NONPARAMETRIC ESTIMATION OF PROBABILITY DENSITY FUNCTIONS*

by

JUGAL KISHORE GHORAI
Purdue University

Department of Statistics
Division of Mathematical Science
Mimeograph Series #503

August 1977

*This research was supported in part by NSF Grant MPS74-07836 A01 at Purdue University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FORTRAN PROGRAM

INTRODUCTION

The topic of nonparametric estimation of probability density functions has been discussed by many authors. Various methods of estimation have been suggested and their properties have been studied. Some of the well known methods are kernel estimates, orthogonal series estimates, spline estimates, fourier-integral estimates, nearest-neighbor estimates. Kernel estimates were studied in detail by Parzen (1962) and since then many properties of these estimates have been derived. Spline estimates were first introduced by Boneva-Kendall-Stefanov (BKS) (1971). Mean square convergence properties were studied by Wahba (1975). She also showed that under appropriate conditions kernel estimates, Kronmal-Tartar type orthogonal series estimate, spline estimates and ordinary histogram-type estimates, all have the same MSE convergence rate.

All these estimates, except the maximum likelihood estimate, are ad-hoc type estimates. For the estimation of a density function given n-observations, the likelihood or its logarithm is relevant. But a naive application of maximum likelihood methods would make the estimate, the mean of a set of Dirac functions at the n observations and gives a value $+\infty$ to the likelihood functional. Hence in any class of functions which has the property that it is possible to construct a sequence of functions which integrate to one, are non-negative and converge point-wise to a Dirac-delta spike, the likelihood will be unbounded and a maximum likelihood estimate will not exist. However if the domain of the likelihood functional is restricted appropriately, a maximum likelihood estimate might exist. For example, if we restrict the domain to the class of unimodal densities with a fixed mode then the maximum likelihood estimate exists. The properties of such estimates were studies by Rao (1969) and Reiss (1973). The fact that in general the nonparametric maximum likelihood estimate does not exist implies that the unrestricted domain must necessarily lead to unsmooth estimates and a numerically ill-posed problem. This leaves the practioners with the following dilemma: for small restricted domain he has no flexibility and the solution will greatly be influenced by the choice of the domain, while for the unrestricted domain the solution must necessarily approximate a linear combination of Dirac-delta spikes, be unsmooth and create numerical problems.

For these reasons and others based on heuristic Bayesian considerations, Good and Gaskin (1971) suggested adding a penalty term to the likelihood which would penalize unsmooth densities. They suggested two specific penalty terms. They also suggested an alternate approach for constructing the penalized maximum likelihood estimate which avoids nonnegativity conditions. But they do not show the equivalence of these two approaches nor do they show the existence of the penalized maximum likelihood estimate.

These problems are considered by Montricher, Tapia and Thompson (1975 a,b). Specifically they establish a general existence and uniqueness theory for a large class of penalized maximum likelihood estimates. They show that a well known class of reproducing kernel Hilbert space (Sobolev Space) leads quite naturally to the penalized maximum likelihood estimate which are polynomial splines with knots at the sample points. They show that in the case of Good's second penalty function, the two approaches need not lead to the same solution.

In Chapter I we derive the solution independently. We also discuss some of the properties of the estimator. It is shown that the penalized maximum likelihood estimate behaves somewhat like a kernel estimate with double exponential kernel. In the rest of Chapter I we discuss the possibility of generalizing this approach to more general type of penalty functions and also to higher-dimensions.

In Chapter II we discuss the problem of Bayesian estimation of density function. We show that under certain conditions it is possible to define posterior measure and posterior mean. It is shown that under very mild condition on the prior the point-wise and integrated Bayes risk of the Bayes estimator with squared error loss tends to zero as $n \to \infty$. As an example, we consider two specific prior distributions for the purpose of Bayes estimation. It is shown that one of them has the property that Bayes estimate is Bayes risk consistent.

CHAPTER I

MAXIMUM PENALIZED LIKELIHOOD ESTIMATION OF DENSITY FUNCTION

The non-existence of maximum likelihood estimates in general in the density estimation problem led Good and Gaskin (1971) to introduce maximum penalized likelihood estimate. They suggested that instead of maximizing the likelihood function directly one should maximize a penalized likelihood. This penalized likelihood can be obtained by adding an appropriate penalty term to the logarithm of the likelihood function. Further research in this line was done by Montricher, Tapia and Thompson (1975 a,b). They derived some general existence and uniqueness theorems.

In section 1.1 we derive the estimate independently of Montricher, Tapia and Thompson by standard calculus of variations methods. We also give a heuristic argument to show that in the case of Good and Gaskin's first penalty function, MPLE can be looked as the posterior mode with respect to an improper prior. This interpretation also suggests that α, the coefficient of the penalty function, should not be fixed if we want an estimator which remains smooth with increasing sample size.

In section 1.2 we present an algorithm for the computation of the estimate. We show also the convergence of the iterative procedure.

In section 1.3 we discuss a Newton-type computational procedure for simultaneous estimation of all the parameters involved. Although we could not show the convergence of this procedure theoretically but in all the numerical examples it gave convergence within 3 to 4 iterations. We have included a FORTRAN program for this algorithm at the end. In section 1.4 we briefly discuss the possibility of a more general penalty function involving only the first derivative. In section 1.5 we discuss the possibility of generalizing to higher dimensional cases. In particular we give an example to show that in the bivariate case penalty function involving only the square of the first partial derivatives (the direct generalization of the Good-Gaskin approach) will not be able to eliminate rough function and hence again MPLE will not exist. We also give an example to show that if the penalty function involves higher derivatives, the MPLE need not be unique. Finally in section 1.6 we give a discussion of the consistency of the estimate which indicates that we have reason to believe that the estimate is consistent, although we could not obtain a rigorous proof.

1.1. Maximum Penalized Likelihood Estimation Using Good's First Penalty Function

Let $X_1, X_2, \ldots X_n$ be iid random variables with common density f. The logarithm of likelihood is given by

$$L(f|X_1 \ldots X_n) = \sum_{i=1}^{n} \log f(X_i)$$

Let Ω be the class of density functions defined over R. A penalty function $\Phi: \Omega \to R$ is a real valued functional defined over the class of density function. The functional $\psi(\cdot|\alpha): f \to (L(f) - \alpha\,\Phi(f))$ is called the logarithm of penalized likelihood function.

<u>Definition 1.1.1</u>. Any measurable mapping $f_n: R^n \to \Omega$, which maximizes $\psi(\cdot|\alpha)$ over Ω is called a maximum penalized likelihood estimate (MPLE).

Let us consider the penalty function $\Phi(f) = \int_{-\infty}^{\infty} \frac{[f'(x)]^2}{f(x)}\,dx$, suggested by Good and Gaskin (1971 a). Then we have the following maximization problem:

Maximize: $\psi(f|\alpha) = \sum_{i=1}^{n} \log f(X_i) - \alpha\,\Phi(f).$

subject to

(1.1.3)
$$\begin{cases} \int_{-\infty}^{\infty} f(x)dx = 1 & \\ f(x) \geq 0 & \text{for all } x \\ f(X_i) > 0 & \text{for all } i = 1, \ldots n. \\ \Phi(f) < \infty & \end{cases}$$

Good suggested the substitution $f = g^2$ and considered the following modified maximization problem:

Maximize: $\psi(g|\alpha) = 2 \sum_{i=1}^{n} \log |g|(X_i) - 4\alpha \int_{-\infty}^{\infty} g'^2(x)dx$

subject to

(1.1.4)
$$\begin{cases} \int_{-\infty}^{\infty} g^2(x)dx = 1 & \\ \int_{-\infty}^{\infty} g'^2(x)dx < \infty & \\ |g|(X_i) > 0 & \text{for all } i = 1, 2 \ldots n. \end{cases}$$

Now using a Lagrange multiplier the above maximization problem can be expressed as:

Maximize: $$\psi_\rho(g|\alpha) = 2 \sum_{i=1}^{n} \log |g|(X_i) - 4\alpha \int_{-\infty}^{\infty} g'^2(x)dx - \rho\int g^2(x)dx$$

subject to (1.1.5)
$$\begin{cases} \int_{-\infty}^{\infty} g'^2(x) < \infty \\ |g|(X_i) > 0 \qquad \text{for } i = 1, 2 \ldots n. \end{cases}$$

Since $\psi_\rho(g|\alpha)$ does not depend on the order of the observations, without loss of generality, we assume that

$$X_1 < X_2 < \ldots < X_n \qquad \text{a.s.}$$

Therefore for given $g(X_1), \ldots, g(X_n)$, the function which maximizes $\psi_\rho(g|\alpha)$ must minimize

$$(1.1.6) \qquad \int_{X_i}^{X_{i+1}} [4\alpha\, g'^2(x) + \rho\, g^2(x)]dx \qquad \text{for } i = 0, 1, \ldots n$$

where $X_o = -\infty$ and $X_{n+1} = \infty$.

<u>Theorem 1.1.1</u>. (a) <u>The function which solves</u> (1.1.4) <u>has the following form,</u>

$$g_\rho(x|\alpha) = \begin{cases} A_o\, e^{\lambda x} & \text{if } x \le X_1 \\ A_i\, e^{\lambda x} + B_i\, e^{-\lambda x} & \text{if } X_i \le x \le X_{i+1} \\ B_n\, e^{-\lambda x} & \text{if } x \ge X_n \end{cases}$$

<u>where</u> $\lambda = \sqrt{\frac{\rho}{4\alpha}}$. <u>Further</u> $g_\rho(x|\alpha)$ <u>can be expressed as</u>

$$g_\rho(x|\alpha) = \sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\, g_\rho(X_j|\alpha)} .$$

(b) The minimizing function $g_\rho(x|\alpha)$ is unique and has same sign for all x.

(In the sequel dependence of g_ρ on α will be suppressed sometimes.)

Proof: Let $I_i = [X_i , X_{i+1}]$

and

$$H_i = \begin{cases} \eta: I_i \to R & \text{such that} \\ \eta(X_i) = \eta(X_{i+1}) = 0 \\ \eta'(t) & \text{exists for all } t \in I_i \\ \int_{I_i} \eta^2(x)dx < \infty \\ \text{and } \int_{I_i} \eta'^2(x)dx < \infty \end{cases}$$

Then consider the difference

$$\int_{I_i} \{4\alpha\, (g'(x) + \varepsilon\, \eta'(x))^2 + \rho(g(x) + \varepsilon\, \eta(x))^2\}dx$$

$$- \int_{I_i} \{4\alpha\, g'^2(x) + \rho g^2(x)\}dx$$

$$(1.1.7) \quad = 2\varepsilon \int_{I_i} \{4\alpha\, g'(x)\, \eta'(x) + \rho g(x)\, \eta(x)\}dx$$

$$+ \varepsilon^2 \int_{I_i} \{4\alpha\, \eta'^2(x) + \rho\eta^2(x)\}dx.$$

If there exists a function g_ρ which extremizes

$\int_{I_i} [4\alpha\, g'^2(x) + \rho g^2(x)]dx$, then for this g_ρ

(1.1.8) $$\int_{I_i} [4\alpha\, g_\rho'(x)\eta'(x) + \rho g_\rho(x)\eta(x)]dx = 0$$

for every $\eta \in H_i$

But $$\int_{I_i} \{4\alpha\, g_\rho'(x)\eta'(x) + \rho g_\rho(x)\eta(x)\}dx$$

$$= 4\alpha\, g_\rho'\eta\Big|_{X_i}^{X_{i+1}} - \int_{I_i} 4\alpha\, g_\rho''(x)\,\eta(x)dx + \int_{I_i} \rho g_\rho(x)\eta(x)dx$$

$$= \int_{I_i} [\rho g_\rho(x) - 4\alpha\, g_\rho''(x)]\eta(x)dx \qquad \text{Since } \eta(X_i) = \eta(X_{i+1}) = 0$$

Thus (1.1.8) implies

$$\int_{I_i} \{\rho g_\rho(x) - 4\alpha\, g_\rho''(x)\}\,\eta(x)dx = 0$$

for every $\eta \in H_i$,

and hence $\rho g_\rho(x) - 4\alpha\, g_\rho''(x) = 0$ for all $x \in I_i$,

(1.1.9) from which $g_\rho''(x) - \frac{\rho}{4\alpha}\, g_\rho(x) = 0$ for all $x \in I_i$

The general solution of the differential equation in (1.1.9) is

(1.1.10) $$g_\rho(x) = A_i\, e^{\lambda x} + B_i e^{-\lambda x}\ ,\quad x \in I_i$$

where $\lambda = \sqrt{\frac{\rho}{4\alpha}}$ and A_i, B_i are constants to be determined to satisfy other side conditions. If $\rho < 0$ then λ is imaginary and the solution will be periodic. This implies that the integral of g_ρ^2 over the end intervals will be infinite and hence g_ρ^2 can not be a density function.

Hence $\rho > 0$. Therefore g_ρ is unique since A_i and B_i are uniquely determined from the conditions

$$(1.1.11) \qquad \begin{aligned} g_\rho(X_i) &= A_i e^{\lambda X_i} + B_i e^{-\lambda X_i} \\ \text{and} \quad g_\rho(X_{i+1}) &= A_i e^{\lambda X_{i+1}} + B_i e^{-\lambda X_{i+1}} \end{aligned}$$

Now $\rho > 0$ implies the second term in (1.1.7) is positive and hence $g_\rho(\cdot|\alpha)$ actually minimizes (1.1.6). The expression for g_ρ in the end intervals can be obtained by letting one end point tend to $-\infty$ or $+\infty$.

Now to get the final form of g_ρ we have from (1.1.11) and

$$(1.1.12) \qquad \begin{cases} g_\rho(X_{i+1}) = A_{i+1}\, e^{\lambda X_{i+1}} + B_{i+1}\, e^{-\lambda X_{i+1}} \\ g_\rho(X_{i+2}) = A_{i+1}\, e^{\lambda X_{i+2}} + B_{i+1}\, e^{-\lambda X_{i+2}} \end{cases},$$

$$A_i = \left[g_\rho(X_i)\, e^{-\lambda X_{i+1}} - g_\rho(X_{i+1})\, e^{-\lambda X_i} \right] / D_i$$

$$B_i = \left[g_\rho(X_{i+1})\, e^{\lambda X_i} - g_\rho(X_i)\, e^{\lambda X_{i+1}} \right] / D_i$$

$$\frac{\partial A_i}{\partial g_\rho(X_{i+1})} = -e^{-\lambda X_i}/D_i$$

$$\frac{\partial B_i}{\partial g_\rho(X_{i+1})} = e^{\lambda X_i}/D_i$$

$$\frac{\partial A_{i+1}}{\partial g_\rho(X_{i+1})} = e^{-\lambda X_{i+2}}/D_{i+1}$$

$$\frac{\partial B_{i+1}}{\partial g_\rho(X_{i+1})} = -e^{\lambda X_{i+2}}/D_{i+1}$$

Now since $\int_{-\infty}^{\infty} (4\alpha\, g'^{2}_{\rho}(x) + \rho\, g_\rho^2(x))^2 dx$

$$= \sum_{i=0}^{n} 4\alpha \int_{I_i} [g'^{2}_{\rho}(x) + \lambda^2 g^2(x)]dx$$

and A_i, B_i appear only in

$$\int_{I_i U I_{i+1}} (g'^{2}_{\rho}(x) + \lambda^2\, g_\rho^2(x))dx$$

$$= \lambda \Big\{ A_i^{\,2} \left[e^{2\lambda X_{i+1}} - e^{2\lambda X_i}\right] + B_i^{\,2} \left[e^{-2\lambda X_i} - e^{-2\lambda X_{i+1}}\right]$$
$$+ A_{i+1}^2 \left[e^{2\lambda X_{i+2}} - e^{2\lambda X_{i+1}}\right] + B_{i+1}^2 \left[e^{-2\lambda X_{i+1}} - e^{-2\lambda X_{i+2}}\right] \Big\}$$

Therefore differentiating

$$2 \sum_{i=1}^{n} \log |g_\rho|(X_i) - 4\alpha \int_{-\infty}^{\infty} \{g'^{2}_{\rho}(x) + \lambda^2 g^2(x)\}dx$$

with respect to $g_\rho(X_{i+1})$ and equating to zero we get,

$$\frac{1}{|g_\rho^{!}(X_{i+1})} = 4\alpha\,\lambda \Big\{ A_i\, e^{-\lambda X_i} \left(e^{2\lambda X_i} - e^{2\lambda X_{i+1}}\right) /D_i$$
$$+ B_i\, e^{\lambda X_i} \left(e^{-2\lambda X_i} - e^{-2\lambda X_{i+1}}\right) /D_i$$

$$(1.1.13)\qquad \begin{aligned} &+ A_{i+1}\, e^{\lambda X_{i+2}} \left(e^{2\lambda X_{i+2}} - e^{2\lambda X_{i+1}}\right) / D_{i+1} \\ &+ B_{i+1}\, e^{\lambda X_{i+2}} \left(e^{-2\lambda X_{i+2}} - e^{-2\lambda X_{i+1}}\right) / D_{i+1} \Big\} \end{aligned}$$

From (1.1.11) and (1.1.10) we get

$$(1.1.14)\qquad e^{\lambda X_{i+1}}(A_{i+1} - A_i) + e^{-\lambda X_{i+1}}(B_{i+1} - B_i) = 0$$

Since the coefficients in the expression of $g_\rho(x|\alpha)$ are unique it is sufficient to exhibit one set of A's and B's which satisfies these equations.

The form of g_ρ, the relation (1.1.12) and $A_n = B_0 = 0$ suggest that A's should be decreasing and B's should be increasing. It turns out that the following choice of A_i, B_i, $i = 1, \ldots n$, satisfies all requirements.

Let
$$A_i = \frac{e^{-\lambda X_{i+1}}}{4\sqrt{\alpha\,\rho}\, g_\rho(X_{i+1}|\alpha)} + A_{i+1}$$

and
$$B_i = B_{i+1} - \frac{e^{\lambda X_{i+1}}}{4\sqrt{\alpha\rho}\; g_\rho(X_{i+1}|\alpha)}.$$

Substituting these in (1.1.13) it can be shown that right hand side of (1.1.13) is equal to the left hand side.

Therefore we can express $g_\rho(x|\alpha)$ as

$$g_\rho(x|\alpha) = \sum_{j=i+1}^{n} \frac{e^{-\lambda X_j}}{4\sqrt{\alpha\rho}} \; \frac{e^{\lambda x}}{g_\rho(X_j|\alpha)} + \sum_{j=1}^{i} \frac{e^{\lambda X_j}}{4\sqrt{\alpha\rho}} \; \frac{e^{-\lambda x}}{g_\rho(X_j|\alpha)}$$

$$= \sum_{j=1}^{n} \frac{e^{-\lambda|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)} .$$

(1.1.15)
$$= \sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{2}}\,|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)}.$$

This completes the proof of part (a).

