On Logistic and Some New Discrimination Rules: Characterizations, Inference and Applications.

Supratik Roy

Thesis submitted to Indian Statistical Institute, in partial fulfillment of requirements for award of Doctor of Philosophy.

September, 1997

Revision, 1999

Contents

ACKNOWLEDGEMENTS

I express my sincere gratitude to Council for Scientific and Industrial Research, Government of India, for granting me a research Fellowship, under the NET program, first as JRF and then advanced to SRF, with continuous support for the whole of my research tenure. I am deeply thankful to Indian Statistical Institute, Calcutta, my alma mater, for permitting me to be affiliated to it as a research fellow, and providing me access to computational facilities.

I am deeply thankful to an anynomous referee who had not asked for any change in the initial version of my thesis. I also appreciate the remarks and suggestions of another anynomous referee which have been incorporated in this current version of the thesis. I am deeply grateful to Professor Sir D.R. Cox for his encouragements and suggestions which have enriched this thesis.

It is my great pleasure to acknowledge, with thanks, the Office of the Dean, I.S.I. - Shri Apurba Guha, in particular, who saw me through the various administrative requirements of both CSIR and ISI. I also thank CSU-ASD for my affiliation with it – Shri Amiya Das, in particular, who bore patiently with my insistent demands for various exotic computational facilities and operational procedures. It is my real pleasure to thank Shri Subhasish Pal of CSSC for helpful discussions on my computer programs.

Finally, I thank my research supervisor, Prof. Ashis SenGupta of CSU-ASD, whose continuous guidance, exhortations and stern reminders to solve the pending problems helped me advance towards my goal, and who almost solely encouraged and urged me to complete my work.

Chapter 1

Introduction

1.1 Introduction and Summary

Consider the problem of classification of an observation into one of two specified populations. Fisher's classification rule, just as several other rules commonly used in practice, depends only on the ratio of the individual densities $f_i(\mathbf{x}), i = 1, 2$. This led Cox [1966],[27] to model the "posterior odds" by a simple function . Specifically ,

$$\ln\left(\frac{q_1(\mathbf{x})}{q_2(\mathbf{x})}\right) = \ln\left(\frac{f_1(\mathbf{x}).\pi_1}{f_2(\mathbf{x}).\pi_2}\right) = \alpha + \beta'\mathbf{x} \tag{1.1}$$

Cox's logistic discrimination (LGD) rule is then based on the statistic $\alpha + \beta'\mathbf{x}$. This has the advantage that individual densities $f_i(\mathbf{x})$ need not be known and we only need to estimate the parameters α and β.

Another advantage , which is claimed , is that the family of densities satisfying (1.1) is "quite wide". It is this richness of the family that we intend to explore, since, beyond the multivariate normal distribution and multivariate discrete distributions following the log-linear model, no parametric description of this class is available. Some of the ideas developed here have been introduced by us, in ([91],[92]).

We introduce the following definitions.

Definition 1.1.1 *Let a distribution possess a density $f_1(\mathbf{x})$ with respect to measure ν. $f_1(\mathbf{x}) \in C$, some specified class, will be said to* *(i) obey the LGD relationship if $\exists$ a density $f_2(\mathbf{x})$ with respect to some measure μ , satisfying (1.1).*

(ii) admit the LGD if $f_2(\mathbf{x})$ given by (1.1) is also in the class C.

Clearly a rule capable of discriminating (in probability) between a pair of populations only when they belong to entirely two different classes is of limited use if the class of such pairs is large , with at least any one component of the pair corresponding to an usually encountered distribution. It is thus essential and important to characterize such pairs (f_1, f_2) or equivalently f_1. In this pursuit, a general method of characterizing families admitting and/ or obeying Cox's [27] LGD for a given density is obtained through functional equations in characteristic type functions. The result is applied to different types and classes of univariate as well as multivariate distributions. This characterization enables us to generalize Cox's LGD to families even not representable by densities with respect to any measure ν , e.g. to the stable and proper infinitely divisible families. We observe that almost all distributions f_1 usually encountered in practice obey the LGD relationship. Of notable exceptions are the Cauchy etc. However the class of distributions for f_1 admitting LGD is somewhat restrictive. We recommend that we should not use LGD when any of these distributions is suspected for f_1 . Rather a preliminary data analysis, should be carried out to identify f_1 more clearly and then accordingly decide on the use of LGD.

Our characterization exposes a functional relationship among the parameters which renders the validity of the usually proposed likelihood based estimation procedures (Anderson, [1982],[5]) theoretically suspect. We present the proper likelihood equations and propose a scheme to evaluate the parameters. The performance of our rule is established to be quite satisfactory through extensive simulations. LGD is illustrated through two well-known real life data sets.

We also suggest and explore some new discrimination rules in the context of stable distributions, directional data, and neural networks. For the case of stable distributions the performance of the rule is studied by using a new real life data set while for that of directional data we use a well-known real-life data set available in the literature. We provide a listing of the source code for various programs developed by us which include both programs written in C as well as in SPLUS.

Chapter 2

The univariate family admitting LGD

2.1 Introduction

We take up the problem of characterization of the family of distributions admitting or obeying the LGD rule in the semiparametric framework developed by Cox [27], Anderson [5] and others, and in a subsequent chapter we consider the problem of estimation in the presence of non-orthogonalizable functional dependence. We also derive results about the form of these dependencies.

As we have already seen, ([92]), that a pair of densities (f_1, f_2), satisfying (1.1) involves normalizing constants which are Laplace transforms, for if $\ln(f_1/f_2) = \alpha + \beta'\mathbf{x}$, then

$$f_1(\mathbf{x}) = \frac{e^{\beta'\mathbf{x}} f_2(\mathbf{x})}{\int e^{\beta'\mathbf{x}} f_2(\mathbf{x}) d\mathbf{x}}$$

Thus LGD is intimately connected with the existence of appropriate Laplace transforms.

With thanks to a Referee, in this connection we briefly review a class of distributions described as "tilted distributions" or "weighted distributions" [83][1985]. For a given probability density $p(x, \theta)$ corresponding to random variable X define

$$p^w(x, \theta, \alpha) = \frac{w(x, \alpha) p(x, \theta)}{E(w(X, \alpha)} \tag{2.1}$$

The weight function $w(x, \alpha)$ can be any arbitrary non-negative function with finite expectation. Rao [83] goes on to describe size-biased distributions where the weight function is $|x|$. Characterizations for this size-biased case have been treated in Janardhan and Rao [57][1983] . However in dealing with LGD we can think of the LGD obeying class as an weighted class with the weight function $\exp(\alpha + \beta' x)$. But results available in their generalities now require further exercises to become specifically applicable to our case. We however, have taken the direct route of attacking this specific case *ab initio*. A characterization for such a specific class contributes to the characterizations of general weighted distributions significantly - for the LGD class coincides with the class of distributions which will admit a representation of a log-linear ratio-of-density (Fisher-type) discriminant function.

2.2 Basic Framework

We will need the following domains of the $z = t + iy$ complex plane,

$$H_{\pm} := \{z : \pm y > 0\}, \quad D := \{z : y = 0\}, \quad S_{ab} = \{z : -a < y < b\}, \tag{2.2}$$

where $0 \le a$, $b \le \infty$. Moreover, we use the following classes of functions

$$M_{\pm} := \{\phi : \text{analytic in} H_{\pm}\text{ , continuous in } H_{\pm} + D, \ |\phi(t)| \le 1\} \tag{2.3}$$

$$N_{\pm} := \{\phi \in M_{\pm} : \ |\phi(z)| \le 1 \text{ in } H_{\pm}\} \tag{2.4}$$

We first prove in the next subsection the following results for univariate analytic characteristic functions (a.c.f.s).

2.2.1 Analytic characteristic functions

Let f stand for the c.f. of the d.f. F. Introducing the functions

$$f_{-}(z) := \int_{-\infty}^{-0} e^{izu} \mathrm{d}F(u) \in N_{-}, \ f_{+}(z) := \int_{-0}^{\infty} e^{izu} \mathrm{d}F(u) \in N_{+} \tag{2.5}$$

we get the decomposition

$$f(t) = f_{-}(t) + f_{+}(t), \quad t \in \mathbb{R}^1 \tag{2.6}$$

If $F(0) = 0$, then $f = f_{+}$. Since F is monotone the following lemma from the theory of Laplace-Stieltjes transforms plays a considerable role.

Lemma 2.2.1 *The line of convergence of the integrals defining $f_\pm$ has a singularity on the imaginary axis. (See [86]).*

Thus a c.f. f is said to be analytic (a.c.f.), if its restriction to some interval $|t| < r$ is continuable into the circle about 0 with radius r. In view of lemma (2.2.1) and equation (2.6) this case occurs if and only if 0 is no singularity for $f_\pm$. Hence we immediately obtain

Theorem 2.2.1 *To every a.c.f. f there corresponds a strip S_{ab}, $0 < a$, $b \leq \infty$, in which it is analytic; the points ib and -ia, if they are finite - are singularities of f. In S_{ab} we have the representation*

$$f(z) = \int_{-\infty}^{\infty} e^{izu} dF(u) \tag{2.7}$$

See [86]. Moreover, such a representation is unique.

Corollary 2.2.1 *The c.f. f is analytic if and only if for some $r > 0$ as $x \rightarrow \infty$*

$$T_F(x) := 1 - F(x) + F(-x) = O(e^{-rx}) \tag{2.8}$$

The proof uses the fact that the intégral (2.7) exists for $z = iy$, $y < r$, if and only if (2.8) is satisfied. (See [86]).

2.2.2 Laplace transforms and a.c.f.s

Let $F(x)$ be a probability distribution function. We call F to be Laplace-transformable or LT (for short), in the strip $[u, v]$, $u < v$, $u, v \in I\!R^1$, if the integral

$$\int_{-\infty}^{\infty} e^{yx} \mathrm{d}F(x) < \infty \ \forall \ y \in [u, v] \tag{2.9}$$

Theorem 2.2.2 *F is LT in the strip $[-b, a]$, b, a real and finite, if and only if f is a.c.f. in S_{ab}.*

Proof: If F is LT in $[-b, a]$ then

$$\int_{-\infty}^{\infty} e^{yx} \mathrm{d}F(x) < \infty, \quad \forall \ y \in [-b, a]$$

Let $z \in S_{ab}$, then -Im $z \in [-b, a]$, and

$$|e^{izu}\,\mathrm{d}F(u)| \leq e^{-\mathrm{Im}\ zu}\mathrm{d}F(u)$$

Therefore

$$|f(z)| \leq \int_{-\infty}^{0} e^{-\mathrm{Im}\ z.u}\mathrm{d}F(u) + \int_{0}^{\infty} e^{-\mathrm{Im}\ z.u}\mathrm{d}F(u)$$

$$\leq \int_{-\infty}^{\infty} e^{-\mathrm{Im}\ z.u}\mathrm{d}F(u) < \infty$$

Therefore

$$f(z) = \int_{-\infty}^{\infty} e^{izu}\mathrm{d}F(u)$$

exists in the strip S_{ab}.

To prove analyticity we proceed as follows :

Since $f(z)$ exists in the strip S_{ab}, letting $z = x + iy$, $x, y \in I\!R$ we have

$$f(z) = \int_{-\infty}^{\infty} e^{i(x+iy)w}\mathrm{d}F(w)$$

$$= \int_{-\infty}^{\infty} e^{-yw}\cos(xw)\mathrm{d}F(w) + i\int_{-\infty}^{\infty} e^{-yw}\sin(xw)\mathrm{d}F(w)$$

$$= R + iI$$

(say). Consider formally the integrals

$$R_x = I_y = \int_{-\infty}^{\infty} e^{-yw}(-\sin(xw))\,w\mathrm{d}F(w)$$

$$R_y = -I_x = \int_{-\infty}^{\infty} e^{-yw}(-\cos(xw))\,w\mathrm{d}F(w)$$

Let δ be such that $0 < \delta < \min(a, b)$. Then

$$\delta\,|w| < e^{\mathrm{sgn}(w)\,\delta w}, \quad w \in I\!R$$

Let $y \in A \subset [-a, b] \ni y + \delta,\ y - \delta \in [-a, b]$ .Then

$$|R_x| = |I_y| \leq \int_{-\infty}^{\infty} e^{-yw}\,|w|\mathrm{d}F(w)$$

$$\leq \frac{1}{\delta}\left(\int_{-\infty}^{0} e^{-(y+\delta)w}\,\mathrm{d}F(w) + \int_{0}^{\infty} e^{-(y-\delta)w}\,\mathrm{d}F(w)\right) < \infty$$

Similarly, $|R_y| = |I_x| < \infty$. Note that since the four integrals exist, and are actually partial derivatives of R, I when the derivatives can be taken under the integral sign (which can be done as the above four integrals exist), we have

$$R_x = I_y,\ R_y = -I_x \ \forall\ (x + iy) \in S_{ab}$$

which implies $f(z)$ is analytic throughout S_{ab}.

If f is a.c.f in S_{ab}, then in S_{ab}, by (2.7) we have the representation

$$f(z) = \int_{-\infty}^{\infty} e^{izu} dF(u)$$

Setting Re $z = 0$, we have

$$f(iy) = \int_{-\infty}^{\infty} e^{-yu} dF(u) < \infty \ \forall\ y \text{ such that } -a < y < b$$

This implies that F is LT in the strip $[-b\,,\, a]$.

Corollary 2.2.2 *The class of non-Laplace transformable distributions is characterized by the class $\mathcal{N}$ where*

$$\mathcal{N} := \{ f \text{ is a c.f., } 0 \text{ is a singularity for } f(z)\,,\ z \text{ complex} \} \qquad (2.10)$$

Theorem 2.2.3 *Let g, f stand for the c.f.s of the d.f.s G, F respectively. Let F be LT in $[-b, a]$, and hence f be a.c.f. in S_{ab}, Let G, F be related as*

$$dG(x) \propto e^{\beta x}\, dF(x) \quad - i\beta \in S_{ab} \qquad (2.11)$$

Then g is a.c.f and is given by

$$g(t) = \frac{f(t - i\beta)}{f(-i\beta)} \qquad (2.12)$$

Proof : Since f is a.c.f., $-i\beta \in S_{ab}$, $f(-i\beta)$ exists and equals

$$\int_{-\infty}^{\infty} e^{\beta u} dF(u)$$

and therefore takes positive real values. Now,

$$dG(u) = c.e^{\beta u}dF(u), \quad c \in \mathbb{R}^1$$

and G a distribution function

$$\Rightarrow \int_{-\infty}^{\infty} dG(u) = c. \int_{-\infty}^{\infty} e^{\beta u}dF(u) = 1$$

$$\Rightarrow c = \frac{1}{\int_{-\infty}^{\infty} e^{\beta u}dF(u)} = \frac{1}{f(-i\beta)}$$

Therefore we have the representation

$$dG(u) = \frac{e^{\beta u}dF(u)}{f(-i\beta)}$$

Let $z_0 = t - i\beta$, $t \in \mathbb{R}^1$. Then $z_0 \in S_{ab}$. Therefore the integral

$$\int_{-\infty}^{\infty} e^{iz_0 u}dF(u) = \int_{-\infty}^{\infty} e^{i(t - i\beta)u}dF(u)$$

exists. Therefore the c.f. g has the representation

$$g(t) = \int_{-\infty}^{\infty} e^{itu}dG(u) = \frac{\int_{-\infty}^{\infty} e^{(it+\beta)u}dF(u)}{f(-i\beta)}$$

$$= \frac{\int_{-\infty}^{\infty} e^{iz_0 u}dF(u)}{f(-i\beta)} = \frac{f(t - i\beta)}{f(-i\beta)}.$$

Note that g is analytic. (See Ramachandran [81]).

2.3 Analytic characteristic functions in L^2

For the following related results see Akhiezer, [1988], ([1]).

Definition 2.3.1 *The function $F(z)$ $(z = x + iy)$ belongs to the Hardy class H_2^+ if it is analytic in a half plane $y > 0$ and satisfies the inequality*

$$M^2 \stackrel{def}{=} \sup_{y>0} \int_{-\infty}^{\infty} |F(x+iy)|^2 dx < \infty \tag{2.13}$$

Definition 2.3.2 *Denote by* $H_2(\alpha,\beta)$ $(-\infty < \alpha < \beta < \infty)$ *the class of functions* $F(z)$ $(z = x + iy)$ *analytic in the strip* $\alpha < y < \beta$ *and satisfying the inequality*

$$\sup_{\alpha<y<\beta} \int_{-\infty}^{\infty} |F(x+iy)|^2 dx < \infty \tag{2.14}$$

Theorem 2.3.1 *Every function* $F(z)$ *of the form*

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{izt} f(t)dt \tag{2.15}$$

where $f \in L^2(0,\infty)$, *belongs to* H_2^+.

Theorem 2.3.2 *For every function* $F(z) \in H_2^+$ *there exists a function* $f \in L^2(0,\infty)$ *such that for any* $y > 0$

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{izt} f(t)dt \quad (z = x + iy) \tag{2.16}$$

Theorem 2.3.3 *For every function* $F(z) \in H_2^+$ *there exists a function* $f \in L^2(0,\infty)$ *such that for any* $y > 0$ *the Laplace transform*

$$\Phi(s) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-st} \phi(t)dt \quad (s = \sigma + i\tau) \tag{2.17}$$

where $\sigma > c$ *and* $e^{-ct}\phi(t) \in L^2(0,\infty)$, *is analytic in the plane* $\sigma > c$, *and* $\Phi(s) \in H_2(c,\infty)$.

Definition 2.3.3 *An entire function* $g(z)$ *is called an* entire function of exponential type $\le A$ *if for any* $\alpha > 0$ *there exists a constant* $B = B(\alpha)$ *such that for any* z

$$|g(z)| \le Be^{(A+\alpha)|z|} \tag{2.18}$$

Theorem 2.3.4 *(Paly-Weiner) The class of entire functions* $g(z)$ *of exponential type* $\le A$ *for which*

$$\int_{-\infty}^{\infty} |g(x)|^2 dx < \infty \tag{2.19}$$

coincides with the class of functions representable in the form

$$g(z) = \frac{1}{\sqrt{2\pi}} \int_A^{A} e^{izt} \phi(t)dt, \quad (z = x + iy) \tag{2.20}$$

where $\phi(t) \in L^2(-A, A)$.

2.4 Conditions guaranteeing analyticity of c.f.s

In order to apply theorems (2.2.2) and (2.2.3), we need to check for the analyticity of the given c.f. f. This can be approached in two different ways :

(a) We can directly verify the analyticity of $f(z)$, z complex, by either verifying the Cauchy-Riemann conditions (for analyticity of a function on the complex domain or tube) on $f(z)$ or checking if (2.8) holds (so that by corollary (2.2.1) $f(z)$ is analytic).

(b) We can verify whether F is LT. Since, if F is LT, then by theorem (2.2.2) $f(z)$ will be analytic in an appropriate domain.

We next present a series of sufficient (and some necessary and sufficient conditions) for F to be LT.

For the sake of completeness we state here the Cauchy-Riemann conditions for analytic functions on the complex domain (a special case of the multivariate holomorphic case).

A function $f(z) = u(x, y) + i v(x, y)$ is differentiable (analytic) at a point $z = x + iy$ if and only if,

(i) both u and v are differentiable at point (x, y) ;

(ii) at point (x, y) the partial derivatives of u and v satisfy the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

2.4.1 Further useful results in a.c.f. theory

For details about the following results see [86].

Result 2.4.1 *Every a.c.f f has a domain of regularity which is symmetric with respect to the imaginary axis; in this domain we have* $f(z) = \overline{f(-z)}$.

Result 2.4.2 *An a.c.f f possesses the ridge property*

$$f(iy) = \max_{-\infty<t<\infty} |f(t + iy)|, \quad t \in I\!R^1 \tag{2.21}$$

From the ridge property we obtain by the maximum modulus principle

Result 2.4.3 *For an a.c.f. f we have for $0 < r < \min(a, b)$*

$$\max_{|z|=r} |f(z)| = \max(f(ir), f(-ir)) \tag{2.22}$$

The a.c.f. f is integral if $a = b = \infty$. This is the case if and only if (2.8) holds for all $r > 0$.

Result 2.4.4 *An integral function of finite order $\varrho > 2$ whose exponent of convergence is less than ϱ cannot be a c.f..*

Let now P_m stand for a polynomial of degree m.

Result 2.4.5 *(Marcinkiewicz) The function* $(\exp(P_m), m > 2)$ *cannot be a c.f..*

An a.c.f. f belongs to the class $\mathcal{D}$ (Dantzig), if

$$\phi(t) := f^{-1}(it), \ |t| < \min(a, b) \tag{2.23}$$

defines another a.c.f.. Note that an a.c.f. f given on a finite interval is uniquely defined on $\mathbb{R}^1$.

It happens that f_- is analytic on a circle K about 0, but f_+ has not this property. Then the c.f. clearly exists in a strip S_{0b} $(0 < b \leq \infty)$. In this case f is called a *boundary* c.f.. Thus many results valid for a.c.f. carry over to boundary c.f. and c.f. of non-negative random variables are either a.c.f. or boundary c.f..

Now consider functions f analytic in S_{ab} (or S_{0b}) satisfying result 2.4.2; they are called *ridge functions*. A.c.f and boundary c.f. form proper partial sets of this class. For integral ridge functions satisfying $f(0) = 1$, the function $f(iy)$, $y \in \mathbb{R}^1$, proves real.

Result 2.4.6 *An entire ridge function f of finite order having only real zeros permits the representation*

$$f(z) = c\exp(-dz^2 + ibz)\prod_k (1 - \frac{z^2}{a_k^2}) \tag{2.24}$$

where $a_k > 0$, $\sum_k a_k^{-2} < \infty$, $Im\, b = 0$, $d \geq 0$, $c \neq 0$ real. This means that f has at most the order 2.

Result 2.4.7 *Let f stand for a ridge function of finite order ϱ defined in H_+; assume that the exponent of convergence for the zeros contained in $\{z : y \geq a > 0\}$ is $\varrho_1 < \varrho$. Then $\varrho \leq 3$.*

2.4.2 Sufficient conditions for Laplace transformability

A sufficient set of conditions for a function $F(s)$ to be a Laplace transform is given by the following theorem :

Result 2.4.8 *Let $F(s)$ be analytic over the half-plane $Re\ s \geq a$, and satisfies the inequality,*

$$|F(s)| \leq \frac{C}{|s|^2} \tag{2.25}$$

where C is a constant. If the integral $\mathcal{L}^{-1}F(s)$ is taken over some vertical line in the half-plane $Re\ s \geq a$, then $\mathcal{L}^{-1}F(s) = f(t)$ exists and is a continuous function for all t. Moreover, $f(t) = 0$ for $t < 0$ and $\mathcal{L}f(t) = F(s)$ at least for $Re\ s > a$.

Result 2.4.9 *Let $f(s)$ be analytic in the strip $\alpha < \sigma < \beta$ and such that*

$$\int_{-\infty}^{\infty} | f(\sigma + i\tau)| \, d\tau < \infty \tag{2.26}$$

Let

$$\lim_{|\tau|\to\infty} f(\sigma + i\tau) = 0 \tag{2.27}$$

uniformly in every closed subinterval of $(\alpha < \sigma < \beta)$, and set

$$\phi(x) = \frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} f(s)\, e^{xs}\, ds \quad (\alpha < \sigma < \beta, -\infty < x \infty) \tag{2.28}$$

Then

$$f(s) = \int_{-\infty}^{\infty} e^{-sx}\, \phi(x)\, dx \quad (\alpha < \sigma < \beta)$$

Definition 2.4.1 *A real function $k(x, y)$ which is continuous in the square $(a \leq x \leq b, a \leq y \leq b)$ is of positive type there if for every real function$\phi(x)$ continuous in $(a \leq x \leq b)$*

$$J(\phi) = \int_a^b\int_a^b k(x, y)\, \phi(x)\, \phi(y)\, dx dy \geq 0 \tag{2.29}$$

Result 2.4.10 *A continuous kernel $k(x, y)$ is of positive type if and only if for every finite sequence $\{x_i\}_0^n$ of distinct numbers of $(a \leq x \leq b)$ the quadratic form*

$$Q_n = \sum_{i=0}^{n}\sum_{i=0}^{n} k(x_i, y_j)\, \xi_i\, \xi_j \tag{2.30}$$

is positive (definite or semidefinite).

Result 2.4.11 *A necessary and sufficient condition that the function $f(x)$ can be represented in the form*

$$f(s) = \int_0^\infty e^{-st}\, d\alpha(t)$$

where $\alpha(t)$ is non-decreasing and the integral converges for $a < x < b$. is that $f(x)$ should be analytic there and that the kernel $f(x + y)$ should be of positive type in the square $(a < 2x < b,\ a < 2y < b)$. See [101, p 265-275].

2.5 Examples and Applications

2.5.1 Univariate examples

Example 2.5.1 *Normal distribution.*

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad |x|, |\mu| < \infty, \sigma \in R^+$$

The corresponding c.f. is

$$\phi_1(t) = \exp(it\mu - \frac{(\sigma t)^2}{2})$$

We denote the c.f. for $f_2(x)$ by $\phi_2(t)$. Therefore

$$\phi_2(t) = \exp(i(t - i\beta)\mu - \frac{(\sigma(t - i\beta))^2}{2}) / \exp(i(-i\beta)\mu - \frac{(\sigma(-i\beta))^2}{2})$$

$$= \exp(it(\sigma^2\beta + \mu) - \frac{(\sigma t)^2}{2}$$

Here there is no need for checking as the resulting c.f. is of the same form as the original c.f. From straightforward comparison we have that $f_2(x)$ is a normal density with mean $\mu' = (\sigma^2\beta + \mu)$ and same variance σ^2 . This can also be checked directly without going through c.f. s.

Example 2.5.2 *Polya Frequency functions*

Polya frequency of type II are IHR. The c.f. can be written as

$$\phi_1(t) = K.\frac{e^{-\gamma t^2 + i\delta t}}{\prod(1 - i\delta_j t)e^{i\delta_j t}}$$

The corresponding analytic extension is

$$\Psi(z) = K.\frac{e^{-\gamma z^2 + i\delta z}}{\prod(1 - i\delta_j z)e^{i\delta_j z}}$$

Therefore the corresponding transformed c.f. can be written as

$$\phi_1(t) = K.\frac{e^{-\gamma t^2 + i\delta t}}{\prod(1 - i\delta_j t)e^{i\delta_j t}}$$

$$\phi_2(t) = K.\frac{e^{-\gamma t^2 + it(\delta + 2\gamma\beta)}}{\prod(1 - i(\frac{\delta_j}{1-\delta_j\beta})t)e^{i\delta_j t}}$$

provided $1 - \delta_j\beta \neq 0 \ \forall \ j$. Therefore the corresponding c.f. $\phi_2(t)$ can be written as

$$\phi_2(t) = K.\frac{e^{-\gamma t^2 + it(\delta + 2\gamma\beta + \beta\sum\frac{\delta_j^2}{1-\beta\delta_j})}}{\prod(1 - i(\frac{\delta_j}{1-\delta_j\beta})t)e^{i\delta_j t}}$$

Therefore for ϕ_2 to be a valid c.f. the only condition to check is:

$$\sum_{j=1}^{\infty}\frac{\delta_j^2}{1 - \beta\delta_j} < \infty, \text{ given } \sum_{j=1}^{\infty}\delta_j^2 < \infty$$

Example 2.5.3 *Mixtures*

We first consider finite mixtures . Let ,

$$f(x) = \sum_{j=1}^{n} a_j f_j(x), \sum_{j=1}^{n} a_j = 1$$

Let the c.f. of $f_j(x)$ and $f(x)$ be $\Psi_j(t)$ and $\Psi(t)$ respectively and the corresponding analytic continuations be $V_j(z)$ and $V(z)$. Let the c.f. of the transformed density be $W(t)$. Then by the linearity of the Fourier transform we have

$$\Psi(t) = \sum_{j=1}^{n} a_j \Psi_j(t)$$

Also for the transformed c.f. $W(t)$ we have

$$W(t) = \sum_{j=1}^{n} a_j \Psi_j(t - i\beta) / \sum_{j=1}^{n} a_j \Psi(-i\beta)$$

or we have a new mixture density with mixing parameters

$$a_j' = \frac{a_j \Psi_j(-i\beta)}{\sum_{j=1}^{n} a_j \Psi_j(-i\beta)},$$

and component densities characterized by the component c.f.s

$$W_j(t) = \frac{\Psi_j(t - i\beta)}{\Psi_j(-i\beta)}$$

Note that the invariance of the new mixture is only dependent on the invariance properties of f_j's.

As an example, consider a normal mixture of n populations :

$$f(x) = \sum_{j=1}^{n} a_j N(\mu_j, \sigma_j^2)$$

$$\Phi_1(t) = \sum_{j=1}^{n} a_j \exp(it\mu_j - \frac{1}{2} t^2 \sigma_j^2)$$

$$\Phi_2(t) = \sum_{j=1}^{n} a_j' \exp(it(\mu_j + (\beta\sigma_j)^2) - \frac{1}{2} t^2 \sigma_j^2)$$

where

$$a_j' = \frac{a_j.\exp(\beta\mu_j + \frac{1}{2}(\beta\sigma_j)^2)}{\sum_{j=1}^{n} a_j \exp(\beta\mu_j + \frac{1}{2}\beta^2\sigma_j^2)}$$

Consider now the simplest case of mixture of two populations. Let , $f = a.f_1 + (1-a).f_2$ and let g be the corresponding transformed density. Let the corresponding parameter for a be c. Then

$$c = 1 + \frac{(1-a).U_2(-i\beta)}{a.U_1(-i\beta)}^{-1}$$

where U_1, U_2 are analytic continuations of Φ_1, Φ_2 Then

$$\frac{\partial c}{\partial a} = \frac{U_1.U_2}{(U_1 - U_2)^2} \cdot \frac{1}{(a + \frac{U_2}{(U_1 - U_2)^2})}$$

i.e., inverse of a quadratic in a and always nonnegative.

For Normal mixtures we find $\frac{\partial c}{\partial \beta}$.

$$\frac{\partial c}{\partial \beta} = \frac{\partial}{\partial \beta}(1 + (\frac{1}{a} - 1)e^{\beta(U_2 - U_1) + ({\sigma_2}^2 - {\sigma_1}^2)^2 \frac{\beta^2}{2}})^{-1}$$

$$= -((\frac{1}{a} - 1)e^{\beta(U_2 - U_1) + (\sigma_2^2 - \sigma_1^2)^2 \frac{\beta^2}{2}})((U_2 - U_1) + (\sigma_2^2 - \sigma_1^2)^2\beta)c^{-2}$$

Therefore $\frac{\partial c}{\partial \beta} > 0$ iff

$$(U_2 - U_1) + (\sigma_2^2 - \sigma_1^2)^2\beta < 0$$

or,

$$\beta < (U_1 - U_2)(\sigma_2^2 - \sigma_1^2)^{-2}$$

Example 2.5.4 *Gamma distribution*

$$f_1(x) = \frac{\alpha^p x^{p-1} e^{-\alpha x}}{\Gamma(p)}, x > 0$$

$$\phi_1(t) = \frac{\alpha^p}{(\alpha - it)^p}$$

$$\phi_2(t) = \frac{(\alpha - \beta)^p}{(\alpha - \beta - it)^p}$$

Note that for $\beta < \alpha$ the transformed c.f. is the c.f. of a gamma distribution.

