{-\gamma+\frac{1}{2k+1}}} h(y)\,\frac{dy}{n^{1/2k+1}}$$

(11)
$$= \frac{n^{\gamma-\frac{1}{2k+1}}}{2d} \int_{-dn^{-\gamma+\frac{1}{2k+1}}}^{dn^{-\gamma+\frac{1}{2k+1}}} h(y)\,dy.$$

Note that if $\gamma > \frac{1}{2k+1}$ as $n \to \infty$ (11) has limit $h(0)$. Hence if $\gamma > \frac{1}{2k+1}$ (10) equals $f_0(0)\,(1 + \frac{h(0)}{\alpha_n} + o(\alpha_n^{-1}))$ as $n \to \infty$ and asymptotically the problem is equivalent to the case of l_p loss. If $\gamma = \frac{1}{2k+1}$ a similar analysis yields (11) equal to

$$f_0(0)\,(1 + \frac{1}{\alpha_n} \cdot \frac{1}{2d}\int_{-d}^{d} h(y)\,dy)\,(1 + o(1)).$$

Hence as in a) we have when $\gamma = \frac{1}{2k+1}$

(12) $$\overline{\lim_{n\to\infty}} \left[\frac{\alpha_n}{f_0(0)}\right]^p \inf_{\delta_n} \sup_{h\in K(c)} E_h \tilde{l}_{n,p}(f_n(h), \delta_n) = \inf_{\delta} \sup_{h\in K(c)} E_h^G \tilde{l}(h, \delta)$$

and

(13) $$\overline{\lim_{n\to\infty}} M^{-\frac{P}{2k+1}} a^{-\frac{Pk}{2k+1}} n^{\frac{Pk}{2k+1}} \inf_{\delta_n} \sup_{f\in F(n,k,M)} E_f \tilde{l}_{n,p}(f, \delta_n)$$

$$\geq \lim_{c\to\infty} \inf_{\delta} \sup_{h\in K(c)} E_h^G l(h, \delta)$$

Example 2: Let $F = \{f: R \to R,\ f \geq 0,\ \int f = 1,\ f$ absolutely continuous, $\int f'^2 \leq 1\}$ Wahba (1975) found a variety of sequences of estimators, say $\{\delta_n\}$, satisfying

(14) $$0 < \overline{\lim_{n\to\infty}} n^{1/2} \sup_{f\in F} E_f(f(0) - \delta_n)^2 < \infty.$$

Furthermore Wahba (1975) showed that for any $\varepsilon > 0$

(15) $$\overline{\lim_{n\to\infty}} n^{\frac{1}{2}+\varepsilon} \inf_{\delta_n} \sup_{f\in F} E_f(f(0) - \delta_n)^2 > 0.$$

We will now use theorem 1 to show that the best asymptotic rate of convergence for a sequence of estimators is $n^{1/2}$. In other words

(16) $$\overline{\lim_{n\to\infty}} n^{1/2} \inf_{\delta_n} \sup_{f\in F} E_f(f(0) - \delta_n)^2 > 0.$$

First take

$$f_0(x) = \begin{cases} 17/8 + x & -17/8 \leq x \leq -15/8 \\ 1/4 & -15/8 \leq x \leq 15/8 \\ 17/8 - x & 15/8 \leq x \leq 17/8 \\ 0 & |x| \geq 17/8 \end{cases}$$

Then $\int f_0(x)\,dx = 1$ and $\int f'^2(x)\,dx = \frac{1}{2}$

$$\text{Let} \quad g(x) = \begin{cases} \frac{1}{8} - |x|, & |x| \leq \frac{1}{8} \\ 0 & ,|x| > \frac{1}{8} \end{cases}$$

Then $\int g'^2(x)\,dx = \frac{1}{4}$. Let $K = \{\theta g:\ 0 \leq \theta \leq 1\}$ and $\alpha_n = n^{1/4}$, $\beta_n = \frac{1}{4} n^{1/2}$. Then $\alpha_n^2 \beta_n = f_0(0)\, n$ and for large n

$$f_n(\theta g; x) \in \mathbf{F}$$

where $f_n(h; x)$ is defined by (4). Moreover $f_n(\theta g; 0) = f_0(0)(1 + \frac{\theta g(0)}{\alpha_n})(1 + o(1))$.

Hence

(17) $$\overline{\lim_{n\to\infty}} \left\{\frac{n^{1/4}}{f_0(0)}\right\}^2 \inf_{\delta_n} \sup_{f \in \mathbf{F}} E_f(f(0) - \delta_n)^2$$

$$\geq \inf_{\delta} \sup_{0 \leq \theta \leq 1} E_\theta^G(\theta g(0) - \delta)^2 > 0$$

Section 4. Proof of Theorem.

Proof.

Let X_i, $i = 1, \ldots, n$ be i.i.d. each with density f_0. Let

(18) $$Q_n = \Sigma\left[\frac{h(\beta_n X_i)}{\alpha_n} - h_n\right]$$

(19) $$R_n = \frac{1}{2}\Sigma\left[\frac{h(\beta_n X_i)}{\alpha_n} - h_n\right]^2.$$

Simple calculations show that

(20) $$\lim_{n\to\infty} E Q_n = 0$$

(21) $$\lim_{n\to\infty} \operatorname{var} Q_n = \int h^2(y)\, dy$$

(22) $$\lim_{n\to\infty} ER_n = \frac{1}{2}\int h^2(y)\, dy$$

(23) $$\lim_{n\to\infty} \operatorname{var} R_n = 0.$$

Finally note that since $h_n = O(n^{-1})$

(24) $$E\left|\frac{h(\beta_n X_i)}{\alpha_n} - h_n\right|^j = O_P\left(\frac{1}{\alpha_n^j \beta_n}\right) = o(n^{-1}) \quad \text{for } j \geq 3 \quad \text{and so}$$

(25) $$\log \prod_1^n \frac{f_n(h; X_i)}{f_0(X_i)} - Q_n - R_n = O_P(1).$$

It then follows immediately from the asymptotic expansion given in (25) and the results in (20)-(23) that the experiments $\{P_h^n : h \in H\}$ converge weakly to the standard Gaussian experiment $\{P_h : h \in H\}$.

Acknowledgements. This work is based on part of the author's doctoral dissertation, which was written at Cornell University under the supervision of Professor L.D. Brown. His encouragement is gratefully acknowledged.

References:

Donoho, D.L. and Liu, R.C. (1988). Geometrizing rates of convergence III. Technical Report, University of California, Berkeley.

Le Cam, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer-Verlag.

Millar, P.W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z.W.* **48**, 233-252.

Romano, J.P. (1989). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16**, 629.

Wahba, G. (1975). Optimal Convergence Properties of Variable Knot, Kernel and Orthogonal Series Methods for Density Estimation. *Ann. Statist.*

Department of Statistics
University of California
Berkeley, California