To show the part (b) we observe that

(1.1.16)
$$\int_{I_i} [4\alpha\, g'^{2}_{\rho}(x) + \rho\, g^2_\rho(x)]dx$$

$$= 4\alpha\lambda \left\{ A_i^2 \left(e^{2\lambda X_{i+1}} - e^{2\lambda X_i}\right) + B_i^2 \left(e^{-2\lambda X_i} - e^{-2\lambda X_{i+1}}\right) \right\}.$$

Also
$$A_i^2 = \left[g^2_\rho(X_{i+1})\, e^{-2\lambda X_i} + g^2_\rho(X_i)\, e^{-2\lambda X_{i+1}} - 2g_\rho(X_i)g_\rho(X_{i+1})\cdot e^{-\lambda(X_i + X_{i+1})}\right] \sinh^{-2}(\lambda\, \Delta X_i)$$

and
$$B_i^2 = \left[g^2_\rho(X_{i+1})\, e^{2\lambda X_i} + g^2{}_\rho(X_i)\, e^{2\lambda X_{i+1}} - 2g_\rho(X_i)g_\rho(X_{i+1})\cdot e^{\lambda(X_i + X_{i+1})}\right] \sinh^{-2}(\lambda\, \Delta X_i)$$

This shows that (1.1.16) is minimized if $g_\rho(X_i)$ and $g_\rho(X_{i+1})$ are of same sign. Therefore $g_\rho(X_1)$, ... $g_\rho(X_n)$ will be of same sign. Now from (1.1.15) we conclude that g_ρ is of same sign for all x.

Remark: 1.1.1. This estimator was derived independently by Montricher, Tapia and Thompson (1975 a,b), using the properties of Reproducing Kernel Hilbert spaces.

<u>Remark 1.1.2</u>. From (1.1.5) it follows that the maximum penalized likelihood estimation of the density f is equivalent to maximizing

$$\psi_\rho^*(g|\alpha) = \log \left[\prod_{i=1}^{n} g^2(X_i) \quad \exp\left(-4\alpha \int_{-\infty}^{\infty} g'^2(x)dx\right)\right]$$

subject to

$$\begin{cases} \int_{-\infty}^{\infty} g'^2(x)dx < \infty & \\ |g|(X_i) > 0 & i = 1, 2 \dots n. \\ \text{and} \quad \int_{-\infty}^{\infty} g^2(x)dx = 1 & \end{cases}$$

where $f = g^2$.

Suppose we take the sample paths of a stationary Gaussian Process, namely Ornstein-Uhlenbeck process, for the g functions then the quantity

$$4\alpha \int_{-\infty}^{\infty} g'^2(x)dx + \rho \int_{-\infty}^{\infty} g^2(x)dx$$

behaves, formally, like the limit of a quadratic form in Gaussian random variable. Hence

$$\exp\left[-4\alpha \int_{-\infty}^{\infty} g'^2(x)dx - \rho \int_{-\infty}^{\infty} g^2(x)\,dx\right]$$

can be looked at as an improper Gaussian prior. To show this we give a heuristic argument as follows.

Let $\{U(t) ; t \in R\}$ be a stationary Gaussian Process with covariance function $\Gamma(s,t) = \sigma^2 \; e^{-\bar{\rho}|t-s|}$. Consider a fixed interval $[a,b]$. Let

$$P = \{a = t_o < t_1 < \dots < t_N = b\}$$

be a partition of the interval [a,b]. Let $||P|| = \max_i |t_{i+1} - t_i|$ denote the norm of the partition P. Then $(U(t_o), U(t_1), \dots U(t_n))' \sim N(0,\Sigma)$ where $\sigma_{ij} = \sigma^2 \exp(-\bar{\rho}|t_j - t_i|)$ and since U is markov, Σ^{-1} is tridiagonal. Let σ^{ij} denote the (i,j) th element of Σ^{-1}, then

$$\frac{1}{\sigma^2} \sigma^{ii} = \begin{cases} \dfrac{1}{1-e^{-2\bar{\rho}\Delta t_1}} & \text{if } i=0 \\[2ex] \dfrac{1}{1-e^{-2\bar{\rho}\Delta t_i}} + \dfrac{e^{-2\bar{\rho}\Delta t_{i+1}}}{1-e^{-2\bar{\rho}\Delta t_{i+1}}} & i=1, \dots N-1 \\[2ex] \dfrac{1}{1-e^{-2\bar{\rho}\Delta t_N}} & i=N \end{cases}$$

$$\frac{1}{\sigma^2} \sigma^{ij} = \begin{cases} -\dfrac{e^{-\bar{\rho}\Delta t_i}}{1-e^{-2\bar{\rho}\Delta t_i}} & j= \quad i+1 \\[2ex] & i=0, \dots N-1 \\[2ex] 0 & j>i+1 \end{cases}$$

and $\sigma^{ji} = \sigma^{ij}$

where $\Delta t_i = |t_i - t_{i-1}|$ $\qquad i=1, \dots . N.$

Now the quadratic form $\underset{\sim}{U}'\Sigma^{-1}\underset{\sim}{U}$ can be expressed as

$$\underset{\sim}{U}'\Sigma^{-1}\underset{\sim}{U} = \frac{1}{\sigma^2} \sum_{i=0}^{N} \sum_{j=0}^{N} U(t_i)U(t_j) \quad \sigma^{ij}$$

$$= \frac{1}{\sigma^2}\left\{ \sum_{i=1}^{N} \frac{1}{1-e^{-2\bar{\rho}\Delta t_i}} (U^2(t_i) + U^2(t_{i-1})) - \sum_{i=1}^{N-1} U^2(t_i) - 2\sum_{i=1}^{N} \frac{e^{-\bar{\rho}\Delta t_i}}{1-e^{-2\bar{\rho}\Delta t_i}} U(t_i)U(t_{i-1}) \right\}$$

$$= \frac{1}{\sigma^2}\left\{ \sum_{i=1}^{N} \frac{e^{-\bar{\rho}\Delta t_i}}{1-e^{-2\bar{\rho}\Delta t_i}} (U(t_i) - U(t_{i-1}))^2 + \sum_{i=1}^{N-1} \frac{1-e^{-\bar{\rho}(\Delta t_i + \Delta t_{i+1})}}{(1+e^{-\bar{\rho}\Delta t_i})(1+e^{-\bar{\rho}\Delta t_{i+1}})} U^2(t_i) + U^2(t_o) + \frac{1}{1+e^{-\bar{\rho}\Delta t_i}} (U^2(t_o) + U^2(t_N)) \right\} .$$

Now if $||P||$ is sufficiently small then the above expression is approximately equal to

$$\frac{1}{\sigma^2}\left\{ \sum_{i=1}^{N} \frac{1}{2\bar{\rho}} \frac{(U(t_i)-U(t_{i-1}))^2}{\Delta t_i} + \sum_{i=1}^{N-1} \frac{\bar{\rho}}{4} (\Delta t_i+\Delta t_{i+1})\, U^2(t_i) + \frac{3}{2} U^2(t_o) + \frac{1}{2} U^2(t_N) \right\} \tag{1.1.17}$$

But $$\int_{t_{i-1}}^{t_i} (U'(t))^2 dt \approx \frac{(U(t_i) - U(t_{i-1}))^2}{\Delta t_i}$$

and $$\int_{t_o}^{t_N} U^2(t)dt = \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} U^2(t)dt$$

$$\approx \sum_{i=1}^{N} (\Delta t_i)\left(\frac{U^2(t_i) + U^2(t_{i-1})}{2}\right)$$

$$= \frac{\Delta t_o}{2} \quad U^2(t_o) + \frac{\Delta t_N}{2} \quad U^2(t_N)$$

$$+ \sum_{i=1}^{N-1} \left(\frac{\Delta t_i + \Delta t_{i+1}}{2}\right) \quad U^2(t_i) \; .$$

Hence if $||P|| \to 0$ then (1.1.17) is

$$\approx \frac{1}{\sigma^2} \Big[\frac{1}{2\bar{\rho}} \int_a^b (U'(t))^2 dt + \frac{\bar{\rho}}{2} \int_a^b U^2(t)dt$$

$$+ \frac{3}{2} U^2(t_o) + \frac{1}{2} U^2(t_N) \Big]$$

Now if we let $a \to -\infty$ and $b \to \infty$ then the above quantity approximates

$$\frac{1}{\sigma^2} \Big[\frac{1}{2\bar{\rho}} \int_{-\infty}^{\infty} (U'(t))^2 dt + \frac{\bar{\rho}}{2} \int_{-\infty}^{\infty} U^2(t)dt \Big] \; ,$$

which is a quadratic form in the process $\{U(t) : t \in R\}$. Therefore if infact g's were sample paths of the process $U(\cdot)$ then

$\exp \left\{ -4\alpha \int_{-\infty}^{\infty} g'^2(t)dt - \rho \int_{-\infty}^{\infty} g^2(t)dt \right\}$ could be taken as an

improper Gaussian Prior. Unfortunately $U^2(t)$ is not integrable. To avoid this we define a new process $\{U_1(t) : t \in R\}$ as

$$U_1(t) = \exp(- \frac{\varepsilon}{2} t^2) \quad U(t) \quad , \quad \varepsilon > 0.$$

Then using the same argument we can show that

$$(1.1.18) \quad \frac{1}{\sigma^2}\left[\ \frac{1}{2\bar{\rho}} \int_{-\infty}^{\infty} e^{\varepsilon t^2}(U_1'(t))^2 dt + \frac{\bar{\rho}}{4} \int_{-\infty}^{\infty} e^{\varepsilon t^2} \quad U_1^2(t)dt\right]$$

approximates a quadratic form of the process $\{U_1(t) : t \in R\}$. Also $EU_1^2(t) < \infty$ implies $U_1^2(t)$ is integrable a.s. Hence the sample paths of $U_1^2(t)$ can be normalized to form a density function.

Now since $\bar{\rho}$ and σ^2 are at our choice we can assume that $U_1^2(t)$ is normalized. Then $\left[-4\alpha \int_{-\infty}^{\infty} g'^2(x)dx - \rho \int_{-\infty}^{\infty} g^2(x)dx\right]$ formally is similar to (1.1.18) except the factor $e^{\varepsilon t^2}$. This formal similarity suggests that

$$\exp\left\{-4\alpha \int_{-\infty}^{\infty} g'^2(x)dx - \rho \int_{-\infty}^{\infty} g^2(x)dx\right\}$$

can be looked as an improper Gaussian prior. One of the reasons for it being improper is that it does not involve the factor $e^{\varepsilon t^2}$, or some other normalizing factor. In fact, the problem is translation invariant. Thus the MPLE can be looked at as the posterior mode with respect to this improper prior.

For the process $\{U(t) : t \in R\}$ direct computation shows that

$$E \sum_{i=1}^{N} (U(t_{i+1}) - U(t_i))^2 \longrightarrow 2\bar{\rho} \quad \sigma^2(b-a)$$

and $\mathrm{Var} \quad [\sum_{i=1}^{N} (U(t_{i+1}) - U(t_i))^2] \longrightarrow 0 \qquad \text{as } N \to \infty$

$$\text{(i.e.) as } ||P|| \longrightarrow 0$$

Therefore for a sufficiently small interval $[a,b]$

$$\int_a^b e^{\varepsilon t^2} \cdot (U_1'(t))^2 dt$$

$$\approx \text{Const.} \sum_{i=1}^{N} [U(t_{i+1}) - U(t_i)]^2/\Delta t_i$$

$$\approx \text{cons.} \quad N \longrightarrow \infty \text{ with } ||P|| \longrightarrow 0$$

This shows that if we consider

$$4\alpha \int_a^b g'^2(x)dx + \rho \int_a^b g^2(x)dx$$

over the interval [a,b] such that $e^{\varepsilon t^2}$ can be considered as constant over [a,b], then the above integral should behave like

$$\frac{1}{2\sigma^2\rho} \int_a^b e^{\varepsilon t^2}(U_1'(t))^2 dt + \frac{\rho}{2\sigma^2} \int_a^b e^{\varepsilon t^2} \quad U_1^2(t)dt$$

apart from a constant arising out of the factor $e^{\varepsilon t^2}$. Since for large n one would expect the estimate to approximate the true density closely, the information $\int_{-\infty}^{\infty} g_n'^2(x)dx$, where g_n^2 is MPLE, should behave like

$$\int_{-\infty}^{\infty} e^{\varepsilon t^2} (U_1'(t))^2 dt.$$

Therefore when the penalty function is looked as an improper prior one would expect $\int_{-\infty}^{\infty} g_n'^2(x)dx$ to tend to ∞ with n.

1.2. Computational Procedure and Some Properties of the Estimate

From the form of the estimate it is clear that it depends upon ρ , which has to be determined in such a way that

$$\int_{-\infty}^{\infty} g_\rho^{\,2}(x)dx = 1 .$$

Also the estimate $g_\rho(x)$, has to satisfy the following relation at the sample points $X_1 \dots X_n$.

$$g_\rho(X_i) = \sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i-X_j|}}{4\sqrt{\alpha\rho}\cdot g_\rho(X_j)} \qquad i = 1, \dots\dots n.$$

The following theorem gives some idea about the computational aspect of these quantities. In next section we will describe a Newton-type iteration procedure for computation of ρ and $g_\rho(X_1), \dots\dots g_\rho(X_n)$.

Theorem 1.2.1. (a) For fixed α, $X_1, \dots, X_n$ and $\rho > n/2$ there exists a unique set of solutions $\{g_\rho(X_i|\alpha\ \ ); i = 1, \dots n\}$ such that

$$g_\rho(X_i|\alpha\ \ ) = \sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i-X_j|}}{4\sqrt{\alpha\rho}\, g_\rho(X_j|\alpha\ \ )} \qquad i = 1, 2, \dots n.$$

(b) For each fixed α and $g_\rho(X_1), \dots\dots g_\rho(X_n)$, there exists a unique $\rho(\alpha, X_1 \dots X_n)$ such that

$$\int_{-\infty}^{\infty} \left[\sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\, g_\rho(X_j)} \right]^2 dx = 1.$$

(c) Any ρ for which (a) and (b) both hold, has the following property

$$\text{i)}\quad n = \rho + 4\alpha \int_{-\infty}^{\infty} g_\rho^{-2}(x)dx \qquad \text{a.s.}$$

$$\text{ii)}\quad \frac{n}{2} < \rho < n. \qquad \text{a.s.}$$

Proof: Let us use the following notation for convenience.

$$\theta_i = g_\rho(X_i|\alpha\ ) \qquad i = 1 \ldots. n$$

and
$$k_{ij} = \frac{1}{4\sqrt{\alpha\rho}} \exp[-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|] \qquad i,j = 1, 2, \ldots n.$$

The equations in (a) can be written as

$$\theta_i = \sum_{j=1}^{n} \frac{K_{ij}}{\theta_j} \qquad i = 1, 2, \ldots n.$$

$$= H_i(\theta_1, \theta_2, \ldots \theta_n) \qquad \text{(say)}.$$

In terms of this notation we want to show that given α, ρ, and $X_1 \ldots X_n$, hence k_{ij}, there exists a unique solution $\bar{\theta}_1, \bar{\theta}_2, .. \bar{\theta}_n$ such that

$$H_i(\bar{\theta}_1, \ldots \bar{\theta}_n) = \bar{\theta}_i \qquad \text{for } i = 1, \ldots n.$$

Let
$$H_{ij} = \frac{\partial H_i(\theta_1, .. \theta_n)}{\partial \theta_j} \qquad i,j = 1, \ldots. n.$$

Now consider the following iteration procedure

$$\underset{\sim}{\theta}^{(\nu+1)} = \Delta^{(\nu)} \underset{\sim}{H}(\underset{\sim}{\theta}^{(\nu)}) + (I-\Delta^{(\nu)})\underset{\sim}{\theta}^{(\nu)}$$

where $\Delta^{(\nu)}$ is a diagonal matrix with positive diagonal elements and $\underset{\sim}{H} = (H_1, H_2, \ldots H_n)'$

Let
$$\underset{\sim}{e}^{(\nu+1)} = \underset{\sim}{\theta}^{(\nu+1)} - \underset{\sim}{\overline{\theta}}$$

$$\text{(i.e.)} \quad e_i^{(\nu+1)} = \delta_i^{(\nu)} H_i(\underset{\sim}{\theta}^{(\nu)}) + (1-\delta_i^{(\nu)})\theta_i^{(\nu)} - \overline{\theta}_i$$

$$= \delta_i^{(\nu)}[H_i(\underset{\sim}{\overline{\theta}}) + H_i(\underset{\sim}{\theta}^{(\nu)}) - H_i(\underset{\sim}{\overline{\theta}})] + (1-\delta_i^{(\nu)})\theta_i^{(\nu)} - \overline{\theta}_i - \delta_i^{(\nu)}\overline{\theta}_i + \delta_i^{(\nu)}\overline{\theta}_i$$

$$= \delta_i^{(\nu)} H_i(\underset{\sim}{\overline{\theta}}) - \delta_i^{(\nu)}\overline{\theta}_i + \delta_i^{(\nu)}[H_i(\underset{\sim}{\theta}^{(\nu)}) - H_i(\underset{\sim}{\overline{\theta}})] + (1-\delta_i^{(\nu)})(\theta_i^{(\nu)} - \overline{\theta}_i)$$

$$= \delta_i^{(\nu)} \sum_{j=1}^{n} (\overline{\theta}_i - \theta_i^{(\nu)}) \frac{k_{ij}}{\theta_j^{(\nu)} \cdot \overline{\theta}_j} + (1-\delta_i^{(\nu)}) e_i^{(\nu)}$$

$$= (1-\delta_i^{(\nu)}) e_i^{(\nu)} - \delta_i^{(\nu)} \sum_{j=1}^{n} e_j^{(\nu)} \cdot \frac{k_{ij}}{\theta_j^{(\nu)} \cdot \overline{\theta}_j}$$

Thus
$$e_i^{(\nu+1)} = (I - \Delta^{(\nu)} + \Delta^{(\nu)}\overline{H}^{(\nu)}) \quad e^{(\nu)}$$

where
$$\overline{H}^{(\nu)} = K\, D_1^{(\nu)} D_2 \quad , \quad K = ((k_{ij}))$$

$$D_1^{(\nu)} = \text{Diag}\left(\frac{1}{\theta_1^{(\nu)}} , \ldots . \frac{1}{\theta_n^{(\nu)}} \right)$$

$$D_2 = \text{Diag}\left(\frac{1}{\overline{\theta}_1} , \frac{1}{\overline{\theta}_2} , \ldots \frac{1}{\overline{\theta}_n} \right).$$

This implies
$$\underset{\sim}{e}^{(\nu+1)} = \prod_{m=0}^{\nu} (I - \Delta^{(m)} + \Delta^{(m)}\overline{H}^{(m)}) \cdot \underset{\sim}{e}^{(0)}$$

Hence the iteration will converge if

$$||I - \Delta^{(\nu)} + \Delta^{(\nu)} \overline{H}^{(\nu)}|| \leq q < 1 \qquad \text{for all } \nu$$

Now

$$I - \Delta^{(\nu)} + \Delta^{(\nu)} \overline{H}^{(\nu)}$$

$$= I - \Delta^{(\nu)}(I - \overline{H}^{(\nu)})$$

$$= I - \Delta^{(\nu)}(I + K D_1^{(\nu)} D_2)$$

Therefore if we choose $\delta_i^{(\nu)} \leq \frac{1}{2} \left(1 + \sum_j \frac{k_{ij}}{\theta_j^{(\nu)} \overline{\theta}_j}\right)^{-1}$ for i = 1, ... n, then the maximum eigen value of $\Delta^{(\nu)}(I + K D_1^{(\nu)} D_2)$ can be made less than one for all ν. Hence the iteration will converge. Since for each ρ, $g_\rho(x|\alpha)$ is unique, the above iteration will converge to the true solution.

Remark: 1.2.1. In all the numerical examples $\delta_i = \frac{1}{2}$ i = 1, 2, ... n always lead to convergence of the iteration procedure.

