THE INSTITUTE
OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

What is Kurtosis?

An Influence Function Approach

David Ruppert
University of North Carolina
Chapel Hill, N.C. 27514

Mimeo Series #1579

June 1985

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

What is Kurtosis?

An Influence Function Approach

David Ruppert*

Abstract: Many textbooks give a short and inadequate discussion of kurtosis, and the literature contains much controversy on what kurtosis measures. In this article Hampel's influence function is used to gain some understanding of the usual kurtosis coefficient, i.e., the standarized fourth moment, as well as some alternative measures of kurtosis.

KEY WORDS: Kurtosis, influence function, peakedness, tail-behavior, bimodality, statistical functional.

*Associate Professor of Statistics, University of North Carolina, Chapel Hill, N.C. 27514. Work supported by NSF Grant DMS-8400602.

1. INTRODUCTION

The concept of kurtosis can be found in statistics books at all levels. For example, Levin's (1984) precalculus text teaches the distinction between mesokurtic, leptokurtic, and platykurtic distributions. The treatment of kurtosis is often quite short and usually fails to give much understanding.

The fundamental problem, as Bickel and Lehmann (1975) have noticed, is that there is no agreement on what kurtosis measures. Some authors, e.g. Levin (1984) and Kendall and Buckland (1971) claim that kurtosis measures "peakedness" versus "flatness". Darlington (1970) emphasizes that the opposite of "peaked" is bimodal. A flat distribution is intermediate between these extremes. The terms "peaked", "flat", and "bimodal" unfortunately under-emphasize the dependence of kurtosis on tail behavior. Chissom (1970) cautions that "a major emphasis must be placed on the tails of the distribution in the determination of the fourth moment." Johnson and Kotz (1985) say that kurtosis is a measure of deviation from normality depending on the relative frequency of values either near the mean or far from it to values an intermediate distance from the mean.

Even an accurate definition such as Johnson and Kotz's (1985) only provides a rough qualitative understanding of kurtosis. What are missing are the specification of the ranges defining the peak, flanks, and tails and an indication of the relative importance of these ranges in determining kurtosis.

In this article, Hampel's (1968, 1974) influence function provides a quantitative understanding of kurtosis. Kurtosis is often thought to measure nonnormality, and the influence function shows precisely how kurtosis changes with slight deviations from the Gaussian distribution. Interestingly, Darlington (1970) independently discovered the influence function of the standard kurtosis coefficient after the influence function was available in an unpublished thesis (Hampel, 1968) but before its publication (Hampel, 1974). In this article, we elaborate on Darlington's (1970) discussion. Moreover, here the influence function is found for some alternative measures of kurtosis and is used to compare them.

Besides quantitatively analyzing the meaning of kurtosis, this article can serve as an introduction to the influence function, an extremely important tool in the theory of robustness (Huber 1981) and diagnostic statistics (Cook and Weisberg 1982 and Belsley, Kuh, and Welsch 1980).

2. THE INFLUENCE FUNCTION

The influence function of a statistical functional was developed by Hampel (1968, 1974) to study the behavior of estimators when the data deviate slightly from an assumed statistical model. A statistical functional is defined as a map from a set of probability distributions to the real numbers. The mean, for example, is the functional $\mu(.)$ on the set of distributions with finite first moment defined by

$$\mu(F) = \int_{-\infty}^{\infty} u \, dF(u) \ .$$

The variance and fourth central moment,

$$\sigma^2(F) = \int_{-\infty}^{\infty} (u - \mu(F))^2 \, dF(u)$$

and

$$\mu_4(F) = \int_{-\infty}^{\infty} (u - \mu(F))^4 \, dF(u) \ ,$$

are also statistical functionals, as is the usual kurtosis coefficient

$$\kappa(F) = \mu_4(F)/\sigma^4(F) \ .$$

Many estimators are obtained by applying a statistical functional to the empirical distribution function. To see how a functional T(.) is changed by minor deviations from F, we compute the "directional derivatives" of T at F. Let G be another distribution and define $F_\epsilon = (1-\epsilon)F + \epsilon G$ for $0 \le \epsilon \le 1$. Then

$$\dot{T}(F;G) = \lim_{\epsilon \downarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}$$

is the derivative of T at F in the direction towards G.

