UNIVERSITY OF ALBERTA

ASYMPTOTIC MINIMAX PROPERTIES OF M-ESTIMATORS FOR SCALE

BY

KA HO EDEN WU

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

SPRING 1990

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

ISBN 0-315-60325-9

UNIVERSITY OF ALBERTA

RELEASE FORM

Name of Author Ka Ho Eden Wu

TITLE OF THESIS Asymptotic Minimax Properties of M-estimators for scale

DEGREE FOR WHICH THESIS WAS PRESENTED Doctor of Philosophy

YEAR OF DEGREE GRANTED Spring 1990

(Signed) Wu Ka Ho

Permanent Address:

Flat 604 Block B
Yee Tsui Court
Chai Wan
Hong Kong

DATED March 6, 1990

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Asymptotic Minimax Properties of M-estimators for Scale submitted by Ka Ho Eden Wu in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisor
D. Wiens

P. Hooper

K. Mehra

J. Sheahan

D. Jobson

External Examiner
J. Collins

Date: Jan 30, 1990

To my family

ABSTRACT

This thesis is concerned with the asymptotic minimax properties of M-estimators for scale when the underlying distribution is only approximately known. It is assumed that the underlying distribution lies within a certain convex set $\mathcal{P}$ of distributions. Two common types of $\mathcal{P}$ which are of interest are the ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$ and the Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$.

Consider Huber's (1981) theory of robust M-estimation of a scale parameter. An M-estimate of scale is defined as $S(F_n)$, where F_n is the empirical distribution function based on a sample $X_1, \ldots, X_n \sim F$, and the functional $S(F)$ is defined implicitly by

$$\int_{-\infty}^{\infty} \chi(\frac{x}{S(F)}) dF(x) = 0.$$

Under certain regularity conditions, $n^{\frac{1}{2}}[S(F_n) - S(F)]$ is asymptotically normally distributed with mean 0 and variance

$$V(\chi, F) = \frac{S^2(F) \int_{-\infty}^{\infty} \chi^2(\frac{x}{S(F)}) dF(x)}{[\int_{-\infty}^{\infty} \chi'(\frac{x}{S(F)})(\frac{x}{S(F)}) dF(x)]^2}$$

A problem which is of interest is to find an estimating function χ_0 such that

$$\sup_F V(\chi_0, F) \leq \sup_F V(\chi, F)$$

for all other χ and for all F in $\mathcal{P}$ satisfying $S_0(F) = 1$ where S_0 is the M-estimator corresponding to the estimating function χ_0. We study the cases when $\mathcal{P}$ equals to the $\mathcal{G}_\varepsilon(G)$ or $\mathcal{K}_\varepsilon(G)$. Note that χ_0 is optimal in a minimax sense as it minimizes the maximum asymptotic variance over a certain neighbourhood of distributions.

Define the asymptotic loss

$$R(\chi, F) = \frac{V(\chi, F)}{S^2(F)}.$$

Huber (1981) showed that when $\varepsilon \leq .04$,

$$R(\chi_0, F_0) > R(\chi_0, F) \qquad (*)$$

for all $F \in \mathcal{G}_\varepsilon(\Phi)$ where $\mathcal{G}_\varepsilon(\Phi)$ is the ε-contamination normal neighbourhood model, $F_0 \in \mathcal{G}_\varepsilon(\Phi)$ minimizes Fisher information for scale over $\mathcal{G}_\varepsilon(\Phi)$ and $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$. Hence the minimax property holds: the maximum (over F) value of $R(\chi, F)$ is minimized by χ_0. Here we show that (*) does not hold for large ε case ($\varepsilon > .2051$). And with the aid of some numerical calculations, (*) does not hold for $.0997 \leq \varepsilon \leq .2051$. We can also see that the role of Φ in the large ε case can be replaced by some non-normal distributions. Moreover, under fairly general conditions, we also show that (*) does not hold when F is assumed to lie in a Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$ and $F_0 \in \mathcal{K}_\varepsilon(G)$ minimizes Fisher information for scale over $\mathcal{K}_\varepsilon(G)$.

Finally, a note on Thall's (1979) paper is provided. Thall (1979) developed a theory of robust estimation of a scale parameter by reformulating Huber's (1964) location parameter results in the scale parameter context. The results were then applied to a particular problem of robust estimation of the parameter θ of the exponential distribution $\Phi_0(\frac{x}{\theta})$ where $\Phi_0(x) = 1 - e^{-x}$, $x > 0$. Unfortunately, a few mistakes were made. Here we explain why those mistakes occured and provide a correct solution to his particular problem.

Acknowledgements

The author would like to express his deepest appreciation for the invaluable guidance and encouragement provided by Professor Douglas P. Wiens during the preparation of this thesis.

Thanks are also due to Mrs. Christine Fischer for the typing of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

MINIMAX ASYMPTOTIC VARIANCE M-ESTIMATORS FOR SCALE OVER ε-CONTAMINATION NEIGHBOURHOOD MODEL $\mathcal{G}_\varepsilon(G)$ SATISFYING $S_0(F) = 1$

1.1 Introduction

This thesis is concerned with the asymptotic minimax properties of M-estimators for scale when the underlying distribution is only approximately known. It is assumed that the underlying distribution lies within a certain convex and vaguely compact set $\mathcal{P}$ of distributions. Two common types of $\mathcal{P}$ which we will discuss in this thesis are the ε-contamination neighbourhood model and the Kolmogorov neighbourhood model.

Consider Huber's (1981) theory of robust M-estimation of a scale parameter. An M-estimate of scale is defined as $S(F_n)$, where F_n is the empirical distribution function based on a sample $X_1, \ldots X_n \sim F$, and the functional $S(F)$ is defined implicitly by

$$\int_{-\infty}^{\infty} \chi\left(\frac{x}{S(F)}\right) dF(x) = 0 \tag{1.1.1}$$

where χ is an arbitrary function. Usually χ is an even function. If the distribution function F is replaced by the empirical distribution function F_n, then (1.1.1) becomes

$$\frac{1}{n} \sum_{i=1}^{n} \chi\left(\frac{x_i}{S}\right) = 0 \tag{1.1.2}$$

from which S is calculated. Note that we are concerned only with estimates that are functionals of the empirical distribution function. In most cases, this kind of consideration is enough since many of the most common estimators depend on the data only through the empirical distribution function. In particular if $\chi(x) = x^2 - 1$, then $S(F)$ is the usual standard deviation if location is zero.

Throughout this thesis, the location is assumed to be known. Without loss of generality, the location is assumed to be zero. In general when the location is unknown, under the assumption that the underlying distribution is symmetric, Huber (1981) points out that the location and scale estimates are asymptotically independent. If the data is subtracted by any consistent estimate for location, the resulting data will asymptotically depend on the scale estimate only. Hence it can be treated as a pure scale problem. However in this case, we have to restrict ourselves only to considering symmetric distributions.

Under certain regularity conditions, the most important of which are the continuity of $\chi(x)$, and the consistency of $S(F_n)$ for $S(F)$, we have that

$$n^{\frac{1}{2}}(S(F_n) - S(F)) \xrightarrow{w} N(0, V(\chi, F))$$

where

$$V(\chi, F) = \frac{S^2(F)\int_{-\infty}^{\infty}\chi^2\left(\frac{x}{S(F)}\right)dF(x)}{\left[\int_{-\infty}^{\infty}\chi'\left(\frac{x}{S(F)}\right)\left(\frac{x}{S(F)}\right)dF(x)\right]^2} \tag{1.1.3}$$

$$(= E_F[IC^2(x, F, S)]).$$

See Theorem 2.2 of Boos and Serfling (1980) or Serfling (1980) Chapter 7 for details. The notation $IC(x, F, S)$ will be used for the influence function or influence curve of the estimator S at F [see Hampel (1974)].

Let ε be fixed $, 0 < \varepsilon \leq 1$,

G be a symmetric known distribution function with finite

Fisher information for scale with density g,

H be an arbitrary unknown distribution function with density h.

Then the ε-contamination neighbourhood may be defined as a set $\mathcal{G}_\varepsilon(G)$ of distribution functions containing all F where $F = (1-\varepsilon)G + \varepsilon H$. That is

$$\mathcal{G}_\varepsilon(G) = \{F : \ F = (1-\varepsilon)G + \varepsilon H\} \tag{1.1.4}$$

The first problem we try to solve here is to find an estimating function χ_0 such that

$$\sup_F V(\chi_0, F) \le \sup_F V(\chi, F) \tag{1.1.5}$$

for all other χ and for all F in $\mathcal{G}_\varepsilon(G)$ satisfying $S_0(F) = 1$ where S_0 is the M-estimator corresponding to the estimating function χ_0.

Note that χ_0 is optimal in a minimax sense as it minimizes the maximum asymptotic variance over a certain neighbourhood of distributions. Furthermore, the solution can be achieved by first finding the least informative (also called least favourable) distribution F_0, that is the distribution which minimizes Fisher information for scale over $\mathcal{G}_\varepsilon(G)$, and then putting $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$ (the score function) where $f_0 = F_0'$. The corresponding M-estimator for scale is then asymptotically efficient at F_0.

Usually, in order to check whether the minimax property (1.1.5) holds, it is common to investigate the saddlepoint property

$$V(\chi_0, F) \le V(\chi_0, F_0) = \frac{1}{I(F_0)} \le V(\chi, F_0) \tag{1.1.6}$$

for all χ and $F \in \mathcal{G}_\varepsilon(G)$ satisfying $S_0(F) = 1$, as (1.1.6) is sufficient for (1.1.5) to hold. The notation $I(F)$ denotes the Fisher information for scale at F with scale being equal to 1. The second inequality in (1.1.6) is essentially the Cramér-Rao Inequality.

Under the restriction $S_0(F) = 1$, $\frac{1}{V(\chi_0, F)}$ is a convex functional of F according to Huber (1981) Lemma 4.4.4. Put $F_t = (1-t)F_0 + tF_1$, where $F_1 \in \mathcal{G}_\varepsilon(G)$ with finite Fisher information for scale, and $0 \le t \le 1$.

As $\int \chi_0(x)dF_0(x) = 0$ and $\int \chi_0(x)dF_1(x) = 0$ imply $\int \chi_0(x)dF_t(x) = 0$, it can be seen that the set $\mathcal{G}_{\varepsilon,0}(G) = \{F \in \mathcal{G}_\varepsilon(G) | \; S_0(F) = 1\}$ is convex. Now

$$\begin{aligned}\left[\frac{d}{dt}\frac{1}{V(\chi_0, F_t)}\right]\Bigg|_{t=0} &= \int_{\infty}^{\infty} [2x\chi_0'(x) - \chi_0^2(x)]d(F_1 - F_0)(x) \\ &= \left[\frac{d}{dt}I(F_t)\right]\Bigg|_{t=0} \\ &\geq 0.\end{aligned}$$

The last inequality follows from the fact that $I(F_t)$ is a convex function of t. Note that for any parametric family of densities $f(x;\theta)$,

$$I(f_t, \theta) = \int_{-\infty}^{\infty} \frac{\left[\frac{\partial}{\partial\theta} f_t(x;\theta)\right]^2}{f_t(x;\theta)} dx$$

is a convex function of t according to Huber (1981) Lemma 4.4.4 where $f_t(x, \theta) = (1-t)f_0(x;\theta) + tf_1(x;\theta)$, $\quad 0 \le t \le 1$, and $I(f;\theta)$ denotes the Fisher information for θ at f. Consequently we have

$$V(\chi_0, F) \le V(\chi_0, F_0) \tag{1.1.7}$$

for all other $F \in \mathcal{G}_{\varepsilon,0}(G)$. This is simply the first inequality in (1.1.6). Hence we can conclude that the saddlepoint property (1.1.6) holds. This in turn implies that the minimax property (1.1.5) holds.

In Chapter II, the same kind of minimax problem will be discussed but with the Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$ instead of $\mathcal{G}_\varepsilon(G)$. The definition of $\mathcal{K}_\varepsilon(G)$ will be shown at the beginning of Chapter II. As we shall see, the approach we use for the Kolmogorov neighbourhood model is significantly different from the ε-contamination neighbourhood model. For the ε-contamination neighbourhood model, we shall rely on a log transformation on the data, changing the scale problem entirely to a location one. But for the Kolmogorov neighbourhood model, a direct approach will be used.

As we can see from expression (1.1.3), the asymptotic variance $V(\chi, F)$ depends on the arbitrary standardization of S, hence it does not provide a good quantity in comparing the performance of the estimators. As in Huber (1981), $S_0(F) = 1$ is a rather serious restriction. A better criterion may be the asymptotic loss $R(\chi, F) = \frac{V(\chi,F)}{S^2(F)}$ which is the asymptotic variance of $\sqrt{n} \log \frac{S(F_n)}{S(F)}$. This quantity $R(\chi, F)$ was first proposed by Daniel (1920) as a measure of accuracy of a scale estimator. Bickel and Lehmann (1976) named it the standardized asymptotic variance. Note that $R(\chi, F)$ does not depend on the arbitrary standardization of S. In Chapter III, we shall discuss the minimax properties of the M-estimators for scale over $\mathcal{G}_\varepsilon(G)$ and $\mathcal{K}_\varepsilon(G)$ by using $R(\chi, F)$ as a criterion in comparing the performance of the estimators and dropping the restriction $S_0(F) = 1$.

In the literature, Huber (1964) developed the classical theory of robust M-estimation of a location parameter. In fact the theory may also be applied to the problem of robust M-estimation of a scale parameter. This is because the problem of estimating a scale parameter for a random variable X can be reduced to that of estimating a location parameter for the random variable $Y = \log X^2$ or $Y = \log |X|$. Huber (1964) carried out this approach in estimating the variance σ^2 of a normal distribution with zero mean, utilizing the transformation $\log X^2$ to estimate $\log \sigma^2$ where σ is a scale parameter of X.

With the same log-transformation technique, Thall (1979) reformulated Huber's location parameter results in the scale parameter context with nonnegative random variables being concerned. He also worked out the most robust M-estimator for the scale parameter of the exponential distribution function where the underlying distribution is assumed to be within a Kolmogorov exponential neighbourhood model. Unfortunately the least informative distribution he obtained is wrong. Detailed discussions will be appearing in Chapter II section 7.

In the past two decades, the problem of finding the least informative distribution within a certain convex set of distributions and the problem of checking whether or not various robust estimators are minimax have undoubtedly received a fairly large amount of attention. For the location case, Huber (1964) first obtained the least informative density f_0 in both the ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$ where G has a strongly unimodal density g and the Kolmogorov normal neighbourhood model $\mathcal{K}_\varepsilon(\Phi)$ with Φ the standard normal distribution function and $\varepsilon \leq .0303$. He also showed that the maximum likelihood estimate corresponding to the score function $\psi_0 = -\frac{f_0'}{f_0}$ is minimax with respect to the asymptotic variance criterion.

Collins and Wiens (1985) worked out the form of the minimax solution for more general ε-contamination models for which the known density g is not necessarily strongly unimodal. Sacks and Ylvisaker (1972) obtained further result for the Kolmogorov normal neighbourhood model for which the minimax solution was found in the extended range of $.0303 \leq \varepsilon \leq .5$. Wiens (1986) extended the result to the general Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$ where G is not necessarily normal. Collins and Wiens (1989) further extended these results to Lévy neighbourhood model in which G satisfies conditions similar to those imposed in Wiens (1986). In both of the Kolmogorov and Lévy neighbourhood model cases, all of the work done require an assumption that G is symmetric.

Moreover, research has extended to investigating whether or not the efficient L- and R-estimators corresponding to the least informative distributions are minimax in a certain class of distributions. To do this, it usually involves a checking of the saddlepoint property. This is because the saddlepoint property implies the minimax property. Definitions and basic properties of L- and R-estimators can be seen in Huber (1981), Lehmann (1983). Jaeckel (1971) provided a positive answer for L- and R-estimators in ε-contamination neighbourhood of symmetric distributions. Sacks and Ylvisaker (1972, 1982) showed that the efficient L-estimator is not minimax in $\mathcal{K}_\varepsilon(\Phi)$ in $\varepsilon \geq .07$ and provided a simple convex class of distributions for which there is neither an L- nor R-estimate that is asymptotically minimax. Collins (1983) established that R-estimation is minimax in $\mathcal{K}_\varepsilon(\Phi)$. Collins and Wiens (1989) extended the result to Lévy neighbourhood model and showed that minimax property holds for R-estimates but fails for L-estimates in Lévy neighbourhood model. Wiens (1990) derived an easily checked necessary condition for L-estimation to be minimax and a related sufficient condition for R-estimation to be minimax. Those cases in the literature in which L-estimation is known not to be minimax and those in which R-estimation is minimax are derived as consequences of these conditions. Besides, Wiens (1987) showed that weighted Cramér-von Mises estimation is also minimax in ε-contamination neighbourhood of symmetric distributions.

For the scale case, the only reported results are those of Huber (1981), who explicitly finds out the least informative distribution for the ε-contamination normal neighbourhood $\mathcal{G}_\varepsilon(\Phi)$ and verifies (with the aid of some numerical calculations) that the minimax property using the loss function $R(\chi, F)$ holds for $\mathcal{G}_\varepsilon(\Phi)$

and $\varepsilon \leq .04$. In this thesis, we try to extend Huber's (1981) result and work on a more general ε-contamination neighbourhood model as well as in Kolmogorov neighbourhood model.

In the next section, we shall also rely on a log transformation of the data, changing the scale problem entirely to a location one and obtaining results for the ε-contamination neighbourhood model in some more general situations.

It is worthy to note that the M-estimate of scale $S(F_n)$ is estimating its own asymptotic value. As the value of $S(F)$ depends on the choice of the estimating function χ and the distribution function F, it may vary from one to another. In this thesis, our concern is to make a suitable choice among those M-estimates of scale rather than to estimate a scale parameter. This same kind of problem has been raised and discussed, especially by Bickel and Lehmann (1975), in a more general context.

1.2 Theory and examples

Definition

Define $F_\sigma(x) = F(\frac{x}{\sigma})$, $\sigma > 0$ and define Fisher information for scale σ of a distribution F on the real line by

$$I(F,\sigma) = \frac{1}{\sigma^2} \sup_{\chi} \frac{[\int_{-\infty}^{\infty} x\chi'(x)dF(x)]^2}{\int_{-\infty}^{\infty} \chi^2(x)dF(x)} \tag{1.2.1}$$

for fixed $\sigma > 0$ where the supremum is taken over the set C_k' of all continuously differentiable functions with compact support, satisfying $\int_{-\infty}^{\infty} \chi^2(x)dF(x) > 0$.

Note that if we restrict the distribution F to be symmetric, the above definition is equivalent to the following:

Definition

Define $F_\sigma(x) = F(\frac{x}{\sigma})$, $\sigma > 0$ and define the Fisher information for scale σ of a symmetric distribution F on the real line by

$$I(F,\sigma) = \frac{1}{\sigma^2} \sup_{\chi} \frac{\left(\int_{-\infty}^{\infty} \chi'(x) dF^*(x) \right)^2}{\int_{-\infty}^{\infty} \chi^2(x) dF^*(x)} \tag{1.2.2}$$

for fixed $\sigma > 0$ where $F^*(x) = 2F(e^x) - 1$ and the supremum is taken over the set C_k' of all continuously differentiable functions with compact support satisfying $\int_{-\infty}^{\infty} \chi^2(x) dF^*(x) > 0$.

Definition (1.2.2) is a natural extension of the classical expression

$$\begin{aligned} I(F,\sigma) &= \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\sigma} \log \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) \right]^2 \frac{1}{\sigma} f(\frac{x}{\sigma}) dx \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left[-x \frac{f'}{f}(x) - 1 \right]^2 f(x) dx \end{aligned}$$

where $f = F'$.

The definition (1.2.2) is motivated by taking the logarithm of the absolute value of the data, so as to convert to a location problem. Making it clearer, let X be distributed as $F(\frac{x}{\sigma})$ where F is a symmetric distribution. Then $\log|X|$ is distributed as $F^*(\log x - \log\sigma)$ and $I(F,\sigma) = \frac{1}{\sigma^2} I_L(F^*, \log\sigma)$. The notation $I_L(F^*, \log\sigma)$ denotes the Fisher information for location $\log\sigma$ at F^* where $F^*(x) = 2F(e^x) - 1$. Thus our definition of the Fisher information for scale σ corresponds to the extended version of the Fisher information for location $\log\sigma$ described as the Huber (1981) Chapter 4.

Theorem 1.1. *(A)The following two statements are equivalent:*

(i) $I(F,\sigma) < \infty$

(ii) F has a density, absolutely continuous on $\mathbb{R} \setminus \{0\}$, *satisfying*

$$\int (-x\frac{f'}{f}(x) - 1)^2 f(x)dx < \infty$$

In either case,

$$I(F, \sigma) = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} (-x\frac{f'}{f}(x) - 1)^2 f(x)dx.$$

It is worth mentioning that $I(F, \sigma) < \infty$ *implies* $xf(x) \to 0$ *as* $x \to 0$ *or* $\pm\infty$.

(B) There is an $F_0 \in \mathcal{P}$ *minimizing* $I(F, \sigma)$.

(C) If $0 < I(F_0, \sigma) < \infty$ *and the set where* $f_0 = F_0'$ *is strictly positive is convex, then* F_0 *is unique.*

Proof: The proof of part (A) is very similar to that of Huber (1981) Theorem 4.2. The major change is that the linear functional A defined in our case is

$$A\chi = -\int x\chi'(x)dF(x).$$

The argument of existence and uniqueness proof of the least informative distribution F_0 in (B) and (C) respectively are similar to Huber (1981) Proposition 4.3 and Propostion 4.5. □

Now if $I(F_0) = \min_{F \in \mathcal{G}_\epsilon(G)} I(F)$, then with $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$, the saddlepoint property (1.1.6) holds. This in turn implies that the minimax property (1.1.5) holds. Therefore in order to solve the above minimax problem, we only have to find the distribution F_0 that minimizes Fisher information for scale under the class $\mathcal{G}_\epsilon(G)$. This is equivalent to finding the distribution F_0^* that minimizes Fisher information for location under the class $\mathcal{G}_\epsilon^*(G)$ which is defined as

$$\mathcal{G}_\epsilon^*(G) = \{F^* : F^*(x) = F(e^x) - F(-e^x), F \in \mathcal{G}_\epsilon(G)\}.$$

Note that $\mathcal{G}_\varepsilon^*(G)$ can also be represented as a form of an ε-contamination neighbourhood model. In fact $\mathcal{G}_\varepsilon^*(G)$ can be written as

$$\mathcal{G}_\varepsilon^*(G) = \{F^* : F^* = (1-\varepsilon)G^* + \varepsilon H^*\}$$

where $G^*(x) = G(e^x) - G(-e^x)$ and $H^*(x) = H(e^x) - H(-e^x)$.

According to Huber (1981) example 5.2, a log transformation of the absolute value of the data gives:

Theorem 1.2. *Suppose $G^* \in \mathcal{G}_\varepsilon^*(G)$ has a twice differentiable density g^* such that $-\log g^*(x)$ is convex on the convex support of G^*, that is g^* is strongly unimodal. Then we have that Fisher information for scale is minimized by that F_0 satisfying:*

Case A: For large ε,

$$f_0(x) = \begin{cases} (1-\varepsilon)g(x_0)\left|\frac{x_0}{x}\right|^{1-k} & |x| \le x_0 \\ (1-\varepsilon)g(x) & x_0 \le |x| \le x_1 \\ (1-\varepsilon)g(x_1)\left|\frac{x_1}{x}\right|^{1+k} & |x| \ge x_1 \end{cases} \tag{1.2.4}$$

$$\chi_0(x) = \begin{cases} -k & |x| \le x_0 \\ -x\frac{g'}{g}(x) - 1 & x_0 \le |x| \le x_1 \\ k & |x| \ge x_1 \end{cases} \tag{1.2.5}$$

where $0 < x_0 < x_1$ are the endpoints of the interval

$$|-x\frac{g'}{g}(x) - 1| \le k$$

and k is related to ε through the relationship

$$2\int_{x_0}^{x_1} g(x)dx + \frac{2x_0g(x_0) + 2x_1g(x_1)}{k} = \frac{1}{1-\varepsilon}. \tag{1.2.6}$$

Here k is always smaller than 1.

Case B: For small ε

$$f_0(x) = \begin{cases} (1-\varepsilon)g(x) & |x| \le x_1 \\ (1-\varepsilon)g(x_1)\left|\frac{x_1}{x}\right|^{1+k} & |x| \ge x_1 \end{cases} \tag{1.2.7}$$

$$\chi_0(x) = \begin{cases} -x\frac{g'}{g}(x) - 1 & x_0 \le x_1 \\ k & |x| \ge x_1 \end{cases} \tag{1.2.8}$$

where $0 < x_1$ is the point satisfying

$$-x\frac{g'}{g}(x) - 1 = k$$

and the constant k is determined by

$$2\int_0^{x_1} g(x)dx + \frac{2x_1 g(x_1)}{k} = \frac{1}{1-\varepsilon}. \tag{1.2.9}$$

Here k is always larger than 1.

Proof: By using the transformation $Y = \log|X|$ and applying the result of Huber (1981) example 5.2, results follow. □

Note that $S_0(F_0) = 1$ follows from the fact that

$$\begin{aligned}\int_{-\infty}^{\infty} \chi_0(x)dF_0(x) &= \int_{-\infty}^{\infty} (-xf_0'(x) - f_0(x))dx \\ &= -xf_0(x)|_0^{\infty} + (-xf_0(x))|_{-\infty}^{0} \\ &= 0.\end{aligned}$$

In order to apply Theorem 1.2, we have to check whether the transformed distribution G^* is strongly unimodal or not. To do this, the following lemma plays a role.

Lemma 1.3. *For any twice differentiable function g, put $g^*(x) = 2e^x g(e^x)$. Then $-\log g^*(x)$ is convex if and only if*

$$\zeta(x) + x\zeta'(x) \geq 0$$

for all $x > 0$ where

$\zeta(x) = -\frac{g'}{g}(x)$.

Proof:

$$\begin{aligned} -\log g^*(y) &= -\log 2 - y - \log g(e^y) \\ [-\log g^*(y)]' &= -1 - e^y \frac{g'}{g}(e^y) \\ &= -1 + e^y \zeta(e^y) \\ [-\log g^*(y)]'' &= e^y[\zeta(e^y) + e^y \zeta'(e^y)] \end{aligned}$$

Therefore

$$-\log g^*(y) \text{ is convex iff } \zeta(x) + x\zeta'(x) \geq 0 \quad \forall x > 0$$

which is the required result. □

Example 1.1. Let $g(x)$ be the density of the Student t distribution with r degrees of freedom. That is

$$\begin{aligned} g(x) &= \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\Gamma(\frac{r}{2})} \left(\frac{x^2}{r} + 1\right)^{-(\frac{r+1}{2})}, \quad -\infty < x < \infty \\ &= c\left(\frac{x^2}{r} + 1\right)^{-(\frac{r+1}{2})} \text{ say} \end{aligned}$$

Then

$$\begin{aligned} g^*(x) &= 2e^x g(e^x) = 2e^x c\left(\frac{e^{2x}}{r} + 1\right)^{-(\frac{r+1}{2})} \\ -\log g^*(x) &= -\log 2c - x + \left(\frac{r+1}{2}\right)\log\left(\frac{e^{2x}}{r} + 1\right) \\ [-\log g^*(x)]' &= -1 + (r+1)\frac{e^{2x}}{(e^{2x} + r)} \\ [-\log g^*(x)]'' &= \frac{(r+1)2re^{2x}}{(e^{2x} + r)^2} \geq 0 \quad \forall x \in (-\infty, \infty) \end{aligned}$$

It follcws that $-\log g^*(x)$ is convex on $(-\infty,\infty)$.

Example 1.2. Consider the density of the form

$$g(x) = Ae^{-|x|^\alpha/\alpha}, \quad -\infty < x < \infty$$

where $\alpha > 0$ and A is the normalizing constant. Then

$$g^*(x) = 2Ae^x e^{-|e^x|^\alpha/\alpha}$$

$$[-\log g^*(x)]' = -1 + e^{\alpha x}$$

$$[-\log g^*(x)]'' = \alpha e^{\alpha x} \geq 0 \quad \forall x \in (-\infty,\infty)$$

Thus $-\log g^*(x)$ is convex on $(-\infty,\infty)$.

Example 1.3. Consider the logistic distribution with density

$$g(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < \infty.$$

We then have

$$g'(x) = -\frac{e^{-x}(1-e^{-x})}{(1+e^{-x})^3}$$

$$\zeta(x) = -\frac{g'}{g}(x) = \frac{e^x - 1}{e^x + 1}$$

$$\zeta'(x) = \frac{2e^x}{(e^x+1)^2}$$

$$\zeta(x) + x\zeta'(x) = \frac{e^{2x} - 1 + 2xe^x}{(e^x+1)^2} > 0 \text{ on } (0,\infty)$$

since $e^{2x} > 1$ on $(0,\infty)$. According to Lemma 1.3, $-\log g^*(x)$ is convex on $(-\infty,\infty)$.

Example 1.4. (An example in which $-\log g^*(x)$ is not convex).

Consider the distribution G with density

$$g(y) = \frac{1}{2\pi|y|(1+\log^2|y|)}, \quad -\infty < x < \infty$$

Note that $G(y) = \int_{-\infty}^{y} g(y)dy$ and $G(\infty) = 1$ implies g is a density. Then

$$g^*(x) = 2e^x g(e^x) = \frac{1}{\pi(1+x^2)}$$

which is the Cauchy density. Hence $-\log g^*(x)$ is not convex. For details, see Collins and Wiens (1985).

From above, the first three examples indicate that the distribution of Student t, logistic and the distribution with density of the form $Ae^{-|x|^\alpha/\alpha}$ $(\alpha > 0)$ all satisfy the required condition of Theorem 1.2. Hence their least informative distributions can be obtained simply by applying Theorem 1.2.

In Example 1.4, since g^* is a Cauchy density and the Cauchy density is not strongly unimodal, Theorem 1.2 is no longer applicable. Further theory will be developed in order to relax the strongly unimodal requirement of g^*.

Note that $\mathcal{G}_\epsilon^*(G)$ is a convex set and

$$I_L(F^*) = \int_{-\infty}^{\infty} \left(\frac{f^{*\prime}}{f^*}\right)^2 (y) f^*(y) dy$$

is a convex functional of F^*. Huber (1981) thus concludes that $F_0^* \in \mathcal{G}_\epsilon^*(G)$ minimizes $I_L(F^*)$ if and only if

$$\frac{d}{dt} I_L(F_t^*)|_{t=0} \geq 0 \tag{1.2.10}$$

for all $F_1^* \in \mathcal{G}_\epsilon^*(G)$ with $I_L(F_1^*) < \infty$ where $F_t^* = (1-t)F_0^* + tF_1^*$. Now for $f_t^* = (1-t)f_0^* + tf_1^*$, we have

$$\begin{aligned}
\frac{d}{dt} I_L(F_t^*)|_{t=0} &= \int_{-\infty}^{\infty} \frac{d}{dt} \frac{(f_t^{*\prime}(x))^2}{f_t^*(x)} dx|_{t=0} \\
&= \int_{-\infty}^{\infty} \frac{2f_1^{*\prime}(x)[f_1^*(x) - f_0^*(x)]' f_t^*(x) - (f_t^*(x))^2 [f_1^*(x) - f_0^*(x)]}{(f_t^*(x))^2}|_{t=0} dx \\
&= \int_{-\infty}^{\infty} \left\{ 2\frac{f_0^{*\prime}(x)}{f_0^*(x)} [f_1^*(x) - f_0^*(x)]' - \left(\frac{f_0^{*\prime}(x)}{f_0^*(x)}\right)^2 [f_1^*(x) - f_0^*(x)] \right\} dx
\end{aligned}$$

Define $\psi_0^*(x) = -\frac{f_0^{*\prime}}{f_0^*}(x)$, then $I_L(F_0^*)$ attains the minimum value in $\mathcal{G}_\varepsilon^*(G)$ if and only if

$$\int_{-\infty}^{\infty} \{-2\psi_0^*(x)[f_1^*(x) - f_0^*(x)]' - (\psi_0^*(x))^2[f_1^*(x) - f_0^*(x)]\}dx \geq 0.$$

Furthermore if ψ_0^* is absolutely continuous and bounded, then

$$\int_{-\infty}^{\infty} -2\psi_0^*(x)[f_1^*(x) - f_0^*(x)]'dx$$
$$= -2\psi_0^*(x)[f_1^*(x) - f_0^*(x)]|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 2\psi_0^{*\prime}(x)[f_1^*(x) - f_0^*(x)]dx$$
$$= \int_{-\infty}^{\infty} 2\psi_0^{*\prime}(x)[f_1^*(x) - f_0^*(x)]dx.$$

The quantity $-2\psi_0^*(x)[f_1^*(x) - f_0^*(x)]|_{-\infty}^{\infty}$ vanishes to zero since $I(F,1) < \infty$ is equivalent to $I_L(F^*) < \infty$ which implies that $f^*(x) \to 0$ as $x \to \infty$ or $-\infty$. Consequently $I_L(F_0^*)$ is a minimum in $\mathcal{G}_\varepsilon^*(G)$ if and only if

$$\int_{-\infty}^{\infty} (2\psi_0^{*\prime}(x) - \psi_0^{*2}(x))(f_1^*(x) - f_0^*(x))dx \geq 0$$

for all $F_1^* \in \mathcal{G}_\varepsilon^*(G)$ with $I_L(F_1^*) < \infty$.