To show part (b) we observe that

$$1 = \int_{-\infty}^{\infty} g_\rho^{\,2}(x|\alpha)dx$$

(1.2.1) $$= \frac{1}{2\rho} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(1 + \frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|\right) \frac{\exp[-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|]}{4\sqrt{\alpha\rho} \quad \theta_i \theta_j}$$

This implies

$$8\alpha\ \rho\sqrt{\rho} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(1 + \frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|\right) \frac{\exp\ [-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|]}{\theta_i \theta_j}.$$

Clearly the left hand side is increasing in $\sqrt{\rho}$ and it is easy to see that right hand side is decreasing in $\sqrt{\rho}$. Hence there is unique ρ. To show part (c) we observe that

$$4\alpha \int_{-\infty}^{\infty} g_\rho^{-2}(x|\alpha)dx$$

(1.2.2) $$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(1 - \frac{1}{2\rho}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|\right) \frac{\exp[-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|]}{4\sqrt{\alpha\rho} \quad \theta_i \theta_j}$$

Therefore $$\rho \int_{-\infty}^{\infty} g_\rho^{\,2}(x|\alpha)dx + 4\alpha \int_{-\infty}^{\infty} g_\rho^{-2}(x|\alpha)dx$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\exp(-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i - X_j|)}{4\sqrt{\alpha\rho} \quad \theta_i \theta_j}$$

$$= n$$

Also (1.2.1) implies $\rho > n/2$ a.s.

Thus $\frac{n}{2} < \rho < n$ a.s.

This completes the proof of the theorem.

Remark: 1.2.2. The heuristic argument in Remark 1.1.2 and the relation (i) in Theorem 1.2.1 (c) suggest that one should choose α depending on sample size n such that $\alpha(n)$ should tend to infinity with n.

1.3. Newton-Type Computational Algorithm

In this section we present an algorithm for solving ρ, θ_1^{-1}, ... θ_n^{-1} simultaneously for given α and $X_1 \ldots X_n$. Since we are going to treat $\frac{1}{\theta_i}$ as the parameter we define $\beta_i = \frac{1}{\theta_i}$, $i = 1 \ldots n$.
Let us define the functions

$$h_i(\beta_1 \ldots \beta_n, \rho) = \frac{1}{\beta_i} - \sum_{j=1}^{n} \beta_j \, k_{ij} \qquad i = 1 \ldots n$$

and

$$h_{n+1}(\beta_1 \ldots \beta_n, \rho) = -\frac{1}{2} + \frac{1}{4\rho} \sum_i \sum_j \left(1 + \tfrac{1}{2}\tfrac{\rho}{\alpha}|X_i - X_j|\right)\beta_i \beta_j k_{ij}$$

Notice that $h_{n+1}(\beta_1 \ldots \beta_n, \rho) = 0$ is equivalent to the condition that $\int_{-\infty}^{\infty} g_\rho^2(x|\alpha)dx = 1$.

Therefore in terms of this notation we can restate our problem as follows:

For fixed α and $X_1 \ldots X_n$, solve

$$\underset{\sim}{h} = \underset{\sim}{0}. \qquad \text{for } \rho \text{ and } \beta_1 \ldots \beta_n$$

where $\underset{\sim}{h}' = (h_1, \ldots h_{n+1})$.

Let $A_{nxn} = ((a_{ij}))$ where $a_{ij} = \dfrac{\partial h_i(\beta_1 \,..\, \beta_n, \rho)}{\partial \beta_j}$ i,j = 1 ... n.

$\underset{\sim}{b}' = (b_1 \,....\, b_n)$ where $b_i = \dfrac{\partial h_i(\beta_1 \,...\, \beta_n, \rho)}{\partial \rho}$

and $c = \dfrac{\partial h_{n+1}(\beta_1 \,....\, \beta_n, \rho)}{\partial \rho}$

It is easy to see that $\dfrac{\partial h_{n+1}(\beta_1, \,..\, ,\beta_n, \rho)}{\partial \beta_i} = b_i$

Also denote by $\underset{\sim}{\beta}$ the vector of parameters $(\beta_1 \,...\, \beta_n, \rho)'$. Let $\underset{\sim}{\beta}_o$ be any initial approximation of $\underset{\sim}{\beta}$. Then expanding $\underset{\sim}{h}$ around $\underset{\sim}{\beta}_o$ and retaining only first order terms we get $0 = \underset{\sim}{h}(\underset{\sim}{\beta}) \doteq \underset{\sim}{h}(\beta_o) + \begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{pmatrix}$ $(\underset{\sim}{\beta}-\underset{\sim}{\beta}_o) + \ldots$.

Hence an improvement over the initial approximation can be obtained as

$$(\underset{\sim}{\beta}-\underset{\sim}{\beta}_o) = - \begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{pmatrix}^{-1} \underset{\sim}{h}(\underset{\sim}{\beta}_o).$$

Take
$$\underset{\sim}{\beta}^{(1)} = \underset{\sim}{\beta}_o - \begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{pmatrix}^{-1} \underset{\sim}{h}(\underset{\sim}{\beta}_o).$$

In general
$$\underset{\sim}{\beta}^{(m+1)} = \underset{\sim}{\beta}^{(m)} + \begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{pmatrix}^{-1} \underset{\sim}{h}(\underset{\sim}{\beta}^{(m)}).$$

This procedure converges very first. For sample of size 50 to 475 it only takes 3 to 4 iterations to give results correct up to 5 decimal places.

Next we will show that the computation of

$$\begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{pmatrix}^{-1} \underset{\sim}{h}(\underset{\sim}{\beta}^{(m)})$$

can essentially be reduced to solving a linear system with tri-diagonal matrix.

Observe that

$$(1.3.1)\quad \begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1}+(e-\underset{\sim}{b}'A^{-1}\underset{\sim}{b})^{-1}A^{-1}\underset{\sim}{b}\underset{\sim}{b}'A^{-1} & -(c-b'A^{-1}b)^{-1}A^{-1}\underset{\sim}{b} \\ -(c-b'A^{-1}\underset{\sim}{b})^{-1}b'A^{-1} & (c-\underset{\sim}{b}'A^{-1}\underset{\sim}{b})^{-1} \end{pmatrix}$$

Now consider $a_{ij} = \dfrac{\partial h_i(\beta_1, \ldots, \beta_n, \rho)}{\partial \beta_j}$

$$= -\left(\frac{\delta_{ij}}{\beta_i^2} + k_{ij}\right)$$

Therefore $$A = -(K + D_\beta^{-2})$$

where $K = ((k_{ij}))$ and $D_\beta = \mathrm{Diag}(\beta_1 \ldots \beta_n)$.

Also define $\rho_i = e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_{i-1}-X_i|}$ $\quad i = 2, \ldots n$

and $\rho_1 = 1$. Then $k_{ij} = \prod_{\ell=i+1}^{j} \rho_\ell$.

Let $$\Lambda = \mathrm{Diag}(1, 1-\rho_2^2, 1-\rho_3^2, \ldots 1-\rho_n^2)$$

and $T = \begin{bmatrix} 1 & -\rho_2 & 0 & 0 & 0 \\ 0 & 1 & -\rho_3 & 0 & 0 \\ & & & & \\ 0 & & & 1 & -\rho_n \\ 0 & & & 0 & 1 \end{bmatrix}$

It is easy to show that

$$K^{-1} = 4/\alpha\rho \quad T \Lambda^{-1} T'$$

Therefore $-A^{-1} = (k + D_\beta^{-2})^{-1} = D_\beta(I + D_\beta^{-1} K^{-1} D_\beta^{-1})^{-1} D_\beta^{-1} K^{-1}$

(1.3.2) $\qquad = D_\beta(I + 4/\alpha\rho \quad D_\beta^{-1} T\Lambda^{-1} T' D_\beta^{-1})^{-1} D_\beta^{-1} (4/\alpha\rho) \quad T\Lambda^{-1} T'$

Now if we write $\underset{\sim}{h} = \begin{bmatrix} \underset{\sim}{h}^{(1)} \\ --- \\ h_{n+1} \end{bmatrix}$, then from (1.3.1) it is clear that essentially we need to compute $A^{-1}\underset{\sim}{b}$ and $A^{-1}\underset{\sim}{h}^{(1)}$. Because of the special form of K^{-1} we can compute $K^{-1}\underset{\sim}{b}$, $K^{-1}\underset{\sim}{h}^{(1)}$ easily. Since K^{-1} is tri-diagonal so is $(I + D_\beta^{-1} K^{-1} D_\beta^{-1})$.

Therefore for any vector $\underset{\sim}{g}_{n\times 1}$, $(I + D_\beta^{-1} K^{-1} D_\beta^{-1})^{-1}\underset{\sim}{g}$, can be obtained by one-step forward and backward solution of the linear system with tri-diagonal matrix.

<u>Remark: 1.3.1</u> In all the examples we tried, no numerical instability occurred.

Remark 1.3.2. Eventhough the algorithm presented in this section gives convergence to a local maximum, in all the numerical examples, we do not have a rigorous proof. Here we give a heuristic argument which shows that asymptotically the matrix

$$\begin{bmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{b}' & c \end{bmatrix}$$

is negative definite. This would imply the convergence of the iteration procedure, if the initial approximation is sufficiently close.

Since A is negative definite it is sufficient to show that $c - \underset{\sim}{b}'A^{-1}\underset{\sim}{b} < 0$.

Let $\dot{k}_{ij} = \frac{\partial k_{ij}}{\partial \rho}$, $\ddot{k}_{ij} = \frac{\partial^2 k_{ij}}{\partial \rho^2}$ $\quad i, j = 1, \ldots n$

and $\underset{\sim}{\beta}_{(1)} = (\beta_1 \ldots . \beta_n)'$.

Then $\qquad c - \underset{\sim}{b}'A^{-1}\underset{\sim}{b}$

$$= -\underset{\sim}{\beta}'_{(1)} \left(\frac{1}{2} \ddot{K} - \dot{K} (K+D_\beta^{-2})^{-1} \dot{K} \right) \underset{\sim}{\beta}_{(1)}$$

where $\dot{K} = ((\dot{k}_{ij}))$ and $\ddot{K} = ((\ddot{k}_{ij}))$

Therefore it is sufficient to show that

$$\underset{\sim}{\beta}'_{(1)} \left(\frac{1}{2} \ddot{K} - \dot{K}(K+D_\beta^{-2})^{-1}\dot{K} \right) \underset{\sim}{\beta}_{(1)} \quad \text{is positive.}$$

But this is true if $\underset{\sim}{\beta}'_{(1)}(\ddot{K} - \dot{K}K^{-1}\dot{K})\underset{\sim}{\beta}_{(1)}$ is positive.

The following heuristic argument shows that this is true asymptotically.

We have $$\dot{k}_{ij} = -k_{ij}\left(\frac{1}{2\rho} + \frac{\Delta_{ij}}{4\sqrt{\alpha\rho}}\right)$$

and $$\ddot{k}_{ij} = k_{ij}\left(\frac{3}{4\rho^2} + \frac{3\Delta_{ij}}{8\rho\sqrt{\alpha\rho}} + \frac{\Delta^2_{ij}}{16\alpha\rho}\right)$$

where $$\Delta_{ij} = |X_i - X_j|$$

Let us denote the matrix $((k_{ij}\Delta^{\ell}_{ij}))$ by $(K\Delta^{\ell})$. Let ϕ be any smooth function. We will denote the vector $(\phi(X_1), \ldots, \phi(X_n))'$ by $\underset{\sim}{\phi}$. Then the i-th element of $(K\Delta^{\ell})\underset{\sim}{\phi}$ can be expressed as

$$((K\Delta^{\ell})\underset{\sim}{\phi})_i = n \int_{-\infty}^{\infty} \frac{1}{4\sqrt{\alpha\rho}} \exp\left[\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_i|\right] |x-X_i|^{\ell}\phi(x)dF_n(x)$$

Since for large n, F_n approximates F closely (heuristically), the above expression behaves approximately as

$$\frac{n}{4\sqrt{\alpha\rho}}\phi(X_i)f(X_i)\int_{-\infty}^{\infty} |u|^{\ell} \exp\left[-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|u|\right] du$$

$$= \frac{n}{4\sqrt{\alpha\rho}} \frac{\phi(X_i)f(X_i)\ 2\ \ell!\ 2^{\ell}\ \alpha^{\frac{\ell}{2}}}{\rho^{\frac{\ell}{2}}}$$

(1.3.3) $$= \frac{n\ \ell^{1}\ 2^{\ell-1}\ \alpha^{\frac{\ell-1}{2}}}{\rho^{\frac{\ell+1}{2}}}\ \phi(X_i)f(X_i)$$

Therefore putting $\ell = 0$, 1 and 2 we get

$$K\underset{\sim}{\phi} \cong \frac{n}{2\sqrt{\alpha\rho}}(\phi\cdot f)$$

$$(K\Delta)\underset{\sim}{\phi} \cong \frac{n}{\rho}\,(\phi\cdot f) \quad \text{and}$$

$$(K\Delta^2)\underset{\sim}{\phi} \cong \frac{4n\sqrt{\alpha}}{\rho\sqrt{\rho}}\,(\phi\cdot f)$$

where $(\phi\cdot f)' = (\phi(X_1)f(X_1), \ldots\ldots, \phi(X_n)f(X_n))$

Using these approximations and the relation

$$K\underset{\sim}{\beta}_{(1)} = D_\beta^{-2}\,\underset{\sim}{\beta}_{(1)}$$

we get $\underset{\sim}{\beta}'_{(1)}(\ddot{K} - \dot{K}\,K^{-1}\,\dot{K})\underset{\sim}{\beta}_{(1)} \approx \frac{1}{2}\,\underset{\sim}{\beta}'_{(1)}\ddot{K}\underset{\sim}{\beta}_{(1)} > 0.$

1.4. A Discussion on the Extension to a More General Penalty Function.

We have seen that in the case of Good's first penalty function the estimate is a spline function with double exponential splines and knots at the every sample point. The nature of this double exponential function makes the estimate some what unsmooth. In this section we investigate the possibility of getting a smooth estimate. The answer seems to be in negative.

Let Q be a positive definite kernel function. As usual we assume $f = g^2$. Define the penalty function as

$$\Phi_Q(g) = 4\alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(t-u)\, g'(t)\, g'(u)\, dt du.$$

Then the logarithm of the penalized likelihood is given by

$$L(g) = 2 \sum_{i=1}^{n} \log g(X_i) - \Phi_Q(g)$$

we want to maximize $L_\rho(g) = L(g) - \rho \int_{-\infty}^{\infty} g^2(x)dx$

subject to: $|g|(x) \geq 0$ for all x

$|g|(X_i) > 0$ for $i = 1, 2 \ldots n$

(1.4.1) and determine ρ such that $\int_{-\infty}^{\infty} g_\rho^{\,2}(x)\, dx = 1.$

The following theorem states the solution of the above maximization problem.

Theorem: 1.4.1. *Let g_ρ be the function which maximizes $L_\rho(g)$, without nonnegativity condition, then $g_\rho(x)$ has the following form:*

$$g_\rho(x) = \sum_{i=1}^{n} \frac{r_i(x)}{g_\rho(X_i)}$$

where $$r_i(x) = r_o(x-X_i)$$

and $\hat{r}_o(s)$, the Fourier transform of r_o, is given by

$$\hat{r}_o(s) = \frac{1}{\rho - 4\alpha\, s^2\, \hat{Q}(s)} \quad .$$

and $\hat{Q}(s)$ is the fourier transform of Q .

Proof: The proof is similar to the proof of Proposition 3.5 Montricher, Tapia and Thompson (1975 b).

Remark: 1.4.1. If r_o is nonnegative then both approaches of Good give the same solution.

Remark 1.4.2. If $\hat{Q}(s)$ has the property that

$$\hat{Q}(s) \longrightarrow \infty \quad \text{as} \quad s \longrightarrow \pm\infty$$

then it might be possible to get a r_o which is smoother than double exponential. But the above condition would require g to have higher derivatives.

1.5. A Discussion on an Extension to Bivariate Case

Let (X,Y) be a random vector with two components. Let f be the density of (X,Y). We again substitute $f = g^2$. We will show that in the bivariate case Good's alternate approach, ignoring the nonnegativity condition of g, may not lead to the same solution. Consider the following penalty function in the case of bivariate density function.

$$\Phi(g) = 4\alpha \cdot \int\int \left(\left[\frac{\partial^2 g(x,y)}{\partial x^2} \right]^2 + \left[\frac{\partial^2 g(x,y)}{\partial y^2} \right]^2 + 2 \left[\frac{\partial^2 g(x,y)}{\partial x \cdot \partial y} \right]^2 \right) dxdy.$$

This penalty function is similar to the one suggested by Good and Gaskin [1972] in the multivariate case. Therefore if (X_1,Y_1), ... (X_n,Y_n) be a sample of size n with common density $f = g^2$, then the logarithm of penalized likelihood function can be expressed as

$$L(g) = 2 \sum_{i=1}^{n} \log g(X_i,Y_i) - \Phi(g).$$

We want to find g such that L(g) is maximized subject to the restriction

$$\int_{-\infty}^{\infty} g^2(x,y)dxdy = 1$$

$$|g|(x,y) \geq 0 \qquad \text{for all } (x,y) \in R^2$$

$$|g|(X_i,Y_i) > 0 \quad , \quad i = 1, \ldots n.$$

Let $$\Phi_\rho(g) = \Phi(g) + \rho \int g^2(x,y)dxdy$$

and $$L_\rho(g) = L(g) - \rho \int g^2(x,y)dxdy.$$

Therefore the above maximization problem is equivalent to the following;

Maximize $\quad L_\rho(g)$

subject to $\quad |g|(x,y) \geq 0$

$\quad |g|(X_i,Y_i) > 0 \qquad i = 1 \ldots n$

and choose ρ such that $\int_{-\infty}^{\infty} g_\rho^2(x,y)dxdy = 1$

Proceeding exactly as before we get the following result:

Theorem 1.5.1. Let g_ρ be the unique function which maximizes $L_\rho(g)$ subject to

$$\int_{-\infty}^{\infty} g_\rho^2(x,y)dxdy = 1$$

Then $g_\rho(x)$ can be expressed as follows

$$g_\rho(x,y) = \sum_{i=1}^{n} \frac{r_0(x,y)}{g_\rho(X_i,Y_i)}$$

where $\hat{r}_0(t,s)$, the fourier transform of r_0, is given by

$$\hat{r}_0(t,s) = \frac{1}{\rho+4\alpha\ (t^2+s^2)^2}$$

Proof of this Theorem is similar to the proof of Proposition 3.5 of Montricher, Tapia and Thompson (1975 b).

Theorem 1.5.2. $r_0(x,y)$ cannot be nonnegative always.

Proof: Suppose $r_0(x,y) \geq 0$ for all $(x,y) \in R^2$. Then $\int_{-\infty}^{\infty} x^2\ r_0(x,y)dxdy > 0$. But

$$\int_{-\infty}^{\infty} x^2\ r_0(x,y)dxdy = \left.\frac{\partial^2\hat{r}_0(t,s)}{\partial t^2}\right|_{t = s = 0.}$$

$$= 0.$$

This completes the proof.

Next we construct an example to show that in higher dimension, penalty function involving only partial derivatives of first order only will not be sufficient for removing the rough densities. Consider only the bivariate case. We will construct a function g(x,y) for which

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g^2(x,y)dxdy < \infty$$

and hence can be normalized to make a density,

Also $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\left(\frac{\partial g(x,y)}{\partial x}\right)^2 + \left(\frac{\partial g(x,y)}{\partial y}\right)^2\right] dxdy < \infty$$

but the likelihood will be unbounded.