If $G = \delta_x$ where δ_x is point mass at the value x, F_ϵ is F with contaminants at x and ϵ is the proportion of contamination. Then $\dot{T}(F;\delta_x)$ is called the influence function of T at F and x;

$$IF(x;F,T) = \dot{T}(F;\delta_x) \ .$$

Typically $\dot{T}(F;G)$ is linear in G and in particular if G is the discete distribution with probability mass $p_1,\ldots,p_M$ at values $x_1,\ldots,x_M$ then

$$\dot{T}(F;G) = \sum_{m=1}^{M} p_m \, IF(x_m;T,F) \ .$$

Therefore, the influence function also shows the effect of contaminants at several values $x_1,\ldots,x_M$.

In some cases it is easier or more desirable to calculate the influence function of log T rather than T itself. Since the influence function is an ordinary one-sided derivative of a function of ϵ, elementary calculus implies that

$$IF(x;F,\log T) = IF(x;F,T)/T(F) ,$$

so that $IF(x;F,\log T)$ measures the relative change in T at F.

3. THE INFLUENCE FUNCTION OF THE KURTOSIS COEFFICIENT

We will confine our discussion to symmetric distributions since asymmetric deviations from normality are better studied by statistics other than kurtosis. Let F and G be symmetric distributions with finite fourth moments and define $F_\epsilon = (1-\epsilon)F + \epsilon G$. Then $\mu(F_\epsilon) = 0$ for all ϵ, so that

$$\log[\kappa(F_\epsilon)] = \log \mu_4(F_\epsilon) - 2 \log \sigma^2(F_\epsilon)$$

$$= \log[\mu_4(F) + \epsilon\{\mu_4(G) - \mu_4(F)\}]$$

$$-2 \log[\sigma^2(F) + \epsilon\{\sigma^2(G) - \sigma^2(F)\}]$$

and

$$\text{(1)} \qquad \frac{d}{d\epsilon_+} \log \kappa(F_\epsilon)\Big|_{\epsilon=0} = \frac{\mu_4(G)-\mu_4(F)}{\mu_4(F)} - 2\,\frac{\sigma^2(G)-\sigma^2(F)}{\sigma^2(F)} .$$

Now define $\eta_x = .5(\delta_x + \delta_{-x})$. η_x represents a contaminant at $\pm x$ with equal probability. Then letting $G = \eta_x$ in (1) we define the symmetric influence function of $\log \kappa$ as

$$\text{(2)} \qquad SIF(x;F,\log \kappa) = \frac{x^4-\mu_4(F)}{\mu_4(F)} - \frac{2(x^2-\sigma^2(F))}{\sigma^2(F)} .$$

Multiplying (2) by $\kappa = \mu_4(F)/\sigma^4(F)$ gives SIF(x;F,κ), which was already obtained by Darlington (1970). Now let F be the standard normal distribution $\underline{\Phi}$. Since $\mu_4(\underline{\Phi}) = 3$ and $\sigma^2(F) = 1$,

$$\text{SIF}(x;\underline{\Phi},\log \kappa) = \frac{x^4-3}{3} - \frac{2(x^2-1)}{1} = \frac{(x^2-3)^2-6}{3},$$

and SIF(x;$\underline{\Phi}$,κ) = $(x^2-3)^2-6$. As noted by Darlington (1970), SIF(x;$\underline{\Phi}$,κ) is negative for $|x|$ in (.7420,2.334) and positive elsewhere. We can identify the peak as the range $|x| < .7420$, the flanks as $.7420 < |x| < 2.334$, and the tails as $|x| > 2.334$. Outliers in both the tails and at the peak increase the kurtosis and outliers in the flanks decrease it.