For the sake of the existence and the uniqueness of $F_0^* \in \mathcal{G}_\varepsilon^*(G)$, we impose some further conditions on G^*:

(G1) $0 < I_L(G^*) < \infty$.

(G2) $g^*(= G^{*\prime})$ is strictly positive on $(-\infty, \infty)$.

(G3) the function $\zeta^* = -\frac{g^{*\prime}}{g^*}$ is absolutely continuous and continuously differentiable.

A slight modification of Theorem 3 part A of Collins and Wiens (1985) provides a further set of necessary and sufficient conditions which makes the explicit determination of minimax ψ_0^* (hence χ_0) possible corresponding to a given G^* satisfying (G1)-(G3).

Theorem 1.4. *Suppose the function G^* in the model $\mathcal{G}_\varepsilon^*(G)$ satisfies (G1)-(G3). Then there exists a unique $F_0^* \in \mathcal{G}_\varepsilon^*(G)$ minimizing $I_L(F^*)$ over $\mathcal{G}_\varepsilon^*(G)$.*

In order that F_0^* minimizes $I_L(F^*)$, the following are necessary and sufficient conditions:

(A1) $B_\lambda = \{x : h_0^*(x) > 0\} \supset A_\lambda = \{x : J(\zeta^*)(x) < -\lambda\}$ *for some* $\lambda > 0$ *where* $J(\zeta^*)(x) = 2\zeta^{*\prime}(x) - (\zeta^*(x))^2$ *and* $\zeta^*(x) = -\frac{g^{*\prime}}{g^*}(x)$.

(A2) For all $x \in B_\lambda$, $J(\psi_0^*)(x) = -\lambda$ *where* $\psi_0^*(x) = -\frac{f_0^{*\prime}}{f_0^*}(x)$

(A3) $\int_{-\infty}^{\infty} f_0^*(x)dx = 1$

(A4) $\psi_0^*(x) = \zeta^*(x) \quad \forall x \in B_\lambda^c$

(A5) $f_0^*(x)$ *and* $\psi_0^*(x)$ *are continuous;* $\psi_0^*(x)$ *is absolutely continuous and bounded, piecewise continuous differentiable on* $(-\infty, \infty)$.

Proof. Parellel to Theorem 3 part A of Collins and Wiens (1985). □

Moreover since B_λ is an open set, it can be represented by the union of nonoverlapping open intervals, i.e. $B_\lambda = \bigcup_{j=1}^{N(\lambda)} B_{\lambda,j}$, then by (A2) and (A4), we can write

$$\psi_0^*(x) = \begin{cases} \zeta^*(x) & x \in B_\lambda^c \\ \xi^*(x; z_j, \lambda) & x \in B_{\lambda,j} \end{cases}$$

where ξ^* satisfies $J(\xi^*)(x; z_j, \lambda) \equiv -\lambda$ on $B_{\lambda,j}$ for any fixed z_j; the z_j is used for tracing the solution of ξ^* on each $B_{\lambda,j}$. Also we have

$$f_0^*(x) = \begin{cases} (1-\varepsilon)g^*(x) & x \in B_\lambda^c \\ (1-\varepsilon)\left\{ \sup_{B_{\lambda,j}} \left[\frac{g^*(x)}{k^*(x; z_j, \lambda)} \right] \right\} k^*(x; z_j, \lambda) & x \in B_\lambda \end{cases}$$

where each k^* satisfies $\xi^* \equiv -\frac{k^{*\prime}}{k^*}$ and $\sup_{B_{\lambda,j}} \left(\frac{g^*}{k^*}\right)(x)$ is attained at each nonzero finite endpoint on $B_{\lambda,j}$. With some modifications of Theorem 6, Collins and Wiens (1985) we obtain the following theorem.

Theorem 1.5. *Suppose that for all $\lambda > 0$, A_λ is of the form $(c_\lambda, d_\lambda) \cup (-d_\lambda, -c_\lambda)$, $0 < c_\lambda < d_\lambda$. Then there exists a unique pair $(z, \overline{\lambda})$ where $z \in [-\infty, \infty]$ and $0 \leq \overline{\lambda} \leq \inf\{\lambda |\ A_\lambda = \phi\}$ such that the pair (ψ_0^*, f_0^*) defined below, satisfies the conditions (A1)-(A5).*

(i) $$\psi_0^*(x) = \begin{cases} \zeta^*(x) & x \in (0, a_{\overline{\lambda}}] \cup [b_{\overline{\lambda}}, \infty) \cup [-a_{\overline{\lambda}}, 0) \cup (-\infty, -b_{\overline{\lambda}}] \\ \xi^*(x, z, \overline{\lambda}) & x \in [a_{\overline{\lambda}}, b_{\overline{\lambda}}] \cup [-b_{\overline{\lambda}}, -a_{\overline{\lambda}}] \end{cases}$$

where

$$J(\xi^*)(x) = -\overline{\lambda}$$

$$a_{\overline{\lambda}} = \sup\{x \leq c_{\overline{\lambda}} |\ (\xi^* - \zeta^*)(x) = 0\}$$

$$b_{\overline{\lambda}} = \inf\{x \geq d_{\overline{\lambda}} |\ (\xi^* - \zeta^*)(x) = 0\}$$

(ii) $$f_0^*(x) = \begin{cases} (1-\varepsilon)g^*(x) & x \in (0, a_{\overline{\lambda}}] \cup [b_{\overline{\lambda}}, \infty) \cup [-a_{\overline{\lambda}}, 0) \cup (-\infty, -b_{\overline{\lambda}}] \\ (1-\varepsilon)sk^*(x) & x \in [a_{\overline{\lambda}}, b_{\overline{\lambda}}] \cup [-b_{\overline{\lambda}}, -a_{\overline{\lambda}}] \end{cases}$$

where

$$\xi^* = -\frac{k^{*\prime}}{k^*}$$

and

$$s = \sup_{[a_{\overline{\lambda}}, b_{\overline{\lambda}}]} \left(\frac{g^*}{k^*}\right)(x) = \begin{cases} \left(\frac{g^*}{k^*}\right)(a_{\overline{\lambda}}) & \text{if } b_{\overline{\lambda}} = 0 \\ \left(\frac{g^*}{k^*}\right)(b_{\overline{\lambda}}) & \text{if } a_{\overline{\lambda}} = 0 \\ \left(\frac{g^*}{k^*}\right)(a_{\overline{\lambda}}) = \left(\frac{g^*}{k^*}\right)(b_{\overline{\lambda}}) & \text{if } 0 < a_{\overline{\lambda}} < b_{\overline{\lambda}} < \infty \end{cases}$$

(iii) $\frac{1-\varepsilon}{\varepsilon} \int_{a_{\overline{\lambda}}}^{b_{\overline{\lambda}}} [sk^*(x) - g^*(x)]dx \leq \frac{1}{2}$ with equality holding when $\overline{\lambda} > 0$.

We return to consider the situation in example 1.4. As in Collins and Wiens (1985) example 3.2,

$$\zeta^*(x) = \frac{2x}{1+x^2},$$
$$J(\zeta^*)(x) = \frac{4(1-2x^2)}{(1+x^2)^2}$$
$$A_\lambda = \begin{cases} (c_\lambda, d_\lambda) \cup (-d_\lambda, -c_\lambda) & \lambda < \frac{4}{3} \\ \phi & \lambda \geq \frac{4}{3} \end{cases}$$

where $c_\lambda^2 = \frac{4-\lambda-2\sqrt{4-3\lambda}}{\lambda}$, $\quad d_\lambda^2 = \frac{4-\lambda+2\sqrt{4-3\lambda}}{\lambda}$.

In this case we can apply Theorem 1.5 and the optimal pair is given by

$$f_0^*(x) = \begin{cases} (1-\varepsilon)\frac{1}{\pi(1+x^2)} & x \notin [a_{\overline{\lambda}}, b_{\overline{\lambda}}] \cup [-b_{\overline{\lambda}}, -a_{\overline{\lambda}}] \\ (1-\varepsilon)s\ \cosh\Big[-\frac{\sqrt{\overline{\lambda}}}{2}(x-w)\Big] & \text{otherwise} \end{cases}$$

where $s = \sup_{(a_{\overline{\lambda}}, b_{\overline{\lambda}})}\left(\frac{h^*}{k^*}\right)(x) = \left(\frac{h^*}{k^*}\right)(a_{\overline{\lambda}}) = \left(\frac{h^*}{k^*}\right)(b_{\overline{\lambda}})$

$$\frac{1-\varepsilon}{\varepsilon}\int_{a_{\overline{\lambda}}}^{b_{\overline{\lambda}}}[sk^*(x) - h^*(x)]dx = \frac{1}{2}$$

$$\psi_0^*(x) = \begin{cases} \frac{2x}{1+x^2} & x \notin [a_{\overline{\lambda}}, b_{\overline{\lambda}}] \cup [-b_{\overline{\lambda}}, -a_{\overline{\lambda}}] \\ \sqrt{\overline{\lambda}}\tanh\Big[-\frac{\sqrt{\overline{\lambda}}}{2}(x-w)\Big] & \text{otherwise} \end{cases}$$

and the five constants $\overline{\lambda}$, w, s, $a_{\overline{\lambda}}$, $b_{\overline{\lambda}}$ are determined by the side conditions that both ψ_0^*, f_0^* be continuous at $a_{\overline{\lambda}}$ and $b_{\overline{\lambda}}$ and that $\int_{-\infty}^{\infty} f_0^*(x)dx = 1$.

Using the relationship $f_0^*(x) = 2e^x f_0(e^x)$, $f_0^{*\prime}(x) = 2e^{2x}f_0'(e^x) + 2e^x f_0(e^x)$ and $\psi_0^*(x) = -\frac{f_0^{*\prime}}{f_0^*}(x) = -e^x\frac{f_0'}{f_0}(e^x) = \chi_0(e^x)$, we have

$$f_0(y) = \begin{cases} (1-\varepsilon)\frac{1}{2\pi|y|(1+\log^2|y|)}, & y \notin [e^{a_{\overline{\lambda}}}, e^{b_{\overline{\lambda}}}] \cup [e^{-b_{\overline{\lambda}}}, e^{-a_{\overline{\lambda}}}] \\ & \cup[-e^{b_{\overline{\lambda}}}, -e^{a_{\overline{\lambda}}}] \\ & \cup[-e^{-a_{\overline{\lambda}}}, -e^{-b_{\overline{\lambda}}}] \\ (1-\varepsilon)\frac{1}{2\pi|y|}s\cosh^2\Big[-\frac{\sqrt{\overline{\lambda}}}{2}(\log|y| - w)\Big], & \text{otherwise;} \end{cases}$$

$$\chi_0(y) = \begin{cases} \frac{2\log|y|}{1+\log^2|y|}, & y \notin [e^{a\frac{1}{\sqrt{\lambda}}}, e^{b\frac{1}{\sqrt{\lambda}}}] \\ & \quad \cup [e^{-b\frac{1}{\sqrt{\lambda}}}, e^{-a\frac{1}{\sqrt{\lambda}}}] \cup [-e^{b\frac{1}{\sqrt{\lambda}}}, -e^{a\frac{1}{\sqrt{\lambda}}}] \\ & \quad \cup [-e^{-a\frac{1}{\sqrt{\lambda}}}, -e^{-b\frac{1}{\sqrt{\lambda}}}] \\ \sqrt{\lambda}\tanh\left[-\frac{\sqrt{\lambda}}{2}(\log|y| - w)\right], & \text{otherwise} . \end{cases}$$

CHAPTER II

MINIMAX ASYMPTOTIC VARIANCE M-ESTIMATORS FOR SCALE OVER KOLMOGOROV NEIGHBOURHOOD MODEL $\mathcal{K}_\varepsilon(G)$ SATISFYING $S_0(F) = 1$

2.1 Introduction

Let G be a fixed but arbitrary symmetric distribution, $\varepsilon > 0$ being fixed ($\varepsilon \leq .25$). Then the Kolmogorov neighbourhood model, $\mathcal{K}_\varepsilon(G)$, may be defined as a set of distribution functions containing all F satisfying $\sup_{x\in\mathbb{R}} |F(x) - G(x)| \leq \varepsilon$. That is

$$\mathcal{K}_\varepsilon(G) = \{F : \sup_{x\in\mathbb{R}} |F(x) - G(x)| \leq \varepsilon\}. \tag{2.1.1}$$

Again, let the functional S be an M-estimator for scale. The problem we are interested in here is to find the minimax asymptotic variance M-estimator for scale over the Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$ satisfying $S_0(F) = 1$.

To solve this minimax problem, as described in Chapter 1, we just have to find the least informative distribution which minimizes Fisher information for scale over the class $\mathcal{K}_\varepsilon(G)$. In the ε-contamination neighbourhood model case $\mathcal{G}_\varepsilon(G)$, as we have seen, one can rely on the transformation $Y = \log|X|$. It is found to work in quite general situation as mentioned in Chapter 1.

In the Kolmogorov neighbourhood model case, although it is easily checked that $F \in \mathcal{K}_\varepsilon(G)$ if and only if $F^* \in \mathcal{K}^*_{2\varepsilon}(G^*)$ where

$$F^*(x) = F(e^x) - F(-e^x),$$

$$\mathcal{K}^*_{2\varepsilon}(G^*) = \{F^* : \sup_{x\in\mathbb{R}} |F^*(x) - G^*(x)| \leq 2\varepsilon\},$$

where $G^*(x) = G(e^x) - G(-e^x)$, there is no corresponding location theory to solve the problem. This is because although the distribution function G is assumed to be symmetric, the transformed distribution function G^* may not be and the existing location theory for the Kolmogorov case can only deal with a symmetric "central" distribution function. This fact is quite different from the ε-contamination neighbourhood model case in which Huber (1981) example 5.2 gives the least informative distributions in ε-contamination neighbourhood models even in the asymmetric case.

In this chapter, a direct approach to the problem will be introduced without using the usual log transformation technique. First of all, we obtain some properties which are necessary and sufficient conditions for a distribution function F_0 to minimize Fisher information for scale over $\mathcal{K}_\varepsilon(G)$.

2.2 Necessary and sufficient properties of F_0 in $\mathcal{K}_\varepsilon(G)$

Note that $\mathcal{K}_\varepsilon(G)$ is a convex class of distributions. If we define $F_t = (1-t)F_0 + tF_1$, $t \in (0,1)$, then $F_0, F_1 \in \mathcal{K}_\varepsilon(G)$ implies $F_t \in \mathcal{K}_\varepsilon(G)$. As $I(F_t)$ is a convex function in t, we can conclude that $F_0 \in \mathcal{K}_\varepsilon(G)$ attains the minimum Fisher information for scale if and only if

$$\frac{d}{dt} I(F_t)|_{t=0} \geq 0$$

for all $F_1 \in \mathcal{K}_\varepsilon(G)$ satisfying $I(F_1) < \infty$. Now

$$\begin{aligned} I(F_t) &= \int_{-\infty}^{\infty} (-x\frac{f_t'}{f_t}(x) - 1)^2 f_t(x)dx \\ &= \int_{-\infty}^{\infty} (x^2\frac{f_t'^2}{f_t}(x) + 2xf_t'(x) + f_t(x))dx \end{aligned}$$

$$\frac{d}{dt}I(F_t)|_{t=0} = \int_{-\infty}^{\infty}\{x^2[2\frac{f_0'}{f_0}(x)(f_1'-f_0')(x) - (\frac{f_0'}{f_0})^2(x)(f_1-f_0)(x)] + 2x(f_1'-f_0')(x) + (f_1-f_0)(x)\}dx$$
$$= \int_{-\infty}^{\infty}\{2x(x\frac{f_0'}{f_0}(x)+1)(f_1'-f_0')(x) - [x^2(\frac{f_0'}{f_0})^2(x)-1](f_1-f_0)(x)]dx$$

Put $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$ and suppose that χ_0 is absolutely continuous and bounded. Then

$$\frac{d}{dt}I(F_t)|_{t=0} = \int_{-\infty}^{\infty} -2x\chi_0(x)(f_1'-f_0')(x)dx - \int_{-\infty}^{\infty}(\chi_0^2(x)+2\chi_0(x))(f_1-f_0)(x)dx. \qquad (2.2.1)$$

Consider

$$\int_{-\infty}^{\infty} -2x\chi_0(x)(f_1'-f_0')(x)dx$$
$$= -2x\chi_0(x)(f_1-f_0)(x)|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 2(f_1-f_0)(x)(x\chi_0'(x)+\chi_0(x))dx$$
$$= \int_{-\infty}^{\infty} 2(x\chi_0'(x)+\chi_0(x))(f_1-f_0)(x)dx. \qquad (2.2.2)$$

Here we use the facts that χ_0 is bounded, $xf_1(x) \to 0$ as $x \to \pm\infty$ and $xf_0(x) \to 0$ as $x \to \pm\infty$.

Combining (2.2.1) and (2.2.2), we can conclude that $F_0 \in \mathcal{K}_\epsilon(G)$ attains the minimum Fisher information for scale if and only if

$$\frac{d}{dt}I(F_t)|_{t=0} = \int_{-\infty}^{\infty} J(\chi_0)(x)d(F_1-F_0)(x) \geq 0 \qquad (2.2.3)$$

for all $F_1 \in \mathcal{K}_\epsilon(G)$ with $I(F_1) < \infty$ where

$$\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$$

and

$$J(\chi)(x) = 2x\chi'(x) - \chi^2(x)$$

provided that χ_0 is absolutely continuous and bounded. Extend J by left continuity where χ' is discontinuous.

Assumptions on G:

(A1) G is symmetric and strictly increasing on $(-\infty, \infty)$ with $G(\infty) = 1$.

(A2) G has an absolutely continuous density g which is strictly decreasing on $(0, \infty)$.

(A3) $\xi(x) = -x\frac{g'}{g}(x) - 1$ is absolutely continuous on $(-\infty, \infty)$, with an absolutely continuous derivative $\xi'(x)$; $\xi(0) = -1$ and $\lim_{x\to 0} x\xi''(x) = 0$.

(A4) $0 < I(G) < \infty$

(A5) For arbitrary $a \in \mathbb{R}$

$$\inf_{[a,\infty)} [2x\xi'(x) - \xi^2(x)] < 0$$

Note that for most of the common cases , (A1) - (A5) are satisfied.

2.3 Motivation of getting the least informative distribution F_0

Partition the support B of f_0 in $(0, \infty)$ into three parts, say B_0, B_L, B_U where

$$B_0 = \{x : \max(0, G(x) - \varepsilon) < F_0(x) < \min(1, G(x) + \varepsilon)\}$$

$$B_L = \{x : F_0(x) = G(x) - \varepsilon\}$$

$$B_U = \{x : F_0(x) = G(x) + \varepsilon\}$$

$$B = B_0 \cup B_L \cup B_U$$

Since B_0 is an open set, it can be represented as a union of countably many disjoint open intervals, that is $B_0 = \bigcup_{i\in I} B_i$ where the $B_i = (a_i, b_i)$ are disjoint open intervals and I is at most countable. Recall that $J(\chi)(x) = 2x\chi'(x) - \chi^2(x)$. If χ' is absolutely continuous, then define $J'(\chi)(x) = \frac{d}{dx}J(\chi)(x)$. It turns out that F_0 satisfies

$$J(\chi_0)(x) = \text{ constant for each } x \in B_0. \qquad (2.3.1)$$

Note that the constant may not be the same for each B_i. In general, if $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$, then

$$f_0(x) = \frac{e^{-\int_c^x \frac{\chi_0(t)}{t} dt}}{x}$$

where c is a constant determined by

$$\int_{-\infty}^{\infty} f_0(x) dx = 1$$

To obtain all the possible forms of χ_0 to the above equation (2.3.1), it is necessary to split into two cases.

Case (i) Suppose that $J(\chi_0)(x) = \lambda^2$, $\lambda > 0$, then

$$2x\chi_0'(x) - \chi_0^2(x) = \lambda^2$$

which implies

$$\chi_0(x) = \lambda tan\ (\frac{\lambda}{2} \log x + c)$$

and the corresponding f_0 is of the form

$$f_0(x) = \frac{c_1 cos^2(\frac{\lambda}{2} \log x + c)}{x}.$$

for some constants c and c_1.

Case (ii) Suppose $J(\chi_0)(x) = -\lambda^2$, $\lambda > 0$ then

$$\chi_0(x) = \lambda,\ -\lambda \text{ or } \frac{\lambda[1 + c_2 x^\lambda]}{[1 - c_2 x^\lambda]}$$

and the corresponding f_0 are of the forms

$$f_0(x) = c_3 x^{-(\lambda+1)},\ c_4 x^{\lambda-1} \text{ or } c_5 \frac{(1 - c_2 x^\lambda)^2}{x^{\lambda+1}}.$$

for some constants c_2, c_3, c_4 and c_5.

2.4 General result

Theorem 2.1. *If F_0 possesses the following properties, then it is the unique member of $\mathcal{K}'_\varepsilon(G)$ minimizing $I(F)$ over $\mathcal{K}_\varepsilon(G)$, where $\mathcal{K}'_\varepsilon(G) = \{F \in \mathcal{K}_\varepsilon(G) : I(F) < \infty\}$.*

(S1) $F_0 \in \mathcal{K}_\varepsilon(G)$, *$F_0$ symmetric,* $F_0(\infty) = 1$

(S2) *F_0 has an absolutely continuous density f_0 and* $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$ *is absolutely continuous on* $(-\infty, \infty)$

(S3) *There exists a sequence* $0 < b_1 \le a_2 < b_2 \le \ldots \le a_{n-1} < b_{n-1} \le a_n$ *and constants* $\lambda_1, \lambda_2, \ldots, \lambda_n$*, such that*

(i) $B_U \cup B_L = \{\bigcup_{i=1}^{n-1} [b_i, a_{i+1}]\} \cup \{\bigcup_{i=1}^{n-1} [-a_{i+1}, -b_i,]\}$

(ii)
$$J(\chi_0)(x) = \begin{cases} \lambda_1, & 0 \le x < b_1 \\ \lambda_i, & a_i < x \le b_i \quad i = 2, \ldots, n-1 \\ \lambda_n < 0, & a_n < x \\ J(\xi)(x), & b_i < x \le a_{i+1} \quad i = 1, \ldots n-1 \end{cases}$$

where

$$J(\xi)(x) = 2x\xi'(x) - \xi^2(x)$$

and

$$\xi(x) = -x\frac{g'}{g}(x) - 1$$

(iii) If $a_i \in B_L$*, then* $J(\chi_0)(a_i^+) = \lambda_i \le J(\chi_0)(a_i^-)$

If $a_i \in B_U$*, then* $J(\chi_0)(a_i^+) = \lambda_i \ge J(\chi_0)(a_i^-)$

If $b_i \in B_L$*, then* $J(\chi_0)(b_i^-) = \lambda_i \ge J(\chi_0)(b_i^+)$

If $b_i \in B_U$*, then* $J(\chi_0)(b_i^-) = \lambda_i \le J(\chi_0)(b_i^+)$

(iv) If (b_i, a_{i+1}) is nonempty and contained in $B_L[B_U]$, then $J(\xi)(x)$ is weakly decreasing [increasing] there.

Proof. (The proof is a modification of Wiens (1986) Theorem 1). We first show that if (S1) (S2), (S3) hold, then χ_0 is a bounded function. We then establish the inequality in (2.2.3) and show that $0 < I(F_0) < \infty$.

For $J(\chi_0)(x) = -\lambda^2$ $(\lambda > 0)$ on (a, ∞), the only form of χ_0 is $\chi_0(x) = \lambda$, the corresponding $f_0(x) = |\frac{a}{x}|^{1+\lambda} f_0(a)$. If we take $\chi_0(x) = -\lambda$, the corresponding $f_0(x) = c_4 x^{\lambda-1}$ on (a, ∞). Integrating from a to ∞ leads to a contradiction as f_0 should be a density. If we take $\chi_0(x) = \frac{\lambda(1+c_2x^\lambda)}{(1-c_2x^\lambda)}$, the corresponding $f_0(x) = c_5 \frac{(1-c_2x^\lambda)^2}{x^{\lambda+1}}$. When $\lambda > 1$, $\lim_{x\to\infty} f_0(x) \neq 0$ which leads to contradiction. When $0 < \lambda \leq 1$, integrating f_0 from a to ∞ gives a contradiction as f_0 should be a density. Since χ_0 is absolutely continuous on $(-\infty, \infty)$ and equals a constant on $(a, \infty) \cup (-\infty, -a)$ for some a, χ_0 is bounded.

On the interval $B_U \cup B_L$, $f_0(x) = g(x) > 0$. On the set $B_0, f_0 > 0$ since no f_0 corresponding to a solution to $J(\chi_0)(x) =$ constant can be zero while χ_0 remains bounded. Thus $f_0 > 0$ on $(-\infty, \infty)$, so that the support of the density f_0 is convex.

If $0 < I(F_0) < \infty$ and $F_0 \in \mathcal{K}_\epsilon(G)$ minimizes $I(F)$ in $\mathcal{K}_\epsilon(G)$, then by (c) of Theorem 1.1, F_0 is the unique member of $\mathcal{K}_\epsilon(G)$ minimizing $I(F)$ in $\mathcal{K}_\epsilon(G)$. This is because the set where the density f_0 is strictly positive is convex. To show $F_0 \in \mathcal{K}_\epsilon(G)$ minimizes $I(F)$ in $\mathcal{K}_\epsilon(G)$, it suffices to verify

(i) $\int_{-\infty}^{\infty} J(\chi_0)(x) d(F_1(x) - F_0(x)) \geq 0 \quad \forall F_1 \in \mathcal{K}_\epsilon(G),\ I(F_1) < \infty$

(ii) χ_0 is absolutely continuous and bounded.

Recall that, as at (2.2.2), (ii) is necessary in order to justify the integration by parts which leads to (i). To check (i), it is enough to check for symmetric

$F_1 \in \mathcal{K}_\varepsilon(G)$ with $I(F_1) < \infty$. This is because, since $I(F)$ is a convex functional of F, we have

$$I(\tilde{F}) \leq I(F) \text{ for all } F \in \mathcal{K}_\varepsilon(G).$$

Here, $\tilde{F}$ is the symmetrization of F, that is $\tilde{F} = \frac{1}{2}(F + \overline{F})$ where $\overline{F}(x) = 1 - F(-x)$. Thus $\tilde{F} \in \mathcal{K}_\varepsilon(G)$ for G is symmetric.

Put $H(x) = F_1(x) - F_0(x)$, then

$$\int_{-\infty}^{\infty} J(\chi_0)(x)d(F_1(x) - F_0(x)) = 2\int_0^{\infty} J(\chi_0)(x)dH(x).$$

With integration by parts on those nondegenerate intervals in $B_L \cup B_U$, we have

$$\begin{aligned}
\int_0^\infty J(\chi_0)(x)dH(x) &= \lambda_1 \int_0^{b_1} dH(x) + \sum_{\substack{i=1 \\ b_i < a_{i+1}}}^{n-1} \int_{b_i}^{a_{i+1}} J(\xi)(x)dH(x) \\
&\quad + \sum_{i=2}^{n-1} \lambda_i \int_{a_i}^{b_i} dH(x) + \lambda_n \int_{a_n}^{\infty} dH(x) \\
&= \sum_{i=1}^{n-1} \lambda_i H(b_i) - \sum_{i=1}^{n-1} \lambda_{i+1} H(a_{i+1}) \\
&\quad + \sum_{\substack{i=1 \\ b_i < a_{i+1}}}^{n-1} \int_{b_i}^{a_{i+1}} J(\xi)(x)dH(x) + \lambda_n H(\infty) \\
&= \sum_{\substack{i=1 \\ b_i < a_{i+1}}}^{n-1} \{\lambda_i H(b_i) - \lambda_{i+1} H(a_{i+1}) \\
&\quad + J(\xi)(a_{i+1})H(a_{i+1}) - J(\xi)(b_i)H(b_i) \\
&\quad - \int_{b_i}^{a_{i+1}} J'(\xi)(x)dH(x)\} \\
&\quad + \sum_{\substack{i=1 \\ b_i = a_{i+1}}}^{n-1} (\lambda_i - \lambda_{i+1})H(b_i) + \lambda_n H(\infty) \\
&= \sum_{\substack{i=1 \\ b+i < a_{i+1}}}^{n-1} \{[\lambda_i - J(\xi)(b_i)]H(b_i) + [J(\xi)(a_{i+1}) \\
&\quad - \lambda_{i+1}]H(a_{i+1}) - \int_{b_i}^{a_{i+1}} J'(\xi)(x)dH(x)\} \\
&\quad + \sum_{\substack{i=1 \\ b_i = a_{i+1}}}^{n-1} \{J(\chi_0)(b_i^-) - J(\chi_0)(b_i^+)\}H(b_i) + \lambda_n H(\infty).
\end{aligned}$$

By (S1) and (S3), all the above terms are nonnegative. Obviously

$$I(F_0) = \int_{-\infty}^{\infty} \chi_0^2(x)dF_0(x) > 0.$$

Point (ii) follows from the assumption (S2) and the above proof. Also

$$\begin{aligned} I(F_0) &= \int_{-\infty}^{\infty} J(\chi_0)(x)dF_0(x) \\ &\leq \sup_{x\in\mathbb{R}} J(\chi_o)(x) \\ &< \infty \end{aligned}$$

Therefore $0 < I(F_0) < \infty$. □

Theorem 2.2. *If F_0 is the unique member of $\mathcal{K}'_\varepsilon(G)$ that minimizes $I(F)$ over $\mathcal{K}_\varepsilon(G)$, then it is necessary for F_0 to possess the following properties.*

(N1) $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$ *is absolutely continuous and bounded on* $(-\infty, \infty)$.

(N2) There exists a sequence $\{\lambda_i\}_{i\in I}$ *of constants such that*

$$J(\chi_0)(x) = \begin{cases} \lambda_i & \text{on } B_i,\ i \in I \\ J(\xi)(x) & \text{on each non-degenerate interval in } B \setminus B_0 \end{cases}$$

(N3) $\int_{-\infty}^{\infty} J(\chi_0)(x)d(F(x) - F_0(x)) \geq 0$ *for all* $F \in \mathcal{K}'_\varepsilon(G)$

(N4) (i) If $x_0 \in B_L \cup B_U$, $f_0(x_0) = g(x_0)$

(ii) If $x_0 \in B_L[B_U]$, $\chi_0(x_0) \leq \xi(x_0)[\chi_0(x_0) \geq \xi(x_0)]$

(N5) Every "regular" point of B_L is a point of decrease of $J(\chi_0)(x)$, and every such point of B_U is a point of increase of $J(\chi_0)(x)$.