Let g(x,y) be spherically symmetric around the origin. Let

$$x = r\cos\theta$$

$$y = r\sin\theta$$

Then $$\int_{-\infty}^{\infty}\int_{\infty}^{\infty} g^2(x,y)dxdy$$

$$= 2\pi\int_{0}^{\infty} r\, h^2(r)\, dr.$$

Since by assumption g(x,y) is spherically symmetric g(r,θ) is a function of r only. We denote this function by h(r).

Similarly $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\left(\frac{\partial g(x,y)}{\partial x}\right)^2 + \left(\frac{\partial g(x,y)}{\partial y}\right)^2\right] dxdy$$

$$= 2\pi\int_{0}^{\infty} r\, h'^2(r)\, dr.$$

Now define

$$h_k(r) = \begin{cases} A\,(\log(\frac{1}{r}))^{\alpha} & \text{if } r < e^{-m} \\ \dfrac{A\, m^{\alpha} e^{-mk}}{r^k} & \text{if } r > e^{-m} \end{cases}$$

where $0 < \alpha < 1/2$, m, α, k, A are constants.

Then

$$h_k'(r) = \begin{cases} -A\alpha\,(\log(\frac{1}{r}))^{\alpha-1}\,\frac{1}{r} & \text{if } r < e^{-m} \\ \dfrac{-k\, A\, m^{\alpha}\, e^{-2mk}}{r^{k+1}} & \text{if } r > e^{-m} \end{cases}$$

Also

$$\int_0^{e^{-m}} r\, h_k^2(r)dr = A^2 \int_0^m e^{-2y}\, y^{2\alpha} dy \quad ,$$

$$\int_{e^{-m}}^{\infty} r\, h_k^2(r)dr = A^2\, m^{2\alpha}/2(k-1) \quad ,$$

$$\int_0^{e^{-m}} r\, h_k'^2(r)dr = -\alpha^2\, A^2\, m^{2\alpha-1}/(2\alpha-1)$$

and

$$\int_{e^{-m}}^{\infty} r\, h_k'^2(r)dr = kA^2\, m^{2\alpha}/2$$

Therefore

$$\int_0^{\infty} r\, h_k{}^2(r)dr = \frac{A^2\, m^{2\alpha}}{2(k-1)} + A^2 \int_0^m e^{-2y} y^{2\alpha}\, dy$$

and $$\int_0^\infty r\, h_k^{-2}(r)dr = \frac{A^2}{2} \cdot m^{2\alpha} \cdot \left[k + \frac{\alpha^2}{m(1-2\alpha)}\right]$$

Therefore if we take

$$g_k(\underset{\sim}{Z}) = c \cdot \sum_{i=1}^{n} h_k(||\underset{\sim}{Z}-(X_i,Y_i)||)$$

This function g_k will have all the properties we need.

Remark 1.5.1. (a) The nonnegativity condition cannot be ignored in the case of multivariate penalty function suggested by Good and Gaskin (1972). Their alternate approach may not lead to the same solution.

(b) Even if $r_o(x,y)$ is positive for all $(x,y) \varepsilon R^2$ the exact form of r_o will not be any easier to determine. The penalty function approach seems necessarily complicated.

1.6. A Discussion on Consistency

In the density estimation problem usually the consistency, asymptotic bias and asymptotic distributions are studied. In their paper Good and Gaskin (1972) gave a "proof" of the consistency. Their proof cannot be considered as a rigorous proof. In fact they stated "we shall give a'physicist's proof' of the following theorem. A rigorous proof might require a further constraint on the allowable density functions." They tried to show that

$$\int_a^b f_n(x)dx \xrightarrow{P} \int_a^b f_o(x)dx \qquad \text{for any } a < b$$

f_n is the MPLE and f_0 is the true density. Their arguments are based on intuitive grounds. Our attempt in this direction to prove consistency has not been very successful. The global nature of the estimate makes it very complicated for mathematical calculations. Even though we have not been able to prove the consistency in a rigorous manner we present some results which indicate what ought to hold.

Let $X_1,\ldots,X_n$ be iid random variables with common density f. Suppose $f(x) = g^2(x)$ and g(x) satisfies the following properties.

(a) $g(x) \geq 0$ for all $x \in R$

(b) $|g(x) - g(y)| \leq M \ |x-y|^\delta$ for some $\delta > 0$ and $M > 0$.

Let

$$S_1(x) = \{j : |\frac{x-X_j}{h}| \leq k \quad \log n\} ,$$

$$S_2(x) = \{j : |\frac{x-X_j}{h}| > K \quad \log n\} ,$$

$$f_n(x) = \sum_{j=1}^{n} \frac{e^{-|\frac{x-X_j}{h}|}}{2nh} , \quad h(n) = 0(n^{-\frac{1}{2}}) ,$$

and

$$g_\rho(x|\alpha) = \sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\ g_\rho(X_j|\alpha)} ,$$

This shows

$$g_\rho^{\,2}(x|\alpha) > \frac{1}{4\sqrt{\alpha\rho}}$$

Now
$$Q_n(x) = \sum_{j\varepsilon S_2(x)} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)}$$

$$\le \frac{n}{4\sqrt{\alpha\rho}} \quad n^{-k} \quad 2(\alpha\rho)^{\frac{1}{4}}$$

$$\approx O(n^{1-\frac{1}{4}-k}) \quad \text{since} \quad \rho = O(n^{\frac{1}{2}})$$

$$\longrightarrow 0 \quad \text{if } k > 3/4$$

Also
$$P_n(x) = \sum_{j\varepsilon S_2(x)} \left[\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|\right] \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)}$$

$$= 2 \sum_{j\varepsilon S_2(x)} \left[\frac{1}{4}\sqrt{\frac{\rho}{\alpha}}|x-X_j|\right] \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)}$$

$$\le \frac{2\; n\; n^{-\frac{k}{2}}}{4\sqrt{\alpha\rho}} \quad 2\; (\alpha\rho)^{\frac{1}{4}}$$

$$= O(n^{1-\frac{1}{4}-\frac{k}{2}})$$

Before proceeding further we state a theorem, due Rubin (1977), which will be useful in the following discussion.

Theorem 1.6.1. Let X_1, X_2, ... X_n be iid random variables with bounded density f. Then for bounded kernels K of bounded variation, $(\hat{f}_n(x) - E\hat{f}_n(x))^2$, where

$\hat{f}_n(x) = \int_{-\infty}^{\infty} K(\frac{z}{h(n)})\, dF_n(x-z)/h(n)$, converges to zero uniformly a.s. if $h(n) \longrightarrow 0$ and $nh(n) \longrightarrow \infty$

In particular the above result is true if

$$K_1(Z) = \frac{1}{2}\, e^{-|Z|}$$

and

$$K_2(Z) = \frac{1}{2}\, |Z|\, e^{-|Z|}$$

This shows that if we replace $g_\rho(x|\alpha)$ by a smooth uniformly strongly consistent estimator $\tilde{g}_n(x)$

then

$$\sum_{j\varepsilon S_1(x)} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)}$$

and

$$\sum_{j\varepsilon S_1(x)} \left[\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|\right] \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j|\alpha)}$$

should approximately behave like

(1.6.1)

$$\sum_{j\varepsilon S_1(x)} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; \tilde{g}_n(X_j)}$$

$$\approx \frac{1}{\tilde{g}_n(x)} \sum_{j\varepsilon S_1(x)} \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}}$$

and

(1.6.2) $$\sum_{j\varepsilon S_1(x)} \left[\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|\right] \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\,\tilde{g}_n(X_j)}$$

$$\approx \frac{1}{\tilde{g}_n(x)} \sum_{j\varepsilon S_1(n)} \left[\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|\right] \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}}$$

respectively.

Then the theorem 1.6.1 would imply that quantities in (1.6.1) and (1.6.2) should converge to g(x). This would imply that the equations for solving ρ and $g_\rho(X_1|\alpha) \ldots g_\rho(X_n|\alpha)$ should hold simultaneously.

These and other numerical results suggest that the following conjecture should hold.

<u>Conjecture</u>:

(a) $$\sup_x \left| g_\rho(x) - \sum \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|x-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_j)} \right| \longrightarrow 0 \quad \text{a.s.}$$

(b) $$\frac{1}{2\rho} \sum_i \sum_j \left(1 + \frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i-X_j|\right) \frac{e^{-\frac{1}{2}\sqrt{\frac{\rho}{\alpha}}|X_i-X_j|}}{4\sqrt{\alpha\rho}\; g_\rho(X_i)g_\rho(X_j)}$$

$$\longrightarrow 1. \quad \text{a.s.}$$

<u>Remark 1.6.1</u>. Since the numerical evidences and other arguments presented in this chapter suggest that $g_\rho^2(x|\alpha)$ behaves essentially like a kernel estimate with double exponential kernel

we should choose $h(\alpha,\rho) = \frac{2\sqrt{\alpha}}{\sqrt{\rho}}$ as the optimal choice of $h(n)$ in kernel estimate. This would imply that α should be like a fractional power of n. This agrees with the remark 1.2.2.

We used Newton-type iteration procedure to compute the estimate. We mentioned that it gives results correct up to 5 decimal places in 3 to 4 iterations. Observations were taken from standard normal distribution. In Table 1.1 we present values of ρ and ρ/n, where ρ is chosen such that $\int_{-\infty}^{\infty} g_\rho^{2}(x)dx = 1$. For each sample size n and α we took 10 different samples and the value of ρ in the table is actually the average of these 10 estimates of ρ. Table 1.1 shows that the ratio $E(\rho/n)$ should converge to 1 as $n \to \infty$. Expectation is taken with respect to $X_1, \ldots, X_n$.

In Table 1.2 we present the information computed from the estimated density. This numerical evidence supports our heuristic argument of remark 1.1.2, that the sample information tends to infinity with n for fixed α. We also present a few graphs which

show that MPLE essentially behaves like kernel estimate with double-exponential kernel. Last two graphs correspond to estimate derived by Montricher, Tapia and Thompson using Good and Gaskin's second penalty function:

$$\Phi(g) = 4\alpha \int_{-\infty}^{\infty} g'^{2}(x) + \beta \int_{-\infty}^{\infty} g''^{2}(x)dx$$

where $f = g^2$.

The smoothness of these estimates is due to the assumption of existence of second derivative. It was pointed out by Montricher, Tapia and

Thompson (1975 b) that maximizing g may not be of same sign always and hence Good's method of substitution $f = g^2$ may not always lead to the correct solution. For samples from standard normal population in all the examples we always got positive values of g. This suggest that maximizing f and maximizing g, may not be very much different, at least in large samples.

Table 1.1. Values of ρ and ρ/n for each α and n^*.

α	n = 25	75	125	175	275	375	475
1	21.827 .87038	70.226 .93635	119.030 .95224	167.404 .95659	266.169 .96789	365.294 .97412	463.674 .97616
2	21.237 .84948	70.102 .93469	119.200 .95360	167.841 .95909	266.972 .97081	366.353 .97694	465.178 .97932
3	20.669 .82676	69.558 .92744	118.757 .95006	167.518 .95725	266.783 .97012	366.257 .97669	465.270 .97952
4	20.175 .80700	68.914 .91885	118.135 .94508	166.958 .95405	266.283 .96830	365.793 .97545	464.906 .97875
5	19.749 .78996	68.248 .90997	117.446 .93957	166.294 .95025	265.643 .96597	365.161 .97376	464.332 .97754
6	19.381 .77524	67.589 .90117	116.732 .93386	165.580 .94617	264.929 .96338	364.436 .97183	463.644 .97609
7	19.059 .76236	66.947 .89265	116.012 .92810	164.842 .94195	264.174 .96063	363.656 .96975	462.886 .97450
8	18.775 .75100	66.333 .88444	115.296 .92237	164.094 .93764	263.395 .95780	362.842 .96785	462.085 .97281

* Each entry is an average of 10 estimates obtained from 10 independent samples.

Table 1.2. Values of $4 \int_{-\infty}^{\infty} g_\rho^{-2}(x)dx$ *

α	n = 25	75	125	175	275	375	475
1	3.1736	4.7740	5.9703	7.5959	8.8306	9.7062	11.3265
2	1.88165	2.4492	2.9002	3.5798	4.0142	4.3234	4.9109
3	1.4437	1.8139	2.0811	2.4939	2.7388	2.9144	3.2434
4	1.2062	1.5215	1.7162	2.0106	2.1792	2.3016	2.5235
5	1.0501	1.3504	1.5108	1.7413	1.8714	1.9678	2.1336
6	.9365	1.2355	1.3780	1.5700	1.6785	1.7606	1.8927
7	.8488	1.1501	1.2840	1.4511	1.5466	1.6205	1.7305
8	.7782	1.0834	1.2130	1.3632	1.4506	1.5197	1.6143

* Each entry is an average of 10 estimates obtained from 10 independent samples.

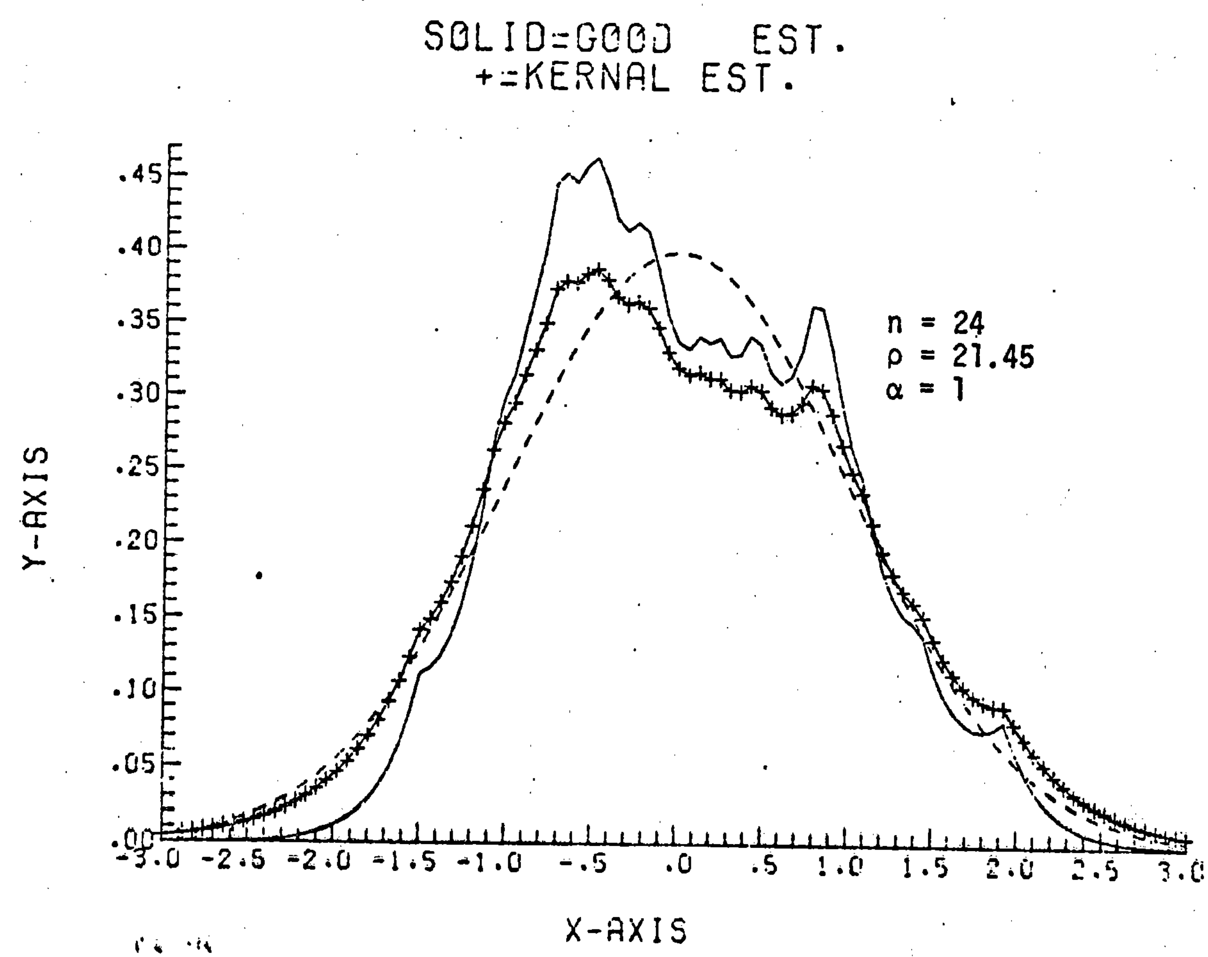
SOLID=GOOD EST.
+=KERNAL EST.
n = 24
ρ = 21.45
α = 1
Y-AXIS
.00
.05
.10
.15
.20
.25
.30
.35
.40
.45
X-AXIS
-3.0
-2.5
-2.0
-1.5
-1.0
-.5
.0
.5
1.0
1.5
2.0
2.5
3.0

Figure 1.1(a)

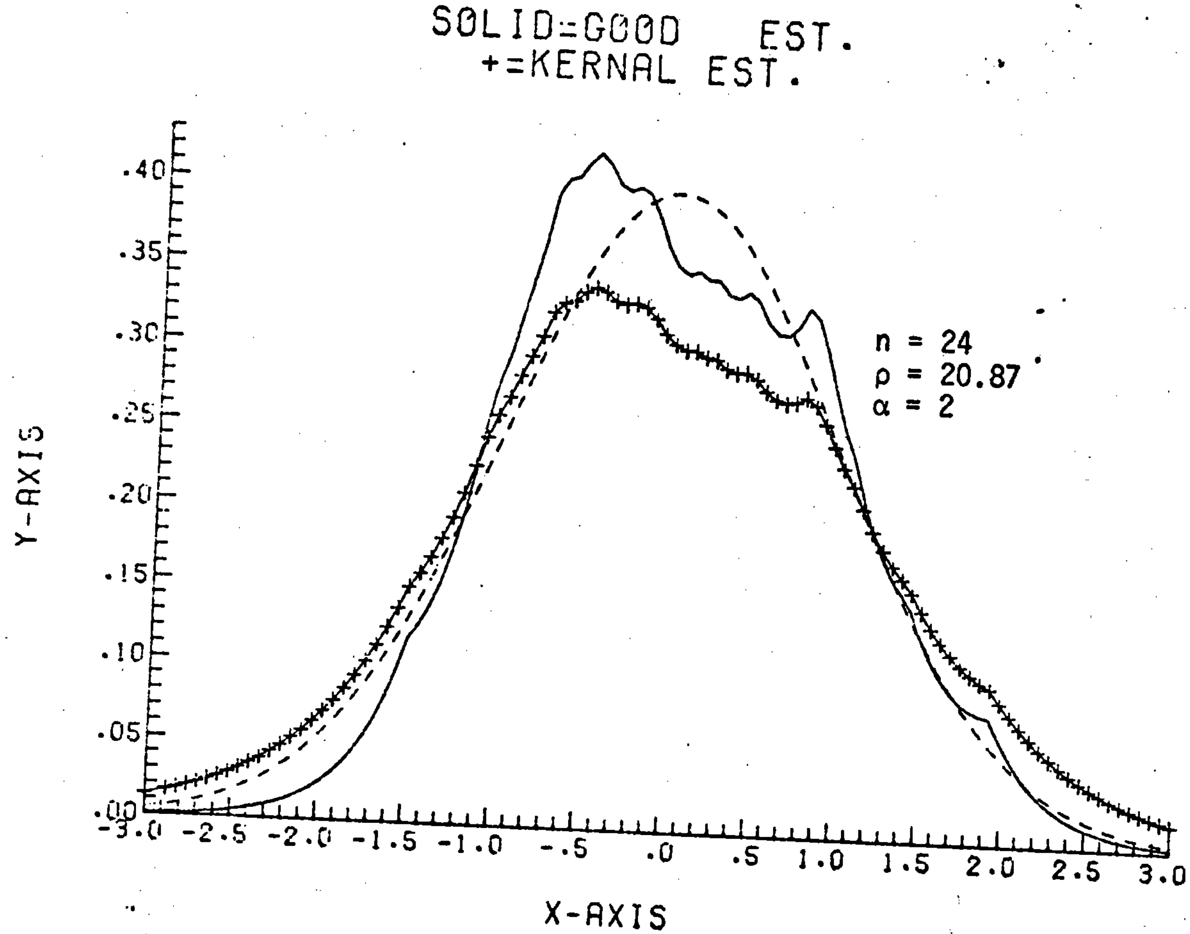
SOLID=GOOD EST.
+=KERNAL EST.
n = 24
ρ = 20.87
α = 2
Y-AXIS
.00
.05
.10
.15
.20
.25
.30
.35
.40
X-AXIS
-3.0
-2.5
-2.0
-1.5
-1.0
-.5
.0
.5
1.0
1.5
2.0
2.5
3.0