Example 2.5.5 *Beta distribution*

$$f_1(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}$$

$$\phi_1(t) = \frac{\Gamma(p+q)}{\Gamma(p)} \sum_{j=0}^{\infty} \frac{(it)^j \Gamma(p+j)}{\Gamma(p+q+j)\Gamma(j+1)}$$

Here we get the transformed c.f. as a power series expansion which however is no longer a beta distribution. Without evaluating the coefficients explicitly in terms of β and gamma functions, we find out the normalizing constant for the new density of the functional form :

$$f_2(x) = \frac{e^{\beta x} x^{p-1}(1-x)^{q-1}}{c}, \quad 0 < x < 1$$

where

$$c = \sum_{j=0}^{\infty} \frac{\beta^j \Gamma(p+j)}{\Gamma(p+q+j)\Gamma(j+1)}$$

Example 2.5.6 *Inverse Gaussian*

The c.f. is given by

$$\exp(\frac{\lambda}{\mu}(1 - (1 - \frac{2i\mu^2 t}{\lambda})^{1/2}))$$

Hence, the corresponding complex extension is analytic (entire).

$$f_1(x; d, \beta, \nu) = \frac{1}{\sqrt{2\pi\beta x^3}}(d. \exp(-\frac{(d - vx)^2}{2\beta x}), x > 0$$

Then $f_2(x) = f_1(x; d, \beta, \sqrt{v^2 - \beta})$. The importance of this distribution in reliability is due to its similarity to Hadwiger's reproduction rate function.

Example 2.5.7 *Location family*

The location family is characterized by the functional relationship $F(Y) = F(X-a)$ where F(.) denotes a distribution function , $a \in R^1$. In terms of random variables this is equivalent to $Y = X - a$ a.e. . Let the c.f. of X be g(t) , $t \in R^1$. Then $E(\exp(itY)) = E(\exp(itX - ita)) = \exp(-ita).E(\exp(itX))$. Therefore any member of the family will have c.f. of the form $h(t) = \exp(-ita).g(t)$. From (2.2.3) we have that the general form of the transformed c.f. for $f_2(x)$ is

$$\exp(-ita).\frac{g(t-i\beta)}{g(-i\beta)}$$

subject to the condition that $g(t)$ is an a.c.f. Thus if the density corresponding to $g(t)$ is LT the resultant transformed class is also a location family. If $g(t)$ belongs to some invariance class then the new location class is also invariant w.r.t. the invariance class.

Example 2.5.8 *Proper infinitely divisible class*

Starting with the Kolmogorov representation ; ie, if $v(t)$ is the c.f. of a proper infinitely divisible distribution having finite second moments then

$$\log(v(t)) = iat + \int \frac{e^{itu} - 1 - itu}{u}.d(K(u))$$

where $a \in R$, and $K(u)$ is a distribution function unique up to a numerical constant. Using (2.2.3) we have the transformed c.f. as

$$\log(g(t)) = iat + \int \frac{e^{itu}.e^{u\beta} - itu - e^{u\beta}}{u}.d(K(u))$$

$$\log(g(t)) = iat + \int \frac{e^{itu} - 1 - itu}{u}.e^{u\beta}d(K(u)) + it\int e^{u\beta}d(K(u))$$

Note therefore that if the integral

$$c(\beta) = \int e^{u\beta}.d(K(u))$$

exists, i.e., the distribution characterized by $K(u)$ is LT then $g(t)$ remains the c.f. of a i.d. distribution and the resultant log c.f. is :

$$\log(g(t)) = i(a + c(\beta))t + \int \frac{e^{itu} - 1 - itu}{u}.d(M(u))$$

If K, M belong to some invariance class then g, v both belong to this same invariance class.

For the Levi-Khintchine representation we have :

$$\ln \phi_1(t) = ibt + \int (e^{itu} - 1 - \frac{itu}{1+u^2})\frac{1+u^2}{u^2} d(G(u))$$

The transformed c.f. is

$$\ln \phi_2 = ibt + \int (e^{itu}.e^{\beta u} - e^{\beta u} - \frac{itu}{1+u^2})\frac{1+u^2}{u^2} d(G(u))$$

The right hand side can be written as $ibt + I_1 + I_2$ where,

$$I_1 = \int (e^{itu} - 1 - \frac{itu}{1+u^2})\frac{1+u^2}{u^2}.e^{\beta u} d(G(u))$$

Note that I_1 will exist if $G(u)$ is LT .

$$I_2 = it \int \frac{e^{\beta u} - 1}{u} d(G(u))$$

Let $h(u) = (\exp(\beta u) - 1)/u$. Then

$$|h(u)| \ \leq \ e^{|\beta|} \ - 1 \ \geq \ 0 \ \ \forall |u| \leq 1$$

$$|h(u)| \ < \ e^{\beta u} \ + 1 \ \ \forall |u| > 1$$

Therefore a sufficient condition for the integral to exist is that G be LT.

Example 2.5.9 *Stable distributions*

The c.f. of non-degenerate members of this class can be written in the form, (See Zolotarev,[103]) :

$$\ln g(t) = \lambda(it\gamma - |t|^{\alpha}\omega_B(t, \alpha, \beta))$$

$$\omega_B(t, \alpha, \beta) = \begin{cases} \exp(-i\frac{\pi}{2}\beta K(\alpha) sgn(t)) & \text{if } \alpha \neq 1 \\ \frac{\pi}{2} + i\beta \ln |t| sgn(t) & \text{if } \alpha = 1 \end{cases}$$

$$0 < \alpha \leq 2 \ , \ |\beta| \leq 1 \ , \ |\gamma| < \infty \ , \ \lambda > 0, \ K(\alpha) = \alpha - 1 + sgn(1 - \alpha)$$

The two-sided Laplace transforms exist for $\beta = 1$, $\alpha < 2$. In the half-plane Re s ≥ 0, the log- Laplace transform can be written as:

$$\ln \Lambda(s, \alpha, 1, \gamma, \lambda) = -\lambda(s\gamma - \epsilon(\alpha)s^{\alpha}) \ if \ \alpha \neq 1$$

$$= \lambda(-s\gamma + s \ln s) \ if \ \alpha = 1$$

where, $\epsilon(\alpha) = sgn(1 - \alpha)$. Therefore the transformed c.f.s can be written as :

$$\ln \phi_2(t) = -\lambda(it\gamma + \epsilon(\alpha)|b|^{\alpha}(\left(1 - \frac{t^2}{b^2}\right)^{\frac{\alpha}{2}} e^{i\alpha \tan^{-1}\frac{t}{b}} - 1) \ if \ \alpha \neq 1$$

$$= \lambda(-it\gamma + \ln\sqrt{1 + \frac{t^2}{b^2}}(1 - it\ b) + (ib - t)\tan^{-1}\frac{t}{b})$$

Note that these are no longer stable.

Example 2.5.10 *Koopman-Darmois Exponential Families*

These distributions are characterized by the density with canonical parameter vector $\underline{\eta}$ and given by

$$p(x, \underline{\eta}) = \exp[\sum_{i=1}^{s} \eta_i T_i(x) - A(\underline{\eta})]h(x)$$

where $T_i(x)$ are real valued functions of x or in usual terminology the statistics.

The following two cases are possible:

(a) Let $T_i(x) \equiv x$ for at least one i.

Let w.l.o.g. $T_1(x) \equiv x$. Then if $(\eta_1 - \beta, \eta_2, \ldots, \eta_s)$ is an interior point of the valid region of the canonical parameter set then the transformed density exists and belongs to the same family. Therefore this case admits LGD.

(b) There exists no i for which $T_i(x) \equiv x$.

If the integral $\int \exp[\beta x + \sum \eta_i T_i]h(x)dx$ exists then the transformed density exists but belongs to a $s + 1$ dimensional exponential family. (See [58] .)

Example 2.5.11 *Circular distributions for Directional data*

We now take up a special class of distributions which violates the usual structure for distributions used for linear data as have been dealt with so far here. These distributions are periodic, have finite support $[0, 2\pi)$ and model data observed on a circle. These additional constraints of periodicity and same finite support $(0, 2\pi)$ are thus to be demanded of $f_2(\mathbf{x})$ given by (1.1) also, so that it qualifies at least as a circular distribution. We first explore to what extent the usual LGD satisfies such constraints. On one hand, this exposes the shortcomings of the use of the LGD directly, while on the other hand, reveals the possibilities of generalizations of the LGD as well as the need for new rules specific to directional data.

We first recall the following results for some common densities defined on the unit circle. For a detailed treatment see Mardia, [71].

a) The c.f. in these cases are of integer arguments, k, say.

b) let c.f. $g(k) = r(k) + ic(k)$ where $r(k) = Re(g(k))$ and $c(k) = Im(g(k))$

If the infinite series

$$\sum_{j=1}^{\infty}(r^2(j) + c^2(j))$$

is convergent, then $g(k)$ corresponds to a density with support $0 < \theta < 2\pi$ defined a.e. by

$$f(\theta) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} g(j)e^{-ij\theta}$$

$$= \frac{1}{2\pi}[1 + 2\sum_{j=1}^{\infty}(r(j).cos(j\theta) + c(j).sin(j\theta))]$$

i) Point distribution. $f_1(\theta) = P(\theta = \theta_0) = 1$. c.f.

$$\phi_1(k) = e^{ik\theta_0}$$

transformed c.f.

$$\phi_2(k) = e^{ik\theta_0}$$

ie identical to the original.

ii) Lattice distribution: For r,m, positive integers ,

$$P(\theta = \nu + \frac{2\pi r}{m}) = p_r, r = 0, \ldots, m-1$$

W.l.o.g. letting $\nu = 0$, we have the c.f. as

$$\phi_1(k) = \sum_{r=0}^{m-1} p_r.e^{\frac{2\pi irk}{m}}$$

$$\phi_2(k) = \sum_{r=0}^{m-1} p_r.e^{\frac{2\pi ir(k-i\beta)}{m}} / \sum_{r=0}^{m-1} p_r.e^{\frac{2\pi ir(-i\beta)}{m}}$$

i.e. ,

$$= \sum_{r=0}^{m-1} p_r^{'}.e^{\frac{2\pi irk}{m}}$$

This is again lattice with $p_k^{'} =$

$$\frac{p_k.e^{\frac{2\pi k\beta}{m}}}{\sum_{r=0}^{m-1} p_r.e^{\frac{2\pi r\beta}{m}}}$$

iii) Offset normal distribution.

Let $\phi(x, y; \underline{\mu}, \Sigma)$be the p.d.f. of the bivariate normal distribution with mean vector $\underline{\mu} = (\mu, \nu)$ and covariance matrix Σ. Let ρ denote the correlation between the variables and let σ_1^2 and σ_2^2 be their variances. Suppose that $\Phi(x)$ is the d.f. of $N(0,1)$. Substituting $x = r\cos\theta$,; $y = r\sin\theta$ we find that the p.d.f. of θ is given by

$$p(\theta; \mu, \nu, \sigma_1, \sigma_2, \rho) = (C(\theta))^{-1}(\phi(\mu, \nu; \underline{0}, \Sigma) + aD(\theta)\Phi(D(\theta))$$
$$\times\phi(a(C(\theta))^{-1}(\mu\sin\theta - \nu\cos\theta)))$$

where

$$a = (\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}$$
$$C(\theta) = a^2(\sigma_2^2\cos^2\theta - \rho\sigma_1\sigma_2\sin 2\theta + \sigma_1^2\sin^2\theta)$$
$$D(\theta) = a^2\sqrt{C(\theta)}(\mu\sigma_2(\sigma_2\cos\theta - \rho\sigma_1\sin\theta) + \nu\sigma_1(\sigma_1\sin\theta - \rho\sigma_2\cos\theta))$$

and $\phi(x)$ is the p.d.f. of $N(0,1)$. Therefore the transformed density is proportional to $e^{\beta\theta}p(\theta; \mu, \nu, \sigma_1, \Sigma_2, \rho)$.

iv) Uniform on the unit circle. $f_1(\theta) = \frac{1}{2\pi}$

$$\phi_1(k) = \frac{e^{2\pi ik} - 1}{2\pi ik}$$

$$\phi_2(k) = \frac{e^{2\pi ik+2\pi\beta} - 1}{\beta + ik} \frac{\beta}{e^{2\pi\beta} - 1}$$

$$f_2(\theta) = \frac{\beta e^{\beta\theta}}{e^{2\pi\beta} - 1}$$

v) Cardioid .

$$f_1(x) = \frac{1}{2\pi}.(1 + 2r\cos(x - x_0)), 0 < x < 2\pi, |r| < 1$$

$$f_2(x) = \frac{\beta e^{\beta x}}{e^{2\pi\beta} - 1}.\frac{1 + 2r\cos(x - x_0)}{1 + 2r\cos(a)\cos(a + x_0)}$$

Note that this has a discontinuity at 0.

vi) Triangular .

$$f_1(x) = \frac{1}{8\pi}.(4 - \pi^2 r + 2\pi r|\pi - x|), 0 < x < 2\pi, r < \frac{4}{\pi^2}$$

$$f_2(x) = \frac{\beta^2 e^{\beta x}}{e^{\pi\beta} - 1}.\frac{(4 - \pi^2 r + 2\pi r|\pi - x|)}{e^{\pi\beta}(4\beta + \pi^2 r\beta - 2\pi r) + (4\beta + \pi^2 r\beta + 2\pi r)}$$

vii) Wrapped Poisson:

$$\phi_1(k) = \exp(\lambda.e^{\frac{2\pi ik}{m}} - \lambda)$$

then ,

$$\phi_2(k) = \exp(\lambda.e^{\frac{2\pi\beta}{m}}(\lambda.e^{\frac{2\pi ik}{m}} - 1))$$

ie the new density is again a wrapped Poisson with parameter

$$\lambda' = \lambda.e^{\frac{2\pi\beta}{m}}$$

Remarks. Observe that (i), (ii) and (vi) admit LGD while the others do not. Further, it is obvious that the most commonly used circular distribution, the von Mises or Circular normal distribution, also does not admit LGD, though it is a member of the two parameter regular exponential family. However, interestingly we note in passing that generalizations or rather transformations of data can make observations from a class of circular distributions amenable to LGD. The unique properties of the circular distributions coupled with the above interesting observation, motivated us to study the classification problem for directional data in slightly more details in a later chapter.

Example 2.5.12 *Generalized Pearsonian system*

Consider the system characterized by the differential equation,

$$\frac{d\ln f_1(x)}{dx} = \frac{P(x)}{Q(x)}$$

The transformed system can be characterized by the system

$$\frac{d\ln f_2(x)}{dx} = \frac{P(x) + \beta Q(x)}{Q(x)}$$

If $\deg(Q) \leq \deg(P)$ then g is in the system. If $\deg(Q) > \deg(P)$ then g is not in the system. Note that in this case the support of the original density has to be finite. However there is no change of support on transformation.

In the classical univariate setup the transformed density does not admit LGD, ie., goes out of the system for all types except Type III .

Example 2.5.13 *Rectangular*

The density is given by

$$f_1(\theta) = \frac{1}{c}, \ a < \theta < a + c$$

.

$$\phi_1(t) = \exp(ita).\frac{e^{itc} - 1}{itc}$$

$$\phi_2(t) = \exp(ita).\frac{e^{itc+c\beta} - 1}{\exp(c\beta) - 1}\frac{\beta}{\beta + it}$$

This gives us a truncated version of an exponential density.

Example 2.5.14 *Triangular distribution*

The density is given by

$$f_1(x) = \frac{1}{a}(1 - \frac{|x|}{a}), a > 0, |x| < a$$

$$\text{c.f.}\phi_1(t) = 2\frac{1 - \cos(at)}{a^2t^2} = (\frac{\exp(\frac{1}{2}iat) + \exp(-\frac{1}{2}iat)}{\frac{1}{2}at})^2$$

Transformed c.f.

$$\phi_2(t) = \left(\frac{\exp(\frac{1}{2}(iat + a\beta)) + \exp(-\frac{1}{2}(iat + a\beta))}{\exp(\frac{1}{2}a\beta) + \exp(-\frac{1}{2}a\beta)}\frac{\beta}{\beta + it}\right)^2$$

2.5.2 Some consequences of the characterization

We observe the following two important consequences of the formal relationship arising from theorem (2.2.3), i.e.,

(1) Let c_1, c_2 be the normalizing factors of the densities f_1, f_2, and $\psi(z)$ be the complex extension for the corresponding c.f., then $c_2/c_1 = \Psi(-i\beta)$ or equivalently

$$f_2(x) = e^{\beta x}.f_1(x)/\Psi(-i\beta)$$

Proof: Set $t = 0$ in the complex integral $\Psi(t - i\beta)$. This gives us a simple way of evaluating the normalizing constant of the new density or straightaway writing the form of the new density without going through the c.f. of the new density.

(2) The class of non L-transformable densities . Consider as an example the standard Cauchy density for f_1 , i.e. ,

$$f_1(x) = \frac{1}{\pi(1+x^2)}, x \in R^1$$

The corresponding c.f. $\Phi_1(t) = \exp(-|t|)$, is not analytic at 0. By corollary (2.2.2) all c.f.s having 0 as a point of singularity are non-LT.

(3) An important consequence of the above characterization is that the actual c.f.s of distributions admitting LGD would be analytic at the origin, and hence "smooth" in appearance, in graphical representations. This can serve as graphical tool in identifying whether a given set of data is appropriate for LGD or not. We have plotted the e.c.f s (empirical characteristic functions) of distributions, namely standard normal and Cauchy, in a small region about the origin, to show that normal indeed shows "smoother" or "flatter" surfaces for the generalized Laplace transforms (i.e., with complex arguments) than Cauchy (Figures 1 to 2).

Chapter 3

The multivariate family admitting LGD

We extend the results in the one-parameter case to the multiparameter, multivariate case to characterize families that admit LGD of Cox [27].

3.1 Introduction

We take up the problem of characterizing multiparameter, multivariate families that admit LGD of Cox [27]. The results here provide extensions to those in the chapter on the univariate case. However, the problem now becomes both statistically and mathematically more interesting. Further, our results here are also useful for practitioners in applied multivariate work and this is illustrated by several standard multivariate multiparameter examples given at the end. Some of these ideas were expounded by us earlier in ([91], [94]).

3.2 Holomorphic functions of several complex variables

Define

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - \mathrm{i}\frac{\partial f}{\partial y}\right) \quad \frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + \mathrm{i}\frac{\partial f}{\partial y}\right)$$

Let G be a domain in the space $\mathbb{C}^n$ of n complex variables $z_1, \ldots, z_n$. A continuous (complex-valued) function defined in G is called holomorphic, if it depends holomorphically on every separate variable z_j.

The assumption of continuity is unnecessary, because by Hartogs Theorem every separately holomorphic function must be continuous.

Holomorphic functions Φ in $\mathbb{C}^n$ fulfill the Cauchy-Riemann system

$$\frac{\partial \Phi}{\partial \bar{z}_j} = 0, \qquad j = 1, \ldots, n \tag{3.1}$$

In the case $n \geq 2$ this system is overdetermined.

3.3 Holomorphic multivariate c.f.s

With the complex vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ we define the scalar product

$$(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} x_j \overline{y_j}, \; |\mathbf{x}| = (\mathbf{x}, \mathbf{x})$$

and the vector $\mathrm{Re}\,\mathbf{y} = (\mathrm{Re}\,y_1, \ldots, \mathrm{Re}\,y_n)$, $\mathrm{Im}\,\mathbf{y} = (\mathrm{Im}\,y_1, \ldots, \mathrm{Im}\,y_n)$. The Fourier-Stieltjes transform of an n-variate probability law p

$$f(t) = \int_{\mathbb{R}^n} e^{\mathrm{i}(t,\mathbf{x})} p(\mathrm{d}\mathbf{x}), \quad t \in \mathbb{R}^n \tag{3.2}$$

is called the c.f. of p. For f, the uniqueness theorem, the convolution theorem, and the continuity theorem can be derived.

Definition 3.3.1 *Let G be a region in $\mathbb{C}^n$ containing the point $\mathbf{t} = 0$. A c.f. $\phi(\mathbf{t}, p)$ is said to be analytic in the region G if there exists a function $f(\mathbf{t})$, analytic in G, such that $\phi(\mathbf{t}, p) = f(\mathbf{t})$ for $\mathbf{t} \in G \cap \mathbb{R}^n$.*

The analytical properties of n-variate c.f. are determined to a great extent by the analytical properties of projections. Let $\mathbf{e} \in \mathbb{R}^n$, $|\mathbf{e}| = 1$ and $B \subset \mathbb{R}$ an arbitrary Borel set. Then the univariate law corresponding to the n-variate random variable X is

$$p_{\mathbf{e}}(B) = \mathrm{P}((X, \mathbf{e}) \in B) \tag{3.3}$$

is called the projection of X on the vector $\mathbf{e}$. Obviously, denoting by f and $f_{\mathbf{e}}$ the c.f. of p and $p_{\mathbf{e}}$ respectively, we have

$$f_{\mathbf{e}}(t) = f(t\mathbf{e}), \quad t \in I\!R \tag{3.4}$$

The univariate c.f. $f_{\mathbf{e}}$ is called the projection of f on the vector $\mathbf{e}$. In many cases we can restrict ourselves to the study of projections. It is to be noted that an n-variate d.f. F $(n > 1)$ need not be determined by the projections on n fixed linearly independent vectors, but the projections on all n-dimensional vectors determine F uniquely.

For the study of a.c.f we need the space $\mathbb{C}^n$ of all complex vectors $\mathbf{z} = (z_1, \ldots, z_n)$. Let $G \subset I\!R^n$ be a convex open set. Then we say that

$$(\mathbf{z} \in \mathbb{C}^n \mid \operatorname{Im} \mathbf{z} \in G) \tag{3.5}$$

is the convex tube with basis G.

Theorem 3.3.1 *Let f be an a.c.f.. Then it is analytic on a convex tube and admits the representation*

$$f(\mathbf{z}) = \int_{I\!R^n - \{0\}} e^{i(\mathbf{z},\mathbf{x})} d(p(\mathbf{x})) \tag{3.6}$$

on this tube. G is the interior of the set

$$(Im\,\mathbf{z} \mid \int_{I\!R^n - \{0\}} e^{i(\mathbf{z},\mathbf{x})} dp(\mathbf{x})) < \infty)$$

If γ is a point of the boundary of G, then $i\gamma$ is a singular point of f.

From Linnik and Ostrovsky [66] we have

Theorem 3.3.2 *Let $\mathbf{e}_1, \ldots, \mathbf{e}_n \in I\!R^n$ be n- linear independent vectors and p an n-variate probability law with c.f. f. Assume that the projections $f_{\mathbf{e}_1}, \ldots, f_{\mathbf{e}_n}$ of f are analytic in the strips $a_j < Im\,t < b_j\,(a_j < 0 < b_j\,, j = 1, \ldots, n)$, respectively. Then f is an a.c.f. whose basis is the interior of the convex hull of the vectors $a_i\mathbf{e}_i$ and $b_i\mathbf{e}_1$, $(i = 1, \ldots, n)$.*

Moreover, the integral in the unique representation (3.6) is uniformly and absolutely convergent on every compact subset of G.

Corollary 3.3.1 *Let $f(\mathbf{z}) = f(z_1, \ldots, z_n)$ be the c.f. of an n-variate distribution p. If the functions*

$$f(z_1, 0, \ldots, 0),\ f(0, z_2, \ldots, 0), \ldots, f(0, 0, \ldots, z_n)$$

are analytic in the respective regions

$$|Im\, z_1| < r_1, |Im\, z_2| < r_2, \ldots, |Im\, z_n| < r_n$$

then $f(\mathbf{z})$ is analytic in the tube

$$\{\mathbf{z} \mid \mathbf{z} \in \mathbb{C}^n, \sum_{j=1}^{n} |Im\, z_j|/r_j < 1\}$$

Theorem 3.3.3 *Let H be some region in $\mathbb{R}^n$ containing $\mathbf{z} = 0$. If the c.f. $f(\mathbf{z})$ is analytic in some region containing the set $\{\mathbf{z} \mid Re\, \mathbf{z} = 0,\ Im\, \mathbf{z} \in H\}$, then $f(\mathbf{z})$ is analytic in the convex tubular region G whose base is the convex hull of H.*

Moreover the representation (3.6) is valid throughout G, with the integral converging absolutely and uniformly on every compact subset of G.

In the multivariate case we call F to be Laplace-transformable or LT in the domain $G \subset \mathbb{R}^n$, if the integral

$$\int_{\mathbb{R}^n} e^{\mathbf{y}^T\mathbf{x}} dF(\mathbf{x}) < \infty, \quad \forall\ \mathbf{y} \in G \tag{3.7}$$

Theorem 3.3.4 *f is an a.c.f. for n-variate distribution function F, analytic on a convex tube with basis G if and only if F is Laplace transformable on G.*

Proof: If f is an a.c.f. then by theorem (3.3.1) we have the integral representation (3.6) is valid, and putting $Re\, \mathbf{z} = 0$, we have that F is Laplace transformable.

If F is LT, then the univariate projections of F are also LT. Therefore by theorem (2.2.2),we have that the c.f. of each univariate projection is an a.c.f.. By theorems (3.3.3) and (3.3.2) we have that f is an n-variate a.c.f.

Finally we have the multivariate analog of theorem (2.2.3).

Theorem 3.3.5 *Let g, f stand for the c.f.s of the n-variate d.f.s G, F respectively. Let F be LT in $H \subset I\!R^n$, and hence f be a.c.f. in a tube with basis H. Let G, F be related as*

$$dG(\mathbf{x}) = c.e^{\beta' \mathbf{x}} \, dF(\mathbf{x}), \quad -i\beta \in H, \; c \in I\!R \tag{3.8}$$

Then g is a.c.f and is given by

$$g(\mathbf{t}) = \frac{f(\mathbf{t} - i\beta)}{(-\mathrm{i}\beta)} \tag{3.9}$$

Since f is an a.c.f., theorem (3.3.1) is applicable and the integral representation is valid. Equation (3.9) thus when substituted in the right hand side of equation (3.9) gives us $g(t)$.

3.4 Examples and Applications

We will need the following results from classical analytic function theory.

(a) If the function $f(z)$ is analytic in the domain G, then it is continuous in G.

(b) If function $f_1(z)$ and $f_2(z)$ are analytic functions in G, then their sum and product are also analytic in G, and the function

$$\phi(z) = \frac{f_1(z)}{f_2(z)}$$

is an analytic function wherever $f_2(z) \neq 0$.

(c) If $\omega = f(z)$ is an analytic function in the domain G of the complex plane, and an analytic function $\Xi = \phi(\omega)$ is defined in the range H of its values in the ω plane, then the function $F(z) = \phi(f(z))$ is an analytic function of z in G.

Therefore we have the following results:

(1) $f(z) = z$, $f(z) = a$(constant), $f(z) = e^z$, are analytic.

Proof: Apply the Cauchy-Riemann conditions.

(2) By (a) and (b), $a_1 z^n$ for a_1, z complex, and $n < \infty$ is analytic in z, and hence any finite degree polynomial $P_n(z)$ in z is analytic.

(3) $e^{P_n(z)}$ is analytic by (a), (b) and (c).

3.4.1 Some standard multivariate distributions

We now illustrate theorems (3.3.4) and (3.3.5) by the following examples. See [58]).

Example 3.4.1 *Multivariate normal distribution.*

$$f_2(\mathbf{x}) = \frac{1}{(\sqrt{2\pi|\Sigma|})^{n/2}} e^{-(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)} \ , \ |\mathbf{x}|, |\mu| < \infty$$

where Σ is a real, positive definite symmetric matrix. The corresponding c.f. is

$$f\ (\mathbf{t}) = \exp(i\mathbf{t}'\mu) - \frac{1}{2}\mathbf{t}'\Sigma^{-1}\mathbf{t})$$

We have now to check the analyticity of $f(\mathbf{z})$.

$$f\ (\mathbf{z}) = \exp(i\mathbf{z}'\mu) - \frac{1}{2}\mathbf{z}'\Sigma^{-1}\mathbf{z})$$

Now $f(\mathbf{z})$ is of the form $e^{P_2(\mathbf{z})}$ and hence is analytic.
We denote the c.f. for $f_1(\mathbf{x})$ by $g\ (\mathbf{t})$. Therefore

$$g\ (\mathbf{t}) = \exp(i(\mathbf{t} - i\beta)'\mu - \frac{1}{2}(\mathbf{t} - i\beta)'\Sigma^{-1}(\mathbf{t} - i\beta)$$

$$-((i(-i\beta)'\mu - \frac{1}{2}(-i\beta)'\Sigma^{-1}(-i\beta))$$

$$= \exp(i(\mathbf{t}'(\Sigma^{-1}\beta + \mu) - \frac{1}{2}(\mathbf{t})'\Sigma^{-1}(\mathbf{t})$$

The resulting c.f. is of the same form as the original c.f. . From straightforward comparison we have that $f_2(x)$ is a multivariate normal density with mean $\mu = (\Sigma^{-1}\beta + \mu)$ and same covariance matrix Σ .

Example 3.4.2 *Wishart distribution (See [58]).*

Let, $\mathbf{x}_i \in \mathbb{R}^m$ be an m-variate random vector, with joint multinormal density :

$$p_{\mathbf{x}_i}(\mathbf{x}_i) = (2\pi)^{-(1/2)m}|\mathbf{V}|^{-1/2}\exp[-\frac{1}{2}(\mathbf{x}_i - \xi)'\mathbf{V}^{-1}(\mathbf{x}_i - \xi)]$$

The m.l.e. of $\mathbf{V}$ is $n^{-1}\mathbf{S}$ where

$$S_{jk} = \sum_{i=1}^{n}(x_{ij} - \overline{x}_{.j})(x_{ik} - \overline{x}_{.k})$$

$$\overline{x}_{.j} = n^{-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_{.j})$$

The m.g.f. (moment generating function) exists and is given by

$$E[\exp \operatorname{tr}(\boldsymbol{\Theta}\,\mathbf{S})] = \left[\frac{|\mathbf{V}^{-1}|}{|\mathbf{V}^{-1} + \boldsymbol{\Theta}|}\right]^{(1/2)\nu} \tag{3.10}$$

where

$$-\boldsymbol{\Theta} = \begin{pmatrix} -2\theta_{11} & \theta_{12} & \dots & \theta_{1n} \\ .. & -2\theta_{22} & \dots & \theta_{2n} \\ .. & .. & .. & .. \\ \theta_{n1} & .. & .. & -2\theta_{nn} \end{pmatrix}$$

The characteristic function is

$$\left[\frac{|\mathbf{V}^{-1}|}{|\mathbf{V}^{-1} + i\boldsymbol{\Theta}|}\right]^{(1/2)\nu} \tag{3.11}$$

Since m.g.f. exists, the distribution is Laplace Transformable, and hence its c.f. is analytic, in some domain of a complex tube. The complex extension is

$$\left[\frac{|\mathbf{V}^{-1}|}{|\mathbf{V}^{-1} + i\mathbf{Z}|}\right]^{(1/2)\nu} \tag{3.12}$$

This is a strictly negative power of a multivariate polynomial in elements of $\mathbf{Z}$. Therefore the extension is holomorphic in some domain and the roots of these polynomials denote the singularities of the extended c.f., - the convex hull of these singularities is the domain of regularity.

The set of singularities is the set of distinct elements of $\mathbf{Z}$, such that 1 is an eigenvalue of $\mathbf{V}\,\mathbf{Z}$.

Within the domain of regularity, the c.f. of a Laplace-transformed density is given as

$$\left[\frac{|\mathbf{V}^{-1}|}{|\mathbf{V}^{-1} + i(\boldsymbol{\Theta} - i\,\mathbf{B})|}\right]^{(1/2)\nu} \Big/ \left[\frac{|\mathbf{V}^{-1}|}{|\mathbf{V}^{-1} + i(-i\mathbf{B})|}\right]^{(1/2)\nu}$$

$$\left[\frac{|\mathbf{V}^{-1} + \mathbf{B}|}{|\mathbf{V}^{-1} + \mathbf{B} + i\Theta|}\right]^{(1/2)\nu} \tag{3.13}$$

where $\mathbf{B}$ is p.d. symmetric real matrix.

Note that equation (3.13) is of the same form as equation (3.11) . Thus the Wishart distribution obeys and admits LGD.

Example 3.4.3 *Multivariate Chi-square or Generalized Rayleigh.*

(See [58]).

