Invariance and Rescaling of Infinite Dimensional Gaussian Shift Experiments

By

Mark G. Low

Technical Report No. 222
October 1989

Department of Statistics
University of California
Berkeley, California

Invariance and Rescaling of Infinite Dimensional Gaussian Shift Experiments

By

Mark G. Low

Introduction.

Finite dimensional Gaussian shift experiments often arise as limiting experiments in many classical parametric statistical problems. Infinite dimensional Gaussian shift experiments have been exploited as limit experiments in a variety of nonparametric problems. See Millar [1979] for examples arising from estimating distribution functions.

Certain pointwise estimation problems arising in density estimation and nonparametric regression have also been shown to have an infinite dimensional Gaussian shift experiment as a natural limiting experiment. The two most important cases are given by

1)
$$dX_t = f(t)\,dt + dW_t \quad 0 \le t < \infty$$
$$f \in \mathbf{F} \subseteq L_2[0,\infty]$$

and where W_t is Brownian motion

2)
$$dX_t = f(t)\,dt + dW_t \quad -\infty < t < \infty$$
$$f \in \mathbf{F} \subseteq L_2(-\infty,\infty)$$

and $W_t = \begin{cases} W_t^1 & t \ge 0 \\ W_{-t}^2 & t < 0 \end{cases}$ where W_t^1 and W_t^2 are independent Brownian motion. In both cases the estimation problem is to estimate $f(0)$. This estimation problem is also of course of interest in its own right. Related problems arising in other nonparametric situation can be found in work of Brown and Farrell [1987], Donoho and Liu [1988] and Romano [1989].

This paper takes a new look at the signal estimation problem given by 1). 2) can be handles entirely analogously. In particular we show how to reduce the problem of estimating $f(0)$ from n observations to the problem of estimating $f(0)$ from one observation. As a trivial consequence asymptotic minimax rates for a large class of **F**'s and many loss functions can be found. Estimators achieving these rates can be given. Examples are given in Section 3. They can be compared to Farrell's rates for density estimation problems Farrell [1972]. This paper is however not primarily concerned with rate questions. In fact it is the exact relationship between n observations and one observation that is important. In Section 2 Theorem 1 gives a precise statement of the invariance hinted at above. Since the assumptions made on the parameter spaces in Theorem 1 are somewhat unusual we give some concrete applications in Section 3. Included in this section we show how Theorem 1 can also yield the functional dependence of the minimax risk over a whole class of parameter spaces $F(M)$. We leave a proof of Theorem 1 to Section 4.

Section 2.

Our main interest is to compare the problems of one and n observations from the model given in (1). However if $X_t^1, X_t^2, \ldots, X_t^n$ are independent then

$$Y_t = \frac{X_t^1 + \cdots + X_t^n}{n} \tag{1}$$

is a stochastic process satisfying

$$dY_t = f(t)\,dt + \frac{1}{\sqrt{n}}\,dW_t \quad [0 \le t < \infty). \tag{3}$$

Hence we may instead just compare the problems of observing one observation from 3) for different values of n. (3) induces a statistical experiment (for each n) when we let $f \in \mathbf{F} \subseteq L^2[0,\infty)$.

It will also be convenient to introduce a second sequence of statistical experiments given by

$$dY_t^n = \frac{f(\beta_n t)}{\alpha_n}\,dt + \frac{1}{\sqrt{n}}\,dW_t, \quad [0 \le t < \infty] \tag{4}$$

where $\alpha_n^2 \beta_n = n$ and $f \in \mathbf{F}$. We shall sometimes write $Y^n(t)$ for Y_t^n when the resulting expression is easier to read.

Remark: If we take $\alpha_1 = \beta_1 = 1$ then (3) and (4) are the same for $n = 1$, but we do not only restrict to that case.

The importance of this second sequence of experiments will be clear from the lemma given below and the remarks following it. Its proof is clear and so is left to the reader.