More precisely put $B_i = (a_i, b_i)$, $i \in I$.

(i) If a_i is not an accumulation point of $\{a_j :\ a_j < a_i\}$, and if $a_i \in B_L$, then $J(\chi_0)(a_i^-) \geq J(\chi_0)(a_i^+)$. The inequality is reversed if $a_i \in B_U$

(ii) If b_i is not an accumulation point of $\{b_j :\ b_j > b_i\}$, and if $b_j \in B_L$, then $J(\chi_0)(b_i^-) \geq J(\chi_0)(b_i^+)$. The inequality is reversed if $b_j \in B_U$.

(iii) If x_0 is an interior point of $B_L[B_U]$, then $J(\chi_0)(x) = J(\xi)(x)$ is decreasing [increasing] in a neighbourhood of x_0.

(N6) There exists $c > 0$ such that

$$\chi_0(x) = \lambda^{\frac{1}{2}} > 0 \text{ for } |x| \geq c$$

(N7) $F_0(\infty) = 1$

(N8) $B_L \cup B_U \neq \emptyset$, *hence by symmetry of* F_0,

$$B_L \neq \emptyset,\ B_U \neq \emptyset.$$

Proof. (The proof is a modification of an unpublished manuscript of Wiens in the location case).

(i) The first step of the proof is to show that χ_0 is absolutely continuous on each open interval $B_i = (a_i, b_i)$, $i \in I$, and satisfies $J(\chi_0)(x) = \lambda_i$ there. For his, let $[a, b] \subset (a_i, b_i)$ be arbitrary, and set

$$\phi(t) = 2t\chi_0(t) - \int_0^t [2\chi_0(x) + \chi_0^2(x)]dx, t \in [a, b].$$

It suffices to show that ϕ is linear, since then it is absolutely continuous with $\phi'(t) = J(\chi_0)(t) = \text{const}.$

Let $\mathcal{K}_{[a,b]}$ be the set of distributions $F \in \mathcal{K}'_\varepsilon(G)$ which agree with F_0 off of (a, b). Note that then f agrees with f_0 off of (a, b). An integration by part then gives

$$\begin{aligned} \int_a^b \phi(t)(f_0 - f)'(t)dt &= -\int_a^b (f_0 - f)(t)[2t\chi_0'(t) - \chi_0^2(t)]dt \\ &= \int_a^b J(\chi_0)(t)d(F - F_0)(t) \geq 0 \end{aligned} \tag{2.4.1}$$

for any $F \in \mathcal{K}_{[a,b]}$.

Suppose that for some $F_1 \in \mathcal{K}_{[a,b]}$, the above inequality is strict. Since

$[a,b] \subset B_0$, we can find $\alpha > 0$ such that $F_\alpha = (1+\alpha)F_0 - \alpha F_1 \in \mathcal{K}_{[a,b]}$, $f_\alpha = (1+\alpha)f_0 - \alpha f_1$. Then

$$\begin{aligned} &\int_a^b \phi(t)(f_0 - f_\alpha)'(t)dt \\ &= -\alpha \int_a^b \phi(t)(f_0 - f_1)'(t)dt \\ &< 0 \end{aligned}$$

a contradiction to (2.4.1). Thus equality holds in (2.4.1).

Now define $\phi_1(t)$ to be a linear function joining $(a, \phi(a))$ to $(b, \phi(b))$ and set $\phi_2 = \phi - \phi_1$. The linearity of ϕ_1 ensures that $\int_a^b \phi_1(t)(f_0 - f)'(t)dt = 0$ for $F \in \mathcal{K}_{[a,b]}$, so that equality in (2.4.1) implies $\int_a^b \phi_2(t)(f_0 - f)'(t)dt = 0$ for all $F \in \mathcal{K}_{[a,b]}$. This together with $\phi_2(a) = \phi_2(b) = 0$ implies $\phi_2 = 0$ on $[a,b]$ and so $\phi \equiv \phi_1$ on $[a,b]$. Hence ϕ is linear.

(ii) By definition, on $B \setminus B_0$ we have $f_0 > 0$ and $|F_0 - G| = \varepsilon$, so that

$$f_0 = g,\ \chi_0 = \xi \text{ and } J(\chi_0)(x) = J(\xi)(x)$$

on each non-degenerate interval in $B \setminus B_0$. Thus χ_0 is piecewise absolutely continuous on B and hence on all of the whole real line. If χ_0 is discontinuous at some point, (2.4.1) could be violated by choosing $[a,b]$ to contain a discontinuity. If χ_0 is unbounded, $\frac{d}{dt}I[(F_t)]|_{t=0} \geq 0$ could be violated. Thus χ_0 is continuous and bounded on the whole real line, hence everywhere absolutely continuous and bounded. As in Theorem 2.1, (N3) is easily seen.

(iii) To establish (N4). For (N4) (i) suppose $x_0 \in B_L$ but that $f_0(x_0) > g(x_0)$. Then there exists $\delta > 0$ such that $f_0(x) > g(x)$ on $[x_0 - \delta, x_0]$. Integrating this relationship and using the fact that $F_0(x_0) = G(x_0) - \varepsilon$ gives $F_0(x_0 - \delta) < G(x_0 - \delta) - \varepsilon$, which implies $F_0 \notin \mathcal{K}_\varepsilon(G)$, a contradiction. The same kind of contradiction would occur if we assume $f_0(x_0) < g(x_0)$. The proof for B_U

is identical.

For (N4)(ii), suppose that $x_0 \in B_L$ but that $\chi_0(x_0) > \xi(x_0)$, then there exists $\delta > 0$ such that $\chi_0(x) > \xi(x)$ on $[x_0, x_0 + \delta]$ which implies

$$-\frac{f_0'(x)}{f_0(x)} > -\frac{g'(x)}{g(x)}$$

on $[x_0, x_0 + \delta]$. Integrating this relationship over $[x_0, x]$, $x < x_0 + \delta$ yields

$$\log \frac{f_0(x_0)}{f_0(x)} > \log \frac{g(x_0)}{g(x)},$$

so that $g(x) > f_0(x)$ on $(x_0, x_0 + \delta)$. Integrating this last relationship and using $F_0(x_0) = G(x_0) - \varepsilon$ gives $G(x_0 + \delta) - \varepsilon > F_0(x_0 + \delta)$, a contradiction. The proof for B_U is identical.

(iv) We prove (N5)(i) and (iii), the proof of (N5)(ii) being essentially identical to that for (N5)(i). The assumptions of (N5)(i) imply that for some $\delta \in (0, b_i - a_i)$, $J(\chi_0)(x)$ is absolutely continuous on $(a_i - \delta, a_i)$, and that $J(\chi_0)(x) = \lambda_i$ on $(a_i, a_i + \delta)$. Let $F \in \mathcal{K}_{[a_i-\delta, a_i+\delta]}$ have

$$\sup_{[a_i-\delta, a_i+\delta]} |F - F_0|(x) = |F - F_0|(a_i) > 0.$$

Then (N3) becomes

$$\begin{aligned} 0 &\leq \int_{a_i-\delta}^{a_i} J(\chi_0)(x) d[F(x) - F_0(x)] + \int_{a_i}^{a_i+\delta} J(\chi_0)(x) d[F(x) - F_0(x)] \\ &= [J(\chi_0)(a_i^-) - \lambda_i](F - F_0)(a_i) - \int_{a_i-\delta}^{a_i} (F - F_0)(x) J'(\chi_0)(x) dx \end{aligned}$$

using an integration by parts. If $a_i \in B_L$,

$$J(\chi_0)(a_i^-) - \lambda_i \geq \frac{\int_{a_i-\delta}^{a_i} (F - F_0)(x) J'(\chi_0)(x) dx}{(F - F_0)(a_i)}$$

for any such F. The inequality is reversed if $a_i \in B_U$. In absolute value, the right hand term above cannot exceed

$$\frac{\delta \sup_{[a_i-\delta,a_i]} |F-F_0|(x) \sup_{[a_i-\delta,a_i]} |J'(\chi_0)(x)|}{(F-F_0)(a_i)}$$
$$= \delta \sup_{[a_i-\delta,a_i]} |J'(\chi_0)(x)|$$

which may be made arbitrarily small. This proves (N5)(i). For (N5)(iii), we give the proof for B_L. Let $\delta > 0$ be small enough that $(x_0-\delta, x_0+\delta) \subset B_L$. Let $F \in \mathcal{K}_{[x_0-\delta,x_0+\delta]}$ and choose $c \in (x_0-\delta, x_0+\delta)$. Let $F_* \in \mathcal{K}_\epsilon(G)$ agrees with F_0 off of $[c, x_0+\delta)$ and have point mass of $F_0([c, x_0+\delta))$ at c. Although $F_* \notin \mathcal{K}_{[x_0-\delta,x_0+\delta]}$ since $I(F_*) = \infty$, it is the pointwise limit of a sequence $F_n \in \mathcal{K}_{[x_0-\delta,x_0+\delta]}$. For these, (N3) gives

$$0 \le \int_{x_0-\delta}^{x_0+\delta} J(\xi)(x)d(F_n-F_0)(x) \overset{n\to\infty}{\longrightarrow} \int_{x_0-\delta}^{x_0+\delta} J(\xi)(x)d(F_*-F_0)(x)$$
$$= J(\xi)(c)F_0([c, x_0+\delta]) - \int_c^{x_0+\delta} J(\xi)(x)dF_0(x)$$

Thus

$$J(\xi)(c) \ge \frac{\int_c^{x_0+\delta} J(\xi)(x)dF_0(x)}{F_0([c, x_0+\delta])}$$
$$\ge \min_{[c,x_0+\delta]} J(\xi)(x)$$

as $F_0([c, x_0+\delta]) = 1$, and this last inequality is strict unless $J(\xi)(x)$ is constant on $[c, x_0+\delta]$. But this can only be the case for all c if $J(\xi)(x)$ is decreasing on $(x_0-\delta, x_0+\delta)$.

(v) To get (N6), we first note that B_L cannot contain a half-infinite interval $[b, \infty)$. Suppose for contradiction that it did. Let F be a symmetric distribution and equal to F_0 on $[-b, b]$ so that by (N3),

$$0 \le \int_b^\infty J(\xi)(x)d(F-F_0).$$

By assumption (A5), $J(\xi)(x)$ must assume negative values of (b, ∞). Take an interval $(x_0, x_1) \subset (b_1, \infty)$ on which $J(\xi)(x) < 0$, and impose the following further conditions on F:

(a) $F_0(x) = F(x)$, $|x| \leq x_0$,

(b) $f(x) > f_0(x)$, $x \in (x_0, x_1)$,

(c) $f(x) = f_0(x)$, $x \in (x_1, \infty)$.

These imply F_0 is substochastic and this may be possible. But then

$$\int_0^{\infty} J(\xi)(x) d(F - F_0)(x) = \int_{x_0}^{x_1} J(\xi)(x)(f(x) - f_0(x))dx < 0$$

which leads to contradiction.

Now consider the behaviour of F_0 on $[G^{-1}(1-\varepsilon), \infty)$, on which interval the upper boundary of the Kolmogorov strip is unity. If F_0 attains this upper boundary at a finite point x_0, then f_0 has finite support and is then discontinuous. This follows upon noting that $g(x) > 0$ for all x, and that no f_0 corresponding to a solution to $J(\chi_0)(x) =$ constant can descend to zero while χ_0 remains bounded. This argument also shows that f_0 is everywhere positive, so that F_0 is unique.

Thus F_0 cannot attain the upper boundary on $[G^{-1}(1-\varepsilon), \infty)$, and cannot remain on the lower boundary on this interval. There then either exists a point c such that

$$G(x) - \varepsilon < F_0(x) < 1 \; \forall \, x \geq c,$$

or F_0 returns infinitely often to the lower boundary. In the first case, $J(\chi_0)(x) = \lambda$ for $x \geq c$, so that either $\lambda < 0$ or $\chi_0 \equiv 0$ on $[c, \infty)$ by assumption (A5). If $\lambda < 0$, then $\chi_0 \equiv (-\lambda)^{\frac{1}{2}} > 0$ on $[c, \infty)$ which is the desired conclusion. If $\chi_0 \equiv 0$, then $-\frac{f_0'}{f_0}(x) = \frac{1}{x}$ which implies $f_0(x) = \frac{k}{x}$ for some constant k. Integrating f_0 from c to ∞ gives the contradiction as

$$\int_c^\infty f_0(x)dx = \int_c^\infty \frac{k}{x}dx = \infty \cdot k$$

which is impossible.

We then need only rule out the possibility of infinitely many returns to the lower boundary. Suppose this is the case on (c, ∞) for some c. At each point in (c, ∞), either $J(\chi_0)(x) = \lambda_i$ for some λ_i or $J(\chi_0)(x) = J(\xi)(x)$ and $J'(\xi)(x) \leq 0$ as at (N5). Any interval on which $J(\chi_0)(x) = J(\xi)(x)$ is bordered by points at which $J(\chi_0)(x) = \lambda_i$, and the values of $J(\xi)(x)$ lies between these λ_i's. We claim that each λ_i is nonnegative. This implies $J(\chi_0)(x) \geq 0$ on (c, ∞), contradicting (A5).

To establish the claim, let (a, b) be one of the intervals with

(i) $F_0(a) = G(a) - \varepsilon$,

(ii) $F_0(b) = G(b) - \varepsilon$,

(iii) $F_0(x) > G(x)$ on (a, b). Let $F \in \mathcal{K}_\varepsilon'(G)$ be symmetric, agrees with F_0 on $[-a, a]$ and with G on $[b, \infty)$. Then by (N3),

$$\begin{aligned} 0 &\leq \int_0^\infty J(\chi_0)(x)d(F - F_0)(x) \\ &= \int_a^b J(\chi_0)(x)d(F - F_0)(x) + \int_b^\infty J(\chi_0)(x)d(G - F_0)(x) \end{aligned}$$

The integral over $[a, b]$ is $\varepsilon\lambda_i$ for some λ_i, and the second term is zero by (N4)(i) and the assumption that F_0 returns infinitely often to the lower boundary. Thus $\lambda_i \geq 0$ and (N6) is proved.

(iv) (N7) follows easily as if F_0 is substochastic, (N3) would be violated by any fully stochastic F agreeing with F_0 on $[-c, c]$.

For (N8), if $B_L \cup B_U = \emptyset$, F_0 would remain strictly between the Kolomogorov strips on $(-\infty, \infty)$. Hence by (N2) and (A5), $J(\chi_0)(x) \equiv \lambda < 0$ on $(-\infty, \infty)$. But then

$$I(F_0) = \int_{-\infty}^{\infty} J(\chi_0)(x) dF_0(x) < 0$$

which is a contradiction. □

2.5. Some graphs of $J(\xi)(x)$

In section 2.6, we shall try to characterize the pair (χ_0, F_0) completely when $J(\xi)(x)$ is of a certain shape. At present, we look at the graphs of $J(\xi)(x)$ for some given underlying distributions G.

Recall that $J(\xi)(x) = 2x\xi'(x) - \xi^2(x)$, $\xi(x) = -x\frac{g'}{g}(x) - 1$ and g is the density of G.

Example 2.1

Consider the distribution $G_\ell(x)$ with density of the form

$$g(x) = Ae^{-|x|^\alpha/\alpha}, \quad 1 \leq \alpha \leq 2, \quad -\infty < x < \infty$$

where A is a normalizing constant. For $x > 0$,

$$\xi(x) = x^\alpha - 1$$

$$\xi'(x) = \alpha x^{\alpha - 1}$$

$$J(\xi)(x) = 2x\xi'(x) - \xi^2(x)$$

$$= -x^{2\alpha} + 2(1+\alpha)x^\alpha - 1$$

$$J(\xi)(0) = -1,$$

$$J(\xi)(\pm\infty) = -\infty \text{ and}$$

$$J(\xi)(x) = J(\xi)(-x).$$

Figure 1 shows the graph of $J(\xi)(x)$ vs. x for normal case (i.e. $\alpha = 2$).

Example 2.2

Consider the Student's t distribution with k degree of freedom where the p.d.f. is given by

$$g(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}(1+\frac{x^2}{k})^{-(\frac{k+1}{2})}, \quad -\infty < x < \infty$$

then

$$\xi(x) = \frac{k(x^2-1)}{x^2+k},$$

$$\xi'(x) = (k+1)\frac{2kx}{(x^2+k)^2}$$

$$J(\xi)(x) = \frac{-k^2x^4 + (6k^2+4k)x^2 - k^2}{(x^2+k)^2}$$

$$J(\xi)(0) = -1,$$

$$J(\xi)(\pm\infty) = -k^2 \text{ and}$$

$$J(\xi)(x) = J(\xi)(-x).$$

Figure 2 shows the graph of $J(\xi)(x)$ vs. x for Cauchy case (i.e. $k = 1$).

Example 2.3.

Consider the logistic distribution where the p.d.f. is given by

$$g(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < \infty$$

Then

$$\xi(x) = x\left(\frac{e^x-1}{e^x+1}\right) - 1,$$

$$\xi'(x) = \frac{2xe^x}{(e^x+1)^2} + \frac{e^x-1}{e^x+1}$$

$$J(\xi)(x) = \frac{x^2}{(e^x+1)^2}[4e^x - (e^x-1)^2] + 4x\left(\frac{e^x-1}{e^x+1}\right) - 1,$$

$$J(\xi)(0) = -1,$$

$$J(\xi)(\pm\infty) = -\infty \text{ and}$$

$$J(\xi)(x) = J(\xi)(-x).$$

Fig. 1. Graph of $J(\xi)(x)$ vs. x for normal case

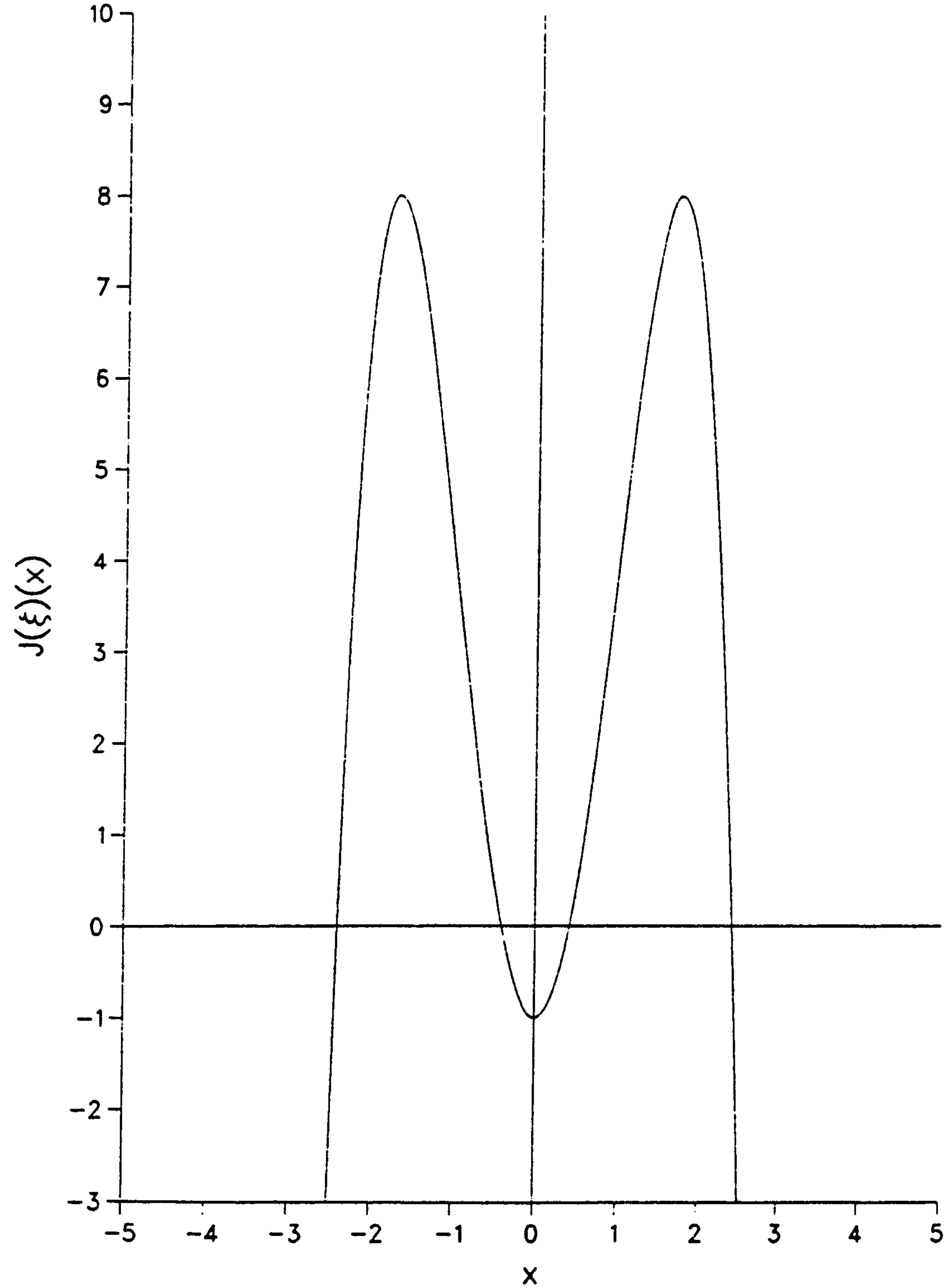


Fig. 2. Graph of $J(\xi)(x)$ vs. x for Cauchy case

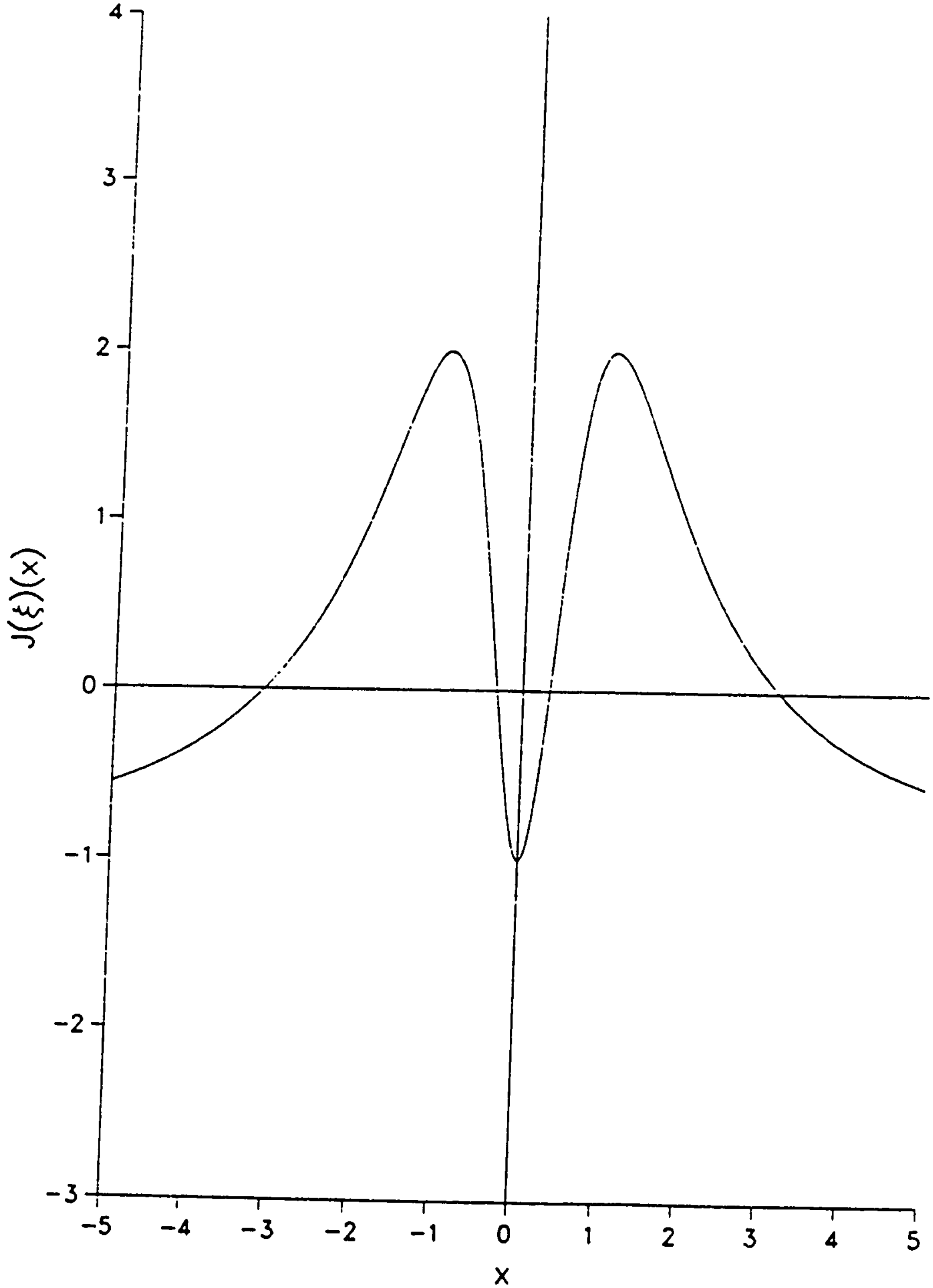


Fig. 3. Graph of $J(\xi)(x)$ vs. x for logistic case

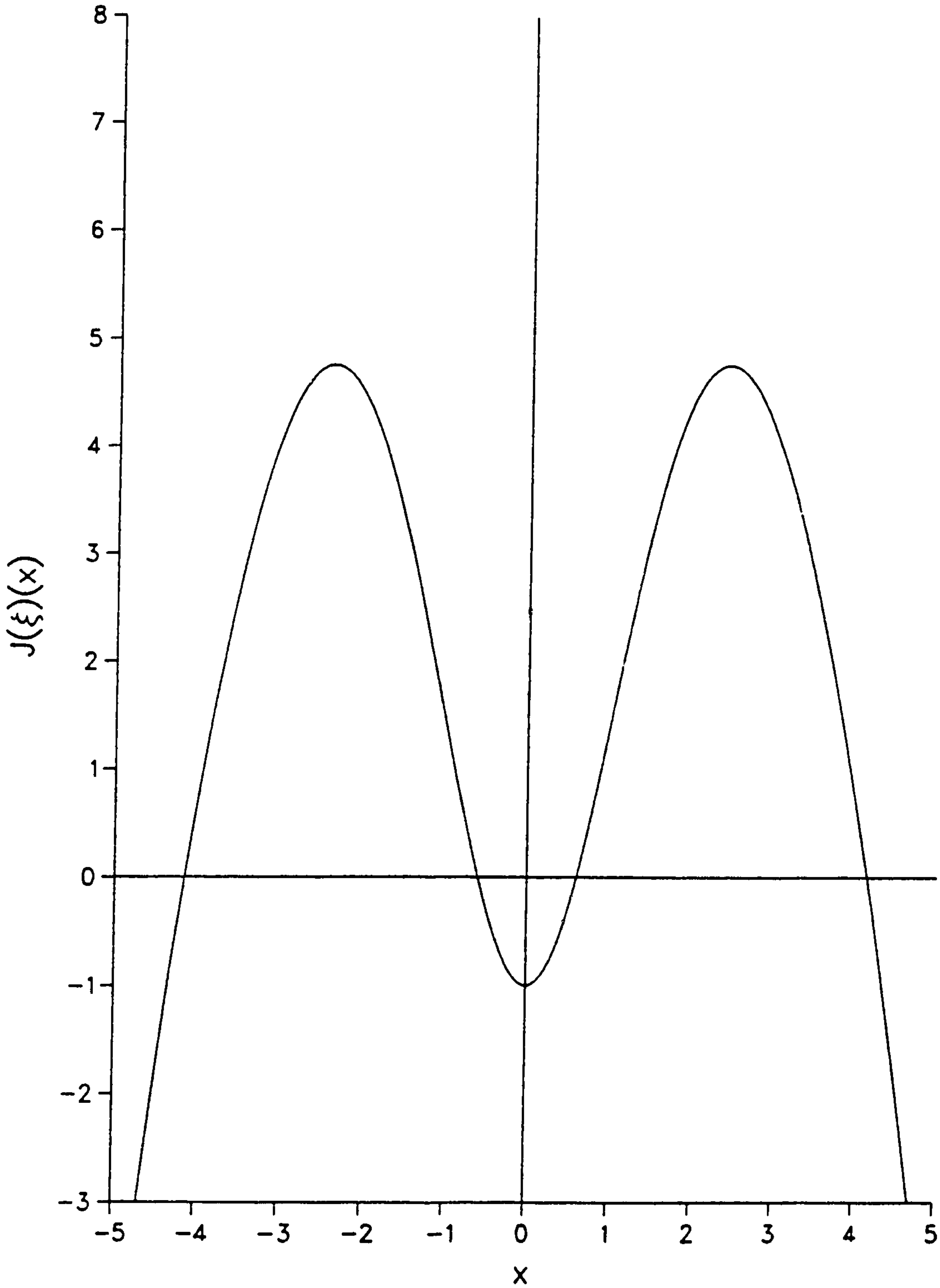


Figure 3 shows the graph of $J(\xi)(x)$ vs. x for logistic case.

2.6 Determination of (χ_0, F_0) for a type of $J(\xi)(x)$

Theorem 2.3 (Small ε case). *Suppose G is a distribution satisfying (A1) - (A5) with $J(\xi)(x) = 2x\xi'(x) - \xi^2(x)$ satisfying the following:*

(J1) $J(\xi)(x) = J(\xi)(-x)$

(J2) $J(\xi)(0) = -1,\ \frac{d^2}{dx^2} J(\xi)(x)|_{x=0} > 0$

(J3) $J(\xi)(x) \to -k$ *as* $x \to \pm\infty$ *where* k *may take any value in the interval* $[1, \infty]$.

(J4) There exists exactly one local maximum point, say $(m, J(\xi)(m))$, $m \in (0, \infty)$, *with* $J(\xi)(m) > 0$.

(J5) $J(\xi)(x)$ *is strictly increasing on* $(0, m)$ *and strictly decreasing on* (m, ∞).