Figure 1.1(b)

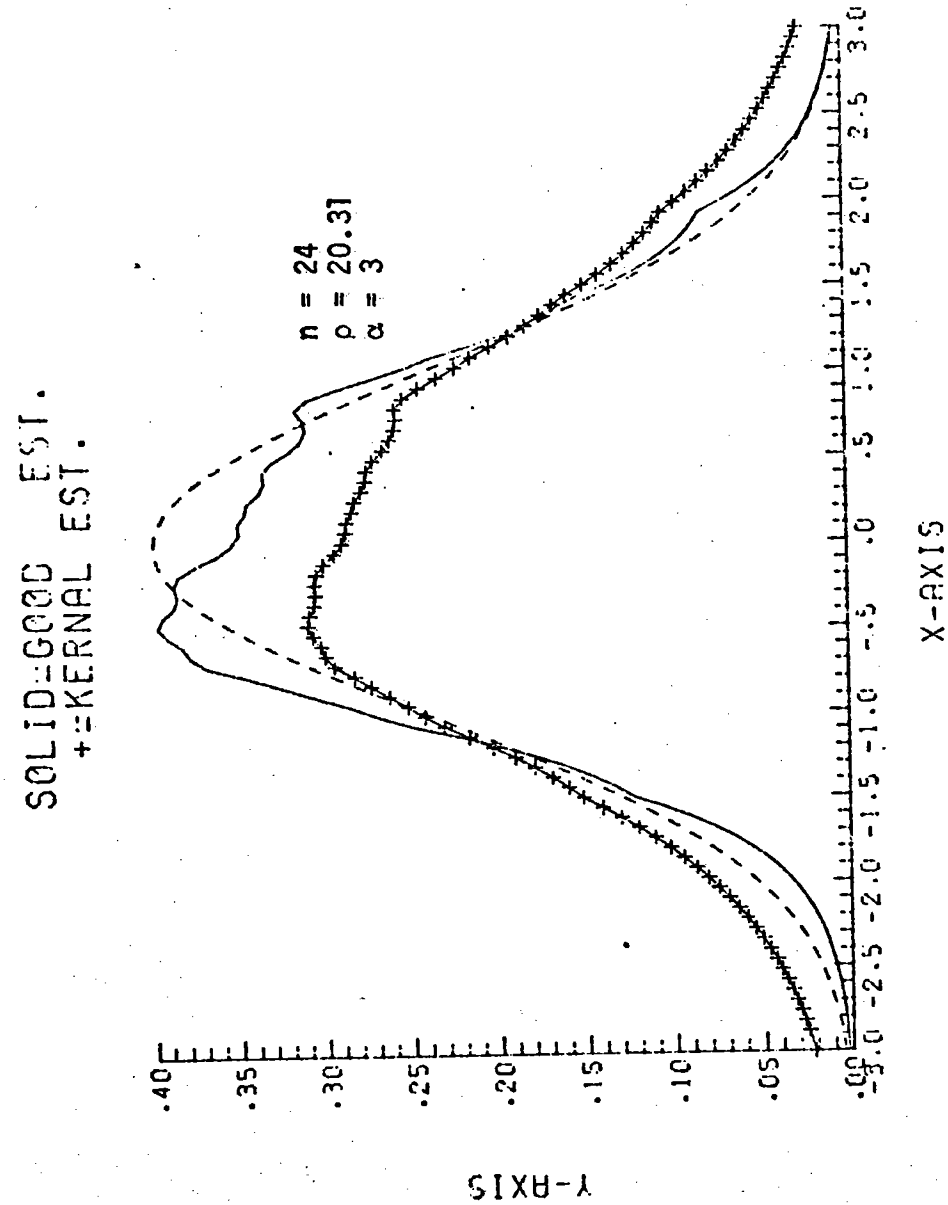

Figure 1.1(c)

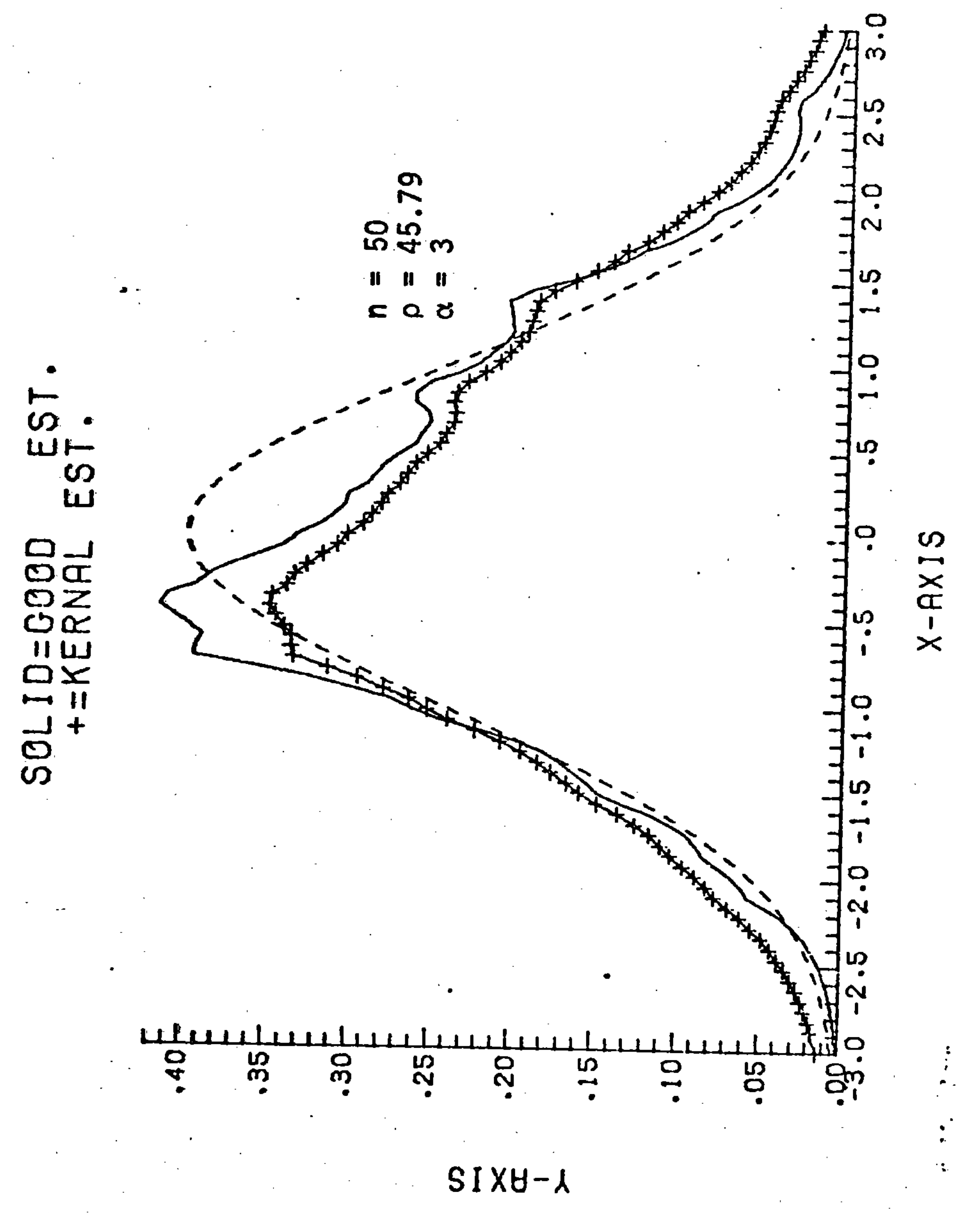
SØLID=GØØD EST.
+=KERNAL EST.
n = 50
ρ = 45.79
α = 3
X-AXIS
Y-AXIS
-3.0 -2.5 -2.0 -1.5 -1.0 -.5 .0 .5 1.0 1.5 2.0 2.5 3.0
.00 .05 .10 .15 .20 .25 .30 .35 .40

Figure 1.1(d)

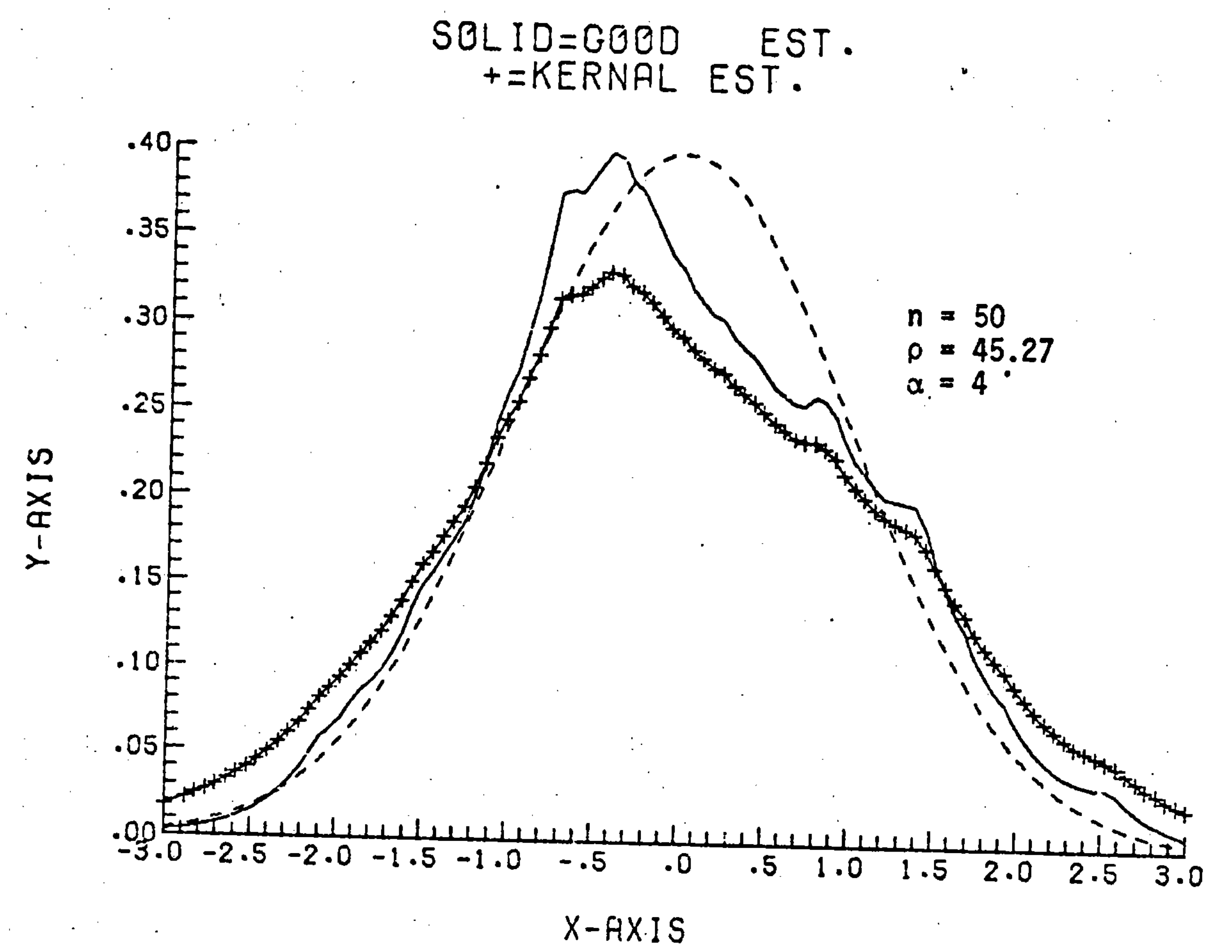
SØLID=GØØD EST.
+=KERNAL EST.
n = 50
ρ = 45.27
α = 4
.40
.35
.30
.25
.20
.15
.10
.05
.00
-3.0 -2.5 -2.0 -1.5 -1.0 -.5 .0 .5 1.0 1.5 2.0 2.5 3.0
X-AXIS
Y-AXIS

Figure 1.1(d)

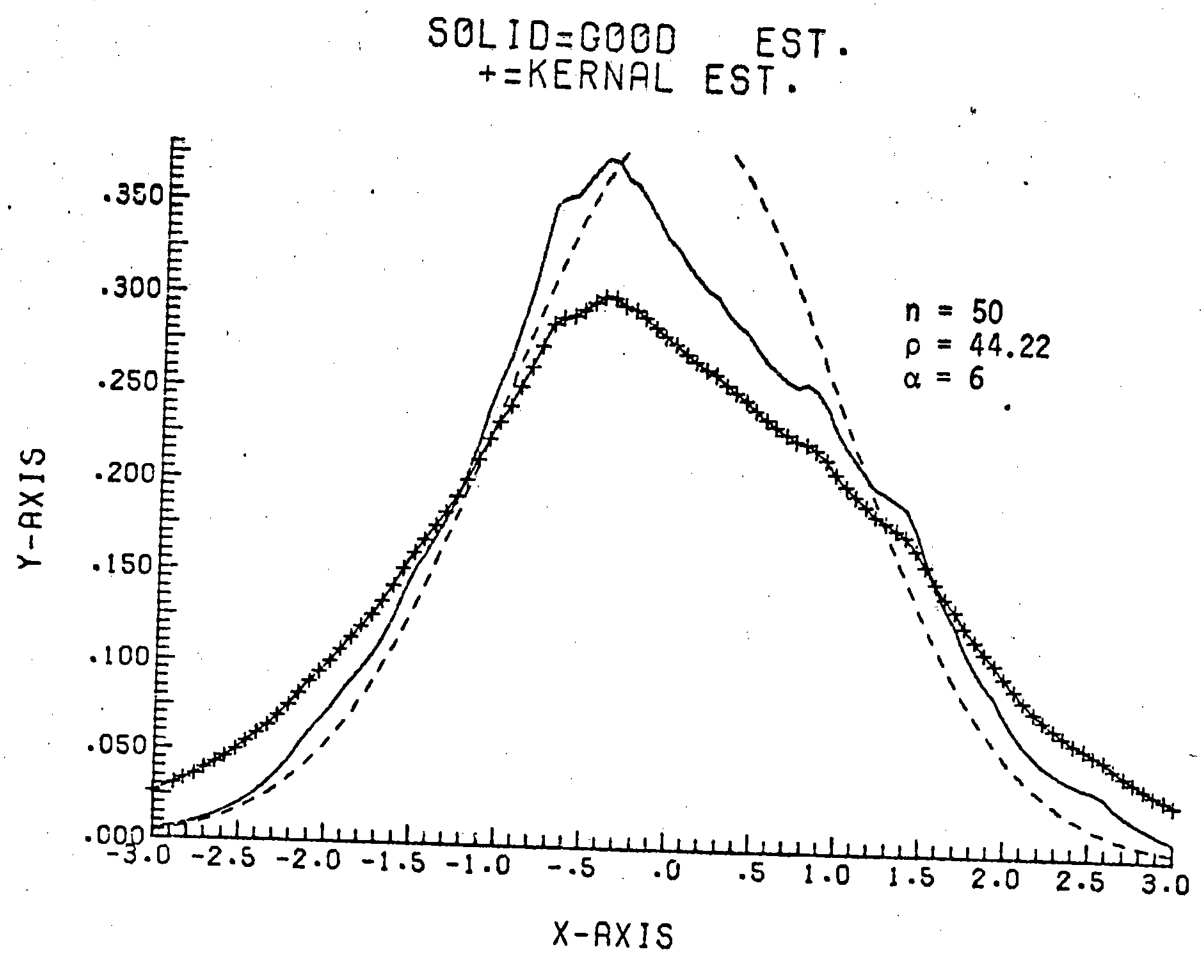
SØLID=GOOD EST.
+=KERNAL EST.
n = 50
ρ = 44.22
α = 6
Y-AXIS
.000
.050
.100
.150
.200
.250
.300
.350
-3.0 -2.5 -2.0 -1.5 -1.0 -.5 .0 .5 1.0 1.5 2.0 2.5 3.0
X-AXIS

Figure 1.1(e)

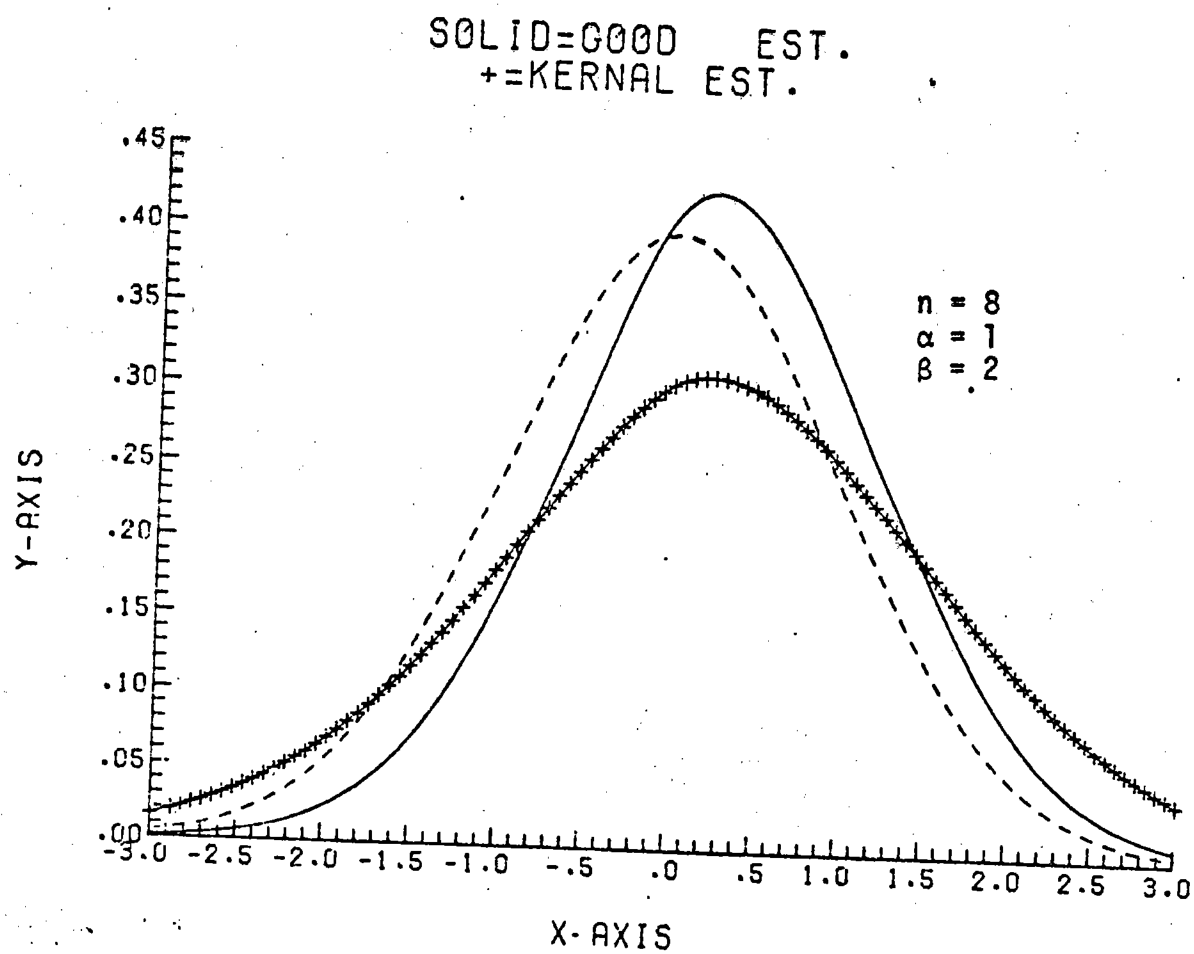
SOLID=GOOD EST.
+=KERNAL EST.
n = 8
α = 1
β = 2
.45
.40
.35
.30
.25
.20
.15
.10
.05
.00
Y-AXIS
-3.0 -2.5 -2.0 -1.5 -1.0 -.5 .0 .5 1.0 1.5 2.0 2.5 3.0
X-AXIS