(a) Joint distribution of $S_1, \ldots, S_m$

The c.f. is given by

$$E\,[\exp\,(\,i\sum_{j=1}^{m} t_j S_j\,)\,] \;=\; |\,\mathbf{I} \;-\; 2i\mathbf{V}\,\mathbf{D_t}\,|^{-\nu/2} \tag{3.14}$$

where $\mathbf{D_t} \;=\; \text{diag}\,(t_1, \ldots, t_m)$.

(b) Joint distribution of $Y_j \;=\; \text{tr}\;\mathbf{S}_{jj}\mathbf{V}_{jj}^{-1} \quad (\,j \;=\; 1, \ldots, m)$ where $\mathbf{S} \;=\; \sum_{j=1}^{m} \mathbf{X}_j\mathbf{X}_j'$ is the Wishart matrix, and $\mathbf{V}$ is the variance-covariance matrix.

The c.f. is

$$E\,[\exp\,(\,i\mathbf{t}'\mathbf{Y}\,)\,] \;=\; |\,\mathbf{I}_p \;-\; 2i\,\mathbf{D}\,(\mathbf{t})\,\mathbf{V}\,|^{-\nu/2} \tag{3.15}$$

where $\mathbf{D}\,(\mathbf{t}) \;=\; \text{diag}\,(t_1\mathbf{V}_{11}^{-1}, \ldots, t_m\mathbf{V}_{mm}^{-1})$ is a block-diagonal matrix.

Example 3.4.4 *Spherically symmetric class.*

Given a random vector $\mathbf{y} \in I\!R^n$ define the class of distributions $\mathcal{F}(\mathbf{y})$ as follows:

$$\mathcal{F}(\mathbf{y}) \;=\; \{\,\mathcal{L}(\mathbf{x})\,|\,\mathbf{x} \;=\; r\mathbf{y}\} \tag{3.16}$$

where r is a positive real. Letting the c.f. of $\mathbf{y}$ be denoted as $f(\mathbf{t})$ we immediately have from equation (3.16) that the corresponding class of c.f.s are given by

$$\mathcal{C}(\mathbf{y}) \;=\; \{g(\mathbf{t})|g(\mathbf{t}) \;=\; f(r\mathbf{t})\} \tag{3.17}$$

Equation (3.17) implies that if f is analytic and hence Laplace Transformable, the entire class $\mathcal{C}$ is analytic and hence Laplace Transformable.

3.4.2 Some familiar Reliability models

Given that an abundance of bivariate reliability distributions does exist to model such life distributions in practice, it seems quite natural to seek techniques to identify a distribution or a class of distributions that may be well advocated for a new observation. Classification techniques for reliability models however seem to have been not much dealt with. The fact that most bivariate reliability models are not members of regular, or even curved, exponential families makes the usual Fisher's rule unattractive. Further, one would prefer a "robust" rule in the face of uncertainty with the competing spectrum of distributions. With these facts in hand, we started to explore this area beginning with ([93]). As applications, we consider the bivariate exponential distributions of Moran, Freund, the semi-parametric bivariate exponential family of Kariya and Bilodeau (1993, JRSA), bivariate exponential conditional distributions of Arnold and Strauss (JRSS, B, 1988), mixture distributions, etc.

The most important distribution as evidenced in reliability literature is the exponential distribution, which plays somewhat an analogous role to that of the Normal Distribution. It forms a kind of boundary between the IHR and DHR classes. It is easy to verify that the Exponential distribution admits LGD. We now present some further examples.

Example 3.4.5 *Moran's bivariate exponetial distribution (See [58]).*

The c.f. is given by

$$\phi(t_1, t_2) = [(1 - it_1)(1 - it_2) + \omega^2 t_1 t_2]^{-\alpha} \tag{3.18}$$

where α is strictly positive. The marginals are each gamma with parameter α. The complex extension is

$$\phi(z_1, z_2) = [(1 - i z_1)(1 - i z_2) + \omega^2 z_1 z_2]^{-\alpha} \tag{3.19}$$

$$= [1 - i z_1 - i z_2 - (1 - \omega^2) z_1 z_2]^{-\alpha}$$

The right-hand side of equation (3.19) is a quadratic in z_1, z_2, raised to a strictly negative power, and hence is holomorphic, except at the singularities i.e. at the roots of the quadratic.

The singularities are given by

$$z_1 = \frac{i}{\pm\sqrt{1 - \omega^2}} \qquad z_2 = \frac{-i}{\mp\sqrt{1 - \omega^2}} \tag{3.20}$$

Therefore within the domain of regularity, the c.f. of the transformed density is

$$\left[1 - it_1 \frac{1 - \beta_2 + \beta_2\omega^2}{1 - \beta_1 - \beta_2 + (1 - \omega^2)\beta_1\beta_2} - it_2 \frac{1 - \beta_1 + \beta_1\omega^2)\,\beta_2}{1 - \beta_1 - \beta_2 + (1 - \omega^2)\beta_1\beta_2} \right.$$
$$\left. - \frac{(1 - \omega^2)t_1t_2}{1 - \beta_1 - \beta_2 + (1 - \omega^2)\beta_1\beta_2}\right]^{-\alpha} \tag{3.21}$$

Equating the coefficients of it_1, it_2 to 1 we have that if $\beta_i \neq 0$, then,one solution is $\beta_1 = \beta_2 = \beta$ (say), with $\beta = (1 + \omega^2)/(1 - \omega^2)$, and for this value of β the transformed density is of the same form as the original with $\omega' = \pm\frac{1}{\omega}$.

Example 3.4.6 *Freund's Bivariate Exponential (See [58]).*

The c.f. is given by

$$\frac{1}{\alpha + \beta - it_1 - it_2}\left[\frac{\beta}{1 - \frac{it_1}{\alpha'}} + \frac{\alpha}{1 - \frac{it_2}{\beta'}}\right] \tag{3.22}$$

The corresponding complex extension is a rational form (in polynomials) and hence analytic except at the roots of the denominator. If a, b denote the shift in the scale parameters, then their relationship is given by

$$\frac{\beta\beta'(\alpha - \alpha')}{\alpha\beta' - \alpha'\beta} - \frac{(\beta - \beta')(\alpha\alpha'\beta)}{\alpha\beta' - \alpha'\beta}.\frac{1}{a} = b \tag{3.23}$$

Example 3.4.7 *Bivariate Exponential Conditional Distribution (Arnold-Strauss, 1988 [12])*

The density for the bivariate random vector (X_1, X_2) is given by

$$f(x_1, x_2) = C(\lambda_3)\lambda_1\lambda_2 \exp\left[-\left(\sum_{i=1}^{2}\lambda_i x_i + \lambda_1\lambda_2\lambda_3 x_1 x_2\right)\right] \tag{3.24}$$

Note that the subfamily defined by

$$\lambda_1\lambda_2\lambda_3 = \text{constant} \tag{3.25}$$

admits LGD.

Example 3.4.8 *Bivariate Semi-parametric Exponential Dependency Parameter Family (Bilodeau and Kariya, 1993, [23])*

The density function is given by

$$f_\theta(x_1, x_2) = \lambda_1\lambda_2 \exp[-(\lambda_1 x_1 + \lambda_2 x_2)]g(\lambda_1 x_1, \lambda_2 x_2; \theta) \qquad (3.26)$$

The subfamily characterized by the condition that for every fixed (x_1, x_2), $g(\lambda_1 x_1, \lambda_2 x_2; \theta) = \text{constant}$, is the subfamily which admits LGD.

Chapter 4

The Pseudo Maximum Likelihood for LGD

4.1 Introduction

We have already seen in chapters 2 and 3, how a pair of densities admitting LGD are constrained by a functional relationship among the parameters of the two densities. Characterization when the exact form of density function is known has already been accomplished in the previous chapters. The analytical results obtained hide a significant statistical problem - that of estimation of parameters on the basis of two training samples in the presence of non-orthogonalizable functional dependencies among the parameters. Moreover, in the absence of knowledge about the actual functional form of the densities, a normalizing constant, which is an integral (in fact a Laplace transform), involving the parameters of discrimination, has to be estimated. This led us to investigate Pseudo ML methods beginning with ([90]), but we later had to take an entirely different approach in ([92]).

In general, the estimation for the parameters of the LGD rule are obtained through three types of sampling procedures, which are noted below. As we show, Anderson's [5] model neglects the dependency referred to above, to arrive at parameter estimates which are consistent under appropriate sampling procedures but less efficient (in terms of asymptotic variances), than when the dependencies are considered. A general lemma is proved from which consistency of Anderson's [5] estimates follow.

As this chapter proceeds on a generalization of the Pseudo Maximum

Likelihood (PML) procedure of Gong and Samaniego, [1981] ([46]) the procedure developed is called Generalized Pseudo Maximum Likelihood (GPML).

we assume the following:

1) Let f_1, f_2 be univariate densities.

2) Let f_1 , f_2 admit D(g) (Discrimination w.r.t g(x)) i.e.,

$$\frac{f_1}{f_2} = \frac{e^{g(\theta,x)}}{\int e^{g(\theta,x)} f_2(x)dx} \tag{4.1}$$

and f_1 , f_2 belong to the same family, and $\theta \in \Omega \subset I\!R^m$, for some m, and Ω is an open set.

3) Let $B = \{\theta | \int_{-\infty}^{\infty} e^{g(\theta,x)} f_2(x)dx < \infty\}$ be non-singleton.

4) Denote the first three partial derivatives of $g(\theta, x)$ w.r.t. θ by $\dot{g}, \ddot{g}, \dddot{g}$ respectively. Let the above partial derivatives exist as also their expectations over Ω.

5) The observations are $\underline{X} = (X_1, \ldots, X_n)$, where the X_i are independent with probability density either $f(x_i, \theta_1)$ or $f(x_i, \theta_2)$ with respect to Lebesgue measure μ.

4.2 Sampling procedures

In LGD, commonly three types of sampling rules are employed, to yield data appropriate for the construction of the rule. In the following we give a brief overview of the three approaches.

There are three common sampling designs (See Mclachlan [72]) that yield data suitable for estimating β : (1) mixture sampling in which a sample of $n = n_1 + n_2$ members is randomly selected from the total population $\mathcal{P}$ so that the n_i are random (Day and Kerridge [1967], [33]); (2) separate sampling in which for $i = 1, 2$ a sample of fixed size n_i is selected from G_i (Anderson [5], Prentice and Pyke [1979],[79]); and (3) conditional sampling in which, for $j = 1, 2, \ldots, m$, $n(x_j)$ members of $\mathcal{P}$ are selected at random from all members of $\mathcal{P}$ with $x = x_j$ (e.g., in bioassay where the two groups refer to "response" and "no response"). In all three cases we adopt the notation used for (3) and assume that the n observations take m different values x_j ($j = 1, 2, \ldots, m$) with frequency $n_i(x_j)$ in group G_i.

4.2.1 Conditional sampling

Here $n(x_j) = n_1(x_j) + n_2(x_j)$ is fixed and $n_i(x_j)$ is random. The likelihood function for this design is

$$L_C = \prod_{j=1}^{m} \{ \Pr [x \in G_1 \mid x = x_j] \}^{n_1(x_j)} \{ \Pr [x \in G_2 \mid x = x_j] \}^{n_2(x_j)}$$

$$= \prod_{i=1}^{2} \prod_{j=1}^{m} \{ q_i (x_j) \}^{n_i x_j} \tag{4.2}$$

From (4.2) we see that L_C is a function of α_0 and β that can be maximized to obtain maximum likelihood estimates.

4.2.2 Mixture sampling

In this case n is fixed and the x_j and $n(x_j)$ are now random variables. Let $L(x, G_i) = \pi_i f_i(x)$ and

$$f(x) = \pi_1 f_1(x) + \pi_2 f_2(x)$$

Then the likelihood function is

$$L_M = \prod_{i=1}^{2} \prod_{j=1}^{m} \{ L(x_j, G_i) \}^{n_i(x_j)}$$

$$= \left[\prod_{i=1}^{2} \prod_{j=1}^{m} \{ \frac{L(x_j, G_i)}{f(x_j)} \}^{n_i(x_j)} \right] \left[\prod_{j=1}^{m} \{ f(x_j) \}^{n(x_j)} \right]$$

$$= \left[\prod_{i=1}^{2} \prod_{j=1}^{m} \{ q_i (x_j) \}^{n_i(x_j)} \right] L$$

$$= L_C \, L \tag{4.3}$$

Since $f_i(x)$ is unspecified, no assumptions are made about the form of $f(x)$, the density function for the above sampling scheme. We can therefore assume that $f(x)$ contains no useful information about α_0 and β. Even if we knew something about $f_i(x)$, the extra information about α_0 and β in L would be small compared with that contained in L_C. Therefore, as in conditional sampling maximum likelihood estimates are again found by maximizing L_C.

In practice we frequently have $m = n$, that is, all the observations are different and $n_i(x_j)$ is 0 or 1.

4.2.3 Separate sampling

Separate sampling is generally the most common sampling situation and the likelihood function is

$$L_S = \prod_{i=1}^{2}\prod_{j=1}^{m} \{ L(x_j|\, G_i) \}^{n_i(x_j)}$$

$$= \prod_{i=1}^{2}\prod_{j=1}^{m} \{ f_i(x_j) \}^{n_i(x_j)}$$

$$= \prod_{i=1}^{2}\prod_{j=1}^{m} \{ \frac{L(x_j,\, G_i)}{\pi_i} \}^{n_i(x_j)}$$

$$= L_M\, \pi_1^{-n_1} \pi_2^{-n_2} \tag{4.4}$$

If π_1 and π_2 are known, then this model is equivalent to (4.3). However, if π_1 and π_2 are unknown, we proceed as follows (Anderson and Blair [1982], [6]). We assume that x is discrete so that the values of f may be taken as multinomial probabilities. From (1.1) we have

$$f_1(x) = f_2(x) \exp(\alpha + \beta' x)$$

and from (4.4),

$$L_S = \prod_{j=1}^{m} p_{x_j}^{n(x_j)} \exp\{n_1(x_j)\ [\alpha + \beta' x]\} \tag{4.5}$$

where $p_x = f_2(x)$. The problem is to maximize L_S subject to the constraints that f_1 and f_2 are probability functions, namely,

$$\sum_x p_x = 1 \tag{4.6}$$

and

$$\sum_x p_x \exp(\alpha + \beta' x) = 1 \tag{4.7}$$

Using Lagrange multipliers it can be shown that the answer is given by estimating p_x by

$$\hat{p_x} = \frac{n(x)}{n_1 \exp(\alpha + \beta' x) + n_2}$$

substituting $\hat{p_x}$ into L_S of (4.5), and then maximizing the resulting expression

$$L'_S = L'_C n_1^{-n_1} n_2^{-n_2} \prod_{j=1}^{m} [n(x_j)]^{n(x_j)}$$

where

$$L'_C = \prod_{i=1}^{2} \prod_{j=1}^{m} \{ \tilde{q}_i (x_j) \}^{n_i(x_j)} \tag{4.8}$$

$$\tilde{q}_1(x) = \frac{n_1 \exp(\alpha + \beta' x)}{n_1 \exp(\alpha + \beta' x) + n_2}$$

$$= \frac{\exp(\alpha + \log(n_1/n_2) + \beta' x)}{\exp(\alpha + \log(n_1/n_2) + \beta' x) + 1}$$

and

$$\tilde{q}_2(x) = 1 - \tilde{q}_1(x)$$

we note that $\tilde{q}_1$ is the same as $q_1(x)$ of (4.2), except that $\alpha_0 = \alpha + \log(\pi_1/\pi_2)$ is replaced by $\alpha + \log(n_1/n_2)$. Hence with a correct interpretation of α_0, maximizing L'_C is equivalent to maximizing L_C. We have therefore reduced the problem of maximizing L_S subject to (4.6) and (4.7) to maximizing L_C again. Prentice and Pyke [79] claim that the restriction to discrete variables can be dropped, as the above estimates will still have satisfactory properties for continuous variables. However, Anderson and Blair [1982],[6] show that for continuous variables, the estimates are no longer technically maximum likelihood and suggest an alternative method called penalized maximum likelihood estimation.

4.3 Consistent and Efficient Estimation

4.3.1 Introduction

Gong and Samaniego [46] gave an estimation procedure which took into account functional dependencies which were orthogonalizable (in the sense that parameters take values in a certain Cartesian product). Here generalization of their ideas to the non-orthogonalizable case is constructed.

4.3.2 Generalized Pseudo-ML

Let $X_1, \ldots, X_n$ be a sequence of random variables as defined above with a corresponding sequence of $\sigma-$algebras $\mathcal{F}_n$ such that X_n is $\mathcal{F}_n$ adapted.

Theorem 4.3.1 *Let ϕ, ψ and ϕ_n be functions such that*

$$\phi(\theta)\ ,\ \phi_n(\theta) : \Omega \longrightarrow I\!R^m, \text{ and } \psi : \ (X_1, \ldots, X_n, \phi_n(\theta), \theta) \to I\!R$$

Let there exist for every positive integer i a sequence of random vectors $Z_{ik} \in I\!R^m$, $k = 1, 2, \ldots$ such that

$$\psi(X_1, \ldots, X_n, \phi_n(\theta), \theta) - \psi(X_1, \ldots, X_n, \phi(\theta), \theta) = Z'_{in}(\phi_n(\theta) - \phi(\theta)) \quad (4.9)$$

$$Z_{in} \text{is bounded a.s. or (p)}, \quad \phi_n(\theta) - \phi(\theta) \to 0 \text{ a.s. or (p)} \quad (4.10)$$

then

$$\psi(X_1, \ldots, X_n, \phi_n(\theta), \theta) \ - \ \psi(X_1, \ldots, X_n, \phi(\theta), \theta) \longrightarrow 0 \text{ a.s. or (p)} \quad (4.11)$$

Proof: Since Z'_{in} is a.s. bounded, the convergence of ϕ_n to ϕ implies that the right hand side of equation 4.9 converges to 0.

Theorem 4.3.2 *Let*

$$\phi(\theta)\ ,\ \phi_n(\theta) : \Omega \longrightarrow I\!R^m$$

$\phi_n(\theta)$ is $\mathcal{F}_n$ -measurable, and $\phi_n(\theta) \to \phi(\theta)$ a.s. Let there exist a function h such that

$$\psi(X_1, \ldots, X_n, \phi_n(\theta), \theta) = \frac{1}{n}\sum_i h\left(X_i, \phi_n(\theta), \theta\right)$$

$$\psi(X_1, \ldots, X_n, \phi(\theta), \theta) = \frac{1}{n}\sum_i h\left(X_i, \phi(\theta), \theta\right) \quad (4.12)$$

For given X_i, θ, one of the three following conditions hold,

(a) $\exists$ a function d such that

$$h(X_i, \phi_n(\theta), \theta) - h(X_i, \phi(\theta), \theta) = d(X_i, \phi_n(\theta), \phi(\theta), \theta)(\phi_n(\theta) - \phi(\theta)) \quad (4.13)$$

$\exists$ a real valued function $G(x, \theta)$ with finite expectation, such that

$$|d(X_i, \phi_n(\theta), \phi(\theta), \theta)| \leq G(X_i, \theta) \tag{4.14}$$

(b) $h(X_i, u, \theta)$ is differentiable w.r.t. u, $\forall u \in \omega \subset \Omega$. There exists a real valued function $M(x, \theta)$ with finite expectation, such that

$$|\frac{\partial}{\partial u} h\,(X_i, u, \theta)| \;\leq\; M(x, \theta) \forall\, i, n \tag{4.15}$$

where $u_{ni} \sim u_{ni}(X_1, \ldots, X_n, \theta)$ is $\mathcal{F}_n$-measurable, and $\phi(\theta) \leq u_{ni} \leq \phi_n(\theta)$.

Then the conclusion of theorem (4.3.1) holds. In particular

$$\psi(X_1, \ldots, X_n, \phi_n(\theta), \theta) \to E\, h(X_1, \phi(\theta), \theta) \tag{4.16}$$

(c) Condition (4.15) can be replaced by the condition that $h(X_i, u_{ni}, \theta)$ is uniformly integrable.

proof:

(a) Equation 4.14 implies that the function d is almost surely bounded (by the Strong Law of Large Numbers).

(b) Since h is differentiable, a Taylor expansion upto the first term is valid, and from equation 4.15 the derivative is a.s. bounded.

(c) Since h is differentiable, h is continuous, hence, $h(X_i, u_{ni}, \theta) \longrightarrow h(X_i, \phi(\theta), \theta)$ a.s., since $u_{ni} \to \phi(\theta)$ a.s. With uniform integrability, $E(h(X_i, u_{ni}, \theta)) \longrightarrow E(h(X_i, \phi(\theta), \theta))$. Therefore by Cezaro limit $\frac{1}{n} \sum_i E(h(X_i, u_{ni}, \theta)) \longrightarrow E(h(X_i, \phi(\theta), \theta)) < \infty$

4.4 LGD and its generalizations

We need the following two Martingale convergence and central limit theorems : (See Hall and Heyde, [1980],[49]).

Theorem 4.4.1 *Let* $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ *be a martingale and* $\{U_n, n \geq 1\}$ *a nondecreasing sequence of positive r.v. such that* U_n *is* $\mathcal{F}_{n-1}$*-measurable for each n. If* $1 \leq p \leq 2$ *then*

$$\lim_{n \to \infty} U_n^{-1} S_n \;=\; 0 \quad \text{a.s.} \tag{4.17}$$

on the set $\{\lim_{n \to \infty} U_n = \infty, \sum_{i=1}^{\infty} U_i^{-p} E(|X_i|^p | \mathcal{F}_{i-1}) < \infty\}$.

Theorem 4.4.2 *Suppose that the probability space $(\Omega_n, \mathcal{F}_n, P_n)$ supports square-integrable r.v.s $S_{n1}, S_{n2}, \ldots, S_{nk_n}$, and that the S_{ni} are adapted to the σ-fields $\mathcal{F}_{ni}$, where $\mathcal{F}_{n1} \subset \mathcal{F}_{n2} \subset \ldots \mathcal{F}_{nk_n} \subset \mathcal{F}_n$. Let $X_{ni} = S_{ni} - S_{n,i-1}$ $(S_{n0} = 0)$ and $U_{ni}^2 = \sum_{j=1}^{i} X_{nj}^2$. If $\mathcal{G}_n$ is a sub-σ-field of $\mathcal{F}_n$, let $\mathcal{G}_{ni} = \mathcal{F}_{ni} \cup \mathcal{G}_n$, (the σ-field generated by $\mathcal{F}_{ni} \cup \mathcal{G}_n$) and let $\mathcal{G}_{n0} = \{\Omega_n, \phi\}$ denote the trivial σ-field. Suppose further that*

$$\max_i |X_{ni}| \xrightarrow{p} 0, \quad \mathrm{E}\,(\max_i X_{ni}^2) \text{ is bounded in } n \tag{4.18}$$

There exists σ-fields $\mathcal{G}_n \subset \mathcal{F}_n$ and $\mathcal{G}_n$- measurable r.v.s u_n^2 such that

$$U_{nk_n}^2 - u_n^2 \xrightarrow{p} 0 \tag{4.19}$$

$$\sum_i \mathrm{E}\,(X_{ni}|\mathcal{G}_{n,i-1}) \xrightarrow{p} 0, \sum_i |\mathrm{E}\,(X_{ni}|\mathcal{G}_{n,i-1})|^2 \xrightarrow{p} 0, \tag{4.20}$$

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \mathrm{P}\,(U_{nk_n} > \delta) = 1 \tag{4.21}$$

Then

$$\frac{S_{nk_n}}{U_{nk_n}} \xrightarrow{d} \mathrm{N}(0,1) \tag{4.22}$$

4.4.1 Estimators for normalizing constant and derivatives

Let $\exists,\ i,\ 1 \le i \le n,\ 1_{[i]} = 1$

$$\begin{aligned} X_i &\sim Z_1 \text{ if } 1_{[i]} = 0, \quad P(g(Z_2) = -\infty) = 0 \\ &\sim Z_2 \text{ if } 1_{[i]} = 1, \quad E(1_{[i]}) = 1 - \pi > 0 \end{aligned}$$

Define

$$\phi_0 = E(e^{g(Z_2)}),\ \phi_1 = E(e^{g(Z_2)}\dot{g}(Z_2)),\ \phi_2 = E(e^{g(Z_2)}(\ddot{g}(Z_2) + \dot{g}^2(Z_2))) \tag{4.23}$$

Let ϕ_0, ϕ_1, ϕ_2 exist. Define,

$$\phi_{0,n} = \frac{\sum_i 1_{[i]} e^{g(X_i)}}{\sum_i 1_{[i]}},\ \phi_{1,n} = \frac{\sum_i 1_{[i]} e^{g(X_i)} \dot{g}}{\sum_i 1_{[i]}}$$

$$\phi_{2,n} = \frac{\sum_i 1_{[i]} e^{g(X_i)} (\ddot{g} + \dot{g}^2)}{\sum_i 1_{[i]}} \tag{4.24}$$

Theorem 4.4.3 *Let equations (4.23) and (4.24) hold. Then*
(i) $P(\phi_{0,n} > 0) = 1$.
(ii) and

$$\phi_{0,n} \to \phi_0, \ \phi_{1,n} \to \phi_1, \ \phi_{2,n} \to \phi_2, \ a.s. \tag{4.25}$$

$$\frac{\phi_{1,n}}{\phi_{0,n}} \to \frac{\phi_1}{\phi_0}, \ \frac{\phi_{2,n}}{\phi_{0,n}} \to \frac{\phi_2}{\phi_0}, \ a.s. \tag{4.26}$$

Proof: (i) w.l.o.g. let $1_{[1]} = 1$. Therefore,

$$P(\phi_{0,n} = 0) \leq P(1_{[1]}e^{g(Z_2)} = 0) = P(g(Z_2) = -\infty) = 0$$

and $\phi_0 > 0$ by nonnegativeness of exponential function.
(ii) Consider

$$\sum_i 1_{[i]}(\phi_{0,n} - \phi_0) = \sum_i 1_{[i]}(e^g - \phi_0)$$

$$\sum_i 1_{[i]}(\phi_{1,n} - \phi_1) = \sum_i 1_{[i]}(e^g \dot{g} - \phi_1)$$

$$\sum_i 1_{[i]}(\phi_{2,n} - \phi_2) = \sum_i 1_{[i]}(e^g \ddot{g} - \phi_2)$$

$$\sum_i 1_{[i]}(\phi_{1,n} - \phi_1) - \frac{\phi_1}{\phi_0}\sum_i 1_{[i]}(\phi_{0,n} - \phi_0)$$

$$\sum_i 1_{[i]}(\phi_{2,n} - \phi_2) - \frac{\phi_2}{\phi_0}\sum_i 1_{[i]}(\phi_{0,n} - \phi_0)$$

are all martingales. Hence by theorem (4.4.1) each of them normalized by n, converge a.s. to 0. Since $\sum_i 1_{[i]}/n$ converge a.s. to π, part (ii) of theorem follows.

4.4.2 Construction of the PML equations

For our model the log-likelihood is given as

$$L_n = \sum 1'_{[i]}(\alpha + g - \ln(\phi_{0,n} + e^{(\cdot)})) - \sum 1_{[i]} \ln(\phi_{0,n} + e^{(\cdot)}) + 1_{[i]} \ln \phi_{0,n} \tag{4.27}$$

and its derivatives as

$$X_{ni} = \frac{1}{n}(1'_{[i]}.\frac{\phi_0\dot{g} - \phi_1}{\phi_0 + e^{(\cdot)}} - 1_{[i]}.\frac{\phi_0\dot{g} - \phi_1}{\phi_0 + e^{(\cdot)}}.\frac{e^{(\cdot)}}{\phi_0}) \tag{4.28}$$

$$Y_{ni} = \frac{1}{n}(1'_{[i]}.\frac{\phi_0}{\phi_0 + e^{(\cdot)}} - 1_{[i]}.\frac{e^{(\cdot)}}{\phi_0 + e^{(\cdot)}}) \tag{4.29}$$

Lemma 4.4.1 *The following convergences are a.s.*

$$\frac{1}{n}\left(\sum \frac{\phi_{0,n}\dot{g} - \phi_{1,n}}{\phi_{0,n} + e^{\alpha+g}} - \sum \frac{\phi_0\dot{g} - \phi_1}{\phi_0 + e^{\alpha+g}}\right) \to 0$$

$$\frac{1}{n}\left(\sum \frac{\phi_{0,n}\dot{g} - \phi_{1,n}}{\phi_{0,n} + e^{\alpha+g}}\frac{e^{(\cdot)}}{\phi_{0,n}} - \sum \frac{\phi_0\dot{g} - \phi_1}{\phi_0 + e^{\alpha+g}}\frac{e^{(\cdot)}}{\phi_0}\right) \to 0$$

$$\frac{1}{n}\Big(\sum 1'_{[i]}(\alpha + g - \ln(\phi_{0,n} + e^{(\cdot)})) - \sum 1_{[i]}\ln(\phi_{0,n} + e^{(\cdot)}) + 1_{[i]}\ln\phi_{0,n}$$
$$- \sum 1'_{[i]}(\alpha + g - \ln(\phi_0 + e^{(\cdot)})) - \sum 1_{[i]}\ln(\phi_0 + e^{(\cdot)}) + 1_{[i]}\ln\phi_0\Big) \to 0$$

Proof:

$$\frac{\phi_{0,n}\dot{g} - \phi_{1,n}}{\phi_{0,n} + e^{\alpha+g}} - \frac{\phi_0\dot{g} - \phi_1}{\phi_0 + e^{\alpha+g}}$$
$$= \frac{e^{(\cdot)}\dot{g} + \phi_1}{(\phi_{0,n} + e^{(\cdot)})(\phi_0 + e^{(\cdot)})}(\phi_{0,n} - \phi_0) - \frac{\phi_{1,n} - \phi_1}{\phi_{0,n} + e^{(\cdot)}}$$

$$\frac{\phi_{0,n}\dot{g} - \phi_{1,n}}{\phi_{0,n} + e^{\alpha+g}}\frac{e^{(\cdot)}}{\phi_{0,n}} - \frac{\phi_0\dot{g} - \phi_1}{\phi_0 + e^{\alpha+g}}\frac{e^{(\cdot)}}{\phi_0}$$
$$= e^{(\cdot)}\left[\frac{1}{\phi_{0,n} + e^{(\cdot)}}\left(\frac{\phi_{1,n}}{\phi_{0,n}} - \frac{\phi_1}{\phi_0}\right) + \frac{(\phi_0\dot{g} - \phi_1)(\phi_{0,n} - \phi_0)}{\phi_0(\phi_0 + e^{(\cdot)})(\phi_{0,n} + e^{(\cdot)})}\right]$$

Note that

$$\frac{1}{(\phi_{0,n} + e^{(\cdot)})} \leq \frac{1}{e^{(\cdot)}}, \quad \frac{1}{(\phi_0 + e^{(\cdot)})} \leq \frac{1}{\phi_0}$$

$$1'_{[i]}(\alpha + g - \ln(\phi_{0,n} + e^{(\cdot)})) - 1_{[i]} \ln(\phi_{0,n} + e^{(\cdot)}) + 1_{[i]} \ln \phi_{0,n}$$

$$= 1'_{[i]}(\alpha + g) - (1'_{[i]} + 1_{[i]}) \ln(\phi_{0,n} + e^{(\cdot)})) + 1_{[i]} \ln \phi_{0,n} \qquad (4.30)$$

Replacing $\phi_{0,n}$ by the dummy variable u and differentiating w.r.t. u we can take the Taylor expansion of equation (4.30) as

$$- (1'_{[i]} + 1_{[i]}) \frac{1}{u + e^{(\cdot)}} + 1_{[i]} \frac{1}{u}\bigg|_{u_{ni}} (\phi_{0,n} - \phi_0) \qquad (4.31)$$

where

$$\phi_0 \leq u_{ni} \leq \phi_{0,n}$$

$$\frac{1}{u_{ni}} \leq \max\left(\frac{1}{\phi_0}, \frac{1}{\phi_{0,n}}\right) \to \frac{1}{\phi_0}$$

$$\text{if } u > 0, \; \frac{1}{u + e^{(\cdot)}} \leq e^{-(\cdot)}$$

Therefore by theorems (4.4.3), (4.3.1), and (4.3.2) the proof of the lemma is complete.