Then for sufficiently small ε and subject to inequalities (ii), (iii), on p. 52, the Fisher information for scale $I(F)$ is minimized over $\mathcal{K}_\varepsilon(G)$ by F_0 with

$$\chi_0(x) = \begin{cases} \lambda & x \in [d, \infty) \\ \xi(x) & x \in [c, d) \\ \lambda_1 tan(\frac{\lambda_1}{2} \log x + \omega) & x \in [b, c) \\ \xi(x) & x \in [a, b) \\ -\lambda_2 & x \in [0, a) \\ \chi_0(-x) & x < 0 \end{cases} \tag{2.6.1}$$

$$f_0(x) = \begin{cases} k_1 x^{-(1+\lambda)} & x \in [d, \infty) \\ g(x) & x \in [c, d) \\ k_2 \frac{cos^2(\frac{\lambda_1}{2} \log x + \omega)}{x} & x \in [b, c) \\ g(x) & x \in [a, b) \\ k_3 x^{\lambda_2 - 1} & x \in [0, a) \\ f_0(-x) & x < 0 \end{cases}$$

where $\lambda, \lambda_1, \lambda_2, k_1, k_2, k_3, \omega, a, b, c, d$ *are constants determined by the continuity of* χ_0 *(B1 - B4 below), the continuity of* f_0 *(C1 - C3 below) and by (D1)-(D3) below.*

Moreover

$$\begin{aligned} I(F_0) = 2[&-\lambda_2^2(G(a) - \frac{1}{2} + \varepsilon) \\ &+ \lambda_1^2(G(0) - G(b) - 2\varepsilon) - \lambda^2(1 - G(d) + \varepsilon) \\ &+ \int_a^b J(\xi)(x)dG(x) + \int_c^d J(\xi)(x)dG(x). \end{aligned} \tag{2.6.3}$$

Note that the density of the form $g(x) = Ae^{|x|^\alpha/\alpha}$, $\alpha \in [1, 2]$, *the Student's t, and the logistic distribution all have* $J(\xi)(x)$ *satisfying (A1) - (A5) and (J1) - (J5).*

To ensure the continuity of χ_0, *we require*

(B1) $\lambda = \xi(d)$

(B2) $\lambda_1 tan(\frac{\lambda_1}{2} \log c + \omega) = \xi(c)$

(B3) $\lambda_1 tan(\frac{\lambda_1}{2} \log b + \omega) = \xi(b)$

(B4) $\xi(a) = -\lambda_2$

To ensure the continuity of f_0 *on* $(0, \infty)$, *we require*

(C1) $k_1 = d^{\lambda+1} g(d)$

(C2) $k_2 = \frac{cg(c)}{cos^2(\frac{\lambda_1}{2} \log c + \omega)} = \frac{bg(b)}{cos^2(\frac{\lambda_1}{2} \log b + \omega)}$

(C3) $k_3 = a^{1-\lambda_2} g(a)$

Also we require

(D1) $\int_0^a f_0 = G(a) - \frac{1}{2} + \varepsilon$

(D2) $\int_b^c f_0 = G(c) - G(b) - 2\varepsilon$

(D3) $\int_d^\infty f_0 = 1 - G(d) + \varepsilon$

Using (B1)-(B4), (C1)-(C4), (D1)-(D3), the constants $\lambda, \lambda_1, \lambda_2, k_1, k_2, k_3, a, b, c, d, \omega$ *can be determined.*

Proof of Theorem 2.3. (The proof of this theorem is lengthy and rather technical.) To prove the above theorem we employ the following procedures. First assume that those eleven constants exist, then show that (recall Theorem 2.1)

(i) $F_0 \in \mathcal{K}_\varepsilon(G)$

(ii) χ_0 is absolutely continuous and bounded

(iii) $\int_{-\infty}^{\infty} J(\chi_0)(x)d(F_1 - F_0) \geq 0$, $\forall\, F_1 \in \mathcal{K}_\varepsilon(G)$ with $I(F_1) < \infty$ where

$$\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1 \text{ and}$$
$$J(\chi_0)(x) = 2x\chi_0'(x) - \chi_0^2(x).$$

After showing these, we verify that those eleven constants really exist for sufficiently small ε.

Before the proof, we first prove two useful lemmas.

Lemma 2.4. *Suppose G satisfies (A1)-(A5) and* $J(\xi)(x) = 2x\xi'(x) - \xi^2(x)$ *satisfies (J1)-(J5) where* $\xi(x) = -x\frac{g'}{g}(x) - 1$. *Then* $\xi(x)$ *is increasing on* $(0, \infty)$.

Proof. Let $t_1 < t_2$ be the two positive roots of $J(\xi)(x) = 0$. First we claim that there exists a root k to the equation $\xi(x) = 0$ where $k \in (0, \infty)$. Let $T = \sup\{x > 0 :\ \xi(x) < 0\}$. If $T < \infty$, we have $\xi(T) = 0$ and we are through. Now assume $0 < t < s < T$, then

$$\xi(x) = -x\frac{g'}{g}(x) - 1 < 0 \text{ on } (t, s)$$

which implies

$$\frac{g'}{g}(x) > -\frac{1}{x} \text{ on } (t, s).$$

Integrating both side from t to s gives

$$\log \frac{g(s)}{g(t)} > \log |\frac{t}{s}|$$

This in turn implies $sg(s) > tg(t)$. If $T = \infty$, we let $s \to T$. Then $0 > tg(t)$ which is a contradiction. Therefore there exists a $k \in (0, \infty)$ such that $\xi(k) = 0$.

Let $a > 0$ be the smallest number such that $\xi(a) = 0$. Then

$$\xi'(a) \geq 0, \quad J(\xi)(a) = 2a\xi'(a) \geq 0$$

Thus $a \in [t_1, t_2]$. Suppose $x_0 > 0$ is the smallest number such that $\xi'(x) = 0$. Then

$$J(\xi)(x_0) = -\xi^2(x_0) \leq 0,$$

and

$$J'(\xi)(x_0) = 2x_0\xi''(x_0) + 2\xi'(x_0) - 2\xi(x_0)\xi'(x_0) = 2x_0\xi''(x_0) \leq 0$$

since x_0 must be a local maximum.

With $J(\xi)(x_0) \leq 0$ and $J'(\xi)(x_0) \leq 0$, we can conclude that

$$\xi(x_0) \geq 0, \quad \xi'(x_0) = 0, \quad \xi''(x_0) \leq 0$$

and ξ is increasing on $(0, x_0]$. By assumption, $g'(x) < 0$ on $(0, \infty)$, then

$$\xi(x) = -x\frac{g'}{g}(x) - 1 > -1 \text{ for all } x \in (0, \infty).$$

Since ξ has a lower bound -1 and ξ has a local maximum x_0, there must exist at least one local minimum, say, $x_1 > x_0$ or if not there must exist a point say $x_2 > x_0$ such that

$$\xi(x_2) < \xi(x_0),\ \xi'(x_2) < 0, \quad \xi''(x_2) = 0.$$

If such x_1 exists, then $\xi'(x_1) = 0$, $\xi''(x_1) \geq 0$ which implies

$$J'(\xi)(x_1) = 2x_1\xi''(x_1) + 2\xi'(x_1)[1 - \xi(x_1)] - 2x_1\xi''(x_1) \geq 0.$$

This is a contradiction as $J'(\xi)(x) \leq 0$ for all $x \geq x_0$.

If such x_2 exists, then there must exist another point $x_3 > x_2$ such that $\xi'(x_3) = 0$, $\xi''(x_3) > 0$. If no such x_3 exists, then $\lim_{x\to\infty}\xi(x) = c$ where c is a finite constant greater than -1. If $c = -1$, then $\lim_{x\to\infty}\xi(x) = -1$ which is equivalent to $\lim_{x\to\infty} x\frac{g'}{g}(x) = 0$. Then

$$\begin{aligned}
&\forall\varepsilon > 0,\ \exists N > 0\ s.t. \\
&\text{whenever } x > N,\ |x\frac{g'(x)}{g(x)}| < \varepsilon \\
&\text{so } -\varepsilon g(x) < xg'(x) \text{ whenever } x > N \\
&(1-\varepsilon)\int_N^\infty g(x)dx < -Ng(N) \text{ whenever } x > N
\end{aligned}$$

which is impossible.

Now since

$$\lim_{x\to\infty}\xi(x) = c > -1$$

we have

$$\lim_{x\to\infty} x\xi'(x) = 0.$$

If not, put $\lim_{x\to\infty} x\xi'(x) = s \neq 0$. Then

$$\forall\varepsilon > 0,\ \exists N > 0\ s.t.$$

whenever $x > N$,

$$\begin{aligned}
&|x\xi'(x) - s| < \varepsilon \\
&-\frac{\varepsilon + s}{x} < \xi'(x) < \frac{\varepsilon + s}{x} \\
&(-\varepsilon + s)\int_N^\infty \frac{dx}{x} < \int_N^\infty \xi'(x)dx < (\varepsilon + s)\int_N^\infty \frac{dx}{N} \\
&(-\varepsilon + s)[\log\infty - \log N] < c - \log N < (\varepsilon + s)(\log\infty - \log N)
\end{aligned}$$

which is impossible for a suitably chosen ε. But then

$$\lim_{x\to\infty} J(\xi)(x) = \lim_{x\to\infty}(2x\xi'(x) - \xi^2(x)) - c^2 > -1$$

which is again impossible. However if such x_3 exists, we have

$$J(\xi)(x_3) = 2x_3\xi''(x_3) + 2\xi'(x_3)[1 - \xi(x_3)] = 2x_3\xi''(x_3) > 0$$

which is a contradiction as $J(\xi)(x) \leq 0$ for all $x > x_0$. Thus we conclude that no such x_0 exists. Hence ξ is increasing on $(0,\infty)$. □

Lemm 2.5. *Put* $\xi(x) = -x\frac{g'}{g}(x) - 1$ *and* $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$. *Suppose on an interval* (a,b), $0 \leq a < b$, *we have* $\chi_0(x) \leq \xi(x)$. *Then* $\frac{g}{f_0}$ *is non-increasing on* (a,b). *In particular*

If $g(a) = f_0(a)$, *then* $g(x) \leq f_0(x)$ *on* (a,b).

If $g(b) = f_0(b)$, *then* $g(x) \geq f_0(x)$ *on* (a,b).

Proof. If $\chi_0(x) \leq \xi(x)$ on (a,b), $0 \leq a < b$, then

$$\begin{aligned}
0 &\leq \xi(x) - \chi_0(x) \\
&= -x[\frac{g'}{g}(x) - \frac{f_0'}{f_0}]
\end{aligned}$$

$$0 \geq [\frac{g'}{g}(x) - \frac{f_0'}{f_0}]$$
$$= \frac{d}{dx} \log[\frac{g(x)}{f_0(x)}]$$

therefore $\log[\frac{g(x)}{f_0(x)}]$ is non-increasing on (a,b). Thus $\frac{g(x)}{f_0(x)}$ is non-increasing on (a,b). □

Now returning to the proof of (i) $F_0 \in \mathcal{K}_\varepsilon(G)$ in Theorem 2.3. In order to prove $F_0 \in \mathcal{K}_\varepsilon(G)$, we separate $(0,\infty)$ into five regions $(0,a],(0,b],(b,c],(c,d],(d,\infty)$. In each region, it suffices to show that $|F_0(x) - G(x)| \leq \varepsilon$. Note that in the regions $(a,b]$ and $(c,d]$, we have $F_0(x) = G(x)+\varepsilon$ and $F_0(x) = G(x)-\varepsilon$ respectively. So it is enough to prove that

$$1^\circ \quad f_0(x) \geq g(x) \qquad x \in (0,a]$$
$$2^\circ \quad f_0(x) \geq g(x) \qquad x \in (d,\infty]$$
$$3^\circ \quad f_0(x) \leq g(x) \qquad x \in (b,c)$$

1° On $(0,a]$:

Put $\eta(x) = \xi(x) - \chi_0(x)$. First we show that $\eta(0^+) < 0$ which is equivalent to showing $-\lambda_2^2 > -1$. This is because

$$-\lambda_2^2 > -1$$
$$\Leftrightarrow J(\chi_0)(0^+) > J(\xi)(0^+)$$
$$\Leftrightarrow -\chi_0^2(0^+) > -\xi^2(0^+)$$
$$\Leftrightarrow [\xi(0^+) + \chi_0(0^+)][\xi(0^+) - \chi_0(0^+)] > 0$$
$$\Leftrightarrow \xi(0^+) - \chi_0(0^+) < 0$$

as $\xi(0^+) + \chi_0(0^+) = -1 - \lambda_2 < 0$ which is then equivalent to $\eta(0^+) < 0$. Note that $-\lambda_2^2 > -1$ is equivalent to $J(\chi_0)(0) > J(\xi)(0)$. Since $J(\chi_0)(x) = -\lambda_2^2$ is in a neighbourhood of 0 and $J(\xi)(x)$ is a continuous function in x, $-\lambda_2^2 > -1$ is further equivalent to $J(\chi_0)(x) > J(\xi)(x)$ on $(-a, a)$ for some $a > 0$. Now suppose by contradiction $-\lambda_2^2 \le -1$, then $J(\chi_0)(x) \le J(\xi)(x)$ on $(-a, a)$.

For $0 < \delta < a$,

$$\begin{aligned}
0 &\le \int_{-\delta}^{\delta} [J(\xi)(x) - J(\chi_0)(x)] f_0(x)dx \\
&= \int_{-\delta}^{\delta} 2x[\xi'(x) - \chi_0'(x)] f_0(x)dx - \int_{-\delta}^{\delta} [\xi^2(x) - \chi_0^2(x)] f_0(x)dx \\
&= 2x f_0(x)[\xi(x) - \chi_0(x)]|_{-\delta}^{\delta} - \int_{-\delta}^{\delta} 2[\xi(x) - \chi_0(x)][x f_0'(x) + f_0(x)]dx \\
&\quad - \int_{-\delta}^{\delta} [\xi^2(x) - \chi_0^2(x)] f_0(x)dx \\
&= 2x f_0(x)[\xi(x) - \chi_0(x)]|_{-\delta}^{\delta} + \int_{-\delta}^{\delta} 2[\xi(x) - \chi_0(x)] f_0(x)\chi_0(x)dx \\
&\quad - \int_{-\delta}^{\delta} [\xi^2(x) - \chi_0^2(x)] f_0(x)dx \\
&= 2x f_0(x)[\xi(x) - \chi_0(x)]|_{-\delta}^{\delta} - \int_{-\delta}^{\delta} [\xi(x) - \chi_0(x)]^2 f_0(x)dx
\end{aligned}$$

Therefore

$$\int_{-\delta}^{\delta} [\xi(x) - \chi_0(x)]^2 f_0(x)dx \le 2x f_0(x)[\xi(x) - \chi_0(x)]|_{-\delta}^{\delta}.$$

When $\delta \to a$, $RHS = 0$ as $\xi(a) = \chi_0(a)$. It follows that $\xi(x) = \chi_0(x)$ on $(-a, a)$ which leads to a contradiction since χ_0 is constant on $(-a, a)$. Hence $-\lambda_2^2 > -1$ which is equivalent to $\eta(0^+) < 0$.

Note that we also have $\eta(a) = 0$ and $J(\chi_0)(x) = -\lambda_2^2$ on $(0, a)$.

Next we prove that $\eta(x) < 0$ on the whole region $(0, a)$, since then we can apply Lemma 2.5 to conclude $g(x) < f_0(x)$ on $(0, a)$. Note that

$$2x\eta'(x) = [(J(\xi)(x) + \xi^2(x)) - (\chi_0^2((x) + \lambda_2^2)]$$

Differentiating with respect to x, we have

$$\begin{aligned}2[x\eta''(x)+\eta'(x)] &= J'(\xi)(x)+2\xi'(x)\xi(x)-2\chi_0'(x)\chi_0(x)\\ &= J'(\xi)(x)+2\xi(x)[\xi'(x)-\chi_0'(x)]+2\chi_0'(x)[\xi(x)-\chi_0(x)]\\ &= J'(\xi)(x)+2\xi(x)\eta'(x)+2\chi_0'(x)\eta(x) \qquad (2.6.4)\end{aligned}$$

If there exists $x_0 \in (0,a)$ such that

$$\eta(x_0)>0,\ \eta'(x_0)=0,\quad \eta''(x_0)\leq 0,$$

then (2.6.4) implies

$$2x_0\eta''(x_0)=J'(\xi)(x_0)+2\chi_0'(x_0)\eta(x_0).$$

As

$$\chi_0'(x_0)=0,\quad J'(\xi)(x_0)=2x_0\eta''(x_0)\leq 0$$

which is a contradiction. Thus $\eta(x)<0$ on $(0,a)$.

Remark: In the above proof, it also follows that

$$\eta'(a+)>0$$

which is equivalent to

$$\chi_0'(a+)<\xi'(a).$$

2° On (d,∞):

Note that

$$\chi_0(d)=\xi(d),\quad \chi_0(x)=\lambda \text{ on } (d,\infty)$$

and by Lemma 2.4, ξ is increasing on $(0,\infty)$. Hence we have

$$\chi_0(x)\leq\xi(x) \text{ on } (d,\infty).$$

Together with the property $g(d) = f_0(d)$, we conclude

$$g(x) \le f_0(x) \text{ on } (d, \infty)$$

by Lemma 2.5.

3° On (b, c):

It suffices to prove that there exists a unique $b_0 \in (b, c)$ such that

$$\chi_0(x) \ge \xi(x) \text{ on } (b, b_0);$$

$$\chi_0(x) \le \xi(x) \text{ on } (b_0, c).$$

If this is true, then with the fact that

$$f_0(b) = g(b), \quad f_0(c) = g(c)$$

and the result of Lemma 2.5 we can conclude that

$$g(x) \ge f_0(x) \text{ on } (b, b_0)$$

$$\text{and } g(x) \ge f_0(x) \text{ on } (b_0, c).$$

This implies

$$g(x) > f_0(x) \text{ on } (b, c)$$

as b_0 is a continuity point of both g and f_0. Since

$$\chi_0^2(b) = \xi^2(b) \text{ and } \chi_0^2(c) = \xi^2(c),$$

we have

$$(\chi_0 - \xi)'(b+) > 0 \text{ provided that } J(\chi_0)(b+) > J(\xi)(b)$$

$$(\chi_0 - \xi)'(c-) > 0 \text{ provided that } J(\chi_0)(c-) > J(\xi)(c)$$

Hence if

$$J(\chi_0)(c-) > J(\xi)(x) \quad (i.e. \lambda_1^2 > J(\xi)(c))$$

and

$$J(\chi_0)(b+) > J(\xi)(b) \quad (i.e. \lambda_1^2 > J(\xi)(b)),$$

then there exists $b_0 \in (b,c)$ such that $(\chi_0 - \xi)(b_0) = 0$.

To prove b_0 is unique, it is enough to prove that if there exists $z \in (b,c)$ such that $(\chi_0 - \xi)(z) = 0$ then $(\chi_0 - \xi)'(z) < 0$. Now if $(\chi_0 - \xi)(z) = 0$, then

$$(\chi_0 - \xi)'(z) = \frac{J(\chi_0)(z) - J(\xi)(z)}{2z}.$$

Again suppose $\lambda_1^2 > J(\xi)(b)$ and $\lambda_1^2 > J(\xi)(c)$. Let $0 < p < q$ be such that

$$J(\xi)(p) = J(\xi)(q) = \lambda_1^2.$$

If $z \in (p,q)$, then $J(\chi_0)(z) < J(\xi)(z)$ which implies $(\chi_0 - \xi)'(z) < 0$. We now claim that there is no root on $(b,p]$ or $[q,c)$ of the equation $(\chi_0 - \xi)(z) = 0$.

On $(b,p]$, suppose there exists at least one solution of the equation $(\chi_0 - \xi)(z) = 0$. Let $z_1 \in (b,p]$ be the smallest root then

$$(\chi_0 - \xi)(z_1) = 0 \text{ and } (\chi_0 - \xi)'(z_1) < 0$$

which implies

$$0 > (\chi_0 - \xi)'(z_1) = \frac{J(\chi_0)(z_1) - J(\xi)(z_1)}{2z_1} \geq 0$$

which is also a contradiction. Hence there is no solution in $[b,p)$ to the equation $(\chi_0 - \xi)(z) = 0$.

Similarly on $[q,c)$, suppose z_2 is the greatest number in $[q,c)$ such that $(\chi_0 - \xi)(z_2) = 0$. Then

$$0 > (\chi_0 - \xi)'(z_2) = \frac{J(\chi_0)(z_2) - J(\xi)(z_2)}{2z_2} \geq 0$$

which is also a contradiction. Hence there is no solution in $[q, c)$ to the equation $(\chi_0 - \xi)(z) = 0$.

(ii) of Theorem 2.3 follows immediately once f_0 is determined.

To prove (iii) of Theorem 2.3, put $H = F_1 - F_0$, then $H(0) = 0$ and $H(\infty) \leq 0$.

$$\begin{aligned}\int_0^\infty J(\chi_0)(x)dH(x) &= \int_0^a (-\lambda_2^2)dH(x) + \int_a^b J(\xi)(x)dH(x) + \int_b^c \lambda_1^2 dH(x) \\ &\quad + \int_c^d J(\xi)(x)dH(x) + \int_d^\infty (-\lambda^2)dH(x) \\ &= -\lambda_2^2 H(a) + \lambda_1^2[H(c) - H(d)] - \lambda^2[H(\infty) - H(d)] \\ &\quad + \int_a^b J(\xi)(x)dH(x) + \int_c^d J(\xi)(x)dH(x) \\ &= -\lambda_2^2 H(a) + \lambda_1^2[H(c) - H(b)] - \lambda^2[H(\infty) - H(d)] \\ &\quad + J(\xi)(x)H(x)|_a^b - \int_a^b H(x)dJ(\xi)(x) \\ &\quad + J(\xi)(x)H(x)|_c^d - \int_c^d H(x)dJ(\xi)(x) \\ &= [-\lambda_2^2 - J(\xi)(a)]H(a) + [J(\xi)(b) - \lambda_1^2]H(b) \\ &\quad + [\lambda_1^2 - J(\xi)(c)]H(c) + [\lambda^2 + J(\xi)(d)]H(d) \\ &\quad - \int_a^b H(x)dJ(\xi)(x) - \int_c^d H(x)dJ(\xi)(x)\end{aligned}$$

On (a,b).

$J'(\xi)(x) > 0,\ H(x) < 0$

therefore $\int_a^b H(x)dJ(\xi)(x) < 0$.

On (c,d),

$J'(\xi)(x) < 0,\ H(x) > 0$

therefore $\int_c^d H(x)dJ(\xi)(x) < 0$

Hence if

(i) $-\lambda_2^2 < J(\xi)(a)$

(ii) $\lambda_1^2 > J(\xi)(b)$

(iii) $\lambda_1^2 > J(\xi)(c)$

(iv) $\lambda^2 > J(\xi)(d)$

then

$$\int_0^\infty J(\chi_0)(x)dH(x) \geq 0$$

as

$$H(a) \leq 0, \quad H(b) \leq 0, \quad H(c) \geq 0, \quad H(d) \geq 0.$$

To prove (i), note from the remark on p. 50 that

$$\chi_0'(a+) > \xi'(a).$$

As $\xi(a) = \chi_0(a)$, we have

$$\begin{aligned} J(\xi)(a) &= 2a\xi'(a) - \xi^2(a) \\ &> 2a\chi_0'(a-) - \chi_0^2(a) \\ &= J(\chi_0)(a-) \\ &= -\lambda_2^2 \end{aligned}$$

which is (i).

Since ξ is increasing on $(0, \infty)$, in particular on (d, ∞),

$$\begin{aligned} J(\xi)(d) &= 2d\xi'(d) - \xi^2(d) \\ &> -\xi^2(d) \\ &= -\chi_0^2(d) - J(\chi_0)(d+) \\ &= -\lambda^2. \end{aligned}$$

Now (iv) follows. This leaves (ii) and (iii), which we check numerically in each individual case.

To prove the existence of the eleven constants, as k_1, k_2, k_3 are explicitly determined by (C1)-(C3) and $\lambda, \lambda_1, \lambda_2, \omega$ can be determined once a, b, c, d are known, it is enough to show the existence of a, b, c, d with $0 < a < b < c < d < \infty$ for sufficiently small ε. Rewrite $\chi_0(x), f_0(x)$ as

$$\chi_0(x) = \begin{cases} \xi(d) & x \in [d, \infty) \\ \xi(x) & x \in [c, d) \\ \lambda_1 tan(\frac{\lambda_1}{2}\log x + \omega) & x \in [b, c) \\ \xi(x) & x \in [a, b) \\ \xi(a) & x \in [0, a) \\ \chi_0(-x) & x < 0 \end{cases} \tag{2.6.5}$$

$$f_0(x) = \begin{cases} g(d)(\frac{d}{x})^{\lambda+1} & x \in [d, \infty) \\ g(x) & x \in [c, d) \\ \frac{cg(c)}{cos^2(\frac{\lambda_1}{2}\log c+\omega)} \frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{x} & x \in [b, c) \\ (or \frac{bg(b)}{cos^2(\frac{\lambda_1}{2}\log b+\omega)} \frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{x}) & \\ g(x) & x \in [a, b) \\ g(a)(\frac{a}{x})^{1-\lambda_2} & x \in [0, a) \\ \infty & x = 0 \\ f_0(-x) & x < 0 \end{cases} \tag{2.6.6}$$

From (D3)

$$\int_d^\infty f_0 = 1 + \varepsilon - G(d)$$

$$\Leftrightarrow dg(d) = \xi(d)(1 + \varepsilon - G(d)).$$

Put

$$\alpha(d) = dg(d) - \xi(d)[1 + \varepsilon - G(d)]$$

Then
$$\alpha'(d) = [dg'(d) + g(d)] - \xi'(d)[1 + \varepsilon - G(d)] + \xi(d)g(d)$$
$$= -\xi'(d)[1 + \varepsilon - G(d)] < 0 \text{ as } \xi'(d) > 0$$

Let x_0 be the number such that $\xi(x_0) = 0$, then

$$\alpha(x_0) - x_0 g(x_0) > 0,$$

$$\alpha(\infty) = -\lim_{d\to\infty} \xi'(d)[1 + \varepsilon - G(d)] < 0.$$

Hence there exists $d_0 \in (x_0, \infty)$ such that $\alpha(d_0) = 0$. As $\alpha'(d) < 0$, d_0 is unique. Thus for each $\varepsilon > 0$ $(0 < \varepsilon < 0.25)$, there exists a unique $d_0 \in (x_0, \infty)$ such that $\alpha(d_0) = 0$.

From (D1)
$$\int_0^a f_0 = G(a) - \frac{1}{2} + \varepsilon$$
$$\Leftrightarrow ag(a) = -\xi(a)[G(a) - \frac{1}{2} + \varepsilon].$$

Put

$$\beta(a) = ag(a) + \xi(a)[G(a) - \frac{1}{2} + \varepsilon]$$

Then

$$\beta(0) = -\varepsilon < 0,$$

$$\beta(x_0) = x_0 g(x_0) > 0.$$

Hence there exists $a_0 \in (0, x_0)$ such that $\beta(a_0) = 0$. Also

$$\beta'(a) = ag'(a) + g(a) + \xi'(a)[G(a) - \frac{1}{2} + \varepsilon] + \xi(a)g(a)$$
$$= -[\xi(a) + 1]g(a) + g(a) + \xi'(a)[G(a) - \frac{1}{2} + \varepsilon] + \xi(a)g(a)$$
$$= \xi'(a)[G(a) - \frac{1}{2} + \varepsilon] > 0.$$

Therefore a_0 is unique. Thus for each $\varepsilon > 0$ $(0 < \varepsilon < .25)$, there exists a unique $a_0 \in (0, x_0)$ such that $\beta(a_0) = 0$.

From (C2) and (C3)

$$\frac{cg(c)}{cos^2(\frac{\lambda_1}{2}c + \omega)} = \frac{bg(b)}{cos^2(\frac{\lambda_1}{2}b + \omega)}.$$

From (B2)

$$sec^2(\frac{\lambda_1}{2}\log c + \omega) = (\frac{\xi(c)}{\lambda_1})^2 + 1$$

From (B3)

$$sec^2(\frac{\lambda_1}{2}\log b + \omega) = (\frac{\xi(b)}{\lambda_1})^2 + 1$$

therefore

$$\lambda_1^2 = \frac{bg(b)\xi^2(b) - cg(c)\xi^2(c)}{cg(c) - bg(b)}. \tag{2.6.7}$$

From (D2)

$$\int_b^c f_0 = G(c) - G(b) - 2\varepsilon.$$

Also

$$\begin{aligned}\int_b^c f_0 &= k_2 \int_b^c \frac{cos^2(\frac{\lambda_1}{2}\log x + \omega)}{x} dx \\ &= k_2[\frac{1}{2}\log(\frac{c}{b}) + \frac{1}{2\lambda_1}\sin 2(\frac{\lambda_1}{2}\log x + \omega)|_b^c] \\ &= \frac{1}{2}\log c \frac{cg(c)}{cos^2(\frac{\lambda_1}{2}\log c + \omega)} - \frac{1}{2}\log b \frac{bg(b)}{cos^2(\frac{\lambda_1}{2}\log b + \omega)} \\ &\quad + \frac{1}{\lambda_1}cg(c)tan(\frac{\lambda_1}{2}\log c + \omega) - \frac{1}{\lambda_1}bg(b)tan(\frac{\lambda_1}{2}\log b + \omega)\end{aligned}$$

therefore

$$\begin{aligned}G(c) - G(b) - 2\varepsilon = \frac{1}{2}cg(c)[1 + \left(\frac{\xi(c)}{\lambda_1}\right)^2]\log c - \frac{1}{2}bg(b)[1 + \left(\frac{\xi(b)}{\lambda_1}\right)^2]\log b \\ + \frac{1}{\lambda_1^2}cg(c)\xi(c) - bg(b)\xi(b)]\end{aligned} \tag{2.6.8}$$

Eliminating ω from (B2) and (B3), we have

$$tan^{-1}\left(\frac{\xi(c)}{\lambda_1}\right) - tan^{-1}\left(\frac{\xi(b)}{\lambda_1}\right) = \frac{\lambda_1}{2}\log\frac{c}{b}. \tag{2.6.9}$$

Before showing that for sufficiently small $\varepsilon > 0$, there exists a unique pair (b, c) with $0 < b < c < \infty$ which satisfies (2.6.7), (2.6.8) and (2.6.9), we state a useful lemma.

Lemma 2.6 (Implicit Function Theorem) [see e.g. *O'Neil (1975)]. Suppose that $F, \frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n}$ are continuous in some δ neighbourhood of $(a_1, \ldots, a_n)$. Suppose also that $F(a_1, \ldots, a_n) = 0$, but $\frac{\partial F(a_1,\ldots,a_n)}{\partial x_n} \neq 0$. Then there exists a positive number η and a function g defined for $|x_1 - a_1| < \eta, \ldots, |x_n - a_n| < \eta$ such that $a_n = g(a_1, \ldots, a_{n-1})$ and*

(i) $F(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) = 0$ for $|x_1 - a_1| < \eta, \ldots, |x_n - a_{n-1}| < \eta$

(ii) g is continuous on $|x_1 - a_1| < \eta, \ldots, |x_n - a_{n-1}| < \eta$

(iii) For $|x_1 - a_1| < \eta, \ldots, |x_n - a_{n-1}| < \eta$ and $1 \leq j \leq n-1$

$$\frac{\partial g(x_1, \ldots x_{n-1})}{\partial x_j} = -\frac{\frac{\partial F(x_1,\ldots,x_{n-1},g(x_1,\ldots,x_{n-1}))}{\partial x_j}}{\frac{\partial F(x_1,\ldots,x_{n-1},g(x_1,\ldots x_{n-1}))}{\partial x_n}}$$

Returning to the proof of the existence and uniqueness of such a pair (b, c) for sufficiently small $\varepsilon > 0$. Put

$$P(b,c,\varepsilon) = G(c) - G(b) - 2\varepsilon - \frac{1}{2}cg(c)[1 + \left(\frac{\xi(c)}{\lambda_1}\right)^2]\log c$$
$$+ \frac{1}{2}bg(b)[1 + \left(\frac{\xi(b)}{\lambda_1}\right)^2]\log b - \frac{1}{\lambda_1^2}[cg(c)\xi(c) - bg(b)\xi(b)] \tag{2.6.10}$$

$$Q(b,c,\varepsilon) = tan^{-1}\left(\frac{\xi(c)}{\lambda_1}\right) - tan^{-1}\left(\frac{\xi(b)}{\lambda_1}\right) - \frac{\lambda_1}{2}\log(\frac{c}{b}) \tag{2.6.11}$$

where $\lambda_1 > 0$ and λ_1^2 satisfies (2.6.7).