SIF(x;$\underline{\Phi}$,log κ) is graphed in Figure 1 over the range $0 \leq x \leq 3$. SIF(x;$\underline{\Phi}$,log κ) grows quarticly in x and has values 54.33, 159.33, and 298.00 at x = 4,5, and 6, respectively. It is clear that the location of tail outliers is much more important than their frequency. For example, one percent contamination at $|x| = 4$ will increase kurtosis by roughly the same amount as five percent contamination at $|x| = 3$. Also, in the peak range, $|x| < .7420$, SIF(x;$\underline{\Phi}$,log κ) has a maximum of 1 at x = 0. Outliers at the peak have far less influence than those in the extreme tails, and κ is primarily a measure of tail behavior, and only to a lesser extent of peakedness or its opposite bimodality as Darlington (1970) suggests. Darlington did not comment on the tail behavior of SIF(x;$\underline{\Phi}$,κ), but as we can see from (2) tail behavior is essential for an understanding of κ.

4. OTHER MEASURES OF KURTOSIS

The usual definition (see Johnson and Kotz 1985) that kurtosis is whatever κ measures is clearly inadequate if one seeks alternative measures of kurtosis. Bickel and Lehmann (1975) define a measure of kurtosis as any "suitable" ratio of scale functionals. A nonnegative functional T is a scale functional if $T(G) = |\theta|T(F)$ when $G(x) = F((x-\mu)/\theta)$, $\theta \neq 0$, i.e., if T is location invariant and scale equivariant. Any ratio of scale functionals is location and scale invariant. The question of what is "suitable" is not addressed by them, though we will propose an answer later.

Hogg (1974) has suggested a measure of tail weight based on "outer means". Let $q_p(F) = F^{-1}(p)$ be the p-th quantile of F, and for $0 < p \leq .5$ define

$$\bar{U}_p(F) = \int_{q_{1-p}(F)}^{\infty} u \, dF(u)$$

and

$$\bar{L}_p(F) = \int_{-\infty}^{q_p(F)} u \, dF(u) .$$

$\bar{U}_p$ and $\bar{L}_p$ are the tail means and measure the weight of the upper and lower tails. $\bar{U}_p(F) - \bar{L}_p(F)$ is a scale functional that is particularly sensitive to tail weight when p is small, and

$$Q_p(F) = \{\bar{U}_p(F) - \bar{L}_p(F)\}/\{\bar{U}_{.5}(F) - \bar{L}_{.5}(F)\}$$

is a location and scale invariant measure of tail weight. Hogg (1974) uses $Q_{.2}(F)$, which he calls Q_1, to chose the trimming fraction of an adaptively trimmed mean.

Since F is assumed to be symmetric about 0, $Q_p(F) = \bar{U}_p(F)/\bar{U}_{.5}(F)$. Huber (1981) shows that the influence function of q_p is

$$IF(x;F,q_p) = \frac{p - I(x < q_p(F))}{f(q_p(F))}$$

where $f(x) = (d/dx)F(x)$, whence the symmetric influence function for $x \geq 0$ is

$$SIF(x;F,q_p) = .5\{IF(x;F,q_p)) + IF(-x;F,q_p)\}$$

$$= \{(p-.5) - .5\ I(x < q_p(F))\}/f(q_p(F))$$

for $p \geq .5$. Then by a standard calculus argument

$$SIF(x;F,\bar{U}_p) = -SIF(x;F,q_{1-p})q_{1-p}(F)f(q_{1-p}(F)) + \int_{q_{1-p}(F)}^{\infty} u\ d(\eta_x-F)(u)$$

$$= .5\max\{x,q_{1-p}(F)\} - \{(.5-p)q_{1-p}(F) + \bar{U}_p(F)\} .$$

Therefore for $x > 0$ the symmetric influence function of $\log Q_p$ is

(3) $$SIF(x;F,\log Q_p) = [.5/\bar{U}_p(F)]\max\{x,q_{1-p}(F)\} -[.5/\bar{U}_{.5}(F)]x - (.5-p)q_{1-p}(F)/\bar{U}_p(F).$$