4.4.3 Consistency of the PMLE

Lemma 4.4.2 *For fixed $y > 0$ let*

$$\psi = y \ln \frac{x}{1 + x} - \ln(1 + x) \quad x > 0 \qquad (4.32)$$

Then ψ reaches its maximum at $x = y$.

Proof:

$$\frac{\partial \psi}{\partial x} = \frac{y}{x} - \frac{1 + y}{1 + x}$$

$$\frac{\partial^2 \psi}{\partial x^2} = -\frac{y}{x^2} + \frac{1 + y}{(1 + x)^2}$$

$$\frac{\partial \psi}{\partial x} = 0 \Rightarrow x = y$$

At $x = y$

$$\frac{\partial^2 \psi}{\partial x^2} = -\frac{1}{x} + \frac{1}{1 + x} = -\frac{1}{x(1 + x)} < 0$$

Lemma 4.4.3 *Let*

$$\lim_{n\to\infty} \frac{1'_{[i]}}{n} = \rho$$

Let u, α^0, θ^0 *fixed. Put* $\rho/(1-\rho) = \exp(c)$. *Consider*

$$(1-\rho)\psi = \rho \frac{e^{g(\theta^0,u)}}{\phi(\theta^0)} \log \frac{e^{\alpha+g(\theta,u)}}{\phi(\theta)+e^{\alpha+g(\theta,u)}} + (1-\rho)\log\frac{\phi(\theta)}{\phi(\theta)+e^{\alpha+g(\theta,u)}}$$

Then ψ is maximized at

$$\frac{e^{\alpha\,+\,g(\theta,x)}}{\phi_0(\theta)} = \frac{e^{c\,+\,g(\theta^0,x)}}{\phi_0(\theta^0)}$$

Proof: Enough to put $y = \dfrac{e^{\alpha_0\,+\,g(\theta^0,x)}}{\phi_0(\theta^0)}, \quad x = \dfrac{e^{\alpha\,+\,g(\theta,x)}}{\phi_0(\theta)}$

Define

$$L_i = 1'_{[i]}(\alpha + g - \log(\phi_{0,n} + e^{\alpha+g}) + 1_{[i]}(\log\phi_{0,n} - \log(\phi_{0,n} + e^{\alpha+g})) \quad (4.33)$$

$$L_{0,i} = 1'_{[i]}(\alpha + g - \log(\phi_0 + e^{\alpha+g}) + 1_{[i]}(\log\phi_0 - \log(\phi_0 + e^{\alpha+g})) \quad (4.34)$$

Let

$$L_n = \frac{1}{n}\sum_i L_i, \quad L_{0,n} = \frac{1}{n}\sum_i L_{0,i}$$

Then by lemma (4.4.1)

$$L_n(\alpha,\theta) - L_n(\alpha^0,\theta^0) - (L_{0,n}(\alpha,\theta) - L_{0,n}(\alpha^0,\theta^0)) \to 0 \text{ a.s.}$$

Now,

$$(L_{0,n}(\alpha,\theta) - L_{0,n}(\alpha^0,\theta^0)) \to$$
$$E_{\alpha^0,\theta^0}\left[(L_{0,n}(\alpha,\theta) - L_{0,n}(\alpha^0,\theta^0)\right] \to d,\ d<0$$

by lemma (4.4.3), and the convexity of $-\ln(x)$.

Hence the local maximums of L_n converge in probability (a.s.) to α^0, θ^0.

Replacing $\phi_{0,n}, \phi_0$ by 1 we find that in Andersons form the corresponding local maximums tend not to α^0, θ^0 but to $\alpha^0 - \log\phi_0, \theta^0$.

Thus in case of LGD, the conditional and mixture sampling will yield consistent solutions but the separate sampling scheme, may not.

4.4.4 Asymptotic normality

Lemma 4.4.4 *Let $X_1, \ldots, X_n$ be a sequence of i.i.d random variables. Let Z_n denote the maximum of $X_1, \ldots, X_n$. Let $E(|X_1|^k) < \infty$. Then*

$$P\left(|Z_n| > \epsilon\right) \leq \frac{n\, E\left(|X_1|^k\right)}{\epsilon^k}, \quad k \in N^+ \tag{4.35}$$

Proof. Since Z_n is nothing but the n-th order statistic,

$$E(|Z_n|^k) \leq n\, E(|X_1|^k)$$

Applying the general form of Markov Inequality, the lemma follows.

For L_α we have

$$|Y_{ni}| \leq \frac{1}{n} \forall\, i \rightarrow \max_i |X_{ni}| \xrightarrow{p} 0 \tag{4.36}$$

For L_θ we have

$$E(\max_i |X_{ni}|^2) \leq nE\left(X_{n1}^2\right) = \frac{nc}{n^2}, \quad c \in I\!R, \tag{4.37}$$

implies $E(\max_i |X_{ni}|^2)$ is bounded in n.

It is easily verified that

Lemma 4.4.5

$$E_{\alpha^0,\theta^0}\left[\frac{\partial nX_{ni}}{\partial\theta} + (nX_{ni})(nX_{ni})'\right] = 0 \tag{4.38}$$

$$E_{\alpha^0,\theta^0}\left[\frac{\partial nX_{ni}}{\partial\alpha} + nY_{ni}(nX_{ni})\right] = 0 \tag{4.39}$$

$$E_{\alpha^0,\theta^0}\left[\frac{\partial nY_{ni}}{\partial\theta} + (nY_{ni})(nX_{ni})\right] = 0 \tag{4.40}$$

$$E_{\alpha^0,\theta^0}\left[\frac{\partial nY_{ni}}{\partial\alpha} + (nY_{ni})^2\right] = 0 \tag{4.41}$$

From lemma (4.4.5) we have that

$$E\,(\ddot{L}_n \;+\; \dot{L}_n \dot{L}_n') \;=\; 0$$

Let $L_n, \dot{L}_n, \ddot{L}_n$ denote the log-likelihood, the first and second derivative with respect to θ of the log-likelihood. Then

$$\dot{L}_n(\theta_n, \phi_n(\theta_n)) - \dot{L}_n(\theta^0, \phi_n(\theta^0))$$

$$= \ddot{L}_n(\theta^0, \phi_n(\theta^0))(\theta_n - \theta^0) \;+\; (\ddot{L}_n(\theta_n^\star, \phi_n(\theta_n^\star)) - \ddot{L}_n(\theta^0, \phi_n(\theta^0)))(\theta_n - \theta^0)$$

where $\theta^0 \le \theta_n^\star \le \theta_n$. If

$$\theta_n \longrightarrow \theta^0 \text{ a.s. or in (p)}$$

then

$$(\ddot{L}_n(\theta_n^\star, \phi_n(\theta_n^\star)) - \ddot{L}_n(\theta^0, \phi_n(\theta^0))) \longrightarrow \mathbf{0}$$

i.e., converges to the identically zero matrix. Let $E_{\alpha_0,\theta_0} 1_{[1]}' = \pi,\ 0 < \pi < 1$. Then,

$$\mathrm{E}_{\theta^0,\alpha^0}(X_{ni}) \;=\; \frac{1}{n}(\pi\ \mathrm{E}_1(\frac{\phi_0 \dot{g} - \phi_1}{\phi_0 + e^{(\cdot)}}) \;-\; e^{\alpha_0}(1-\pi)\mathrm{E}_1(\frac{\phi_0 \dot{g} - \phi_1}{\phi_0 + e^{(\cdot)}})) \;=\; 0 \quad (4.42)$$

$$\mathrm{E}_{\theta^0,\alpha^0}(Y_{ni}) \;=\; \frac{1}{n}(\pi \mathrm{E}_1(\frac{\phi_0}{\phi_0 + e^{(\cdot)}}) \;-\; e^{\alpha^0}(1-\pi)\mathrm{E}_1(\frac{\phi_0}{\phi_0 + e^{(\cdot)}})) \;=\; 0 \quad (4.43)$$

$$\mathrm{E}_{\theta^0,\alpha^0}(X_{ni}X_{ni}') \;=\; \frac{1}{n^2}(\pi\ \mathrm{E}_1(\frac{(\phi_0 \dot{g} - \phi_1)(\phi_0 \dot{g} - \phi_1)'}{\phi_0(\phi_0 + e^{(\cdot)})}) \quad (4.44)$$

$$\mathrm{E}_{\theta^0,\alpha^0}(X_{ni}Y_{ni}) \;=\; \frac{1}{n^2}(\pi\ \mathrm{E}_1(\frac{(\phi_0 \dot{g} - \phi_1)}{\phi_0 + e^{(\cdot)}}) \quad (4.45)$$

$$\mathrm{E}_{\theta^0,\alpha^0}(Y_{ni}^2) \;=\; \frac{1}{n^2}\pi \mathrm{E}_1(\frac{\phi_0}{\phi_0 + e^{(\cdot)}}) \quad (4.46)$$

For Andersons model

$$X_{ni} \;=\; \frac{1}{n}(1_{[i]}'.\frac{\dot{g}}{1 + e^{(\cdot)}} \;-\; 1_{[i]}.\frac{\dot{g}.e^{\alpha+g}}{1 + e^{(\cdot)}}) \quad (4.47)$$

$$Y_{ni} \;=\; \frac{1}{n}(1_{[i]}'.\frac{1}{1 + e^{(\cdot)}} \;-\; 1_{[i]}.\frac{e^{\alpha+g}}{1 + e^{\alpha+g}} \quad (4.48)$$

At $\theta^0, \alpha^0 - \ln \phi_0(\theta^0)$

$$X_{ni} = \frac{1}{n}(1'_{[i]}.\frac{\phi_0 \dot{g}}{\phi_0 + e^{\alpha_0+g}} - 1_{[i]}.\frac{\dot{g}.e^{\alpha_0+g}}{\phi_0 + e^{\alpha_0+g}}) \tag{4.49}$$

$$Y_{ni} = \frac{1}{n}(1'_{[i]}.\frac{\phi_0}{\phi_0 + e^{\alpha_0+g}} - 1_{[i]}.\frac{e^{\alpha_0+g}}{\phi_0 + e^{\alpha_0+g}}) \tag{4.50}$$

$$\mathrm{E}_{\theta^0,\alpha^0}(X_{ni}) = 0 \quad \mathrm{E}_{\theta^0,\alpha^0}(Y_{ni}) = 0 \tag{4.51}$$

$$\mathrm{E}_{\theta^0,\alpha^0}(X_{ni}X'_{ni}) = \frac{1}{n^2}\pi\,\mathrm{E}_1(\frac{(\phi_0\dot{g})(\phi_0\dot{g})'}{\phi_0(\phi_0 + e^{(.)})}) \tag{4.52}$$

$$\mathrm{E}_{\theta^0,\alpha^0}(X_{ni}Y_{ni}) = \frac{1}{n^2}\pi\,\mathrm{E}_1(\frac{(\phi_0\dot{g})}{\phi_0 + e^{(.)}}) \tag{4.53}$$

$$\mathrm{E}_{\theta^0,\alpha^0}(Y_{ni}^2) = \frac{1}{n^2}\pi\mathrm{E}_1(\frac{\phi_0}{\phi_0 + e^{(.)}}) \tag{4.54}$$

By lemma (4.4.1)

$$\dot{L}_n(\theta^0, \phi_n(\theta^0)) - \dot{L}_n(\theta^0, \phi_0(\theta^0)) \to 0\text{a.s.}$$

$$\ddot{L}_n(\theta^0, \phi_n(\theta^0)) - \ddot{L}_n(\dot{\theta}^0, \phi_0(\theta^0)) \to 0\text{a.s.}$$

By theorem (4.4.2)

$$\dot{L}_n(\theta^0, \phi_0(\theta^0)) \sim N(0, \Sigma) \tag{4.55}$$

where $\Sigma = E(\ddot{L}_n)$

For the asymptotic variance-covariance matrix for $\dot{L}$, we have when $\theta \in I\!R^m$,

$$\Sigma = \frac{\pi}{n}\begin{pmatrix} E\frac{\phi_0}{\phi_0+e^{(.)}} & E\frac{\phi_0\dot{g}-\phi_1}{\phi_0+e^{(.)}} \\ & E\frac{(\phi_0\dot{g}-\phi_1)(\phi_0\dot{g}-\phi_1)^T}{\phi_0(\phi_0+e^{(.)})} \end{pmatrix} \tag{4.56}$$

Note that the Σ for Andersons form (say Σ_A) can be found out simply by substituting $\phi_1 = 0$ in the Σ (say Σ_{GPML}) in equation (4.56). Thus $\Sigma_{GPML} - \Sigma_A$ is non-negative definite, if the matrix

$$\Sigma = \frac{\pi}{n}\begin{pmatrix} -E\log\frac{n_1}{n_2} & E\frac{-\phi_1}{\phi_0+e^{(.)}} \\ & E\frac{-\phi_1\dot{g}'-\dot{g}\phi_1'+\phi_1\phi_1'}{\phi_0(\phi_0+e^{(.)})} \end{pmatrix} \tag{4.57}$$

is non-negative definite. Hence in this case, since Σ_A is positive definite, $\Sigma_{GPML}^{-1} - \Sigma_A^{-1}$ is non-positive definite, by C.R.Rao,[1974], ([82], p 70).

Hence, the above result can be summarized as follows :

Theorem 4.4.4 *Let θ_n denote the solution of the GPML equations, (i.e., $\dot{L}_n(\theta_n, \phi_n(\theta_n)) = 0$. Then θ_n is strongly consistent for θ^0 and*

$$\sqrt{n}(\theta_n - \theta^0) \sim N(\mathbf{0}, \Sigma) \tag{4.58}$$

where

$$\Sigma^{-1} = E_{\alpha^0, \theta^0}(\dot{L}_n \dot{L}_n')$$

Moreover the GPML solution θ_n has asymptotic variance less than that in Andersons procedure under appropriate conditions.

4.5 Computational considerations

4.5.1 Existence of solution

One advantage of the GPML method as developed here is that with appropriate choice of the sample group, on whose basis the estimator for the normalizing constant is constructed, always ensures a solution, i.e., a finite solution, regardless of any hyperplane separability or other data singularity that causes the ML method to break down.

Thus if the estimator for the normalizing constant is constructed on the sample, which for a given β (i.e., in $\alpha + \beta x$) gives a greater estimate (in magnitude) than the other group, then a finite solution for the ML always exists. This is because the GPML equations are finite for $||\beta|| = \mathbf{0}$, and approach 0 as $||\beta|| \longrightarrow \infty$.

4.5.2 Methods of Estimation

For LGD we have two forms : LGDP given as follows

$$\alpha + \beta'\mathbf{x} - \log \phi(\beta), \quad \phi(\beta) = \int_{\mathbb{R}^m} e^{\beta'\mathbf{x}} f_2(\mathbf{x}) d\mathbf{x}, \quad \alpha = \log \frac{\pi_1}{\pi_2} \tag{4.59}$$

Case 1. $\phi(\beta)$ known as a function of β. Estimate $\hat{\alpha}$, $\hat{\beta}$ by minimizing

$$\prod_{i=1}^{n} \left(\frac{e^{\alpha+\beta'\mathbf{x}}}{\phi(\beta) + e^{\alpha+\beta'\mathbf{x}}} \right)^{n_1(\mathbf{x}_i)} \left(\frac{\phi(\beta)}{\phi(\beta) + e^{\alpha+\beta'\mathbf{x}}} \right)^{n_2(\mathbf{x}_i)} \tag{4.60}$$

Case 2. Form of $\phi(\beta)$ is unknown as a function of β, but some estimator $\hat{\phi}$ of ϕ is available. Estimate $\hat{\alpha}$, $\hat{\beta}$ by minimizing

$$\prod_{i=1}^{n}\left(\frac{e^{\alpha+\beta'\mathbf{x}}}{\hat{\phi}(\beta)+e^{\alpha+\beta'\mathbf{x}}}\right)^{n_1(\mathbf{x}_i)}\left(\frac{\hat{\phi}(\beta)}{\hat{\phi}(\beta)+e^{\alpha+\beta'\mathbf{x}}}\right)^{n_2(\mathbf{x}_i)} \tag{4.61}$$

The other form is as given by Anderson [5], et al. which we denote by LGDA, as follows :

$$\alpha+\beta'\mathbf{x} \tag{4.62}$$

Estimate $\hat{\alpha}$, $\hat{\beta}$ by minimizing

$$\prod_{i=1}^{n}\left(\frac{e^{\alpha+\beta'\mathbf{x}}}{1+e^{\alpha+\beta'\mathbf{x}}}\right)^{n_1(\mathbf{x}_i)}\left(\frac{1}{1+e^{\alpha+\beta'\mathbf{x}}}\right)^{n_2(\mathbf{x}_i)} \tag{4.63}$$

4.5.3 Simulation studies and AER

In the following tables we give in summarized form the results of simulation studies undertaken on distributions. The underlying relationship is assumed to be of the form $\ln f_1/f_2 = \alpha + \beta' x$ in the case of LGDA. We denote the LGDP as $\ln f_1/f_2 = \alpha - \ln\phi + \beta' x$ where ϕ is the appropriate normalizing constant, which is estimated.

Misclassification probabilities are in general difficult to compute as the estimates cannot be obtained in closed form. They need computation intensive estimation.

In the tables, α_P denotes $\alpha - \ln\phi$. We also consider apparent error rates, (AERs) i.e., estimated errors of misclassification when the estimated rule is used to classify observations from the training samples, (see McLachlan [72]). $\hat{\alpha}$ denotes estimated α and $\hat{\beta}$ denotes estimated β. The subscripts A and P denotes estimators corresponding to Anderson's [5], and the GPML (as developed by us) respectively.

Table I. (Normal, $\alpha = 0.5$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_A$ (average)	$\hat{\beta}_A$ (average)	AER (av)	No. of Iter.(av)
15	50	1.2	1.6	0.51	29
30	50	0.89	1.45	0.5	24
40	50	0.8	1.34	0.45	24
60	50	0.7	1.3	0.43	22

Table II. (Normal, $\alpha = 0.7$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_A$ (average)	$\hat{\beta}_A$ (average)	AER (av)	No. of Iter.(av)
15	50	1.1	1.61	0.5	30
30	50	0.8	1.4	0.5	26
40	50	0.79	1.34	0.46	25
60	50	0.79	1.3	0.44	24.2

Table III. (Normal, $\alpha = 0.7$, $\beta = 1.2$)

Sample Size	No. of Runs	$\hat{\alpha}_A$ (average)	$\hat{\beta}_A$ (average)	AER (av)	No. of Iter.(av)
15	50	1.22	1.62	0.52	30
30	50	0.9	1.42	0.5	26
40	50	0.82	1.4	0.45	24
60	50	0.71	1.4	0.43	23

Table IP. (Normal, $\alpha = 0.5$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_P$ (average)	$\hat{\beta}_P$ (average)	AER (av)	No. of Iter.(av)
15	50	0.8	1.4	0.41	21
30	50	0.7	1.25	0.39	17
40	50	0.56	1.12	0.35	12
60	50	0.5	1.1	0.35	11

Table IIP. (Normal, $\alpha = 0.7$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_P$ (average)	$\hat{\beta}_P$ (average)	AER (av)	No. of Iter.(av)
15	50	0.9	1.45	0.41	21
30	50	0.8	1.26	0.4	17
40	50	0.7	1.13	0.35	13
60	50	0.7	1.1	0.35	12

Table IIIP. (Normal, $\alpha = 0.7$, $\beta = 1.2$)

Sample Size	No. of Runs	$\hat{\alpha}_P$ (average)	$\hat{\beta}_P$ (average)	AER (av)	No. of Iter.(av)
15	50	0.8	1.52	0.43	20
30	50	0.7	1.4	0.4	16
40	50	0.55	1.35	0.4	13
60	50	0.5	1.33	0.33	11

Table I. (Exponential, $\alpha = 0.5$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_A$ (average)	$\hat{\beta}_A$ (average)	AER (av)	No. of Iter.(av)
15	50	1.9	1.8	0.6	30
30	50	1.2	1.6	0.5	24
40	50	0.9	1.5	0.45	24
60	50	0.7	1.5	0.43	22

Table II. (Exponential, $\alpha = 0.7$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_A$ (average)	$\hat{\beta}_A$ (average)	AER (av)	No. of Iter.(av)
15	50	1.3	1.6	0.57	29
30	50	1.2	1.45	0.52	26
40	50	1.2	1.4	0.48	25
60	50	0.79	1.35	0.46	24

Table III. (Exponential, $\alpha = 0.7$, $\beta = 1.2$)

Sample Size	No. of Runs	$\hat{\alpha}_A$ (average)	$\hat{\beta}_A$ (average)	AER (av)	No. of Iter.(av)
15	50	0.8	1.72	0.5	29
30	50	0.6	1.64	0.4	25
40	50	0.5	1.6	0.4	24
60	50	0.5	1.5	0.37	19

Table IP. (Exponential, $\alpha = 0.5$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_P$ (average)	$\hat{\beta}_P$ (average)	AER (av)	No. of Iter.(av)
15	50	0.8	1.4	0.41	21
30	50	0.7	1.25	0.39	17
40	50	0.56	1.12	0.35	12
60	50	0.5	1.1	0.35	11

Table IIP. (Exponential, $\alpha = 0.7$, $\beta = 1$)

Sample Size	No. of Runs	$\hat{\alpha}_P$ (average)	$\hat{\beta}_P$ (average)	AER (av)	No. of Iter.(av)
15	50	0.9	1.45	0.41	21
30	50	0.8	1.26	0.4	17
40	50	0.7	1.13	0.35	13
60	50	0.7	1.1	0.35	12

Table IIIP. (Exponential, $\alpha = 0.7$, $\beta = 1.2$)

Sample Size	No. of Runs	$\hat{\alpha}_P$ (average)	$\hat{\beta}_P$ (average)	AER (av)	No. of Iter.(av)
15	50	0.95	1.7	0.45	17
30	50	0.84	1.3	0.45	15
40	50	0.72	1.3	0.4	14
60	50	0.71	1.24	0.4	10

4.6 Real life examples

Applying GPML on some of the real life examples cited by Cox [27] the results are as follows :

(i) For the data referred to in table 4 of Cox [1962],[26] obtains $\hat{\alpha} = 0.413$ and $\hat{\beta} = 0.904$. GPML yields $\hat{\alpha} = 0.4$ and $\hat{\beta} = 0.94$.

(ii) For the data referred to in table 1 of Hodges [1958],[52] and referred to in Cox, ([27])) obtains $\hat{\beta} = 0.62$. GPML yields $\hat{\beta} = 0.64$. Hodges obtains for table 2 in the same paper, $\hat{\beta} = 1.21$. GPML yields $\hat{\beta} = 1.22$.

Although the comparisons have been made under the assumption that the underlying distributions admit LGD, initial tests to check admission have not been carried out. The "closeness" of the stimators may be due to poor separation of the two groups as also departures from the criteria for admitting LGD.

Chapter 5

GPML through Amarts

5.1 Introduction

In this chapter, we introduce amarts and exhibit that the results obtained in the earlier chapters can be concisely and elegantly represented in terms of amarts. Towards this end we recall below certain properties and results on amarts. For further details, see Edgar and Sucheston ([35]).

5.1.1 Amarts : definitions and results

Let D denote either N or $-N$, i.e. the set of positive or negative integers. Let $\{\mathcal{F}_n\}_{n\in D}$ be an increasing family of sub-sigma algebras of $\mathcal{F}$. The set of bounded stopping times will be denoted by T_D.

Definition 5.1.1 *Let $\{X_n\}_{n\in D}$ be an integrable family of random variables which is adapted to $\{\mathcal{F}_n\}_{n\in D}$. We call $\{X_n, \mathcal{F}_n\}_{n\in D}$ an amart iff the net $(E\, X_\tau)_{\tau\in T}$ is convergent.*

Definition 5.1.2 *Let $\{X_n\}_{n\in D}$ be adapted to $\{\mathcal{F}_n\}_{n\in D}$. We say that $\{X_n\}_{n\in D}$ is T-uniformly integrable if the set $\{X_\tau\}_{\tau\in T}$ is uniformly integrable, i.e. if for any given $\epsilon > 0$ there exists λ_0, such that*

$$\sup_{\tau} E\,|X_\tau|.I\{|X_\tau| > \lambda\} < \epsilon\, \forall\, \lambda > \lambda_0 \tag{5.1}$$

Remark Every T-uniformly integrable sequence is also uniformly integrable. Every uniformly integrable amart is T-uniformly integrable.

Lemma 5.1.1 *Let $\{X_n\}_{n\in D}$ be an amart. Then $E(X_\tau)_{\tau\in D}$ is bounded.*

Proposition 5.1.1 *A linear combination, maximum, minimum, and cutoffs of L^1 bounded amarts (in the case of $D = N$ are amarts. If $\sup_N E|X_n| < \infty$, then*

$$|X_n|,\ X_n^+,\ X_n^-,\ X_n.1_{[-\lambda,\lambda]}$$

are L^1 bounded amarts.

$$\sup_T E|X_\tau| < \infty, \ \ \sup |X_n| < \infty, \ a.e.$$

Proposition 5.1.2 *Let $\{X_n\}_{n\in D}$ be an amart for the increasing family $\{\mathcal{F}_n\}_{n\in D}$ of σ-algebras. Let $\{\mathcal{G}_n\}_{n\in D}$ be another increasing family of σ-algebras with $\mathcal{G}_n \subset \mathcal{F}_n \forall n \in D$. Then Y_n $E(X_n|\mathcal{G}_n)$ is an amart for $\mathcal{G}_n$.*

Proposition 5.1.3 *Let $\{Z_n\}_{n\in D}$ be an amart for $\{\mathcal{F}_n\}_{n\in D}$, such that $\lim_{n\to\infty} E((Z_n|\mathcal{F}_m) = 0$, a.e. for all $m \in N$. Then:*

$$E \sup_n |E(Z_n|\mathcal{F}_m)| < \infty \ \forall\, m$$

$$E(Z_n|\mathcal{F}_m) \longrightarrow 0 \ in\ L^1 \,\forall\, m$$

$$\lim_{\tau\in T} E|Z_\tau| = 0$$

$$Z_n \longrightarrow 0 \ a.e.\ and\ in\ L^1$$

$$(Z_\tau)_{\tau\in T} \text{ is uniformly integrable}$$

Proposition 5.1.4 *Suppose $X_n = \sum_{i=1}^n Y_i$ is an amart. Assume that*

$$\sup_i E(Y_i^2) < \infty$$

Then $(1/n)\, X_n$ converges a.e..

Result 5.1.1 *Every martingale is an amart.*

Result 5.1.2 *If $\{X_n, \mathcal{F}_n\}_{n\in D}$ is an amart, then $\{|X_n|, \mathcal{F}_n\}_{n\in D}$ is also an amart.*

Result 5.1.3 *Let $\{X_n\}_{n\in D}$ be an adapted sequence and suppose that $E \sup_n |X_n| < \infty$. The following statements are equivalent:*
(i) X_n converges a.s. as $|n| \to \infty$.
(ii) $\{X_n\}_{n\in D}$ is an amart.

Result 5.1.4 *Let $\{X_n\}_{n\in D}$ be an adapted sequence and T-uniformly integrable. The following statements are equivalent:*
(i) X_n converges a.s. as $|n| \to \infty$.
(ii) $\{X_n\}_{n\in D}$ is an amart.

Result 5.1.5 *Let $\{X_n, \mathcal{F}_n\}_{n\in D}$ be an amart. If $D = N$ assume in addition, that $\{X_n\}_{n\in N}$ is L^1-bounded. Let $\phi : R \to R$ be a function such that*
(i) ϕ is continuous and
(ii) $\lim_{x\to\infty} \frac{\phi(x)}{x}$ and $\lim_{x\to-\infty} \frac{\phi(x)}{x}$ exist and are finite. Then, $\{\phi(X_n), \mathcal{F}_n\}_{n\in D}$ is an L^1-bounded amart.

Result 5.1.6 *(Riesz decomposition for amarts) Let $\{X_n, \mathcal{F}_n\}_{n\in N}$ be an amart. Then X_n can be uniquely written as $X_n = Y_n + Z_n$, where $\{Y_n, \mathcal{F}_n\}_{n\in N}$ is a martingale and $\{Z_n, \mathcal{F}_n\}_{n\in N}$ is a T-uniformly integrable amart, such that $Z_n \to 0$ a.s. and in L^1.*

Further, the unique martingale $\{Y_n, \mathcal{F}_n\}_{n\in N}$ is the L^1 limit of $\{E(X_m|\mathcal{F}_n)\}$ as $m \to \infty$.

5.2 PML with estimated functionals

5.2.1 The basic framework

Let $\{X_n(\theta, f(\theta)), \mathcal{F}_n\}_{n\in N}$ be a martingale, for all $\theta \in \Theta$. Let θ_n, $f_n(\theta_n)$ be $\{\mathcal{F}_n\}_{n\in N}$ adapted.

5.2.2 Convergence and related results

Theorem 5.2.1 *Let for all $\theta \in \Theta$,*

(i) $n(f_n(\theta) - f(\theta))$ be a $\mathcal{F}_n$ adapted martingale, with $E f_n(\theta) = f(\theta)$.

(ii) The following decomposition holds :

$$X_n(\theta, f_n(\theta)) - X_n(\theta, f(\theta)) = Y_n(\theta).(f_n(\theta) - f(\theta)) \qquad (5.2)$$

with Y_n an adapted sequence uniformly in θ.

(iii) There exists a $\mathcal{F}_n$ adapted martingale Z_n such that $|Y_n| < Z_n$ a.s., and $\lim_n E\, Z_n/n < \infty$.

Then $X_n(\theta, f_n(\theta))$ is an amart. Further, the difference of the unique martingale in the Riesz decomposition of the given amart from the martingale $X_n(\theta, f(\theta))$ is a martingale which a.s. converges to 0.

Proof: Since $X_n(\theta, f(\theta))$ is a martingale, and hence an amart by result (5.1.1). Now from (5.2)

$$Y_n(\theta).(f_n(\theta) - f(\theta)) = \frac{Y_n(\theta)}{n}.n(f_n(\theta) - f(\theta))$$

$$\leq \frac{Z_n(\theta)}{n}.n(f_n(\theta) - f(\theta))$$

But since Z_n is a martingale, Z_n/n converges by the strong law for martingales to its expectation which by (iii) is finite. This same argument repeated with $-Z_n$ as a lower bound, shows that $\frac{Y_n(\theta)}{n}$ a.s. bounded.

On the other hand $n(f_n(\theta) - f(\theta))$ being a martingale converges a.s. by the martingale convergence theorem. (We could also treat it as a particular case of amarts and hence deduce the above from the amart convergence theorem.).

Since the right hand side of (5.2) converges a.s., we have that the sum of the right hand side with a martingale (and hence an amart) also converges a.s. and hence that $X_n(\theta, f_n(\theta))$ converges a.s. as $|n| \to \infty$.

This also shows that $E(\sup_n |X_n|(\theta, f_n(\theta))) < \infty$. Hence by proposition (5.1.3), the proof of amart is completed.

Once the amart proof is complete, we have from the Riesz decomposition of the amart (result 5.1.6), that $X_n(f_n)$ can be uniquely decomposed as the sum of a martingale and a Doob potential. Taking the limits of conditional expectations as given in result (5.1.6), we have that the unique martingale in the Riesz decomposition is the sum of the martingale $X_n(f)$ and a potential bounded by a 0-sum martingale, and which therefore converges a.s. 0.