Suppose $m > 0$ is the maximum point of $J(\xi)(x)$, $x > 0$. Put

$$h(x) = k(x)\xi^2(x)$$

where

$$k(x) = xg(x),\ \xi(x) = -x\frac{g'}{g}(x) - 1,$$

then

$$h'(x) = 2k(x)\xi(x)\xi'(x) + \xi^2(x)k(x)$$

$$k'(x) = xg'(x) + g(x) = -g(x)\xi(x)$$

therefore

$$\begin{aligned} h'(x) &= 2xg(x)\xi(x)\frac{J(\xi)(x) + \xi^2(x)}{2x} + \xi^2(x)[-g(x)\xi(x)] \\ &= J(\xi)(x)g(x)\xi(x) \end{aligned}$$

From (2.6.7),

$$\lambda_1^2 = \frac{h(b) - h(c)}{k(c) - k(b)} = \frac{\frac{-h(b)-h(c)}{b-c}}{\frac{k(c)-k(b)}{c-b}}.$$

When $c, b \to m$,

$$\begin{aligned} \lambda_1^2 \to -\frac{h'(m)}{k'(m)} &= -\left[\frac{k'(m)\xi^2(m) + 2k(m)\xi(m)\xi'(m)}{k'(m)}\right] \\ &= -\xi^2(m) - \frac{2\xi(m)\xi'(m)}{[\log k(m)]'} \end{aligned}$$

$$\begin{aligned} [\log k(x)]' &= [\log x + \log g(x)]' \\ &= \frac{1}{x} + \frac{g'}{g}(x) \\ &= -\frac{\xi(x)}{x} \end{aligned}$$

therefore

$$[\log k(m)]' = -\frac{\xi(m)}{m}$$

so

$$\begin{aligned} \lambda_1^2(b,c) &\overset{b,c\to m}{\longrightarrow} -\xi^2(m) - \frac{2m\xi(m)\xi'(m)}{-\xi(m)} \\ &= J(\xi)(m) \quad (> 0). \end{aligned} \tag{2.6.12}$$

Obviously

$$P(m,m,0)=0,\ Q(m,m,0)=0$$

and $P, \frac{\partial P}{\partial b}, \frac{\partial P}{\partial c}, \frac{\partial P}{\partial \varepsilon}$ are continuous in some neighbourhood about $(m,m,0)$. Also

$$\frac{\partial P(m,m,0)}{\partial \varepsilon} = -2 \neq 0.$$

Applying Lemma 2.6, the Implicit Function Theorem, we can then solve $P(b,c,\varepsilon) = 0$ for $\varepsilon = f(b,c)$ in some region $|b-m| < \delta$, $|c-m| < \delta$ for some δ. Substitute this into Q to obtain

$$Q(b,c,\varepsilon) = Q(b,c,f(b,c)) = H(b,c)$$

says. Note that $f(m,m) = 0$, so $H(m,m) = 0$. Again observe that $H, \frac{\partial H}{\partial b}, \frac{\partial H}{\partial c}$ are continuous about (m,m). Now we try to verify that $\frac{\partial H(m,m)}{\partial c} \neq 0$ so that we can then solve $c = g(b)$ such that $Q(b, g(b), f(b, g(b)) = 0$ in some interval $|b-m| < \delta_1$, for some δ_1, $\varepsilon = f(b, g(b))$, $c = g(b)$.

To verify $\frac{\partial H(m,m)}{\partial c} \neq 0$.

$$\begin{aligned}\frac{\partial H(b,c)}{\partial c} &= \frac{-1}{\left[1+\left(\frac{\xi(b)}{\lambda_1}\right)^2\right]}\left(\frac{\xi(b)}{\lambda_1^2}\right)\frac{\partial \lambda_1}{\partial c} - \frac{\log b}{2}\left(\frac{\partial \lambda_1}{\partial c}\right) \\ &\quad - \frac{1}{\left[1+\left(\frac{\xi(b)}{\lambda_1}\right)^2\right]}\frac{\lambda_1 \xi'(c) - \xi(c)\frac{\partial \lambda_1}{\partial c}}{\lambda_1^2} + \frac{\lambda_1}{2c} + \frac{\log c}{2}\frac{\partial \lambda_1}{\partial c} \\ &= \frac{\partial \lambda_1}{\partial c}\left[-\frac{\xi(b)}{\lambda_1^2+\xi^2(b)} + \frac{\xi(c)}{\lambda_1^2+\xi^2(c)} + \frac{1}{2}\log\frac{c}{b}\right] - \frac{\lambda_1\left[\frac{J(\xi)(c)+\xi^2(c)}{2c}\right]}{\lambda_1^2+\xi^2(c)}.\end{aligned}$$

When $b, c \to m$,

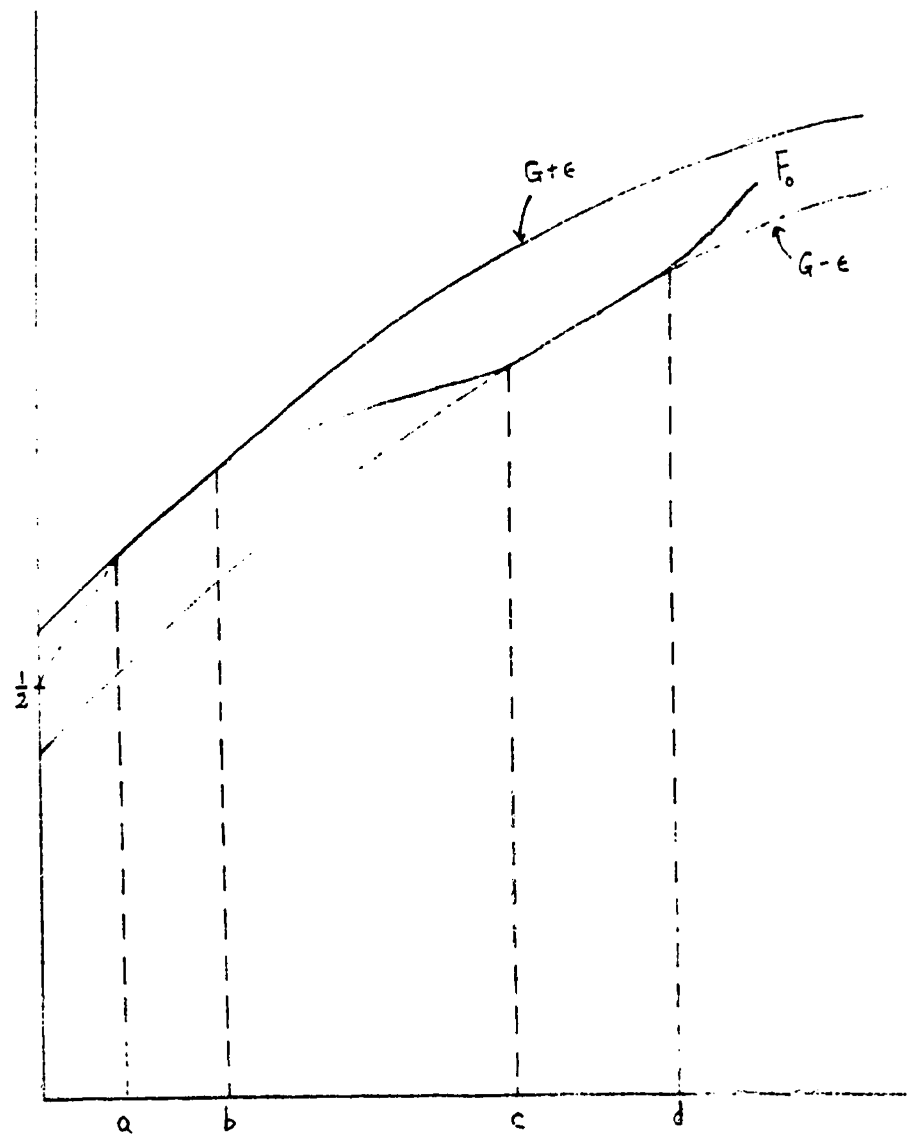
$$\frac{\partial H(b,c)}{\partial c} \to -\sqrt{\frac{J(\xi)(m)}{2m}} \neq 0.$$

From the above, we have shown that for sufficiently small ε, there exists a pair (b,c) satisfying (2.6.7), (2.6.8) and (2.6.9), with $0 < b < c < \infty$. Moreover, since

$\mathcal{K}_\epsilon(G)$ is convex and vaguely compact, the $F_0 \in \mathcal{K}_\epsilon(G)$ that minimizes $I(F)$ is unique, and hence the pair (b,c) is unique.

Figure 4 shows the least informative distribution F_0 in $\mathcal{K}_\epsilon(G)$ for small ϵ when G satisfies (A1)-(A5) and $J(\xi)(x)$ satifies (J1)-(J5).

Fig. 4: The least informative distribution F_0 in $\mathcal{K}_\epsilon(G)$ for small ϵ when G satisfies (A1)-(A5) and $J(\xi)(x)$ satisfies (J1)-(J5).



Theorem 2.7 (Large ε case). *Suppose G is a distribution satisfying (A1)-(A5) with $J(\xi)(x)$ satisfying (J1)-(J5) and subject to the inequalities (*) on p. 63. Then for sufficiently large ε ($\varepsilon \leq .25$), the Fisher information for scale $I(F)$ is minimized over $\mathcal{K}_\varepsilon$ by F_0 with*

$$\chi_0(x) = \begin{cases} \lambda_1 tan(\frac{\lambda_1}{2}\log d + \omega) & x \in [d,\infty) \\ \lambda_1 tan(\frac{\lambda_1}{2}\log x + \omega) & x \in [a,d) \\ \lambda_1 tan(\frac{\lambda_1}{2}\log a + \omega) & x \in [0,a) \\ \chi_0(-x) & x < 0 \end{cases} \tag{2.6.13}$$

$$f_0(x) = \begin{cases} g(d)(\frac{d}{x})^{\lambda_1 tan(\frac{\lambda_1}{2}\log d+\omega)+1} & x \in [d,\infty) \\ \frac{dg(d)}{cos^2(\frac{\lambda_1}{2}\log d+\omega)} \frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{x} & x \in [a,d) \\ \left(or \frac{ag(a)}{cos^2(\frac{\lambda_1}{2}\log a+\omega)} \frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{x} \right) & \\ g(a)(\frac{a}{x})^{\lambda_1 tan(\frac{\lambda_1}{2}\log a+\omega)+1} & x \in (0,a) \\ \infty & x = 0 \\ f_0(-x) & x < 0 \end{cases} \tag{2.6.14}$$

where a, d, λ_1, ω are determined by

$$\int_0^a f_0 = \int_0^a g + \varepsilon$$
$$\int_d^\infty f_0 = \int_d^\infty g + \varepsilon$$
$$\int_a^d f_0 = \int_a^d g - 2\varepsilon$$
$$\frac{dg(d)}{ag(a)} = \frac{cos^2(\frac{\lambda_1}{2}\log d + \omega)}{cos^2(\frac{\lambda_1}{2}\log a + \omega)}.$$

Minimum information is

$$I(F_0) = 2[-\lambda_1^2 tan^2(\frac{\lambda_1}{2}\log a + \omega)(G(a) - \frac{1}{2} + \varepsilon) + \lambda_1^2(G(d) - G(a) - 2\varepsilon)$$
$$- \lambda_1^2 tan^2(\frac{\lambda_1}{2}\log d + \omega)(1 - G(d) + \varepsilon)]$$

The existence proof of the constant a, d, λ and ω is not provided but numerical checking has been done for the cases where G is normal, Cauchy, logistic and some t-distributions with different degrees of freedom. That F_0 minimizes $I(F)$ over $\mathcal{K}_\varepsilon(G)$ follows from the fact that

$$\begin{aligned}\int_0^\infty J(\chi_0)(x)dH(x) &= \int_0^a -\lambda_1^2 tan^2(\frac{\lambda_1}{2}\log a+\omega)dH(x)+\int_a^d \lambda_1^2 dH(x)\\ &\quad + \int_d^\infty -\lambda_1^2 tan^2(\frac{\lambda_1}{2}\log d+\omega)dH(x)\\ &= -\lambda_1^2 sec^2(\frac{\lambda_1}{2}\log a+\omega)H(a)+\lambda_1^2 sec^2(\frac{\lambda_1}{2}\log d+\omega)H(d)\\ &\quad - \lambda_1^2 tan^2(\frac{\lambda_1}{2}\log d+\omega)H(\infty)\\ &\geq 0\end{aligned}$$

since

$$H(a) \leq 0,\ H(d) \geq 0 \text{ and } H(\infty) \leq 0$$

where $H = F - F_0$ and $F \in \mathcal{K}_\varepsilon(G)$. Moreover if we have checked

$$(*) \qquad \chi_0(a) > \xi(a) \text{ and } \chi_0(d) < \xi(d),$$

then we can conclude that $F_0 \in \mathcal{K}_\varepsilon(G)$. The reason is:

$$\chi_0(a) > \xi(a) \text{ implies } \chi_0(x) > \xi(x)$$

for $x \in (0, a)$ since χ_0 is constant on $(0, a)$ and ξ is increasing. With the property that $f_0(a) = g(a)$, we can get $f_0(x) \geq g(x)$ for all $x \in (0, a)$ by applying Lemma 2.5. Similarly

$$\chi_0(d) < \xi(d) \text{ implies } \chi_0(x) < \xi(x)$$

for $x \in (d, \infty)$. As $f_0(d) = g(d)$, we get

$$f_0(x) \geq g(x) \text{ for all } x \in (d, \infty)$$

by applying Lemma 2.5 again.

Furthermore since

$$\chi_0(a) > \xi(a) \text{ and } \chi_0(d) < \xi(d),$$

there must exist at least one $z \in (a,d)$ such that $\chi_0(z) = \xi(z)$ because χ_0 is continuous. To prove such a z is unique, it suffices to prove $(\chi_0 - \xi)(z) = 0$ implies $(\chi_0 - \xi)'(z) < 0$. Note that

$$J(\chi_0)(z) = 2z\chi_0'(z) - \chi_0^2(z)$$

and

$$J(\xi)(z) = 2z\xi'(z) - \xi^2(z).$$

Hence

$$\chi_0'(z) - \xi'(z) = \frac{J(\chi_0)(z) - J(\xi)(z)}{2z}.$$

If $J(\chi_0)(z) < J(\xi)(z)$, result follows.

If $J(\chi_0)(z) > J(\xi)(z)$, then we let p and q where $0 < p < q$ be the two points such that

$$J(\chi_0)(z) = J(\xi)(p) = J(\xi)(q).$$

Certainly $z \in (a,p] \cup [q,d)$. Note that one of the intervals $(a,p]$ or $[q,d)$ may be an empty set. Now we try to show that there is no solution to the equation $(\chi_0 - \xi)(z) = 0$ in the region $(a,p] \cup [q,d)$.

On $(a,p]$, suppose there exists at least one solution to the equation $(\chi_0 - \xi)(z) = 0$. Then there must exist an $z_1 \in (a,p]$ such that

$$(\chi_0 - \xi)(z_1) = 0 \text{ and } (\chi_0 - \xi)'(z_1) < 0$$

which implies

$$\begin{aligned} 0 &> (\chi_0 - \xi)'(z_1) \\ &= \frac{J(\chi_0)(z_1) - J(\xi)(z_1)}{2z_1} \\ &\geq 0 \end{aligned}$$

which is a contradiction.

Similarly on $[q, d)$, suppose there exists at least one solution to the equation $(\chi_0 - \xi)(z) = 0$. Then there must exist an $z_2 \in [q, p)$ such that

$$(\chi_0 - \xi)(z_2) = 0 \text{ and } (\chi_0 - \xi)'(z_2) < 0$$

which implies

$$\begin{aligned} 0 &> (\chi_0 - \xi)'(z_2) \\ &= \frac{J(\chi_0)(z_2) - J(\xi)(z_2)}{2z_2} \\ &\geq 0 \end{aligned}$$

which again is a contradiction. Hence it is not possible to have a solution to the equation $(\chi_0 - \xi)(z) = 0$ in the region $(a, p] \cup [q, d)$.

Now let k be the unique solution to $(\chi_0 - \xi)(z) = 0$. On (a, k), we have $\chi_0(x) \geq \xi(x)$ which implies $\frac{f_0}{g}$ is decreasing. As $f_0(a) = g(a)$, we have

$$f_0(x) \leq g(x) \text{ on } (a, k).$$

On (k, d), we have $\chi_0(x) \leq \xi(x)$ which implies $\frac{f_0}{g}$ is increasing. As $f_0(d) = g(d)$, we have

$$f_0(x) \leq g(x) \text{ on } (k, d).$$

Consequently, we have

$$f_0 \geq g \text{ on } (0, a) \cup (d, \infty)$$

and

$$f_0 \leq g \text{ on } (a, d)$$

which implies $F_0 \in \mathcal{K}_\varepsilon(G)$.

Now since $\mathcal{K}_\varepsilon(G)$ is convex and vaguely compact, we conclude that F_0 is unique.

Finally, we state the two possible "medium ε" forms of (χ_0, F_0).

Form 1.

$$\chi_0(x) = \begin{cases} \lambda_1 \tan(\frac{\lambda_1}{2} \log d + \omega) & x \in [d, \infty) \\ \lambda_1 tan(\frac{\lambda_1}{2} \log x + \omega) & x \in [b, d) \\ \xi(x) & x \in [a, d) \\ \xi(a) & x \in [0, a) \end{cases} \tag{2.6.15}$$

$$f_0(x) = \begin{cases} g(d)(\frac{d}{x})^{\lambda_1 tan(\frac{\lambda_1}{2} \log d+\omega)+1} & x \in [d, \infty) \\ \frac{dg(d)}{cos^2(\frac{\lambda_1}{2} \log d+\omega)} \frac{cos^2(\frac{\lambda_1}{2} \log x+\omega)}{x} & x \in [b, d) \\ \left(or \frac{bg(b)}{cos^2(\frac{\lambda_1}{2} \log b+\omega)} \frac{cos^2(\frac{\lambda_1}{2} \log x+\omega)}{x} \right) & \\ g(x) & x \in [a, d) \\ g(a)(\frac{a}{x})^{\xi(a)+1} & x \in (0, a) \\ \infty & x = 0 \\ f_0(-x) & x < 0 \end{cases} \tag{2.6.16}$$

where $a, b, d, \lambda_1, \omega$ are determined by

$$\int_0^a f_0 = \int_0^a g + \varepsilon$$

$$\int_d^\infty f_0 = \int_d^\infty g + \varepsilon$$

$$\int_a^d f_0 = \int_a^d g - 2\varepsilon$$

$$\frac{dg(d)}{bg(b)} = \frac{cos^2(\frac{\lambda_1}{2} \log d + \omega)}{cos^2(\frac{\lambda_1}{2} \log b + \omega)}$$

$$\lambda_1 \tan(\frac{\lambda_1}{2} \log b + \omega) = \xi(b)$$

Minimum information is

$$I(F_0) = 2[-\xi^2(a)(G(a) - \frac{1}{2} + \varepsilon) + \lambda_1^2(G(d) - G(b) - 2\varepsilon) - \lambda_1^2 tan^2(\frac{\lambda_1}{2}\log d + \omega)(1 - G(d) + \varepsilon) + \int_a^b J(\xi)(x)dG(x)]$$

Form 2

$$\chi_0(x) = \begin{cases} \xi(d) & x \in [d, \infty) \\ \xi(x) & x \in [c, d) \\ \lambda_1 tan(\frac{\lambda_1}{2}\log x + \omega) & x \in [a, c) \\ \lambda_1 tan(\frac{\lambda_1}{2}\log a + \omega) & x \in [0, a) \end{cases} \quad (2.6.17)$$

$$f_0(x) = \begin{cases} g(d)(\frac{d}{x})^{\xi(d)+1} & x \in [d, \infty) \\ g(x) & x \in [c, d) \\ \frac{cg(c)}{cos^2(\frac{\lambda_1}{2}\log c+\omega)} \frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{x} & x \in [a, c) \\ \left(or \frac{ag(a)}{cos^2(\frac{\lambda_1}{2}\log a+\omega)} \frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{x}\right) & \\ g(a)(\frac{a}{x})^{\lambda_1 tan(\frac{\lambda_1}{2}\log a+\omega)+1} & x \in (0, a) \\ \infty & x = 0 \\ f_0(-x) & x < 0 \end{cases} \quad (2.6.18)$$

where $a, c, d, \lambda_1, \omega$ are determined by

$$\int_0^a f_0 = \int_0^a g + \varepsilon$$

$$\int_d^\infty f_0 = \int_d^\infty g + \varepsilon$$

$$\int_a^c f_0 = \int_a^c g - 2\varepsilon$$

$$\frac{cg(c)}{ag(a)} = \frac{cos^2(\frac{\lambda_1}{2}\log c + \omega)}{cos^2(\frac{\lambda_1}{2}\log a + \omega)}$$

$$\lambda_1 \tan(\frac{\lambda_1}{2}\log c + \omega) = \xi(c)$$

Minimum information is

$$I(F_0) = 2[-\lambda_1^2 tan^2(\frac{\lambda_1}{2}\log(\frac{\lambda_1}{2}\log a + \omega))(G(a) - \frac{1}{2} + \varepsilon)$$
$$+ \lambda_1^2(G(c) - G(a) - 2\varepsilon)$$
$$- \xi^2(d)(1 - G(d) + \varepsilon) + \int_c^d J(\xi)(x)dG(x)]$$

For the Kolmogorov normal or logistic distributions the explicit form of (χ_0, F_0) corresponds to the type of Form 1. For the Kolmogorov Cauchy distributions, the explicit form of (χ_0, F_0) corresponds to the type of Form 2. In general, the form of (χ_0, F_0) depends on that of G. And in particular it depends on the degree of freedom when G is a t-distribution. This phenomenon was noted as well-again numerically - in the study of minimum information distributions for location in Kolmogorov neighbourhood models, Wiens (1986).

Table I, II and III provide some numerical results for the least informative distribution F_0 when G is normal, Cauchy and logistic respectively. All the numerical calculations were done using the IMSL library. Figures 5, 6 and 7 show the graphs of $\chi_0(x)$ against x for small, medium and large ε-normal cases respectively. Figures 8, 9 and 10 show the graphs of $\chi_0(x)$ against x for small, medium and large ε-Cauchy cases respectively.

2.7. A note on Thall's (1979) paper.

Thall (1979) developed a theory of robust estimation of a scale parameter by reformulating Huber's (1964) location parameter results in the scale parameter context. The results were then applied to a particular problem of robust estimation of the parameter θ of the exponential distribution $\Phi_0(\frac{x}{\theta})$ where $\Phi_0(x) = 1 - e^{-x}$, $x > 0$. Unfortunately, a few mistakes were made.

Table I. The Kolmogorov normal distributions that are least informative for

scale ε	a	b	c	d	λ_1	ω	$J(\xi)(b)$	$J(\xi)(b)$	$I(F_0)$
10^{-4}	-.0720	1.320	2.152	3.278	2.710	-.1098	6.419	5.338	1.971
.0005	.1228	1.168	2.314	2.887	2.610	-.0639	5.321	2.456	1.905
.0008	.1434	1.114	2.373	2.760	2.568	-.0450	4.904	1.017	1.867
.001	.1544	1.090	2.399	2.702	2.549	-.0362	4.719	.4150	1.845
.0018	.1871	1.009	2.489	2.546	2.478	-.0039	4.079	-2.213	1.769
.00205	.1952	.9911	2.511	2.511	2.461	-.0038	3.929	-2.910	1.748
.004	.2424	.8794		2.407	2.341	.0539	3.042	.2058	1.609
.01	.3240	.6819		2.274	2.071	.1436	1.574	3.290	1.318
.02	.4003	.5103		2.197	1.787	.2087	.4964	4.667	1.009
.025	.4276	.4529		2.180	1.683	.2251	.1888	4.931	.8947
.0267	.4359	.4359		2.176	1.651	.2293	.1042	4.990	.8600
.03	.4159			2.170	1.593	.2361	.0070	5.075	.7927
.04	.3645			2.166	1.441	.2490	-.2200	5.142	.6387
.1	.1933			2.305	.8567	.2406	-.7770	2.659	.1721
.2	.0458			2.883	.2847	.1534	-.9874	-20.19	.0067
.23	.0152			3.278	.1384	.1057	-.9986	-52.01	.0007
.249	.0005			4.257	.0187	.0288	-1.000	-220.5	6×10^{-7}

Table II. The Kolmogorov logistic distributions that are least informative for scale

ε	a	b	c	d	λ_1	ω	$J(\xi)(b)$	$J(\xi)(c)$	$I(F_0)$
10^{-4}	.1061	1.748	3.184	7.365	2.110	-.481	3.95	3.681	1.417
10^{-5}	.2257	1.413	3.731	5.445	2.008	-.471	2.958	1.851	1.355
.002	.2857	1.286	3.975	4.900	1.956	-.384	2.507	.799	1.304
.003	.3260	1.205	4.144	4.592	1.919	-.360	2.206	-.002	1.261
.0043	.3663	1.129	4.321	4.321	1.880	-.335	1.916	-.906	1.210
.0005	.3845	1.095		4.270	1.863	-.324	1.786	-.641	1.186
.01	.4794	.9092		4.054	1.743	-.255	1.075	.403	1.036
.02	.5939	.6886		3.892	1.556	-.170	.2825	1.174	.8171
.0023	.6194	.6410		3.869	1.509	-.153	.1268	1.272	.7638
.00239	.6272	.6272		3.864	1.495	-.147	.0827	1.296	.7478
.025	.6184			3.858	1.480	-.142	.0550	1.319	.7309
.04	.5145			3.841	1.292	-.083	-.2507	1.396	.5316
.01	.2830			4.213	.789	.023	-.7640	- .347	.4515
.15	.1623			4.814	.502	.057	-.9215	-3.748	.0406
.20	.0681			5.841	.265	.069	-.9861	-11.10	.0058
.24	.0101			8.062	.079	.050	-.9997	-33.61	.0002

Table III. The Kolmogorov Cauchy distributions that are least informative for scale

ε	a	b	c	d	λ_1	ω	$J(\xi)(b)$	$J(\xi)(c)$	$I(F_0)$
10^{-4}	.1954	.6689	1.495	16.20	1.361	$.14 \times 10^{-5}$	1.563	1.563	.4986
.0007	.2106	.5435	1.840	8.490	1.300	-.6458	1.126	1.113	.4919
.001	.2172	.5168	1.935	7.540	1.283	-.8255	.996	.996	.4888
.005	.2758	.3811	2.629	4.440	1.174	$-.2 \times 10^{-6}$	.329	.329	.4515
.0096	.3055	.3055	3.14	3.600	1.109	$.2 \times 10^{-3}$	-.063	.003	.4356
.0100	.3040		3.180	3.550	1.104	$.9 \times 10^{-4}$	-.070	-.017	.4285
.0110	.3005		3.278	3.440	1.904	$-.1 \times 10^{-4}$	-.088	-.007	.4108
.0118	.2979		3.359	3.360	1.085	$-.6 \times 10^{-6}$	-.102	-.102	.3970
.02	.2745			3.643	1.010	$-.964 \times 10^{-5}$	-.218	-.218	.3393
.1	.1420			7.041	.5334	$-.412 \times 10^{-5}$	-.767	-.767	.0654
.20	.0344			29.07	.1729	$-.82 \times 10^{-5}$	-.986	-.986	.0025
.22	.0183			54.50	.1128	$-.12 \times 10^{-4}$	-.996	-.996	.0005
.249	.0004			28.53	.0110	$-.52 \times 10^{-6}$	-.100	-1.00	$.2 \times 10^{-6}$

Fig. 5. Graph of $\chi_0(x)$ vs. x
for small ε−normal case (ε=.001)

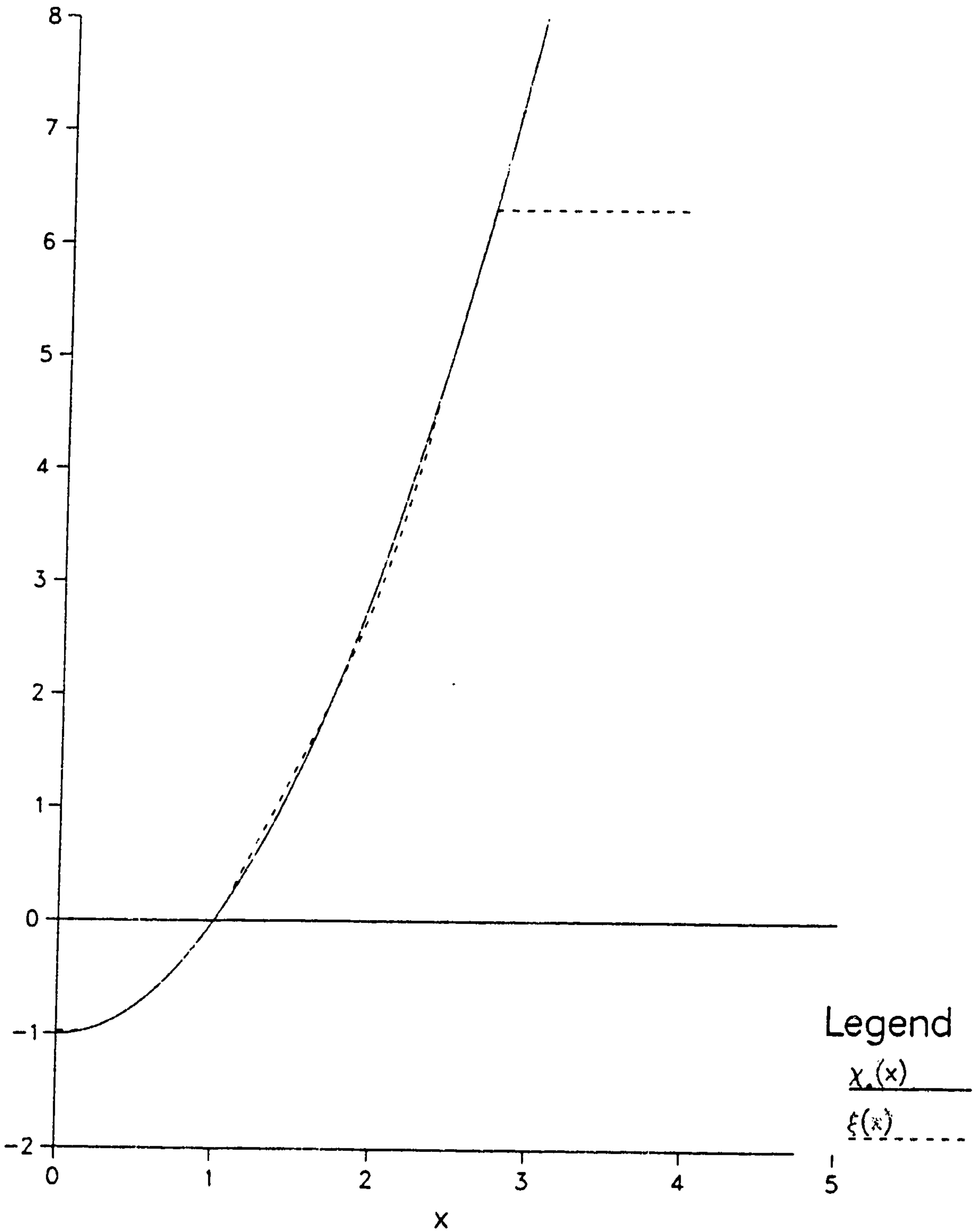


Fig. 6. Graph of $\chi_0(x)$ vs. x for medium ε−normal case (ε=.01)