Figure 1.2(a)

Program for Computing the MPLE

```
      SUBROUTINE DENSGD(N,ALFA,X,INDPR,EPS1,EPS2,SX,ESTSX,
     1 NITID,G,ALDA,SINFOR)
C     ESTIMATIOM BY GOOD≠S FIRST PENALTY FUNCTION.
C     *** N= SAMPLE SIZE
C     *** ALFA= COEFF. OF PENALTY FUNCTION
C     *** X= ORDERED SAMPLE VECTOR
C     *** INDPR=  1 IF SAMPLE INFORMATION(SINFOR) IS TO BE PRINTED.
C               =  0  OTHERWISE
C     *** EPS1= ERRCR BOUND FOR DENSITY ESTIMATES
C     *** EPS2= ERROR BOUND FOR RHO ESTIMATF
C     *** G= OUTPUT VECTOR OF SQUARE ROOT OF ESTIMATES
C     *** ALDA= ESTIMATE OF LAGRANGE MULTIPLIER (RHC)
C     *** SINFOR= INFORMATION COMPUTED FROM THE ESTIMATED DENSITY
C     *** SX= VALUE AT  WHICH DENSITY IS TO BE COMPUTED
C     *** ESTSX= VALUE OF THE ESTIMATED DENSITY
      DIMENSION X(N),G(N),AF(N),CRN(N),GG(N),G1(N),Y(N),
     1 XDIF(N),EDIF(N),HL(N),RL(N),HLS(N),RLS(N),RHC(N),
     2 R(N),S(N),T(N)
      AN=N
      ALDA=AN-5.
      N1=N-1
      DO 230 I=1,N1
  230 XDIF(I)=X(I+1)-X(I)
      IND=-1
      NOITN=0
  198 P=1./(4.*SQRT(ALDA*ALFA))
      Q=-SQRT(ALDA/ALFA)/2.
C      *****CALCULATE THE RHO-VECTOR
      DO 77 I=2,N
      D=-Q*XDIF(I-1)
      DD=D*D
      IF( D .GT. ALOG(2.)) GO TO 83
      APQ=.5*D+DD/(12.+DD/(5.+DD/(28.+DD/(9.+DD/(44.+DD/
     1(13.+DD/60.)))))))
      RHO(I)=1.-D/(1.+APQ)
      GO TO 77
   83 RHO(I)=  EXP(-D)
   77 CONTINUE
      DO 333 I=1,N1
  333 EDIF(I)=EXP(Q*XDIF(I))
      IF( IND .GT. 0) GO TO 99
C     GENERATE THE INITIAL VALUES
      HL(1)=0.
      RL(N)=0.
      DO 334  I=1,N1
      NI=N-I
      HL(I+1)=(1.+HL(I))*EDIF(I)
  334 RL(NI)=(1.+RL(NI+1))*EDIF(NI)
```

```
      DO 335 I=1,N
      G(I)=SQRT(P*(1.+HL(I)+RL(I)))
  335 GG(I)=1./G(I)
C     COMPUTE THE IMPROVED VALUES
   99 CONTINUE
      HL(1)=0.
      RL(N)=0.
      DO 336 I=1,N1
      NI=N-I
      I1=I+1
      HL(I+1)=(GG(I)   +HL(I))*EDIF(I)
  336 RL(NI)=(GG(NI+1)   +RL(NI+1))*EDIF(NI)
      DO 340 I=1,N
      AF(I)=P*(GG(I)   +HL(I)+RL(I))
  340 CRN(I)=AF(I)-G(I)
C      COMPUTE G1=SUM((1-Q*ABS(X(I)-X(J)))*K(I,J)*GG(J))
      HLS(1)=0.
      RLS(N)=0.
      DO 339 I=1,N1
       NI=N-I
      I1=I+1
      HLS(I1)=XDIF(I)*HL(I1)+HLS( I)*EDIF(I)
  339 RLS(NI)=EDIF(NI)*RLS(NI+1)+XDIF(NI)*RL(NI)
      DO 338 I=1,N
  338 G1(I)=(GG(I)+HL(I)+RL(I)-Q*(HLS(I)+RLS(I)))*P
C     COMPUTE CRN(N+1)
      CRN(N+1)=0.
      DO 65 I=1,N
   65 CRN(N+1)=CRN(N+1)+G1(I)*GG(I)
      CRN(N+1)=.5-CRN(N+1)/(4.*ALDA)
C     COMPUTE G1(N+1)=LAST ELE. OF THE DERIVATIVE MATRIX
      HL(1)=0.
      RL(N)=0.
      HLS(1)=0.
      RLS(N)=0.
       DO 66 I=1,N1
       NI=N-I
      HL(I+1)=(GG(I)*X(I)+HL(I))*EDIF(I)
      HLS(I+1)=(GG(I)*X(I)*X(I)+HLS(I))*EDIF(I)
      RL(NI)=(GG(NI+1)*X(NI+1)+RL(NI+1))*EDIF(NI)
   66 RLS(NI)=(GG(NI+1)*X(NI+1)*X(NI+1)+RLS(NI+1))*EDIF(NI)
      APQ=0.
      G1(N+1)=0.
      DO 67 I=1,N
      G1(N+1)=G1(N+1)+GG(I)*G1(I)
   67 APQ=APQ+GG(I)*(X(I)*X(I)*AF(I)+P*(GG(I)*X(I)*X(I)+
     1HLS(I)+RLS(I))-2.*X(I)*P*(GG(I)*X(I)+HL(I)+RL(I)))
      G1(N+1)=3.*G1(N+1)/(8.*ALDA*ALDA)+APQ/(32.*ALFA*ALDA)
C     *****COMPUTE HL,   RL
C     ***** HL CONTAINS THE DIAGONAL TERMS
```

```
C       ***** RL CONTAINS THF OFFDIAGONAL TERMS
         HL(1)= P+G(1)*G(1)
        R(1)=G1(1)
         DO 68 I=2,N
        HL(I)=P +G(I)*G(I)+RHC(I)*RHO(I)*(G(I-1)*G(I-1)-P )
        RL(I-1)=-RHO(I)*G(I-1)*G(I-1)
   68   R(I)=G1(I)-RHO(I)  *G1(I-1)
         DO 69 INT=1,2
         DO 70 I=1,N
         S(I)=HL(I)
   70   T(I)=RL(I)
        DO 71 I=2,N
        S(I)=S(I)-T(I-1)*T(I-1)/S(I-1)
        R(I)=R(I)-T(I-1)*R(I-1)/S(I-1)
         T(I-1)=T(I-1)/S(I-1)
   71   R(I-1)=R(I-1)/S(I-1)
        R(N)=R(N)/S(N)
        DO 72 I=1,N1
        NI=N-I
   72   R(NI)=R(NI)-T(NI)*R(NI+1)
        IF(INT .GT. 1) GO TO 74
        DO 79 I=1,N1
   79   Y(I)=R(I)-RHO(I+1)*R(I+1)
        Y(N)=R(N)
        R(1)=CRN(1)
        DO 80 I=2,N
   80   R(I)=CRN(I)-RHO(I)  *CRN(I-1)
        GO TO 69
   74   DO 81 I=1,N1
   81   AF(I)=R(I)-RHO(I+1)*R(I+1)
        AF(N)=R(N)
   69   CONTINUE
         SA3=0.
        D=0.
        DO 84 I=1,N
        SA3=SA3+G1(I)*Y(I)
   84   D=D+CRN(I)*Y(I)
        SA3=1./(G1(N+1)-SA3/(4.*ALDA*ALDA))
         SA4=SA3*(CRN(N+1)+D/(2.*ALDA))
        APO=SA3*(CRN(N+1)+D/(2.*ALDA))/(2.*ALDA)
         DO 88 I=1,N
        CRN(I)=AF(I)+APO*Y(I)
         GG(I)=GG(I)-CRN(I)
        G(I)=1./GG(I)
   88   CRN(I)=ABS(CRN(I))
        ALDA=ALDA-SA4
        CRN(N+1)=ABS(SA4)
C       NOW PROCEED TO DO THE TESTING
        DO 17 I=1,N
        IF(CRN(I)-EPS1    )17,17,23
```

```
 17 CONTINUE
    IF( CRN(N+1)  -EPS2    )87,87,29
 29 CONTINUE
    IND=1
    NOITN=NOITN+1
    GO TO 198
 87 CONTINUE
    SINFOR=(AN-ALDA)/ALFA
    ESTSX=0.
    DO 668 I=1,N
668 ESTSX=ESTSX+EXP(Q*ABS(SX-X(I)))
    ESTSX=P*P*ESTSX*ESTSX
    IF(INDPR .EQ. 0) GO TO 777
    WRITE(6,73) SINFOR
 73 FORMAT(10X,  38HINTEGRAL OF SQUAREOF 1ST DERIVATIVE
   1  ,E17.10)
777 IF(NITID .EQ. 0) GO TO 735
    WRITE(6,775) NOITN
775 FORMAT(10X,18HNO. OF ITERATION=  ,I2)
735 STOP
    END
    RETURN
```

CHAPTER II

BAYESIAN ESTIMATION OF PROBABILITY DENSITY FUNCTIONS

In this chapter we discuss the Bayesian aspect of estimation of probability density functions. Even though the literature on density estimation is large, the literature on Bayesian estimation of density function is relatively small. The reason is the lack of a suitable prior over the space of density functions. There have been attempts to define priors over the space of probability measures but they have not yielded any workable priors for the purpose of density estimation. Dubins and Freedman (1963) have defined random distribution functions which are singular with probability one. Kraft (1964) has defined a class of distribution function processes which have derivatives, but not continuous derivatives and hence these are not quite suitable for density estimation. The only really convenient prior is Dirichlet process prior due to Ferguson (1973), but unfortunately this prior concentrates all its mass over the discrete distributions.

In section 2.1 we discuss the existence of posterior distributions and conditional expectations for arbitrary priors over the space of continuous density functions. In section 2.2 we discuss the Bayes estimate with respect to squared error loss. We show that the Bayes procedure is Bayes risk consistent. In section 2.3 we discuss the construction of a prior for a particular Gaussian process. Using the result of section 2.2 we show that posterior mean is Bayes risk consistent with respect to squared error loss function. In section 2.4 the possibility of constructing a prior through a given absolutely continuous distribution function is discussed. This avoids the normalization of sample paths as is done in section 2.3. But these again do not give any easy method for computing Bayes estimate. In section 2.5 we give an example to show that the normalization adopted in 2.3 can not always be ignored, i.e., post normalization of sample paths may lead to undesirable estimates.

2.1. Computation of Posterior Measure and Posterior Expectation

Let Ω be the class of all continuous probability density function over the real line. Let F be a σ-field of subsets of Ω such that the random variable $\int_u^v f(x)$, defined over Ω, is F-measurable for all $u < v$ and let μ be a probability measure defined over F. Let I be the class of all intervals with rational end points. Let $X_1, X_2, \ldots X_n$ be iid random variables with common density $f \in \Omega$. Let P_X^f denote the product probability measure defined over $B(R^n)$ induced by f. We see that $P_X^f(B)$, $B \in B(R^n)$ is μ-measurable on F. Then the product measure theorem implies that there exists a unique measure on the product space $(\Omega \times R^n, F \times B(R^n))$ which give rise to μ and P_X^f respectively. Let us denote this measure on $F \times B(R^n)$ by Q. Now for any interval $I \in I$ and $f \in \Omega$ define.

$$I^* : \Omega \longrightarrow [0,1]$$

by $$I^*(f) = \int_I f(x)dx.$$

Then $\{I^* : I\epsilon\mathcal{I}\} \cup \{X_1, \ldots X_n\}$ defines a countable family of random variables. Now let $\{\varepsilon_k\} \downarrow o$ be a sequence of real numbers. For each fixed ε_k and fixed i choose a_{ik} and M_{ik} such that $\mu(A_{ik}) < \frac{\varepsilon_k}{2^i}$ where

$$A_{ik} = \left\{ f: \begin{array}{l} \text{Either} \int_{-M_{ik}}^{M_{ik}} f(x)dx < 1 - \frac{1}{i} \text{ or } |f(x)-f(y)| > \frac{1}{i} \text{ for some} \\ x,y \text{ such that } |x-y| < a_{ik} \text{ and } |x|,|y| \le M_{ik} \end{array} \right\}$$

we may assume that M_{ik} is non-decreasing and a_{ik} is non-increasing in i and k. Now define

$$\Omega_k = \left\{ f: \begin{array}{l} \int_{-M_{ik}}^{M_{ik}} f(x)dx \ge 1 - \frac{1}{i} \text{ and } |f(x)-f(y)| \le \frac{1}{i} \\ \text{whenever } |x-y| \le a_{ik},\ |x|,|y| \le M_{ik} \text{ for all } i. \end{array} \right\}$$

Then clearly $\mu(\Omega_k) > 1 - \varepsilon_k$, Ω_k are increasing and $\Omega = \bigcup_{k=1}^{\infty}\Omega_k \cup \Omega_o$ where $\mu(\Omega_o) = 0$.

Let $C_k = \Omega_k - \Omega_{k-1}$, $k = 1,2, \ldots.$

Define $\nu_k(f) = \begin{cases} 1 & \text{if } f\epsilon C_k \\ 0 & \text{otherwise.} \end{cases}$

Then by Theorem A, Loeve (1963) pp. 361, there exists version of conditional distribution of $\{I_1^*, I_2^*, \ldots\ldots ; \nu_1, \nu_2, \ldots\ldots\}$ given the σ-field generated by $X_1 \ldots. X_n$.

Now for each fixed $t \varepsilon R$ and $f \varepsilon \bigcup_{k=1}^{\infty} \Omega_k$ we have

$$f(t) = \lim_{j} \frac{I_j^*(f)}{\ell(I_j)}$$

where $\ell(\cdot)$ is the Lebesque measure and $\{I_j\} \downarrow \{t\}$, $I_j \varepsilon \mathcal{I}$, $i = 1. 2, \ldots .$

Therefore the conditional distribution defined above determines a conditional joint distribution of $\{f(t), t\varepsilon R\}$ a.e. P_X given the σ-field generated by $X_1 \ldots X_n$, where P_X is the marginal distribution of $(X_1 \ldots X_n)$.

<u>Lemma 2.1.1</u>. <u>Let</u> μ <u>be the prior measure discussed above</u>. <u>Then</u>

$$\text{(a)} \quad \int_\Omega \prod_{i=1}^n f(X_i)\, \mu(df) < \infty \qquad \text{a.e. } P_X$$

and

$$\text{(b)} \quad \int_\Omega f(t) \prod_{i=1}^n f(X_i)\, \mu(df) < \infty \qquad \text{a.e. } P_X$$

<u>Proof</u>. Follows from Fubini's Theorem.

Let CxB be any measurable rectangle in $\mathcal{F} \times \mathcal{B}(R^n)$. Let $I_{C\times B}$ denote the indicator of CxB. Then by Fubini's theorem (Ash, 1972) $\int_{\Omega \times R^n} I_{C\times B}(f,x)\, Q(df, dx)$ exists, where Q is the product measure defined earlier. Also

$$\int_\Omega I_{C\times B}(f, x)\, Q(df, dx)$$

$$= \int_C \left[\int_B \prod_{i=1}^n f(X_i)\, dX_i \right] \mu(df)$$

Since we have already shown the existence of a conditional measure a.e. P_X, the above integral can be written as

$$\int_B \mu_X(C)\, dP_X = \int_B \mu_X(C)\left[\int_\Omega \prod_{i=1}^n f((X_i)\, \mu(df)\right] \prod_{i=1}^n dX_i$$

Therefore $\mu_X(C)$ can be expressed as

$$\mu_X(C) = \frac{\int_C \Pi f(X_i)\, \mu(df)}{\int_\Omega \Pi f(X_i)\, \mu(df)} \qquad \text{a.e. } P_X$$

provided the denominator is strictly positive. This holds for any measurable subset C of Ω.

Therefore

$$\mu_X(f) = \frac{\Pi f(X_i)\, \mu(df)}{\int_\Omega \Pi f(X_i)\, \mu(df)} \qquad \text{a.e. } P_X$$

Also for any fixed $t\varepsilon R$, define the posterior mean

$$\tilde{f}(t) = E(f(t)\,|\underset{\sim}{X}) = \frac{\int_\Omega f(t) \prod_{i=1}^n f(X_i)\, \mu(df)}{\int_\Omega \prod_{i=1}^n f(X_i)\, \mu(df)}$$

$$= \frac{E_X\left[f(t) \prod_{i=1}^n f(X_i)\right]}{E_X\left[\prod_{i=1}^n f(X_i)\right]} \qquad \text{a.e. } P_X.$$

Lemma 2.1.2. (a) $\tilde{f}(t) \geq 0$ for all $t \varepsilon R$, (b) $\int_{-\infty}^{\infty} \tilde{f}(t)\, dt = 1$, a.e. P_X, (c) Assume that for any fixed $t_0 \varepsilon R$ and $X_1, X_2, \ldots X_n$, there exists a neighborhood N_0 of t_0 such that

$$\int_\Omega \left(\sup_{t \varepsilon N_0} f(t)\right) \prod_{i=1}^n f(X_i)\, \mu(df) < \infty$$

then $\tilde{f}$ is continuous at t_0.

