THE MIDRANGE ESTIMATOR IN SYMMETRIC DISTRIBUTIONS

by

RICHARD ALLEN SUNDHEIM

B. S., Kansas State University, 1971

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1974

Approved by:

Ray O. Waller

Major Professor

TABLE OF CONTENTS

1. INTRODUCTION

When the center, θ, of a symmetric population is unknown, we have several alternatives for an estimator of θ. The almost habitual assumption that the population is normally distributed or approximately so has led to the wide use of the sample mean. If the population is in fact normally distributed, the sample mean is considered the optimal estimator since it is consistent, unbiased, sufficient, and efficient. However, the sample mean being very sensitive to extreme outliers is not the most desirable estimator when the parent population has "heavy-tails".

The alternative estimator we consider in this paper is the midrange. After defining and characterizing heavy-tailed distributions, we investigate the behavior of the midrange in a variety of heavy-tailed and light-tailed distributions. The classical properties of consistency, unbiasedness, sufficiency, and efficiency are then used to compare the relative merits of the midrange estimator with that of the sample mean and sample median.

The hypothesis put forth is that the shape of the parent distribution's tail is the determining factor in the choice of the estimator. Only when the tails are light is the sample mean preferred. When the tails are "short and heavy" the midrange is found to be a very desirable estimator while the median is preferred when sampling from populations with "long and heavy tails".

Bryson [2] contends there is evidence for placing more emphasis on the use of heavy-tailed distributions when modeling data. Therefore when the form of the distribution is unknown, rather than assuming normality, the sample informtion should be used to determine the tail-shape of the distribution before choosing the estimator. This is a concept of robust estimation and is briefly discussed in Chapter 5.

2. HEAVY-TAILED DISTRIBUTIONS

To define and characterize heavy-tailed distributions we consider a symmetric distribution with a density function $f(\cdot)$, distribution function $F(\cdot)$, and tail function $G(\cdot) = 1 - F(\cdot)$. The notion of heavy tails is a relative one which is used to compare the tail weights of different distributions. Intuitively, a heavy-tailed distribution assigns greater likelihood to extreme (tail) values in the range of positive density than does a light-tailed distribution. If the density function is positive over a finite range (i.e. $F(x) = 1$ for some finite x) the distribution is said to be bounded and if $F(x) < 1$ for all finite x, then the distribution is unbounded. This distinction leads us to adopt a separate characterization of heavy tails for the two cases.

2.1 Unbounded Distributions.

The amount of weight in the tails of the distribution is described by the rate at which $F(x)$ converges to one as x goes to infinity. This is equivalent to the rate at which the tail function $G(x)$ and the density function $f(x)$ converge to zero as x goes to infinity. So if $F(x)$ converges "slowly" to one, the distribution is considered to have heavy tails. For example, let

$$f_1(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \quad , \quad -\infty < x < \infty \tag{2.1}$$

$$f_2(x) = \lambda e^{-\lambda x} \quad , \quad x > 0 \ , \ \lambda > 0 \tag{2.2}$$

$$f_3(x) = \frac{1}{\pi\,(1 + x^2)} \quad , \qquad -\infty < x < \infty \tag{2.3}$$

represent the density functions for the standard normal, exponential, and Cauchy distributions, respectively. Since as x goes to infinity, e^x approaches infinity faster than any power of x, it follows that $f_1(x)$ approaches zero faster than $f_2(x)$ which in turn approaches zero faster than $f_3(x)$. Therefore we can conclude the Cauchy distribution has heavier tails than the other two distributions, while the normal distribution has light tails when compared to the exponential and Cauchy distributions.

To characterize heavy-tailed distributions we need some borderline distribution ϕ, such that every distribution with tails heavier than ϕ is considered to be a member of the class of heavy-tailed distributions. Bryson [2] argues that ϕ should be the exponential distribution. He does this by considering the conditional mean exceedence, CME, which is defined as

$$CME_c = E\,(\,X - c \mid X \geq c\,). \tag{2.4}$$

If c is sufficiently large, we obtain the following characterization of tail weights:

(i) A distribution is said to be heavy-tailed if the CME_c increases as c increases.

(ii) A distribution is said to have light tails if the CME_c decreases as c increases.

(iii) A distribution is a borderline case if the CME_c is constant as c increases.

To apply this characterization, we put (2.4) in a more tractable form:

$$CME_c = \frac{1}{1 - F(c)} \int_c^{\infty} (x - c)\, dF(x) \tag{2.5}$$

Letting $y = x - c$, (2.5) becomes

$$CME_c = \frac{1}{1 - F(c)} \int_0^{\infty} y\, dF(y)$$

$$= \frac{-1}{1 - F(c)} \int_0^{\infty} y \frac{d}{dy}[1 - F(y)]\, dy \quad .$$

Integrating by parts yields

$$CME_c = \frac{1}{1 - F(c)} \int_c^{\infty} [1 - F(x)]\, dx$$

or

$$CME_c = \frac{1}{G(c)} \int_c^{\infty} G(x)\, dx \quad . \tag{2.6}$$

To illustrate (2.6) consider the exponential distribution in (2.2).

Then

$$F(x) = 1 - e^{-\lambda x} \quad ,$$

$$G(x) = e^{-\lambda x} \quad ,$$

and

$$CME_c = \frac{1}{e^{-\lambda c}} \int_c^{\infty} e^{-\lambda x}\, dx = \frac{1}{\lambda} \quad .$$