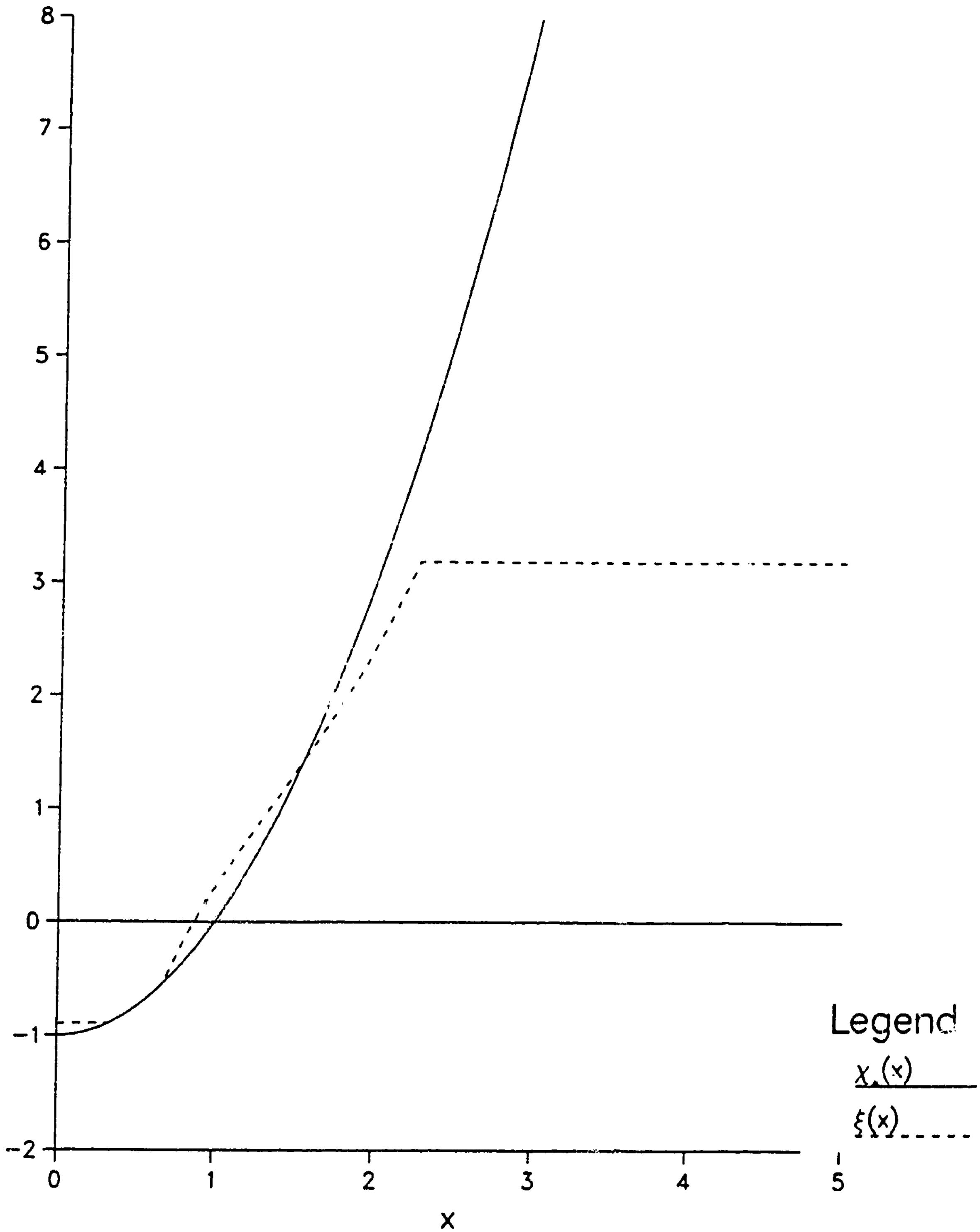


Fig. 7. Graph of $\chi_0(x)$ vs. x
for large ϵ−normal case (ϵ=.1)

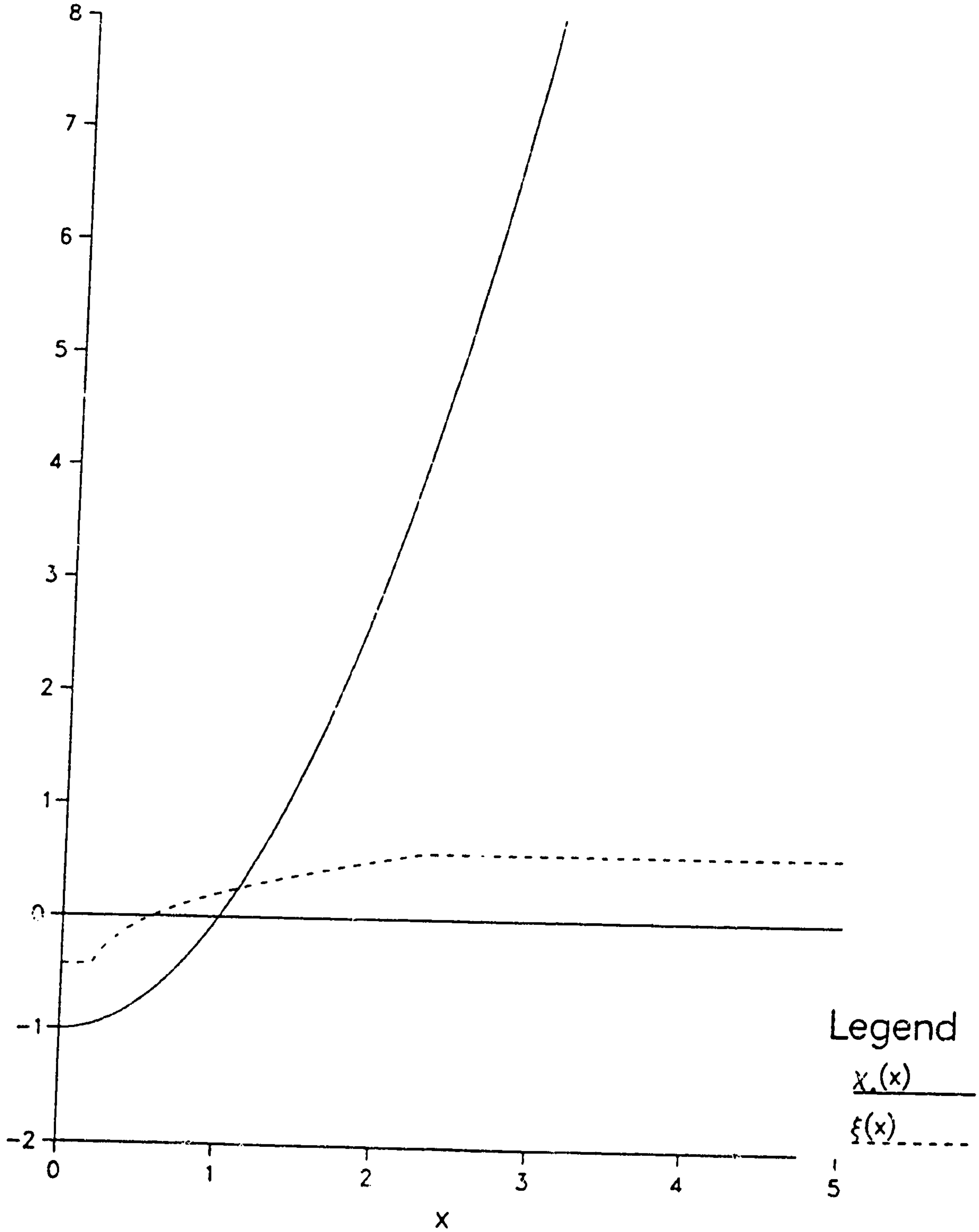


Fig. 8. Graph of $\chi_0(x)$ vs. x for small ϵ−Cauchy case (ϵ=.005)

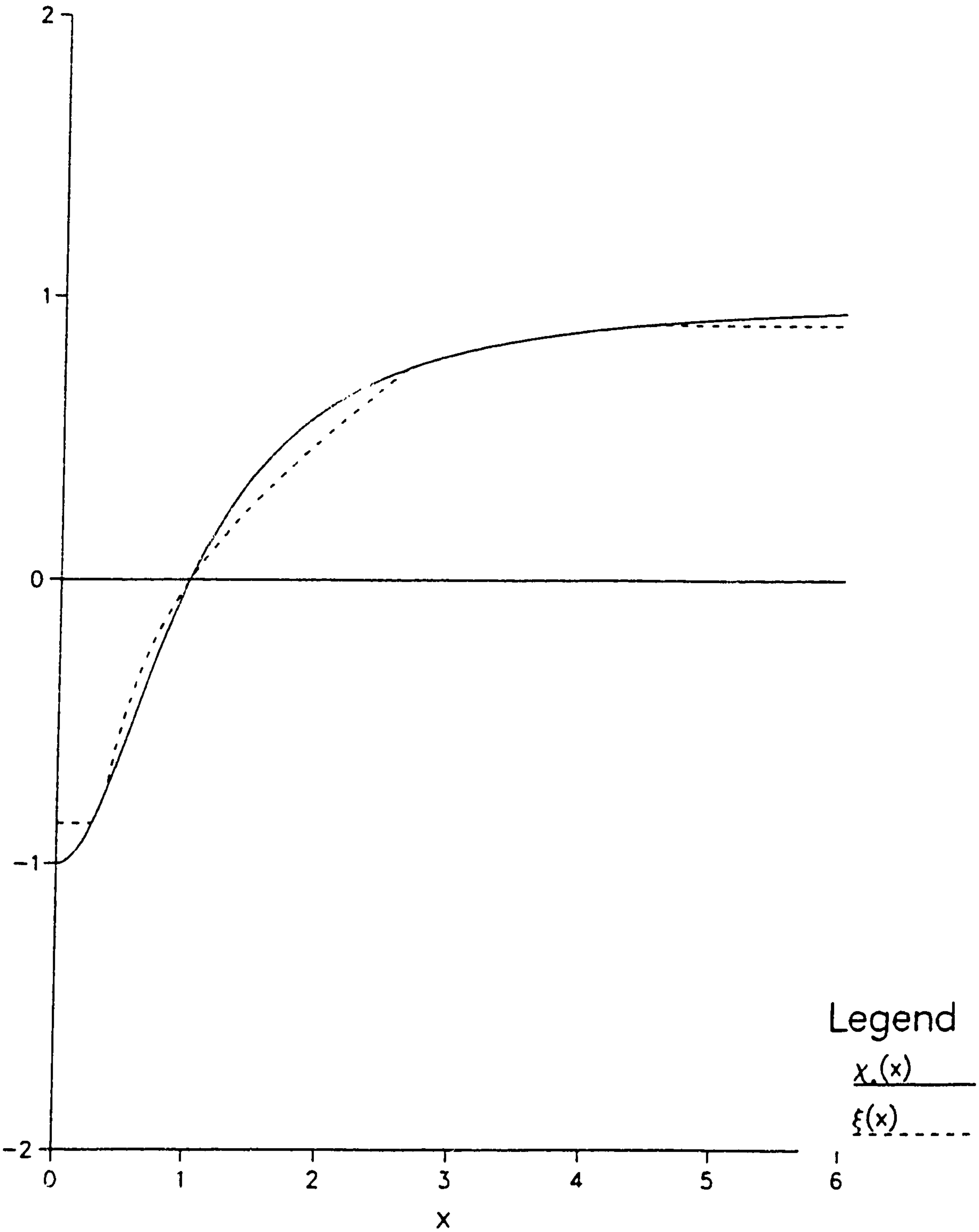


Fig. 9. Graph of $\chi_0(x)$ vs. x
for medium ϵ−Cauchy case (ϵ=.01)

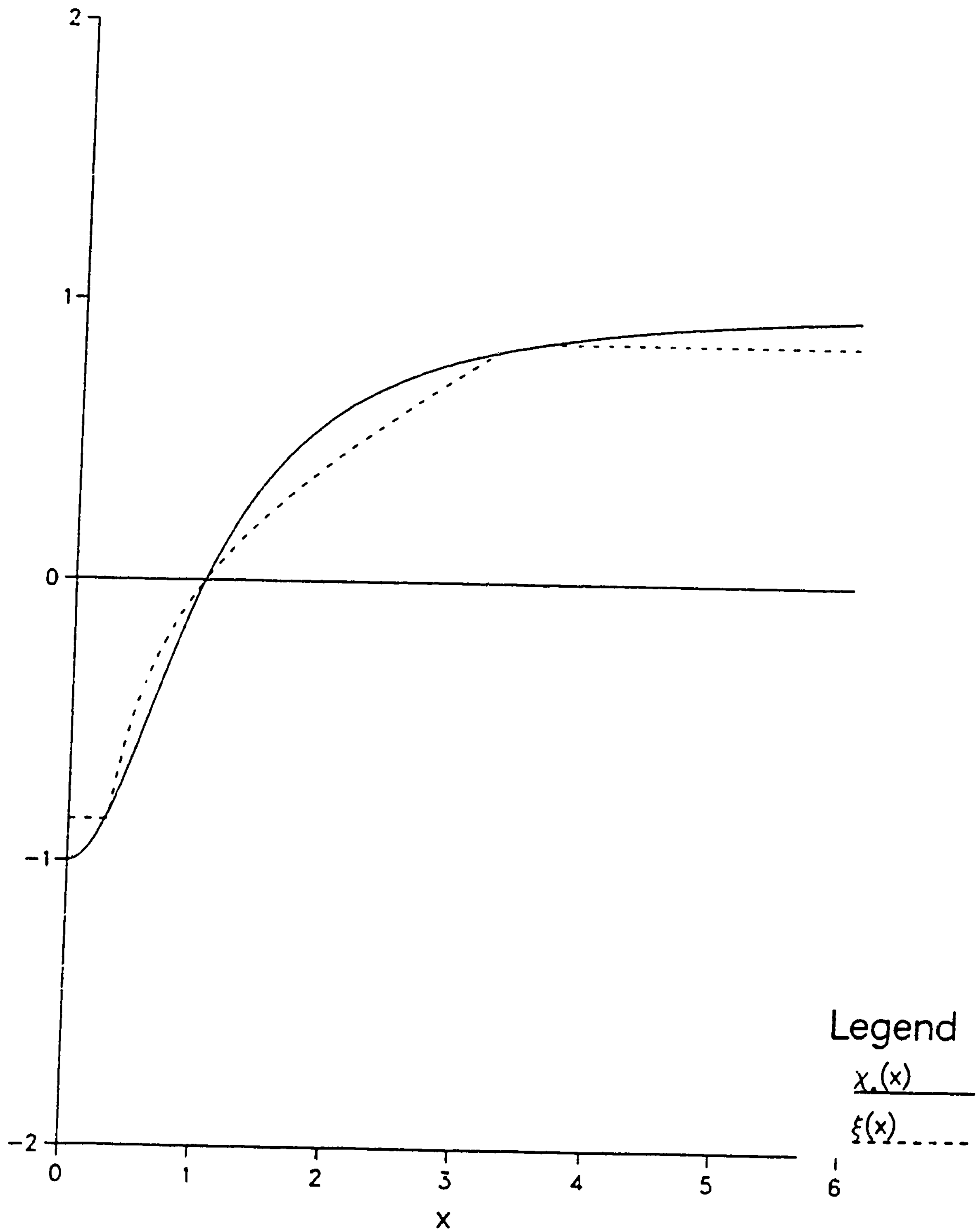
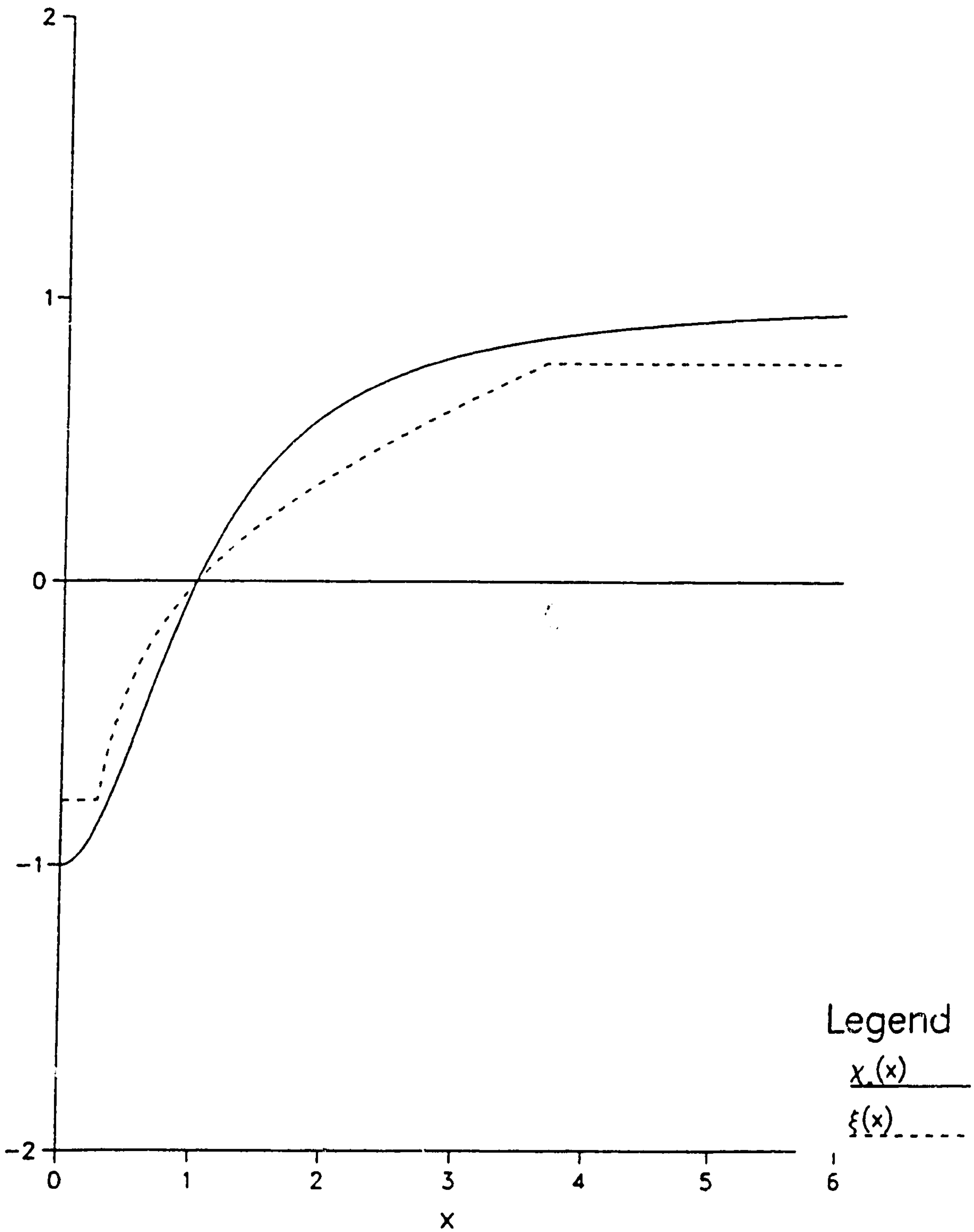


Fig. 10. Graph of $\chi_0(x)$ vs. x
for large ϵ−Cauchy case (ϵ=.02)



Thall stated that the maximum likelihood estimate for scale based on the least informative distribution is minimax in the sense that

$$\sup_{F\in C} V(\chi_0, F) = \inf_{\chi} \sup_{F\in C} V(\chi, F) \tag{2.7.1}$$

where $C = C(\varepsilon) = \{F : \sup_{x>0} |F(x) - \Phi_0(x)| \leq \varepsilon\}$, χ is taken over the set of all continuous and nondecreasing functions on $(0,\infty)$, $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$, $f_0 = F_0'$ and F_0 minimizes Fisher information for scale over C.

$$V(\chi, F) = \frac{S^2(F)\int_0^\infty \chi^2(\frac{x}{S(F)}dF(x)}{(\int_0^\infty \chi'(\frac{x}{S(F)})(\frac{x}{X(F)})dF(x))^2}$$

and $S(F)$ is defined by

$$\int_0^\infty \chi(\frac{x}{S(F)})dF(x) = 0.$$

In fact, as mentioned in Huber's (1981) page 122 or the introduction of chapter 1 in this thesis, the result of (2.7.1) is far more than what we can say. In fact, we can at most conclude that

$$\sup_{F\in C^*} V(\chi_0, F) = \inf_{\chi} \sup_{F\in C^*} V(\chi, F) \tag{2.7.2}$$

where $C^* = \{F \in C : \int_0^\infty \chi_0(x)dF(x) = 0\}$.

In the same paper, Thall also wrote down the form of (χ_0, F_0) where F_0 minimizes Fisher information for scale over C. Although he could show that the form of F_0 he obtained really minimizes Fisher information for scale over C, F_0 itself is not in C. This fact is easily seen if we look at the side condition (3.3) in his paper which is

$$\int_0^a f_0(x)dx = \int_0^a e^{-x}dx + \varepsilon. \tag{2.7.3}$$

(2.7.3) indicates that $F_0(a) = \Phi_0(a) + \varepsilon$. This implies that

$$f_0(a) = e^{-a}$$

if F_0 is required to remain in C. The form of F_0 obtained by Thall does not satisfy this condition. One can easily check this by using his numerical results.

At the end of section 3 of Thall's paper, Thall mentioned that extending his solution to larger values of ε was impossible since there were too few parameters to simultaneously satisfy the resulting set of constraints on f_0. According to his obtained solution, this probably involves overlooking the general solution of χ_0 to the differential equation

$$2x\chi_0'(x) - \chi_0^2(x) = \lambda^2,\ \lambda > 0.$$

The general solution of χ_0 to this differential equation should be

$$\chi_0(x) = \lambda tan(\frac{\lambda}{2}\log x + \omega)$$

rather than

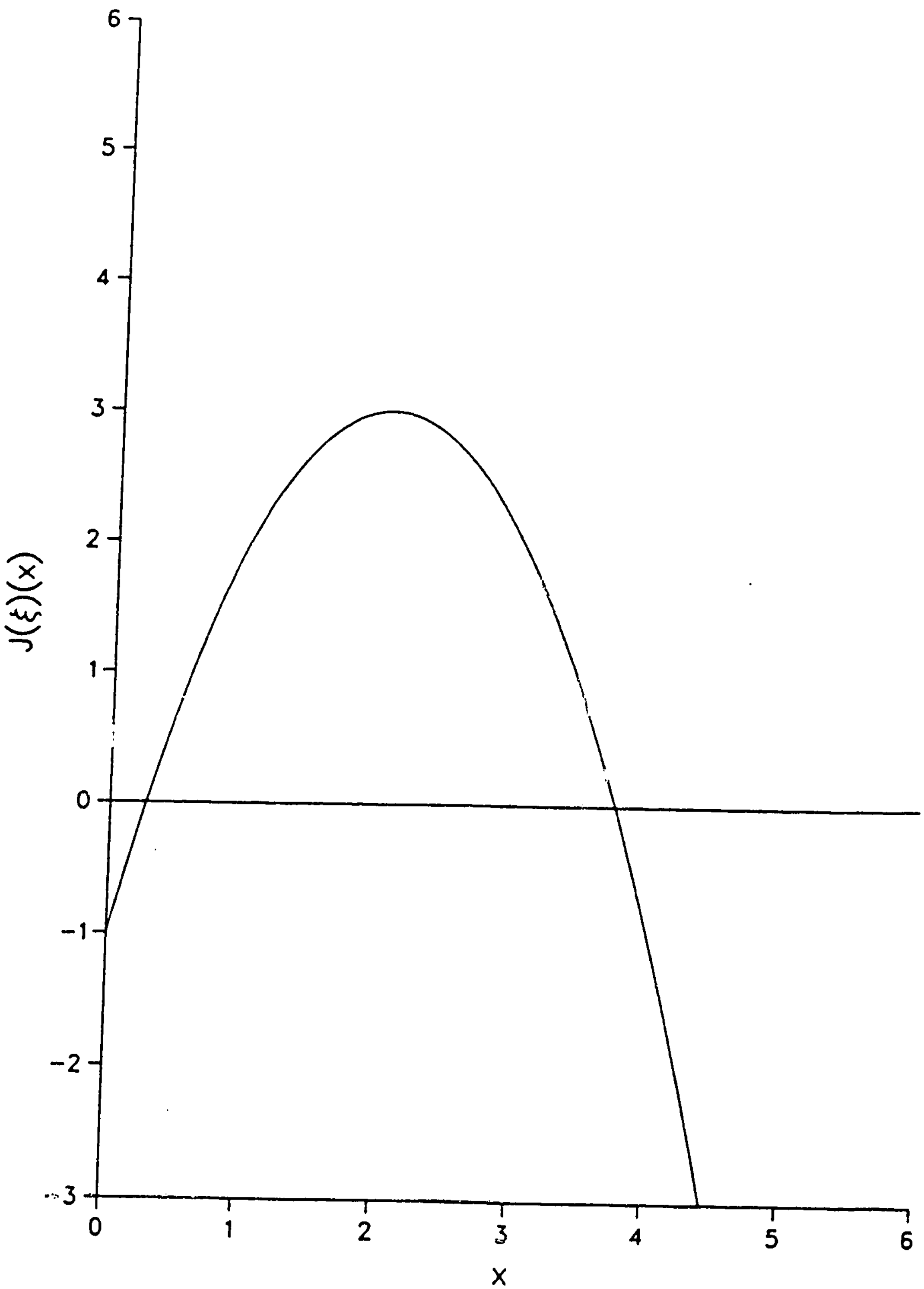
$$\chi_0(x) = \lambda tan(\frac{\lambda}{2}\log x),$$

the one obtained by Thall.

Moreover, there should be totally three cases (small ε, medium or large ε) for the form of (χ_0, F_0) instead of only two cases (small ε, large ε) mentioned by Thall. The correct form of (χ_0, F_0), for all possible range of ε, can be obtained in a similar way as described in section 2.6. It is worthy to note that the form of $J(\xi)(x)$ in this case looks just like the right half of our earlier ones and the theory is analogous. Figure 11 shows the graph of $J(\xi)(x)$ vs. x for exponential cases.

Case (i). For small ε:

Fig. 11. Graph of $J(\xi)(x)$ vs. x for exponential case



$$\chi_0(x) = \begin{cases} d-1 & x \in [d,\infty) \\ x-1 & x \in [c,d) \\ \lambda_1 \tan(\frac{\lambda_1}{2}\log x + \omega) & x \in [b,c) \\ x-1 & x \in [a,b) \\ a-1 & x \in [0,a) \end{cases} \tag{2.7.4}$$

$$f_0(x) = \begin{cases} e^{-d}(\frac{d}{x})^d & x \in [d,\infty) \\ e^{-x} & x \in [c,d) \\ \frac{ce^{-c}}{x}\frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{cos^2(\frac{\lambda_1}{2}\log c+\omega)} & x \in [a,c) \\ \left(or \frac{be^{-b}}{x}\frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{cos^2(\frac{\lambda_1}{2}\log b+\omega)} \right) & \\ e^{-x} & x \in [a,b) \\ e^{-a}(\frac{a}{x})^a & x \in (0,a) \\ \infty & x = 0 \end{cases} \tag{2.7.5}$$

where $a, b, c, d, \omega, \lambda_1$ are determined by

$$\int_0^a f_0 = \int_0^a e^{-x} + \varepsilon \tag{2.7.6}$$

$$\int_b^c f_0 = \int_b^c e^{-x} - 2\varepsilon \tag{2.7.7}$$

$$\int_d^\infty f_0 = \int_d^\infty e^{-x} + \varepsilon \tag{2.7.8}$$

$$c - 1 = \lambda_1 tan(\frac{\lambda_1}{2}\log c + \omega) \tag{2.7.9}$$

$$b - 1 = \lambda_1 \tan(\frac{\lambda_1}{2}\log b + \omega) \tag{2.7.10}$$

$$\frac{ce^{-c}}{be^{-b}} = \frac{cos^2(\frac{\lambda_1}{2}\log c + \omega)}{cos^2(\frac{\lambda_1}{2}\log b + \omega)} \tag{2.7.11}$$

Minimum information is

$$I(F_0) = [-(a-1)^2(1 - e^{-a} + \varepsilon) + \lambda_1^2(e^{-b} - e^{-c} - 2\varepsilon)$$
$$-(d-1)^2(e^{-d} + \varepsilon) + \int_{[a,b]\cup[c,d]} (-x^2 + 4x - 1)e^{-x}dx$$

Case (ii) For medium ε:

$$\chi_0(x) = \begin{cases} \lambda_1 \tan(\frac{\lambda_1}{2}\log d + \omega) & x \in [d, \infty) \\ \lambda_1 \tan(\frac{\lambda_1}{2}\log x + \omega) & x \in [c, d) \\ x - 1 & x \in [a, c) \\ a - 1 & x \in [0, a) \end{cases} \tag{2.7.12}$$

$$f_0(x) = \begin{cases} e^{-d}(\frac{d}{x})^{\lambda_1 \tan(\frac{\lambda_1}{2}\log d+\omega)+1} & x \in [d, \infty) \\ \frac{ce^{-c}}{x}\frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{cos^2(\frac{\lambda_1}{2}\log c+\omega)} & x \in [c, d) \\ \left(or\, \frac{de^{-d}}{x}\frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{cos^2(\frac{\lambda_1}{2}\log d+\omega)}\right) & \\ e^{-x} & x \in [a, c) \\ e^{-a}(\frac{a}{x})^a & x \in (0, a) \\ \infty & x = 0 \end{cases} \tag{2.7.13}$$

where $a, c, d, \omega, \lambda_1$ are determined by (2.7.6), (2.7.7), (2.7.8), (2.7.9) and (2.7.11) with $a = b$.

Minimum information is

$$I(F_0) = [-(a-1)^2(1 - e^{-a} + \varepsilon) + \lambda_1^2(e^{-c} - e^{-d} - 2\varepsilon) - \lambda_1^2 tan^2(\frac{\lambda_1}{2}\log d + w)(e^{-d} + \varepsilon) + \int_a^c (-x^2 + 4x - 1)e^{-x}dx]$$

Case (iii) For large ε:

$$\chi_0(x) = \begin{cases} \lambda_1 \tan(\frac{\lambda_1}{2}\log d + \omega) & x \in [d, \infty) \\ \lambda_1 \tan(\frac{\lambda_1}{2}\log x + \omega) & x \in [a, d) \\ \lambda_1 \tan(\frac{\lambda_1}{2}\log a + \omega) & x \in (0, a) \end{cases} \tag{2.7.14}$$

$$f_0(x) = \begin{cases} e^{-d}(\frac{d}{x})^{\lambda_1 \tan(\frac{\lambda_1}{2}\log d+\omega)} & x \in [d, \infty) \\ \frac{ae^{-a}}{x}\frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{cos^2(\frac{\lambda_1}{2}\log a+\omega)} & x \in [a, d) \\ \left(or\, \frac{de^{-d}}{x}\frac{cos^2(\frac{\lambda_1}{2}\log x+\omega)}{cos^2(\frac{\lambda_1}{2}\log d+\omega)}\right) & \\ e^{-a}(\frac{a}{x})^{\lambda_1 tan\frac{\lambda_1}{2}\log a+\omega)} & x \in (0, a) \\ \infty & x = 0 \end{cases} \tag{2.7.15}$$

where $a, c, d, \omega, \lambda_1$ are determined by (2.7.6), (2.7.7), (2.7.8), and (2.7.11) with $a = b$ and $c = d$.

Minimum information is

$$I(F_0) = [-\lambda_1^2 tan^2(\frac{\lambda_1}{2}\log a + w)(1 - e^{-a} + \varepsilon) + \lambda_1^2(e^{-a} - e^{-d} - 2\varepsilon) - \lambda_1^2 tan^2(\frac{\lambda_1}{2}\log d + w)(e^{-d} + \varepsilon)$$

Some numerical results for the least informative distribution F_0 when G is exponential are provided in Table IV. Figures 12, 13 and 14 show the graphs of $\chi_0(x)$ against x for small, medium and large ε-exponential cases respectively.