Proof: (a) is clear,

(b) follows from Fubini's theorem.

$$\int_R \int_\Omega f(t) \prod_{i=1}^n f(X_i)\, \mu(df)\, dt$$

$$= \int_\Omega \int_R f(t)\, dt \prod_{i=1}^n f(X_i)\, \mu(df).$$

$$= \int_\Omega \prod_{i=1}^n f(X_i)\, \mu(df)$$

Hence $\int_R \tilde{f}(t)\, dt = 1$ a.e. P_X; (c) for fixed ε and $X_1, \ldots. X_n$ choose ε_X such that

$$\frac{\varepsilon_X}{\int_\Omega \prod_{i=1}^n f(X_i)\, \mu(df)} < \varepsilon/4$$

By assumption $\int_\Omega \left(\sup_{t \varepsilon N_0} f(t)\right) \prod_{i=1}^n f(X_i)\, \mu(df) < \infty$

Therefore for given ε_X, there is $\delta(\varepsilon_X)$ such that for any $B \subset \Omega$ with $\mu(B) < \delta(\varepsilon_X)$,

$$\int_B \left(\sup_{t \varepsilon N_0} f(t)\right) \prod_{i=1}^{n} f(X_i)\ \mu(df) < (\varepsilon_X/2).$$

After choosing $\delta(\varepsilon_X)$, choose i such that

$$\mu(A_i) < \frac{\delta(\varepsilon_X)}{2^i} \qquad \text{where}$$

$$A_i = \left\{ f: \begin{array}{l} \text{Either} \int_{-M_i}^{M_i} f(x)dx < 1 - \frac{1}{i} \text{ or } |f(x)-f(y)| > \frac{1}{i} \\ \text{for some } x, y \text{ such that } |x|,|y| \leq M_i \text{ and} \\ |x-y| \leq a_i \end{array} \right\}$$

Let

$$\Omega_{\delta(\varepsilon_X)} = \bigcap_{i=1}^{\infty} (\sim A_i)$$

Clearly

$$\int_{\sim\Omega_{\delta(\varepsilon_X)}} f(t) \prod_{i=1}^{n} f(X_i)\ \mu(df)$$

$$\leq \int_{\sim\Omega_{\delta(\varepsilon_X)}} \left(\sup_{t\ N_0} f(t)\right) \cdot \prod_{i=1}^{n} f(X_i)\ \mu(df)$$

$$< \varepsilon_X \qquad \text{for all } t \varepsilon N_0.$$

Also from the definition of $\Omega_{\delta(\varepsilon_X)}$ it is clear that this set is a set of equicontinuous functions. Choose i_0 such that $\frac{1}{i_0} < \frac{\varepsilon}{2}$. Then

$$|t-t_o| < a_{i_o} \Longrightarrow |f(t)-f(t_o)| < \frac{\varepsilon}{2} \quad \text{for all} \quad f \in \Omega_{\delta(\varepsilon_X)}$$

$$\int_{\Omega_{\delta(\varepsilon_X)}} |f(t)-f(t_o)| \prod_{i=1}^{n} f(X_i)\ \mu(df)$$

$$\leq \frac{\varepsilon}{2} \cdot \int_{\Omega_{\delta(\varepsilon_X)}} \prod_{i=1}^{n} f(X_i)\ \mu(df)$$

$$\leq \frac{\varepsilon}{2} \int_{\Omega} \prod_{i=1}^{n} f(X_i)\ \mu(df)$$

Hence the existence of the neighborhood N_o of t_o with stated property implies

$$\frac{\left| \int_{\Omega} f(t) \prod_{i=1}^{n} f(X_i)\ \mu(df) - \int_{\Omega} f(t_o) \prod_{i=1}^{n} f(X_i)\ \mu(df) \right|}{\int_{\Omega} \prod_{i=1}^{n} f(X_i)\ \mu(df)} \leq \varepsilon$$

for all $t \in (t_o - a_{i_o}\ ,\ t_o + a_{i_o}) \cap N_o$.

This proves the continuity of $\tilde{f}(t)$.

<u>Remark: 2.1.1</u>. The condition

$$\int_{\Omega} \left(\sup_{t \in N_o} f(t) \right) \prod_{i=1}^{n} f(X_i)\ \mu(df) < \infty$$

is much weaker than the condition

$$\int_{\Omega} \left(\sup_{t} f(t) \right)^m \mu(df) < \infty$$

for all m which has been used by Montricher, Tapia and Thompson (1975 a).

2.2. Bayes Estimate and It's Properties

In this section we consider the squared error loss function. Then the pointwise and integrated risk of assuming ϕ_n as an estimate of f is given by

$$R(\phi_n,f)\ (t) = \int_{R^n} (\phi_n(t)-f(t))^2 \prod_{i=1}^{n} f(X_i)dX_i$$

and

$$R(\phi_n,f) = \int_{R} R(\phi_n,f)\ (t)dt$$

Then the corresponding Bayes risks are given

by

$$r_\mu[\phi_n](t) = \int_{\Omega} R(\phi_n,f)(t)\ \mu(df)$$

and

$$r_\mu[\phi_n] = \int_{\Omega} R(\phi_n,f)\ \mu(df)$$

Theorem 2.2.1. Either pointwise or globally, the estimator that minimizes Bayes risk is the posterior mean, i.e.,

$$\inf_{\phi} r_\mu[\phi]\ (t) = r_\mu[\tilde{f}](t)$$

and

$$\inf_{\phi} r_\mu[\phi] = r_\mu[\tilde{f}].$$

where $\tilde{f}$ is the posterior mean.

Next theorem shows that under certain regularity conditions on the prior measure the posterior mean is Bayes risk consistent.

Theorem 2.2.2. Let (Ω, F, μ) be the measure space defined earlier. Let $r_\mu[\tilde{f}](t)$ and $r_\mu[\tilde{f}]$ be Bayes risk and integrated Bayes risk of $\tilde{f}$ respectively. Then

(a) $r_\mu[\tilde{f}](t) \longrightarrow 0$ if there exists a neighborhood $N(t)$ of t

$$\text{such that } \int_\Omega \left(\sup_{u \in N(t)} f(u) \right)^2 \mu(df) < \infty$$

(b) $r_\mu[\tilde{f}] \longrightarrow 0 \quad \text{if} \quad \int_\Omega \left(\int_R f^2(t)dt \right) \mu(df) < \infty$

Proof: Let $K(\cdot)$ be a continuous real valued function satisfying the following conditions

i) $\int K(y)dy = 1$

ii) $\int K^2(y)dy < \infty$,

iii) $K(y) = 0$ if $y \notin [a,b]$ for some a,b.

Define the kernel estimate

$$f_n(x) = \frac{1}{nh(n)} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h(n)}\right)$$

where $h(n) \longrightarrow 0$ and $nh(n) \longrightarrow \infty$ as $n \longrightarrow \infty$. We will show that $r_\mu[f_n](t)$ and $r_\mu[f_n]$ tend to zero as $n \longrightarrow \infty$.

Now $R(f_n,f)(t) = E[f_n(t)-f(t)]^2$

$$= \frac{1}{nh^2(n)} \text{ Var}\left[K\left(\frac{t-x}{h(n)}\right)\right] + \left[Ef_n(t)-f(t)\right]^2$$

Also $$\frac{1}{h(n)} \operatorname{Var}\left[K\left(\frac{t-X}{h(n)}\right)\right] \le \frac{1}{h} \int K^2\left(\frac{t-y}{h(n)}\right) f(y)dy$$

$$= \int_a^b K^2(z)\, f(t-hz)dz$$

(2.2.1) $$\le \left(\sup_{u \in N_h(t)} f(u)\right) \int K^2(z)dz \quad \text{where } N_h(t) = (t-bh,\ t-ah)$$

and $$(E\, f_n(t) - f(t))^2$$

$$= \left[\frac{1}{h(n)} \int K\left(\frac{t-y}{h(n)}\right) f(y)dy - f(t)\right]^2$$

(2.2.2) $$\le 2\left[f^2(t) + \left(\int_a^b K(z)\, f(t-hz)dz\right)^2\right]$$

$$\le 2\left[f^2(t) + \left(\sup_{u \in N_h(t)} f(u)\right)^2 \left(\int |K(y)|dy\right)^2\right]$$

By assumption we have $nh(n) \longrightarrow \infty$ as $n \longrightarrow \infty$. Also $N_h(t) \subset N_o$ if n is sufficiently large. Therefore (2.2.1) and (2.2.2) along with the condition in (a) imply that $R(f_n,f)(t)$ is bounded by an μ-integrable function. Since f is continuous and $nh(n) \longrightarrow \infty$ by assumption, we know that $R(f_n,f)(t) \longrightarrow 0$ for every fixed t and f. Hence the Dominated Convergence Theorem implies (a).

Now to show (b) first observe that

$$\int_{-\infty}^{\infty} \frac{1}{h(n)} \operatorname{Var}\left[K\left(\frac{t-x}{h(n)}\right)\right] dt$$

$$\leq \int_{-\infty}^{\infty} \left[\int_a^b K^2(z)\ f(t-hz)dz \right] dt$$

$$\leq \int_a^b k^2(z)dz$$

Therefore $\frac{1}{nh(n)} \int_{-\infty}^{\infty} \frac{1}{h(n)}$ Var $\left[K\left(\frac{t-x}{h(n)}\right)\right]$ dt $\leq \frac{1}{nh(n)} \int_a^b K^2(z)dz$

implies the above quantity is bounded by a constant which for each fixed f converges to zero.

For the second term in the expression of $R(f_n,f)(t)$ we have

$$\int_{-\infty}^{\infty} \left(E\ f_n(t) - f(t)\right)^2 dt$$

$$\leq 2 \int_{-\infty}^{\infty} \left[f^2(t) + (Ef_n(t))^2 \right] dt.$$

Therefore we need to show that $\int_{-\infty}^{\infty} (Ef_n(t))^2 dt$ is bounded by an integrable function. But

$$\int_{-\infty}^{\infty} (Ef_n(t))^2 dt$$

$$= \int_{-\infty}^{\infty} \left[\int_a^b \int_a^b K(u)\ K(z)\ f(t-hz)\ f(t-hu)dzdu \right] dt$$

$$\leq \frac{1}{2} \int_{-\infty}^{\infty} [\int_a^b \int_a^b |K(z)| f^2(t-hz)\ |K(u)|\ dzdu$$

$$+ \int_a^b \int_a^b |K(u)|\ f^2(t-hu)|K(z)|\ dudz]dt$$

$$= [(\int_a^b |K(z)|dz)^2 (\int f^2(t)dt)]$$

This and the assumption in (b) implies that $\int_{-\infty}^{\infty} (Ef_n(t)-f(t))^2dt$ is bounded by a μ-integrable function. Therefore for any given ε there exists $\delta(\varepsilon)$ such that for any set $B \subset \Omega$ with $\mu(B) < \delta(\varepsilon)$, we have

$$\int_B [\int_{-\infty}^{\infty} (Ef_n(t)-f(t))^2dt]\mu(df) < \varepsilon.$$

Now for this $\delta(\varepsilon)$ choose $\Omega_{\delta(\varepsilon)}$ as before with $\mu(\Omega_{\delta(\varepsilon)}) > 1 - \varepsilon$. Hence

(2.2.3) $$\int_{\sim\Omega_{\delta(\varepsilon)}} \int_{-\infty}^{\infty} (Ef_n(t)-f(t))^2dt\ \mu(df) < \varepsilon$$

Also $$\int_{-N}^{N} \left(Ef_n(t) \right)^2 dt$$

$$\leq \left(\int_a^b |K(z)|dz \right)^2 \int_{-N-bh(n)}^{N-ah(n)} f^2(t)dt$$

Therefore for any fixed M

$$\int_{\Omega_{\delta(\varepsilon)}} \int_{|t|>M} \left(Ef_n(t) - f(t) \right)^2 dt \; \mu(df)$$

$$\leq \text{Const.} \int_{\Omega_{\delta(\varepsilon)}} \left(\int_{-M-bh(n)}^{M-ah(n)} f^2(t)dt \right) \mu(df)$$

Now $$\int_\Omega \int_R f^2(t)dt \; \mu(df) < \infty$$

and the sets $\Omega_{\delta(\varepsilon)} x[-M-bh(n), M-ah(n)] \longrightarrow \phi$ implies $\int_{\Omega_{\delta(\varepsilon)}} \int_{|t|>M} (Ef_n(t) - f(t))^2 dt \; \mu(df)$ can be made as small as we want by taking M sufficiently large. Let $M(\varepsilon)$ be such that

(2.2.4) $$\int_{\Omega_{\delta(\varepsilon)}} \int_{|t|>M(\varepsilon)} \left(Ef_n(t) - f(t) \right)^2 dt \; \mu(df) < \varepsilon.$$

Now consider

(2.2.5) $$\int_{\Omega_{\delta(\varepsilon)}} \int_{-M(\varepsilon)}^{M(\varepsilon)} \left(Ef_n(t) - f(t) \right)^2 dt \; \mu(df)$$

Since f is uniformly continuous on $[-M(\varepsilon), M(\varepsilon)]$ we see that

$$\int_{-M(\varepsilon)}^{M(\varepsilon)} \left| Ef_n(t) - f(t) \right|^2 dt \longrightarrow 0$$

for every $f \varepsilon \Omega_{\delta(\varepsilon)}$. Hence by Dominated Convergence Theorem we conclude that (2.2.5) converges to zero. Also (2.2.3), (2.2.4) and 2.2.5) together imply (b).

Remark: 2.2.1. Our assumptions are much weaker than the assumption, $\int (\sup_t f(t))^m \mu(df) < \infty$ for all $m > 0$, made by Montricher, Tapia and Thompson (1975 a).

Even though we have proved a desirable property of the Bayes estimate, the estimate itself is far from being computable. It is worth mentioning the only "easily computable" Bayesian type estimate due to Wahba [1976]. We call it Bayesian type because the class of functions over which the prior measure is considered, is not a class of probability density functions.

2.3. Construction of a Prior Through a Gaussian Process

In this section we will take for the prior measure a particular stochastic process. It will be shown that the Bayes estimator with respect to this prior is Bayes risk consistent. Define the stochastic process $\{z(t) : t \varepsilon R\}$ in the following way:

$$z(t) = \exp[-\lambda|t| + W(t)] \quad \text{where}$$

$W(t)$ is a Gaussian process with independent increments and with

(i) $W(0) = 0$

(ii) $E\,W(t) = 0$ for all t,

(iii) $Cov[W(t_1), W(t_2)] = \delta(t_1, t_2) \min(|t_1|, |t_2|)$

$$\text{where } \delta(t_1, t_2) = \begin{cases} 1 & \text{if } t_1 t_2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

Lemma 2.3.1. $Z(\cdot)$ is integrable a.s.

Proof: Since $\{W(t) : t>0\}$ is a Brownian Motion, for any $\lambda > 0$ (Breinman, pp 289)

$$P\left\{ W(t) < a + \frac{1}{2}\lambda t \quad \text{for all} \quad t > 0 \right\} = 1 - e^{-a\lambda}$$

Hence $$P\left\{ Z(t) < e^{a - \frac{1}{2}\lambda |t|}, \quad \text{for all } t \right\} > 1 - 2e^{-a\lambda}$$

implies $$P\left\{ \int_{-\infty}^{\infty} Z(t)dt < 4\frac{e^a}{\lambda} \right\} > 1 - 2e^{-a\lambda}$$

$$\longrightarrow 1 \text{ as } a \longrightarrow \infty.$$

Now define $\phi(t) = \dfrac{Z(t)}{\int Z(u)du}$. Then

i) $\phi(t) \geq 0$ for all $t \in R$,

ii) $\int \phi(t)dt = 1$, a.e.

iii) $\phi(t)$ is continuous.

Therefore the normalized sample functions $\phi(\cdot)$ can be considered as random probability density functions. Next we present a lemma, due to Rubin, which will be used latter.

Lemma 2.3.2 (Rubin, 1976). Let

$$\frac{1}{X} = \int_0^\infty e^{-\lambda t + W(t)} dt \qquad \lambda > 0$$

where $W(\cdot)$ is the Gaussian process defined above.
Then $X \sim$ Gamma $(2,2\lambda)$.

Now we develope some lemmas which will be used to show that posterior mean is Bayes risk consistent with respect to this prior and squared error loss.

Define t^* such that $\phi(t^*) = \sup_{t > 0} \phi(t)$. Then

$$\phi(t^*) = \frac{Z(t^*)}{\int Z(u)dn} \leq \frac{Z(t^*)}{\int_{t^*}^\infty Z(u)du} = Z_0 \text{ (say).}$$

Now for any fixed α, define

$$t_\alpha = \inf\{t : -\lambda(t-t^*) + W(t) - W(t^*) = -\alpha \quad \text{and} \quad t \geq t^*\}$$

and $$\frac{1}{Z_\alpha} = \int_{t_\alpha}^\infty e^{-\lambda(t-t_\alpha) + W(t) - W(t_\alpha)} dt \text{ , then } Z_0 \leq e^\alpha Z_\alpha.$$

The following lemma shows that Z_α is a well behaved random variable in the sense that its moment generating function exists.

Lemma 2.3.3. Let Z_α be as defined above. Then for α sufficiently large the moment generating function of Z_α exists.

<u>Proof</u>: For any fixed $a > 0$

$$P\left\{Z_\alpha > a\right\}$$

$$= P\left\{\int_{t_\alpha}^{\infty} e^{-\lambda(t-t_\alpha) + W(t) - W(t_\alpha)} dt < \frac{1}{a} \;\middle|\; \begin{array}{l} -\lambda(t-t_\alpha)+W(t)-W(t_\alpha) \leq \alpha \\ \text{for all } t \geq t_\alpha \end{array} \right\}$$

$$\leq \frac{P\left\{\int_{t_\alpha}^{\infty} e^{-\lambda(t-t_\alpha) + W(t) - W(t_\alpha)} dt < \frac{1}{a}\right\}}{P\left\{-\lambda(t-t_\alpha) + W(t) - W(t_\alpha) \leq \alpha, \quad \text{for all } t \geq t_\alpha\right\}} .$$

$$= \frac{P\left\{\int_0^{\infty} e^{-\lambda t + W(t)} dt < \frac{1}{a}\right\}}{(1 - 2e^{-\alpha\lambda})} \qquad \text{Using strong Markov property}$$

$$\leq 2 \; P\left\{\int_0^{\infty} e^{-\lambda t + W(t)} dt < \frac{1}{a}\right\} \qquad \text{if } \alpha > \frac{1}{2\lambda}.$$

But $P\left\{\int_0^{\infty} e^{-\lambda t + W(t)} dt < \frac{1}{a}\right\}$

$$= \frac{2^{2\lambda}}{\Gamma(2\lambda)} \int_a^{\infty} e^{-2x} \; x^{2\lambda-1} \; dx$$

$$\leq C(\lambda) \frac{a^{2\lambda} \; e^{-2a}}{a - \max(0, \lambda-\frac{1}{2})} \qquad \text{if } a > \lambda-\frac{1}{2}$$

where $C(\lambda)$ is a constant depending only on λ.

$$\text{Now} \quad E(e^{tZ_\alpha}) = \int_1^\infty P\left\{e^{tZ_\alpha} > c\right\} dc$$

$$= \int_1^\theta P\left\{e^{tZ_\alpha} > c\right\} dc$$

$$+ \int_\theta^\infty P\left\{e^{tZ_\alpha} > c\right\} dc \quad , \quad \theta > e^{t(\lambda - \frac{1}{2})}.$$

Hence it is sufficient to show that second term in the above expression is finite.

$$\text{But} \quad \int_\theta^\infty P\left\{e^{tZ_\alpha} > c\right\} dc$$

$$= \int_\theta^\infty P\left\{Z_\alpha > \frac{\log c}{t}\right\} dc$$

$$= \int_{\frac{\log \theta}{t}}^\infty P\left\{Z_\alpha > a\right\} \quad t \quad e^{at} \quad da$$

$$\leq \int_{\frac{\log \theta}{t}}^\infty t \quad e^{at} \quad \frac{c(\lambda) \quad a^{2\lambda} \quad e^{-2a}}{a - \max(0, \lambda - \frac{1}{2})} \quad da \quad < \infty$$

$$\text{if} \quad t < 2$$

This completes the proof of the lemma.