This function is piecewise linear with a corner at $q_{1-p}(F)$ where it reaches its minimum. For $p = .2$ and $F = \Phi$ we have

$$SIF(x;\Phi,Q_{.2}) = 1.7865\max(x,.842) - 1.2539x - .90256,$$

which is negative on $.480 < |x| < 1.701$ and positive elsewhere and has a minimum at $x = .842 = \Phi^{-1}(.8)$; see figure 2.

Both κ and $Q_{.2}$ are increased by outliers in the tails and the peak and are decreased by outliers in the flanks. However, as we have seen the flanks of Φ are $.7420 < |x| < 2.334$ for κ and $.480 < |x| < 1.701$ for $Q_{.2}$.

The most important difference between κ and $Q_{.2}$ is that as $x \to \infty$ $SIF(x;\Phi,\kappa)$ is of order x^4 but $SIF(x;\Phi,Q_{.2})$ is only linear in x. Both κ and $Q_{.2}$ are sensitive to the location of tail outliers as well as their frequency, but κ is much more sensitive.

A kurtosis measure with a bounded SIF is desirable from a robustness viewpoint, especially if one is primarily interested in measuring peakedness or its opposite, bimodality. As Hildebrand (1971) points out, κ does measure bimodality for distributions whose tail behavior cannot dominate κ, such as the symmetric beta distributions. However, among the symmetric gamma distributions Hildebrand found a bimodal distribution with kurtosis equal to 3, the same as Φ. A little reflection shows that for a kurtosis measure with unbounded SIF, one can construct a bimodal distribution with an arbitrarily large value of that measure.

To construct a bounded-influence kurtosis measure, one can take the ratio of two robust scale functionals, for example two interfractile ranges. For $0 < p < \Delta < .5$ we define

$$R_{\Delta,p}(F) = \{q_{1-p}(F) - q_p(F)\}/\{q_{1-\Delta}(F) - q_{\Delta}(F)\} .$$

If F is symmetric about 0, then

$$R_{\Delta,p}(F) = q_{1-p}(F)/q_{1-\Delta}(F) .$$

Horn (1983) proposed

$$mt_p = 1 - p/(q_{p+.5}(F)f(0)) ,$$

$0 < p < .5$, as a measure of "peakedness". Since $f(0) \cong \epsilon/F^{-1}(\epsilon + .5) = \epsilon/q_{\epsilon+.5}(F)$ if ϵ is small,

$$mt_p \cong 1 - p/[\Delta\, R_{.5-\Delta,.5-p}]$$

so that Horn's measure is almost a monotonic function of $R_{.5-\Delta,.5-p}$ if Δ

is close to .5 . The symmetric influence function of $\log R_{\Delta,p}$ is

$$\mathrm{SIF}(x;F,\log R_{\Delta,p}) =$$

(4) $$\mathrm{SIF}(x;F,q_{1-p})/q_{1-p}(F) - \mathrm{SIF}(x;F,q_{1-\Delta})/q_{1-\Delta}(F) =$$

$$\frac{.5-p-.5\ I(x<p_{1-p}(F))}{f(q_{1-p}(F))q_{1-p}(F)} - \frac{.5-\Delta-.5\ I(x<q_{1-\Delta}(F))}{f(q_{1-\Delta}(F))q_{1-\Delta}(F)} .$$

At $F = \Phi$, $\Delta = .3$, and $p = .1$ this becomes

$$\mathrm{SIF}(x;\Phi,\log R_{.3,.1}) = .6814 + 2.744\ I(x \leq .524) - 2.2236\ I(x \leq 1.282) ,$$

which is graphed in figure 3. $R_{\Delta,p}$ and therefore Horn's mt_p are measures of both peakedness and tail weight, but the influence of tail outliers is bounded, so as intended they are much better measures of peakedness/bimodality than κ or Q_p.