Theorem 5.2.2 *Let* $\{X_n(\theta), \mathcal{F}_n\}_{n\in D}$ *be an amart (martingale), uniformly in* θ. *Let there exist an adapted sequence* Y_n, *such that* $\{Y_n(\theta), \mathcal{F}_n\}_{n\in D}$ *is an amart, and*

$$\frac{\partial X_n(\theta)}{\partial\theta} \leq Y_n(\theta) \ a.s. \tag{5.3}$$

then $\{\frac{\partial X_n(\theta)}{\partial\theta}, \mathcal{F}_n\}_{n\in D}$ *is also an amart (martingale).*

Proof: The inequality in (5.3) is the condition for interchangeability of the partial derivative and the integral sign. From elementary properties of conditional expectations, we have

$$E\left(\frac{\partial X_n(\theta)}{\partial\theta}|\mathcal{F}_m\right) = \frac{\partial E\left(X_n(\theta)|\mathcal{F}_m\right)}{\partial\theta} = \frac{\partial(X_m(\theta)|\mathcal{F}_m)}{\partial\theta}$$

Hence if $X_n(\theta)$ is an amart (martingale) $\frac{\partial X_n(\theta)}{\partial\theta}$ is also an amart (martingale).

The above two stated theorems show that by replacing $X_n(\theta)$ in the above, by the log-likelihoods (as under GPML) with estimated functionals of parameters and their derivatives, the entire GPML theory developed so far can be expressed in terms of amarts.

Chapter 6

On Quantile Based Discrimination In Stable Distributions

6.1 Introduction.

We consider the problem of classification of an observation from one of two strictly α-stable populations ($\alpha \neq 1$), based on a training sample composed of observations (with known group inclusions) from the two populations.

We require the following adaptations from results in Zolotarev, [103].

Theorem 6.1.1 *For an α-stable ($\alpha \neq 1$) random variable Y with admissible parameter quadruple $(\alpha, \beta, \gamma, \lambda)$, there exists an unique representation such that*

$$Y(\alpha, \beta, \gamma, \lambda) \stackrel{d}{=} \lambda^{1/\alpha} Y(\alpha, \beta, 0, 1) + \lambda\gamma \tag{6.1}$$

Theorem 6.1.2 *Let $Y(\alpha, \beta_1, \gamma_1, \lambda_1)$ and $Y(\alpha, \beta_2, \gamma_2, \lambda_2)$ be two independent α-stable random variables. Then $Y(\alpha, \beta_1, \gamma_1, \lambda_1) - Y(\alpha, \beta_2, \gamma_2, \lambda_2)$ is distributed as $Y(\alpha, \beta, \gamma, \lambda)$ where $\lambda = \lambda_1 + \lambda_2$,$\lambda\beta = \lambda_1\beta_1 - \lambda_2\beta_2$,$\lambda\gamma = \lambda_1\gamma_1 - \lambda_2\gamma_2$*

6.2 Construction of the classification rule

$$\text{Let } d_{ij} = Y(\alpha, \beta_i, \gamma_i, \lambda_i) - Y(\alpha, \beta_j, \gamma_j, \lambda_j) \tag{6.2}$$

$$\text{Therefore } d_{ii} = Y(\alpha, 0, 0, 2\lambda_i) \tag{6.3}$$

$$d_{ij} = Y(\alpha, \frac{\lambda_i\beta_i - \lambda_j\beta_j}{\lambda_1 + \lambda_2}, \frac{\lambda_i\gamma_i - \lambda_j\gamma_j}{\lambda_1 + \lambda_2}, \lambda_1 + \lambda_2),\ i \neq j \tag{6.4}$$

For $\alpha \neq 1$, The tail probability of a standard stable distribution at 0, $F(0)$ has a closed form expression, i.e.,

$$F(0) = 1 - \frac{1}{2}(1 - \beta\frac{K(\alpha)}{\alpha}) \tag{6.5}$$

Let Y_i's be strictly stable. Let $\beta = (\lambda_1\beta_1 - \lambda_2\beta_2)/(\lambda_1 + \lambda_2)$. Then, by equation (6.1)

$$d_{ii} = (2\lambda_i)^{1/\alpha}Y(\alpha, 0, 0, 1),\ d_{ij} = (\lambda_1 + \lambda_2)^{1/\alpha}Y(\alpha, (j-i)\beta, 0, 1)$$

$$\text{Therefore } P(d_{ii} > 0) = P(Y(\alpha, 0, 0, 1) > 0) = \frac{1}{2}$$

$$\text{Let } p_{ij} = P(d_{ij} > 0). \text{ Then } p_{22} = \frac{1}{2} = p_{11} \tag{6.6}$$

$$P(d_{ij} > 0) = \frac{1}{2}(1 + (j-i)\beta\frac{K(\alpha)}{\alpha}) = p_{ij} \tag{6.7}$$

Note that $p_{12} + p_{21} = 1$.

6.2.1 The statistic $G_{n,i}$

Let, for a given real r.v. Z, $1_{[Z]}$ be the indicator function of the event $\{Z > 0\}$. Let $\{X_{ik},\ i = 1, 2; k = 1, \ldots, n_i\}$, be two samples from strictly α-stable distributions - with admissible parameter quadruples $(\alpha, \beta_i, 0, \lambda_i), i = 1, 2$. Let Y be a new observation (unclasified) from one of the two given populations. Define,

$$G_{n,i} = \frac{1}{n_i}\sum_{k=1}^{n_i} 1_{[X_{ik} - Y_{n+1}]} \tag{6.8}$$

The classification rule can now be formulated as follows:

$$\text{If } |\frac{1}{2} - G_{n,1}| < |\frac{1}{2} - G_{n,2}| \text{ assign to Population I}$$

o.w. assign to Population 2

A corresponding modification to account for inclusion probabilities is

$$\text{If } \pi_1 | \frac{1}{2} - G_{n,1} | < \pi_2 | \frac{1}{2} - G_{n,2} | \text{ assign to Population I}$$

$$\text{o.w. assign to Population 2}$$

Remark

At this point although nothing can be said about non-strictly stable distributions in general, a few simple manipulations with the integral representations of the distribution functions of stable distributions as given by Zolotarev, shows that the above procedure can be applied to the non strictly stable distributions given by $\alpha < 1$, $\beta < 1$ and $\alpha > 1$, i.e. the classes for which $p_{12} \neq 0.5$.

6.2.2 The distribution of the $G_{n,i}$

From equation (6.8), we have that each indicator in the defining equation follows a binomial distribution, taking the value 1 with probability p, where

$$p = E(1_{[X_{i1} - Y]}) \; i = 1, 2 \tag{6.9}$$

Note that p takes one of the values p_{ij}'s.

The exact distribution is found out by considering the distribution of the number of X_{ik}'s that are greater than Y.

$$P(n_i G_{n,i} = k) = \binom{n_i}{k} p^k (1-p)^{n_i - k} \tag{6.10}$$

$$P(|G_{n,i} - \frac{1}{2}| = h) = \binom{n_i}{j} (p^k (1-p)^{n_i - k} + (1-p)^k p^{n_i - k}), \; k = \left[\frac{n_i}{2}\right] + j \tag{6.11}$$

6.3 A Real Life Example

6.3.1 Description of the data set

The original data set provided to us by Prof. S.T. Rachev, consisted of share prices of a certain category, over a period of time, which are transformed by taking sequential differences of natural logarithms of each share price as is usually done in econometric studies of price phenomena (See Mittnik, Rachev [1993],[73]). The justification for dividing the set into two samples is on the basis of a "changepoint" from "bullish" to "bearish" tendencies.

6.3.2 Construction of the rule

A brief description of the SPLUS functions and program parts used is given below :

Total samples size is 614. nsamp1 gives the breakpoint at which the sample is partitioned. This is set to be 320. prlog is the transformed data set which is created by taking the successive differences of the log of the input vector of raw data.

```
f <- function(u,sam,nsam)
{q <- length(sam[sam > u])/nsam
q}
```

This actually gives the empirical tail probability at u based on sample sam . p1, p2 are estimates of population inclusion probabilities.

```
fq <- function(x,d1,d2)
{q1 <- f(x,samp1,nsamp1)

q2 <- f(x,samp2,nsamp2)
val <- (d1*abs(q1-0.5)) - (d2*abs(q2-0.5))
val}
```

The function fq actually constructs the rule which is of the form

$$r(x) = d1.|q_1(x) - \frac{1}{2}| - d2.|q_2(x) - \frac{1}{2}| \tag{6.12}$$

The first plot (Figure 3) shows a comparison of the constructed rule with and without considering inclusion probabilities. The lighter line, Rule 1, shows the rule without and the darker one, Rule 2, shows it with the inclusion probabilities.

The second plot (Figure 4) shows a comparison of the log-tails of the two empirical distribution functions. From a well known result in stable distributions, (See Samorodnitsky and Taqqu, [1994],[88]) the negative of the log-tail asymptotically goes to the index α of the underlying α-stable distribution. This shows a difference in the indices of the two distributions. The lighter line shows the first sample and the darker one shows the second sample.

Rough estimates were obtained by considering the upper extreme 1/10-th of each sample.

AERs were obtained using `esterr` function. This function first computes samplewise error estimates `prop1`, `prop2` and then combines them to obtain the combined sample AER (`prop3`).

6.3.3 Further comparison w.r.t. sample homogeneity

Incorporating a C-routine supplied by Prof. S.T.Rachev ([80]) to estimate the parameter quadruples of a stable distribution, a C-routine developed by the author, calculates the AER obtained from the rule constructed, and the corresponding parameter estimates for the assumed underlying stable populations, given the choice of subsamples (input by the user).

Rachev's estimator fails for certain ranges, for which however the rule can still be constructed but no comparisons can be made as to sample homogeneity. Incidentally some of these regions appear to give the lowest AER's.

Varying these subsample choices, we report the following : (c is the transformed "scale" factor, ($\lambda^{1/\alpha}$ of Zolotarev [103]) and δ is the transformed "shift", as in Rachev, [80].

<u>First pair of subsamples</u>

subsample-1 lower limit = 288 (i.e., starting from data no.288 of original set), upper limit = 429

subsample-2 lower limit = 430, upper limit = 500

error estimate for 1-st sample = 0.414286, for 2-nd sample = 0.528571 ,for combined sample = 0.312796.

parameter estimates for subsample-1

$\alpha = 1.694558$,$\beta = 0.227588$, $c = 0.021124$, $\delta = -0.006073$.

parameter estimates for subsample-2

$\alpha = 1.641819$, $\beta = -0.094336$, $c = 0.017767$, $\delta = -0.002327$.

<u>Second pair of subsamples</u>

subsample-1 lower limit = 250, upper limit = 450

subsample-2 lower limit = 451, upper limit = 600

error estimate for 1-st sample = 0.483221, for 2-nd sample = 0.469799 ,for combined sample = 0.406877.

parameter estimates for subsample-1

$\alpha = 1.585061$, $\beta = -0.191576$, $c = 0.017243$,$\delta = -0.003982$.

parameter estimates for subsample-2
$\alpha = 1.637159$, $\beta = -0.270756$, $c = 0.016563$, $\delta = 0.001094$.

6.3.4 Conclusions

Shifting the breakpoint back from 320 to 200 reduces the AER to about 0.33, which combined with the corresponding graphical comparison of the indices of the two subsamples, indicates that retrogression of the breakpoint reduces the index heterogeneity between the two groups and hence the rule performs better.

6.4 Fisher Type Discrimination rules

Since, in general closed form expressions for the stable densities in terms of elementary functions are not available, it is difficult to visualize density ratio type discriminant functions. However, Zolotarev [103] gives an analytical form containing definite integrals, which can be used as basis for constructing approximations to the actual density ratios. This method, though comptation intensive, suffers from theoretical difficulties - notably absence of conclusive results about the existence of expectations of derivatives of log-densities of stable distributions.

From the computational viewpoint, using the GPML framework as developed by us, we can also incorporate this density-ratio in a LGD-type GPML framework.

6.4.1 Representation of Stable densities

The following representations are from Zolotarev [103] :

Theorem 6.4.1 *Let*

$$\epsilon(\alpha) = sgn(1-\alpha), \theta = \beta\frac{K(\alpha)}{\alpha}, \theta^* = \theta \; sgn \; x$$

Let,

$$U_\alpha(\phi,\theta) = \left(\frac{\sin(\frac{\pi\alpha}{2})(\phi+\theta)}{\cos\frac{\pi\phi}{2}}\right)^{\frac{\alpha}{1-\alpha}} . \frac{\cos\frac{\pi}{2}((\alpha-1)\phi+\alpha\theta)}{\cos\frac{\pi\phi}{2}}$$

$$U_1(\phi,\beta) = \frac{\pi(1+\beta\phi)}{2\cos\frac{\pi\phi}{2}}.\exp(\frac{\pi}{2}(\phi+\frac{1}{\beta})\tan\frac{\pi\alpha}{2})$$

Then the densities of standard stable distributions can be written as follows:

1) If $\alpha \neq 1, |\beta| \leq 1, x \neq 0$ *then,*

$$g(x,\alpha,\beta) = \frac{\alpha|x|^{\frac{1}{\alpha-1}}}{2|1-\alpha|}.\int_{-\theta^*}^{1} U_\alpha(\phi,\theta^*)\exp(-|x|^{\frac{\alpha}{\alpha-1}}U_\alpha(\phi,\theta^*))d\phi \qquad (6.13)$$

2) If $\alpha = 1, |\beta| \neq 0, \forall x,$

$$g(x,1,\beta) = \frac{e^{-\frac{x}{\beta}}}{2|\beta|}.\int_{-1}^{1} U_1(\phi,\beta)\exp(-e^{-\frac{x}{\beta}}U_1(\phi,\beta))d\phi \qquad (6.14)$$

3)$\alpha \neq 1, x = 0,$

$$g(0,\alpha,\beta) = \frac{1}{\pi}\Gamma(1+\frac{1}{\alpha}\cos(\frac{\pi\beta K(\alpha)}{\alpha}) \qquad (6.15)$$

4) $\alpha = 1, \beta = 0$ *corresponds to the Cauchy distribution.*

6.4.2 Construction of the Discrimination rules

Let $h(x;\alpha_1,\beta_1,\alpha_2,\beta_2) \equiv h(x;\underline{\alpha},\underline{\beta}) \equiv h(x)$

$$h(x) = \frac{g(x,\alpha_1,\beta_1)}{g(x,\alpha_2,\beta_2)}$$

$$\underline{\alpha} = \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}\ \underline{\beta} = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}\ \underline{b} = \begin{pmatrix}b\\ b\end{pmatrix}\ ,\ b \in I\!R$$

We first consider standard stable distributions and classify them into three major groups by the form of their density representations under $h(x)$ within these groups.

Next we replace the integrals in the numerator and denominator by composite quadrature formulae i.e., divide the range of integration into several subintervals and applying a simple quadrature formula to each of them.

If the underlying distributional parameters are already known we simply substitute these values in the computed ratios, or use estimated values of the parameters from the sample derived from either empirical c.f.s or other methods.

Since the quadrature formulae occurring in the numerator and denominator can be computed with arbitrary accuracy, we have a simple though computationally intensive method of discrimination in stable distributions. It thus only remains to present $h(x)$ explicitly.

Denote

1) for $\alpha \neq 1, x \neq 0, |\beta| \leq 1$

$$V_\alpha(\phi) = U_\alpha(\phi, \theta^\star) \exp(-|x|^{\frac{\alpha}{(\alpha-1)}} U_\alpha(\phi, \theta^\star)) \tag{6.16}$$

where, $-\theta^\star \leq \phi \leq 1$

2) for $\alpha = 1, \beta \neq 0, \forall x,$

$$V_1(\phi) = U_1(\phi, \beta) \exp(-e^{-\frac{x}{\beta}} U_1(\phi, \beta)) \tag{6.17}$$

We sometimes use the notation V_α or V_1 for short , equivalently.

Denote :

$$I_{\alpha,\theta^\star}(x) \equiv \int_{-\theta^\star}^{1} V_\alpha(\phi) d\phi, \; I_{1,\beta}(x) \equiv \int_{-1}^{1} V_1(\phi) d\phi$$

We now present $h(x)$ explicitly. Note that several cases are possible:

1) $\underline{\alpha} \neq \underline{1}$, $x \neq 0$, $|\beta_1| \leq 1$, $|\beta_2| \leq 1$

$$h_1(x) = \begin{cases} \frac{\alpha_1|1-\alpha_2|}{\alpha_2|1-\alpha_1|} |x|^{\frac{\alpha_1}{\alpha_1-1} - \frac{\alpha_2}{\alpha_2-1}} . \frac{I_{\alpha_1,\theta_1^\star}(x)}{I_{\alpha_2,\theta_2^\star}(x)} & \text{if } x \neq 0 \\ \frac{\Gamma(1+\frac{1}{\alpha_1}) \cos(\frac{\pi}{2}\theta_1)}{\Gamma(1+\frac{1}{\alpha_2}) \cos(\frac{\pi}{2}\theta_2)} & \text{if } x = o \end{cases} \tag{6.18}$$

2) $\underline{\alpha} = \underline{1}$, $\underline{\beta} \neq \underline{0}$, $\forall x$

$$h_2(x) = \frac{|\beta_2|}{|\beta_1|} e^{-x(\frac{1}{\beta_1} - \frac{1}{\beta_2})} . \frac{I_{1,\beta_1}(x)}{I_{1,\beta_2}(x)} \tag{6.19}$$

3) $\alpha_1 \neq 1$, $\alpha_2 = 1$, $|\beta_1| \leq 1$, $\beta_2 \neq 0$

$$h_3(x) = \frac{\alpha_1|\beta_2|}{|1-\alpha_1|} . |x|^{\frac{1}{\alpha_1-1}} . e^{\frac{x}{\beta_2}} \frac{I_{\alpha_1,\theta_1^\star}(x)}{I_{1,\beta_2}(x)} \text{ if } x \neq 0 \tag{6.20}$$

$$= \frac{1}{2\pi|\beta_2|}\Gamma\left(1+\frac{1}{\alpha_1}\right)\cos\left(\frac{\pi}{2}\theta_1\right)\frac{1}{I_{1,\beta_2}(0)} \text{ if } x=0 \tag{6.21}$$

Remarks

We observe that by using the form of the ratio of the densities, and estimating in the usual GPML iterative schemes, we can arrive at estimators of stable parameter quadruples as well as the rule simultaneously.

Chapter 7

Directional data

7.1 Introduction

Suppose we have observations as directional data [see e.g., Mardia [71]; Rao (currently Jammalamadaka, S.R.) [84]] from two distinct (identifiable) populations on the unit circle. We need to classify a new observation as belonging to one of these two populations, using the given data as needed.

Let the past observations be denoted by

$$\theta_{ij}, \quad i = 1, 2, \; j = 1, \ldots, n_i$$

Let the new observation be denoted by θ. Assume that the sample mean directions are given by $\overline{\theta}_i$, $i = 1, 2$.

In the usual linear setup, for the multivariate or univariate Normal distribution the Fisher type discriminant (which coincides with the LGD for same variances and the Quadratic LGD (Anderson [3]) if variances are different) can be viewed as a quadratic distance function i.e., with variance-covariance matrix playing the role of the metric tensor. As we show in the folowing, a similar coincidence occurs in a class of directional distributions.

The basic idea is to find out the average "distance" (in an appropriate sense) from the new observation to the observations in the two known groups. If the distance from one group is less than from the other, then the new observation is classified as belonging to the "closer" population. These ideas formed the nucleus of an invited lecture (Roy [87]), and were later elaborated by us.

7.1.1 The distance measure

The simplest distance that can be used is the arc-length, which in the case of the unit circle is equivalent to the radian measure subtended at the center of the circle, i.e., the value of the observation in radians.

But to be a proper distance on the circle, the distance measure must be rotationally invariant, both in terms of magnitude as well as sense of rotation. Thus if we have to consider the arc-length in terms of radian measure, we have to transform it in a suitable way, i.e., take absolute value of the difference in angles, modulus 2π. We may also have to consider the minimum of the two arclengths into which two points on the circle divides a circle.

These problems do not arise if instead of the arc-length we consider the length of the chord cut off by the two points on the circle. This is always non-negative, invariant under rotation, both in magnitude and displacement. As we shall see, this particular form has also other attractive properties due to its similarity to known descriptive measures in circular distributions.

We appreciate the remarks of a Referee and observe that the use of chord length as a descriptive measure is quite natural and may have been in use for long. However, our extensive search for a chord-based classification rule in the prevailing literature proved futile and thus, to our knowledge, the approach in the following section seems to be the maiden attempt in this direction.

7.2 Construction of the rule

7.2.1 Properties of the chord length as a distance measure

Let two points on the unit circle be denoted by θ_1, θ_2. Then the square of the chord-length between the two is given by $2\,(1 - \cos(\theta_1 - \theta_2))$. Based on this we take the distance measure as

$$d_{ij} = 1 - \cos(\theta_i - \theta_j) \tag{7.1}$$

Note that d_{ij} has the following properties : It is always non-negative, symmetric in its indices and is invariant under rotation.

7.2.2 The average distance of a point from a group

The average distance $d_i(\theta)$ of θ from the group i, is given by

$$d_i(\theta) = 1 - \frac{1}{n_i}\sum_j \cos(\theta_{ij} - \theta) \tag{7.2}$$

Note that this is similar to the sample circular variance with a shift in the mean direction. Let

$$\overline{C}_i = \frac{1}{n_i}\sum_j \cos\theta_{ij},\ \overline{S}_i = \frac{1}{n_i}\sum_j \sin\theta_{ij},\ R_i = \sqrt{\overline{C}_i^2 + \overline{S}_i^2},\ \tan\overline{\theta}_i = \frac{\overline{S}_i}{\overline{C}_i}$$

7.2.3 The rule

Let the new observation to classify be θ. Let d_{0i} be the distance of θ from the group i circular mean θ_i. Define $D(\theta) = d_{01}(\theta) - d_{02}(\theta)$. Let c be areal constant. The classification rule is given by

$$\text{If } \quad D(\theta) > c \quad \text{assign to Population 1}$$

$$\text{assign to population 2 o.w.} \tag{7.3}$$

$$\text{Now } D(\theta) = (\cos(\bar{\theta}_2) - \cos(\bar{\theta}_1))\cos\theta + (\sin(\bar{\theta}_2) - \sin(\bar{\theta}_1))\sin\theta \tag{7.4}$$

$$\text{Let } \tan\theta_0 = \frac{\sin(\bar{\theta}_2) - \sin(\bar{\theta}_1)}{\cos(\bar{\theta}_2) - \cos(\bar{\theta}_1)} \tag{7.5}$$

Then equation (7.4) can be written as

$$D(\theta) = \sqrt{2 - 2\cos(\bar{\theta}_1 - \bar{\theta}_2)}\cos(\theta - \theta_0) \tag{7.6}$$

Note that by equation (7.5), there will be two sloutions for θ_0.

The classification rule (7.3) can now be now given in an equivalent form as

$$\text{If } \quad \cos(\theta - \theta_0) > c' \quad \text{assign to Population 1}$$

$$\text{assign to population 2 o.w.} \tag{7.7}$$

where c' is an appropriate constant.

The rule as given by equation (7.7) simply partitions the circle into sectors (of width 180° if $c = 0$. In this case, explicitly, the sectors can be specified as one semicircle having θ_0 as its midpoint, and the complementary arc.

Remark 1. Note that if the sample mean directions are equal, unequal variances have no effect on the rule. In this case θ_0 is simply the mean direction itself.

Remark 2. When the sample mean directions are not equal, the variances affect θ_0.

7.3 A modification of the rule

For a brief review of the background of this section, see Mardia,[1973]([71]). The $A(\kappa)$ as used here is the ratio of the Bessel functions $I_0(\kappa)$ and $I_1(\kappa)$.

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$$

For a complete derivation of $A(\kappa)$, its relevance for the Von-Mises distribution, and some useful properties, see Mardia,[1973] as referred to above. In fact for the Von Mises population, R is asymptotically $A(\kappa)$.

Note that $1/2\kappa R^2$ approximates the integral $\int A(\kappa)\mathrm{d}\kappa$ by a triangle. A better approximation can be obtained by directly considering the difference $\int_{\kappa_1}^{\kappa_2} A(\kappa)\mathrm{d}\kappa$ which can be taken in terms of the trapezoidal rule as $1/2\ (A(\kappa_1) + A(\kappa_2))(\kappa_2 - \kappa_1)$.

Let $V_1 = D(\bar{\theta}_1, V_2 = D(\bar{\theta}_2$, i.e., V_i is the average intragroup "distance" from each other of observations in group or sample i. Note that V_i is nothing but the sample circular variance for sample i. Define the intragroup average d_{ii} from the sample mean direction as

$$d_{ii} = 1 - \frac{1}{n_i}\sum \cos(\theta_i - \theta_0) \tag{7.8}$$

Note that $d_{ii} = 1 - R_i = V_i$.

$$\text{Let} \quad D_1(\theta) = \alpha_1(d_{01}(\theta) - \frac{d_{11} + d_{22}}{2}) - \alpha_2(d_{02} - \frac{d_{11} + d_{22}}{2}) + \beta$$

The classification rule is given by

$$\text{If} \quad D_1(\theta) > 0 \quad \text{assign to Population 1}$$

$$\text{assign to population 2 o.w.} \tag{7.9}$$

Now $D_1(\theta)$ reduces to

$$(\alpha_2 \cos(\bar{\theta}_2) - \alpha_1 \cos(\bar{\theta}_1)) \cos\theta + (\alpha_2 \sin(\bar{\theta}_2) - \alpha_1 \sin(\bar{\theta}_1)) \sin\theta$$

$$+ \frac{1}{2}(\alpha_1 - \alpha_2)(R_1 + R_2) \tag{7.10}$$

Let

$$\tan\theta_0 = \frac{\alpha_2 \sin(\bar{\theta}_2) - \alpha_1 \sin(\bar{\theta}_1))}{\alpha_2 \cos(\bar{\theta}_2) - \alpha_1 \cos(\bar{\theta}_1)} \tag{7.11}$$

Note that by equation (7.11), there will be two sloutions for θ_0.

Assuming an underlying population of the Von Mises family, (i.e., $M(\kappa_1, \mu_1)$, and $M(\kappa_2, \mu_2)$, we note that given the parameters, the standard Fisher type rule would have the form

$$\ln I_0(\kappa_1) - \ln I_0(\kappa_2) + (\kappa_1 \cos(\mu_1) - \kappa_2 \cos(\mu_2)) \cos(\theta)$$

$$+ (\kappa_1 \sin(\mu_1) - \kappa_2 \sin(\mu_2)) \sin(\theta) + \beta \tag{7.12}$$

$$\ln I_0(\kappa_1) - \ln I_0(\kappa_2) + \sqrt{\kappa_1^2 + \kappa_2^2 - 2\kappa_1\kappa_2 \cos(\mu_1 - \mu_2)}$$

$$\times \cos(\theta - \tan^{-1}\frac{\kappa_2 \sin\mu_2 - \kappa_1 \sin\mu_1}{\kappa_2 \cos\mu_2 - \kappa_1 \cos\mu_1}) + \beta \tag{7.13}$$

Putting $\alpha_i = \kappa_i$ in equation (7.10) and observing that

$$\frac{d}{d\kappa} \ln I_0(\kappa) = A(\kappa)$$

we have,

$$\ln I_0(\kappa) = \int A(\kappa) d\kappa$$

Note also that

$$\frac{d}{d\kappa} A(\kappa) = 1 - A^2(\kappa) - \frac{A(\kappa)}{\kappa}$$

and hence that for small change in κ, the order of change in $A(\kappa)$ is less than that of κ. Therefore,

$$\ln I_0(\kappa_1) - \ln I_0(\kappa_2) = \int_{\kappa_1}^{\kappa_2} A(\kappa) d\kappa \sim \frac{1}{2}[A(\kappa_1) + A(\kappa_2)](\kappa_2 - \kappa_1)$$

Note that, if κ_1 and κ_2 are very close to each other, asymptotically equation (7.10) approximates (strongly converges) to the corresponding portion of equation (7.13). Thus although we have kept the rule flexible by introducing the constants α_i's, a recommended choice is that which is found by substituting the pairs

$$\hat{\kappa}_i = A^{-1}(R_i); \quad A(\hat{\kappa}_i) = R_i$$

7.4 Exact distribution of $D(\theta)$

The distribution of $\bar{\theta}$ conditional on R is Von Mises with mean direction μ and concentration parameter κR. The joint distribution of $C = R\cos\bar{\theta}, S = R\sin\bar{\theta}$ is given by

$$f(C,S) = \frac{1}{I_0^n(\kappa)} e^{\kappa(\cos(\mu)C+\sin(\mu)S)}\phi_n(C^2+S^2) \tag{7.14}$$

Here ϕ_n is the uniform distribution on the n--dimensional hypersphere [71]. The joint distribution of is $U = \alpha_1\cos\bar{\theta}_1 - \alpha_2\cos\bar{\theta}_2$, $V = \alpha_1\sin\bar{\theta}_1 - \alpha_2\sin\bar{\theta}_2$, given $R_1, R_2, \alpha_1, \alpha_2$ is given by

$$f(U,V) = \frac{e^{\kappa_1 R_1\cos(\mu_1)\frac{U}{\alpha_1}+\kappa_2 R_2\cos(\mu_2)\frac{V}{\alpha_2}}}{(2\pi)^2 I_0(\kappa_1 R_1) I_0(\kappa_2 R_2)}$$

$$\times \int_{\bar{\theta}_2} e^{(\kappa_1 R_1\cos(\mu_1)\frac{\alpha_2}{\alpha_1}+\kappa_2 R_2\cos(\mu_2))\cos(\bar{\theta}_2)+(\kappa_1 R_1\sin(\mu_1)\frac{\alpha_2}{\alpha_1}+\kappa_2 R_2\sin(\mu_2))\sin(\bar{\theta}_2)}$$

Combining this with equation 7.14, we have the joint distribution of (C,S,U,V) (where $\cos(\theta) = C$ and $\sin(\theta) = S$, given $R_1, R_2, \alpha_1, \alpha_2$ to be

$$f(C,S,U,V) = \frac{e^{\kappa_1 R_1\cos(\mu_1)\frac{U}{\alpha_1}+\kappa_2 R_2\cos(\mu_2)\frac{V}{\alpha_2}+\kappa\cos(\theta-\mu)}}{(2\pi)^3 I_0(\kappa) I_0(\kappa_1 R_1) I_0(\kappa_2 R_2)}$$

$$\times\phi_1(1)\int_{\bar{\theta}_2} e^{(\kappa_1 R_1\cos(\mu_1)\frac{\alpha_2}{\alpha_1}+\kappa_2 R_2\cos(\mu_2))\cos(\theta_2)}$$

$$e^{(\kappa_1 R_1\sin(\mu_1)\frac{\alpha_2}{\alpha_1}+\kappa_2 R_2\sin(\mu_2))\sin(\theta_2)} \tag{7.15}$$

To get the distribution of the statistic, the conditional density in equation (7.15) multiplied by the joint distribution $h_{n_1}(R_1)h_{n_2}(R_2)$ (for definition of $h_n(R)$ see Mardia [71]) has to be integrated over regions of the form $d = aCU + (1-a)SV$.

7.5 Efficiency of the Rule

As is apparent, closed form expressions for error probabilities do not exist and the actual values have to be numerically computed for each pair of training samples.

In the following tables $\mu_1 = 0$. ERR_1 denotes the calculated error probability from the exact distribution of the modified statistic as given above, ERR_2 denotes the calculated error probability from the Fisher type ratio of density discrimination rule. AER_1 denotes the apparent error rate from the modified statistic as given above, AER_2 denotes the apparent error rate from the Fisher type ratio of density discrimination rule.