Corollary 2.3.1. Let $\phi(t)$ be as defined above. Then $(\sup_t \phi(t))$ has finite moments of all orders.

Corollary 2.3.2. Let $r_Z(\tilde{f})(t)$ and $r_Z(\tilde{f})$ be pointwise and integrated Bayes risk of the Bayes estimator, $\tilde{f}$, the posterior mean, with respect to the prior induced by $\{Z(t), t \in R\}$. Then

$$r_Z(\tilde{f})(t) \longrightarrow 0$$
$$r_Z(\tilde{f}) \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty$$

Proof: Follows from Corollary 2.3.1, lemma 2.2.3 and Theorem 2.2.2.

Under the setup of this section the Bayes estimate of the density at a point t given n-observations $X_1, X_2, \ldots X_n$ can be expressed as

$$\tilde{\phi}(t) = \frac{E_X\left[\phi(t) \prod_{i=1}^{n} \phi(X_i)\right]}{E_X\left[\prod_{i=1}^{n} \phi(X_i)\right]}$$

where
$$\phi(u) = \frac{Z(u)}{\int Z(\nu)d\nu}$$

The Bayes estimate in the above form looks simple but the evaluation of the numerator and the denominator in the expression of $\tilde{\phi}(t)$ will be very difficult. We will derive a computable expression for the prior mean $E\ \phi(t)$. This will demonstrate the magnitude of difficulty one will have in attempting to evaluate the expressions for

$$E_X\left[\phi(t) \prod_{i=1}^{n} \phi(X_i)\right] \text{ and } E_X\left[\prod_{i=1}^{n} \phi(X_i)\right].$$

Without loss of generality let us assume that $t > 0$. Let $-\lambda t + W(t) = u$.

Then $$E\,\phi(t) = \int_{-\infty}^{\infty} E\left\{ \frac{e^{u}}{\int_{-\infty}^{\infty} e^{-\lambda|\nu| + W(\nu)}d\nu} \;\middle|\; -\lambda t + W(t) = u \right\} \frac{e^{-\frac{1}{2t}(u + \frac{t}{2})^2}}{\sqrt{2\pi t}}\, du.$$

We will derive a computable expression for the conditional expectation under the integral sign. Let us introduce the following notations.

$$X^- = \int_{-\infty}^{0} e^{-\lambda|\nu| + W(\nu)}d\nu$$

$$X^+ = \int_{0}^{\infty} e^{-\lambda\nu + W(\nu)}d\nu$$

$$Y = \int_{0}^{t} e^{-\lambda\nu + W(\nu)}d\nu \qquad \text{for } t > 0$$

$$V = \int_{t}^{\infty} e^{-\lambda\nu + W(\nu)}d\nu.$$

$$= e^{u} \cdot V'$$

X^- and V' are identically and independently distributed. Before we proceed to compute the conditional expectations we will state two known results.

Let X^- and X^+ be as defined. Then the following result is known.

Result 1 $\quad X^-/(X^- + X^+) \sim Be(2\lambda, 2\lambda)$ on [01] and is independent of $(X^- + X^+)$.

The special case with $\lambda = \frac{1}{2}$ is mentioned in Paranjape and Rubin (1975).

Lemma 2.3.4. (Paranjape and Rubin 1975) Let Y be as defined above. If $Y_1 = e^{u/2}/Y$, then the Laplace transform of Y_1, viz;

$$E\left[e^{-\theta Y_1}\right] = \psi(\theta) \quad , \text{ is given by}$$

$$\psi(\theta) = \exp\left[\frac{1}{2}\frac{u^2}{t}\right] \exp\left[-\frac{2}{t} \text{ arc } \text{Cosh}^2\left(\text{Cosh}\frac{u}{2} + \frac{\theta}{4}\right)\right] \quad \text{for } \theta \geq -4.$$

Define $B = \frac{1}{V'} + \frac{1}{X^-}$ and $A = \frac{X^-}{X^- + V'}$. Also from now on we will omit the conditioning event $-\lambda t + W(t) = u$. Hence the conditional expectation can be expressed as

$$E\left[\frac{e^u}{X^- + Y + e^u V'}\right]$$

$$(2.3.1) \quad = E\left[\frac{B e^u}{\frac{1}{1-A} + \frac{e^u}{A}}\right] - E\left[\frac{B^2 e^{3u/2}}{\frac{1}{1-A} + \frac{e^u}{A}} \quad \frac{1}{Be^{u/2} + Y_1\left(\frac{1}{1-A} + \frac{e^u}{A}\right)}\right]$$

Also we can write

$$\frac{1}{B e^{u/2} + Y_1\left(\frac{1}{1-A} + \frac{e^u}{A}\right)} = \int_0^\infty e^{-x\left[B e^{u/2} + Y_1\left(\frac{1}{1-A} + \frac{e^u}{A}\right)\right]} dx$$

Now taking the expectation with respect to Y_1 we get

$$E\left[B e^{u/2} + Y_1\left(\frac{1}{1-A} + \frac{e^u}{A}\right)\right]^{-1}$$

$$= \int_0^\infty e^{-xB e^{u/2}} \int_0^\infty e^{-x\left(\frac{1}{1-A} + \frac{e^u}{A}\right) y_1} f(y_1) \, dy_1 \, dx$$

$$= \int_0^\infty e^{-x\, Be^{u/2}} \psi\left[x \left(\frac{1}{1-A} + \frac{e^u}{A}\right)\right] dx$$

Hence the second term in (2.3.1) can be expressed as

$$\exp[3u/2] \int_0^\infty \left[\int_0^1 \left(\frac{b^2}{\frac{1}{1-a} + \frac{e^u}{a}} \int_0^\infty e^{-x\, b\, e^{u/2}} \psi\left[x \left(\frac{1}{1-a} + \frac{e^u}{a}\right)\right] dx \right.\right.$$

$$(2.3.2) \qquad \left.\left. \cdot \frac{e^{-2b}}{\Gamma(4\lambda)} (2b)^{4\lambda-1} \right) d(2b) \quad q(u) \right] da.$$

where $\quad q(a) = \frac{1}{B(2\lambda,\, 2\lambda)} \; a^{2\lambda-1} \, (1-a)^{2\lambda-1}$.

Also

$$(2.3.3) \qquad E\left[\frac{B\, e^u}{\frac{1}{1-A} + \frac{e^u}{A}}\right] = (2\lambda) \quad e^u \int_0^1 \left(\frac{1}{1-a} + \frac{e^u}{a}\right) q(a) \quad da.$$

Now substituting (2.3.2) and (2.3.3) back into (2.3.1) we can get a computable expression for

$$E\left[\frac{e^{u/2}}{X^- + Y + e^u \quad V'} \;\middle|\; -\lambda t + W(t) = u\right]$$

Using this the unconditional expectation

$$E\left[\frac{Z(t)}{\int Z(u)du}\right] \text{ can be computed.}$$

2.4. Construction of a Prior Through a Given Distribution Function

In this section we will describe another way of constructing a random probability density function. Let $\{Y(t) : t \in R\}$ be a stochastic process such that $Y^2(t)$ is integrable a.s.

Define
$$Z(t) = \int_a^t Y^2(u)du \qquad \text{if } t>a$$
$$= -\int_t^a Y^2(u)du \qquad \text{if } t<a$$

Hence $Z(t)$ is well defined for all $t \in R$. Let F be any fixed absolutely continuous distribution function. Then $F(Z(t))$ defines a random distribution function and corresponding density function is given by

$$\phi(t) = \frac{dF(Z(t))}{dt} = f(Z(t)) \quad Y^2(t)$$

where
$$f(t) = \frac{dF(t)}{dt}$$

$\phi(\cdot)$ defined in this way is a proper density function. Hence the posterior expectation can be expressed as

$$\tilde{\phi}(t) = \frac{E_X\left[\phi(t) \quad \prod_{i=1}^n \phi(X_i)\right]}{E_X\left[\prod_{i=1}^n \phi(X_i)\right]} .$$

So far the Bayes estimate looks good but getting a computable expression is difficult.

Example 2.4.1. Let us consider two specific density functions.

(a) $$f(y) = \frac{1}{2|y|^2} e^{-\frac{1}{|y|}}$$

and (b) $$f(y) = \frac{1}{2} e^{-|y|}.$$

As before let $Y(t) = e^{-\lambda|t|} + W(t)$

and $$Z(t) = \int_0^t Y(u)du \qquad t>0$$

$$= -\int_t^0 Y(u)du \qquad t<0$$

where $W(\cdot)$ is the Gaussian process considered earlier. Suppose $t>0$. Then the prior mean of $\phi(t)$ in case (a) is given by

$$E\,\phi(t) = \frac{1}{2} E\left[e^{-\frac{1}{|Z(t)|}} \frac{Y(t)}{(Z(t))^2}\right]$$

$$= \frac{1}{2} \int e^u \; E\left[\frac{e^{-\frac{1}{Z(t)}}}{(Z(t))^2} \,\middle|\, -\lambda t + W(t) = u\right] \cdot \frac{e^{-\frac{1}{2t}(u+\lambda t)^2}}{\sqrt{2\pi t}} du$$

Let $Y_1 = \frac{e^{\frac{u}{2}}}{Z(t)}$. Then

$$E\left[\frac{e^{-\frac{1}{Z(t)}}}{(Z(t))^2} \,\middle|\, -\lambda t + W(t) = u\right]$$

$$= \int_0^\infty e^{-\frac{u}{2} y_1} \quad e^{-u} \quad y_1^2 \quad g(y_1) \quad dy_1$$

$$= e^{-u} \quad \psi'' \left(\frac{u}{2} \right)$$

Where $g(\cdot)$ is the density of Y_1.

Hence $$E\,\phi(t) = \frac{1}{2} \int_{-\infty}^{\infty} \psi'' \left(\frac{u}{2} \right) \frac{e^{-\frac{1}{2t}(u+\lambda t)^2}}{\sqrt{2\pi t}} \, du.$$

Again computation of posterior mean presents no less difficulty.

It seems that the computation of the posterior mean in case (b), i.e., with $f(t) = \frac{1}{2} e^{-|t|}$ may be less difficult.

Let $X_1 < X_2 < \ldots < X_n$ be the ordered sample. Then we want to evaluate the following expectation.

$$E\left[\prod_{i=1}^{n} Y(X_i) \quad \exp\left(-\frac{1}{2} \sum_{i=1}^{n} Z(X_i)\right)\right]$$

Therefore first conditioning on $-\lambda|X_i| + W(X_i) = u_i$, we need to evaluate conditional expectation of the following form;

$$\text{(2.4.1)} \qquad E\left[\exp[-\tfrac{1}{2}(Z(X_{i+1})-Z(X_i))] \,\middle|\, -\lambda X_j + W(X_j) = u_j,\ j = i,\ i+1\right]$$

But $$Z(X_{i+1}) - Z(X_i) = \int_{X_i}^{X_{i+1}} e^{-\lambda u + W(u)} du$$

$$= e^{u_i} \int_{0}^{(X_{i+1}-X_i)} e^{-\lambda u + W(u)} du$$

Hence the distribution of $Z(X_{i+1}) - Z(X_i)$ given $-\lambda X_j + W(X_j) = u_j$, $j = i, i + 1$, is same as that of

$$e^{u_i} \int_{0}^{(X_{i+1} - X_i)} e^{-\lambda \nu + W(\nu)} d\nu$$

given that $-\lambda(X_{i+1} - X_i) + W(X_{i+1} - X_i) = u_{i+1} - u_i = \Delta u_i$,

But $\exp[\frac{\Delta u_i}{2}] / \int_0^{\Delta X_i} e^{-\lambda u + W(u)} du$ has the conditional density

$$f_i(y) = \frac{1}{y^2} \exp\left[- \frac{\Delta u_i}{2\Delta X_i} - 4y \quad \mathrm{Cosh}\left(\frac{\Delta u_i}{2}\right) + \psi(y)\right]$$

Therefore the expectation in (2.4.1) can be written as

$$\int_0^{\infty} \exp\left[- \frac{1}{2} \, e^{u_i} \quad e^{(\Delta u_i/2)} \quad \frac{1}{y}\right] \quad f_i(y)dy$$

Of course this expectation involves the unknown function ψ, where the Laplace transform of $\frac{e^{\psi(a)}}{a^2}$ is given by $\exp[- \frac{2}{t} \;\; \mathrm{arc\ Cosh}^2(\frac{\theta}{4})]$. Once these individual expectations are obtained then a "computable" expression for unconditional expectation can be obtained.

2.5. Some Negative Results

In this section we gave an example to show that in the stochastic process formulation of the prior distribution it is necessary to consider the normalized sample functions. In the absence of such normalization the estimator could be a bad one. Specifically let $\{Y(t) : t \in R\}$ be a Gaussian process with mean function $m(t)$ and covariance function $\sigma(\cdot,\cdot)$. The functions $m(\cdot)$ and $\sigma(\cdot,\cdot)$ can be chosen to make $Z(t) = \exp[Y(t)]$ integrable a.s. In section 2.3 we have considered the normalized sample functions

$$\phi(t) = \frac{Z(t)}{\int Z(u)du}$$

as the random density function and the Bayes estimate was

$$\tilde{\phi}(t) = \frac{E_X\left[\phi(t) \quad \prod_{i=1}^{n} \phi(X_i)\right]}{E_X\left[\prod_{i=1}^{n} \phi(X_i)\right]}$$

Now suppose we consider the prior measure over the class of functions $\{Z(t) : t \in R\}$, induced by $\{Y(t) : t \in R\}$.
Then the pseudo Bayes estimate will be given by

$$\phi^*(t|X) = \frac{E_X\left[Z(t) \quad \prod_{i=1}^{n} Z(X_i)\right]}{E_X\left[\prod_{i=1}^{n} Z(X_i)\right]}$$

$$= \exp[-m(t) + \frac{1}{2}\sigma(t,t) + \sum_{i=1}^{n} \sigma(t,X_i)].$$

Also for given $X_1, X_2, \ldots X_n$, $\int \phi^*(t|X)dt < \infty$ for appropriate choice of $m(\cdot)$ and $\sigma(\cdot,\cdot)$.

Therefore $\tilde{\phi}(t) = \dfrac{\phi^*(t|X)}{\int \phi^*(t|X)dt}$. can be taken as an estimate of density function. The following lemma follows from Theorem 7.2.5 Ash (1972) pp. 275.

Lemma 2.5.1. Let $Z_1, Z_2, \ldots\ldots Z_n$ be a sequence of iid random variables. Define

$$S_n = Z_1 + \ldots\ldots + Z_n$$

Then either (a) S_n diverges to $+\infty$ or $-\infty$

or (b) S_n oscillate between $-\infty$ and $+\infty$

or (c) $S_n \equiv 0$ a.s.

The following theorem illustrates that the post-normalized estimate $\tilde{\phi}(t)$ based on n-observations is not a good estimate. For any t and u, let

$$L(X_1 \ldots X_n|t, u) = \log \frac{\tilde{\phi}(t)}{\tilde{\phi}(u)}$$

Theorem 2.5.1. Let $L(X_1 .. X_n|t, u)$ be as defined above. Then

either (a) $L(X_1 .. X_n|t, u)$ diverges to $+\infty$ or $-\infty$

or (b) $L(X_1 .. X_n|t, u)$ oscillates between $+\infty$ and $-\infty$

or (c) $L(X_1 .. X_n|t, u)$ is constant (depending on t and u only) a.s. for all n.

Proof: Define $T_i = \sigma(t, X_i) - \sigma(u, X_i)$

and $S_n = T_1 + \ldots + T_n$.

Then $(X_1 \ldots X_n | t,u) = -m(t) + m(u) + \frac{1}{2}\sigma(t,t) - \frac{1}{2}\sigma(u,u) + S_n$.

Since $T_1, \ldots T_n$ are iid, S_n will have the properties stated in the lemma 2.5.1.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ash, R. B. (1972). Real Analysis and Probability Theory. Academic Press Inc.

Boneva, L., Kendall, D. and Stefanov, I. (1971). Spline transformations: Three new diagnostic aids for the statistical data analyst. J. Roy. Statist. Soc. 33: 1-70.

Dubins, L. E. and Freedman, D. A. (1963). Random distribution functions. Bull. Amer. Math. Soc. 69: 548-551.

Ferguson, T. S. (1973). A Bayesian Analysis of some nonparametric problems. Ann. Statist. 1: 209-230.

Good, I. J. and Gaskin, R. A. (1971). Nonparametric roughness penalties for probability densities. Biometrika, 58: 255-277.

Good, I. J. and Gaskin, R. A. (1972). Global nonparametric estimation of probability densities. Virginia J. Sc. 23: 171-193.

Grenander, U. (1956). On the theory of mortality measurements, Part-II. Skan. Aktuarictidskr. 39: 125-153.

Iaasacson, E. and Keller, H. B. (1966). Analysis of Numerical Methods. John Wiley and Sons Inc., New York.

Kraft, C. H. and VanEden, C. (1964). Bayesian bio-assay. Ann. Math. Statist. 35: 886-890.

Kraft, C. H. (1964). A class of distribution function processes which have derivatives. J. Appl. Prob. 1: 385-388.

Kronmal, R. and Tarter, M. (1968). The estimation of probability density and cumulatives by Fourier Series methods. J. Amer. Statist. Assoc. 63: 925-952.

Loeve, M. (1963). Probability Theory. Van Nostrand. East-West Press.

Montricher, G. F., Tapia, R. A. and Thompson, J. R. (1975 a). Nonparametric Bayesian estimation of probability densities by function space techniques. Rice Univ. Memeograph.

Montricher, G. F., Tapia, R. A. and Thompson, J. R. (1975 b). Nonparametric maximum likelihood estimation of probability densities by penalty function methods. Ann. Statist. 3: 1329-1348.

Paranjape, S. R. and Rubin, H. (1975). Special case of the distribution of the median. Ann. Statist. 3: 251-256.

Parzen, E. (1962). On the estimation of probability density function and mode. Ann. Math. Statist. 33: 1065-1076.

Rao, B. L. S. P. (1969). Estimation of unimodal density. Sankhya Ser. A. 31: 23-36.

Reiss, R. D. (1973). On the measurability and consistency of maximum likelihood estimates for unimodal densities. Ann. Statist. 5: 888-901.

Robertson, T. (1967). On estimating a density which is measurable with respect to a σ-lattice. Ann. Math. Statist. 38: 482-493.

Rosenblatt, M. (1956). Remark on some nonparametric estimates of density functions. Ann. Math. Statist. 27: 832-837.

Rubin, H. (1976). Personal communication.

Rubin, H. (1977). Uniform convergence of the random component of kernel density estimators. IMS Bulletin, Vol-6, No. 2, Issue No. 31: 77t-25.

Wahba, G. (1975). Interpolating spline methods for density estimation-I. Equispaced knots. Ann. Statist. 3: 30-48.

Wahba, G. (1976). A survey of some smoothing problems and the methods of generalized cross validation for solving them. Univ. of Wisconsin, Department of Statistics Technical Report #457.

Wegman, E. J. (1972 a). Nonparametric probability density estimation-I. A survey of available methods. Technometrics 14: 533-546.

Wegman, E. J. (1972 b). Nonparametric probability density estimation-II. A comparison of density estimation methods. J. Statist. Comput. simul. 1: 225-245.