5. DISCUSSION

It should now be clear why there has been little agreement on what kurtosis represents. Many writers have sought a simple explanation of kurtosis, but kurtosis is complex. κ and its alternatives Q_p and $R_{\Delta,p}$ are large for peaked and/or heavy-tailed distributions but are decreased by excessive flanks, with flatness and bimodality corresponding to moderate and extremely heavy flanks, respectively. Among the functionals we have studied, there appear to be no pure measures of either peakedness or tail-behavior, only measure combining both.

We are now in a position to suggest a general definition of measures of kurtosis: a kurtosis measure is any location and scale invariant measure T such that for any symmetric F at which T(F) and SIF(x;F,T) are defined, there exist positive numbers $a < b$ depending on F and T with the property that SIF(x;F,T) is negative if $a < |x| < b$ and positive if $|x| < a$ or $|x| > b$. By equations (2) and (3), K and Q_p are kurtosis measures according to this definition. By equation (4), SIF($x;F,R_{\triangle,p}$) jumps twice, first downward and then upward, as x goes from 0 to ∞. However, whether the jumps are from positive to negative and then back to positive depends on the value of f at $q_{1-p}(F)$ and $q_{1-\triangle}(F)$. Therefore, our definition of kurtosis does not include $R_{\triangle,p}$.

It is important that T be scale invariant or else one ends up measuring dispersion as well as peakedness and tail behavior. Birnbaum (1943) made such an error in defining the notion that one random variable is "more peaked" than another.

Ratios of scale functionals are location and scale invariant. If "suitable" means having a SIF with the above property, then our general definition includes that of Bickel and Lehmann (1975). Ours but not theirs also includes monotonic increasing transformations of such ratios such as κ and Horn's measure of peakedness.

Acknowledgment. I thank Steve Marron and Ed Carlstein for their comments on an earlier draft.

REFERENCES

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), Regression Diagnostics, New York: John Wiley.

Bickel, P.J. and Lehmann, E.L. (1975), Descriptive Statistics For Nonparametric Models. I. Introduction, The Annals of Statistics, 3, 1038-1044.

Birnbaum, Z.W. (1948), On Random Variables With Comparable Peakedness, The Annals of Mathematical Statistics, 19, 76-81.

Chisson, Brad S. (1970), Interpretation of the Kurtosis Statistic, The American Statistician, 24 (4), 19-23.

Cook, R.D. and Weisberg, S. (1982), Residuals and Influence in Regression. New York and London: Chapman and Hall.

Darlington, R.B. (1970), Is Kurtosis Really Peakedness?, The American Statistician, 24 (2), 19-20.

Hampel, F.R. (1968). Contributions to the Theory of Robust Estimation. Unpublished Ph.D. dissertation. Department of Statistics, University of California, Berkeley.

Hampel, F.R. (1974), The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association, 62, 1179-1186.

Hildebrand, D.K. (1971), Kurtosis Measures Bimodality?, The American Statistician, 25 (1), 42-43.

Hogg, R.V. (1974), Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory. Journal of the American Statistical Association, 69, 909-927.

Horn, Paul S. (1983), A Measure of Peakedness. The American Statistician, 37, 55-56.

Huber, P.J. (1981), Robust Statistics, New York: John Wiley.

Johnson, N.L. and Kotz, S. (1985), Encyclopedia of Statistical Sciences, Volume 5, New York: John Wiley.

Kendall, M.G. and Buckland, W.R. (1971), A Dictionary of Statistical Terms, Third Edition, New York: Hafner.

Levin, R.I. (1984). Statistics for Management, Third Edition, Englewood Cliffs: Prentice Hall.