$n_1 = 10$											
n_2	μ_2	κ_1	κ_2	$\hat{\mu}_1$	$\hat{\mu}_2$	R_1	R_2	ERR_1	AER_1	ERR_2	AER_2
10	0.1	0.1	0.1	0.05	0.12	0.09	0.09	0.15	0.19	0.13	0.25
10	0.2	0.1	0.1	0.05	0.17	0.091	0.089	0.14	0.16	0.14	0.25
10	0.3	0.1	0.1	0.05	0.26	0.091	0.09	0.137	0.15	0.146	0.24
10	0.4	0.1	0.1	0.055	0.36	0.092	0.091	0.135	0.148	0.145	0.24
10	0.5	0.1	0.1	0.05	0.48	0.093	0.092	0.13	0.14	0.148	0.21

$n_1 = 10$											
n_2	μ_2	κ_1	κ_2	$\hat{\mu}_1$	$\hat{\mu}_2$	R_1	R_2	ERR_1	AER_1	ERR_2	AER_2
10	0.1	0.1	0.2	0.05	0.12	0.09	0.18	0.15	0.17	0.25	0.3
10	0.2	0.1	0.2	0.05	0.17	0.091	0.19	0.14	0.16	0.22	0.26
10	0.3	0.1	0.2	0.05	0.26	0.091	0.19	0.132	0.145	0.2	0.25
10	0.4	0.1	0.2	0.055	0.36	0.092	0.192	0.129	0.14	0.19	0.24
10	0.5	0.1	0.2	0.05	0.48	0.093	0.192	0.126	0.134	0.18	0.22

$n_1 = 10$											
n_2	μ_2	κ_1	κ_2	$\hat{\mu}_1$	$\hat{\mu}_2$	R_1	R_2	ERR_1	AER_1	ERR_2	AER_2
10	0.1	0.1	0.3	0.05	0.12	0.09	0.28	0.16	0.17	0.25	0.3
10	0.2	0.1	0.3	0.05	0.17	0.091	0.289	0.14	0.16	0.22	0.26
10	0.3	0.1	0.3	0.05	0.26	0.091	0.29	0.132	0.145	0.2	0.25
10	0.4	0.1	0.3	0.055	0.36	0.092	0.291	0.129	0.14	0.19	0.24
10	0.5	0.1	0.3	0.05	0.48	0.093	0.292	0.126	0.134	0.18	0.22

7.6 A Real-life Example

We now consider the data on pigeon-homing, as referred to in Mardia [71], pp 156-157, in which the internal clocks of 10 birds were reset by 6 hours clockwise while the clocks of 9 birds were left unaltered. Assuming that the underlying distributions are Von-Mises with equal concentration parameters (as in Mardia, [71] p 157), we run a SPLUS program that classifies each observation in the two samples on the basis of the remaining observations, by comparing the average chord-length distance from each group. The program is included in chapter 10.

The output is apparent error rate (AER), which is 0.0 for control group, 0.25 for experimental group, 0.117 for combined sample.

The apparent error rates (sample misclassification probabilities) show that the rule correctly classifies all the observations in the control group, and 75% in the experimental group.

7.7 Similarity with LGD

In the above discussion, the modified rule can easily be identified as a semi-parametric rule which approaches the Fisher type rule (ratio of densities) when the underlying populations are Von Mises and they are close to each other in terms of population parameters.

Since LGD models the ratio of densities, in the case, when the log ratio is linear, we find that a simple generalization of LGD in particular can be used to discriminate between two Von Mises populations, since the log-ratio in this case is linear on the sine and cosine transformation of θ. This also bypasses the rather computationally tricky problem of having to estimate κ, $A(\kappa)$ and their logarithms, as the constant term in the expression of the LGD subsumes all the Bessel function terms. This can also be approached from GPML. The rule as given above by us based on chord lengths, however need not assume independence of the linear components of the Logistic rule.

Chapter 8

E.c.f. based estimation

8.1 Introduction

One important consequence of the characterization of pairs of densities admitting and obeying LGD through c.f.s is that the nature of the characterization makes available well-developed procedures of estimation through empirical c.f.s (e.c.f). Although the literature mainly is concerned with estimation for stable distributions the theory follows through for other classes of distributions (see DuMouchel [1973][34], Feurvergher and McDunnough [1981] [38], Brockwell [1981][24]).

We start with Kellermeir's [1980] [59] work and proceed to construct e.c.f. based estimators for LGD parameters.

8.2 Some useful results

Given a sequence of stochastic processes Z_n and a function μ such that $\sqrt{n}\,(Z_n - \mu)$ converges weakly to a 0-mean process. Consider the statistics :

$$S_n \;=\; \sup_{t\in[a,b]} |Z_n\,(t)| \text{ and } T_n \;=\; \int |Z_n(t)|^2 \mathrm{d}G(t)$$

where G is a distribution function. Let $X_1, \ldots, X_n$ be i.i.d. e.d.f. is $F_n(x) = N(x)/N$, $x \in I\!R$, $N(x) \;=\; \#\{X_j \leq x\}$ for $1 \leq j \leq n$.

$$\text{e.c.f.} \;=\; c_n(t) \;=\; \int e^{itx}\mathrm{d}F_n(x) \;=\; \frac{1}{n}\sum e^{itX_j},\; t \in I\!R$$

Take $C^2(a,b]$ to be the Banach space of continuous complex functions on $[a,b]$ with the norm $||f||_\infty = \sup_{t\in(a,b)}|f(t)|$.

Result 8.2.1 *Let Z, $\{Z_n\}$ be random elements of $C^2(a,b]$ such that $\sqrt{n}Z_n \xrightarrow{D} Z$. Let $E \subset [a,b]$ and define*

$$S_n = \sup_{t\in E}|Z_n(t)|, and S = \sup_{t\in E}|Z(t)|$$

then $\sqrt{n}S_n \xrightarrow{D} S$.

Result 8.2.2 *Let Z, $\{Z_n\}$ be random elements of $C^2(a,b]$. Let $\mu \in C^2[a,b]$. such that $\forall t \in [a,b]$ $\mu(t) \neq 0$ and $\sqrt{n}(Z_n - \mu) \xrightarrow{D} Z$. then $\sqrt{n}(|Z_n| - |\mu|) \xrightarrow{D} Re(Z/\mu).|\mu|$.*

Result 8.2.3 *Let Z, $\{Z_n\}$ be random elements of $C^2(a,b]$. Let $\mu \in C^2[a,b]$ and $\sqrt{n}(Z_n - \mu) \xrightarrow{D} Z$. Suppose $\{t \in [a,b] \; |\mu(t) = ||\mu||_\infty\} = \{t_1,\dots,t_k\}$. then $\sqrt{n}(||Z_n||_\infty - \max_{1\le i\le k}|Z_n(t_i)|) \xrightarrow{P} 0$ as $n \longrightarrow \infty$.*

Result 8.2.4 *With the same notations and assumptions as in result (8.2.3), we have*

$$\sqrt{n}(||Z_n||_\infty - ||\mu||_\infty) \xrightarrow{D} \max_{1\le i\le k} Re\frac{Z(t_i)}{\mu(t_i)}||\mu||_\infty$$

Result 8.2.5 *Let $\{Z_n(t),\ t \in [0,1]\}$ be a sequence of real Stochastic Processses, such that $\sqrt{n}(Z_n - \mu) \xrightarrow{D} Z$, where μ is continuous and Z is a 0-mean Gaussian process in $[0,1]$ with covariance function $E\,Z(t)Z(s) = \Gamma(t,s)$.*

Let f be a continuously differentiable function with derivative f'. Let G be a distribution function with support in $[0,1]$. Define

$$T_n = \int f(Z_n(t))\,dG(t)$$

$$M = \int f(\mu(t))\,dG(t)$$

$$\sigma^2 = \int\int f'(\mu(t))f'(\mu(s))\Gamma(t,s)\,dG(t)\,dG(s)$$

and assume all integrals are well defined. Then $\sqrt{n}(T_n - M) \xrightarrow{D} \Xi$ where $\Xi \asymp N(0,\sigma^2)$ if $\sigma^2 \neq 0$ and Ξ is degenerate at 0 if $\sigma^2 = 0$.

Result 8.2.6 *A sufficient condition that result (8.2.5) holds for an arbitrary distribution function G is that* $\exists \alpha \geq 1$ *and* $Q : \mathbb{R} \longrightarrow [0, \infty)$ *such that*

$$\int Q(t)dG(t) < \infty \text{ and}$$

$$E|\sqrt{n}(f(Z_n(t)) - f(\mu(t)))|^\alpha \leq Q(t)\forall n$$

Result 8.2.7 *Let* $\{Y_n(t),\ t \in \mathbb{R}\}$, $\{Y(t),\ t \in \mathbb{R}\}$ *be real Stochastic Processes, such that* $\sqrt{n}Y_n \xrightarrow{D} Y$, *in every finite interval. Let* f *be a continuous function on* $\mathbb{R}$ *and let* G *be a distribution function. Define*

$$T_n = \int f(Y_n(t))dG(t)$$

$$M = \int f(Y(t))dG(t)$$

Assume T_n *and* T *are well defined with probability 1. Furthermore suppose* $\exists \alpha \geq 1$ *such that the sequence of functions* $E|(f(Y(t))|^\alpha$ *is uniformly integrable w.r.t. G. Then*

$$T_n \xrightarrow{D} T \text{ as } n \longrightarrow \infty$$

8.3 E.c.f theory for LGD

Note that from the form of the functional relationships as derived in chapter 2, we can have an e.c.f. based estimator for the LGD by minimizing a certain integral as follows :

$$\min_\beta L(\beta) = \int \left\| \hat{f}_2(t) - \frac{\hat{f}_1(t - \mathrm{i}\beta)}{\hat{f}_1(-\mathrm{i}\beta)} \right\| \mathrm{d}\omega(t) \ \text{ w.r.t } \beta \qquad (8.1)$$

where $\omega(t)$ is a distribution function, or equivalently,

$$\min_\beta L(\beta) = \int \left\| \hat{f}_1(t)\hat{f}_2(-\mathrm{i}\beta) - \hat{f}_2(t - \mathrm{i}\beta) \right\| \mathrm{d}\omega(t) \ \text{ w.r.t } \beta \qquad (8.2)$$

By theorem 2.1 of Feuerverger and Mureika [1977] ([38]), with probability 1, the e.c.f. converges uniformly to the characteristic function on every bounded interval.

The random complex process $Y_n(t) = \sqrt{n}(c_n(t) - c(t))$, in t has $E(Y_n(t)) = 0$, $E(Y_n(t_1)Y_n(t_2) = c(t_1 + t_2) - c(t_1)c(t_2)$. This latter term fully determines the covariance structure of $Y_n(t)$. Define $Y(t)$ to be a zero mean complex valued Gaussian process satisfying $\overline{Y}(t) = Y(-t)$ and having some covariance structure as $Y_n(t)$.

$$Cov\, Re\, Y(t_1), Re\, Y(t_2)) = \frac{1}{2}[Re\, c(t_1 + t_2) + Re\, c(t_1 - t_2)] - Re\, c(t_1) Re\, c(t_2)$$

$$Cov\, Re\, Y(t_1), Im\, Y(t_2)) = \frac{1}{2}[Im\, c(t_1 + t_2) + Im\, c(t_1 - t_2)] - Re\, c(t_1) Im\, c(t_2)$$

$$Cov\, Im\, Y(t_1), Im\, Y(t_2)) = \frac{1}{2}[-Re\, c(t_1 + t_2) + Re\, c(t_1 - t_2)] - Im\, c(t_1) Im\, c(t_2)$$

By theorem 3.1 of Feurverger and Mureika [38], the process $Y_n(t)$ converges weakly to $Y(t)$ in every finite interval.

Now let $\theta \in \Theta$ be an $l \times 1$ parameter and T_n be a $k \times 1$ statistic where $k \geq l$. Let the random implicit equation $F(\theta, T_n) = 0$, $F : I\!R^{l \times k} \rightarrow I\!R^l$ hold. Assume that F is continuously differentiable, and θ_0 denote the actual θ, and let Θ be an open rectangle. Then, we have from Feuerverger and McDunnough [37],

Result 8.3.1 *(a) If* $T_n \rightarrow \lambda(\theta_0), a.s.$, $F(\theta_0, \lambda(\theta_0)) = 0$, *and* $\frac{\partial F(\theta_0, \lambda(\theta_0))}{\partial \theta}$ *is invertible, then there exists* $\hat{\theta} \rightarrow \theta_0$, *a.s., which is an asymptotic random root for* F.

(b) If $T_n \rightarrow N(\lambda(\theta_0), \frac{\Sigma}{n})$, *a.s.,* $F(\theta_0, \lambda(\theta_0)) = 0$, *and* $\frac{\partial F(\theta_0, \lambda(\theta_0))}{\partial \theta}$ *is invertible, and* $\hat{\theta} \rightarrow \theta_0$, *in pr., is a root of* F, *then*

$$\hat{\theta}_n - \theta_0 \rightarrow N\left(0, \left[\frac{\partial F(\theta_0, \lambda(\theta_0))}{\partial \theta}\right]^{\prime -1} \frac{\Sigma}{n} \left[\frac{\partial F(\theta_0, \lambda(\theta_0))}{\partial \theta}\right]^{-1}\right) \quad (8.3)$$

The random implicit equation (8.2), is continuously differentiable as the integrand is a second degree polynomial in exponential and trigonometric functions. The second derivative of the integral at true value θ_0 is invertible, and hence by part (a) of result (8.3.1) we have that there exists a consistent root of equation (8.2). By part (b) of the same we have that this consistent root is also asymptotically Normal with mean 0 and variance covariance matrix given by

$$\Sigma^{-1} = L_{\beta\beta}$$

Chapter 9

Neural Networks

9.1 Introduction

LGD has been used in the recent years extensively in Artificial Neural Networks (ANN). The ANN approach towards pattern recognition and classification starts with the simplest models which basically use the linear discriminant function. This naturally lends itself to the LGD as a semi-parametric framework in which at least apparently "less" need be known about the underlying distributions and the network can be kept concerned only with the linear discriminant (coefficients) parameters of interest.

The standard practice is to assume an external trainer, which adjusts the different parameters to minimize some sort of penalty for misclassification. We show that by considering a more "natural" goal-seeking behaviour (brain-like) the LGD as used in statistical practice can be implemented as an ANN.

9.2 The ML-training of Gish

We first review, some results in posterior probability estimation through ANNs [50].

Let $x_i, i = 1, \ldots, N$, denote a set of feature vectors, and that each x_i belongs to either class C_1 or C_2. We consider the classifier network as being described by the mapping $f(x, \theta)$, which is parametrized by θ and maps each x_i onto the interval $[0, 1]$. Our goal is the selection of parameters θ

such that $f(x,\theta)$ is an estimate of $P_\theta(C_1|x)$, the probability that class C_1 has occurred given that we have observed x.

In order to model $f(x,\theta)$, i.e. the posterior probabilities, we can use the LGD approach, as follows :

The output of our network $f(x,\theta)$ will be a sigmoidal transformation of a function $z(x,\theta)$, given by the relation

$$f(x,\theta) = \frac{1}{1 + e^{-z(x,\theta)}} \tag{9.1}$$

Inverting the sigmoid we have

$$z(x,\theta) = \log \frac{f(x,\theta)}{1 - f(x,\theta)} \tag{9.2}$$

Therefore

$$z(x,\theta) = \log \frac{P_\theta(C_1|x)}{P_\theta(C_2|x)} \tag{9.3}$$

The above equations show that the ANN is in effect modelling the log-likelihood ratio of the two classes. The direct modeling of the ratio allows the ANN to be efficient in its use of parameters since each of the probability density functions is not modeled separately as is done with other types of classifiers.

If we restrict $z(x,\theta)$ to be linear we are in the LGD setup. When viewing this type of model as an ANN, it constitutes a network with no hidden layers. In an ANN with a single hidden layer we have (with $\theta = (\alpha,\beta)$)

$$z(x_i,\theta) = \beta_0 + \sum_{j=1}^{m} \beta_j \Lambda(\alpha_{0,j} + \sum_{k=1}^{p} \alpha_{k,j} x_{i,k}) \tag{9.4}$$

where $\Lambda(v) = (1 + \exp(-v))^{-1}$, $x_{i,k}$, is the kth of the p components in input vector x. Also each of the Λ-components of the m terms of the sum represent the output of a node of the hidden layer, with the input weights to a hidden node being the α terms and the weights of the output node being the β terms. Thus ANN's with layers comprise significantly more complex models than are usually considered in LGD. The significant consequence of this additional complexity is that of having more complex partitions of the feature space.

Another alternative is the inclusion of more complex functional forms in $z(x,\theta)$ itself. This can give rise to multiple partitions with a single layer.

9.2.1 Algorithm for ML-training of ANN's

The usual strategy is to apply the scoring method to different subsets of the parameters in the ANN while keeping the remaining weights fixed, i.e., the log-likelihood is maximized in stages, a procedure generally referred to as a Gauss-Siedel iterative method.

With the ANN in its initial state a node from the hidden layer is selected and its weights adjusted so as to maximize the likelihood. This change of wights alters the output that will be generated by this node. Thus the changes in the weights of the node in the hidden layer is followed by updating the weights of the output node. Then the next node in the hidden layer is considered and so on, until all the nodes in the hidden layer have been updated and then the procedure is repeated. Thus two iterations go on simultaneously - the scoring to update the weights, and the cycling through the nodes of the ANN.

9.3 The Perceptron

Let $X \subset I\!R^2$. Let $\psi(X)$ be a mapping from X onto the set $\{0, 1\}$. A predicate is simply a variable statement whose truth or falsity depends on choice of X, i.e., on $\psi(X)$. A predicate ψ is said to be conjunctively local of order k if it can be computed by a set Φ of predicates ϕ such that each ϕ depends upon no more than k points of $I\!R^2$:

$$\psi(X) = \begin{cases} 1 & \text{if}\phi(X) = 1\text{for every }\phi\text{in}\Phi \\ 0 & \text{otherwise} \end{cases} \tag{9.5}$$

Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ be a family of predicates. Then ψ is linear with respect to Φ if there exists a number θ and a set of scalars $\{a_{\phi_1}, \ldots, a_{\phi_n}\}$ such that $\psi(X) = 1$ if and only if $\sum_1^n a_{\phi_i}\phi_i(X) > \theta$.

A perceptron can also be defined as a device capable of computing all predicates that are linear in some given set Φ of partial predicates.

9.3.1 The Gradient Search

Define a gradient descent vector $J(a)$ that is minimized when an appropriate a is found. We start with some arbitrarily chosen weight vector a_1 and compute the gradient vectr. The iteration proceeds in the direction of steepest descent i.e., $a_{k+1} = a_k - \rho_k \nabla J(a_k)$, where ρ_k is a positive scale factor that sets the step size. Small step sizes lead to very slow convergences, while large step sizes can cause overshooting and divergence. If a quadratic expansion is valid, i.e.,

$$J(a) \approx J(a_k) + (\nabla J)^t(a - a_k) + \frac{1}{2}(a - a_k)^t D(a - a_k)$$

where D is the matrix of second derivatives $\frac{\partial 2J}{\partial a_i \partial a_j}$ evaluated at $a = a_k$. Then combining the above two we have

$$J(a_{K+1}) \approx J(a_k) - \rho_k ||(\nabla J)||^2 + \frac{1}{2}\rho_k^2 (\nabla J)^t D \nabla J \qquad (9.6)$$

Therefor to minimize the criterion function $J(a)$, choose $\rho_k = \frac{||\nabla J||^2}{(\nabla J)^t D \nabla J}$.

Rosenblatt's Perceptron Convergence theorem states that given an elementary α-perceptron, a stimulus world W, and any classification $C(W)$ for which a solution exists; let all stimuli in W occur in any sequence, provided that each stimulus must reoccur in finite time; then beginning from an arbitrary initial state, an error correction procedure will always yield a solution to $C(W)$ in finite time.

9.3.2 The LGD as a Perceptron

Consider a perceptron with a multivariate input $X \in \mathbb{R}^n$. The perceptron contains a single neuron that takes weighted inputs, sums them, transforms it by a sigmoidal (logistic) transfer function, and finally puts it through a hard-limiting output function, i.e., an indicator.

The descent function is taken to be the negative of log-likelihood of the corresponding LGD model as follows:

Let the transfer function be the logistic sigmoidal transformation

$$g(\mathbf{x}) = \frac{1}{1 + e^{\mathbf{w}^t\mathbf{x}}}$$

Note that $g(\mathbf{x})$ is monotone and one-to-one in $\mathbf{w}^t\mathbf{x}$. Thus a discrimination rule based on $\mathbf{w}^t\mathbf{x}$ can be equivalently formulated in terms of $g(\mathbf{x})$. Thus the input to the single neuron is the weighted sum of multivariate components. The perceptron generates an action potential y of value 1 if $g(\mathbf{x}) > c$, 0 otherwise. The error function and hence the gradient descent vector is taken to be

$$\sum(y_i - 1_{[g(\mathbf{x}_i)>c]})^2 \tag{9.7}$$

Note that by minimizing (9.7) we obtain the hyperplane separating the separable regions.

9.3.3 Apparent paradoxes

By the perceptron convergence theorem of Rosenblatt, the above perceptron converges in finite time (in finite passes) In the usual perceptron classifier, the descent function is a linear combination of weighted inputs over samples that are misclassified. This leads to a hyperplane separating the data. This is exactly the situation where the ordinary LGD of Anderson [5] fails.

9.4 Klopf's heterostat

9.4.1 Mathematical model

In Klopf's model [61] the neuron generates an action potential if

$$\sum_{i=1}^{n} w_i(t) f_i(t) \geq \theta(t_0) \tag{9.8}$$

where

n = number of synaptic inputs to the neuron;
$w_i(t)$ = synaptic transmittance associated with the ith input;
$f_i(t)$ = frequency measure of the input intensity at the ith synapse;
$\theta(t_0)$ = neuronal threshold;
t = time; and
t_0 = time elapsed since the generation of the last action potential. 8 The heterostatic variable μ is maximized when 'heterostasis' is achieved (i.e., the neuron experienced maximal polarization relative to environmental and adaptive mechanism induced constraints). Depolarization represents the positive aspect of polarization, hyperpolarization the negative. A neuron is in hetrostasis for the time t to $t+\tau$ if the quantity $\mu_1^{t,t+\tau}$ is maximized:

$$\mu_1^{t,t+\tau} = D_{t,t+\tau} - H_{t,t+\tau}$$

$$= \int_t^{t+\tau} (v_p(t) - v_r)dt - \int_t^{t+\tau} (v_n(t) - v_r)dt$$

$$= \int_t^{t+\tau} (v(t) - v_r)dt \qquad (9.9)$$

where

$D_t^{t+\tau}$ = amount of depolarization experienced during t to $t + \tau$
$H_t^{t+\tau}$ = amount of hyperpolarization experienced during t to $t + \tau$
$v(t)$ = potential difference across the neuronal membrane,
v_r = neuronal resting potential,
$v_p(t) = v(t)$ if $v(t) \geq v_r$, otherwise $v_p(t) = v_r$, and
$v_n(t) = v(t)$ if $v(t) \leq v_r$, otherwise $v_n(t) = v_r$.

9.4.2 ML and the Heterostat

From the structure of the heterostatic variable, an obvious similarity to ordinary maximum likelihood is immediately apparent. Think of a situation where the input signals are arriving in dicrete time and the active(non-rest component) neuronal potential is changed instantaneously by the input and remains at the changed level until the arrival of the next. Also assume that the total time alloted for the reception of signals is bounded. Take the neuronal potential difference $v(t)$ to be the log density (conditional) of the input component, and additionally that the heterostasis is considered over the entire period of signal arrival.

Let total number of samples be n, with n_i of population type Π_i, $i = 1, 2$. Without loss of generality assume total time of preliminary (training) signal arrival to be 1. Assume constant rate of arrival. Thus time difference between the arrival of consecutive signals is $1/n$. Let z_i be the indicator of inclusion taking values in the set $\{1, 0, -1\}$. Here 0 indicates no prior knowledge about the input class. Unlike other neuronal models since the heterostat is a goal-seeking unit, we can consider the z_i's as inputs to the neuron. However, these z_i's will perform the role of 'switches' that alter the state or configuration of the processing topology within the neuron, they modify the form of the polarization potentials existing at a given point of time within the neuron. Mathematically the potential can be written in one of several equivalent forms. However to bring out the similarity between the theory developed so far in LGD and the heterostat approach, we consider the following form:

$$v(t) = \log\left(\frac{1}{1+e^{-z\mathbf{w}^t\mathbf{x}}}\right)$$

Assume that the neuron maximizes its 'pleasure', i.e., the heterostatic variable at the end of regular intervals, in this case 1, and let v_r be 0. Thus the heterostatic variable μ becomes

$$\mu = \int v(t)dt = \int \log\left(\frac{1}{1+e^{-z\mathbf{w}^t\mathbf{x(t)}}}dt\right)$$
$$= \sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{1+e^{-z\mathbf{w}^t\mathbf{x}_i}}\right)$$

9.5 Consequences of the implementation

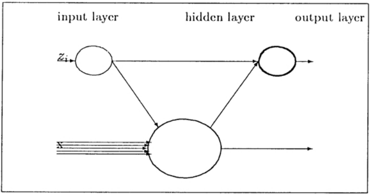
Assuming that at the end of its optimizing cycle, since the optimum solutions correspond to maximization of μ - the tendency of the neuron will be to resist changes in the weight configuration i.e., the neuron will "expect" any further change in thestate space to be zero. Mathematically this will take the form of the constraint :

$$E_{\mathbf{w}_{opt}}\left(\frac{\partial\mu}{\partial\mathbf{w}}|_{\mathbf{w}_{opt}}\right) = 0 \quad (9.10)$$

Note statistically this is equivalent to assuming that the expected derivatives of the log-likelihood are 0's.

Biologically the capacity of a neuron to speed up any internal processes, such as weight optimizations has to be limited. Thus in the model the rate of change of μ with respect to weight changes have to be bounded. The log-logistic potential satisfies this requirement.

The log-logistic potential is a generalization of possible actual functional potentials in two distinct ways. First, noting that the trigonometric polynomials are compact in $\mathcal{C}[0,1]$, and the cos function can be approximated (to arbitrary accuracy) by linear combinations of logistic functions. Hence a heterostat with log-logistic potential is indeed representative of general functional potentials. Secondly, as we have already pointed out, that the class admitting LGD is in fact exactly the class in which log-linear (Fisher Type) discrimination is allowed.



9.6 An extended model

The above model works by implementing a two pass cycle. In the first pass optimization occurs in the first hidden layer. In this cycle, the final outputs and error estimates are based on previously existing weights. In the second cycle, the same input signal is repeated to obtain correct error estimates. In this cycle, since the same signal pattern repeats no further optimization takes place, but the error estimator in the second hidden layer, correctly estimates the apparent error probabilities.

Note that in each pass the doubtful cases or unclassified cases, i.e., $z_i = 0$, do not enter the optimization process although they do get classified at the output end.

The proof for error estimator node is as follows:

$$\mu = \int v(t)dt = \int \log\left(\frac{1}{1+e^{-z\mathbf{w}^t\mathbf{x(t)}}}dt\right)$$

$$= \sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{1+e^{-z(wy_i)}}\right)$$

$$= -\sum_{C} \log(1+e^{w}) + \sum_{W} \log(1+e^{-w})$$

$$= -m_1 \log(1 + e^w) + m_2 \log(1 + e^{-w})$$

where C indicates the class of signals correctly classified , i.e., $yz = 1$, and W the wrong classifications, m_1 and m_2 the corresponding class sizes.

On optimization, we find that

$$w = \log \frac{m_1}{m_2}$$

Together with the restriction that there were no initial doubtful inputs, i.e., $m_1 + m_2 = n_1 + n_2 = n$, we find that non-hard limited output of the error estimator gives the correct classification estimates for tha corresponding signals.

Chapter 10

Programs and Computational support

10.1 Introduction

Here the various computer programs used have been detailed with the appropriate source code being supplied. Almost all the programs that have been developed for the specific purpose of computational work required in this work has been written in C and compiled in TURBO C. Apart from programs written in Turbo-C, the statistical package SPLUS (Windows Version 3.3) has been used. Graphs have been drawn using GNUPLOT 2.2. The GNUPLOT programs here have been designed to output LATEX output which need to be included in a LATEX document. Commenting out the set terminal and set output options in the programs will generate screen outputs.

Here is a brief contents list of the included programs.

3.combin1.c combin1.c is a C-routine that incorporates the program to estimate stable parameter quadruples as also the corresponding discrimination rule based on quantiles. page 108.

4.stabfig.c The C-program that generates tables for the GNUPLOT programs "stab1.inp" and "stab2.inp" for stable discrimination. page 123.

5.ecfcomr.inp GNUPLOT program to plot the graph of real parts of ecf of Normal and Cauchy. The program for imaginary part can be formed by just replacing the cos terms by sin. page 127.

6.stab1.inp GNUPLOT program to plot the two rules i.e., with and without considering inclusion probabilities. page 128.

7.stab2.inp GNUPLOT program to plot the estimated log-quantiles for the two subsamples. page 128.