Table IV. The Kolmogorov exponential distributions that are least informative for scale

ε	a	b	c	d	λ_1	ω	$J(\xi)(b)$	$J(\xi)(c)$	$I(F_0)$
10^{-4}	.0141	1.391	2.729	7.360	1.690	-.052	2.630	[illegible].469	.9940
10^{-3}	.0440	1.100	3.216	5.421	1.629	-.015	2.17[illegible]	[illegible]	.9631
.003	.754	.9111	3.570	4.544	1.577	.0171	1.81[illegible]	[illegible]	.9227
.005	.0966	.8207	3.775	4.151	1.545	.037	1.60[illegible]	-.15[illegible]4	.8896
.006	.1054	.7875	3.856	4.013	1.532	.045	1.530	-.4433	.8746
.0065	.1096	.7725	3.893	3.953	1.525	.649	1.493	-.5842	.8662
.0683	.1122	.7635		3.916	1.522	.051	1.471	-.6731	.8625
.008	.1210	.7331		3.852	1.509	.0591	1.395	-.4299	.8323
.01	.1345	.6868		3.763	1.486	.0715	1.276	-.1072	.7983
.02	.1860	.5279		3.504	1.389	.1163	.8329	.7392	.7309
.05	.2807	.3116		3.252	1.191	.1703	.1493	1.432	.5131
.0546	.2916	.2920		3.237	1.168	.1736	.0829	1.470	.4893
.1	.2213			3.220	.9742	.1870	-.1638	1.513	.3092
.2	.1330			3.535	.6448	.1754	-.4856	.6449	.1090
.3	.0751			4.123	.4315	.1495	-.7053	-1.508	.0309
.4	.0312			5.145	.2341	.1141	-.8761	-6.899	.0016
.45	.0136			6.127	.1374	.0877	-.9458	-14.03	8×10^{-4}
.49	.0021			8.269	.0456	.0463	-.9915	-36.31	1.9×10^{-4}

Fig. 12. Graph of $\chi_0(x)$ vs. x
for small ϵ−exponential case (ϵ=.005)

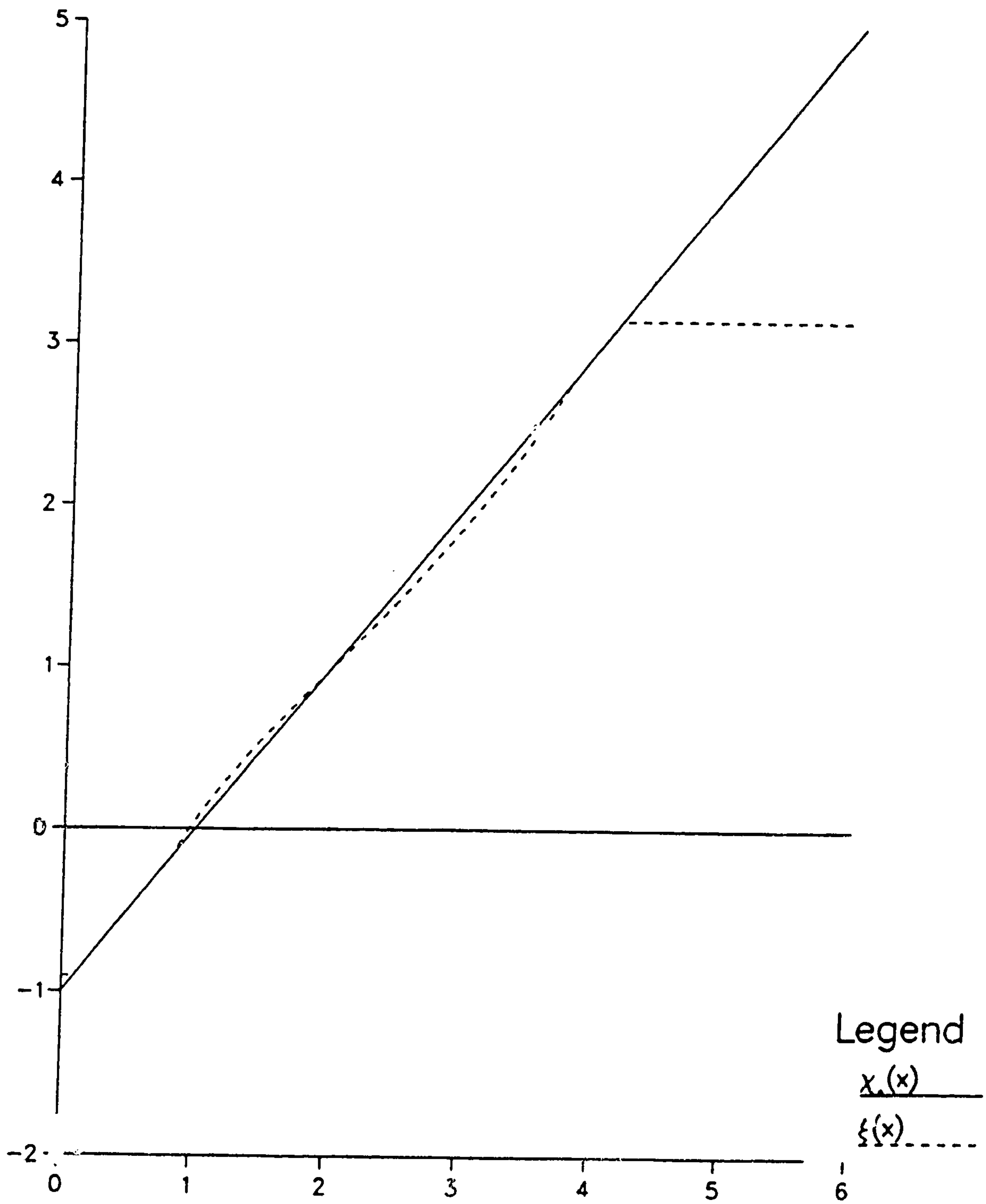


Fig. 13. Graph of $\chi_0(x)$ vs. x
for medium ϵ−exponential case (ϵ=.01)

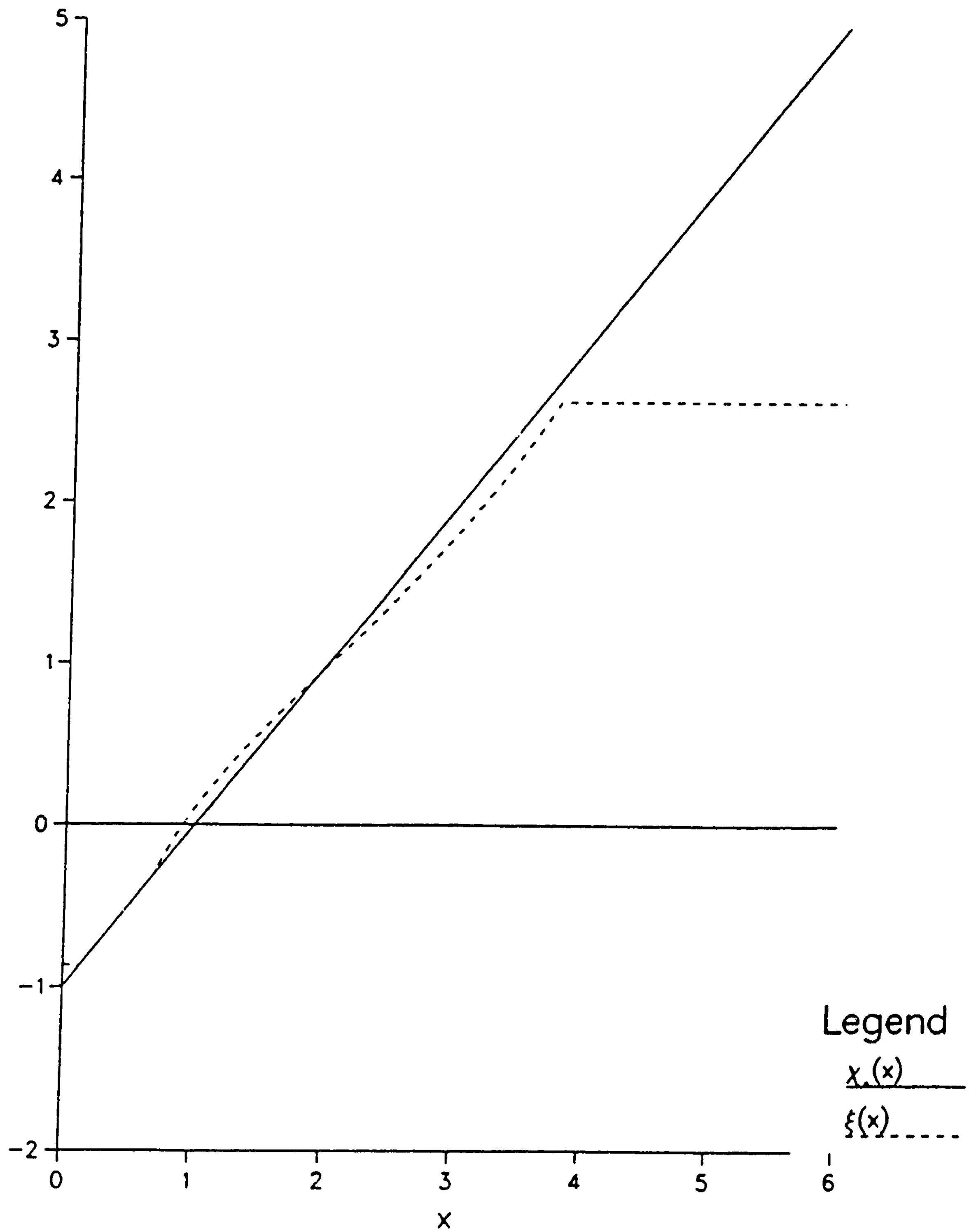
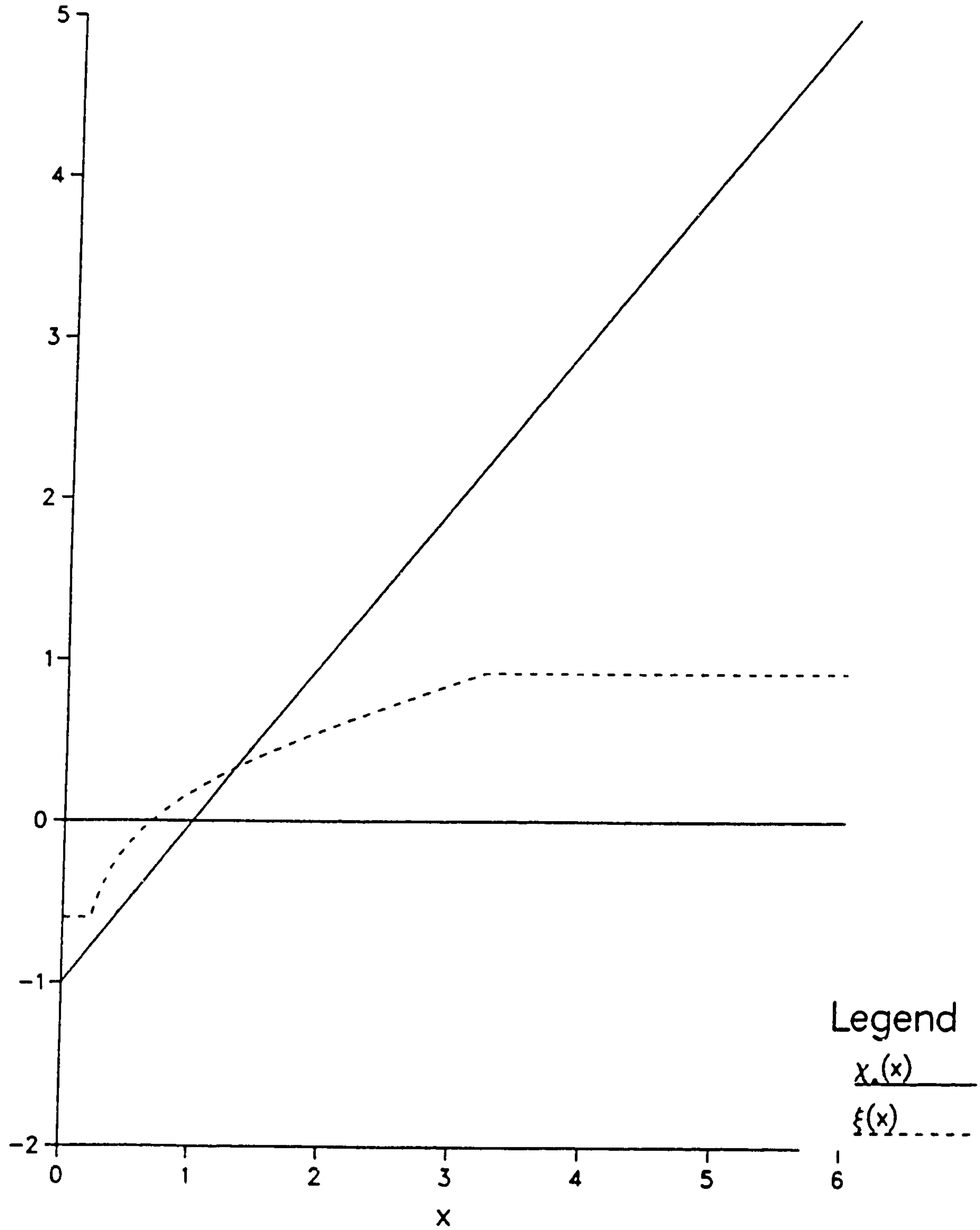


Fig. 14. Graph of $\chi_0(x)$ vs. x
for large ε-exponential case (ε=.1)



CHAPTER III

MINIMAX ASYMPTOTIC PROPERTIES OF M-ESTIMATORS FOR SCALE

3.1 Introduction

The results obtained in the previous two chapters show that the M-estimator $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$, where $f_0 = F_0'$ and F_0 minimizes Fisher information for scale over an ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$, [resp. a Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$], is minimax with regard to the asymptotic variance $V(\chi, F)$ over those members of $\mathcal{G}_\varepsilon(G)$ [resp. $\mathcal{K}_\varepsilon(G)$] satisfying $S_0(F) = 1$. We note that the condition $S_0(F) = 1$ is a rather serious restriction. Also the asymptotic variance $V(\chi, F)$ is not a good quantity to use in comparing the performance of the estimators since $V(\chi, F)$ is a quantity depending on the arbitrary standardization of S.

We thus define the asymptotic loss

$$R(\chi, F) = \frac{V(\chi, F)}{S^2(F)}.$$

As $R(\chi, F)$ is the asymptotic variance of $\sqrt{n}\log\frac{S(F_n)}{S(F)}$ and does not depend on the arbitrary standardization of S, it provides a good quantity for comparing the performance of the estimators.

3.2 Minimax asymptotic properties of M-estimators for scale over ε-contamination normal neighbourhood model $\mathcal{G}_\varepsilon(\Phi)$ for small ε $(0 < \varepsilon \leq 2.051)$

Huber (1981, chapter 5) shows that when $\epsilon \leq .04$,

$$R(\chi_0, F_0) \geq R(\chi_0, F) \tag{3.2.1}$$

among all the distributions F in $\mathcal{G}_\varepsilon(\Phi)$ where Φ is the standard normal distribution. $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$, $f_0 = F_0'$ and F_0 minimizes Fisher information for scale over $\mathcal{G}_\varepsilon(\Phi)$. It follows that the minimax property

$$\sup_{F \in \mathcal{G}_\varepsilon(\Phi)} R(\chi_0, F) = \inf_{\chi} \sup_{F \in \mathcal{G}_\varepsilon(\Phi)} R(\chi, F), \tag{3.2.2}$$

holds for at least this range of ε.

In this section, we shall show that (3.2.1) fails for the range of ε between .0997 and .2051. Note that $\varepsilon = .2051$ is the boundary case between the large and small ε. By definition,

$$\begin{aligned}\frac{1}{R(\chi_0, F)} &= \frac{S^2(F)}{V(\chi_0, F)} \\ &= \frac{[\int \chi_0'(\frac{x}{S(F)})(\frac{x}{S(F)})dF(x)]^2}{\int \chi_0^2(\frac{x}{S(F)})dF(x)}\end{aligned} \tag{3.2.3}$$

and S satisfies the condition

$$\int \chi_0(\frac{x}{S(F)})dF(x) = 0. \tag{3.2.4}$$

For each $k \in [0, 1]$, define

$$\mathcal{G}_k = \{F \in \mathcal{G}_\varepsilon(\Phi) : S(F) = k\}.$$

Note that $\mathcal{G}_k$ is convex, $\bigcup_k \mathcal{G}_k = \mathcal{G}_\varepsilon(\Phi)$ and by Huber (1981) Lemma 4.4.4, $\frac{1}{R(\chi_0,F)}$ is a convex functional of F. Furthermore, $\frac{1}{R(\chi_0,F)}$ is minimized by a contamination εH putting all its mass on $\{0\} \cup [-kx_1, kx_1]^c$. Note that Huber (1981) page 123 states that $\frac{1}{R(\chi_0,F)}$ is minimized by a contamination εH putting all its mass on $\{0\} \cup [-x_1, x_1]^c$ which is sufficient but not necessary. The reason is as follows: Define

$$\rho_k(t, \lambda) = \frac{[\int \chi_0'(\frac{x}{k})(\frac{x}{k})dF_t(x)]^2}{\int \chi_0^2(\frac{x}{k})dF_t(x)} + \lambda \int \chi_0(\frac{x}{k})dF_t(x) \tag{3.2.4}$$

for $0 \leq t \leq 1$ where

$F_t = (1-t)F_* + tF_1$, $F_1 \in \mathcal{G}_k$ with $I(F_1) < \infty$,

F_* minimizes $\frac{1}{R(\chi_0, F)}$ among all the distributions F in $\mathcal{G}_k$ and

λ is a Lagrange multiplier.

Note that F_* minimizes $\frac{1}{R(\chi_0, F)}$ in $\mathcal{G}_k$ if and only if $\rho_k(t,\lambda)$ is minimized at $t = 0$ for some λ. Put

$$N_t = \int \chi_0'(\frac{x}{k})(\frac{x}{k})dF_t(x) \tag{3.2.5}$$

and

$$D_t = \int \chi_0(\frac{x}{k})dF_t(x). \tag{3.2.6}$$

Then (3.2.4) becomes

$$\rho_k(t,\lambda) = \frac{N_t^2}{D_t} + \lambda \int \chi_0(\frac{x}{k})dF_t(x). \tag{3.2.7}$$

Differentiating (3.2.7) with respect to t, we have

$$\rho_k'(t,\lambda) = \frac{\partial}{\partial t}\rho_k(t,\lambda) = \frac{N_t}{D_t^2}[2N_t'D_t - N_tD_t'] + \lambda \int \chi_0(\frac{x}{k})d(F_1 - F_*)(x)$$

Set $t = 0$ and put

$$A_{F_*}(x) = 2(\frac{x}{k})\chi_0'(\frac{x}{k}) \int \chi_0^2(\frac{x}{k})dF_*(x) - \chi_0^2(\frac{x}{k}) \int (\frac{x}{k})\chi_0'(\frac{x}{k})dF_*(x)$$

then

$$\rho_k'(0,\lambda) = \int [\frac{N_0}{D_0}A_{F_*}(x) + \lambda\chi_0(\frac{x}{k})]d(F_1 - F_*)(x)$$

Hence we have

$$0 \leq \rho_k'(0,\lambda) \Leftrightarrow 0 \leq \int [\frac{N_0}{D_0}A_{F_*}(x) + \lambda\chi_0(\frac{x}{k})]d(F_1 - F_*)(x)$$

Thus F_* must put all its available mass on the place where

$$\frac{N_0}{D_0}A_{F_*}(x) + \lambda\chi_0(\frac{x}{k})$$

is a minimum. Put

$$\alpha_* = \int \chi_0^2(\frac{x}{k})dF_*(x)$$

and

$$\beta_* = \int (\frac{x}{k})\chi_0'(\frac{x}{k})dF_*(x),$$

then

$$\frac{N_0}{D_0}A_{F_*}(x) + \lambda\chi_0(\frac{x}{k}) = \begin{cases} \frac{N_0}{D_0^2}\{4\alpha_*(\frac{x}{k})^2 - (\beta_* - \lambda)[(\frac{x}{k})^2 - 1]\} & |x| \le x_1 k \\ -\frac{N_0}{D_0^2}(\beta_* - \lambda)[x_1^2 - 1] & |x| > x_1 k \end{cases}$$

When $\beta^* \ge \lambda$,

$$\frac{N_0}{D_0^2}A_{F_*}(x) + \lambda\chi_0(\frac{x}{k}) \text{ attains minimum at } |x| > x_1 k.$$

When $\beta^* < \lambda$,

$$\frac{N_0}{D_0^2}A_{F_*}(x) + \lambda\chi_0(\frac{x}{k}) \text{ attains minimum at } x = 0.$$

From here it is clear that $\frac{1}{R(\chi_0,F)}$ is minimized in $\mathcal{G}_k$ by a contamination εH putting all its mass on $\{0\} \cup [-kx_1, kx_1]^c$. Note that $\mathcal{G}_k$ is a nonempty set only if there exists an $F \in \mathcal{G}_\varepsilon(\Phi)$ satisfying $\int \chi_0(\frac{x}{k})dF(x) = 0$. Obviously $\mathcal{G}_1$ is nonempty since $S_0(F_0) = 1$. Moreover, if F is a distribution in $\mathcal{G}_\varepsilon(\Phi) = \bigcup_k \mathcal{G}_k$ that minimizes $\frac{1}{R(\chi_0,F)}$, it follows from the above arguments that it should place all its contaminating mass on $\{0\} \cup [-S(F)x_1, S(F)x_1]^c$ and satisfy

$$\int \chi_0(\frac{x}{S(F)})dF(x) = 0. \tag{3.2.8}$$

Now we view $R(\chi_0, F)$ as a function of S and define

$$\pi(S) = \frac{1}{R(\chi_0,F)} = \frac{[\int(\frac{x}{S})\chi_0'(\frac{x}{S})dF(x)]^2}{\int \chi_0^2(\frac{x}{S})dF(x)}. \tag{3.2.9}$$

Using the relationship $\chi_0(x) = \psi_0^2(x) - 1$ where $\psi_0(x) = min[x, max(-x_1, x)]$ and assuming that εH puts mass $\varepsilon - \varepsilon_1$ on $\{0\}$ and mass ε_1 on $[-x_1 S, x_1 S]^c$, (3.2.9) and (3.2.8) then reduce to

$$\pi(S) = \frac{[2(1-\varepsilon)\int_{-x_1 S}^{x_1 S}(\frac{x}{S})^2 d\Phi(x)]^2}{(1-\varepsilon)\int \chi^4(\frac{x}{S})d\Phi(x) + \varepsilon_1 x_1^4 - 1}. \tag{3.2.10}$$

and

$$(1-\varepsilon)\int \psi_0^2(\frac{x}{S})d\Phi(x) + \varepsilon_1 x_1^2 = 1. \tag{3.2.11}$$

Note that (3.2.10) and (3.2.11) are equivalent to (7.7) and (7.8) in Huber (1981). However, Huber treats (3.2.10) as a function of ε_1 whereas we treat (3.2.10) as a function of S.

Now define

$$\gamma(S) = \int \psi_0^2(\frac{x}{S})d\Phi(x) = \int_{-x_1 S}^{x_1 S}(\frac{x}{S})^2 d\Phi(x) + 2x_1^2(1 - \Phi(x_1 S)) \tag{3.2.12}$$

Notice that $\gamma(S)$ has the following properties:

(i) $\gamma(0) = x_1^2 > 2$

(ii) $\gamma(1) = \frac{1-\varepsilon x_1^2}{1-\varepsilon}, 0 < \gamma(1) < 1$

(iii) $\gamma'(S) = -\frac{4}{S^3}\int_0^{x_1 S} x^2 d\Phi(x) < 0$

(iv) $\frac{1-\varepsilon_1^2}{1-\varepsilon} \leq \gamma(S) \leq \frac{1}{1-\varepsilon}$.

(i), (ii) and (iii) follow directly from the definition as in (3.2.12). From (3.2.11), letting ε_1 be zero and ε respectively, (iv) follows.

Define S_{min} to be the smallest possible value of S such that (3.2.8) satisfied. Note that (3.2.8) and (3.2.11) are equivalent. By the above properties (iii) and (iv), we can see that S_{min} is a function of ε. In fact, S_{min} and ε are related by

$$\gamma(S_{min}) = \frac{1}{1-\varepsilon}. \tag{3.2.13}$$

It is worthy to note that

$$S = S_{min} \text{ is equivalent to } \varepsilon_1 = 0$$

and

$$S = 1 \text{ is equivalent to } \varepsilon_1 = \varepsilon.$$

Moreover ε and x_1 are related by the equation

$$2[\Phi(x) - \frac{1}{2}] + \frac{2x_1\phi(x_1)}{x_1^2 - 1} = \frac{1}{1-\varepsilon} \tag{3.2.14}$$

where $\phi = \Phi'$. Thus x_1 can also be viewed as a function of ε. After some calculations, we have

$$\begin{aligned} \pi(S_{min}) &= \frac{[2(1-\varepsilon)\int_{-x_1 S_{min}}^{x_1 S_{min}} (\frac{x}{S_{min}})^2 d\Phi(x)]^2}{(1-\varepsilon)\int \psi_0^4(\frac{x}{S_{min}}) d\Phi(x) - 1} \\ &= \frac{4\{1 - 2(1-\varepsilon)x_1^2[1 - \Phi(x_1 S_{min})]\}^2}{(1-\varepsilon)\{\int_{-x_1 S_{min}}^{x_1 S_{min}} (\frac{x}{S_{min}})^4 d\Phi(x) + 2x_1^4[1 - \Phi(x_1 S_{min})]\} - 1} \end{aligned} \tag{3.2.15}$$

and

$$\pi(1) = \frac{[2(1-\varepsilon)\int_{-x_1}^{x_1} x^2 d\Phi(x)]^2}{(1-\varepsilon)\int \psi_0^4 d\Phi(x) + \varepsilon x_1^4 - 1}. \tag{3.2.16}$$

(3.2.15) and (3.2.16) show that both $\pi(S_{min})$ and $\pi(1)$ are functions of ε. For $\varepsilon = .2051$, we find numerically that $S_{min} = .5194$, $\pi(1) = .6796$ and $\pi(S_{min}) = .3604$. For this S_{min}, we can always find a corresponding distribution in $\mathcal{G}_\varepsilon(\Phi)$, say F_{min}, such that

$$\int \chi_0(\frac{x}{S_{min}}) dF_{min}(x) = 0$$

and F_{min} places all its contamination mass on $\{0\}$.

This shows that (3.2.1) fails when $\varepsilon = .2051$ in the ε-contamination normal neighbourhood since $\pi(S_{min}) < \pi(1)$ which is equivalent to $R(\chi_0, F_{min}) > R(\chi_0, F_0)$. Note that (3.2.1) holds if and only if $\pi(S) \geq \pi(1)$ for all possible S.

Moreover, since $\pi(S_{min}) - \pi(1)$ is a continuous function of ε, there must exist an ε_* such that (3.2.1) fails on the range of $\varepsilon_* \leq \varepsilon \leq .2051$.

Table V provides some numerical calculations for the quantities x_1, S_{min}, $\pi(1)$ and $\pi(S_{min})$ for some given ε. Note that it is much easier to determine $x_1(\geq 1.414)$ first , then ε and the rest of the other quantities. Numerical calculations show that for $.0997 \leq \varepsilon \leq .2051$, we can find an $F_{min} \in \mathcal{G}_\varepsilon(\Phi)$ such that $R(\chi_0, F_0) < R(\chi_0, F_{min})$ since $\pi(S_{min}) < \pi(1)$ in this range of ε.

3.3 Minimax asymptotic properties of M-estimators for scale over ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$ for large ε.

In this section, our goal is to show that in the ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G) = \{F|\ F = (1-\varepsilon)G + \varepsilon H\}$, large ε case, there exists an $F_* \in \mathcal{G}_\varepsilon(G)$ such that $R(\chi_0, F_0) < R(\chi_0, F_*)$. Here the distribution G is assumed to satisfy the property that the derivative of $G^*(x) = 2G(e^x) - 1$ is strongly unimodal. Note that $\mathcal{G}_\varepsilon(\Phi)$ is only a special case of $\mathcal{G}_\varepsilon(G)$.

The strategy of tackling the problem is as follow: Define

$$u(t) = \frac{1}{R(\chi_0, F_t)} = \frac{N_t^2}{D_t} \tag{3.3.1}$$

where

$F_t = (1-t)F_0 + tF_1$, $0 \leq t \leq 1$, $F_1 \in \mathcal{G}_\varepsilon(G)$ with $I(F_1) < \infty$,

F_0 minimizes Fisher information for scale over $\mathcal{G}_\varepsilon(G)$,

$N_t = \int \frac{x}{S(F_t)} \chi_0'(\frac{x}{S(F_t)}) dF_t(x)$,

and

$D_t = \int \chi_0^2 \frac{x}{S(F_t(x))} dF_t(x)$.

We then show that

(i) $u'(0) \geq 0$ for any $F_1 \in \mathcal{G}_\varepsilon(G)$ with $I(F_1) < \infty$

Table V To provide some numerical calculations for the quantities x_1, S_{min}, $\pi(1)$, $\pi(S_{min})$ for some given ε.

x_1	ε	S_{min}	$\pi(1)$	$\pi(S_{min})$
1.414	.2051	.5194	.6796	.3604
1.45	.1801	.5831	.7356	.4804
1.46	.1738	.5988	.7510	.5136
1.47	.1672	.6138	.7663	.5466
1.48	.1617	.6281	.7816	.5793
1.49	.1560	.6417	.7967	.6117
1.50	.1505	.6547	.8118	.6438
1.51	.1453	.6672	.8268	.6754
1.52	.1402	.6791	.8418	.7065
1.53	.1353	.6905	.8566	.7373
1.54	.1305	.7013	.8713	.7675
1.55	.1260	.7118	.8860	.7972
1.56	.1216	.7218	.9006	.8263
1.57	.1173	.7314	.9150	.8550
1.58	.1132	.7406	.9294	.8831
1.59	.1093	.7494	.9437	.9106
1.60	.1055	.7579	.9579	.9376
1.61	.1018	.7661	.9720	.9640
1.615	.1000	.7700	.9760	.9718
1.616	.0997	.7707	.9804	.9794
1.617	.0993	.7716	.98189	.98191
1.62	.0983	.7739	.9861	.9898

(ii) there exists a subset $\mathcal{G}_\varepsilon^*(G)$ of $\mathcal{G}_\varepsilon(G)$ such that

$u'(0) = 0$ and $u''(0) < 0$ for any $F_1 \in \mathcal{G}_\varepsilon^*(G)$.

By Taylor series expansion,

$$\begin{aligned} u(t) &= u(0) + tu'(0) + \frac{t^2}{2}u''(0) + o(t^2) \\ &= u(0) + \frac{t^2}{2}u''(0) + o(t^2) \text{ for } F_1 \in \mathcal{G}^*(G) \\ &< u(0) \text{ for } F_1 \in \mathcal{G}_\varepsilon^*(G) \text{ and sufficiently small } t \end{aligned}$$

Since $\mathcal{G}_\varepsilon(G)$ is convex, $F_t = (1-t)F_0 + tF_1 \in \mathcal{G}_\varepsilon(G)$, $0 \le t \le 1$. Put $F_* = F_t$ where t satisfies $u(t) < u(0)$. By definiton of $u(t)$, we conclude that

$$R(\chi_0, F_0) < R(\chi_0, F_*)$$

which is what we would like to prove.

To prove (i) and (ii), first note that

(a) $D_0 = \int \chi_0^2 dF_0 = \int x\chi_0' dF_0 = N_0 = I(F_0)$

(b) $S(F)$ is defined by $\int \chi_0(\frac{x}{S(F)})dF(x) = 0$

By (a), we have

$$u'(t)|_{t=0} = (2N_t' - D_t')|_{t=0} \tag{3.3.2}$$

and

$$u''(t) = (2N_t'' - D_t'')|_{t=0} + 2\frac{N_t'^2|_{t=0}}{I(F_0)} - u'(0)\frac{2D_t'|_{t=0}}{I(F_0)} \tag{3.3.3}$$

Put

$$\dot{S} = \frac{d}{dt}S(F_t)|_{t=0} \text{ and } \ddot{S} = \frac{d^2}{dt^2}S(F_t)|_{t=0}$$

From (b), we have

$$\dot{S} = \frac{\int \chi_0(x)d(F_1 - F_0)(x)}{\int x\chi_0'(x)dF_0(x)}$$

Using (a), it becomes

$$\dot{S} = \frac{\int \chi_0(x) dF_1(x)}{I(F_0)}. \tag{3.3.4}$$

Note that χ_0 is even, χ_0' has a finite number of jump points, say $-d_{m-1}, -d_{m-2}, \ldots,$ $-d_1, d_1, \ldots, d_{m-1}$ and χ_0' is continous and at least twice differentiable on the intervals $(-d_m, -d_{m-1}), \ldots, (-d_i, -d_{i+1}), \ldots, (-d_1, d_1), \ldots, (d_{m-1}, d_m)$ where $d_m = \infty$. Then $\chi_0'(\frac{x}{S(F_t)})$ is continuous and at least twice differentiable on the intervals $(-d_m S(F_t), -d_{m-1}S(F_t)), \ldots, (-d_i S(F_t), -d_{i+1}S(F_t)), \ldots,$ $(-d_1 S(F_t), d_1 S(F_t)), \ldots (d_{m_1} S(F_t), d_m S(F_t))$. Write

$$\begin{aligned} N_t = \sum_{k=2}^{m} [\int_{-d_k S(F_t)}^{-d_{k-1}S(F_t)} \frac{x}{S(F_t)} \chi_0'(\frac{x}{S(F_t)}) dF_t(x) + \int_{-d_{k-1}S(F_t)}^{d_k S(F_t)} \frac{x}{S(F_t)} \chi_0'(\frac{x}{S(F_t)}) dF_t(x)] \\ + \int_{-d_1 S(F_t)}^{d_1 S(F_t)} \frac{x}{S(F_t)} \chi_0'(\frac{x}{S(F_t)}) dF_t(x) \end{aligned} \tag{3.3.5}$$

$$\begin{aligned} D_t = \sum_{k=2}^{m} [\int_{-d_k S(F_t)}^{-d_{k-1}S(F_t)} \chi_0^2(\frac{x}{S(F_t)}) dF_t(x) + \int_{-d_{k-1}S(F_t)}^{d_k S(F_t)} \chi_0^2(\frac{x}{S(F_t)}) dF_t(x)] \\ + \int_{-d_1 S(F_t)}^{d_1 S(F_t)} \chi_0^2(\frac{x}{S(F_t)}) dF_t(x). \end{aligned} \tag{3.3.6}$$

Differentiate (3.3.5) and (3.3.6) with respect to t and set $t = 0$. We then have

$$\begin{aligned} N_t'|_{t=0} = & -\dot{S}[\int x\chi_0'(x) dF_0(x) + \int x^2 \chi_0''(x) dF_0(x)] \\ & + \int x\chi_0'(x) d(F_1 - F_0)(x) \\ & - 2\dot{S} \sum_{k=1}^{m-1} d_k^2 f_0(d_k) \delta(d_k) \end{aligned} \tag{3.3.7}$$

and

$$\begin{aligned} D_t'|_{t=0} = & -2\dot{S} \int x\chi_0(x)\chi_0'(x) dF_0(x)] \\ & + \int \chi_0^2(x) d(F_1 - F_0)(x) \end{aligned} \tag{3.3.8}$$

where $\delta(d_k) = \chi_0'(d_k^+) - \chi_0'(d_k^-)$ which is the jump point of χ_0' at d_k.