List of figures

Figure 1

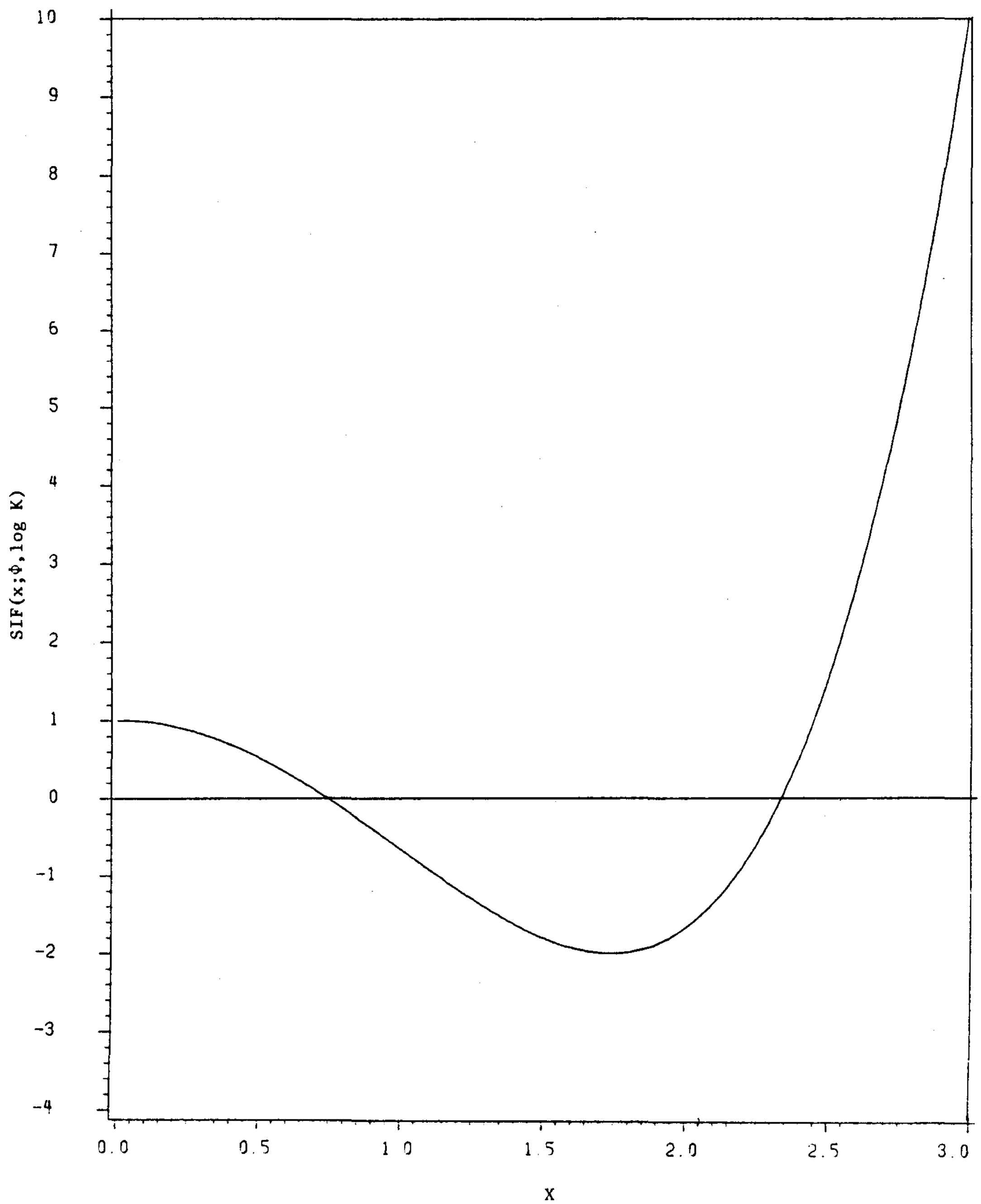