8.dird.in The SPLUS program segment **dird.in** which computes Apparent Error Rates (AERs) for the first simple chord rule when applied to the vanishing angles of the homing pigeons as given by the data cited in Mardia [71] pp 156-157. page 129.

uni.in

```
#sink("outfile.dat",append=TRUE)
#sink() #output to the terminal
#warnings("outfile1.dat")
n <- 20; dimen <- 2; pr <- 0.5; gpsiz <- 1

   theta <- array(1:dimen,c(1));beta <- array(1:dimen,c(1))
 thetain <- array(1:dimen,c(1));ebeta <- array(1:dimen,c(1))
 x <- array(1:n,c(1));ind <- array(1:n,c(1))

thetain[1] <- 0;thetain[2] <- 1.2
 pr <- exp(thetain[1]); pr <- pr/(1+pr)

#rexp(n, rate=lambda) : exponential with mean 1/lambda
#rgamma(n, shape) :sample of size n with shape
#runif(n, min=0, max=1) # uniform from [min,max]
```

```
   for(i in 1:n) { ind[i] <- rbinom(1,size=gpsiz,pr)}
# ind = 0 for 1-st popn, 1 for 2-nd popn

mean2 <- 0; sd2 <- 1
mean1 <- mean2 + (sd2 * sd2 * thetain[2]);
sd1 <- sd2
for(i in 1:n)
{if( ind[i] == 1)
{ x[i] <- rnorm(1, mean2, sd2)
 }
else
{ x[i] <- thetain[2]+ rnorm(1,mean1,sd1)}}
for(i in 1:n){ l <- c(ind[i],x[i])
 write(l,file="outfile1.dat",append=T)};
n2 <- 0
for(i in 1:n){ n2 <- n2 + ind[i]}
# computation of likelihood in PML form
uscnt <- 0;assign("x",x,frame=1)
liklihdus <- function(theta)
{ uscnt <- uscnt + 1; ealpha <- exp(theta[1])
    for(i in 1:n){ ebeta[i] <- exp(x[i]*theta[2])
  };
      phi <- 0
       for(i in 1:n){if(ind[i] == 1)
        {phi <- phi + ebeta[i]}}
 phi <- phi/n2; like <- 1; pbyeal <- phi / ealpha
 # computation of likelihood starts here
  for(i in 1:n){ term <- ebeta[i]/(pbyeal + ebeta[i])
       term <- ((1 - (2*ind[i]))*term) + ind[i]
 like <- like*term}
like <- (-1) * like;like}

for(i in 1:dimen) {theta[i] <- 0}
   minvalus <- nlmin(liklihdus, theta, print.level=1,
     max.fcal=50, max.iter=50)

cat("alphahat betahat in PML form \n",
```

```
 file="outfile1.dat",append=TRUE)
write(minvalus,file="outfile1.dat", append=TRUE)
soln <- minvalus[1]
sol <- unlist(soln,recursive=T,use.names=F)

for(i in 1:n){ebeta[i] <- exp(x[i]*sol[2])
  };
phi <- 0
       for(i in 1:n){if(ind[i] == 1)
      {phi <- phi + ebeta[i]}}
       phi <- phi/n2;
lnphi <- log(phi)
errind <- 0;solnphi <- sol[1] - lnphi
for (i in 1:n){ erind <- 1
   if(solnphi + (sol[2]*x[i]) > 0) {erind <- 0}
     if(erind + ind[i] == 1){errind <- errind +1}
 };
aerus <- errind/n
cat("\n AER in PML = ",file="c:/spr/tt/outfile1.dat",
 append=TRUE)
write(aerus,file="c:/spr/tt/outfile1.dat",append=TRUE)
cat("\n The PML rule is  \n",
 file="c:/spr/tt/outfile1.dat",append=TRUE)
write(c(solnphi, sol[2]),
  file="c:/spr/tt/outfile1.dat",append=TRUE)
# computation of likelihood in Andersons form
ancnt <- 0; assign("x",x,frame=1)
liklihdan <- function(theta)
{ ancnt <- ancnt + 1; ealpha <- exp(theta[1])
    for(i in 1:n){ ebeta[i] <- exp(x[i]*theta[2])}
     like <- 1;
  pbyeal <- 1.0 / ealpha
 # computation of likelihood starts here
  for(i in 1:n)
{ term <- ebeta[i]/(pbyeal + ebeta[i])
       term <- ((1 - (2*ind[i]))*term) + ind[i]
 like <- like*term}
```

```
like <- (-1) * like; like
}
for(i in 1:dimen) {theta[i] <- 0}
   minvalan <- nlmin(liklihdan, theta, print.level=1,
     max.fcal=50, max.iter=50)
cat("alphahat betahat in Andersons form\n",
 file="c:/spr/tt/outfile1.dat",append=TRUE)
    write(minvalan,"c:/spr/tt/outfile1.dat",append=TRUE)
soln <- minvalan[1]
sol <- unlist(soln,recursive=T,use.names=F)

errind <- 0
for (i in 1:n)
 { erind <- 1
   if(sol[1] + (sol[2]*x[i]) > 0) {erind <- 0}
     if(erind + ind[i] == 1){errind <- errind +1}
 }
  aerus <- errind/n
      cat("\n AER in Anderson's = ",
 file="c:/spr/tt/outfile1.dat",append=TRUE)
write(aerus,file="c:/spr/tt/outfile1.dat",append=TRUE)
```

logis.c

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#include"supdir.h"
#include"supdir.c"
#include"optim.h"
#include"optim.c"

main(int argc,char *argv[])
{
/* void real_dat(int argc,char *argv[]);*/
void simul_dat(int argc,char *argv[]);
simul_dat(argc,argv);
}
```

```
/*

void real_dat(int argc,char *argv[])
{ FILE *datfile,*outfile;
double temp,**par;
int i,count,*ncol;
   char *data;

double funand(double *x,double **par);
void dfunand(double *xx, double *df,double **par);
double funsupr(double *x,double **par);
void dfunsupr(double *xx, double *df,double **par);
void init_estim(double **par,double (*fun)(double *x,double **par),
void (*dfun)(double *xx,double *df,double **par),char head[]);

data = argv[1];
if (data == NULL)
 {
 printf("\n%s [datafile] [outfile]\n",argv[0]);exit(0);
      }
    datfile = fopen(data,"r");
   if (datfile == NULL)
   { printf("No such file exists. Try again.\n");
   exit(0);}
   count=0;
   while( fscanf(datfile,"%lf",&temp) != EOF)
   {count++;}
    fclose(datfile);
printf("\ninput data size = %d\n",count);
ncol=v_ialloc(count+2);
for(i=0;i<count+2;i++)*(ncol+i)=1;
par=mat_alloc(count+2,ncol);
printf("\nType 1-st sample size \n");
scanf("%lf",&temp);
*(*par)=temp;
*(*(par+1))=((double)count)-(*(*par));
printf("\n*(*par) = %lf *(*(par+1)) = %lf\n",*(*par),*(*(par+1)));
```

```
datfile = fopen(data,"r");
for(i=0;i<count;i++)
{
fscanf(datfile,"%lf",&temp);
*(*(par+i+2)) = temp;
}
fclose(datfile);

outfile=fopen("logis.out","w");
fprintf(outfile,"\nComparison of Anderson's and Roy's Method\n");
fclose(outfile);
init_estim(par,funand,dfunand,"By Anderson's Method");
init_estim(par,funsupr,dfunsupr,"By Roy's Method");
}
*/

void simul_dat(int argc, char *argv[])
{ FILE *datfile,*outfile;
double temp,**par,beta;
int i,count,n1,*ncol;
   char *data;

double funand(double *x,double **par);
void dfunand(double *xx, double *df,double **par);
double funsupr(double *x,double **par);

void dfunsupr(double *xx, double *df,double **par);
void init_estim(double **par,double (*fun)(double *x,double **par),
void (*dfun)(double *xx,double *df,double **par),char head[]);

   data = argv[1];
   if (data == NULL)
   {
      printf("\n%s [datafile] [outfile]\n",argv[0]);
   exit(0);      }
    datfile = fopen(data,"r");
   if (datfile == NULL)
```

```
    {printf("No such file exists. Try again.\n");
    exit(0);}
    count=0;
    while( fscanf(datfile,"%lf",&temp) != EOF)
    {count++;}
     fclose(datfile);
printf("\ninput data size = %d\n",count);
ncol=v_ialloc(count+2);
for(i=0;i<count+2;i++) *(ncol+i)=1;
par=mat_alloc(count+2,ncol);

printf("\nType 1-st sample size \n");
scanf("%lf",&temp);
*(*par)=temp;
*(*(par+1))=((double)count)-(*(*par));
printf("\nn1=%lf, n2=%lf\n",*(*par),*(*(par+1)));
datfile = fopen(data,"r");
for(i=0;i<count;i++)
{fscanf(datfile,"%lf",&temp);
*(*(par+i+2)) = temp;
printf("\n data [%d] =%lf\n",i,*(*(par+i+2)));
}
fclose(datfile);

n1=(int)(*(*par));
printf("\nType beta value \n");
scanf("%lf",&beta);
for(i=0;i<n1;i++)
*(*(par+i+2)) += beta;

outfile=fopen("logis.out","w");
fprintf(outfile,"\nComparison of Anderson's and Roy's Method\n");
fclose(outfile);
init_estim(par,funand,dfunand,"By Anderson's Method");
init_estim(par,funsupr,dfunsupr,"By Roy's Method");
}
```

```
void init_estim(double **par,double (*fun)(double *x,double **par),
void (*dfun)(double *xx,double *df,double **par),char head[])
{
FILE *outfile;
double *p,ftol,*fret;
int i,n,*iter;
ftol=1.0e-6;
p=v_alloc(2);
fret=v_alloc(1);
iter=v_ialloc(1);
n=2;

outfile=fopen("logis.out","a");
fprintf(outfile,"\n%s\n",head);

for(i=0;i<n;i++) *(p+i)=0.0;
frprmn(p,n,ftol,iter,fret,(*fun),(*dfun),par);
for(i=0;i<n;i++){
printf("\n theta[%d] = %lf\n",i,*(p+i));
fprintf(outfile,"\n theta[%d] = %lf\n",i,*(p+i));
}
printf("\n No.of iterations = %d\n",*iter);
fprintf(outfile,"\n No.of iterations = %d\n",*iter);
fclose(outfile);
free(fret);
free(iter);
free(p);
}

double funand(double *x,double **par)
{int n1,n2,i,n;
double val,temp;
n1 =(int) (*(*par));
n2 =(int)(*(*(par+1)));
val=0.0;n=n1+n2;
for(i=0;i<n;i++){
 temp = *x + (*(x+1) *(*(*(par+i+2))));
```

```
 if(i<n1) val += temp;
 val -= log(1.0+exp(temp));}
return(-val);}

void dfunand(double *x, double *df,double **par)
{int n1,n2,i,n;
double temp,temp1;
n1 =(int) (*(*par));
n2 =(int)(*(*(par+1)));
n=n1+n2;
*df=0.0;
*(df+1)=0.0;
for(i=0;i<n;i++){
 temp = *x + (*(x+1) *(*(*(par+i+2))) ) ;
temp1= 1.0/(1.0+exp(temp));
 if(i<n1){
  *df += temp1;
  *(df+1) += (*(*(par+i+2)))*temp1;
  } else {
 *df -= exp(temp)*temp1;
 *(df+1) -= exp(temp)*(*(*(par+i+2)))*temp1;
 } }
*df= -(*df);
*(df+1)= -(*(df+1));}

double funsupr(double *x,double **par)
{int n1,n2,i,n;
double val,temp,*phi,*samp1;
void phi_hat(double *samp,int samp_siz,double *thet,double *phi);

n1 =(int) (*(*par));
n2 =(int)(*(*(par+1)));
samp1=v_alloc(n2);
for(i=0;i<n2;i++)
*(samp1+i)=*(*(par+n1+2+i));
val=0.0;
n=n1+n2;
```

```
phi=v_alloc(2);
phi_hat(samp1,n2,x,phi);

for(i=0;i<n;i++){
 temp = *x + (*(x+1) *(*(*(par+i+2))));
 if(i<n1) {val += temp;}
 else
 {val += log(*phi);}
 val -= log((*phi)+exp(temp));
 } free(samp1);
return(-val);
}

void phi_hat(double *samp,int sampsiz,double *thet,double *phi)
{
int i;
double val1,val2,temp;
val1=val2=0.0;
for(i=0;i<sampsiz;i++){
temp = exp(*(thet+1) *(*(samp+i)));
val1 += temp;
val2 += temp*(*(samp+i));
}
val1/=(double)sampsiz;
val2/=(double)sampsiz;
*phi=val1;
*(phi+1)=val2;}

void dfunsupr(double *x, double *df,double **par)
{int n1,n2,i,n;
double temp,*phi,*samp1;
void phi_hat(double *samp,int samp_siz,double *thet,double *phi);
n1 =(int) (*(*par));
n2 =(int)(*(*(par+1)));
n=n1+n2;
*df=0.0;
*(df+1)=0.0;
```

```
samp1=v_alloc(n2);
for(i=0;i<n2;i++)
*(samp1+i)=*(*(par+n1+2+i));
phi=v_alloc(2);

phi_hat(samp1,n2,x,phi);
for(i=0;i<n;i++){
 temp = exp(*x + (*(x+1) *(*(*(par+i+2)))));
 if(i<n1){
  *df += (*phi)/((*phi)+temp);
  *(df+1) += ((*phi)*(*(*(par+i+2)))-(*(phi+1)))/((*phi)+temp);
  } else {
  *df -= temp/((*phi)+temp);
 *(df+1) -= temp*((*phi)*(*(*(par+i+2)))
-(*(phi+1)))/((*phi)+temp);
 }}
printf("\n*df = %lf , *(df+1) = %lf\n",*df,*(df+1));
 *df=-(*df);
 *(df+1) = -(*(df+1));
 free(phi);
}
```

combin1.c

```
#include<stdio.h>
#include<math.h>
#include<stdlib.h>
#include<time.h>

main(argc,argv)
int argc;
char *argv[];
{ FILE *datfile;
 double *stk, *stock,*samp[2],*rnstab,q[2];
double *param,*prop,*est_par,ind,x,val,*test_smp[2];
float temp;
int i,j,k,count,cnt,lim,range;
```

```
   int *limits[2],length;
   char *data;
void estimate(int smp_no, double *stk,int range,
double *est_par);

void stabgen(double *rnstab,double *par,int length);

void error_calc(double *samp[2],
  double *test[2],int *limits[2],
int *limits1[2], double *prop, double *param);

void err_print(int argc,char *argv[],double *prop,
int *limits[2]);

void est_print(int argc,char *argv[],
  double *est_par,int j);

double absol(double x, double y);
double square(double x, double y);

   data = argv[1];
   if (data == NULL)
   {
      printf("combin [datafile] [outfile]\n");
   exit(0);
      }
    datfile = fopen(data,"r");
   if (datfile == NULL)
   {
   printf("No such file exists. Try again.\n");
   exit(0);
      }
   count=0;
   while( fscanf(datfile,"%f",&temp) != EOF)
   {count++;}
   rewind(datfile);
   stock = (double *)malloc(count*sizeof(double));
```

```
   for(i=0;i<count;i++)
   { fscanf(datfile,"%f",&temp);
     *(stock+i) = (double) temp;
      }
 printf("\nTotal no. of data points in file = %d\n",count
for(i=0;i<2;i++){
limits[i] = (int *)malloc(3*sizeof(int));
}
*limits[0]=*limits[1]=1;
*(limits[0]+1)=*(limits[1]+1)=count;
if(argc > 2)
{
printf("\nType 1-st sample starting lower limit\n");
scanf("%d",&k);
*limits[0]=k;
printf("\nType 1-st sample end, upper limit\n");
scanf("%d",&k);
*(limits[0]+1)=k;
printf("\nType 2-nd sample starting lower limit\n");
scanf("%d",&k);
*limits[1]=k;
printf("\nType 2-nd sample end, upper limit\n");
scanf("%d",&k);
*(limits[1]+1)=k;
};

for(j=0;j<2;j++){
*(limits[j]+2)=*(limits[j]+1)-*limits[j];
samp[j] = (double *) malloc(*(limits[j]+2)*sizeof(double));
}

for(j=0;j<2;j++)
{
for(i=0;i< *(limits[j]+2);i++)
{ *(samp[j]+i)=log(*(stock+i+1+ *limits[j]))
  -log(*(stock+i+ *limits[j]));}
}
```

```
fclose(datfile);

param = (double *)malloc(3*sizeof(double));
*param=0.5;
*(param+1)=1.0;
*(param+2)=1.0;

/* param pints to (0) - pivot value (default 0.5)
    (1) - scale factor of 1-st component
    (2) - scale factor of 2-nd component*/
prop= (double *)malloc(3*sizeof(double));
*(prop+2)=0.0;
error_calc(samp,samp,limits,limits,prop,param);
err_print(argc,argv,prop,limits);
est_par = (double *)malloc(4*sizeof(double));

for(j=0;j<2;j++)
{
range=*(limits[j]+2);
stk=(double *)malloc(range*sizeof(double));

for(i=0;i<range;i++){*(stk+i)=*(samp[j]+i);}
estimate(j,stk,range,est_par);
est_print(argc,argv,est_par,j);

free(stk);
length=range;
rnstab = (double *)malloc(length*sizeof(double));
stabgen(rnstab,est_par,length);
estimate(j+2,rnstab,length,est_par);

test_smp[j] = (double *)malloc(length*sizeof(double));
for(i=0;i<length;i++){*(test_smp[j]+i)=*(rnstab+i);}
free(rnstab);
est_print(argc,argv,est_par,j+2);
}
```

```
error_calc(samp,test_smp,limits,limits,prop,param);
err_print(argc,argv,prop,limits);
}

void err_print(int argc,char *argv[],double *prop,
int *limits[2])
{ FILE *outfile;
printf("\n");
     printf("error estimate for
 1-st sample = %lf\n",*prop);
     printf("error estimate for
2-nd sample = %lf\n",*(prop+1));
  printf("error estimate for
combined sample = %lf\n",*(prop+2));
     printf("sample_1 lower limit = %d\n",*limits[0]);
printf("sample_1 upper limit = %d\n",*(limits[0]+1));
printf("sample_2 lower limit = %d\n",*limits[1]);
printf("sample_2 upper limit = %d\n",*(limits[1]+1));

if(argc == 3)
{ outfile = fopen(argv[2],"a"); .
     fprintf(outfile,"\n");
     fprintf(outfile,"error estimate for
1-st sample = %lf\n",*prop);
     fprintf(outfile,"error estimate for
2-nd sample = %lf\n",*(prop+1));
     fprintf(outfile,"error estimate for
combined sample = %lf\n",*(prop+2));
     fprintf(outfile,"sample_1 lower limit
= %d\n",*limits[0]);
fprintf(outfile,"sample_1 upper limit
= %d\n",*(limits[0]+1));
fprintf(outfile,"sample_2 lower limit
= %d\n",*limits[1]);
fprintf(outfile,"sample_2 upper limit
= %d\n",*(limits[1]+1));
fclose(outfile);
```

```
 };
 }

 void est_print(int argc,char *argv[],
 double *est_par,int j)
 {FILE *outfile;
  printf("\n parameter estimates for sample_%d\n",j);
      printf("alpha = %lf\n",*est_par);
      printf("beta  = %lf\n",*(est_par+1));
   printf("c      = %lf\n",*(est_par+2));
      printf("delta = %lf\n",*(est_par+3));

 if(argc == 3)
   { outfile = fopen(argv[2],"a");
      fprintf(outfile,"\n parameter estimates
 for sample_%d\n",j);
      fprintf(outfile,"alpha = %lf\n",*est_par);
      fprintf(outfile,"beta  = %lf\n",*(est_par+1));
  fprintf(outfile,"c      = %lf\n",*(est_par+2));
      fprintf(outfile,"delta = %lf\n",*(est_par+3));
   fclose(outfile);
  };
}

void error_calc(double *samp[2],
 double *test[2],int *limits[2],
int *limits1[2],double *prop, double *param)
{int j,i,jj,ii,cnt,lim,lim1;

double q[2],x,ind,level=0.0,val;
double absol(double x, double y);

for(j=0;j<2;j++)
{
ind = (double) (1-(2*j));
cnt=0;lim=*(limits1[j]+2);
for(i=0;i<lim;i++)
```

```
  { x=*(test[j]+i);

for(jj=0;jj<2;jj++){
q[jj]=0; lim1=*(limits[jj]+2);
 for(ii=0;ii<lim1;ii++)
 {if(x > *(samp[jj]+ii)){q[jj] +=1.0;};
     }
  q[jj] /= (double) lim1;
  }
 val = absol(q[0],*param)-absol(q[1],*param);
if(ind*(val-level) > 0.0)
{cnt++;};}
*(prop+j)= ((double) cnt)/((double) lim);
*(prop+2) += (double) cnt;
}
*(prop+2) /= (double) (*(limits[0]+2)+ *(limits[1]+2));
}

double absol(double x, double y)
{return(fabs(x-y));}

double square(double x, double y)
{return( (x-y)*(x-y));}

void stabgen(double *rnstab,double *par,int length)
{FILE *rstab;
int i;
 double u1,u2,v1,v2,v3,w,phi,phi0;
 double kalpha,val,pi_by2,pi=3.14159;
  pi_by2=pi/2.0;
   randomize();
kalpha=*par;
     if(kalpha > 1.0){kalpha -= 2.0;};
     phi0=(-1.0)*pi_by2* *(par+1) *kalpha/(*par);
rstab= fopen("stabvar","a");
if(*par != 1.0)
{for(i=0;i<length;i++)
```

```
    {
 u1= rand();
 u1 /= (double) 32767;
 u2= rand();
 u2 /= (double) 32767;
    w=(-1.0)*log(u1);
     phi=pi*(u2-0.5);
        val=*par *(phi-phi0);
   v1 = cos(phi-val)/w;
    v2 = sin(val);
     v3 = cos(phi);
    val = exp( (log(v1*v3))/(*par));
      val /= v1; val *=v2;
    val *= *(par+2);
  val += *(par+3)/(*(par+2));
      *(rnstab+i)=val;
        fprintf(rstab,"\n *rnstab = %lf\n",*(rnstab+i));
      }
 };
if(*par == 1.0)
{
for(i=0;i<length;i++)
   {
   u1= rand();
   u1 /= (double) 32767;
   u2= rand();
   u2/= (double) 32767;
    w=(-1.0)*log(u1);
     phi=pi*(u2-0.5);
      v1=2.0* *(par+1)/pi;
       v2= 1.0+(v1*phi);
  *(rnstab+i)=(v2*tan(phi))-(v1*log(w*cos(phi)/v2));
     }
};
}

void estimate(int smp_no, double *stk,
```

```
 int range,double *est_par)
{     double x_95,x_75,x_50,x_25,x_05;
      int i_95,i_75,i_50,i_25,i_05;
   double na,nb,nc,nd,x;
   float temp;
   int i,j,k,count;
   double a,b,c,d,*aa;
int mycompare(double *x, double *y);
double interpolate(double, double,
  double,double,double);

void table_1(double x, double y, double *aa);
double table_2(double, double);
double table_3(double, double);

aa = (double *) malloc(2*sizeof(double));
      count=range;
     qsort(stk,count,sizeof(double),mycompare);
    x = (double) (2.0*count*.05+1.0)*.5;
     i_05 = (int) x;
  x_05 = interpolate((double)i_05,*(stk+i_05-1),
      (double)i_05+1,*(stk+i_05),x);
        x = (double) (2.0*count*.25+1.0)*.5;
     i_25 = (int) x;
x_25 = interpolate((double)i_25,*(stk+i_25-1),
(double)i_25+1,*(stk+i_25),x);
x = (double) (2.0*count*.5+1.0)*.5;
     i_50 = (int) x;
x_50 = interpolate((double)i_50,*(stk+i_50-1),
 (double)i_50+1,*(stk+i_50),x);
     x = (double) (2.0*count*.75+1.0)*.5;
    i_75 = (int)x;
x_75 = interpolate((double)i_75,*(stk+i_75-1),
(double)i_75+1,*(stk+i_75),x);
   x  = (double) (2.0*count*.95+1.0)*.5;
      i_95 = (int) x;
      x_95 = interpolate((double)i_95,*(stk+i_95-1),
```

```
        (double)i_95+1,*(stk+i_95),x);
   na = (x_95 - x_05)/(x_75 - x_25);
        nb = (x_95 + x_05 - 2.0*x_50)/(x_95 - x_05);
        table_1(na,nb,aa);
        a=*aa; b=*(aa+1);
        nc = table_2(a,b);
       c = (x_75 - x_25)/nc;
        nd = table_3(a,b);
        d = x_50 + c*nd - b*c*tan(3.14159265*a/2.0);
  printf("\n parameter estimates for sample_%d\n",smp_no);
     printf("alpha = %lf\n",a);
     printf("beta  = %lf\n",b);
  printf("c      = %lf\n",c);
     printf("delta = %lf\n",d);
*est_par=a;
*(est_par+1)=b;
*(est_par+2)=c;
*(est_par+3)=d;
}

int mycompare (double *x,double *y)
{
     if ( (*x - *y) > 0 ) {return(1);}
   else {if ((*x -*y) <0) return (-1);};
   return(0);
     }
double interpolate(x1,y1,x2,y2,x)
double x1,y1,x2,y2,x;
{      double tmp;
       tmp = y1 + (x- x1)*(y2 - y1)/(x2 - x1);
      return(tmp);}

void table_1( double x, double y, double *aa)
{
        int i,j,pt_x=14,pt_y=6;
        double a_1,a_2,b_1,b_2,a,b;
```

```
double data_1[] = {2.0    ,2.0    ,2.0     ,2.0    ,
2.0   ,2.0   ,2.0    ,
1.916  ,1.924  ,1.924   ,1.924  ,1.924  ,1.924 ,1.924  ,
1.808  ,1.813  ,1.829   ,1.829  ,1.829  ,1.829 ,1.829  ,
1.729  ,1.730  ,1.737   ,1.745  ,1.745  ,1.745 ,1.745  ,
1.664  ,1.663  ,1.663   ,1.668  ,1.676  ,1.676 ,1.676  ,
1.563  ,1.560  ,1.553   ,1.548  ,1.547  ,1.547 ,1.547  ,
1.484  ,1.480  ,1.471   ,1.460  ,1.448  ,1.438 ,1.438  ,
1.391  ,1.386  ,1.378   ,1.364  ,1.337  ,1.318 ,1.318  ,
1.279  ,1.273  ,1.266   ,1.250  ,1.210  ,1.184 ,1.150  ,
1.128  ,1.121  ,1.114   ,1.101  ,1.067  ,1.027 ,0.973  ,
1.029  ,1.021  ,1.014   ,1.004  ,0.974  ,0.935 , .874  ,
.896   , .892  , .887   , .883  , .855  , .823 , .769  ,
.818   , .812  , .806   , .801  , .780  , .756 , .691  ,
.698   , .695  , .692   , .689  , .676  , .656 , .595  ,
.593   , .590  , .588   , .586  , .579  , .563 , .513  };

double data_2[] = {0.0,2.160,1.0,1.0,1.0,1.0,1.0    ,
     0.0,1.592,3.39,1.0,1.0,1.0,1.0    ,
 0.0     , .759  ,1.8     ,1.0    ,1.0    ,1.0    ,1.0    ,
 0.0     , .482  ,1.048   ,1.694 , ,1.0   ,1.0    ,1.0    ,
 0.0     , .360  , .76    ,1.232  ,2.229  ,1.0    ,1.0    ,
 0.0     , .253  , .518   , .823  ,1.575  ,1.0    ,1.0    ,
 0.0     , .203  , .41    , .632  ,1.244  ,1.906 ,1.0     ,
 0.0     , .165  , .332   , .499  , .943  ,1.56  ,1.0     ,
 0.0     , .136  , .271   , .404  , .689  ,1.23  ,2.195   ,
 0.0     , .109  , .216   , .323  , .539  , .827 ,1.917   ,
 0.0     , .096  , .19    , .284  , .472  , .693 ,1.759   ,
 0.0     , .082  , .163   , .243  , .412  , .601 ,1.596   ,
 0.0     , .074  , .147   , .22   , .377  , .546 ,1.482   ,
 0.0     , .064  , .128   , .191  , .33   , .478 ,1.362   ,
 0.0     , .056  , .112   , .167  , .285  , .428 ,1.274   };

double nu_x[] = {2.439,2.5,2.6,2.7,2.8,3.0,3.2,
   3.5,4.0,5.0,6.0,8.0,10.0,15.0,25.0};
double nu_y[] = {0.0,0.1,0.2,0.3,0.5,0.7,1.0};
```

```
        if (x < 2.439)
        {  a = 2.0;    b = y; return;}
        else if ( fabs(y) > 1.0)
        {
    b = (y/fabs(y));
         i=0;
    while(i < 15)
         {
  if (x < nu_x[i])
  { pt_x = i-1;
    break;
    }
i++;        }
 a = interpolate(nu_x[pt_x],data_1[7*pt_x+6],
        nu_x[pt_x +1],
    data_1[7*pt_x+13],x);
    return;
}
       else
{  i=0;
    while (i<15)
    {
        if (x < nu_x[i])
        {
   pt_x = i-1;
   break;
        }
        i++;
    }
    j=0;
    while (j< 7)
    {
        if (fabs(y) < nu_y[j])
        {   pt_y = j-1;
    break;
        }
        j++;
```

```
    }
a_1=interpolate(nu_x[pt_x],data_1[7*pt_x+pt_y],
  nu_x[pt_x+1],data_1[7*pt_x+pt_y+7],x);
a_2=interpolate(nu_x[pt_x],data_1[7*pt_x+pt_y+1],
  nu_x[pt_x+1],data_1[7*pt_x+pt_y+8],x);
a=interpolate(nu_y[pt_y],a_1,nu_y[pt_y+1],a_2,fabs(y));
b_1=interpolate(nu_x[pt_x],data_2[7*pt_x+pt_y],
  nu_x[pt_x+1],data_2[7*pt_x+pt_y+7],x);
  b_2=interpolate(nu_x[pt_x],data_2[7*pt_x+pt_y+1],
  nu_x[pt_x+1],data_2[7*pt_x+pt_y+8],x);
b=interpolate(nu_y[pt_y],b_1,nu_y[pt_y+1],b_2,fabs(y));
        if (b > 1.0) b =1.0;
        b = (y/fabs(y))* b;
        *aa=a; *(aa+1)=b;
        }}
double table_2(x,y)
double x,y;
{
      int i,j,pt_x=15,pt_y=4;
      double nc1,nc2,nc3;
double data[] = {1.908 ,1.908  ,1.908  ,1.908  ,1.908  ,
   1.914 ,1.915  ,1.916  ,1.918  ,1.921  ,
   1.921 ,1.922  ,1.927  ,1.936  ,1.947  ,
   1.927 ,1.930  ,1.943  ,1.961  ,1.987  ,
   1.933 ,1.940  ,1.962  ,1.997  ,2.043  ,
   1.939 ,1.952  ,1.988  ,2.045  ,2.116  ,
   1.946 ,1.967  ,2.022  ,2.106  ,2.211  ,
   1.955 ,1.984  ,2.067  ,2.188  ,2.333  ,
   1.965 ,2.007  ,2.125  ,2.294  ,2.491  ,
   1.980 ,2.040  ,2.205  ,2.435  ,2.696  ,
   2.000 ,2.085  ,2.311  ,2.624  ,2.973  ,
   2.04  ,2.149  ,2.461  ,2.886  ,3.356  ,
   2.098 ,2.244  ,2.676  ,3.265  ,3.912  ,
   2.189 ,2.392  ,3.004  ,3.844  ,4.775  ,
   2.337 ,2.635  ,3.542  ,4.808  ,6.247  ,
   2.588 ,3.073  ,4.534  ,6.636  ,9.144  };
double nu_a[]= {2.0,1.9,1.8,1.7,1.6,1.5,1.4,1.3,1.2,
```

```
1.1,1.0,.9,.8,.7,.6,.5};
double nu_b[]= {0.0,0.25,0.5,0.75,1.0};
      i=0;
       while (i<16)
       {   if( x > nu_a[i])
       {   pt_x = i-1; break;
       }
       i++;
       }
      j=0;
       while (j<5)
       {
       if ( fabs(y) < nu_b[j])
       { pt_y=j-1;
       break;
         }
  j++;
        }
nc1 =interpolate(nu_a[pt_x],data[5*pt_x+pt_y],
  nu_a[pt_x+1],data[5*pt_x+pt_y+5],x);
     nc2 =interpolate(nu_a[pt_x],data[5*pt_x+pt_y+1],
      nu_a[pt_x+1],data[5*pt_x+pt_y+6],x);
nc3 =interpolate(nu_b[pt_y],nc1,nu_b[pt_y+1],nc2,fabs(y));
   return(nc3);}

double table_3(x,y)
double x,y;
{       int i,j,pta=15,ptb=4;
        double nd1,nd2,nd3;
 double dat[]={0.0 ,0.0    ,0.0      ,0.0      ,0.0     ,
        0.0 ,-.017  ,-.032    ,-.049    ,-.064    ,
        0.0 ,-.030  ,-.061    ,-.092    ,-.123    ,
        0.0 ,-.043  ,-.088    ,-.132    ,-.179    ,
        0.0 ,-.056  ,-.111    ,-.170    ,-.232    ,
        0.0 ,-.066  ,-.134    ,-.206    ,-.283    ,
        0.0 ,-.075  ,-.154    ,-.241    ,-.335    ,
        0.0 ,-.084  ,-.173    ,-.276    ,-.390    ,
```

```
        0.0 ,-.090  ,-.192    ,-.310    ,-.447    ,
        0.0 ,-.095  ,-.208    ,-.346    ,-.508    ,
        0.0 ,-.098  ,-.223    ,-.383    ,-.576    ,
        0.0 ,-.099  ,-.237    ,-.424    ,-.652    ,
        0.0 ,-.096  ,-.250    ,-.469    ,-.742    ,
        0.0 ,-.089  ,-.262    ,-.520    ,-.853    ,
        0.0 ,-.078  ,-.272    ,-.581    ,-.997    ,
        0.0 ,-.061  ,-.279    ,-.659    ,-1.198  };

double nua[]={2.0,1.9,1.8,1.7,1.6,1.5,1.4,1.3,1.2,1.1,1.0,
         .9,.8,.7,.6,.5};
double nub[]={0.0,0.25,0.50,0.75,1.00};

        i=0;
        while(i<16)
        {
        if( x > nua[i])
        { pta = i-1;
        break;
        }
        i++;
        }
        j=0;
        while(j<5)
        { if( fabs(y) < nub[j])
        { ptb =j-1; break; }
    j++;
        }
  nd1 = interpolate(nua[pta],dat[5*pta+ptb],
   nua[pta+1],dat[5*pta+ptb+5],x);
   nd2 = interpolate(nua[pta],dat[5*pta+ptb+1],
    nua[pta+1],dat[5*pta+ptb+6],x);
     nd3 = interpolate(nub[ptb],nd1,nub[ptb+1],nd2,fabs(y));
      nd3 = (y/fabs(y))*nd3;
    return(nd3);
}
```

stabfig.c