Lemma: Suppose $Y^n(t)$ has a distribution given by (4). Then

$$Z(t) = \alpha_n \beta_n Y^n(t/\beta_n) \tag{5}$$

follows a distribution given by

$$dZ(t) = f(t)\,dt + dW_t. \tag{6}$$

Similar if $Z(t)$ has distribution given by (6) then $Y^n(t)$ defined by (5) has distribution given by (4).

Remark:

The lemma establishes a precise equivalence between every pair of experiments in the sequence of experiments given in (4). This equivalence can be thought of in terms of Le Cam's theory of experiments or in terms of Blackwells sufficiency of two experiments. Although these points of view put the lemma into context the reader does not have to be familiar with these ideas to follow this paper.

To connect the lemma with the more interesting sequence of experiments given by (3) we need to take a more decision theoretic viewpoint and introduce loss functions. In fact we will allow the loss function to depend on n subject to the following condition.

Assumption A: We restrict attention to a sequence of loss functions l_n and a fixed loss function l such that

i) $l_n : L_2[0,\infty] \times R \rightarrow R^+$
$l : L_2[0,\infty] \times R \rightarrow R^+$

ii) There is a function $g : R \rightarrow R$ such that if $\alpha_n^2 \beta_n = n$ then

$$l_n\left[\frac{f(\beta_n t)}{\alpha_n}, \frac{a}{\alpha_n}\right] = g(\alpha_n)\, l\,(f, a).$$

Now let δ be an estimator, $\delta : l_2[0,\infty) \rightarrow R$. $E_g^m L_n(f, \delta(X(t))$ is then to be interpreted as taking the expectation, under the model

(7) $$dX_t = g(t)\,dt + \frac{1}{\sqrt{m}} dW_t$$

of the random function $l_n(f, \delta)$. When $g = f$ and $m = n$ this is the risk of the estimator δ with loss function l_n under model (3).

Theorem: Let $(T_n f)(t) = \dfrac{f(\beta_n t)}{\alpha_n}$ where $\alpha_n^2 \beta_n = n$ and for each estimator δ_n let $\tilde{\delta}_n$ be defined by $\dfrac{\tilde{\delta}_n(Z(t))}{\alpha_n} = \delta_n\left[\dfrac{Z(\beta_n t)}{\alpha_n \beta_n}\right]$. Then

(8) $$E_{T_n f}^n\, l_n(T_n f, \delta_n) = g(\alpha_n) E_f^1\, l\,(f, \tilde{\delta}_n)$$

(9) $$\sup_{f_n \in T_n F} E_{f_n}^n\, l_n(f_n, \delta_n) = g(\alpha_n) \sup_{f \in F} E_f^1\, l\,(f, \tilde{\delta}_n)$$

(10) $$\inf_{\delta_n} \sup_{f_n \in T_n F} E_{f_n}^n\, l_n(f_n, \delta_n) = g(\alpha_n) \inf_{\delta} \sup_{f \in F} E_f^1\, l\,(f, \delta)$$

Cor 1: If $T_n \mathbf{F} = \mathbf{F}$ and $T_1 f = f$ for all $f \in \mathbf{F}$ then

(11) $$\inf_{\delta_n} \sup_{f \in F} E_f^n\, l_n(f, \delta_n) = g(\alpha_n) \inf_{\delta} \sup_{f \in F} E_f^1\, l\,(f, \delta)$$

Cor 2:

(12) $$\inf_{\delta_1} \sup_{f \in T_1 F} E_f^1\, l_1(f, \delta_1) = g(\alpha_1) \inf_{\delta} \sup_{f \in F} E_f^1\, l\,(f, \delta).$$

Section 3. Applications.

Example 1. Write $f^k(x)$ for the k^{th} derivative of f. Let $\mathbf{F}(k, M) = \{f \in L_2(0,\infty): |f^k(x)| \le M \vee x\}$ and let

$$l_n \equiv l \text{ satisfying } l(f,a) = (f(0)-a)^q.$$

Furthermore take $\alpha_n = n^{k/2k+1}$, $\beta_n = n^{+1/2k+1}$. Then $\alpha_n^2\beta_n = n$, $T_n\mathbf{F} = \mathbf{F}$ and $l\left[\frac{f(\beta_n t)}{\alpha_n}, \frac{a}{\alpha_n}\right] = \frac{1}{\alpha_n^q} l(f,a)$.