Figure 2

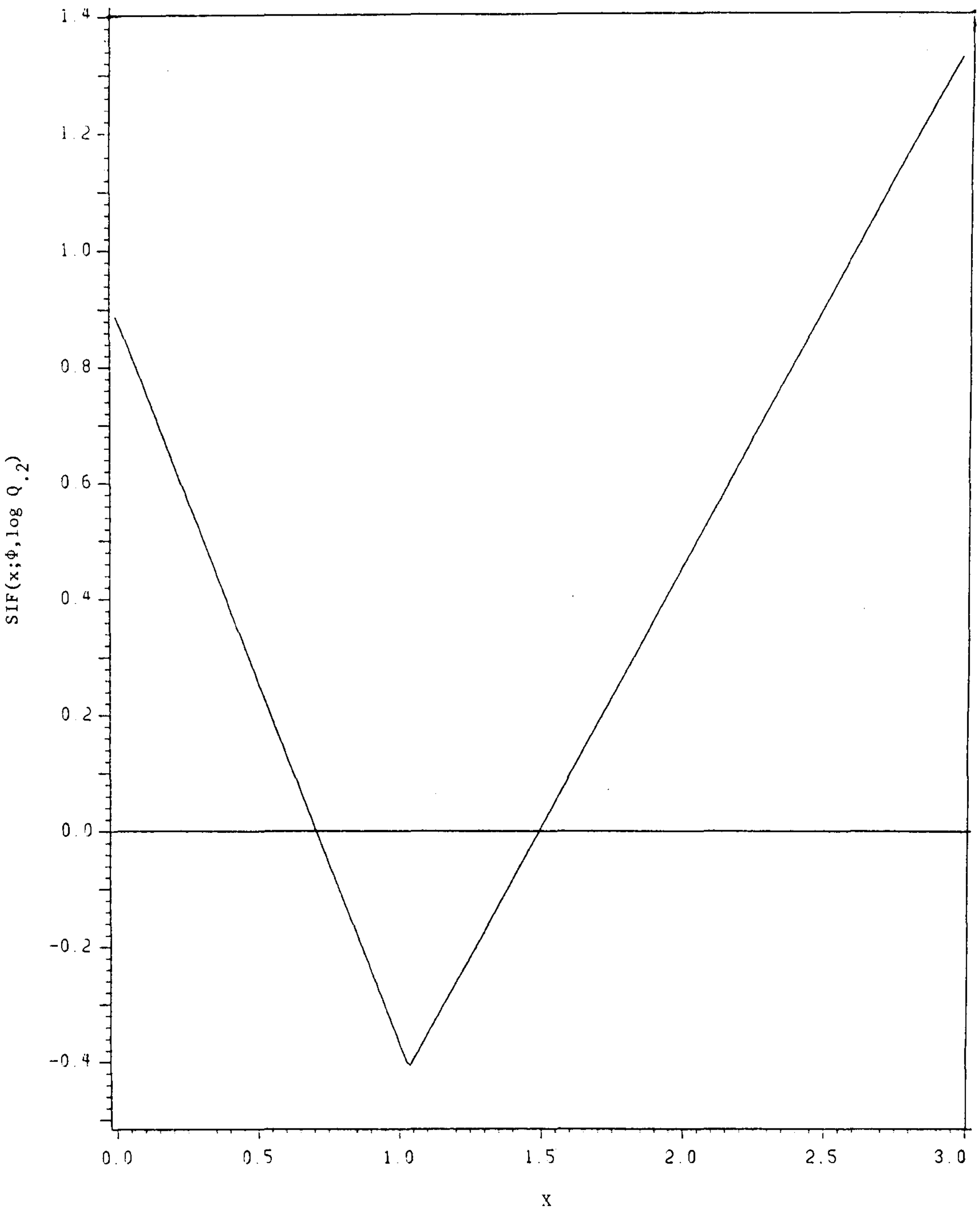

Figure 3