Note that

$$\begin{aligned}\int x\chi_0(x)\chi_0'(x)dF_0(x) - \int x\chi_0'(x)dF_0(x) - \int x^2\chi_0''(x)dF_0(x) \\ = 2\sum_{k=1}^{M-1} d_k^2 f_0(d_k)\delta(d_k). \qquad (3.3.9)\end{aligned}$$

Using the relationship (3.3.9) and putting (3.3.7), (3.3.8) into (3.3.2), we then have

$$\begin{aligned}u'(0) &= \int [2x\chi_0'(x) - \chi_0^2(x)]d(F_1 - F_0)(x) \\ &= \int J(\chi_0)(x)d(F_1 - F_0)(x). \qquad (3.3.10)\end{aligned}$$

As we know that F_0 minimizes Fisher information for scale over $\mathcal{G}_\varepsilon(G)$, we conclude $u'(0) = \int J(\chi_0)(x)d(F_1 - F_0)(x) \geq 0$ which is (i).

Now assume (a, b) to be one of the intervals $(-d_m, -d_{m-1})$, $\ldots, (-d_1, d_1)$, $\ldots, (d_{m-1}, d_m)$. Then

$$\frac{d^2}{dt^2}\int_{aS(F_t)}^{bS(F_t)}\frac{x}{S(F_t)}\chi_0'(\frac{x}{S(F_t)})dF_t(x)$$

$$
\begin{aligned}
&=\int_{aS(F_t)}^{bS(F_t)}\{\frac{2x}{S^3(F_t)}(\frac{\partial S(F_t)}{\partial t})^2\chi_0'(\frac{x}{S(F_t)})f_t(x)-\frac{x}{S^2(F_t)}\frac{\partial^2 S(F_t)}{\partial t^2}\chi_0'(\frac{x}{S(F_t)})f_t(x)\\
&+\frac{x^2}{S^4(F_t)}(\frac{\partial S(F_t)}{\partial t})^2\chi_0''(\frac{x}{S(F_t)})f_t(x)-\frac{x}{S^2(F_t)}\frac{\partial S(F_t)}{\partial t}\chi_0'(\frac{x}{S(F_t)})(f_1-f_0)(x)\\
&+\frac{3x^2}{S^4(F_t)}(\frac{\partial S(F_t)}{\partial t})^2\chi_0''(\frac{x}{S(F_t)})f_t(x)+\frac{x^3}{S^5(F_t)}\chi_0'''(\frac{x}{S(F_t)})(\frac{\partial S(F_t)}{\partial t})^2 f_t(x)\\
&-\frac{x^2}{S^3(F_t)}\chi_0''(\frac{x}{S(F_t)})\frac{\partial^2 S(F_t)}{\partial t^2}f_t(x)-\frac{x^2}{S^2(F_t)}\chi_0''(\frac{x}{S(F_t)})\frac{\partial S(F_t)}{\partial t}(f_1-f_0)(x)\\
&-\frac{x}{S^2(F_t)}\frac{\partial S(F_t)}{\partial t}\chi_0'(\frac{x}{S(F_t)})(f_1-f_0)(x)\\
&-\frac{x^2}{S^2(F_t)}\chi_0''(\frac{x}{S(F_t)})\frac{\partial S(F_t)}{\partial t}(f_1-f_0)(x)\}dx\\
&+b\frac{\partial S(F_t)}{\partial t}\{-\frac{b}{S(F_t)}\frac{\partial S(F_t)}{\partial t}\chi_0'(b)f_t(bS(F_t))-\frac{b^2}{S(F_t)}\chi_0''(b)\frac{\partial S(F_t)}{\partial t}f_t(bS(F_t))\\
&+b\chi_0'(b)(f_1-f_0)(bS(F_t))\}\\
&-a\frac{\partial S(F_t)}{\partial t}\{-\frac{a}{S(F_t)}\frac{\partial S(F_t)}{\partial t}\chi_0'(a)f_t(aS(F_t))-\frac{a^2}{S(F_t)}\chi_0''(a)\frac{\partial S(F_t)}{\partial t}f_t(aS(F_t))\\
&+a\chi_0'(a)(f_1-f_0)(aS(F_t))\}\\
&+b^2\frac{\partial^2 S(F_t)}{\partial t^2}\chi_0'(b)f_t(bS(F_t))\\
&+b^2\frac{\partial S(F_t)}{\partial t}\chi_0'(b)[(f_1-f_0)(bS(F_t))+b\frac{\partial S(F_t)}{\partial t}f_t'(bS(F_t))]\\
&-a^2\frac{\partial^2 S(F_t)}{\partial t^2}\chi_0'(a)f_t(aS(F_t))\\
&-a^2\frac{\partial S(F_t)}{\partial t}\chi_0'(a)[(f_1-f_0)(aS(F_t))+a\frac{\partial S(F_t)}{\partial t}f_t'(aS(F_t))]
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\frac{d^2}{dt^2}\int_{aS(F_t)}^{bS(F_t)}\frac{x}{S(F_t)}\chi_0'(\frac{x}{S(F_t)})dF_t(x)|_{t=0}\\
&=A_1+A_2+A_3 \qquad (3.3.11)
\end{aligned}
$$

where

$$A_1 = \ddot{S} \int_a^b [x\chi_0'(x)f_0(x) + x^2\chi_0'(x)f_0'(x)]dx = -\ddot{S} \int_a^b x\chi_0(x)\chi_0'(x)f_0(x)dx$$

$$A_2 = -(\dot{S})^2 \int_a^b \{2x\chi_0'(x)[\chi_0(x)+1]f_0(x) + x^2(\chi_0'(x))^2 f_0(x) + x^2(\chi_0(x)+2)f_0'(x)\chi_0'(x)\}dx$$

$$A_3 = 2\dot{S}\{-\int_a^b [x\chi_0'(x) + x^2\chi_0''(x)(f_1 - f_0)(x)]dx + x^2\chi_0'(x)(f_1 - f_0)(x)|_a^b\}$$

Also

$$\begin{aligned}
&\frac{d^2}{dt^2}\int_{aS(F_t)}^{bS(F_t)} \chi_0^2(\frac{x}{S(F_t)})dF_t(x)\\
&= \int_{aS(F_t)}^{bS(F_t)} \{\frac{4x}{S^3(F_t)}\chi_0(\frac{x}{S(F_t)})\chi_0'(\frac{x}{S(F_t)})(\frac{\partial S(F_t)}{\partial t})^2 f_t(x)\\
&+ \frac{2x^2}{S^4(F_t)}(\chi_0'(\frac{x}{S(F_t)}))^2(\frac{\partial S(F_t)}{\partial t})^2 f_t(x)\\
&+ \frac{4x^2}{S^4(F_t)}\chi_0(\frac{x}{S(F_t)})\chi_0''(\frac{x}{S(F_t)})(\frac{\partial S(F_t)}{\partial t})^2 f_t(x)\\
&- \frac{2x}{S^2(F_t)}\chi_0(\frac{x}{S(F_t)})\chi_0'(\frac{x}{S(F_t)})\frac{\partial^2 S(F_t)}{\partial t^2} f_t(x)\\
&- \frac{4x}{S^2(F_t)}\chi_0(\frac{x}{S(F_t)})\chi_0'(\frac{x}{S(F_t)})\frac{\partial S(F_t)}{\partial t}(f_1 - f_0)(x)\}dx\\
&+ b\frac{\partial S(F_t)}{\partial t}[-2b\chi_0(b)\chi_0'(b)\frac{\partial S(F_t)}{\partial t}f_t(bS(F_t)) + \chi_0^2(b)(f_1 - f_0)(bS(F_t))]\\
&- a\frac{\partial S(F_t)}{\partial t}[-2a\chi_0(a)\chi_0'(a)\frac{\partial S(F_t)}{\partial t}f_t(aS(F_t)) + \chi_a^2(a)(f_1 - f_0)(aS(F_t))]\\
&+ b\chi_0^2(b)[\frac{\partial^2 S(F_t)}{\partial t^2}f_t(bS(F_t)) + \frac{\partial S(F_t)}{\partial t}(f_t'(bS(F_t))b\frac{\partial S(F_t)}{\partial t} + (f_1 - f_0)(aS(F_t))]\\
&- a\chi_0^2(a)[\frac{\partial^2 S(F_t)}{\partial t^2}f_t(aS(F_t)) + \frac{\partial S(F_t)}{\partial t}(f_t'(aS(F_t))a\frac{\partial S(F_t)}{\partial t} + (f_1 - f_0)(aS(F_t))]
\end{aligned}$$

Therefore

$$\begin{aligned}
&\frac{d^2}{dt^2}\int_{aS(F_t)}^{bS(F_t)} \chi_0^2(\frac{x}{S(F_t)})f_t(x)|_{t=0}\\
&= B_1 + B_2 + B_3 \qquad (3.3.12)
\end{aligned}$$

where

$$B_1 = \ddot{S}[\int_a^b -2x\chi_0(x)\chi_0'(x)f_0(x)dx + x\chi_0^2(x)f_0(x)|_a^b]$$

$$B_2 = (\dot{S})^2\{\int_a^b [4x\chi_0(x)\chi_0'(x)f_0(x) + 2x^2(\chi_0'(x))^2 f_0'(x) + 2x^2\chi_0(x)\chi_0''(x)f_0(x)]dx$$
$$- 2x^2\chi_0(x)\chi'(x)f_0(x)|_a^b + x^2\chi_0'(x)f_0'(x)|_a^b\}$$

$$B_3 = \dot{S}\{-\int_a^b 4x\chi_0(x)(f_1 - f_0)(x)dx + 2x^2\chi_0^2(x)(f_1 - f_0)(x)|_a^b\}.$$

After some calculations,

$$\begin{aligned}
&2(A_1 + A_2 + A_3) - (B_1 + B_2 + B_3) \\
&\qquad = -\ddot{S}x\chi_0^2(x)f_0(x)|_a^b \\
&\qquad - \frac{(\dot{S})^2}{2}\{\int_a^b [J^2(\chi_0)(x) + \chi_0^4(x)]f_0(x)dx + x\chi_0^2(x)f_0(x)|_a^b\} \\
&\qquad - \dot{S}\{\int_a^b 2xJ'(\chi_0)(x)(f_1 - f_0)(x)dx \\
&\qquad - 2xJ(\chi_0)(x)(f_1 - f_0)(x)|_a^b\}
\end{aligned} \tag{3.3.13}$$

Recall from the result of Theorem 1.2 on page 14 that for large ε, χ_0 is of the form

$$\chi_0(x) = \begin{cases} -k & |x| \le x_0 \\ \xi(x) & x_0 \le |x| \le x_1 \\ k & |x| > x_1 \end{cases}$$

for $0 < x_0 < x_1 < \infty$ where $\xi(x) = -x\frac{g'}{g}(x) - 1$. Thus

$$J(\chi_0)(x) = 2x\chi_0'(x) - \chi_0^2(x) = \begin{cases} -k^2 & |x| < x_0 \\ 2x\xi'(x) - \xi^2(x) & x_0 \le |x| \le x_1 \\ k^2 & |x| > x_1. \end{cases}$$

Define

$$\mathcal{G}_\varepsilon^1(G) = \{F_1 \in \mathcal{G}_\varepsilon(G) | F_1(x_0) = (1-\varepsilon)G(x_0) + \varepsilon,\ F_1 \text{ symmetric },I(F_1) < \infty,$$
$$f_1 = F_1' \text{ exists and } \int \chi_0(x)dF_1(x) \neq 0\}.$$

Note that each element F_1 in $\mathcal{G}_\varepsilon^1(G)$ satisfies

(i) $f_1(x) = (1-\varepsilon)g(x) = f_0(x)$ on (x_0, x_1)

(ii) $F_1(x) = (1-\varepsilon)G(x) + \varepsilon$ on (x_1, ∞)

(iii) $F_1(x_0) - F_0(x_0) - F_1(x_1) + F_0(x_1) = 0$

(iv) $\int_{x_0}^{x_1} J(\xi)(x) d(F_1 - F_0)(x) = 0$

and $\mathcal{G}_\varepsilon^1(G)$ is nonempty.

Thus for each $F_1 \in \mathcal{G}_\varepsilon^1(G)$,

$$\begin{aligned} u'(0) &= 2\int_0^\infty J(\chi_0)(x) d(F_1 - F_0)(x) \\ &= 2\int_{x_0}^{x_1} J(\chi_0)(x) d(F_1 - F_0)(x) \\ &= 2k^2[F_1(x_1) - F_0(x_1) - F_1(x_0) + F_0(x_0)] = 0. \end{aligned}$$

Furthermore if $F_1 \in \mathcal{G}_\varepsilon^1(G)$, from (3.3.13), (3.3.7) and (3.3.8), we have

$$(2N_t'' - D_t'')|_{t=0} = -\frac{(\dot{S})^2}{2}\int_a^b [J^2(\chi_0)(x) + \chi_0^4(x)] f_0(x) dx \qquad (3.3.14)$$

and

$$\begin{aligned} N_t'|_{t=0} &= \frac{D_t'}{2}|_{t=0} \\ &= \dot{S}\int x\chi_0(x)\chi_0'(x) f_0(x) dx \\ &= \frac{\dot{S}}{2}\int \chi_0^3(x) f_0(x) dx \end{aligned} \qquad (3.3.15)$$

By substituting (3.3.14) and (3.3.15) into (3.3.3), if $F_1 \in \mathcal{G}_\varepsilon^1(G)$, we have

$$\begin{aligned} u''(0) &= -\frac{(\dot{S})^2}{2}\{\int [J^2(\chi_0)(x) + \chi_0^4(x)] f_0(x) dx - \frac{(\int \chi_0^3(x) f_0(x) dx)^2}{I(F_0)}\} \\ &= -\frac{(\dot{S})^2}{2}\{\int J^2(\chi_0)(x) f_0(x) dx + \frac{I(F_0)(\int \chi_0^4(x) f_0(x) dx - (\int \chi_0^3(x) f_0(x) dx)^2}{I(F_0)}\} \end{aligned}$$

By Cauchy Schwarz inequality,

$$\begin{aligned} I(F_0)\int \chi_0^4(x) f_0(x) dx &= \int \chi_0^2(x) f_0(x) dx \int \chi_0^4(x) f_0(x) dx \\ &\geq (\int \chi_0^3(x) f_0(x) dx)^2 \end{aligned}$$

Hence

$$u''(0) \le -\frac{(\dot{S})^2}{2}\int J^2(\chi_0)(x)f_0(x)dx$$
$$< 0$$

since $F_1 \in \mathcal{G}_\varepsilon^1(G)$ This completes the proof of (ii) by setting $\mathcal{G}_\varepsilon^*(G) = \mathcal{G}_\varepsilon^1(G)$. Note that if we define

$$\mathcal{G}_\varepsilon^2(G) = \{F_1 \in \mathcal{G}_\varepsilon(G) |\, F_1(-x_1) \doteq (1-\varepsilon)G(-x_1) + \frac{\varepsilon}{2}, F_1 \text{ symmetric },$$
$$I(F_1) < \infty, f_1 = F_1' \text{ exists and } \int \chi_0(x)dF_1(x) \neq 0\}$$

we can prove, in a similar way, that for each $F_1 \in \mathcal{G}_\varepsilon^2(G)$, $u'(0) = 0$ and $u''(0) < 0$. Thus in general we can let $\mathcal{G}_\varepsilon^*(G) = \mathcal{G}_\varepsilon^1(G) \cup \mathcal{G}_\varepsilon^2(G)$. Roughly speaking, $\mathcal{G}_\varepsilon^1(G)$ contains all the distributions in $\mathcal{G}_\varepsilon(G)$ which put all their contaminating mass in $[-x_0, x_0]$ and $\mathcal{G}_\varepsilon^2(G)$ contains all the distributions $\mathcal{G}_\varepsilon(G)$ which put all their contamination mass in $[-x_1, x_1]^c$.

3.4 Minimax asymptotic properties of M-estimators for scale over Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$

The problem we try to solve in this section is to show that there exists an $F_* \in \mathcal{K}_\varepsilon(G)$ such that $R(\chi_0, F_0) < R(\chi_0, F_*)$ where $\mathcal{K}_\varepsilon(G) = \{F : \sup_{x\in\mathbb{R}} |F(x) - G(x)| \le \varepsilon\}$, the Kolmogorov neighbourhood model and G satisfies the conditions (A1)-(A5) and (J1)-(J5) as in Theorem 2.3 Chapter II.

The way to prove the existence of such an F_* is quite similar to that described as in the last section. Define

$$w(t) = \frac{1}{R(F_t)} = \frac{N_t^2}{D_t} \tag{3.4.1}$$

where

$$F_t = (1-t)F_0 + tF_1, 0 \le t \le 1, F_1 \in \mathcal{K}_\varepsilon(G) \text{ with } I(F_1) < \infty,$$

F_0 minimizes Fisher information for scale over $\mathcal{K}_\varepsilon(G)$

$$N_t = \int \frac{x}{S(F_t)} \chi_0'(\frac{x}{S(F_t)}) dF_t(x)$$

$$D_t = \int \chi_0^2(\frac{x}{S(F_t)}) dF_t(x).$$

With similar arguments as in section 3.3, we have

$$w'(0) = \int J(\chi_0)(x) d(F_1 - F_0)(x) \qquad (3.4.2)$$

(a) For the Kolmogorov neighbourhood model (large ε case):

Put

$$\mathcal{K}_{\varepsilon,1}(G) = \{F_1 \in \mathcal{K}_\varepsilon(G) : (F_1 - F_0)(a) = 0, \ (F_1 - F_0)(d) = 0,$$
$$F_1 \text{ symmetric }, F_1' = f_1 \text{ exists }, I(F_1) < \infty$$
$$\int \chi_0(x) dF_1(x) \ne 0\}.$$

Note that

(i) $\mathcal{K}_{\varepsilon,1}(G) \ne \emptyset$

(ii) $w'(0) = 0$ if $F_1 \in \mathcal{K}_{\varepsilon,1}(G)$

(iii) $w''(0) < 0$ if $F_1 \in \mathcal{K}_{\varepsilon,1}(G)$.

To prove (i), take a symmetric F_1 with density f_1 such that $I(F_1) < \infty$, $F_1(a) = F_0(a)$, $F_1(d) = F_0(d)$ and $F_1(x) < F_0(x)$ on (a, d). Then $F_1 \in \mathcal{K}_{\varepsilon,1}(G)$ which implies $\mathcal{K}_{\varepsilon,1}(G) \ne \emptyset$. For $F_1 \in \mathcal{K}_{\varepsilon,1}(G)$,

$$w'(0) = 2\int_0^\infty J(\chi_0)(x) d(F_1 - F_0)(x)$$
$$= 2[\int_0^a (-\lambda_2^2) d(F_1 - F_0)(x) + \int_a^d \lambda_1^2 d(F_1 - F_0)(x) + \int_d^\infty (-\lambda^2) d(F_1 - F_0)(x)$$
$$= 0.$$

which is (ii). (iii) follows exactly from the same argument in proving $u''(t) < 0$ as in section 3.3. Now using Taylor series expansion,

$$\begin{aligned} w(t) &= w(0) + tw'(0) + \frac{t^2}{2}w''(0) + o(t^2) \\ &= w(0) + \frac{t^2}{2}w''(0) + o(t^2) \text{ for } F_1 \in \mathcal{K}_{\varepsilon,1}(G). \end{aligned}$$

Thus we can conclude that $w(t) < w(0)$ for all sufficient small t if $F_1 \in \mathcal{K}_{\varepsilon,1}(G)$ and $F_t = (1-t)F_0 + tF_1$. Because $w(t) = \frac{1}{R(F_t)}$, we have $R(F_t) > R(F_0)$ which is what we desire. This is because if we put $F_* = (1-t)F_0 + tF_1$ where $F_1 \in \mathcal{K}_{\varepsilon,1}(G)$ for sufficient small t, we have

$$R(\chi_0, F_*) = R(F_*) > R(F_0) = R(\chi_0, F_0).$$

Note that the condition

$$\int \chi_0(x)dF_1(x) \neq 0$$

implies

$$\int_a^d \chi_0(x)dF_1(x) \neq \int_a^d \chi_0(x)dF_0(x).$$

Since χ_0 is nonconstant only in the region $(-d,-a) \cup (a,d)$, we require $F_1 \neq F_0$ for at least a subinterval of (a,d).

(b) For the Kolmogorov neighourhood model (small ε case):

Put

$$\begin{aligned} \mathcal{K}_{\varepsilon,2}(G) = \{F_1 \in \mathcal{K}_\varepsilon(G) :\ & F_1(x) = F_0(x) \text{ on } [a,b] \cup [c,d], \\ & F_1 \text{ symmetric }, F_1' = f_1 \text{ exists }, I(F_1) < \infty \\ & \text{and } \int \chi_0(x)dF_1(x) \neq 0\}. \end{aligned}$$

Here $a = b$ or $c = d$. By the same argument as in case (a), since

(i) $\mathcal{K}_{\varepsilon,2}(G) \neq \emptyset$

(ii) $w'(0) = 0$ if $F_1 \in \mathcal{K}_{\varepsilon,2}(G)$

(iii) $w''(0) < 0$ if $F_1 \in \mathcal{K}_{\varepsilon,2}(G)$

there exists a $F_* \in \mathcal{K}_\varepsilon(G)$ such that $R(\chi_0, F_0) < R(\chi_0, F_*)$.

(c) For the Kolmogorov neighourhood model (medium ε case):

Put

$$\begin{aligned}\mathcal{K}_{\varepsilon,3}(G) = \{F_1 \in \mathcal{K}_\varepsilon(G) :\ & F(x) = F_0(x) \text{ on } x \in [a,b] \cup [c,d], \\ & F_1 \text{ symmetric }, F_1' = f \text{ exists }, I(F_1) < \infty \\ & \text{and } \int \chi_0(x) dF_1(x) \neq 0\}\end{aligned}$$

Here $a = b$ or $c = d$. By the same argument as in case (a) or case (b) there exists an $F_* \in \mathcal{K}_\varepsilon(G)$ such that $R(\chi_0, F_0) < R(\chi_0, F_*)$.

CHAPTER IV

CONCLUSION

4.1 Summary

In this thesis, we have considered the problem of minimax, robust M-estimation of scale, when the distribution generating the observations is assumed to be approximately known. It is assumed that the underlying distribution lies within a certain convex class $\mathcal{P}$ of distributions. Two common classes of distributions which we have studied are ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$ and Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$.

In chapter I, an M-estimate S of scale is defined. It corresponds to the choice of an estimating function χ. We denote $I(F)$ to be the Fisher information for scale at F with scale equal to 1 and F_0 to be the distribution which minimizes $I(F)$ over $\mathcal{P}$. We put $\chi_0(x) = -x\frac{f_0'}{f_0}(x) - 1$ where $f_0 = F_0'$. Then we note that χ_0 is minimax in the sense that it minimizes the maximum asymptotic variance over the distributions $F \in \mathcal{P}$ satisfying $S_0(F) = 1$ where S_0 is the M-estimators corresponding to the estimating function χ_0.

We first obtained the least informative distribution F_0 and the corresponding χ_0 in the ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$ where G is symmetric with finite Fisher information for scale and that the density $g^*(x) = 2e^x g(e^x)$ is strongly unimodal, $g = G'$. We derived an easily checked necessary and sufficient condition for g^* to be strongly unimodal. We found that typical distributions G satisfying these conditions are the normal, logistic, Student's t and Laplace. We also provided a result for the case when g^* is not strongly unimodal. All of the obtained results for the least informative distribution in the ε-contamination neighbourhood model rely on a log transformation of the absolute value of the data, changing the scale problem entirely to a location one.

In Chapter II, we first pointed out that although the Kolmogorov neighbourhood structure is maintained under the log transformation of the absolute value of the data, there is no corresponding location theory to obtain the least informative distribution in $\mathcal{K}_\varepsilon(G)$. This is because although we assume the distribution function G to be symmetric, the transformed distribution function G^* defined by $G^*(x) = G(e^x) - G(-e^x)$ may not be and the existing location theory for Kolmogorov case can only deal with symmetric G^*. A direct approach to the problem was then introduced.

Under certain mild assumptions on the distribution function G, we obtained some conditions which are necessary and sufficient in order that F_0 have minimum Fisher information for scale in $\mathcal{K}_\varepsilon(G)$. We note that the above mentioned results are not sufficient to characterize the forms of F_0 and the corresponding χ_0 since the form of (χ_0, F_0) depends on the shape of the graph $J(\xi)(x) = 2x\xi'(x) - \xi^2(x)$ versus x where $\xi(x) = -x\frac{g'}{g}(x) - 1$ and $g = G'$. By imposing some further restrictions on the graph of $J(\xi)(x)$, the explicit forms of (χ_0, F_0) were obtained. These results are general enough to include the cases where G are normal, logistic, Student's t and Laplace. Besides, our results are not restricted to random variables on the whole real line - the real line may be replaced by its non-negative half throughout, with no essential changes in the theory.

In the last section of Chapter II, we made some corrections on Thall's (1979) paper and provided a correct solution to his particular problem which is essentially to find (χ_0, F_0) when the Kolmogorov exponential neighbourhood model (i.e. $\mathcal{K}_\varepsilon(G), G(x) = 1 - e^{-x}, x \geq 0$) is concerned.

In Chapter III, we define the asymptotic loss $R(\chi, F)$ and use it as a quantity in comparing the performance of the scale estimators. We ask whether or not the saddlepoint property

$$R(\chi_0, F) \le R(\chi_0, F_0) = \frac{1}{I(F_0)} \le R(\chi, F_0)$$

holds in ε-contamination neighbourhood model $\mathcal{G}_\varepsilon(G)$ and Kolmogorov neighbourhood model $\mathcal{K}_\varepsilon(G)$ with G satisfying certain conditions. If it does, the minimax properties holds. We showed that the above saddlepoint property fails in

(i) $\mathcal{G}_\varepsilon(G)$, large ε case where G is assumed to be symmetric and satisfy the property that the derivative of $G^*(x) = 2G(e^x) - 1$ is strongly unimodal.

(ii) $\mathcal{G}_\varepsilon(\Phi), .0997 \le \varepsilon \le .2051$ (= the boundary between the "small ε" and "large ε" form, if $G = \Phi$).

(iii) $\mathcal{K}_\varepsilon(G)$, all ε where G satisfies certain mild conditions.

4.2 Future Research

One question that arises in this thesis is that if the saddlepoint property fails, does the minimax property still hold? Is it possible to find an M-estimator of scale which is minimax within a given convex class of distributions? This same qucstion may also be asked in location case.

Throughout this thesis, we use the asymptotic variance or asymptotic loss as the criterion in comparing the performance of M-estimators for scale. But if our interest is not merely to find a 'best' estimator among all the M-estimator of scale but to estimate a scale parameter for a random variable X, other possible criterion that serve may be the asymptotic bias or mean square error. Martin and Zamar (1989) discussed the asymtotically min-max bias robust M-estimates of scale for positive random variables. Another interesting problem may be to find the M-estimator of scale that minimizes the maximum mean square error in a given convex class of distributions.

Moreover, it might be possible to extend the research to investigating whether or not the efficient L-estimators for scale corresponding to the least informative distributions satisfy the saddlepoint property. Indeed, we conjecture that the answer is negative.

BIBLIOGRAPHY

Bickel, P.J. and Lehamnn, E.L. (1975). Descriptive statistics for nonparametric models I. Introduction. Ann Statist.3, 1038-1044.

Bickel, P.J. and Lehmann, E.L. (1976). Descriptive Statistics for nonparametric models. III. Dispersion, Ann. Statist. 4, 1139-1158.

Boos, D.D. and Serfling, R.J. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates. Ann. Statist. 8, 618-624.

Collins, J.R. (1983). On the minimax property for R-estimators of location. Ann. Statist. 11, 1129-1195.

Collins, J.R. and Wiens, D.P. (1985). Minimax variance M-estimators in ε-contamination models. Ann. Statist. 13, 1078-1096.

Collins, J. and Wiens, D. (1989). Minimax properties of M-, R- and L-estimators of location on Lévy neighbourhoods. Ann. Statist. 17, 327-336.

Daniell, P.J. (1920). Observations weighted according to order, Amer. J. Math. 42, 222-236.

Hampel, F.R. (1974). The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 62, 1179-1186.

Huber, P.J. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35, 73-101.

Huber, P.J. (1981). Robust statistics, Wiley, New York.

Ibragimov, I.A. (1956). On the composition of unimodal distribution. Theor. Probab. Applic. 1, 255-260.

Jaeckel, L.A. (1971). Robust estimates of Location: symmetry and asymmetric contamination, Ann. Math. Statist. 42, 1020-1034.

Lehmann, E.L. (1983). Theory of Point Estimation, Wiley, New York.

Martin, R.D. an Zamar, R.H. (1989). Asymptotically min-max bias robust M-estimates of scale for positive random variables. J. Amer. Statist. Assoc. 84, 494-501.

O'Neil, P.V. (1975). Advanced Calculus, pure and applied, Macmillan, New York.

Sacks, J. and Ylvisaker, D. (1972). A note on Huber's robust estimation of a location parameter, Ann. Math. Statist. 43, 1068-1075.

Sacks, J. and Ylvisaker, D. (1982). L- and R-estimation and the minimax property. Ann. Statist. 10, 643-645.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York.

Thall, P.F. (1979). Huber-Sense Robust M-estimation of a scale parameter, with Application to the Exponential Distribution, J. Amer. Statist. Assoc. 74, 147-152.

Wiens, D.P. (1985). Minimizing Fisher information for location over Kolmogorov neighbourhoods, unpublished manuscript.

Wiens, D. (1986). Minimax variance M-estimators of location in Kolomogorov neighbourhoods. Ann. Statist. 13, 724-732.

Wiens, D. (1987). Robust weighted Cramér-von Mises estimators of location, with minimax variance in ε-contamination neighbourhoods. Canad. J. Statist. 15, 269-278.

Wiens, D. (1990). Minimax variance L- and R-estimators of location, Canad. J. Statist. 18, No. 1 or 2.