The assumptions of Cor 1 then clearly hold and yields

$$\inf_{\delta_n} \sup_{f\in \mathbf{F}(k,M)} E_f^n(f(0)-\delta_n)^q = \frac{1}{n^{qk/2k+1}} \inf_{\delta} \sup_{f\in \mathbf{F}(k,M)} E_f^1(f(0)-\delta)^q.$$

Also let $\alpha_1 = M^{-\frac{1}{2k+1}}$, $\beta_1 = M^{2/2k+1}$. Then $T_1\mathbf{F}(k,1) = \mathbf{F}(k,M)$ and cor 2 yields

$$\inf_{\delta} \sup_{f\in \mathbf{F}(k,M)} E_f^1(f(0)-\delta)^q = M^{1/2k+1} \inf_{\delta} \sup_{f\in \mathbf{F}(k,1)} E_f^1(f(0)-\delta)^q$$

Example 2. As a simple case of an application with a varying l_n take the parameter space to be

$$\mathbf{F}(1,M),\ l(f,a) = (f(1)-a)^2,\ \alpha_n = n^{1/3},\ \beta_n = n^{1/3},$$

and $l_n(f,a) = (f(n^{-1/3})-a)^2$. Then Cor 1 gives

$$\inf_{\delta_n} \sup_{f\in \mathbf{F}(1,M)} E_f^n(f(n^{-1/3})-\delta_n)^q = \frac{1}{n^{2/3}} \inf_{\delta} \sup_{f\in \mathbf{F}(1,M)} E_f^1(f(1)-\delta)^2$$

Example 3. Let $\mathbf{G}(M) = \{f \in L^2[0,\infty): \int f'^2 \le M\}$. Take $l_n \equiv l$ such that $l(f,a) = (f(0)-a)^q$. Let $\alpha_n = n^{1/4}$, $\beta_n = n^{1/2}$. Then $T_n\mathbf{G}(M) = \mathbf{G}(M)$ and Cor 1 yields

$$\inf_{\delta_n} \sup_{f\in \mathbf{G}(M)} E_f^n(f(0)-a)^q = \frac{1}{n^{q/4}} \inf_{\delta} \sup_{f\in \mathbf{G}(M)} E_f^1(f(0)-a)^q.$$

Section 4. Proof of Theorem.

Proof.

The lemma immediately yields

$$E_{T_n f}^n l_n(T_n f, \delta_n(X(t))) = E_f^1 l_n\left(T_n f, \delta_n\left[\frac{X(\beta_n t)}{\alpha_n\beta_n}\right]\right) \tag{13}$$

since $\frac{\delta_n(X(t))}{\alpha_n} = \delta_n\left[\frac{X(\beta_n t)}{\alpha_n\beta_n}\right]$ (13) is equal to

(14) $$E_f^1\, l_n\left[T_n f, \frac{\tilde{\delta}_n(X(t))}{\alpha_n}\right].$$

By assumption A

$$l_n\left[T_n f, \frac{\tilde{\delta}_n(X(t))}{\alpha_n}\right] = g(\alpha_n)\, l\,(f, \tilde{\delta}_n(X(t)))\,.$$

Hence (14) is equal to

(15) $g(\alpha_n)\, E_f^1\, l\,(f, \tilde{\delta}_n)$ which is the same as (8).

This establishes (8). (9) follows immediately upon taking sup's. Likewise (10) follows on taking inf's.

Proof (Cor 1 and Cor 2).

Cor 1 is just a rewriting of (10) under the assumption $T\mathbf{F} = \mathbf{F}$. Cor 2 is just a rewriting of (10) for $n = 1$.

References

Brown, L.D. and Farrell, R.H. (1987). A lower bound for the risk in estimating the value of a probability density. Cornell Statistics Center preprint series.

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

Farrell, R.H. (1972). On the best obtainable rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43**, 170-180.

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**, 629-648.

Department of Statistics
University of California
Berkeley, California