```
#include<stdio.h>
#include<math.h>
#include<stdlib.h>
#include<time.h>
#include"supdir.h"
#include"supdir.c"

main(int argc,char *argv[])
{     FILE *datfile,*outfile1,*outfile2,*outfile3;
   double *stack,*stock,*samp[2];
   double *param,x,val;
   double temp,left,rt,h,xx;
   int i,j,k,start,stop,count,cnt,len,range,*sampdim;
   int x_lim;
   char *data;
double rule1(double *samp[2],double x,double *par);
double rule2(double *samp[2],double x,double *par);
void up_order(double *samp,int sampdim);
double qtile(double *samp,double x,int sampdim);

   data = argv[1];
   if (data == NULL)
   {
      printf("combin [datafile] [outfile]\n");
   exit(0);
      }
    datfile = fopen(data,"r");
   if (datfile == NULL)
   {
   printf("No such file exists. Try again.\n");
   exit(0);
      }

   count=0;
   while( fscanf(datfile,"%lf",&temp) != EOF)
```

```
   {count++;}
   rewind(datfile);
   stock = v_alloc(count);
   for(i=0;i<count;i++)
   { fscanf(datfile,"%lf",&temp);
     *(stock+i) = log(temp);
      }
 printf("\nTotal no. of data points in file = %d\n",count);

 stack=v_alloc(count-1);
for(i=0;i<count-1;i++)
 *(stack+i)=*(stock+i+1)-(*(stock+i));
 free(stock);

 sampdim=v_ialloc(2);
for(i=0;i<2;i++){
printf("\nType %d-st sample starting lower limit\n",i+1);
scanf("%d",&start);
printf("\nType %d-st sample end, upper limit\n",i+1);
scanf("%d",&stop);
len=stop-start+1;
*(sampdim+i)=len;
samp[i]=v_alloc(len);
for(j=0;j<len;j++) *(samp[i]+j)=*(stack+start-1+j);
}
fclose(datfile);
up_order(samp[0],*sampdim);
up_order(samp[1],*(sampdim+1));
for(i=0;i<*sampdim;i++)
printf("\nsamp[0] +%d =%lf\n",i,*(samp[0]+i));

left= *samp[0] > *samp[1] ? *samp[0] : *samp[1];
rt= *(samp[0]+(*sampdim)-1) < *(samp[1]+(*(sampdim+1))-1) ?
*(samp[0]+(*sampdim)-1) : *(samp[1]+(*(sampdim+1))-1);
printf("\nleft = %lf, rt = %lf\n",left,rt);

param=v_alloc(2);
```

```
*param = (double) *sampdim;
*(param+1)=(double) *(sampdim+1);
x_lim=50;
h=(rt-left)/((double)x_lim);
xx=left;
outfile1=fopen("stplot1.dat","w");
for(i=0;i<x_lim;i++)
{
fprintf(outfile1,"%lf %lf %lf\n",xx,rule1(samp,xx,param),
 rule2(samp,xx,param));
xx+=h;
}
outfile2=fopen("stplot2.dat","w");
for(i=0;i<*sampdim-1;i++)
{
xx=*(samp[0]+i);
fprintf(outfile2,"%lf %lf\n",xx,-log(qtile(samp[0],xx,*sampdim)));
}

outfile3=fopen("stplot3.dat","w");
for(i=0;i<*(sampdim+1)-1;i++)
{
xx=*(samp[1]+i);
fprintf(outfile3,"%lf %lf\n",xx,-log(qtile(samp[1],xx,*(sampdim+1))));
}}

double rule1(double *samp[2],double x,double *par)
{
int i,n1,n2;
double val,np1,np2,n;
n1=(int) (*par);
n2=(int) (*(par+1));
n= (*(par+1))+(*par);
np1=np2=0.0;
for(i=0;i<n1;i++)
np1 += *(samp[0]+i) >x ? 1.0 : 0.0;
for(i=0;i<n2;i++)
```

```
np2 += *(samp[1]+i) >x ? 1.0 : 0.0;
np1 /=n;
np2 /=n;
val=fabs(np2-0.5)-fabs(np1-0.5);
return(val);
}

double rule2(double *samp[2],double x,double *par)
{
int i,n1,n2;
double val,np1,np2,n,pi1,pi2;

n1=(int) (*par);
n2=(int) (*(par+1));
n= (*(par+1))+(*par);
pi1= *par/n;
pi2=*(par+1)/n;

np1=np2=0.0;
for(i=0;i<n1;i++)
np1 += *(samp[0]+i) >x ? 1.0 : 0.0;
for(i=0;i<n2;i++)
np2 += *(samp[1]+i) >x ? 1.0 : 0.0;
np1 /=n;
np2 /=n;
val=(pi2*fabs(np2-0.5))-(pi1*fabs(np1-0.5));
return(val);
}

double qtile(double *samp,double x,int sampdim)
{
int i;
double val=0.0;
for(i=0;i<sampdim;i++)
val += *(samp+i) >x ? 1.0 : 0.0;
val /=(double)sampdim;
return(val);
```

```
}
void up_order(double *samp,int sampdim)
{
int i,j,k;
double temp;
for(i=0;i<sampdim;i++)
{for(j=i+1;j<sampdim;j++)
 {if(*(samp+j) < *(samp+i)){temp=*(samp+j);
       *(samp+j)=*(samp+i);
       *(samp+i)=temp;
      }}}}
```

ecfcomr.inp

```
set terminal latex
set output "ecfr.tex"
set size 4.7/5.,3.8/3.
#set format xy "$%g$
set title "Figure 1. Plot of real parts of e.c.f of Normal and Cauchy."
set xlabel "$x$" -1,-1
set ylabel "$y$" 1,-1
set zlabel "$Re \phi (x+\mbox{i}y)$" -1
set nokey
set label "Normal" at 0.25,0.25,1.15
set label "Cauchy" at 0.25,0.25,1.1
#set xtics
x1=0.150289049572354
x2=1.44738274462592
x3=-1.2834786708003
x4=-0.924228233319889
x5=0.82362118260394
x6=0.399613675386185
x7=-0.384365934425452
x8=-1.21272501905165
x9=0.729908355095443
x10=1.26577318735458
y1=-0.384118949439741
y2=0.779717426386718
```

```
y3=-0.931442798351007
y4=0.89859960887145
y5=1.45644941243947
y6=2.33519940686864
y7=-0.553710415077411
y8=0.690911363405423
y9=2.17043384864147
y10=-1.89370528544057
normal(x,y)=((exp(-y*x1)*cos(x*x1))+(exp(-y*x2)*cos(x*x2))\
+(exp(-y*x3)*cos(x*x3))+(exp(-y*x4)*cos(x*x4))\
+(exp(-y*x5)*cos(x*x5))+(exp(-y*x6)*cos(x*x6))\
+(exp(-y*x7)*cos(x*x7))+(exp(-y*x8)*cos(x*x8))\
+(exp(-y*x9)*cos(x*x9))+(exp(-y*x10)*cos(x*x10)))/10.0
cauchy(x,y)=((exp(-y*y1)*cos(x*y1))+(exp(-y*y2)*cos(x*y2))\
+(exp(-y*y3)*cos(x*y3))+(exp(-y*y4)*cos(x*y4))+\
(exp(-y*y5)*cos(x*y5))+(exp(-y*y6)*cos(x*y6))\
+(exp(-y*y7)*cos(x*y7))+(exp(-y*y8)*cos(x*y8))\
+(exp(-y*y9)*cos(x*y9))+(exp(-y*y10)*cos(x*y10)))/10.0
splot [-0.1:0.1] [-0.1:0.1] normal(x,y), cauchy(x,y)
```

stab1.inp

```
set terminal latex
set output "stabr1.tex"
set size 4.5/5.,3.8/3.
#set format xy "$%g$
set title "Figure 1. Plot of two classification rules"
set xlabel "$x$" 0,-1
set ylabel "$rule\;value$" -6
set nokey
set label "Rule 1 ----" at 0.08,0.12
set label "Rule 2 {\bf ----}" at 0.08,0.1
set xtics  -0.14,0.04
plot "stplot1.dat" using 1:2 w l,"stplot1.dat" using 1:3 w l
```

stab2.inp

```
set terminal latex
```

```
set output "stabr2.tex"
set size 4.5/5.,3.5/3.
#set format xy "$%g$
set title "Figure 4. Plot of -log(quantile) for two samples"
set xlabel "$x$" 0,-2
set ylabel "$-log(quantile)$" -7
set nokey
set label "1st sample (1-320) \ \ ---- " at 0.18,6 left
set label "2nd sample (321-614) {\bf ----}" at 0.18,5.5  left
#set xtics
plot "stplot2.dat" w l,"stplot3.dat" w l
```

dird.in

```
n1 <- 9
n2 <- 8
x1<- rep(0,n1)
x2<- rep(0,n2)
 chord<-function(theta,x)
{1-mean(cos(x-rep(theta,length(x))))}

x1[1]<-75
x1[2]<-75
x1[3]<-80
x1[4]<-80
x1[5]<-80
x1[6]<-95
x1[7]<-130
x1[8]<-170
x1[9]<-210
x2[1]<-10
x2[2]<-50
x2[3]<-55
x2[4]<-65
x2[5]<-90
x2[6]<-285
x2[7]<-325
x2[8]<-355
```

```
x1 <- x1*3.14159/180
x2 <- x2*3.14159/180

cbar<- function(x)
{mean(cos(x))}
sbar<- function(x)
{mean(sin(x))}
rbar<- function(x)
{sbar(x)/cbar(x)}
popind<-function(theta,x1,x2,a,b,d)
{if( a*chord(theta,x1) > b*chord(theta,x2)+d ) u<-0 else u <-1
u }
aer<-function(x1,x2,a,b,d)
{errpro<-rep(0,3)
n1 <- length(x1)
n2 <- length(x2)
for(i in 1:n1)
 {
   y <- x1[x1 != x1[i]]
    u<-popind(x1[i],y,x2,a,b,d)-1
    errpro[1] <- errpro[1]+(u*u),
 }
for(i in 1:n2)
 {
   y <- x2[x2 != x2[i]]
    errpro[2] <- errpro[2]+popind(x2[i],x1,y,a,b,d)
 }
errpro[3] <- (errpro[1]+errpro[2])/(n1+n2)
errpro[1] <- errpro[1]/n1
errpro[2] <- errpro[2]/n2

errpro
}
aer(x1,x2,1,1,0)
```

Chapter 11

Figures

11.1 Introduction

The figures shown here were all drawn using GNUPLOT programs which can be used to produce LATEX files. Figure 1 shows a 3-d surface plot of the real parts of e.c.f.s (extended to the complex plane i.e., the generalized Fourier Transform) of Normal and Cauchy (scale factors 1 and shift parameters 0). Figure 2 shows the imaginary part for the same set of variates generated. The underlying variates were generated using the SPLUS rnorm and rstab functions. Just 10 data points show the difference between the two distributions.

Figure 3 compares the quantile based rules with and without considering inclusion probabilities.

Figure 4 shows that the initial breakpoint to separate the two samples from the original dataset is not quite justified taking the two samples to be homogeneous with respect to the index or α.

Figure 1. Plot of real parts of c.c.f of Normal and Cauchy

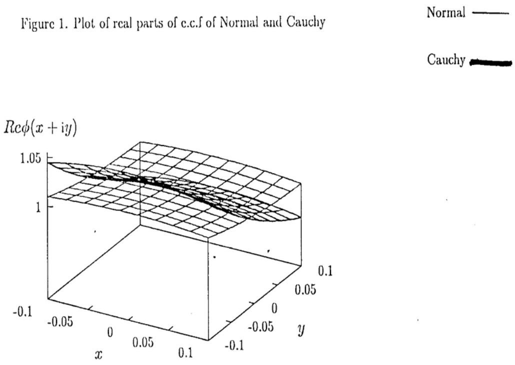
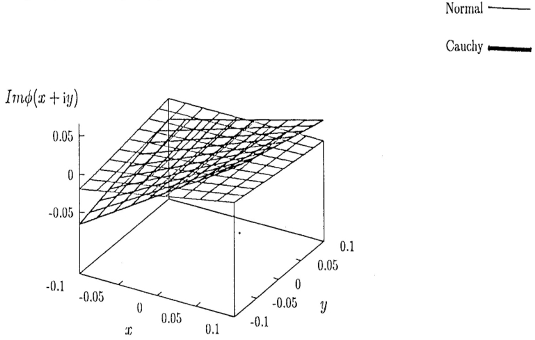


Figure 2. Plot of imaginary parts of c.c.f of Normal and Cauchy



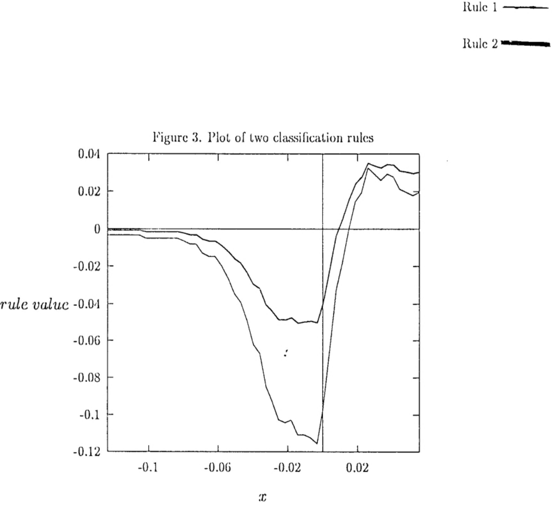
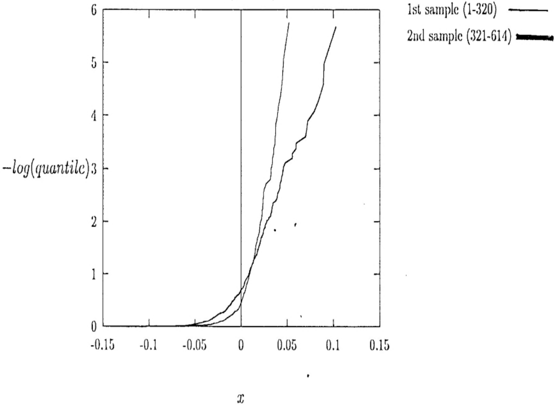


Figure 3. Plot of two classification rules

Figure 4. Plot of -log(quantile) for two samples



Bibliography

[1] Akhiezer, N.I. [1988] *Lectures on Integral Transforms*, AMS, Providence.

[2] Allgoweer, E. and Georg, K., [1980] Simplical and continuation methods for approximating fixed points and solutions to sysytems of equations. SIAM Rev., **22**, pp 28-85.

[3] Anderson,J.A. [1975] Quadratic Logistic Discrimination. Biometrika, **62**, pp 149-154.

[4] Anderson,J.A. [1979] Multivariate Logistic compounds. Biometrika **66**, pp 7-16.

[5] Anderson, J.A. [1982] *Logistic discrimination.* In Handbook of Statistics, Vol.II (Eds. P.R.Krishnaiah and L.Kanal) . North Holland, Amsterdam, pp 28-85.

[6] Anderson, J.A. and Blaire, V. [1982] Penalized Maximum Likelihood estimation in logistic regression and discrimination. Biometrika **69**, pp 123-136.

[7] Anderson, J.A. and Richardson, S.C. [1979] Logistic Discrimination and bias correction in maximum likelihood estimation. Technometrics **21**, pp 71-78.

[8] Andrews, D.F., [1973] A general method for the approximation of tail areas. Ann. Stat. **1**, pp 367-372.

[9] Andrews, D.F., Bickel, P.J., Hampel, F.K., Huber, P.J., Rogers, W.H., and Tukey J.W., [1972] Robust estimates of Location : Survey and Advances, Princeton Univ. Press, Princeton.

[10] Arad, R. [1980] Parameter estimation for symmetric stable distributions. International. Econ. Review. **21**, pp 209-220.

[11] Arnold, B.C., [1988] Pseudo-likelihood estimation for the Pareto Conditionals Distribution. Tech. Report. # 170, Dept. of Stat., Univ. of California., Riverside.

[12] Arnold, B.C. and Strauss, D. [1988] Bivariate distributions with exponential conditionals. J. Amer. Statist. Assoc., **83**, pp 522-527.

[13] Arnold, B.C., and Strauss, D.[1988] Pseudo-likelihood estimation. Tech. Report. # 164, Dept. of Stat., Univ. of California., Riverside

[14] Ashikaga, T., and Chang, P.C. [1981] Robustness of Fisher's linear discriminant function under two-xomponentmixed normal models. J. Amer.Statist. Assoc. **76**, pp 676-680.

[15] Bard, Y., [1974] *Nonlinear Parameter Estimation.* Academic Press, New York.

[16] Bahadur, R.R. [1960] Asymptotic efficiency of tests and estimates, Sankhya A, **22**, pp 229-252.

[17] Bahadur, R.R. [1967] Rates of Convergence of estimates and Test Statistics, Ann. Math. Statist. **38**, pp 303 - 324.

[18] Bahadur, R.R. [1980] On large deviations of maximum likelihood and realted estimates, Tech. Rep. # 121, Univ. of Chicago.

[19] Bahadur, R.R., Gupta, J.C., and Zabell, S.L.,[1980] Large deviations, tests and estimates., In *Asymptotic theory of statistical tests and estimation.* (J.M. Chakravarti, Ed.), Academic Press, New York.

[20] Barndorff-Nielsen, O. and Cox, D.R. [1979] Edgeworth and saddle-point approximations with statistical applications, J. Roy. Statist. Soc., B **41** pp 279-312.

[21] Barndorff-Nielsen, O.E., and Sorensen, M. [1991] Information quantities in non-classical settings Comp. Stat and Data Analysis, **12**, pp 143-158.

[22] Billingsley, P. [1968] *Convergence of Probability Measures.*, Wiley, Int., New York.

[23] Bilodeau, M. and Kariya, T. [1993] LBI tests of independence in a bivariate exponential distribution. J. Roy. Statist. Assoc., **68**, pp 704-706.

[24] Brockwell, B. [1981] High efficiency estimation for the positive stable laws. J. Amer. Statist. Assoc. **76**, pp 626-631.

[25] Bryant, J.L. and Paulson, A.S. [1979] Some comments on characteristic function based estimators. Sankhya A **41**, pp 109-116.

[26] Cox, D.R. [1962] Further Results on Tests of Separate Families of Hypotheses, J.Roy.Statist.Soc.(B), **24**, pp 406-424.

[27] Cox , D.R. [1966] Some procedures associated with the logistic qualitative response curve . In *Research Papers in Statistics.* (Ed. F.N. David), pp 55-71. Wiley, New York.

[28] Cox, D.R. [1972] The analysis of multivariate binary data. Appl. Statist. bf 21, pp 113-120.

[29] Crawley, D.R. [1979] Logistic Discrimination as an alternative to Fisher's Linear Discrimination. N.Z. Statist. **14**(2), pp 21-25.

[30] Daniels, H.E. [1954] Saddlepoint approximations in statistics. Ann. Math. Statist. **25**, pp 631-650.

[31] Daniels, H.E., [1976] Approximation to the distribution of Huber's robust M estimators (Abstract), European Meeting of Statisticians, Universite de Grenoble, pp 180.

[32] Dawid, A.P. [1976] Properties of diagnostic data distributions. Biometrics **32**, pp 647-658.

[33] Day, N.E. and Kerridge, D.F. [1967] A general maximum likelihood discriminant. Biometrics **23**, pp 313-323.

[34] DuMouchel, W.H. [1973] On the asymptotic normality of the MLE when sampling from a stable distribution. Ann. Statist., **1**, pp 948-957.

[35] Edgar, G.A. and Sucheston, L. [1976] Amarts : A class of Asymptotic Martingales. A. Discrete Parameter. J. Mult. Anal., **6**, pp 193-221.

[36] Efron, B., and Hinkley, D.V. [1978] Assessing the accuracy of the maximum likelihood estimator : Observed versus expected Fisher information, Biometrika, **65**, pp 457-487.

[37] Feuerverger, A and McDunnough, P. [1981] On some Fourier methods for inference. J. Amer. Statist. Assoc., **76**, pp 379-387.

[38] Feuerverger, A and Mureika, R.A. [1977] The empirical characteristic function and its applications. Ann. Statist. **5**, pp 88-97.

[39] Fu, J.C., [1973] On a theorem of Bahadur on the rate of convergence of point estimators. Ann. Statist. **1**, pp 745-749.

[40] Fu, J.C., [1975] The rate of convergence of consistent point estimators. Ann. Statist. **3**, pp 234-240.

[41] Fu, J.C., [1982] Large sample point estimation : a large deviation theory approach. Ann. Statist. **10**, pp 762-771.

[42] Geman, S., and Hwang, C. R. [1979] Nonparametric maximum likelihood estimation by the method of sieves, Reports in Pattern Analysis, no. 80, Div. Appl. Math., Brown University, Providence.

[43] Gill, P.E. and Murray, W.J. [1972] Quasi-Newton methods for unconstrained optimization. Inst. Math. Appl., **9**, pp 91-108.

[44] Godambe, V.P. and Heyde, C.C. [1987] Quasi-likelihood and optimal estimation, Intnl. Statist. Review., **55**, pp 231-244.

[45] Goldstein, M., and Dillon, W. R. [1978] *Discrete Discriminant Analysis.* Wiley, New York..

[46] Gong, G and Samaniego, F. [1981]Pseudo Maximum Likelihood Estimation : Theory and Applications. Ann. Statist., **4**, pp 861-869.

[47] Gut, Allan and Schimdt, Klaus. [1980] *Amarts and set function processes* Lecture Notes in Mathematics, No. 1042, Springer Verlag, New York.

[48] Haff. L.R. [1986] On linear log-odds and estimation of discriminant coefficients. Commun. Statist. Theory. Meth. **15**, pp 2131-2144.

[49] Hall. P. and Heyde, C.C. [1980] *Martingale Limit Theorems and Their Application*. Academic Publishers, New York.

[50] Hand,D.J. [1993] *Artificial Intelligence : Frontiers in Statistics. AI and Statistics, III.* Chapman and Hall, Madras.

[51] Heyde, C. C. [1987] On combining quasi-likelihood estimating functions, Stochastic Process. Appl. **25**, pp 281-287.

[52] Hodges, H.L. Fitting the Logistic by the maximum likelihood, Biometrics, **14**, pp 453-461.

[53] Holevo, A.S. [1973] On the general problem of mean estimation J. Mult. Anal., **3**, pp 262-275.

[54] Huber, P.J. [1967] The behaviour of maximum likelihood estimates under non-standard conditions, Proc. 5-th Berkeley Symp., **I**, pp 221 - 233.

[55] Hutton, J.E. and Nelson, P.I. [1986] Quasi-likelihood estimation for semimartingales, Stochastic. Process. Appl. **22**, pp 245-257.

[56] Ibrakhalilov, I.S. [1975] On estimates of functionally related parameters. Theor. Prob. Math. Statist. **6**, pp 59-72.

[57] Janardhan,K.G. and Rao, B.R. [1983]. Lagrange Distrinutions of the second kind and weighted distributions. SIAM J. Appl. Math., **43**, pp 302-313.

[58] Johnson, N.L. and Kotz, S. [1970] *Distributions in Statistics.* Wiley., New York.

[59] Kellermeier, John. [1980] The empirical characteristic function and large sample hypothesis testing. J. Mult. Anal. **10**, pp 78-87.

[60] Klecka, W.R. [1980] *Discriminant Analysis.* Sage Publ., Beverly Hills, California.

[61] Klopf, A.H., [1982] *The Hedonistic Neuron : A theory of Memory, Learning and Intelligence*, Hemisphere Publ. Corp., New York.

[62] Kraft, C. and LeCam, L., [1956] A remark on the roots of the maximum likelihood equation, Ann. Math. Stat., **27**, pp 1174 - 1177.

[63] Lachhenbruch, P.A. [1975] *Discriminant Analysis*, Hafner. New York.

[64] LeCam, L. [1970] On the assumptions used to prove asymptotic normality of maximum likelihood estimates, Ann. Math. Stat. **41**, pp 802 - 828.

[65] Lee, T. S. and Kozin, F. [1977] Almost sure asymptotics likelihood theory, J. Appl. Prob., **14**, pp 527 - 537.

[66] Linnik, Ju. V. and Ostrovskii, I. V. [1977] Decomposition of Random Variables and Vectors. AMS Trans. of Math Mono. **48**, Providence.

[67] Lukacs, E. [1967] Analytical methods in probability theory. In *Symposium on Probability Methods in Analysis.* Lecture Notes in Mathematics, **31**. Springer, New York, pp 208-238.

[68] Lukacs, E. [1970] *Characteristic functions.* 2nd ed. Darien, Hafner, New York.

[69] Lukacs, E. [1972] A survey of the theory of characteristic functions. Adv. Appl. Prob. **4**, pp 1-38.

[70] Lukacs, E. [1983] *Development in characteristic function theory.* Griffith, London.

[71] Mardia, K.V. [1972] *Statistics of Directional Data.* Academic Press, London.

[72] McLachlan, G.J. [1992] *Discriminant Analysis and Statistical Pattern Recognition.*, Wiley, New York.

[73] Mittnik, S., Rachev, S.T. and Paolella, M.S. [1993] Modeling Asset Returns with Alternative Stable Distributions, Econometric Reviews, **12**, pp 261-330.

[74] Morris, J.E. and Laycock, P.J. [1974] Discriminant Analysis of directional data. Biometrika, **61**, pp 335-341.

[75] Mukherjee, S.P. and Chatterjee, A. [1987] Estimation of parameters in finite mixtures of exponential families from censored data. Commun. Statist. Theor. Meth., **16**, pp 3133-3147.

[76] Perlman, M. D. [1972] On the strong consistency of approximate maximum likelihood estimators, Proc. 6th Berkeley Symp. Math. Statist. Prob., pp 263 - 281.

[77] Perng, S.S. [1978] Rate of convergence of estimators based on sample mean. Ann. Statist. **6**, pp 1048-1056.

[78] Pfanzagl, J. [1990] *Estimation in Semiparametric Models.* Lecture notes in statistics, **63**, Springer, New York.

[79] Prentice, R.L. and Pyke, R. [1979] Logistic disease incidence models and case-control studies. Biometrika **66**, pp 403-411.

[80] Rachev, S.T., [1997] A program for computing the estimates of a 4-parameter stable distribution, (Private Communication).

[81] Ramachandran, B. [1967] *Advanced Theory of Characteristic Functions.* Statistical Publishing Socirty, Calcutta.

[82] Rao, C.R. [1974] *Linear Statistical Inference and its Applications.* Wiley, New York.

[83] Rao, C.R. [1985] *A Celebration of Statistics.* The ISI Centenary Volume. Eds. Atkinson, A.C. and Fienberg, S.E. Springer-Verlag, New York. pp 543-569.

[84] Rao, J.S. [1984] Nonparametric methods in directional data. Chapter 31 in *Handbook of Statistics,* Vol. 4, Eds. P.R. Krishnaiah & P.K. Sen. Elsevier Science Publishers, Amsterdam, pp 755-770.

[85] Rosenfeld, A., Hummel, R.A., and Zucker, S.W. [1976]. Scene labeling by relaxation operations. IEEE Trans. Syst. Man. Cybern. **SMC-6**, pp 420-433.

[86] Rossberg, H.J. and Jesiak, B. [1985] In *Complex Analysis : Methods, Trends and Applications.* (Lanckau, E., and Tutschke, W. ed), North Oxford Academic, Oxford, pp 309-335.

[87] Roy, S. [1995] Classification Rules for Directional Data , Invited lecture in *Winter School on Analysis of Directional Data*, Indian Statistical Institute, Calcutta.

[88] Samorodnitsky, G. and Taqqu, M.S. [1994] *Stable Non-Gaussian Random Processes.*, Chapman and Hall, New York.

[89] SenGupta, A., and Roy, S. [1992] On Families Admitting Cox's Logistic Discrimination: Caharacterization and Estimation. Tech. Report No. 231, Dept. of Statistics and AP, University of California, Santa Barbara.

[90] SenGupta, A., and Roy, S. [1994] On Consistent and Efficient Estimation of Cox's Semiparametric Logistic Discrimination Rule. Presented at *Second International Calcutta Symposium on Probability and Statistics*, Calcutta, Dec. 1994 - Jan. 1995. Also published as Tech. Rep. No. ASC/95/5, February 1995.

[91] SenGupta, A., and Roy, S. [1995] On Families Admitting Cox's Logistic Discrimination : Characterizing the Multiparameter Multivariate Case. Tech. Report. published at I.S.I. Calcutta. Technical Report No. ASC/95/4, January 1995.

[92] SenGupta, A., and Roy, S. [1995] On Consistent and Efficient Estimation of Cox's Semiparametric Logistic Discrimination Rule-II. Tech. Report. published at I.S.I. Calcutta. Technical Report No. ASC/95/6, March 1995.

[93] SenGupta, A., and Roy, S. [1995] On a Robust Classification Rule for Reliability Models. Paper presented at *National Seminar on Survival Analysis and Reliability*, Calcutta University, June, 1995.

[94] SenGupta, A., and Roy, S. [1996] On Families Admitting Logistic Discrimination Rule : Characterizing the Multivariate, Multiparameter Case. *Under revision following favorable comments for possible publication in J. Mult. Anal.*

[95] SenGupta, A., and Roy, S. [1999] On Consistent and Efficient Estimation of Cox's Logistic Discrimination Rule. Paper submitted to Ann. Statist.

[96] SenGupta, A., and Roy, S. [1999] On Classification Rules for Stable Distributions. Submitted as Invited paper in a Special Volume on Stable Distributions, (under preparation) (S.T.Rachev and A. Vollert. eds.).

[97] Sorensen, M. [1989] Some asymptotic properties of QL estimators for semimartingales, In *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics.*, (P. Mandl and M. Huskova, Eds.) Charles University, Prague, pp 469-479.

[98] Sorensen, M. [1990] On QL for semimartingales. Stoch. Process. Appl. **35**, 1990, pp 331-346.

[99] Stein, C. [1986] Approximate computation of expectations. Inst. of Math. Statist. Lecture notes **7**, Hayward, California.

[100] Wald, A. [1949] Note on the consistency of the maximum likelihood estimate, Ann. Math. Stat. **20**, pp 30 -48.

[101] Widder, D. V., [1966] *The Laplace Transform* 7-th printing, Princeton University Press, Princeton.

[102] Zemanian, A.R. [1965] *Distribution theory and Transform analysis.* , McGrawHill, London.

[103] Zolotarev, V.M. [1986] *One Dimensional Stable Distributions.* (AMS translation of mathematical monographs), Providence.

Note : The bibliography cited above contains reading material in addition to those cited in the text of the thesis.

INDIAN STATISTICAL INSTITUTE LIBRARY

Processing slip

Source............................P/E/G date. 23.12.65 Acc. no. T 157

Author... Roy, Saptatik

Title.. On logistic and some new discrimination rules.

CHECKING :

1 In library/earlier ed./later ed.

2 Earlier volumes/nos. of the series in lib.

3 New title ✓

Suggested/Approved by

Note :

23.12.65

Signature & date

CLASSIFICATION :

Class no..............................x-ref...

Scrutiny report : Signature & date

Class no..............................x-ref...

Note :

Signature & date

CATALOGUING ;

1 Main Card

2 Supp. Cards (Subject/Title/Series/Added entries/Shelf List) [Total no. of cards]...........................

3 Mechanical processing completed on.............................

4 Cards filed

Note :

Signature & date